-

COMPARATIVE ANALYSIS OF CONTENT: HUMAN EXPERT VS. ARTIFICIAL INTELLIGENCE IN AN EXPLORATORY STUDY

By Lakshmi Panguluri

A Thesis

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Science

in Computer Science

Northern Arizona University

December 2023

Approved:

Ashish Amresh, Ph.D., Chair

Morgan Vigil-Hayes, Ph.D., Co-chair

Jared Duval, Ph.D.

ABSTRACT

COMPARATIVE ANALYSIS OF CONTENT: HUMAN EXPERT VS. ARTIFICIAL INTELLIGENCE IN AN EXPLORATORY STUDY

LAKSHMI PANGULURI

The field of Artificial Intelligence (AI) is currently experiencing a significant transformation characterized by notable advancements in education, healthcare, commerce, banking, agriculture, and transportation. One of the most crucial areas is Generative AI, which represents a fundamental shift in the field. The evolution of large language models within Generative AI is particularly noteworthy, as they demonstrate an impressive ability to generate text that closely resembles human-like language in terms of coherence and relevance. Chat Generative Pre-trained Transformer(ChatGPT), a prominent example of a generative model, has received widespread recognition for its adeptness in understanding and generating text, showcasing the advancements achieved in natural language processing. There has been limited investigation into the ability of ChatGPT to generate text that could be used in places requiring significant human effort. This research aims to initiate an exploratory comparative analysis of the content created by human experts and ChatGPT, employing a set of expert-provided keywords tailored for scripting a meditation activity. As a case study, we generate customized content for the meditation activity within the SUNRISE mobile health application. To ensure a meaningful comparative analysis of the content of a meditation activity, it is crucial to conduct a usability study of the application. This study evaluates the application's effectiveness and user-friendliness, allowing for a comprehensive understanding of its performance. A usability study is employed, utilizing the uMARS (Mobile Application Rating Scale) protocol specifically designed to assess the quality of mobile health applications. This evaluation

aims to ensure that the application is user-friendly and meets the necessary standards for effective usability. This research was structured as an empirical exploratory study aimed at understanding user perceptions and experiences. Specifically, the study delved into how individuals with varying levels of familiarity with meditation activities perceived content created by ChatGPT. By strategically employing expert-provided keywords, the research aimed to provide insights into users' nuanced responses and preferences when presented with content generated by human experts and AI, thereby shedding light on the evolving landscape of text generation by ChatGPT. According to the uMARS findings, the app demonstrates usability with a 3.93 out of 5 rating. Initial results suggest that proficient users(N=6) of meditation rated both human expert-generated and AI-generated content equally. In contrast, novice users(N=6) exhibited a preference for AI-generated content over human expert-generated content. It is critical to note that this exploratory study serves as an initial initiative, and further research is required to delve deeper into the utilization of AI-generated content in healthcare contexts.

COMPARATIVE ANALYSIS OF CONTENT: HUMAN EXPERT VS. ARTIFICIAL INTELLIGENCE IN AN EXPLORATORY STUDY

by

Lakshmi Panguluri

Acknowledgements

This work was funded by NSF 2224014.

Contents

List of Tables

List of Figures

Acronyms

AI Artificial Intelligence

GAI Generative Artificial Intelligence

GPT Generative Pre-Trained Transformer

MHA Mobile Health Application

NLP Natural Language Processing

Chapter 1

Introduction

The Artificial Intelligence (AI) landscape is constantly evolving, and significant strides have been made in the realm of Generative Artificial Intelligence (GAI)[2.1]. The development of advanced language models within this trajectory, exemplified by OpenAI's ChatGPT[1.1.3], demonstrates its notable ability to generate text that exhibits human-like coherence. This study provides an exploratory comparative analysis of content created by human experts and ChatGPT. By examining user perceptions and experiences, the research seeks to understand how individuals engage with text generated by ChatGPT in the context of meditation activities.

1.1 Background

1.1.1 SUNRISE Application

Mental health is important in every stage of life and mental health issues are more prevalent among Indigenous communities[2.5] due to the sustained impacts of colonization [1]. According to the National Institute of Mental Health, 22.8% US adults suffer from mental health illnesses and 26.6% of these individuals are from the American In-

dian/Alaskan Natives community [2]. Developing mental health interventions for Native American tribal communities necessitates a culturally sensitive approach that acknowledges their deep-rooted traditions [3, 4]. At the time of our experiments, SUNRISE was in its preliminary stage of development, with a particular emphasis on enhancing mental health and well-being in a Native American tribal community. The SUNRISE team emphasizes a comprehensive strategy that integrates mHealth technology, educational gaming, and curricular scaffolding to promote mental wellness among adolescents. Grounded in community-based participatory research, the SUNRISE application incorporates mindfulness activities. The task of developing the mobile health application, which includes the author, falls under the purview of the development team. Comprising four members, the development team actively collaborates with SUNRISE application researchers, receiving crucial inputs to inform the design and development process.

The application operates on a team-based approach that comprises two distinct user roles: mentee and mentor. A mentee is an individual who seeks to improve their mental well-being, while a mentor is responsible for overseeing and supervising the actions of mentees. The app strongly promises anonymity for all users by allocating alias names, thereby promoting privacy and safety. The app offers a range of mindfulness activities, such as breathing exercises, running, walking, meditation, and art, all of which are geared towards improving mental wellness. While some of these activities are designed for individual use, others are intended for team-based engagement. Apart from the aforementioned activities, the app encompasses features such as messaging, a gallery, rewards, and emergency aid to augment user engagement and support. As a part of this research, the author focuses on the meditation activity in the application.

1.1.2 Artificial Intelligence

Artificial Intelligence(AI) emerged as a transformative technology in 1956 [5], wielding considerable influence across various economic sectors. AI is defined as "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation [6]." The recent advancements and expansion in machine learning have paved the way for more sophisticated technologies in digital content generation, such as Generative Artificial Intelligence (GAI). Generative AI refers to the utilization of Artificial Intelligence tools or technologies that are designed to produce various forms of content, such as text, images, audio, or video. The primary aim of generative AI is to generate unique content that exhibits a high degree of creativity by leveraging the knowledge and patterns extracted from the data on which the AI system has been trained [7] This framework operates on unsupervised or partially supervised machine learning principles, employing statistics and probabilities to generate artificial artifacts. The potential of Generative AI is vast and can be applied to various sectors such as commerce, academia, medical care, and content creation [8]. Generative AI tools are constructed upon foundational AI models, notably Large Language Models (LLMs), which constitute the text-generating component of this technology [9]. Large Language Models (LLMs) are a classification of Artificial Intelligence (AI) that have undergone extensive training using extensive public text datasets obtained from articles, books, and online content. These models have the ability to identify patterns within available data, enabling them to offer information or generate responses based on a wide range of inputs, leveraging their comprehensive training [10].

1.1.3 ChatGPT

A large language model named ChatGPT (Chat Generative Pre-Trained Transformer) launched by OpenAI has gained much attention by responding promptly to human requests provided in natural language [11]. It originates from the original GPT (Generative Pre-trained Transformer) model, which has undergone iterative updates from GPT-1 to GPT-4 [12]. In less than three months of launching ChatGPT in January 2023, it has gained more than 100 million monthly active users, according to the report by UBS [13, 14]. The ChatGPT system combines diverse technologies: deep learning, unsupervised learning, fine-tuning through instructions, multitasking, contextual learning, and reinforcement learning.

GPT-1, which was developed in 2018, primarily focuses on training a generative language model using unsupervised learning within the Transformer framework [15]. In 2019, GPT-2 emerged with the introduction of multi-task learning, incorporating more network parameters and data than its predecessor. This enhancement allows the pre-trained generative language model to generalize across supervised subtasks without the need for additional fine-tuning [16]. GPT-3 takes a step further by integrating meta-learning and in-context learning, significantly improving its generalization abilities. It outperforms many existing methods across various downstream tasks [17]. The most recent version, GPT-4, builds upon this progress by increasing the number of network parameters, enabling predictable scaling, and gaining the capability to process both image and text inputs while generating text outputs [18]. ChatGPT derives its foundation from the GPT-3.5 architecture, a modification of the GPT-3 model unveiled by OpenAI in 2020. Essentially, GPT-3.5 serves as a scaled-down iteration of GPT-3, characterized by a reduced parameter count of 6.7 billion as opposed to GPT-3's 175 billion parameters [19, 20, 21]. The vast amount of data has empowered ChatGPT to gain a deep

understanding of patterns and connections between words and phrases in human language on an unprecedented level. This characteristic plays a crucial role in enhancing its ability to produce responses that are not only logical but also relevant to the context of user queries.

The proficiency of ChatGPT in mental health scenarios is examined due to the emphasis on mental health in the SUNRISE application. ChatGPT has undergone testing as a virtual counselor and companion for mental health, providing a spectrum of supportive services to users dealing with different mental health difficulties [12]. ChatGPT offers resources on coping strategies, details about mental health conditions, and self-help techniques [22]. ChatGPT assists users in managing anxiety and stress by providing mindfulness exercises, relaxation techniques, positive affirmations, and self-reflection exercises [23]. ChatGPT's ability to generate language allows it to provide therapeutic interventions based on proven techniques. It helps users explore their thoughts and behaviors, promoting self-awareness and gaining insights into their mental well-being [24].

1.1.4 Content Generation in Meditation Activity

The focus of our experimentation is on the meditation activity offered by the SUNRISE app, which provides content to enhance the experience of practicing meditation. The practice of meditation has become increasingly popular as a method to improve the mental well-being of adolescents [25]. This technique helps individuals develop self-compassion, mindfulness, and meta-cognitive abilities [26]. Moreover, meditation has been proven effective in reducing anxiety and stress, promoting relaxation, and improving mood [27]. By incorporating meditation into health interventions for adolescents, it becomes evident that it has the potential to enhance mental health in various ways, such as increasing self-awareness, regulating emotions, and building overall psychological

resilience.

Initially, the content of the meditation activity in the SUNRISE application was carefully crafted by a human expert using the Quiet Ego model [28, 29]. Our research focuses on the integration of ChatGPT content into the SUNRISE application as an experimental version to generate a tailored meditation script. We assess participants' perceptions and user experiences regarding the ChatGPT-generated script, comparing it to the meditation script created by a human expert.

1.2 Problem Statement

The utilization of Artificial Intelligence (AI) has made remarkable progress, as evidenced by the widespread use of ChatGPT. Generative AI like ChatGPT could be particularly useful for generating content for mobile mental health applications, but its novelty requires us to understand the perceptions of the ChatGPT-generated content. There is a minimal amount of research on utilizing AI, especially ChatGPT, for mobile health interventions. The exploration of the role of generative AI in mental healthcare applications is a complex and nuanced undertaking that necessitates careful consideration and in-depth analysis. This complexity arises from the intricate interplay of various factors, including inconsistent and potentially harmful responses, privacy risks, and ethical considerations [30, 31]. However, despite these obstacles, the urgent need for digital mental health content and the limited availability of human resources for content creation prompt a comprehensive exploration of the potential benefits that may arise from the responsible use of Generative AI in this domain. ChatGPT's ability to generate language allows it to provide therapeutic interventions based on proven techniques. It helps users explore their thoughts and behaviors, promote self-awareness, and gain insights into their mental well-being [24].

The inclusion of content generated by ChatGPT in the SUNRISE application is being carried out as a case study into an experimental version of the application. This undertaking plays a crucial role in methodically investigating the possible applications of ChatGPT in the realm of mobile health interventions, thus adding to a detailed exploration of its abilities in this particular context. Although humans are considered the ideal content creators, practical challenges such as the time-intensive nature of collaborating with human experts may necessitate alternative approaches. ChatGPT's efficiency and rapid script development capabilities make it a practical solution to expedite the process and improve accessibility to culturally sensitive content. This approach acknowledges the limitations of relying solely on human-generated content, particularly in situations where time constraints and the need to address mental health disparities require a more agile solution. By analyzing the SUNRISE application as a case study, we explore the utilization of ChatGPT in creating meditation scripts and evaluate its effectiveness in generating content for the meditation activity. The evaluation of user experience involves a comparison between the content produced by ChatGPT and that generated by human experts.

1.3 Research Questions

In examining the acceptability of ChatGPT-generated content, a comparison is drawn between the content generated by ChatGPT and that produced by human experts. The assessment of AI-generated meditation content is centered on users' perceptions and experiences. Additionally, the evaluation of AI-generated scripts considers the users' familiarity with the meditation experience in assessing the perceptions and experiences of users. The research questions are formulated to address the perspectives of novice and proficient practitioners in meditation practices. This research aims to capture insights

into user perceptions and experiences within the context of the meditation activity. The research questions are framed as follows:

RQ1: How does the user's background (experience/expertise) influence their perception of the meditation scripts developed by human experts and AI?

RQ2: How do AI-generated scripts perform in terms of user experience compared to human expert-generated scripts?

In addressing Research Question 1 (RQ1), the study examines user attitudes toward both AI-generated and human expert-generated content. This investigation considers user behaviors in relation to their prior meditation experience and delves into their perceptions of both the content associated with meditation. The evaluation of the vocal aspect linked to meditation has been incorporated in RQ1 in order to assess user perspectives and establish a uniform experiential standard for content produced by both human experts and AI systems. The second research question explores user experiences with AI-generated content as opposed to human expert-generated content. The parameters for assessing user experience include content complexity, clarity, mood impact, and overall activity experience. The responses to these research questions aim to offer an initial evaluation of the effectiveness and user reception of AI-generated content in the context of mobile health applications. The study provides insights into the efficacy of AI in generating content for mobile mental health interventions, offering the first characterization of ChatGPT's role for users with diverse experience levels.

1.4 Research Objective

The primary objective of this study is to evaluate the potential of Artificial Intelligence, specifically ChatGPT, in creating content for mobile health applications. The investigation focuses on comprehending how users with varying degrees of familiarity with

meditation practices perceive AI-generated content in mobile health interventions. The study aims to gain insights into the perceptions and user experiences of content generated by ChatGPT among both novice and proficient meditation practitioners. This phase is crucial in ensuring that AI-generated content does not unintentionally include inappropriate or harmful elements. The feedback obtained will play a vital role in optimizing the utilization of ChatGPT and improving its acceptance among individuals with different levels of familiarity with meditation practices. Having a profound understanding of this knowledge is of utmost importance in tailoring and optimizing AI-generated content to effectively cater to diverse user groups. This leads to improved accessibility and impact of mental health interventions. Furthermore, this study contributes significantly to the ongoing discourse on the integration of AI technologies in healthcare, shedding light on their potential role in supporting and augmenting traditional practices.

1.5 Scope of Research

The scope of the current investigation is to examine the audio components of an application, with a specific focus on the text-based content of the audio script. By utilizing a voice synthesizer, we can modulate the voice of an audio file, which allows us to focus on the script's content rather than the voice's particular characteristics. It is important to note that the study does not consider image, audio voice, or video elements but concentrates solely on text-based information. The primary variable being studied is the perceptions and user experience associated with content generated by ChatGPT, particularly among individuals with varying levels of meditation experience. Although this methodology has certain limitations, it provides valuable insights into the significance of AI-generated content in mobile health applications. This preliminary study sheds light on the effectiveness of AI-generated content and is expected to encourage further research

into the integration of ChatGPT-generated content in mobile health interventions. The key findings of this research have the potential to pave the way for similar initiatives aimed at improving accessibility and efficacy in mobile health support through innovative applications of AI technology.

1.6 Positionality Statement

The SUNRISE Project is an ongoing initiative focused on developing a sociotechnical behavioral health intervention for a Native Amercian tribal community in the Southwestern United States. The project was established in 2019 and the project team comprises academic and community researchers, community advisors, and software developers. The project uses a participatory action research approach to address complex behavioral health challenges that have been identified by various community stakeholders. The SUNRISE app is one component of this project and has been developed through iterative participatory design involving computer science researchers specializing in serious games for health, cross-cultural psychology for Indigenous communities, networking in challenged environments, public health, and education leadership.

The author plays a role within the SUNRISE mobile health app team as a software developer and usability researcher who helps implement new features and helps facilitate quality assurance and usability testing that conforms with visions identified through participatory design sessions.

1.7 Thesis Outline

This thesis is organized into seven chapters, each serving a unique purpose in building the argument for a comparative analysis between human expertise and Artificial Intelli-

gence among individuals with varied familiarity with meditation practices. In Chapter 2, we explore the meanings of various terms mentioned in the thesis. Existing literature concerning Artificial Intelligence in mental health applications, Generative Artificial Intelligence in health care, Indigenous mental health applications, and the intersection of these are explored in Chapter 3. The process of generating content is explored in Chapter 4. The sampling strategy and data collection techniques used in this study are described in detail in Chapter 5. Chapter 5 also examines the research-related ethical issues and data analysis methods. The thesis takeaways are discussed in Chapter 6, followed by the conclusion and future work in Chapter 7.

Chapter 2

Terminology

2.1 Generative Artificial Intelligence

Generative AI, also known as Generative Artificial Intelligence or GenAI, refers to the ability of Artificial Intelligence to generate text, images, or other types of media using generative models [32]. These models learn patterns and structures from their training data, enabling them to create new data that exhibits similar characteristics [33].

2.2 Content

The term "content" within an application encompasses various elements, including text, graphics, images, audio, video, and information. Text refers to the information and data presented and displayed within the application. It forms textual content that users can read and interact with. Graphics represent the visuals of the ideas, such as icons, typography, and page layout. Audio pertains to the spoken or recorded content that plays in the application's background. Images encompass the pictures and backgrounds within the application. When we mention displaying a set of images along with audio, we

refer to video content, where a combination of visuals and audio elements come together to convey information or create an engaging experience for the user.

Our primary focus is audio content, specifically its utilization within the meditation activity featured in the mental health application. The process of creating audio content for meditation involves two key steps. First and foremost, the foundation of this audio content is a carefully crafted script. This script serves as the narrative for the meditation experience, providing instructions to the user throughout the session. Secondly, to bring this script to life, an individual must lend their voice by reading out the script. This narration process is crucial in creating a soothing and immersive meditation experience for the users.

2.3 Mobile Health Application

In order to grasp the concept of a "mobile health application," it is crucial to clarify two fundamental terms: health and mobile application. A mobile application is specifically designed to function on a mobile phone, such as an Android or an iPhone, and can be installed through dedicated platforms like the Google Play Store for Android or the Apple App Store for iPhones. Within the realm of a mobile health application, the emphasis lies on enhancing particular health behaviors. Our research delves into a mobile health application called SUNRISE, which has been developed using the Flutter framework and Firebase technology. The SUNRISE team is dedicated to promoting mental well-being by offering support and a range of activities to assist individuals in effectively managing their mental health. Among the various mindfulness activities offered in mental health applications, our specific focus centers on meditation activity. Being researchers actively participating in the development of ChatGPT-generated content for meditation in the SUNRISE application, our positionality drives our dedication to improving accessible

mental health interventions. While our goal is to maintain objectivity, we acknowledge that our roles may influence the interpretation and objectives of the study.

2.4 Usability study

Usability is defined as the "degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [34]." Here, we conduct a usability study of a mobile health application named SUNRISE as a sub-study for our research.

2.5 Indigenous

The word "Indigenous" has different definitions defined by various organizations. When we refer to the term "Indigenous", we refer to the definition coined by the World Bank. According to the World Bank, Indigenous People are "distinct social and cultural groups that share collective ancestral ties to the lands and natural resources where they live, occupy or from which they have been displaced. The land and natural resources on which they depend are inextricably linked to their identities, cultures, livelihoods, and physical and spiritual well-being. They often subscribe to their customary leaders and organizations for representation that is distinct or separate from those of the mainstream society or culture [35]."

Chapter 3

Literature Review

A thorough literature review is required to achieve the primary goal of this research, which is to analyze the potential role of Artificial Intelligence in developing mental health applications for Indigenous communities. This review would be performed through well-established databases such as Google Scholar, ACM Digital Library, IEEE for computer science and informatics perspectives, and PubMed for insights into mental health.

In the initial segment of the review, we will navigate the role of Artificial Intelligence in mental health from the official beginning of AI in the 1950s to the present day. This wide timeline allows a detailed study of the progression and maturation of AI technologies. It helps understand how AI has gradually found its place in healthcare, especially in mental health domains. Moreover, it would foster discussions on the ethical considerations of embedding AI in mental health care, providing a rounded view of the domain. Only research papers published in English language constraints are considered for data reliability. This part of the review promises to unveil how AI technologies have transformed over decades to become vital tools in mental healthcare. An extensive review of the role of Generative Artificial Intelligence in healthcare is conducted to understand the extent of usage of Generative AI. Next, the focus is to understand the current land-

scape of mental health applications tailored for Indigenous people, drawing from articles published in English. Knowing how these applications have previously integrated crucial aspects such as the historical background and diversity of Indigenous communities is vital. This phase is expected to provide an in-depth understanding of how technology has been incorporated into the longstanding traditions of these groups.

3.1 Artificial Intelligence in Mental Health Applications

Artificial Intelligence can help assess, diagnose, and treat mental health diseases [36]. Using Artificial Intelligence in mental health can improve the quality and level of services such as continuous monitoring and accurate treatment [36]. It has the potential to be employed in psychotherapy using Artificial Intelligence and natural language processing technology. In addition, machine learning techniques can also detect mental health status, offering respective treatment [37]. Rasouli, Ghafurian, and Dautenhahn conducted an online survey to identify the preferences of undergraduate and graduate students in using intelligent agents for their mental well-being [38]. Students preferred personalized coaching of relaxing activities and social skills training through humanoid robots and conversational agents [38]. A multitude of chatbots, functioning as conversational agents, have been devised to bolster mental well-being by prioritizing accountability, healthcare provision, and user comprehension. Each chatbot reviewed is custom-made to serve specific purposes, such as diagnosis, prevention, and therapy [39]. Koulouri, Macredie, and Olakitan conducted a survey among young adults, which revealed that they prefer self-help techniques to cope with frequent stress. The researchers suggested enhancing mental health methodologies by utilizing interactive and adaptive content facilitated by

conversational chatbots. However, the study's limitations include a small sample size and the use of a convenience sample [40].

Individuals prefer to receive mental health services online to maintain anonymity and are voting for chatbots. They help save time, maintain confidentiality, are always available for immediate assistance, relieve geographical isolation, and are easy to use [41]. Thieme et al. proposed a collaborative and design-centered approach to develop an AI application using internet-based Cognitive Behavioral Therapy for depression and anxiety. The team created mockups for meaningful conversations with AI and generated expected outputs. Interviews were conducted to analyze the final design, and participants regarded it as a helpful indicator for patient treatment and monitoring. The study emphasizes using human-AI interactions to strengthen the outcomes in clinicians' prediction and treatment of mental health issues instead of replacing them [42]. Rani, Vishnoi, and Mishra created a mental health application portal with Flutter called Saarthi to offer information about and treatment for anxiety and depression. The application uses AI techniques such as Natural Language Processing, Term Frequency – Inverse Document(TF-IDF), N-Gram, Tokenization, NLTK, and sentimental analysis to answer the user's customized questions based on their stress levels. The application is satisfactory according to 685 clinical mental health centers and plans on providing online counseling in the future in an efficient way [43].

A chatbot was designed by Dharmapuri et al. to detect suicide and prevent it. The algorithm used is trained with actual data between patients and psychologists of ten generations and sends emails to authorities if suicide intent is detected [44]. A web application named Moderated Online Social Therapy(MOST) was developed for mental health recovery using social media by D'Alfonso et al. The applications suggest content to the users based on the comments mentioned in social media using advanced artificial computational methods. Developing individual therapy sessions based on linguistic anal-

ysis is the future scope of research for the project [45]. A chatbot is presented within the software system by Anand et al. to assess various factors affecting the mental health of the individual and provide mental health education. This chatbot is built by importing packages and dialogues, developing Artificial Intelligence models, data training, and prototype model development [46]. Using machine learning algorithms for data training, Wang et al. proposed a chatbot to assess the mental health of perinatal women in real-time by accessing their health data through human-robot interaction [47].

Mody and Mody conducted a comparative analysis of techniques and algorithms used for monitoring mental health, such as counseling, therapy, and diagnosis. Subsequently, they devised a composite system for customized mental health care by combining the best characteristics of the above systems [48]. A remote patient monitoring system is developed using non-invasive technology such as RFID and NCS to supervise patients' vitals in mental health facilities. The monitoring uses AI techniques such as prediction modeling and classification modeling, followed by evaluation. This system was developed by Shaik et al. to eliminate any unforeseen clinical disasters [49]. Wilbourne, Dexter, and Shoup assessed of the effectiveness of a mobile self-help intervention called Sibly, which incorporates real-time assistance, established scientific principles, and AI. Artificial Intelligence used in Sibly ensures the quality of the applications and accessibility to mental health tools. Sibly utilizes cognitive behavioral tools and demonstrates improvement in the well-being of 95% of the users [50].

Nayar et al. proposed and evaluated a mental health assistant called Dost, a chatbot built on the Rasa framework. The chatbot uses tokenizers, featurizers, intent classifiers, and entity extractors to predict the output of a given input. The bot is deployed on Telegram and successfully identified mental health disorders with an accuracy of 83.5% [51]. Tess, a psychologically customizable AI chatbot, uses emotional algorithms and machine learning techniques to support mental health. Due to its emotional Intelligence, Tess

adjusts itself based on the user's needs and provides meaningful support. A study was conducted by Fulmer et al. to identify the feasibility and efficacy of Tess for college students. In the study, a group of participants received new messages, while the other group received the same content biweekly for four weeks. Both groups exhibited a significant reduction in depression and anxiety [52]. Liu et al. performed a comparative study between chatbot therapy and bibliotherapy(self-help intervention) in providing self-help to university students. Surveys were conducted, and the results showed that chatbot therapy is efficient compared to bibliotherapy [53]. Dev et al. conducted a comparative study of chatbots used for healing depression, such as CARO, XiaoIce, DEPRA, PRERONA, and Eviebot. There is a reduction of depression using all these chatbots; however, the chatbots have not unlocked their full potential [54]. Chatbots such as MYLO, Bonobot, and Vivobot, developed for self-help and stress reduction, were evaluated, and the participants provided positive feedback and satisfaction in using these chatbots [40]. The Cirkus design probe is being utilized to explore the potential of Machine Learning as a source of play and motivation. The focus is on identifying movements among children with and without Sensory-Based Motor Disorders. Cirkus boasts an adaptable design that encourages participatory game development and movement data collection in diverse settings. Additionally, the participatory machine learning approach, which involves children in choreographing animal movements, holds promise for generating valuable datasets for assistive technology [55]

3.1.1 ChatGPT in Mental Health

Artificial intelligence (AI) language models have seen massive development and are currently used in daily-use services and activities. These models enable AI systems to converse with people organically and successfully since they are built to understand, in-

terpret, produce, and respond in language like human speech. One of the largest language models that gained huge popularity is ChatGPT. The beginning of a new age in online interventions, highlighted by the incorporation of Chat-GPT and AI-assisted therapist supervision, highlights the potential of these technologies to improve and expand mental health assistance [56]. ChatGPT can pave the way to a new section of AI, opening many opportunities in the mental health field [57]. Mental health practitioners can use Chat-GPT as a resource and enhance proactive treatment adjustments as it is trained using large data sets. This study suggests integrating ChatGPT-enabled apps in health devices for child mental health care. ChatGPT can also help patients easily access information and reduce stress without any potential bias during therapy [58]. Using chatbots based on ChatGPT, anyone worldwide can easily access personalized mental health support, coping strategies, and self-help. However, there are limitations like emotional Intelligence in conversation, customization, privacy, and security concerns [59].

3.2 Generative AI in Healthcare

Generative AI uses deep learning and modeling techniques to produce diverse content by appropriating existing media like text, audio, video, and graphics. Generative modeling is an Artificial Intelligence technique that generates synthetic artifacts by analyzing training examples, learning their patterns and distribution, and then creating realistic facsimiles [60]. Deep learning is an Artificial Intelligence method that teaches computers to process data in a way that is inspired by the human brain [61]. Jovanovic et al. discussed ubiquitous generative AI techniques that can be leveraged: generative adversarial networks, generative pre-trained transformers (GPTs), and generative diffusion models. For our research, we delved more into GPT, one of the large language models (LLMs), especially how it contributes to natural language. Volker Bilgram et al. discuss how

GPT can augment the early phase of innovation, particularly exploration, ideation, and digital prototyping [62]. The versatility of LLMs makes them applicable in diverse scenarios. Pretrained LLMs often outperform models specifically trained for a particular use case [62]. This is helpful as we do not have to create labeled training data, which is difficult to obtain given the sensitivity of healthcare data.

Despite the considerable advantages that come with the utilization of GPTs, it is of utmost importance to exercise caution and approach their application with care, placing a strong emphasis on thoroughly assessing the social and ethical implications involved. As Kenthapadi et al. mentions in their paper, these considerations include lack of interpretability, bias, discrimination, privacy, lack of model robustness, fake and misleading content, copyright implications, plagiarism, and environmental impact associated with training and inference of generative AI models [63]. To handle this, the author suggested adopting responsible AI principles during the development and deployment of LLMs. When it comes to healthcare, data maintenance is a critical task one should account for. Nova discusses using Generative AI to enhance electronic health records(EHRs) through automated conversation summarization, simplifying complex medical language into patient-friendly summaries, and providing personalized care recommendations using data from smartwatches and wearables [64]. The paper proposed a generative AI technical framework for recording a conversation, summarizing it, and adding the summary to the EHR after converting it to patient-friendly language. This framework involves data collection, preprocessing, training, evaluation, fine-tuning, deployment, user interaction, summary generation, output presentation, and feedback iteration. They also used generative AI models to generate personalized care recommendations [64].

ChatGPT, a generative AI that utilizes LLMs, utilizes transformer neural networks pre-trained on unlabelled large text corpora to acquire a comprehensive understanding of language patterns [65]. As ChatGPT generates human-like text better than most natural

language processing tools, we can use it to fine-tune the technical frameworks' summarization capability. While ChatGPT is beneficial, we should recognize the challenge that comes with using ChatGPT-based systems. These systems heavily rely on the training data that needs to be relevant for specific purposes. Also, when the system has to address a situation it is not sufficiently trained on, it still would suggest with the same certainty as in cases where it is adequately trained. This could be risky, especially when it comes to giving personalized healthcare recommendations. Hence, it is essential to have elements in place to address this issue when using ChatGPT or any other generative AI framework [65].

Mental Healthcare applications are a subset of healthcare applications that must be customized for a user group. While someone in physical pain would be willing to navigate an application that does not appeal to them, the same cannot be said for people going through mental health issues. Smits and Sophie discuss the same in their paper 'Customization in e-(mental) Health [66].' Cheng states that despite the convenience of ChatGPT, this advanced chatbot currently has limited practical applications in psychiatry. As mentioned in the paper, for using ChatGPT in mental healthcare, we require future advances that would inject empathy, emotion recognition, personality assessment, and detection of mental health warning signs into ChatGPT [67]. This can be achieved using clinically validated algorithms such as voice comprehension modules, and emotion discrimination algorithms to integrate real-world data and AI user and patient interfaces. ChatBeacon [68], is a ChatGPT-powered mental health assistance, an all-inclusive customer service platform. Another example is a free therapy program, Koko [69], which tests a GPT-3 mental health intervention.

3.3 Indigenous Mental Health Applications

Digital storytelling, e-mental health apps, and social media are the major technologies Indigenous people use for mental health access. These technologies have shown a positive impact in improving the mental health of these individuals, and thus, creating culture-specific Indigenous mental health applications could significantly improve their mental health [70]. Apps with culturally relevant content targeting mental well-being are strongly accepted by Aboriginal and Torres Strait Islander people in the study conducted by Povey et al. Personal factors, environmental factors, and app characteristics are the major concerns affecting the acceptability of mental health by these Indigenous people. Mobile apps are making mental health services accessible and cost-effective to the user [71] Tighe et al. conducted a randomized trial to test the effectiveness of a suicide prevention app named ibobbly with Indigenous Australians aged 18-35 years. The usage of the app for six weeks helped in the reduction of distress and depression among the participants. However, there are no signs of a significant decrease in suicidal ideation and impulsivity [72]. A study conducted by Perdacher et al. with practitioners and clients of the Indigenous Mental Health Intervention Program to test the feasibility of the Stay Strong application. Qualitative Interviews were conducted, and participants exhibited satisfaction with the app and marked it to be culturally safe for Aboriginal and Torres Strait Islander people. In addition to this, they also provided inputs for further customization of the application with cultural appropriateness [73]. A web-based application named StayinOnTrack utilizing SMS, MMS, email, and a companion mobile application is designed for the mental health support of new Indigenous fathers. The application consists of testimonials of former users, informational resources, a mood tracker, and web pages about the project. Positive responses and encouragement were received after the production of the application developed by Skinner, Fletcher, and May [74]. An

online survey was conducted with historically black colleges and universities(HBCUs) to understand the barriers to accessing mental health care services. School work, job search, work-life balance, and relationships are the major reasons for stress, and an online community is created by grouping people with the same interests to provide mental health support. The community is named SoulGlow by Williams and Washington and uses vlogging techniques to make the users feel included. The future work of this application is to embed computer vision techniques and natural language processing to automatically find the mental health disorders of users from the vlogs [75].

Mobile applications supporting the well-being of Indigenous youth are reviewed by Noronha et al [76]. A mobile application named Joypop is under development for mental health and resilience based on insights from the review. Mobile applications such as ibobbly and HABITS were reviewed for this study. Ibobbly app is designed for suicide prevention of Aboriginal and Torres Strait Islander people of Australia by taking in inputs from their community and is currently available in the market [72]. HABITS, developed based on gamified cognitive behavioral therapy, is a mental health application focusing on Maori and Pacific communities and is also available in the market [76]. A wellness model was proposed for integrating Indigenous and non-Indigenous science based on insights from the review of the mobile applications [76]. According to the study by Hensel, Ellard, and Kotek, web-based applications, digital storytelling, and social media are the major patient-directed solutions for Indigenous mental well-being in Australia [77]. Mindspot, ibobbly, and Australian Integrated Mental Health Initiative (AIMhi) StayStrong, were designed to meet the needs of targeted populations using Information and Communication Technology (ICT) such as web and mobile apps [77]. Mindspot offers a virtual evaluation of mental health and Indigenous self-management courses that are web-based. ibobbly is used for suicide prevention among Indigenous communities [72], and the Australian Integrated Mental Health Initiative (AIMhi) Stay Strong app is used for evaluating factors

affecting mental health and providing support in overcoming them [77]. Reilly et al. conducted a review of web-based therapeutic interventions for Indigenous populations. Six interventions related to mental health are discussed: R U Appy, DM-DS, Sparx, Ibobbly, Stay Strong, and Stayin on track. The study highlighted the importance of developing applications specific to target communities [78].

Toombs et al. conducted a systematic review of electronic mental health technologies developed for Indigenous people [79]. This study reviews mobile applications such as Ibobbly [72], The Australian Integrated Mental Health Initiative (AIMhi) Stay Strong [73], Cedar Project WeTel, SMS4dads, Rapid City Hospital's Child and Adolescent Telepsychiatry Service, and TeleLink Mental Health Program38 and websites such as Tertiary Student Health Project and SmokingZine for American Indian Youth. Increased exposure to cultural content, coordination, and consultation with community members would resolve the barriers faced by Indigenous populations in accessing mental health [79]. Blackdeer and Wolf explored the mental health interventions for suicide prevention using evidence mapping for American Indian and Alaskan Native youth. They categorized the research-supported interventions into four categories: school-based, cultural adaptations, culture as treatment, and culture as community involvement. This study emphasizes the necessity of cultural inclusivity and community involvement in developing mental health interventions [80]. Antonio et al. analyzed culturally tailored mental health interventions utilizing the Positive Youth Development Framework for American Indian and Alaskan Native populations. The interventions reviewed are related to mental health resilience, suicide prevention, substance use, alcohol use, depression, and anxiety. The research revealed that most applications have positive outcomes in improving mental health and reducing substance use among AI/AN individuals [81].

Table 1: Unique Applications in Literature Review

Category	Applications	Overview
Artificial Intelligence in MHA	ChatBots	Dost, Tess, CARO, XiaoIce, DEPRA, PRERONA, Eviebot, MYLO, Bonobot, Vivobot
	Saarthi	Information about and treatment for anxiety and depression
	MOST	Mental Health recovery based on Social Media
	Serena	Counselling application for Mental Health
	Sibly	Intervention used to assess Mental Health Tools
Generative AI in HealthCare	ChatBots	ChatBeacon, Koko
Indigenous MHA	iBobbly	Suicide prevention app for Indigenous Australians
	ALMhi StayStrong	Mental Healthcare app for Aboriginal and Torres Strait Islander people
	StayinOnTrack	Mental Health support app for Indigenous Fathers
	HABITS	Mental Health support app for Maori and Pacific People

3.4 Findings in Literature Review

This literature review highlights the potential benefits of integrating AI, particularly Large Language Models such as ChatGPT, with mobile health interventions. ChatGPT's efficiency and rapid script development capabilities can significantly reduce the workload of human experts, making it a valuable tool in content creation. However, it is important to recognize that human expertise remains essential in creating meaningful and ethical healthcare solutions. While ChatGPT can function autonomously in routine scenarios, a collaborative approach is necessary in critical cases that require human understanding and ethical considerations. The successful integration of AI in mobile health interventions depends on the careful balance between ChatGPT's strengths and the unique creative

and ethical abilities of human content creators. By leveraging this synergy, the field can advance mobile health interventions while maintaining the importance of human expertise. Table 1 presents a visual representation of the widely accepted belief that Artificial Intelligence, particularly Generative AI, is primarily linked to chatbot usage in the healthcare field. Nevertheless, it is imperative to acknowledge the vast potential of Generative AI, which extends well beyond chatbots. Our investigation introduces a different viewpoint, emphasizing the adaptable utilization of Generative AI in content creation across various applications within the healthcare industry. In this thesis, we compare the usage of content [2.2] generated by ChatGPT and humans from the perspective of an Indigenous mental health application.

Chapter 4

Content Development

The application and the original meditation scripts are tailored with cultural references unique to a specific Native American tribal community, centering its attention on adolescents' mental well-being. The app offers various mindfulness activities designed for the mental wellness of the Indigenous youth, including breathing exercises, running, walking, meditation, and art creation. Our primary focus is on the meditation activity, which follows Wayment and Bauer's Quiet Ego Model [29], that has been carefully iterated through a community-based participatory design process. The author is involved in the software development of the application and collaborates in generating ChatGPT-based content for the meditation activity. This Chapter delves into the processes of script development from human and Generative AI perspectives for the meditation activity. The meditation activity has 4 versions with the four seasons of the year.

4.1 Script Development by Human Expert

The meditation activity was crafted with a human expert in cross-cultural psychology as a part of a community-based participatory design approach. The SUNRISE

Team diversified the meditation activities by associating each season with a unique technique of the Quiet Ego model such as mindfulness(detached awareness), interconnectedness(inclusive identity), compassion(perspective taking), and growth mindset(growth) [28]. The original human expert-generated meditation scripts had been carefully crafted to pair each season with one of the primary Quiet Ego model approaches: mindfulness, interconnectedness, compassion, and a growth mindset. This intentional alignment ensures that the essence of each season is in sync with a specific Quiet Ego model technique, thereby enhancing the overall user experience. For instance, in the SUNRISE App, mindfulness is linked with the renewal of spring plantings, while interconnectedness is associated with the flourishing of plants in summer. Compassion is interwoven during the harvest time of fall, and a growth mindset is integrated into the tranquility of winter. This amalgamation of seasonal aesthetics with corresponding Quiet Ego model techniques introduces variety and offers users a nuanced and comprehensive journey through different aspects of meditation practice.

4.1.1 Expertise

A proficient expert with the credentials of a licensed clinical psychologist has been engaged in developing human expert-generated content for a meditation script. The human expert, though non-Indigenous, brings significant experience in collaborating with Native American tribal communities to develop and tailor psychological interventions. This individual possesses a wealth of knowledge and expertise, having served as a seasoned researcher for more than three decades in their respective field.

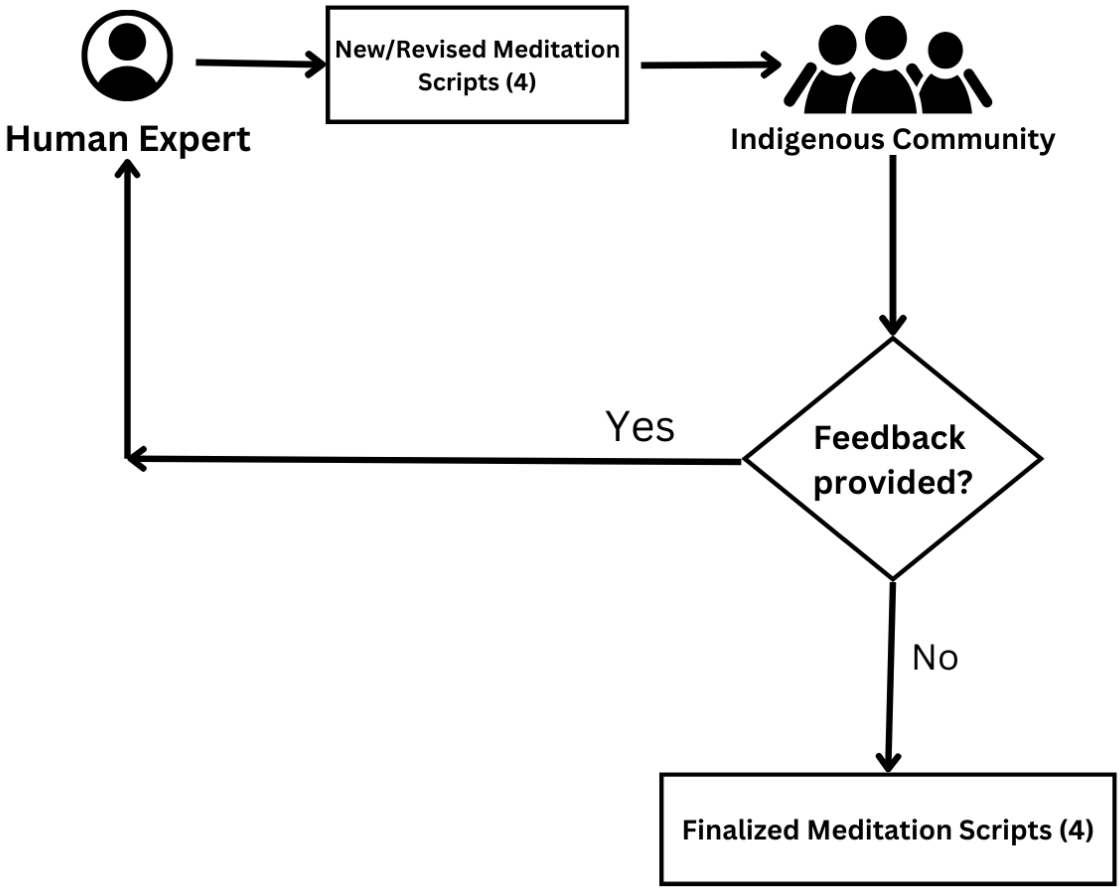

Figure 1: Script Development by Human Expert

4.1.2 Procedure

The procedure followed for the development of meditation scripts is depicted in Figure 1. The process of developing human-expert scripts followed an iterative approach, which involved a participatory design that was community-based. This collaborative process brought together a human expert and a specific Indigenous community, with the goal of customizing the scripts to align with the community's unique perspectives and values. By actively involving the community, the aim was to ensure that the script creation process was guided by their input and insights. The partnership between the human expert and the Indigenous community played a significant role in shaping the foundation of the script, ensuring that it resonated with the community's distinct perspectives and values. The development of meditation scripts involved the utilization of techniques derived from the Quiet Ego Model. This entailed the integration of mindfulness, interconnectedness, compassion, and a growth mindset into the process of script creation. These components from the Quiet Ego Model were purposefully employed to steer the content of the meditation scripts, with the objective of establishing a comprehensive and nurturing framework for individuals participating in the practice of meditation.

4.2 Expert-provided keywords

To enhance the similarity between ChatGPT-generated scripts and those authored by humans, the author sought the assistance of a human expert who crafted the original meditation scripts. The human expert identified essential keywords to generate similar content using ChatGPT. This collaborative effort aimed at the content generation of ChatGPT, ensuring a higher level of alignment with the quality found in scripts created by humans. The human expert identified the keywords forming the basis for Generative Artificial Intelligence (GAI), specifically applied to ChatGPT. The identified keywords

served as guiding elements for GAI to generate new scripts, ensuring theoretical comparability with the original human-expert scripts. These words guide the construction of the script, ensuring that the message is coherent and focused. As the keywords are provided by the same human expert who developed scripts for human-generated content, consistency is maintained across the scripts.

The expert-provided keywords consist of essential components, such as the season's name, activities for the season such as agriculture, celebrations, and dances, the Quiet Ego model technique utilized, the name of the specific Native American Tribal community, the age of the youth participants, the readability level of the words used, and the word count. Furthermore, these prompts include vocabulary related to meditation, such as "beginner", "meditation", and "slow deep breathing technique." The prompts have been thoughtfully designed to enhance user engagement and ensure a coherent flow of information, incorporating words like "logical flow" and "engage the user." Capturing participants' attention and fostering sustained involvement in the meditation activity is crucial for ensuring user engagement thereby promoting mental well-being. This strategic integration of diverse elements and language choices aims to create prompts that are both accessible and engaging for youth participants while maintaining a logical structure in the content presentation.

4.3 Script Development by ChatGPT

ChatGPT is a sibling model to InstructGPT [82], which is trained to follow instructions in a prompt and provide a detailed response.

4.3.1 Expertise

ChatGPT, constructed using the GPT-3.5 framework, is an enhanced version of OpenAI's GPT-3 model that was initially introduced in 2020 [19, 20, 21]. It excels in comprehending language, generating text proficiently, and facilitating effective machine translation. This iteration of the model has been extensively trained on publicly available large text datasets and has undergone fine-tuning to specifically enhance its ability to generate conversational responses. Consequently, ChatGPT showcases an exceptional capability to produce responses that closely resemble human-like conversations [22, 23, 83, 84].

4.3.2 Procedure

The author, who lacks expertise in generating prompts and meditation practices, is responsible for creating the prompts. A generic prompt is formulated by incorporating all the given expert-provided keywords, except for "season" and "Quiet Ego technique." After this initial stage, both "season" and "Quiet Ego technique" are included in the generic prompt, leading to the development of four unique versions. The purpose of incorporating both season and quiet ego techniques is to add variety and improve the quality of the prompts by making them more specific. Once the prompts are inputted into ChatGPT, scripts are generated.

The generic prompt created by the author is as follows: "Generate a meditation script for [Spring/ Summer/ Fall/ Winter] season for [Native American tribal community] youth of age 14-24 years. Use the technique of [Mindfulness/ Interconnectedness/ Compassion/ Growth Mindset]. The words used in the script should be understandable by 6th-grade students or lower. This meditation is for beginners and should focus on slow, deep breathing. The content should follow a logical flow and engage the user in meditation. The content should describe the [Spring/ Summer/ Fall/ Winter] seasonal

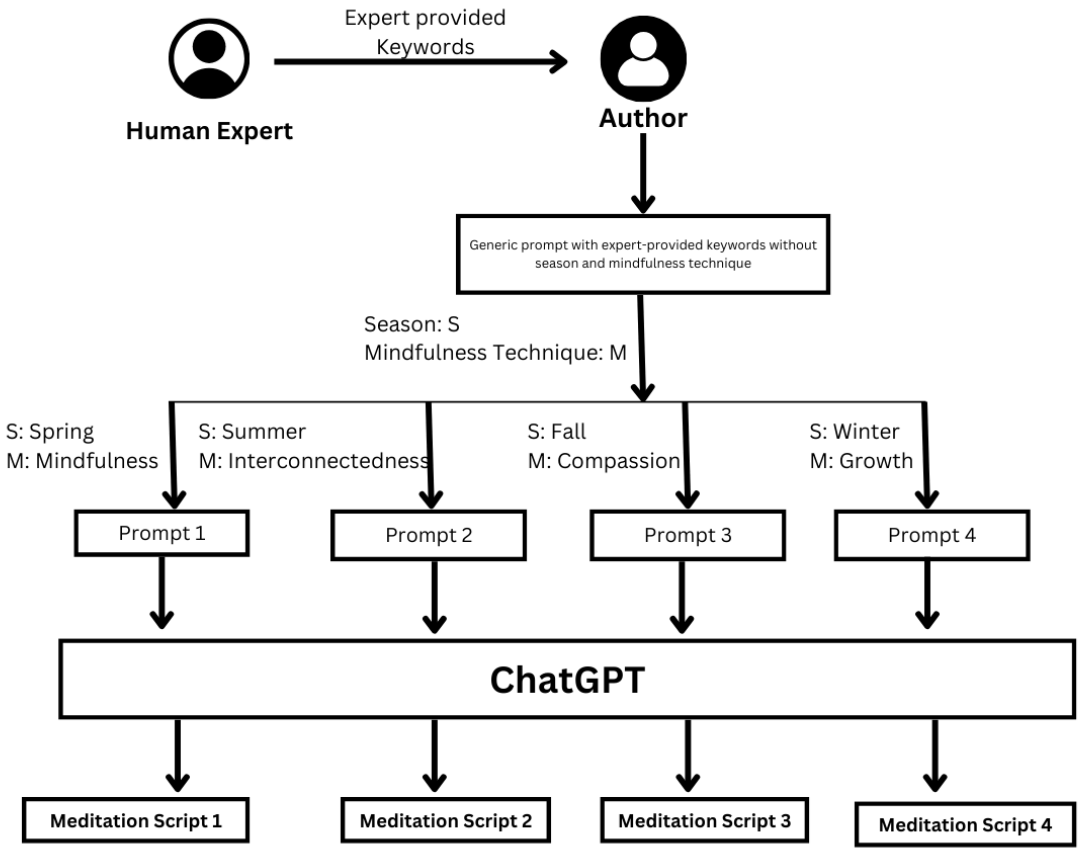

Figure 2: Script Development by AI

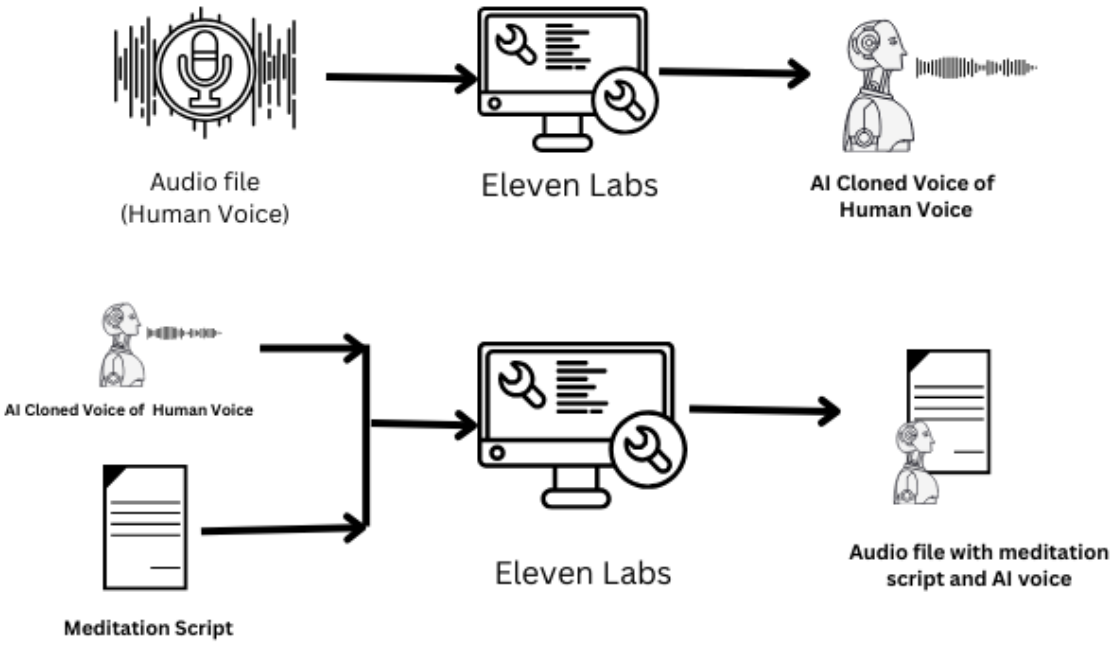

Figure 3: Audio Development for Scripts.

activities related to agriculture, celebrations, and dances. The script should use simple sentences and contain less than 500 words." Distinct scripts for the meditation activity are generated by incorporating the relevant Quiet Ego model technique, which is refreshed for each season. Figure 2 visually represents the procedure for the generation of AI content using ChatGPT.

4.4 Audio Development for Scripts

4.4.1 Procedure

As our primary focus is on the textual content, maintaining uniformity in the audio between human-expert and AI-generated scripts ensures consistency of experience. We employ two voices to narrate the meditation activities—a male and a female. In order to

create a well-rounded and inclusive atmosphere, the utilization of both male and female voices is implemented, catering to a diverse audience and promoting gender inclusivity within the meditation activity. We have utilized Artificial Intelligence for voicing all the scripts, relying on a tool called ElevenLabs for text-to-speech technology [85]. ElevenLabs specializes in Artificial Intelligence and deep learning for creating natural-sounding speech synthesis and text-to-speech software. By employing advanced algorithms, this system utilizes contextual analysis of text to effectively identify emotions such as anger, sadness, happiness, or alarm. The primary aim is to achieve a more genuine and human-like tone in the generated output through the process of training AI [86]. Figure 3 represents the process for the development of an audio file using AI voice for the meditation script. The tool processes the audio files to generate the voice, and once available, we input the script for speech conversion, selecting the desired voice. An issue arises when the text contains over 200 words, which affects the resemblance to the original inputted voice. To resolve this, the script is divided into three sections and later joined after audio generation. For merging the audios, we use Audacity, a free and open-source digital audio editor and recording application available across various operating systems [87]. Using free tools like ElevenLabs and Audacity is deliberate to ensure easy accessibility. The audio files employed in the research are pre-generated prior to the collection of data, ensuring uniformity across all participants, and are not produced in real-time. These pre-generated audio files are seamlessly incorporated only once into the meditation exercises featured in the SUNRISE application.

4.4.2 Ethical considerations

In the realm of Artificial Intelligence, it is of utmost importance to consider the ethical implications of generating AI voices, which necessitates obtaining consent from individ-

uals. This process involves informing individuals and seeking their explicit permission before utilizing their voice data for training or developing AI-generated speech. By adhering to this practice, transparency is maintained, and the autonomy of individuals is upheld, acknowledging their right to control the usage of their personal data. Consent mechanisms typically outline the intended purpose for which the voice may be used, along with any potential implications that may arise. Consent is obtained from the individuals to generate audio using ElevenLabs [85].

Chapter 5

Research Methodology

This Chapter examines the research methodology that includes the process of selecting samples, collecting data, and analyzing it in relation to the creation of two different versions of a meditation script. One version is created by a human expert in the field, while the other is generated by ChatGPT. The thorough examination of these methodological aspects offers valuable insights into the meticulous and systematic approach used to compare the two script versions, thereby highlighting the strength and reliability of the research design.

5.1 Usability Study

The meditation activity constitutes an integral component of the mobile health application (SUNRISE). Before delving into a detailed analysis of the meditation activity's content, it is crucial to thoroughly evaluate the application as a whole. This initial assessment is essential in determining the overall usability of the application. If the application proves to be user-friendly in this initial evaluation, then conducting a comparative analysis of the meditation activity's content becomes a more relevant and meaningful un-

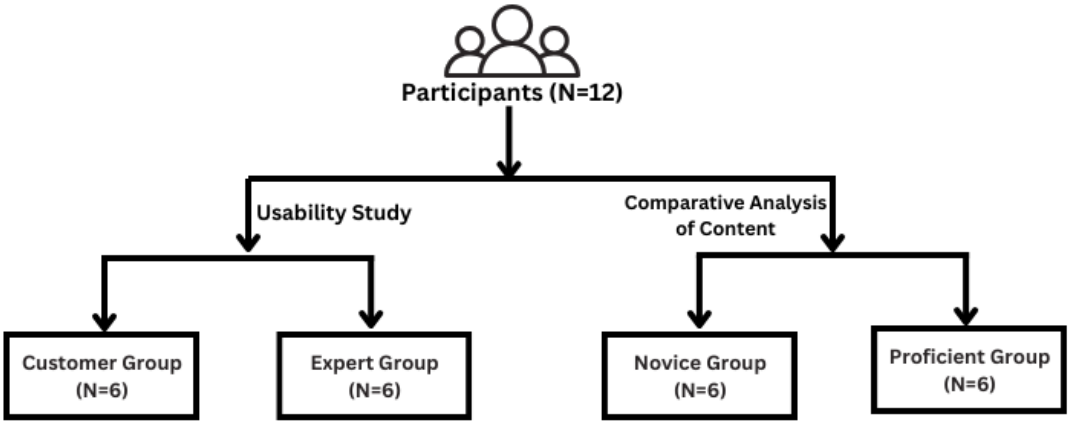

Figure 4: Sample Selection

dertaking. On the other hand, if the application shows any weaknesses or ineffectiveness during the initial evaluation, solely examining the meditation content in isolation may not provide the necessary strength and importance for comprehensive understanding. Therefore, a preliminary study on usability is carried out as a pre-study before conducting a comparative analysis of the content within the meditation activity of the application. This methodological approach guarantees that the subsequent examination of meditation content takes place within the wider framework of a mobile health application that functions efficiently and is user-friendly.

5.2 Sample Selection

Figure 4 represents the selection of participants for the usability study as well as a comparative analysis of content.

5.2.1 Usability Study

For the purpose of our research, we have established two main groups: the customer group(N=6) and expert group(N=6). The customer group consists of individuals aged 18 to 27 who are potential users of the app, irrespective of their background. The expert group comprises individuals who are 18 years old or older and possess specialized knowledge in either software development or behavioral health interventions. The term "expert" is deliberately used to encompass individuals who exhibit advanced proficiency in their respective fields. We discuss the limitations of using the term "expert" in the Limitations section 7.1. To ensure a comprehensive and influential study, we employed a purposive sampling technique, specifically targeting 12 participants who were evenly distributed between the consumer and expert categories. This deliberate selection process prioritizes participants whose experiences and insights can significantly impact our research outcomes [88].

To attract a diverse range of potential participants, we have implemented a recruitment strategy that involves various methods. This includes circulating emails, reaching out through social media platforms, and distributing university flyers. These approaches aim to engage individuals from different backgrounds and perspectives. Additionally, we have directly contacted potential expert participants via email to incorporate academic rigor and specialized viewpoints into our study. By employing these strategies, we aim to gather a well-rounded and knowledgeable group of participants to enhance the quality and validity of our research findings. Table 2 outlines the participant distribution details.

5.2.2 Content Analysis of Meditation Activity

During the content comparison phase, we involved the same participants from the usability study, including both expert and customer groups. This decision was driven by convenience, as these participants were readily accessible and willing to contribute to the additional assessment of the content. While the purposive sampling in the usability study ensured specific insights, extending this sample to the content comparison phase allowed for a seamless transition and efficient utilization of available resources. However, it is important to acknowledge that this convenience sampling approach may introduce potential biases due to the non-random selection of participants which we discuss in the Limitations section [7.1]. Participants are segregated into two groups, namely the novice and proficient groups, based on their self-evaluated familiarity with meditation. Novice individuals are those who rate their familiarity as three or lower, while proficient individuals are those who rate higher. The categorization of proficient users as those participants with ratings above 3 stems from the expectation that individuals placing themselves in these higher categories possess a greater degree of familiarity with meditation practices. Such classification enables the evaluation of ChatGPT-created meditation content across participants with differing levels of experience in meditation activities.

5.3 Data Collection

The assessment of research heavily relies on the data collection process, which consists of two distinct steps: the usability study and the evaluation of ChatGPT-generated content in the meditation activity within the application. Each step encompasses a 90-minute session, with the initial usability study and the subsequent comparative content analysis. In order to maintain consistency, all participants are equipped with mobile devices that have the application pre-installed, courtesy of the department. This stan-

Table 2: Participant Demographics.

Participant Code	**Age**	**Gender**	**Race**	**Ethnicity**
C1	22	Male	Mix of multiple races	Non-Hispanic
C2	21	Male	White	Non-Hispanic
C3	18	Female	White	Non-Hispanic
C4	18	Female	White	Non-Hispanic
C5	21	Female	Asian	Hispanic
C6	21	Female	Asian	Non-Hispanic
E1	30	Male	White	Non-Hispanic
E2	45	Male	White	Non-Hispanic
E3	33	Female	NA	NA
E4	56	Male	White	Non-Hispanic
E5	29	Male	American Indian or Alaskan Native	Non-Hispanic
E6	47	Male	Asian	Non-Hispanic

dardization of the user experience ensures that feedback primarily focuses on the app's content rather than variations in devices. Employing the same devices for all interviews minimizes potential errors arising from device discrepancies. All sessions are conducted in a private setting at Northern Arizona University, Flagstaff, and participants are sent reminder emails prior to the study outlining the session details and location. Customer users are provided with a $25 Amazon gift card for each session, while expert involvement is voluntary and uncompensated.

5.3.1 Usability Study

The application caters to two categories of users: mentees and mentors. The mentee version is evaluated by customers, while experts assess both the mentor and mentee versions. The evaluation process involves collecting data through surveys after the participants have explored the application. A thorough check of the mobile devices is conducted before app exploration to ensure their functionality. Users are required to follow

a talk-aloud protocol, expressing their actions and thoughts as they navigate through the application. Upon completion of the app exploration, participants are requested to complete a survey to evaluate the application's overall quality.

The survey employs the User version of the Mobile Application Rating Scale (uMARS), a well-established and reliable tool designed to evaluate mobile health application quality. uMARS is specifically tailored for assessing mobile health applications (MHAs) and is a comprehensive tool developed through semantic analysis and literature synthesis [89]. In accordance with the recommendations of uMARS developers, participants dedicate more than 10 minutes to testing the application on their mobile devices [90]. Subsequently, they evaluate the app's quality by completing a Google Form created using uMARS, which is a simplified user version of the Mobile Application Rating Scale (MARS). This user version consists of 20 items, condensed from the original 23. MARS, widely utilized for assessing app quality, has recently undergone a multicenter study that confirmed its construct validity, reliability, and objectivity [89]. Both rating scales assess five key dimensions: engagement, functionality, aesthetics, information, and subjective quality, employing 5-point scales. In contrast to the original MARS, uMARS consolidates the seven information quality items into four: quality of information, quantity of information, visual information, and credibility of the source. The subjective quality dimension is based on four questions, each utilizing a five-point scale, which covers the participant's willingness to recommend the app, anticipated frequency of app usage, willingness to pay for the app, and an overall rating. The assessment of the application encompasses its influence on the user's understanding, beliefs, and intentions pertaining to the specific health behavior. This evaluation is carried out by means of the perceived impact segment of the uMARS protocol [91].

The mentee version of the application was navigated by customer users who then proceeded to complete the survey. In contrast, expert users were provided with a signup

Table 3: Order of evaluation in Comparative Analysis

Participant Code	**First Version**	**Second Version**
C1	Human Version	AI Version
C2	AI Version	Human Version
C3	AI Version	Human Version
C4	Human Version	AI Version
C5	Human Version	AI Version
C6	AI Version	Human Version
E1	Human Version	AI Version
E2	AI Version	Human Version
E3	AI Version	Human Version
E4	AI Version	Human Version
E5	Human Version	AI Version
E6	Human Version	AI Version

process briefing before delving into the mentee version. Upon exploring the mentee version, experts were given access to the mentor version. After exploring both versions, expert users completed the survey.

5.3.2 Comparative Analysis of Content

The meditation activity consists of two different versions of content: one generated by human experts and the other generated by ChatGPT. The ChatGPT-generated content is prepared in advance of gathering data, rather than being created in real-time. All four scripts for the meditation activity are developed simultaneously and then uploaded onto the SUNRISE app. All participants, encompassing customers and experts alike, are granted access to an identical collection of pre-generated ChatGPT scripts. Subsequently, both the content generated by humans and that generated by AI are methodically presented to them through a random sampling technique, guaranteeing an equitable and unbiased distribution for evaluation. Table 3 displays the order of the human-generated content and AI-generated content provided to the participants. This ensures that the

evaluation of the meditation content is fair and unbiased. Participants were given mobile devices to engage in the meditation activity within the app. We used two devices, each loaded with a different app version to ensure consistency. This approach helps minimize any potential variations in user experiences that may arise from engaging with different activities or functionalities during the evaluation process. The comparative content analysis is conducted in a similar manner for both customers and experts through interview sessions. These sessions aim to gather detailed insights from participants about their experiences and preferences for the two versions of the app. The interviews include a combination of open-ended and closed-ended questions, creating a structured conversation that encourages participants to provide both detailed and spontaneous responses.

5.4 Data Analysis

The analysis of the data collected from the uMARS survey and interviews involves evaluating the app and comparing the content of the meditation activity. The usability study utilizes quantitative research methods, while a blend of quantitative and qualitative analyses is used to compare the content of the meditation activity. During interviews for comparative analysis of content, open-ended questions play a crucial role in obtaining qualitative data as they allow participants to articulate their opinions in a detailed manner. On the other hand, closed-ended questions aid in gathering quantitative data by providing a set of predetermined response options.

5.4.1 Usability Study

Statistical measures such as mean and standard deviation are employed in the evaluation of app features. Descriptive statistics play a pivotal role in gaining insights into the distribution of various app characteristics. Specifically, boxplots are utilized to analyze

the score distributions for engagement, functionality, aesthetics, information, total score, subjective quality, and perceived impact based on participants' feedback. It is noteworthy that each domain encompasses a range of subdomains such as: 1. Engagement: entertainment, interest, customization, interactivity, target group. 2. Functionality: performance, ease of use, navigation, gestural design. 3. Aesthetics: layout, graphics, visual appeal. 4. Information: quality of information, quantity of information, visual information, and credibility of the source.

Box plots are utilized as a graphical representation to depict the spread of various sub-domains and primary domains, including engagement, functionality, aesthetics, information, total score, subjective quality, and perceived impact. The comparison of box plots is carried out individually for customer and expert groups. Furthermore, mean and standard deviation scores are calculated for each domain within both the customer and expert groups.

5.4.2 Comparative Analysis of Content

The content analysis involved comparing the transcription of interview recordings, with the goal of exploring various elements related to the audio and script of both human experts and AI versions of meditation. The interview methodology focused on factors such as respondents' comfort with vocal tones, the soothing quality of the voice, the clarity and understandability of the language used, and the pace of the meditation, particularly in terms of its auditory aspects. For instance:

- **Comfort with vocal tones**: Understanding the spoken words in the meditation.

- **Calmness of voice**: Suitability of the voice for a meditative experience.

- **Clarity and understandability of language**: Assessing the cadence and delivery of the scripts.

- **Pace of the meditation**: Evaluating the timing and rhythm of speech.

Content evaluation included ease of understanding, engagement, and structuredness. For instance:

- **Ease of understanding**: Clarity in the words used for the meditation content.

- **Engagement of content**: The participant's attention and interest in the content.

- **Structuredness of content**: The logical flow and coherence in the material.

Valuable insights into participant perceptions contingent on their expertise were obtained through the combination of various factors, including the duration of the meditation. The parameters related to audio were presented in a tabular format, while bar graphs were utilized to compare the customer and expert groups for other parameters, thereby enhancing visual clarity. Additionally, RQ1 was addressed by examining how participants perceived these characteristics in relation to different content versions.

To assess the user experience in meditation, various factors such as the overall experience, word complexity, content clarity, and emotional impact are taken into account. Bar graphs are utilized to compare these factors between customer and expert groups, except for the emotional impact, which is analyzed thematically. The thematic analysis involves a multi-stage process where coders read and re-read transcripts to identify potential themes. These themes are then shared with the lead author for input. Coders then conduct a comprehensive review to ensure diversity in initial codes while developing overarching elements and higher-level sub-themes. This process is guided by Research Question 2, which aims to understand the emotional impact of meditation content. In the third stage, coders analyze to identify quotes that align with the overarching themes. The authors then review these themes, refining and naming them before finalizing the thematic analysis.

5.5 Ethical Considerations

Prior to commencing the research, participants will be provided with comprehensive informed consent forms that explicitly outline the study's objectives and procedures, emphasizing the voluntary nature of their involvement. This ensures that they fully comprehend their right to withdraw from the study at any point without facing any negative consequences. Our commitment to confidentiality involves protecting participants' identities and personal information through anonymization. The study brief clearly states that participants are free to withdraw from the study at any time if they so choose. Participants have given their consent for interview recordings, which are subsequently anonymized and transcribed. Ethical approvals were obtained through an Institutional Review Board (IRB) to collect data from participants, involving both audio and video recordings. The study anticipated no risks for participants engaging in the research activities. The usability study and comparative analysis in this research utilize a consistent group of participants. However, the utilization of this shared cohort introduces the potential for variations in participant representation. To address this issue, a comprehensive analysis of the limitations related to participant representation is presented in the Limitations section 7.1.

Chapter 6

Findings

This chapter delves into the outcomes derived from our comprehensive analysis, uncovering valuable insights. The app's usability is evaluated through the uMARS survey ratings. Additionally, our focus in the comparative analysis centers on assessing AI-generated content from both novice and expert user perspectives.

6.1 Participant Characteristics

The demographic composition of the 12 participants(N=12) is outlined in Table 2. The participants of the study exclusively associate themselves with the conventional gender binary, acknowledging their identity solely as either male or female. Among these participants, 7 (58.3%) were male, while 5 (41.7%) were female. The racial breakdown among the participants is as follows: 6 (50%) identified as White, 3 (25%) as Asian, 1 (8.3%) as American Indian or Alaskan Native, 1 (8.3%) identified with multiple races, 10 (83.3%) as Non-Hispanic, 1 (8.3%) as Hispanic, and 1 (8.3%) chose not to specify their background. Additionally, the mean comfort level in mobile application usage, presented in Figure 5, is 4.83, with a standard deviation of 0.91.

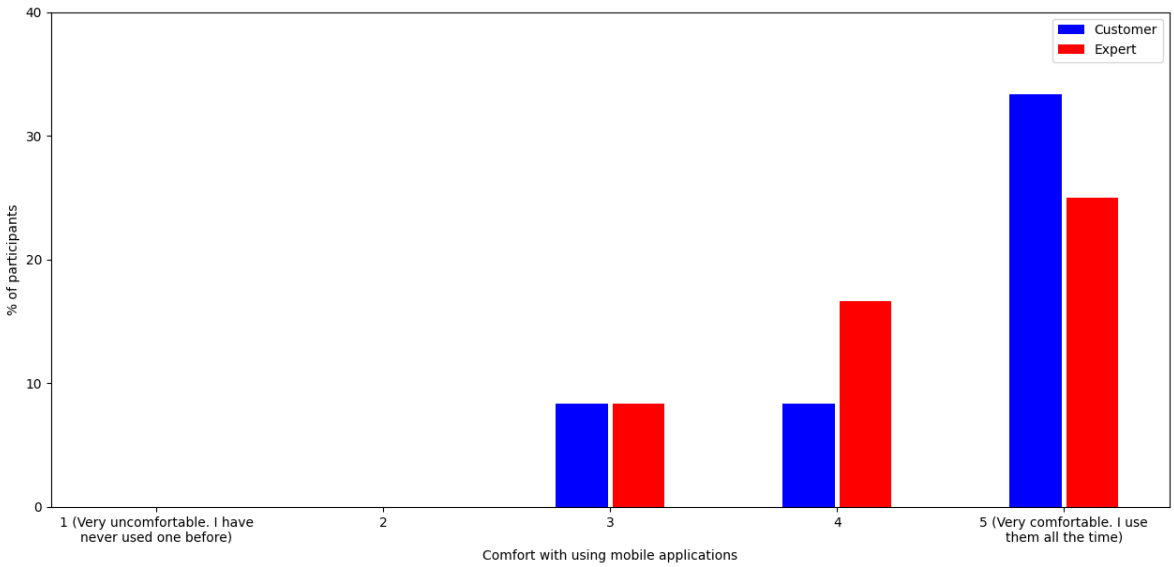

Figure 5: Comfort of the participant with mobile applications.

6.2 uMARS Rating of SUNRISE Application

Table 4 displays mean scores and standard deviations for each subdomain within the domains of engagement, functionality, aesthetics, information, subjective quality, and perceived impact on the uMARS scale. In the overall ratings, Aesthetics attained the highest mean rating of 4.35, followed by Information at 3.92, Functionality at 3.71, and Engagement at 3.75. For the customer group, Aesthetics and Information obtained the highest ratings with means of 4.5, followed by Functionality at 4.21 and Engagement at 4.1. Comparatively, the expert group rated Aesthetics highest at 4, followed by Engagement at 3.4, Information at 3.33, and Functionality at 3.21. Notably, the mean total score of the app was higher for the Customer group than the Expert group. This delineates a divergence in the app's overall quality perception between the two user categories.

The following sections show the uMARS ratings of every parameter in the subdomain, along with the application's subjective quality and perceived impact. Each domain and its subdomain are explored in detail with the analysis.

Table 4: uMARS rating of the app with a scale of 1 to 5(1-Low and 5- High)

uMARS parameter	**Overall Mean Rating (std dev.)**	**Customer Mean Rating(std dev.)**	**Expert Mean Rating(std dev.)**
Engagement	3.75 (2.65)	4.1 (1.62)	3.4 (2.10)
Functionality	3.71 (2.12)	4.21 (1.20)	3.21 (1.75)
Aesthetics	4.25 (1.56)	4.5 (1.15)	4 (1.05)
Information	3.92 (2.09)	4.5 (1.17)	3.33 (1.73)
Total Score	3.91 (0.21)	4.33 (0.18)	3.49 (0.31)
Subjective Quality	3.06 (3.31)	3.25 (2.26)	2.88 (2.41)
Perceived Impact	3.81 (2.49)	4.36 (1.62)	3.25 (1.89)

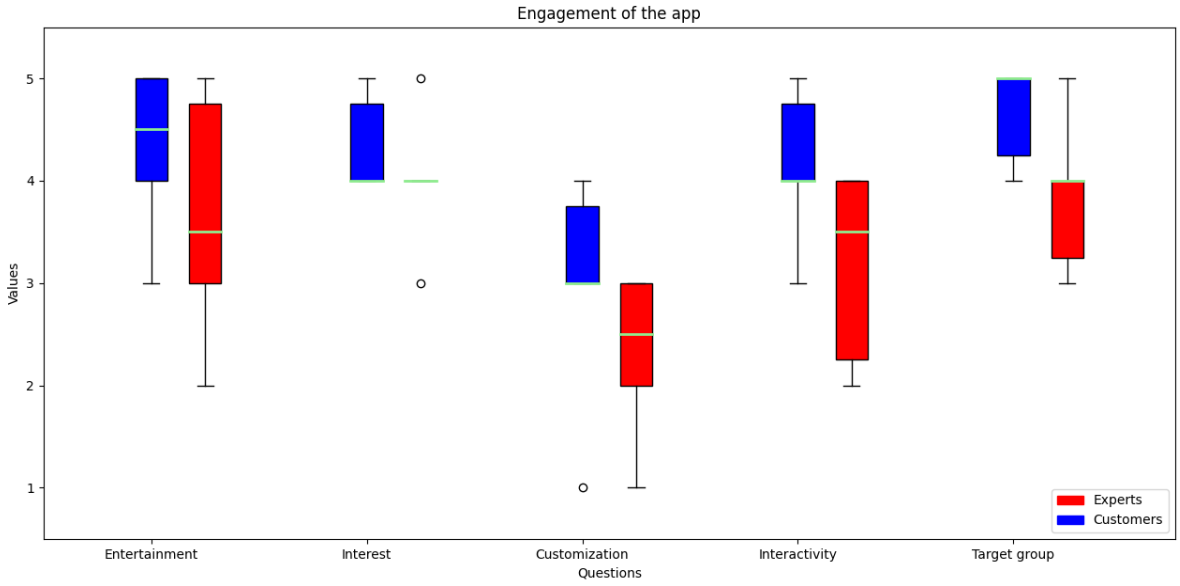

Figure 6: Engagement of the app.

6.2.1 Engagement

Figure 6 illustrates the overall engagement levels within the app, comparing both the customer and expert groups. Examining the entertainment subdomain, the median rating for entertainment among the customer group notably surpasses that of the expert group. The broader width of the box for expert entertainment suggests a more diverse range of opinions among experts. Although there is a slight overlap between the boxes, indicating some shared thoughts, there are discernible differences in entertainment ratings

between the groups, marked by negative skewness.

Regarding interest, both the customer and expert groups showcase almost equal median levels, suggesting similar levels of satisfaction in this domain. However, outliers within the expert group signify stronger opinions, while the spread indicates varied opinions among the customer group. Notably, the positive skew in the customer group implies a higher tendency for customer satisfaction regarding their interest in the app.

Customers and experts display similar median satisfaction levels in the customization domain, indicating comparable contentment. However, the lack of overlap between the boxes signifies a notable opinion discrepancy between the two groups, suggesting a broader range of perspectives. An outlier within the customer group illustrates an extreme viewpoint regarding customization. In contrast, the negative skew in the expert group suggests a lower level of satisfaction in this domain.

Regarding interactivity, the median in the customer group slightly exceeds that of the expert group, suggesting a higher level of satisfaction among customers in the app's interactive aspects. Similar to customization, the lack of overlap between customer and expert groups indicates differing opinions. Negative skewness is observed in both groups, implying lower overall satisfaction with the app's interactivity.

When considering the app's alignment with the target audience, the customer group perceives it as specifically tailored for their demographic, while experts hold a slightly differing opinion. Positive skewness in the expert group suggests that most experts moderately agree with the app's alignment with the target group.

6.2.2 Functionality

Figure 7 displays the overall functionality ratings provided by both the expert and customer groups. The median functionality rating from the customer group surpasses

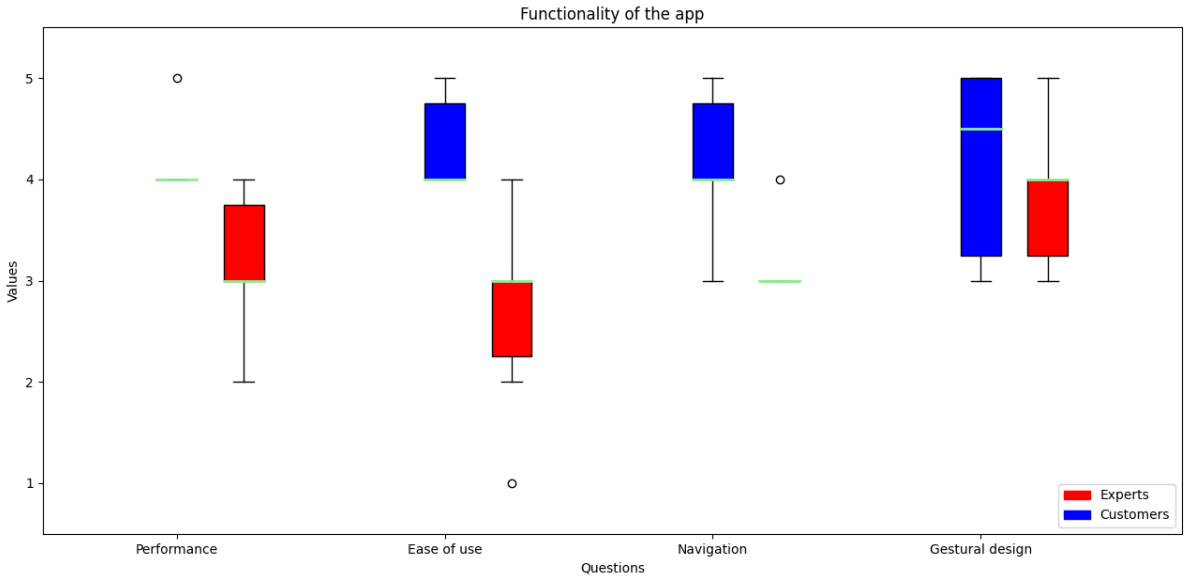

Figure 7: Functionality of the app.

that of the expert group, signaling higher satisfaction levels with the app's performance among customers. An outlier above the mean in the customer group indicates an inclination towards a higher perceived app performance. Conversely, the negative skew in the expert group's box plot suggests a slightly lower level of satisfaction with the app's performance among experts.

Regarding usability, the customer group found the app easier to use, displaying higher satisfaction than the expert group. An outlier below the mean in the expert group highlights strong opinions about the app's ease of use. The positive skew further signifies varying views among the experts regarding the app's ease of use.

The analysis of navigation satisfaction between the customer and expert groups highlights a notably higher satisfaction level within the customer group. However, negative skewness within the customer group implies that most customers rated the app's navigation as easy, although with varied opinions. An outlier above the mean in the expert group indicates strong opinions and higher satisfaction levels regarding the app's navigation.

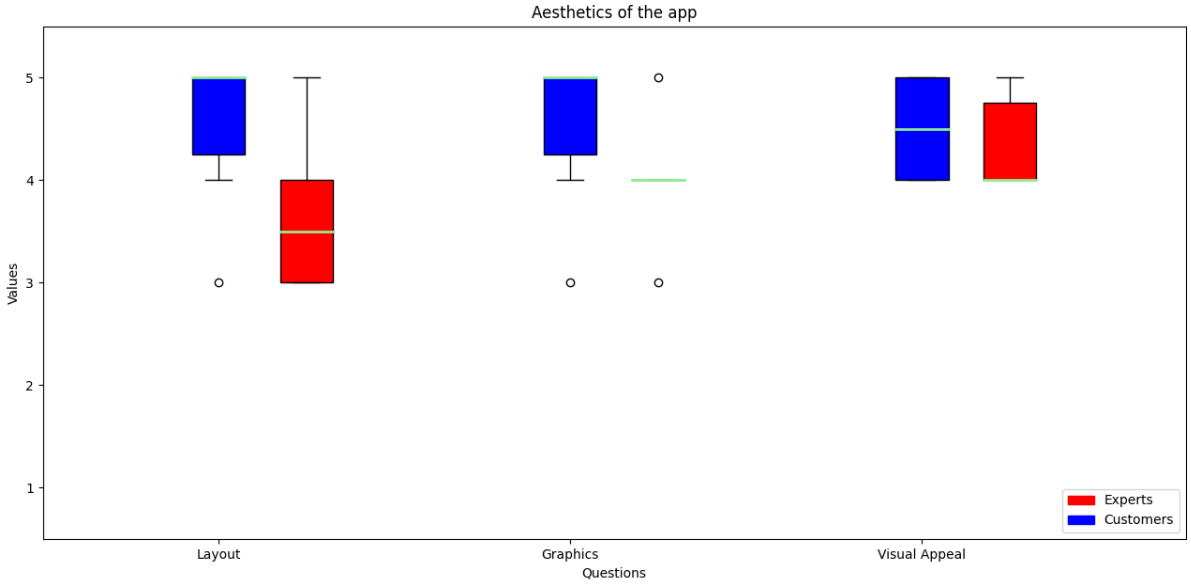

Figure 8: Aesthetics of the app.

Regarding gestural design, the customer group demonstrates a notable difference of opinion, with their median rating surpassing that of the expert group. There is significant variability in views about the gestural design of the app among the customer group. Meanwhile, the positive skew within the expert group indicates that a higher proportion of experts moderately rated the gestural design positively.

6.2.3 Aesthetics

Figure 8 depicts the ratings of the app's aesthetics from both customers and experts. The customer group showcases notably higher satisfaction with the app's layout than the expert group, with a significantly higher median. Strong opinions and more positive ratings were observed within the customer group, although most inclined towards better ratings.

Similar trends are observed in the appraisal of the app's graphics, with the customer group indicating a stronger preference than the expert group. However, extreme opinions

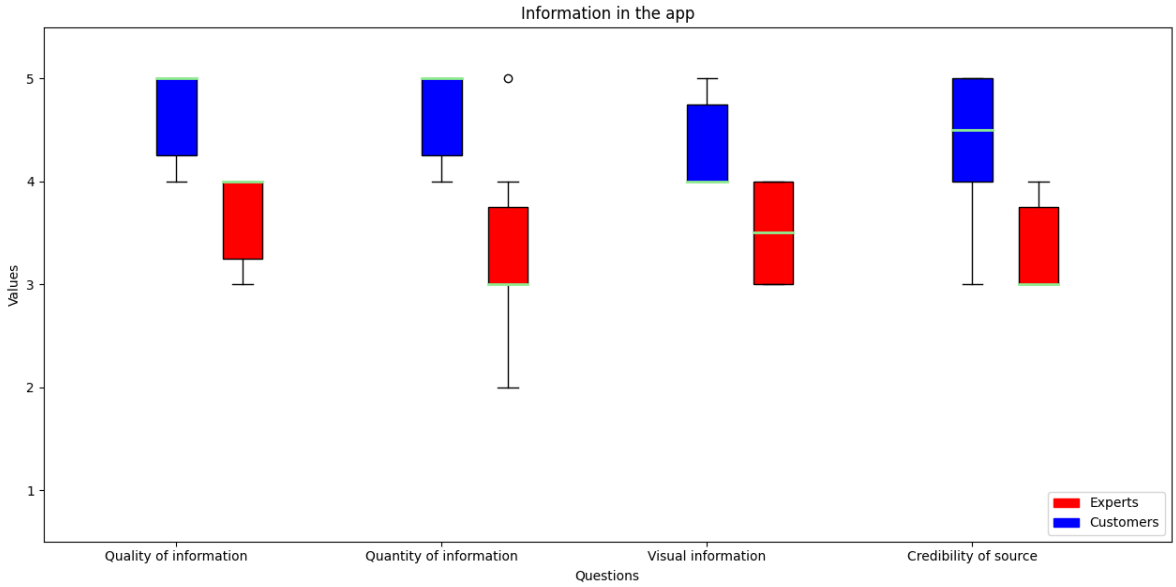

Figure 9: Information in the app.

regarding the graphics were noted among some experts and customers. Positive skewness among the customers represents varied opinions about the app's graphics.

When considering the visual appeal, both customers and experts rated it positively. Symmetry in the box plot of customers denotes a consistent opinion about the visual appeal. At the same time, the expert group shows a positive skew, suggesting that very few rated it as exceptional.

6.2.4 Information

The ratings regarding the app's information are depicted in Figure 9 from both the customer and expert groups. The customer group displayed a stronger inclination towards the high quality of the information in comparison to the expert group. There is a lack of overlap between the boxes, signifying a slight opinion difference. Both groups exhibit a positive skew, suggesting a higher number of ratings in the lowest quartile for the quality of information in the app.

A noticeable disparity exists in the ratings for the quantity of information between the customer and expert groups. Customers found the information comprehensive and concise, while the experts indicated a moderate level of comprehensiveness and conciseness. An outlier in the expert group signifies extreme opinions about comprehensive and concise information. The customer group shows a positive skew, while the expert group demonstrates a negative skew, indicating that most experts believe the information in the app is concise and comprehensive.

The visual information ratings between the expert and customer groups reveal a slightly lower median in the expert group compared to the customer group. While the expert group shows symmetry in their ratings, indicating consistent variability, there is no overlap between the groups, suggesting a slight difference in opinion. The customer group displays a positive skew, signifying a higher number of customers in the lower quartile for ratings of visual information.

In terms of the credibility of the information source, substantial differences emerge between the customer and expert groups. Customers believe the source is highly credible, while experts rate it moderately credible. The absence of overlap between the groups indicates a complete disparity in opinions. The expert group's positive skew aligns with the majority position in the lower quartile. In contrast, the customer group displays a negative skew, with most providing high ratings for the credibility of the information source.

6.2.5 Subjective Quality

The customer and expert groups display some differences in the subjective quality domain as shown in Figure 10. For recommending the app, customers lean more toward suggesting it to others than the expert group. Although experts show a wide range of

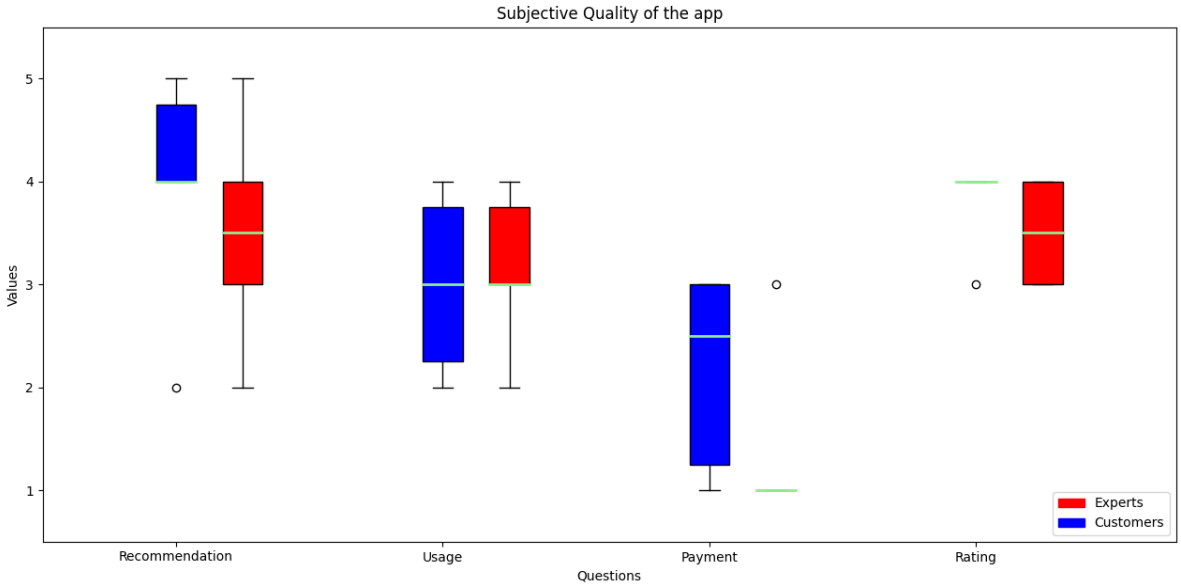

Figure 10: Subjective Quality of the app.

balanced opinions, there is a significant disparity between the groups with no overlapping data. An outlier in the customer group represents an extreme opinion against recommending the app. However, the positive skew among customers indicates that most are inclined toward recommending the app to more people.

Regarding the app's usage, both groups rate it similarly, sharing the same median. The boxes overlap, reflecting common opinions between the upper quartile of customers and all the experts. While customer ratings are symmetrically spread, the expert group exhibits a negative skew, suggesting a majority leaning toward using the app around 3 to 10 times a year.

The expert group was reluctant to pay for the app, believing that mental health wellness apps should be freely available. An outlier among the experts represented an extreme opinion against paying. However, among the customers, there is a wide range of views, reflected by the extensive box spread ranging from 1 to 3. The positive skew among the customers shows that the majority rated the willingness for payment in the lower quartile.

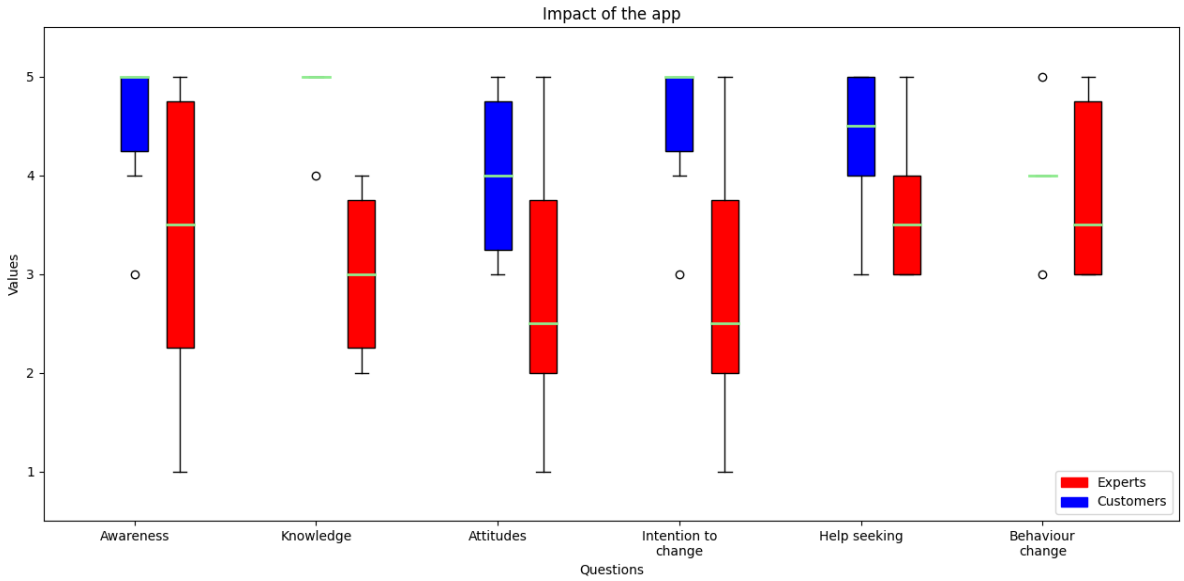

Figure 11: Perceived Impact of the app.

Regarding the app's star ratings, most customers leaned towards a 4-star rating out of 5, with a minority providing a 3-star rating—illustrated by an outlier at the bottom. Contrarily, the expert group exhibited a symmetric spread of ratings, indicating a balanced range of opinions, with a mean rating of 3.5 out of 5.

6.2.6 Perceived Impact

In the perceived impact on user knowledge, attitudes, and intentions, displayed in Figure 11, the customer group indicated a higher awareness of health behavior than the expert group. The spread in the expert group's ratings displays diverse opinions, which is evident from the widespread box. An outlier exists in the customer group, signifying a strong belief about a moderate level of awareness. Most customers strongly agreed that the app increased their knowledge of health behavior, although an outlier denotes an extreme opinion within the customer group. In contrast, the expert group shows varied opinions rather than strong agreement on increased knowledge. Experts have a more

balanced range of views when rating knowledge increment.

Both customer and expert groups showcased balanced yet slightly different opinions on the attitudes parameter. The customer group leaned toward change in attitudes compared to the expert group. For intentions to change, the expert group displayed more balanced opinions on the app's potential to influence intentions. In contrast, the customer group strongly agreed that the app could change their intentions regarding health behavior. Extreme opinions in the customer group indicated neutrality in the app's ability to affect intentions. Both customers and experts found the app encouraging to seek further help regarding health behaviors. The customers strongly agreed, while the experts predominantly agreed. This was represented by a positive skew among experts and a negative skew among customers, indicating their opinions. The customers largely decided that the app leads to behavior change, with some expressing neutral or strong agreement. The expert group showcased diverse opinions, although the majority leaned towards agreement on behavior change due to the app.

6.3 Content of Meditation Activity

Table 5 illustrates the participants' varying familiarity with mobile health applications and meditation practices. The level of familiarity that participants have with mobile health applications and meditation is assessed using a comprehensive rating scale that spans from 1 to 5. Each numerical value on this scale corresponds to different degrees of familiarity. A rating of 1 indicates a high level of unfamiliarity, while a rating of 2 suggests partial unfamiliarity. A rating of 3 represents a neutral standpoint, indicating neither familiarity nor unfamiliarity. On the other hand, a rating of 4 signifies a moderate level of familiarity, and a rating of 5 reflects a high level of familiarity.

Based on this detailed evaluation, participants are then categorized into two distinct

Table 5: Familiarity with Mobile Health Applications and Meditation practices (Scale of 1 to 5 1-not familiar and 5-highly familiar.

Participant code	**Familiarity with mobile health applications**	**Familiarity with Meditation practices**
C1	2 - Moderately unfamiliar	2 - Moderately unfamiliar
C2	4 - Moderately familiar	3 - Neither familiar nor unfamiliar
C3	3 - Neither familiar nor unfamiliar	4 - Moderately familiar
C4	4 - Moderately familiar	3 - Neither familiar nor unfamiliar
C5	3 - Neither familiar nor unfamiliar	4 - Moderately familiar
C6	4 - Moderately familiar	3 - Neither familiar nor unfamiliar
E1	5 - Highly familiar	3 - Neither familiar nor unfamiliar
E2	1 - Highly unfamiliar	4 - Moderately familiar
E3	3 - Neither familiar nor unfamiliar	5 - Highly familiar
E4	5 - Highly familiar	2 - Moderately unfamiliar
E5	2 - Moderately unfamiliar	5 - Highly familiar
E6	4 - Moderately familiar	4 - Moderately familiar
Mean	3.33	3.67
Std. Dev.	1.12	0.94

groups based on their level of familiarity with meditation activity: novices and proficient users. Novices are identified by ratings of 3 or lower, indicating a range of unfamiliarity to neutral familiarity. In contrast, proficient users are characterized by ratings higher than 3, encompassing ratings of 4 and 5. The rationale behind classifying participants with ratings above 3 as proficient users is rooted in the expectation that individuals who position themselves within these higher categories are likely to possess a heightened familiarity with meditation practices. This method ensures a nuanced differentiation, effectively capturing the varying levels of familiarity within the participant cohort. The mean and standard deviation for familiarity with mobile health applications are 3.33 and 1.12, respectively. Similarly, the mean and standard deviation are 3.67 and 0.92 for familiarity with meditation practices. Based on their ratings, the participants are segregated into two groups: The novice group (C1, C2, C4, C6, E1, and E4) and the Proficient group (C3, C5, E2, E3, E5, and E6).

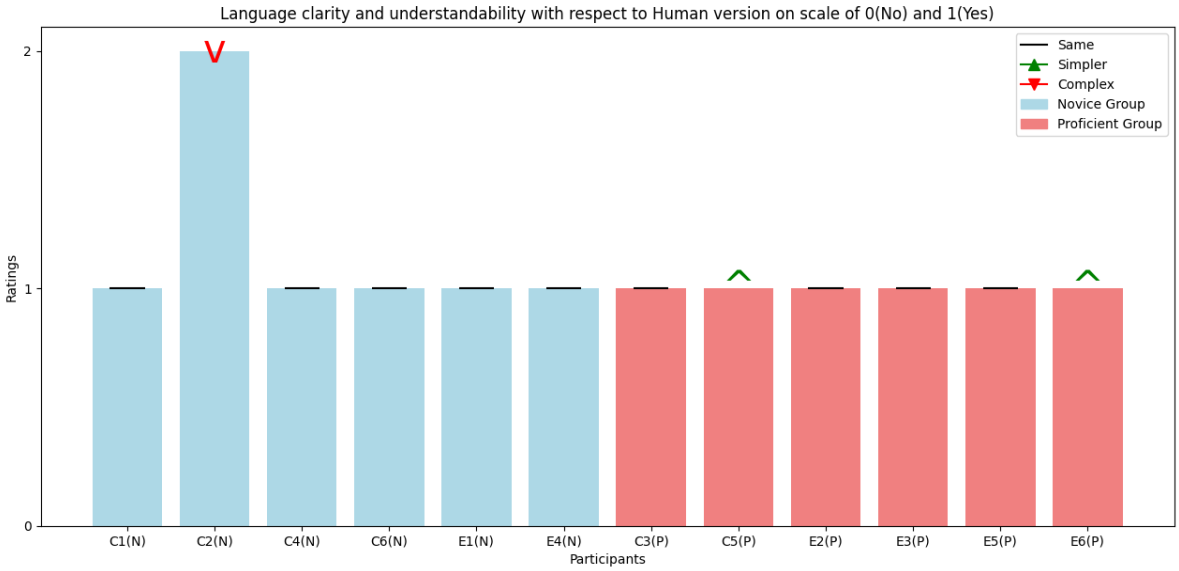

Figure 12: Language clarity and understandability

6.3.1 Perceptions on Content

Evaluating the content produced by a human expert and ChatGPT involves examining users' perceptions based on their expertise, focusing on the audio's voice and the script. The consistent use of the same voices across the app involves scripts being read by AI. In Table 6, various parameters concerning the voice of the meditation activity are depicted with participant ratings. The novice and proficient groups generally share similar perspectives on the voice utilized in the AI and human-expert versions of the activity. For the human-expert version, characteristics like comfort and calmness receive high satisfaction ratings from both novice and proficient groups. Comparatively, the ratings for the AI version show a parallel level of satisfaction for comfort and calmness across both groups. Yet, a slight disparity emerges between the groups regarding the pace of the AI version. 100% of the proficient group and Novice group expressed satisfaction with the pace of meditation. The consistency in voice aims to provide a similar experience across versions, directing attention to the content. Refer to Figure 12 for a visual repre-

Table 6: Perceptions of content - voice

Parameter	**% comfortable with AI version (N=12)**	**% comfortable with Human expert version(N=12)**
Novice Group		
Comfort with vocal tones	100%(6)	100%(6)
Calmness of voice	100%(6)	100%(6)
Pace of meditation	100%(6)	100%(6)
Proficient Group		
Comfort with vocal tones	100%(6)	100%(6)
Calmness of voice	100%(6)	100%(6)
Pace of meditation	100%(6)	100%(6)

sentation of the user's perception of language clarity and understandability with respect to human-generated content. It can be observed that one from the Novice group found the language clarity and understandability to be clearer on the human version while two from the proficient group found AI content to be clearer.

To assess user experience, we've explored factors such as ease of understanding, engagement, and structure of the content. Bar graphs have been employed to visualize these content-related characteristics. In Figure 13, the users' perception of the ease of understanding between the human expert and AI versions is represented, segmented by novice and proficient users. When comparing novice and proficient user's ratings for both human expert and AI versions, novices found the AI-generated content to be more easily understandable than human expert-generated content. However, proficient users deemed both human and AI versions easy to understand. For the human expert version, the overall ease of understanding averaged 91.7%, with individual ratings of 83.3% from novices and a full 100% from proficient users. Regarding the AI version, both novices and proficient users rated the ease of understanding 100%, resulting in an overall score of 100%.

Both novice and proficient users perceived AI-generated content as more engaging

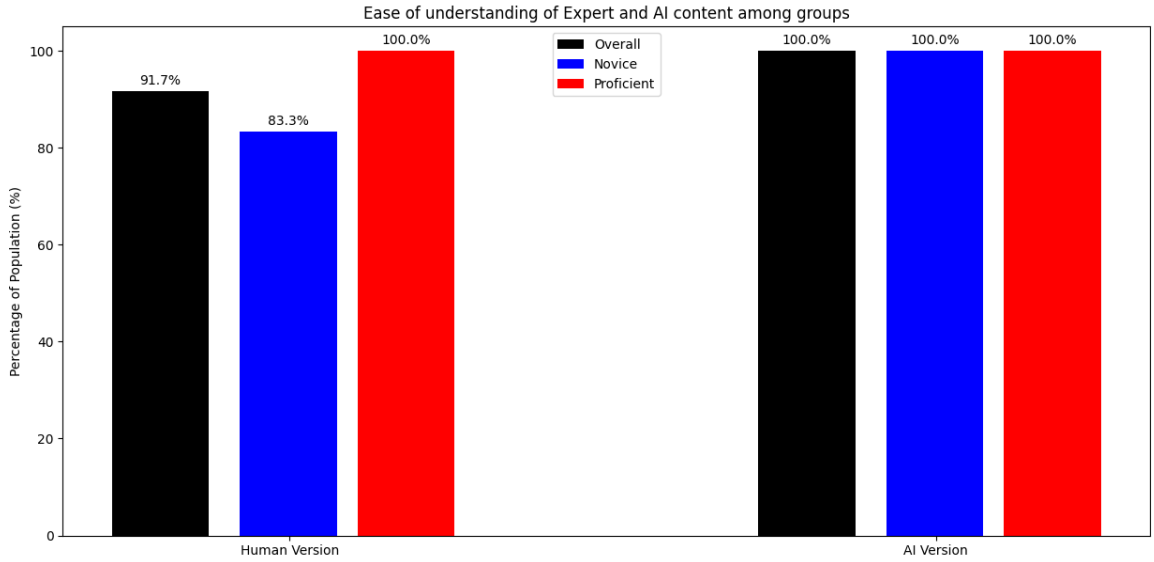

Figure 13: Ease of understanding of meditation content

than human expert-generated content. Among the novice users, 66.7% found the expert-generated content engaging, while 83.3% of the proficient users found the human expert-generated content engaging. Regarding the human expert-generated version, the overall rate of engagement reached 75%, comprised of 66.7% from novices and 83.3% from proficient users. In contrast, the overall engagement with the AI-generated content was notably higher, averaging 91.7%, with 83.3% from novices and 100% from proficient users. All proficient users found the AI-generated content engaging. Refer to Figure 14 for a more comprehensive view of the user's perception on engagement of content in the meditation activity.

From Figure 15, we observe that both novice and proficient users rated the content as structured, regardless of whether it was AI-generated or human expert-generated. The structuredness of content was equally rated by proficient and novice users, yielding a perfect 100% rating for human expert-generated content, reflecting 100% agreement from both proficient and novice users. Similarly, the AI content received a 100% rating, contributed by both novice and proficient users.

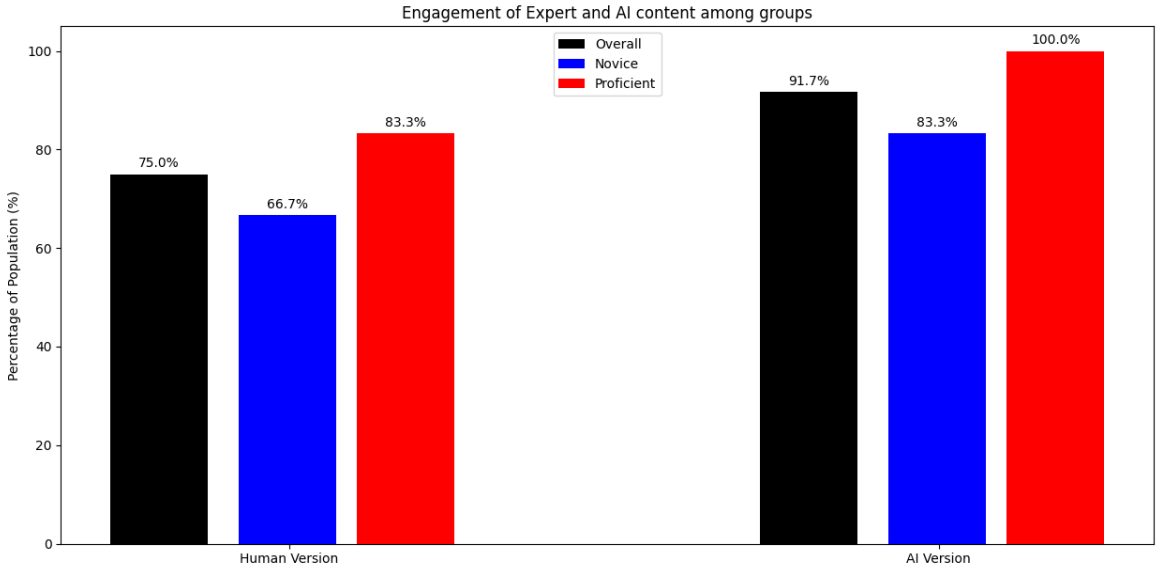

Figure 14: Engagement of meditation content

Overall, both novice and proficient users consistently perceived the characteristics related to voice. Regarding content-related attributes, novices preferred AI-generated content over human expert-generated content across all aspects. Proficient users found the ease of understanding and the structuredness of content to be similar in both AI and human expert versions. However, proficient users favored AI-generated content for its increased engagement.

6.3.2 User Experience of Content

Figure 16 illustrates that novice users slightly prefer AI-generated content to human expert-generated content, while proficient users rated both versions equally. The proficient users had consistent opinions about human expert-generated and AI-generated content, while the novice users had slightly different views on AI-generated content than human expert-generated content. In Figure 17, we can observe the opinions of novice and proficient users regarding AI-generated and human expert-generated content. The

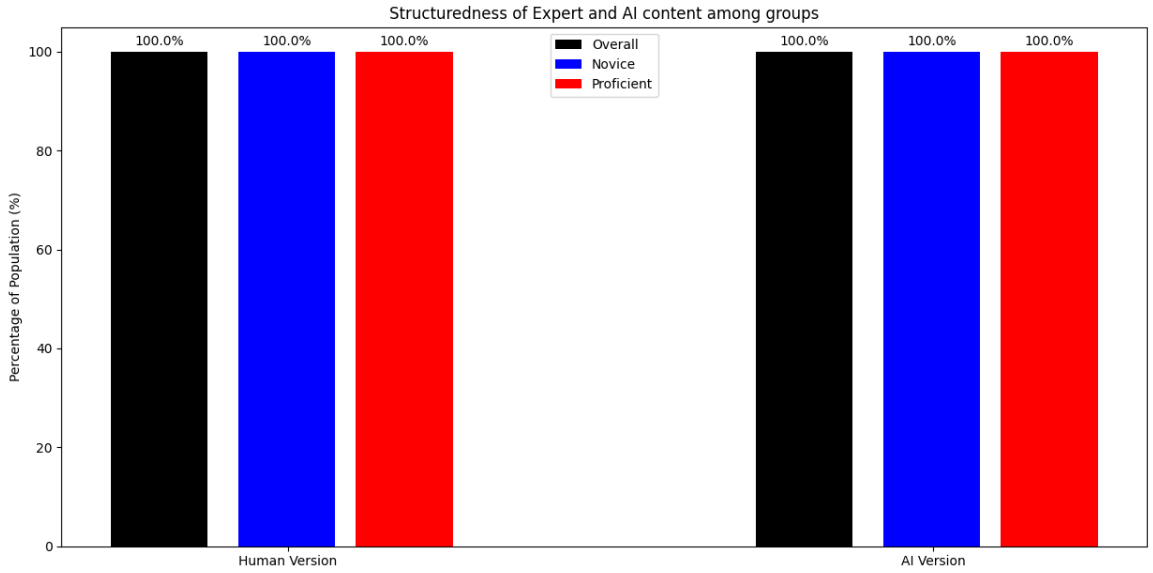

Figure 15: Structuredness of meditation content

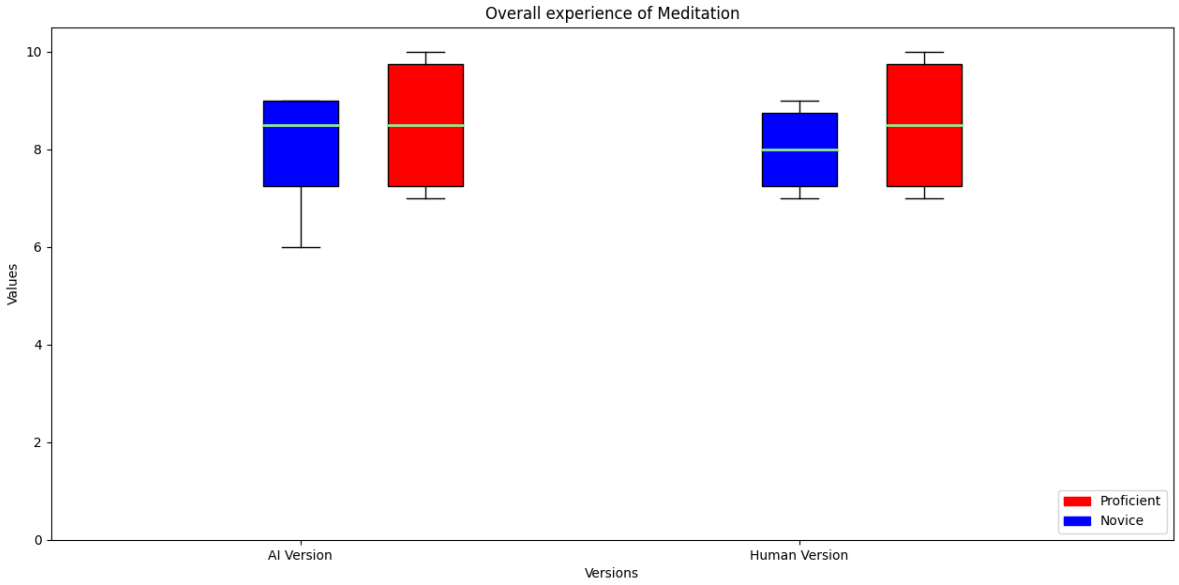

Figure 16: Overall meditation experience - Novice Group vs Proficient

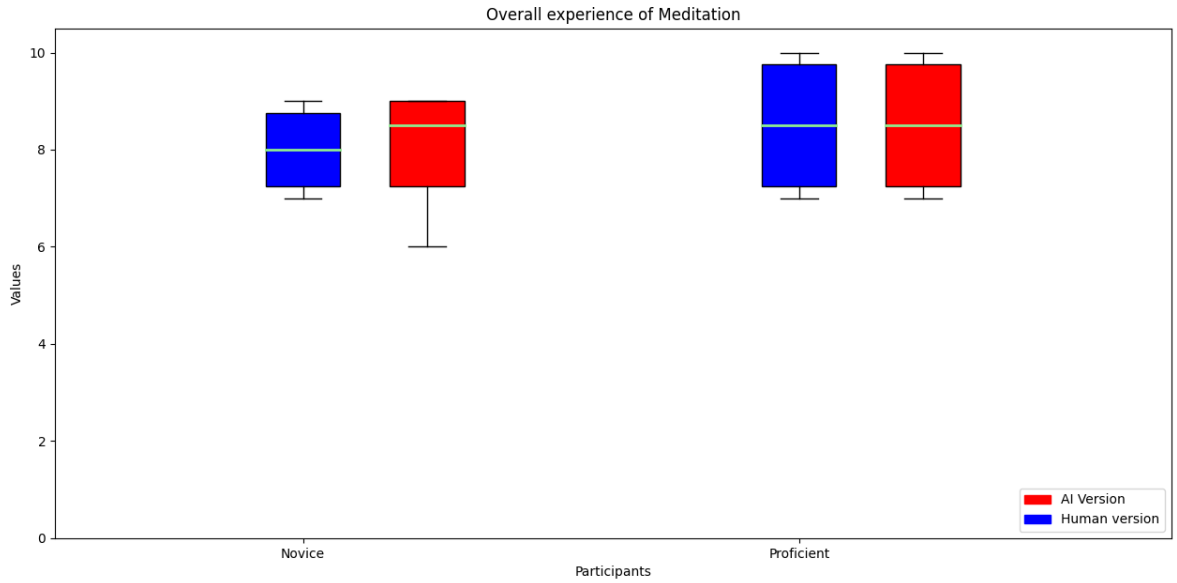

Figure 17: Overall meditation experience - AI content vs Human Expert content

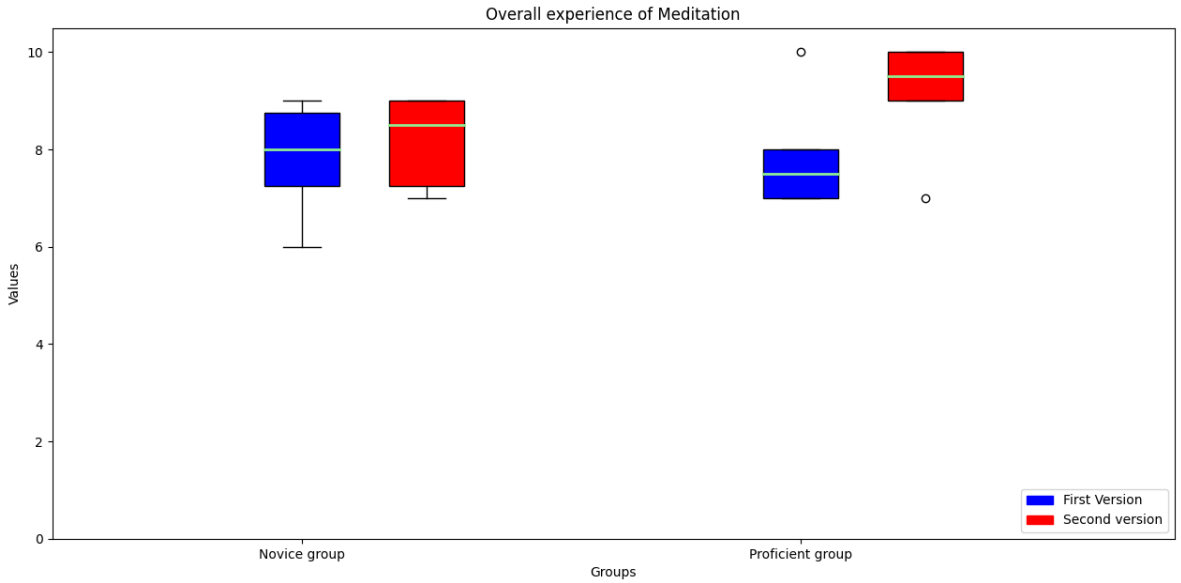

Figure 18: Overall meditation experience - First Version vs Second Version

proficient and novice groups rated the AI version equally, but the proficient group had a slight inclination toward human expert content compared to the novice group. The novice group displayed varied opinions on AI content, with the majority in the lower quartile, while the proficient group had a balanced set of views. Both proficient and novice groups had balanced opinions of human expert content, with proficient groups showing a wider range and novices having a slightly narrower range.

To assess the overall meditation experience of two versions, human and AI, participants were exposed to each version randomly. It is crucial to determine if participants consistently favor the second version in order to evaluate potential biases in preferences. The utilization of box plots allowed for the visualization and analysis of patterns of differences in rankings for both versions. Figure 18 depicts that both novice and proficient users have a higher median for the second version compared to the first version. There has been a similar range of opinions on the first versions for the novice group whereas they tended towards a higher end for the second version. For the proficient group, outliers exist for both versions, indicating strong opinions on the versions. The range of opinions remains equally distributed among the proficient group for both the first version and the second version. Some quotes from the qualitative data that support this are as follows:

Human-generated content:

C1: "I felt like calming down. I was able to visualize the different seasons and imagine the nice scenery and community aspects."

C2: "My emotions or state of mind are impacted in a better way compared to the previous version(AI version). Because it focuses more on the happy memories instead of just tracking the lifecycle of corn."

C3: "I like the content of this one(Human), it feels more personal. My emotions or state of mind are impacted in a slightly better way compared to the previous version(AI version). I find this(Human) a little bit better because it feels more personal and then

more reflecting."

C4: "The meditation content impacted me greatly. It really made me calm and thankful for my surroundings and all of life."

C5: "I would say I feel very relaxed right now."

C6: "My emotions or state of mind are impacted in a little worse way compared to the previous version(AI version). I didn't feel as relaxed in this one(Human). They(Meditation) didn't tell me to close my eyes. I was just kind of sitting there. I think that was the biggest difference. It felt like I was just kind of listening."

E1: "The human version is more introspective. The set of seasons made me go back into my memories. It makes me want to take a nap. It's a good thing to realize your energy levels are low."

E2: "My emotions or state of mind are impacted in a better way compared to the previous version(AI version). I like it a lot better."

E3: "The meditation ended kind of abruptly. The concept of belly breathing was brought up, but it wasn't actually taught. My meditation experience is the same, both versions had their pros and cons."

E4: "I like the content of this one(Human version), it feels more personal. I do like this version better, primarily because most of them except for Winter start with a directed breathing exercise."

E5: "The content(Human) was good. Very simple and straightforward. I enjoyed the meditation. I wouldn't say I was lost during the session. But there were times when I kind of forgot that I was meditating. It impacted my state and my emotions to some degree. It brings me back to living there in my home place. This is kind of what my home looks like. I got really into this version. I felt like the human one might have had more cultural significance."

E6: "I really liked the content of this version(AI-generated version), and they're all

engaging."

AI-generated content:

C1: "I really liked the content this one(AI version). I am fully immersed in the scene that they're imagining. I like the experience of comparing your own self to another life form like a plant."

C2: "I feel so much better after doing the activity(AI version). It didn't really affect me too much because I didn't really have any negative feelings coming in here. They(Meditation) were talking about gratitude and stuff and were trying to correlate that with crops. I thought that was pretty nice."

C3: "I was kind of following a storyline throughout the activity. I like that it is very calming. I liked that I was going through the seasons. I felt really relaxed during it. The voices are relaxing, felt long enough that I can actually enter that state."

C4: "My emotions or state of mind are impacted in a little better way compared to the previous version(Human). It made me feel really relaxed. I love self-reflection, and it(meditation) really does that. I think it was really nice. Closing your eyes and being able to visually picture was nice. I just liked that it reminded you to be really thankful for your people and your surroundings."

C5: "Content-wise, it(AI content) was better. It was really helpful when they said, now breathe, close your eyes, and then get to the part where you envision where you're supposed to be. And then at the end, it says, open your eyes when you are ready because the meditation is about to end."

C6: "I felt more relaxed. They(Meditation) told me to close my eyes in this(AI) one, and I like it. I feel a lot calmer. I feel like I needed that. I was not stressed, but I do not know what it is called. But I feel more relaxed. It(Meditation) was pretty good."

E1: "This one is(AI) more imaginative. This one was asking me to sort of simulate an experience in the present as if I were planting something now."

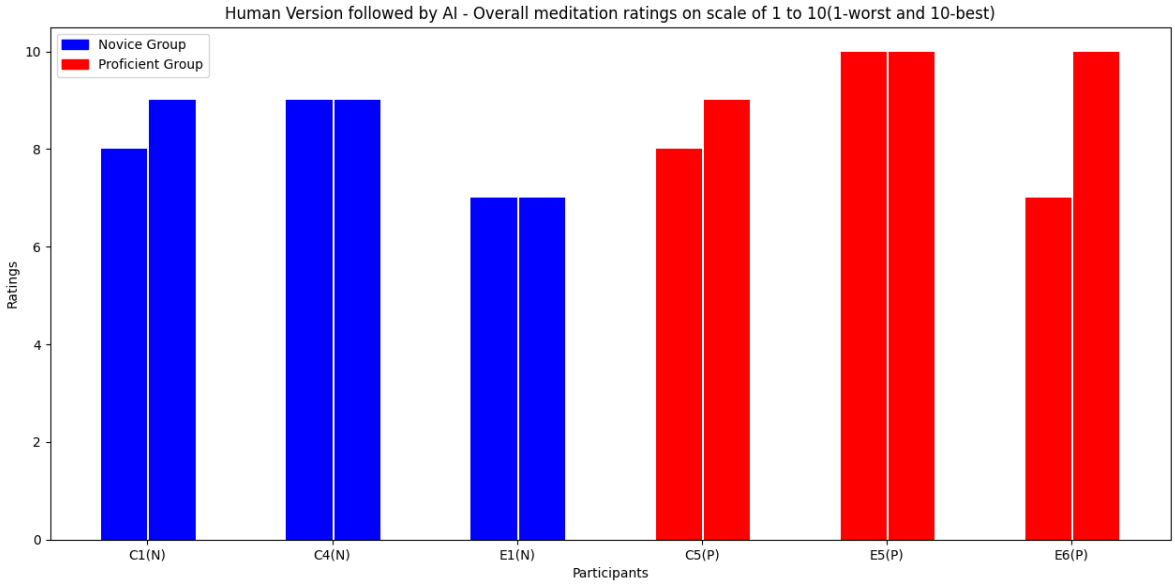

Figure 19: Human version followed by AI version

E2: "It would certainly relax me. And put me in a good space because it's nature-oriented. I could do just one of those(versions) and feel and probably feel a bit relaxed or all of them if I really felt really need it."

E3: "I think I came in calm and I remained calm. It didn't necessarily change my state of mind based on where my state of mind was prior to recording."

E4: "The content did not impact my emotions or state of mind."

E5: "Content was equally as good as the first one(Human Version). I didn't feel like I reached that same level, compared to the Human version."

E6: "I liked it(the AI version) a lot better because it was more rooted. I thought the logical structure of the content was better than the previous one. It was connected to each other a lot better. My emotions or state of mind are impacted in a better way compared to the previous version(Human). It helped me visualize how the plants are growing and how the leaves are absorbing energy. I thought it was really well done."

As we employed a method of random sampling to provide the AI version and Human version to participants, we have conducted an investigation into the responses of users

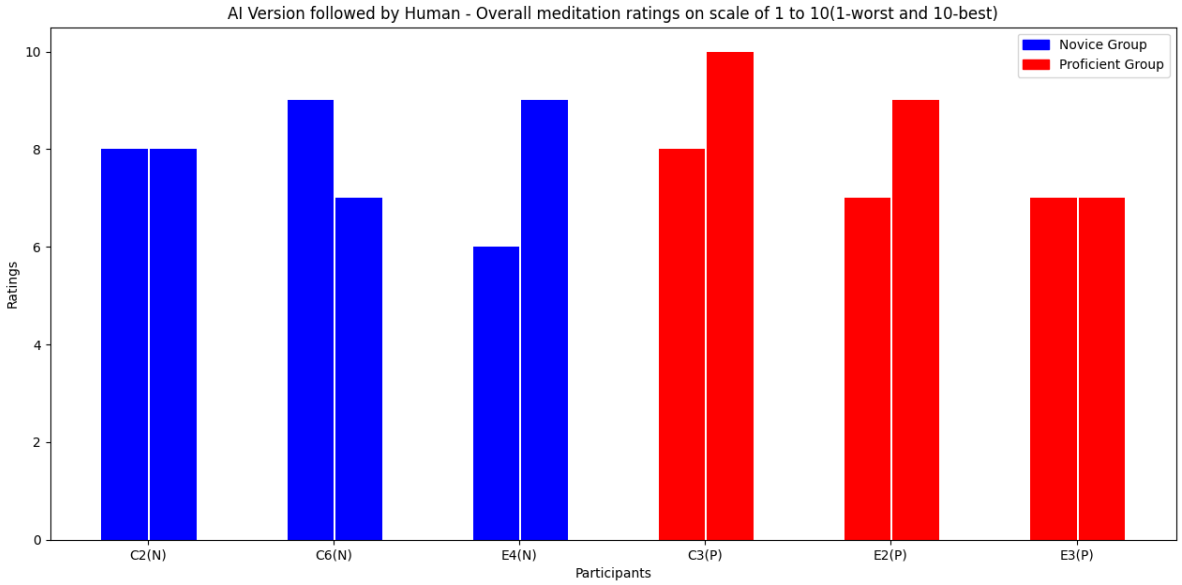

Figure 20: AI version followed by Human version

towards the order in which they were presented. The results, as illustrated in Figure 19, indicate that participants, regardless of their level of expertise, generally rated their meditation experience as either equal to or higher when they were initially exposed to the human version. More specifically, three participants maintained their initial ratings, while three participants increased their ratings for the AI version after encountering the human version first.

In contrast, Figure 20 portrays a diverse range of experiences when participants were first presented with the AI version. Two participants maintained their initial ratings, one participant decreased their rating, and three participants increased their ratings. Novice users displayed a wide range of ratings, whereas proficient users showed a stronger preference for the human version when initially exposed to the AI version.

In considering Figure 21, all novice users reported similar levels of complexity for both AI-generated and human expert-generated content. However, one of the proficient users felt that the complexity of words is simpler in the AI-generated content. Notably, the ratings for word complexity ranged from “very clear” to “moderate” in both versions.

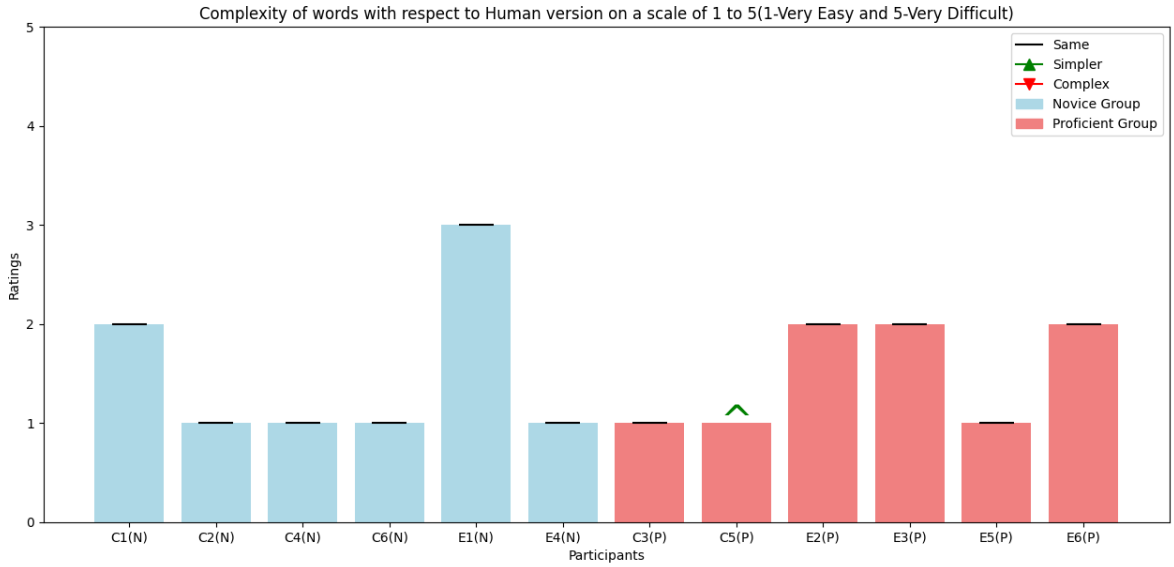

Figure 21: Complexity of words

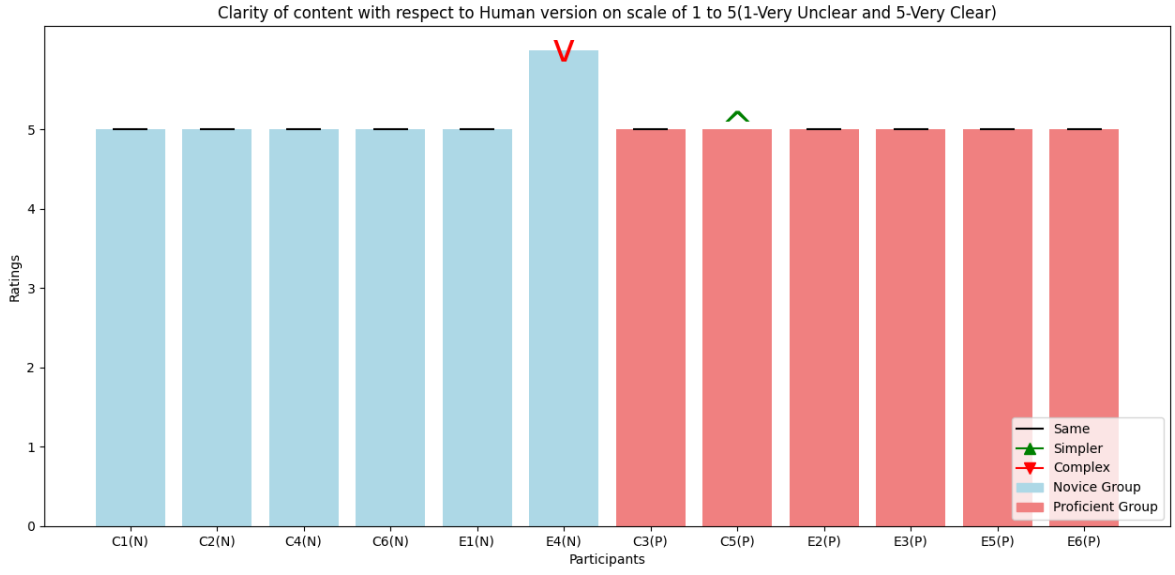

Figure 22: Clarity of content

The qualitative data related to the complexity of words are:

Human-generated content:

C2: "The complexity of words in this version is the same as the previous version. I mean the words are not very complex. The words are as simple as the previous version(AI)."

E1: "I rate the complexity of words as three out of five which is moderate."

E2: "The complexity of words in this version is the same as the previous version. But I think adding in those [Indigenous] tribe specifics was really good."

E5: "I rate the complexity of words as one out of five which is very easy. The AI version seemed more descriptive in terms of what you should be imagining."

AI-generated content:

C5: "The complexity of words in this version is simpler than the previous version. I feel like this one(AI version)was easier to understand."

E1: "The complexity of words in this version is the same as the previous version. I think the words themselves are pretty much the same, the complexity is a little bit easier because it is not asking you to remember it is asking you to imagine."

E4: "I rate the complexity of words as one out of five which is very easy. I didn't really have a problem with the clarity of this version."

About content clarity, as shown in Figure 22, most novices and proficient users rated both AI-generated and human expert-generated versions similarly. One of the novice users perceived the human-generated content to be clearer than the AI-generated content which is contradicted by a proficient user who tended that AI-generated content is clearer. Across participants, the clarity ratings ranged from "clear" to "very clear." Supporting statements about the clarity of content are:

Human-generated content:

C2: " The content is just as clear as the previous version(AI)."

E4: "The clarity of content in this version is better than the previous version. I mean, it's more directed."

AI-generated content:

C5: "The clarity of content in this version is better than the previous version(Human). I think it was more clear with the instructions like, when to start breathing and when the meditation was going to end was really helpful."

E2: "I rate the clarity of content as four out of five which is clear. I mean, because they're using universal things (seasons). I don't give it a five that it's extremely clear because I think those things we've talked about the I'm trying to think if I'm a [Indigenous] youth, there was some language in there, the cadence, the things that can be brought about with the [Indigenous] Language integrated into it."

Analyzing the impact of content involves categorizing it as positive, neutral, negative, or mixed, employing an inductive approach for sub-theme identification. Specific sub-themes emerged within the "positive" category, including emotional state, gratitude, and reflection. Participants' statements under the "emotional state" sub-theme, expressing feelings of being engaged, calm, relaxed, and emotionally connected, contribute to this category. Comments from the participants supporting these perspectives include:

C3: "I like that it(the meditation session) is very calming"

C5: "The meditation content has impacted me positively. I feel very nice."

E5: "I really enjoyed it(the meditation session)."

E1: "I think the last one(meditation session) made me go back into my memories and relive my childhood."

In the "gratitude and reflection" category, codes include "mindful", "inner reflection", and "thankfulness". Here are statements from participants that exemplify mindful enjoyment:

C3: "I find it(The meditation session) a little bit better because it feels more personal

and then more reflecting and thinking."

C4: "It(The meditation session) really made me calm and thankful for my surroundings and all of life."

E6: "It(Meditation session) helped me be more present in a moment and think about and not think about what I have to do the rest of the day."

Under the "negative" code, two specific categories have been identified: "not relaxed" and "depressing." Here are the supporting statements for these categories:

E4: "The speaker talked quite a bit about regeneration and so forth in the first half. It was depressing."

C6: "I didn't feel relaxed in this one(meditation session)."

Under the code "neutral", we have the codes such as no change. The supporting statements are:

E3: "It(The meditation session) didn't change my state of mind."

C2: "It(The meditation session) didn't really affect me too much."

The examination of the influence on mood reveals distinct patterns in the user experience. Initially, when considering the content generated by human experts, four proficient users and four novice users displayed positive responses. In contrast, only one novice user exhibited a negative reaction to the content generated by human experts. Shifting the focus to the content generated by AI, four proficient users responded positively, while one user remained neutral. Among the novice users, three participants demonstrated a positive impact, one had a negative reaction, and another responded neutrally.

In the broader study that involved 12 participants, only a subset of individuals provided comprehensive details regarding the impact of the content, resulting in a partial representation of the data. Out of the 12 participants, 8 expressed a positive impact on the content generated by humans, while one participant reported a negative impact. Three participants did not provide extensive feedback. For the content generated by AI,

7 out of 12 participants reported a positive impact, one reported a negative impact, and two remained neutral. Two participants did not elaborate on the impact of their mood due to their engagement in meditation activities. Proficient users consistently displayed a positive impact for both versions, with no observed negative effects. On the other hand, novice users tended to exhibit a more positive impact on the human-generated version compared to the AI-generated version. Interestingly, both versions equally influenced novice users negatively. Additionally, one proficient user and one novice user reported a neutral impact from the AI-generated version. This comprehensive assessment underscores the nuanced emotional responses within both proficiency groups, shedding light on the diverse impacts stemming from the different types of content.

Chapter 7

Conclusion & Future Work

The SUNRISE app has received a positive assessment with an overall rating of 3.93 out of 5, reflecting customers' rating at 4.1 out of 5 and experts at 3.4 out of 5. This evaluation encompasses four key factors: engagement, functionality, aesthetics, and information. Of these, the aesthetics section stands out with the highest rating of 4.25 out of 5. Studies emphasize the significance of aesthetics alongside functionality in user experience [92]. Both customers (rating 4.5 out of 5) and experts (rating 4 out of 5) have highlighted the app's aesthetics as highly satisfactory. The specific subdomains within aesthetics are individually placed in the range of 4 to 4.5, with visual appeal ranking at the top with a rating of 4.42 out of 5.

This analysis reveals the vital role of participant background in interpreting meditation script evaluations. While voice comfort, calmness, and pace are uniformly perceived across novice and proficient groups, variations emerge in understandability and language clarity. Here, a minority within the proficient group leans toward the AI-generated content's voice. Notably, parameters like ease of understanding and content engagement receive higher or equal ratings for AI-generated content than human-generated content across both proficient and novice groups. Interestingly, both groups rate content struc-

turedness at 100% across both content versions.

The expert group perceived both the AI-generated and human expert-generated content similarly, whereas novice users tended to favor the AI-generated content. Proficient users and most novice users found that the complexity of words was similar for both versions, while a minority of customers thought the human expert-generated content was simpler. Regarding content clarity, the proficient users perceived both versions as equal, but the novice group had slightly different opinions. Most customers rated the clarity as similar, although some mentioned that the AI-generated content is clearer than the human expert-generated one. Human expert-generated and AI-generated content impacted proficient and novice users positively, neutrally, and negatively. More than half of the proficient and novice groups had a positive impact from both content versions. However, the human expert-generated content had a greater effect, especially among the proficient group.

The preliminary findings that proficient practitioners of meditation rated both human expert-generated and AI-generated content equally, while novice practitioners showed a preference for AI-generated content, have significant implications. Firstly, this indicates that the acceptance and perception of AI-generated content may vary depending on the user's level of familiarity with the subject matter. Experienced practitioners, who may be accustomed to nuanced and personalized guidance, may find both sources equally effective, demonstrating the potential for AI-generated content to match human expertise in certain contexts. The preference for AI-generated content among novice practitioners suggests that the technology has the potential to provide accessible and acceptable guidance to individuals new to meditation.

Initial trends in this research significantly enhance our understanding of how users perceive and accept AI-generated content in the realm of meditation. The study emphasizes the importance of considering users' expertise levels when evaluating the effectiveness of

AI-generated guidance, offering a nuanced perspective that enriches ongoing discussions on the role of AI in promoting personal well-being. The results suggest that AI-generated content demonstrates adaptability and the potential to be inclusive in guiding individuals through meditation. Interestingly, proficient users equally appreciate both AI and human-generated content, while novice users show a clear preference for AI-generated content, highlighting the accessibility and engagement potential of this technology for newcomers to the practice. These insights can inform the development of personalized meditation tools that cater to the diverse needs and proficiency levels of users, thereby contributing to a more inclusive and effective integration of AI in the field of mental health care and well-being.

7.1 Limitations

Given the constraints of our sample size, it's important to recognize that our observations capture trends within this specific dataset, urging the need for broader exploration. Utilizing a larger sample size in the comparative analysis of AI-generated content can provide deeper insights into more resilient and representative trends, thereby bolstering statistical power and precision. It's essential to highlight the lack of substantial participation of various groups, which limits the generalizability of our findings to wider communities. In the usability study, a group of expert users was selected from a wide range of backgrounds, including app development and behavioral health interventions. The inclusion of individuals with different areas of expertise in a single category played a role in the current findings. However, it is possible that if these participants were classified based on their specific areas of expertise, different outcomes may have been obtained.

Considering that the meditation content is specifically tailored for a mental health

application, a more cautious strategy would have involved taking into account the participants' proficiency in both mental health applications and meditation practices. This alternative approach could have led to diverse outcomes and offered a more intricate comprehension of the relationship between familiarity with mental health applications, meditation practices, and the study's findings. Additionally, the insufficient representation of individuals with mental health experiences raises concerns regarding the adequacy of the sample and implies that this disparity might have influenced the outcomes.

The meticulously developed human script, tailored for a specific Indigenous community through community-based participatory design methods, exhibited a limitation in participant representation. The implementation of convenience sampling during data collection can result in different outcomes compared to those attainable under prevailing conditions. Only one participant belonged to the targeted community, possessing a nuanced understanding of the human script, while the remaining participants lacked this cultural familiarity. This discrepancy may constrain the robustness of our conclusions. Conducting the experiment with individuals from the intended Indigenous community could yield disparate outcomes. Alternatively, if the human expert had crafted the script for a broader, mainstream population, results might differ. These considerations emphasize the need for a more comprehensive exploration of contextual factors and participant characteristics to enhance the validity and generalizability of our findings.

The content generated by the AI was authored by an individual who lacks proficiency in prompt-generation and meditation practices. In addition to this, there is no contribution from a prompt-generation expert to formulate prompts for ChatGPT. This may potentially lead to suboptimal utilization of ChatGPT for generating content with expert-provided keywords. Additionally, the human expert responsible for creating the human-version scripts is not involved in generating prompts for ChatGPT, nor are they involved in editing or reevaluating the scripts generated by ChatGPT. As a result, it

is possible that AI-generated scripts may have a reduced focus on psychological aspects compared to scripts generated by humans. The ChatGPT-generated scripts' trustworthiness relies on the input prompts' effectiveness, as the prompts' attributes can exert a considerable effect on the ultimate outputs. Moreover, it's vital to acknowledge that preceding language models, like GPT-2 preceding ChatGPT, might display biases and require supervision [93].

In order to compare the content of meditation, participants were divided into two groups: novices and proficient users. This categorization was based on their own assessment of their familiarity with meditation. However, the significance of this distinction may be compromised due to the lack of comprehensive information regarding participants' previous experiences with meditation, including the duration, type, and method of their meditation practices. The levels of familiarity were evaluated by the participants themselves, which introduces a subjective element as individuals may reassess their perceived highest level of familiarity with meditation over time. This subjective self-assessment introduces variability into the study and raises concerns about the reliability and generalizability of the results. It is crucial to acknowledge and address this limitation when interpreting the findings, highlighting the fluid nature of participants' self-assessment of familiarity with meditation, which could potentially impact the overall outcomes of the study.

7.2 Future Work

The primary objective of the study was to compare the content generated by artificial intelligence (AI) with that produced by human experts in the context of integrating AI technology into mobile health applications. To ensure the accuracy of findings, it is crucial to carefully select participants who possess the necessary expertise. By examining the

usability of the applications within a specific group of experts, valuable insights into the effectiveness of the application can be achieved. Furthermore, future studies could explore the technical proficiency of participants in meditation practices, which are utilized for comparative analysis in the study. In order to further investigate this topic, it would be beneficial to establish a clear distinction between novice and proficient groups, as well as consider participants' expertise in mental health applications. This strategic approach aims to focus the study and enhance our comprehension of how participant expertise influences the effectiveness of integrating AI technology.

While the use of AI in the field of medicine is steadily increasing, its integration into mobile health apps has been relatively slow. In order to achieve this goal, the researchers employed content generated by ChatGPT and compared it with content generated by human experts using audio files based on user experiences. To further advance this research, future endeavors could involve the inclusion of human experts in the content editing process of ChatGPT is identified as a prospect to enhance the streamlining of the refinement process. By utilizing the insights and proficiency of human specialists, this approach can facilitate the production of more precise, contextually relevant, and polished content. In addition to this, the inclusion of human experts to provide iterative feedback after the AI has generated scripts. This approach would enable real-time critique and refinement of the AI-generated content, leveraging the expertise of human evaluators to enhance its overall quality. Additionally, incorporating the input of community experts in critiquing AI scripts introduces a collaborative element, allowing for diverse perspectives to evaluate and improve the content. Another avenue for future exploration is the utilization of real-time ChatGPT/AI-generated content. This entails evaluating the dynamic and evolving nature of the content as it is generated in real-time, enabling a more responsive and adaptive approach to content generation.

The research also proposes exploring the utilization of human voices in AI-generated

content and human expert-generated content. The objective is to improve the user experience and establish a stronger connection with the AI-generated content by incorporating human-like vocal characteristics. Furthermore, it is crucial to incorporate user feedback mechanisms, accessibility features, and cultural sensitivity in voice generation to ensure a user-centric approach. Collaborating with speech experts and integrating real-time voice synthesis can further enhance the auditory aspects, resulting in AI-generated content that is more dynamic, culturally conscious, and responsive to evolving health information.

The study highlights the significance of prompt-generation techniques, specifically in enhancing the effectiveness and efficiency of the content generated by ChatGPT. The key to achieving this objective lies in the skillful crafting of prompts that strategically enhance the quality of outputs. To further enrich and diversify the content produced by ChatGPT, the exploration of generating multiple prompts under the guidance of prompt creation experts is suggested as a promising approach. This collaborative method, which incorporates human expertise, aims to infuse creativity and domain knowledge into the process of content generation. These directions collectively aim to enhance the integration of ChatGPT into mobile health applications.

Appendix A

Data collection documents

A.1 Background Survey

Usability Testing Background Survey

* Required

1. In what year were you born? *

_____

2. Which best describes your gender? *

Mark only one oval.

- ◯ Female
- ◯ Male
- ◯ Prefer not to say
- ◯ Other: _____

3. Which of the following best describes your race? (Select all that apply) *

Check all that apply.

- ☐ African American or Black
- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ White
- ☐ Other: _____

4. Which ethnicity best describes you? *

Mark only one oval.

- Hispanic
- Non-Hispanic

5. How comfortable are you with using mobile applications (apps)? *

Mark only one oval.

Very uncomfortable. I have never used one before.	
1	
2	
3	
4	
5	
Very comfortable. I use them all the time.	

A.2 uMARS Survey

Instructions for use:

Raters should:

1. Use the app and trial it thoroughly for at least 10 minutes;
2. Determine how easy it is to use, how well it functions and does it do what it purports to do;
3. Review app settings, developer information, external links, security features, etc.

Scoring

A: Engagement Mean Score = _____

B: Functionality Mean Score = _____

C: Aesthetics Mean Score = _____

D: Information Mean Score* = _____

* Exclude questions rated as “N/A” from the mean score calculation.

App quality mean score _____ = A + B + C + D / 4

The *App subjective quality* scale can be reported as individual items or as a mean score, depending on the aims of the research.

The *Perceived impact* items can be adjusted and used to obtain information on the perceived impact of the app on the user’s knowledge, attitudes and intentions related to the target health behaviour.

Mobile Application Rating Scale: user version (uMARS)

App Name: _____

Circle the number that most accurately represents the quality of the app you are rating. All items are rated on a 5-point scale from "1.Inadequate" to "5.Excellent". Select N/A if the app component is irrelevant.

App Quality Ratings

SECTION A

Engagement – fun, interesting, customisable, interactive, has prompts (e.g. sends alerts, messages, reminders, feedback, enables sharing)

1. **Entertainment: Is the app fun/entertaining to use? Does it have components that make it more fun than other similar apps?**

   1 Dull, not fun or entertaining at all
   2 Mostly boring
   3 OK, fun enough to entertain user for a brief time (< 5 minutes)
   4 Moderately fun and entertaining, would entertain user for some time (5-10 minutes total)
   5 Highly entertaining and fun, would stimulate repeat use

2. **Interest: Is the app interesting to use? Does it present its information in an interesting way compared to other similar apps?**

   1 Not interesting at all
   2 Mostly uninteresting
   3 OK, neither interesting nor uninteresting; would engage user for a brief time (< 5 minutes)
   4 Moderately interesting; would engage user for some time (5-10 minutes total)
   5 Very interesting, would engage user in repeat use

3. **Customisation: Does it allow you to customise the settings and preferences that you would like to (e.g. sound, content and notifications)?**

   1 Does not allow any customisation or requires setting to be input every time
   2 Allows little customisation and that limits app's functions
   3 Basic customisation to function adequately
   4 Allows numerous options for customisation
   5 Allows complete tailoring the user's characteristics/preferences, remembers all settings

4. **Interactivity: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)?**

   1 No interactive features and/or no response to user input
   2 Some, but not enough interactive features which limits app's functions
   3 Basic interactive features to function adequately
   4 Offers a variety of interactive features, feedback and user input options
   5 Very high level of responsiveness through interactive features, feedback and user input options

5. **Target group: Is the app content (visuals, language, design) appropriate for the target audience?**

    1 Completely inappropriate, unclear or confusing
    2 Mostly inappropriate, unclear or confusing
    3 Acceptable but not specifically designed for the target audience. May be inappropriate/ unclear/confusing at times
    4 Designed for the target audience, with minor issues
    5 Designed specifically for the target audience, no issues found

SECTION B

Functionality – app functioning, easy to learn, navigation, flow logic, and gestural design of app

6. **Performance: How accurately/fast do the app features (functions) and components (buttons/menus) work?**

    1 App is broken; no/insufficient/inaccurate response (e.g. crashes/bugs/broken features, etc.)
    2 Some functions work, but lagging or contains major technical problems
    3 App works overall. Some technical problems need fixing, or is slow at times
    4 Mostly functional with minor/negligible problems
    5 Perfect/timely response; no technical bugs found, or contains a 'loading time left' indicator (if relevant)

7. **Ease of use: How easy is it to learn how to use the app; how clear are the menu labels, icons and instructions?**

    1 No/limited instructions; menu labels, icons are confusing; complicated
    2 Takes a lot of time or effort
    3 Takes some time or effort
    4 Easy to learn (or has clear instructions)
    5 Able to use app immediately; intuitive; simple (no instructions needed)

8. **Navigation: Does moving between screens make sense; Does app have all necessary links between screens?**

    1 No logical connection between screens at all /navigation is difficult
    2 Understandable after a lot of time/effort
    3 Understandable after some time/effort
    4 Easy to understand/navigate
    5 Perfectly logical, easy, clear and intuitive screen flow throughout, and/or has shortcuts

9. **Gestural design: Do taps/swipes/pinches/scrolls make sense? Are they consistent across all components/screens?**

    1 Completely inconsistent/confusing
    2 Often inconsistent/confusing
    3 OK with some inconsistencies/confusing elements
    4 Mostly consistent/intuitive with negligible problems
    5 Perfectly consistent and intuitive

SECTION C

Aesthetics – graphic design, overall visual appeal, colour scheme, and stylistic consistency

10. Layout: Is arrangement and size of buttons, icons, menus and content on the screen appropriate?

1 Very bad design, cluttered, some options impossible to select, locate, see or read
2 Bad design, random, unclear, some options difficult to select/locate/see/read
3 Satisfactory, few problems with selecting/locating/seeing/reading items
4 Mostly clear, able to select/locate/see/read items
5 Professional, simple, clear, orderly, logically organised

11. Graphics: How high is the quality/resolution of graphics used for buttons, icons, menus and content?

1 Graphics appear amateur, very poor visual design - disproportionate, stylistically inconsistent
2 Low quality/low resolution graphics; low quality visual design – disproportionate
3 Moderate quality graphics and visual design (generally consistent in style)
4 High quality/resolution graphics and visual design – mostly proportionate, consistent in style
5 Very high quality/resolution graphics and visual design - proportionate, consistent in style throughout

12. Visual appeal: How good does the app look?

1 Ugly, unpleasant to look at, poorly designed, clashing, mismatched colours
2 Bad – poorly designed, bad use of colour, visually boring
3 OK – average, neither pleasant, nor unpleasant
4 Pleasant – seamless graphics – consistent and professionally designed
5 Beautiful – very attractive, memorable, stands out; use of colour enhances app features/menus

SECTION D

Information – Contains high quality information (e.g. text, feedback, measures, references) from a credible source

13. Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?

N/A There is no information within the app
1 Irrelevant/inappropriate/incoherent/incorrect
2 Poor. Barely relevant/appropriate/coherent/may be incorrect
3 Moderately relevant/appropriate/coherent/and appears correct
4 Relevant/appropriate/coherent/correct
5 Highly relevant, appropriate, coherent, and correct

14. Quantity of information: Is the information within the app comprehensive but concise?

N/A There is no information within the app
1 Minimal or overwhelming
2 Insufficient or possibly overwhelming
3 OK but not comprehensive or concise
4 Offers a broad range of information, has some gaps or unnecessary detail; or has no links to more information and resources
5 Comprehensive and concise; contains links to more information and resources

15. **Visual information: Is visual explanation of concepts – through charts/graphs/images/videos, etc. – clear, logical, correct?**

N/A There is no visual information within the app (e.g. it only contains audio, or text)
1 Completely unclear/confusing/wrong or necessary but missing
2 Mostly unclear/confusing/wrong
3 OK but often unclear/confusing/wrong
4 Mostly clear/logical/correct with negligible issues
5 Perfectly clear/logical/correct

16. **Credibility of source: does the information within the app seem to come from a credible source?**

N/A There is no information within the app
1 Suspicious source
2 Lacks credibility
3 Not suspicious but legitimacy of source is unclear
4 Possibly comes from a legitimate source
5 Definitely comes from a legitimate/specialised source

App subjective quality

SECTION E

17. **Would you recommend this app to people who might benefit from it?**

1	Not at all	I would not recommend this app to anyone
2		There are very few people I would recommend this app to
3	Maybe	There are several people I would recommend this app to
4		There are many people I would recommend this app to
5	Definitely	I would recommend this app to everyone

18. **How many times do you think you would use this app in the next 12 months if it was relevant to you?**

1 None
2 1-2
3 3-10
4 10-50
5 >50

19. **Would you pay for this app?**

1 Definitely not
2
3
4
5 Definitely yes

20. **What is your overall (star) rating of the app?**

1	★	One of the worst apps I've used
2	★★	
3	★★★	Average
4	★★★★	
5	★★★★★	One of the best apps I've used

Perceived impact

SECTION F

1. **Awareness – This app has increased my awareness of the importance of addressing the health behaviour**

Strongly disagree				Strongly Agree
1	2	3	4	5

2. **Knowledge – This app has increased my knowledge/understanding of the health behaviour**

Strongly disagree				Strongly Agree
1	2	3	4	5

3. **Attitudes – The app has changed my attitudes toward improving this health behaviour**

Strongly disagree				Strongly Agree
1	2	3	4	5

4. **Intention to change – The app has increased my intentions/motivation to address this health behaviour**

Strongly disagree				Strongly Agree
1	2	3	4	5

5. **Help seeking – This app would encourage me to seek further help to address the health behaviour (if I needed it)**

Strongly disagree				Strongly Agree
1	2	3	4	5

6. **Behaviour change – Use of this app will increase/decrease the health behaviour**

Strongly disagree				Strongly Agree
1	2	3	4	5

Further comments about the app?

THANK YOU!

A.3 Comparative Analysis Questionnaire

Version 1 - Questions on meditation activity

Content - complexity of words, clarity of content, ease of understanding, impact on emotions, logical structure, overall experience,

Pre-test questions

1. How familiar are you with mobile health applications on a scale of 1 to 5. 1 being least familiar and 5 being highly familiar
2. On a scale of 1 to 5, please indicate your current level of familiarity with meditation practices, with 1 being least familiar and 5 being highly familiar.
3. Are there specific features or elements you look for in meditation content?
4. How do you generally find new meditation content or apps to try?
5. Do you have any preferences in terms of meditation duration?

In-test questions

1. What was your overall meditation experience like on a scale of 1 to 10. 1 being worst and 10 being best
2. Are the vocal tones comfortable for you?
3. Is the language clear and understandable?
4. Did you find the voice to be calming and soothing?
5. How is the pace of meditation? Is it too slow or too fast?
6. Do you think the length of meditation is too short or too long?
7. Are you able to follow the instructions easily?
8. Did you find yourself lost during any point in the session? – Yes/No
9. Do you find content to be logically structured?
10. How do you rate complexity of words, on a scale of 1 to 5. 1 being very easy and 5 being very difficult
11. How do you rate clarity of content, on a scale of 1 to 5. 1 being very unclear and 5 being very clear
12. How did the content impact your emotions or state of mind?
13. Is there any season that you felt really engaging? If yes, why?
14. Do you have any suggestions to improve the content?

Version 2 - Questions on meditation activity

1. What is your overall meditation experience for this version on a scale of 1 to 10 . 1 being worst and 10 being the best
2. Are the vocal tones comfortable for you?
3. Compared to the previous version, is the language clear and understandable? - worse/same/better
4. Did you find the voice to be calming and soothing?
5. How is the pace of meditation? Is it too slow or too fast?
6. Do you think the length of meditation is too short or too long?
7. Are you able to follow the instructions easily?
8. Did you find yourself lost during any point in the session? – Yes/No
9. Do you find the content to be logically structured?
10. Compared to the previous version, what do you think about complexity of words - simpler/same/complex
11. Compared to the previous version, what do you think about clarity of content- less clear/same/more clear
12. Compared to the previous version, how did the content impact your emotions or state of mind? – worse/same/better
13. Is there any season that you felt really engaging ? If yes, why?
14. Do you have any suggestions to improve content?

A.4 ChatGPT prompts

CHATGPT PROMPTS USED FOR GENERATING THE SCRIPTS

Generate a meditation script for Spring season for Hopi youth of age 14-24 years. Use the technique of Mindfulness. The words used in the script should be understandable by 6th-grade students or lower. This meditation is for beginners and should focus on slow, deep breathing. The content should follow a logical flow and engage the user in meditation. The content should describe the Spring seasonal activities related to agriculture, celebrations, and dances. The script should use simple sentences and contain less than 500 words.

Generate a meditation script for Summer season for Hopi youth of age 14-24 years. Use the technique of Interconnectedness. The words used in the script should be understandable by 6th-grade students or lower. This meditation is for beginners and should focus on slow, deep breathing. The content should follow a logical flow and engage the user in meditation. The content should describe the Summer seasonal activities related to agriculture, celebrations, and dances. The script should use simple sentences and contain less than 500 words.

Generate a meditation script for Fall season for Hopi youth of age 14-24 years. Use the technique of Compassion. The words used in the script should be understandable by 6th-grade students or lower. This meditation is for beginners and should focus on slow, deep breathing. The content should follow a logical flow and engage the user in meditation. The content should describe the Fall seasonal activities related to agriculture, celebrations, and dances. The script should use simple sentences and contain less than 500 words.

Generate a meditation script for Winter season for Hopi youth of age 14-24 years. Use the technique of Growth Mindset. The words used in the script should be understandable by 6th-grade students or lower. This meditation is for beginners and should focus on slow, deep breathing. The content should follow a logical flow and engage the user in meditation. The content should describe the Winter seasonal activities related to agriculture, celebrations, and dances. The script should use simple sentences and contain less than 500 words.

Bibliography

[1] S. E. Nelson and K. Wilson, *The Mental Health of Indigenous Peoples in Canada: A Critical Review of Research*, *Social Science & Medicine* **176** (2017) 93–112.

[2] "Mental Illness — nimh.nih.gov." https://www.nimh.nih.gov/health/statistics/mental-illness: :text=PrevalenceMar, 2023.

[3] A. H. Alsswey, H. Al-Samarraie, F. A. El-Qirem, A. I. Alzahrani, and O. Alfarraj, *Culture in the Design of mHealth UI: An Effort to Increase Acceptance Among Culturally Specific Groups*, *The Electronic Library* **38** (2020), no. 2 257–272.

[4] M. Vigil-Hayes, A. F. Collier, S. Hagemann, G. Castillo, K. Mikkelson, J. Dingman, A. Muñoz, J. Luther, and A. McLaughlin, *Integrating Cultural Relevance into a Behavioral mHealth Intervention for Native American Youth*, *Proceedings of the ACM on human-computer interaction* **5** (2021), no. CSCW1 1–29.

[5] H. A. Simon, *Artificial Intelligence: Where has it Been, and Where is it Going?*, *IEEE Transactions on Knowledge and Data Engineering* **3** (1991), no. 2 128–136.

[6] A. Kaplan and M. Haenlein, *Siri, Siri, in my Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence*, *Business Horizons* **62** (2019), no. 1 15–25.

[7] Ö. Aydın and E. Karaarslan, *Is ChatGPT Leading Generative AI? What is Beyond Expectations?*, *Academic Platform Journal of Engineering and Smart Systems* **11** (2023), no. 3 118–134.

[8] F. Fui-Hoon Nah, R. Zheng, J. Cai, K. Siau, and L. Chen, *Generative AI and ChatGPT: Applications, Challenges, and AI-Human Collaboration*, *Journal of Information Technology Case and Application Research* **25** (2023), no. 3 277–304.

[9] H. S. Sætra, *Generative AI: Here to Stay, but for Good?*, *Technology in Society* **75** (2023) 102372.

[10] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, *Large Language Models in Medicine*, *Nature Medicine* **29** (2023), no. 8 1930–1940.

[11] F. Farhat, *ChatGPT as a Complementary Mental Health Resource: A Boon or a Bane*, *Annals of Biomedical Engineering* (2023) 1–4.

[12] A. Nazir and Z. Wang, *A Comprehensive Survey of ChatGPT: Advancements, Applications, Prospects, and Challenges*, *Meta-Radiology* **1** (2023), no. 2 100022.

[13] B. Wodecki, "UBS: ChatGPT is the Fastest Growing App of All Time — aibusiness.com." https://aibusiness.com/nlp/ubs-chatgpt-is-the-fastest-growing-app-of-all-time, Feb, 2023.

[14] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, and Y. Tang, *A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development*, *IEEE/CAA Journal of Automatica Sinica* **10** (2023), no. 5 1122–1136.

[15] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, *et. al.*, *Improving Language Understanding by Generative Pre-Training*, .

[16] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *et. al.*, *Language Models are Unsupervised Multitask Learners*, *OpenAI blog* **1** (2019), no. 8 9.

[17] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et. al.*, *Language Models are Few-shot Learners*, *Advances in Neural Information Processing Systems* **33** (2020) 1877–1901.

[18] OpenAI, *GPT-4 Technical Report*, *ArXiv* **abs/2303.08774** (2023).

[19] S. Frieder, L. Pinchetti, R.-R. Griffiths, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, and J. Berner, *Mathematical Capabilities of ChatGPT*, *arXiv preprint arXiv:2301.13867* (2023).

[20] H. Alkaissi and S. I. McFarlane, *Artificial Hallucinations in ChatGPT: Implications in Scientific Writing*, *Cureus* **15** (2023), no. 2.

[21] A. Borji, *A Categorical Archive of ChatGPT Failures*, *arXiv preprint arXiv:2302.03494* (2023).

[22] D. R. Cotton, P. A. Cotton, and J. R. Shipway, *Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT*, *Innovations in Education and Teaching International* (2023) 1–12.

[23] T. Y. Zhuo, Y. Huang, C. Chen, and Z. Xing, *Exploring AI Ethics of ChatGPT: A Diagnostic Analysis*, *arXiv preprint arXiv:2301.12867* (2023).

[24] M. Cascella, J. Montomoli, V. Bellini, and E. Bignami, *Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios*, *Journal of Medical Systems* **47** (2023), no. 1 33.

[25] N. Vijayaraghavan and M. Chandran, *Effect of Meditation on Psychological Well-Being*, *The International Journal of Indian Psychology* **7** (2019), no. 2.

[26] M. C. Pascoe, M. de Manincor, J. Tseberja, M. Hallgren, P. A. Baldwin, and A. G. Parker, *Psychobiological Mechanisms Underlying the Mood Benefits of Meditation: A Narrative Review*, *Comprehensive Psychoneuroendocrinology* **6** (2021) 100037.

[27] M. Ireland, *Meditation and Psychological Health and Functioning: A Descriptive and Critical Review*, *The Scientific Review of Mental Health Practice* **9** (2012), no. 1 4–19.

[28] H. A. Wayment, J. J. Bauer, and K. Sylaska, *The Quiet Ego Scale: Measuring the Compassionate Self-Identity*, *Journal of Happiness Studies* **16** (2015) 999–1033.

[29] H. A. Wayment and J. J. Bauer, *The Quiet Ego: Motives for Self-Other Balance and Growth in Relation to Well-Being*, *Journal of Happiness Studies* **19** (2018) 881–896.

[30] D. R. King, G. Nanda, J. Stoddard, A. Dempsey, S. Hergert, J. H. Shore, and J. Torous, *An Introduction to Generative Artificial Intelligence in Mental Health Care: Considerations and Guidance*, *Current Psychiatry Reports* (Nov, 2023).

[31] C. Blease and J. Torous, *ChatGPT and Mental Healthcare: Balancing Benefits with Risks of Harms*, *BMJ Mental Health* **26** (2023), no. 1 e300884.

[32] W. H. Pinaya, M. S. Graham, E. Kerfoot, P.-D. Tudosiu, J. Dafflon, V. Fernandez, P. Sanchez, J. Wolleb, P. F. da Costa, A. Patel, *et. al.*, *Generative AI for Medical Imaging: Extending the MONAI Framework*, *arXiv preprint arXiv:2307.15208* (2023).

[33] A. Karpathy, P. Abbeel, G. Brockman, P. Chen, V. Cheung, Y. Duan, I. Goodfellow, D. Kingma, J. Ho, R. Houthooft, and et al., "Generative Models." https://openai.com/research/generative-models, Jun, 2016.

[34] P. Weichbroth, *Usability of Mobile Applications: A Systematic Literature Study*, *IEEE Access* **8** (2020) 55563–55577.

[35] "Indigenous Peoples — worldbank.org." https://www.worldbank.org/en/topic/indigenouspeoples, Apr, 2023.

[36] F. Yang, T. Han, K. Deng, and Y. Han, *The Application of Artificial Intelligence in the Mental Diseases*, in *Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare*, (Taiyuan, China), p. 36–40, Association for Computing Machinery, 2020.

[37] X. Feng, M. Hu, and W. Guo, *Application of Artificial Intelligence in Mental Health and Mental Illnesses*, in *Proceedings of the 3rd International Symposium on Artificial Intelligence for Medicine Sciences*, (Amsterdam, Netherlands), p. 506–511, Association for Computing Machinery, 2022.

[38] S. Rasouli, M. Ghafurian, and K. Dautenhahn, *Students' Views on Intelligent Agents as Assistive Tools for Dealing with Stress and Anxiety in Social Situations*, in *Proceedings of the 10th International Conference on Human-Agent Interaction*, pp. 23–31, 2022.

[39] M. Jovanović, M. Baez, and F. Casati, *Chatbots as Conversational Healthcare Services*, *IEEE Internet Computing* **25** (2020), no. 3 44–51.

[40] T. Koulouri, R. Macredie, and D. Olakitan, *Chatbots to Support Young Adults' Mental Health: An Exploratory Study of Acceptability*, *ACM Transactions on Interactive Intelligent Systems* **12** (Jun, 2022) 1–39.

[41] A. Kapoor and S. Goel, *Applications of Coversational AI in Mental Health: A Survey*, in *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1013–1016, IEEE, 2022.

[42] A. Thieme, M. Hanratty, M. Lyons, J. Palacios, R. F. Marques, C. Morrison, and G. Doherty, *Designing Human-Centered AI for Mental Health: Developing Clinically Relevant Applications for Online CBT Treatment*, *ACM Transactions on Computer-Human Interaction* **30** (2023), no. 2 1–50.

[43] K. Rani, H. Vishnoi, and M. Mishra, *A Mental Health Chatbot Delivering Cognitive Behavior Therapy and Remote Health Monitoring Using NLP and AI*, pp. 313–317, May, 2023.

[44] C. Dharmapuri, A. Agarwal, F. Anwer, and J. Mahor, *AI Chatbot: Application in Psychiatric Treatment and Suicide Prevention*, pp. 41–44, Mar, 2022.

[45] S. D'Alfonso, O. Santesteban-Echarri, S. Rice, G. Wadley, R. Lederman, J. Gleeson, and M. Alvarez-Jimenez, *Artificial Intelligence-Assisted Online Social Therapy for Youth Mental Health*, *Frontiers in Psychology* **8** (06, 2017) 796.

[46] N. Anand, L. Pant, T. Alam, S. Pundir, L. Thomas, and R. Renukaprasad, *Artificial Intelligence's Contribution to Mental Health Education*, pp. 462–467, Mar, 2023.

[47] R. Wang, J. Wang, Y. Liao, and J. Wang, *Supervised Machine Learning Chatbots for Perinatal Mental Healthcare*, pp. 378–383, Dec, 2020.

[48] V. Mody and V. Mody, *Mental Health Monitoring System using Artificial Intelligence: A Review*, in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pp. 1–6, IEEE, 2019.

[49] T. Shaik, X. Tao, N. Higgins, H. Xie, R. Gururajan, and X. Zhou, *AI Enabled RPM for Mental Health Facility*, in *Proceedings of the 1st ACM Workshop on Mobile and Wireless Sensing for Smart Healthcare*, pp. 26–32, Association for Computing Machinery, 2022.

[50] P. Wilbourne, G. Dexter, and D. Shoup, *Research Driven: Sibly and the Transformation of Mental Health and Wellness*, pp. 389–391, May, 2018.

[51] A. Nayar, Z. Attar, S. Kachwala, T. Biswas, and S. Wagh, *Dost-Mental Health Assistant Chatbot*, pp. 252–257, Dec, 2022.

[52] R. Fulmer, A. Joerin, B. Gentile, L. Lakerink, M. Rauws, *et. al.*, *Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial*, *JMIR Mental Health* **5** (2018), no. 4 e9782.

[53] H. Liu, H. Peng, X. Song, C. Xu, and Z. Meng, *Using AI Chatbots to Provide Self-help Depression Interventions for University Students: A Randomized Trial of Effectiveness*, *Internet Interventions* **27** (Jan, 2022) 100495.

[54] P. Dev, S. Haque, A. Noor, A. Srabon, M. Wasik, S. Mim, S. Sharife, F. Rahman, R. Syara, S. Iqbal, M. H. K. Mehedi, and A. A. Rasel, *A Comparative Study of Chatbot Catered Toward Mental Health*, in *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, pp. 1–6, IEEE, 2022.

[55] J. Duval, L. Turmo Vidal, E. Márquez Segura, Y. Li, and A. Waern, *Reimagining Machine Learning's Role in Assistive Technology by Co-Designing Exergames with Children Using a Participatory Machine Learning Design Probe*, in *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, (New York, NY, USA), pp. 1–16, Association for Computing Machinery, 2023.

[56] P. Carlbring, H. Hadjistavropoulos, A. Kleiboer, and G. Andersson, *A New Era in Internet Interventions: The Advent of Chat-GPT and AI-Assisted Therapist Guidance*, *Internet Interventions* **32** (Apr, 2023) 100621.

[57] R. Bhattacharyya, K. Chakraborty, and R. Neogi, *ChatGPT and its Application in the Field of Mental Health*, *Journal of SAARC Psychiatric Federation* **1** (Jan, 2023) 6.

[58] N. Imran, A. Hashmi, and A. Imran, *Chat-GPT: Opportunities and Challenges in Child Mental Healthcare*, *Pakistan Journal of Medical Sciences* **39** (Jun, 2023).

[59] R. Villarino and M. Villarino, *ChatGPT (Generated Pre-trained Transformer) as an Adjunct to Mental Health Interventions: A Commentary*, *Int J Psychol Psychoanal* (Apr, 2023).

[60] M. Jovanovic and M. Campbell, *Generative Artificial Intelligence: Trends and Prospects*, *Computer* **55** (2022), no. 10 107–112.

[61] "What is Deep Learning." https://aws.amazon.com/what-is/deep-learning/: :text=Deepaccurate

[62] V. Bilgram and F. Laarmann, *Accelerating Innovation with Generative AI: AI-Augmented Digital Prototyping and Innovation Methods*, *IEEE Engineering Management Review* **51** (2023), no. 2 18–25.

[63] K. Kenthapadi, H. Lakkaraju, and N. Rajani, *Generative AI Meets Responsible AI: Practical Challenges and Opportunities*, in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, (Long Beach, CA, USA), p. 5805–5806, Association for Computing Machinery, 2023.

[64] K. Nova, *Generative AI in Healthcare: Advancements in Electronic Health Records, Facilitating Medical Languages, and Personalized Patient Care*, *Journal of Advanced Analytics in Healthcare Management* **7** (Apr, 2023) 115–131.

[65] J. Varghese and J. Chapiro, *ChatGPT: The Transformative Influence of Generative AI on Science and Healthcare*, *Journal of Hepatology* (2023).

[66] I. S. Smits, "Customisation in E-(Mental) Health." http://essay.utwente.nl/81964/, Jul, 2020.

[67] S.-W. Cheng, C.-W. Chang, W.-J. Chang, H.-W. Wang, C.-S. Liang, T. Kishimoto, J. P.-C. Chang, J. S. Kuo, and K.-P. Su, *The Now and Future of ChatGPT and GPT in Psychiatry*, *Psychiatry and Clinical Neurosciences* **77** (2023), no. 11 592–596, [https://onlinelibrary.wiley.com/doi/pdf/10.1111/pcn.13588].

[68] ChatBeacon, "ChatBeacon." https://www.chatbeacon.io/industry-chatgpt/mental-health-chatbot.

[69] Koko, "Koko." https://gpt3demo.com/apps/koko-ai.

[70] J. Li and A. Brar, *The Use and Impact of Digital Technologies for and on the Mental Health and Wellbeing of Indigenous People: A Systematic Review of Empirical Studies*, *Computers in Human Behavior* (Aug, 2021) 106988.

[71] J. Povey, P. P. J. R. Mills, K. M. Dingwall, A. Lowell, J. Singer, D. Rotumah, J. Bennett-Levy, and T. Nagel, *Acceptability of Mental Health Apps for Aboriginal and Torres Strait Islander Australians: A Qualitative Study*, *J Med Internet Res* **18** (Mar, 2016) e65.

[72] J. Tighe, F. Shand, R. Ridani, A. Mackinnon, N. De La Mata, and H. Christensen, *Ibobbly Mobile Health Intervention for Suicide Prevention in Australian Indigenous Youth: A Pilot Randomised Controlled Trial*, *BMJ Open* **7** (2017), no. 1 e013518.

[73] E. Perdacher, D. Kavanagh, J. Sheffield, K. Healy, P. Dale, and E. Heffernan, *Using the Stay Strong App for the Well-being of Indigenous Australian Prisoners: Feasibility Study*, *JMIR Form Res* **6** (Apr, 2022) e32157.

[74] G. Skinner, R. Fletcher, and C. May, *An ICT eHealth Mobile Web Application for Young Indigenous Fathers – StayinOnTrack*, in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 103–108, IEEE, Jul, 2017.

[75] L. Williams and G. Washington, *Soul Glow: An Application for Helping to Understand and Improve Mental Health Care of HBCU Students*, pp. 358–360, Jun, 2018.

[76] N. Noronha, S. Smith, D. Martin Hill, L. Davis Hill, S. Smith, A. General, C. Mcqueen, M. Horse, A. Drossos, C. Lokker, N. Bilodeau, and C. Wekerle, *The Use of Mobile Applications to Support Indigenous Youth Wellbeing in Canada*, *International Journal of Child and Adolescent Resilience* **8** (May, 2021).

[77] J. M. Hensel, K. Ellard, M. Koltek, G. Wilson, and J. Sareen, *Digital Health Solutions for Indigenous Mental Well-Being*, *Current Psychiatry Reports* **21** (Jul, 2019) 1–9.

[78] S. J. M. J. Reilly, R., *Use and Uptake of Web-based Therapeutic Interventions Amongst Indigenous Populations in Australia, New Zealand, the United States of America and Canada: a Scoping Review*, *Systematic Reviews* **9** (May, 2020) 1–17.

[79] E. Toombs, K. Kowatch, L. Dalicandro, S. McConkey, C. Hopkins, and C. Mushquash, *A Systematic Review of Electronic Mental Health Interventions for Indigenous Youth: Results and Recommendations*, *Journal of Telemedicine and Telecare* **27** (Jan, 2020) 1357633X1989923.

[80] A. BlackDeer and D. Wolf, *Evidence Mapping: Interventions for American Indian and Alaska Native Youth Mental Health*, *Journal of Evidence-Based Social Work* **17** (Jan, 2020) 49–62.

[81] M. Antonio and J. Chung-Do, *Systematic Review of Interventions Focusing on Indigenous Adolescent Mental Health and Substance Use*, *American Indian and Alaska Native Mental Health Research (Online)* **22** (Oct, 2015) 36–56.

[82] A. Chan, *GPT-3 and InstructGPT: Technological Dystopianism, Utopianism, and "Contextual" Perspectives in AI Ethics and Industry*, *AI and Ethics* **3** (2023), no. 1 53–64.

[83] D. Baidoo-Anu and L. O. Ansah, *Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning*, *Journal of AI* **7** (2023), no. 1 52–62.

[84] A. Howard, W. Hope, and A. Gerada, *ChatGPT and Antimicrobial Advice: the End of the Consulting Infection Doctor?*, *The Lancet Infectious Diseases* **23** (2023), no. 4 405–406.

[85] "Elevenlabs." https://elevenlabs.io/.

[86] "ElevenLabs." https://en.wikipedia.org/wiki/ElevenLabs, Feb, 2023.

[87] "Audacity (Audio Editor)." https://en.wikipedia.org/wiki/Audacity$_(audio_editor)$.

[88] M. Q. Patton, *Purposeful sampling and case selection: Overview of strategies and options*, in *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Sage, 2014.

[89] Y. Terhorst, P. Philippi, L. B. Sander, D. Schultchen, S. Paganini, M. Bardus, K. Santo, J. Knitza, G. C. Machado, S. Schoeppe, *et. al.*, *Validation of the Mobile Application Rating Scale (MARS)*, *Plos one* **15** (2020), no. 11 e0241480.

[90] S. R. Stoyanov, L. Hides, D. J. Kavanagh, and H. Wilson, *Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS)*, *JMIR mHealth and uHealth* **4** (2016), no. 2 e5849.

[91] A. Lambrecht, N. Vuillerme, C. Raab, D. Simon, E.-M. Messner, M. Hagen, S. Bayat, A. Kleyer, T. Aubourg, G. Schett, *et. al.*, *Quality of a Supporting Mobile App for Rheumatic Patients: Patient-based Assessment Using the User Version of the Mobile Application Scale (uMARS)*, *Frontiers in Medicine* **8** (2021) 715345.

[92] A. Türkyilmaz, S. Kantar, M. E. Bulak, O. Uysal, *et. al.*, *User Experience Design: Aesthetics or Functionality*, *Managing Intellectual Capital and Innovation for Sustainable and Inclusive Society: Managing Intellectual Capital and Innovation* (2015) 559–565.

[93] M. Nadeem, A. Bethke, and S. Reddy, *Stereoset: Measuring Stereotypical Bias in Pretrained Language Models*, *arXiv preprint arXiv:2004.09456* (2020).