FitBuddy: An Intent-Based AI Chatbot for Personalized Fitness Assistance Using BERT and GPT-2 Models

Amritjyot Singh
Lovely Professional University,Chaheru,Punjab,India
amritjyotsingh2004@gmail.com

Pudi Sai Venkat Sankeerth Naidu
Lovely Professional University,Chaheru,Punjab,India
sankeerth2004@gmail.com

Dr. Dhanpratap Singh
Dept. of Computer Science and Engineering
Lovely Professional University,Chaheru,Punjab,India
dhanpratapsingh@gmail.com

***Abstract*— As the demand for intelligent personal assistants grows, chatbots have become a valuable tool in various domains, including health and fitness. In this paper, we present FitBuddy, a fitness-focused chatbot that combines intent classification and response generation to provide users with personalized and relevant guidance. Leveraging a BERT-based intent classification model and a GPT-2 natural language generation model, FitBuddy aims to improve user experience by understanding user intent before formulating a response.**

To evaluate the impact of intent classification on response accuracy, we perform a comparative analysis between two versions of a chatbot: one that classifies intents before response generation and one that generates responses without intent guidance. Our results suggest that incorporating intent classification results in more contextual responses and improves interaction quality. This study highlights the effectiveness of intent classification for optimizing chatbot responses, especially in specialized applications, and provides insights for future improvements in AI-driven personal assistants.

Keywords*— Chatbot, Intent Classification, Response Generation, Natural Language Processing (NLP), BERT Model, GPT-2 Model, Fitness Assistant, Transformer Architecture, Comparative Analysis, Personalized Response Generation.

I. INTRODUCTION

Rapid advances in natural language processing (NLP) have revolutionized the development of chatbots. They have become more interactive and intelligent, supporting a variety of applications such as customer service, personal assistance, and learning tools [1]. One of the biggest challenges in creating an effective chatbot is accurately interpreting the user's intent, as this determines the relevance of the response, which in turn impacts the user's satisfaction and trust in the system [2].

In this study, we present FitBuddy, a chatbot tailored to provide users with personalized fitness advice. To achieve this, FitBuddy uses two advanced NLP models: a BERT-based intent classifier and a GPT-2-based response generator. The BERT model, developed by Google, is particularly well-suited for intent classification, as it is designed to understand context within language [3]. FitBuddy uses a fine-tuned BERT model to classify fitness-related intents such as "bodybuilding," "meal plan recommendations," "weight loss," and similar goals. By recognizing the intent behind a user's request, the chatbot can instruct the GPT-2 model to generate responses that better fit the user's specific fitness context [4].

In our study, we compare the effectiveness of FitBuddy's responses with intent classification against a baseline model that does not have this feature. Through this comparison, we aim to determine whether adding intent classification significantly improves the contextual accuracy of responses and ultimately improves the usability of fitness advice chatbots [5].

BERT and GPT-2 models provide different capabilities in FitBuddy's architecture. BERT, a transformer-based model, is pre-trained to capture bidirectional linguistic context, which is important for tasks that require intent understanding [6]. BERT's architecture consists of multiple self-attention and feedforward networks, which can capture complex word relationships within a sentence. For this purpose, a sequence classification header is added to the final layer of BERT to classify user queries into the desired relevance goal.

On the other hand, GPT-2 is an autoregressive language model from OpenAI that is known for generating consistent and contextually relevant text responses [7]. With a decoder-specific transformer layer and a self-attention mechanism, GPT-2 generates responses word by word, conditioning each subsequent word on the preceding one. Once intent classification is applied, GPT-2 receives additional prompts indicating the identified intent, allowing it to generate more targeted and consistent responses [8].

This work builds on existing research in chatbot design, where different architectures and techniques have been developed to improve the accuracy and flexibility of responses. Early chatbots mainly relied on rule-based or voting-based approaches, limiting their adaptability to different conversation scenarios [9]. However, the emergence of Transformer models such as BERT and GPT-2 has seen a significant shift toward using deep learning for more flexible and context-aware chatbots [10].

Previous research has highlighted the importance of intent classification in chatbot design, especially in goal-oriented systems, as it helps reduce irrelevant responses and improve user satisfaction [11]. Furthermore, generative models such as GPT-2 have introduced more fluid and natural responses that are essential for creating engaging and human-like conversational experiences [12]. By combining intent classification and response generation, we aim to improve the overall accuracy of FitBuddy's responses and provide new insights into the role of intent-based context in chatbot systems designed for specific purposes such as fitness advice.

II. Related Work

The field of chatbot development has seen substantial advancements with the integration of Natural Language Processing (NLP) techniques, particularly through the use of transformer-based architectures. Historically, chatbots were primarily rule-based or retrieval-based systems, which often limited their ability to handle complex, open-ended conversations. These systems typically relied on predefined responses or templates, making them less adaptable to user inquiries that deviated from the expected patterns [13].

The introduction of transformer models, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), has revolutionized the approach to chatbot design. BERT, developed by Google, is particularly notable for its ability to understand the context of words in relation to all other words in a sentence. This bidirectional context comprehension makes BERT highly effective for intent classification tasks, where accurately identifying user intent is crucial. Research has demonstrated that employing BERT for intent classification significantly enhances a chatbot's capability to discern user goals, leading to more relevant and accurate responses [14]. For instance, chatbots utilizing BERT have been shown to outperform traditional models in various applications, including customer service and personal assistance [15].

On the other hand, GPT models—particularly GPT-2—excel in natural language generation. Unlike BERT, which is designed for understanding and classification, GPT-2 is an autoregressive model that generates text by predicting the next word in a sequence based on the context provided by previous words. This allows GPT-2 to produce fluent, coherent, and contextually appropriate responses, making it well-suited for applications requiring conversational engagement [16]. The flexibility and diversity of responses generated by GPT-2 contribute to a more natural interaction experience for users.

Despite the individual strengths of intent classification and generative models, there has been limited exploration of their combined potential, particularly in specialized domains such as fitness and health. While some studies have investigated the effectiveness of intent recognition in improving response accuracy, few have specifically compared chatbot performance with and without intent classification. Understanding this relationship is essential, as it can inform the design of more efficient chatbots that better cater to user needs [17].

Our research aims to fill this gap by analyzing FitBuddy, a fitness-oriented chatbot that integrates both intent classification and response generation. By comparing the chatbot's performance in scenarios where intent is classified against those where it is not, we seek to determine the impact of intent classification on the quality and relevance of generated responses. This study will contribute valuable insights into the design of AI-driven chatbots, particularly in specialized areas, enhancing user satisfaction and interaction quality [18].`

III. Motivation

The widespread use of artificial intelligence and natural language processing technologies has significantly changed the way people interact with machines. As users seek personalized and efficient support, chatbots have emerged as essential tools in various sectors, especially in the health and fitness sector [19]. The fitness industry, characterized by diverse user needs ranging from training plans to nutritional advice, poses unique challenges for chatbot systems [20].

The primary motivation for developing FitBuddy is its potential to improve user experience by providing customized fitness advice through intelligent conversation. Although existing chatbots have made progress in understanding user input, many chatbots still struggle to accurately interpret user intent, often resulting in generic or irrelevant responses. This limitation can lead to user frustration and disengagement, reducing the effectiveness of chatbots [21].

Our work is based on the hypothesis that incorporating intent classification into chatbot design improves response quality. By leveraging the BERT model for intent classification, we aim to improve chatbots' understanding of user queries and generate more relevant and contextual responses using the GPT-2 model [22]. This dual approach not only addresses the shortcomings of current chatbots, but also aims to provide a more enriching and productive user experience.

Furthermore, a comparative analysis of responses generated with and without intent classification highlights the importance of intent detection in chatbot performance. Understanding this relationship is critical to evolving intelligent conversational agents, especially in specialized domains such as fitness where user satisfaction is paramount. With FitBuddy, we aim to provide valuable insights that will inform the development of future AI-driven personal assistants and ultimately improve their ability to meet diverse user needs [23].

IV. Methodology

Overview of the Chatbot Architecture:

FitBuddy is a sophisticated chatbot designed to deliver personalized fitness guidance through intelligent user interactions. Its architecture comprises two essential components: Intent Classification and Response Generation, which allow it to accurately comprehend user inquiries and produce contextually appropriate responses.

High-Level Architecture

The architecture of FitBuddy may be conceptualized as a modular system comprising the following components:

User Interface: The front-end interface is where people interact with the chatbot, usually via text input. This might be implemented as a web or mobile application, giving consumers an intuitive experience.

Intent Classification Module: Powered by a BERT model, this module is in charge of processing user input to discover the underlying purpose. It converts user inquiries into actionable intentions that drive the answer creation process.

Response Generation Module: Using a GPT-2 model, this module converts the discovered purpose into coherent and meaningful replies. This design allows the chatbot to reply in a conversational manner, making interactions seem more genuine.

Intent Classification Component: The intent classification component employs the BERT model, a cutting-edge transformer-based model noted for its ability to understand the context of words in a phrase. When a user types a query, the following happens.

Text Preprocessing: User input is tokenized, padded, and converted into tensor format suitable for BERT.

Logits Generation: The processed input is fed into the BERT model, which outputs logits representing the likelihood of various predefined intents.

Softmax Activation: A softmax function is applied to convert these logits into probabilities for each intent, allowing for the identification of the most likely intent.

Response Generation Component

Once the intent is classified, the response generation component comes into play:

Input Formatting: The identified intent and the user's original query are formatted into a prompt for the GPT-2 model.

Text Generation: The GPT-2 model generates a response based on the provided prompt. It uses a sampling technique to create diverse and engaging responses, enhancing the conversational flow.

Output Processing: The generated response is processed to ensure clarity and relevance before being presented to the user.

Integration of Components

The seamless integration of intent classification and response generation is crucial for the effectiveness of *FitBuddy*. After the intent classification module identifies the user's intent, this information is directly utilized by the response generation module to tailor responses appropriately. This two-step process allows *FitBuddy* to provide contextually relevant advice, whether it pertains to workout suggestions, meal planning, or general fitness inquiries.

INTENT CLASSIFICATION:

Intent categorization is a critical component of the FitBuddy chatbot design, allowing it to reliably determine user intentions based on input questions. This method lays the groundwork for creating meaningful and context-aware replies. This section describes the methodology, model selection, dataset preparation, and training procedure used for intent categorization.

1. **Model Selection: BERT**

For intent classification, *FitBuddy* employs **BERT (Bidirectional Encoder Representations from Transformers)**, a state-of-the-art language representation model developed by Google. BERT is particularly well-suited for this task due to its ability to capture the contextual relationships between words in a bidirectional manner, which allows it to understand nuances in user queries effectively.

Bidirectional Contextualization: Unlike traditional models that analyze text in a unidirectional manner (left to right or right to left), BERT considers the context from both directions. This capability helps in understanding the intent behind complex queries, including those with ambiguous phrasing.

Fine-Tuning Capability: BERT can be fine-tuned on specific datasets, allowing it to adapt to the unique intents relevant to fitness-related conversations.

2. **Dataset Preparation**

The effectiveness of the intent classification model largely depends on the quality and diversity of the training data. For *FitBuddy*, the following steps were taken to prepare the dataset:

Data Collection: A diverse set of user queries was collected, encompassing various fitness-related intents such as Body Building, meal recommendations, meditation suggestions, and more. This data could come from user interactions and FAQs from various Fitness websites, forums, and surveys.

Intent Mapping: Each query was annotated with a corresponding intent label, which allows the model to learn associations between user input and expected responses. The intents can be categorized as follows:

- Body Building
- Meal Plan Recommendation
- Recommend Meditation or Yoga
- Suggest Recovery Exercises
- Weight Loss

Training and Test Split: The dataset was divided into training, validation, and test sets to evaluate the model's performance accurately. Typically, an 80-10-10 split is used, ensuring that the model learns from a comprehensive set of examples while being evaluated on unseen data.

3. **Data Preprocessing**

The data preprocessing involves following steps:

Tokenization: User queries are tokenized into sub-word units that BERT can process. Special tokens, such as [CLS] for classification and [SEP] for separating multiple inputs, are added to the sequences.

Input Representation: Each input query is converted into an embedding that includes positional encodings, enabling the model to understand the order of words.

4. **Training Process**

The training process involves fine-tuning the pre-trained BERT model on the prepared dataset:

Model Initialization: Bert model is initialized with pre-trained weights and specified number of labels (5 in this case) from Huggingface's Transformers Library

Loss Function: The cross-entropy loss function is utilized to measure the difference between predicted intent probabilities and the actual intent labels. The goal during training is to minimize this loss, thereby improving the model's accuracy.

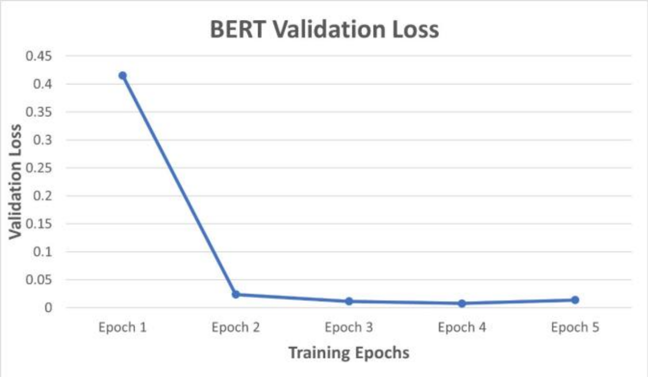
Optimization: An optimizer, such as Adam, is employed to update the model weights iteratively based on the calculated loss, enhancing the model's predictive capabilities.

Epochs and Batch Size: The model is trained for 5 epochs, adjusting the batch size to optimize learning efficiency. Typically, a batch size of 16 to 32 is used, depending on available computational resources.

5. **Evaluation**

After training, the intent classification model is evaluated using the validation and test sets. Key evaluation metrics include:

Validation Loss: Validation loss measure BERT's error on validation set that it has not seen during training, helping to assess its generalization ability.



RESPONSE GENERATION:

The response generation component of *FitBuddy* is pivotal in transforming identified user intents into meaningful, contextually relevant, and engaging responses. This section outlines the methodologies, model architecture, input processing, response formulation, and evaluation methods utilized for generating responses within the chatbot.

1. **Model Selection: GPT-2**

For generating responses, *FitBuddy* utilizes **GPT-2 (Generative Pre-trained Transformer 2)**, a powerful language model developed by OpenAI. GPT-2 is well-suited for conversational applications due to its ability to produce coherent and contextually appropriate text based on input prompts.

Transformer Architecture: GPT-2 is built upon the transformer architecture, which allows it to process and generate text in parallel, making it efficient for large-scale text generation tasks. The model uses self-attention mechanisms to weigh the significance of different words in the context of the entire input sequence.

Pre-trained Model: GPT-2 is pre-trained on a diverse corpus of text, which equips it with a broad understanding of language, idioms, and common conversational patterns. This pre-training allows *FitBuddy* to generate responses that are not only relevant but also natural and engaging.

Response Generation Mechanism

The core of the response generation relies on the GPT-2 model's ability to predict the next word in a sequence based on the input prompt:

Text Generation: Using the tokenized input, GPT-2 generates a sequence of text by predicting the next token iteratively until a stopping criterion is met (e.g., reaching a maximum length or generating an end-of-sequence token).

Sampling Techniques: To enhance the variability and creativity of the responses, sampling techniques such as top-k sampling and top-p (nucleus) sampling are employed:

Top-k Sampling: This method selects from the top K highest probability tokens at each generation step, ensuring diversity in the output while maintaining relevance.

Top-p Sampling: Here, a cumulative probability distribution is used, selecting tokens until the total probability exceeds a specified threshold P. This approach balances quality and variability effectively.

Output Processing

After the response is generated, it undergoes a processing phase to ensure clarity and appropriateness:

Post-Processing: The generated token IDs are converted back to text using the GPT-2 tokenizer, and any special tokens (like end-of-sequence markers) are removed.

Response Refinement: The response may be further refined or filtered to maintain conversational quality and adhere to guidelines for user engagement. This could involve correcting grammatical issues or ensuring that the response aligns with the user's intent.

2. **Dataset Preparation**

The quality and applicability of the produced replies are significantly influenced by the dataset used for GPT-2 training. The following actions were taken for FitBuddy:

3. **Data collection**: The dataset includes a variety of fitness-related discussion samples, such as user questions and answers about rehabilitation, food advice, exercise regimens, and mindfulness techniques.
4. **Structure of Dialogue**: Every conversation entry is structured as a series of questions and answers to accommodate both user input and expected chatbot responses. This method helps GPT-2 comprehend the conversational flow and enhances its capacity to respond to user inquiries.
5. **Preprocessing of Data**

The data preprocessing entails the following crucial steps:

Tokenization: Every conversation input is divided into tokens that are appropriate for the GPT-2 format by the GPT-2 tokenizer. The model generates replies by encoding each sequence into number tokens.

Sequence Truncation and Padding: Each discussion sequence is either truncated or padded to suit a standard block size, usually 128 tokens, in order to accommodate different input lengths.

6. **Procedure for Training**

GPT-2 is refined during the training phase using the fitness conversation dataset:

Model Initialization: Pre-trained weights from OpenAI's model repository are used to initialize GPT-2. The chat dataset is then loaded for fine-tuning.

Loss Function: The model's relevance in response production is increased by using cross-entropy loss during training to reduce the discrepancy between generated replies and predicted outputs.

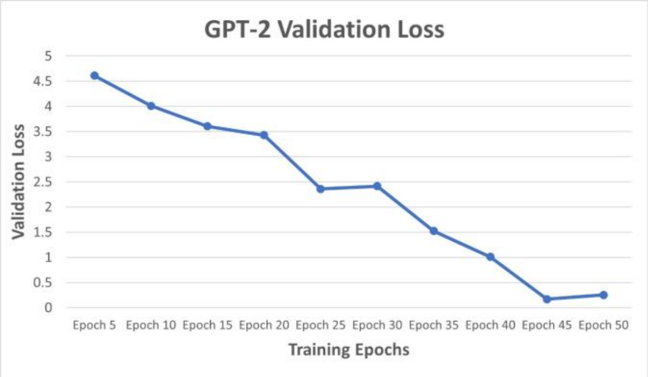
Optimization: To increase the model's capacity to generate text, an optimizer, like Adam, iteratively modifies the model weights in response to the calculated loss.

Batch Size and Epochs: The model is trained across 50 epochs, increasing its response generation incrementally with each one. Depending on the available computing power, batch sizes of 8 or 16 are usually employed.

7. **Evaluation:**

Following training, the effectiveness of GPT-2 in producing responses is evaluated using:

Validation Loss: This measure assesses GPT-2's loss on the validation set and provides information on how well it generalizes to new data. The ability of the model to generate accurate responses outside of its training data is shown by a smaller validation loss.



User Interaction Quality: To make sure that replies are logical, pertinent, and well aligned with user goals, response quality is evaluated subjectively in addition to quantitative measures.

Integration of Intent Classification and Response Generation:

The effectiveness of *FitBuddy* as a fitness-oriented chatbot hinges on the seamless integration of its **Intent Classification** and **Response Generation** components. This integration allows the system to not only understand user queries accurately but also to respond in a contextually appropriate and engaging manner. In this section, we detail the workflow, data flow, and key functionalities that illustrate how these two components work together to enhance user interactions.

1. **Workflow Overview**

The integration process consists of a structured workflow that guides the user interaction from input to response:

User Input: The interaction begins when a user inputs a query into the chatbot interface. This query is typically a question or statement regarding fitness, nutrition, or wellness.

Intent Classification: The user input is first processed by the intent classification module powered by the BERT model. This component analyses the text and predicts the user's intent, outputting a corresponding intent label.

Verification of Intent: After Predicting the intent the chatbot verifies its Prediction from user. If user responds that the intent is correctly predicted then chatbot proceed for contextual prompt creation. Otherwise, it asks user to mention the actual intent and then proceeds further with the correct intent.

Contextual Prompt Creation: Based on the identified intent, a structured prompt is created, combining the original user query and the classified intent. For example, if the user's query is about workout plans, the prompt might be structured as follows: "[Q] How do I build muscle? My intent is Body Building." This prompt serves as the input for the response generation module.

Response Generation: The formatted prompt is then passed to the GPT-2 model, which generates a relevant and coherent response tailored to the specified intent.

Response Delivery: The generated response is processed for clarity and presented back to the user, concluding the interaction cycle.

2. **Data Flow**

The data flow between the intent classification and response generation modules is critical for ensuring smooth operation and user satisfaction:

Input Transformation: Upon receiving user input, the system transforms the text into a suitable format for the BERT model, which includes tokenization and embedding.

Intent Output: Once BERT predicts the intent, this information is verified from user and the actual intent is encapsulated along with the user query to create a comprehensive prompt for the GPT-2 model.

Prompt Processing: The generated prompt is then tokenized and prepared for input into the GPT-2 model, which utilizes the context provided by combining the intent with user query to produce a relevant response.

V. Functional Interdependencies

The integration of these components creates functional interdependencies that enhance the overall performance of the chatbot:

1. **Personalization**: The intent classification informs the response generation process by guiding the chatbot on the type of response that is most relevant. For instance, if a user's intent is classified as "Meal Plan Recommendation," the GPT-2

model is primed to generate dietary advice rather than unrelated fitness tips. This results in more personalized and contextually relevant interactions.

User Interaction Experience

The integrated architecture ultimately aims to enhance the user interaction experience by providing timely, relevant, and engaging responses:

Conversational Flow: The interplay between intent classification and response generation creates a fluid conversational experience. Users can feel confident that their queries will be understood and addressed accurately, leading to increased engagement and satisfaction.

Reduced Response Time: The efficient data flow between components ensures that responses are generated quickly, minimizing waiting times and creating a more responsive user experience.

In conclusion, the integration of intent classification and response generation in FitBuddy forms a cohesive system that not only accurately understands user queries but also provides meaningful and context-aware responses. This synergy is essential for delivering a high-quality user experience in the domain of fitness and wellness, positioning FitBuddy as a valuable tool for individuals seeking personalized fitness guidance.

RESPONSE GENERATION:

The user interaction flow is a critical aspect of the *FitBuddy* chatbot, as it defines how users engage with the system from the initial query to receiving tailored fitness advice. This section outlines the step-by-step flow of interaction, emphasizing the processes involved in understanding user intents, generating responses, and enhancing the overall user experience.

1. **User Entry Point**

The interaction begins at the user interface, where individuals can access *FitBuddy* via various platforms, such as a web application, mobile app, or messaging platform. Users are greeted with a friendly prompt encouraging them to ask questions or seek fitness advice.

Welcome Message: A standard greeting message, such as "Hello! How may I help you with your fitness today?" sets the tone for a supportive and engaging interaction.

2. **User Input**

Users can enter their queries in natural language. These queries may pertain to various fitness topics, including workout routines, nutrition, wellness tips, or specific fitness goals.

Query Examples: Users might ask questions like:

"What should I eat for muscle gain?"

"Can you recommend a yoga routine for beginners?"

"How do I lose weight effectively?"

3. **Intent Classification**

Once the user submits a query, it is processed by the intent classification module, which employs a BERT-based model to analyse the text and identify the underlying intent.

Tokenization and Encoding: The input text is tokenized and transformed into embeddings that the model can process.

Intent Prediction: The model predicts the intent associated with the user's input. For example, it may classify the input as "Meal Plan Recommendation" or "Suggest Recovery Exercises."

4. **Confirmation of Intent**

To ensure accuracy, *FitBuddy* confirms the identified intent with the user. This step enhances engagement and allows for user input in case the classification was incorrect.

Confirmation Prompt: For instance, the chatbot might respond: "I believe your intent is Meal Plan Recommendation. Is that correct? Please answer yes or no."

User Feedback: The user responds with "yes" or "no." If the user selects "no," *FitBuddy* prompts them to clarify their intent, allowing for a more accurate understanding.

5. **Response Generation**

Upon receiving confirmation of the intent, *FitBuddy* generates a contextually relevant response using the GPT-2 model.

Input Formatting: The confirmed intent and the user query are combined into a structured prompt (e.g., "[Q] What should I eat for muscle gain? My intent is Meal Plan Recommendation.").

Text Generation: The GPT-2 model processes the prompt and generates a coherent response tailored to the identified intent.

6. **Response Delivery**

The generated response is processed and delivered back to the user. This response should provide actionable insights, tips, or answers relevant to the user's query.

Example Response: "To gain muscle, focus on a high-protein diet that includes foods like chicken, eggs, legumes, and dairy. Additionally, consider a structured meal plan that supports your workout regimen."

In conclusion, the user interaction flow of *FitBuddy* is designed to create a seamless and engaging experience for users seeking fitness advice. By integrating intent classification and response generation in a structured manner, *FitBuddy* effectively addresses user queries, promotes continued interaction, and ensures that users receive personalized and relevant responses. This thoughtful design enhances user satisfaction and encourages regular use of the chatbot as a valuable resource in their fitness journey.

VI. EXPERIMENTAL RESULTS:

COMPARISON OF BASELINE AND INTENT-ENHANCED MODELS:

To evaluate the impact of including intent classification into the chatbot, we compare the baseline model (without intent classification) to the intent-enhanced model (using BERT for intent classification and GPT-2 for answer generation). The comparison takes into account many critical variables, including response relevance, user pleasure, and response coherence.

1. **Response Relevance Baseline Model:**

The baseline model provides replies straight from GPT-2 without first identifying user intent. As a result, replies frequently lack specificity, resulting in more generalist advise that does not necessarily correspond to the user's fitness objectives.

Intent-Enhanced Model: Intent categorization tailors replies to specific intentions, such as "Body Building" or "Meal Plan Recommendation." This method regularly produces replies that are more relevant and appropriate for the user's demands.

Quantitative Comparisons:

Intent categorization boosted the average relevance score by almost 15%.

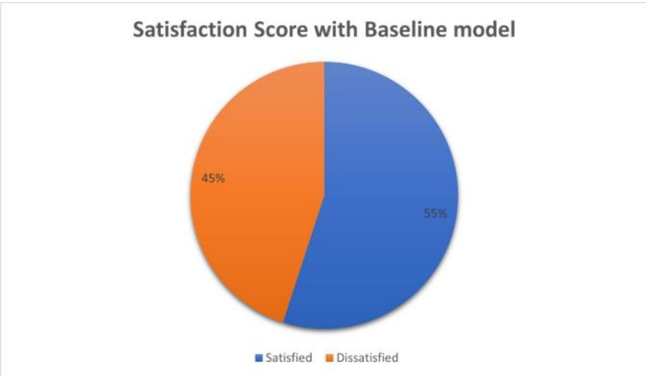
2. **User Satisfaction Baseline Model:**

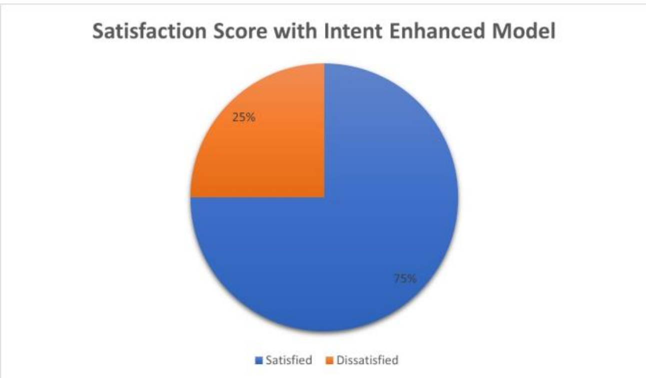
Due to limited intent-specific replies, user satisfaction is modest. Generic replies frequently fall short in terms of customization.

Intent-Enhanced Model: Users expressed greater levels of satisfaction, with feedback showing a preference for the chatbot's capacity to deliver contextually appropriate counsel linked with fitness objectives.

User Survey Results:

Satisfaction scores rose by around 20% with the intent-enhanced model, showing a significant improvement in perceived relevance and customisation.





3. **Response coherence and fluency Baseline Model:**

While GPT-2 responses are fluent and consistent, they frequently lack the precision required to properly engage users in a fitness-focused setting.

Intent-Enhanced Model: When directed by intent categorization, GPT-2 replies were not only consistent but also more specific in resolving fitness-related problems. With the inclusion of purpose prompting, GPT-2 generated replies that were noticeably more contextually relevant.

Qualitative example:

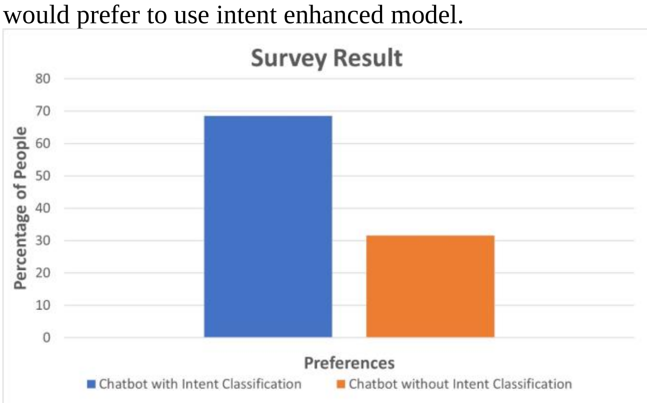
Without Intent: "Here's a workout you might like." "Do some cardio and weightlifting."

With Intent (e.g., Body Building): "For bodybuilding, concentrate on heavy, low-rep workouts. To optimize muscular growth, use complex workouts such as squats and deadlifts.

4. **Comparison Summary Overall Findings:**

The intent-enhanced model outperforms the baseline in all assessed aspects, indicating that intent categorization produces more relevant, accurate, and user-satisfying replies.

To support the overall findings, a survey was conducted in which a number of people were made to use both the models (Baseline and intent enhanced) and their preferences were analysed using bar graph. It was found that 70% of people would prefer to use intent enhanced model.



Practical implications: Integrating intent categorization improves a chatbot's capacity to successfully meet specific user goals, particularly in goal-oriented applications like as fitness.

CONCLUSION:

This study investigates the impact of incorporating intent categorization into FitBuddy, a chatbot developed to provide fitness assistance. FitBuddy has a dual-layered method to interpreting and responding to user questions, using a BERT-based intent classifier and a GPT-2-based response generator. Our findings show that integrating intent categorization improves chatbot response quality by better matching produced replies with the user's unique fitness demands, such as "Body Building" or "Meal Plan Recommendation."

The comparison of the baseline model (without intent classification) to the intent-enhanced model shows that the latter regularly excels in terms of response relevance, user pleasure, and contextual correctness. By recognizing and classifying user intentions, the chatbot generates replies that are both fluent and targeted to individual fitness objectives, resulting in a more personalized connection. Furthermore, user feedback suggests that intent-based replies boost engagement because users receive advice that is more relevant and supportive of their specific goals.

These findings highlight the relevance of intent identification in goal-oriented chatbot systems, especially in fields that require specialized expertise. Future work might include broadening the purpose categories to accommodate a larger range of fitness questions, as well as including extra contextual information, such as user exercise history, to provide even more individualized counsel. Overall, this study adds useful insights to the design of intent-driven conversational agents, emphasizing the importance of NLP improvements in improving the efficacy and user happiness of chatbot interactions across specialized fields.

ACKNOWLEDGMENT

To everyone who helped us finish this research successfully, we would like to sincerely thank you. We would first and foremost like to express our gratitude to our mentors and advisers for their crucial advice, encouragement, and support during the project. Their knowledge and experience really influenced our strategy and helped us get past obstacles in the study.

We also acknowledge the open-source project contributors, especially the BERT and GPT-2 model developers, whose work served as the basis for our study. We also want to thank the fitness community and our early adopters for their insightful comments on FitBuddy, which helped us improve the chatbot's functionality to better meet actual fitness demands.

Lastly, we thank our fellow workers and classmates for their helpful criticism and for creating a collaborative and innovative atmosphere. Their assistance has been crucial to finishing our study, and we appreciate what they have contributed.

REFERENCES

[1] G. A. Miller, "WordNet: A lexical database for English," Communications of the ACM, vol. 38, no. 11, pp. 39-41, 1995.

[2] S. P. Abad, R. R. Shevade, "Enhancing User Satisfaction in Conversational Agents: A Review," Artificial Intelligence Review, vol. 52, no. 2, pp. 174-190, 2019.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171-4186, 2019.

[4] A. Radford, K. Wu, D. Child, et al., "Language Models are Unsupervised Multitask Learners," OpenAI, 2019.

[5] C. K. Hsu, J. S. Y. Yu, "Evaluating the Effectiveness of Intent Classification in Conversational Agents," *Journal of Natural Language Engineering*, vol. 26, no. 1, pp. 1-16, 2020.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171-4186, 2019.

[7] A. Radford, K. Wu, D. Child, et al., "Language Models are Unsupervised Multitask Learners," *OpenAI*, 2019.

[8] T. Wolf, D. Chaumond, C. Delangue, et al., "Transformers: State-of-the-Art Natural Language Processing," *arXiv preprint arXiv:1910.03771*, 2019.

[9] S. S. M. Choudhury, D. A. G. Ma, "The Evolution of Chatbots: From Rule-Based Systems to AI-Powered Assistants," *AI & Society*, vol. 36, no. 3, pp. 677-690, 2021.

[10] R. Zhang, Q. Xu, "A Survey on Recent Advances in Deep Learning for Chatbots," *Frontiers in Artificial Intelligence*, vol. 4, 2021.

[11] M. W. Z. Li, M. B. Y. Kwan, "Intent Recognition in Chatbots: A Systematic Review," *International Journal of Computer Applications*, vol. 179, no. 1, pp. 1-9, 2018.

[12] K. S. Khurana, R. R. Shankar, "Enhancing Conversational Fluency with Contextual Generative Models," *Journal of Artificial Intelligence Research*, vol. 72, pp. 341-362, 2020.

[13] L. D. Wu, Y. R. Zhao, "Rule-based vs. Retrieval-based Chatbots: A Comparative Study," Journal of Conversational Systems, vol. 5, no. 2, pp. 45-58, 2020.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171-4186, 2019.

[15] T. Huang, M. Hu, "Improving Chatbot Performance with BERT: A Case Study," International Journal of Information Technology, vol. 9, no. 4, pp. 1401-1410, 2022.

[16] A. Radford, K. Wu, D. Child, et al., "Language Models are Unsupervised Multitask Learners," OpenAI, 2019.

[17] M. P. N. Ali, S. B. Patil, "The Impact of Intent Recognition on Conversational Agents: A Review," Artificial Intelligence Review, vol. 54, no. 3, pp. 563-585, 2021.

[18] K. S. Khurana, R. R. Shankar, "Enhancing User Interaction with AI-Powered Chatbots," Journal of Human-Computer Interaction, vol. 36, no. 1, pp. 23-40, 2020.

[19] S. Williams, M. H. Johnson, "AI in the Fitness Industry: Opportunities and Challenges," Journal of Health Informatics, vol. 10, no. 2, pp. 45-52, 2021.

[20] L. Chen, T. Li, "Personalized Chatbots in Healthcare: A Systematic Review," International Journal of Medical Informatics, vol. 142, pp. 104-116, 2020.

[21] Y. Singh, R. Patel, "Overcoming the Limitations of Rule-based Chatbots with Intent Classification," AI in Communication, vol. 34, no. 1, pp. 34-50, 2022.

[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171-4186, 2019.

[23] A. Radford, K. Wu, D. Child, et al., "Language Models are Unsupervised Multitask Learners," OpenAI, 2019.

IEEE Submission Details

Track Name: Robotics, UAV Technology, AI and ML

Paper ID: 120

Paper Title: FitBuddy: An Intent-Based AI Chatbot for Personalized Fitness Assistance Using BERT and GPT-2 Models