© 2025 IJRTI | Volume 10, Issue 4 April 2025 | ISSN: 2456-3315

FitZen: AI-based Fitness and Wellbeing Application

Shivom S. Hatalkar[1], Aditya U. Apte[2], Praharsh S. Shinde[3], Chaitanya U. Barapatre[4], Priyanka O. Bhoir[5]
[1,2,3,4]*Student, Dept of Artificial Intelligence and Machine Learning, Universal College of Engineering, Mumbai, India*
[5]*Professor, Dept of Artificial Intelligence and Machine Learning, Universal College of Engineering, Mumbai, India*

***Abstract*— FitZen is an AI-driven web application designed to promote holistic fitness and well-being through a multi-faceted approach. It encompasses four core sections: Mental Health, Workout Recommendations, Diet Planning, and Yoga Guidance. At its heart lies Elixir-Health-Llama3B, a fine-tuned LLM chatbot based on Llama-3.2-3B-Instruct, which provides empathetic mental health support and personalized fitness-related guidance. The application integrates advanced natural language processing, rule-based filtering, and dynamic recommendation systems. Leveraging the capabilities of the LLM alongside a robust database stored on Supabase, FitZen offers personalized workout routines, tailored diet plans, and curated yoga asanas. This paper outlines the design, methodology, technical infrastructure, and evaluation metrics used to deliver an intelligent, user-centric health platform. FitZen also offers a future-ready architecture to support additional features such as fitness tracking, meditation, and reward-based engagement strategies.**

***Keywords*— Holistic Fitness, Wellbeing, AI Chatbot, LLM Fine-Tuning, Personalized Recommendations, Mental Health, Conversational AI Chatbot**

I. INTRODUCTION

Health and wellness have become paramount in today's fast-paced world, where individuals strive to maintain a balance between physical fitness, mental well-being, and dietary practices. While numerous applications address these domains individually, they lack an integrated approach that acknowledges the interconnectedness of fitness, mental health, diet, and yoga. FitZen addresses this gap by combining state-of-the-art language models with rule-based filtering and data management to offer a comprehensive wellness assistant.

Existing solutions in the market often exhibit fragmentation, focusing solely on either fitness routines, mental health guidance, dietary plans, or yoga practices. Consequently, users are required to switch between multiple applications, leading to a disjointed experience. FitZen innovates by providing a unified platform that not only consolidates these aspects but also dynamically adapts recommendations based on user input and context. This paper outlines the methodology, and results of developing FitZen, highlighting its potential to transform holistic health management.

Key objectives of FitZen:

- **Mental Health Support**: Provide empathetic and accurate mental health assistance via the fine-tuned Elixir-Health-Llama3B model.
- **Workout Recommendation**: Generate personalized workout plans based on user parameters such as age, BMI, and fitness goals.
- **Diet Planning**: Deliver customized diet plans tailored to users' nutritional needs and health goals.
- **Yoga Guidance**: Offer curated lists of yoga asanas and, in the future, personalized sequences for well-being.

The system employs a hybrid architecture combining advanced LLM-based natural language processing with rule-based data retrieval, providing both flexibility and accuracy. This paper explores the technical framework, dataset composition, training methodologies, and the resulting performance evaluation of the FitZen platform.

II. LITERATURE REVIEW

The intersection of fitness and mental health applications has gained substantial traction in recent years, with platforms such as MyFitnessPal, Calm, and Headspace emerging as leaders in their respective domains [5]. MyFitnessPal focuses primarily on caloric tracking and exercise logging, while Calm and Headspace emphasize guided meditation, mindfulness, and stress management. Despite their popularity, these applications tend to operate in isolation, failing to provide an integrated experience that considers the holistic well-being of users. This gap in personalization and adaptability highlights the need for solutions that leverage advancements in artificial intelligence (AI) and machine learning (ML) to offer a context-aware and personalized user experience.

The development of FitZen draws upon extensive research in the fields of natural language processing (NLP), health informatics, and recommendation systems. Below is a comprehensive breakdown of the related works that inspired and informed the project.

1. Large Language Models (LLMs) in Context-Aware Conversational AI

Recent advancements in large language models (LLMs) have significantly improved the ability of AI systems to engage in context-aware, human-like conversations. Models such as Llama-3.2-3B-Instruct have demonstrated promising results in healthcare, mental health, and fitness-related applications. However, challenges persist, particularly in ensuring response accuracy, maintaining empathetic communication, and reducing hallucinations. Studies indicate that generic LLMs often fail to provide reliable advice in specialized domains due to their broad, uncurated training data.

2. Mental Health Chatbots and LLM Hallucination Challenges

AI-powered mental health chatbots have garnered significant attention in recent years, with platforms such as Woebot and Wysa demonstrating the potential of LLM-driven therapy assistants [2]. However, studies indicate that LLM-based chatbots face three primary challenges:

- Hallucination Issues: LLMs sometimes generate misleading or fabricated responses, which can be problematic in mental health contexts where accuracy is critical.

- Lack of Emotional Intelligence: While LLMs can recognize emotional cues, they often struggle to provide truly empathetic responses.
- Generic Responses: Without domain-specific fine-tuning, LLMs default to generic, surface-level advice, which may not effectively support users in mental health crises.

FitZen tackles these challenges by leveraging a curated mental health dataset comprising 75k rows of data, including validated psychological assessments, user interactions, and therapy dialogue scripts. This fine-tuned dataset ensures that responses remain empathetic, contextually relevant, and tailored to user concerns.

3. Mental Health Chatbots and AI-based Assistance
   - Woebot (Fitzpatrick et al., 2017): Demonstrated the efficacy of AI-driven conversational agents in delivering cognitive behavioral therapy (CBT) and reducing symptoms of depression.

     Relevance to FitZen: We incorporated a similar conversational interface but enhanced with a large-scale fine-tuned model trained on 75K rows of mental health and medical data.
   - TherapAI (Ravichandran et al., 2022): Combined LLMs with therapeutic frameworks to deliver evidence-based mental health interventions.

     Relevance to FitZen: Inspired the use of multi-turn dialogues for better conversational depth and contextual understanding.
4. LLM Applications in Healthcare
   - BioGPT (Luo et al., 2022): Specialized in biomedical text generation using transformer models.

     Relevance to FitZen: Informed our decision to use a specialized dataset and LoRA fine-tuning to enhance Elixir-Health-Llama3B's medical accuracy.
   - MedPaLM (Singhal et al., 2023): Integrated medical knowledge graphs with LLMs for accurate patient responses.

     Relevance to FitZen: Emphasized the importance of reliable data sources (e.g., NHS, peer-reviewed datasets) during the curation of our 75K dataset.
5. Fitness and Nutrition Recommendation Systems
   - Hybrid Recommendation Systems (Ricci et al., 2015): Demonstrated the effectiveness of combining content-based and collaborative filtering.

     Relevance to FitZen: We utilize rule-based filtering for structured outputs, ensuring precision in workout and diet recommendations.
   - Personalized Health Guidance (Yu et al., 2021): Explored multi-modal approaches to personalized health systems integrating physiological parameters.

     Relevance to FitZen: Influenced our approach to use dynamic user inputs for adaptive recommendations.

III. PROPOSED SYSTEM

FitZen is designed as a comprehensive, modular wellness assistant aimed at promoting holistic well-being. The system integrates four key modules, each focused on a specific dimension of health: mental wellness, physical fitness, dietary guidance, and yoga practices.

1. Mental Health Module
   - Conversational AI Backbone: The core of FitZen's mental health module is powered by the Elixir-Health-Llama3B model, fine-tuned specifically to understand, process, and respond empathetically to mental health-related queries.
   - Fine-Tuning Details:
     a. Model Architecture: Llama-3.2-3B-Instruct.
     b. Dataset: 75,000 rows comprising 42,000 mental health dialogues and 33,000 general medical advice entries [15].
     c. Method: LoRA (Low-Rank Adaptation) fine-tuning with 8-bit quantization for resource optimization [13][14].
     d. Cloud GPU (Runpod.io): Fine-tuned on an RTX 6000 Ada GPU (48GB VRAM).
     e. Duration: 27 hours

The conversational interface supports multi-turn dialogues, ensuring that user interactions are coherent and contextually aware, simulating the continuity typical in therapy sessions.

2. Workout Recommendation Module
   - User parameters are dynamically encoded as structured prompt strings.
   - The LLM interprets the prompts to generate preliminary suggestions.
   - Rule-based filtering mechanisms retrieve suitable exercise routines from a Supabase database, ensuring recommendations align with recognized fitness guidelines and the user's objectives.
3. Diet Recommendation Module
   - User Input Parameters:
     a. Dietary Restrictions (e.g., vegan, keto, gluten-free)
     b. Health Goals (e.g., calorie deficit, muscle gain, maintenance)
   - System Logic:
     a. Users' dietary preferences and goals are processed as prompt strings.
     b. The model dynamically generates personalized meal plans.
     c. Rule-based checks validate nutritional balance (e.g., macronutrient consistency, avoidance of allergens), retrieving data from Supabase's verified meal records.
4. Yoga Module
   - Current Functionality: Provides a static listing of yoga asanas, complete with descriptions and benefits.

IV. METHODOLOGY

The core methodology behind FitZen lies in its multi-component system, where user inputs are processed and transformed into meaningful health and wellness advice through a combination of fine-tuned language modeling and deterministic rule-based logic.

A. Mental Health Chatbot: The mental health chatbot serves as one of the most crucial components of FitZen, aimed at providing empathetic, context-aware responses to users seeking mental and general health guidance. This system leverages a custom fine-tuned version of the Llama-3.2-3B-Instruct model, named Elixir-Health-Llama3B.

a) Fine-Tuning Process: To optimize the model for deployment under limited hardware constraints while maintaining high-quality responses, LoRA (Low-Rank Adaptation) was employed [13]. LoRA allows training only low-rank matrices inserted into the transformer architecture, reducing the number of trainable parameters [14]. In addition, 8-bit quantization was applied to reduce GPU memory consumption, enabling larger batch sizes and longer sequences without compromising model accuracy.

Training was executed on a high-performance NVIDIA RTX 6000 Ada GPU with 48GB VRAM using a Python-based Jupyter Notebook on RunPod infrastructure. The training process spanned approximately 27 hours, fine-tuning the model on a dataset containing 75,000 curated instruction-response pairs, including 20,000 multi-turn conversations and 55,000 single-turn mental health and medical queries [15].

b) System Workflow:

1. User Interface: A responsive front-end interface, built using Next.js and styled with Aceternity and optionally ShadCN components, is responsible for capturing user input.
2. Backend Integration: The front-end communicates with a Flask-based backend that acts as a bridge between the user interface, the LLM (Elixir-Health-Llama3B), and the Supabase database. Alternatively, in deployment scenarios using the GGUF quantized model format, the chatbot can be integrated directly with the frontend using the Ollama inference engine via API endpoints, eliminating the need for intermediary backend logic.
3. LLM Interaction: The structured inputs from the user are converted into prompt templates compatible with the LLM's training format. The API or Ollama interface forwards the prompt to the Elixir-Health-Llama3B model and receives a rich, context-sensitive natural language response.
4. Data Retrieval and Recommendation: Supabase serves as the primary data storage system for static and semi-structured content such as yoga asanas, workout routines, and dietary plans. Rule-based logic is applied on top of the database to fetch the most appropriate entries that match the user's profile, model responses, and fitness goals.

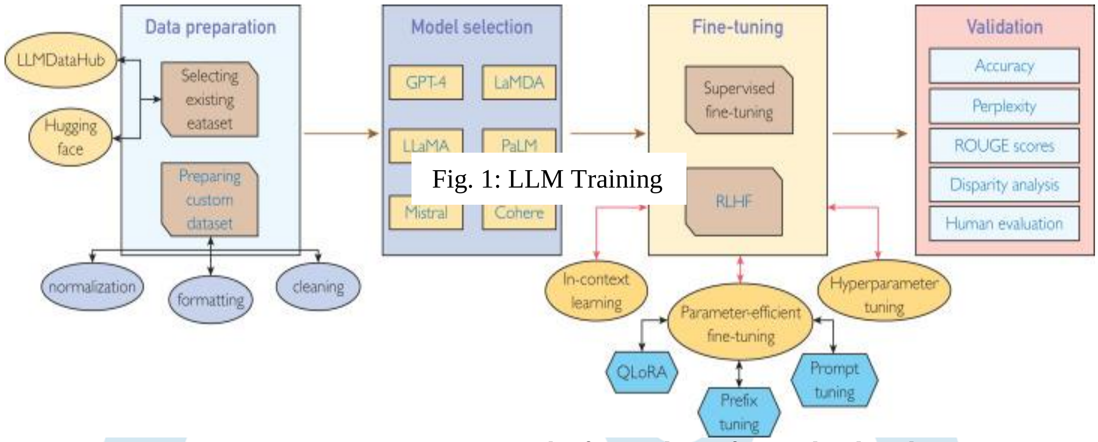

Fig. 1: LLM Training

B. System Architecture

The architecture of FitZen is designed with modularity and scalability in mind, comprising five primary components: the frontend interface, backend services, model integration layer, Supabase database, and the recommendation engine.

The frontend interface is developed using Next.js and styled with Aceternity UI and optionally ShadCN components. It serves as the user's primary interaction layer, facilitating navigation and input collection. Users authenticate via Clerk and provide parameters such as age, BMI, and personal goals, which are forwarded to the backend APIs.

The backend is developed using Flask and JavaScript, and is responsible for handling API requests, managing routing, input validation, session tracking, and establishing communication between the frontend, database, and model integration layer. It acts as the logical controller of the system, ensuring efficient data flow and decision-making.

The model integration layer incorporates the fine-tuned Elixir-Health-Llama3B model, trained on a diverse 75k-row dataset consisting of mental health and general medical data [12][15]. This layer processes user prompts and returns context-aware responses, including personalized workout, diet, and mental health recommendations. Responses generated by the model are post-processed through a rule-based filtering mechanism for enhanced relevance and safety.

The Supabase database serves as the central data hub,

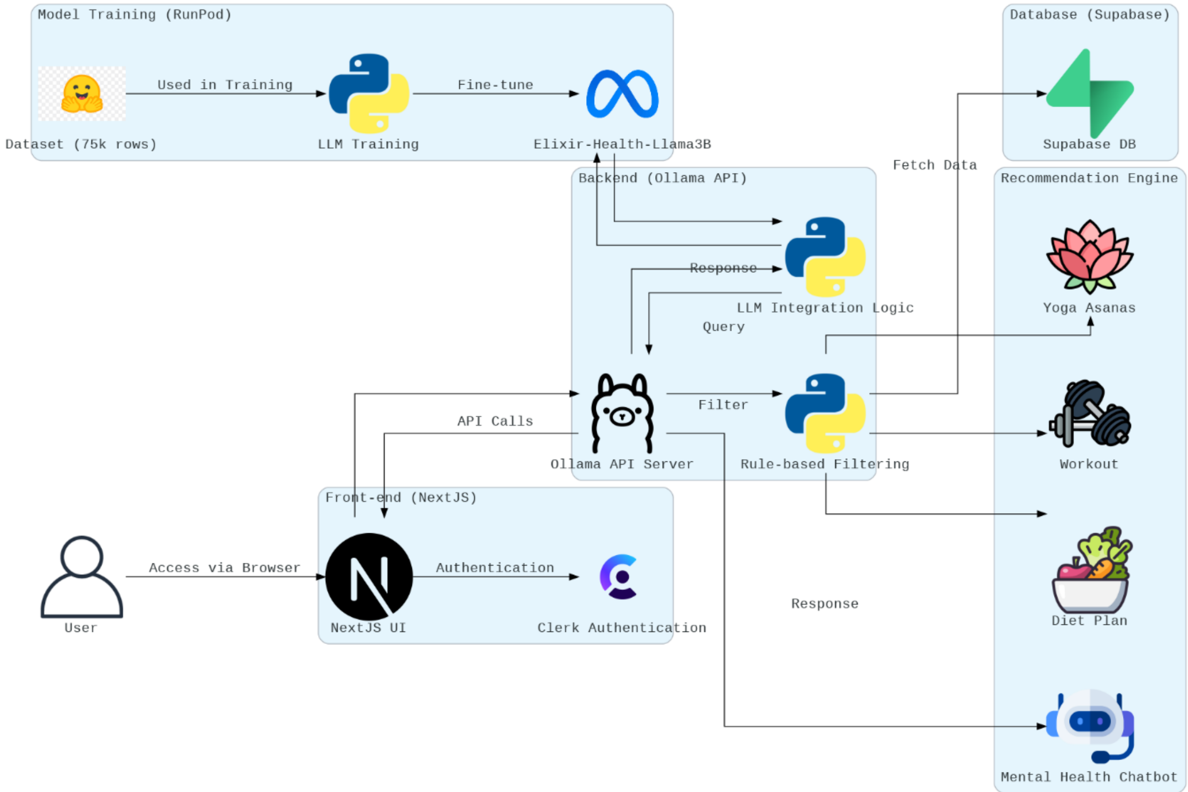

Fig. 2: FitZen Architecture

storing user profiles, structured fitness and dietary datasets, yoga asanas, and logs. It is tightly coupled with the backend and rule-based filtering engine to ensure data persistence and retrieval efficiency.

The recommendation engine is a composite module containing both the LLM-powered generation unit and a rule-based logic component. The LLM handles natural language generation tasks such as mental health support and dynamic workout or diet recommendations. In contrast, the rule-based filtering system retrieves static and semi-structured content from the Supabase database, particularly for exercises, yoga asanas, and dietary suggestions that require deterministic logic.

FitZen follows a modular, scalable architecture consisting of a Next.js-based frontend for user interaction, a Flask-JavaScript backend for handling API logic, a fine-tuned LLM (Elixir-Health-Llama3B) for generating personalized responses, a Supabase database for data management, and a dual-layered recommendation engine. The engine combines LLM generated advice with rule-based filtering to ensure bot personalization and reliability in workout, diet, yog and mental health modules. Together, these componen deliver a seamless, AI-powered holistic healt experience.

V. RESULTS & ANALYSIS

A. Chatbot Performance

To evaluate the effectiveness and quality of the fine tuned Elixir-Health-Llama3B model compared to its base counterpart, Llama-3.2-3B-Instruct, a comprehensive set of NLP evaluation metrics was employed. The following table shows the results:

Metric	Elixir-Health-Llama3B	Llama-3.2-3B-Instruct
BLEU	0.0480	0.0384
ROUGE-1	0.3167	0.3160
ROUGE-2	0.0793	0.0713
ROUGE-L	0.1591	0.1538
METEOR	0.2535	0.2574
BERTScore	0.8345	0.8323
Perplexity	3.4097	3.3593
Toxicity	0.0147	0.0079
Readability	5.7673	6.8177
Distinct-1	0.0753	0.0599
Distinct-2	0.3633	0.2895
Self-BLEU	0.5452	0.6847

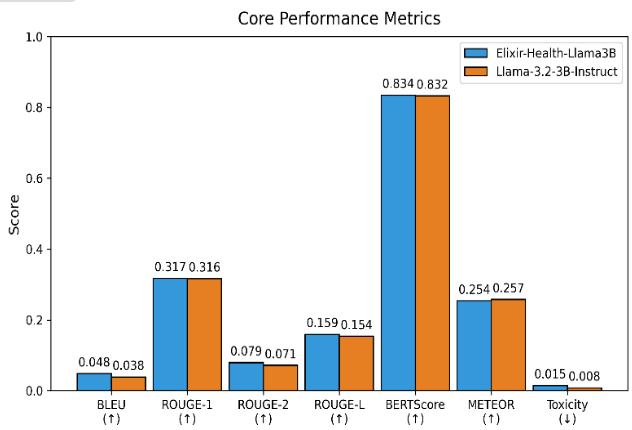

Fig. 3: Core Performance Metrics

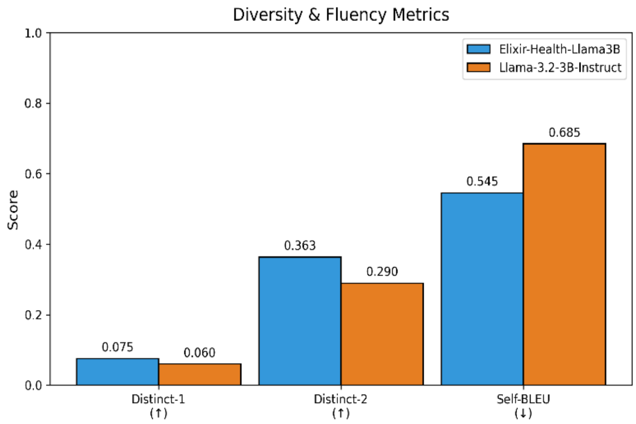

Fig. 4: Diversity & Fluency Metrics

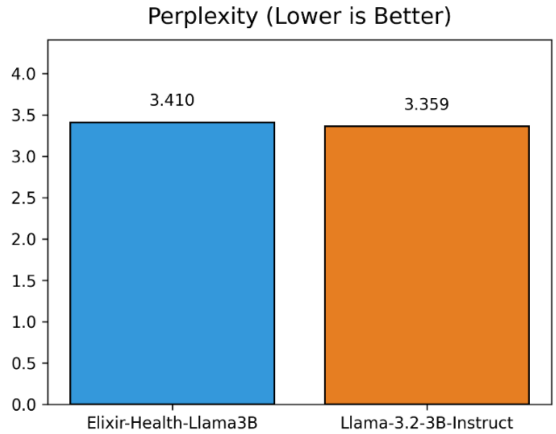

Fig. 5: Perplexity Graph

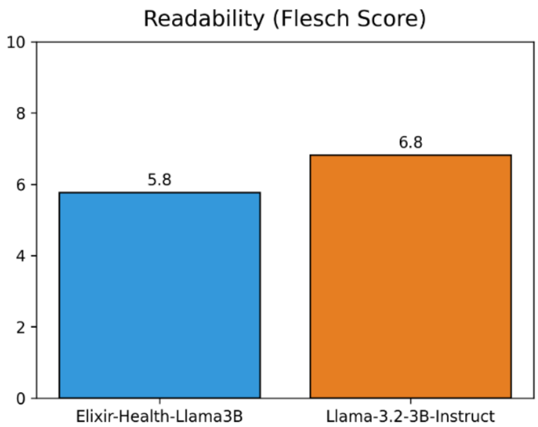

Fig. 6: Readability Graph

VI. DISCUSSION

Observations:

1. BLEU Score:
   The BLEU score for Elixir-Health-Llama3B stands at 0.0480, outperforming the base model's 0.0384. This indicates an improved ability to produce n-gram overlaps with the reference outputs, which is crucial for generating coherent and grammatically correct responses.
2. ROUGE Metrics:
   Both models perform similarly in ROUGE-1 and ROUGE-L, with Elixir-Health-Llama3B holding a slight edge. In ROUGE-2, the fine-tuned model demonstrates a noticeable improvement (0.0793 vs 0.0713), reflecting better capability to capture bi-gram level semantic information and maintaining consistency across longer sequences.
3. METEOR Score:
   Interestingly, the base model shows a marginally higher METEOR score (0.2574 vs 0.2535). This could be due to the broader linguistic diversity. However, the difference is minimal and does not significantly impact the overall evaluation.
4. BERTScore:
   Elixir-Health-Llama3B achieves a higher BERTScore (0.8345), indicates stronger semantic similarity between generated and reference responses.
5. Perplexity:
   The fine-tuned model exhibits a slightly higher perplexity (3.4096) than the base model (3.3593), suggesting marginally less confidence in its output probabilities. However, the difference is negligible and within acceptable bounds.
6. Toxicity:
   A critical evaluation parameter, toxicity levels are slightly higher in Elixir-Health-Llama3B (0.0147) compared to the base model (0.0079), this could be mainly due to the presence of raw mental health conversations in the training dataset. However, it's still well within the safety range required for mental health chatbots (< 3%).
7. Diversity Metrics (Distinct-1 & Distinct-2):
   The fine-tuned model substantially outperforms in terms of diversity:
   - Distinct-1: 0.0753 vs 0.0599
   - Distinct-2: 0.3633 vs 0.2895

   Higher values indicate that Elixir-Health-Llama3B generates more varied and non-repetitive responses, improving conversational naturalness and avoiding generic replies.
8. Self-BLEU:
   Lower Self-BLEU scores are desirable, indicating less repetitiveness. The fine-tuned model achieves a lower Self-BLEU (0.5452) compared to the base model (0.6847), reinforcing its superior diversity and reducing redundancy.

VII. FUTURE SCOPE

While FitZen already offers a robust wellness platform, several avenues for future enhancements have been identified to expand its capabilities:

1. **Personalized Yoga Sequences**: Currently, FitZen provides static yoga asana listings. Future iterations aim to incorporate **dynamic yoga routines**, tailored based on individual user inputs such as stress levels, flexibility, fitness goals, and physical limitations.
2. **Fitness and Sleep Tracking Integration**: Incorporating wearable device data (e.g., from smartwatches) will allow FitZen to offer real-time fitness tracking and sleep monitoring.
3. **Meditation and Healing Music Integration**: Adding curated meditation guides and healing music playlists will enhance the mental wellness offerings of FitZen [3].

VIII. CONCLUSION

FitZen successfully delivers a comprehensive AI-powered health and wellness platform, integrating mental health support, personalized workout and diet plans, and yoga recommendations into a seamless user experience. By harnessing the capabilities of the fine-tuned Elixir-Health-Llama3B model, FitZen achieves significant improvements in conversational relevance, diversity, and safety over baseline models. The system's architecture, grounded in state-of-the-art NLP techniques, LoRA fine-tuning, and hybrid recommendation systems, offers scalable, personalized, and adaptive health guidance to users.

With its demonstrated potential and room for further enhancements such as fitness tracking and personalized meditation, FitZen paves the way for future innovations in AI-driven personalized healthcare technology.

IX. REFERENCES

[1] David, D.; Cristea, I.; Hofmann, S.G. Why cognitive behavioral therapy is the current gold standard of psychotherapy. Front. Psychiatry (2018).
[2] Fitzpatrick, K.K.; Darcy, A.; Vierhile, M. Delivering cognitive behaviour therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot) (2017).
[3] Iqbal, N., Singh, A. & Aleem, S. Effect of Dynamic Meditation on Mental Health (2016).
[4] D. Subhasri, Sayantani Dutta, M. Maria Leena, J.A. Moses, C. Anandharamakrishnan. Gastronomy: An extended platform for customized nutrition (2022).
[5] Zhijun Guo, Alvina Lai, Johan H Thygesen, Joseph Farrington, Thomas Keen, Kezhi Li. Large Language Models for Mental Health Applications: Systematic Review (2024).
[6] Charles F. Reynolds 3rd, Dilip V. Jeste, Perminder S. Sachdev, Dan G. Blazer. Mental health care for older adults: recent advances and new directions (2022).
[7] Patrick J. Smith, Rhonda M. Merwin. The Role of Exercise in Management of Mental Health Disorders: An Integrative Review (2021).
[8] Henning Wackerhage, Brad J. Schoenfeld. Personalized, Evidence-Informed Training Plans and Exercise Prescriptions (2021).
[9] Gerhard Andersson. The latest developments with internet-based psychological treatments for depression (2024).
[10] Justine Fortin, Mélissandre Leblanc, Guillaume Elgbeili, Matthew J. Cordova, Marie-France Marin, Alain Brunet. The mental health impacts of receiving a breast cancer diagnosis: A meta-analysis (2021).
[11] Sandra, Mulkens, Glenn,Waller. New developments in cognitive-behavioural therapy for eating disorders (CBT-ED) (2021).
[12] D.M. Anisuzzaman PhD, Jeffrey G. Malins PhD, Paul A. Friedman MD, Zachi I. Attia PhD. Fine-Tuning Large Language Models for Specialized Use Cases (2024).
[13] Rupak (Bob) Roy - II. Fine-Tuning Large Language Models with PEFT (LoRA) and Rouge Score: A Comprehensive Hands-On Guide (2024).
[14] Shalini Dhote. Parameter-Efficient Fine-Tuning of Large Language Models with LoRA and QLoRA (2023).
[15] ShivomH/Medical-and-Mental-Health