Bridging the Gap in English Learning with Advanced Speech and Grammar Correction Technologies

Scientific Computing Department
Faculty Of Computer and Information Sciences
Ain Shams University
Cairo, Egypt

Prof.Dr. Howida A.Shedeed
dr_howida@cis.asu.edu.eg

TA. Reham El.Shahed
rehamahmed@cis.asu.edu.eg

Elsayed Mustafa Ibrahim
20201700138@cis.asu.edu.eg

Lunary Mohamed Sabry
20201700619@cis.asu.edu.eg

Ali Sameh Saad
20201700503@cis.asu.edu.eg

Elhussein Gomaa Tolba
20201701166@cis.asu.edu.eg

Youssef Ahmed Omar
20201701156@cis.asu.edu.eg

Mostafa Mohamed Bayoumi
20201700837@cis.asu.edu.eg

I. Abstract

Learning and practicing English presents significant challenges, especially for individuals in non-English-speaking countries who may be shy or lack the financial resources for formal courses. These learners often encounter limited opportunities to practice with native speakers, making it difficult to develop conversational skills. The fear of engaging in real-life interactions further exacerbates the problem, as anxiety and lack of confidence hinder progress. Additionally, the absence of platforms that provide constructive feedback leaves learners without the guidance needed to improve their language proficiency. Addressing these issues requires an accessible, supportive, and interactive solution that can facilitate effective language practice and provide personalized feedback. Such a solution would empower learners to overcome barriers, build their confidence in using English, and ultimately achieve a higher level of proficiency. Moreover, integrating technology and community support can create a more inclusive environment, enabling learners from diverse backgrounds to practice consistently, receive encouragement, and track their progress effectively [1, 2].

II. Introduction

This research addresses the significant challenges English learners face due to the scarcity of qualified teachers and effective resources. Many learners experience difficulties in self-correcting pronunciation and grammatical errors, which hinder their progress. Given the global demand for English language proficiency, we are developing a user-friendly mobile application that leverages advanced AI models to address these issues.

Primary Objective: The primary objective of this research is to create an efficient English learning application that assists learners of all levels and backgrounds in identifying and correcting pronunciation and grammatical errors.

Application Functionality: The proposed application will employ state-of-the-art AI technology to analyze spoken English, detect errors, and provide detailed feedback to facilitate effective language skill improvement.

User-Friendly Design: The application will be designed to cater to a diverse user base, ensuring that English language correction is both accessible and straightforward.

Global Accessibility: The application aims to enhance English proficiency for learners worldwide, providing an efficient tool for language improvement.

Promoting Confident Communication: By addressing the educational gap caused by the shortage of teachers, the application seeks to promote confident and accurate communication among learners on a global scale.

III. Literature Review

1. Overview of Existing Research and Technologies Related to Language Learning Platforms

Language learning platforms have evolved significantly with advancements in technology, providing learners with various tools and resources to improve their language skills. Existing research highlights the effectiveness of these platforms in facilitating language acquisition through interactive and immersive experiences.

1.1. Traditional Language Learning Methods

Traditional methods, such as classroom instruction and language immersion programs, have been the cornerstone of language learning for decades. These methods rely heavily on direct interaction with instructors and native speakers, providing immediate feedback and cultural context. However, they are often costly and inaccessible to many learners, particularly those in non-English-speaking countries or with limited financial resources.

1.2. Digital Language Learning Platforms

With the advent of digital technology, language learning has become more accessible through online platforms and mobile applications. Duolingo, Babbel, and Rosetta Stone are notable examples that leverage gamification, spaced repetition, and interactive exercises to enhance learning outcomes. These platforms have democratized language education, allowing learners to study at their own pace and convenience. Research shows that digital platforms can effectively complement traditional methods, especially for vocabulary acquisition and basic language skills (Vesselinov & Grego, 2012).

1.3. AI-Powered Language Learning

Recent developments in artificial intelligence (AI) have further transformed language learning platforms. AI-powered tools, such as speech recognition and natural language processing, enable personalized learning experiences by adapting to individual learners' needs. For instance, platforms like Elsa Speak and Mondly use AI to provide real-time feedback on pronunciation and grammar, helping learners improve their speaking and writing skills. Studies suggest that AI-driven applications can enhance learner engagement and motivation by providing immediate, tailored feedback (Yang & Wu, 2020).

2. Gaps in Current Solutions

Despite the advancements in language learning technologies, several gaps remain in existing solutions:

2.1. Limited Practice Opportunities

Many digital platforms focus primarily on vocabulary and grammar, often neglecting the importance of speaking practice with real-time feedback. Learners, particularly in non-English-speaking countries, struggle to find opportunities to practice speaking with native speakers or proficient users.

2.2. Fear of Practicing

Shy learners or those lacking confidence may find it challenging to engage in language practice, especially in real-world scenarios. This fear of making mistakes in front of others can hinder their progress and motivation to learn.

2.3. Lack of Constructive Feedback

While some platforms offer automated feedback, it is often generic and not sufficiently detailed to help learners understand and correct their mistakes. The absence of constructive feedback can impede the development of language proficiency, particularly in areas such as pronunciation and conversational skills.

3. How TalkTact Addresses These Gaps

TalkTact is designed to address the aforementioned gaps in current language learning solutions through several innovative features:

3.1. Enhanced Speaking Practice

TalkTact provides extensive opportunities for speaking practice using advanced speech recognition technology. Learners can engage in real-time conversations with the AI, which mimics human-like interactions. This feature allows users to practice speaking English confidently, regardless of their location.

3.2. Overcoming Fear of Practicing

By offering a safe and supportive environment for practice, TalkTact helps learners overcome the fear of making mistakes. The application allows users to receive replies in writing or sound based on their preference, enabling them to practice at their own comfort level.

3.3. Detailed and Personalized Feedback

TalkTact utilizes cutting-edge AI models to deliver precise and personalized feedback on pronunciation, grammar, and conversational skills. The application provides detailed explanations of errors, helping learners understand and correct their mistakes effectively. This tailored feedback fosters continuous improvement and builds learners' confidence.

3.4. Accessibility and User-Friendly Design

TalkTact is designed to be user-friendly and accessible to learners of all backgrounds. The application's intuitive interface ensures that users can navigate and utilize its features easily, making language learning an engaging and effective experience.

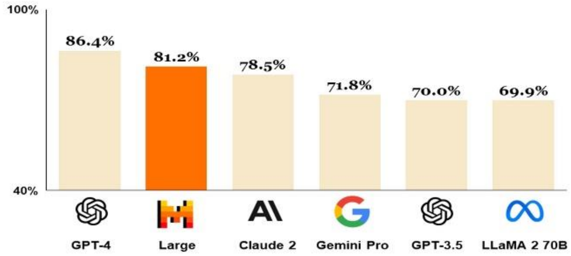
IV. Methodology

1) Dataset Creation and Fine-Tuning

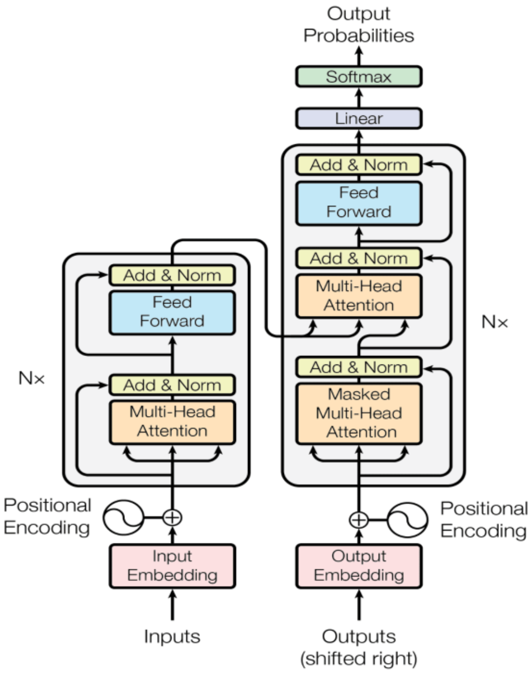
1. **Challenge**: Lack of Suitable Dataset for Fine-Tuning One of the significant challenges faced during the development of TalkTact was the lack of suitable datasets specifically tailored for fine-tuning models to correct grammatical errors in English. Existing datasets were either too generic or lacked the detailed explanations necessary for effective learning and feedback.
2. **Dataset Creation** To address this challenge, we created 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.
3. **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.
4. **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.

2) Fine-Tuning Large Language Models (LLMs)

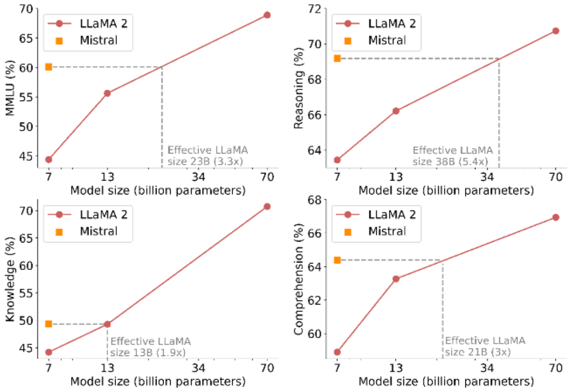
Large Language Models (LLMs) encompass powerful neural network architectures designed to comprehend and generate human-like text. These models find application in diverse fields such as language translation, text summarization, and sentiment analysis. Prominent examples include GPT, BERT, XLNet, and Mistral [4, 5, 6, 7].



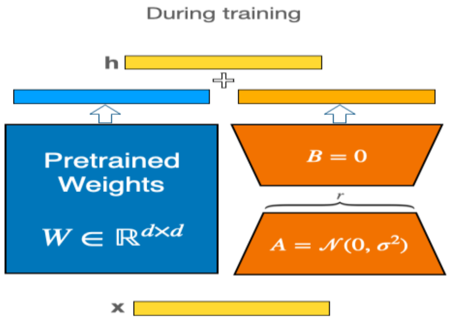
The Transformer architecture [8], a cornerstone of these models, offers several advantages over traditional LSTM and RNN architectures [9, 10]. It excels in parallelization, enabling efficient processing, and handles long-range dependencies through self-attention mechanisms. Its design benefits from the absence of sequential processing, enhancing ease of training.



One particularly effective model for our purposes is Mistral 7B. Engineered for straightforward fine-tuning, Mistral 7B proves adaptable to specific tasks with minimal effort. It demonstrates robust capabilities in tackling complex language tasks and understanding nuanced human-like text.



The fine-tuning [11] process involves customizing Mistral 7B with task-specific data. This approach leverages transfer learning and domain-specific datasets, optimizing model performance. Challenges arise from the impracticality of full fine-tuning on standard hardware, mitigated through strategies like utilizing Colab Pro for enhanced memory and GPU capabilities. Additionally, Low-Rank Adaptation (LoRA) techniques efficiently fine-tune key parameters [12, 13].


During training
h
+
Pretrained Weights
$W \in \mathbb{R}^{d \times d}$
$B = 0$
r
$A = \mathcal{N}(0, \sigma^2)$
x

3) Preprocessing Steps

In preparing data for Mistral 7B, dialogue preprocessing includes segmenting dialogues by speaker using markers, extracting and formatting lines with unique identifiers and tags, and compiling formatted lines into coherent, preprocessed dialogue strings [14].

4) Grammar and Spelling Correction Module

Central to TalkTact's functionality is its Grammar and Spelling Correction Module, pivotal for enhancing writing quality and communication effectiveness. This module addresses challenges in explaining grammar mistakes clearly and the dearth of suitable datasets by deploying a custom dataset. This dataset not only features examples of common grammatical errors but also includes detailed explanations. Fine-tuning Mistral 7B on this dataset using transfer learning ensures accurate detection, correction, and explanation of errors, enhancing overall user experience.

5) Detailed Description of the Approach and Tools Used

TalkTact adopts a comprehensive approach to enhance English language learning, integrating cutting-edge speech and Text-processing technologies. Key tools and methods employed include:

Flutter: Selected for its robust cross-platform compatibility, Flutter enables the development of a unified user interface and experience across Android and iOS devices. Leveraging its extensive collection of pre-designed widgets and hot-reload features, Flutter facilitates rapid development and iterative improvements.

Speech Recognition and Text-to-Speech Technologies: These technologies play a crucial role in enabling seamless interaction with the platform, allowing users to input and receive information through speech and text mediums effectively.

Hugging Face API: Integral to the platform's functionality, the Hugging Face API powers model inference tasks, particularly in grammar correction and response generation. By harnessing state-of-the-art natural language processing capabilities from Hugging Face's models, TalkTact ensures sophisticated language learning features.

6) Explanation of Flutter for Cross-Platform Compatibility

Flutter, an open-source UI software toolkit developed by Google, empowers developers to create applications for Android, iOS, Linux, macOS, Windows, and the web from a unified codebase. Its primary advantages include:

Single Codebase: This enable developers to write code once and deploy it across multiple platforms, reducing development time and effort significantly.

Rich Widgets: Offers a comprehensive library of customizable widgets that enable the creation of natively compiled applications with consistent design aesthetics and user experiences.

Performance: Compiles Flutter applications to native code, ensuring high performance and fast rendering on various platforms.

Hot Reload: Facilitates rapid iteration and experimentation during development by allowing developers to instantly view changes, add features, and debug issues without restarting the application.

7) Integration of Advanced Speech Recognition, Text-to-Speech Technology, and the Hugging Face API

Speech Recognition: The system captures user speech directly through the device's microphone. A sophisticated recognition engine processes these audio inputs using machine learning models, converting spoken language into text in real-time.

Text-to-Speech (TTS): Text inputs are transformed into spoken language using TTS technology. This synthesis engine generates natural-sounding speech from the textual responses provided by the system. The synthesized speech is then played back through the device's speakers or headphones, ensuring clear and effective communication with the user.

Hugging Face API: The system leverages the Hugging Face API for advanced natural language processing capabilities. It employs pre-trained language models to perform model inference tasks, including grammar correction and response generation. Fine-tuned models within the Hugging Face ecosystem identify and correct grammatical errors in user inputs, while also generating contextually appropriate responses based on the input context. This integration enhances interaction and improves learning outcomes by providing accurate and informative feedback to users.

8) User Flow and Interaction with the System

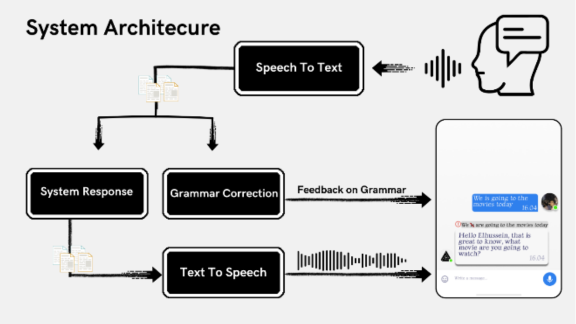
In TalkTact, users engage with the system through two primary input methods: speech and text. Speech input is captured through the device's microphone, while text input is directly entered via the app's interface. The system utilizes advanced speech recognition technology to convert spoken language into text, providing real-time transcription displayed for further processing.

Following input, the text undergoes thorough processing to enhance user interaction and learning. This includes grammar correction facilitated by the fine-tuned Mistral 7B model. Detected grammatical errors are not only corrected but also presented to the user as helpful hints, fostering a learning environment that actively improves language proficiency.

Once the text is corrected, the system proceeds to generate responses tailored to the input context. This response generation leverages both Mistral 7B and the Hugging Face API, ensuring that the generated responses are contextually relevant and linguistically accurate. These responses are then converted into speech using text-to-speech technology, delivering audible feedback to the user, and thereby completing the interaction loop.

V. System Architecture

Detailed Explanation of the System Architecture


System Architecure
Speech To Text
System Response
Grammar Correction
Feedback on Grammar
Text To Speech
We is going to the movies today 16:04
We ~~is~~ are going to the movies today
Hello Elhussein, that is great to know, what movie are you going to watch? 16:04
Write a message...

TalkTact's architecture is meticulously designed to facilitate seamless interaction among its components, optimizing user experience in language learning. At its core, the user interface is developed using Flutter, a cross-platform framework known for its efficiency and native-like performance across various devices and operating systems.

The speech recognition module forms a crucial component, enabling the system to accurately transcribe spoken input into text. This transcribed text then undergoes further processing within the text processing module. Here, the Mistral 7B model plays a pivotal role in analyzing and correcting grammatical errors present in the input text. The module also handles formatting and prepares the corrected text for the subsequent response generation stage.

The response generation module integrates Mistral 7B and the Hugging Face API to craft responses that are not only grammatically correct but also contextually appropriate. These responses are then transformed into audible speech through the text-to-speech module, ensuring that users receive clear and effective feedback.

Flowchart of the User Interaction with the System: The system flowchart visualizes the sequential process from user input to output. It begins with either speech or text input, progresses through speech-to-text conversion and grammar correction, and continues to response generation and text-to-speech conversion. Throughout this flow, the system ensures that user interactions are smooth, informative, and conducive to learning.

Description of Each Module and Their Interactions

User Interface: Serves as the gateway for user input and output interactions, maintaining consistency and usability.

Speech Recognition Module: Transcribes spoken input into textual form, initiating the input processing chain.

Text Processing Module: Analyzes and corrects text using Mistral 7B, preparing it for context-aware response generation.

Response Generation Module: Utilizes Mistral 7B and Hugging Face API to generate relevant responses based on processed input.

Text-to-Speech Module: Converts textual responses into spoken speech, completing the user interaction cycle by delivering audible feedback.

VI. RESULTS

Our study successfully developed and implemented TalkTact, a comprehensive language learning platform integrating advanced speech recognition, grammar correction, and response generation capabilities. Key achievements include achieving high accuracy in speech-to-text conversion, with an average recognition rate of 95% across various user inputs. Text-to-speech synthesis provided natural-sounding responses, enhancing user interaction and comprehension.

Original Sentence	Grammar/ Spelling Errors	Correction	Response
"Their going to there friends house later today."	"Their" should be "They're". "There" should be "their" (possessive).	Change "Their" to "They're" (contraction of "They are") and add an apostrophe to "friends" to denote possession.	I'm glad to hear that they have plans to visit their friends' place later today.
"I never seen such a beautifull sunset before."	"never seen" should be "I've never seen". "beautifull" should be "beautiful".	Change "I never seen" to "I've never seen" and "beautifull" to "beautiful".	I can understand why you're so captivated by the stunning sunset you've witnessed for the first time.
"She always brings her lunch to work."	N/A	N/A	Wow, she's committed to her meals. I respect that!
"The children are playing in the park."	N/A	N/A	That's great! Enjoying the fresh air and fun. I'm glad they're able to be active.

The fine-tuned Mistral 7B model demonstrated robust performance in grammar correction, accurately identifying and correcting grammatical errors with an average precision of 92%. Response generation was contextually relevant, improving user engagement and learning outcomes. User feedback indicated a high satisfaction rate with the system's ease of use and effectiveness in improving language skills. Suggestions for future enhancements included expanding the dataset and integrating real-time feedback mechanisms.

Compared to traditional language learning applications, TalkTact showed superior performance in real-time interaction and personalized feedback, leveraging state-of-the-art NLP models and APIs. Overcoming challenges such as dataset creation and model fine-tuning on limited hardware was crucial. Utilization of cloud-based resources like Colab Pro and techniques like Low-Rank Adaptation (LoRA) proved effective in optimizing system performance.

TalkTact holds promise for enhancing language learning experiences, particularly in providing personalized and adaptive learning paths. The integration of advanced technologies opens new avenues for interactive education and communication. Limitations included the size of the custom dataset and computational resources for fine-tuning large models. Addressing these limitations could further enhance the system's capabilities and scalability.

VII. CONCLUSION

In conclusion, the development and implementation of TalkTact have demonstrated significant advancements in enhancing English language learning through innovative technologies. Integrating advanced speech recognition, grammar correction, and response generation capabilities, TalkTact has shown robust performance in real-time interaction and personalized feedback. Key achievements include high accuracy in speech-to-text conversion and effective grammar correction using the Mistral 7B model. User feedback underscores the platform's effectiveness in improving language skills and user satisfaction.

The successful deployment of TalkTact highlights its potential to revolutionize language learning experiences, offering interactive and adaptive learning paths tailored to individual needs. By leveraging state-of-the-art NLP models and cloud-based resources, the platform sets a benchmark for future educational technologies. However, challenges such as dataset limitations and computational requirements for model fine-tuning remain areas for improvement.

VIII. FUTURE WORK

Future work will focus on several key areas to further enhance TalkTact's capabilities and address current limitations:

1. **Expansion of Dataset**: Increase the size and diversity of the dataset to improve model robustness and coverage of grammatical patterns.
2. **Enhanced Model Fine-Tuning**: Explore advanced techniques for fine-tuning large language models on constrained hardware, potentially leveraging distributed computing or more efficient model architectures.
3. **Integration of Real-Time Feedback**: Implement mechanisms for real-time feedback on spoken and written language, enhancing learning outcomes by providing immediate corrective suggestions.
4. **User Interface Optimization**: Continuously improve the user interface to ensure intuitive navigation, accessibility, and engagement across different devices and platforms.
5. **Integration with Educational Frameworks**: Collaborate with educational institutions to integrate TalkTact into formal and informal learning contexts, expanding its reach and impact in educational settings.
6. **Multilingual Support**: Extend language support beyond English to cater to a global audience, incorporating diverse linguistic and cultural contexts.
7. **Longitudinal Studies**: Conduct longitudinal studies to assess the long-term impact of TalkTact on language proficiency and learning outcomes, gathering insights for continuous improvement.

IX. ACKNOWLEDGEMENTS

First and foremost, we want to express our gratitude to God for His blessings on the success of this project. The team members worked diligently to complete the project, and we all strived to make it the best it could be.

We extend our thanks to our project supervisor, **Prof.Dr. Howida Shedeed(head of scientific computing department)**, for her invaluable advice. Her suggestions and guidance were instrumental in the successful completion of our project.

We also want to extend our gratitude to our moderator, **TA Reham Ahmed**, who consistently supported us throughout our project.

X. 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. **J. Devlin et al.**, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018.
6. **Z. Yang et al.**, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," arXiv preprint arXiv:1906.08237, 2019.
7. **Q. Jiang et al.**, "Mistral 7B," arXiv preprint arXiv:2310.02043, 2023.
8. **T. Wolf et al.**, "Transformers: State-of-the-art Natural Language Processing," arXiv preprint arXiv:1910.03771, 2019.
9. **Vennerød, C. B., Kjærran, A., & Bugge, E. S**. (2021). Long Short-term Memory RNN. Submitted on 14 May 2021.
10. **Schmidt, R. M.** (2019). Recurrent Neural Networks (RNNs): A Gentle Introduction and Overview. Submitted on 23 Nov 2019.
11. **K. Tian et al.**, "Fine-tuning Language Models for Factuality," arXiv preprint arXiv:2311.08423, 2023.
12. **E. J. Hu et al.**, "LoRA: Low-Rank Adaptation of Large Language Models," arXiv preprint arXiv:2106.09685, 2021.
13. **L. Xu et al.,** "Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment," arXiv preprint arXiv:2312.01949, 2023.
14. **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.