Storyteller Bot V1: Rhino Reader
Bridging the Third Grade Literacy Gap

Aneekah Uddin
Adviser: Radhika Nagpal

Abstract

This paper details the design, development, and evaluation of a computer-vision-powered product, Storyteller Bot, to help students learn how to read. The literacy gap disproportionately affects Black, Hispanic, differently abled, first-generation, low-income, and English as a Second Language students, thus affecting their future success. In response, the device uses computer vision, other ML models, and kid-friendly design patterns to power a stuffed animal that can read to students in real-time or from an audio book. This application aims to bridge the widening literacy gap, by reinforcing literacy skills and engaging in storytelling to children while maintaining accessibility and inclusivity in design practices.

1. Introduction

82 percent of students eligible for free or reduced lunches do not read at or above proficient levels by the time they enter fourth grade. Education studies experts consider meeting third-grade reading level benchmarks to be the most crucial milestone for academic success [12]. By falling behind in reading, a child risks falling behind in every single subject– children are no longer "learning to read," they are "reading to learn." [15] Failing to meet this milestone results in students dropping out of high school and unemployment. Reading is an essential part of participation in society yet, is often framed as a privilege.

If 82 percent of students from lower socioeconomic statuses fail to meet this benchmark, society is choosing to accept that 82 percent of people from lower socioeconomic backgrounds shouldn't have access to "success," or influence society. The NAEP (National Assessment for Educational Progress) has found that Black, Hispanic, English as a Second Language, and differently-abled students have much lower literacy rates than their peers. [14] It's unacceptable that our system of education does not support these demographics.

Though I was eligible for free lunch in fourth grade, I was fortunate enough to be part of the "18 percent" that was meeting literacy benchmarks. I attest this to my upbringing. From an early age, my mother ensured that I understood the importance of reading. She diligently read to me every day and as I got older, she helped me practice the skill of reading. She would help me sound out the words. And she would progressively bring more challenging books from the library. I wouldn't have the English skills I have now nor would I be attending a leading university today without her support.

But not every child has the privilege of a parent who can read, has time to spend, or has access to resources. Yet this extra support can be transformative to a child's trajectory and support their literacy journey. While there are numerous other reasons why a student might fail to meet literacy levels, such as undiagnosed learning disabilities, hearing or vision loss, etc., extra support employing reading techniques learned in the classroom promotes literacy.

I plan to build a technological device that can help children learn how to read and address the literacy gap, especially for students of different socioeconomic statuses and other marginalized identities. Current solutions for literacy are commercialized with large sticker prices. Nontechnological alternatives including private tutoring, are costly and not always frequently available. Because of the current landscape's challenges, I am designing a solution that will not require reliance on others, allowing children to explore reading on their own, and engage in storytelling.

Even well-intentioned, technological solutions can cause more harm than good. Technology has powerful implications if used properly– leading to the emergence of intersectional, critical theory on how to build technology for our future. Some key scholars include Sasha Constanza-Chock, author of "Design Justice," [13] and Professor Ruha Benjamin, author of "Race After Technology" and "Viral Justice." These frameworks and perspectives have shaped how I approach the design process and what I choose to create. Following with the seminar's theme of reimagining robotics through different art forms, I hope to use the art form of storytelling to redefine what "robotics" means. By deeply considering how to design the solution with principles of justice and equity, I hope this project has an impact on increasing literacy rates for children.

2. Background and Related Work

This project requires interdisciplinary, research from educational studies, product design, computer science, etc. In this section, I outline relevant research in these fields and dive into the current landscape of solutions and alternatives.

2.1. Understanding the Literacy Gap

The NAEP has identified a major literacy gap across different demographics, demonstrating that there is a lack of support for those students. However, what causes these gaps? Some educators point to the rising use of technology as a form of entertainment for children, detracting from reading books which would ultimately strengthen reading skills. Many others state that students are not proficient in "phonics" or educators aren't required to teach phonics skills in early elementary school [14]. Phonics is a technique used to teach reading by having students "sound out" the word, allowing them to "decode the word." Reinforcing phonics as a practice is critical for students to be able to learn new vocabulary. Often times, students who struggle to properly decode will guess words based on the first letter of a word and the picture provided [9]. This demonstrates strong language comprehension. Language comprehension is the act of understanding a word, sentence, or phrase in isolation and in context. While language comprehension is critical to reading comprehension, the skill alone is not enough. In addition to strengthening word decoding, students must sharpen language comprehension to meet reading comprehension benchmarks. If struggling readers do not have explicit intervention where they are taught decoding skills and language comprehension, their reading skills will be slow to improve and might not reach a satisfactory level for their grade level.

2.2. Importance of Storytelling and Representation in Storytelling

Stories are often compartmentalized as a creative endeavor. But they are part of our everyday lives. It's how we entertain ourselves. It's how we express ourselves. It's how we recount and experience life. Our stories impact how we carry ourselves in the world.

Representation in storytelling is critical for empowering children. Historically, marginalized identities do not see themselves in stories– whether that's media, movies, books, theater, etc. When these identities have been written in, they are often sidekicks, villains, or stereotypically caricaturized. Recently there has been more of a push to include stories that celebrate difference and diversity in all forms of media, including children's books. These stories are inspirational to people who hold these identities as they are able to relate to the characters. Additionally, children are more likely to be attached to stories that showcase their identities as they may see themselves represented on a page, thus encouraging them to read and engage in storytelling more.

Without understanding how to craft stories, people, often those from marginalized identities, fall into what Chinua Achebe describes as the danger of a single story. Their perspectives are erased from the narrative, thus impeding a legitimate dialogue. While stories are important for students to engage in learning to read, their power in our society is so much more.

2.3. Kid-Centric Design

Designing a universal device for all children captures a large audience but often neglects children who may need the most support. Universal Design for Play Toys Guidelines oversimplify accessibility guidelines through vague, incomplete understandings of accessibility [11]. When designing for children, there are several factors that must be accounted for, to make the device inclusive and accessible, especially for children of different abilities. There are several ways to incorporate this intention into the physical design [3], including:

- **Sensory Exploration** - Students with different abilities often struggle with engaging their senses in different ways. Adding a range of textures, colors, and noises can stimulate sensory exploration.
- **Fine Motor Skills** - Toys such as building blocks, manipulative puzzles, etc., that require specificity in movement help sharpen fine motor skills.
- **Social Interaction** - Games or toys that allow for a multiplayer option or promote some mode of social interaction will help children develop social skills and deeper connections with their peers.

- **Assistive Technology** - Having specific features built into the product to support needs of children with different abilities allows them to engage more meaningfully with toys and feel supported.
- **Personalization** - Going beyond specific features, personalization accounts for ability and preference, among other relevant factors, to generate a custom experience for the child.

Designing products using these tactics has the potential to create an inclusive product. However, extensive testing and understanding the needs of users are necessary to design a final product that truly addresses the gap. Given the scope of the project, an initial set of assumptions will be made to help develop the physicality of the device, which will be tested.

2.4. Relevant Solutions

Figure 1: Relevant Academic and Commercial Solutions in the Landscape

An academic project that stems from a similar motivation is MIT Media Lab's Companion AI. This project began over 5 years ago. Companion AI is housed in an abstract, colorful, fuzzy robot and develops personalized games to help children at this benchmark, develop literacy skills [18]. The team is focused on personalizing the AI to teach children reading and other early education skills. While playing games can be engaging, the project doesn't highlight aspects of storytelling– a large portion of early reading that promotes creativity. I'm hoping to address that gap in my project.

In the current commercial landscape, I found a few alternatives: Luka Reads, developed by a Singapore-based Company, is an alien robot that can read select titles in English and Chinese. The bot scans the books and seems to play an audiobook. Luka can also provide oral stories from over 70K+ titles. During "parent" mode, a parent can type things for Luka to say into the companion app.

Luka will read these out in real time to their child, encouraging them to complete activities such as brushing teeth or chores. Additionally, the company offers a “pen” product that can be used to read on the go [5]. While this product does engage children, especially as a storytelling device, it falls short in actually reading alongside the child since it seems to access audiobook titles and can only read a select number of them. Further, it is over $300 for the bot, which is an unmanageable price point for many families.

Similarly, there is Codi Bot, a robot with preprogrammed audio file tracks of classic stories. These are quite limited and only in English [2]. While the price point is lower, its value add does not match the price, making the product feel like a “luxury” product.

There are a few other, commercially available AI Companions that are designed for children. Miko[6] and Moxie[7] are two examples. These however are meant to engage children in games and do not focus specifically on literacy or education. These bots solely operate in English, which misses out on a wider audience.

Beyond hardware-based alternatives, reading software is often used in classrooms to promote literacy skills. These are online or desktop software platforms where students can practice reading skills[4]. These often test reading comprehension by different grade levels.

Other nontechnological solutions that address the problem of closing the literacy gap include private tutoring, early reader programs at libraries and schools, etc. These are all forms of human labor– that can become costly for families and may not be frequently available.

Despite all of this prior work, the current landscape isn’t built to address the core problem of the literacy gap across different marginalized identities, while maintaining a commitment to “storytelling.” Additionally, many solutions create a “tech-centric” product design– housing the technology in cartoonish robots– which is unaligned with what kids need. Further, cost and accessibility have not been a major concern of discussion. For my independent work, I hope to address these missing concerns in the current landscape.

3. Approach

To approach designing a solution to address the literacy gap, I planned to develop physical prototypes. This approach has a more effective end product than just researching the related academic landscape and/or drawing up blueprints because I may discover new realizations in the hacking process. The first version of Storyteller Bot, named Rhino Reader, is a stuffed animal (a rhino) that houses a wireless camera and speaker. A phone or laptop controls the hardware. The stuffed animal, using CV, ML models and the speaker, will read to the child any book placed in front of the rhino's eyes. I chose this to be a first iteration for several reasons:

- **Housing the Device in a Stuffed Animal** - Stuffed animals are known for promoting empathy, developing socio-emotional skills, strengthening communication and decision-making, and assuading anxiety. Especially for students with different abilities, stuffed animals promote social interaction and provide a safe sensory experience [10].
- **A Computer Vision System** - The device is tailored to reading and reacting to what is presented in front of it, providing a sense of personalization, while still allowing for storytelling.
- **Cost** - While the current iteration might be at a much higher price point than I originally intended, this can be refined after a proof of concept.

These reasons begin to address the gaps in the current landscape.

3.1. Prototyping Plan:

While interdisciplinary research may impact the design, there are three major facets to building a prototype. Each of these parts has many questions that will be answered through attempting implementation.

- **Developing Hardware Loop:** This includes identifying the correct technology, setting up the hardware to see how to best position the camera and speaker, how to hook up wires, and how to encase all of this inside of the product.

- **Developing Software:** The software is the major part of this project and requires multiple substeps, including:
  - *Connection between hardware and software:* How to connect the software system and hardware? Could I develop a "phone app" or should I run the program in the mobile browser or connect it to a laptop for the first prototype? How would I be able to access the camera and the speaker easily?
  - *Software Framework:* This includes creating a chart of how the child should interact with Rhino Reader, from the moment it turns on to every single one of its features. What decisions can the child make? What is required of their parent? What design patterns support inclusivity? This will serve as a blueprint for the MVP of the project.
  - *Computer Vision:* There are numerous ways of how to integrate computer vision into the software system. Does the system use real-time image or video data or does the bot take images and process them? How does the bot recognize text? How does it "find" the book? This will include research into what APIs are available and what might need to be developed on my own.
  - *Speech:* There are multiple times when Rhino Reader is in conversation with the child who is reading. How should the bot respond to the child? How can we add intonation and other vocal techniques to the speech? This will include research into what APIs, tools, or ML models are available. If the bot listens and responds to the child, how do they parse what is being said, especially with students of different speech abilities or languages?
  - *Further Extensions:* How can we personalize the literacy experience for every child? How can we include aspects of oral tradition? How can we bring a library to children who might not have books? What role does translation play in this? How can we encourage reading and reading comprehension?
- **The Physical Product:** Building the physical product consists of combining the software and hardware loop into a final product that is kid-friendly and kid-empowering by design. Some questions that might guide this part of the journey are: How do we put the product together in a way that encourages kids to connect to it? How do we "kid-proof" the device? What other

intersectional identities can we consider in the design of the product?

Using this plan and approach, I will develop a prototype or proof of concept of my project to address the widening literacy gap, across different marginalized groups, at a young age.

To determine whether my project has reached the stated goal, I will look at both the research-based design ideas and the prototype. My design success is determined on building a working, proof-of-concept solution and demonstrated potential on further iterations– enough for another student to continue the project. Further, to quantitatively assess this project, I will develop a cost breakdown scheme to determine whether this is a reasonable threshold for schools, libraries, or parents to provide.

4. Implementation

I used the three-part plan outlined in the approach to implement my solution.

4.1. Connecting Hardware and Software

Following my approach, I began with understanding how to connect the hardware and software components, to see in which ways my software implementation would be limited. I settled on using Python since it would be easy to use with APIs and could connect well to OpenCV and other Machine Learning libraries. Using online tutorials I found different ways to connect the camera to the code and receive data from it. Based on how the data was collected, the best implementation was using the VideoCapture method in OpenCV's library. This reads from video and captures a "frame" of data when called. Given how often frame data is captured, this can imperfectly simulate real-time video.

Python has many different ways of playing sound, but if given an audio file, it can use the OS library to play the file aloud. The program cannot exit during the playback of an audio file, or else the file is not played to completion.

Regardless of wire or wireless devices, connecting to the main program is the same. For MacBook users with iPhones, the two can easily connect in a way that an iPhone can be used as a wireless camera.

I chose to keep the first iteration of the software on a program running on my laptop, however, in future iterations, creating a separate brain for the robot or housing the program in a "Companion App" on smartphones could simplify the connection between hardware and software.

4.2. Software

The software design and implementation are a large component of Rhino Reader. Before implementing anything, I researched what APIs, tools, or libraries were available for different aspects of the project, such as translating image data to text and text data to speech.

4.2.1. Tools to Translate Image Data to Text Possibilities

- **OpenCV** has a few relevant libraries for character recognition. Many of these are older techniques and did not have a great level of accuracy. Many children's books have different types of font characters, which makes it difficult to use some of these libraries that aren't trained to recognize differently designed characters.
- **Hugging Face Models** is a huge open-source collection of different machine learning models, many of which are developed by research labs. There were a few possible models (MoE LLaVA, LLaVA, MiniGPT V4) that were designed by academics to replicate their commercially named counterparts. The models had "demos" which allowed users to test the model for accuracy. I could submit my own photos and ask a question to the model. After a few test cases, the best-performing model was LLaVA. Since these all have Gradio APIs, I could easily connect this to my code.
  - *MoE LLaVA* [16]
  - *LLaVa* [17]
  - *MiniGPT V4* [19]
- **GPT 4 Turbo** [8] is another reliable vision language model. There is no online demo so there was no possibility of testing without purchase. Since there were free alternatives, I decided to return to this possibility if the CV capabilities needed upgrading. With that in mind, I chose to modularize my code to allow for easily changing between models, especially since this field is rapidly changing.

- **PyTorch/ML Libraries** had a few possible vision models such as the PyTesseract library and MMOCR or other OCR models. These did not have great performance on a few sample tests.
- **Google Cloud Vision API** is the technology that underlies the Google Translate image processing use case. Given the complexity of understanding different characters in several languages and in images (not necessarily standard text), this use case was most aligned with my project's use case. It would allow me the possibility to expand the project to different languages. While it is a paid option, the free trial would cover enough for testing and demos. Additionally, the API was incredibly simple to set up, with extensive documentation, unlike Gradio APIs that power the Hugging Face Models. Further, Google's APIs work more naturally together, without much transformation of data, allowing for more possible features.

Because of the high success rate, ease of setup, and the possibility of expanding to different languages, I chose to use Google Cloud Vision API to translate image data captured into text data.

4.2.2. Tools to Translate Text Data into Speech

- **gTTS** is a Python library that converts text into mp3 files using the standard OS voice on a device. The voice sounds very robotic and has no specific intonation.
- **Commercially Available Text-to-Speech Models - Open AI, Google, etc.,** all have models that translate text into speech, adding "human" sounding qualities to these voices. However, since this is a "newer" problem than vision, there are less free models available to the standard of quality necessary. Further, there is lots of controversy surrounding how these voices are generated and trained, especially if these larger tech companies have given credit to the voices that the products are trained on.

Rhino Reader uses the gTTS library but will transition to Google's Text-To-Speech API to improve the quality of storytelling since storytelling is a main value of the project. However, if more ethical model options are possible, this would more strongly align with principles of design justice and ultimately the project.

Further, more research will be done to determine whether there are certain voice qualities that are friendlier for children to engage with. A feature that provides optionality in the types of voices available could be built into the project.

4.2.3. File Structure After determining the relevant tools, I devised the following map to structure my code and determine where decisions from the user would be made. With the exception of dark

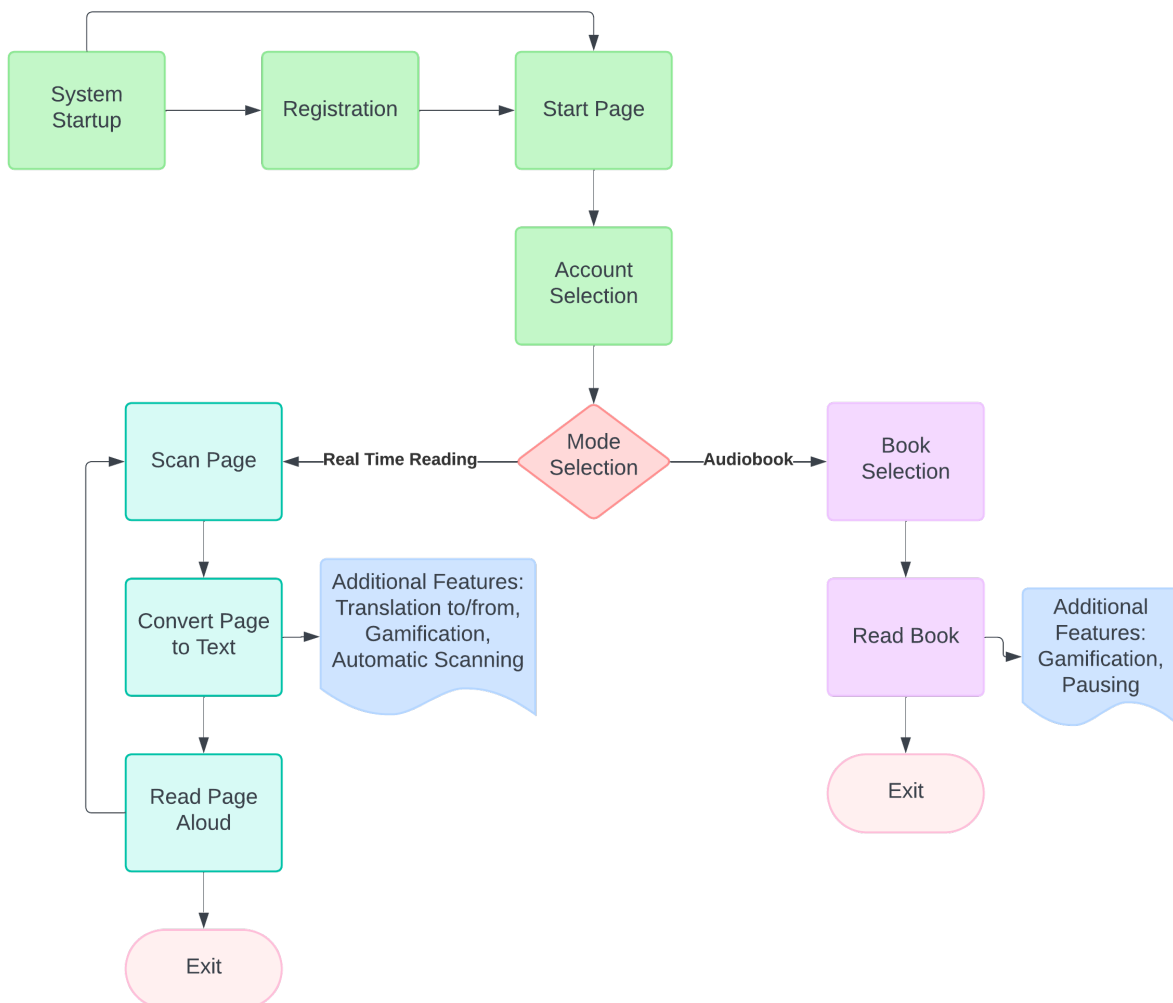

Figure 2: Rhino Reader System Framework Map

blue and light red, each color represents a different file in the program.

Green handles the logistics of setting up and starting up the program. When the bot is first run, this guides the user through account creation and basic setup (any logistical information that might be relevant in the future – accessibility needs– or preferences – voice types, etc.) While the option to add another account will remain for future uses, the bot will default to the original account in the future.

Red represents the "home page" of the application after a user has created an account/logged in successfully. The user can choose how they want to engage with the bot.

Purple represents an option where the user can listen to an audiobook. Upon clicking this, they can pick from a list of available options to hear a pre-recorded audiobook. This promotes oral tradition in storytelling and increase language comprehension.

Light blue represents the use case where the user presents a book to the bot to read aloud. In the first iteration, user input will trigger image data capture. The image is processed and read aloud to the user. This is repeated until the user finishes reading.

Dark blue represents the potential features that could be built into that portion of the project. The steps that are outlined are preliminary steps for the first iteration. With proof of concept, more features to expand the user experience will be added.

Light red demonstrates moments where the program exits.

In addition to these files, there are extra files that store relevant data (reused sound clips, audiobook files) and modular code (helper functions for easy exchange of ML models).

4.3. The Physical Product

The original implementation of the product included taking different types of stuffed animals and adding a camera and a speaker inside the animal. While this is possible, it became challenging to find a good spot on the stuffed animal to house the camera, where it could easily see the book. This would change drastically with any type of animal and would require a way of repositioning the camera and adjusting how the code reacted to the different angles. Additionally housing the camera inside of the animal felt intrusive and took away from the "cuteness" of the stuffed animal. Further, it felt techno-omniscient to house a camera inside of a toy that children often trust. In the case of a "hack," children's bedrooms and safe spaces would be at risk.

In order to address all of these problems, instead of placing the webcam into the stuffed animal, I added it to a headband that the stuffed animal can wear. Depending on how the kid holds the stuffed animal, they can place the webcam headband in a specific position. This also allows them to move

Figure 3: Rhino Reader Prototype

it to different stuffed animals, creating different characters. When not in use, the webcam headband can be stored in a different place for safety and safekeeping.

With this change, adjusting the positioning of the camera became much easier and allowed for more degrees of flexibility with the size of the book and its placement. While testing the product, I centered stories of diverse and different identities. I tried to strike a balance between "classic" stories and these stories. I found many titles available in the Princeton Public Library and in Firestone's Children's Library, including Princeton Professor Dr. Carolyn Choi's Intersectional Allies.

5. Evaluation

To evaluate my device, I considered both qualitative observations (ease of use, addressing the problem, inclusivity and accessibility, and kid-centric design) and quantitative observations (cost). There are far more qualitative measures since the values that lead this project are not numbers-based and the idea is in the beginning stages of development.

Due to the time constraints of the seminar, the current implementation is a prototype of the real

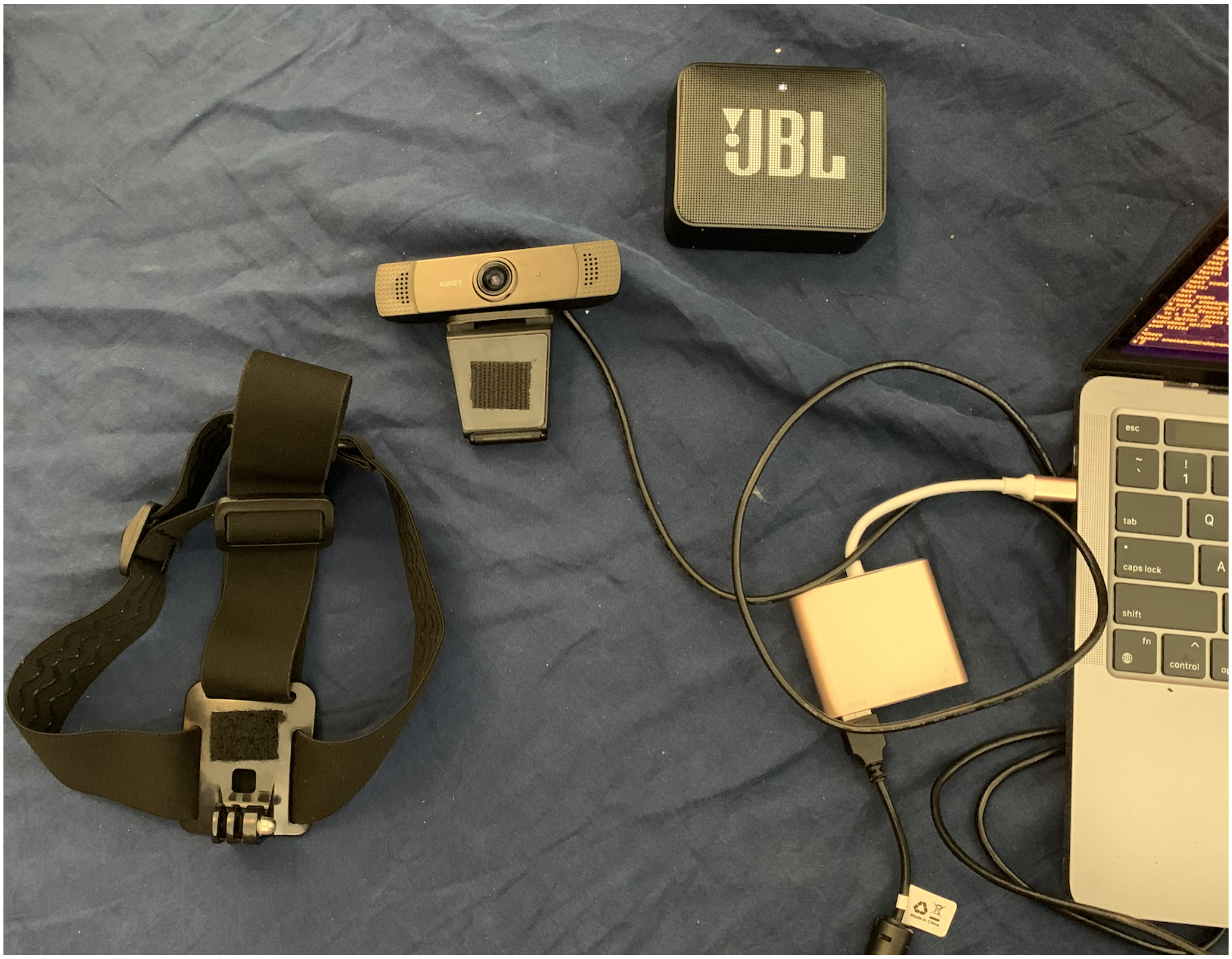

Figure 4: Headband, Camera (Securd Via Velcro), Speaker and Relevant Wiring

product. After evaluation of the implementation, there are certain limitations that can be expanded upon in further development:

5.1. Ease of Use:

The user experience follows a logical sense of flow that makes sense. Additionally, camera ports and speakers are automatically connected and synced. However, since this prototype has a barebones user experience, the product has a lot to improve on, especially in terms of "ease of use."

In the real-time reading functionality, the current working implementation relies on the user to "scan" the page which captures the relevant frame of data. This causes a choppy user experience that is heavily reliant on students being able to listen closely to directions and pulls them away from the experience of reading, after each page. However, in future extensions of the project, this data collection could be automated. As frames are constantly being read, if the portion of the frame that has the book, drastically changes, this could signal a change in the "page" of the book we are reading. The only caveat is that children are not still and the portion of the frame that has the book in one scanning of a page may not be the same in another. This makes basic comparison of image data difficult, however, more advanced ML techniques could be used to recognize new pages, books falling out of frame, closing of books, etc.

Another issue that arises is understanding what portions of the page should be read, which is highly dependent on how the student places the camera. The Google Cloud Vision API works almost too well at optical character recognition. For example, when testing, if there was an "EXIT" sign in the background, the API also read this aloud. A possible solution to this would be creating a "bounding" box around what the system identifies as the book. The previously mentioned vision language models and OpenCV models might be possible tools, however, their accuracy is not guaranteed. In a previous iteration of the project, I had implemented a version of this, by using an online tutorial [1]. However, the accuracy drops at different angles and also begs the question of how to handle frames with multiple books.

5.2. Addressing the Problem

The current functionality of the implementation supports reading stories and telling stories. Both of these functions bolster language comprehension skills. However, a large portion of why third graders fail to meet the benchmark before fourth grade is due to a lack of phonics lessons, help, and reinforcement.

In further iterations, the real-time reading functionality should expand to support the child leading the reading process, where the child reads and if they are struggling, the bot helps them "sound out" the word. This would require a means of understanding the child's speech (there are so many complicated child speech patterns to account for), checking the speech against the read text, and the ability to phonetically break down a word. More research is needed to find a model, or likely create a speech-to-text model for children's voices. Yet this brings up ethical issues since the model would have to be trained on children. To produce an mp3 file where a bot breaks down how to sound out a word would require finding a way to connect to pronunciations for words.

A large motivation was considering how the literacy gap affected marginalized communities, especially those from lower-income communities. The cost of the current device includes a $40 speaker, a $30 webcam, a $15 headband, and the cost of using the Google API calls. This assumes the kid has their own stuffed animal (which can run from $5-100). The current setup also uses a laptop or computer! Assuming a family has access to a computing device, this is still cheaper than most of the robots and personal assistants developed by commercial products. Yet, it's still a high price point for the communities that these devices are built for. To cut costs, in later versions, we can replace the camera and speaker with a smartphone, which could also allow us to eliminate the need for a laptop. The whole product could be housed on an app. The only costs remaining would be the headband, stuffed animal, and Google API calls, making the product much more affordable.

By cutting costs and hardware down to make it more widely accessible, the product looks and seems less like a "robot." What consists of a robot? Is using a smartphone still not robotic? Does our concept of robotics come from "complex machinery" beyond the everyday mechanisms we see? Why is over-engineering a solution seen as more reliable? With these perspectives, our understanding of robotics might be harmfully over-platforming the field and further gatekeeping it from new minds and communities. Additionally, it increases our reliance on technological systems, since we assume we can trust it. Having more technology is not always the answer. In fact, building out the least technically advanced solution for a specific problem is critical to innovating from it. Cutting where possible allows us to preserve resources for when it's really needed. These critical reflections of technology result from considering the Design Justice framework and truly building a solution that fits the audience you've chosen. With this in mind, additional features should pass this framework and similar lines of critical questioning.

5.3. Inclusivity and Accessibility

In addition to cost, other factors contribute to the accessibility and inclusivity of the product. While the stuffed animal promotes social interaction, supports sensory exploration, and reduces anxiety, there are other design choices in the software that can be added in to make the experience more accessible.

The current implementation uses the command line terminal for user input. There are audio directions that are also printed in the terminal. However, if a kid is reading on their own, they are likely unable to read the terminal and might miss the voiced instructions. Building out a simple GUI for the students to use that uses emojis or pictures for the buttons, developing a "remote," or using voice commands, are other ways of user interaction that could make the project easier to use and more accessible for younger students.

Further, many children get distracted when listening to a book, especially if they don't have the book in front of them. In those cases, having pause and play functionality on audio files could be useful. Students could take breaks and are not required to listen to the whole book in one sitting. Especially as the student grows older and are listening to longer audiobooks, this would allow them to listen over numerous sessions.

Another critical step to increase accessibility is translation, both for translating texts in a non-English language to speaking English stories and instructions in a non-English language. Many children read books in different languages or encounter books with words in different languages. Moreover, children who are fluent in another language or English is not their first language can benefit from having another option to hear instructions. Further, hearing stories in another language may be a way of learning that language. Adding in translation capabilities would enhance the usability of the product, address the needs of often marginalized students, and maintain a commitment to celebrating diversity in storytelling.

5.4. Kid-Centric Design

Many choices in the Rhino Reader prototype demonstrate a commitment to designing for an audience of children, without caricaturizing their needs. Stuffed animals are a classic way to support children in their growth, emotional health, and social skills. Another clear design choice to support children is the Voice Instruction UI. For early readers, voice based instructions are easier to follow than text based instructions. Also when Rhino Reader speaks to the child, it encourages them to read. To further build excitement, the software could gamify how long children read by providing statistics and keeping track of activity over time. This can be shared more widely with friends and family who can encourage the student to keep reading!

All in all, the current iteration has room to grow, but demonstrates “proof of concept,” reaching the goal of the seminar. To further evaluate how the product affects the target user group, testing needs to be done on younger children with the product. Before then, a few more features should be built to streamline the testing process and focus the study. Further, accessibility needs should be prioritized in the next iteration to allow for testing of how the device supports students with different needs. At the current moment, almost everyone I’ve spoken to about the project who is a parent, children’s educator, babysitter, or at that age level, is excited about the impact this could have on children’s literacy.

6. Conclusions and Future Work

The literacy gap in third grade disproportionately affects children of marginalized identities. This is unacceptable. Our current landscape of solutions falls short in designing for children of different identities, preserving the art of storytelling, and maintaining a fair and accessible price point. Storyteller Bot aims to address this gap through its low-tech design, kid-friendly design, and conscious effort to maintain Design Justice values throughout the development process. While the current version is a barebones “proof of concept,” the paper lays out numerous pathways to create a more robust version, including features like translation, gamification, adjustment to movement, etc.

With this new version implemented in the future, more rigorous performance studies with children can be conducted, determining whether the robot will have an effect on strengthening literacy skills. Given successful performance studies, this product could be expanded into school districts across the country and world.

In the near future, the project will be featured at events within the Princeton community– as I continue to work with the Children's Librarian in Firestone and partake in the seminar's exhibition. At these events, I hope to raise interest and awareness for the project so that it can continue into the future and become a transformative tool for literacy! Further, by expanding literacy, I hope to spread the power of storytelling and inspire those to continue to expand representation in storytelling.

7. Appendix

To take a look at the software behind the code and a demo of the working prototype, check out:

Github Repository Link:

https://github.com/aneekahuddin/IW-Spring-2024

References

[1] "Book detection tutorial." [Online]. Available: https://yasoob.me/2015/03/11/a-guide-to-finding-books-in-images-using-python-and-opencv/

[2] "Codibot." [Online]. Available: https://www.pillarlearning.com/products/codi

[3] "Designing inclusive toys." [Online]. Available: https://www.linkedin.com/pulse/designing-inclusive-toys-empowering-students-george/

[4] "Elementary school reading apps and websites." [Online]. Available: https://www.commonsense.org/education/lists/elementary-school-reading-apps-and-websites

[5] "Luka reads." [Online]. Available: https://www.lukareads.com/

[6] "Miko." [Online]. Available: https://miko.ai/

[7] "Moxie bot." [Online]. Available: https://moxierobot.com/products/ai-robot?nbt=nb%3Aadwords%3Ax%3A20303082944%3A%3A&nb_adtype=pla&nb_kwd=&nb_ti=&nb_mi=219573267&nb_pc=online&nb_pi=101300&nb_ppi=&nb_placement=&nb_li_ms=&nb_lp_ms=&nb_fii=&nb_ap=&nb_mt=&utm_source=google&utm_medium=cpc&utm_target=&utm_content=&utm_campaign=Performance_Max_ShoppingOnlyTest&utm_term=&utm_campaignid=20303082944&gad_source=1&gclid=Cj0KCQiAoKeuBhCoARIsAB4WxtcEPh9XmBKIqHDXchnWQcVIuGINegjwBByuFGZI93UPq6WH4_p9rwYaAh1FEALw_wcB

[8] "Openai." [Online]. Available: https://openai.com/product

[9] "The simple view of reading." [Online]. Available: https://www.readingrockets.org/topics/about-reading/articles/simple-view-reading

[10] "The benefits of stuffed animals in child development." [Online]. Available: https://bunniesbythebay.com/blogs/how-to-delight/the-benefits-of-stuffed-animals-in-child-development

[11] "Universal play guidelines (adapted)." [Online]. Available: https://www.teachingei.org/technology/documents/3.5%20UDP%20Guidelines.pdf

[12] "Why is reading on grade level by 3rd grade so important." [Online]. Available: https://commcenter.bsu.edu/message/why-is-reading-on-grade-level-by-3rd-grade-so-important

[13] S. C. Chock, *Design Justice*. Boston, Massachusetts: MIT Press, 2020.

[14] E. Group, “Narrowing the third-grade reading gap,” EAB, Tech. Rep., 2020.

[15] J. Lesnick *et al.*, “Reading on grade level in third grade: How is it related to high school performance and college enrollment? a longitudinal analysis of third-grade students in chicago in 1996-97 and their educational outcomes.” Chapin Hall at the University of Chicago, Tech. Rep., 2010.

[16] B. Lin *et al.*, “Moe-llava: Mixture of experts for large vision-language models,” 2024.

[17] H. Liu *et al.*, “Improved baselines with visual instruction tuning,” 2023.

[18] H. W. Park *et al.*, “Telling stories to robots: The effect of backchanneling on a child’s storytelling,” in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 100–108. Available: https://doi.org/10.1145/2909824.3020245

[19] D. Zhu *et al.*, “Minigpt-4: Enhancing vision-language understanding with advanced large language models,” 2023.