

POLITECNICO

MILANO 1863

ARTYQUEST — Elevating Speech Therapy Through Personalized Exercises with LLM Integration

Course:

Advanced User Interfaces (077685)
Computer Science & Engineering

Authors:

Anuntaya Kitiporn
Nuno Costa
Frederik Andersen

Supervisor:

Pietro Covari

Date:

January 12, 2024

Abstract

This report examines *ArtyQuest*, a groundbreaking web application designed to enhance speech therapy for children using Generative AI. The project addresses the need for personalized, technology-enhanced speech therapy, with a focus on early intervention and collaboration between therapists and patients. *ArtyQuest* achieves this through a conversational agent-based interface that empowers therapists to create customized exercises while also providing children with engaging exercises tailored to their interests and speech conditions. The technical implementation relies on a ReactJS frontend and a Python backend powered by Flask, integrating external APIs from OpenAI and Azure for exercise generation and audio assessment.

ArtyQuest offers personalized and engaging exercises, a child-friendly UX/UI, AI-powered exercise generation, accessibility, and empowerment of therapists. Initial user feedback highlights positive impressions of the user interface, audio feedback, and exercise generation process. However, future work involves gathering user feedback, introducing new exercise types, refining the app based on analytics, and advanced AI integration to address evolving user needs.

In summary, *ArtyQuest* represents a promising advancement in speech therapy for children, offering a user-centric, technologically advanced solution that empowers and supports their speech and language development journey.

Team



Anuntaya Kitiporn
10987357@mail.polimi.it
Management Engineering



Nuno Costa
10988150@mail.polimi.it
Computer Science and
Engineering



Frederik Andersen
10984837@mail.polimi.it
Management Engineering

Member Contributions

Report

Role/Contribution	Name
Abstract/Executive summary	Frederik Andersen
Requirements	Frederik Andersen
State of the art	Frederik Andersen
Solution (UX Design)	Frederik Andersen
Solution (Implementation)	Nuno Costa
Empirical Evaluation	Nuno Costa
Value Proposition	Anuntaya Kitiporn
Discussion and future work	Anuntaya Kitiporn

Solution

Name	Role/Contribution	Description
Nuno Costa	Technical Architecture & Backend Implementation	Defined the system's underlying architecture, developed the API connections and backend, including data storage. Also responsible for prompt engineering and integration with the frontend.
Anuntaya Kitiporn & Nuno Costa	Front end implementation	Implemented the visual components of the application, focusing on user interaction and workflow.
Frederik Andersen & Anuntaya Kitiporn	UX Design	Focused on UI/UX ideation and created all necessary mockups. Played a pivotal role in the design of the user interface and user experience.

Contents

Abstract	2
Team	3
Member Contributions	4
1 Executive Summary	8
2 Workflow Activities	10
2.1 Conceptualization	10
2.2 Low-Fidelity Prototyping	10
2.3 Cyclic Evaluation and Development	11
2.3.1 Evaluation	11
2.3.2 Development	11
3 Requirements	12
3.1 Main Target Groups/Stakeholders	12
3.2 Needs of the Stakeholders	12
3.3 Context	13
3.4 Constraints	13
3.5 Goals	13
4 State of the Art	15
4.1 Speech Therapy	15
4.2 Speech Therapy in Children vs Adult	16
4.3 Automatic Exercise Generation through LLMs	16
5 UX Design Solution	17
5.1 General Interaction Paradigm	17

5.2	User Workflow	17
5.2.1	Therapists' Interface	18
5.2.2	Children's Interface	18
5.3	Scenarios of Use	19
5.3.1	Scenario for Children	19
5.3.2	Scenario for Therapists	19
5.4	Design Motivations	19
6	Technological Solution	23
6.1	General Technological Approach	23
6.2	Software Architecture	23
6.2.1	Frontend with React	23
6.2.2	Backend with Flask	24
6.2.3	OpenAI's API Integration	24
6.2.4	Azure's API Integration	24
6.3	Exercise Generation Process	24
7	Evaluation	26
7.1	Usability and User Experience	26
7.2	Accessibility, Performance and Reliability	27
7.2.1	Accessibility	27
7.2.2	Accuracy of Content Generation	27
7.2.3	Audio Evaluation Mechanism	31
7.2.4	Application Reliability	32
7.3	Integration and Compatibility	32
8	Value Proposition	34
8.1	Personalized and Engaging Exercises	34
8.2	Child-friendly UX-UI	34
8.3	AI-Powered Exercise Generation	35
8.4	Accessible Anytime, Anywhere	35
8.5	Empowering Therapists	35
9	Discussion and Future Work	36
9.1	Critical Reflection	36
9.2	Future Directions	36
9.2.1	Short-Term Enhancements (0-6 months)	37

9.2.2	Mid-Term Development (6-12 months)	37
9.2.3	Ongoing Refinement	37
9.2.4	Long-Term Vision (12+ months and beyond)	37
9.3	Concluding Thoughts	37

Chapter 1

Executive Summary

The *AstyQuest* project represents a pioneering web application that aims to transform speech therapy for children by harnessing generative AI technology. This executive summary offers a comprehensive overview of the project, covering its core objectives, technical architecture, key features, initial user feedback, and plans for future development.

AstyQuest addresses the critical need for personalized, technology-integrated speech therapy, with a particular emphasis on early intervention and fostering collaboration between therapists and their young patients. At its heart, the project introduces a conversational agent-based interface, empowering speech therapists to create tailored exercises while simultaneously offering children engaging activities that cater to their specific speech conditions and personal interests. The project's technical foundation comprises a ReactJS frontend and a Python backend powered by Flask. Seamless integration with external APIs from OpenAI and Azure enables exercise generation and audio assessment, ensuring a responsive and efficient user experience.

AstyQuest offers a range of essential features, including personalized and engaging exercises, a child-friendly user interface, AI-driven exercise generation, accessibility, and tools to support therapists. Initial user feedback has been positive, particularly regarding the user interface, audio feedback, and exercise generation process. Moving forward, *AstyQuest* has laid out a strategic roadmap for further development. This roadmap spans short-term, mid-term, and long-term objectives, all rooted in the principles of user-centricity and continuous improvement. These objectives include systematic user feedback collection, the introduction of new exercise types, data-driven refinements based on user analytics, and the integration of advanced AI capabilities to meet the evolving needs of the user community. In summary,

ArtyQuest represents a significant advancement in pediatric speech therapy. It offers a user-centric, technologically advanced solution that empowers and supports children's speech and language development. With its user-friendly interface, AI-powered exercises, and collaborative approach, *ArtyQuest* holds great promise for enhancing the speech therapy experience for children.

Chapter 2

Workflow Activities

The development of *AstyQuest* was characterized by a dynamic and iterative process, where evaluation and development phases operated in a cyclical fashion, informing and refining each other. This chapter outlines the sequence of workflow activities, from the initial market research to the development of a functional prototype, with continuous iterations shaped by user feedback and usability testing. The cyclical nature of evaluation and development ensured that the final product was not only aligned with user needs but also adhered to the highest standards of user experience and technical implementation.

2.1 Conceptualization

The journey of *AstyQuest* began with a thorough market analysis aimed at identifying the unmet needs within the realm of speech therapy applications. This research informed the development of a comprehensive concept, delineating the core features, the primary audience, and the unique value proposition of *AstyQuest*. This concept served as the cornerstone of the project, directing the design and development efforts towards creating a solution that would stand out in the market.

2.2 Low-Fidelity Prototyping

With the foundational concept in place, the team proceeded to create low-fidelity wireframes and initial design concepts. This stage focused on ensuring that the navigation was user-friendly and that the visuals were captivating and engaging. The low-fidelity prototypes were instrumental in visualizing the solution and provided a tangible reference for user testing and

feedback.

2.3 Cyclic Evaluation and Development

The core of *AstyQuest*'s development process was the cyclic nature of evaluation and development. Usability testing was conducted rigorously with the tutor and later on with other users to gather feedback, which was then swiftly acted upon to make necessary adjustments, thereby enhancing the application's usability.

2.3.1 Evaluation

During the evaluation phase, the team asked for and gathered feedback from the tutor on the application's functionality. This feedback was critical in identifying areas that required refinement and ensuring that the application met the users' needs and expectations. Later on, in the *Demo Madness* event, we also took the opportunity to gather feedback from other users.

2.3.2 Development

In parallel, the development phase was set in motion with the creation of a functional prototype. Regular sprint meetings guided the implementation of core features to ensure that the development was in sync with the project's objectives and user feedback.

The iterative loop between evaluation and development allowed the team to remain agile, making iterative improvements to the application. This approach tried to ensure that the final product was a result of a user-centric design process, even though we did not have access to either therapists or children to test our application.

In summary, the workflow activities of *AstyQuest* reflect a meticulous balance between user experience design and technical development, with a strong emphasis on iterative improvement and user satisfaction.

Chapter 3

Requirements

The inception of "Experiencer", the initial idea for a *tool that, given a set of speech therapy exercises and information about the patient's profile, creates an engaging experience around such exercises, wrapping them within automatically generated conversations and multimedia contents*, set the foundation for what would evolve into "ArtyQuest," an innovative tool designed to enhance the field of speech therapy. This chapter details the requirements of our project, reflecting a synthesis of stakeholder needs, context, constraints, and system goals as outlined in the User-Needs-Goals (UNG) diagram.

3.1 Main Target Groups/Stakeholders

ArtyQuest caters to two primary groups of stakeholders: children with language impairments and the speech therapists dedicated to their care. Children, as the direct beneficiaries, require engaging and effective therapy that holds their attention and fosters consistent practice. Speech therapists, integral in the therapeutic process, need a platform to craft exercises that are not only tailored to each child's linguistic challenges but also to their interests and that, somehow, are also measurable in terms of progress.

3.2 Needs of the Stakeholders

The needs of these stakeholders are paramount. Children look for captivating exercises that can make their therapy sessions enjoyable and stimulating, ensuring that they remain engaged throughout the process. Speech therapists seek an easy-to-navigate platform that provides a seamless exercise generation process, allowing them to assist in children's measurable speech

articulation progress over time. Furthermore, they require access to tools that enable the tracking of the child's advancement in speech therapy, a vital component for successful treatment.

3.3 Context

The application of *AntyQuest* spans various contexts. It is designed for flexibility, accommodating use during therapy sessions at clinics or schools, as well as at home, where therapy can continue in a comfortable and familiar setting. The adaptability of the platform is central to its design, ensuring that no matter the context — be it during therapy or in daily life — children can continue to practice and improve their speech with the support of their therapists.

3.4 Constraints

Within the project's scope, several constraints are acknowledged: time limitations that dictate the optimal length of exercises, the attention span of young users, physical limitations related to the device in use, and technological requirements for modern browsers and stable internet connections. Each of these constraints is carefully considered in the design of "AntyQuest."

3.5 Goals

The goals of *AntyQuest* are distinct and well-defined. The system aspires to offer a wide array of dynamic content, including videos, images, sounds, and colors, all tailored to each exercise. The exercises are designed to be as effective as or more effective than traditional therapy methods, with an emphasis on intuitive navigation and a user-friendly interface. Speech therapists are provided with tools to create, edit, delete, and manage exercises with flexibility, ensuring the system meets the evolving needs of each child.

In conclusion, *AntyQuest* emerges from the "Experiencer" project with a focused vision: to bridge the capabilities of Large Language Models (LLMs) with the personalized needs of speech therapy. By aligning our project with the detailed requirements laid out in the UNG diagram, we are committed to delivering a platform that will transform the therapeutic experience for children and empower therapists to create enriching, effective, and personalized speech therapy experiences.

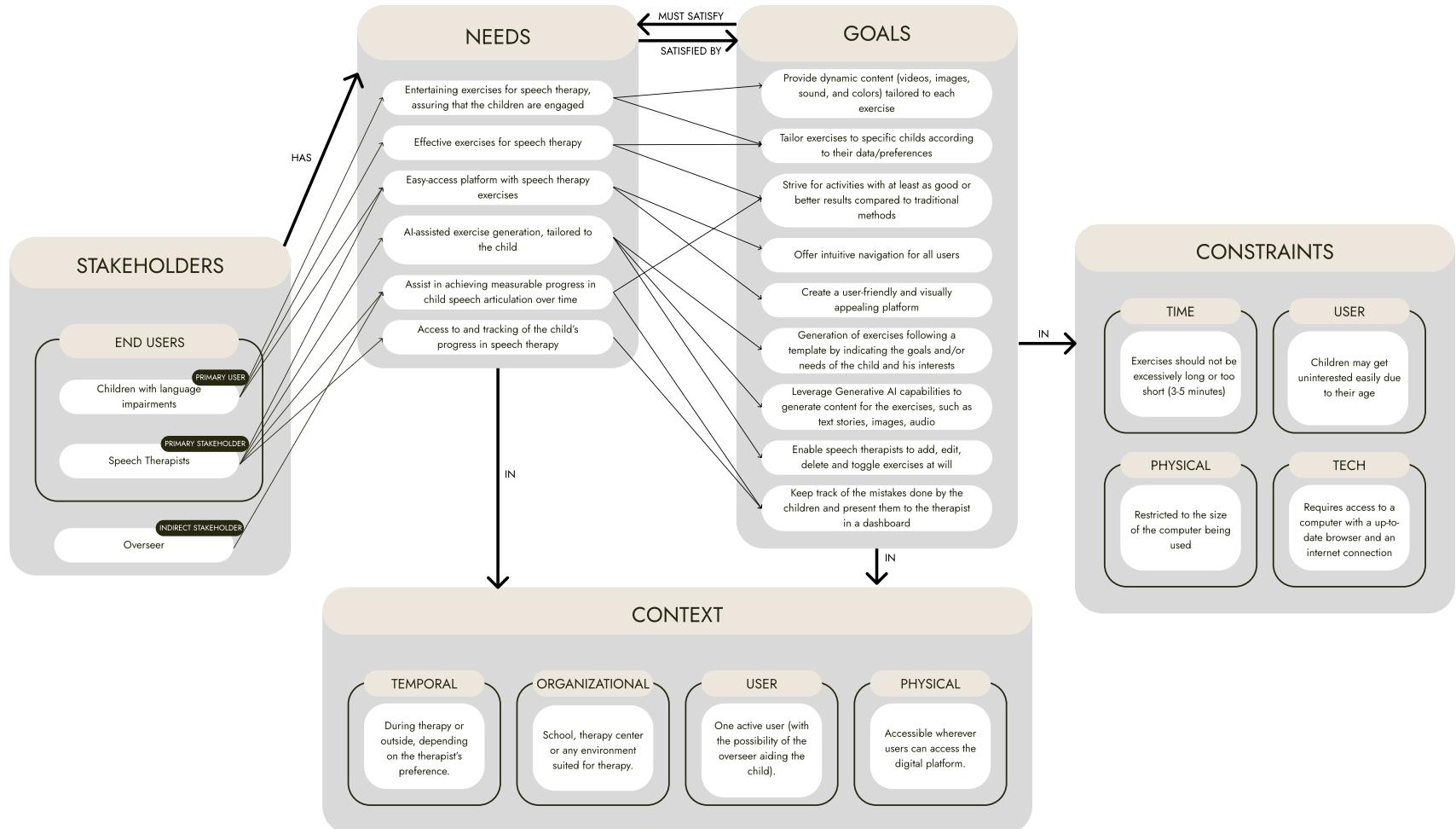


Figure 3.1: The User-Needs-Goals (UNG) diagram illustrating the intricate relationship between stakeholders' needs, the diverse contexts of those needs, and the ambitious goals of the *ArtyQuest* system.

Chapter 4

State of the Art

This section unfolds three interconnected segments, each offering a distinctive perspective on the realm of speech therapy. The initial subsection delves into the historical foundations and essential definitions of speech therapy, with a specific focus on children. Following this, the analysis shifts to age-specific considerations, contrasting the natural evolution of speech disorders in children with the nuanced challenges faced in adult populations. The review culminates with a forward-looking exploration of technological integration, spotlighting the use of Large Language Models (LLMs) for Automated Exercise Generation. Through this coherent structure, the review navigates from foundational insights to age-stratified approaches and, ultimately, to the technological frontiers shaping the future of speech therapy.

4.1 Speech Therapy

The field of speech therapy has evolved significantly, addressing various aspects of speech disorders and rehabilitation. Current research and practices highlight the importance of tailored approaches, early intervention, and the integration of technology in speech therapy. The speech therapy rehabilitation process, particularly for post-surgery patients dealing with voice and swallowing issues, emphasizes the significance of location, early psychological support, and personalized therapeutic exercises based on the surgery type. This holistic approach spans in-hospital and outpatient stages, prioritizing voice restoration and the gradual reintroduction of foods [1].

Enderby and Emerson's investigation into the effectiveness of speech and language therapy reveals diverse outcomes across different disorders [2]. Intensive targeted therapy shows positive results for dysphasia, especially when tailored to specific patient needs. Challenges

persist in determining the effectiveness of therapy for children's speech disorders, dysarthria, apraxia, laryngectomy, learning disabilities, and stammering, indicating a need for more diverse research methods and comprehensive evaluations. The American Speech-Language-Hearing Association delves into the details of various speech disorders, their symptoms, and the most effective treatment approaches [3]. The research by Australian Speech-Language Pathologists underscores the value of hybrid therapies tailored to individual needs [4].

The Cleveland Clinic emphasizes exercises to strengthen tongue and mouth muscles, facial movements for improved motor skills, reading aloud to enhance mouth and tongue coordination, and word games to maintain cognitive function [5]. Popovici and Buică-Belciu highlight the transition from preset exercises to contemporary systems offering comprehensive tools, including real-time feedback [6].

4.2 Speech Therapy in Children vs Adult

The landscape of speech therapy manifests distinctive nuances when applied to children versus adults. Pediatric speech therapy focuses on early intervention, addressing developmental milestones, and fostering age-appropriate communication skills during the formative years. Adult speech therapy centers on retraining or improving existing communication skills, recovering from injuries or medical conditions and managing acquired speech or language disorders [7, 8].

4.3 Automatic Exercise Generation through LLMs

The intersection of LLMs and speech therapy tools heralds a transformative era in personalized rehabilitation. Research in automatic exercise generation through LLMs has yielded promising outcomes across various domains [9, 10]. A systematic review explores online speech therapy systems for childhood communication disorders, underscoring the crucial role of OST systems in supporting speech-language pathologists [11].

In summary, the state-of-the-art in speech therapy recognizes the need for personalized, technology-enhanced approaches and emphasizes early intervention and collaborative decision-making between therapists and patients.

Chapter 5

UX Design Solution

The *AstyQuest* platform represents a confluence of user-centric design and advanced technology, aiming to create an engaging journey for users through interactive speech therapy exercises. This chapter delves into the UX design aspect of *AstyQuest*, providing an in-depth analysis from a user-centric perspective, including our general interaction paradigm, detailed user workflows, and illustrative scenarios of the solution's practical applications.

5.1 General Interaction Paradigm

AstyQuest is a web application that leverages the conversational agent paradigm to facilitate interaction between the system and its users, which include both speech therapists and children with speech conditions. The system employs natural language processing (NLP) to understand and respond to user inputs, creating a seamless flow of communication through both written and spoken language. This allows for a broad range of topics to be covered and for subtle nuances and ambiguities to be effectively managed, thanks to the integration of Large Language Models (LLMs) like ChatGPT, which enhances the conversational agent's proficiency.

5.2 User Workflow

The application is structured into two main sections, each catering to the different needs of our stakeholders:

5.2.1 Therapists' Interface

Speech therapists have the ability to input each child's conditions and personal interests, creating personalized exercises. The system processes these inputs using its NLP capabilities to generate tailored content for the children. They can also easily view and manage their patient's information, improvement and exercises through an intuitive dashboard.

5.2.2 Children's Interface

Children interact with the system through exercises based on the therapist's inputs. They are presented with an animated narrative and are tasked with articulating specific sentences that target their speech impairments. The system assesses their responses and provides feedback through a user-friendly interface.

A graphical representation of this user workflow is depicted below:

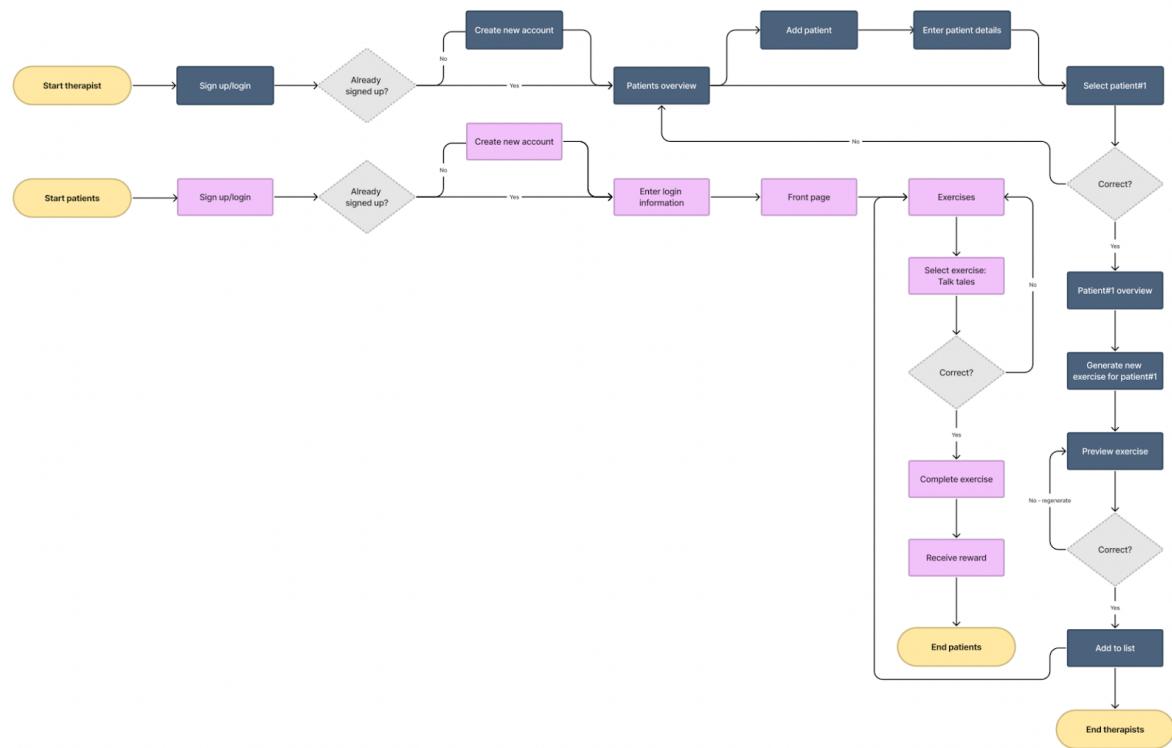


Figure 5.1: User Workflow Diagram depicting the sequential steps and interactions within "ArtyQuest."

5.3 Scenarios of Use

To further understand the interactions with "ArtyQuest," we present detailed user scenarios:

5.3.1 Scenario for Children

A child logs into *ArtyQuest* and chooses one of the available exercises. In the exercise, they are presented with an engaging story tailored to their needs. As they progress, they are prompted to articulate sentences targeting their speech condition. The system provides feedback, visually highlighting any misarticulated or missing words, thereby facilitating a fun and educational experience. (Figure 5.2)

5.3.2 Scenario for Therapists

Monotoring

A speech therapist logs in and sees a list of all his patients. He then chooses to overview one particular patient of interest. On that patient's profile page, he can choose what exercises he'll have available. With real-time access to the child's progress, the therapist can easily plan the sessions accordingly.

Ideation

A speech therapist logs in and sees a list of all his patients. He then proceeds to see his newest patient's page and decides he wants to create a new exercise for him. As such, he goes over to the exercises overview page, where he can proceed to create a new exercise. He types in the condition he seeks to tackle with that exercise and the theme to meet the patient's need. After a story, accompanying image, and phrases are generated, he decides to tweak them to fit the therapy session better. He then saves the exercise and makes it available for the new patient. (Figure 5.3)

5.4 Design Motivations

Our design choices are motivated by the desire to create an engaging, intuitive, and therapeutic environment. The application of conversational AI and multimodal interactions serve to make speech therapy more accessible and enjoyable, thus increasing the likelihood of consistent practice and improvement.

In summary, the UX design of *AstyQuest* reflects our commitment to delivering a solution that not only meets the therapeutic needs of children with speech impairments but also empowers therapists to provide personalized care. Through the use of conversational agents and LLMs, we provide a platform that is both innovative and empathetic, catering to the intricate needs of speech therapy in a digital age.

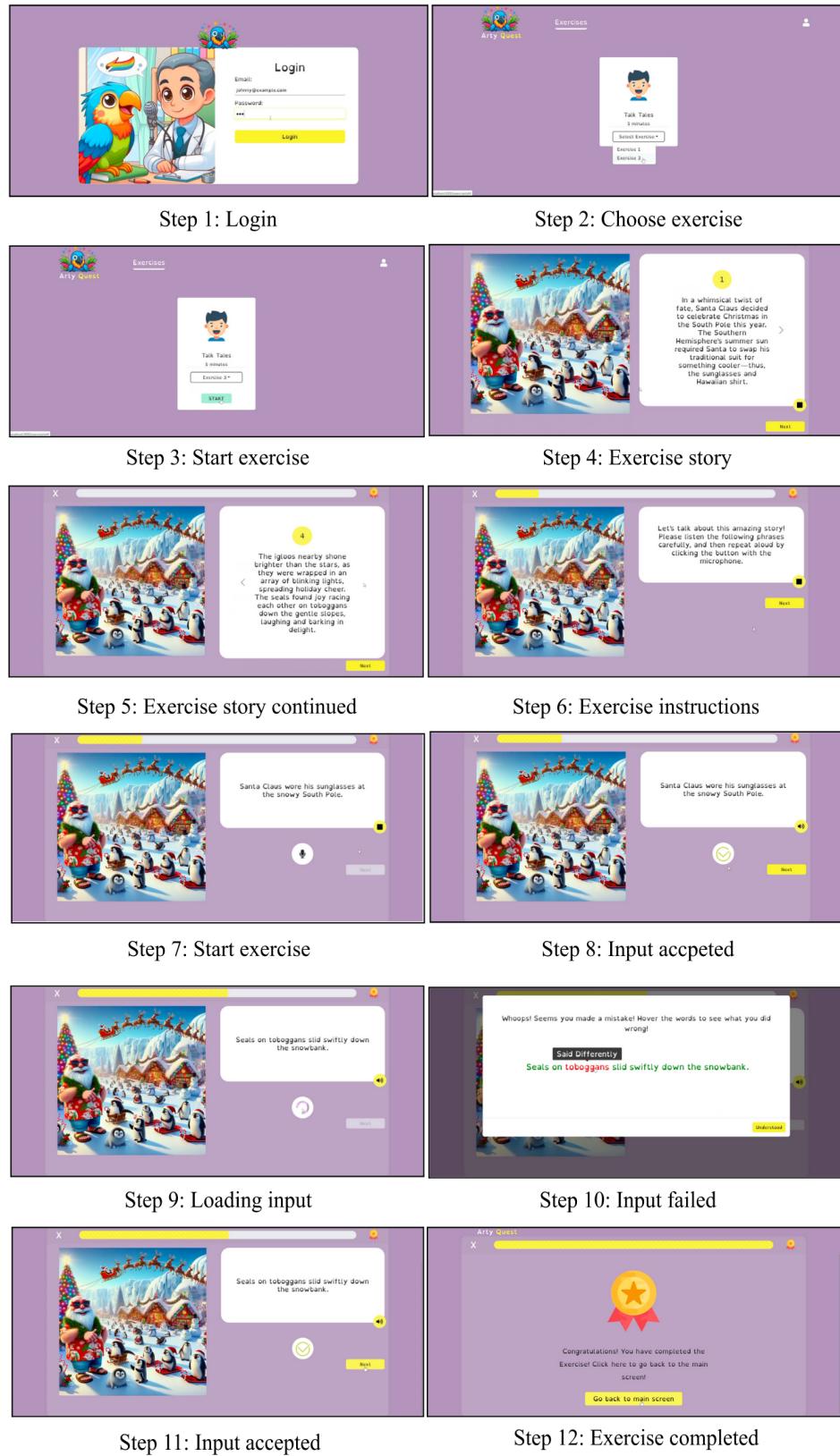


Figure 5.2: User Scenario for Children

Step 1: Login

Step 2: Patient overview

Step 3: Choose patient

Step 4: Exercise overview

Step 5: Add new exercise

Step 6: Generating new exercise

Step 7: Preview new exercise

Step 8: Saving new exercise to list

Figure 5.3: User Scenario for Therapists — Ideation

Chapter 6

Technological Solution

The *AartyQuest* platform is a culmination of modern web technologies and advanced AI integrations, which come together to create a seamless and interactive experience for speech therapy. This chapter describes the general technological approach and the specific hardware and software architectures that constitute the backbone of our application, along with the prompt engineering methods used to interact with generative AI models.

6.1 General Technological Approach

Our technological strategy is anchored in creating a responsive and intuitive application that leverages the latest advancements in AI to support speech therapy. By integrating a conversational interface powered by AI, we offer an innovative approach to speech therapy that is both engaging and effective.

6.2 Software Architecture

The software architecture of *AartyQuest* is a combination of a ReactJS frontend and a Python backend using the Flask framework. These choices represent our commitment to using cutting-edge, yet well-supported technologies to deliver a robust solution.

6.2.1 Frontend with React

The frontend of our application is built using ReactJS, known for its efficiency and the capability to build dynamic user interfaces. It caters to the needs of both patients and therapists,

enabling functionalities such as user authentication, profile management, and the presentation of speech exercises which include a blend of text, audio, and visual elements. The real-time audio evaluation mechanism is a crucial feature that integrates with the backend to provide immediate feedback on speech exercises.

6.2.2 Backend with Flask

The backend is developed in Python and uses the Flask framework to provide a lightweight yet powerful server environment. It consists of several components:

- **Storage Compartment:** Handles the storage of exercises and user profiles, using a local storage system with JSON, WAV, and JPG files.
- **Requests API:** Acts as a conduit between the frontend and backend, managing data transfer and processing user requests.
- **External API Integration:** Integrates with external APIs from OpenAI and Azure to enhance the system's capabilities in content generation and audio assessment.

6.2.3 OpenAI's API Integration

The integration with OpenAI's API is pivotal for generating the text, images, and audio for the therapy exercises. Utilizing models like GPT-4 for text, DALL-E-3 for images, and a specialized model for audio, we ensure the content is engaging and tailored to the therapy's needs.

6.2.4 Azure's API Integration

Azure's Speech Cognitive Services are employed for audio assessment, comparing the patient's spoken audio with the expected outcome to provide a detailed analysis via JSON of the analyzed speech, which is then reflected back to the user through the frontend interface after parsing.

6.3 Exercise Generation Process

The process of generating therapy exercises involves three steps:

1. A story is created using OpenAI's text generation API, informed by prompts that include the therapist's input on goals and themes.

2. Descriptive elements from the story guide the generation of corresponding images via the DALL-E-3 model.
3. Audio for the exercises is produced, completing the set of materials needed for each exercise.

In summary, the implementation of *AstyQuest* synergizes the flexibility of React with the efficiency of Flask, alongside the creative power of AI through OpenAI and Azure. This blend of technologies provides a personalized and effective solution for children undergoing speech therapy.

Chapter 7

Evaluation

In this chapter, we present an empirical evaluation of the project. This evaluation critically examines the application's usability, functionality, and technological robustness, focusing on user experience, performance under various conditions, and the accuracy of content generation.

7.1 Usability and User Experience

During the Demo Madness event, *AartyQuest* received its initial round of usability feedback from adult users without therapeutic expertise. We mainly asked colleagues about their experience with the tool (positive or negative) and to explain what they liked about the tool and what they would see as possible improvements. We collected around 10 distinct answers. The general response was positive, with users expressing satisfaction with the application's interface and navigation. They found the design simple yet effective, fitting well for the target audience of children. The intuitive nature of the application's navigation was particularly appreciated.

A standout feature for users was the audio feedback provided during exercises. The users also appreciated the ease of editing and generating exercises, noting the application's flexibility and creative output on the Therapist component. However, it's important to acknowledge that these users were not actual therapists, which limits the relevance of this feedback. One area for improvement identified was the mechanism for speech detection. Users expected automatic speech detection rather than manual activation, indicating a need for more intuitive voice interaction features.

Overall, the feedback on the satisfaction with the application was largely positive, with minimal negative points. Users suggested additions like enhanced control over exercise generation and better user management. Following this feedback, the feature to regenerate images was added, leaving other suggested enhancements for future development.

7.2 Accessibility, Performance and Reliability

7.2.1 Accessibility

Regarding accessibility, the application's design on the therapist's end does not heavily focus on inclusivity features. However, for the children's user interface, color coding, and tooltips in the error assessment pop-up have been implemented to enhance inclusivity, which the users noted. Different colors were used to indicate various types of speech errors, and the language used in tooltips was simplified for better comprehension by children.

7.2.2 Accuracy of Content Generation

A series of iterative developments marked the journey to achieve better accuracy of content generation. Initially, the focus was on creating roleplay scenarios for speech therapy, guided by a structured prompt designed to generate a narrative involving a character named ARTYQUEST and the patient. The initial prompt sought to create a therapy session disguised as a story, but this approach encountered obstacles. The generated conversations felt unnatural and overly focused on being part of a speech therapy context, lacking creativity and character depth. This initial prompt, after our attempts to fix said problems, was the following:

```
f"""Generate a role play scenario for a speech therapy patient to practice with
    ↵ a bot named ARTYQUEST. Be creative and fit the description to one or more of
    ↵ the patient's interests (randomly selected), which include: ',',
    ↵ '.join(self.interests)}. Stick to the following format, where the NARRATOR
    ↵ field is optional according to the story, may be replaced with another name,
    ↵ and they don't acknowledge that it is a therapy session, they only serve to
    ↵ narrate the story:
NARRATOR: ---
PATIENT: ---
ARTYQUEST: ---
(repeat)"""
```

As an example, when using the interests football, tennis, this role-playing exercise would be outputted:

NARRATOR: Our story unfolds on a sunny afternoon at a local tennis court, where the patient, Alex, a passionate tennis enthusiast, meets ARTYQUEST for a friendly game.

PATIENT: Hey there, ARTYQUEST! Ready for a game of tennis?

ARTYQUEST: Absolutely, Alex! I've been brushing up on my tennis skills. Let's rally and have some fun.

NARRATOR: The patient, Alex, serves the first ball with enthusiasm, starting the match.

PATIENT: Here comes the serve!

ARTYQUEST: (responding to the serve) Nice one, Alex! You're really getting the hang of this.

NARRATOR: As the game progresses, ARTYQUEST provides feedback and engages in tennis-related conversation with Alex, helping improve both speech and tennis skills.

PATIENT: (during a rally) So, ARTYQUEST, have you ever watched a professional tennis match?

ARTYQUEST: (continuing the rally) Oh, yes! I've seen some exciting matches featuring players like Roger Federer and Serena Williams. They are incredible athletes.

NARRATOR: The game continues, with both Alex and ARTYQUEST enjoying their tennis session and working on speech therapy goals without even realizing it.

PATIENT: (celebrating a point) Yes, I got that one!

ARTYQUEST: (encouraging) Great job, Alex! Your game and **speech** are improving with every point.

NARRATOR: The match goes on, providing a fun and engaging way for Alex **to practice speech therapy** while indulging in their love for tennis.

As we can see from this example, while it refrains from mentioning speech-therapy for most of the exercise (which wouldn't happen many times), it still does so at the end of the exercise. In our view, this would break the immersion of the user in the exercise.

Recognizing these issues, we pivoted to a different exercise format. The revised approach revolved around generating a story based on an image, with the patient required to repeat specific phrases aloud. This new format led to the development of a second prompt. This second prompt directed the AI to create a story with simple language from a given exercise focus and

patient interests, which the therapist would define. This change significantly improved the engagement and relevance of the exercises. The final prompt structure used was:

```
f"""Generate the description of an image suited for a very creative story for a
    ↵ speech therapy patient to describe, with various objects and/or characters
    ↵ to describe.

The patient will have to repeat aloud at least 10 phrases. The first 6 phrases
    ↵ should be a statement about the image. The next 4 should be tongue twisters as
    ↵ compound nouns that emphasize the patient's difficulties, and reflect the image
    ↵ description.

The phrases should tackle: {speech_focus}. Try to include the patient's
    ↵ interests, such as: {interests}.

Do not mention the patient's difficulties in any of the phrases nor in the
    ↵ story. The language of the story should be simple and easy to understand. The
    ↵ story should not focus on describing the image, but rather on telling a story
    ↵ that includes the image.

Output the detailed description and story containing all necessary info for the
    ↵ phrases; output those same phrases as well. Make the description brief, while
    ↵ keeping all relevant information. Strictly follow this template:

DESCRIPTION: --
STORY: --
PHRASES:
1:...
2:...
3:...
4:...
5:...
6:...
COMPOUND NOUNS:
1:
2:
3:
4:"""
```

We underwent several iterations of the prompt to refine the exercise content. A key realization was the need for simplicity in language, as some phrases turned out to be overly complex. Incorporating the directive, 'The language of the story should be simple and easy to understand,' made a noticeable difference. Additionally, changing the focus of the story from directly describing the image to incorporating the image into a broader narrative ('The story should not focus on describing the image, but rather on telling a story that includes the image') made the exercises more engaging and resolved issues encountered with our initial approach. Ensuring the prompt strictly adhered to a set template also helped in achieving reliable and consistent

outputs suitable for parsing.

Regarding image generation, we observed interesting nuances. With DALL-E-2, using either the description or the story as input didn't significantly impact the quality of the generated images. However, upon upgrading to DALL-E-3, which produces higher-quality images, we found that using the 'story' as input was impractical because it often resulted in the image containing text elements. Consequently, we opted to use the 'description' for image generation prompts to avoid this issue.

As an example of our solution, for the speech focus "**'t' sounds**", and with the interests being "**the spaceman who read books**", the output of the GPT-4 model would be the following:

DESCRIPTION:

A spaceman in a shiny silver suit stands next to a towering rocket on the red Martian surface. In his gloved hand, he holds an open, antique book with golden page edges that glint in the sunlight. Scattered around his feet are unusual Martian rocks, metallic and sparkling. In the background, a trio of alien creatures with turquoise tentacles peek curiously from behind a space rover with twinkling lights and antenna. Overhead, a brilliant Earth hangs in the dark sky amidst twinkling stars.

STORY:

On Mars, where the dust whispers tales of the cosmos, there was a spaceman named Tom who embraced adventure and knowledge. Unlike other astronauts, Tom had brought his most treasured antique book, 'Tales of Tenacious Trailblazers,' to read in the quiet moments between explorations. On a peaceful Martian afternoon, as the orange sun dipped closer to the horizon, Tom read stories to the local three-eyed aliens. They were entranced by the tales of daring and discovery, their tentacles softly swaying to the rhythm of his voice. Tom, in return, had grown fond of his peculiar audience, often sharing tidbits about Earth and its wonders. Their gatherings had become an intergalactic tradition, every time beneath the gentle gaze of Earth, cultivating an extraordinary friendship bound by stories and the stars.

PHRASES:

- 1: The spaceman stood still on the Martian terrain.
- 2: Tom's silver suit shimmered under the sun.

- 3: His fingers flipped through the antique book's pages.
- 4: The curious Martian rocks lay near his boots.
- 5: Turquoise tentacles timidly touched the rover.
- 6: Above, Earth shined like a celestial beacon.

COMPOUND NOUNS:

- 1: Time-tested trailblazer tutored tentacled trackers.
- 2: Tentacle-touched technology twinkled teasingly.
- 3: Tough-traveling Titan toured the terrain.
- 4: Tome-toting traveler told tremendous tales.

As we can see, this exercise captures the goal perfectly: it talks about a spaceman who reads a book, 'Tales of Tenacious Trailblazers', and keeps its focus very well on 't' sounds — for example, the book name or the **COMPOUND NOUNS**. The generated image by DALL-E-3, with the description field as input, also captures very well the situation at hand:



Figure 7.1: The generated image for the exercise — *A spaceman in a shiny silver suit stands next to a towering rocket on the red Martian surface...*

7.2.3 Audio Evaluation Mechanism

The audio evaluation mechanism stands out as one of the most effective features of the application, though we identified two primary concerns in its implementation.

The first issue relates to the 'Accuracy' metric of Azure's recognition system, which might

be overly stringent for our application's context. For instance, a 60% accuracy rating could indicate either a slightly mumbled but correct pronunciation or a word with clear mispronunciations. Fine-tuning this metric to suit therapists' future preferences proved challenging due to our development team's limited expertise in speech assistance. Additionally, the system's performance appeared to diminish in noisy environments, suggesting the need for quieter settings to ensure optimal functioning.

The second issue is associated with the alignment algorithm used in the assessment. This algorithm attempts to match the user's spoken words with the expected text, sometimes leading to inconsistent evaluations. For example, the system might interpret stuttering as repeated words, resulting in erroneous assessments. Moreover, the identified errors might not perfectly align with what the user actually said but rather with the closest match to the expected text. Despite these issues, the feedback mechanism provided by the audio evaluation is invaluable. The benefits of this mechanism, such as its ability to pinpoint specific speech errors, significantly outweigh the drawbacks and enhance the overall user experience with the application.

7.2.4 Application Reliability

The application's performance heavily depends on a stable internet connection and the availability of OpenAI's API for exercise generation and Azure's API for exercise execution. In situations where the APIs are at full capacity, exercise generation and execution might be hindered. Aside from this dependency, the application showed no significant reliability issues.

7.3 Integration and Compatibility

The application achieves seamless integration with both OpenAI's and Azure's APIs, presenting a transparent experience to the user where the underlying connections are imperceptible.

However, integrating Azure's API posed significant challenges. A notable amount of development time was dedicated to resolving issues related to audio format compatibility. Initially, we attempted to use Azure's real-time assessment API but faced obstacles that prevented its successful implementation. Consequently, we reverted to using Azure's standard assessment API, which relies on comparing the user's spoken audio with the pre-defined 'ground truth' text.

The solution to the audio format issue involved a multi-step process. We temporarily stored the audio file and then converted its format to WAV before feeding it into Azure's API, as this

format was found to be compatible with the API’s requirements. While this workaround was effective, it was not the ideal solution we had hoped for.

Furthermore, we discovered this solution worked reliably on Windows-based systems but encountered compatibility issues on Mac laptops. The exact cause of this discrepancy—whether it relates to Azure’s API or the audio file format used—is still unclear. This platform-specific challenge highlights an area for future technical refinement to ensure consistent functionality across different operating systems.

Chapter 8

Value Proposition

The *AartyQuest* application is designed with a clear vision: to transform the experience of speech therapy for children and empower therapists with innovative tools. This chapter outlines the unique value proposition of our application, highlighting how it stands apart in delivering personalized, engaging, and accessible speech therapy solutions.

8.1 Personalized and Engaging Exercises

The Speech Therapy App for Children offers personalized exercises generated by AI algorithms based on each child's unique interests. This tailored approach ensures an engaging and motivating therapy experience, fostering active participation. By integrating the child's preferences, the app provides an enjoyable learning environment, making speech therapy effective and personalized.

8.2 Child-friendly UX-UI

The app boasts an intuitive and colorful user interface, purposefully designed for children. This thoughtful design creates a comfortable and enjoyable environment during therapy sessions. With easy navigation and interactive features seamlessly integrated, the interface not only encourages active participation but also makes the learning experience more enjoyable for young users.

8.3 AI-Powered Exercise Generation

The cutting-edge AI technology meticulously tailors exercises based on a child's unique needs, ensuring precise improvement in speech and language skills. The app's adaptable content evolves in sync with the child's progress, delivering a dynamic and highly effective learning experience. This dynamic adaptation ensures that the exercises remain aligned with the child's evolving capabilities, maximizing the impact of each session for comprehensive skill enhancement.

8.4 Accessible Anytime, Anywhere

The app provides flexibility for therapy sessions to take place at any time and anywhere, making it convenient for both therapists and parents to integrate speech exercises into the child's daily routine.

8.5 Empowering Therapists

The app empowers therapists with intelligent insights into a child's performance, facilitating the development of personalized therapy plans. Therapists can then focus on delivering quality guidance and support, knowing the app augments their efforts with tailored exercises. This collaborative approach ensures an effective and comprehensive speech therapy experience.

Chapter 9

Discussion and Future Work

9.1 Critical Reflection

The journey of developing the "ArtyQuest" Speech Therapy application has been both challenging and enlightening. Integrating AI into the application has proven successful, as evidenced by the positive feedback on its child-friendly design and intuitive user interface. However, this project's complexity cannot be understated. The need for continual therapist input and the hurdles encountered in refining content generation and speech detection mechanisms have been significant challenges. Moreover, the compatibility issues with Azure's API underscored the importance of rigorous testing and transparency in our development process.

Despite these obstacles, "ArtyQuest" stands as a testament to the potential of technology in addressing the needs of speech therapy. Our solution has shown real promise in meeting the requirements of our target group, and there is potential for broader application in the future.

9.2 Future Directions

The path ahead for the "ArtyQuest" Speech Therapy application, while not pursued by the current team, opens up a myriad of possibilities for future development and innovation. These potential directions are outlined as guidelines based on the current landscape of technology and user needs.

9.2.1 Short-Term Enhancements (0-6 months)

In the short term, the focus could be on enhancing user feedback mechanisms. Implementing in-app surveys and user reviews would be essential for gathering valuable insights into user experiences. Additionally, the establishment of a customer support channel could address immediate user concerns and queries, thereby improving overall user satisfaction and engagement.

9.2.2 Mid-Term Development (6-12 months)

The mid-term strategy could involve deeper user engagement and collaboration with professionals in speech therapy. Conducting detailed user surveys and interviews would provide a clearer understanding of user preferences and expectations. Collaborating with speech therapists could ensure that the exercises remain aligned with emerging trends and best practices in speech therapy. Analytics could also play a crucial role in refining the app's content strategy to better meet user needs.

9.2.3 Ongoing Refinement

Continual refinement of the app would benefit from a responsive customer support system and regular analysis of user data. This process would help in identifying areas that require improvement or expansion. Collaboration with user experience (UX) experts could further enhance the usability of the app, making it more intuitive and user-friendly.

9.2.4 Long-Term Vision (12+ months and beyond)

Over the long term, potential advancements could include forming partnerships with renowned speech therapy institutions. This would help in staying abreast of industry advancements and integrating them into the app. Investing in research and development to enhance AI capabilities could lead to more sophisticated and varied exercise generation. Additionally, exploring cultural nuances and linguistic diversity might enable the customization of exercises for different regions, potentially broadening the app's global appeal and applicability.

9.3 Concluding Thoughts

In conclusion, the "ArtyQuest" project, though met with challenges, has laid a strong foundation for an effective and comprehensive speech therapy experience. Our comprehensive

plan, ranging from foundational exercises in the short term to global collaboration and AI advancements in the long term, positions "ArtyQuest" as a dynamic and evolving solution in the field of speech therapy.

Bibliography

- [1] M. P. Luppi, F. Nizzoli, G. Bergamini, A. Ghidini, and S. Palma, “Speech therapy rehabilitation,” *Acta otorhinolaryngologica Italica*, vol. 30, no. 5, pp. 244–247, 2010.
- [2] P. Enderby and J. Emerson, “Speech and language therapy: does it work?,” *The BMJ*, vol. 312, no. 7047, pp. 1655–1658, 1996.
- [3] “Childhood apraxia of speech.” <https://www.asha.org/practice-portal/clinical-topics/childhood-apraxia-of-speech/>. Accessed: 2024-01-11.
- [4] L. Furlong, M. E. Morris, T. Serry, and S. Erickson, “Treating childhood speech sound disorders: Current approaches to management by australian speech-language pathologists,” *Language Speech and Hearing Services in Schools*, vol. 52, no. 2, pp. 581–596, 2021.
- [5] “Speech therapy.” <https://my.clevelandclinic.org/health/treatments/22366-speech-therapy>. Accessed: 2024-01-11.
- [6] D. Popovici and C. Buică-Belciu, “Professional challenges in computer-assisted speech therapy,” *Procedia - Social and Behavioral Sciences*, vol. 33, pp. 518–522, 2012.
- [7] “Adult speech therapy: Who it helps, and why it’s worth it,” January 26 2022.
- [8] E. Seder, “Speech therapy for children: What are the benefits?,” May 18 2023.
- [9] T. Felske, S. Bader, and T. Kirste, “Automatic generation of personalized and context-dependent textual interventions during neuro-rehabilitation,” *KI - Künstliche Intelligenz*, vol. 36, no. 2, pp. 189–193, 2022.
- [10] Z. Liang, W. Yu, T. Rajpurohit, P. Clark, X. Zhang, and A. Kaylan, “Let gpt be a math tutor: Teaching math word problem solvers with customized exercise generation,” *arXiv*, 2023.

- [11] G. A. Attwell, K. E. Bennin, and B. Tekinerdogan, “A systematic review of online speech therapy systems for intervention in childhood speech communication disorders,” *Sensors*, vol. 22, no. 24, p. 9713, 2022.