Ain Shams University
Faculty of Computer & Information Sciences

Scientific Computing Department

TalkTact

By

Elhussein Gomaa Tolba [Scientific Computing]
Elsayed Mostafa Ibrahim [Scientific Computing]
Lunary Mohamed Sabry [Scientific Computing]
Ali Sameh Saad [Scientific Computing]
Mostafa Mohamed Bayoumi [Scientific Computing]
Youssef Ahmed Omar [Scientific Computing]

Under Supervision of
[Prof.Dr. Howida A. shedeed]
[Professor],
Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

[TA. Reham Ahmed]
[Teaching Assistant],
Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

June 2024

Acknowledgements

First and foremost, we would like to extend our deepest gratitude to everyone who played a role in helping us complete our graduation project. We owe a special debt of gratitude to our esteemed supervisor, Dr. Howida Shedeed. Her support and encouragement were essential in guiding us through every challenge we faced. Dr. Shedeed's generous dedication of her time and effort, coupled with her invaluable advice and insights played a huge role in the success of our project. Her mentorship not only helped us navigate the complexities of our work but also inspired us to strive for excellence.

We would also like to express our heartfelt appreciation to our friends and family. Their constant support and encouragement were a source of inspiration throughout this journey. Their faith in our abilities and their ,belief in our potential fueled our determination and motivated us to overcome obstacles and achieve our goals.

In addition, we are immensely grateful to everyone who contributed to our project in any capacity. Whether through direct involvement, offering feedback, or providing resources, your contributions were vital to the realization of our project.

Lastly, we extend our sincere thanks to all those who made this project a success. Your guidance, support, and involvement have been incredibly valuable, and we appreciate the opportunity to have worked on this project with your help. This experience has been profoundly enriching, and we are deeply thankful for the collective effort that brought our project to fruition.

Abstract

In today's globalized world, knowing English well is crucial for communication and career growth. Many people in non-English speaking countries struggle to become fluent. This project creates a mobile app to help with that. It lets users practice English remotely in realistic ways.

The app tackles common problems like fear of mistakes and lack of confidence in speaking with native speakers. It does this with three main features: real-life conversations, live voice chats, and helpful feedback. Using real data, the app simulates real conversations to make learning more natural. The live voice chat helps users practice speaking confidently in real-time. Plus, it gives instant feedback on grammar and mistakes so users can learn without feeling embarrassed.

To achieve this, we'll build an easy-to-use voice chat, add speech recognition and text-to-speech features, and fine-tune a smart model for better language understanding. We'll use a diverse set of real conversations to create lifelike

practice scenarios. The app will work globally, considering different cultures and languages.

Our plan includes using powerful computers to train the model and a clear schedule to stay on track. This app aims to make learning English easier and more effective for everyone, helping users worldwide speak fluently and confidently.

Table of Contents

List of Abbreviations

NLP - Natural Language Processing

AI - Artificial Intelligence

STT - Speech-to-Text

TTS - Text-to-Speech

LLM - Large Language Model

RNN - Recurrent Neural Network

LSTM - Long Short-Term Memory network

PEFT - Parameter Efficient Fine-Tuning

LORA - Low-Rank Adaptation

GPT - Generative Pre-trained Transformer

BERT - Bidirectional Encoder Representations from Transformers

List of Figures

Chapter One:

Introduction

1.1 Problem Definition:

English is the most spoken language in the whole world with around 1.452 million speakers [1]. It is an essential language that we see daily in our lives even if we don't live in a country where English is a first language. English Is needed for communication between people from different nations.

A student may graduate with Top-notch skills in his/her field but face trouble landing a job in a global firm Due to a lack of English proficiency.

Knowledge of English is necessary if one wants to come up in life. It is the major window of the modern world. This is all the more true where the advanced countries have opened their doors for recruiting technically qualified persons. Only those who have a command of the English language are given a job [2].

1.2 Motivation:

Building an application that helps users practice the language could play a huge role in increasing the number of English speakers. Users will be able to use the app remotely without needing to attend any courses. It will be a good practice outlet that the user can return to whenever they feel like they want to practice or enhance their English.

Here are the following motivation points for our application:

I. Real-world Conversations:

Our models are trained on real-life data to give users realistic conversations. This helps create a better language learning experience by mimicking real human interactions. With these lifelike chats, users can practice their language skills in a natural and engaging way, making it easier to learn and understand.

II. Interactive Voice Chats:

Our platform focuses on improving speaking skills through live voice chats. Users can practice real-time conversations, which are crucial for effective communication in different situations. By offering interactive sessions, our platform helps users gain confidence and proficiency in speaking, preparing them to handle various communication challenges confidently and skillfully.

III. Constructive feedback:

English learners often find it hard to talk with native speakers because they worry about making mistakes and being corrected. This fear prevents them from having meaningful conversations, which are important for getting better at the language.

They also worry about feeling embarrassed. They might be scared that they won't be understood or sounding awkward This fear can make them avoid talking to native speakers altogether and only practice with other learners like themselves.

Another thing that makes them nervous is feeling like native speakers are better and know more. This makes them feel shy about starting or keeping up conversations.

Because of these worries, learners don't always feel confident using English in real-life situations. To help them get better, it's important to create friendly places where they can practice without feeling scared of making mistakes. The app will be there for the learner as it will help take down that fear barrier which is a huge obstacle for learners.

1.3 Objectives:

To address the challenges posed by learning English, this research aims to fulfill the following objectives:

I. Developing a Functional Voice Chat Interface:

Create a user-friendly interface that facilitates real-time voice conversations.

II. Incorporating Grammar Correction Module:

Integrate an automated feedback system on grammar and syntax errors during conversations so users can learn from their mistakes and store those errors in a tab where the users can come back to it and check what stopped in their way before.

III. Building and Fine-tuning Transformer Model:

Develop and optimize a Transformer-based model to enhance language understanding and generation capabilities.

IV. Implementing Speech-to-Text & Text-to-Speech Functionality:

Enable users to convert spoken language into text and vice versa, enhancing accessibility and interaction as if we are simulating a real life

conversation where the user speaks (STT comes in) and then the AI model will return the output which is played to the user as if someone is talking back (TTS comes in).

V. Utilizing conversational-dataset Dataset for Model Training:

Utilize a comprehensive dataset of conversational data to train machine learning models, ensuring realistic and effective language practice scenarios. It is essential for us to use a real life dataset so the user feels like he is talking to an actual person and not just a bot. Having this sense of reality will enhance the general user experience and learning journey.

VI. Ensuring Global Accessibility:

Design the application to be accessible worldwide, considering language support, cultural sensitivity, and ease of use across diverse regions and user demographics.

1.4 Methodology:

The methodology used to create this system includes the use of Colab Pro that provides us with extra memory and GPU power to be able to Fine-tune our Large language model(LLM).

1.5 Time Plan:

Tasks	*From*	*To*	*Duration*
Finalize project objectives, team roles, and responsibilities.	14th of October 2023	28th of October 2023	2 weeks
Set up the development environment and communication tools.	28th of October 2023	11th of November 2023	2 weeks
Conduct market research to understand the language learning app landscape & develop and launch user surveys to gather feedback on potential user needs.	11th of November 2023	25th of November 2023	2 weeks
Design and Prototyping: **Weeks 1-2**: Create wireframes and prototypes of the user interface. **Weeks 3-4**: Gather feedback on prototypes and make necessary revisions. **Weeks 4-5**: Finalize the user interface design.	1st of December 2023	5th of January 2024	5 weeks
Backend Development and Database Setup: **Weeks 1-4**: Set up the server infrastructure and database. **Weeks 4-10**: Develop the backend functionality for audio conversations, grammar correction, and progress tracking (The NLP section).	5th of January 2024	15th of March 2024	10 weeks
Frontend Development: **Weeks 1-4**: Start developing the frontend of the mobile application using Flutter. **Weeks 3-8**: Begin implementing real-time grammar correction and voice recognition features. **Weeks 8-10**: Start integrating the chatbot and progress tracking functionality.	15th of March 2024	24th of May 2024	10 weeks
Testing and reviewing feedback.	24th of May 2024	21st of June 2024	4 weeks

Figure 1.1: Time Plan

Developing an application aimed at helping users practice English language skills remotely requires careful planning and execution. The following processes are structured to ensure smooth progression and integration. For example, feedback gathered during the design phase will be promptly incorporated into subsequent stages. Here is a time plan paragraph description that outlines the key steps and estimated timelines for building the application:

I. Finalize project objectives, team roles, and responsibilities (2 Weeks):

- **Week 1:** Define the project's main goals and create a project charter.
- **Week 2:** Assign team roles and responsibilities and set up initial meetings to align everyone on the project vision.

II. Set up the development environment and communication tools (2 Weeks):

- **Week 1:** Install and configure development tools (IDEs, version control, etc.).
- **Week 2:** Set up communication platforms (Discord, email, project management tools) and ensure everyone is connected.

III. Conduct market research and gather user feedback (2 Weeks):

- Research current language learning apps, identify key competitors, and analyze their features.

IV. Design and Prototyping (5 Weeks):

- **Weeks 1-2:** Create wireframes and initial prototypes of the user interface using design tools like Figma or Sketch.
- **Weeks 3-4:** Conduct usability testing sessions with potential users to gather feedback on the prototypes. Make revisions based on the feedback to improve the design.
- **Week 5:** Finalize the user interface design, ensuring it is user-friendly and meets the project's requirements.

V. Backend Development and Database Setup (10 Weeks):

- **Weeks 1-2:** Set up the server infrastructure, including selecting the appropriate server and setting up the operating system.
- **Weeks 3-4:** Design and implement the database schema, ensuring it can handle user data, progress tracking, and other necessary information.
- **Weeks 4-10:** Develop the backend functionality, focusing on:
    - Audio conversations: Implement APIs for recording and processing audio.

- Grammar correction: Integrate NLP models to correct grammar in real-time.
- Progress tracking: Develop systems to track user progress and provide feedback.

VI. Frontend Development (10 Weeks):

- **Weeks 1-2:** Start developing the mobile app's basic structure using Flutter, focusing on navigation and UI components.
- **Weeks 3-4:** Implement user registration, login, and profile management features.
- **Weeks 5-6:** Develop the interface for audio conversations, ensuring smooth recording and playback functionality.
- **Weeks 7-8:** Add real-time grammar correction and voice recognition features, integrating them with the backend.
- **Weeks 9-10:** Integrate the chatbot and progress tracking features, ensuring they work seamlessly within the app.

VII. Testing and reviewing feedback (4 Weeks):

- **Weeks 1-2:** Conduct thorough testing of the app, including functional, usability, and performance testing. Identify and fix any bugs or issues.
- **Weeks 3-4:** Gather feedback from a group of beta testers to identify any remaining issues or areas for improvement. Make necessary adjustments to enhance the user experience and functionality.

1.6 Thesis Outline:

- **Chapter 1: Introduction**

This chapter provides an overview of the research project or system development, setting the context for the entire document. It outlines the problem statement, objectives, and research questions, highlighting the significance and relevance of the study. The chapter introduces the scope and limitations of the project and presents an outline of the subsequent chapters.

- **Chapter 2: Literature Review**

In this chapter, a comprehensive review of existing literature and related works is presented. It involves an in-depth analysis of scholarly articles, research papers, and relevant sources to identify gaps, trends, and insights related to the topic. The literature review provides a theoretical framework and supports the research methodology adopted in the study.

- **Chapter 3: System Architecture**

This chapter delves into the technical aspects of the system architecture. It outlines the overall design, components, and infrastructure of the developed application. The chapter discusses the selection of technologies, frameworks, and tools used in the development process.

- **Chapter 4: System Implementation & Results**

This chapter focuses on the practical implementation of the system. It details the development process, including coding practices, algorithms, and data structures utilized. The chapter presents the results of the system implementation, such as performance metrics, accuracy rates, and validation tests.

- **Chapter 5: Run the Application**

This chapter provides instructions on how to run and utilize the developed application. It guides users through the installation process, system setup, and configuration steps. The chapter explains the application's user interface, features, and functionalities.

- **Chapter 6: Conclusion & Future Work**

The final chapter summarizes the key findings, contributions, and implications of the research or system development. It reflects on the objectives stated in the introduction and addresses the research questions. The chapter discusses the significance of the results and their potential impact on the field. Additionally, it suggests areas for future research,

enhancements, or improvements to extend the system's capabilities and address any limitations identified during the project.

These chapter descriptions provide a general understanding of the content covered in each chapter.

Chapter Two:

Literature Review

2.1 Introduction:

The rapid advancements in Natural Language Processing (NLP) have paved the way for innovative applications that enhance language learning experiences. One such application is the "Talk Tact" project, an NLP-based chatbot designed to facilitate English language practice. This project leverages state-of-the-art technologies to provide an interactive platform where users can practice speaking English, receive real-time grammar and spelling corrections, and understand the reasons behind their mistakes. The integration of speech-to-text and text-to-speech functionalities further enriches the user experience by enabling seamless verbal communication.

The primary objective of this literature review is to examine the underlying technologies and methodologies employed in the Talk Tact project. By exploring the current state of NLP, large language models (LLMs), transformer neural

networks, and other related technologies, this review aims to provide a comprehensive understanding of the project's technical foundation. Additionally, this review will compare related works to highlight the unique contributions and advancements made by Talk Tact.

In the following sections, we will go through the technological background, evaluate related works, and conclude with insights and future directions for research. This literature review serves as a foundational reference for understanding the complex interplay of technologies that drive the Talk Tact project and its potential impact on language learning.

2.2 Technology Background:

Natural Language Processing (NLP) has become a cornerstone in the development of advanced language models and conversational agents. The "Talk Tact" project incorporates various cutting-edge technologies to create a sophisticated chatbot capable of facilitating English language practice. This section delves into the key technologies that underpin the project, providing a detailed understanding of their functionalities and applications.

NLP and Transformer Models

NLP is a field of artificial intelligence that focuses on the interaction between computers and human languages. It encompasses a range of tasks including speech recognition, text generation, translation, and sentiment analysis.

One of the pivotal advancements in NLP is the development of transformer neural networks, introduced by Vaswani et al. in 2017. These models utilize self-attention mechanisms to process and generate language, significantly outperforming previous architectures like recurrent neural networks (RNNs) and long short-term memory networks (LSTMs).

Speech-to-Text and Text-to-Speech

Speech-to-Text (STT) and Text-to-Speech (TTS) are integral components of the Talk Tact project. STT technology converts spoken language into text, enabling the chatbot to understand and process user inputs. TTS, on the other hand, transforms text responses generated by the chatbot back into spoken language, providing an interactive and natural user experience. These technologies rely on deep learning models trained on large datasets of speech and text pairs.

Grammar and Spelling Correction

The project also incorporates a grammar and spelling correction module, which identifies and rectifies errors in user inputs. This module not only corrects mistakes but also provides explanations, helping users understand their errors and improve their language skills. Techniques such as rule-based systems, statistical models, and neural networks are employed to achieve high accuracy in error detection and correction.

Large Language Models (LLMs)

LLMs, such as GPT-3 and text-davinci-003, are at the heart of the Talk Tact project. These models are pre-trained on vast amounts of text data, enabling them

to generate coherent and contextually relevant language. GPT-3, developed by OpenAI, is one of the largest and most powerful LLMs, containing 175 billion parameters. The project utilizes a fine-tuned version of this model, text-davinci-003, which has been optimized for specific conversational tasks.

Mistral AI 7B Parameters and Fine-Tuning

To further enhance the chatbot's performance, the project employs Mistral AI's 7B parameter model. This model is fine-tuned using a process known as transfer learning, where a pre-trained model is adapted to a new task by retraining it on a smaller, task-specific dataset. This approach leverages the general language understanding capabilities of the pre-trained model while tailoring it to the specific needs of the project.

LORA and Transfer Learning

Parameter Efficient Fine-Tuning (PEFT) techniques, such as LORA (Low-Rank Adaptation), are used to overcome the challenges associated with fine-tuning large models. LORA involves freezing most of the pre-trained parameters and only updating a small subset of them. This reduces the computational resources required and improves the efficiency of the fine-tuning process.

In summary, the Talk Tact project is built on a robust foundation of advanced NLP technologies, transformer models, and innovative fine-tuning techniques. These technologies work in concert to create a powerful and versatile language learning tool that offers real-time feedback and interaction.

2.3 Related Works:

The development of conversational agents and language learning tools has been an area of active research, with numerous projects and studies contributing to the field. This section reviews related works, comparing them with the Talk Tact project and identifying insights and gaps.

Project/Approach	Dataset Used	Model/Technique	Challenges/Limitations	Key Insights/Gains
Initial Attempt with Custom Transformer	Shakespeare Plays	Custom Transformer Model	Limited dataset size	Importance of large, diverse datasets for effective language models.
OpenWebText Dataset and GPT-2 Training	OpenWebText (55.21 GB)	GPT-2	Complexity and scope beyond project capabilities	Highlighted the challenges of adapting general-purpose models.
PolyAi Conversational Datasets	Reddit, OpenSubtitles, Amazon QA	Various NLP Models	Size and computational resource requirements	Emphasized balancing data richness with feasibility.
Large Language Models (LLMs) - Mistral 7B	Not specified	Mistral 7B	Trade-offs between model size and performance	Need for further fine-tuning to achieve human-like conversations.
Fine-Tuning Challenges and Solutions	Not specified	LORA, PEFT Techniques	Vast number of parameters	Enhanced model performance by focusing on key parameters.
Real-Life Scenario Dialogue Summarization	DialogSum	Fine-Tuned LLM	Initial fine-tuning challenges	Improved responses and natural conversational flow.

Initial Attempt with Custom Transformer (Sheksbeer Plays)

Our initial attempt involved developing a custom transformer model using the Shakespeare Plays dataset. Despite the innovative approach, the limited size of the dataset posed significant challenges, resulting in suboptimal performance in real-life conversation scenarios. This experience underscored the importance of large, diverse datasets for training effective language models [3].

Exploration of OpenWebText Dataset and GPT-2 Training

Recognizing the need for more expansive data, we explored the OpenWebText dataset, which contains 55.21 GB of data. Although the dataset held immense potential, it was beyond the scope of our project. Instead, we leveraged GPT-2, a model pre-trained on OpenWebText. Despite its capabilities, we encountered difficulties in achieving the desired language skill improvement, highlighting the complexity of adapting general-purpose models to specific tasks [5].

Discovery of Conversational Datasets by PolyAi

We then explored conversational datasets curated by PolyAi, including sources like Reddit, OpenSubtitles, and Amazon QA. These datasets offered rich information but also presented practical challenges due to their size and the extensive computational resources required for processing. This led us to reconsider our approach and seek alternative solutions that balance data richness with feasibility.

Exploration of Large Language Models (LLMs)

In pursuit of more effective models, we investigated Large Language Models (LLMs) and selected Mistral 7B. Although promising, the model did not fully meet our expectations for human-like conversational responses and academic information. This exploration highlighted the trade-offs between model size and performance, and the need for further fine-tuning to achieve optimal results.

Fine-Tuning Challenges and Solutions

The process of fine-tuning LLMs presented significant challenges due to the vast number of parameters. To address this, we employed Parameter Efficient Fine-Tuning (PEFT) techniques such as LORA. By freezing pre-trained parameters and retraining only key parameters, we streamlined the fine-tuning process and enhanced model performance.

Fine-Tuning with Real-Life Scenario Dialogue Summarization Dataset (DialogSum)

Our continuous pursuit of refining the model's language skills led us to the DialogSum dataset. This real-life scenario dialogue summarization dataset allowed us to fine-tune the model to our specific task, resulting in improved responses and a more natural conversational flow. This emphasizes the importance of selecting appropriate datasets for achieving project goals.

Comparison and Insights

Comparing the Talk Tact project with related works reveals several key insights. First, the use of diverse and large datasets is crucial for training effective conversational agents. Second, the balance between model size and performance is a persistent challenge that requires innovative fine-tuning techniques. Finally, the integration of real-time feedback mechanisms, such as grammar and spelling correction, distinguishes Talk Tact from other language learning tools by providing immediate and actionable insights to users.

2.4 Conclusion:

The Talk Tact project represents a significant advancement in the field of NLP and language learning. By leveraging state-of-the-art technologies such as transformer neural networks, LLMs, and innovative fine-tuning techniques, the project addresses key challenges in developing effective conversational agents. The integration of real-time feedback mechanisms enhances the learning experience, providing users with immediate and actionable insights.

The review of related works highlights the unique contributions of the Talk Tact project and underscores the importance of large, diverse datasets and efficient fine-tuning techniques. Moving forward, further research and development are needed to refine these technologies and explore new approaches to enhancing language learning tools. The insights gained from this project provide a solid foundation for future advancements in the field.

Chapter Three:

System Architecture

This chapter provides an in-depth overview of the system architecture and technical methods used in TalkTact, an innovative language learning platform. It outlines the flow of user interactions, speech-to-text conversion, text processing, response generation, and feedback mechanisms, integrating advanced technologies for an enhanced learning experience.

3.1 System Architecture:

The system architecture of TalkTact involves several key components and workflows to ensure efficient and accurate language learning experiences. The primary components include:

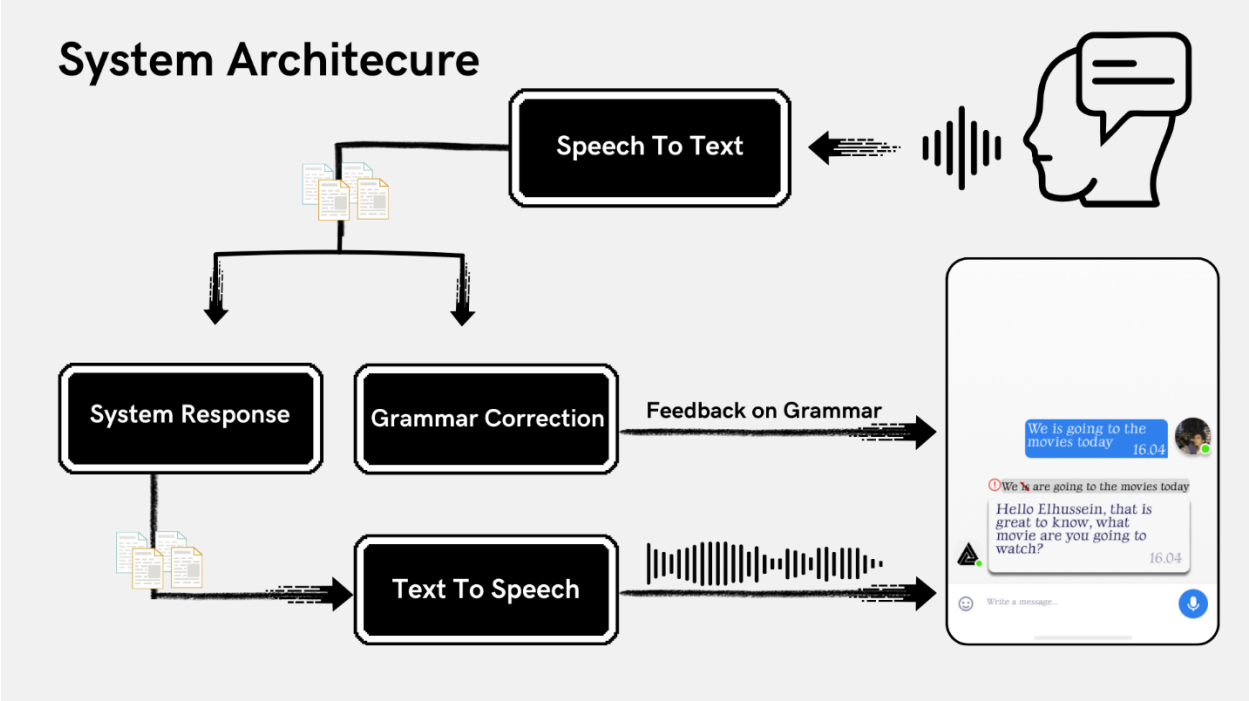

Figure 3.1: System Architecture

A. User Interaction

Users interact with the system through a voice chat interface or by typing their inputs. The interface captures audio or text input and initiates the processing workflow.

B. Speech-to-Text Conversion

The captured audio input is processed through a speech recognition engine, which converts the audio signals into text. This step involves:

1. Audio signal preprocessing.
2. Application of speech recognition algorithms.
3. Transcription of spoken words into text.

C. Text Processing and Grammar Correction

The transcribed text or user-typed input undergoes several processing steps:

1. **Text Normalization:** Standardizes the text format.

2. **Phoneme Conversion:** Converts text into phonemes for detailed linguistic analysis.
3. **Syllabification:** Segments the text into syllables for better linguistic processing.
4. **Syllable Matching:** Matches the segmented syllables against a pre-established database to ensure accuracy.
5. **Grammar Correction:** Utilizes a large language model to provide contextual grammar corrections and feedback.

D. Response Generation

Based on the processed and corrected text, the system generates appropriate responses. This involves:

1. Contextual understanding of the input text.
2. Generation of accurate and contextually relevant responses using a fine-tuned large language model.
3. Conversion of the generated response into text.

E. Text-to-Speech Conversion

The generated text response is then converted back into speech using text-to-speech technology. This step involves:

1. Conversion of text into lifelike speech.
2. Utilization of Flutter TTS for seamless and natural-sounding speech output.

F. Feedback Mechanisms

The system provides real-time feedback to users, displaying grammar corrections as hints and delivering audio responses. This enhances the interactive learning experience and helps users improve their language skills effectively.

In summary, the TalkTact platform leverages advanced speech recognition, text processing, and text-to-speech technologies to create an interactive and efficient language learning environment. The integration of these technologies within a robust system architecture ensures that users receive accurate, contextual, and real-time feedback, aiding in their language learning journey.

I. Technical Methods

A. Developing a Functional Voice Chat Interface

1) Speech Recognition

Speech recognition is a critical component of TalkTact, enabling the system to capture and interpret spoken language input from users. The process involves several detailed steps:

- **Start:** The system begins by capturing audio input from the user through the device's microphone.
- **Preprocess:** The captured audio is converted to a suitable format for processing, such as adjusting the sample rate and normalizing the audio signal.
- **Decoding:** Advanced speech recognition algorithms, often powered by neural networks, are employed to transcribe spoken words into text. This involves the use of acoustic models, language models, and decoding algorithms to accurately recognize speech.

- **Check Likelihood of Words:** The transcribed words are subjected to probability checks to ensure accuracy. This step involves comparing the likelihood of different word sequences and selecting the most probable transcription.
- **Output Text:** The final transcribed text is displayed to the user or utilized in subsequent processes for further analysis and response generation.

2) Speech-to-Text

The speech-to-text functionality in TalkTact allows for the instant transcription of spoken words into written text. This feature is powered by state-of-the-art speech recognition technology and offers several key advantages:

- **Instant Transcription:** Spoken words are transcribed into written text in real-time, providing immediate feedback to the user.
- **Cutting-Edge Technology:** Leveraging advanced algorithms and models, the system ensures high accuracy in transcription.

- **Hands-Free Dictation:** Users can dictate messages, notes, and documents without needing to type, enhancing convenience and accessibility.
- **Cross-Platform Compatibility:** Integrated with Flutter, the speech-to-text feature works seamlessly across multiple platforms, including iOS and Android.
- **User-Friendly Interface:** The interface is designed for effortless transcription, making it easy for users to interact with the system.

3) Text-to-Speech

The text-to-speech (TTS) functionality in TalkTact converts written text into lifelike speech, enhancing the user experience by providing audible responses. Key aspects of the TTS feature include:

- **Lifelike Speech Conversion:** The system generates natural-sounding speech from text input, making interactions more engaging and realistic.

- **Flutter TTS Integration:** Utilizing the Flutter TTS plugin, the system ensures smooth and efficient text-to-speech conversion.
- **Versatile Applications:** The generated speech can be used in various applications, such as reading out text messages, providing verbal feedback, and more.
- **Cross-Platform Consistency:** The TTS feature is compatible with Flutter, ensuring consistent functionality across different platforms.
- **Enhanced User Experience:** Interactive audio output enhances user engagement and provides a more dynamic learning environment.

4) Text Processing

Text processing is a crucial step in preparing and analyzing text input for grammar correction and response generation. The process involves several stages:

- **Start:** The system receives text input from either the speech recognition module or directly from user writing.

- **Text Normalization:** The input text is standardized, ensuring consistency in formatting and structure.
- **Phoneme Conversion:** The normalized text is converted into phonemes for linguistic processing, facilitating more accurate analysis.
- **Syllabification:** The text is segmented into syllables, which are essential for phonological analysis and pronunciation correction.
- **Find Syllable:** The segmented syllables are matched with a pre-established database to ensure accuracy in phonetic representation.
- **Compare with Syllables Database:** The system compares the input syllables with the database entries to validate the matches.
- **Match/Not Match:** Based on the comparison, the system finalizes the text processing, determining whether the input matches the expected phonetic patterns.
- **End:** The processed text is outputted for further use in grammar correction and response generation modules.

B. Fine-Tuning Large Language Models (LLMs

1) Overview of Large Language Models (LLMs)

Large Language Models (LLMs) are advanced neural network architectures that are designed to understand and generate human-like text. These models have revolutionized natural language processing (NLP) by enabling a wide range of applications through their ability to comprehend context, generate coherent sentences, and perform various language-related tasks.

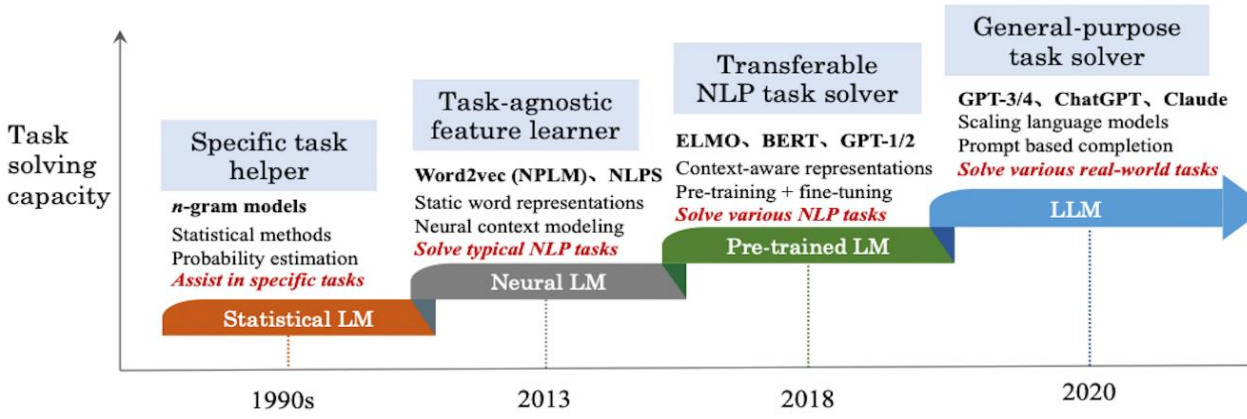

Figure 3.2: Evolution of Language Models

Definition: LLMs are deep learning models that consist of millions or even billions of parameters. They are trained on vast

corpora of text data to predict the next word in a sentence, understand context, and generate text that mimics human writing.

Applications: LLMs are versatile and can be applied to numerous NLP tasks, including:

- **Language Translation:** Translating text from one language to another with high accuracy.
- **Text Summarization:** Condensing long documents into concise summaries.
- **Sentiment Analysis:** Determining the sentiment expressed in a piece of text.
- **Question Answering:** Providing answers to questions based on contextual understanding.
- **Conversational Agents:** Powering chatbots and virtual assistants with the ability to hold natural and meaningful conversations.

Popular LLMs: Several well-known LLMs have set benchmarks in the field of NLP, including:

- **GPT (Generative Pre-trained Transformer):** Developed by OpenAI, GPT models are known for their ability to generate coherent and contextually relevant text.
- **BERT (Bidirectional Encoder Representations from Transformers):** Created by Google, BERT excels in understanding the context of words in a sentence, making it highly effective for tasks like question answering and language inference.
- **XLNet:** An extension of BERT that uses permutation-based training to capture bidirectional context more effectively.

  Mistral: A newer model known for its efficiency and effectiveness in handling complex language tasks

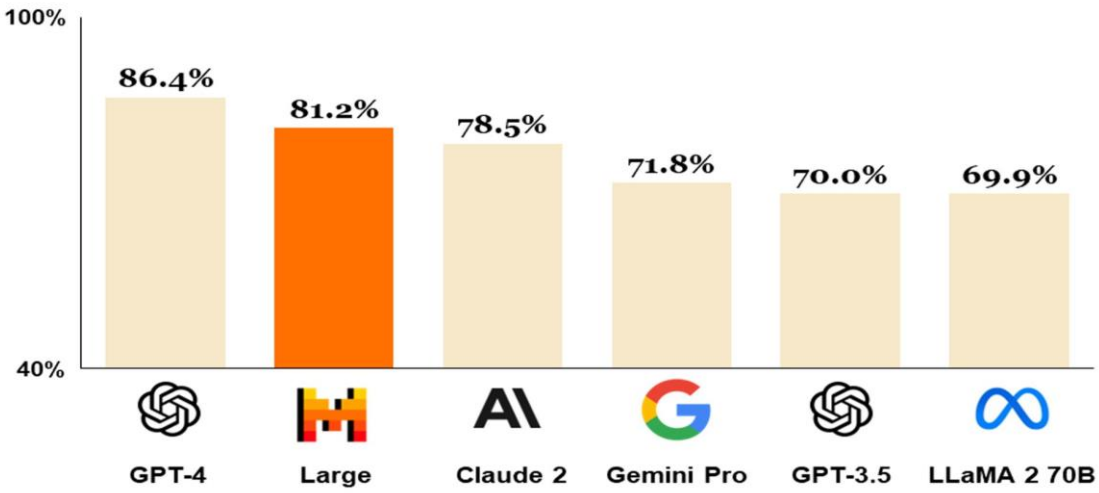

Figure 3.3: LLMs Rankings.

2) Transformer Architecture

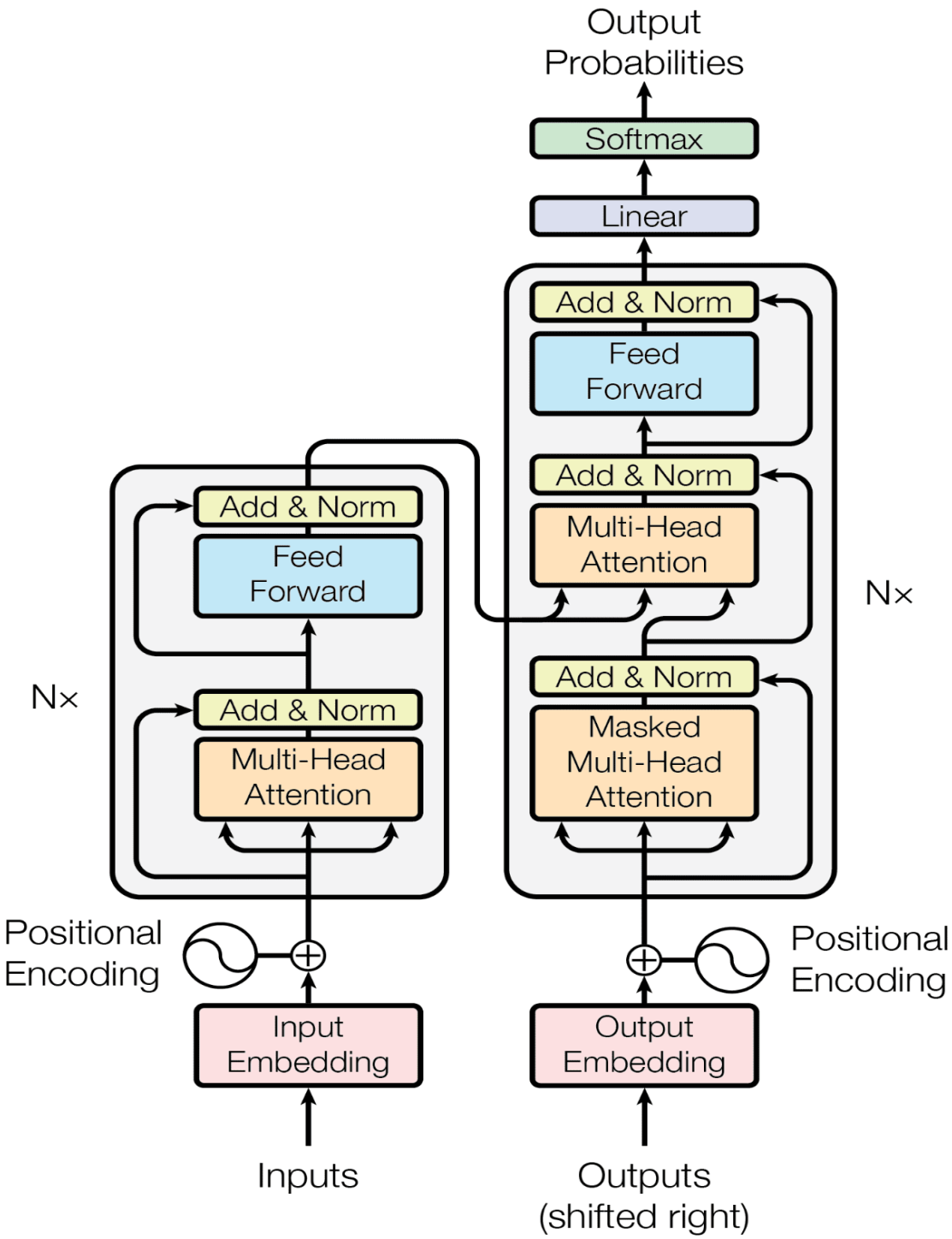

Figure 3.4: Transformer Architecture

The Transformer architecture, introduced by Vaswani et al. in 2017, has become the foundation for many modern LLMs due to its ability to handle long-range dependencies and parallelize training.

Advantages: Transformers address the limitations of previous architectures like LSTM (Long Short-Term Memory) and RNN (Recurrent Neural Network), which struggled with long-range dependencies and sequential processing.

Parallelization: Transformers use self-attention mechanisms that allow for parallel processing of input data, significantly improving training efficiency and scalability.

Long-Range Dependencies: The self-attention mechanism enables the model to focus on different parts of the input sequence, capturing long-range dependencies and contextual relationships more effectively.

Ease of Training: Transformers benefit from the absence of sequential processing, which reduces training time and computational resources. This makes them easier to train on large datasets compared to traditional RNNs.

3) Chosen Model for Fine-Tuning: Mistral 7B

Introduction: Mistral 7B is a powerful language model designed for straightforward fine-tuning, making it adaptable to specific tasks with minimal effort. It is optimized for both efficiency and performance, handling complex language tasks with ease.

Capabilities: Mistral 7B excels in understanding and generating human-like text, making it suitable for a wide range of applications, including grammar correction, response generation, and more. Its architecture allows it to capture nuanced language

patterns and provide accurate and contextually relevant outputs.

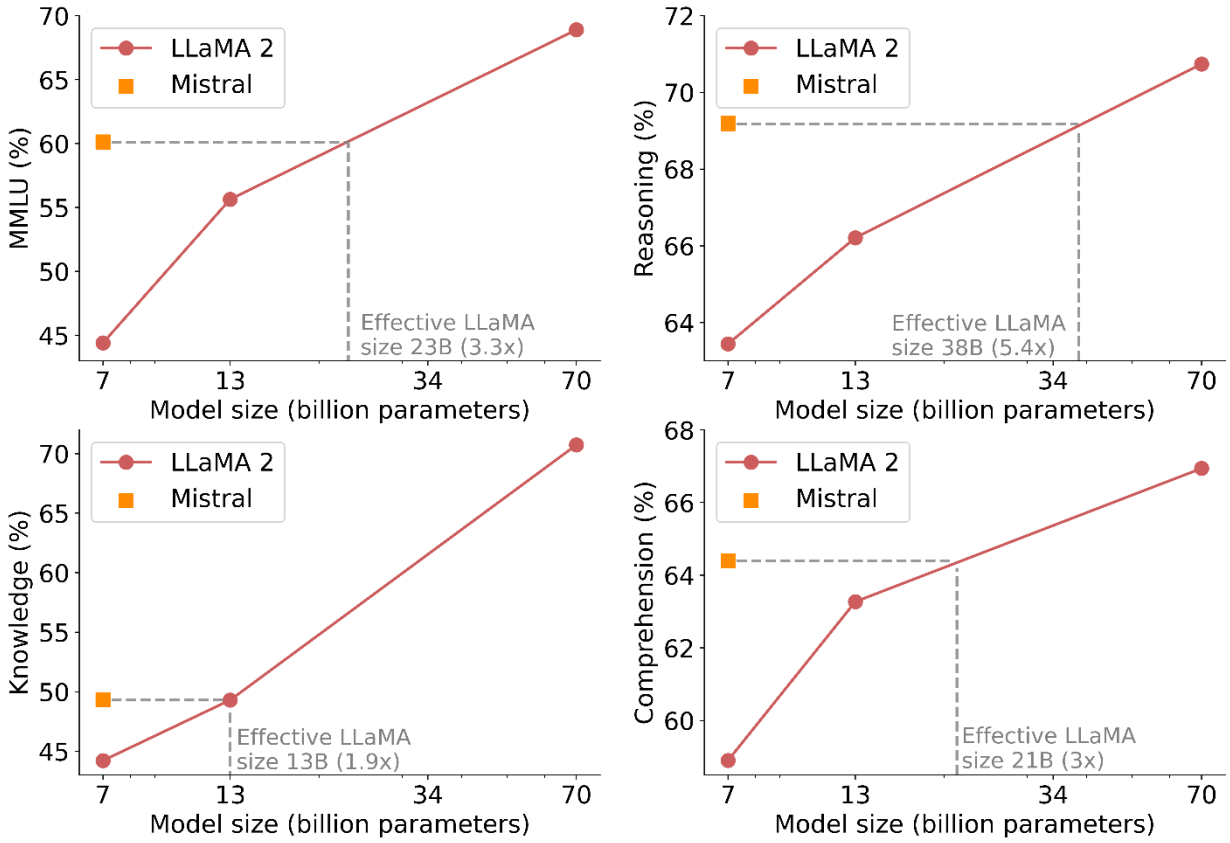

Figure 3.5: Results Comparisons

4) Fine-Tuning Process

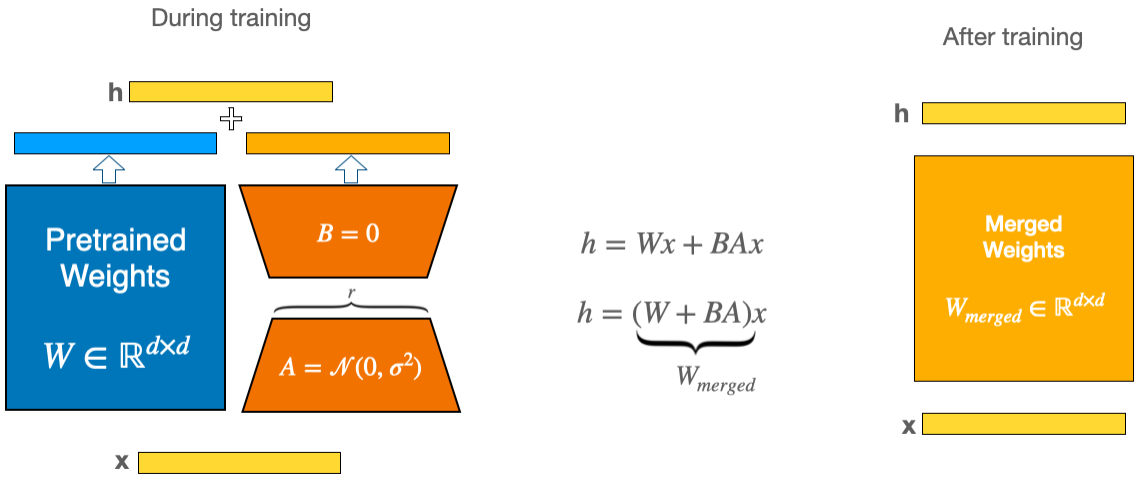

Figure 3.6: Fine-Tuning Process

Definition: Fine-tuning is the process of adapting a pre-trained LLM like Mistral 7B to specific tasks by training it on task-specific data. This process enhances the model's ability to perform well on the target task.

Project Approach: The fine-tuning process for TalkTact involves employing transfer learning, where the pre-trained Mistral 7B model is further trained on domain-specific datasets. This approach leverages the model's pre-existing knowledge and

adapts it to the unique requirements of the language learning platform.

5) Dataset and Preprocessing

Dataset: The primary dataset used for fine-tuning is the Dialogsum dataset, which consists of 13,460 dialogues. This dataset provides a rich source of conversational data that is essential for training the model to handle dialogue-based tasks effectively.

Challenges: Full fine-tuning of large models like Mistral 7B on regular hardware can be impractical due to high computational and memory requirements.

Solutions:

- **Utilized Colab Pro:** To overcome hardware limitations, Google Colab Pro is used, providing enhanced memory and GPU

capabilities to manage the computational demands of fine-tuning large models.

- **Applied Low-Rank Adaptation (LoRA):** LoRA is a technique that fine-tunes key parameters of the model efficiently, reducing the computational burden and enabling effective fine-tuning on available hardware.

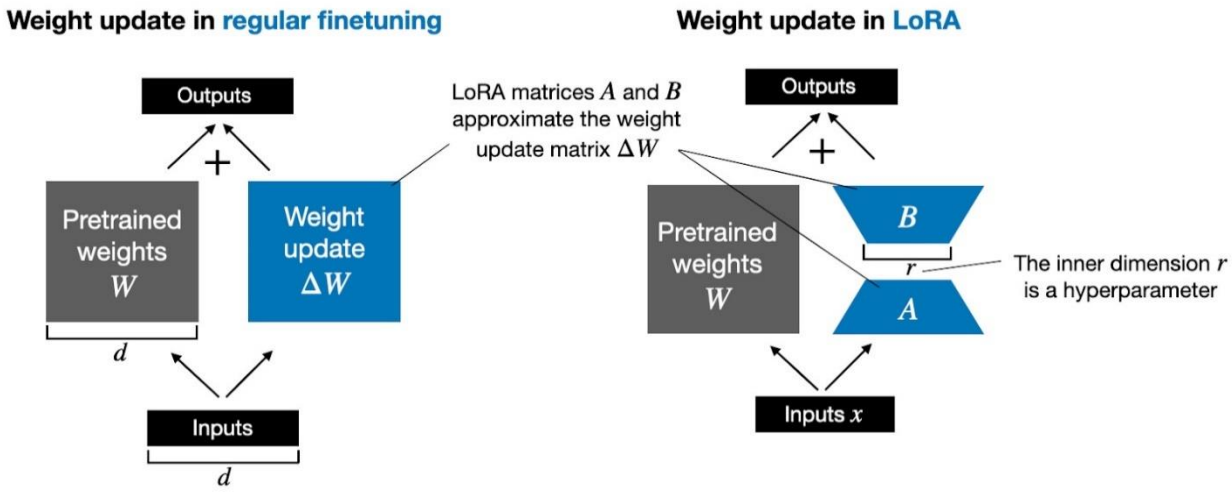

Figure 3.6: LoRa Architecture

6) Preprocessing Steps

Split Dialogue: The dialogues are divided by speaker using markers to separate individual turns in the conversation. This helps in organizing the data for better analysis and processing.

Iterate Over Lines: Each line of dialogue is iterated over to extract the speaker and spoken text. This involves parsing the dialogue to identify who is speaking and what is being said.

Format Lines: Formatting tags and unique IDs are added to each line to ensure the data is structured and easily accessible during processing. This step standardizes the data format for consistency.

Add to Preprocessed Dialogue: The formatted lines are compiled into a preprocessed dialogue string, creating a structured dataset that is ready for model training.

7) Grammar and Spelling Correction Module

Definition: The grammar and spelling correction module identifies and corrects grammatical errors in text. It is crucial for enhancing the quality of writing and communication by ensuring grammatical correctness.

Importance: Improving grammar and spelling not only enhances writing quality but also aids in effective communication. This is particularly important in a language learning platform where accurate feedback is essential for learner development.

Challenge: Lack of Suitable Dataset for Fine-Tuning: One of the significant challenges faced during the development of TalkTact was the lack of suitable datasets specifically tailored for fine-tuning models to correct grammatical errors in English. Existing datasets were either too generic or lacked the detailed explanations necessary for effective learning and feedback.

Dataset Creation: To address this challenge, we undertook the creation of a custom dataset specifically designed to meet the needs of our language learning platform. The goal was to develop a comprehensive dataset that not only included sentences with grammatical mistakes but also provided clear explanations for each error, facilitating more effective learning and correction.

Solution: Custom Dataset with Examples of Grammatical Mistakes Our custom dataset comprises a wide range of sentences that include various types of grammatical errors. Each sentence is carefully crafted to represent common mistakes made by English learners, ensuring that the dataset is relevant and practical for real-world language learning scenarios.

Focus: Dataset Includes Explanations for Each Grammatical Error To enhance the learning experience, each sentence in the dataset is accompanied by a detailed explanation of the grammatical error it contains. These explanations provide insights into why the sentence is incorrect and how it can be corrected, helping users understand the underlying grammatical rules and improve their language skills more effectively.

Approach: Mistral 7B was fine-tuned using transfer learning on the custom dataset. This process involved training the model to recognize common grammatical errors and provide corrections along with explanations.

8) Fine-Tuning Process

Methodology: The fine-tuning process focused on detecting, correcting, and explaining errors. The model was trained on the custom dataset to enhance its ability to identify grammatical mistakes, correct them, and provide explanations to the user. This approach ensures that the model not only corrects errors but also educates the user, contributing to their learning process.

Conclusion

The system architecture and technical methods underpinning TalkTact showcase the integration of advanced technologies to facilitate effective and interactive English language learning. By leveraging state-of-the-art speech recognition, text-to-speech conversion, and large language models, TalkTact provides users with a seamless experience for both spoken and written interactions.

The architecture emphasizes a robust workflow, beginning with user interaction and progressing through speech-to-text conversion, comprehensive text processing, grammar correction, response generation, and ultimately, delivering feedback through text and speech. The fine-tuning of the Mistral 7B model, coupled with custom datasets, ensures that the platform offers precise and contextual feedback, essential for language learners.

The technical methods employed, such as the development of a functional voice chat interface and the implementation of advanced language models, underline the innovative approach of TalkTact. These methods collectively enhance the accuracy, efficiency, and user-friendliness of the platform, making it a powerful tool for mastering English.

In summary, the integration of cutting-edge technologies within a meticulously designed system architecture enables TalkTact to provide an engaging, accurate, and effective language learning environment. This innovative approach ensures that users receive real-time, contextual feedback, significantly aiding their journey to fluency and confidence in English.

Chapter Four:

System Testing & Results

This chapter will present a detailed analysis of our model's performance by simulating normal user interactions with the application. We will conduct a series of thorough tests to evaluate how well the model works in real-world scenarios. These evaluations will provide useful insights into the model's effectiveness when used in the application, giving a realistic assessment of its strengths and areas that may need improvement.

4.1 Prompt with grammar and spelling errors:

1. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "Their going to there friends house later today."

- **Grammar or spelling errors:**

    - "They're going to their friends' house later today."
- **Correction:**
    - Change "Their" to "They're" (contraction of "They are") and add an apostrophe to "friends" to denote possession.
- **Response:**
    - I'm glad to hear that they have plans to visit their friends' place later today.

2. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "I never seen such a beautifull sunset before."

- **Grammar or spelling errors:**
    - "I've never seen such a beautiful sunset before."
- **Correction:**
    - Change "I never seen" to "I've never seen" and "beautifull" to "beautiful."

- **Response:**
    - I can understand why you're so captivated by the stunning sunset you've witnessed for the first time.

3. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "She dont likes the color blue."

- **Grammar or spelling errors:**
    - "She doesn't like the color blue."
- **Correction:**
    - Change "dont" to "doesn't" and "likes" to "like."
- **Response:**
    - I'm here for you, and I'll support you even when your favorite color isn't mine, like blue in this case.

4. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "He was running fastly to catch the bus."

- **Grammar or spelling errors:**
    - "He was running quickly to catch the bus."

- **Correction:**
    - Change "fastly" to "quickly."
- **Response:**
    - It's amazing how determined he is to not miss the bus by running quickly.

5. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "Their going to the movies tonight."

- **Grammar or spelling errors:**
    - "They're going to the movies tonight."
- **Correction:**
    - Change "Their" to "They're."
- **Response:**
    - I'm thrilled for you and your group of friends to enjoy a night out at the movies together tonight.

6. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "Its raining outside but I forgot my umbrella."

- **Grammar or spelling errors:**
    - "It's raining outside, but I forgot my umbrella."
- **Correction:**
    - Change "Its" to "It's" and add a comma after "outside."
- **Response:**
    - Despite the rain, I'm here to offer you a raincoat or a ride, my friend.

7. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "Can I has some more cake?"

- **Grammar or spelling errors:**
    - "Can I have some more cake?"
- **Correction:**
    - Change "has" to "have."
- **Response:**
    - Of course, my friend! I'd be happy to help you satisfy your sweet tooth with an additional slice of cake.

8. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "He dont know where he's going."

    - **Grammar or spelling errors:**
        - "He doesn't know where he's going."
    - **Correction:**
        - Change "dont" to "doesn't."
    - **Response:**
        - It's okay, my friend, let's figure it out together and help him reach his destination.

9. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "There are too many peoples in the room."

    - **Grammar or spelling errors:**
        - "There are too many people in the room."
    - **Correction:**
        - Change "peoples" to "people."

- **Response:**
    - It seems the room is teeming with the presence of many individuals today.

10. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "The cats is sleeping on the couch."

- **Grammar or spelling errors:**
    - "The cat is sleeping on the couch."
- **Correction:**
    - Change "cats" to "cat."
- **Response:**
    - I see the cat enjoying a peaceful slumber on the couch.

4.2 Prompt without grammar and spelling errors:

11. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "She always brings her lunch to work."

- **Grammar or spelling errors:**
  - N/A
- **Response:**
  - Wow, she's committed to her meals. I respect that!

12. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "The children are playing in the park."

- **Grammar or spelling errors:**
  - N/A
- **Response:**
  - That's great! Enjoying the fresh air and fun. I'm glad they're able to be active.

13. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "He enjoys reading science fiction novels."

- **Grammar or spelling errors:**
    - N/A
- **Response:**
    - Cool! Expanding his imagination with the unknown and futuristic worlds.

14. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "She speaks three languages fluently."

- **Grammar or spelling errors:**
    - N/A
- **Response:**
    - Impressive! Being multilingual opens up a whole new world of communication and connection.

15. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "They bought a new house in the suburbs."

- **Grammar or spelling errors:**
    - N/A
- **Response:**
    - Exciting times ahead! A new home is a fresh start and a place to build memories.

16. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "He wakes up early every morning."

- **Grammar or spelling errors:**
    - N/A
- **Response:**
    - Good for him! An early start to the day means more time for goals and pursuits.

17. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "She loves to paint landscapes."

- **Grammar or spelling errors:**
    - N/A

- **Response:**
    - Beautiful! Capturing nature's essence through art is truly inspiring.

18. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "They are preparing for their final exams."

- **Grammar or spelling errors:**
    - N/A
- **Response:**
    - Best of luck to them! Focused study and preparation is key to success.

19. **Prompt:** As a best friend, please respond to the following sentence in one sentence only: "He enjoys playing the guitar in his free time."

- **Grammar or spelling errors:**
    - N/A
- **Response:**

- That's awesome! Creatively expressing oneself through music is a wonderful hobby.

Screenshots:

Here are some Screenshots from the app to see how the user will be shown the results:

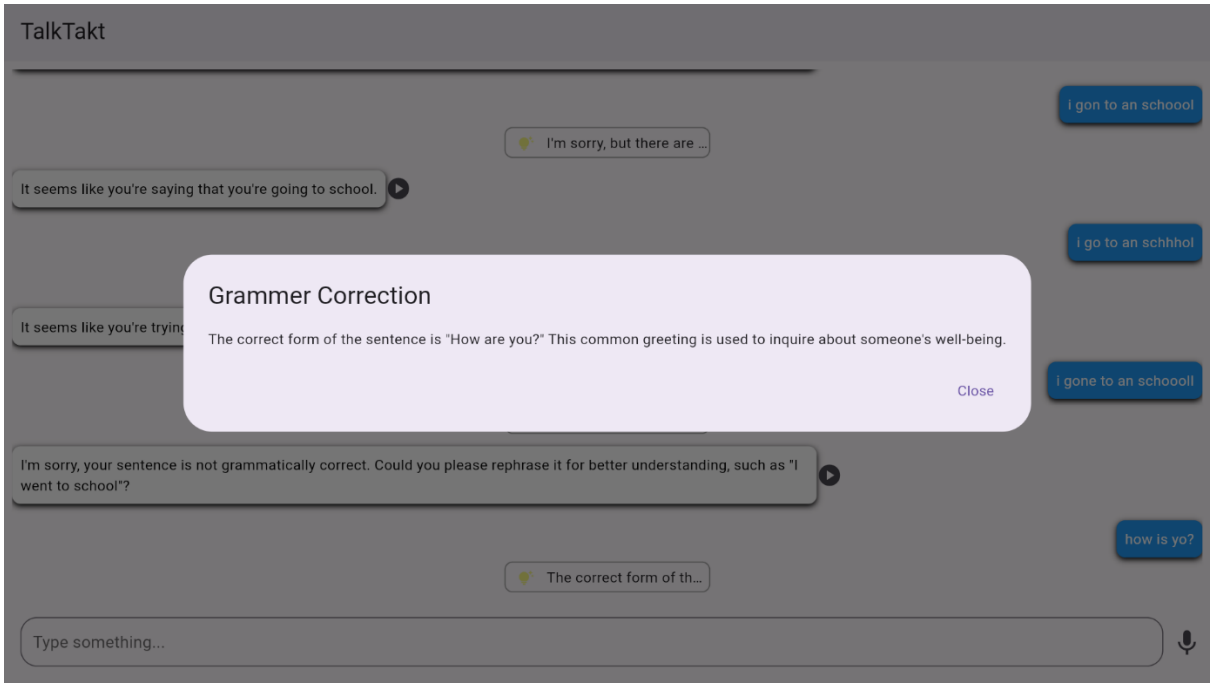

Figure 4.1: WebApp Grammer Correction Screen

- This is how the user is shown the correction if he makes a grammatical mistake

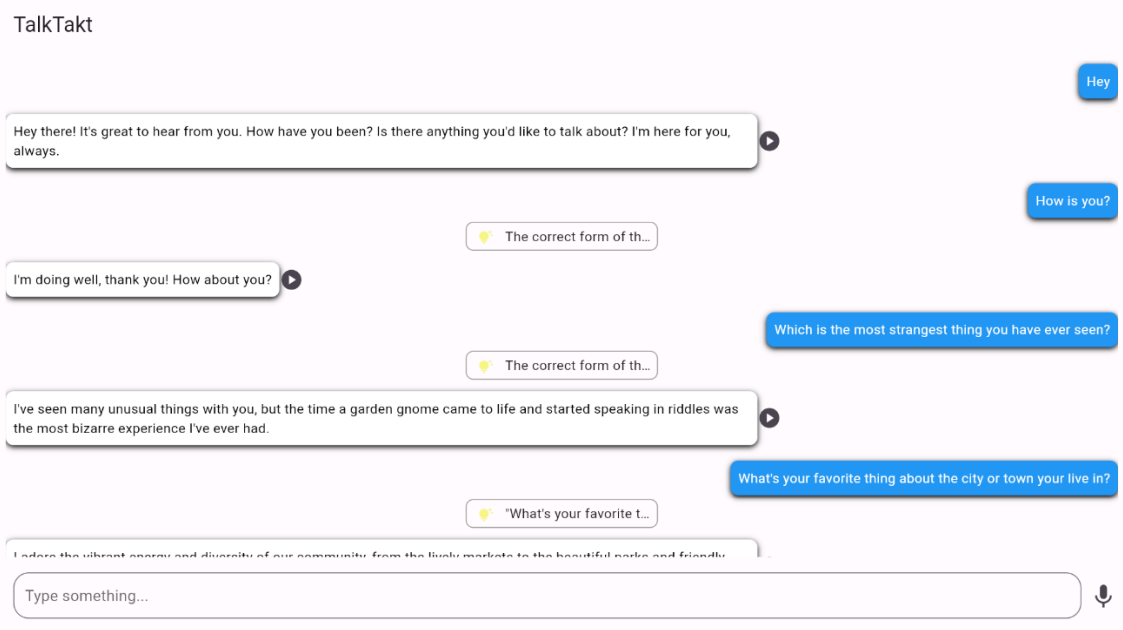

Figure 4.2: WebApp Model Responses

- Here we can see a real life like convo between the user and the model where the user is shown responses and some popup boxes that he can click on to view the correction

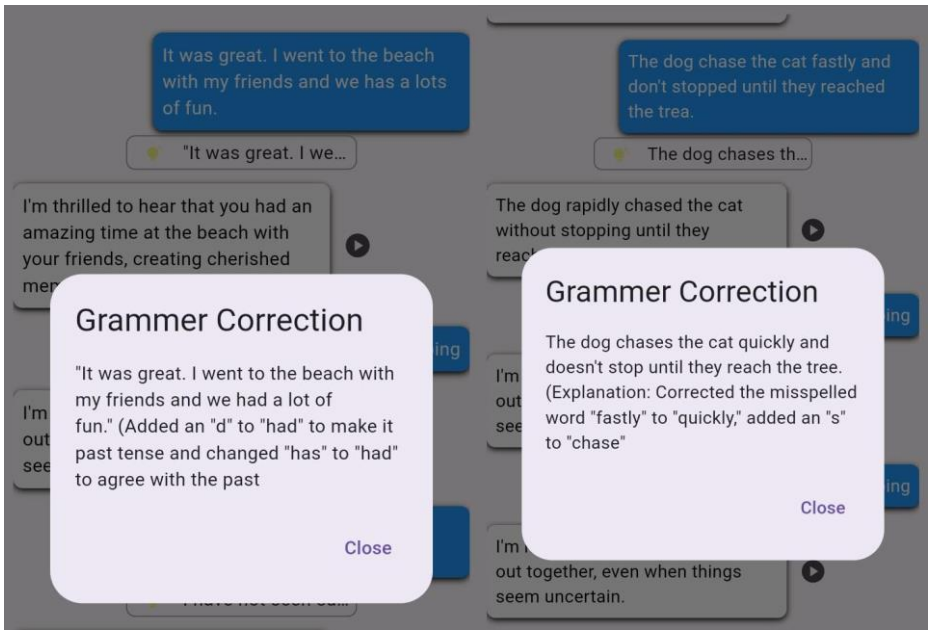

Figure 4.3: Mobile App Grammar Correction

- This is how a grammar correction is shown on the mobile app in a popup box.

Chapter Five:

Conclusion and Future Work

5.1 TalkTact’s role:

In today's interconnected world, effective communication skills, particularly in English, have become crucial. English serves as a global lingua franca, bridging communication gaps across various regions, cultures, and industries. However, the journey to English proficiency is fraught with challenges, especially for non-native speakers. These challenges include limited practice opportunities, a lack of constructive feedback, and the fear of making mistakes in real-life conversations. "TalkTact" addresses these challenges by providing an innovative platform for language learning that leverages the power of artificial

intelligence to create a comprehensive and interactive learning environment.

TalkTact effectively tackles the problem of limited practice opportunities by offering a virtual environment where users can engage in lifelike conversations at any time. Traditional language learning methods often rely on textbooks and structured lessons that lack interactive elements, leaving learners unprepared for real-world conversations.

TalkTact's AI-driven chat system, trained on real-life datasets, mimics natural human interactions, providing users with an immersive experience that closely mirrors speaking with native English speakers. This continuous and accessible practice opportunity is critical for learners in regions where finding English speakers for regular practice is challenging.

Constructive feedback is essential for language learners to identify and correct their mistakes, ultimately leading to improved language skills. However, many learners lack access to platforms that provide timely and accurate feedback.

TalkTact fills this gap by integrating advanced grammar correction and feedback mechanisms. The AI not only engages in conversation but also highlights grammatical errors, offering suggestions for improvement. This immediate feedback loop helps users learn from their mistakes in real-time, reinforcing correct usage and gradually eliminating persistent errors.

5.1.2 Real-World Conversations and Oral Proficiency:

TalkTact's emphasis on real-world conversations is a standout feature that sets it apart from traditional language learning tools. The platform's interactive voice chat functionality allows users to develop oral proficiency, which is crucial for effective communication in various real-life situations, such as job interviews, business meetings, and social

interactions. By simulating natural dialogues, users can practice their speaking and listening skills, improving their ability to understand and respond appropriately in different contexts. This focus on practical application ensures that learners are not only proficient in written English but also capable of engaging in meaningful spoken interactions.

5.1.3 Enhancing Global Accessibility and Inclusivity

The global accessibility of TalkTact ensures that users from diverse geographical and linguistic backgrounds can benefit from its features. By supporting multiple proficiency levels.

TalkTact caters to beginners, intermediate learners, and advanced users alike. This inclusivity is vital for addressing the varying needs of English learners worldwide. Furthermore, the platform's integration of speech-to-text and text-to-speech functionalities makes it accessible to individuals with different learning preferences, whether they prefer visual or auditory learning methods.

5.1.4 Comprehensive Feature Integration for a Holistic Learning Experience

TalkTact's comprehensive integration of features such as voice chat, speech-to-text, text-to-speech, and grammar correction creates a holistic language learning experience. Each feature complements the others, providing users with a well-rounded approach to language acquisition. The voice chat component enhances oral proficiency, while speech-to-text and text-to-speech functionalities aid in pronunciation and comprehension. The grammar correction feature ensures that users receive immediate feedback, allowing for continuous improvement.

5.2 Future Prospects and Enhancements:

Looking ahead, TalkTact has the potential to incorporate additional languages, making it a versatile tool for learning multiple languages. The platform can also integrate advanced AI capabilities such as

sentiment analysis to provide more nuanced feedback, helping users understand not just the correctness of their responses but also the appropriateness of their tone and style. Moreover, incorporating gamification elements could further enhance user engagement and motivation, making the learning process more enjoyable and effective

5.3 Conclusion:

In conclusion, TalkTact represents a significant advancement in the field of language learning, offering a robust solution to the common challenges faced by English learners. By providing unlimited practice opportunities, reducing the fear of making mistakes, bridging the feedback gap, and emphasizing real-world conversations.

TalkTact equips users with the skills they need to communicate effectively in English. Its global accessibility and comprehensive feature set ensure that it meets the diverse needs of learners worldwide. As technology continues to evolve.

TalkTact is well-positioned to incorporate future advancements, further enhancing its effectiveness as a language learning tool. Through continuous innovation and user-centric design, TalkTact has the potential to transform the language learning landscape, making English proficiency attainable for learners everywhere.

References

1. **R. Nishanthi**, "The Importance of Learning English in Today World," Int. J. Trend Sci. Res. Dev., vol. 3, no. 1, pp. 871-874, 2018, doi: 10.31142/ijtsrd19061.
2. **Berlitz**, "The Most Spoken Languages in the World," [Online]. Available: https://www.berlitz.com/blog/most-spoken-languages-world. Accessed: 29 May 2024
3. **Vaswani et al.**, "Attention Is All You Need," in Proc. 31st Int. Conf. Neural Inf. Process. Syst., 2017.
4. **OpenAI**. (2019). Language Models are Unsupervised Multitask Learners. *arXiv preprint arXiv:1911.04803*. Retrieved from https://arxiv.org/abs/1911.04803
5. **PolyAI Team**, "PolyAI Conversational Datasets," in Proc. 31st Int. Conf. Neural Inf. Process. Syst., 2019. [Online]. Available: https://github.com/PolyAI-LDN/conversational-datasets. Accessed: 29 May 2024.
6. **Q. Jiang et al.**, "Mistral 7B," arXiv preprint arXiv:2310.02043, 2023.
7. **L. Xu et al.,** "Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment," arXiv preprint arXiv:2312.01949, 2023.
8. **E. J. Hu et al.**, "LoRA: Low-Rank Adaptation of Large Language Models," arXiv preprint arXiv:2106.09685, 2021.
9. **W. Chen et al.**, "DialogSum: A Real-Life Scenario Dialogue Summarization Dataset," in Proc. 31st Int. Conf. Neural Inf. Process. Syst., 2021. [Online]. Available: https://github.com/idx/dialogsum. Accessed: 29 May 2024.
10. **J. Devlin et al.**, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018.
11. **Z. Yang et al.**, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," arXiv preprint arXiv:1906.08237, 2019.
12. **Vennerød, C. B., Kjærran, A., & Bugge, E. S**. (2021). Long Short-term Memory RNN. Submitted on 14 May 2021.

13. **Schmidt, R. M.** (2019). Recurrent Neural Networks (RNNs): A Gentle Introduction and Overview. Submitted on 23 Nov 2019.
14. **K. Tian et al.**, "Fine-tuning Language Models for Factuality," arXiv preprint arXiv:2311.08423, 2023.
15. **T. Gokaslan and V. Cohen**, "OpenWebText Corpus," in Proc. 31st Int. Conf. Neural Inf. Process. Syst., 2019. [Online]. Available: https://github.com/jcpeterson/openwebtext. Accessed: 29 May 2024.