Proof of clinical feasibility of natural language learning with machines

Mihir Mulye1,4,*, Stefan Conrad1,† and Stefan Knecht2,†

1Institut für Informatik, Heinrich Heine University, Düsseldorf
2Institut für Klinische Neurowissenschaften und Medizinische Psychologie, Universitätsklinikum Düsseldorf

Abstract
Aphasia is a condition resulting from a stroke or injury to the brain. This condition renders the patient unable to use language efficiently. Machine learning and artificial intelligence automate the assessment and diagnosis of aphasia to a certain extent. However, solutions based on machine learning that are patient-centric, self-paced, and help with therapy are lacking. In this work, we present a rehabilitation framework that can be implemented to address this deficit. As a part of an ongoing project, we aim to implement these ideas to facilitate the rehabilitation of aphasic patients by benefiting from the developments in machine learning. It is also anticipated that these ideas can be extended to help people wanting to acquire a new language.

Keywords
aphasia, stroke, machine learning (ML), artificial intelligence (AI), language rehabilitation, people with aphasia (PWA).

1. Introduction
The rapid advancements in machine learning and artificial intelligence have made it possible to solve diverse problems. A particular use case in healthcare that has attracted the attention of researchers is the application of machine learning approaches to manage aphasia. Aphasia is an acquired language disorder that can be caused by an injury to the brain, a stroke, or a result of degenerative processes [1, 2]. This condition generally impairs the affected person’s ability to speak, comprehend, read, and write often to varying degrees [2, 3]. People suffering from aphasia experience life-altering psychosocial consequences such as reduced participation in social settings, and diminished quality of life possibly leading to psychological conditions such as anxiety and depression [4]. Some patients are unable to function as independent individuals as a consequence of this condition [5].

Therapy and recovery to some extent are possible in most cases. However, constraints such as a shortage of therapists, availability of frequent therapy appointments for sustained recovery, and limited funding of support infrastructure for long-term rehabilitation programs prove to be a hurdle [2].

In this work, we discuss ideas as a part of an ongoing project, that could be used to rehabilitate People With Aphasia (PWA). Our objective is to implement a rehabilitation framework that is easy to use, patient-centric, self-paced, and attempts to utilize the advances in machine learning. We aim to design the framework and offer the patients the possibility to practice as often and for as many iterations as they would like. We expect to eventually reduce the patients’ dependence on therapists and provide more flexibility during rehabilitation.

The subsequent sections in this paper present the related research, discussions regarding the approach planned to be implemented throughout this project, and the extent of implementation followed by future work. The final section concludes the discussion conducted in this paper.

2. Related work
An overview of machine learning in aphasia management has been provided in [2, 6]. The discussions in these papers conclude that the majority of the works use machine learning for the diagnosis and assessment of aphasia. The authors highlight the need for exploring the possibility of utilizing the advances in machine learning to improve aphasia rehabilitation as well.

Kohlschein et al., 2017 [7] propose an approach using Bag of Audio Words and Long Short-Term Memory (LSTM) neural networks to automate the detection and classification of aphasia speech. In this work, neural networks are trained to identify the features corresponding to aphasic speech and then classify the severity of disability based on the input audio sample. Similar to this work, are approaches proposed by Hirsch et al., 2023 [8] and Barbera et al. 2021 [9]. In these works, the authors utilize automatic speech recognition networks to categorize if the uttered word is correct or incorrect compared to the ground truth. Mahmoud et al., 2021 [10] compare the
An analysis (LDA) among others to classify speech. Aphasic

Though not directly related to aphasia management, this

(visual and audio streams) to learn the word-referent map-

programs to categorize and predict the responsiveness to

post-stroke aphasia patients with machine learning al-

grating the behavioral and brain variables obtained from

and might improve the efficiency of the rehabilitation.

needed for preparing patient-specific therapy material

that such quantitative analysis could minimize the time

and their frequency of occurrence. The authors claim

such quantitative analysis could minimize the time

needed for preparing patient-specific therapy material

and might improve the efficiency of the rehabilitation.

Gu et al., 2020 [15] explored the possibility of integ-

rating the behavioral and brain variables obtained from

post-stroke aphasia patients with machine learning al-

gorithms to categorize and predict the responsiveness to

the therapy. The goal of this study was to forecast the

individualized rehabilitation results.

Das et al., 2022 [16] discuss the possibility of using

extended reality to help rehabilitate stroke patients. In

particular, the authors predict that delivering speech ther-

apy via extended reality could be a helpful solution for

people with aphasia in an immersive setting.

From the discussions of existing research work con-

ducted in this section, we find that machine learning

approaches have been primarily used to assess, diagnose,

and facilitate understanding of aphasia. However, a sys-


tematic framework that is easy to use and focuses on

helping aphasic patients recover functionally in a user-

centric and self-paced manner using machine learning

approaches is lacking [6].

3. Proposed approach

In this section, we discuss an approach designed to help

rehabilitate PWA. We provide an overview of the planned

approach, discuss the rehabilitation tasks required for

therapy, establish the necessity for implementing adap-

tive difficulty, and finally the possibility of expanding

this framework to acquire a new language.

3.1. Language learning framework

We propose incorporating the tasks and activities [17]

that PWAs accomplish during their traditional therapy,

into an online framework. To this end, we use the Flutter

toolkit based on the Dart programming language to im-

plement this framework. This affords the possibility to
develop applications that run across multiple platforms.

We intend to implement this framework as a gamified

platform. The basic Graphical User Interface (GUI) de-

pends on the task being presented to the patients. The

GUIs for two tasks (of the four) are presented in Figure

1 and Figure 2. The figures show a tablet emulator with

the interface visible to the patients. These figures show

a virtual therapist who serves as a guide to the patients

throughout this therapy experience. The tasks are pre-

sented as questions that the patient answers to progress.

The gamified aspect of the framework is implemented

by including a computerized rival that aids in demon-

strating the workings of the game and also serves as a

benchmark for the patients to compare their performance

against. We have implemented additional components of

uncertainty and randomness into the framework. These

include elements such as a random probability generator
to determine whether to award points for correctly an-

swering a task. Including these aspects should make the

rehabilitation experience more engaging.
3.2. Tasks for rehabilitation

Currently, we plan on integrating four tasks that are involved in the traditional therapy for aphasia. These tasks focus on improving individual aspects of language ranging from single-word level to sentence-level complexity. These tasks also focus on improving skills such as retrieving a desired word, coherent speech, and fluency. An overview of the planned tasks is provided here:

- **Comprehension:**
  This activity helps the patient at a single-word level and strengthens the word-object mapping. The patient is provided with two images and an audio prompt. The patient has to select the image corresponding to the spoken word prompt. The GUI for this exercise is depicted in Figure 1. As a demonstration, the computerized rival exhibits how to answer this task with several audio prompts. As the computerized rival is implemented to select the correct options with a preset probability, it is also possible for the rival to commit errors. This uncertainty aids in exhibiting to the patients, how not to solve the task. Success and failure of the rival are appropriately indicated by providing video (via gifs) and audio (via notification tone) feedback. After this demonstration, the patient is provided the audio prompt to solve the task. Identical video and audio feedback are provided to the patient to inform them how they perform the task.

- **Repetition:**
  This activity helps improve the patient’s ability to understand spoken words and produce coherent speech. The patient is provided with an audio input in this task. The patient has to then reproduce that particular word by speaking. In this task, the patient might answer in several ways. The patient might repeat the name correctly. The patient can also incorrectly repeat the given word. It is also possible that the patient might hesitate before answering or produce incoherent speech. It is therefore imperative to use a speech recognition system and corresponding evaluation metric that handles the listed cases robustly. We plan to record the patient’s speech and use speech recognition systems similar to [7, 8, 9] to assess the patient response.

- **Naming:**
  This task aims to improve the patient’s ability to retrieve words from their vocabulary and produce coherent speech. In this task, the patient is provided a visual prompt as seen in Figure 2. The patient has to speak the name of the object. As the patient output modality is similar to the previous task (audio/speech), it would be logical to use a similar processing methodology. The speech systems we select to use should be able to determine if the patient’s speech corresponds to the given prompt or not.

- **Multi-word comprehension:**
  This task helps to improve the patient’s ability to understand the context of a given situation, formulate their thoughts, and articulate them in words. In this task, the patient is given an input image similar to the one shown in Figure 3 and is asked questions about this image. The patient should then answer the questions based on what they observe in this image. For instance, for the
depicted image, the patient might be asked to answer questions such as:

(i) What do you see in this image?
(ii) How many people are in this image?
(iii) What is kept on the table?
(iv) Could you point to the [lamp, windows, table, chair]?

The answer provided by the patient will naturally depend on the type of question being asked. For instance, for (i) and (iii), the answer might be descriptive, whereas it might primarily be numeric for (ii). Also, it is possible that for (iv), the patient might either answer verbally, with gestures, or point to the relevant entity in the image. This necessitates a robust evaluation strategy for this task.

### 3.3. Adaptive difficulty adjustment

In addition to the tasks discussed above, we plan to incorporate the principle of adaptive difficulty adjustment into our framework. This idea has been explored in video games to tune the difficulty level of the game based on the skill and performance of the player [18, 19, 20, 21, 22, 23]. If the player finds the game too difficult, the game difficulty is lowered. However, if the player finds the game too easy, the difficulty level is increased. This concept has been visualized in Figure 4. The optimum zone of operation where the player is engaged and upgrades their skill is in the flow zone. Anywhere other than this zone would make the game either too difficult/too easy or not substantially rewarding for the player.

This principle can be utilized in aphasia rehabilitation tasks as well [4]. If the patient finds the given task too difficult, the patient might be frustrated. To avoid this, the difficulty level of the task should be lowered. Once the patient is comfortable, the difficulty level can be gradually increased. This is done to avoid the situation in which the patient finds the tasks too easy to solve and loses interest. Many approaches exist to implement the adaptive difficulty adjustment as is discussed in [23, 24]. Additional research will be conducted to identify which adaptive difficulty adjustment approaches are suitable to integrate into our framework. It also needs to be determined, which parameters are necessary for implementing the chosen algorithm. Quantities such as the patient’s response time, the difficulty level of the task being solved, and the patient’s performance/score in the tasks of varying difficulties among others can be utilized to tune the dynamic difficulty algorithm correspondingly. Naturally, the choice of these parameters depends on the selection of the algorithm.

### 3.4. Acquisition of a new language

The framework discussed in previous sections can theoretically be used by healthy individuals who want to learn a new language. Starting from a single-word level complexity, the subject aiming to learn a new language can acquire the ability to understand the situational context with the aid of multi-modal data such as images, audio, and videos. To this end, the framework must be expanded to support multiple languages. This process will involve acquiring audio data of spoken words, and a description of visual data in the target language among other upgrades. These modifications can be implemented by harnessing the power of generative AI as is explored in the works of [11].

### 4. Discussions and Future Work

In this section, we discuss the state of implementation of the ideas discussed in this paper. Additionally, we provide the potential research directions to be explored.

- **Implementation of tasks**: The discussions concluded above provide a high-level overview of the task description. The implementation of these tasks is in progress. The first iteration of the comprehension task is implemented and the subsequent tasks will be targeted in the future.
- **Evaluation criteria**: For the comprehension task, the accuracy of the patient response can be used as the evaluation metric. The patient’s
5. Conclusion

Advances in machine learning and generative artificial intelligence can be used to create training material for rehabilitation. Also, implementing concepts from the domain of game design such as adaptive difficulty adjustment algorithms into our proposed therapy framework can make it engaging and rewarding. Finally, the scope of the rehabilitation framework can be extended with upgrades for additional languages. This will make the framework useful for healthy individuals who want to acquire a new language.

Acknowledgments

We acknowledge support from the Federal Ministry of Education and Research (BMBF) under grant number 13GW0481C.

References


A. Online Resources

- AI Hotpot: https://hotpot.ai/blog/ai-art-and-image-generator-guide