Personalized Speech Analysis for Wernicke's Aphasia Patients

Sreekar Baddepudi

Project Summary: Wernicke's aphasia (WA) is a neurological disorder arising from brain lesions, affecting comprehension and production of meaningful speech. fMRI is typically used to find the location of a lesion, but it is invasive and expensive, and does not give an indication of symptoms or optimal modes of therapy. Conversely, speech-language therapists assess and monitor the severity of language impairment in patients through live speech or recordings. However, manually analyzing patients' responses is a tedious process that relies heavily on the expertise of individual therapists. To address this issue, I have developed a personalized application which leverages generative AI technology, large language model (LLM) and custom prompts using retrieval-augmented generation (RAG) techniques to automate WA patients' speech analysis. The model provides consistent linguistic assessment and suggests a customized set of treatment options. As such, this application can provide automation, precision, and objectivity for patient assessments and the design of therapeutic regimens.

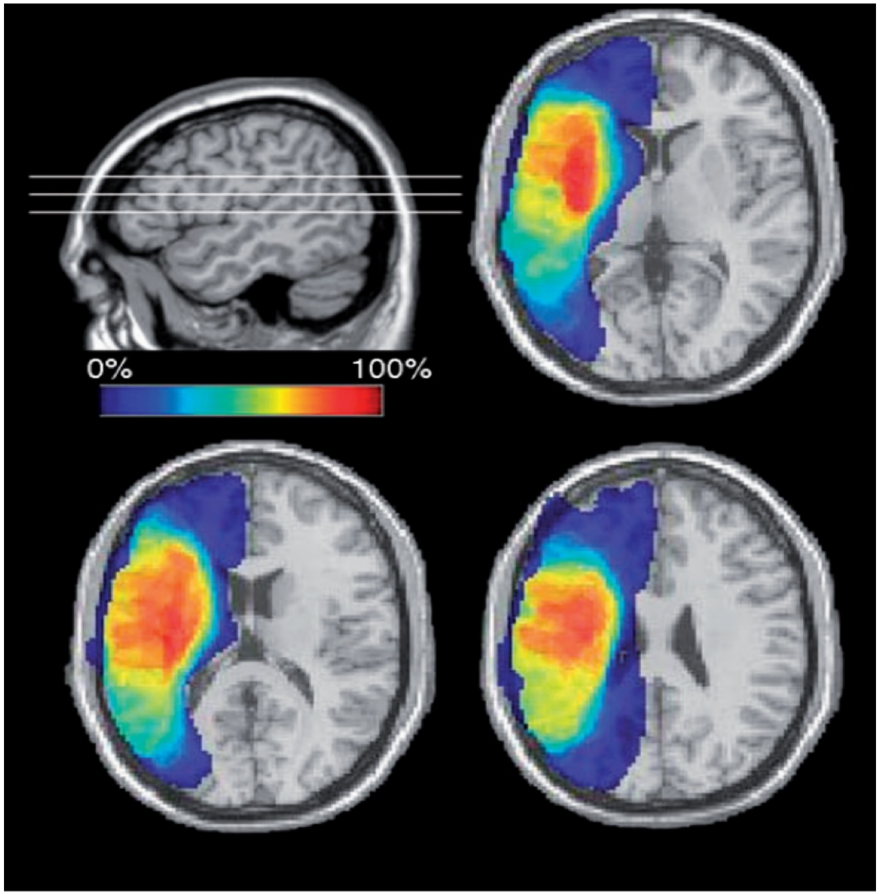

Fig. 1: fMRI scans of lesion locations in the brain for Wernicke's Aphasia Adelman & Smith, 2009

Introduction

Research Question: Can large language models (LLM) be used diagnose and treat Wernicke's aphasia via personalized syntactical, grammatical, and content-specific analyses?

Hypothesis: By analyzing language-dependent and universal symptoms associated with Wernicke's aphasia, the large language model can automate speech analysis of the individuals and suggest targeted set of customized treatments, eliminating the manual effort, potential bias, and inconsistency inherent in the process

Project Motivation: Last summer, I visited my grandmother in India for the first time since her recent stroke and learnt that she had Wernicke's aphasia. Around the world, fifteen million strokes occur annually, with up to 40% of stroke survivors diagnosed with aphasia [1]. The National Institute on Deafness and Other Communication Disorders states that approximately 2 million Americans have some aphasia and 16% of people with acute aphasia have Wernicke's aphasia (ASHA, 2023).

Wernicke's Aphasia: Wernicke's area translates auditory signals into comprehensible linguistic input. Wernicke's aphasia (WA) results from cerebral lesions to Wernicke's area in the left temporal lobe. Lesions vary in location, leading to diverse symptoms and affected neural networks. Individuals affected with WA can speak at normal speed and rhythm, but their verbal expressions are incomprehensible, lacking meaningful content and context.

Challenges: There is no cure for WA. Speech therapies can help improve the patient's language abilities, however

- There is a shortage of funding for aphasia intervention relative to the number of people living with aphasia (Adikari et al, 2023). The lack of available speech pathologists is detrimental to the patient's recovery and health
- Transcribing speech is manual, time-consuming, and subjective.
- Different therapists have different approaches to analysis, leading to inconsistencies in assessments

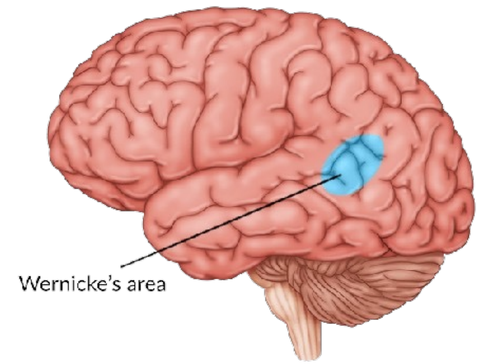

Fig. 2: Image of Wernicke's Area
Neuroscience Bite: Wernicke's Area, October 2022

Wernicke's Aphasia Example
Examiner: Hi, how are you doing today?
Patient: I'm happy, are you pretty.
Examiner: What are you doing today?
Patient: We stayed with the water today.

Fig3. https://www.ncbi.nlm.nih.gov/books/NBK559315/

Introduction

Generative artificial intelligence (AI) has been gaining popularity for its ability to generate new, relevant content in natural language by adapting to user's inputs. **Large language models (LLM)** are a type of generative AI model trained on vast corpus of text and can automate tasks that involve understanding and responding to language. LLMs and natural language processing (NLP) technologies have offered new avenues to study language and analyze the complex linguistic patterns in the speech of people with aphasia.

Work by others

- The majority of studies (69%) used supervised machine learning models for predictions and classifications, such as diagnosing aphasia subtypes and severity. The review concludes that the adoption of AI in aphasia management has been relatively slow. (Adikari et al, 2023).
- Two studies explored using LLM as assistive tool for people with aphasia: 1. Used LLM in spontaneous word retrieval during conversations. The LLM successfully identified the intended word in 11 out of 12 instances (92% accuracy). (Purohit et al, 2023), and 2. Leveraged LLMs to predict the intended target words for paraphasia (unintended word errors) made by people with aphasia. (Salem et al, 2023)

Limitations

- Utilizing AI and natural language processing (NLP) techniques, specifically large language models (LLMs), for automated speech analysis of aphasia patients is an area that has not been extensively explored or implemented in practice.
- Most of the studies related to LLM are recent (2023) and are just limited to just one symptom of aphasia, not allowing for the true scope of the disorder to be understood by the model. However, this is an emerging field and with the accessibility of LLM, we will positively see more studies focused on aphasia.

Benefits of this study

- Holistic as it focuses on comprehensive aphasia assessment and thereby suggestive treatment options
- Adaptive, personalized therapy requires less clinician input, addressing the availability challenges due to shortage speech-language pathologists

Previous work: None.

Methods

Data collection

- The samples in this study were carefully selected to ensure a representative sample of individuals with WA, obtained from **AphasiaBan**k database.
- All individuals studied here had received a clinical diagnosis of Wernicke's aphasia from medical professionals
- Each of these patients had performed the same protocol, which had been outlined and peer-reviewed by the Talkbank.
- A total of 13 transcripts of WA patients across **English, French, Japanese** and **Cantonese** languages were evaluated for this study.

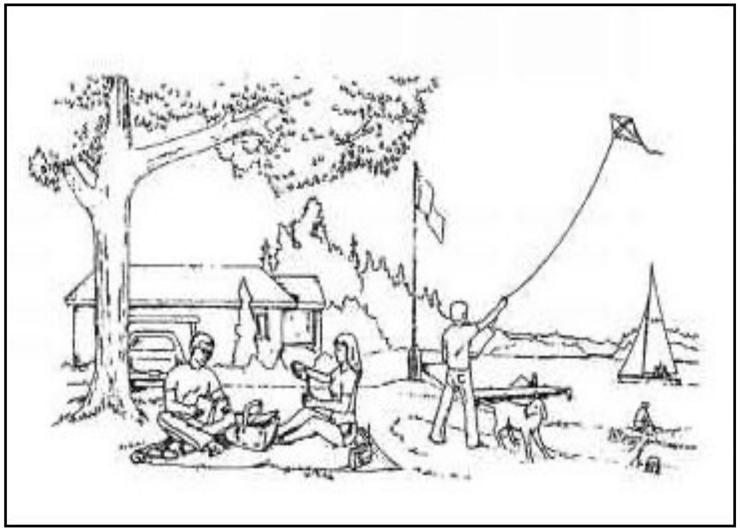

Fig. 3: This shows an image that would be used to test an aphasic patients' image recognition.
Western Aphasia Battery-Revised, July 1, 2019

Data analysis

- **Step 1:** Focused different ways of formatting prompts and configuring inferential hyper-parameters to get the LLM to provide the most relevant data possible
- Each of the WA patients' transcripts were analyzed individually using these prompts to compare the identified speech components and rules across languages.

  - **Anthropic's Claude 2.1,** an advanced large language model (LLM), has been used to analyze the unique and universal speech patterns in transcripts from AphasiaBank. Claude supports multiple languages allowing for easier analysis comparing speech characteristics across patients who speak various languages .

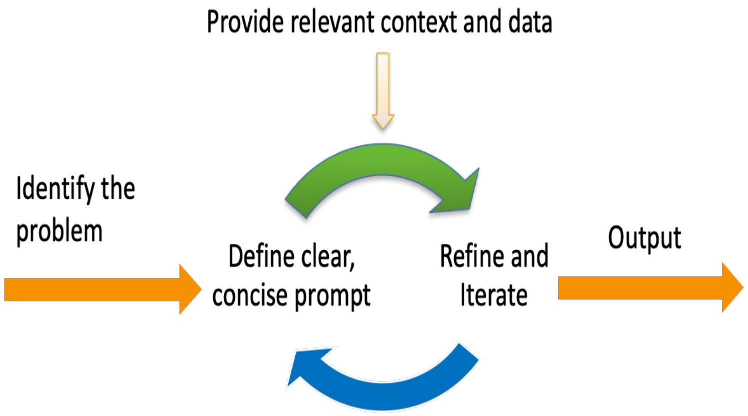

Fig. 4: Stages of Prompt Engineering
Chart by Sreekar Baddepudi

Methods

Data analysis

- **Step 2 :** Claude was prompted to emulate WAB-R assessments on the speech transcripts and the responses were compared to see if they aligned with the WAB-R diagnoses by the clinicians
    - **Western Aphasia Battery-Revised (WAB-R)** is the gold standard for assessing and diagnosing aphasia. By measuring performance across these linguistics areas, the WAB-R provides vital data to determine the presence, type, and severity of aphasia. It also localizes likely neural regions disrupted to spur particular profile deficits. The speech samples under consideration are already benchmarked using the Western Aphasia Battery-Revised (WAB-R) by the clinicians.

Building the application

- Test the LLM using an application interface that takes WA patient's speech transcript as an input and provides linguistic analysis, WAB score, aphasia severity, and suggestive treatment as output
- Validate the suggestive treatment options with a licensed speech therapist
    - **Amazon Bedrock** is a machine learning platform that helps developers build generative AI applications on Amazon's cloud computing platform. The application interface was built using Python (Boto3) with Amazon Bedrock.
    - **Retrieval augmented-generation (RAG) framework using Titan embedding model and vector database** augments the prompts with the patient's latest and historical assessments so that the LLM can suggest treatment options based on their progress over time, enabling personal and context-aware recommendations

Results: Validating LLM

Goal: Test if Claude could analyze the common linguistic construction among WA patients.

Results

- Claude was able to diagnose common issue of word-finding difficulties, where patients used filler words, partial words, and had trouble retrieving specific vocabulary.
- Claude was able to identify common grammar issues in aphasic patients, including the frequent omission of function words like articles, prepositions, and auxiliaries, as well as problems with verb tense, aspect, and agreement.
- Additionally, Claude was able to pinpoint patients' struggles with pronoun usage, plurals, possessives, word order, and sentence termination words.

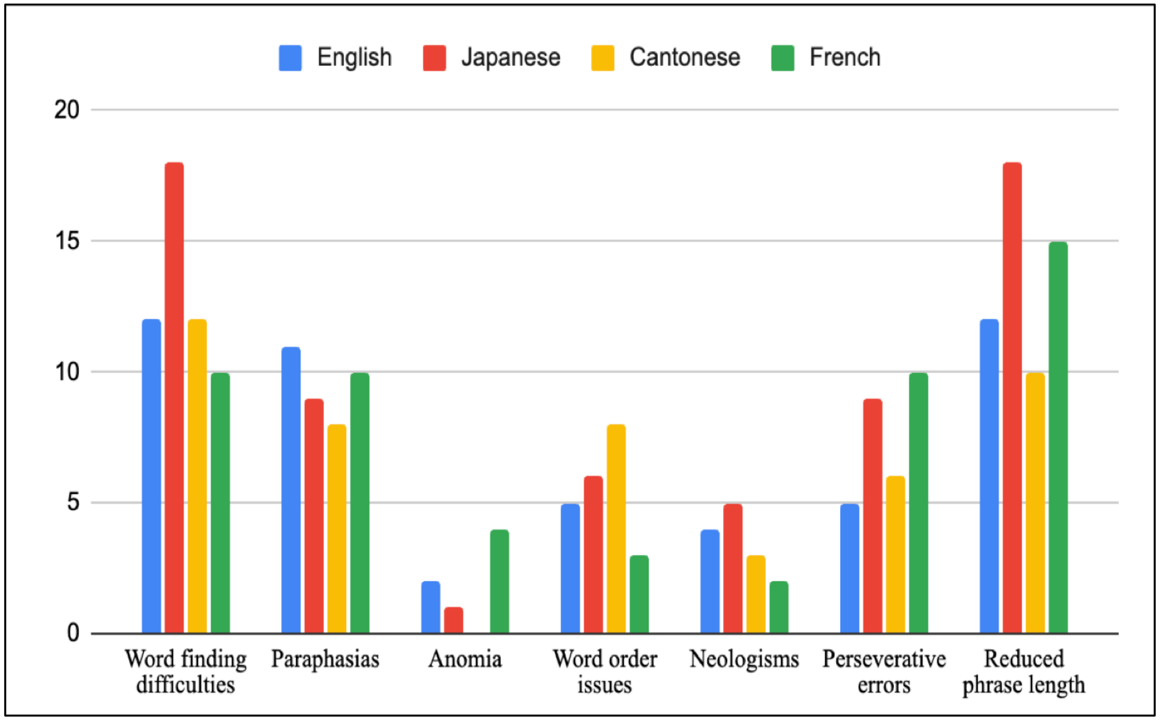

Fig. 5: Visualization of elements of Wernicke's aphasia identified by Claude during the initial analysis of Wernicke's Aphasia Speech
Graph by Sreekar Baddepudi

Word finding difficulties : Using filler words (e.g. "um", "uh"), circumlocution (describing an object rather than naming it directly), semantic paraphasia (using an incorrect but related word), or leaving words unfinished
Paraphasia : Unintended substitutions of sounds, syllables or words, like saying "supper" instead of "slipper." This could be a phonemic, verbal or semantic paraphasia.
Anomia: Difficulty producing names of objects, people or places. Speakers may use generic terms like "thing" instead
Word order issues: Words coming out in the wrong order,
Neologisms: Made up words that the speaker intends as real words, like "spate" instead of "spread"
Perseverative errors: Repeating parts of words or phrases over and over, getting "stuck" on sounds
Reduced phrase length: Shorter phrases and sentences compared to typical speech

Results: Linguistic analysis

Background: The investigation into language complexity dimensions and associated patterns of impairment severity among aphasia patients of diverse native language backgrounds has revealed intriguing insights.

Cross-Linguistic Analysis

- Common errors found across WA patients in all languages included word finding difficulties, par aphasias, neologisms, perseverative errors, and reduced phrase length. Language-specific differences were also noted.
- The manner in which these rule deviations were followed was consistent across all audio files studied, spanning all four languages
- The most common grammar issues were omission of function words like articles and prepositions, and problems with verb tense and agreement. Word order issues were also observed.

Language-to self Analysis:

Language-to-self comparisons showed WA speech deviates from each language's typical grammar conventions in systematic ways, while retaining some core language structure.

1. English: English WA patients produced sentence fragments, omitted function words, made verb tense errors, and displayed word order issues while retaining basic vocabulary and phrases.
2. French: French WA speakers over-applied verb endings, confused gender categories, and used informal word order, but maintained complex verb patterns and described objects indirectly when struggling with retrieval.
3. Japanese: Japanese WA patients omitted particles, dropped sentence-final elements, and produced simplified utterances, but adhered to standard verb positioning when present.
4. Cantonese: Cantonese WA speech featured incomplete phrases, misapplied suffixes, and fragmentation, but followed tonal contour principles and exhibited code-mixing reminiscent of informal dialects

This groundwork offers a valuable platform for in-depth exploration and understanding of language impairments, opening doors to research into how these deficits extend to different linguistic contexts and cognitive processes, thereby enriching our knowledge of language disorders and their underlying mechanisms.

Results: WAB simulation

Goal: Compare Claude LLM prompted emulated WAB-R assessments on the speech transcripts with the WAB-R diagnoses by clinicians

Results

Overall, strong agreement emerged between the patient severity grades computationally derived by the model versus the genuine scores from manual administration of the Western Aphasia Battery (WAB) by specialists, demonstrating promising simulation capability.

Patient ID	Language	Sim Score	WAB Score	Accuracy
P001	English	59	62	94.9
P002	English	45	44	97.8
P003	French	74	78	94.6
P004	French	64	68	93.8
P005	Japanese	42	41	97.6
P006	Japanese	58	58	100
P007	Cantonese	67	64	95.5
P008	Cantonese	51	47	92.2

Fig. 6: Table showing simulated Western Aphasia Battery Scores(WAB) in comparison to clinician determined WAB scores
Chart by Sreekar Baddepudi

- Across the 8 patients spanning English, French, Japanese and Cantonese languages, the model-predicted total WAB scores reached over 90% accuracy relative to the recorded evaluations by trained clinicians.
- Differences for the all of the patients was under 5 points on the full scale.
- This suggests precise capture of pertinent markers within the speech transcripts aligned with expert judgements based on interactive multi-faceted analysis.

Application

Optimization of the model: Careful optimization of hyper-parameters plays a crucial role in utilizing the language model, ensuring it produces the most relevant, consistent, and reliable outputs. The hyper-parameters that produced the most consistent and predictable outputs were a Temperature of 1, Top P of 0.5, K value of 250, and a maximum response length of 2048 tokens. Setting the Temperature to 1 and Top P to 0.5 helped reduce randomness and increase consistency in the model's responses. A K value of 250 optimized the number of candidate options considered at each step, which improved the relevance of responses. The maximum response length to 2048 tokens led to more focused responses.

RAG framework: The application used Amazon Bedrock's Titan embeddings and OpenSearch as vector database.

Building the User Interface: StreamLit: Streamlit is an open-source Python library that is used for building interactive web applications. Streamlit simplifies the process of creating interactive web apps, making data exploration and model deployment accessible to Python developers of all skill levels.

```python
def call_bedrock(prompt, assistant):
    # Call the bedrock client
    bedrock = boto3.client(service_name='bedrock-runtime',
                           region_name='us-east-1',
                           aws_access_key_id=access_key,
                           aws_secret_access_key=access_secret)

    # Tweak your preferred model parameters, prompt and assistant information
    body = json.dumps({
        "prompt": f"\n\nHuman:{prompt}\n\nAssistant:{assistant}",
        "max_tokens_to_sample": 4096,
        "temperature": 0.1,
        "top_p": 0.7,})

    # Define the type of model that will be used
    modelId = 'anthropic.claude-v2'
    accept = 'application/json'
    contentType = 'application/json'

    # Call the Bedrock API
    response = bedrock.invoke_model(body=body, modelId=modelId, accept=accept, contentType=contentType)
    response_body = json.loads(response.get('body').read())
    completion = response_body.get('completion')

    return completion
```

Fig. 7: Snippet of code
Code by Sreekar Baddepudi

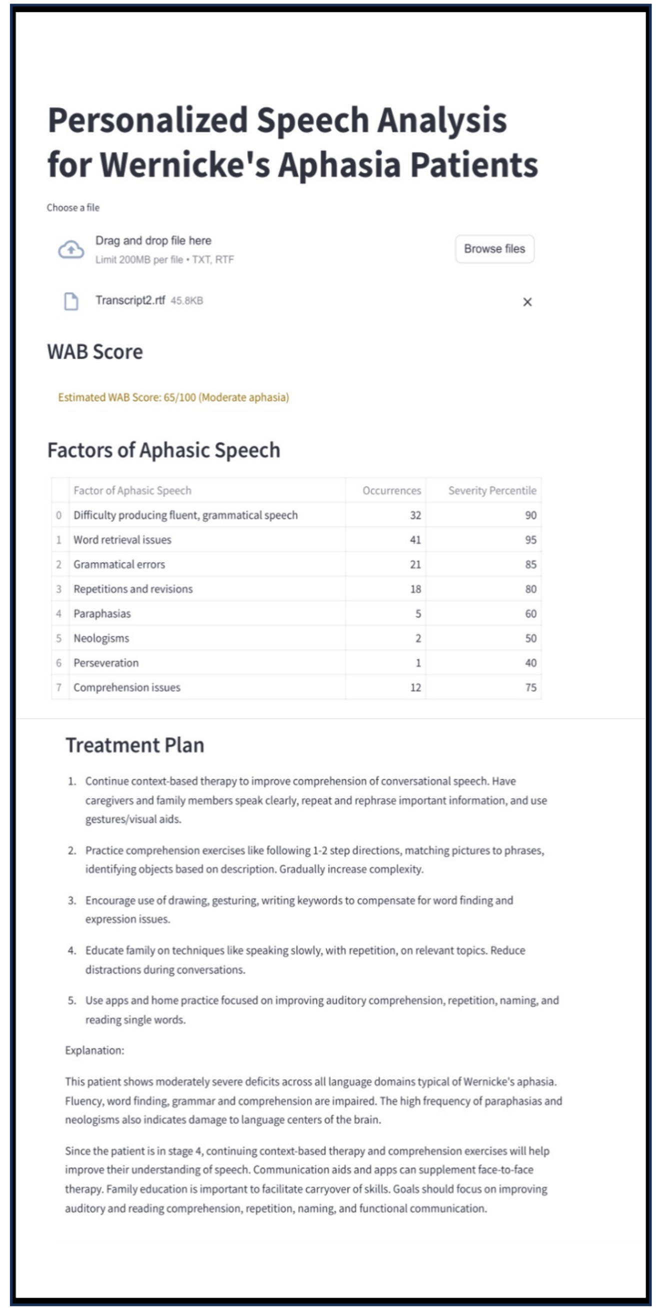

Fig. 8: An example of what the app interface that was built with StreamLit looks like
Image by Sreekar Baddepudi

Discussion

Findings: The LLM-assisted analysis of Wernicke's aphasia yielded promising results, demonstrating the potential for this approach to revolutionize research and clinical practices in the field of aphasia. The key findings highlight the efficiency gains and the ability to generalize the techniques aphasia patients from various backgrounds.

- The LLM's ability to process large amounts of data rapidly and its multilingual capabilities were instrumental in accelerating the analysis process for this project.
- The efficiency gains highlight the potential for LLMs to revolutionize research and clinical practices in the field of aphasia and beyond.
- A skilled clinician armed with the insights derived from the LLM-assisted analysis can effectively differentiate between various aphasia types, which is crucial for crafting personalized treatment plans.

Application to other forms of aphasia

The successful application of the analysis techniques to Broca's aphasia suggests that the approach may be effective for a wide range of aphasia syndromes. By leveraging the insights gained from these techniques, clinicians can develop personalized treatment plans that cater to the specific needs of patients with different aphasia types.

- The analysis techniques used in this project, which focused on Wernicke's aphasia, were also applied to Broca's aphasia with similarly impressive results.
- The success of the analysis techniques in both Wernicke's and Broca's aphasia highlights the potential for this approach to be adapted and utilized in the assessment and management of a wide range of aphasia types.

Limitations

While the LLM-assisted analysis has shown promising results, the project also revealed some limitations in the current diagnostic approach.

- The discrepancy in deficit severity determination for one of the Frencacross diverse languages indicates that the algorithms may be biased towards certain error markers based on the structure of the native language.
- The bias could lead to an overemphasis on some error patterns while undervaluing others that may be clinically significant.
- The current model's reliance on a limited set of error markers may not capture the full complexity of language impairments in aphasia.

Conclusions

This application demonstrates successfully modeling of the trusted gold standard WAB testing toolkit that can be used by the clinicians for

- Identifying universal linguistic biomarkers that transcend language barriers and those that lie within them
- Minimizing human error in treatment and diagnosis, and speed up the analysis of complex diseases
- Personalizing aphasia treatment by generating language exercises tailored to a patient's unique deficits and native language.
- Achieved a 95% accuracy rate compared to clinical WAB assessments
- Assisting large-scale automated assessments to measure trends across various demographics
- Multilingual capabilities of the application expands the reach of aphasia care, ensuring that patients from diverse backgrounds have access to cutting-edge diagnostic and therapeutic resources

Future work

Application

Create a comprehensive understanding of each patient's condition by combining speech pattern analysis with real-time brain activity measurements.

- Integration of speech analysis data with EEG and fMRI brain imaging to capture real-time brain activity and functional connectivity patterns during language processing.
- Include aphasia diagnosis like motor ability and image recognition parameters in the multi-modal AI model to analyze and assist with increased accuracy

Analysis

- Extend analyses for all forms of aphasia, such as Broca's and Anomic Aphasia
- Expand the languages used in this analysis for addressing linguistic diversity in the diagnosis and treatment of Wernicke's aphasia.

References

1. Acharya, Aninda B., and Michael Wroten. "Wernicke Aphasia." StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing LLC, 2023 Jan. 2023 Aug 8. PMID: 28722980.
2. Binder JR. Current Controversies on Wernicke's Area and its Role in Language. Curr Neurol Neurosci Rep. 2017 Aug;17(8):58. [PubMed: 28656532].
3. Fridriksson J, den Ouden DB, Hillis AE, Hickok G, Rorden C, Basilakos A, Yourganov G, Bonilha L. Anatomy of aphasia revisited. Brain. 2018 Mar 01;141(3):848-862. [PMC free article: PMC5837461] [PubMed: 29360947].
4. American Speech-Language-Hearing Association. Aphasia. Available at: https://www.asha.org/public/speech/disorders/aphasia/. Accessed November 2, 2023.
5. Ropper AH, Samuels MA, Klein J et al. Adams and Victor's Principles of Neurology 11th Edition. (2019) ISBN: 9780071842617.
6. Kassebaum NJ, Murray CJ, Lopez AD. Burden of disease and injury. In: Murray CJ, Lopez AD, eds. The Global Burden of Disease: A Comprehensive Assessment 1990-2010. Cambridge, MA: Harvard University Press; 2012: Chapter 1.
7. Benjamin N, Leveque C, Howard D, Lindamer L, Turk A, Haaland K. Communication disorders in adults: Stroke.2018;49(10):2736-2744.
8. Rizzi, L. (1997). The fine structure of the left periphery. In: L. Haegeman (Ed.), Elements of grammar: A handbook of generative syntax (pp. 281-337). Kluwer Academic Publishers.
9. Lazard, G. (1994). Grammaire comparée des langues romanes. Presses universitaires de France.
10. Huddleston, R., & Pullum, G. K. (2002). The Cambridge grammar of the English language. Cambridge University Press.
11. Matthews, S., & Yip, V. (1994). Cantonese: A comprehensive grammar. Routledge.
12. Shibatani, M. (1990). The languages of Japan. Cambridge University Press.
13. Goodglass, H., & Kaplan, E. (1968). The Assessment of Aphasia and Related Disorders. Lea & Febiger, Philadelphia, p. 248.
14. MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. Aphasiology, 25, 1286-1307.
15. Robson, H., Zahn, R., Keidel, J.L., Binney, R.J., Sage, K., & Lambon Ralph, M.A. (2014). The anterior temporal lobes support residual comprehension in Wernicke's aphasia. Brain, 137(Pt 3), 931–943. https://doi.org/10.1093/brain/awt373