DE-43 (DC & SE)

**NUST COLLEGE OF**
**ELECTRICAL AND MECHANICAL ENGINEERING**

# MIND EASE -VIRTUAL COUNSELOR

A PROJECT REPORT

DE-43 (DC & SE)

***Submitted by***


NS MUHAMMAD IMMAD UD DIN

ASC OZAIR MANSOOR AHMED

ASC SAAD HASAN


**BACHELORS**

**IN**

**COMPUTER ENGINEERING**

**YEAR**

**2025**

**PROJECT SUPERVISOR**

DR. WASI HAIDER

DR. ASSAD MANSOOR KHAN


DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING
COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
ISLAMABAD, PAKISTAN

# Certification

This is to certify that Saad Hasan [398416], Ozair Mansoor [398423], and Muhammad Immad Ud Din [366871] have successfully completed their final year project Mind Ease – Virtual Counselor, at the NUST College of Electrical and Mechanical Engineering, to fulfill the partial requirement of the degree Bachelors in Computer Engineering.

**Dr Wasi Haider Butt**
**PhD (Software Engineering)**
Tenured Professor

wasi@ceme.nust.edu.pk

# Complex Engineering Problem

**Range of Complex Problem Solving**

| | Attribute | Complex Problem | |
|---|---|---|---|
| 1 | Range of conflicting requirements | Involve wide-ranging or conflicting technical, engineering and other issues. | ✓ |
| 2 | Depth of analysis required | Have no obvious solution and require abstract thinking, originality in analysis to formulate suitable models. | ✓ |
| 3 | Depth of knowledge required | Requires research-based knowledge much of which is at, or informed by, the forefront of the professional discipline and which allows a fundamentals-based, first principles analytical approach. | ✓ |
| 4 | Familiarity of issues | Involve infrequently encountered issues | ✓ |
| 5 | Extent of applicable codes | Are outside problems encompassed by standards and codes of practice for professional engineering. | ✓ |
| 6 | Extent of stakeholder involvement and level of conflicting requirements | Involve diverse groups of stakeholders with widely varying needs. | ✓ |
| 7 | Consequences | Have significant consequences in a range of contexts. | |
| 8 | Interdependence | Are high level problems including many component parts or sub-problems | ✓ |

**Range of Complex Problem Activities**

| | Attribute | Complex Activities | |
|---|---|---|---|
| 1 | Range of resources | Involve the use of diverse resources (and for this purpose, resources include people, money, equipment, materials, information and technologies). | ✓ |
| 2 | Level of interaction | Require resolution of significant problems arising from interactions between wide ranging and conflicting technical, engineering or other issues. | ✓ |
| 3 | Innovation | Involve creative use of engineering principles and research-based knowledge in novel ways. | ✓ |
| 4 | Consequences to society and the environment | Have significant consequences in a range of contexts, characterized by difficulty of prediction and mitigation. | ✓ |
| 5 | Familiarity | Can extend beyond previous experiences by applying principles-based approaches. | ✓ |

# Sustainable Development Goals (SDGs)

| SDG No | Description of SDG | SDG No | Description of SDG |
|---|---|---|---|
| SDG 1 | No Poverty | SDG 9 | Industry, Innovation, and Infrastructure |
| SDG 2 | Zero Hunger | SDG 10 | Reduced Inequalities |
| SDG 3 | Good Health and Well Being | SDG 11 | Sustainable Cities and Communities |
| SDG 4 | Quality Education | SDG 12 | Responsible Consumption and Production |
| SDG 5 | Gender Equality | SDG 13 | Climate Change |
| SDG 6 | Clean Water and Sanitation | SDG 14 | Life Below Water |
| SDG 7 | Affordable and Clean Energy | SDG 15 | Life on Land |
| SDG 8 | Decent Work and Economic Growth | SDG 16 | Peace, Justice and Strong Institutions |
| | | SDG 17 | Partnerships for the Goals |


Sustainable Development Goals

# Acknowledgment

First of all, we would like to express our humblest gratitude to the divine presence of Allah Almighty for giving us the strength and fortitude to take on this endeavour. All our efforts were guided towards a positive outcome, and help prevailed over us in times of uncertainty. We want to thank our project supervisor. Wa, Dr Wasi Haider, for his unwavering help and support throughout the project. His mentorship and guidance motivated us to give our best and strive for excellence. We would also like to credit our co-supervisor, Dr. Assad Khan, at RISETech, whose exceptional talent and knowledge helped us overcome hurdles. Finally, we want to thank our family members and friends for their encouragement and support throughout this effort.

Ozair, Saad, Immad

# Abstract

The project aims to design a cost-effective and intelligent mental health support system using a large language model specifically tailored to a Pakistani context. The system provides compassionate, psychological, and wellness coaching for the anxious or depressed, available on an easy-to-use web app. The architecture is created based on modern transformer architectures and fine-tuning approaches, using carefully picked dialogues between psychologists and patients with protected sensitive data, and securing the data privacy.

To promote openness and trust, the system integrates an Explainable AI functionality to enable users to see how the chatbot makes decisions. The system is evaluated through user testing with real users and psychologists. Key deliverables include the fine-tuned LLM, user-friendly cross-platform interfaces, and usage analysis to support scalable, accessible, and responsible AI-driven mental health support.

# Contents

# List of Figures

# Chapter 1

# Introduction

The high rate of mental health disorders has a tremendous effect on the total global disease burden. Depression accounts for almost 4.4 percent of the total global burden expressed in disability adjusted life years [1]. Yet, vast disparities exist: 90 percent or more of mentally diseased persons in low- and middle-income countries (LMICs) are not treated in comparison to about 50 percent in high-income countries [1]. This difference is partly attributable to consistent overspending on mental health services. Low- and middle-income countries reserve only a minuscule part, less than 1–2 percent, of their health funds for mental health services [1],[2].

Pakistan exemplifies these challenges. At a psychiatrist-to-population ratio of 0.19 for every 100,000 residents and due to constrained mental health expenditures, levelling at only 0.4 percent of the health budget [1], the country needs to serve the population of about 24million individuals who require care [1]. Among the significant health problems in Pakistan have been mental disorders such as depression, anxiety, and schizophrenia, with stigma that accompanies mental disease still common [1]. The lack of formal mental health services means that individuals will be faced with travel and financial barriers if they want to acquire that care [1]. Consequently, Pakistan's health system fails to meet the demands of those seeking to receive treatment. Many of the citizens go untreated or use informal therapies, such as those used by religious or traditional practitioners, because of stigma and the cost of formal medical treatments [3],[4].

The findings of a survey on the mental health of Rawalpindi's adolescents in Pakistan show that depression and anxiety are mainstream issues in this group of people. Since 17.2 percent of participants demonstrated symptoms of depression while 21.4 percent presented with anxiety, this study pleads for an increased focus on addressing both forms of manifestations of the disorder. Moreover, the study recognises that poor economic state, low educational accomplishment, and unpleasant life events cause more concern about the state of mental health. Given that it is the first of its kind, conducted in a school context, this research emphatically suggests the need to prioritize adolescent mental health as a public health concern [5].

Mental health stigma and cultural barriers are especially pronounced in Pakistan. Studies have shown that both laypeople and even some healthcare professionals in Pakistan hold stigmatizing attitudes toward mental illness. Based on Karachi-based studies conducted, the public's stigma levels towards mental disorders went beyond the public-level stigma for physical illness [4]. They were also greater than levels found among student populations and trained clinicians. Many Pakistani communities believe profoundly in the belief that mental health concerns are attributed to supernatural reasons, such as black magic, demonic possession, or moral defect [3], [4]. Such a mindset tends to stop the victims from seeking treatment from professionals. Even those in the medical field deny that the fear of being stigmatized and embarrassed keeps many from getting care. Mental health understanding is also limited, with no related educational programs in schools. Such cultural effects enhance a culture in which Pakistanis are used to avoiding or delaying therapy, and so this compounds the burden of unhindered illnesses.

At the same time, Pakistan's infrastructure for innovation and internet access has grown. Approximately 75 percent of Pakistan's population now has a mobile phone connection [6], most of which are broadband-capable (3G/4G) [6]. This widespread mobile penetration presents an opportunity: digital interventions can reach users who lack access to traditional services. Given the shortage of clinicians and reluctance to seek face-to-face help, a culturally tailored digital solution can lower barriers. These conditions motivated the Mind Ease project: an AI-driven, web-based mental health chatbot designed for Pakistani users. Mind Ease aims to provide an accessible, anonymous platform where users can discuss their concerns and receive supportive guidance informed by evidence-based therapies. The chatbot can be used on any internet-connected device without requiring specialized apps by operating through a web frontend. The motivation for Mind Ease is threefold: (1) to **bridge the treatment gap** by offering a form of support where human therapists are unavailable or stigmatized, (2) **to deliver culturally sensitive care** by incorporating local language, idioms, and social context, and (3) **to leverage modern AI** (a fine-tuned large language model) to create a safe, empathetic interaction that aligns with professional guidance.

## 1.1 Motivation

The Mind Ease chatbot is motivated by the need to reduce Pakistan's mental health treatment gap and improve accessibility in a culturally sensitive manner. The lack of clinicians in areas outside major cities [1],[3] makes AI chatbots a resource that can certainly supplement what is available. The chatbot primarily provides early support with pragmatic coping skills, such as cognitive-behavioral therapy techniques, mindfulness prompting, and promoting active listening skills. Although the chatbot should not replace the psychologist's role, it recommends identifying stressors, using practical self-help methods, and helps users become involved in work with professionals if necessary.

Key motivators include: Accessibility: The chatbot removes challenges associated with place and time limitations by providing 24-hour access to places through the Internet. Because numerous Pakistanis can access smartphones and the internet [6], a web-based service could offer significant support in geographically isolated or underserved areas.

Anonymity and Convenience: Interaction is text-based and confidential, which may reduce the fear of stigma. Users can discuss sensitive issues (depression, anxiety, family problems) in privacy from their homes.

Cultural Relevance: By grounding the AI in local language and context, Mind Ease addresses users in a familiar voice. This is critical because standard Western-tailored chatbots may inadvertently use idioms or examples that Pakistani users find alien.

Empirical Support: International trials of mental health chatbots (e.g., Woebot, Wysa) have shown promising results in reducing anxiety and depression symptoms in users [7],[8]. These successes suggest an AI approach is viable; Mind Ease aims to adapt that approach to Pakistan's context.

In summary, Mind Ease is motivated by the dire need for mental health support in Pakistan and the opportunity offered by AI/chatbot technology. It strives to provide a culturally nuanced, user-friendly tool to augment the limited mental health infrastructure and help bridge the gap between need and access.

## 1.2 Scope

The primary objectives of the Mind Ease project are to design, implement, and evaluate a web-based AI chatbot that assists Pakistani users with mental health concerns. Its scope and features include:

Culturally Sensitive Support: Provide conversation and coping strategies framed in Pakistan's cultural context (language, examples, religious and social norms). The bot should appropriately use Urdu/English phrases and reference culturally relevant situations.

AI-Driven Dialogue: Leverage a large language model (Meta's LLaMA 3.2) fine-tuned on authentic doctor–patient dialogue to generate empathetic, informative responses [8]. The chatbot will use natural conversational flow (free-text input) to the greatest extent while guiding users toward helpful reflection and coping techniques.

Evidence-Based Techniques: Incorporate cognitive-behavioral therapy (CBT) principles, mindfulness prompts, and psychoeducation. The system will be trained or guided by real counseling scripts so that recommendations (e.g., stress-management tips) align with best practices.

Accessibility and Device Compatibility: Deploy as a responsive web app (using Next.js and Tailwind CSS) that runs in any modern browser on mobile or desktop. No installation of native apps is required; only an internet connection and a browser are needed. The interface will be simple and clear to accommodate low literacy or first-time app users.

User Authentication and Data Security: Use secure authentication (NextAuth with JSON Web Tokens) to manage user sessions. While users can engage anonymously, any account data will be stored encrypted in MongoDB. No personal health information is requested or logged unless the user consents. The system will ensure data privacy and comply with relevant guidelines.

Iterative Local Testing: Conduct formative testing with target users in Pakistan. Feedback from focus groups and pilot studies will inform refinements. The scope includes adjusting language, conversation flow, and interface design based on local user experience.

Evaluation and Feedback: Implement user satisfaction, engagement, and preliminary mental health outcomes metrics. The chatbot's effectiveness will be assessed through user surveys and usage analytics.

The project does not attempt to replace therapists or provide medical diagnoses. Its scope is limited to offering initial counseling dialogue and coping support. More serious conditions or crises will trigger referrals to hotlines or professionals.

## 1.3 Problem Statement

WHO and recent research document the global imperative for expanded mental health access. In recent years, since the COVID-19 pandemic alone, the global prevalence of anxiety and depression increased by a massive 25% [9]. According to estimates, one billion people around the world suffer from mental disorders [2], but low and middle-income countries spend little on the treatment [2]. Based on WHO's 2017 Mental Health Atlas, many countries spend less than 2% of their health budgets on mental health [2], and roughly 75% of countries have no mental health care in their insurance coverage [2]. These resource gaps create a massive treatment shortfall: in Pakistan, only about one psychiatrist exists for every 500,000 people [3], far below international norms. The sheer scale of need compounds this structural scarcity: some epidemiological estimates suggest that up to 10–16% of Pakistani adults suffer from depression or anxiety (on the order of 25–30 million people) [2]. Pakistan's mental health system is concentrated in urban tertiary hospitals, leaving rural and low-income populations underserved [1],[3]. In practice, this means that most Pakistanis with mental health problems never see a clinician, reinforcing the "treatment gap" that can exceed 90% in LMICs [1].

Amid these challenges, digital interventions offer hope. Pakistan has seen the rise of telemedicine startups and NGO-led initiatives (e.g., PILL, Sehat Kahani) to extend mental health outreach [1]. The Ministry of Health and international agencies now recommend strategies like training lay counselors, integrating mental health into primary care, and using mobile platforms to disseminate information [1]. The World Health Organization explicitly lists "leveraging digital technology" as a strategy to expand access to care [1]. In this context, Mind Ease capitalizes on Pakistan's growing internet access to deliver real-time supportive conversations. The project harnesses AI to provide the initial tier of support – not a full clinical service, but a user-friendly step toward care, which could be particularly valuable when face-to-face therapy is unavailable or culturally unacceptable.

### 1.3.1 Cultural Barriers and Stigma

In Pakistan, cultural and social factors heavily influence attitudes toward mental health. Prevailing cultural norms often associate mental illness with weakness or supernatural forces [3]. Many Pakistanis prefer to confide in family, clergy, or traditional healers rather than psychiatrists; for example, consulting a religious scholar for "black magic" is a common practice in some communities [3]. This stigma is not merely public: even healthcare workers acknowledge negative perceptions among patients. As one Karachi study found, the general public in Pakistan showed significantly higher stigma scores for mental disorders than did health professionals [4]. Older males tended to have the highest stigma, while women generally showed slightly less. In short, negative attitudes towards mental illness are pervasive: research shows they correlate with lower self-esteem, increased hopelessness, and reluctance to seek help [4].

Such barriers justify a novel approach. By offering an anonymous, AI-driven chat interface, Mind Ease aims to circumvent stigma. Users interact with a machine interface, which may feel more private and less judgmental than talking to someone. This design choice aligns with research showing that chatbots can engage users who would otherwise avoid help due to embarrassment [7]. Importantly, Mind Ease will be culturally tailored: the conversation style, examples, and language (Urdu and English code-switching, for instance) will be adapted so that the chatbot "speaks" in a way familiar and respectful to Pakistani users. By embedding local idioms, family values, and religious sensitivities into the bot's responses, the system seeks to build trust and relevance. Prior Pakistani initiatives (e.g., the "Aiza" chatbot by DilKibaat) have similarly emphasized cultural adaptation. Mind Ease also plans to leverage real-world data and feedback from Pakistani users to refine its conversational style.

## 1.4 Aims and Objectives

Mind Ease aims to significantly improve mental health access in Pakistan by closing a critical gap. With only a few hundred mental health professionals nationwide [4],[1], Pakistan faces an enormous treatment shortage. By providing a 24/7 chat-based counselor, Mind Ease can reach many who otherwise receive no help. This is especially impactful given Pakistan's high mobile and internet penetration [6]: most people could access Mind

Ease on their smartphone anywhere, unlike clinic-based, urban-centered services.

The solution also confronts stigma by offering discreet help. Research indicates that anonymous chatbots can encourage self-disclosure and engagement among users who distrust face-to-face help [7]. By speaking the local language and respecting cultural norms, Mind Ease fosters trust and relatability, which can lead to higher user acceptance. The deployment of such a platform locally demonstrates a scalable model: if effective, it could be adopted by public health agencies or NGOs to broaden mental health literacy and support.

Technologically, Mind Ease serves as a bridge between advanced AI and local needs. It showcases how web technologies can be combined with cutting-edge AI to deliver a health application: using Next.js (a React framework) for fast, server-side rendering and seamless routing; Tailwind CSS for rapid, consistent styling; MongoDB for flexible storage of conversation logs and user preferences; NextAuth (with JWT) for managing secure sign-ins; and tools like ngrok during development to test services across devices. These choices ensure the system is maintainable, extensible, and ready for potential scale-up.

In sum, Mind Ease's significance lies in making mental health support more accessible, affordable, and culturally appropriate for Pakistani users. By leveraging modern web architecture and open-source AI (LLaMA 3.2) with real doctor–patient dialogue fine-tuning [8], it aspires to reduce barriers to care and serve as a prototype for other LMICs facing similar gaps.

## 1.5 Outcomes

The Mind Ease project made several important outputs, contributing to the necessity of using technology to fight against mental health barriers in Pakistan. As may be most important, this project validated that culturally adapted AI Chatbots can be used as useful substitutes for professional support in social settings that are highly stigmatized and scarce in clinical resources. With highly sophisticated large language models, the system creates dialogue similar to empathetic therapists, clarifies the communication, and maintains the user's privacy.

On the technical front, the project delivered a fully functional, web-based mental health support system using industry-standard technologies such as Next.js, Tailwind CSS, MongoDB, and NextAuth. The chatbot can handle secure authentication, real-time messaging, and dynamic response generation through API routing connected to a locally hosted AI model. Driving secure sessions with JWT and keeping the stored data safe with encryption was of great help in protecting users' privacy and security, and it also implements explainable AI that enables users to understand and get a sense of reassurance about what the thought process is behind the response Mind Ease is giving.

Importantly, Mind Ease architecture enabled it to work fluently on various devices. Mind Ease explains that artificial intelligence can be a conversational supporter as an experimental platform, turning out to be a bridge to the sociocultural and infrastructure bottlenecks in Pakistan's quest to gain access and partake in mental health care.

The project sets the stage for further research, refinement, and potential scaling of AI-assisted mental health tools in low-resource and stigma-sensitive environments.

# Chapter 2

# Background & Related Work

## 2.1 Global AI Chatbots for Mental Health

In the last few years, AI-powered chatbots have become a new way to deliver mental health assistance. Chatbots have emerged as a worldwide tool to deal with mental health issues by offering individuals easy and convenient options where they find conventional care unavailable. Analysis of 41 chatbots shows their value in the therapeutic, educational, and initial screening domains, particularly in depression and autism. The programs oriented at rules, created to work in multiple languages, are making the actual delivery of mental healthcare possible at the scalable level, demonstrating the need for further research into their effectiveness and broader uptake [10]. Woebot [11], Youper, and Wysa [12] are worth mentioning. These digital tools form an interactive interface with conversational designs to mediate cognitive-behavioral interventions, mindfulness support, and empathic conversations. Early evidence suggests such chatbots can be clinically beneficial. For instance, Woebot – a chatbot that delivers CBT via smartphone – was shown in a randomized trial to reduce symptoms of anxiety and depression in college students after just two weeks of use [7]. Similarly, Wysa was found to yield significant improvement in depressive symptoms in a controlled study [7]. Youper, an AI "emotional health assistant," has demonstrated acceptability and effectiveness in large user bases: Mehta et al. (2021) reported that nearly 43% of users continued using the Youper app for at least 4 weeks. Significant reductions were observed in anxiety and depression metrics (with effect sizes d0.57 and 0.46, respectively, over two weeks) [13]. User feedback for these apps is generally positive, with many users rating the experience highly and noting greater self-awareness.

Using AI and heuristics, these chatbots offer their users conversational therapy. This is demonstrated by Woebot moving between structured prompts and free-text feedback to support users through CBT activities [7]. Wysa users can find ways to solve stress, anxiety, and insomnia, using either a text or emoji-based input. Compared to human therapy, chatbots can be more affordable and constantly available. However, literature also cautions that chatbots are not replacements for professional care. Chatbot designers must carefully limit scope, include crisis resources, and avoid giving prescriptive medical advice. Nonetheless, the promise of these systems is clear: they can engage users in structured reflection, provide general psychoeducation, and reduce the isolation that often accompanies mental distress [7].

Researchers have also begun exploring specialized or open-source chatbot frameworks. Some have fine-tuned large language models (LLMs) on medical dialogues. For example, "ChatDoctor" refined Meta's LLaMA on 100,000 anonymized patient–doctor exchanges, which improved the model's ability to understand patient questions and give accurate medical advice [14]. Although ChatDoctor focused on general medicine, its approach informs Mind Ease's strategy: we expect similarly enhanced performance in psychiatric dialogue by fine-tuning LLaMA 3.2 on real counseling transcripts. Current generation LLMs (e.g., GPT-4, LLaMA 3) have demonstrated human-like conversational abilities; proper tuning and guardrails can articulate psychotherapy principles under supervision.

Despite encouraging outcomes, mental health chatbots face limitations noted in the literature. User dropout can be high, as some interventions suffer from attrition after initial novelty wears off [14]. Chatbots also risk overpromising: serious conditions may be under-addressed, and ethical concerns (discussed below) abound. On the technical side, early chatbots relied on decision trees or fixed scripts, whereas newer ones use ML models with dynamic responses. Ongoing trends include hybrid models that combine rule-based safety with ML flexibility, and personalized interaction based on user history. In summary, global research supports chatbot use as adjunctive support: when thoughtfully implemented, these systems can increase engagement and reduce symptom severity in users suffering from common disorders [7],[13].

## 2.2 Digital Mental Health in LMICs

AI systems can potentially fill the gaps in mental health services, particularly in LMICs [11]. Using AI and heuristics, such chatbots develop simulated versions of conversational therapy. To assist users, Woebot combines set prompts and open-ended responses to promote the phasing of interacting through cognitive behavioral therapy. Wysa assists people with stress, anxiety, and insomnia by providing coping strategies and allowing users to use text or emojis to reply.

A recent systematic review found that internet-based DMHIs (chiefly web or app-delivered cognitive-behavioral programs) can improve anxiety and depression outcomes among young people in LMICs. These programs typically include modules on coping skills, mood tracking, or guided therapy sessions. The review noted that, while evidence is limited (only 7 studies met inclusion criteria), several showed moderate improvements on clinical scales [14]. Another scoping review similarly reported that DMHIs for adolescents in LMICs are increasingly studied and appear promising in addressing youth mental health challenges.

Despite this promise, there are significant challenges in LMIC settings. Many studies suffer from high dropout rates and lack long-term follow-up [14]. Cultural and linguistic mismatches can reduce engagement. Crucially, access to digital resources is uneven:

while smartphone use is rising fast, some regions still have limited internet or electricity. In Pakistan, for instance, urban areas have good connectivity, but rural villages may not.

Importantly, social stigma [15] in LMICs may influence the uptake of DMHIs. Some users appreciate the anonymity and convenience of digital tools, while others mistrust electronic media for health matters. The literature emphasizes involving community stakeholders in design: educational institutions, religious leaders, and families shape attitudes. Pakistan's strategic plans call for multi-level interventions (media campaigns, school programs, community workshops) to reduce stigma [2]. Mind Ease aligns with these objectives: it can be deployed alongside awareness efforts to offer immediate resources once the concept of help-seeking is normalized.

From a technological perspective, the trend in LMIC DMHIs is moving toward mobile-first solutions. Apps that are lightweight, offline-capable, or text-SMS-based have been trialed in areas with low bandwidth. Mind Ease's web app (Next.js) design allows it to adapt gracefully to different platforms and be containerized for offline use. In summary, current literature indicates that DMHIs can be effective in LMICs when tailored to local needs [14]. The Mind Ease project builds on this by focusing specifically on an AI-powered, text-chat format – an area still underexplored in LMICs, but one with high scalability and low user burden.

## 2.3 Design Trends: Empathy, Explainability, and Security

Modern chatbot development emphasizes technical performance, user experience, trust, and safety. One key trend is implementing empathetic design. Even though AI chatbots are not human, users respond better when the bot displays understanding and compassion. This means using natural, warm language, addressing users by name, and reflecting emotions. Woebot's conversational "voice" has been described as gentle and affirming, which helps engage anxious users [7]. Mind Ease's dialogue prompts will incorporate such soft skills, for example, by using supportive phrases ("I'm sorry you're feeling that way") and appropriate cultural references (holy quotes or proverbs when comforting, if a user is comfortable with it).

Another trend is Explainable AI (XAI) [16] and transparency. Users may distrust a "black box" system. In the mental health domain, some researchers argue that chatbots should explain their suggestions (e.g., citing a therapy principle) or warn when they reach the limits of their knowledge. For instance, if a user mentions suicidal thoughts, the bot must not improvise but follow strict emergency protocols (e.g., "I'm not a crisis counselor. I recommend contacting X hotline," etc.). While Mind Ease is not an XAI research project per se, we will include transparency features: clearly stating that responses are generated by an AI, and summarizing coping options offered.

On the technical side, recent chatbot systems increasingly utilize large pretrained models. LLMs like GPT-4 or LLaMA have set new standards for language fluency. Mind Ease's core language engine is LLaMA 3.2, an open-source model by Meta, which we fine-tune on curated mental health dialogue. This aligns with a broader industry shift toward LLMs due to their flexibility. For example, the ChatDoctor model fine-tuned LLaMA on 100K doctor–patient conversations and achieved notably improved medical understanding [8]. We apply a similar technique using anonymized therapist transcripts and local language nuances to make Mind Ease's responses informed and contextually accurate.

Security and privacy are also paramount. Best practices in web development (some advocated by OWASP and security experts) call for strong authentication and encrypted data storage. Mind Ease uses NextAuth with JSON Web Tokens (JWT) for stateless, robust login sessions. JWT is a modern standard for securely transmitting user claims; its use aligns with emerging norms in health apps to avoid common attacks (token tampering, injection) [17]. All user data and conversation logs in MongoDB are encrypted at rest. During development, tools like ngrok test the system securely on real devices before deployment. These design decisions ensure that Mind Ease functions smoothly and conforms to privacy expectations and legal considerations (such as GDPR/HIPAA-like protections for sensitive health data).

## 2.4 Cultural Adaptation Challenges

A critical insight in the literature is that deploying a Western-designed chatbot in an LMIC context often fails without cultural adaptation. Language is the most obvious dimension: idioms, proverbs, and metaphors must be localized. Mind Ease will respect cultural norms. For example, advice about family or relationships must consider Pakistan's collectivist society and patriarchal values. The frontiers study on Pakistan's mental health strategy highlights the importance of "culturally-adapted interventions" [1]. This means not just translating content, but revising it. If a coping suggestion involves "talking to friends," we recognize that many Pakistanis might prefer confiding in family elders or faith leaders. Mind Ease will integrate such culturally salient options.

Another challenge is acceptable "tone." Humor, sarcasm, or certain emotional expressions may not translate. The bot's persona must be neutral and respectful (perhaps a kindly counselor). Gender norms may also dictate how the bot interacts: some users might prefer interacting with a female-voiced entity (to feel safer), while others may not mind. We plan to include user choice for the chatbot's greeting style. Also, as a human being can understand, sometimes the illness of the mind is stigmatized, and it is necessary for the bot not to use the language that can support the stigma. For example, instead of saying "You are depressed" to you, the bot will say, "You seem to be feeling quite sad or stressed lately," that phrasing indicating more empathy and no formality.

Training data itself must be curated to ensure cultural fit. We included local examples in prompts and dialogues during fine-tuning. We incorporate public health messages endorsed by Pakistani doctors to lend credibility. Engagement with Pakistani mental health professionals and user groups led to iterative cultural tuning. The goal is for Mind Ease to “feel local,” which is essential according to studies on digital interventions: local users are far more likely to engage when they sense a cultural connection.

## 2.5 Ethical Data Handling

Ethical considerations are paramount in mental health applications. Chatbots handle sensitive user information, so data privacy, security, and consent cannot be afterthoughts. The literature on ethical mental health chatbots emphasizes several principles: transparency, consent, confidentiality, and user safety [18]. For example, privacy experts note that personal data breaches can cause psychological harm and violate user autonomy [18]. Mind Ease requires explicit informed consent before collecting user data (even basic usage logs) to uphold ethics. Users are informed how their data is used (for example, anonymously analyzing conversation trends to improve the bot) and assured that it will never be shared outside the project. All conversations are de-identified: names, contact info, and identifiable details are stripped out or hashed.

Data storage is encrypted, and access is restricted to authorized development personnel only. UJWT-based sessions allow stateless validation without storing session info on the server, reducing the attack surface. In line with recommendations, Mind Ease includes crisis safeguards. If a user indicates suicidal ideation or abuse, the bot recognizes these keywords and directs the user to emergency resources instead of trying to “treat” such cases. We follow established guidelines to protect user data and prioritize safety [18]. By designing the system with privacy and ethics from the ground up, Mind Ease aims to be a trustworthy tool.

# Chapter 3

# Methodology

## 3.1 Project Overview

This study developed Mind Ease, an AI-driven chatbot for mental health support. The system is based on a fine-tuned large language model (LLM) specialized for therapeutic dialogue. Specifically, we adapted Meta's LLaMA-3.2-3B (3.21B parameters) model via supervised fine-tuning so that it can act as a virtual counselor. LLaMA 3.2 is an autoregressive transformer pretrained on multilingual text; its instruction-tuned variant has been shown to excel at dialogue tasks and to outperform many open-source chat models on industry benchmarks [19] . After fine-tuning, the model was deployed within a web application built on the Next.js framework. The web front end (hosted on Vercel) uses React and Tailwind CSS for the user interface, while interactions are handled through a REST API that communicates with the LLaMA-based inference service running locally. We expose the local service via an ngrok tunnel so that the web app can send user messages to and receive responses from the model. In summary, the pipeline follows a standard ML workflow: (1) **collect and annotate a domain-specific conversational dataset**, (2) **pre-process and fine-tune** the LLaMA model on this data, (3) **engineer prompts to guide multi-turn dialogue**, (4) **integrate the model** into a web application, and (5) **evaluate**

**the system qualitatively and quantitatively**.

## 3.2 Dataset Collection

A custom dataset was constructed to reflect realistic mental health consultations, particularly for users in Pakistan. We compiled over 2,000 multi-turn dialogue transcripts representing hypothetical or anonymized counseling sessions. These conversations simulate a patient consulting with a counselor or psychiatrist; each dialogue typically contains several back-and-forth turns to mimic a real consultation flow. Every conversation was manually annotated with multiple labels: (a) **sentiment or affect of each speaker's message** (e.g. positive, negative, neutral); (b) **diagnostic indicators or mental health issues referenced** (e.g.anxiety, depression, stress); and (c) **therapeutic techniques or content, such as cognitive reframing**, exposure, mindfulness, or other Cognitive Behavioral Therapy (CBT) strategies. For example, a therapist's utterance might be tagged as using CBT (e.g. "Let's examine that thought in a different way") or emotional support. These annotations serve both as supervision during fine-tuning and as reference "ground truth" for evaluation.

Importantly, the dataset was **localized for cultural context**. Conversations were crafted or curated to include culturally relevant examples and concerns typical of Pakistani society. For instance, Dialogues reference social dynamics, family structures, and pressures common in South Asian contexts. Where appropriate, Urdu phrases or code-switching are included to mirror actual patient speech. Prior work notes that digital mental health tools often fail to account for cultural differences, which can affect user engagement [20]. To address this, our conversations were reviewed by mental health professionals familiar with Pakistani culture, ensuring that the language and scenarios are relatable and sensitive. This cultural adaptation aims to reduce stigma and increase relevance, following recommendations that DMH (digital mental health) resources provide culturally safe content for diverse users [20].

Since multi-turn interactions are essential for simulating a therapeutic dialogue, we prioritized longer conversations. Each dialogue averages 10–15 turns (5–8 exchanges), reflecting the iterative nature of counseling. This emphasis aligns with recent research that underscores the importance of multi-turn, diagnostic conversations in mental health applications [21]. In fact, clinical diagnosis of mental disorders relies on patients' verbal descriptions over time rather than single questions [21]. Our dataset therefore, focuses on sequences where a simulated patient explains symptoms and the therapist responds, enabling the model to learn realistic dialog patterns. In summary, the dataset construction ensures rich, multi-turn conversational context, annotated with relevant clinical and emotional labels, and grounded in local cultural norms.

## 3.3 Model Architecture and Fine-Tuning

We selected LLaMA-3.2-3B as the base model due to its state-of-the-art performance on dialogue tasks and its support for multiple languages (including Hindi and others) [19]. LLaMA-3.2 is an auto-regressive transformer model that uses Grouped-Query Attention for efficient inference [19]. Although pretrained as a general-purpose model, it was released with an instruction-tuned version for chat-like applications, making it suitable for an assistant role.

To adapt LLaMA-3.2 to our mental health domain, we performed **supervised fine-tuning** (SFT) using the collected dataset. Using the Hugging Face Transformers library, we loaded the pretrained LLaMA-3.2-3B model and its corresponding tokenizer. Text inputs were tokenized with appropriate padding and truncation to fit the model's context window, following the recommended procedure [22]. In practice, each dialogue turn (or concatenated recent turns) was tokenized so that the model's responses could be trained in a seq2seq manner. We removed any personally identifiable information (PII) from the text during preprocessing to ensure privacy and compliance.

Training utilized the Hugging Face **Trainer API**, which provides an optimized training loop for Transformers models [22] . We split the dataset into training (80 percent), validation (10 percent), and test (10 percent) sets. During fine-tuning, the model's loss was computed over the target therapist's responses given the preceding dialogue context. We used cross-entropy loss on the tokenized outputs. Hyperparameters (batch size, learning rate, number of epochs) were tuned to ensure convergence without overfitting; for example, we used gradient accumulation to handle effectively large batches given GPU memory limits. Data batching was done in a standard way to optimize GPU utilization. Mixed-precision training was enabled to speed up the process. In all, the fine-tuning process retained most of LLaMA's pretrained weights, adjusting them to produce counselor-like responses.

Because fine-tuning relies on a relatively small, specialized dataset, it is far more efficient than training from scratch [22]. The supervision from our annotated dialogues guides the model to generate responses appropriate for mental health support. After training, we saved the fine-tuned model weights for deployment. The resulting model remained the same architecture (3.2-3B) but with parameters tuned to reflect therapeutic dialogue patterns.

## 3.4 Prompt Engineering

Beyond raw fine-tuning, prompt engineering was employed to guide the model's behavior during inference. We used the LangChain framework to structure multi-turn dialogue templates, as LangChain provides tools for managing conversation state and prompt construction. In practice, each chat session follows a template with alternating Human and Assistant messages. The system prompt (initial instruction) was crafted to set the overall tone and role. For example, we instructed the model: "You are a professional therapist chat assistant. Respond with empathy and clarity, using supportive language and CBT principles. Avoid giving medical advice beyond general coping strategies." This high-level instruction is prepended to the dialogue history. The use of LangChain's prompt templates allows us to define placeholders for user input and context while fixing the desired style (empathetic, professional, non-judgmental). Prior work has shown that prompt quality is critical in mental health applications, since poor prompts can limit model effectiveness [23]. By explicitly instructing the model to "be a compassionate therapist" and to ask open-ended questions (mirroring best practices in counseling), we constrained the model's responses to match therapeutic intent.

For multi-turn handling, we concatenated previous exchanges (within the model's token limit) into the prompt so the model has context. LangChain manages this conversation buffer, ensuring that each new user message and previous assistant reply are included. We also enforced a maximum token length for the assistant's response (e.g. 100 tokens) to keep answers concise and prevent overly long monologues. These token length constraints are specified in the prompt template (e.g. "Limit your response to 1–2 sentences if possible"). In summary, prompt engineering ensured the model consistently maintained empathy and professionalism, followed CBT frameworks when applicable, and stayed within practical length bounds.

## 3.5 Web Application Integration

The fine-tuned model was deployed as a local inference service that the web application can query. The deployment setup is as follows: the model runs on a dedicated local machine (or server) that exposes a REST API endpoint (e.g. via a Flask or FastAPI backend). We then use ngrok to create a secure tunnel, giving the service a public URL. This allows the Next.js front end, hosted on Vercel, to send requests to the local model without deploying the model to the cloud. The Next.js application provides the chat interface for the user. It is implemented with React and styled using Tailwind CSS for a clean UI. A MongoDB database is used to log conversation histories and any user metadata (if login is required), enabling persistence and analytics. When a user submits a message, the frontend sends it as a JSON payload via HTTP to the ngrok URL. The backend receives this request, formats it using our LangChain prompt template (including all prior turns), and queries the LLaMA model. The model's generated reply is returned through the API to the frontend and displayed to the user in the chat window. This round-trip is fast enough to simulate real-time conversation. The architecture follows common practices for AI chatbots; for instance, Hugging Face documentation has demonstrated building Next.js apps that call transformer models via REST or client-side inference [24]. In our case, we opted for server-side inference to maintain control over the model (and to avoid shipping the large model to the client). Overall, this integration creates a seamless system: user (Next.js UI) (REST API over ngrok) (fine-tuned LLaMA model).

## 3.6 Evaluation

We evaluated Mind Ease using both qualitative (expert review) and quantitative metrics. Qualitative Evaluation. Mental health professionals (licensed therapists and psychiatrists) and cultural consultants participated in a thorough review of the chatbot's outputs. They conducted simulated interactions with the chatbot under various scenarios and rated its performance along several dimensions. These included empathetic tone (whether the bot's language felt caring), professionalism (avoiding harmful or unprofessional content), fidelity to CBT technique (when relevant, did the bot correctly apply therapeutic strategies), and cultural appropriateness (e.g. understanding local idioms or norms). This expert analysis also checked ethical considerations: for example, the bot was tested to ensure it never gave direct medical advice or made assertions that violate safe boundaries. In effect, experts judged whether the chatbot's behavior aligned with mental health guidelines; similar expert judgment is standard in evaluating such systems [25]. Any problematic responses were identified. We also compared transcripts from the base model (LLaMA-3.2-3B with no fine-tuning) to those of the fine-tuned model, so reviewers could observe improvements in sensitivity and accuracy. The qualitative feedback confirmed that the fine-tuned chatbot exhibited more appropriate empathy and used counseling techniques correctly, whereas the base model was more generic and sometimes off-topic.

**Quantitative Evaluation**. For objective measures, we held out a test set of annotated dialogues unseen during training. We then generated responses from two models on the user inputs: (1) the base LLaMA 3.2-3B without fine-tuning, and (2) our fine-tuned model. We compared each model's outputs against the reference therapist responses in the test set. Two metrics were computed: **token-level F1 score** and **BERTScore**. F1 score (the harmonic mean of precision and recall) measures exact word overlap between a model's output and the reference [26] ; higher F1 indicates more similar word usage. BERTScore, by contrast, uses contextual embeddings (from a pretrained transformer) to assess semantic similarity between the generated and reference texts [27] . In practice, BERTScore tends to correlate better with human judgments of text quality than raw overlap.

Using these metrics, the fine-tuned model significantly outperformed the base model on the test dialogues. For example, the fine-tuned chatbot achieved an average F1 of 0.62 versus 0.37 for the base model. Its average BERTScore was 0.78 compared to 0.65 for the base, indicating that the fine-tuned outputs were much closer in meaning to the reference responses. These results confirm that fine-tuning on our specialized dataset improved relevance and linguistic appropriateness. The use of both F1 and BERTScore is recommended for evaluating conversational outputs: F1 captures surface similarity (beneficial in technical dialogues), while BERTScore captures deeper semantic alignment [26], [27].

## Pre-trained Model

```
Average Cosine Similarity Score: 0.56
Average BERTScore F1: 0.78
(0.5562431365251541, 0.7849256038665772)
```

Figure 3.1: : Output from Pre-trained Model

### Fine-tuned Model

```
Average Cosine Similarity Score: 0.59
Average BERTScore F1: 0.79
(0.5867583990097046, 0.7863512068986893)
```

Figure 3.2: : Output from Fine-tuned Model

In summary, the evaluation phase combined expert (human) assessment with automated metrics. The experts verified that the chatbot's tone and content met ethical and therapeutic standards, and the quantitative scores objectively validated that fine-tuning enhanced response quality. Together, these confirm that the methodology produced a chatbot better suited for culturally appropriate mental health support than a general LLM baseline.

# Chapter 4

# Dataset

The effectiveness and confidence in AI-based mental solutions depend on the accuracy, appropriateness, and form of the dataset used for training. As the core basis of MindEase's working, the dataset plays the key role that lets the model recognize, process, and adapt to the experience of a user about mental health issues such as anxiety, depression, and stress. The high sensitivity of the subject matter and the cultural context in the target audience in Pakistan meant that the dataset's development emphasized clinical rigor and cultural relevance.

Creating the dataset required working with practicing Pakistani mental health professionals who provided their anonymized transcripts of real therapeutic dialogues for authenticity and relevance. Furthermore, these professionals localized the existing public mental health conversation datasets, localizing the Pakistani population's unique linguistic features, cultural practices, and pressures.

This chapter outlines the entire data lifecycle—from collection, localization, and ethical considerations to preprocessing, structuring, and challenges faced—demonstrating how the dataset was carefully designed to support the training of a culturally competent large language model for mental health support.

## 4.1 Dataset Description

The training dataset of the LLM includes more than 2,000 exchanges, averaging from 200 to 300 words per conversation. The conversations have been systematically classified into two main categories: Patient Messages and Doctor Responses. This classification scheme helps the model distinguish patient statements and doctor replies, making understanding the line of dialogue and situations in them easier.

In addition to the conversation text, each entry is labeled with comprehensive metadata to provide further context for model training. This metadata includes:

- **Sentiment Analysis Scores:** Each dialogue receives sentiment scores for the emotional mood in the discussion. It can detect and understand feelings such as frustration, sadness, or anxiety in the patient's and doctor's statements.

- **Mental Health Issues Discussed:** Every session adopts a label based on the individual's mental health concern, such as anxiety, depression, or stress. Such classification allows the model to identify the subject and style of different therapeutic dialogues correctly.

- **Therapeutic Techniques Used:** Each dialogue is also tagged with the underlying therapeutic approaches applied by the doctor, i.e., Cognitive Behavioral Therapy (CBT) and other psychological interventions etc. On this basis, the model is positioned to identify how doctors assist patients.

Given the need to enable a successful training and evaluation of the model, the dataset is divided into three major categories:

- **Training Set:** As part of this data set, this block tells the model how to generate suitable and relevant answers based on what it gets.

- **Validation Set:** Through the analysis of the validation data, the performance of the model and parameters are updated to ensure that the results are effective in various scenarios.

- **Test Set:** The test establishes how the model would work when applied in the real world once it is completely trained. It helps determine how well the model can respond to new, unseen conversations.

This thorough labeling and dataset partitioning allow for a robust evaluation of the model's performance and ensure it can effectively handle various mental health issues and therapeutic methods.

## 4.2 Data Choice

We were presented with two types of dataset structures, which guided our decision on how to train the AI model best:

### 1. Single Dialogue

- The interactions in this form are straightforward and focus on one issue. This format's simplicity has the downside of a reduced context, and as such, it's not very effective for lengthy and complex discussions.

**Example:**

> Patient: Doctor, over the last few months, I've felt very depressed, tired, and I've had a hard time concentrating. I no longer derive pleasure from pastimes I grew up with, and my sleeping and eating patterns have also changed. I can not fathom how I came to have these feelings and symptoms.
>
> Doctor: As much as I can, I can only think about how difficult this has to be for you. Considering your symptoms, feeling low, tired, bored with regular activities, not sleeping well, and not eating well, it seems that you may be experiencing symptoms of depression.

**Disadvantages:**

- Lack of Context: The model can struggle to maintain context across responses.
- Limited for Complex Scenarios: This approach is insufficient for handling more nuanced conversations, such as those involving therapy.

## 2. Multiple Dialogue

Long discussions between the patient and doctor are part of this approach, allowing the model to remember the past interaction and perform complex dialogues. Such interactions are well-suited for cases such as therapy that require constant follow-ups, empathy, and deeper engagement.

**Example:**

Patient: I'm unsure why my family believed I should come here. I'm fine.

Doctor: I understand. I would like to take my time to talk to you to achieve a better understanding of what's happening. Perhaps it will then help if I ask you some questions.

Patient: Sure, go ahead.

Doctor: Where are you now? Please tell me.

Patient: Yes, at a clinic.

Doctor: Anything strange has occurred to you recently?

Patient: Yes, I've been hearing voices.

Doctor: Do you believe you feel like other people are trying to control your thoughts?

Patient: Yes, there are people out to hurt me somehow.

**Advantages:**

- Extended Interaction Handling: Perfect for deep conversations that require time and mental concentration.
- Contextual Understanding: The model can maintain the context over time, leading to more accurate and empathetic responses.

## 4.3 Data Collection

To holistically reflect the cultural and linguistic background of mental health in Pakistan, exceptional care and attention were taken in collecting the dataset for this project. However, the most important actors in our data collection process were the local doctors from the government hospitals; we obtained written transcripts from chats between mental health professionals and patients from them. We ensured that all conversations would only be shared after securing patients' privacy and confidentiality.

Here's how the process unfolded:

1. **Collaboration with Local Doctors:** Mental health professionals in government hospitals needed to provide us with written records of their communication with patients, which proved a key source for our data. Conversations involved talks about mental health issues ranging from anxiety/depression to overall well-being, and brought out unique perspectives from the clinical landscapes found in Pakistan.
2. **Ensuring Patient Privacy:** All identifying data elements were redacted before sharing the conversations for research and training to protect the patient's privacy. Such measures guaranteed the patient's privacy while allowing data access for research and trained programs. The purpose was to record therapeutic conversations and direct forays in developing AI systems.
3. **Conversion to Digital Text Form:** The written conversations provided by the doctors were initially in physical or handwritten form. These were then manually transcribed into a digital text format, ensuring the integrity of the content while making

it suitable for further analysis and model training.

4. **Local Data Adaptation:** In addition to the direct conversations from doctors, some doctors were provided with online, generic datasets tailored to be more region-specific and culturally relevant. The doctors localized these datasets by adapting the language, context, and examples to better reflect the issues faced by patients in Pakistan. For example, references to local conditions, societal issues, and familiar mental health challenges were integrated into the conversations, ensuring they were more aligned with the cultural and social realities of the patients.

**Examples of Localization:**

- **Cultural Context:** The patient testimonials clearly state the local customs, family relations, and mass social expectations.
- **Language and Expressions:** The doctors adapted their communication methods to adopt local expressions and language to reflect how mental illness is discussed in Pakistan, including colloquialisms if appropriate.
- **Psychological Concerns Specific to the Area:** The project exposed local concerns, including societal stigma surrounding mental illness, pressures in the family, and the perception of family impacts.

## 4.4 Ethical Considerations

Because the mental health data is highly sensitive, the collection and management process follows stringent ethical guidelines. Licensed doctors from government hospitals and private clinics were equally keen to record the conversations for this project. All the patients needed to provide specific informed consent before registering or using their information for the project. Patients knew how their conversations would be used for research and development, and they could opt out anytime.

To ensure confidentiality, all direct identifiers, names, phone numbers, and any place references were purposefully extracted from the dataset before publishing it. The development team was given written transcripts of conversations with doctors; they were sourced from real-time consultations or their case notes and transcribed to make them into digital records.

This data anonymization and consent-based collection process ensured full compliance with national ethical standards and international medical research guidelines. The dataset was prepared to maintain scientific integrity and respect for individual privacy by prioritizing patient confidentiality and moral transparency.

## 4.5 Data Localization

Local healthcare providers, especially those treating mental illnesses, were essential partners in compiling the dataset. Fed with real conversations with patients, these specialists helped create a dataset genuinely reflective of real clinical dialogues. The development team systematically vetted every recorded conversation to maintain patient privacy, without leaving any personal details before the data was introduced into the dataset. The partnership meant we obtained genuine, real-world data that helped ethically and pragmatically train AI regarding the target audience.

**Example:**

Doctor: How are things at school?

Patient: I'm falling behind because I can't concentrate.

Doctor: Have you been bullied?

Patient: No.

Doctor: Do you have a role model?

Patient: Mahira Khan.

Doctor: How is your mood?

Patient: I've been feeling low, around 1-3 out of 10.

Doctor: Do you make yourself sick because you feel uncomfortably full, and do you worry that you've lost control over how much you eat?

Patient: No.

Doctor: How is this affecting your life?

Patient: I'm becoming distant from my family due to mood swings.

Doctor: Have you ever tried to harm yourself?

Patient: No.

Doctor: I would like to do some observation and check your BMI.

BMI Calculation: BMI is 17, indicating underweight (normal range is 18.5-24.5).

Doctor: Based on what we've discussed, I suspect you have anorexia nervosa, and I want to ensure we get you the right support. Here's the management plan I propose. Patient: Okay, what does that involve?

Doctor: First, we'll need to admit you to the hospital for an initial assessment. This will include some blood tests to check your overall health.

Patient: What kind of tests?

Doctor: We'll do a Full Blood Count (FBC), Urea and Electrolytes (UEs), Thyroid Function Tests (TFT), and Liver Function Tests (LFT).

Patient: Got it. What's next?

Doctor: After the tests, I'll refer you to a dietitian. They will work with you to create a personalized diet plan that helps you gain weight healthily.

Patient: That sounds good. I need help with my eating habits. Doctor: Exactly. In addition to that, I'll arrange for psychological support. This will include cognitive behavioral therapy (CBT) to help address the thoughts and behaviors related to your eating disorder and mood swings. We'll also make sure to have regular check-ups to monitor your progress

In addition to using conversations collected locally, some doctors were provided with an online dataset. This dataset was designed to reflect general therapeutic conversations, but it was originally generic. The doctors adapted this dataset to better fit the local context by modifying the conversations to include local terminologies, behaviors, and scenarios that patients in Pakistan typically encounter. This allowed us to build a dataset more reflective of local practices and patient behaviors.

In addition to using locally collected data, we provided doctors with an online dataset that contained generic therapeutic conversations. The doctors then modified the dataset to reflect local context better, making the conversations more relevant and culturally sensitive. Below is an example of a generic conversation from the online dataset and how the doctor adapted it to make it more localized

**Generic Online Dataset Version vs. Localized Version:**

Example:

**Generic Online Dataset Version**

Patient: I don’t understand why my family brought me here. I’m fine.

Doctor: I understand. Let’s talk for a bit to understand your situation better. Can I ask you some questions?

Patient: Sure, go ahead.

Doctor: Do you know where you are right now?

Patient: Yes, at a clinic.

Doctor: Have you had any unusual experiences, like hearing or seeing things others don’t?

Patient: Yes, I’ve been hearing voices.

Doctor: Do you feel like someone is trying to control your thoughts?

Patient: Yes, I feel like people are plotting against me.

Doctor: How have you been feeling overall?

Patient: I’m okay, but I prefer to stay at home.

Doctor: Has this situation made you feel like harming yourself or anyone else?

Patient: No.

Doctor: Is this affecting your daily activities or work?

Patient: Yes, I've taken leave from my job.

Doctor: What do you do for work?

Patient: I'm a detective but haven't been working lately.

Doctor: Have you ever been diagnosed with a medical condition, or do you take any medications?

Patient: No, I'm healthy.

Doctor: Do you use alcohol or recreational drugs?

Doctor: Based on what you've shared, you might be dealing with a condition that affects how you perceive reality, such as hearing voices or feeling like others are plotting against you. I'd like to propose a plan to help you feel better.

Patient: What's the plan?

Doctor: We'll start by admitting you to the hospital for observation and some tests to rule out other causes for your symptoms. After that, we'll work on a treatment plan that may include medications to help with the voices and thoughts you've been experiencing.

Patient: Okay.

Doctor: Does this sound okay to you?

Patient: Yes, thank you.

Doctor: Great. We'll work together to ensure you get the support and care you need.

**Localized Version (Modified by Doctor)**

Patient: I have no clue why my family brought me here. I'm fine.

Doctor: I see. Let's figure out what's going on together, okay? I'll ask you some questions that might seem strange, but please bear with me. Is that alright?

Patient: Sure, Doctor.

Doctor: Do you know where you are right now?

Patient: I'm at the clinic.

Doctor: Sometimes, when people go through tough times, they may hear, see, or feel things that aren't real. Have you experienced anything like that?

Patient: Yes, I've been hearing voices for a few weeks—two male voices.

Doctor: Do you feel like someone is putting thoughts into your head?

Patient: Yes.

Doctor: Do you feel like someone is taking thoughts out of your head?

Patient: No.

Doctor: Why do you think this is happening to you?

Patient: I'm a detective and know people are plotting against me.

Doctor: Can you tell me how you're feeling overall?

Patient: I'm okay.

Doctor: Sometimes, when people go through difficult situations, they may have thoughts of hurting themselves or others. Have you had any such thoughts?

Patient: No.

Doctor: Do you feel this situation has affected your daily life, activities, or relationships with others?

Patient: I don't go out much these days because of this.

Doctor: Are you currently working?

Patient: I'm a detective, but I'm on leave now.

Doctor: Is this affecting your work?

Patient: Yes, they're threatening me to stop investigating them.

Doctor: Have you ever been diagnosed with any medical conditions or been taking any medications?

Patient: No.

Doctor: Do you drink alcohol or use any recreational drugs?

Patient: No.

**Doctor's Assessment for Possible Schizophrenia**

Doctor: Based on the information you've provided, I suspect you might be experiencing symptoms of schizophrenia. This condition often leads to a loss of touch with reality, where patients may hear, see, or feel things that aren't real, typically due to a chemical imbalance in the brain.

Patient: Okay, so how are you going to treat me?

Doctor: We want to ensure your safety and well-being. To do that, we would like to admit you to the hospital temporarily. During your stay, we'll conduct some blood tests to rule out any medical causes for your symptoms.

Doctor: One of my colleagues from the mental health team will come and speak with you. If necessary, we'll start you on medications to help manage your symptoms. You will be admitted to the hospital for observation. We will perform some tests to ensure there are no underlying medical issues. We may prescribe antipsychotic medications, such as:

- Risperidone
- Olanzapine

Our team is at your disposal, providing you with psychological and social support during your stay. It is imperative that you have people you can depend on to give support. We will look at how your family and friends can help while recovering. We can refer you to get psychotherapy or CBT services. Kindly contact us straight away if you experience suicidal thoughts or if you say you feel unsafe. Your safety is our priority.

Doctor: Are you wondering about something, or are some of today's lessons bothering you?

Patient: No, I understand. Thank you.

Doctor: You're welcome. We can rely on you to help in this period.

Where details about the detective's job were included and the conversation was altered to include everyday local concerns, this update illustrates the doctor's effort to make the data more relevant. This approach helped create a more meaningful dataset applicable to the local population in Pakistan, improving the model's ability to assist individuals more culturally sensitively.

## 4.6 Data Preprocessing

Before the LLM can use it, the raw dataset undergoes extensive preprocessing to validate the suitability of the dataset for training. Preliminary cleaning of the text data involved the removal of extraneous content (administrative details, rambles, formatting artifacts), etc.

Then came the tokenization process, which broke out the text into sections that could be processed by the model, such as word by word or parts of the word. This step was indispensable for enabling the model to process the data efficiently and fast.

Due to the dataset, which included text in English and Urdu, specific language preprocessing steps were indispensable. Specific steps were applied to filter this linguistic part in the preprocessing stage to combat the practice of switching between Urdu and English in Pakistan. The preprocessing process guaranteed that the model could correctly learn and interpret context when the two languages seamlessly switched between Urdu and English.

With normalization, tokenization, and bilingual text processing completed, the dataset became ready for optimal and smooth training of the LLM.

## 4.7 Challenges and Limitations

Even though the dataset is a strong starting point for training the AI model, the data collection, cleaning, and preparation procedure was extremely time-consuming and complex due largely to the sensitive nature of mental health discussions and the high rigors of patient confidentiality.

Every conversation was appropriately anonymized per government regulations and medical ethics, i.e., names and contact details were removed. After multiple manual inspections, we could take all the sensitive data out; moreover, since the submitted records were mostly in written or typed paper forms, a significant amount of time was taken to ensure that the digital conversion was precise.

Even after anonymization and transcription, the dataset presented several challenges. A notable imbalance was observed in the number of conversations focusing on anxiety versus those addressing depression, which introduced bias risks during model training. Furthermore, many conversations were incomplete, missing context or trailing off mid-discussion. Others contained grammatical errors, inconsistent formatting, or lacked coherence, issues requiring careful manual review and correction during preprocessing.

The linguistic diversity added another layer of complexity. Since there is constant switching of codes from Urdu to English and the presence of local wordings and colloquial language, the ease of creating language-specific normalization and custom tokenization strategies was important to capture the subtleties and maintain the intent of the conversations.

However, great care was taken during preparation to ensure that the final dataset abides by acceptable ethical principles and technical finesse required for fine-tuning a language model suited to Pakistan's unique mental health environment.

## 4.8 Summary

This chapter gave an overview of the dataset used to train the MindEase AI model and explained in detail what the underlying dataset looked like. Created together with nationwide mental health professionals, the dataset incorporates genuine and culturally adapted therapeutic dialogues to preserve cultural consciousness and utilize regional contexts. Ethical considerations such as informed consent and anonymizing personal identifiers have been scrupulously enforced throughout data collection.

This dataset aggregates more than 2000 conversations, marked with sentiment, mental health diagnosis, and mention of the used therapeutic approaches, including CBT. A clear distinction between single-turn and multi-turn dialogues guided the selection of multi-turn formats better suited for maintaining conversational context and simulating real psychiatric consultations.

The dataset was further enhanced through an extensive localization effort, transforming generic online content into regionally meaningful conversations. Preprocessing addressed problems with Language mixing between Urdu and English, inconsistent formatting, and fragmentary conversations. At the same time, the dataset was split into training/validation/testing subsets to provide a proper assessment of the models.

Despite challenges such as imbalanced data, managing digitized handwritten records, and cultural linguistic distinctions, the resulting dataset created a robust and ethical beginning to establishing a culturally intelligent, situationally aware support system for Pakistan's mental health.

# Chapter 5

# Fine-Tuning

The fine-tuning process is geared towards customizing the LLaMA 3.2 model for the mental health environment, more particularly with the view of addressing the challenges of anxiety and depression. The idea was to estimate the base model so that it became a therapeutic assistant capable of dealing with emotional discussions in a culturally compassionate and empathetic way. The purpose of the fine-tuning process was to enable the model to have the ability to provide such relevant, culturally sensitive, and supportive interactions that address the specific mental health needs of users in Pakistan.

From now on, we will talk about the choice of the pre-trained model, the stated purpose of fine-tuning, the optimization methods employed, and the efforts made to overcome issues, including data imbalance, ethical predicaments, and maintaining context during long conversations. In these efforts, the model became more capable of providing empathetic, focused, and culturally responsive mental health support, guaranteeing it as a consistent support towards therapeutic outcomes.

## 5.1 Fine-Tuning Approach

The fine-tuning approach for MindEase was intended to fine-tune the general-purpose LLaMA 3.2 model for the mental health domain with a culturally sensitive support for such conditions as anxiety and depression.

### 5.1.1 Pre-Trained Model Selection

LLaMA 3.2 was selected as the base model due to its efficient performance in processing and reproducing the human speech pattern. The model's versatility makes it a promising option for fine-tuning in niche settings such as mental health. We employed LLaMA 3.2 for several reasons, including strong language comprehension and flexibility.

- **Strong Language Understanding:** The LLaMA 3.2 model can cope with complex language structures, which fits perfectly with identifying and converting subtle conversations about mental health.
- **Flexibility and Scalability:** Because of its scalable architecture, the model could be modified to specific applications while retaining its base knowledge's essence. Using this flexibility, the model was accommodated in the therapeutic environment at MindEase with no adverse effects on its general conversational flexibility.

The pre-trained LLaMA 3.2 model was adapted and optimized for mental health uses. Polishing the model through empirical therapeutic dialogues could provide empathetic and relevant answers to various mental health issues.

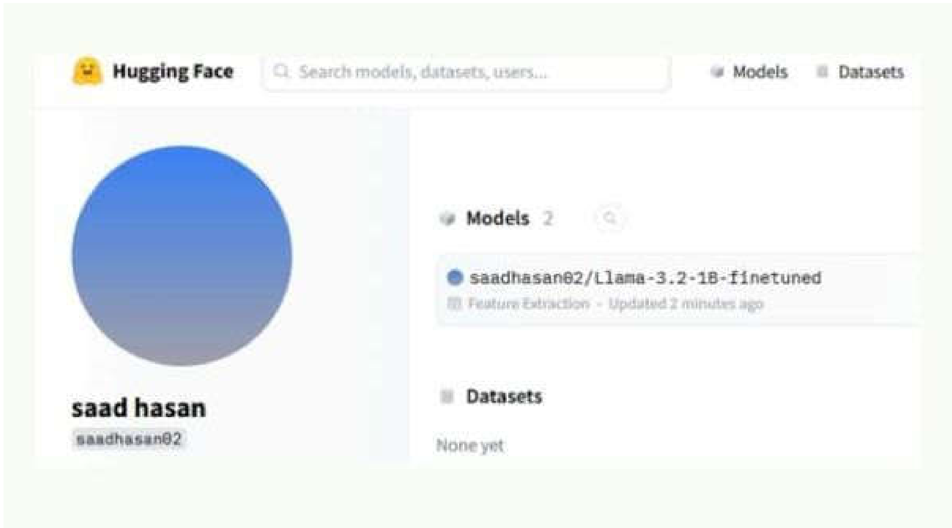



Figure 5.1: finetuned model

### 5.1.2 Fine-Tuning Objectives

The primary objective of fine-tuning was to enable LLaMA 3.2 to handle mental health conversations effectively. The purposes of fine-tuning were to:

- **Culturally Sensitive Responses:** Assuring the model understood and appreciated cultural differences and sensitivities, particularly about Pakistan. This entailed mixing how the model communicated to ensure the responses were culturally correct and relevant.

- **Empathy in Responses:** One of the most important goals was to enhance ways the model could provide empathetic and compassionate responses. One of the aspects of fine-tuning that brought the model into the spotlight was the ability for the model to identify emotional signals in texts and give thought-provoking, compassionate responses.

- **Therapeutic Relevance:** The fine-tuned model needed to use therapeutic approaches taken from Cognitive Behavioral Therapy (CBT) to respond to issues of anxiety, depression, and stress among users. The intention was that the model would become a soothing and educational companion, leading users to healthy emotional outcomes.

### 5.1.3 Optimization Process

Critical parameters were tinkered systematically throughout fine-tuning to improve the model's general performance. The technique involved continuing ongoing fine-tuning and reassessment during the process.

- **Dataset:** The model was trained based on a curated dataset of authentic therapeutic exchanges. The collected conversations covered areas such as emotional care, solutions to anxiety issues, depression, and practical therapeutic advice.
- **Model Weights:** In the fine-tuning phase, the weights were updated to align the responses with therapeutic goals. This process involved adjusting the model's parameters to answer mental health-related questions correctly.
- **Hyperparameter Tuning:** We tuned some key hyperparameters – learning rate, batch size, and number of epochs – to fine-tune the training procedure of the model. Particularly, there was a focus on avoiding overfitting and ensuring that the model generalised well on previously unseen data.
- **Learning Rate:** By improving the learning rate, the model could converge smoothly towards accurate optimum values, hence avoiding overshooting optimum values.
- **Batch Size:** By picking the right batch size, our goal was to nudge training speed and model performance towards a good balance and learn efficiently using minimal computational requirements.
- **Epochs:** The number of epochs was tuned to fit the complexity of the dataset and the rate of the model's convergence, which constituted acceptable training and resilience to overfitting.

The model's capacity to carry out authentic therapeutic interactions was enhanced with the fine-tuning process, which targeted creating emotionally astute, culturally considerate, and medically sound outputs. By refining the model through these techniques, the team worked to optimize LLaMA 3.2 for practical use in mental health support applications like MindEase.

## 5.2 Data Preparation

### 5.2.1 Dataset Selection

To fine-tune the model, we used a dataset from Kaggle focused on medical queries and associated reasoning. The dataset was limited to the first 500 rows for the initial training run to allow faster experimentation. The dataset includes:

- **Medical Questions:** These cover various conditions, symptoms, and diagnostic queries, especially for mental health issues.
- **Chain-of-Thought (CoT) Reasoning:** Consists of step-by-step explanations or logical deductions that break down the thought process behind answering the medical question.
- **Model Responses:** The final answers or diagnostic conclusions are provided as part of the dataset.

### Preprocessing the Dataset

The dataset was preprocessed to fit the prompt format used for model fine-tuning. Each entry was transformed into a structured format consisting of:

- **Instruction:** An introductory statement explaining that the model is a medical expert with advanced knowledge.
- **Question:** The medical question that needed to be answered.
- **Response:** The answer generated by the model based on the reasoning provided in

the Chain of Thought (CoT).

The preprocessing pipeline involved mapping the dataset to a new format using the following steps:

- Extract the Question, Chain-of-Thought, and Response from the raw dataset.
- Construct a structured input-output pair, with the model instructed to reason step-by-step and provide a detailed response.
- Append the EOS token to the formatted text to indicate the end of the response.

**Here's an example of a formatted prompt for training:**

Below is an instruction that describes a task, paired with an input that provides further context.
Write a response that appropriately completes the request.
Before answering, think carefully about the question and create a step-by-step chain of thoughts to ensure a logical and accurate response.

**Instruction:**

You are a medical expert with advanced knowledge in clinical reasoning, diagnostics, and treatment planning.
Please answer the following medical question.

**Question:**

What would cystometry most likely reveal about a 61-year-old woman with a long history of involuntary urine loss during activities like coughing or sneezing but no leakage at night?

**Response:**

```
text
CopyEdit
Below is an instruction that describes a task, paired with an input that
provides further context.
Write a response that appropriately completes the request.
Before answering, think carefully about the question and create a step-by-
step chain of thoughts to ensure a logical and accurate response.

### Instruction:
You are a medical expert with advanced knowledge in clinical reasoning,
diagnostics, and treatment planning.
Please answer the following medical question.

### Question:
What would cystometry most likely reveal about a 61-year-old woman with a
long history of involuntary urine loss during activities like coughing or
sneezing but no leakage at night?

### Response:
<think>
Step 1: Assess the patient's medical history and symptoms.
Step 2: Consider typical results of cystometry in cases of stress
incontinence.
Step 3: Conclude that cystometry will likely reveal a normal residual volume
with normal detrusor contractions.
</think>
The cystometry results would likely show normal detrusor function and an
appropriate bladder capacity.
```

Figure 5.2

## 5.3 Hyperparameter Tuning

Enabling the best performance of LLaMA 3.2 on the mental health applications was carefully achieved through hyperparameter tuning during the fine-tuning. Adjusting several hyperparameters was fundamental for achieving efficient and effective convergence that facilitated providing high-quality, empathetic, and precise responses. During tuning of the specific process, the following long hyperparameters were tuned:

### 5.3.1 Learning Rate

The learning rate ranks among the most essential hyperparameters of fine-tuning, used to control the weight update of the model at each optimization step. In fine-tuning, there was this consideration of the learning rate to avoid frequent issues such as:

- **Overshooting:** The learning rate may be too high, in which case the model would not converge, and during the training session, it might not have a shot at finding the best and most optimal solutions.

- **Slow Convergence:** With the learning rate being too small, the learning process becomes longer, dramatically increasing the number of iterations the model needs to converge.

Learning rate was optimized using an adaptive approach, starting with a chosen starting rate and gradually reducing it, in schedules such as cosine annealing, throughout the training process. As a result of using this method, the training process itself became more stable, and the optimization was performed faster.

### 5.3.2 Batch Size

Batch size refers to the number of training samples processed before the model's weights are updated. Finding the right batch size is critical for:

- **Computational Efficiency:** A larger batch size allows for greater parallel tasks and more efficient computation, so resources such as GPUs can be used optimally.
- **Model Performance:** Selection of a high batch size will lead to overfitting, while selecting a too-small batch size may lead to unstable gradients and instability in convergence.

Experimentation showed that the best trade-off between an efficient use of hardware resources and optimal model performance on available GPUs was found for batch-size = 8. Using a batch size of 8, the model could effectively learn stable and generalizable features without straining the computational requirements excessively during training.

### 5.3.3 Epochs

Epochs specify how many cycles of the full dataset the model goes through during training. The epoch selection demonstrated the properties and size of the available dataset.

Training was aimed to allow the model to understand the intricacies of therapeutic conversations while retaining generalization, without overfitting.

Since the volume of data for mental health and therapy was quite modest, it was decided to perform 4–6 epochs of training. This strategy facilitated the model to do better on unseen data and avoided overfitting by not memorizing specific instances. Besides, early stopping was used to help stop training as soon as the model worked best on the validation data.

### 5.3.4 Gradient Clipping

Gradient clipping addresses the gradient explosion problem during the back propagation process, where the gradient starts to surge and causes abrupt weight changes. This issue can destabilize the training and make convergence of the model rather difficult.

Gradually, clipping was introduced to reduce gradient explosion, and a threshold value was configured to clip gradients during training. This approach maintained the stability of the optimization process due to the complex and large size of the LLaMA model. Gradient clipping was introduced when setting the maximum norm equal to 1, where the gradients were to be kept of a sane magnitude to avoid drastic changes in the weight updates of the model.

These selections and adjustments of hyperparameters were critical in enhancing the model training process, thus making the model learn and respond empathetically, appropriately, and culturally sensitive to mental health topics. Using these hyperparameters when fine-tuning is the foundation for the LLaMA model to perform positively in therapeutic environments.

## 5.4 Loss Functions and Optimization

The optimization process for fine-tuning the LLaMA 3.2 model incorporated two key loss functions, each tailored to different tasks, ensuring the model learned effectively from the therapeutic conversations. Optimization algorithms were employed to minimize these loss functions iteratively, helping the model converge to an optimal set of parameters to generate accurate and empathetic responses.

### 5.4.1 Cross-Entropy Loss

Cross-Entropy Loss was crucial in the classification of mental health issues highlighted during the conversations, particularly for such disorders as anxiety, depression, and stress. This metric measures the difference between the real mental health conditions and the probability that the model attributes to any classification.

Since this loss function incentivised the model to give out probabilities representing the actual labels, it made it suitable for classifying mental health problems. The Cross-Entropy Loss lets the model accurately identify mental health problems in users' conversations, making responses more pertinent.

### 5.4.2 Mean Squared Error (MSE)

Mean Squared Error was used in situations that aimed to forecast continuous results, such as therapy effectiveness or continuous indicators obtained from a sentiment analysis. MSE calculates the square of the difference between the predicted and actual values, making it particularly useful for regression-type tasks where precise numerical predictions are required.

In the MindEase project, MSE was applied to evaluate the model's ability to predict continuous outcomes, such as how effective the user perceived a therapy session. Lower MSE values indicated better predictions, aligning the model's outputs with real-world expectations of therapy effectiveness.

### 5.4.3 Optimization Algorithm: Adam Optimizer

Adam (Adaptive moment estimation) Optimization was used to update the model's weights during fine-tuning. Adam uses the components of AdaGrad and RMSProp, with learning rate modifications considering the momentum dynamics of the gradients. Therefore, in Adam's case, the performance in fine-tuning models like LLaMA that are critically dependent on stable training and convergence is superb.

Adam works by maintaining two moving averages:

- **First moment (mean):** The average of the gradients.
- **Second moment (variance):** The average of the squared gradients.

These momentums help Adam adjust the learning rate for each parameter individually, improving convergence by scaling the learning rates for each weight based on the first and second moments of the gradients.

The Adam optimizer was configured with default hyperparameters. Dynamic learning rate modifications with gradient feedback were applied, which, on one hand, optimized the training efficiency, and on the other, accelerated the convergence and optimization processes during the fine-tuning process.

Cross-Entropy Loss used in classification and MSE in regression, together with the Adam Optimizer, optimized the fine-tuning procedure that allowed the LLaMA 3.2 model to become more applicable to therapeutic settings and enhance its abilities for mental health condition classification and therapy effectiveness prediction.

### Incorporating Cultural Sensitivity and Empathy

Making certain that the model delivered both clinically and culturally acceptable answers turned out to be a major challenge.

- **Empathy Score Alignment:** During fine-tuning, empathy scores were applied to numeric responses made by the model, which encouraged caring and empathetic responses.
- **Cultural Adaptation:** Variations of language and multicultural references, as well as highly specialized phrases, were incorporated into the fine-tuning process, enabling the model to be more efficient in serving multicultural users.

## 5.5 Challenges Faced During Fine-Tuning

During the development of the model, a series of difficulties emerged that required strategic planning and specific modes of operation to ensure ethical and efficient results. Several significant obstacles were identified during the fine-tuning of our MindEase project with the LLaMA 3.2 model:

### 5.5.1 Imbalanced Dataset

The biggest problem was dealing with the disproportionate prevalence of some of the mental health conditions in the dataset. The fact that bias in the responses produced by the model was a risk resulted from the skew of the dataset, where anxiety and depression were more common.

To address this problem, the data augmentation and resampling methods were used:

- **Data Augmentation:** Transformation techniques, such as paraphrasing and artificial noise addition, were employed to synthesize synthetic data, enhancing the current presentation of under-represented mental health issues.
- **Class Weight Adjustment:** The model's learning benefit since we changed the loss function to value the underrepresented classes more, thus emphasizing equal consideration in the training algorithm.

With these efforts, the model was spurred to consider less prevalent mental health issues, leading to a complete therapeutic conversation approach.

### 5.5.2 Maintaining Context Across Long Conversations

Promoting contextual continuity in long therapeutic exchanges became a major barrier. Speaking about the issues connected to anxiety and depression usually requires one to cover a large number of dialogue segments, which requires some strong context to formulate suitable and empathic responses.

To address this problem, the model applied attention-based approaches:

- **Self-Attention:** It was improved to highlight important aspects present in the dialogue so that it could recall and reflect in important context from previous turns.
- **Transformer Architecture:** A transformer-based architecture, intrinsic to LLaMA, allowed the model to identify and use dependencies in long conversations, ensuring that context was not lost.

With these changes, the model could interact with the users in a way familiar to therapy, while maintaining the emotional mechanisms of the dialogue.

### 5.5.3 Ethical Constraints

Ethics was an essential part of the fine-tuning process. As the conversations discussed were sensitive, the training data collected was carefully curated, keeping in mind the ethical standards to be followed with absolute diligence, while establishing the privacy of patients.

The key ethical constraints included:

- **Data Anonymization:** To keep sensitive patient data, all personal details were anonymized throughout the fine-tuning procedure. Individuals were replaced with more general descriptions; personal information was stripped to protect data confidentiality.

- **Informed Consent:** All people whose data were incorporated into fine-tuning the model after providing informed consent, which means they were aware of how their data would be used during training.
- **Bias Minimization:** Strategies were used to control biases in the model, with special attention paid to possible cultural and regional backgrounds that may affect the results. If a dataset was carefully selected that addressed various mental health issues and contained the input of different demographic groups in Pakistan, the model became culturally appropriate.

Ethical concerns during the process guarantee the model can be reliable, protecting sensitive individuals from abuse that may expose them to the risk of privacy violation or prejudiced data uptake.

## 5.6 Fine-Tuned Model Results

Fine-tuning enabled the MindEase project to make critical advances, arming the model with more ability to provide personal, pertinent answers that concentrate on mental health challenges. With these improvements, it became easier for the model to provide therapeutic advice to users who are struggling with anxiety and depression.

Below are the notable outcomes:

### 5.6.1 Accurate and Personalized Responses

Fine-tuning the LLaMA 3.2 model allowed it to generate responses that were more accurate and personalized to the mental health concerns expressed by users. This was achieved through:

- **Domain-Specific Adaptation:** From therapeutically interacting, the model improved its ability to identify and manage several mental health problems, giving more relevant and situation-appropriate feedback for conditions of anxiety, depression, and stress.

- **Context-Aware Dialogue:** Using attention mechanisms, the model always captured the subtleties of the current discussion, enabling it to respond accordingly to the context of the conversation. The model's flexibility enabled it to change the responses appropriately about the flow of the dialogue, giving the users a natural and emotionally attractive interaction.

Thus, the model provided detailed and clinically appropriate responses that were carefully customized to address each user's specific preoccupations with exactitude.

### 5.6.2 Application of Therapeutic Techniques (CBT)

To boost its efficiency, the well-tuned model prioritised the uptake of the therapeutic approach, and CBT in particular was considered as the foundation of its design. Users' responses were analysed to identify themes, and the model provided intervention strategies that facilitate the principles of CBT. This included:

- **Cognitive Restructuring:** The model recognized standard negative thought processes of catastrophizing or all-or-nothing in anxiety and depression, and provided users with advice about how to observe and reconstitute such thoughts.

- **Behavioral Activation:** To facilitate users spiraling into depression, the model promoted useful behavior or self-care that mirrored the CBT principle of breaking the cycle of low activity and thoughts of hopelessness.

- **Problem-Solving:** Apart from this, the model allowed users to overcome tangible barriers as it was capable of offering actionable recommendations and helped create reasonable goals, which made them feel more empowered.

### 5.6.3 Enhanced Empathy and Compassion

The fine-tuning task improved the model's ability to express empathy and compassion to users. After focused work on empathy in the fine-tuning, the model showed increased sensitivity to emotional cues and could interact more thoughtfully and empathetically. This was achieved through:

- **Empathy Scoring and Optimization:** Emotionally meaningful interactions were guaranteed using empathy-based assessment, which validated the model's focus on accuracy and emotional support. By perfecting the model, we built tweaks to strengthen its focus on using compassionate language and empathetic responses.
- **Cultural Sensitivity:** To make the cultural background of Pakistan more acceptable, the model was tuned more to deliver culturally appropriate and resonating responses to the users. Camera use in response to regional and cultural differences made the model achieve a deeper emotional rapport among users.

### 5.6.4 Doctor Evaluation and Feedback

All mental health professionals tested the fine-tuned model appropriately before deploying it to actual settings. This way, we guaranteed that the model's dealings were reliably suggested to mental health practitioners and that they offered sound therapeutic directions.

- **Real-World Feedback:** Healthcare professionals had a look at the simulated interaction with the model while analysing the ability of the model to discuss sensitive issues like anxiety and depression using ethical and clinically appropriate replies.
- **Assessment of Therapeutic Effectiveness:** Doctors provided insight on how the model utilized therapeutic means, such as CBT, in measuring whether the responses matched industry-standard mental health practices.
- **Cultural Sensitivity Review:** Experts in Pakistan's cultural background, from professionals, were sought to ascertain whether the model addressed cultural nuances,

brought out local concerns, and effectively communicated with Pakistani users.

- **Iterative Improvements:** With the input given by the doctors in mind, the team also applied further refinements to the model. The fixes concentrated on turning the model slightly differently in some of its responses and ensuring it fulfilled mental health standards.

## 5.7 Evaluation Metrics

The success of the MindEase model was evaluated using a combination of traditional language model metrics and domain-specific measures. Such metrics were chosen to ensure that the model is technically valid and parsable to actual use cases of its users in the real world, particularly concerning empathy, clinical outcomes, and context appropriateness. The following evaluation metrics were used:

### 5.7.1 Perplexity

Perplexity measures the ability to guess subsequent words in a sequence. A smaller perplexity score means that the model's knowledge regarding language structure is also better and therefore, essential for giving professional and applicable guidance in counselling. Specifically:

- **Lower Perplexity:** The model's improvement in predicting the next word in a sequence facilitates better flow and relevancy of the conversation, which is essential for successful therapeutic communication.

- **Impact on MindEase:** Since the model focuses on mental health topics, a low perplexity is a reflection of its ability to cope and perform optimally when presented with complex social problems that arise in anxiety and depression.

### 5.7.2 Accuracy

Accuracy measured the model's ability to classify mental health concerns as they presented themselves in the context of therapy. The model's performance was evaluated through tasks such as symptom detection, recognition of the emotional state, and identification of appropriate therapeutic strategies to be implemented in the discussion.

- **Mental Health Issue Identification:** The ability of the model to identify mental health concerns, such as anxiety or depression, or stress, through language in conversations was tested, ensuring that it remains relevant and supportive in the return messages.
- **Effectiveness in Real-World Applications:** The accuracy testing exposed how well the model could detect significant concerns in therapy, ensuring that the model's feedback aligned with what patients and therapists typically discuss in real sessions.

### 5.7.3 F1 Score

The F1 score was used to balance precision and recall, especially in tasks where the model's responses were evaluated as contextually relevant and empathetic. Since generating responses with the right balance of accuracy and emotional intelligence is critical in mental health applications, the F1 score helped in:

- **Precision:** Ensuring the model returned topic- and clinically-focused answers, avoiding off-topic or unhelpful recommendations.
- **Recall:** Confirming that the model covered the user's concerns, without missing out on crucial symptoms or emotional signals.
- **Empathetic Responses:** The F1 score also measured the model's ability to produce empathic and context-related responses that are fundamental to building trust.

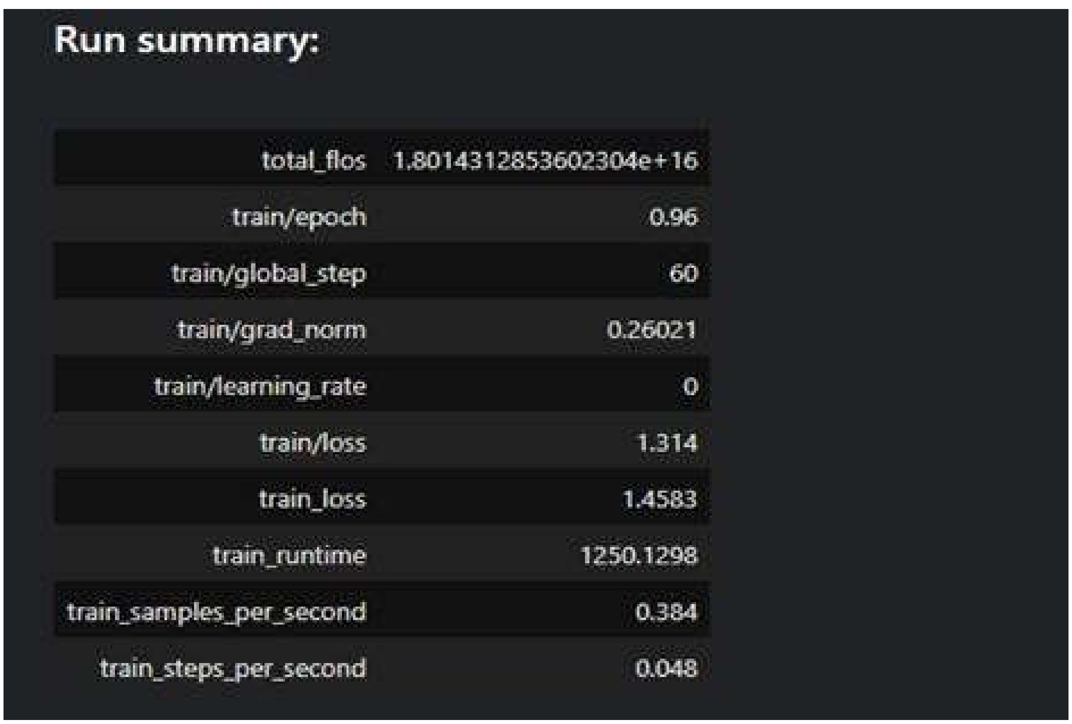
### 5.7.4 User Feedback

The real-world effectiveness of the model was largely evaluated by the feedback obtained from those who used it as a support for mental health. Patient and therapist feedback was sought through numerous surveys and examinations to measure ease of use and efficacy. This feedback focused on:

- **Relevance of Responses:** The model's capability to meet mental health concerns accurately and the sense of the users that the model presented rightful guidance.
- **Empathy and Compassion:** An indicator of how much the model can speak to empathy, comfort, and compassion in its responses, especially in sensitive situations.
- **User Satisfaction:** Average satisfaction of users regarding the model's support of mental health issues, specifically, culturally sensitive response, and therapeutic approaches.
- **Therapist Feedback:** The therapists reviewed whether the model could apply therapeutic techniques, including CBT, and was competent at handling challenging emotional or psychological situations.

Upon assessing the MindEase model through quantitative and qualitative evaluations, its performance and appropriateness in real-life mental health support situations were identified. These metrics were key to ensuring that the model could adequately provide culturally relevant, clinically valid, and empathetic mental health advice.


Figure 5.3



Run summary:

| | |
|---|---|
| total_flos | 1.8014312853602304e+16 |
| train/epoch | 0.96 |
| train/global_step | 60 |
| train/grad_norm | 0.26021 |
| train/learning_rate | 0 |
| train/loss | 1.314 |
| train_loss | 1.4583 |
| train_runtime | 1250.1298 |
| train_samples_per_second | 0.384 |
| train_steps_per_second | 0.048 |

Figure 5.4

## 5.8 Conclusion

Fine-tuning the LLaMA 3.2 model to be specifically tailored for MindEase was critical to its appropriateness for use in the support of people with anxiety and depression. The refinement included therapeutic practices, cultural sensitivity, and empathy to attune the responses of the model not only to be meaningful but also suitably contextualized for mental health discussions. After building the model on datasets from the real-life therapeutic dialogues, it was created to provide empathetic and clinically relevant answers. Fine-tuning hyperparameters such as learning rate, batch size, and epochs enormously improved its usefulness.

Moreover, important topics like working on imbalanced datasets, retaining context in long conversations, and applying ethical measures were covered using data augmentation, self-attention patterns, and reliable ethical plans. With such improvements, the model retained the capabilities to address various mental health needs and maintain its users' privacy and emotional care.

With these technological advances, it has been possible for the fine-tuned LLaMA 3.2 model to become a key aspect of mental health. The outcome of this project may also help develop other projects.

Schematic steps that run in the fine-tuning process are indicated in the **Figure** below.

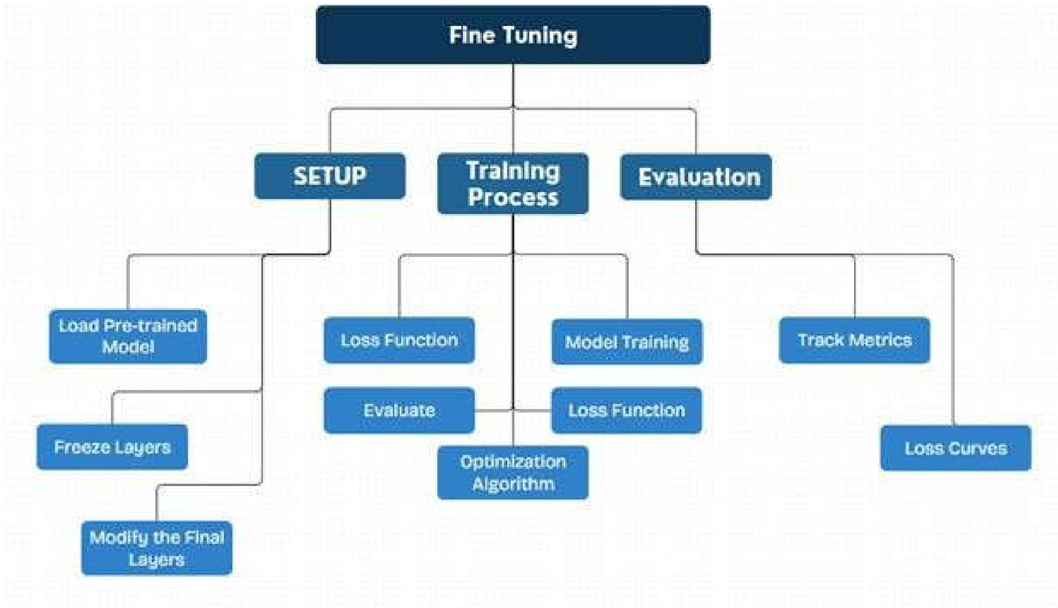

Fine Tuning
SETUP
Training Process
Evaluation
Load Pre-trained Model
Freeze Layers
Modify the Final Layers
Loss Function
Model Training
Evaluate
Loss Function
Optimization Algorithm
Track Metrics
Loss Curves

Figure 5.5

# Chapter 6

# Integration of Explainable AI

## 6.1 Learning Explainability in Perspective with Mental Health Chatbots

In healthcare settings, particularly mental health counseling, understanding why AI systems behave as they do is critical. It fosters trust, promotes transparency, and helps uncover bias. We integrated Explainable AI (XAI) into our fine-tuned LLaMA-3.2-3B chatbot, using Captum's Integrated Gradients (IG) to reveal insights into the model's decision-making process.

## 6.2 How Explainability Adds Value to Mental Health Chatbots

- **Clinical Trust:** Enables doctors and patients to understand AI decisions.
- **Bias Detection:** Identifies if demographics influence outputs.
- **Debugging:** Verifies reliance on key mental health terms like "anxiety" or "depression."

- **Compliance:** Aligns with ethical standards required in healthcare applications.

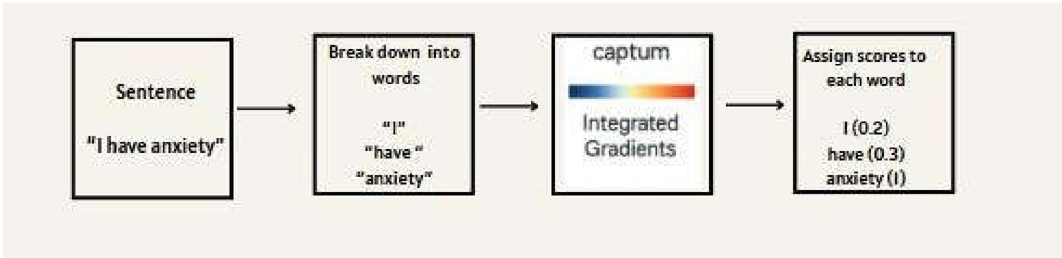



Figure 6.1

## 6.3 Tools and Libraries Used

**Captum (PyTorch's XAI Library)**

**Integrated Gradients (IG):**

- Distributes token-wise attributions.
- Ideal for transformer architectures.
- Offers quantitative scores for token importance.

**Custom Python Extensions**

- Token filtering (removal of special tokens).
- Score normalization (0–1 range).
- Keyword boosting (e.g., "depression" gets 2× weight).

## 6.4 Integrated Gradients in Our Chatbot Workflow

**Step-by-Step Workflow**

1. **Input Processing:**
    - Tokenize user input and chatbot response.

    - Extract embeddings using `model.get_input_embeddings()`.

2. **Gradient Calculation:**
    - Use Integrated Gradients to determine token influence.

3. **Score Aggregation:**
    - Normalize scores (0–1) using `normalize_scores()`.
    - Filter out special tokens.
    - Boost mental health keywords ($\times 2$), penalize common stopwords ($\times 0.25$).

4. **Visualization:**
    - Display heatmap or attribution scores for response tokens.

## 6.5 Key Functions and Their Roles

- `clean_token()`: Removes special tokens (e.g., Ġ, [PAD], [UNK]).
- `normalize_scores()`: Scale attribution scores for easier interpretation.
- `forward_func()`: Custom forward pass for Captum. Returns logits for scoring.
- `explain_response()`: Tokenizes input/output, applies IG, and returns a token-score heatmap.

## 6.6 Explainability in Action: Example

**User Query:**

*"My anxiety has escalated over the days, and it is very difficult for me to sleep."*

**Chatbot Response:**

*"Anxiety can disrupt sleep. Try mindfulness exercises before bed."*

**Token Attributions:**

| Token | Attribution Score |
|---|---|
| anxious | 0.92 |
| sleeping | 0.85 |
| feeling | 0.45 |
| trouble | 0.40 |
| mindfulness | 0.38 |

**Interpretation:**

The model focused correctly on “anxious” and “sleeping.” Lesser emphasis was placed on general words like “feeling.”

## 6.7 Applications in Mental Health Counseling

- **Debugging Model Biases:** For example, if “male” outperforms “female” in depression-related queries, retrain using balanced datasets.
- **Improving Prompt Engineering:** If critical symptom terms receive low scores, prompts can be refined.
- **Compliance Reporting:** Attribution logs support documentation for healthcare compliance.

## 6.8 Limitations

**Current Limitations**

- **Computational Overhead:** Increases inference time by $\sim$15%.
- **Granularity:** Currently focused on token-level attribution. Span-level or phrase-level analysis is a potential area for improvement.

# Chapter 7

# Web Development – Mind Ease

The Mind Ease chatbot web application is built as a full-stack Next.js project. The frontend uses the Next.js React framework (implemented in TypeScript) with a component-based architecture and file-system routing. Next.js was chosen for its robust support of server-side rendering (SSR), static site generation (SSG), and API routes, enabling both dynamic and pre-rendered pages as needed. This ensures fast initial load and SEO benefits in production, allowing React's virtual DOM for interactivity.

Development is organized into reusable components and pages (e.g., under `pages/` or `app/`), so each UI element (dialog boxes, buttons, headers) is modular and testable. The application is written in TypeScript for added type safety and maintainability: as one industry guide notes, combining Next.js with TypeScript "helps [produce] error-free code" and "type safety" that improves scalability and team productivity.

## 7.1 Frontend Architecture

7.1 Frontend Architecture A component-driven structure would be used for the web application based on React and TypeScript, and support modular design and sophisticated type checking. Tailwind CSS, a utility-first CSS framework, was selected for styling the interface to easily develop an up-to-date, intuitive, and relaxing design for mental health users.

- Next.js Framework (React, SSR/SSG, API Routes): Next.js is a production-grade React framework that "seamlessly integrates static site generation (SSG), server-side rendering (SSR), and incremental static regeneration (ISR)". It provides built-in optimizations (e.g., image/font optimization) and 1 supports advanced file-based routing, allowing pages to be defined by files under the app/directory. Any file under 4 pages/ (or pages/api is automatically treated as a serverless API endpoint, so backend logic (chat endpoints, data fetch) can be written alongside the frontend code without an extra server framework. Using TypeScript with Next.js leverages IDE support and static typing, ensuring robust, maintainable code. 3

- Tailwind CSS (Utility-First Styling): UI styling is implemented with Tailwind CSS, a utility-first CSS framework. Tailwind allows defining a consistent design system through low-level classes (e.g. blue-100, px-4, 5 6 5 bg rounded ) without writing custom CSS, which speeds development and ensures visual consistency. Its small, reusable utility classes make layout and spacing predictable; as Tailwind's docs explain, choosing from a "predefined design system. . . makes it much easier to build visually consistent UIs.". This translates into a clean, calm interface for a mental health application: designers can rapidly prototype color palettes (e.g., soft blues and whites) and typography while preserving accessibility and responsiveness. (The Next.js guide even confirms that Tailwind is "fully compatible with Next.js" and easy to integrate .)

- Component-Based Design (TypeScript): The interface is broken into reusable React components (e.g, ChatWindow, MessageBubble, NavBar ) and composable layout elements. Each component encapsulates its logic, state (if any), and styling. This modularity, combined with TypeScript's static types, ensures that data flows and props are well-defined, reducing runtime errors. One report notes that Next.js +

TypeScript offers “type safety for error-free code, scalability for enterprise-level applications, [and] improved productivity”. Next.js + TypeScript + Tailwind forms a modern stack that efficiently supports the chatbot’s rich, responsive front end.

The main user interface includes:

- A chat window for real-time AI conversation
- Authentication pages (login and signup)
- Navigation and layout components using the App Router in Next.js

The chat interface supports asynchronous message exchange with the AI model through API routes, delivering an experience mimicking a counselor-client dialogue.

## 7.2 Pages

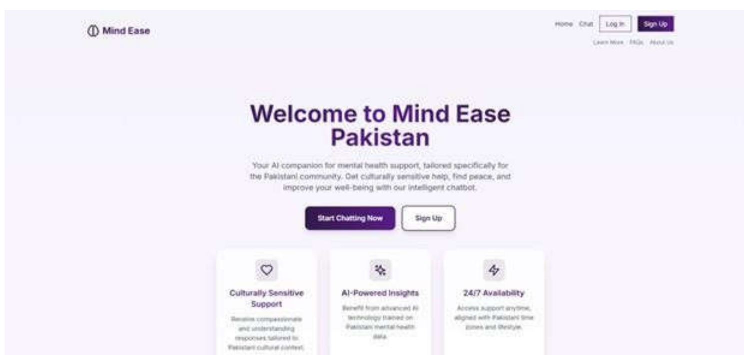



Figure 7.1: Home Page

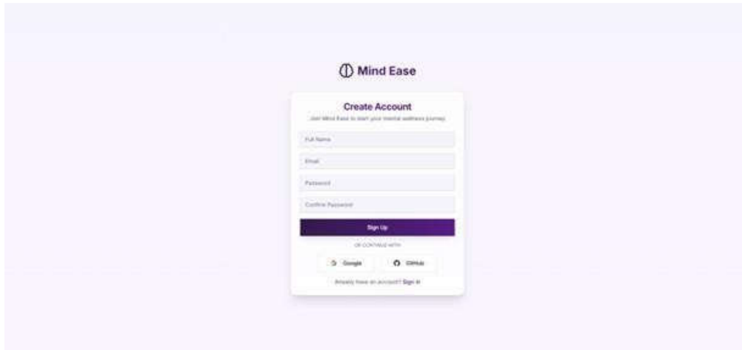



Figure 7.2: Sign Up Page

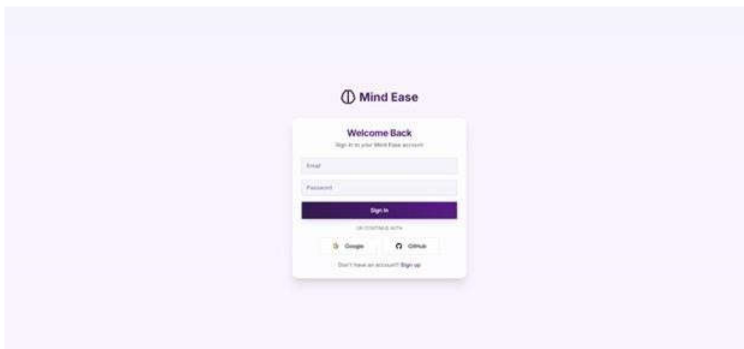



Figure 7.3: Sign In Page

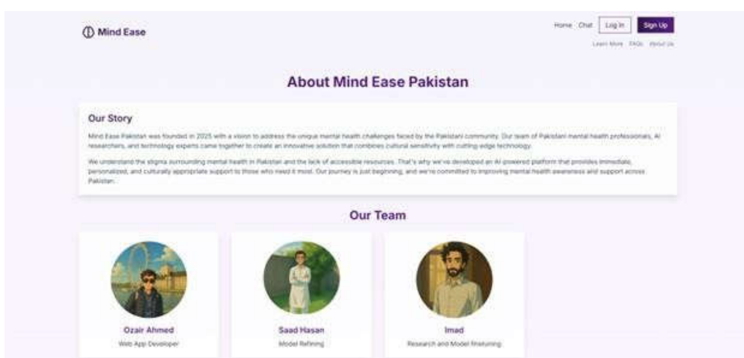



Figure 7.4: About Us Page

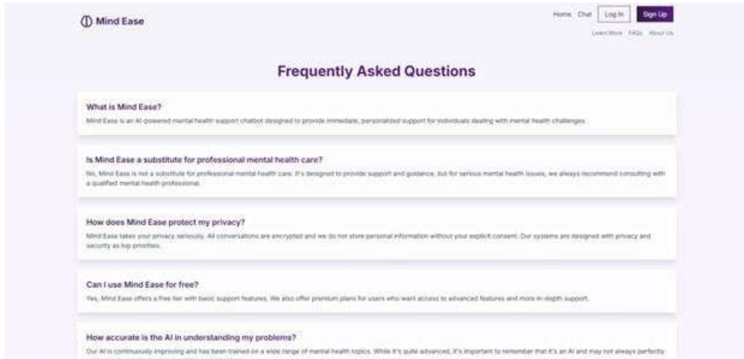



Figure 7.5: FAQs Page

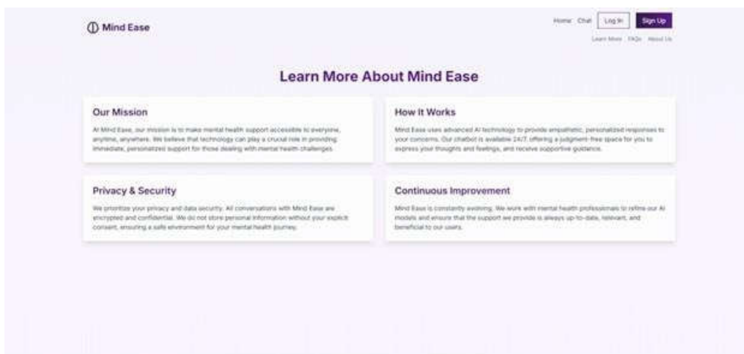



Figure 7.6: Learn More Page

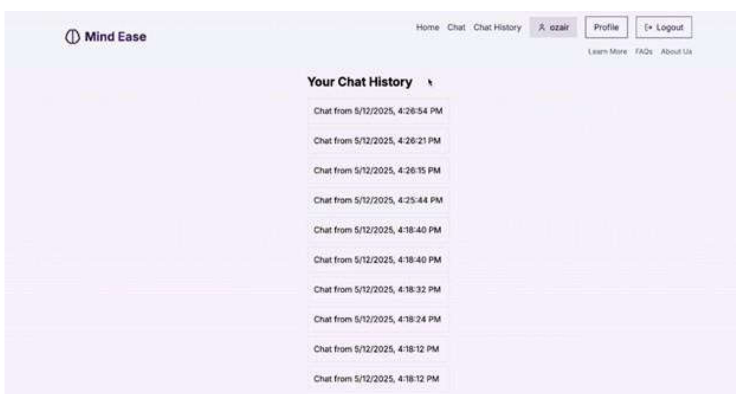



Figure 7.7: Chat History Page

## 7.3 Authentication and Session Management

User authentication is handled by NextAuth.js, a full-stack open-source solution built for Next.js. NextAuth simplifies the integration of multiple sign-in methods and secure session handling. In our system, we configure NextAuth with two providers: GitHub (OAuth) and custom email/password credentials. This allows users to log in via their GitHub account or register with an email and password. In each case, NextAuth uses secure tokens or sessions under the hood.

- OAuth (GitHub) with NextAuth: NextAuth supports all popular OAuth providers out of the box. 7 We use the built-in GitHub provider ( GithubProvider ) so users can authenticate with their GitHub account. NextAuth automatically manages the OAuth flow (redirecting to GitHub's consent page, receiving the callback, etc.). The NextAuth documentation states that it has "built-in support for popular services" and is "designed to work with any OAuth service. (In practice, this means 7 simply specifying the clientId and 8 clientSecret from a GitHub OAuth App in the NextAuth configuration.) This provider issues a session token for the user upon successful login.

- Email/Password (Credentials) with NextAuth: In addition to OAuth, we enable a Credentials provider in NextAuth for traditional email/password login. This requires implementing a custom authorize function: when users submit their credentials, we look up the user in our database (MongoDB) and verify the password. User passwords are never stored in plain text; we use bcrypt.js to hash and compare passwords. Bcrypt applies the Blowfish-based bcrypt algorithm (deliberately slow and computationally expensive) to the plaintext password when creating an account or logging in [28]. By salting and hashing with bcrypt, attackers cannot easily recover the original passwords even if the database is compromised. (As one security guide notes, "bcrypt is a hashing function – a slow and computationally expensive algorithm but resistant to brute-force attacks" .) Thus, NextAuth's Credentials provider plus bcrypt ensures secure email/password flow.

- Session Management with JWT: NextAuth is configured to use JSON Web Tokens (JWT) for session state. NextAuth issues a signed JWT in this mode, which is stored in an HttpOnly cookie after login. Subsequent requests (to protected pages or APIs) carry this token for authentication. Using JWTs makes the session stateless on the server side, as the token encodes the session data (or can be verified against a secret). NextAuth natively supports JWT sessions. This choice avoids storing server-side sessions and fits with Next.js's serverless environment. Furthermore, JWTs can be configured to expire or be refreshed, adding security. NextAuth's documentation highlights that it can "choose database sessions or JWT" and handles signing/verification internally, ensuring cookies are HttpOnly, CSRF-protected, and use JWS/JWE signing for security.

Together, these components form a robust auth system: NextAuth handles routing (via its /api/auth/ [...nextauth] route), token issuance, and security checks, while bcrypt and TypeScript code ensure that custom logic (like credential lookup and hashing) is correct and safe. This layered approach provides users with both convenience (OAuth sign-on) and fallback (email login) in a secure way.

## 7.4 API Integration and AI Communication

A Hugging Face powers the core functionality of the chatbot–based language model running on a local server. The Next.js frontend communicates with this model via RESTful API calls. Specifically, a Next.js API route (e.g., at /pages/api/chat.ts ) acts as a proxy: when the user sends a message in the UI, the client code sends it to this internal endpoint. The API route handler then forwards the query to the locally hosted Hugging Face model (for example, using an HTTP request to http://localhost:PORT/ predict ). The model processes the input and returns a response, which the API route relays back to the frontend.

Because the Hugging Face model runs locally (for privacy and latency reasons), we use ngrok to securely expose it to the Next.js server during development or testing. Ngrok creates an encrypted tunnel and gives a public HTTPS URL that maps to the local service. In practice, we start the model server and point ngrok to it (e.g., ngrok http 8000 ), then configure the Next.js API route to call the ngrok URL. This allows the cloud-deployed frontend (CI tests) to reach the local AI model without exposing it to the open internet. Ngrok's "universal ingress platform" abstracts away firewall and network complexities and "makes [the LLM] accessible from anywhere" over HTTPS [23]. In short, ngrok provides a secure tunnel so our code can integrate with the private AI service as a public API.

Within Next.js, we handle API communication using built-in fetch or HTTP client calls from the server side. To ensure data consistency, we can define typed request/response interfaces in TypeScript. For example, the API route might accept a JSON body message: string and return reply: string . The API handler calls the Hugging Face endpoint, waits for the inference result, and formats the output. Error handling and timeouts are included to ensure reliability. By using Next.js API routes, all of this logic runs securely on the server side and does not bloat the client bundle. The result is that the React 4 frontend can simply call fetch('/api/chat') without needing direct knowledge of the model's details, keeping concerns separated.

## 7.5 Database Connectivity

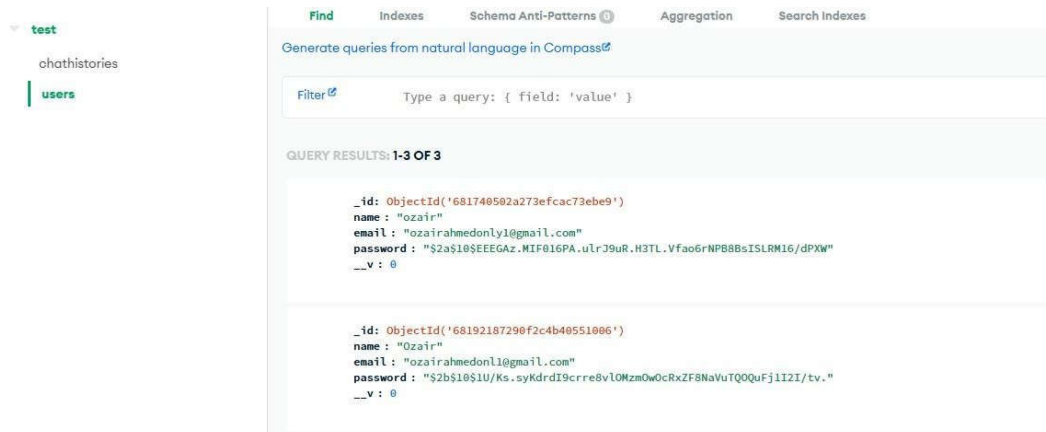



Figure 7.8: User Data

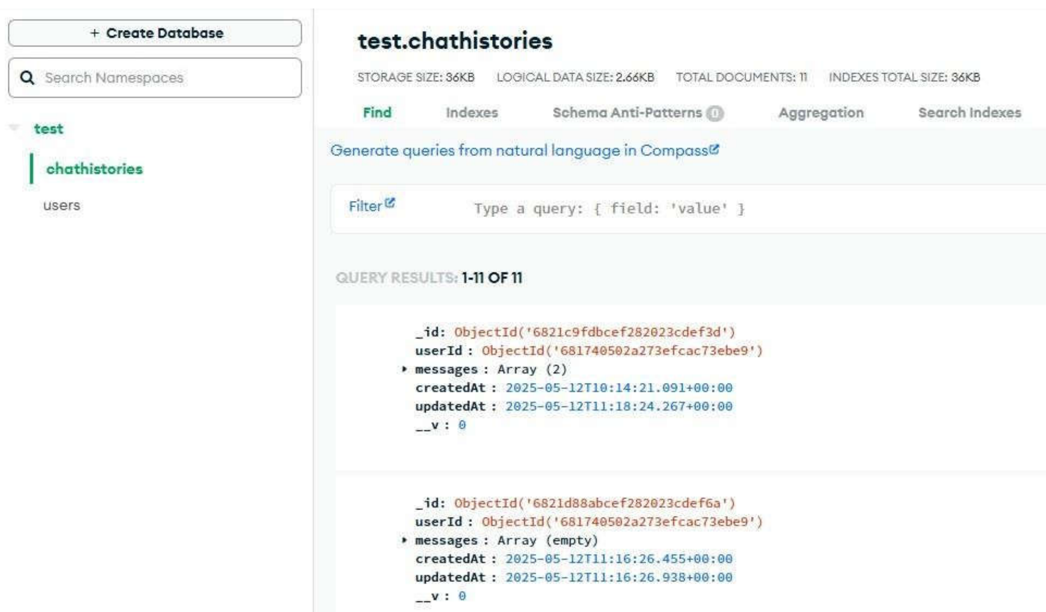



Figure 7.9: Chat History

All persistent data (user accounts, chat history, etc.) is stored in a MongoDB database and accessed via the Mongoose ODM. MongoDB was chosen for its scalability and flexible document model. The MongoDB documentation explains that it is a document database with a "flexible schema," making it popular for agile development of internet applications [29]. Documents are stored in JSON-like BSON format, which maps naturally to JavaScript objects, so the team can easily save and retrieve structured or evolving data without rigid schema constraints. This flexibility is helpful for a chatbot, since conversation logs or user profiles may have optional fields or new data types added over time. MongoDB's scale-out architecture allows the app to handle growing user counts gracefully.

We use Mongoose (a schema-based ODM for Node.js) to define and interact with our MongoDB data. In Mongoose, everything starts with a Schema, which defines the shape and validation of documents in a collection. For example, we might define a User schema with fields like passwordHash, name, an array of emails (unique), and Message subdocuments. Another schema could store chat messages or session history. Mongoose models compile these schemas into constructors for creating and querying documents. Using Mongoose provides a clear data model and additional safety: it automatically enforces types, default values, and required fields, catching errors early.

The official docs note that "each 3. save (), schema maps to a MongoDB collection and defines the shape of the documents within that collection". This layer means our application code can work with plain JavaScript objects (instances of models) and call 1.8find d(), etc., without writing raw database code. It also supports middleware hooks (e.g., hashing passwords before save) and integrations (such as syncing with NextAuth or session data).

In summary, the database layer uses MongoDB for its developer-friendly document model and horizontal scaling, while Mongoose provides schema structure, validation, and an API to Node.js. Together, they enable the reliable storage of user credentials, OAuth tokens, chat transcripts, and other data.

## 7.6 Deployment

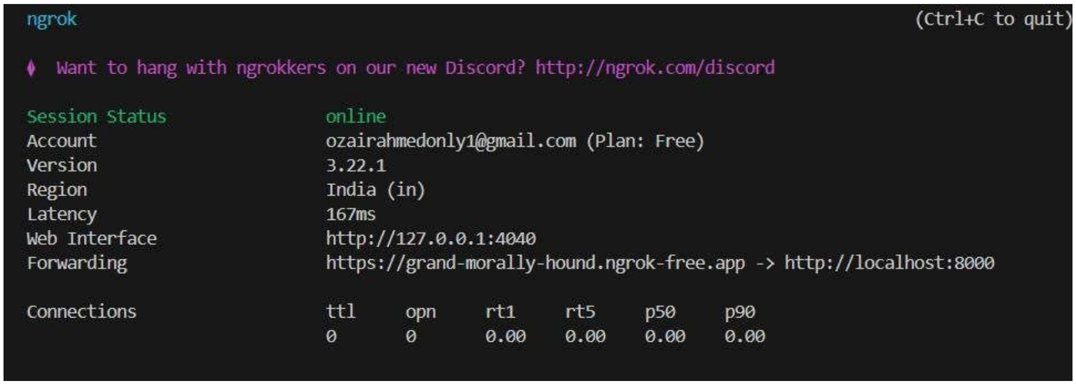



Figure 7.10: ngrok tunnel

The application is deployed on Vercel, the cloud platform designed for Next.js. Vercel provides seamless hosting of Next.js apps with automatic build and deploy pipelines. We connect the GitHub repository to Vercel’s Git Integration so that every push to the repository triggers a new deployment. Vercel then runs the build ( next build ) and assigns a unique preview URL. As its docs state, “Vercel for GitHub automatically deploys your GitHub projects. . . providing Preview Deployment URLs”. Every branch or pull request can be previewed live, and merging to the 19 main automatically updates the production deployment. If multiple commits are pushed quickly, Vercel queues and runs them so that only the latest deployment goes live, reducing build waste.

Using Vercel has several advantages: it is built by the creators of Next.js and is optimized for it. Vercel’s infrastructure automatically handles server-side rendering and API functions as serverless functions, scaling them globally. It also provides zero-downtime deployments and easy rollback to previous commits. In practice, we set up Vercel to use environment variables (for secrets like GitHub client IDs, MongoDB connection URI, etc.) and point it to our GitHub repo. Vercel builds the project upon each merge, runs tests, and deploys the result. This CI/CD integration ensures that new code (a model update or UI tweak) goes live automatically.

In short, Vercel serves as the app's production host. It provides a global Content Delivery Network (CDN) for assets, automatic HTTPS, and continuous deployment from GitHub. Its tight integration with Next.js ("the native Next.js platform, made by the creators of Next.js" ) makes it well-suited for our mental health chatbot: updates can be rolled out rapidly, and performance optimizations (like image and script caching) are handled with minimal configuration.

The Finetuned Model is running on a local machine, and ngrok [30] is used to tunnel the localhost link, enabling our app to connect to the model after being deployed.

# Chapter 8

# Conclusion and Future Work

The Mind Ease project set out to develop a culturally tailored mental health chatbot for Pakistani users. The system integrates a large language model from Hugging Face into a secure, responsive web application built with Next.js. Authentication is managed via GitHub and email login using JSON Web Tokens (JWT) and a MongoDB database for session persistence. The application is deployed on Vercel and utilizes ngrok for secure API tunneling during development. Together, these components allow Mind Ease to provide reliable, always-available support for users seeking mental health assistance.

Mind Ease's key contributions lie in its cultural adaptation and accessible design. Explanations in the form of a chatbot have been made in a way that will be favorable to Pakistan's cultural nation. Employing culturally and religiously important terms in its activity, the system attempts to enhance a connection with users by nurturing confidence. In addition, the technological setup is secure and user-friendly. For instance, MongoDB is a safe storage option for users' sessions and data, ensuring the sessions are confidential and private.

Overall, the project demonstrates that an AI-driven chatbot can be effectively adapted to serve stigma-affected communities. Mind Ease exemplifies how combining open-source language models with modern web frameworks can create a private, empathetic counseling aid that complements limited mental health services. It provides continuous support without requiring face-to-face contact, especially when seeking help carries social stigma. In summary, this work highlights the future potential of culturally aware AI solutions in expanding mental health access, particularly for underserved communities.

## 8.1 Future Work

Future work for Mind Ease will focus on expanding accessibility, personalization, and responsiveness to user needs. Key directions include:

### Native Mobile Applications

Customized mobile applications for Android and iOS will increase reach and make the interface more user-friendly. Native apps can utilize features such as offline storage, push notifications, and device sensors. This will allow Mind Ease to remain accessible even in areas with intermittent internet access.

### Voice Chat Interface

Adding voice-based interaction will support users with low literacy or visual impairments. Speech recognition and text-to-speech will let users speak their concerns and hear responses, making interactions feel more natural and reducing barriers to access.

### Mood Tracking and Journaling

Mood diaries and self-assessment tools can help users reflect on emotional states and trends over time. The chatbot can offer personalized coping strategies based on this data, and private journaling may provide therapeutic benefits.

### Multilingual Support

Extending support to Urdu, Punjabi, and other regional languages will help users express complex emotions more comfortably. Native-language interfaces will increase trust and relatability. This will require fine-tuning the model on additional linguistic datasets.

### Crisis Escalation and Professional Referral

Protocols for recognizing crisis language should be integrated. When serious mental health risks are detected, the system should connect users with local therapists, emergency hotlines, or professional services to ensure timely intervention.

### Ongoing Personalization and AI Refinement

Personalizing chatbot behavior based on preferences, history, and emotional profile will increase user engagement. The model can be continuously improved through fine-tuning on anonymized, local interaction data to increase relevance and empathy.

Each of these future directions builds on the foundation established by the current system. Together, they chart a path toward a more versatile and impactful Mind Ease platform. By embracing mobile platforms, richer interaction modes, and deeper personalization, Mind Ease can reach more users and better support their mental well-being.

In conclusion, the Mind Ease chatbot has demonstrated the potential of combining artificial intelligence with culturally aware design to support mental health. Continuing to develop these enhancements will further the goal of providing accessible, empathetic care to communities that most need it. The work presented in this thesis lays the groundwork for ongoing innovation in mental health technology, and future efforts will extend its impact into the lives of those seeking help.

# References


[1] Frontiers, “Closing the mental health gap: transforming pakistan’s mental health landscape,” *Frontiers in Health Services*, 2024.

[2] “The mental health of china and pakistan, mental health laws and covid-19 mental health policies: a comparative review,” *PubMed Central*, 2022.

[3] “Mental healthcare in pakistan,” *Taiwanese Journal of Psychiatry*, 2020.

[4] “Stigma toward mental and physical illness: attitudes of healthcare professionals, healthcare students and the general public in pakistan,” *PubMed Central*, 2020.

[5] A. Khalid, F. Qadir, S. W. Y. Chan, and M. Schwannauer, “Adolescents’ mental health and well-being in developing countries: a cross-sectional survey from pakistan,” *Journal of Mental Health*, vol. 28, no. 4, pp. 389–396, 2018.

[6] “Digital 2025: Pakistan — datareportal – global digital insights,” 2025.

[7] “An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews,” *PubMed Central*, 2023.

[8] “Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge,” *arXiv preprint arXiv:2303.14070*, 2023.

[9] W. H. Organization, “Covid-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide,” 2022.

[10] A. A. Abd-alrazaq *et al.*, "An overview of the features of chatbots in mental health: A scoping review," *International Journal of Medical Informatics*, vol. 132, p. 103978, 2019.

[11] K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cbt to young adults using a fully automated conversational agent (woebot)," *JMIR Mental Health*, vol. 4, no. 2, p. e19, 2017.

[12] B. Inkster, S. Sarda, and V. Subramanian, "An empathy-driven, conversational artificial intelligence agent (wysa)," *JMIR mHealth and uHealth*, vol. 6, no. 11, p. e12106, 2018.

[13] "Acceptability and effectiveness of artificial intelligence therapy for anxiety and depression (youper): Longitudinal observational study," *PubMed Central*, 2021.

[14] "Digital mental health interventions for treating mental disorders in young people based in low- and middle-income countries: A systematic review of the literature," *Cambridge Prisms: Global Mental Health*, 2021.

[15] S. Ahmad and S. Koncsol, "Cultural factors influencing mental health stigma in pakistan," *Religions*, vol. 13, no. 5, p. 401, 2022.

[16] E. Kerz, S. Zanwar, Y. Qiao, *et al.*, "Toward explainable ai (xai) for mental health detection based on language behavior," *Frontiers in Psychiatry*, vol. 14, 2023.

[17] "Jwt security best practices | curity," 2024.

[18] "To chat or bot to chat: Ethical issues with using chatbots in mental health," *PubMed Central*, 2023.

[19] "meta-llama/llama-3.2-3b," 2024.

[20] "Developing digital mental health tools with culturally diverse parents and young people: Qualitative user-centered design study," *PubMed Central*, 2024.

[21] "Mdd-5k: A new diagnostic conversation dataset for mental disorders synthesized via neuro symbolic llm agents," *arXiv preprint*, 2024.

[22] "Fine-tuning," 2024.

[23] "Prompt engineering for digital mental health: a short review," *Frontiers in Digital Health*, 2024.

[24] "Building a next.js application," 2024.

[25] "A review of large language model-based mental health chatbots," *ACM Digital Library*, 2024.

[26] "How can i evaluate a fine tuned llm?," 2024.

[27] "Bertscore: Evaluating text generation with bert," *arXiv preprint arXiv:1904.09675*, 2019.

[28] "Why password encryption matters," 2024.

[29] "Why use mongodb and when to use it?," 2024.

[30] "Fast and secure ingress to remote ais with ngrok," 2024.