A Financial Literacy AI-Enabled Voice Assistant System for Educational Use

A Technical Report submitted to the Department of Systems and Information Engineering

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Vineeth Sai Thota

Spring, 2022

Technical Project Team Members

Jesilyn Gopurathingal

Candace Miu

Matthew Thompson

Niels van Beek

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Tariq Iqbal, Department of Engineering Systems and Environment

A Financial Literacy AI-Enabled Voice Assistant System for Educational Use

Candace Miu[1], Jesilyn Gopurathingal[1], Vineeth Thota[1], Matthew Thompson[1], Niels van Beek[1], Jennifer Kuczynski[2], Jyotirmay Gadewadikar[2], and Tariq Iqbal[1]

Abstract—**Financial literacy is crucial for saving money, avoiding debt, establishing strong credit, and many other skills that help build wealth throughout an individual's life. A very large percentage of Americans from various demographics and backgrounds do not have the basic financial and economic knowledge to sustain themselves financially. Our proposed solution to tackle financial illiteracy is by ensuring students are taught the foundational expertise at a young age so that they make wise financial choices by the time they reach adulthood. We have developed a virtual voice assistant that will improve financial literacy by offering lessons that will cover all topics within the National Standards in K-12 Personal Finance Education educational curricula. Data was collected and analyzed in order to assess the effectiveness, robustness, and engagement of the voice assistant. While further analysis on engagement should be conducted, the bot met baseline goals of effectiveness and robustness which can further be improved through more intent training and testing on potential users.**

Index Terms—**financial literacy, voice assistants, education, conversational mapping, artificial intelligence**

I. Introduction

Financial literacy refers to the skills, information, and tools that enable people to make individual economic decisions and achieve their objectives. Studies show that only 57% of adults in the United States are considered financially literate [1]. Financial literacy is often referred to as financial competence, particularly when combined with access to financial products and services [2]. This is extremely valuable as it equips one with the knowledge to manage and utilize money effectively. With the growing and complex nature of the American financial system, the need for a call to action to improve the national financial literacy levels is more imperative than ever. Many Americans are tasked with paying off their cars, homes, or student loans. In 2019, about 42% of adults in the United States said they had a budget and used it to keep track of their spending [3].

If taught at a young age, personal finance would help one prepare to make major financial decisions after high school. Results have shown that about 1 in every 5 American teenagers lack basic financial literacy skills [4]. After high school, students often follow various paths, such as attending college or joining the workforce immediately. A personal finance education delivered in high school would increase the likelihood of young adults having a better understanding of financial topics. Currently, personal finance and economics classes are offered in most high schools; however, only 21 states require high school students to take a personal finance class before they graduate [5], and these classes are seldom taught in middle and elementary schools. Even high schools struggle to find teachers and resources to adequately prepare their students since a large percentage of American adults themselves lack a sufficient level of financial understanding; a little more than half of adults are considered financially literate [6]. If nothing is done to improve the financial literacy rates in the U.S., many Americans will continue to struggle with managing their money, which is crucial to ensure financial stability and well-being for many. Without this basic knowledge and skill set, students will be more prone to making irresponsible financial decisions, resulting in severe consequences like unpaid debt, bankruptcies, and foreclosures [7]. We intend to implement a virtual voice assistant that will guide students through various lesson plans outside of the classroom to achieve higher financial literacy rates; this will help to introduce these important skills to students at a young age and prepare them to make future financial decisions.

Our plan of developing a virtual voice assistant includes lesson plan creation, cloud platform implementation and system evaluation. Through this, we intend to create a dynamic collaboration between education and technology to gradually build knowledge among students. Our goal is to help students build off of their foundational knowledge and practice their financial literacy skills at a reasonable pace. This will prepare them to tackle more advanced concepts and set them up for economic success in the future, allowing them to be financially independent.

Data from 100 calls — 50 for each lesson — were used in the assessment of the voice assistant. This data was analyzed in order to determine how effectively the assistant achieved the goals laid out for it. The four primary goals for the assistant were: effectiveness of the voice assistant, robustness of the system, engagement of the conversations, and robustness of parameters. The voice assistant yielded strong results, meeting thresholds for all goals, but still showing room for improvement in areas. The Kindergarten Spending and Saving average conversation duration was within the expected duration range while the Fourth Grade Credit and

[1]*Department of Engineering Systems and Environment, University of Virginia*, Charlottesville, USA {cpm6ug, jmg8aw, vst7ta, mlt2we, nhv5bq, tiqbal}@virginia.edu.

[2]*MITRE*, McLean, USA {jkuczynski, jgadewadikar}@mitre.org.

Debit average conversation duration was longer than the expected duration range. Additionally, the assistant was able to correctly detect the users' speech on the first attempt 81.8% of the time and longer interactions were usually due to misunderstanding user input. Understanding a wider range of possible responses through increased testing and training for the voice bot will improve the system's ability to handle any user input.

II. Prior Work

Voice assistants and their educational role have been examined in regards to their existing uses, as well as their future potential throughout various levels of education. A 2021 study found that voice assistants are known to be popular for simple everyday tasks such as asking what the weather is like and playing music but seem to be less used within the realm of education [8]. Nonetheless, the markets for A.I. and chatbots, which are the backbone of technology for voice assistants, are growing in size and beginning to garner interest in classrooms [9]. Other current and educational applications for voice assistants include language learning [10] and teacher-to-student communication [11], which have received overall positive feedback from students. Studies have also been conducted to test future voice assistant capabilities, like the ability to improve web page information accessibility [12], assist students with autism in the classroom [13], and ease the responsibilities of teachers for online classes [14]. While these studies offer insight into the educational uses of voice assistants, they do not explore the technical methodology of developing a voice assistant and implementing it onto a cloud platform. This is crucial to advancing the capabilities of virtual voice assistants for future use in various learning environments. For this reason, we explore the process of building a voice assistant, specifically an assistant that addresses financial literacy.

In their 2019 study, Reyes et al. [15] walk through the methodology used to create a virtual voice assistant in an educational setting. The voice assistant was developed using Google Dialogflow, with the purpose of allowing undergraduate students to access learning material for a course about artificial intelligence. Through their data gathering phase, a foundational knowledge of the course itself was established through literature review, with subsequent questions that were formed and grouped into categories for developers to structure the content. The conversational flow was another area of focus, as it defines the interactions between the voice assistant and the user. Two types of conversational flows were identified: linear [16] and nonlinear [17]. Linear flows represent the basic question and answer steps, while nonlinear flows allow for a more diverse structure and focus heavily on suggestions. Gadewadikar and Vatatmaja reconfirmed the importance of the stochastic nature of human conversation and reinforced that voice assistants must then follow a nonlinear approach in conversation to best reflect this [18].

In addition to choosing the right type of conversational flow, visualizing and mapping out each flow is imperative. To facilitate this process, decision trees were utilized to construct the framework of the voice assistant [15], with nodes that are assigned conditionals that activate when a user's query matches the response stored in the node; these nodes were translated as intents in Dialogflow. This study details the methodology of building and implementing a virtual voice assistant on Google Dialogflow; however, there is much to explore in regards to content generation for students, such as different modes of passing on knowledge. This could come in the form of independent learning, direct interactions between teachers and students, etc. Building an effective assistant is an interactive process with an emphasis on training, testing, and tuning, otherwise known as the three Ts of conversational AI [19]. By continuously performing the three Ts, the assistant can become more conversational and realistic.

In collaboration with MITRE, a team from the Massachusetts Institute of Technology established the functional, non-functional, and technical requirements of the virtual voice assistant. We utilized the work above as foundational knowledge to design our own voice assistant in a more specific realm of financial literacy and generate lesson plan content tailored for kindergarten to twelfth grade students.

III. Methodology

The creation of the prototype was divided into three main components: lesson plan content generation, cloud platform implementation, and system evaluation.

A. Phase 1

For lesson plan content generation, the first task was to determine the recommended standards for students to be considered financially literate. The National Standards in K-12 Personal Finance Education provided benchmarks for students in the following topics: employment and income, spending and saving, investing, credit and debt, risk management and insurance, and financial decision making [2]. The National Standards was used to tailor each lesson plan to the students' expected level and personal progress. Benchmarks were provided for kindergarten, fourth, eighth, and twelfth graders and lesson plans were created for each benchmark of the first standard for all six units.

B. Phase 2

For the second component, the next task was to choose a public cloud platform that would best fit the project's scope. The three initial candidates included Amazon Web Services [20], Microsoft Azure [21], and Google Cloud Platform [22]. Google was chosen for its user-friendly capabilities, specifically with Google Dialogflow CX being the most useful platform that Google Cloud offered. Some advantages of Dialogflow included an interface with a visual conversational flow builder, easy to understand documentation, and multiple platforms for telephone integration. The first set of lesson plans were imported into a platform called Voiceflow [23]

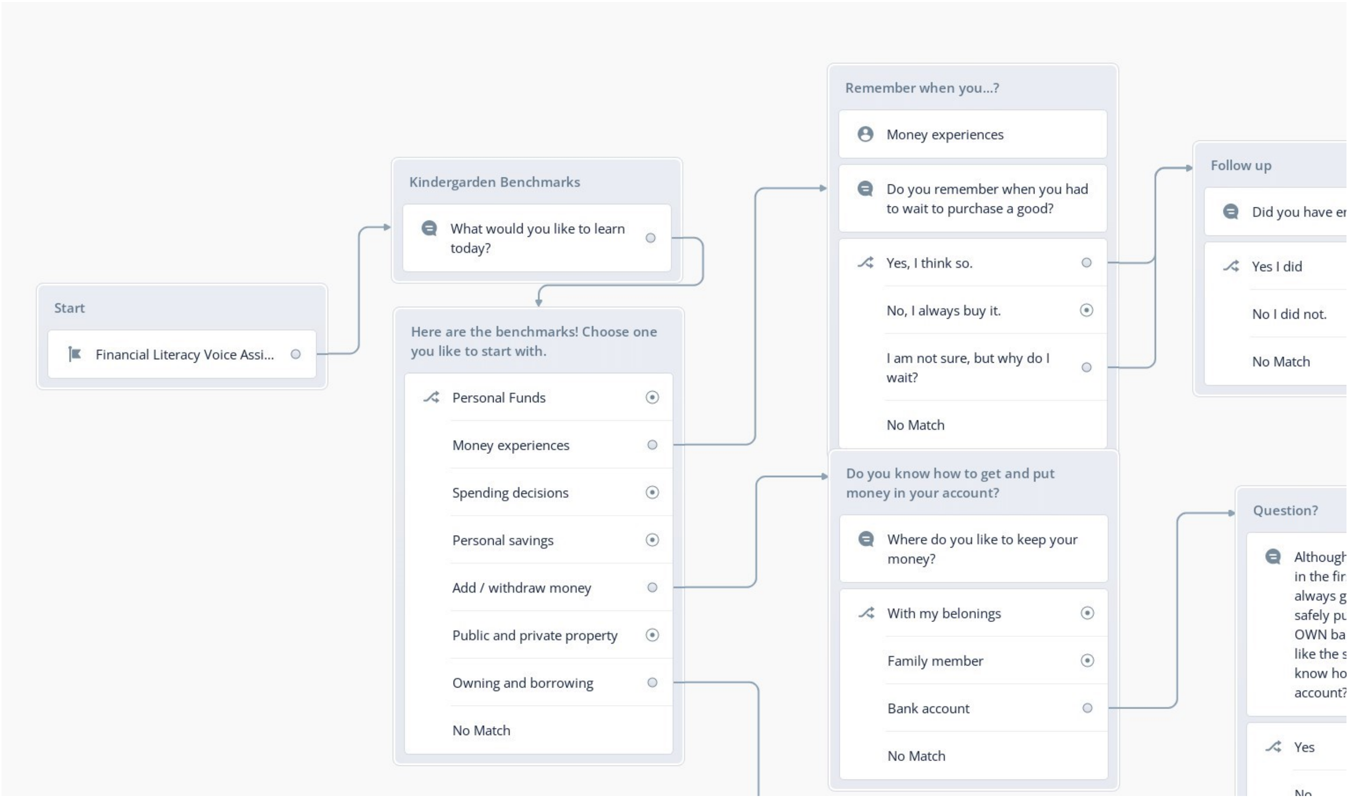

Fig. 1. An example kindergarten lesson mapped out in VoiceFlow.

to visualize the conversational flows of the first few lessons before they were transferred to Dialogflow, along with the remainder of the lesson plans (See Fig. 1). The initial flows built out in Voiceflow helped to show the complexity of each lesson plan, as well as the different routes that diverged from each topic. It also enabled us to clearly list out all the options of user responses for each prompted question and how each response would transition into the next segment. Once each lesson plan was completed in Google Dialogflow, the project was assigned an agent called "Finance Guru", which served as our chatbot. To implement a lesson, a new flow was created. Each flow had a start page that greeted the user and listened for a response before continuing on with the lesson. As the lesson progressed, each question that the agent prompted was given its own page, with an entry fulfillment that acted as the agent's response when the page was activated. On each page, there were routes that represented the different response options of the user. Some routes were as simple as a "yes" or "no", while others represented more complicated responses.

Additionally, every route required an intent or something that represented what the users wanted to do during the conversation with the agent. While Dialogflow provided default intents, we added our own intents with personalized training phrases, which made our voice assistant more compatible with various user inputs and made the conversations more natural and personable. Once Dialogflow matched the user's response to an intent to the intent, another fulfillment was activated, so the agent could respond to the user - usually confirming or dissenting the user's answer.

The final step to create a page was to create the transition, which connected the current page to either a new page or a new flow. Each page was connected to a new page that represented the next question in most cases. For routes where the user said the incorrect answer, the transition connected again to the same page, creating a loop so that the agent repeated the question for the user to try again. In a more simply structured lesson plan, the flow consisted of each page connected to one other page, with an option to loop back for every question. However, some lesson plans that contained more complex structures had flows in which certain pages could connect to two or more pages, with routes that diverged from each other. Finally, each flow ended with an "End Flow" page, which signaled to Dialogflow that the flow should be terminated.

The team utilized the CX Phone Gateway [24] (Google's built-in telephony service) and AudioCodes [25] (a third-party telephony service compatible with Dialogflow) to allow users to access the various lesson plans via telephone. CX Phone Gateway provided the most conversational voice options but failed to store important data from the phone calls. On the other hand, AudioCodes had fewer conversational

TABLE I
KEY VARIABLES WITH THEIR ASSOCIATED METRIC TO EVALUATE THE PERFORMANCE OF THE VOICE ASSISTANT SYSTEM

Goal	Metric	Reason
(1) Effectiveness of the voice assistant in delivering the content of the lesson plans	Lesson duration: track conversation length of new conversation against some baseline	See if lessons are running longer or shorter than expected. This can show room for improvement of the language our lessons use.
	Page duration: Track page durations normalized to each interaction to be able to compare all interactions together	See if certain pages need to be trained more or if language at a certain page can be improved.
	Lesson duration standard deviation: spread of duration for each conversation	Track if a lesson has inconsistent call times.
(2) Robustness of the voice assistant's understanding of the users' interactions	Intent/Match detection confidence value: an AudioCodes-provided metric indicating how confident the AI is in its matching of user intent to trained intents	See if we need to train a given page more.
	Number of repetitions: Average number of times the bot repeated itself throughout the course of delivering a lesson plan	See if we need to train a given page more.
(3) Level of engagement from the voice assistant	Conversation completeness: if a conversation made it to the end of the lesson	See if users are sticking with conversation.
(4) Overall strength of the parameters	Parameter extraction accuracy: this can be measured by looking at the repetitions in lessons that use parameters (e.g. Kindergarten Spending and Saving lesson)	See if more parameter training needs to be done.

options but provided abundant data during each phone call. Because of this, the final product used the CX Phone Gateway, but all testing, data collection, and evaluation was done through AudioCodes. Additionally, the dialogue was framed in Speech Synthesis Markup Language (SSML), a language that allows for more customization of the bot's speech. This allowed for the addition of pauses and emphasis to increase the conversationality of the bot.

C. Phase 3

For the third and final phase, a system evaluation approach was used to test the effectiveness of the voice assistant. The major goals of the evaluation infrastructure were to understand the (1) effectiveness of the voice assistant in delivering the content of the lesson plans, (2) the robustness of the voice assistant's understanding of the users' interactions, (3) the level of engagement from the voice assistant, and (4) the overall strength of the parameters. The specific metrics used are shown in (See Table I).

AudioCodes was then used to collect real-time data on the phone calls; the software provides its own text-to-speech bots and its own speech-to-text software, using Dialogflow on the back end to process the users' words to return a response. AudioCodes is directly integrated into the conversation, so Dialogflow and AudioCodes send information directly back and forth during each interaction, and AudioCodes logs this data. Data is organized with each row representing an individual interaction with the assistant (e.g., the user answers a question asked by the voice assistant), and the metrics listed above are provided both directly and indirectly. Data is exported from AudioCodes and parsed in JavaScript to provide a cleaned dataset in the form of a JSON file. The JavaScript code takes in a file as input, iterates through each interaction, and tracks and stores the metrics (See Table I). This dataset is then converted into a CSV, and an R script is used to provide visual and numeric analysis on the system.

IV. RESULTS AND DISCUSSION

In order to conduct our analysis, a dataset of summarized conversations were used and a dataset of the individual interactions. The summarized dataset tells a better story for the overall performance of the virtual assistant, whereas individual interactions dataset allows for a more detailed analysis into the strengths and weaknesses of the system. Two key lesson plans were analyzed: Kindergarten Spending and Saving and Fourth Grade Credit and Debit. Data from 50 calls was collected for each lesson via AudioCodes.

A. Goal 1: Effectiveness of the Voice Assistant

The Credit and Debit lesson shows a right-skewed distribution for complete conversations with most incomplete calls being much shorter (see Fig. 2). This is expected, as this lesson was more complex, leading to longer conversations and more errors with the voice bot. The complexity also caused more incomplete phone calls. The Spending and Saving lesson was much more normal around the mean with a lower incompleteness rate, which was expected due to the lesson plans' simpler format (see Fig. 2). Two single sample t-tests were performed on the total conversation duration for complete conversations against expected means of 620 seconds for Credit and Debit and 360 seconds for Spending and Saving. These expected durations are based on the duration of a perfect call with no repetitions or errors, simulating the ideal lesson length. The Credit and Debit

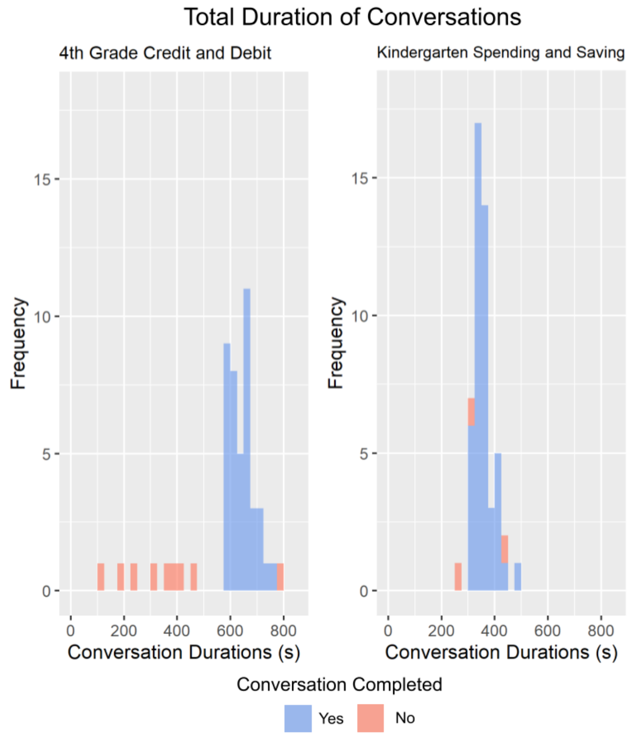

Fig. 2. Histograms of the total durations of both lessons

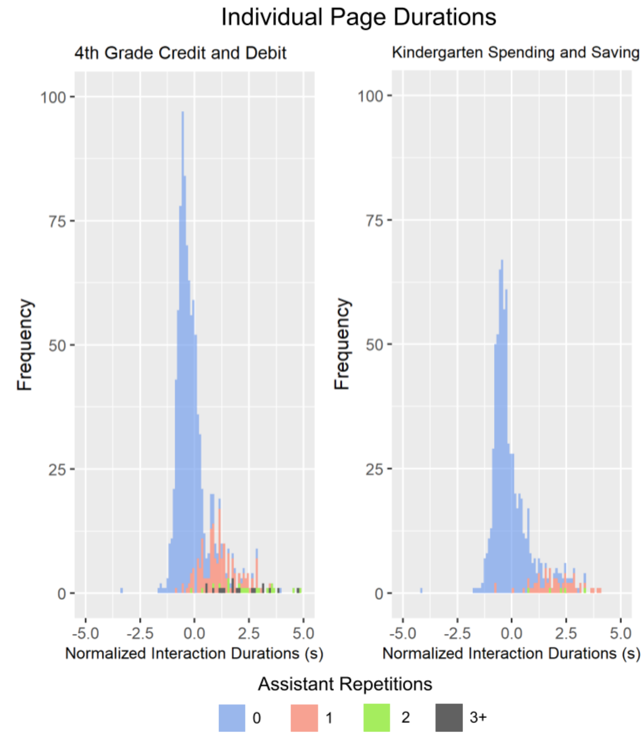

Fig. 3. Histograms of the page durations of both lessons

data had a mean of 642.9 seconds and was higher than the expected mean at a 99% confidence level (p-value < 0.01). The Spending and Saving data had a mean of 358.5 seconds and was not significantly different from the expected mean. This shows that the Credit and Debit lesson could be less effective at delivering the content than needed. Looking at the spread of conversation lengths, the Credit and Debit data had a standard deviation of 46.5 seconds while the Spending and Saving data had a standard deviation of 36.0 seconds. The spread should continue to be tracked as the bot is improved since we want future lessons to be tighter around the mean, indicating more consistent performance from the assistant.

B. *Goal 2: Robustness of the System*

An important metric to measure the robustness of the system is the duration of each interaction (question) throughout the lessons. This is important to measure because it indicates if the user is struggling on a certain question within the conversation. We normalized the lengths of each individual interaction while grouping the interactions by each question. Grouping the interactions by each question is necessary before normalizing, as it allows us to compare page durations across all questions, no matter the length of question or how open-ended the question is. The lengths of each singular interaction across all conversations were grouped and graphed as described to visualize the robustness of the system (See Fig. 3).

The vast majority of interactions with no repetitions remained near the mean, which indicates good system performance by the voice assistant in terms of its ability to consistently deliver the questions to the user and have the user understand and respond in a timely manner. We can also see that many of the interactions that went beyond the mean were a result of the assistant repeating the question, whether it was due to misunderstanding the user, or the user answered incorrectly. This is important because it shows that longer interactions can be attributed to pages being repeated. Our system should be able to handle misunderstandings better and be able to understand the user on the first try without any repetitions. The assistant was able to correctly detect the users' speech on the first attempt 81.8% of the time. This shows a need to better train the intents of each page and incorporate stronger error handling so our bot can be more robust to unknown inputs. In order to increase this confidence, we can increase the number of training phrases in responses to its questions so that it is able to handle a wider range of possible responses. The Credit and Debit data shows an average interaction confidence of 87.3% for complete conversations and 83.8% for incomplete conversations. The Spending and Saving data shows an average interaction confidence of 88.8% for complete conversations and 89.3% for incomplete conversations. These results indicate that the bot's confidence level has no impact on whether or not a conversation is complete, showing that its ability to understand user input is not a large factor in increasing conversation completeness.

C. Goals 3 and 4: Engagement of the Conversations and Robustness of Parameters

The best data available that gives potential insight on engagement of users is the rate of incompleteness for the lessons. For the Spending and Saving data, 6% of the calls were incomplete while for Credit and Debit, 18% of the calls were incomplete. Incompleteness rate can be attributed to other factors, such as errors in the system. An opportunity to better measure engagement is adding a survey at the end of a lesson. The Spending and Saving conversation used parameters to increase the conversationality of the lesson. The bot was built to pull out certain keywords that the user gave and incorporate them into the conversation. This lesson saw an average repetition rate of 1.6 repetitions per conversation (for reference, the Credit and Debit lesson had 5.6 repetitions per conversation). This low level of repetitions shows the robustness in error handling of the Spending and Saving lesson, including the bot's ability to understand parameters.

V. Conclusion and Future Work

Smart technologies such as virtual voice assistants have the power to change how students learn. In this paper, we have presented a model that focuses heavily on two main lesson plans, with each being from a different category and age group of the National Standards. This model aims to improve financial literacy for students in kindergarten to twelfth grade. Our experimental results indicate that our current voice bot has a good baseline for our goals of effectiveness and robustness with room for improvement via more bot training and improved language used in the lessons. A future area of improvement for testing is adding a survey to our lessons so we can better track the engagement of users.

Usage of this technology in an educational setting is still fairly new, which means there is access to limited data. Future work for progressing the prototype will require several iterations of user testing in order to acquire this data. Given the addition of more lesson plans and users, a more thorough analysis of the effectiveness of the prototype on a student's educational experience can be performed as well. A student perspective would be extremely valuable due to the knowledge gap between the team developers and student users. This could also indicate changes to be made in the pace of the voice, type of voice, and the appropriate vocabulary for each grade level. The data collected from this will highlight issues in the prototype to be addressed, thus ensuring the prototype fits the user's needs more adequately.

Acknowledgment

We would like to thank MITRE for supporting the project.

References

[1] Klapper, L., Lusardi, A., & Oudheusden, P. van. (2015). Financial literacy around the world: Insights from the Standard & Poor's Ratings Services Global Financial Literacy Survey.
[2] U.S. National Strategy for Financial Literacy 2020. (2020). 28.
[3] Harris On Demand. (2019, March 25). 2019 Consumer Financial Literacy Survey.
[4] OECD. (2017). PISA 2015 Results (Volume IV): Students' Financial Literacy. OECD.
[5] Carrns, A. (2020, February 7). More states require students to learn about money matters. The New York Times.
[6] Contreras, O., Bendix, J. (2021). Financial Literacy in the United States.
[7] Pollack, J., Rigoglioso, C., Fagan, J. (n.d.). The Financial Illiteracy of America's Youth and its Impact on the American Economy.
[8] Tsourakas, T., Terzopoulos, G., Goumas, S. (2021). Educational use of voice assistants and smart speakers. Journal of Engineering Science and Technology Review
[9] Osetskyi, V., Vitrenko, A., Tatomyr, I., Bilan, S., Hirnyk, Y. (2020). Artificial intelligence application in education: financial implications and prospects. In: Financial and credit activity: problems of theory and practice
[10] Terzopoulos, G., Satratzemi, M. (2020). Voice Assistants and Smart Speakers in Everyday Life and in Education. Informatics in Education
[11] Şerban, C., Todericiu, I. A. (2020). Alexa, What classes do I have today? The use of Artificial Intelligence via Smart Speakers in Education. Procedia Computer Science
[12] Bortoli, M., Furini, M., Mirri, S., Montangero, M., Prandi, C. (2020). Conversational interfaces for a smart campus: A case study. In Proceedings of the AVI
[13] Bakhai, A., Constantin, A., Alexandru, C. A. (2020). Motivateme!: An Alexa skill to support higher education students with autism. In: IHCI
[14] Zhao, J., Bhatt, S., Thille, C., Zimmaro, D., Gattani, N., Walker, J. (2020). Introducing Alexa for E-learning. In: Proceedings of the Seventh ACM Conference on Learning@ Scale
[15] Reyes, R., Garza, D., Garrido, L., De la Cueva, V., Ramirez, J. (2019). Methodology for the implementation of virtual assistants for education using Google Dialogflow. Advances in Soft Computing
[16] Tiwari, A., Talekar, R., Patil, S. M. (2017). College Information Chat Bot System.
[17] Bhushan, R., Kulkarni, K., Pandey, V.K., Rawls, C., Mechtley, B., Jayasuriya, S., Ziegler, C. (2020) ODO: Design of Multimodal Chatbot for an Experiential Media System. Multimodal Technologies and Interaction.
[18] Gadewadikar, J., Vatatmaja. (2020). Best practices in developing conversational systems. Conversational Interaction Conference
[19] Vatatmaja, Gadewadikar, J., Sherry, C., Timothy, M. (2019). How the three Ts of conversational AI build better voice assistants.
[20] Amazon Web Services (2022). Available at https://aws.amazon.com/
[21] Cloud computing services: Microsoft Azure (2022). Available at https://azure.microsoft.com/en-us/
[22] Dialogflow Google Cloud. (2022). Available at https://cloud.google.com/dialogflow
[23] VoiceFlow. (2022). Avaiable at https://www.voiceflow.com/
[24] Voximplant. (2022). Available at https://voximplant.com/
[25] AudioCodes. (2022). Available at https://www.audiocodes.com/