This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3443592

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

LLM-Based Text Prediction and Question Answer Models for Aphasia Speech

Shamiha Binta Manir[1], PhD candidate, K M Sajjadul Islam[2], PhD student, Praveen Madiraju[3], PhD, Priya Deshpande[4], PhD

[1]Department of EECE, Marquette University, Milwaukee, WI 53233, USA
[2]Department of CS, Marquette University, Milwaukee, WI 53233, USA
[3]Department of CS, Marquette University, Milwaukee, WI 53233, USA
[4]Department of EECE, Marquette University, Milwaukee, WI 53233, USA

Corresponding author: Shamiha Binta Manir (e-mail: shamihabinta.manir@marquette.edu).

ABSTRACT
Aphasia, a brain injury-related linguistic problem, hinders communication. Current techniques generally struggle to handle aphasic speech's intricacies. BERT, short for Bidirectional Encoder Representations from Transformers, is a pre-trained natural language model that utilizes contextual information from both preceding and succeeding words in a sentence to predict the target word. This study uses BERT models to predict and fill in sentences for people with aphasia, using the AphasiaBank dataset. The patients' transcripts were thoroughly preprocessed, with nonverbal clues and redundant phrases removed. Because of the lack of control data, the accuracy of BERT in predicting masked tokens in aphasic speech was evaluated using a manual rating system with four raters. In addition, BERT was used for question-answering to increase context comprehension, underlining its ability to aid communication for those with aphasia. The preprocessing pipeline used advanced text-cleaning algorithms to ensure input data quality. The evaluation of BERT performance yielded satisfactory results with strong inter-rater reliability. The inter-rater correlation was remarkably strong, overall coefficients ranging from 0.61 to 0.74, suggesting a substantial level of agreement (Fleiss' Kappa Score: 0.32). BERT's predictions demonstrated a significant degree of contextual relevance and grammatical accuracy, as proven by ratings that were primarily above 3.0. The box plots also suggested a minimal number of outliers. The goal of this method is to improve the accuracy of speech prediction, which is beneficial for caregivers and speech therapists. BERT shows its nuanced capability in Aphasia sentence completion tests by exhibiting exceptional performance in terms of contextual appropriateness and grammatical correctness, as confirmed by manual evaluation.

INDEX TERMS Aphasia, BERT models, Natural Language Processing (NLP), Speech prediction, Sentence completion, AphasiaBank, Transformer models, Patient-spoken transcripts, Communication aids, Speech therapy, Text preprocessing

I. INTRODUCTION
Aphasia is a language impairment caused by brain injury, and speech assessment is important in determining its severity [1]. Narrative spontaneous speech is used for assessment, but this speech is atypical due to language impairment. Automatic recognition of aphasic speech is difficult due to various impairments and limited training data [2]. AphasiaBank, a shared database primarily used by clinicians to study aphasia, provides valuable data for modeling speech recognition using deep neural networks [3]. This research utilizes Natural Language Processing (NLP) techniques on the AphasiaBank database to develop speech prediction models for individuals with aphasia, addressing the challenges posed by this language impairment caused by brain injury. The objective is to train NLP models on conversational transcripts from AphasiaBank, employing data preprocessing and model training using transformer models like BERT (Bidirectional Encoder Representations from Transformers) [4]. Performance evaluation on validation and test sets contributes to enhancing communication and understanding for aphasia patients, benefiting caregivers and speech therapists.

Additionally, this research aims to analyze the language impairments exhibited by individuals with aphasia using NLP techniques to extract linguistic features such as word

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

usage, syntactic patterns, and discourse coherence. By identifying patterns specific to different types and severities of aphasia, valuable insights into the nature of these language impairments are gained. Furthermore, exploring text classification techniques enables the classification of aphasia types (e.g., Broca's aphasia and Wernicke's aphasia) and the identification of associated symptoms like word-finding difficulty or speech comprehension problems [5]. This approach provides objective assessments of aphasia subtypes and symptomatology.

By integrating language impairment analysis and text classification, a comprehensive understanding of aphasia is achieved, fostering the development of effective speech prediction models. Moreover, it advances the field of aphasia research by uncovering linguistic patterns and offering diagnostic and clinical insights. Leveraging the AphasiaBank dataset, NLP models assist individuals with aphasia by predicting the next word or generating complete sentences based on partial input, reducing communication effort [6]. This method enhances the communication capabilities of individuals with aphasia by facilitating the creation of speech prediction models that are more contextually appropriate and precise. Additionally, it offers a comprehensive understanding of the various varieties and severity levels of aphasia, which is beneficial for the classification and diagnosis of aphasia subtypes. Furthermore, the incorporation of deep neural network techniques like BERT paves the way for personalized language models that adapt to individual speech patterns and improve over time with more data.

The project intends to improve aphasia intervention tactics by building prediction techniques that can adapt to the varying demands of patients. This adaptable methodology enables the ongoing enhancement of communication aids, making them progressively more effective as they acquire knowledge from each interaction. The primary objective is to develop advanced, personalized methods that not only enhance the accuracy of speech pattern prediction but also enable caregivers and speech therapists to employ superior interpretation and response strategies.

These initiatives are intended to result in a number of practical applications that will significantly benefit people with aphasia. By achieving these goals, the project will enhance current communication methods and offer essential resources that enhance independence and quality of life for individuals with aphasia. This research not only improves the technical domains of NLP and speech therapy, but also significantly influences healthcare practices by combining advanced computational models with therapeutic understanding.

II. BACKGROUND

Advancements in the field of speech prediction models and aids for individuals with aphasia have made great progress, while there are still certain issues that need to be addressed. Prior research has employed diverse methodologies to improve the precision and efficiency of these models, frequently making use of the AphasiaBank database due to its extensive collection of aphasic speech data.

Le et al. employed the AphasiaBank database to enhance the precision of speech recognition for individuals suffering from aphasia. Their methodology consisted of choosing speakers from various sub-databases, extracting distinctive characteristics, and building a baseline for automated speech recognition (ASR). The researchers examined both conventional and contemporary methods to enhance the precision of recognition, emphasizing the possibility of utilizing larger and more varied datasets to enhance the performance of the model [7].

Jothi et al. created a speech evaluation system that utilized a combination of machine learning classifiers and textual/acoustic characteristics to classify the extent of aphasia. Their technique employed Mel-frequency cepstral coefficients (MFCCs) and ensemble classifiers for audio analysis, enabling accurate categorization of aphasia severity while reducing overfitting by selecting relevant features [8].

Qin et al. conducted a study that specifically examined the utilization of text-based characteristics to evaluate aphasia in patients who speak Cantonese. The researchers utilized a continuous bag of words (CBOW) model that incorporated syllable-level embeddings to acquire word vectors and supra-segmental length features. They showed that these text features were successful in differentiating between speech that was impaired and speech that was unimpaired [9].

MacWhinney and Fromm conducted an extensive analysis of 45 studies that utilized the AphasiaBank database. The studies included several aspects like speech, grammar, lexicon, gesture, fluency, syndrome classification, social factors, and therapy effects. Their research emphasized the practicality of using the database to analyze different elements of aphasic language productions and indicated the possibility of increasing the database's size by collecting recordings on-site and through teletherapy sessions [10].

Clough and Gordon examined the elements that influence fluency assessments in individuals with aphasia by comparing classifications made using the Western Aphasia Battery (WAB-R) with clinical evaluations. The researchers examined many aspects of spontaneous speech variables in the AphasiaBank database, with a specific emphasis on grammatical proficiency, lexical recall, and speech generation. Their research highlighted the intricate nature of evaluating fluency and proposed adjustments to current measurement tools in order to improve the consistency between clinical and standardized assessments [11].

Qin, Le, and Kong introduced a speech evaluation system for Cantonese-speaking aphasic patients. They used an ASR system to extract reliable text characteristics from incorrect outputs. The researchers integrated textual data with supra-segmental duration features to analyze unusual prosody patterns and successfully demonstrated the system's efficacy in binary classification and automated prediction of subjective assessment scores [9].

Park and Park conducted an investigation into several neural language models for sentence completion, ultimately attaining exceptional outcomes by refining pre-trained BERT models. Their research showcased the capabilities of transformer-based models in phrase prediction tasks, indicating that comparable methodologies could be employed to enhance text prediction in individuals with aphasia [12].

Jothi et al. conducted a comprehensive analysis of different methodologies used in automatic speech assessment systems that categorize the intensity of aphasic speech. The researchers assessed several characteristics and classifiers, emphasizing the efficacy of deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) in accomplishing this objective [13].

Le and Mower Provost conducted a study on the AphasiaBank database, where they set a baseline for large-vocabulary continuous speech recognition (LVCSR). Their findings showed that discriminative pretraining can greatly enhance recognition rates for speech affected by aphasia. Their investigation unveiled the diverse advantages of distinct adaptation techniques, contingent upon the severity of aphasia, indicating that customized procedures could further augment recognition precision [14].

Shih-Hsuan Chiu and Berlin Chen proposed a new method called TPBERT (Task-specific Pre-trained BERT), which is based on BERT, to reorder the N-best hypotheses in automated speech recognition (ASR) problems. TPBERT utilizes task-specific global topic information to improve the BERT model's capacity to efficiently embed N-best hypotheses. The study conducted a series of comprehensive experiments on the AMI benchmark corpus to assess the efficacy and viability of the suggested methods in contrast to traditional autoregressive models such as RNN and other BERT-based reranking methods [15].

The SpeechBERT model, an audio-and-text jointly learnt language model for end-to-end spoken question answering (SQA), was proposed by Yung-Sung Chuang et al. In order to generate embeddings for audio words that are aligned with text BERT embeddings, SpeechBERT is pre-trained on audio and text datasets. Through information extraction from audio data, the model showed higher performance in SQA tasks, especially when ASR errors were present [16].

Chia-Chih Kuo et al. developed an audio-enhanced framework based on BERT for answering multiple-choice questions in spoken form (SMCQA). This framework combines acoustic-level data with text-level data to enhance the resilience and precision of SMCQA systems. Their trials demonstrated significant enhancements in accuracy compared to chosen reference points and cutting-edge systems on a publicly available Chinese SMCQA dataset [17].

Torre et al. investigated innovative semi-supervised learning techniques applied to the AphasiaBank dataset to tackle obstacles in identifying aphasic speech. This methodology showcased enhancements for the English language and set benchmarks for Spanish, emphasizing the efficacy of semi-supervised techniques in improving ASR systems used in clinical speech disorders [18].

Mahmoud et al. conducted a comparison between specific machine learning algorithms and pre-made platforms for aphasia assessment batteries. The study indicated that CNN-based speech recognition algorithms perform better than LDA and commercial platforms such as Microsoft Azure and Google. The research in question highlights the capacity of customized machine learning methods to enhance the accuracy of recognizing speech in individuals with aphasia [19].

Zha et al. introduced BERTRL (BERT-based Relational Learning) as a method for completing knowledge graphs. The method was shown to perform very well in both inductive and transductive scenarios. This work highlights the capabilities of pre-trained language models such as BERT in effectively dealing with intricate relational tasks [20]. Such models could be modified and utilized for predicting and analyzing aphasic speech.

Chen et al. developed SST-BERT (Structured Sentences with Time-enhanced BERT) to predict temporal relations in knowledge graphs by integrating time-specific improvements. The success of this model in predicting temporal tasks indicates possibilities of using it to boost temporal comprehension in circumstances featuring aphasic speech [20].

Garrido-Merchan et al. conducted a comparison between BERT and standard machine learning methods for text classification, showcasing BERT's superiority and adaptability. This empirical evidence validates the utilization of BERT as an established technique in natural language processing (NLP) tasks, including those that relate to aphasic communication [21].

Islam et al. investigated the application of BioGPT for automatically completing Chief Complaints in Electronic Health Records. The study discovered a direct relationship between the magnitude of a corpus and the perplexity score of a Language Model. Increased corpus size generally leads to higher scores, indicating enhanced performance. Deep learning models typically experience advantages when exposed to larger amounts of training data [22]. This research provides evidence for the effectiveness of utilizing extensive language models such as BERT in applications that involve speech impairments caused by aphasia.

These works demonstrate the progress made in using the AphasiaBank database and several machine learning approaches to enhance speech detection and evaluation for individuals with aphasia. Nevertheless, there are still

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3443592

deficiencies, namely in the incorporation of sophisticated NLP models such as BERT for tasks like text prediction and question answering [23].

Our research focuses on filling these gaps by utilizing BERT, an advanced transformer model, to predict patterns in aphasic speech and produce coherent sentences with incomplete information. Our study focuses on improving communication aids for aphasia patients by strengthening text prediction and question answering capabilities. This seeks to provide more accurate and reliable support to caregivers and speech therapists, eventually improving the capacity of aphasia patients to speak successfully.

III. DATASET

AphasiaBank is an open-access repository that has video and audio recordings of individuals with aphasia participating in multiple language-related activities. This resource, created by experts in the field of aphasiology, is specifically developed for the study of language problems related to aphasia. All the recordings are accompanied by comprehensive transcripts and commentaries, which provide detailed information into the linguistic difficulties experienced by individuals with aphasia. This dataset is heterogeneous, comprising people with different types and degrees of aphasia, and consists of speakers of numerous languages (Currently 12).

The recordings encompass many language challenges: including picture descriptions, narrative retellings, and conversational encounters. Researchers can investigate a broad range of aphasic speech patterns becasue of this variability. AphasiaBank is regularly updated with fresh recordings and annotations, rendering it an indispensable resource for academics and clinicians who are specifically interested in the development and evaluation of interventions [24].

A. Explanation of the dataset

The AphasiaBank transcripts include linguistic annotations that offer comprehensive details about the structure and meaning of each line of dialogue. As an illustration:

1. *Dialogue Line*: *INV: I'm going to be asking you to do some talking.*

2. *Morphological Annotation (%mor)*: The Morphological Annotation (%mor) is a process that breaks out the morphological components of the utterance. The program provides extensive tags for each grammatical category, encompassing subject pronouns, verb tenses, and infinitive markers. As an example:

The abbreviation "pro:sub|I" represents the subject pronoun 'I'. The notation "aux|be&1S" represents the first person singular present tense of the verb 'be'. The term "part|go-PRESP" denotes the progressive tense of the verb 'go'. The term "inf|to" represents the infinitive marker 'to'. The term "part|ask-PRESP" denotes the progressive tense of the verb 'ask'.

3. *Syntactic Annotation (%gra)*: This annotation represents the syntactic organization of the phrase using a format derived from dependency grammar. Each word is assigned a numerical value, and the relationships between words are depicted using arrows. For example: The notation "1|3|SUBJ" indicates that the subject "I" is connected to the verb "going". The notation "2|3|AUX" indicates that the auxiliary verb "be" is connected to the verb "going". The notation "6|3|COMP" signifies that the complementizer "to" is connected to the verb "going". The notation "7|6|OBJ" indicates that the object "talking" is connected to the complementizer "to".

These annotations provide a visual and detailed representation of the grammatical structure of the sentences, aiding in the analysis and understanding of aphasic speech [25].

B. Linguistic Features of Aphasia

Aphasia presents itself in diverse manners, with distinct linguistic characteristics frequently seen in the speech of individuals with aphasia. Anomia is a condition marked by difficulties in finding and recalling words, which can be observed through symptoms including extended pauses, incomplete words, filler words, incomplete sentences, and displays of irritation. Circumlocution is the use of indirect and roundabout words to explain a phrase or concept when the precise term cannot be recalled. Conduit d'approche refers to a series of tries made to pronounce a target word, with each attempt becoming closer to the correct phonetic pronunciation. The ultimate outcome of the project is uncertain.

A compilation of filler words was created by meticulously analyzing the .cha files and the AphasiaBank transcripts, which are specifically annotated to capture the subtle linguistic features of aphasic speech. The detected repeated non-lexical elements, such as "gram," "nk," "pn," "jar," "mm," "exc," "xxx," "uh," "um," "er," "uhm," "ah," and "umh," are filler words that disturb the fluidity of speech. By filtering these terms, it becomes possible to concentrate on the fundamental substance and structure of the aphasic speech, which is key for precise analysis and modeling.

Jargon is a coherent and phonetically accurate result that imitates the structure and intonation of English, but consists mostly of speech that lacks significant content. Occasionally, it can be comprehensible and able to be written down, but frequently it is incomprehensible. Neologism refers to the act of inventing new words to replace existing ones, typically with less than 50% similarity in sounds between the newly created term and the original word. The target term may be familiar or unfamiliar.

Perseveration refers to the act of repeating a word or phrase that was used before but is no longer suitable for the current situation. Phonemic paraphasia refers to the replacement, addition, removal, or rearrangement of speech sounds, with a minimum of 50% similarity between the incorrect pronunciation and the intended word. The error has the potential to be self-corrected, but it is not guaranteed. Semantic paraphasia refers to the act of replacing a specific term with another word, whether or not they are connected in

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3443592

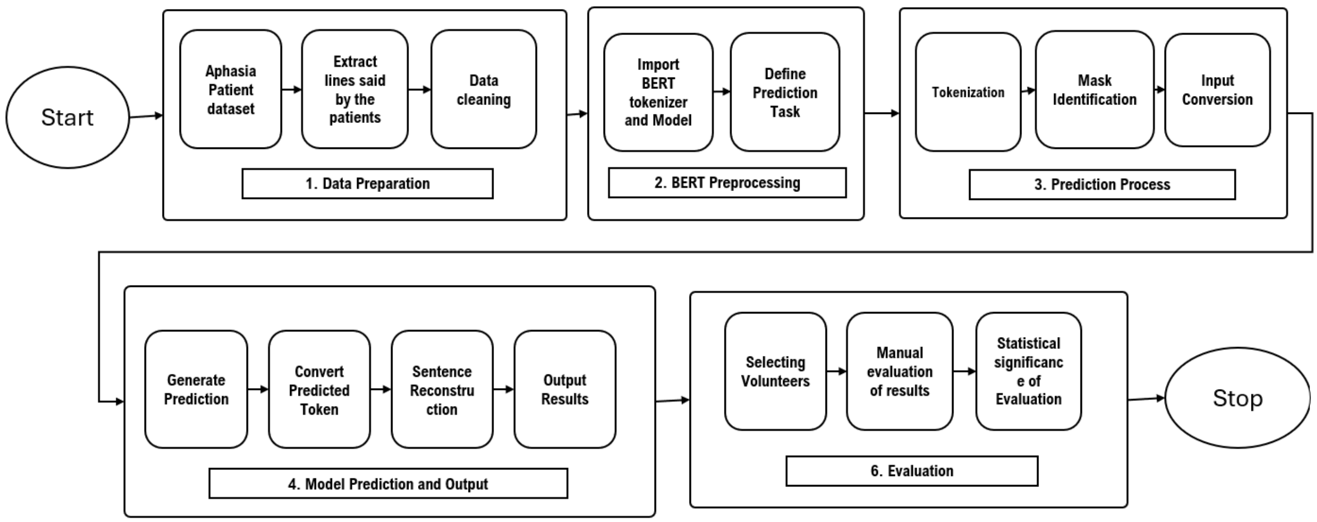


Figure 1. **Block diagram of Methodology.**

meaning. The error has the potential to be self-corrected, but it is not guaranteed.

Stereotypy refers to the regular and repetitive occurrence of a syllable, word, or phrase within a given sample, which can consist of either real words or made-up words. The understanding of the many expressions of aphasia relies heavily on these linguistic characteristics, which in turn aid in making precise diagnoses and customized therapeutic approaches. Through the examination of these characteristics, experts and medical professionals can get useful knowledge about the language deficiencies linked to aphasia and create improved tools for communication assistance and therapeutic approaches [26].

IV. METHODOLOGY

A. Data Collection and Preprocessing

The AphasiaBank database is a public resource that comprises a large collection of video and audio recordings of people with aphasia completing a variety of linguistic tasks. This dataset is significant for aphasiology research, as it contains thorough transcripts and annotations that shed light on the language challenges that aphasia sufferers have. The dataset comprises recordings of people with various forms and severity levels of aphasia, as well as speakers from multiple languages. Several language skills are covered, including picture description, story retelling, and conversational exchanges. The dataset is routinely updated with fresh recordings and annotations, making it a dynamic resource for researchers and clinicians working to create and assess aphasia therapies.

To prepare the data for analysis, we created a complete preprocessing pipeline that cleaned and organized transcripts from the AphasiaBank database. Several measures were taken during preprocessing to guarantee that the text data was ready for analysis and modeling.

Initially, we concentrated on utterances spoken by patients, denoted by the prefix "PAR" in the transcripts. This decision was made to guarantee that the data utilized for analysis accurately reflected the speech patterns and difficulties experienced by people with aphasia. To ensure the text's readability, we deleted non-alphanumeric characters while preserving spaces. Specific words and fillers, such as "gram," "nk," "pn," "jar," "mm," "exc," "xxx," "uh," "um," "er," "uhm," "ah," and "umh," were deleted to focus on the speech's essential content.

Next, we removed motions and particular gesture phrases from the text, such as "nods," "laughs," "sighs," "clapschest," "patschest," "raisesarm," "pointshigh," and "pointspicture," This phase was critical to ensuring that the transcriptions accurately represented the spoken words without interference from nonverbal cues. We also eliminated words beginning with &=ges: until a space appeared to eliminate any residual gesture-related annotations.

To improve the text, we eliminated all single letters except "I" and "a," which are legitimate words in English. To improve the integrity of the text formatting, we converted whitespace to single spaces and eliminated any trailing numbers that may have been included in the transcriptions. Only sentences longer than five words were chosen for further analysis, ensuring that the text data utilized for modeling retained enough context to be meaningfully analyzed.

B. Feature Extraction and Model Training

When it comes to feature extraction, we initially explored methods like Bag-of-Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), and Word Embeddings. Nevertheless, we ultimately utilized BERT (Bidirectional Encoder Representations from Transformers) because of its exceptional capability to comprehend contextual connections within text. The BERT model underwent fine-tuning to enhance its comprehension and prediction of aphasic speech patterns.

The optimization technique employed AdamW, an improved

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3443592

variant of the basic Adam optimizer, specifically engineered to successfully manage weight decay regularization [27]. The training approach consisted of a thorough loop that covered three whole epochs, carefully iterating through the entire dataset in each epoch. Prior to processing each batch, we employed uniform gradient resets to guarantee compatibility with the computer device.

The aphasia paragraphs were used to extract fragments of text, which were then processed individually using the model. During the training process, the main emphasis was placed on computing the loss at masked spots, hence guaranteeing the model's precision in predicting words that were omitted. The process of backpropagation was introduced to compute gradients based on the computed loss. This is followed by an optimization step to improve the model parameters, resulting in the minimization of the loss.

After establishing a solid strategy for collecting and preparing data, as well as training the model, we then assessed the usefulness of our method by conducting a series of experiments and analyses.

V. EXPERIMENTS AND RESULTS

A. Initial Sentence Prediction Using Masked Language Modeling

In order to predict the missing words in the sentences of individuals with aphasia, we employed the BERT tokenizer and a pre-trained BERT model. The technique entailed a series of intricate and specific processes. The text was initially tokenized using the BERT tokenizer, which transforms the content into a format that is appropriate for the model. Every sentence was defined and tokenized to determine the position of the masked token ([MASK]) [28]. The tokenized sentences were subsequently translated into input IDs, which were then converted into tensors for input into the BERT model. Utilizing the pre-trained BERT model, predictions were generated to extract the index of the masked token. The model's output was decrypted using the tokenizer to restore it to its original form. Subsequently, the predicted token was used to replace the masked token in the original sentence, so completing the sentence.

In Table 1, the wording "and she [MASK] supposed to be somewhere else" was revised as "and she was supposed to be somewhere else." This study successfully showcased the capabilities of BERT in accurately anticipating and filling in phrases using the contextual information derived from the speech of individuals with aphasia.

Table 1 shows snippets of the original Aphasia Dataset and the pre-processed and cleaned dataset.

TABLE I
SNIPPETS OF THE ORIGINAL AND INITIAL PRE-PROCESSED APHASIA TRANSCRIBED DATA

Original Dataset	Pre-processed Dataset
and she &=claps:chest rose [: supposed] [* s:ur] to be somewhere else . [+ gram] 652111_657621	and she [MASK] supposed to be somewhere else
&-uh well ‡ Cinderella was a poor child in &-uh [/] in &-uh &-uh +... 503723_512113	[MASK] well Cinderella was a poor child in [MASK] in [MASK] [MASK]
and &-um the mice (.) &-um give her the key . 601012_606482	and [MASK] the mice [MASK] give her the key
&=points:picture and the man is getting &-uh &-uh a &-uh (.) fireman to come get him . 312284_322024	and the man is getting [MASK] [MASK] a [MASK] fireman to come get him
well ‡ she's <going (.) have a> [//] &-uh take [: give] [*] it him [//] to him . 1592720_1598670	well shes going have a [MASK] take give it him to him

TABLE II
RESULT OF USING BERT FOR SENTENCE PREDICTION OF APHASIA PATIENTS

Original Sentence	and she [MASK] supposed to be somewhere else
Completed Sentence	and she was supposed to be somewhere else

B. Fine-Tuning BERT Model

To enhance the model's performance, we fine-tuned the pre-trained BERT model on the aphasia-specific dataset. The fine-tuning technique comprised multiple sequential steps. We chose the AdamW optimizer, which is an enhanced iteration of the Adam optimizer, specifically engineered to efficiently manage weight decay regularization. The training loop consisted of three full epochs, with diligent iteration across the entire dataset in each epoch. Gradients were reset prior to processing each batch to guarantee uniformity.

The aphasia paragraphs were used to extract parts of text, which were then processed individually by the model. The primary emphasis was placed on computing the loss at concealed spots to guarantee the model's precision in forecasting omitted words. The backpropagation algorithm was introduced to compute gradients based on the computed loss, which is then followed by an optimization step to fine-tune the model parameters and minimize the loss.

The extensive training and refinement process enabled the BERT model to get a deeper understanding of the subtle intricacies of aphasic speech, hence enhancing its capacity to accurately anticipate and fill in phrases in a contextually suitable manner. By prioritizing these meticulous procedures, we guaranteed the resilience and precision of the model, rendering it a powerful instrument for comprehending and forecasting aphasic speech patterns.

C. Sentence Vs Paragraph Level Training

To attempt to discover the most effective method for dealing with aphasic speech patterns, we conducted an investigation on the benefits of both sentence-level and paragraph-level training.

Sentence-level training entails analyzing individual sentences as the primary unit of analysis. This method offers a concentrated context, which can result in enhanced comprehension within sentences and increased computing efficiency. Nevertheless, it is possible that the model may not

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3443592

fully consider the wider context and is more likely to make errors when dealing with short or varied sentences. Paragraph-level training, on the other hand, involves analyzing full paragraphs to improve understanding of context between phrases. This method enhances the ability to handle unexpected situations and apply knowledge to new scenarios, but it requires more computer power and may unintentionally incorporate irrelevant information.

Due to the fragmented and context-dependent characteristics of aphasic language patterns, we determined that it would be advantageous to begin with sentence-level instruction during the earliest phases. This methodology enabled the model to concentrate on well defined and succinct situations. As the model progressed, we systematically integrated paragraph-level training to strengthen its capacity to comprehend and predict in wider contexts, ultimately resulting in a more resilient and adaptable model. An example of paragraph level results using from UCLASS is given in the table below [29].

TABLE III
PARAGRAPH LEVEL RESULTS FOR PREDICTION USING BERT

Original text	MAA PLI, I got my playstation in the [MASK] holidays, yeah AA AND W when you open it, IT TAKES CD IT TAKES CD games. And I've got one game which is a karate game. And it's got twenty [MASK] in it and they give you ten. And, you have to get the other twenty, and um you have to get the other ten. And I've got nineteen of them so I need one more, but he's really hard to get. SA and DI is got, um, it's got lots, I think it's got five different stages, THERE'S THERE'S time attack, there's [MASK], there's practice mode, there's um freeze mode, there's um option mode and there's arcade mode, there's six. And um, I'm looking forward to this game in about a month, and it's called Batman and Robin. Um, I think it's a two player game. And um and, um, I'm looking forward to that and after that I'm getting A A different game, it's called ASHOAW ASOAWGX and I it's, you call it, IT'S A K IT'S A [MASK] game I think it's two player. Yeah, when I get it yeah I'm going to be playing with it a lot. Uh and M my uncle he got a Nintendo 64 which, and he's got four games. He's got Super Mario Land 64, he's got Star Wars, he's got Lilac Wars and he's got, I can't riMEM, he's got BRAAraid Racer.
Completed text	ma ##a pl ##i , i got my playstation in the **CHRISTMAS** holidays , yeah aa and w when you open it , it takes cd it takes cd games . and i ' ve got one game which is a karate game . and it ' s got twenty **PEOPLE** in it and they give you ten . and , you have to get the other twenty , and um you have to get the other ten . and i ' ve got nineteen of them so i need one more , but he ' s really hard to get . sa and di is got , um , it ' s got lots , i think it ' s got five different stages , there ' s there ' s time attack , there ' s **MODE** , there ' s practice mode , there ' s um freeze mode , there ' s um option mode and there ' s arcade mode , there ' s six . and um , i ' m looking forward to this game in about a month , and it ' s called batman and robin . um , i think it ' s a two player game . and um and , um , i ' m looking forward to that and after that i ' m getting a a different game , it ' s called ash ##oa ##w as ##oa ##wg ##x and i it ' s , you call it , it ' s a k it ' s a **GOOD** game i think it ' s two player . yeah , when i get it yeah i ' m going to be playing with it a lot . uh and m my uncle he got a nintendo 64 which , and he ' s got four games . he ' s got super mario land 64 , he ' s got star wars , he ' s got lila ##c wars and he ' s got , i can ' t rim ##em , he ' s got bra ##ara ##id racer .

D. BERT Integration for Question-Answering

To further improve our model's capabilities, we integrated BERT for question-answering tasks. The purpose of this integration was to enhance the current masked token prediction by improving the comprehension of context, which is especially advantageous for dealing with the complex linguistic patterns in the AphasiaBank dataset.

The integration started with substantial modifications to the preprocessing phase in order to accommodate both a query and its surrounding context. To achieve this, it was necessary to combine the question and context into a single input sequence. Additionally, the tokenization procedure was adjusted to ensure that it is compatible with BERT. The model inputs were meticulously designed to accommodate this merged sequence, enabling BERT to effectively analyze the complete context and produce coherent replies. The output method was modified to accurately anticipate the beginning and ending positions of the answers inside the context, guaranteeing precise localization of the pertinent information.

The process of incorporating BERT into our question-answering system consisted of multiple essential stages. Initially, we instantiated the BERT tokenizer and model by loading a pre-trained BERT model that has been specifically fine-tuned for question-answering tasks, such as `bert-large-uncased-whole-word-masking-finetuned-squad`.

Subsequently, we employed the BERT tokenizer to tokenize both the input question and context. This stage involved managing lengthy sequences in which, if the total number of tokens beyond the maximum limit of 512 tokens for BERT, the context was either shortened or divided to ensure it fits within this limit. The tokenized sequence was modified by adding special tokens `[CLS]` and `[SEP]` at the start and end, respectively, to indicate the boundaries of the question and context. Next, the tokens were translated into input IDs and then transformed into tensors that are compliant with the requirements of the model.

After preparing the input tensors, they were inputted into the BERT model to generate predictions. The model produces two sets of logits: `start_logits` and `end_logits`. These logits indicate the model's level of certainty on whether each token is the beginning or end of the answer. The bounds of the projected response were determined by identifying the positions with the highest start and finish logits. The tokenizer was used to convert these token places back to their original string form. Ultimately, the function synthesized the solution

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3443592

by merging the recognized tokens and supplied it as the ultimate result.

Data preparation for question-answering entailed manipulating a multiline string that was allocated to a variable called `data`. The data was trimmed of any leading or trailing whitespace and divided into separate lines. The script systematically examined these lines to discover pairs of lines in which the first line concluded with a question mark or the word "do." Every pair that was detected, which includes a question and its subsequent line (context), was appended to a list called `qas_pairs` as a dictionary with the keys "question" and "context."

The code imports the `BertTokenizer` and `BertForQuestionAnswering` modules from the transformers library. This library offers a PyTorch interface for interacting with pre-trained BERT models. The specified BERT model fine-tuned for question-answering tasks and its related tokenizer were successfully loaded. An alert was displayed during this procedure indicating that not all weights from the checkpoint were utilized. This outcome was anticipated as the model was employed for question answering rather than for pretraining tasks such as masked language modeling. The `ask_bert` function utilized the BERT model to extract the response by taking a question and a context as inputs. The input query and context were tokenized, processed via the model, and the start and end logits were retrieved. Subsequently, the function determined the tokens with the most elevated start and end scores, transformed these tokens into a string, and ultimately provided the resulting response.

A set of question-context pairings, named `example_qas_pairs`, was constructed. The script then iterated over these pairs and utilized the `ask_bert` function to retrieve the response. The outcomes were subsequently printed, displaying the posed question, the given context, and the answer ascertained by the BERT model.

The incorporation of BERT into question-answering yielded encouraging outcomes. For example, when inquired about the timing of the individual's stroke, Given the context "do you remember when you had your stroke yes j tracestwo its two thousand traceszero traceszero pointstable two days", the model accurately responded with "two days". Another instance entailed the inquiry "What are the person's sentiments towards China?" within the context "it appears that you had a profound affection for China, yes indeed," and the model provided an accurate response of "greatly adored." The results indicate that the BERT model has improved ability to comprehend and produce contextually suitable responses, greatly enhancing the interaction with and analysis of the AphasiaBank dataset. This integration not only enhanced the current masked token prediction but also established a strong foundation for question-answering, enabling a more profound understanding of the language patterns of individuals with aphasia.

TABLE IV
Q&A RESULTS USING BERT

Question	Context	Answer
When did the person have their stroke?	do you remember when you had your stroke yes j tracestwo its two thousand traceszero traceszero pointstable two days	two days
How does the person feel about China?	sounds like you really loved China y oh yes	really loved
What did you receive for Christmas?	For Christmas I got that bear AND AND also got a small cat bear that's um black and.	bear
What happened to Cinderella at midnight?	dancing around midnight and she clapschest rose supposed to be somewhere else	dancing

E. Evaluation

The evaluation of our BERT-based model encompassed various factors, such as contextual suitability, grammatical accuracy, and the fluency of the generated sentences. We employed a rating scale ranging from 1 to 4, where a rating of 1 denoted forecasts that were completely unsuitable and a rating of 4 denoted predictions that were highly suitable.

In order to ensure a robust evaluation, we performed a manual assessment with four evaluators who were proficient in English and had a strong understanding of grammar and linguistic nuances, qualifying them to accurately evaluate the model's outputs. The raters assessed the model's predictions according to the provided criteria. This manual rating process allowed us to verify the model's performance against expert human judgment, providing a comprehensive assessment of its capabilities.

The inter-rater reliability was assessed using Fleiss' Kappa score [30]. The Fleiss' Kappa score of 0.32 indicates a moderate level of agreement among the raters, suggesting a certain degree of consistency in their evaluations, although not very strong. A Fleiss' Kappa score ranging from 0.21 to 0.40 is commonly understood to suggest a reasonable level of agreement that goes above what would be expected by chance alone. The variation in human assessments stems from the subjective aspect of analyzing linguistic elements, such as contextual appropriateness and grammatical correctness.

In addition, we determined the intra-rater reliability by assessing the standard deviation of the ratings given by each rater. The standard deviations (Rater 1: 0.941357, Rater 2: 0.874423, Rater 3: 0.934797, Rater 4: 0.979796) suggest that the scores given by each rater were generally consistent, but there was some degree of variation. This analysis allowed us to evaluate the consistency of scores offered by individual raters.

To ensure the validity of our predictions, we calculated the NDCG (Normalized Discounted Cumulative Gain) scores for each rater in comparison to the ideal ranking. The NDCG scores varied between 0.954452 and 1.142482, indicating a

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3443592

strong agreement between the ratings given by the evaluators and the expected ideal output.

TABLE V
EVALUATION OF PREDICTED RESULTS

Data ID	Results	Rating 1 to 4 (additional comments if any)
AWCT01a	Original: and she &=claps:chest rose [: supposed] [* s:ur] to be somewhere else . [+ gram] 652111_657621 Processed: and she [MASK] supposed to be somewhere else Completed: and she was supposed to be somewhere else	4
AWCT01a	Original: &-uh well ‡ Cinderella was a poor child in &-uh [/] in &-uh &-uh +... 503723_512113 Processed: [MASK] well Cinderella was a poor child in [MASK] in [MASK] [MASK] Completed: oh well Cinderella was a poor child in poverty in england.	2 "poor" and "in poverty", while grammatically correct, seems redundant
AWCT02a	Original: and &-um the mice (.) &-um give her the key . 601012_606482 Processed: and [MASK] the mice [MASK] give her the key Completed: and then the mice will give her the key	3 "Will" seems to give an odd future sense here. I feel it would make more sense with no word where will is.
AWCT07a	Original: &=points:picture (..) the boy &-uh (..) was getting a &-uh umbrella (.) from his mother . 247996_262506 Processed: the boy [MASK] was getting a [MASK] umbrella from his mother Completed: the boy who was getting a new umbrella from his mother	3

```
Fleiss' Kappa Score: 0.31948790896159324
Intra-rater Reliability (Standard Deviation):
                 Standard Deviation
Rater 1               0.941357
Rater 2               0.874423
Rater 3               0.934797
Rater 4               0.979796
```

Figure 2. Fleiss' Kappa Score.

```
NDCG Scores for Each Rater
Compared to the Ideal Ranking:
     Rater   NDCG Score
0   Rater 1    1.142482
1   Rater 2    1.063404
2   Rater 3    0.984604
3   Rater 4    0.954452
```

Figure 3. NDCG Scores.

F. Algorithm

The algorithm for text cleaning:

Algorithm TextClean

Require: T: Raw text which may contain unwanted characters and filler words.
Ensure: T_{cleaned}: Text that has been cleaned of specified unwanted characters and filler words, with normalized whitespace.

$U \leftarrow \{\&, =, -, ,, ., :, ;, n, ', (,), <, >, \%, *, +, !, @, \#, \$, \hat{}, \&, \_\}$
$F \leftarrow$ {gram, nk, pn, jar, mm, exc, xxx, uh, um, er, uhm, ah, umh}
$T_{\text{modified}} \leftarrow \text{RemoveCharacters}(T, U)$ ▷ Remove all characters in U from T
$T_{\text{modified}} \leftarrow \text{RemoveFillerWords}(T_{\text{modified}}, F)$ ▷ Remove all words in F from T_{modified}
$T_{\text{cleaned}} \leftarrow \text{NormalizeWhitespace}(T_{\text{modified}})$ ▷ Convert multiple spaces to a single space and trim spaces at the ends
return T_{cleaned}

Figure 4. Algorithm for Text Cleaning.

The algorithm for text prediction using BERT:

Algorithm . WordPredict

tokenizer ← Initialize BERT tokenizer
model ← Load BERT model trained for masked language prediction
Cha ← Directory containing .cha files used for analysis
for each file in Cha **do**
  C ← Read content of the file
  if 'PAR' found in C **then**
    $TPAR$ ← Extract text content following the 'PAR' identifier
    $T_{\text{cleaned}} \leftarrow \text{TextClean}(TPAR)$ ▷ Clean text using Algorithm 1
    $Tm \leftarrow \text{MaskStumblesReplace}(T_{\text{cleaned}}, '[MASK]')$ ▷ Replace stutters and filler words with '[MASK]'
    if '[MASK]' exists in Tm **then**
      $Tt \leftarrow \text{Tokenize}(tokenizer, Tm)$ ▷ Convert Tm to a format suitable for the BERT model
      $Pt \leftarrow \text{Predict}(model, Tt)$ ▷ Predict masked words using the BERT model
      $T_{\text{complete}} \leftarrow \text{Replace}('[MASK]', Pt, Tm)$ ▷ Replace '[MASK]' with predicted words in Tm
    else
      $T_{\text{complete}} \leftarrow T_{\text{cleaned}}$
  return T_{complete}

Figure 5. Algorithm for text prediction using BERT.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3443592

The algorithm for question answering using BERT:

Algorithm BERT-based Question Answering

tokenizer ← Initialize BERT tokenizer with a large model fine-tuned for question answering
model ← Load BERT large model fine-tuned for question answering
function ASK_BERT(question, context)
  Input:
    question: Text of the question to be answered
    context: Text containing the information relevant to the question
  Output:
    answer: The text of the answer extracted from the context

  Tq ← Tokenize(question + " " + context) ▷ Combine and tokenize question and context
  IDs, masks ← Convert tokens to input IDs and attention masks
  start_logits, end_logits ← Use model to predict start and end logits
  start_index, end_index ← Determine the maximum logit indices for start and end
  answer_tokens ← Extract tokens from input IDs using start_index and end_index
  answer ← Decode answer tokens to string
  return answer

Figure 6. **Algorithm for question answering using BERT.**

G. Results

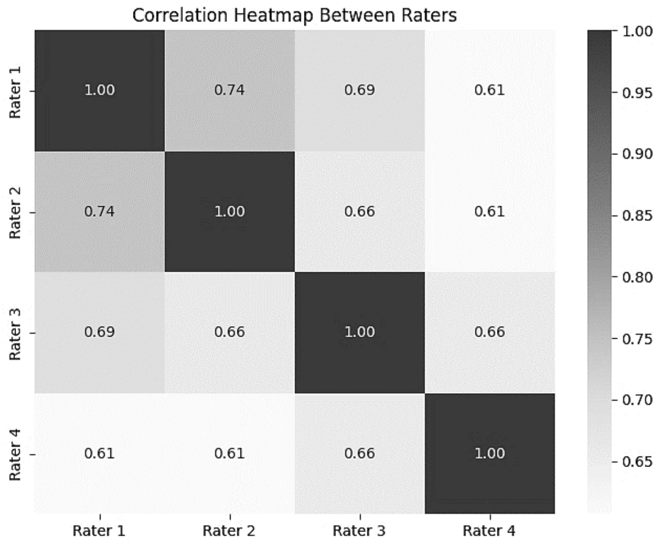


Figure 7. **Correlation Heatmap Between Raters.**

The preliminary results from the BERT modeling and sentence prediction showed promising improvements in contextual comprehension and predictive precision for text related to aphasia. By including question-answering capabilities, the model's performance in real-world situations was improved, resulting in more precise and contextually relevant predictions.

The correlation heatmap among raters displayed the pairwise correlations, indicating a substantial level of agreement among the raters. The minimum correlation observed was 0.61, while the maximum correlation was 0.74. The boxplots illustrating the rating distribution revealed that all raters exhibited a rather similar range of ratings, but with a few exceptional cases for Raters 1 and 2. This indicates that these raters occasionally diverged from their typical rating preferences.

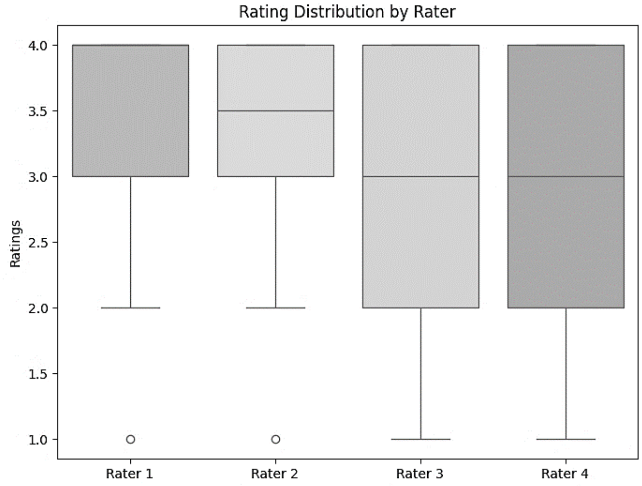


Figure 8. **Rating Distribution by Raters.**

The Fleiss' Kappa score demonstrated a moderate level of concordance among the raters, while the intra-rater reliability tests revealed that the raters' scores were quite stable. The NDCG scores for each rater [31], when compared to the ideal ranking, indicated that Rater 1 showed the highest level of alignment with the ideal ranking. This was followed by Raters 2, 3, and 4.

Finally, the findings demonstrated the efficacy of the BERT model in forecasting aphasic speech patterns. Furthermore, the incorporation of question-answering tasks provided further advantages for practical use cases.

VI. DISCUSSION

Although the utilization of a pre-trained BERT model on the AphasiaBank dataset has demonstrated efficacy in predicting certain outcomes and answering questions in the context of aphasic speech, we recognize that the limited scope of this dataset may restrict the generalizability of our results. The dataset's emphasis on aphasic speech improves our capacity to provide detailed analyses and interventions that are specifically tailored for aphasia therapy. Nevertheless, the distinct attributes of our dataset imply that our findings are most relevant to comparable clinical scenarios. In addition, the manual review procedure, although comprehensive, introduces a subjective factor that might potentially affect the outcomes. Future research should explore the use of other objective, automated evaluation measures in conjunction with evaluation by humans to strengthen the reliability of the findings. In the future, our research will continue to study and enhance these uses, with a specific emphasis on utilizing knowledge from the AphasiaBank dataset to further improve treatment methods for individuals with aphasia. The objective of this focused approach is to enhance comprehension of aphasic speech patterns and enhance intervention techniques, so offering significant advantages to individuals with aphasia while contributing to specialized improvements in speech pathology.

VII. CONCLUSION

While the research approach appears promising at first glance, there is still considerable space for improvement when it comes to data pretreatment and data cleaning. Quantitative measures like accuracy ratings or qualitative comments from people with aphasia or language experts could be incorporated to fully evaluate the outcomes. The importance of this initiative lies in its contribution to the knowledge and management of aphasia. This project uses the sentence prediction , question answering methods and the AphasiaBank database to give therapists and researchers valuable insights into the language issues faced by aphasics. Accurate sentence prediction enhances the quality of life for aphasics by providing them with options for language treatment and efficient communication. In our study, we preprocessed the AphasiaBank transcripts by removing nonverbal cues and redundant sentences, resulting in an enhanced ability of the model to focus on crucial content. We employed BERT to predict masked tokens and perform question-answering tasks, leading to substantial improvements in contextual understanding and predictive accuracy.The assessment of the BERT model yielded a Fleiss' Kappa value of 0.32, indicating a modest degree of consensus among evaluators. The NDCG scores for each rater ranged from 0.954452 to 1.142482, reflecting a high level of agreement between the model's predictions and the ideal outputs.Future research will prioritize expanding the dataset to include a broader range of aphasia types. Additionally, exploring other advanced NLP models will be conducted to enhance the reliability and precision of predictions. This integrated strategy will greatly enhance the treatment and rehabilitation of aphasia and promote deeper comprehension of aphasic speech.

REFERENCES

[1] "What Is Aphasia? — Types, Causes and Treatment," *NIDCD*, Mar. 06, 2017. Available: https://www.nidcd.nih.gov/health/aphasia

[2] C. C. M. Professional, "Aphasia," *Cleveland Clinic*. Available: https://my.clevelandclinic.org/health/diseases/5502-aphasia

[3] "AphasiaBank." Available: https://aphasia.talkbank.org/

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv (Cornell University)*, Jan. 2018, doi: 10.48550/arxiv.1810.04805. Available: https://arxiv.org/abs/1810.04805

[5] H. M. D. P. M. Herath, W. A. S. A. Weraniyagoda, R. T. M. Rajapaksha, P. A. D. S. N. Wijesekara, K. L. K. Sudheera, and P. H. J. Chong, "Automatic Assessment of Aphasic Speech Sensed by Audio Sensors for Classification into Aphasia Severity Levels to Recommend Speech Therapies," *Sensors*, vol. 22, no. 18, p. 6966, Sep. 2022, doi: 10.3390/s22186966. Available: https://doi.org/10.3390/s22186966

[6] D. Beltrami, G. Gagliardi, R. R. Favretti, E. Ghidoni, F. Tamburini, and L. Calzà, "Speech Analysis by Natural Language Processing Techniques: A Possible Tool for Very Early Detection of Cognitive Decline?," *Frontiers in Aging Neuroscience*, vol. 10, Nov. 2018, doi: 10.3389/fnagi.2018.00369. Available: https://doi.org/10.3389/fnagi.2018.00369

[7] D. Le and E. M. Provost, "Improving Automatic Recognition of Aphasic Speech with AphasiaBank," Sep. 2016, doi: 10.21437/interspeech.2016-213. Available: https://doi.org/10.21437/interspeech.2016-213

[8] K. R. Jothi, S. S. Sivaraju and P. J. Yawalkar, "AI based Speech Language Therapy using Speech Quality Parameters for Aphasia Person: A Comprehensive Review," *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2020, pp. 1263-1271,doi: 10.1109/ICECA49313.2020.9297591. Available: https://ieeexplore.ieee.org/document/9297591

[9] Y. Qin, T. Lee, and A. P. H. Kong, "Automatic Speech Assessment for Aphasic Patients Based on Syllable-Level Embedding and Supra-Segmental Duration Features," Apr. 2018, doi: 10.1109/icassp.2018.8461289. Available: https://doi.org/10.1109/icassp.2018.8461289

[10] D. Fromm and B. MacWhinney, "AphasiaBank as BigData," *Seminars in Speech and Language*, vol. 37, no. 01, pp. 010–022, Feb. 2016, doi: 10.1055/s-0036-1571357. Available: https://pubmed.ncbi.nlm.nih.gov/26882361/

[11] S. Clough and J. K. Gordon, "Fluent or nonfluent? Part A. Underlying contributors to categorical classifications of fluency in aphasia," *Aphasiology*, vol. 34, no. 5, pp. 515–539, Feb. 2020, doi: 10.1080/02687038.2020.1727709. Available: https://doi.org/10.1080/02687038.2020.1727709

[12] H. Park and J. Park, "Assessment of Word-Level Neural Language Models for Sentence Completion," *Applied Sciences*, vol. 10, no. 4, p. 1340, Feb. 2020, doi: 10.3390/app10041340. Available: https://www.mdpi.com/2076-3417/10/4/1340

[13] K. R. J. P. Y. Mamatha V. L., "Automatic Speech Assessment System for Aphasia Speech Disorder," May 27, 2021. Available: http://annalsofrscb.ro/index.php/journal/article/view/6425

[14] D. Le and E. M. Provost, "Improving Automatic Recognition of Aphasic Speech with AphasiaBank," Sep. 2016, doi: 10.21437/interspeech.2016-213. Available: https://www.isca-archive.org/interspeech_2016/le16b_interspeech.html

[15] S. -H. Chiu and B. Chen, "Innovative Bert-Based Reranking Language Models for Speech Recognition," *2021 IEEE Spoken Language Technology Workshop (SLT)*, Shenzhen, China, 2021, pp. 266-271, doi: 10.1109/SLT48900.2021.9383557.

[16] Y.-S. Chuang, C.-L. Liu, H.-Y. Lee, and L.-S. Lee, "SpeechBERT: An Audio-and-text Jointly Learned Language Model for End-to-end Spoken Question Answering," *arXiv.org*, Oct. 25, 2019. Available: https://arxiv.org/abs/1910.11559

[17] C.-C. Kuo, S.-B. Luo, and K.-Y. Chen, "An Audio-enriched BERT-based Framework for Spoken Multiple-choice Question Answering," *arXiv.org*, May 25, 2020. Available: https://arxiv.org/abs/2005.12142

[18] I. G. Torre, M. Romero, and A. Álvarez, "Improving Aphasic Speech Recognition by Using Novel Semi-Supervised Learning Methods on AphasiaBank for English and Spanish," *Applied Sciences*, vol. 11, no. 19, p. 8872, Sep. 2021, doi: 10.3390/app11198872. Available: https://doi.org/10.3390/app11198872

[19] S. S. Mahmoud, R. F. Pallaud, A. Kumar, S. Faisal, Y. Wang, and Q. Fang, "A Comparative Investigation of Automatic Speech Recognition Platforms for Aphasia Assessment Batteries," *Sensors*, vol. 23, no. 2, p. 857, Jan. 2023, doi: 10.3390/s23020857. Available: https://doi.org/10.3390/s23020857

[20] H. Zha, Z. Chen, and X. Yan, "Inductive Relation Prediction by BERT," *Proceedings of the ... AAAI Conference on Artificial Intelligence*, vol. 36, no. 5, pp. 5923–5931, Jun. 2022, doi: 10.1609/aaai.v36i5.20537. Available: https://doi.org/10.1609/aaai.v36i5.20537

[21] E. C. Garrido-Merchan, R. Gozalo-Brizuela, and S. Gonzalez-Carvajal, "Comparing BERT against Traditional Machine Learning Models in Text Classification," *Journal of Computational and Cognitive Engineering*, Apr. 2023, doi: 10.47852/bonviewjcce3202838. Available: https://doi.org/10.47852/bonviewjcce3202838

[22] K. M. S. Islam, A. S. Nipu, P. Madiraju, and P. Deshpande, "Autocompletion of Chief Complaints in the Electronic Health Records using Large Language Models," Dec. 2023, doi: 10.1109/bigdata59044.2023.10386778. Available: https://doi.org/10.1109/bigdata59044.2023.10386778

[23] I. G. Torre, M. Romero, and A. Álvarez, "Improving Aphasic Speech Recognition by Using Novel Semi-Supervised Learning Methods on AphasiaBank for English and Spanish," *Applied Sciences*, vol. 11, no. 19, p. 8872, Sep. 2021, doi: 10.3390/app11198872. Available: https://www.mdpi.com/2076-3417/11/19/8872

[24] "AphasiaBank Browsable Database." Available: https://sla.talkbank.org/TBB/aphasia. [Accessed: May 28, 2024]

[25] "Using CLAN." Available: https://dali.talkbank.org/clan/

[26] "Patterns of Connected Speech Features in Aphasia." Available: https://aphasia.talkbank.org/publications/2017/Casilio17b.pdf. [Accessed: May 28, 2024]

[27]"AdamW." Available: https://huggingface.co/docs/bitsandbytes/main/en/reference/optim/adamw

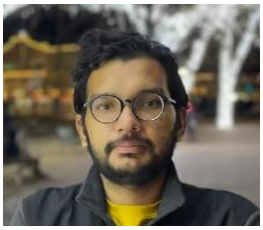
[28]"Masked language modeling." Available: https://huggingface.co/docs/transformers/main/en/tasks/masked_language_modeling

[29] "Release 1," *UCL 1999-2005*. Available: https://www.uclass.psychol.ucl.ac.uk/

[30] "Interrater reliability: the kappa statistic," *PubMed*, 2012. Available: https://pubmed.ncbi.nlm.nih.gov/23092060/

[31] V. Erdem, "NDCG for Ranking Evaluation | Medium," *Medium*, Nov. 06, 2023. Available: https://medium.com/@hvarolerdem/ndcg-for-ranking-evaluation-fd8bf45179fc

SHAMIHA BINTA MANIR is a PhD student in Electrical and Computer Engineering at Marquette University. She studied electrical and electronic engineering for her undergraduate and master's degrees from the University of Dhaka, Bangladesh. Currently, Shamiha's research focuses on creating predictive models for health conditions such as breast cancer and developing BERT-based text prediction tools for aphasia patients. Her contributions to the academic community are highlighted through her published works on topics such as student success in higher education and breast cancer risk analysis. In her free time, she loves to paint in watercolor.



K M SAJJADUL ISLAM is a PhD student in Computer Science at Marquette University, Wisconsin, USA. He completed his undergraduate studies at the Shahjalal University of Science and Technology in Sylhet, Bangladesh. Sajjad's research focuses on Data Analytics and Natural Language Processing (NLP), with a particular interest in developing healthcare solutions. In a previous research project, he employed large language models to develop an autocomplete tool for Electronic Health Records (EHR). Currently, he is working on an AI-enhanced Peer Support System (PSS) for Opioid Use Disorder (OUD).

PRAVEEN MADIRAJU received the Ph.D. degree in computer science from Georgia State University. He directs the Data Science and Text Analytics Laboratory. The Data Laboratory focuses on solving real-world problems by applying techniques from the broad area of data science and data analytics on both structured and unstructured data. The laboratory also conducts research on applying machine learning techniques to analyze textual and social media data. He is currently a Professor with the Department of Computer Science, Marquette University, Milwaukee, WI, USA. He is also the Graduate Program Chair of the Computer Science Program, Marquette University. He has published over 50 peer-reviewed articles and has organized workshops on middleware systems in conjunction with ACM SAC and IEEE COMPSAC. His research interests include data science, healthcare informatics, text analytics, and databases. He regularly serves on NSF panels.

PRIYA DESHPANDE is an Assistant Professor at Electrical and Computer Engineering department at Marquette University. Priya Deshpande completed her PhD in Computer Science from DePaul University, Chicago. Priya is a senior member of the Institute of Electrical and Electronics Engineers (IEEE). Her research interests encompass machine learning, natural language processing, databases, big data analytics, and computer-aided diagnosis. Priya's previous research addressed big data analytics and load balancing in cloud computing. Her current research focuses on biomedical data integration. She integrates public clinical data sources and medical ontologies. She incorporates databases, natural language processing, information retrieval, artificial intelligence, and machine learning techniques to help clinical experts to find supplemental reference information and improve diagnostic accuracy.