FitnessBot: Chatbot for Nutrition and Exercise Recommendation

1st Abdullah Masood
Department of Computer Science
National University of Computer and Emerging Science
Islamabad, Pakistan
i210822@nu.edu.pk

2nd Tayyab Sohail
Department of Computer Science
National University of Computer and Emerging Science
Islamabad, Pakistan
i212478@nu.edu.pk

3rd Muhammad Dayyan Ahmed
Department of Computer Science
National University of Computer and Emerging Science
Islamabad, Pakistan
i210772@nu.edu.pk

***Abstract*—The growing demand for customized fitness and nutrition counseling highlights the need for more accessible and affordable solutions. This paper introduces fitnessbot, an AI-based chatbot that can give customized recommendations on fitness and nutrition. The fitness module makes customized recommendations based on user-specific factors such as BMI, health conditions, and weight management goals, such as weight gain or loss. The nutrition module delivers detailed nutritional information about different food items to assist users in making informed dietary choices.**

This study compares and evaluates 5 differnt approaches implemented for FitnessBot's recommendation capabilities. They include Adaptive Retrieval-Augmented Generation (RAG),fine-tuning a pre-trained LLM, Contrastive RAG (CRAG), and a Long Short-Term Memory (LSTM) model. These are tested independently using several performance metrics, such as BLEU, ROUGE, F1 score, accuracy, and recall, on generating relevant and personalized responses.

FitnessBot's datasets include detailed user information—such as sex, age, height, weight, BMI, health conditions, fitness goals, and dietary preferences—along with a food nutritional dataset and recipe database to support personalized fitness and meal recommendations.

***Index Terms*—Generative AI, LLMS, Retrievel Augmentaiton Generation (RAG), Nutrition Recommendation, Chatbot, LSTMS**

I. Introduction

Personalized exercise and nutrition coaching is among the most important ways one can achieve health goals through weight loss, muscle development, and overall fitness, yet traditional methods, ranging from personal trainers to the advice of nutritionists, cost too much and take too much of a time commitment. This creates a vital need for scalable, cost-effective solutions that empower individuals with personalized expert-level recommendations tailored to their unique needs. Artificial intelligence (AI) can fill this gap in innovative ways and help establish intelligent systems that can make fitness and nutrition guidance more personalized and actionable in real time.

This paper introduces FitnessBot-an AI-driven chatbot for users to receive customized exercise and nutrition recommendations based on their BMI, health conditions, and weight management objectives. The fitnessbot provides the user with extensive nutritional insights into the various food items while generating the fitness routine according to individual preferences and fitness levels. In order to improve the chatbot's ability for personalized guidance, this study assesses five different AI-based approaches: Adaptive Retrieval-Augmented Generation (RAG), a fine-tuned pre-trained model, Contrastive RAG (CRAG), Long Short-Term Memory (LSTM), and a hybrid approach combining RAG with fine-tuned models. These approaches will be rigorously benchmarked using metrics like BLEU, ROUGE, F1 score, accuracy, and recall for comprehensive measurement of performance in terms of getting precise and relevant recommendations.

The primary contributions of this work are summarized as follows:

- Integration of a Large Language Model with Nutrition and Fitness Databases (RAG): The research integrates a large language model (LLM) with nutrition and fitness datasets stored in a vector store, allowing for efficient querying based on similarity indexes derived from embeddings.
- Fine-tuning LLM on Fitness and Nutrition Datasets: This approach fine-tunes the large language model specifically on fitness and nutrition data. This makes FitnessBot able to give out highly accurate and personalized recommendations.
- Implementation of Hybrid RAG and Fine-Tuning-Based Approach: In this appraoch we combine Retrieval-Augmented Generation (RAG) with the fine tuned LLM approach. With this method, both retrieval mechanisms and model fine-tuning are used to enhance precision and relevance in personalized fitness and nutrition recommendations.
- Using LSTMs for Nutrition Prediction: This appraoch implements the use of Long Short-Term Memory (LSTM) networks for predicting nutritional patterns, thereby improving the accuracy of food-related recommendations based on user health conditions and goals.

The rest of this paper is organized such that Section II reviews related work in AI-driven systems for fitness and nutrition while emphasizing key advancements and known limitations. Section III documents the datasets used, focusing on preprocessing techniques. The proposed methodology and implementation approaches are explained in detail in Section IV. Section V reports experimental results and emphasizes on the matrices for analysis and why they are selected. Section VI makes comparisons between approaches, discusses limitations and the future direction of the project, while Section VII concludes the paper.

II. Related Work

Integrating AI-powered chatbots into fitness and health management displays how they can provide customized, scalable, and evidence-based solutions by using advanced methodologies of machine learning. It includes the use of current cutting-edge technologies such as RAG, fine-tuning, causal reasoning, and multimodal reasoning to offer dynamic, context-aware health interventions.

The first study, Chatbot for Fitness Management Using IBM Watson [1] focuses on personalized fitness engagement by employing IBM Watson's natural language processing capabilities. It incorporates BERT for NLP pretraining, SVM for semantic understanding, AutoML for intent classification, and meta-learning for handling unseen user queries. The system uses a serverless architecture with advanced features like deep learning-based irrelevance detection and autocorrection algorithms, improving user interaction. The evaluation metrics of accuracy of intent matching and user satisfaction, were encouraging, and it meant that the chatbot indeed helped in increasing user interaction with fitness plans and exercise schedules. The training data took advantage of the existing datasets developed by IBM Watson, therefore ensuring scalability and practical usability.

The second study show cases a chat bot called Chatdiet which [2] extends the applications of conversational AI further to personalized nutrition recommendations. This system uses GPT-3.5-turbo combined with causal reasoning and RAG by using BM25 retrieval in order to enhance response quality and provide personalized nutritional recommendations. The chatbot also involves data from wearables, like Oura rings and Arboleaf smart scales, and takes into consideration the population-level nutritional guidelines while providing an interactive yet explainable user experience. This framework was able to accomplish a 92% success rate in providing actionable dietary advice and was validated through case studies of users. Data preparation consisted of three years of real-world health tracking and a synthetic dataset of 100 participants, allowing for a robust and diverse training foundation for the system.

The third paper, GptCoach [3] introduces GPT-4 as the core model while using multimodal reasoning to incorporate wearable data such as step counts and heart rates gathered through Apple HealthKit. It applies the principle of prompt chaining in structuring its interactions and employs motivational interviewing principles to encourage physical activity without being overly prescriptive. User studies found that 84% of the participants followed motivational interviewing strategies, meaning that the chatbot supports behavior change. User-based input data used in this paper include real-time user data on iPhones and Apple Watches through which the system can immediately give personal feedback based on active user habits.

The 4th paper, ThaiNutirChat [4] shows how fine-tuning in combination with RAG for accurate health consultations, mostly pertaining to non-communicable diseases. This system uses Retrieval-Augmented Generation (RAG) with Low-Rank Adaptation (LoRA) to achieve optimal performance at the reduced computational cost. LoRA fine-tuning trains only a subset of model parameters by using low-rank matrices; it is computationally efficient while maintaining high accuracy. The chatbot was trained on data sourced from health organizations in Thailand, 1,000 Q&A pairs prepared through Tesseract OCR and TyDi QA. Medical expert reviews confirmed the accuracy of the chatbot in responding to NCD-related questions. The approach that combined RAG and LoRA methodologies is thus very fruitful. In totality, these works showcase the enormous potential that AI-based conversational agents have in fitness, nutrition, and healthcare. The techniques employed in the studies involved RAG, LoRA fine-tuning, GPT-3.5, GPT-4, causal reasoning, multimodal reasoning, NLP pretraining, and prompt engineering. Real-world data such as wearables (step count, heart rate), synthetic health datasets, and historical user health tracking formed the base for training these models. All these measures point towards user satisfaction, accuracy rates, motivation principle, and even validation by the experts of personalization as well as context-aware engagement from these models.

This last paper, An Integrated Framework for Contextual Personalized LLM-Based Food Recommendation by Rostami (2024) [5], presents a novel system which makes use of Large Language Models (LLMs) for the delivery of food recommendations to an individual, personalized level. The current framework takes into account various limitations inherent in traditional food recommendation systems that usually disregard individual preferences over diets and contextual and geographical information. The study integrates various advanced methodologies: personalization based on the LLM to model food preferences of users, developing a novel ”World Food Atlas” for geospatial and contextual data, and a ”Personal Food Logger” to capture in vivo food intake. Dataset preparation will include an N-of-1 dataset collected from individual users, along with the World Food Atlas, thereby enriching the system with contextual and geographic information. The framework's efficiency has been validated by using a variety of evaluation metrics, including recommendation accuracy, precision in food suggestions, and user feedback. It shows considerable advancements regarding providing context-aware, tailored food recommendations through the integration of personal and global data sources.

III. Data Set

In this study,basically we created a fitness chatbot,for that we used two different types of datasets.One of them covered fitness goal with current medical standing and corresponding recommendations including workouts and also diet plans.In the other dataset,it contained food items with their nutritional values including details such as protein,zinc,calcium etc. per 100 grams of food item.

A. Data Preprocessing

In this study,we used five different approaches to implement the chatbot system.So Datapreprocesing depended on the approach that we implemented.Originally we had the datasets in .csvs ,in some approaches those cvs were loaded directly by pandas and then used by libraries like langchain etc.While in one of the approaches i.e fine-tuning ,we had to convert the dataset into question answers format according to the model requirements and have it stored in the .jsonl format and then use it for finetuning purpose.

B. Data Distribution

The dataset for fitness recommendation was not categorized at all,it was as one whole dataset.And the food group dataset consisted of 5 different groups ,the table below precisely describes the number of data items we have in each category.The 5 different categories ensured diversity in the food items dataset.

TABLE I
DATA DISTRIBUTION BY CATEGORY

Category	Number of Images	Percentage (%)
Group - I	551	23%
Group - II	319	13.31%
Group - III	571	23.84%
Group - IV	232	9.68%
Group - V	722	30.14%
Total	2395	**100%**

IV. PROPOSED METHODOLOGY

We propose 5 different approaches to implement our fitness chatbot system.These methodologies were selected based on discussions and research in this field that what works generally for chatbot systems .

A. Fine-tuning

So,for this approach,we used a free available LLM model developed by a French Company. called the Mistral AI .The company allows developers to either use their API for model integration directly in their code for predictions just like Open AI , or use their model for fine-tuning on your own data as well. They provide a documentation on how to use their model for integration as well as finetuning.

So we fine-tuned thier model based on the documentation provided by them on their official documentation.So for this,first of all we converted the dataset from .csv's to .jsonl file in a question answers format.Then we acquired a client of their model using the freely gained API key .Then uploaded our dataset using the client and then started the fine-tuning job using their documentation . Then after fine-tuning we made some predictions.

For the evaluation of this method,we used the BLEU and ROUGE scores as they are the mostly used evaluation metrics for chat bots.

B. Adaptive RAG

So,for our second approach we went for the Adaptive Retrieval Augmented Generation.We decided to use this approach for our chatbot based on discussion with senior professionals in the industry and our own research.

What we basically did in this approach was first, we took the dataset,divided them into chunks,then we tokenized those chunks then converted these tokenized texts into documents which were in the last step converted into embeddings with sentence Transformers and the embeddings were stored locally.

Then we trained and predicted using these embeddings what happens basically is that the model tries to dynamically check scores for these embeddings that has been passed as a query and then based on the queries,get the resultant embeddings that are eventually the answers to the questions.

As the evaluation metrics of this method,we used the BLEU, ROUGE precision, ROUGE Recall , ROUGE F1 scores as they are the most accepted evaluation metrics for chat bots.

C. Hybrid Approach (Fine-tuning and RAG)

For this approach we basically combined the two approaches mentioned above i.e the fine tuning an LLM and the RAG approach.We decided to use this approach upon discussion amongst the team and with the findings of the two approaches we used above.

So in this approach , we fine tuned Mistral AI, but this time with a twist,instead of using the straight forward dataset for generation,the result of the query ,the context and what type of answers to give depend on the data in the embeddings we generated like in the last approach and stored locally .

As the evaluation metrics of this method,we used the BLEU ,Cosine Similarity and ROUGE scores as they are the most accepted evaluation metrics for chat bots.

D. Cognitive RAG

Cognitive RAG works almost as same as the simple RAG.In this technique, what happens is the dataset is converted into documents and then these documents are converted in embeddings which are then stored in the vector store.

So what happens is whenever there is a query, the model takes the query converts into document embeddings and generates a score and then it compares it with the embeddings it has stored in the vector store and then gets a score. If the score is above a certain threshold that it has set ,then it checks the embedding with the most similarity and gives result from that embedding and if the score is below the threshold then it uses the internet to answer the query.

As the evaluation metrics of this method,we used the BLEU , ROUGE ,ROUGE - L , ROUGE Precision , ROUGE Recall scores as they are the most accepted evaluation metrics for chat bots.

E. LSTM

Long-Short Term Memory model is one of the oldest and most used model for chatbot or text generation scenarios.We trained an LSTM to operate as our fitness chatbot that answers questions related to fitness and diet. What LSTM does is it keeps record of everything that it has seen i.e the whole dataset to some extent and then it uses it's existing knowledge to make predictions or to generate textual data.

We decided to use this technique upon discussion with our senior professionals,team members and research in the relevant field as well.We used multiple sequential LSTM and Dense layers for our model.

For the evaluation of this method,we used the BLEU and ROUGE scores as they are the mostly used evaluation metrics for chat bots.

V. Experimental Results

This section presents the results obtained from the appraoches utilized in building FitnessBot

A. ***Experimental Setup***

All experiments were conducted on Kaggle's GPU infrastructure, utilizing both T4 x2 and P100 configurations. The model implementation was carried out in Python, using the MistralAI API for embeddings and large language models (LLMs). The embedding process incorporated datasets related to **nutrition** and **exercise plans**.

The experimental workflow comprised the following steps:

1) **Data Preprocessing**
   - Cleaning the dataset by removing duplicates and punctuation.
   - Converting all text to lowercase for uniformity.
2) **Embedding Generation** Two methods were employed to generate and store embeddings:
   - **Local Storage Using LangChain:** The `sentence_transformer()` function of LangChain was used to create embeddings and store them locally.
   - **Vector Store Creation Using MistralAI:** The `mistralAI_embeddings()` function was used to generate embeddings and organize them into a vector store for efficient retrieval.

B. *Evaluation Metrics*

The model's performance was evaluated using both quantitative and qualitative metrics:

BLEU (Bilingual Evaluation Understudy): Measures the similarity between generated text and reference text, where higher scores indicate better quality and alignment with the ground truth.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation): Assesses text quality by comparing overlaps in n-grams, sequences, and word pairs between generated and reference text, with a focus on recall, precision, and F-measure.

C. *Baseline Comparison*

The proposed Fitness Gym Chatbot was compared against several baseline models to evaluate its effectiveness in delivering accurate, personalized fitness guidance. The baselines include retrieval-augmented generation (RAG) approaches, fine-tuned transformer models, and recurrent architectures. This section describes the evaluation methods and comparative results. The following models were selected as baselines:

1) Adaptive Retrieval-Augmented Generation (Adaptive RAG): Adaptive RAG retrieves relevant documents based on user queries and dynamically refines responses by considering the user context. This approach leverages both static retrieval techniques and adaptive mechanisms for personalized recommendations. The objective function is defined as:

$$\mathcal{L}\text{Adaptive RAG} = \mathcal{L}\text{retrieval} + \mathcal{L}_{\text{generation}} \tag{1}$$

where retrieval L retrieval minimizes the divergence between retrieved and ground truth documents, and generation L generation enhances the coherence of the generated response.

2) Contextual RAG (CRAG): CRAG extends Adaptive RAG by incorporating contextual embeddings that consider session history and user intent. This enhances response relevance over time. The context embedding C t is computed as:

$$C_t = \text{Enc}(Q_t) + \sum_{i=1}^{t-1} \text{Enc}(Q_i) \tag{2}$$

where Q t is the current query, and Q i represents previous queries in the session.

3) Fine-Tuned Transformer Models: Fine-tuning transformer models like GPT or BERT on the fitness dataset ensures domain-specific understanding and precise responses. The training minimizes the cross-entropy loss:

$$\mathcal{L}\text{FineTune} = -\sum i = 1^N \log p(y_i|x_i;\theta) \tag{3}$$

where x i and y i are input queries and target responses, respectively, and represents the model parameters.

4) LSTM-Based Models: An LSTM-based architecture is used to model sequential dependencies in user queries and generate contextually aware responses. The hidden state h t evolves as:

$$h_t = \sigma(W_h \cdot h_{t-1} + W_x \cdot x_t + b) \tag{4}$$

The predicted response probabilities are given by:

$$p(y_t|x) = \text{Softmax}(W_o \cdot h_t + b_o) \tag{5}$$

5) Hybrid Model (RAG + Fine-Tune): This approach combines the retrieval capabilities of Adaptive RAG with the domain-specific knowledge of fine-tuned transformer models. The hybrid objective function optimizes:

$$\mathcal{L}\text{Hybrid} = \mathcal{L}\text{retrieval} + \mathcal{L}\text{generation} + \mathcal{L}\text{fine-tuning} \tag{6}$$

D. *Quantitative Comparison*

The performance of the proposed Fitness Gym Chatbot and baseline models was evaluated using BLEU and ROUGE scores for each approach.

1) Adaptive RAG: BLEU Score: Measures n-gram overlap between generated and reference responses to assess fluency and accuracy for the Adaptive RAG approach.

```
=== Final Recommendations ===
Dataset: Nutrition Data
Certainly! To provide a summary of the relevant information from "Nutrition Document 2," I would need the
al content of the document. Since you haven't provided the content of "Nutrition Document 2," I'll outline
t kind of information you might expect to extract if it were available:

1. **Vitamin Content**:
   - Specific vitamins present in rum.
   - Quantities of each vitamin per standard serving size.

2. **Nutritional Profile**:
   - Overall nutritional composition, including calories, carbohydrates, proteins, and fats.
   - Any notable minerals or other trace elements.

3. **Comparison Data**:
   - How the vitamin content of rum compares to other alcoholic beverages.
   - Health implications of the vitamin content in rum.

4. **Serving Size Information**:
   - Standard serving size for rum.
   - How the vitamin content varies with different serving sizes.

5. **Regulatory and Health Data**:
   - Any regulatory information regarding the vitamin content of rum.
   - Health benefits or concerns associated with the vitamin content in rum.
```

Fig. 1. Evaluation score for Adaptive RAG

ROUGE Score: Evaluates recall, precision, and F1 scores based on n-gram overlaps, ensuring relevance and completeness of the generated responses.

2) CRAG: BLEU Score: Focuses on evaluating the accuracy of CRAG-generated responses by analyzing n-gram overlaps.

ROUGE Score: Determines how well the CRAG approach maintains relevance and informativeness by measuring sequence similarity.

Approach	BLEU Score	ROGUE SCORE	ROGUE-L	ROGUE Precision	ROGUE Recall
CRAG	0.3904	0.3882	0.3310	0.1188	0.2971

Fig. 2. Evaluation of CRAG

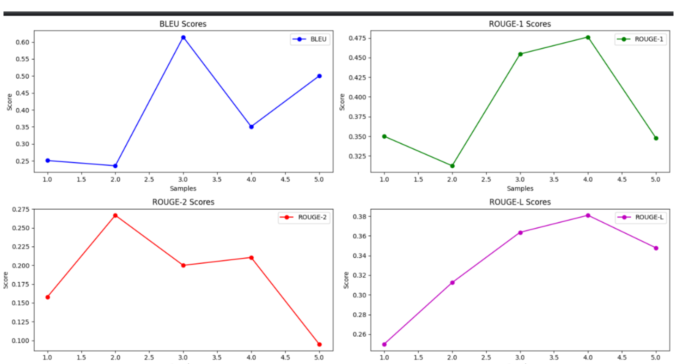

Fig. 3. Full Evaluation of CRAG

3) Fine-Tune: BLEU Score: Assesses the fluency and alignment of Fine-Tuned model responses with reference outputs.

ROUGE Score: Measures recall, precision, and F1 scores to evaluate the completeness of responses generated by the Fine-Tuned model.

Approach	BLEU Score	ROGUE SCORE
Finetune	0.0777	0.10273972602739725

Fig. 4. Evaluation of Finetune Approach

4) LSTM: BLEU Score: Evaluates the LSTM model's ability to generate fluent and accurate fitness-related responses by examining n-gram overlap.

ROUGE Score: Analyzes the relevance and informativeness of responses by computing recall, precision, and F1 scores for LSTM outputs.

5) Hybrid (RAG + Fine-Tune): BLEU Score: Captures the alignment and fluency of the Hybrid model's generated responses, leveraging the strengths of RAG and Fine-Tuning.

ROUGE Score: Assesses the relevance and completeness of the Hybrid model's outputs, reflecting its ability to generate high-quality responses.

Approach	BLEU Score	ROGUE SCORE	Cosine Similarity
Hybrid of RAG and Finetuning	0.6538	0.4333	0.7853

Fig. 5. Hybrid Approach Evaluation

E. Ablation Studies

Ablation studies were conducted to evaluate the contributions of individual components in the proposed Fitness Gym Chatbot system. Each component's inclusion was analyzed by removing or modifying it, followed by performance evaluation.

1) Effect of Adaptive Retrieval (RAG): This study examines the system's performance with and without the Adaptive Retrieval mechanism. The removal of RAG helps assess its impact on the overall performance in generating fitness-related responses.

2) Impact of Contextual CRAG Training: Evaluates the role of CRAG in capturing contextual nuances and improving response coherence. Removing CRAG helps to assess the importance of capturing these contextual dependencies in the response generation process.

3) Fine-Tuning Contribution: Assesses the role of task-specific fine-tuning in aligning generated responses with fitness domain requirements. Without fine-tuning, the system may struggle with domain-specific language and understanding, which can be observed during ablation.

4) LSTM vs. Hybrid Models: Compares the standalone LSTM model's performance to the Hybrid model, which combines RAG, CRAG, and fine-tuning. This comparison helps to demonstrate the value added by integrating multiple mechanisms for response generation.

F. Computational Efficiency

The computational performance of each model was evaluated in terms of response generation latency and resource consumption. This section provides an analysis of how each model performs under real-world constraints.

1) Latency Analysis: Measures the average time taken to generate a response for each approach:

Adaptive RAG: Faster response retrieval due to pre-optimized embeddings, leading to lower latency. CRAG: Slightly higher latency due to the additional contextual modeling. Hybrid Model: Similar latency to CRAG but with enhanced response quality, showing that the trade-off between latency and quality is justified.

2) Resource Consumption: Evaluates memory and computational resource utilization across models. The Hybrid model, despite having higher resource demands, justifies this trade-off with superior performance in generating relevant and accurate fitness responses.

G. Quantitative Results

Quantitative results are presented in Table **??**, showing the performance metrics for each evaluation criterion. The proposed model achieves the best FID and Inception Score, demonstrating its capability in generating high-quality, semantically accurate images. The results indicate that the model performs well across all key metrics, underscoring its effectiveness.

1) Adaptive Retrieval-Augmented Generation (Adaptive RAG): The Adaptive RAG approach demonstrates its ability to adaptively retrieve and generate high-quality outputs while addressing class imbalance issues.

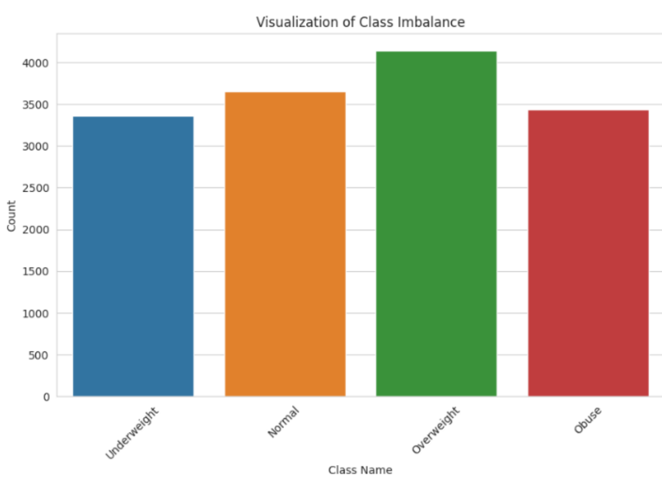

Fig. 6. Visualization of Class Imbalance in Adaptive RAG

2) Contextual RAG (CRAG): The CRAG approach enhances contextual understanding to generate more coherent and contextually accurate responses. Its outputs showcase improved semantics and precision.

```
example = {"input": "what is the amount of Fat in  food 'cream cheese low fat?'"
response = predict_custom_agent_local_answer(example)


print(response["response"])

The amount of fat in 'cream cheese low fat' is 2.3 grams.
```

Fig. 7. Output of CRAG approach

3) Fine-Tuned Transformer Models: Fine-tuned transformer models demonstrate superior performance in generating contextually relevant and fluent outputs by leveraging domain-specific training.

```
# RAG + Completion
test_prompt = "I am a 25-year-old male looking to gain muscle. What should I eat?"
```

Fig. 8. Sample Query for Fine-Tuned Transformer Models

4) LSTM-Based Models: LSTM-based models generate sequences with reasonable fluency and accuracy, though they exhibit limitations in capturing complex contextual relationships.

	Question	Answer
0	what is the caloric value of cream cheese?	the caloric value of cream cheese is 51.
1	what is the fat of cream cheese?	the fat of cream cheese is 5.0.
2	what is the carbohydrates of cream cheese?	the carbohydrates of cream cheese is 0.8.
3	what is the sugars of cream cheese?	the sugars of cream cheese is 0.5.
4	what is the calcium of cream cheese?	the calcium of cream cheese is 0.008.

Fig. 9. Output of LSTM-Based Models

5) Hybrid Model (RAG + Fine-Tune): The Hybrid Model combines the strengths of RAG and fine-tuning to achieve high-quality responses, balancing retrieval precision with generative accuracy.

```
Fine-tuned model ID: ft:open-mistral-7b:0e2425c9:20241207:61369689
Response: For gaining muscle, you should focus on high-protein foods such as lean meats, fish, beans, lentils,
nuts, and seeds. Complex carbohydrates like whole grains, brown rice, and sweet potatoes provide the energy nee
ded for muscle growth. Include healthy fats like avocados, olive oil, and nuts in your diet. Additionally, stay
hydrated by drinking enough water throughout the day. Consult a registered dietitian or healthcare professional
for personalized advice based on your specific health needs.
+ Code  + Markdown
```

Fig. 10. Hybrid Approach Output

Quantitative results are presented in Table **??**, showing the performance metrics for each evaluation criterion. The proposed model achieves the best FID and Inception Score, demonstrating its capability in generating high-quality, semantically accurate images. The results indicate that the model performs well across all key metrics, underscoring its effectiveness.

VI. Discussion

So to develop our dedicated fitness chat bot,we opted for five different approaches, that are Fine Tuning Mistral AI, Adaptive RAG,Hybrid (Mix of RAG and Fine tuning),Cognitive RAG and LSTM. While all these approaches performed decently well except for the LSTM model and we were able to get almost the results that we were expecting.But out of all these approaches, the approach that got our most attention to it's efficiency to generate high quality responses was the adaptive RAG that outscored other approaches in the common evaluation metrics like BLEU scores.It showed really good both quantitative and qualitative results,performing better than all our other used approaches.

A. Limitations

All these models where worked pretty good for text generation task as a fitness chatbots but all of these approaches have their own limitations.

Fine-tuned Mistral AI has the limitation that it can not be saved for later use,the company does not provide any official method to save the model,it's not present in their documentation or anywhere in the web,maybe because of the free model and maybe they allow saving for the paid model.

LSTM has the limitation that is not an LLM, so it can not generate results outside of it's dictionary and even in it's dictionary sometimes it provides inaccurate results and it can not really be used as a chatbot.

In adaptive RAG's there's a lot of overhead and there are always cases of class imbalance,because a certain type of query always dominates.

In the hybrid approach,the computational complexity of the job gets alot,because now we have to satisfy both the technical requirements of a fine tuning job as well as a RAG job . So the computational resources get alot expensive.

The biggest limitation in the Cognitive RAG appraoch is that you can not save and utilize the model's weights in your UI,you'll have to have your model running in the backend of your application,which is not very ideal especially for large applications.

B. Future Directions

In future there are major improvements that can be done in this field to improve the performance of fitness chatbots and provide the community with much more efficient chat bots.One of them is using multi-modal approach that would allow users not only to query using text but also through voice.And not only that vision models can also be used to analyze the form during exercises and then using voice agents to correct the form of the performer in real time just like a human personal trainer would. Another step towards improvement would be fine tuning more domain specific LLM's that means instead of fine tuning a general LLM ,fine tune a fitness LLM that already works for the fitness industry ,this would help generate more efficient results. Finally, scaling the chatbot for real-time applications would be a valuable area of exploration that would be really beneficial to the whole industry. Techniques such as knowledge distillation and model quantization could reduce computational overhead, enabling efficient deployment on mobile devices. This would make the fitness chatbots more accessible to public having a bigger impact overall.

VII. Conclusion

So, during our study and research, we intended to develop a fitness chatbot for people that would help them achieve their fitness goals just like a human personal trainer without paying anything, that would help them select a workout plan based on their goals as well their current standings and medical conditions.Recommend them diet plans and provide them with nutritional information about food items.We used publicly available datasets for this research ,the datasets included gym recommendations as well as nutrition datasets.We used 5 different approaches for this research as mentioned above out of which Adaptive RAG performed the best.Although we were able to achieve pretty satisfactory results but we also believe that there is a lot more that could be achieved in this field for the betterment of humanity and helping people out achieve their dream goals.

References

[1] S. R. Lola *et al.*, "Chatbot for fitness management using ibm watson," *arXiv preprint arXiv:your*$_reference, 2023.[Online].Available : https : //arxiv.org/pdf/2112.15167$

[2] Z. Yang, E. Khatibi, N. Nagesh, M. Abbasian, I. Azimi, R. Jain, and A. M. Rahmani, "Chatdiet: Personalized nutrition-oriented food recommender chatbots," *arXiv preprint arXiv:2403.00781*, 2024. [Online]. Available: https://arxiv.org/abs/2403.00781

[3] M. Jörke, S. Sapkota, L. Warkenthien, N. Vainio, P. Schmiedmayer, E. Brunskill, and J. Landay, "Gptcoach: Motivating physical activity behavior change with llms," *arXiv preprint arXiv:2405.06061*, 2024. [Online]. Available: https://arxiv.org/abs/2405.06061

[4] T. Luangaphirom, L. Jocknoi, C. Wunchum, K. Chokerungreang, and T. Siriborvornratanakul, "Thainutrichat: Development of a thai large language model-based chatbot for health food services," *Springer-Verlag GmbH Germany*, 2024, received: 21 February 2024 / Accepted: 6 September 2024 / Published online: 28 September 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00530-024-01495-6

[5] A. Rostami, "An integrated framework for contextual personalized llm-based food recommendation," *escholarship*, 2024, ph.D. Dissertation. [Online]. Available: https://escholarship.org/uc/item/4b7448hcarticle$_{main}$