DESIGNING FOR YOUTH TO CREATE CONVERSATIONAL AGENTS

By

XIAOYI TIAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2024

ACKNOWLEDGEMENTS

Before I entered the PhD program at UF, I was told that doing a PhD will be a challenging but extremely rewarding experience. Today, I couldn't agree more. I'm really thankful for everyone who helped me on this path. First, a big thanks to my advisor, Kristy Elizabeth Boyer. Your way of organizing, sticking to ethics, staying positive, and making wise life choices has taught me a lot. Your mentorship has transformed me, not just into a better researcher, but a happier person all around. Heartfelt thanks to my committee—Eric Ragan, Jaime Ruiz, and Wanli Xing—your insightful feedback and invested time have been priceless.

A shout out to the Project DIALOGS team — those successful summer camps wouldn't have rocked without you —Yukeyong Song, Gloria Katuka, Timothy Brown, Lydia Pezzullo, Amit Kumar, Sunny Dhama, Tom McKlin, Christine Wise, Joanne Barrett, Maya Israel, and all the fantastic camp facilitators that I have worked with over the last three summers, I couldn't have finished my dissertation without you.

To the LearnDialogue group, I'm blessed with your support. Joseph Wiggins and Mehmet Celepkolu, thanks for the guidance in my first and second PhD years. Amogh Mannekote, your jokes and friendship made those long nights bearable and productive. Yingbo Ma, thanks for generously offering firsthand tips about the qualifying exam and proposal. Carly Solomon, my awesome undergrad, thanks for your two years of hard work reviewing literature, running summer camps, and analyzing data. And to everyone in the lab, I'm so lucky to have all of you around.

And to my personal cheer squad—my friends and family—Jingxi Weng, you're my rock, my constant, and the best role model for work-life balance. To Heting, Jiarui, Yutong, Yujuan, Song, Yingbo, and everyone in my family, my mother, father, aunts, uncles, grandmother and grandfather, and many more—your love and support wrap around me like the warmest blanket. I am incredibly grateful to have all of you in my life.

Thanks also to the National Science Foundation for the graduate research assistantship through project DIALOGS (DRL-2048480) and PRIME (DUE-1626235 and DUE-1625908), which made this dissertation possible.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DESIGNING FOR YOUTH TO CREATE CONVERSATIONAL AGENTS

By

Xiaoyi Tian

August 2024

Chair: Kristy Elizabeth Boyer
Major: Human-Centered Computing

As Artificial Intelligence (AI) becomes increasingly ubiquitous in society, conversational agents such as Siri, Alexa, and ChatGPT are shaping the experiences of younger generations. However, these young users often lack opportunities to learn about the inner workings of these AI technologies. One way to foster such learning is by empowering youth to create AI that is personally and socially meaningful to them.

To address this educational need, my research introduces AMBY (“AI Made By You”), a novel development environment for youth to build their own conversational AI projects. AMBY was iteratively designed for and with youth aged 12-13 through contextual inquiry and usability studies and has been deployed in an AI summer camp over two years. In summer of 2022, I explore the experiences and perceptions of middle school learners interacting with AMBY. The insights gained from these user studies informed the development of AMBY 2.0, which integrates the *entity* feature to enhance learners’ experiences. Preliminary results from the summer of 2023 indicate that this feature aids in the better design of projects and mitigates learners’ frustration with repetitive entry of training data.

My proposed dissertation study seeks to expand upon the summer camp deployment by moving into a formal learning setting, that of middle school science classrooms. I will conduct a new classroom study using the updated AMBY 2.0 interface. The primary goal is to evaluate the impact of the *entity* feature on learners’ interests and experiences with AI. Additionally, I aim to explore how conversational AI learning experiences shape learners’ understanding of AI, as well as their attitudes and identities toward AI in the context of middle school science education.

CHAPTER 1
INTRODUCTION

As Artificial Intelligence (AI) becomes increasingly ubiquitous in society, conversational agents such as Siri, Alexa, and ChatGPT are shaping the experiences of younger generations [12, 17, 13, 112]. Conversational AI applications include virtual agents [105], intelligent personal assistants [12, 87], and chatbots [94]. These applications are powered by complex AI algorithms and bring new opportunities to immerse children in AI-driven experiences. Such innovative use cases include increasing engagement in reading [109], supporting language learning [110, 33], promoting story comprehension and engagement [111], and fostering question-asking behaviors [6, 66].

Although opportunities for young people to *interact with* conversational AI are plentiful, opportunities to *deeply understand* how these technologies work are still scarce. Due to their complex and inaccessible nature of these AI technologies, many people, especially children, find it hard to see how these technologies work, these AI systems remain “black boxes”. This lack of transparency can lead to many misunderstandings [23] and a noticeable gap between using AI and truly understanding it, which brings up important questions about whether future generations will be able to critically interact with and positively contribute to the development of AI. Fostering AI literacy can inspire more students to consider careers in AI, laying a strong foundation for their higher education and professional lives [59].

There have been efforts to establish frameworks for AI education at the K-12 level. Touretzky et al. [95] founded the AI4K12 initiative, which proposed five “big ideas” to navigate the landscape of AI education. These “big ideas” include perception, representation and reasoning, learning, natural interaction and societal impact [95]. For young learners, developing their own personally meaningful conversational agents can serve as a rich learning experience, shaping their perceptions and enhancing their understanding of AI [24, 100, 62]. However, there is a lack of developmentally appropriate tools for *learning to build conversational AI* [30].

Although platforms like Google Dialogflow [4], Rasa [2], IBM Watson [5, 27], and Azure Bot Service [1] offer robust development tools and a myriad of functionalities enabling skilled developers to construct advanced conversational AI applications, they often demand extensive programming knowledge [81, 15]. Many of these features were not conceptualized to facilitate learning about AI in a manner that is authentic and effective for young learners, thereby presenting a barrier to fostering AI comprehension among this demographic.

My dissertation contributes to addressing this need by introducing a novel conversational AI development tool, AMBY ("AI Made By You"), designed for young learners to create their own conversational agents, even without prior programming experience. Through a funded research project, Project DIALOGS[1], AMBY was iteratively designed over 14 months and has been implemented in AI summer camps for the past two years in Gainesville, Florida [50]. Within this period, AMBY has empowered 39 learners to build their own conversational agents. They expressed that AMBY provided them with the autonomy to develop personally meaningful projects. The results from the summer camp demonstrate significant increases in learners' ability beliefs, willingness to share their learning experience, and intent to persist in AI learning [89].

My proposed dissertation study aims to extend our previous work on summer camp implementation [50], conversational AI curriculum [89] and the development interface AMBY [54], to a formal learning environment, that of middle school science classrooms. The primary goal is to evaluate the impact of the interface feature on the learning outcomes. Additionally, I aim to explore how conversational AI learning experiences shape learners' understanding of AI, as well as their attitudes and identities toward AI in the context of middle school science education.

1.1 Research motivation

My research targets middle school-aged youth because this age has been identified as a key developmental period for interest and identity building [36]. A positive AI learning experience during this age could significantly impact learners' interest and attitudes towards AI [59].

I aim to integrate AMBY, along with its conversational AI curriculum, in a formal learning environment because summer camps are opt-in experiences that serve only a subset of learners, whereas classroom experiences have the potential to include all learners in the

1 Camp DIALOGS: Fostering STEM Career Identity and Computer Science Learning through Youth-Led Conversational App Development Experiences (DRL-2048480) PI: Kristy Boyer, Co-PI: Maya Israel.

partner teacher's classroom.

Transitioning the deployment of AMBY to in-school setting brings forth numerous benefits. Currently, there are limited classroom resources available within the schools to teach computing courses [18]. Few teachers have the necessary resources and expertise to introduce CS and AI concepts to their students [70]. Bringing AMBY and the conversational AI curriculum into the classrooms can bridge this gap by providing an accessible platform for learning AI concepts in a relevant and engaging manner. Additionally, classrooms naturally offer structured environments that demand fewer staff, making it feasible to reach a larger number of learners with fewer resources. Given AI's interdisciplinary nature, embedding AMBY within subject-specific curricula, such as science and language, empowers students to explore cutting-edge AI concepts in real-world applications, making their educational experience more relevant and engaging.

I chose the science classroom as the starting point of the transition, as middle school science standards in the state of Florida–and similarly in other states–contain essential AI components as learning objectives. These standards requires students to understand the concept of AI and recognize the responsible use and ethical implications of AI technologies[2]. While there have been efforts to develop AI curriculum at the middle school level [103, 106], a gap still exits in understanding how to integrate AI learning environments and curricula in to the core subjects such as science.

Through my multi-year user studies of AMBY, we have continuously updated AMBY with new features, which the two major versions I name as AMBY 1.0 and AMBY 2.0. Both AMBY 1.0 and 2.0 support users in generating training data and visualizing conversation flow. They also allow both written and spoken input and output modalities and enables users to customize the voice and appearance of their agents. Users can deploy their new conversational agents from AMBY to a website, they can even access their agents by calling them on the phone. We deployed AMBY 1.0 in summer 2022. The results from the previous camp inspired us to design and develop a new set of features for AMBY to support more diverse use cases. In summer 2023, we introduced AMBY 2.0, with the new feature called *entity*. The objective

2 Florida K-12 standards (relevant middle school CS and AI objectives are SC.68.CS): https://www.cpalms.org/public/search/Standard.

of the *entity* feature is to allow users to create more personalized responses for their agents and become more efficient in creating training phrases, thus enhancing the overall user experience.

1.2 Research Questions and Hypotheses

The overarching research question guiding my dissertation is: *How can we provide engaging and authentic AI learning experiences for children?* I split this research question into four research questions:

RQ1 How does the iterative design process with both adults and youth drive the evolution of a learning system that empowers youth with conversational AI development?

RQ2 How do youth engage with a development environment designed to support them in making conversational agents?

RQ3 What features do youth desire in a learning environment to support their educational needs?

To date, my investigations into the first three RQs have been documented through a series of studies. In Chapter 3, I introduce AMBY, a novel interface that empowers children to create their own AI projects. In Chapter 4, I explore the experiences and perceptions of middle school-aged children interacting with AMBY during a summer camp. The study revealed that while learners were engaged and creative, they also encountered challenges, particularly with the labor-intensive data entry required for effective AI performance and a desire for more personalized chatbot responses. These insights informed the development of AMBY 2.0, which integrates the *entity* functionality feature (Chapter 5). Through our summer camp study in 2023, I observed learners' interactions with this new feature, which suggested the necessity of a more rigorous experimental study to evaluate its impact on AI learning experiences. In my proposed work, I plan to investigate RQ4 through a new classroom study.

RQ4 In what ways does the *entity* feature impact students' AI learning experience?

To conduct the new classroom study, I will first describe an updated, science-oriented conversational AI curriculum, which has been developed in partnership with three middle school teachers, containing examples related to previously learned science topics, and activities that are closely tied to students' previous science learning experiences. Partnering with the middle school teacher, I will conduct a classroom study in their science classroom. Students will learn conversational AI concepts, and work with a student partner to develop a conversational agent with relevant science topics using AMBY. The classroom activities will span 10 class sessions of approximately one hour each.

During the classroom study, my goal is to assess the effectiveness of the new entity feature in AMBY 2.0 (RQ4). Preliminary results from summer 2023 (Chapter 5) have indicated that this feature can aid in better design of agents and mitigate learners' frustration due to the repetitive entry of training data.

Hypothesis: I hypothesize that the entity feature will positively affect students' learning experience, as measured by **higher interest in conversational AI** (measured in post-survey), and the chatbots they create will demonstrate **higher project quality** (measured in four aspects: project ideation, conversation design, AI development, and end-user satisfaction), compared to students using AMBY without the *entity* feature.

Experimental Conditions. My goal is to investigate whether the *entity* feature impacts students' learning experiences. Therefore, I will conduct a between-subject experiment with two versions of AMBY: *AMBY with entity* and *AMBY without entity*. The student participants will be randomly assigned to either condition to use AMBY to create their conversational apps.

Outcomes. Research with AMBY has considered many outcomes, such as ability beliefs, identity (change from pre-survey to post-survey) [89] and learning outcomes. In my work, I am interested in interest formation and project quality. For **interest formation**, I will use triggered situational interest and maintained situational interest in conversational AI [3]. Based on Hidi and Renninger [40]'s interest development model, the development of interest contains four phases: 1) triggered situational interest; 2) maintained situational interest; 3) emerging individual interest; and 4) well-developed individual interest. Our classroom intervention specifically addresses phases 1 and 2 and hopes to trigger and then maintain students' interest in conversational AI as they find developing conversational agents to be

3 The interest formation measurements were collaboratively developed within the Camp DIALOGS team, led by the project external evaluator, Tom McKlin.

meaningful. For students' **project quality**, I will use a validated rubric (Cohen's Kappa = 0.751) to rate the project in the following aspects: 1) project ideation; 2) conversational design; 3) AI development; 4) end-user satisfaction. More details about these outcome measurements can be found in Chapter 6. As students in both learning conditions will complete the conversational AI development task, I do not expect there will be any significant difference in students' AI learning between the two groups.

1.3 Proposal Document Overview

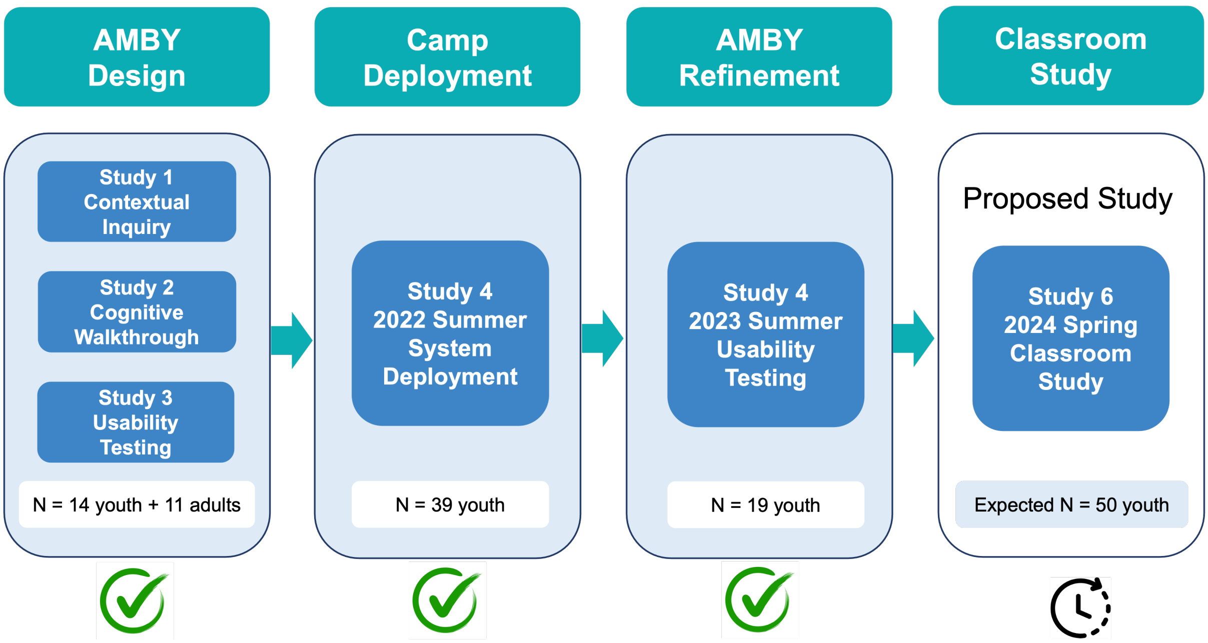


Figure 1-1. Proposed dissertation study overview.

Figure 1-1 provides a overview of my dissertation studies, which contains four distinct phases related to the iterative design and deployment of the learning environment that fosters AI learning: AMBY Design, Camp Deployment, AMBY refinement and the proposed classroom study. The structure of this proposal document is as follows:

Related Work (Chapter 2): I conducted a literature review on digital learning environments that support the teaching of conversational AI and, more broadly, natural language processing in K-12 settings. This chapter also reviews existing conversational AI development tools tailored for both adults and children and offers an overview of fundamental conversational AI concepts and terminology.

AMBY 1.0 Design (Chapter 3)[§]: Over the course of a year, we engaged in a series of

§ Work detailed in Chapters 3 and 4, titled "AMBY: A Development Environment for Youth to Create Converiterative design studies with middle school students and adults to conceptualize and create AMBY from scratch.

AMBY 1.0 Camp Deployment (Chapter 4): This chapter shares insights from the AMBY camp deployment study conducted in the summer of 2022. During this time, 39 middle school learners used AMBY to create 58 conversational AI projects. I explore the experiences and perceptions of middle school-aged children interacting with AMBY.

AMBY 2.0 Design and Refinement (Chapter 5): Stretching from Fall 2022 to Summer 2023, this phase focused on refining the AMBY environment, culminating in AMBY 2.0. Driven by insights from AMBY 1.0 user studies, we incorporated new features, including the *entity* feature, to enhance students' learning experiences. This chapter reports the results from our user study on AMBY 2.0 in Summer 2023.

Proposed Classroom Study for AMBY 2.0 (Chapter 6): The final stage will center around a classroom study, and the *entity* feature experiment.

sational Agents," has been accepted to the International Journal of Child-Computer Interaction, with me as the first author.

CHAPTER 2
RELATED WORK

This work stands at the intersection of AI in K-12 education and conversational AI development. This chapter first presents a literature review of digital learning environments for teaching AI (specifically Natural Language Processing) to youth. This is followed by an examination of existing conversational AI development tools for both youth and adults. Lastly, I introduce the fundamental conversational AI concepts and terminology.

2.1 A Scoping Review of Digital Learning Environments for Teaching Natural Language Processing in K-12 Education

Conversational AI, also known as dialogue systems, is a subset of Natural Language Processing (NLP). NLP is a crucial element in AI education due to its role in facilitating machine understanding, interpretation and generation of human language [96, 41]. The AI4K12 big ideas highlight many NLP tasks and applications for children to grasp. As per Touretzky et al. [95], students should understand basic NLP concepts such as speech recognition (big idea #1), word embeddings (big idea #2), parsing (big idea #3), text generation and sentiment analysis (big idea #4), as well as ethical concerns related to NLP applications (big idea #5). To foster students' growth from AI consumers into AI creators, learning environments must provide authentic, hands-on learning experiences [88, 64]. Such experiences can be facilitated through relatable NLP tasks that simulate real-world applications, such as creating personalized chatbots and exploring sentiment analysis models.

The objective of this study is to investigate the state of the art in digital learning environments for learning NLP in the context of K-12. I aim to characterize and compare the implementation and evaluation of these tools to identify gaps and potential opportunities for future conversational AI research.

2.1.1 Literature Search

The methodology for this scoping review is based on the framework outlined by Arksey and O'Malley [9] to search and review existing literature. Based on the research questions, I identified two main term-categories to include in the literature search: discipline (NLP-related) and target population (K-12). After performing multiple iterations of searches, I derived a list of relevant synonyms for each category. The search terms were applied over both the titles as well as the abstracts of the publications. In the literature search, we used the combined keywords from the two categories (Table 2-1). The complete search string was ("AI Learning" AND "AI Education" AND "AI literacy" AND "NLP" AND "natural language processing" AND "linguistics" AND "conversational AI" AND "dialog* system" AND "chatbot") OR ("K12" AND "middle school" AND "high school" AND "elementary" AND "primary school" AND "secondary education" AND "youth" AND "kid" AND "child*"). The actual string varied based on the restrictions of each database.

Table 2-1. Keyword list for literature search

Category (connected using "and" logic)	**Search terms (connected using "or" logic)**
Discipline	AI Learning, AI Education, AI literacy, NLP/natural language processing, linguistics, conversational AI, dialog(ue) system, chatbot
Target population	K12, middle school, high school, elementary/primary school, secondary education, youth, kid, child*

2.1.1.1 Sources

We searched the main digital databases and libraries in the field of computing, including ACM Digital Library, IEEE Xplore, ScienceDirect and Google Scholar. In addition, we also used Google search to account for the possibility that some educational tools may not yet have been published in scientific databases [76]. Since research on the topic of NLP education is fairly new, recent works are most often published in niche conferences and workshops, including AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI), Special Interest Group on Computer Science Education Technical Symposium (SIGCSE), Interaction Design and Children (IDC), ACM CHI Conference on Human Factors in Computing Systems (CHI), Computer-Supported Cooperative Work (CSCW), IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC). We performed additional searches in these proceedings to minimize the risk of omitting relevant works.

2.1.1.2 Selection criteria

Based on the goal of this literature review, following the criteria described in Tatar and Eseryel [92], the selection criteria for this literature review are shown in Table 2-2.

The initial literature search was conducted in March 2022. A second search was conducted in February 2023 to include any additional papers published between March 2022

Table 2-2. Selection criteria for NLP learning tools for K-12

Inclusion criteria	Exclusion criteria
K12 education (kindergarten through the 12th grade)	Other stages of education such as pre-university level, college, and graduate level
Empirical studies	Theoretical studies
Involves tools or technologies	Studies that do not involve digital tools (e.g., curriculum design, unplugged activities only, workshop design)
Report at least one form of assessment (e.g., learning outcomes, engagement, perception)*	Studies that do not provide assessment*
English publications	Non-English publications

Note. Criteria marked with an asterisk (*) were used for selection of papers answering the research question "How have researchers evaluated these tools in educational contexts?" only.

and February 2023. The searched papers were screened by scrutinizing their titles and abstracts to determine their eligibility based on the selection criteria. Because this field is still new, some learning environments are still works-in-progress and thus lack a published system evaluation. However, it is still important to include these tools for the completeness of this review. Through backwards and forwards snowball sampling, this review ultimately yielded 21 publications describing 11 learning environments.

2.1.2 What digital learning environments are available for NLP learning in K-12 education?

I identified 11 digital learning environments developed for NLP learning in K-12 education. These learning environments and corresponding evaluation studies are found in 21 publications (some systems involve multiple studies published as different papers). Below I will briefly introduce four learning tools that specifically aims to teach conversational AI concepts[1].

1. **Convo [114, 113]:** Two graduate students at MIT developed Convo for middle school students. As a conversational programming agent, Convo enables students to create deep learning-based conversational AI agents. It provides a learning environment that explores AI-driven communication systems and their applications.

2. **Zhorai [62]:** Created by researchers at both Harvard Graduate School of Design and MIT, Zhorai is a conversational agent that teaches AI/ML concepts through interactive

1 A complete set of learning environments for teaching NLP is available as in the preprint https://arxiv.org/abs/2310.01603

dialogue for young users (aged 8-11). It focuses on representation and reasoning, learning, and the social impact of AI.

3. **ConvoBlocks [99, 98, 100, 101]:** ConvoBlocks is a block-based programming interface developed by MIT for learners between the ages of 11-18. It offers a hands-on experience in training, transfer learning, large language models, intents, societal impact and ethics, speech synthesis, and speech recognition.

4. **Build-a-Bot [75]:** Developed by researchers at MIT, Build-a-Bot is an open-source tool designed for classroom environments. It introduces students to the NLP pipeline, which includes data collection and labeling, data augmentation, keyword filtering, intent recognition, and question answering, serving as a valuable resource for teaching AI concepts.

Most of these tools are available as web applications, which makes them easily accessible to anyone with an internet connection. Of the above tools that are publicly available, most are available for free, but some are restricted to registered users (e.g., NLP4All, eCraft2Learn) or require an API key to access them (e.g., Cognimates). 7 out of 11 tools allow users to deploy the artifact (a working application or a trained model) to an external site. Three tools do not offer external integration and one tool does not mention integration. Regarding language support, six tools only support English, while five support at least one more language in addition to English. Among those five tools that support more than one language, three tools offer multiple (10+) choices.

2.1.3 Key Findings, Research Gaps and Implications

Next, I summarize the key findings from this scoping literature review. In total, I identified 11 digital learning environments for NLP learning in K-12 education, with most being accessible as online web apps. However, some tools have limitations such as restricted access or language support, which may affect their usability for beginners.

The 11 digital learning environments for NLP in K-12 education primarily support text classification, speech recognition, and intent recognition tasks, with limited support for other popular NLP tasks. These tools offer varying capabilities for training and deploying NLP models and provide different data input modalities, such as keyboard and speech. However, the majority of the systems lack in-depth scaffolding and explanations for NLP processes, which could be improved for better learner understanding.

A majority of the studies employed mixed methods for evaluating their tools, with moderate sample sizes ranging from 3 to 135 (median = 29.5). Research studies were more often deployed in informal learning contexts than formal contexts.

Most NLP learning activities target middle and high school students, with evaluations focusing on AI knowledge assessment and learning experiences. These tools prove effective for teaching NLP and AI concepts, fostering interest, and improving students' understanding and engagement. However, learning challenges persist in machine learning and ethics concepts, and there is more work to be done in addressing issues of over-trusting technology.

The analysis of my literature review revealed six prominent gaps. First, there is a limited variety of NLP tasks, with a strong focus on natural language understanding (NLU) and limited exposure to other essential NLP tasks. Second, comprehensive evaluation methods for NLP learning tools, particularly for younger students, are still underdeveloped. Third, while many pedagogical systems provide explanations, they often fall short in offering intuitive insights and comprehensive understandings of NLP concepts. Fourth, there is a limited focus on younger children, with most tools targeting middle and high school students. Fifth, there is insufficient personalized learning experiences tailored to diverse learners' unique needs. Lastly, the literature lacks concrete recommendations for effective teaching strategies to incorporate NLP education efficiently in K-12 settings.

My dissertation directly addresses these gaps by developing and deploying a novel educational tool, AMBY, which broadens the variety of NLP tasks accessible to novice students through an engaging and intuitive interface. I have also developed a validated rubric to assess the AI artifact created on AMBY, which enhances the evaluation of NLP learning experiences for young learners. The system provides rich and interactive explanations that fosters deeper understanding of AI. Through the deployment of AMBY in middle school science classrooms, my research offers concrete, empirically-supported instructional resources for integrating conversational AI into K-12 education.

2.2 Conversational Agents and their Role in Learning

Conversational agents, or chatbots, communicate with users in natural language (text, speech, or both) [46]. With rapid advancements in the fields of AI and machine learning, modern conversational AI systems are robust enough to serve users in everyday life. A growing body of research is exploring how these systems can play a role in learning.

Druga et al. [25] specifically investigated young children's perceptions of, and interactions with, conversational agents, and proposed a series of design considerations to engage young children in the interaction. For instance, voice and prosody features were found to be decisive in children's perceptions of friendliness with agents. Hoffman et al. [43] found that children, as reported by their parents, tend to establish meaningful emotional connections with conversational agents, perceiving them as entities capable of feeling and eliciting emotions. Garg and Sengupta [29] explored children's and parents' perceptions of using conversational technologies for in-home learning, finding that children had high expectations for these devices' knowledge and capabilities for naturalistic interaction, and that parents found these technologies' potential role in learning to be desirable, while also wanting to monitor their children's usage.

Lovato and Piper [65] reviewed studies of children's voice-search technology use from developmental and human-computer interaction perspectives, and concluded that since children's question-asking serves a developmentally different and important role than the question-asking of adults, conversational interfaces should be able to identify child users and be prepared to respond to their questions in different, appropriate ways. In this spirit, Oranç and Ruggeri [71] explored how young children of different ages ask questions to conversational agents, finding that while all children could identify when answers were irrelevant, only older children, who were more familiar with conversational agents, tended to adapt their question-asking when an agent's answers were unhelpful. Similarly, Girouard-Hallam and Danovitch [32] investigated how young learners use conversational agents as information sources, and found that children's trust in conversational agents as information sources increased with age.

Some researchers have applied insights such as those described above to implement and evaluate novel interactive learning experiences using conversational AI. For example, Xu and Warschauer [108] embedded conversational agents into animated television programs to help children (ages 4-6) improve science learning by asking questions, providing feedback and offering scaffolding. Lovato et al. [66] engaged young children in creative storytelling with embodied stuffed animal agents to explore playful conversational agent design. These burgeoning efforts demonstrate the potential for conversational AI to support youth learning experiences.

A recurring theme in designing educational technologies for younger audiences is the importance of authenticity and meaningful engagement [88, 86]. Designing for younger demographics often involves direct collaboration with the intended age group, taking into account their needs and desires [20, 21]. In alignment with these prior work, my work with AMBY involved iterative design processes with middle school youth to ensure that the tool aligns with learners' interests and preferences while also addressing their social and educational needs.

2.3 Conversational AI Development Tools

There have been numerous efforts to foster learning *about* conversational AI. Many popular AI education platforms for youth have integrated specific modules that involve some aspects of conversational AI, such as Cognimates [23], LearningML [28, 80], ML4K (Machine Learning for Kids) [55], Zhorai [62] and eCraft2Learn [48]. However, most of these systems only allow users to engage with a subset of conversational AI concepts (*e.g.*, natural language processing) rather than allowing users to engage in building conversational AI applications themselves.

Currently, there are several robust tools developers have access to for creating conversational applications. These tools (e.g., Google Dialogflow [4], Rasa [2], IBM Watson [5, 27], Amazon Lex [68], Azure Bot Service [1], and Wit.ai [3]) offer a plethora of functionalities for skilled developers to create advanced conversational AI applications. However, these tools are not well suited for educational purposes that target young learners. Many features require extensive programming knowledge [81, 15] and were not designed for fostering AI learning in a robust and authentic manner for young learners.

There have been efforts to close this gap, designing systems specifically for young learners to learn about conversational AI by building it. For instance, Van Brummelen [98] introduced conversational AI modules within MIT App Inventor, enabling students to program Alexa Skills in a block-based programming environment. In a five-day workshop involving 47 students aged between 11 to 18, the researchers observed significant learning gains in general AI and conversational AI concepts. Zhu and Van Brummelen [114, 113], on the other hand, developed Convo, a conversational programming agent that enables students to create deep learning-based conversational agents. Through Convo's user study, the authors observed an increase in the participants' confidence in their abilities to build conversational agents.

Despite these advances, these tools still present limitations, particularly in supporting the design of sophisticated, multi-turn conversations, a cornerstone of conversational logic. Our novel interface, AMBY, aims to address this by incorporating dialogue concepts into the design process. Incorporating dialogue concepts into AI learning environments is critical as it gives learners a tangible understanding of conversational AI. This understanding aligns with the principle of natural interaction, one of the "Five Big Ideas for AI Education in K-12" outlined by Touretzky et al. [95]. This principle emphasizes the need for learners to understand how AI systems mimic human communication in an interactive and dynamic manner. Through engaging with these concepts, learners may develop a more nuanced understanding of how AI systems manage complex, multi-turn dialogues. Moreover, this approach may encourage critical thinking and foster communication skills as learners navigate diverse conversational scenarios, ensuring their AI responds appropriately. Additionally, AMBY offers the option to customize the agent's appearance and voice, a feature designed to enhance engagement and learning. Previous studies have indicated the significance of this capability [45, 39]. Another key distinction is that youth were actively involved in AMBY's design process. Unlike previous systems, our method ensured that the users themselves were involved in the iterative design process, allowing us to tailor the tool more effectively to meet learners' needs.

2.4 Conversational AI Development Concepts and Terminology

This section provides an overview of conversational AI development concepts involved in the task of developing conversational agents. A simple conversational AI system consists of several modules. It takes the user's speech and processes it in the *speech understanding module*, which converts the speech signals to text and infers the user's ***intent*** by matching the text with a pre-defined category[2]. For example, when a user says, "Can you suggest a movie to watch?", the *speech understanding module* processes the user input and identifies the user intent as "Request movie recommendation". After recognizing the user's intent, the *speech understanding module* sends this information to the *dialogue manager* to decide what action to take based on the user's intent and select from a list of ***responses*** to return to the user. For example, the "Request movie recommendation" intent might serve responses such as "You might like to watch *Owls of Magic*" or "My suggestion is *Wizards and Armies*". Once the response is selected, the system sends it to the *speech generation module*, which transforms this response into speech output and returns it to the user.

A conversational AI's intent recognition accuracy is largely constrained by the robustness of its training data (also called ***training phrases***). These phrases induce the model to capture different linguistic manifestations of the same intent. As a developer, authoring intents, associated training phrases, and responses are core activities to creating a conversational AI. Additional activities include authoring ***follow-up dialogues*** and creating ***fallback intents*** which are used when no other intent is recognized in the user's utterance.

2 Some conversational systems are textual and omit the speech recognition step as well as the speech generation module mentioned below.

CHAPTER 3
PHASE 1: AMBY 1.0 DESIGN

This chapter[1] describes my work which aimed to answer **RQ1: How can we design a learning environment that empowers children with AI learning through creating their personally relevant conversational AI projects?**

To develop AMBY, we[2] utilized an iterative design approach, working with youth at multiple design stages (Figure 3-1). This process consisted of three studies in total. Study 1, conducted in 2021, was a contextual inquiry (Section 3.1) during a summer camp with 14 youth. The feedback derived from this contextual inquiry and the literature-driven design principles (Section 3.2) informed the initial AMBY prototypes. We conducted two usability studies to pilot the system and identify potential issues. The first usability study, Study 2 (Section 3.3), was a cognitive walkthrough with expert reviewers. The second usability study, Study 3 (Section 3.4), was a think-aloud usability test with youth who had also participated in the contextual inquiry the year prior (Study 1). The AMBY 1.0 features and technical implementation is described in Section 3.5.

3.1 Study 1: Contextual Inquiry

Contextual inquiry is a widely used technique that consists of observing and talking with people in the context of performing specific tasks [78], which can inform the design of a system that will support an improved work experience for the target users [102, 57]. In this contextual inquiry study, our goal was (1) to investigate how youth learners use an existing conversational agent development tool, Dialogflow[3], to create their own conversational agent in a summer camp and (2) to identify their challenges and needs to accomplish their development goals. We chose Dialogflow for the following reasons: 1) it is free and publicly available; 2) it provides detailed documentation and guidance for small and simple

1 Work described in Chapter 3 and 4 has been accepted as a journal article. Reference: Tian, X, Kumar, A., Solomon, C. E., Calder, K. D., Katuka, G. A., Song, Y., Celepkolu, M., Pezzullo, L., Barrett, J., Boyer, K. E., & Israel, M. (2023). AMBY: A Development Environment for Youth to Create Conversational Agents. International Journal of Child-Computer Interaction, In press.

2 This work represents a collaborative endeavor involving all named authors listed above. As the first author, I took the lead in designing the AMBY interface prototypes, outlining the user interaction flow, and specifying the feature functionality. I also developed the research instruments used in the usability studies with adults and children (study 2 and 3) and was responsible for overseeing the data collection (study 2 and 3) and analysis (study 1, 2 and 3). The majority of initial manuscript draft was written by me.

3 https://dialogflow.cloud.google.com/

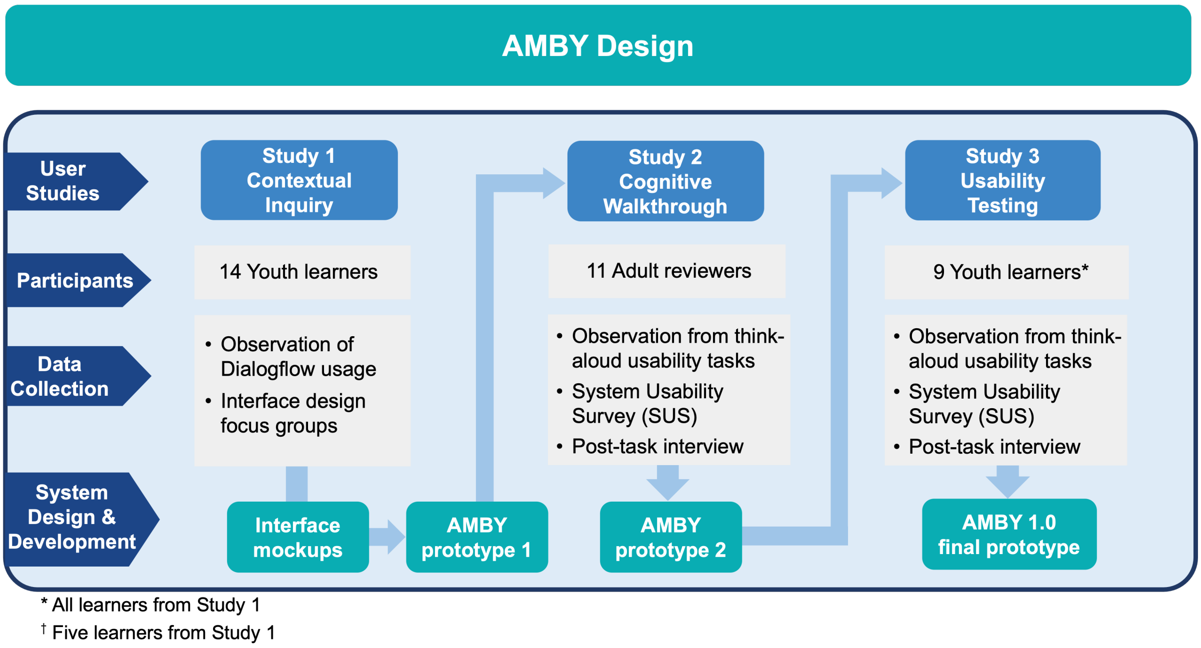


Figure 3-1. Overview of phase 1: AMBY design studies.

agent-development tasks; 3) it utilizes state-of-the-art language training models; and 4) it offers easy integration to other platforms, such as Google Assistant and Google Home devices.

3.1.1 Participants

In the summer of 2021, 14 youths attended the summer camp. Our participants came from a primarily Black community in the southeastern United States. We held the summer camp at no cost to their families at a local community center. Among the 14 participants, 2 identified as female and 12 as male; 11 as Black/African American, and 3 as White/Caucasian. The average age of the participants was 12.3 (SD = 1). Seven participants (50%) reported having no prior coding experience; the remaining seven (50%) reported having block-based coding experience (e.g., Scratch).

3.1.2 Camp Context with Dialogflow

During the two-week summer camp, students learned about foundational principles of artificial intelligence and conversational AI (Figure 3-2). In the first week of the camp, the participants learned about important AI concepts as they applied to Dialogflow, such as machine learning, conversational AI, intents, training phrases, responses, parameters, contexts and follow-up intents (these terms were defined in Section 2.4). In the second week, learners worked in pairs to build a conversational agent using Dialogflow, with a topic or purpose of

Figure 3-2. Left: Summer camp 2021 classroom. Right: Interface design focus group. Learners are presented with paper mockups, guided by a camp facilitator.

their choice. They integrated and tested their conversational agents with Google Assistant, as well as on a Google Home Mini device. The camp also provided CS/AI Unplugged activities [63] and social activities. Eight camp facilitators recruited from the researchers' university worked closely with learners on their project development and also reported daily observations, noting the challenges learners faced while using the Dialogflow interface. Facilitators observed the learners' behavior throughout each day, and documented any issues they noticed in a daily reflection entry. In the reflection entry, the facilitators responded to prompts such as "what went well today," "what can be improved, and how," along with any questions or concerns they had. Facilitators would have been familiar with some of the challenges that learners might be facing as Dialogflow novices, as none of the facilitators had had any Dialogflow experience prior to their own training in the weeks prior to camp. These facilitator reflection entries were carefully noted and examined together by two researchers to extract themes.

3.1.3 Dialogflow Challenges

This section presents our observations from the contextual inquiry with learners using Dialogflow during a summer camp.

- **Limited affordances for conversational AI design**: While Dialogflow can support sophisticated conversational app development, its interface does not support novices in applying conversational AI design concepts (Section 2.3). Learners rarely used the advanced features that were discussed in lessons and mostly used the basic elements of each intent (i.e., training phrases and responses).

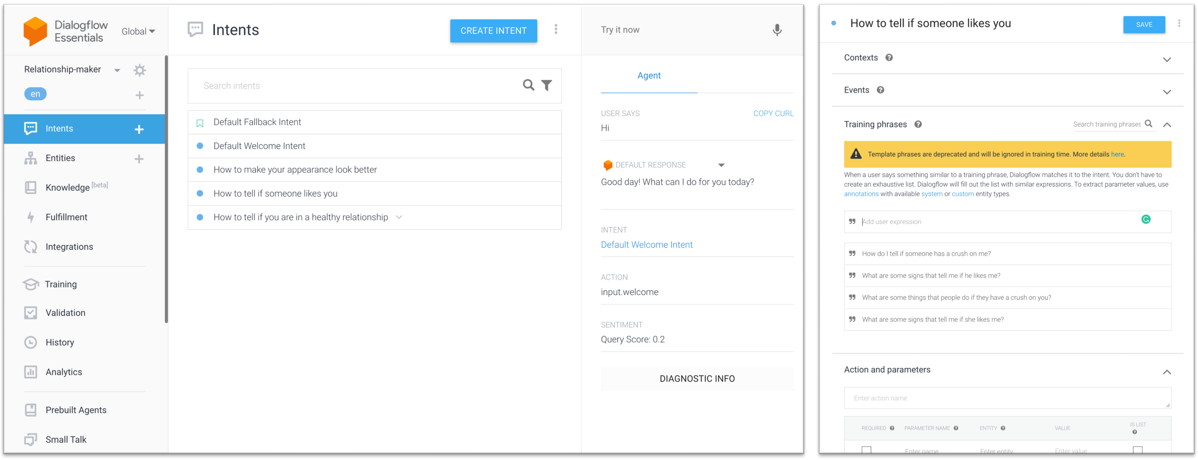


Figure 3-3. Dialogflow interface; Left: main development page for intents. Right: intent editing screen.

- **Overwhelming information from Dialogflow causes frustration**: Dialogflow's screens contain dense text (Figure 3-3), which appeared to contribute to learners becoming bored and frustrated. A substantial amount of their development time was consumed by navigating the interface and locating its relevant features.

- **Difficulty with typing**: Training the conversational AI requires entering a variety of potential user expressions (training phrases) for each intent. We observed that some learners had difficulty typing, which caused frustration and unwillingness to input enough data to effectively train the AI.

3.2 Design Principles and Initial AMBY Interface Mockups

Prior to the contextual inquiry study, we anticipated that young learners would face challenges with Dialogflow. Therefore, in the spring of 2021, in parallel with designing the 2021 summer camp curriculum that utilized Dialogflow, we also worked toward a paper prototype of a novel conversational app development environment for youth. Through a series of discussions within the research team and consultation with external advisory members, we derived four design principles from the existing literature on AI for K-12 and interface design for youth. These design principles guided us throughout the entire design cycle for the alternative interface, which we detail in Sections 3.2 through 3.5. The design principles were as follows:

1. **Foster an accurate conceptualization of conversational AI.** Some related work

suggests strategies to introduce young learners to machine learning [16, 115] and natural language processing concepts [23, 10, 42]. Similarly, as learners create and tinker with conversational AI, the system should represent AI concepts accurately, such as the importance of training data and design of conversational flow [64].

2. **Embodiment of AI agents.** Embodiment of a virtual agent can significantly improve children's engagement in a learning activity [11, 74, 45, 39]. Customization of the agent's embodied characteristics, such as gender, skin tone, and voice, can enhance learner's identity [52] and create a better sense of belonging, thus encouraging youth to engage more with the system [77]. However, agent customization options can also distract from the learning activity itself [60]. We therefore sought to balance the freedom of customization with the core cognitive tasks (e.g., designing the dialogue, creating intents, entering training phrases) afforded by the interface.

3. **Simplicity and age appropriateness.** Younger learners face lower cognitive load and report a better user experience when presented with large design elements [37], simple and intuitive displays [107, 91, 14], and concepts that are conveyed visually rather than with dense text [73, 56]. Thus, we aim to keep interface elements simple and interactive to maintain youth's attention.

4. **Flexible input modalities.** Research finds that interfaces supporting multimodal interaction are preferred over unimodal interfaces because of their flexibility to adapt to user needs [84, 35]. Multimodal interaction is especially beneficial for users developing conversational agents [85]. Our interface follows this path to provide flexible input methods (e.g., typing and voice) to improve input efficiency and adaptivity.

Drawing from the above design principles, especially the agent embodiment and simplicity, we drafted two initial interface mockups (Figure 3-4). The two interface mockups pared down the information in DialogFlow and displayed it in graphical form inspired by Blockly[4]. To elicit feedback from learners on these initial interface designs, we presented them as paper-based mockups in focus groups at the end of the 2021 summer camp. Each

4 https://developers.google.com/blockly

focus group, which comprised 3-4 youth participants, was moderated by one camp facilitator and was audio-recorded. These recordings were subsequently transcribed manually for analysis. Initial open coding of the responses was performed independently by one researcher, who then engaged in a collaborative discussion of the emerging themes with the other researchers during a group meeting.

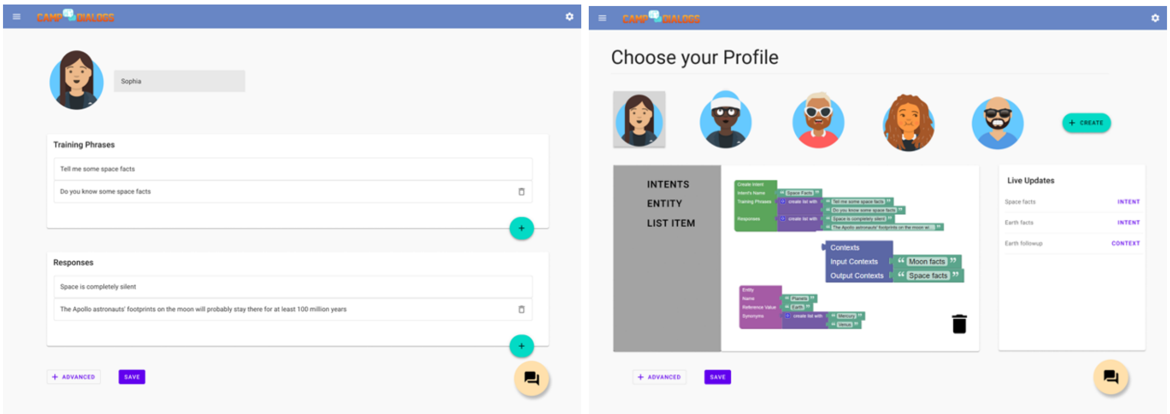


Figure 3-4. The two interface mockups used during the focus group in the contextual inquiry study (Study 1)

3.2.1 Findings from AMBY Paper Prototype Focus Groups

In focus groups, participants spoke to a desire for a streamlined interface that supported agent avatar customization. When discussing the simplified Dialogflow-inspired mockup (Figure 3-4 left), many participants agreed that such a simplified interface would help them focus on creating their agents. When considering the block-based interface (Figure 3-4 right), learners who had prior experience with block-based coding felt the interface could require more time to learn for users without such experience. For both mockups, learners were able to identify key features and functions. All the participants expressed interest in the option to select an avatar to represent their agent.

3.3 Study 2: Cognitive Walkthrough with Adult Reviewers

Based on the findings from the contextual inquiry and paper prototype focus groups, we iteratively refined a series of wireframes using feedback from our entire team, including camp facilitators, K-12 instructional designers, and university researchers in computer science and educational technology. We used these wireframes to implement the first prototype of AMBY, and then conducted a cognitive walkthrough study. A cognitive walkthrough is an expert review method in which interface experts simulate users “walking through” a series of tasks to identify potential issues and new system features [57, 67].

The 11 cognitive walkthrough reviewers included 8 members of the authors’ HCI research lab and 3 researchers specializing in educational technology and computer science education (note that the cognitive walkthrough reviewers’ association with the authors may have limited their willingness to provide honest feedback). Among the educational technology researchers, two had over 20 years of experience in instructional design and technology for youth, and the other had 3 years of experience in the field. The HCI team comprised two senior researchers each boasting 15 and 8 years of experience in HCI and dialogue systems research, three with over 3 years of experience, and another three with more than 1 year of experience. Out of these HCI researchers, 6 had done graduate coursework on dialogue systems and had experience developing conversational agents using modern dialogue system frameworks. Though non-representative users, these reviewers were able to use a conversational agent development interface to perform tasks that a typical interface user would need to accomplish, thereby identifying potential design and usability issues.

The cognitive walkthrough study was conducted online through Zoom and lasted approximately one hour. Each reviewer was guided by one researcher to complete four think-aloud tasks using AMBY. The tasks were as follows: create an agent of their choice; edit an existing system intent (the “greet” intent); create a new intent; and create a follow-up intent. In the post-task interview, participants discussed the challenges they faced during the tasks and provided feedback on different interface elements. After the study, researchers discussed their observational notes until they arrived at a consensus on key user needs.

Users encountered no major issues with the fundamental design of the interface and could complete all development tasks within the study’s timeframe. Reviewer feedback was used to improve the visual design, such as giving the system default intents unique colors and positions for better clarity, and to simplify the interface text and improve linguistic consistency, as well as to improve usability with functionalities like alert messages and a button to “clear” the chat transcript in the testing panel.

3.4 Study 3: Usability Testing with Young Learners

We updated AMBY prototype 1 based on the cognitive walkthrough study. To assess the usability of the updated prototype (prototype 2), we proceeded to conduct a think-aloud usability study with representative users.

The participants were nine middle school learners who had attended the summer camp in 2021. Former participants were recruited because they were familiar with the fundamentals of conversational agent development.

The study was conducted as an after-school, two-hour, in-person workshop located at a youth educational center. The study procedure was similar to the cognitive walkthrough study, with the consideration that the tasks would take more time for youth to complete than for adults. Before starting the usability tests, participants were given a 20-minute refresher lesson that reviewed necessary conversational AI concepts. After the refresher lesson, participants were then divided into small groups to complete the tasks, guided by researchers. Each researcher guided one or two participants during the session. Participants' interactions and post-task interviews were both screen and audio recorded, with parental consent and learner assent.

During the post-task interview, participants reported liking the AMBY interface's aesthetics. They suggested adding more avatar choices including a way to customize the agent's voice to convey an emotion or embody a character. All participants were able to finish the task in the allotted time. We noted a few common difficulties: it was not clear to learners that progress would be lost when exiting the intent editor if "Save" was not clicked, and learners had trouble distinguishing the training phrase and response entry fields from one another. We modified the system's behavior and visual design to alleviate the identified issues.

3.5 AMBY: A Conversational App Development Environment

In this section, we present the final prototype of AMBY. We describe the system features and the technical implementation of the software.

3.5.1 AMBY 1.0 Prototype Features

When users first login to AMBY, they land on the Dashboard (Figure 3-5, left), where they can (1) create a new AI project, (2) import an AI project from local files, (3) open

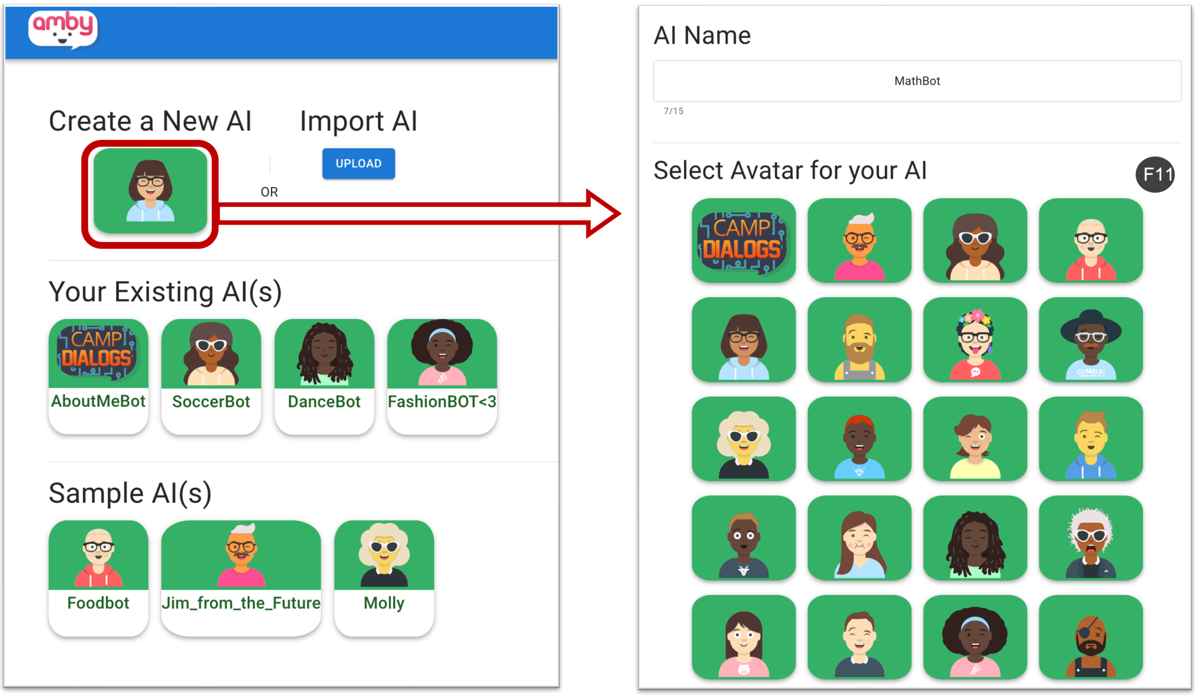


Figure 3-5. Left: AMBY dashboard page. Users can create or import a new agent, select an existing agent, or tinker with sample agents. Right: The agent creation window with a collection of avatars that the learner can choose from. Based on focus group insights, avatars anchor the user's first experiences upon launching AMBY.

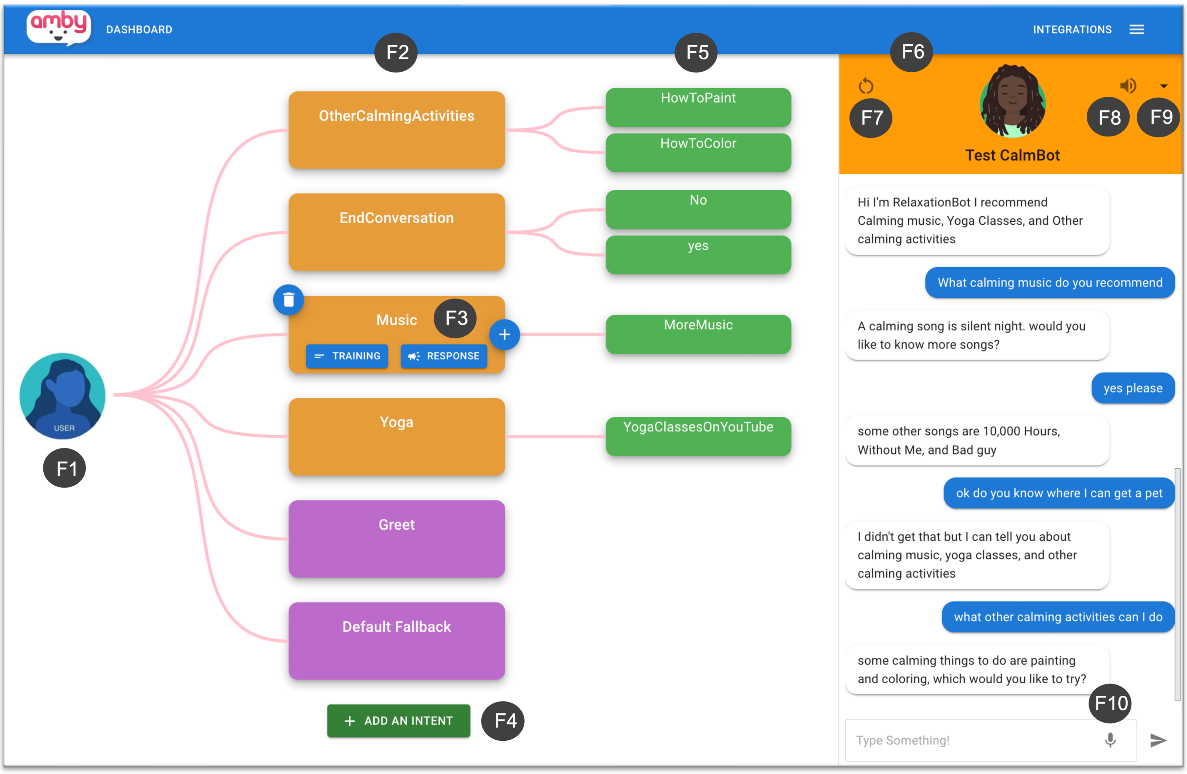


Figure 3-6. AMBY playground page. F1-F10 depict specific interface elements, which are detailed in Section 3.5.1.

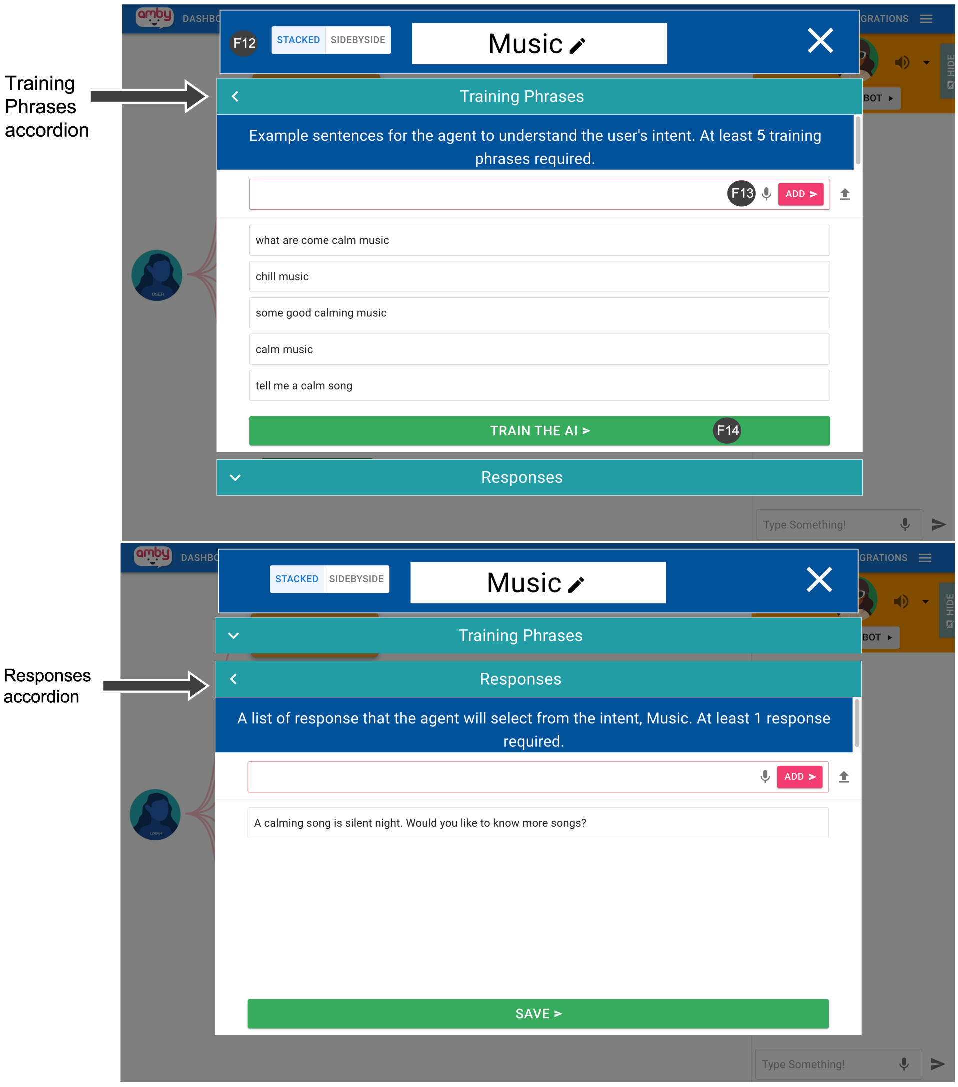


Figure 3-7. Intent editing window (stacked view) for training phrases and responses

previously created projects, and (4) open sample projects available on the website. If they opt to start a new project, they first select an avatar to represent it (Figure 3-5, right). Once the user has selected or created a new project, they are then directed to the Playground page (Figure 3-6), where they can develop, and test their agent. From the Playground page they can also deploy their agent on a Google Assistant-compatible device.

Choice of avatar selection for conversational agents. Although an avatar is not required to deploy a conversational agent on most smart speakers, such embodiment can be helpful for youth to design persona and enhance engagement [11]. AMBY provides a menu of avatars (Figure 3-5) for the users to represent their agents. There are 19 human avatars of different ages and genders and with different skin tones, clothing, accessories, and facial expressions. There is also one non-human avatar, a logo of the summer camp.

Visualization of dialogue flow. AMBY allows users to create a conversational agent simply by specifying intents, training phrases, and responses. The main development page (Figure 3-6) utilizes a card-based tree design to visualize the dialogue structure (as opposed to a block-based development environment). The conversation tree begins at the end user[5] (F1) and branches out first into the main intents (F2), one of which the end user must invoke before any of the follow-up intents (F5) can be activated. By including the app's end user in this representation, we aim to emphasize the conversational AI concept that intents represent the end user's implicit or explicit goal at any moment in the conversation.

Intents in the tree are represented by simple cards labeled with the intent's name. Options for interacting with an intent card (F3) appear on mouseover. User-generated intents are colored yellow (for main intents) and green (for follow-up intents). AMBY is built on Dialogflow, which generates two default intents ("greet" and "default fallback") that serve special purposes and have unique properties, so these intent cards are colored differently (purple). Follow-up intents (F5) can only be added to a main intent by clicking the "+" button on the right. Once these follow-up intents are created, they are visually connected to their parent intent, rather than directly to the end user, indicating a conditional conversational flow. AMBY users can create an unlimited number of main intents and a maximum of three follow-up intents per main intent. We limited the number of follow-up intents to support a simple visual design and encourage learners to be more strategic about designing the flow of their conversational app.

Intent editing window. When the user clicks the "Training" or "Response" button on an intent card (F3), AMBY displays an intent editing pop-up window, or modal (Figure 3-7). Inside the modal, the user can add, edit, or delete training phrases and responses for the specific intent. Users can toggle how training phrases and responses are displayed in the modal (F12): side-by-side or vertically stacked.

AMBY requires users to enter at least five training phrases before the intent can be saved. This is in alignment with our design principles: while Dialogflow has no minimum

5 In this paper, "User" refers to youth who are developing a conversational AI using AMBY. "End user" refers to a person who is interacting with or testing the conversational AI the youth built.

requirement, AMBY seeks to foster AI understanding by encouraging learners to generate multiple variations of potential user expressions, which also helps minimize the frustrating experience of diagnosing under-trained intents. On the other hand, too many required training phrases could create a situation where learners struggle to generate enough linguistic variations. The five-phrase minimum is a compromise between highlighting the importance of good training data and accommodating the language level and patience of youth.

Agent training/learning animation. We use animation to visualize the agent "learning" from the training process. In the intent editing modal (Figure 3-7), once learners have entered at least five training phrases, they can click the "Train the AI" button (F14) to save their changes. When a learner clicks "Train the AI", AMBY shows an animation (Figure 3-8) in which the agent's avatar is gradually encircled by a progress ring. When the ring is filled, a light bulb appears above the avatar's head, conveying that the agent has successfully learned the new training phrases. No animation is shown when saving responses, to illustrate the distinction that the machine learning model *learns* from training phrases to recognize similar expressions, but repeats response(s) exactly as the developer has entered them.

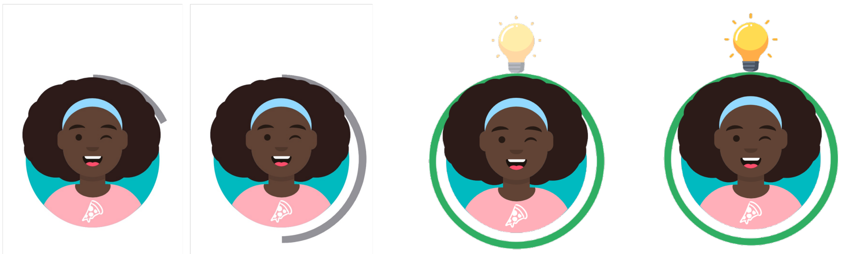


Figure 3-8. The agent learning animation (triggered by the "TRAIN THE AI" button (F14) in Figure 3-7)

Testing panel. Following from common block-based programming environment designs (e.g., Scratch, Snap*!*), the testing panel (similar to an output console or "stage") is on the right of the screen (F6, Figure 3-6). Users can test the agent instantly while editing the intents. The testing panel contains the avatar of the user's agent, a *clear chat history* button (F7), a *mute/unmute* button (F8), and an agent voice customization drop-down menu (F9). In the user text entry box, there is a microphone button (F10) that enables voice-based interaction.

Voice as an input modality. We observed that for some learners, typing was a barrier to using Dialogflow (see Section 3.1.3). Thus, AMBY supports voice-to-text as an input modality. When entering training phrases, system responses, and “user” dialogue for agent testing, learners have the option to use voice-to-text by clicking a microphone button on the screen (F10, Figure 3-6 and F13, Figure 3-7).

Agent voice customization. In response to feedback from usability testing with returning participants (Study 3), where it stood out as a desired feature, AMBY provides features for the user to customize their conversational agent’s voice (Figure 3-9). The voice can be customized along three dimensions: gender (male or female), pitch (-20 to 20 semitones), and speech rate, or speed (0.25 to 4).

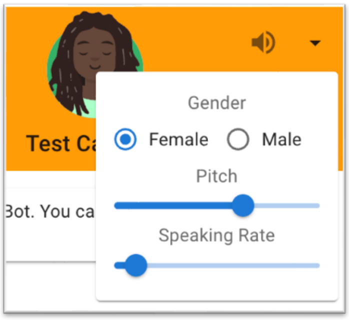


Figure 3-9. Voice customization drop-down menu

3.5.2 Technical Implementation

AMBY is an interactive web application built as a user interface for Google’s Dialogflow, which has a robust natural language understanding model, publicly available APIs to facilitate conversational AI management, features for speech and voice modulation, and connectivity with Google Assistant compatible smart speakers and devices. AMBY is developed using the MERN stack (MongoDB, ExpressJS, ReactJS, and NodeJS) and consists of four main components (Figure 3-10): client-side (front-end), Dialogflow interactions, server-side (back-end), and database[6].

The React-based front end handles user login and allows users to see, manipulate, train and test their conversational app. A user’s conversational app itself is constructed behind the scenes in Dialogflow; AMBY’s front end communicates with Dialogflow using Google’s

6 The technical implementation of AMBY was mainly done by Amit Kumar, John Tran Hoang, and Sunny Dhama, all from the University of Florida. Their technical contributions were invaluable to the development of AMBY and the facilitation of this research. My role primarily involved leading the design of the user interface, interaction workflows, and feature prototypes, alongside architecting the database for interaction logs.

publicly available APIs. Once the user has trained their conversational AI, the app can be deployed to a Google Assistant-compatible device in a few steps.

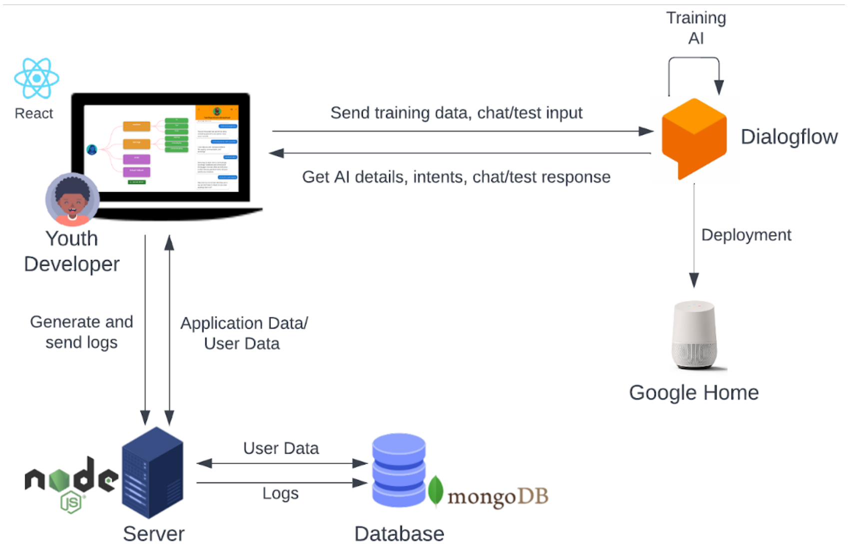


Figure 3-10. Technical implementation architecture of AMBY

CHAPTER 4
PHASE 2: AMBY 1.0 SUMMER 2022 DEPLOYMENT

We deployed AMBY to a two-week AI summer camp where it was extensively used for nine consecutive days. This camp deployment helped us investigate how well AMBY supports youth with little computing background or conversational AI experience as they learn to create their own personally relevant conversational agents, both individually and collaboratively. This chapter is guided by the research question, **RQ2: How do youth engage with a development environment designed to support them in making conversational AI?** I answer this question by analyzing several sources of data: 1) the conversational AI projects learners created using AMBY (Section 4.2.1); 2) learners' experiences using AMBY (Section 4.2.2); 3) learners' usage and perception about the features of the interface (Section 4.2.3); and 4) learners' common challenges using AMBY to develop conversational agents (Section 4.2.4).

4.1 Study Procdure

4.1.1 Participants

In summer 2022, 38 youth (P1-P38) attended the summer camp[1]. Among these 38 participants, 19 identified as female and 19 as male; 31 were Black/African American, five were Hispanic/Latinx, four were White/Caucasian, one was Asian and one prefer not to say[2]. The average age of the participants was 12.7 (SD = 0.7) and all participants were rising seventh or eighth-graders in the upcoming school year. 14 participants (37%) reported having no prior coding experience; 24 participants (63%) reported having experience in at least one type of coding environment such as block-based coding (e.g., Scratch), robotics (e.g., Lego Robots), or text-based coding and app programming (e.g, App Inventor). Among these learners, five had attended the project's summer camp in 2021 (study 1); one attended both the 2021 camp (study 1) and the usability testing (study 3). All parents completed consent forms for data collection prior to camp, and learners provided assent at the start of camp.

1 We host two camp sessions in summer 2022, 17 youth participated camp session A and 21 participated session B.
2 Participants could identify as more than one race/ethnicity.

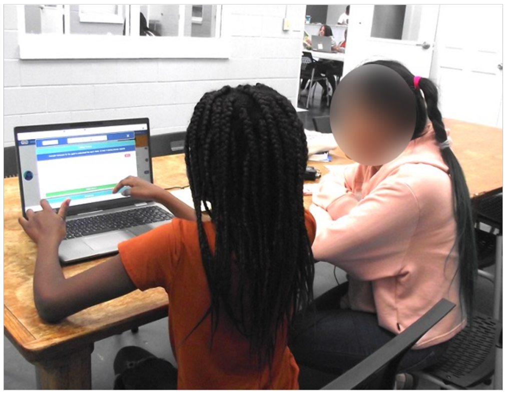
 

Figure 4-1. Left: Learners work on their individual projects, mentored by a camp facilitator. Right: Learners work on their paired project.

4.1.2 Study Description

AMBY learning activities spanned eight days over two weeks of the camp. Learners followed a "use-modify-create" progression approach [58] with AMBY. Specifically, on their first day using AMBY, learners used example projects created by the camp facilitators to become familiar with the AMBY interface. On the second day, they learned to modify an example project, "About Me Bot," so that the bot would tell its users fun facts about themselves (the learner). Then, they were guided step-by-step to create a conversational agent from scratch. On days 3 and 4, the learners developed their individual projects with hands-on help from camp facilitators. Beginning in the second week (days 5-8), they worked in pairs to develop another conversational agent relevant to both partners' interests. At the end of the camp, learners showcased their projects to their peers and family members on the Google Home Mini device.

4.1.3 Data Collection and Analysis

During the camp, learners were introduced to design thinking and engineering design processes [93, 8]. We provided a **design log document** (Appendix A) in which learners were asked to articulate their design ideas in seven steps: empathize, define, ideate/brainstorm, prototype, test, modify, and share. We used these documents to extract the ideas and themes found in the learner-created projects.

AMBY also collected **logs of learners' interactions** with the interface. Relevant log actions reported in the paper included: 'create a new project,' 'create a new intent,' 'press the microphone button to enable voice-to-text,' and 'send messages to the agent.' We used the log data to better understand how learners used AMBY's features and their challenges.

We conducted individual **interviews** with 13 learners from camp session A who attended on day 4 when individual projects were completed. Each interview lasted about 15 minutes and focused on their experience using AMBY for their project and their perception of the embodiment of their agent. On day 8, after learners finished their paired projects, we conducted 30-minute focus groups. We asked 15 learners (three or four per group) about specific features of the interface and solicited suggestions for improvements. Both interviews and focus groups were audio-recorded and manually transcribed by researchers.

We utilized a content analysis approach [44], specifically an inductive coding process [26], to analyze the interview and focus group data. This method is prevalent in HCI literature [19, 47, 51]. First, one researcher (primary coder) conducted open coding on all of the transcripts. Then, the primary coder met multiple times with another researcher (secondary coder) to review and discuss the codes and resolve any disagreements. Finally, the primary and secondary coders worked together to derive themes from the codes until they reached an agreement. The results of this data analysis speak to learners' experiences and their challenges using AMBY.

4.2 Findings

4.2.1 Conversational Agents Created Using AMBY

In total, learners used AMBY to create 25 conversational AI projects, including 18 individual projects and 7 group projects. Each project's name and the description provided by its creator(s) are shown in Appendix B-1. Projects were clustered into themes, using the answers learners wrote in the provided design document template (e.g., Who will use this app? What will this app do?) as well as the conversations their chatbots facilitated. The six major themes were as follows: fashion/shopping, personal/joke, mental health/boredom, educational/knowledge, sports/hobby, and task-oriented. Note that one chatbot may belong to multiple themes. For the scope of this paper, the lead author categorized the projects.

Among these projects, we select two examples that illustrate how learners were able to express themselves using conversational AI.

Example 1. Black History. This conversational agent, named Jerry Berry, was built collaboratively by two African-American male learners, to teach people about black history and influential black figures including Martin Luther King Jr., Barack Obama, Al Green, Harriet Tubman, and Rosa Parks. During their project demo, they shared the motivation for their conversational app idea:

> *"... Our design represents black power. Black power is something we need..."*

In addition to populating intents with historical facts, the learners also effectively utilized conversational markers to achieve a more natural user experience. For example, they broke up the description of each historical figure across multiple intents. The pair used a connecting phrase, "Would you like to know more?", at the end of each agent response, and provided the follow-up intents to handle "Yes" or "No". Their conversational agent also contains intents that handle social utterances, such as "thank you" and "bye", and an intent handling requests for "help" that describes what the chatbot can do and directs the user in how they might start a conversation. These learners showcased their strong conversational design skills in this personally and socially relevant project.

Example 2. Supporting Mental Health. This was a popular theme, addressed by five of the learners' projects. Many of these aimed to talk to people about their feelings and gave advice on coping with different emotions. Learners said they created the projects mainly due to their personal experience dealing with emotions as middle schoolers, but one learner also indicated its relevance to her career goal. P16 (female) created the conversational agent "ReachOutAndGrabaHand" with the capability to talk about negative emotions (e.g., angry, sad) and give advice on communicating with a partner. She stated that

> *"I created a therapy bot because when I grow up, I want to be a therapist ... [People would like] having a robot that's programmed to be a nice human, instead of judging. It's easier to talk to that instead of talking to a person that can go back and tell someone [else]."*

These youth were able to use conversational AI to explore and express empathy and think about solutions to salient problems in their lives.

4.2.2 Learners' Experiences Using AMBY

Here, we report on the learners' comments in focus groups and interviews.

Overall engagement. Overall, learners enjoyed using the tool to create conversational agents on their own. They expressed that AMBY gave them the freedom to create their personally relevant projects. In two participants' words:

> *"It lets you choose the responses ... how it lets you do what you want to and that it doesn't tell you what to do."* - P7 (female)
>
> *"[I like] creating and adding the intents because it's fun to make your chatbot respond to anything."* - P11 (female)

Learners also mentioned that they liked the testing window on the interface, which allows them to test on the fly.

> *"I like that you can add your own intent and you can test it right away to make sure it works."* - P4 (male)

Five learners from this study also attended the camp in the summer of 2021. All felt that using AMBY was easier and more engaging than Dialogflow. One returning participant, P2 (male), created his chatbot to be a representation of his own appearance and personality. Over the course of the camp, he had put significant effort into developing his individual agent and stated that in AMBY, *"the avatar, the voice, everything"* were better than the Dialogflow interface he had used the previous year.

Control over the AI. All the interviewed learners thought the agent they created was intelligent, and that because they were the ones who added (e.g.) *"information"*, *"knowledge"*, *"questions and answers"*, *"A lot of training phrases"*, or *"more intents"*, they were also in control of the agent's intelligence. P15 (female) mentioned that she *"made it smarter by adding wrong spellings of certain words, so it would still recognize it"*. P13 (female) emphasized the agent's machine learning ability and said she believed that *"if you work on it enough, it could be smart enough to work on its own."*

4.2.3 Learners' Usage and Perception About the AMBY Features

Agent embodiment: Voice customization. Of the 13 learners interviewed, 11 had used the voice customization feature. Six reported that customizing the agent's voice was helpful in conveying its personality. P2 (male) said, *"If you want it to be funny, you give it a high pitch voice"*, while P11 said that to show her agent's *"nice and caring personality"* she decided to *"make it a very soft, squeaky voice."* Further personifying her agent, she also represented excitement in her agent by adding emojis to its text responses:

> *"I made it speak with a bunch of emojis so the user knows what the bot is feeling.*
> *"* - P11 (female)

Agent embodiment: Avatar selection. When asked why they chose a specific avatar for their project's agent, 7 learners reported they picked the avatar because it looked similar to themselves; 5 reported they picked the avatar based on their target end user (e.g., P16 chose the "pirate"-styled avatar with an eye patch for her therapy bot, "ReachOutAndGrabaHand", because she thought *"he would need someone to talk to"*). One learner reported that they had picked their avatar at random.

Voice-to-text feature usage. Next, we investigated how learners used the voice-to-text feature in AMBY for authoring and testing the conversational agents (F10, Figure 3-6 and F13, Figure 3-7). Across 18 individual projects, we found 12 projects used voice-to-text for sending testing messages, six for creating responses, and three to create training phrases.

Although the voice-to-text feature was not used by all learners, it did significantly address some specific learners' needs. For example, one learner (P6, male) utilized voice-to-text frequently for training phrases, responses, and chat testing for both his individual and paired projects. Using the voice-to-text feature, he entered almost twice as many testing messages by speaking (65 messages) as he did typing (34 messages).

4.2.4 Common challenges using AMBY to create conversational agents.

While learners enjoyed the creative freedom of their projects, their most commonly reported challenges also stemmed from the creation of content for the agent. For example, P3 (male), who made a boxing coach agent, said, *"I had to search up things about boxing to use it on AMBY."* P7 cited *"the fact you have to write a lot"* as a difficulty: she had made some revisions that required her to rewrite many training phrases and responses. Generating ample, sufficiently varied training data to recognize each intent was also reported as a common difficulty. P8 (female) said her biggest challenge came from,

> *"knowing what the user was gonna say, and word[ing] it a bunch of different ways for training phrases."*

Another challenge for the learners was interpreting the intent classification failure. When the agent cannot confidently match a user utterance to an existing intent, the only output the tester receives is the default fallback response. It is up to the developer (the learner) to infer what has gone wrong, and many learners found the limited feedback to be a frustrating challenge.

Finally, a number of learners reported problems with system instability such as system lagging or no response. In part, this can be attributed to the limitations of the Dialogflow API for handling high-volume request calls as well as to slow internet speeds at the camp location.

4.3 Discussion and Design Implications

The results from our summer camp deployment suggest that youth learners can successfully create personally relevant conversational agents using AMBY: the projects that learners created using AMBY covered a variety of themes and interests, and learners reported positive experiences during interviews and focus groups, despite also facing challenges. In this section, we discuss the design implications from our effort to create a conversational AI development interface for young learners. We hope these implications will stimulate continuing conversation within the research community about future trajectories for learning technologies that support AI education for youth.

4.3.1 Interfaces Should Be Low-entry, But High-ceiling

Numerous studies have emphasized the importance of offering a low barrier of entry to novice learners [38, 34]. The low-entry interface we designed allowed learners with no prior coding experience to create relatively complex conversational agents, compared to those created in the summer of 2021 by learners using Dialogflow, which was not designed for use by novices, for the same task. In summer 2021, using Dialogflow, the average number of intents learners created was 4.71 (*SD* = 1.67), consisting of an average of 3.71 main intents and 1 follow-up intent. In contrast, in summer 2022, using AMBY, the first-time learners[3] made 15.88 intents per project on average (*SD* = 11.5), with an average of 9.88 main intents and 6 follow-up intents, which represents significantly more complex projects.

Interfaces that support conversational agent development should also be high-ceiling. Considering the display size of a laptop screen, AMBY only supported two layers of intent (one layer of main intent and one layer of follow-up intent) in this study. Learners suggested adding the capacity for more levels of follow-up intent to meet their project needs. For example, P17 (male) was an advanced learner who wanted to create a tic-tac-toe game. He calculated that implementing this game would require creating 81 total intents, including at least two layers of follow-up intents, which the AMBY environment could not support.

Some literature suggests that responsive interface elements can be more welcoming [7]. Our participants also spoke to this notion, suggesting that AMBY should allow them to collapse and expand subtrees of follow-up intents, or "*move them [intent cards] anywhere, like [from] a [main] intent to a follow-up intent.* (P14)" To employ another common strategy, the interface could be made more flexible by collapsing the advanced features into a different module, and de-emphasizing the advanced module to novice learners; the module might even be "locked" until the learner has completed certain basic tasks in AMBY.

4.3.2 AI Development Environment for Learners Should Be Transparent

A pedagogical system for conversational AI development should be transparent about how the AI represents knowledge and makes decisions. In our context, we directly represent the agent's knowledge by visualizing the dialogue structure, and we reinforce the agent's way of learning implicitly by scaffolding the intent creation process and explicitly with the learning animation. However, our system can be further improved by adding more transparency to the agent training and intent classification processes. As discussed in the findings (Section 4.2.4), one main challenge the learners faced was understanding intent classification. As P7 (female) said, *"it would be helpful to see exactly what the bot does not understand."* Learners reported that it would be helpful for the system to locate intent

3 excluding returning participants, whose prior experience with Dialogflow would likely impact their projects' complexity

classification mistakes and scaffold their understanding of the AI's decision-making process. This design implication maps to AI literacy competencies, specifically those regarding understanding knowledge representation and how computers reason and make decisions [64]. Literature suggests that graphical visualizations and interactive demonstrations of models can aid a better understanding of AI [53]. For conversational AI development interfaces, specific design considerations for transparency would be to include the intent classification results for learners who desire to inspect it. Similarly to existing interactive tools for exploring natural language processing techniques [42, 10, 31], the interface could also highlight important words or phrases which the system weighted more highly in order to aid in learners' understanding of the computer's representation of natural language [69].

4.3.3 Interfaces Should Foster Users' AI Learning Experience

The findings of this study suggest that interfaces should prioritize the ability of users to showcase their knowledge and skills in relevant and meaningful ways through the projects they create. Prior research has shown that people are more likely to identify with a learning experience that is culturally relevant and reflects their community [22]. The projects created by the learners exemplify this. Design features that enable such personalization, such as agent embodiment with avatar selection and voice customization, has facilitated this user expression. For instance, from our study, a majority of learners customized their agent's voice to convey a certain personality, many chose avatars that resembled themselves or were related to the theme of their project, showing the significance of personalization and its impact on user engagement. Beyond personalization, it's also evident from section 5.4.1 that the choice of project themes can stem from deeper, personal or societal motivations. Voice-to-text feature usage offers another insight: interfaces should provide diverse interaction modes to cater to different learner needs. The primary implication here is not just about embedding personalization features, but about deeply understanding and integrating learners' backgrounds, motivations, and experiences in AI learning tool designs. There is a tremendous opportunity for future research to further investigate how learners' backgrounds shape their interactions with AI tools, and how these tools can be refined to foster a more enriched and engaged learning experience.

4.3.4 Interfaces Should Empower Users To Incorporate Multimedia

In the study interviews, many learners indicated a desire to include multimedia in their agents' responses. For example, one participant wanted their agent to be able to provide images and videos to demonstrate the dance moves it was designed to talk about, and two others, both of whom independently created music recommendation agents, said they would have preferred if their agents could play music, rather than simply naming songs. While these are currently beyond the scope of AMBY, working with multimedia has been shown to foster creativity [97] and learner engagement [79], and there has been some research into multimodal dialogue systems [61, 90, 83]. There are existing tools such as Adaptive Cards[4] which may be easy to implement for adding multimedia support; however, such support has to be adapted to youth needs. Future efforts to create conversational AI development systems for youth should consider enabling users to embed multimodal content into agent responses, or potentially even automating connection to appropriate APIs.

4.3.5 Limitations and Future Work

This study has several limitations. First, due to the nature of the summer camp format, we are unable to measure participants' AI learning as a result of using AMBY alone. Although learners used AMBY extensively throughout the two-week session, they also engaged in other types of learning activities. It would be interesting to see how AMBY could be utilized outside of an informal, camp context to support different learning tasks. For example, a middle school science teacher might introduce AMBY in their classroom to assign students to create quiz bots on science content to support learning objectives.

Another limitation is that we did not evaluate the effectiveness of AMBY in a controlled experiment. As mentioned in section 2.3, currently there is no conversational AI development tool that can achieve the same tasks as AMBY that are developmentally appropriate for youth. Our results have demonstrated the extent to which youth created more sophisticated projects using AMBY compared to DialogFlow, but this direct comparison must be taken lightly because DialogFlow was not designed for novices. Our approach to investigating the effectiveness of AMBY follows best practices (such as extracting themes qualitatively using

4 https://adaptivecards.io/

field notes and observations [49], focus groups and contextual inquiry [82]) within the HCI community when an experimental study is not practical.

4.4 Conclusion

This paper has presented the iterative design and development of a conversational AI development interface, AMBY, that supports learners to create and tinker with their own conversational agents. Working in partnership with 26 youths, the interface was iteratively designed and developed through multiple user studies over 14 months. The interface was deployed to a two-week summer camp, allowing the study to engage learners in an informal setting with limited prior computing experience. Our work offers a new alternative to empower youth without an extensive technical background in building authentic AI applications. With continued research, this line of investigation holds the potential to open authentic AI learning experiences to learners of all backgrounds and ages.

CHAPTER 5

PHASE 3: AMBY REFINEMENT AND SUMMER 2023 USABILITY STUDY

In this chapter, I further improved AMBY with additional features to enhance learners' experiences. This study is guided by RQ3, which states: **What features do youth desire in a learning environment to support their educational needs?** My goal for this iteration is to explore learner's initial reaction and their usage on the updated AMBY features. Section 5.1 describes the new features integrated into AMBY 2.0, Section 5.2 presents the study procedure and findings of the user study in Summer 2023.

5.1 Additional Development of AMBY

In the previous studies, learners have provided important feedback on improving the usability of AMBY. A central goal for this phase was to address these suggestions to further enhance learners' experiences. An overview of the additional features for AMBY 2.0 is in Figure 5-1.

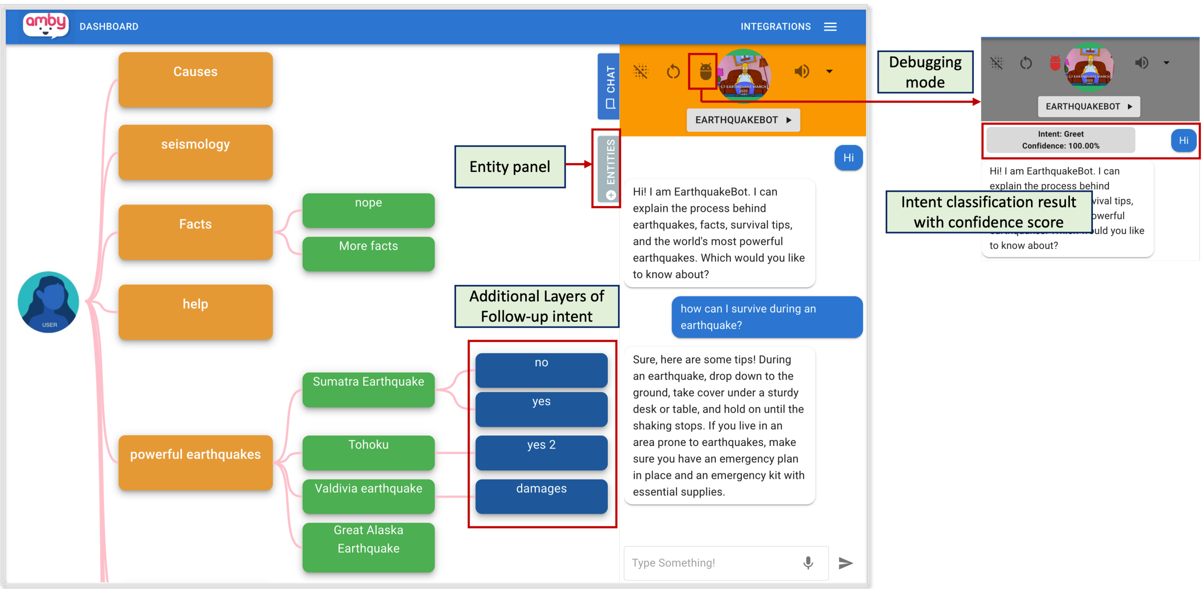


Figure 5-1. Overview of additional features in AMBY 2.0.

Additional Follow-up Intent Layers. Based on feedback from prior user studies (Chapter 4), learners suggested incorporating additional follow-up intent layers to enable more comprehensive conversations. In alignment with the design recommendation of a "high-ceiling" interface (Section 4.3.1), our updated AMBY version now supports up to three intent layers. This enhancement empowers learners to develop more detailed and nuanced conversations for their conversational agents.

Intent Debugging Feature. Drawing from past observations and the design recommendation emphasizing transparency in agent training and intent classification (Section 4.3.2), we have introduced an "intent debugging" feature. This AMBY version provides users with a button to inspect the intent classification outcomes of user utterances, along with the confidence level of such classifications. To activate the debugging mode, users simply click on the bug icon adjacent to the avatar in the chat simulation panel. When this mode is on, the chat panel's top banner changes to a grey shade, highlighting its developer-oriented nature. Hovering over individual user utterances will then display a popover, detailing the intent classification result and its associated confidence level.

Entity feature. A recurring challenge identified by participants in past studies was the tedious process of entering numerous training phrases. They frequently pointed out the monotony of inputting similar training phrases for different intents, which often resulted in user frustration. Some learners expressed interests in further expanding their project visions by incorporating more personalized responses based on user utterances.

In response to these feedbacks, and with an aim to expand the use cases of AMBY, we introduced the "entity" feature. An entity consists of words or phrases extractable from user input[1]. For instance, if a user requests, "Tell me the flight information from Orlando to Atlanta," the intent "flight information" is activated, while "Orlando" and "Atlanta" are identified as the "location" entity. The introduction of the entity feature yields two primary advantages:

1) It significantly streamlines the training process by eliminating the need for repetitive entry of similar phrases. In the absence of entities, developers would need to input every possible combination of locations for the "flight information" intent. Entities introduce a more efficient approach; rather than tediously substituting the noun or verb in each training phrase, developers can concentrate on developing diverse linguistic variations of the phrases. This not only saves time but also enhances the overall quality and diversity of the training dataset.

2) Developers can construct more personalized responses based on user utterances. In

1 In Dialogflow's original definitions, there are two distinctive definitions of 'entity types' and 'entity entry'. To teach middle school learners and consider the use cases of our system, we simplied the two terms to one "entity" term. https://cloud.google.com/dialogflow/es/docs/entities-overview

the past, responses can only be hard-coded by the developers thus might have been generic, such as, "*There are three flights for the cities you mentioned.*" With the new feature, a more specific reply such as, "*There are three flights from Orlando to Atlanta today*" becomes feasible. This enhances the end-user experience, as the agent can directly address and confirm user-specific details.

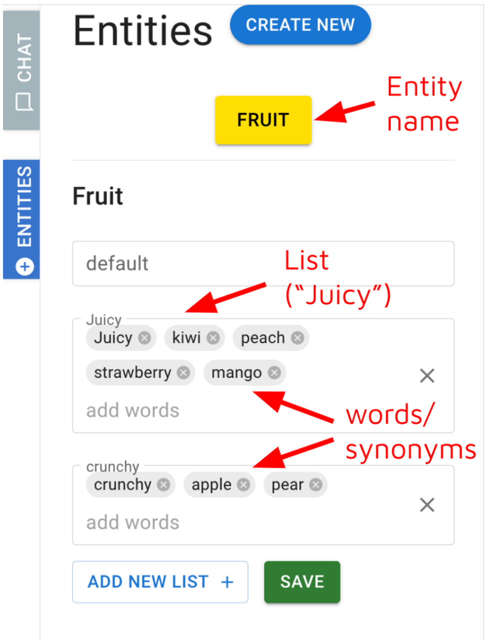


Figure 5-2. AMBY entity creation page.

Developers can navigate to the entity feature by selecting the "entities" button located on the right panel. This position is designed to replace the chat simulation panel during the entity setup phase (expecting few testing requirements during this process). The entity setup interface is shown in Figure 5-2. Here, developers can formulate a new entity and, if desired, append entity lists (entity sub-categories, there is a default list). Within each list, synonymous words or phrases can be grouped (e.g., "kiwi" and "mango" are categorized as "juicy" fruits). After the entity creation, they can be integrated into the intent. Figure 5-3 shows how to incorporate entities within training phrases and responses. In the training phrases window, inputting a $ symbol prompts a dropdown of available entities. Once an entity is selected, it is highlighted in yellow with an underscore, representing its distinctive nature. Any word or phrase within this entity is treated equivalently during training. When the intent is activated based on user input, relevant entity data is extracted and can be incorporated into a customized

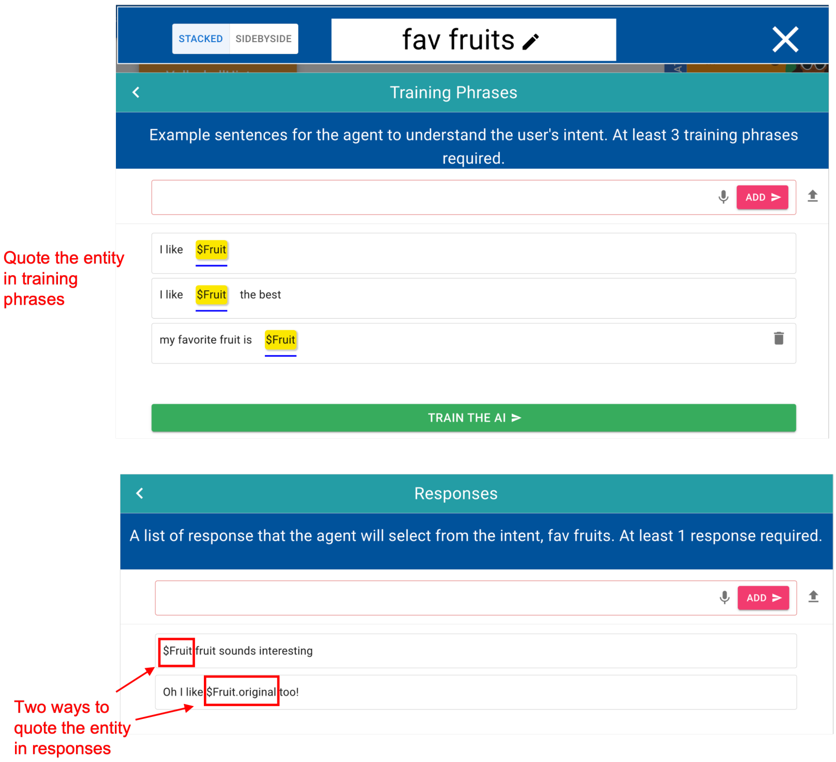


Figure 5-3. Interfaces to quote the entity within an intent. In this example, 'fav fruits' intent can recognize different kinds of fruits through the "Fruit" entity. It can produce personalized response based on the user's utterance. For example, if the user says: "I like *apple* the best. ", The agent would respond either as "*Crunchy fruit* sounds interesting" (because 'apple' was set to fall in the 'crunchy' list of the 'Fruit' entity) or "Oh I like *apple* too!"

response using the "$name" or "$name.original" syntax.

5.2 Summer 2023 Study

To teach the concept of entities at the summer camp, I developed a new lesson centered around the entities. This lesson was first piloted during professional development sessions with the camp facilitators before the camp started. Given that this topic is an advanced learning concept, it is structured to be introduced to learners only after they have acquired some developmental experience with AMBY. In Summer 2023, the "Entity" lesson was introduced to learners on the morning of the seventh day of camp. By this time, the learners had already gone through all the scheduled camp lessons, completed an individual project, and had made some progress towards a collaborative project.

In summer 2023, 19 participants attended the summer camp. Among these 19 participants, 7 identified as female and 12 as male; 8 were White/Caucasian, 6 were Black/African American, 4 were Asian and 2 were Hispanic/Latinx. The average age of the participants was 12.05 (SD = 0.4). 14 were rising 7th-graders and 5 were rising 8th-graders in the upcoming school year. During the camp, we conducted one-on-one interviews with 16 participants at the end of camp day 5, where they just finished their individual projects. We also administered two focus groups on the perception and usage of the entity feature on camp day 7. This was after introducing them to the concept of entity and providing several sessions throughout the day to work on a second project. Both interviews and focus groups were audio-recorded and transcribed for analysis. I used the same content analysis approach for analyzing the interview data as detailed in Chapter 4.1.3.

5.3 Preliminary Findings on Entity Feature Perception and Usage

The majority of participants were familiar with the concept of an entity. When prompted to provide a definition, three of the learners described it as akin to *variables* in programming that can hold various values or words.

"*(It's) like a variable on Scratch. It can hold different meanings for it but using just one thing.*" (camper101)

Most of the participants acknowledged the utility of the entity, noting that it "*allows you to use way less training phrases and make the response more personalized for people.*" (camper 109). They believed it was beneficial in improving the efficiancy of their development work.

Regarding the usage of the feature, among the 17 participants, four (camper 101, 109, 120, 122) stated that they incorporated the entity feature into their group projects. Notably, both teams developed interactive games: one modelled after the popular American quiz show "Jeopardy", and the other named "PokeGameBot", which simulated a Pokémon battle. Both games employed consistent sentence structures across diverse topics (e.g., the answers to the Jeopardy game consistently began with a "What is" phrase). To create more streamlined and diverse inputs, these groups incorporated keywords and phrases into an entity, ensuring they "*don't have to type the same words over and over again.*" (camper123)

"*Instead of typing like, 'Is the answer ___? And/or is it that?' You just put a bunch of stuff for the entity.*" - camper122

Interestingly, both teams found the entity feature to be beneficial primarily for training phrases, rather than for personalized responses. This indicates that learners recognize the value of entities in off-loading the repetitive aspects of training tasks. By reducing manual input, entities allow learners to shift their focus towards more creative aspects of AI development. Additionally, the preference for using entities in training phrases over personalized responses also suggests the critical role of effective training in the success of an AI application.

During the focus group, four participants admitted to finding the entity feature somewhat "*confusing*". This perception might stem from the feature's inherent complexity and its introduction relatively late in the camp (on Day 7). By that time, many learners had already established their developmental approach with AMBY, and incorporating a newly introduced feature—especially one not initially factored into their project's design—became a challenge. Their initial mental model for development had been solidified.

"*It was a bit confusing and I decided that it would take too long to figure it out. It would be simpler to just go on what I had already been going.*" (camper119)

There was a clear desire among learners for more fine-grained control over conversational design. For example, camper109 suggested the possibility of integrating a customizable default fallback for follow-up intents, thus allowing their chatbots to generate more contextually appropriate fallback responses.

In summary, our preliminary findings from Summer 2023 showed the promise of the entity feature in enhancing learner experiences. Majority of the participants demonstrated an understanding of the entity concept and perceived it as useful tool. Many saw its potential in streamlining and personalizing chatbot training phrases, and a subset even leveraged it heavily in their projects. Despite the entity feature being recognized as an advanced learning topic, our results suggest that its introduction should occur earlier in the curriculum, before they start building their own projects. Doing so would better support learners in establishing effective development routines with AMBY and maximize its utility.

5.4 Conclusion

In this chapter, I have described the refinement of AMBY with three additional features: more follow-up intent layers, intent debugging feature and the entity feature. Through our summer camp study in 2023, I gained pivotal insights into learners' reception and engagement with the enhanced features, especially the entity feature. These findings emphasize the need for a more rigorous experimental study to assess the feature's effectiveness on a larger scale. Moving forward in Chapter 6, I will focus on deploying AMBY 2.0 in the middle school classroom context and investigating the influence of the entity feature on students' interests and experiences.

CHAPTER 6
PROPOSED WORK: PHASE 4: AMBY CLASSROOM STUDY

The previous chapters have described the iterative design and development of the learning environment AMBY and findings from the user studies. In my proposed work, I will conduct a new classroom study using the AMBY 2.0 interface. My primary goal is to examine the impact of the entity feature in AMBY on learners' interests and experiences. The research question guided this study is: **RQ4: In what ways does the entity feature impact students' AI learning experience?** Additionally, I will also investigate the students' experiences of learning conversational AI through chatbot development using AMBY in middle school science classrooms.

In this chapter, in Section 6.1, I will first introduce the science-based conversational AI curriculum that I developed for the classroom study. Then, in Section 6.2, I will describe a detailed research design for my proposed classroom study for AMBY 2.0, including paticipants, study procedure, data collection and analysis. In Section 6.3, I will present the plan for post-hoc analysis regarding the outcomes of the classroom intervention. Finally, in Section 6.4, I will describe the timeline for the proposed work.

6.1 Science-Based Conversational AI Curriculum

Building on the foundation of the earlier phases of AMBY and summer camp curriculum, my next step is to extend its application in a formal classroom context. To transition the learning activities from informal summer camp to classroom settings, we will make adjustments to the conversational AI curriculum.

First, I have condensed the curriculum to fit into an approximately 10-hour learning module, which aligns with the intended duration of the classroom study (Section 6.1.1). Second, I have worked with three middle school science teachers to embed science content in the conversational AI curriculum (Section 6.1.2). The new science-based curriculum contains examples related to the science topics that the students learned previously, and the activities are closely tied to their previous science learning experiences.

6.1.1 Classroom AI Learning Modules

The curriculum is focused on teaching students the fundamentals of Artificial Intelligence and Conversational AI in a science context. The course is divided into several modules, with each module covering specific topics related to AI and conversational agents. The learning objectives are shown in Table 6-1.

In the first module (L1 - Intro to AI), students are introduced to the basics of AI and its characteristics. They learn about the different types of AI and are given examples of AI applications such as Siri and self-driving cars. The learning objectives for this module include being able to define artificial and intelligence, identify characteristics of AI, and give examples of AI.

In the second module (L2 - Intro to chatbot), students are introduced to chatbots and conversational AI applications. They learn about the different tasks that chatbots can perform, such as answering questions and making recommendations. The learning objectives for this module include being able to identify chatbot/conversational AI applications and list examples of what chatbots can do. In this learning module, students will also play with the sample AIs in AMBY.

The third module (L3 - Intro to intent) covers the basics of intents and how they are used in conversational AI. Students learn about the different components of intents, how training phrases work, and how responses work. They also learn about the greet intent and the default fallback intent and how to create customized responses for these intents. The learning objectives for this module include being able to differentiate the role of developer/user/agent, define intent, and create customized responses for the greet intent and the default fallback intent.

The fourth module (L4 - Follow-up intents and conversational design principles) covers follow-up intents and good conversational design practices. Students learn about the importance of setting user expectations through chatbot greet responses, designing conversational flow, using conversational markers, and guiding users when there is no match intent through customized default fallback response. The learning objectives for this module include being able to define follow-up intents, identify good conversational design practices, and create many training phrases.

The fifth module (L5 - Intro to entities) introduces the concept of entities in conversational AI. Students learn to define entities and identify them in user utterances.

Lesson	Learning objectives
L1 - Intro to AI	Define artificial and intelligence Identify characteristics of AI (artificial and intelligent) Give examples of AI (such as siri, self-driving cars) and understand what makes the example AI
L2 - Intro to chatbot	Identify chatbot/conversational AI applications (e.g., Siri, google home, Alexa) List examples of what chatbots can do (e.g., answer questions, make recommendations, perform some task)
L3 - Intro to intents	Differentiate the role of developer/user/agent Define intents Identify the components of intents Explain how training phrases work Explain how responses work Identify when the greet intent will be used Identify when the default fallback intent will be used Create customized responses for the greet intent Create customized responses for the default fallback intent
L4 - Follow up intents & Conversational design principles	Define follow-up intents Identify good conversational design practices, including: 1. Setting user's expectations through chatbot greet responses 2. Designing conversational flow through multiple logical follow-up intents in which each follow-up intent is related to its parent intent logically 3. Using conversational markers (such as "Okay", "Thank you") to make naturalistic conversation 4. Guiding the user when there is no match intent through customized default fallback response 5. Having a "help" intent to handle user confusion such as "help", "what can you do" 6. Adding many training phrases
L5 - Intro to Entities	Define entities in the context of conversational AI Identify potential entities from user utterances
L6 - Project development	Demonstrate conversational agent ideas Use AMBY to create a conversational AI project Apply conversational design principles to make naturalistic conversations Test and revise their projects Evaluate others' conversational agents Reflect on what worked well, what did not work well with their conversational agent based on peer testing

Table 6-1. Learning objectives of science-based conversational AI curriculum

Students engage in hands-on practices regarding entities in AMBY to understand the benifits of using entities in their conversation. The key objectives include understanding entities in AI conversations and recognizing their use in enhancing chatbot interactions.

The sixth module covers project development (L6 - Project development), where students demonstrate their understanding of conversational agent ideas, and use AMBY to create a conversational AI project related to the science topics they learned previously. Students also learn how to apply conversational AI design principles to make conversations and test and revise their projects. The learning objectives for this module include being able to demonstrate conversational agent ideas related to science, use AMBY to create a conversational AI project, apply conversational AI design principles, and test and revise their projects. Additionally, after students finish their project, they will engage in peer testing in which students will form small groups and evaluate other's conversational agents. After peer testing and feedback, students will reflect on what worked well, what did not work well with their project and refine their agent based on the peer feedback.

In summary, the curriculum is designed to provide students with a comprehensive understanding of artificial intelligence and conversational AI, with a focus on designing and building conversational agents. Through these lessons, students will learn how to create naturalistic conversations, understand good conversational design practices, and apply their knowledge to create functional conversational agents.

6.1.2 Exemplar Science Chatbots within the Curriculum

For this science learning context, we provide two science agents, one is *EarthquakeBot* and one is *Wavie* as sample AI agents for students to test on. *EarthquakeBot* is in alignmnet with the *plate tectonic* topic in science curriculum. The agent explains the process behind earthquakes, facts, survival tips, and the world's most powerful earthquakes. *Wavie* is relevant to the *waves and electromagnetic radiation* unit. This agent is knowledgeable about different types of waves, the parts of a wave, and the definition of a wave. These two sample agents serves as examplar projects for students to model for their personal projects. These sample agents follows the best conversational design principles introduced in the later part of the curriculum. We also use examples extracted from these sample agents to reinforce the learning objectives in the conversational AI curriculum.

6.2 Study Design and Data Analysis

In this study, I aim to investigate the effectiveness of the entity feature within AMBY.

Experimental Conditions and Hypothesis: To investigate whether the entity feature impacts students' learning experiences, I will conduct a between-subject experiment with two versions of AMBY: *AMBY with entity* and *AMBY without entity*. The student participants will be randomly assigned to either condition to use AMBY to create their conversational apps. I hypothesize that the entity feature will positively affect students' learning experience. Consequently, I hypothesize that students who have access to the entity feature in AMBY will show a **greater interest in conversational AI** (measured in post-survey). Furthermore, the chatbots they produce are expected to demonstrate **higher project quality**, assessed across four dimensions: project ideation, conversation design, AI development, and end-user satisfaction.

Analysis Plan: To answer RQ4, I will investigate the differences in **interest formation** and **project quality** between the two experimental groups using an independent samples t-test. The interest formation construct contains triggered situational interest and maintained situational interest in conversational AI. Based on Hidi and Renninger [40]'s interest development model, the development of interest contains four phases: 1) triggered situational interest; 2) maintained situational interest; 3) emerging individual interest; and 4) well-developed individual interest. Our classroom intervention specifically addresses phases 1 and 2 and hopes to trigger and then maintain students' interest in conversational AI as they find developing conversational agents to be meaningful. There are six statements related to triggered situational interest, specifically related to instant enjoyment, exploration intention and attention demand; there are two statements relevant to maintained situational interest and three statements focusing on sharing. More details about the statements can be found in Appendix D.

I will also compare the quality of students' project between the two groups. The student projects will be evaluated across four dimensions: 1) project ideation; 2) conversational design; 3) AI development; 4) end-user satisfaction. For the first three dimensions, under each dimension, there will be specific project aspects that will be rated on a score of 1 to 4 (1 - little to no evidence of approaching expectations; 2 - approaching expectations; 3 - meets expectations, 4 - exceeds expectations). Our goal is to provide enough support so that learners' projects can meet our pedagogical expectations and achieve a rating of 3 or higher. Table 6-2 details the criteria for a "meets expectations" rating (score = 3) in relation to specific project aspects. For criteria of other scores, see the rubric in Appendix C. The fourth dimension, *end-user satisfaction*, will be assessed through an interaction with the chatbot, mimicking an end-user's experience. This will be rated based on a 5-point Likert scale encompassing the following statements adapted from Walker et al. [104]: 1) The agent was easy to understand; 2) The agent understood what I said in this conversation; 3) In this conversation, it was easy to find the information I wanted; 4) I knew what I could say at each point of the dialogue; 5) The agent worked the way I expected ; 6) I would like to talk to the agent again.

6.2.1 Participants & Recruitment

Teachers will be recruited based on our existing partnerships in Alachua County Public schools. Once the teacher decides the learning module to be beneficial and meets their learning objectives and standards, the students from the teacher's science class will be recruited through in-person announcements and email communications.

The teacher has five sessions of classes, with approximately 20 students in each, totaling 100 potential participants. Because all students will participate in the AI learning activities regardless of their consent/assent status, I anticipate a lower consent rate for research compared to previous classroom studies conducted by our lab. Based on my past experiences of conducting classroom studies, I project that 60% of the students will give consent for research participation. Factoring in potential absences, I estimate a final dataset with complete data from about n=50 students.

6.2.2 Study procedure

The study will be conducted for ten total study days over three weeks during students' regular science class time. Same as the regular class period, the classroom activity will last 50-60 minutes per day. To account for student arrival and departure, I design the learning content of each day to be about 40 minutes. Table 6-3 shows the daily study schedule and data

Dimensions	Project Aspects	Statement for Score of 3 (Meets Expectations)
Project Ideation	Demonstrating purpose	The student has a clear idea of what the bot will do and implements their idea clearly.
	Chatbot Personality design	The agent demonstrates a unique personality through at least two of linguistic and visual choices (avatar, voice, word choice) and demonstrates intentional thought to align with chatbot purpose.
Conversational Design	Overall Intents	Project intents align with its purpose. The project has a balanced overall structure of the intents, has reasonable variation. Some adjustments could be made for streamlined design.
	Main intents	The majority main intents (more than 60%) are mutually exclusive and sufficient in demonstrating the purpose.
	Follow up intents	The agent has multiple logical follow-up intents AND Each follow-up intent is related to its parent intent mostly logically. Most follow-up intents can be triggered properly based on the responses from their parent intents.
	Greet intent	The agent has at least one customized greet response demonstrating its purpose. May not set exact user expectations.
	Default fallback	The response is created by the learner and can redirect the users.
AI Development	Training phrases	Most training phrases are ample, cohesive, and varied within the intent.
	Responses	Most customized intents contain at least one response that is in proper length, logical, and mostly mimic natural conversation.

Table 6-2. Student chatbot project evaluation criteria for the first three dimensions

collection plan for each day.

Each class session will be randomly assigned to either *AMBY with entity* or *AMBY without entity* condition. For the *AMBY with entity* condition, all students from the class period will be using AMBY with the entity feature available during their learning. These students will be introduced to the concept of entity and have a hands-on practice session on Day 4, before their project development. To control for the interaction time on AMBY for the two conditions, the control group (*AMBY without entity*) will engage in a similar hands-on activity on day 4, where they will be guided to create an AI agent from scratch. To avoid disadvantaging students in the control group from learning the concepts of entity, I will introduce entity to this group after they finish the project so both groups will gain the same learning opportunity from this experience.

40-minute period	Daily Tasks	Data collection
Day 1	Assent, Pre-survey, Introduction to AI & Chatbots, Log in AMBY, Play with sample agents	Pre-survey
Day 2	AMBY lesson: Intents and Special intents. Hands-on practice: Modify 'AboutMeBot' on AMBY	No collection
Day 3	AMBY lesson: Follow up intents, Conversational Design Principles, Hands-on practice	No collection
Day 4	Kahoot, (1) entity lesson, hands-on practice on entity or (2) hands-on practice on creating an agent	Kahoot
Day 5	Introduction to Pair Programming, Pair Programming: Brainstorm Project Ideas, Create a New Agent	Video/audio/screen, project design log worksheet
Day 6 & 7	Pair programming: Project Development	Video/audio/screen, AMBY interaction logs, final project
Day 8	Peer Testing, Project refinement, post-survey	Post-survey
Day 9	Post-assessment, entity lesson (control group)	Post-assessment
Day 10	Interview about AMBY feature and AMBY project	Interview

Tasks for the *AMBY with entity* condition and *AMBY without entity* condition

Table 6-3. Classroom study schedule (each day is approximately 40 minutes of content)

In day 1, 2 and 3, the students will learn the relevant concepts of AI and conversational AI. Day 4 will involve a formative assessment using Kahoot (a interactive quiz game commonly used in K-12 classrooms) and introduction of the entity feature. Day 5, 6 and 7 will feature project development where students work in pairs to develop a conversational agent relevant to science topics they learned in class. On day 8, students will engage in a round-robin peer testing, where they test each other's projects, offer feedback and refine their project based on peer's feedback. They will also complete the post survey to reflect the classroom activities. On Day 9, students will engage in a paper-and-pencil test to assess their knowledge gain in AI. On Day 10, I will select some participants (based on their class engagement during previous days) to interview about the AMBY feature and their project to gather their in-depth feedback.

6.2.3 Data collection instruments

To answer the research question, I will collect the following data:

- **Pre-/Post-Survey.** Both pre- and post-survey (Appendix D) includes the AI attitude items adopted from The Barriers and Supports to Implementing Computer Science

(BASICS) questionnaire [72]. The constructs include ability beliefs, persistence, identity, and interest. In pre-survey, there will be question asking students' language background, prior experience in computing and demographic information. In the post-survey, there will be questions prompting students to reflect their experience of the classroom activities.

- **Observational Field notes.** The study team will observe the classroom environment and student conversations and feedbacks about the classroom activities
- **AMBY interaction logs.** The AMBY platform will log participants' interactions with the system, such as adding a training phrase, creating an intent, deleting a response, messaging in the testing panel.
- **Artifacts, worksheets, reflection notes**: As part of the AMBY activity, participants will complete a Project Design Log (Appendix A) as they create their chatbot project. During the learning activities, students may complete worksheets or be prompted to write reflection notes about the learning activities.
- **Post-assessment**: Students will complete a paper-pencil-based post-assessment (Appendix E) to evaluate their learning based on the learning objectives (Table 6-1).

6.3 Post-hoc analysis

In addition to investigating the impact of the entity feature, I will also conduct a post-hoc analysis regarding the outcomes of the classroom intervention. I plan to examine how the integration of conversational AI into middle school curricula fosters students' learning about AI, as well as their attitudes and identities towards AI. This will be done through comparing pre- and post-surveys and the results from the post-assessment. The analyses will aggregate data from all participants, regardless of their experimental conditions. Given that all participants engaged in learning the core conversational AI learning modules, I do not expect significant differences between the two groups in terms of attitudes and identity change relating to AI. However, for the completeness of my analysis, I will still run a comparison of the outcomes between the two conditions.

In the pre- and post-surveys, the AI attitudes items are measured in 4-point likert scale. To investigate whether student's attitudes improves after the classroom intervention, I will compare student pre and post attitude scores using a paired samples t-test. Regarding the interpretation of outcomes, I anticipate a lower attitudes (persistence, ability beliefs, identity) in AI in the pre-survey because they have only heard about AI but have not learned about AI yet. If there is a significant increase in attitude toward AI in the post-survey, that could be attributed to our fun and engaging classroom intervention.

I will examine students' experiences of learning conversational AI in science classrooms by analyzing qualitatively from student post-survey responses collected from the following questions: 1) What did you learn from the conversational AI lessons and activities? 2) Did the conversational AI lessons and activities help you understand science concepts you learn from class? If so, how? 3) What did you like about the conversational AI lessons and activities? 4) How could the conversational AI lessons and activities be improved? I will use content analysis approach [44] to extract codes and themes from the written responses.

To assess students' AI learning from this classroom intervention, I will report the scores calculated from the post-assessment. This assessment comprises 15 questions: 14 are multiple-choice, and one is open-ended. Each question correlates with a specific learning objective, as outlined in Table 6-1.

A pre-assessment will not be conducted. Because there is no comparable AI curriculum at the participants' school, and our learning objectives are closely aligned with the features on AMBY, it is reasonable to presume that the participants will not have prior knowledge in this area. Administering a pre-test could potentially harm the learning experience by causing frustration and diminishing the students' interest in AI, as they have not been exposed to such material before the intervention.

While the lack of a pre-assessment means I cannot measure learning against a "baseline," I will nonetheless analyze the rates of correct answers across different questions. This will help inform the curriculum's effectiveness in achieving the specified learning objectives.

To validate the effectiveness of the entity learning module, I will calculate students' learning for both *overall AI learning* (excluding the two entity questions) and specific *entity learning* (through the two entity questions). Because the students in the *AMBY with entity* condition will be taught the concept about entity and may apply this knowledge for their projects, I expect they will exhibit a higher understanding of the entity concept than their peers in the control condition. However, since students in both conditions will complete the same conversational AI development task, I do not expect significant difference in students' overall AI learning between the two groups.

6.4 Timeline

A timeline of my entire Ph.D. can be seen in Figure 6-1. A more detailed overview of my proposed work, spanning from October 2023 to July 2024, can be seen in Table 6-4.

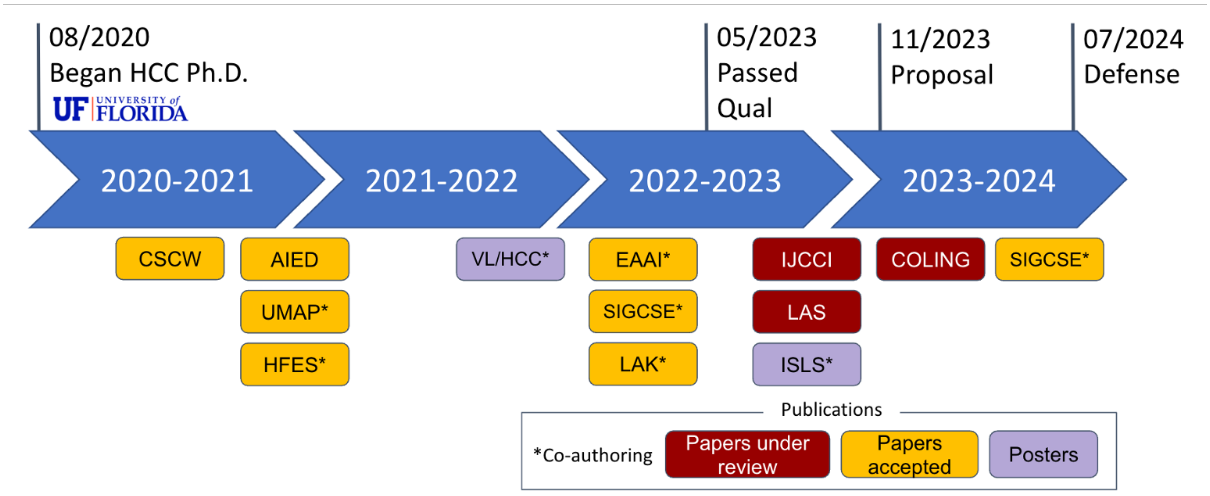


Figure 6-1. A timeline of my Ph.D. and publications.

Table 6-4. Timeline of proposed work

Time Frame	**Task**
Oct 2023	Submitted classroom study IRB for expedited review
Nov 2023	Proposal
Dec 2023	Obtain IRB approval, develop two versions of the system (w/o entity)
Jan & Feb 2024	Conduct the classroom study, collect data
Mar 2024	Analyze data for RQ4
Apr 2024	Post-hoc analysis
May 2024	Write dissertation
June 2024	Defend dissertation

CHAPTER 7
EXPECTED CONTRIBUTIONS

This dissertation will make the following contributions to the field of artificial intelligence education in K-12 settings.

1. A novel development environment for young and novice learners to create AI-powered conversational agents. It uses an innovative card-based tree design, unique among conversational AI development tools, to represent conversational flow. The results from our summer camp deployment suggest that youth learners can successfully create personally relevant conversational agents using AMBY.

2. Design recommendations for building learning technologies that support AI education for youth. The findings shed light on future directions for the design and research of youth-centered AI-authoring tools.

3. A new approach to enhance middle school science education by integrating AI into science classrooms. I will identify best practices for designing and implementing this conversational AI curriculum in science classrooms. The work explores the hypothesis that by integrating conversational AI into science classrooms, students can engage in hands-on learning activities that allow them to apply their knowledge and skills about science in a real-world setting. This has the potential to improve student engagement, motivation, and learning outcomes.

4. Empirical findings on the impact of learning environment design on student outcomes. Through a controlled experiment on usage of different versions of AMBY, the results will draw insights on the effectiveness of conversational AI learning environments in facilitating interests in AI and learning experiences for middle school learners.

APPENDIX A
DESIGN LOG DOCUMENT TEMPLATE

Next page is the design log that learners used to document the process of creating conversational agents.

PROJECT DESIGN LOG

Enter the Name of your App Here!

Student Name

June 2022
Camp Dialogs

List of Sample Project Scopes	Goals
- Reminders/ Task oriented apps (e.g. remind me to do my homework) - Sports (e.g. sports in the olympics, what the record is, what my favorite player on the team, stats on a player) - Educational (e.g. accessing learning resources) - Gaming(e.g. tips, achievement tracker - Recipes(breakfast, lunch, dinner) - Motivational - Entertainment (e.g. movie recs, movie review) - Shopping Assistant - Mental Health	- Helping themselves - Helping family member

Empathy

Step 1: Understanding the people that will be using your app

Who will you be trying to help?	
What do you think they will need?	

Define the Problem

Step 2: Defining the challenges from your point-of-view

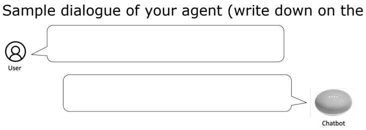
What is the problem you are trying to solve?	
How do you think you can solve this problem?	
Who will use this app?	
What will this app do?	

Ideate

Step 3: Brainstorming ideas and create solutions for your users

What are your ideas to solve this problem?	
Which idea is the most original?	
Which idea do you think is the easiest or most complex?	
What makes it easier or harder?	

Sample dialogue of your agent (write down on the paper sheet)



Prototype

Step 4: Drafting elements of the agent

What is your agent's name?	
What type of intents will your agent have (*Welcome* and *fallback* intents are provided for you)	

Test

Step 5: Testing your prototype and getting feedback

Follow the instructions provided by the camp counselors to set up your Google Home device for testing:

- Test the google assistant/ google home device and get feedback

Answer the following questions:

What worked well with your app?	
What didn't work well with your app?	
What can you change to make it better?	
What other ways might your user use your app?	

Modify

Step 6: Improving your prototype

What can you do to fix what didn't work well with your app?	
How can you make your app better?	

Share

Step 7: Showing your app

How will you show others your app?	

- Design a Logo or Icon that will represent your project:

APPENDIX B
LIST OF CONVERSATIONAL AGENTS THAT LEARNERS CREATED USING AMBY IN SUMMER 2022

Table B-1. Conversational agents that learners created using AMBY in Summer 2022. Descriptions given were written by the learners as they completed a design document for the project. The themes were summarized by myself. Note: Some learners named their agents after themselves; to protect their privacy, these are given as [redacted].

Theme	**Chatbot Name**	**Description**
Mental health	goldiethestressbot	Have a bot to talk about your feeling.
	MentalHealthBot	Provide support and answers to people that struggle with mental health problems.
	CalmBot	Give meditation advice, recommend calming things.
	ReachOutAnd-GrabaHand	A therapy bot that helps about marriage issues and emotions.
	diamond	Help people feel better in life and don't go through stress.
Game	Gaminglogicbot	Give introduction to three games, Fortnite, Roblox, and 2K.
	teacherbot	Teach my parents about the game Madden.
	BlahBot	Inform people about the video games.
	Rox-bot	Tell info about Roblox (and possibly Roblox games).
	Gamebot	Help friends with video game levels.
	Gamebot	Give tips to get past a certain part of the game (Animal Crossing).
	HorizonBot	Help with Animal Crossing New Horizons.
	FortniteBot	Help people get used to the game Fortnite.
Music/movie	boy_bot	Play music.
	JokerTheMusician	Tell people jokes and let them listen to music depending on their mood.
	Musicbot	Give people music recommendations.
	ezmae	Recommend music to people based on the genre they like.
	MusicBot	Recommendations of what music to play.
	Angelbot	Recommend interesting music and tell you facts about singers.
	Esmie	Provide good movies to watch.
	horrorbot	Recommend horror movies.
Personal/joke	hal-9000	Simply be funny. Tell jokes, play simple games, recommend music, and be funny.
	tmx-10000	Be funny. Tell joke, music, food, and stuff about pets.
	jackthejoker	Crack jokes back and forth.
	[redacted]	Tell you jokes and little things about me.

Continued on the next page

Table B-1 Conversational agents that learners created using AMBY (continued)

Theme	Chatbot Name	Description
	PinkusBot	Give song recommendations, giving advice on being happy and making jokes.
Task-oriented	MotiveBot	Help keep my grandad entertained and motivated.
	Home_Gurl	Help the teachers to know about dance.
	AngelGirl	Help with cooking and cleaning.
	Petbot	Reptile store, help with information buying reptiles, arachnids, and prey.
	AssistaBOT	Help me with homework or waking up early.
Recommendations	ShopBot	Suggest the best brands for clothes, shoes, accessories, etc.
	FashionBOT<3	Help people who need fashion advice.
	Gabby	Recommend gifts you can give to certain people of all ages and basic personalities.
	Artist_Helper	Help bored artists figure out what to draw and which styles to draw them in.
	ideabot	Creates ideas for people who are bored.
	FoodBot	Recommend restaurants and grocery stores.
	RecBott	Recommend books of different genres with marginalized people as main characters and properly portrayed.
Sports	Boxingcoach	Help people learn how to box and the rules.
	[redacted]	Tell people the key elements on playing dodgeball.
	BASKETBALL-BOT	Tell about my favorite basketball team, players, and skills.
	Sport	Tell the user about baseball.
	BasketballBob	Give information about basketball.
	BasketBallBen	A cool and fun way to learn about tips, history, and facts about basketball.
	Theyenvy_Nya	Give tips and information about volleyball.
	baldy	Teach about football.
	FootballBot	Tell you facts about football.
	Diamond	Teach about basketball tips and provide information about Steph Curry.
	KingBot	Teach people about LeBron James.
Educational	jerryberry	Teach people with black history, provide information about black influencers.
	OlympicBot	Inform the user about the Summer Olympics.
	ZooBot	Fun and interesting facts about animals.
	VRbot	Explain how VR works, what you can play, and some tips.
	Twinnem	Tell you interesting facts about Fraternal twins.
	Mathbot	Help people understand how to solve math problems.

Continued on the next page

Table B-1 Conversational agents that learners created using AMBY (continued)

Theme	Chatbot Name	Description
	botbot	Quiz on specific math topics.
	Cookie_Cutter	Tell recipes about cookies.
	DanceBot	Tell people about dance tips.

APPENDIX C
AMBY STUDENT PROJECT EVALUATION RUBRIC

Instruction on how to evaluate different dimensions:

1. **Project ideation:** the expert evaluators go through the chatbot content in AMBY (development panel, left side), in combination of looking at the design document from the individual/group, to understand the purpose of the chatbot and target users.

2. **Conversational design:** the expert evaluators go through the chatbot content (intents, training phrases, responses) in AMBY (development panel, left side).

3. **AI development:** same as "conversational design".

4. **End-User Satisfaction:** the expert evaluators test the chatbot as an end user without necessarily knowing/understanding the inner workings of the chatbots.

Table C-1 shows the statement for the end-user satisfaction dimention. Each item, adapted from Walker et al. [104] is rated based on 5 point likert-scale: 1 - Strongly Disagree, 2 - Disagree, 3 - Neither agree nor disagree, 4 - Agree, 5 - Strongly agree.

User Satisfaction Statement	**Aspects of User Perception**
The agent was easy to understand.	NLG Performance
The agent understood what I said in this conversation.	NLU Performance
In this conversation, it was easy to find the information I wanted.	Task Ease
I knew what I could say at each point of the dialogue.	User Expertise
The agent worked the way I expected.	Expected Behavior
I would like to talk to the agent again.	Future Use

Table C-1. End-User Satisfaction Dimension Statement

Table C-2. AMBY Student Project Evaluation Rubric for Project Ideation, Conversational Design, and AI Development Dimensions.

Dimensions	Project aspects	1. Little to no evidence of approaching expectations	2. Approaching Expectations	3. Meets Expectations	4. Exceeds Expectations
Project Ideation	Demonstrating purpose	The purpose of the design is vague / unclear OR The implementation has no clear purpose (the system is random)	The purpose is broad, not fully clear. OR The purpose does not meet the needs of their target audience OR the system implementation doesn't fit the purpose written.	The student has a clear idea of what the bot will do and implements their idea clearly.	The purpose is well-thought out, demonstrating the social connectivity by stating the chatbot is to help a specific group or community.
	Chatbot Personality design	There is no intentional linguistic or visual choices to align with the chatbot's purpose	The agent demonstrates at least one visual and linguistic choices including the following components (avatar, voice, word choice) but not all OR does not fully align with chatbot's purpose	The agent demonstrates a unique personality through at least two of linguistic and visual choices (avatar, voice, word choice) and demonstrates intentional thought to align with chatbot purpose	The agent demonstrates a unique personality through all of linguistic and visual choices (avatar, voice, word choice) And using the unique language consistently throughout and all visual and linguistic choices

Continued on the next page

Table C-2 – continued from previous page

Dimensions	Project aspects	1. Little to no evidence of approaching expectations	2. Approaching Expectations	3. Meets Expectations	4. Exceeds Expectations
Conversational design	Overall Intents (mainly look at the conversation tree, not the testing result)	No progression of the conversation. OR Does not demonstrate the logical conversation patterns (the followup is completely disconnected from the main intent)	The dialogue tree has "gaps" (e.g., main intent only has one followup, or no followups)	Project intents align with its purpose. The project has a balanced overall structure of the intents, has reasonable variation (reasonable means an appropriate ratio and distribution for the main and follow-up intents)	The follow-up intents are well-developed. The overall flow of the dialogue tree is logical and creative
	Main intents	No customized main intents provided OR The customized intents are unrelated to the project purpose	The main intents are not mutually exclusive (some intents could be collapsed) and/or The intents are not comprehensive (not aligned with the project purpose, or lacking important information)	The majority intents (more than 60%) are mutually exclusive and comprehensive in demonstrating the purpose, some adjustments could be made for streamlined design	All intents are mutually exclusive and comprehensive of purpose, no design changes are needed

Continued on the next page

Table C-2 – continued from previous page

Dimensions	Project aspects	1. Little to no evidence of approaching expectations	2. Approaching Expectations	3. Meets Expectations	4. Exceeds Expectations
	Follow up intents	No follow-up intent is provided	The agent has at least one follow-up intent OR most of the follow-up intents do not logically match with its parent intent (or the response cannot trigger the follow-up intent properly) OR most follow-ups are unnecessary or repeated	The agent has multiple logical follow-up intents. AND Each follow-up intent is related to its parent intent mostly logically	All of the follow-up intents are not only logically related to main intent and numerous, they are mutually exclusive
	Greet intent	No customized greet response is provided	The agent has at least one customized greet intent, however the purpose is not clear or actionable (e.g., “Hi, I’m Santa bot.”, “ask me anything you need!”)	The agent has at least one customized greet intent demonstrating its purpose (e.g., “You can ask me about XYZ”) May not set exact user expectations: (“Ask me for song recommendations”, “ask me about NBA tips”, “hey im blah bot do you need any assistance on video games?”)	The response(s) in the greet intent effectively greet the user (e.g., “hello”), introduce the chatbot (e.g., “I am MusicBot”), and demonstrate the purpose (“I can introduce XYZ”). AND Set exact user expectations (e.g., “I can talk about pop rock or hip hop music”) or clearly directs the user for next steps (e.g., “simply state ‘quiz me on math’”)

Continued on the next page

Table C-2 – continued from previous page

Dimensions	**Project aspects**	**1. Little to no evidence of approaching expectations**	**2. Approaching Expectations**	**3. Meets Expectations**	**4. Exceeds Expectations**
	Default fallback	No camper-created fallback response is given	The response is created by the camper, however it cannot not redirect the users (e.g., “I didn’t get that. Try it again.”)	The response is created by the camper and can redirect the users (e.g., “I didn’t get that as I’m still learning. I’m more confident to talk about XYZ instead.”)	The agent has multiple varied, customized and meaningful responses that can redirect the users
	Help intent	No help intent	The project has a “help” or equivalent intent but the training phrases is limited (less than 3) OR the response does not demonstrate the purpose clearly	The help intent can recognize common user expressions such as “I need help”, “what can you do?” AND The project has a “help” or equivalent intent to help the user navigate the chatbot, within the intent, the response introduces the chatbot functions clearly.	The training phrases for the help intent are varied and numerous AND/OR The response allow the users to take actions (e.g., “I can do XYZ, what would you like to start with?”) AND/OR Has multiple, varied, meaningful responses

Continued on the next page

Table C-2 – continued from previous page

Dimensions	**Project aspects**	**1. Little to no evidence of approaching expectations**	**2. Approaching Expectations**	**3. Meets Expectations**	**4. Exceeds Expectations**
AI Development	Training phrases	The amount of training phrases is limited (less than system requirement) OR Most of training phrases are random in the customized intents	The amount of training phrases meet the system requirement, but the content does not show enough linguistic variations (syntactically and lexically) within the intent or topic variations across different intents	Most training phrases are ample, cohesive and varied within the intent; also differ from those in other intents. They present variations in either syntactic structure or lexicon choices	The project contains consistently more varied training phrases than what the system requires, which can capture some edge cases. Training phrases are given and they are unique in both lexical and syntactic structure
	Responses	The responses are random in most of the customized intents	Most Responses (60%+) are provided either too long or too short, or lack of information or contains grammatical errors that impede user's understanding If there are multiple responses, the content is not consistent enough to trigger similar user reactions Example: "Bad Romance by Lady Gaga" - not conversational	Most customized intents contain at least one response that is in proper length, logical, mostly free of grammatical errors, mostly mimic/display natural and conversational, may include some conversational markers.	Intents contain multiple logical, error-free responses OR The responses contain hints to keep the conversation going (e.g., "Alligators are dangerous animals... Now, do you want to learn about other animals?) OR Utilize the conversational markers throughout the customized intents when appropriate

APPENDIX D
PRE- AND POST-SURVEY FOR THE AMBY CLASSROOM STUDY

Pre-survey

1. What is your first name?

2. What is your last name?

3. **Ability Beliefs**

   Prompt: How much do you agree or disagree with the following statements?

   Response Options: 1 - Strongly Disagree; 2 - Disagree; 3 - Agree; 4 - Strongly Agree

   a. I know enough about artificial intelligence (AI) to make a chatbot on my own.

   b. I am confident that I can understand AI.

   c. I can figure out how to solve hard AI problems if I try.

4. **Identity**

   a. If I chose to, I could have a job that uses AI.

   b. I see myself using AI in my future job.

   c. I want to use AI in my job.

5. **Persistence**

   a. I would like to learn more about AI in the future.

   b. I would like to take a class in AI.

   c. I would like to join an AI club.

   d. I think I could do work in AI when I grow up.

6. **Prior programming experience**

   Have you ever written a computer program before?

   a. Yes

   b. No

c. Don't Know

Have you ever used: (select all that apply)

a. Block coding (Examples: Scratch, Scratch Jr., Tynker)
b. Robotics (Examples: Lego Robots, Lego Spike, Hummingbird, Root, PicoCrickets, Sphero, Micro:bit)
c. App Programming (Examples: App Lab, App Inventor, Mad-Learn)
d. Graphics, Javascript or web pages/HTML (Examples: Pencil Code, Vidcode, Python Turtle, Grasshopper, Processing)
e. Text-based coding (Example: Python)
f. Conversational agent programming (Example: Dialogflow, AMBY, Alexa skill blueprints)
g. Other (please specify): ____
h. None of the above

7. **Language Background**

Are you a native English speaker?

a. Yes
b. No
c. Not sure

What language(s) do you speak at home? (select all that apply)

a. English
b. Spanish
c. French Creole/French
d. Portuguese
e. German

f. Tagalog
g. Chinese
h. Korean
i. Vietnamese
j. Not Listed: ____

Post Survey

1. What is your first name?

2. What is your last name?

3. **Ability Beliefs**

   Prompt: How much do you agree or disagree with the following statements?

   Response Options: 1 - Strongly Disagree; 2 - Disagree; 3 - Agree; 4 - Strongly Agree

   a. I know enough about artificial intelligence (AI) to make a chatbot on my own.
   b. I am confident that I can understand AI.
   c. I can figure out how to solve hard AI problems if I try.

4. **Identity**

   a. If I chose to, I could have a job that uses AI.
   b. I see myself using AI in my future job.
   c. I want to use AI in my job.

5. **Persistence**

   a. I would like to learn more about AI in the future.
   b. I would like to take a class in AI.
   c. I would like to join an AI club.
   d. I think I could do work in AI when I grow up.

6. **Interest formation**

    a. I want to learn more about conversational AI.

    b. I want to learn more about how conversational AI apps (like Siri, Alexa, Google Home) work.

    c. Time in this class passed quickly while working with AMBY.

    d. I was focused while using AMBY.

    e. Creating a chatbot was exciting.

    f. Creating a chatbot was enjoyable.

    g. Making a chatbot is meaningful to me.

    h. I am proud of the chatbot I created.

    i. I would like to show my chatbot to my friends.

    j. I would like to show my chatbot to my family.

    k. AMBY is something I would like to use at home.

7. **Reflection of the classroom activity (At least 3 sentences)**

    a. What did you learn from the conversational AI lessons and activities?

    b. Did the conversational AI lessons and activities help you understand Science concepts you learn from class? If so, how?

    c. What did you like about the conversational AI lessons and activities?

    d. How could the conversational AI lessons and activities be improved?

8. How old are you?

    a. 10

    b. 11

    c. 12

    d. 13

e. 14
f. 15
g. Not Listed ____

9. What is your gender?
   a. Female
   b. Male
   c. Not Listed ____
   d. Prefer not to answer

10. Which of the following racial or ethnic groups do you most identify with?
   a. Native American
   b. Asian
   c. Black or African American
   d. Hispanic or Latino
   e. White
   f. Not Listed ____
   g. Prefer not to answer

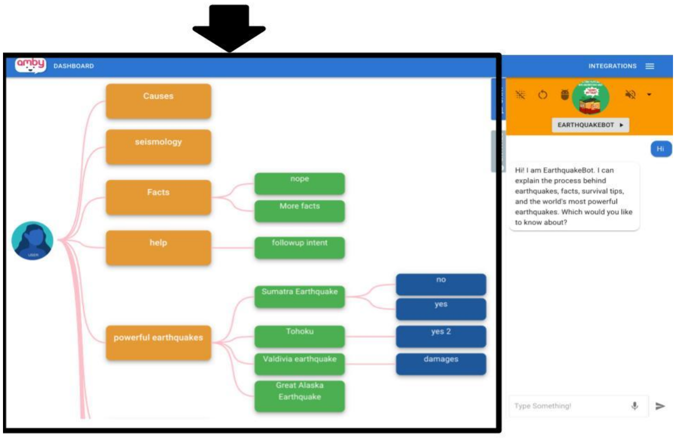
APPENDIX E
CLASSROOM STUDY POST-ASSESSMENT

Next page is the written post-assessment for learners to assess their AI learning from the classroom intervention. Answers highlight in green are the keys.

Dialogs Classroom Study Post Assessment

Instructions: Please read each question carefully and circle the best response.

1. Using the image below, what happens inside the large box?



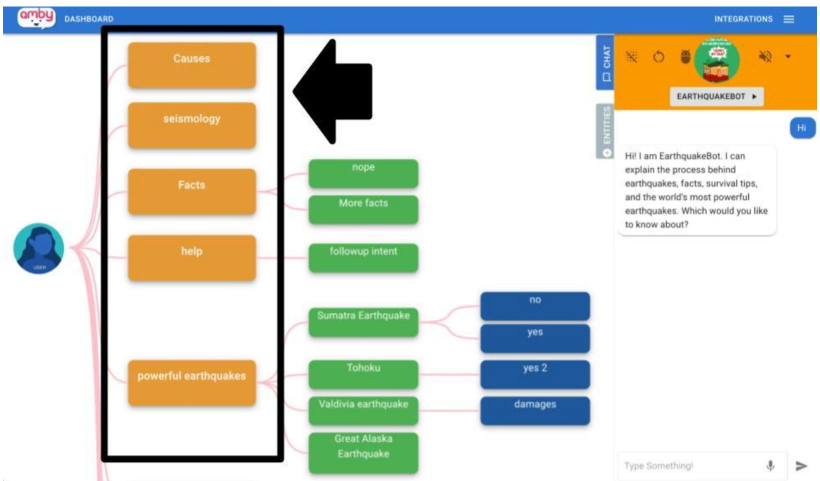
a) The user answers the developer

b) The user talks to the chatbot

c) The developer builds the chatbot

d) The chatbot listens to the user

2. Using the image below to answer the following question. In AMBY, what are the blocks called inside the box?



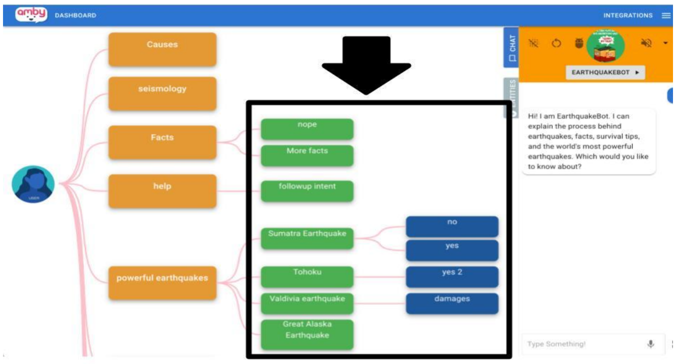
a) Follow-up intents

b) Entities

c) Main intents

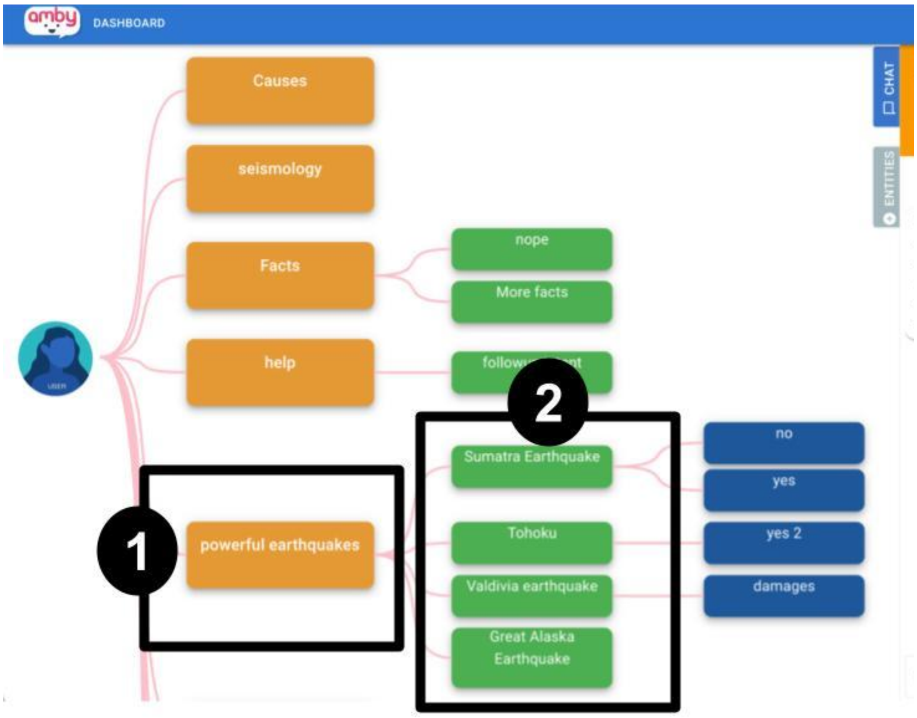
d) Default intent

3. Using the image below, what are the blocks called inside the box?



a) Training phrases
b) Main intents
c) Triggers
d) Follow-up intents

4. Using the image below, what is the relationship between the blocks in Groups 1 and the blocks in Group 2?



a) Group 2 gives more information about Group 1
b) Group 1 is a response for Group 2
c) Group 1 and Group 2 both help the user
d) There is no relationship between the two groups

5. A well- made intent needs as many training phrases as possible. Why does AI need to be trained on multiple training phrases for an intent?

a) Because the system (AMBY) needs two training phrases to work
b) To give the chatbot enough data to trigger the correct intent
c) So that the AI will only understand the phrases that programmer used
d) Chatbots can learn training phrases on their own

Use the following scenario to answer questions 6 & 7:

A developer has the following three training phrases for the intent, "Friend's birthday gift recommendation."

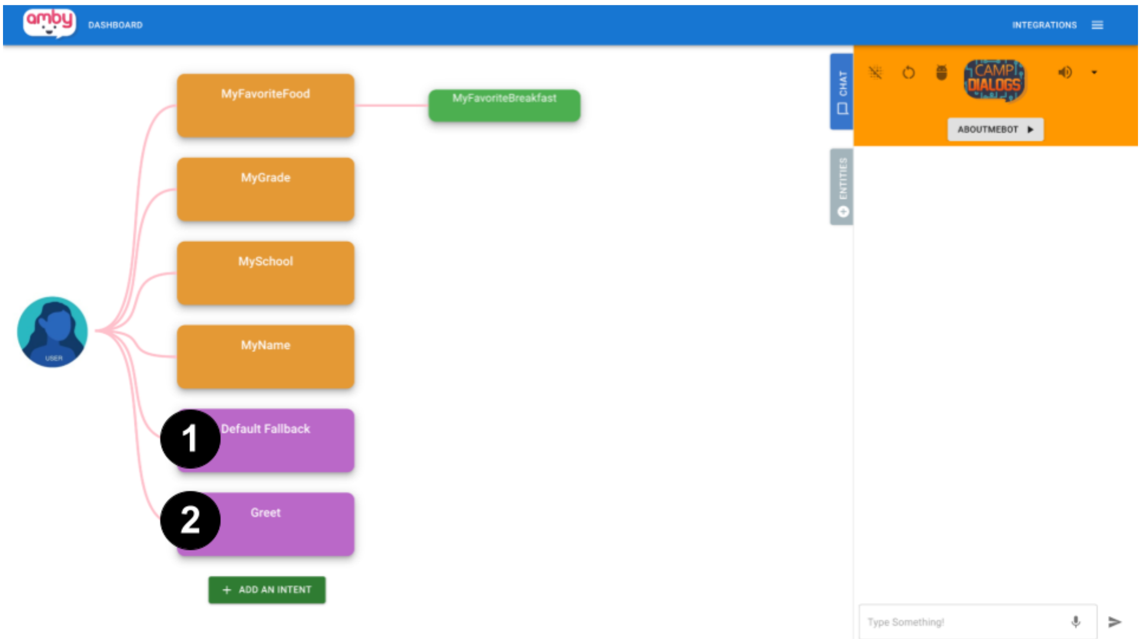
Training Phrases
What gift should I get my friend for their birthday?
I can't decide what to give my friend for their birthday. Any ideas?
I need some ideas to pick out a birthday gift for my friend.

6. What would you suggest as another training phrase for this intent? Write your answer in the box below.

7. Now the intent "Friend's birthday gift recommendation" has been triggered. What would be a good response for this intent?

a) For a gift, would your friend choose video games, books, or gift cards?
b) I don't know how I can help with that.
c) That's exciting! What are you doing for Earth Day?
d) Sure, my favorite basketball player is Lebron James!

<u>Use the following visual for questions 8 - 9.</u>



8. When would the "*Default Fallback" Intent* (represented by 1 on the image) be triggered?

a) When the agent starts the conversation
b) When information about a main intent is provided
c) When there is no intent matching what the user said
d) When the system (AMBY) does not load properly

9. When would the agent use the *"Greet" intent* (represented by 2 on the image above)?

a) When the user asks something the AI does not have information about

b) When the agent starts the conversation with the user

c) When the user needs help

d) When the developer has to end the conversation with the user

10. What is the purpose of follow-up intents?

a) To inform the user of the chatbots abilities

b) AMBY does not have enough room for too many main intents

c) To greet the user

d) To allow users to discuss the same topic further

Here are two possible chatbot responses to a user saying "I need help." Use them to answer question 11.

1 Bot: "I can help you bake delicious cookies!"

2 Bot: "I can help you with anything you want!"

11. Which one is a better conversational design and why?

a. 1, because the bot provides food recipes to the user

b. 1, because it says what the chatbot can help with

c. 2, because it props the user to keep the conversation moving

d. Neither, because they do not offer help

12. Using the information in the box below, which of the following chatbot responses for the *"Default Fallback"* intent is better?

1. Bot: "Sorry, I didn't get that. Can you rephrase?"
2. Bot: "I'm sorry, I don't like that kind of music. Try asking me about country, jazz or pop music."

a) 2, because it redirects the users

b) 1, because it because it is shorter

c) 1, because it is a question

d) 2, because it shows the chatbot likes music

13. You are designing a chatbot that will be personalized and friendly, here are two possible chatbot responses to a user asking "Can you recommend a song?" Which one has a better conversational design? Why?

1. Sure! What kind of music do you like? Country, jazz or pop songs?

2. I recommend 'Break Free' by Ariana Grande.

a. 2, because the user gets an artist recommendation.

b. 1, because the conversation is more interactive and customized to the user.

c. 2, because it does not ask the likes of the users.

d. 1, because it is a longer reply to the user.

14. In a sentence like "I want to book a flight to Paris for tomorrow", which word represents a potential entity related to destination?

a. Tomorrow
b. Book
c. Flight
d. Paris

15. What is an entity in the context of conversational AI?

a. A complete chatbot application.
b. A phrase that the chatbot is trained on.
c. Specific pieces of information that users might provide.
d. The sentiment of a user's message.

LIST OF REFERENCES

[1] Azure bot service – conversational ai application: Microsoft azure, 2021. URL https://azure.microsoft.com/en-us/services/bot-services/.

[2] Rasa: Open source conversational ai, 2021. URL https://rasa.com/.

[3] wit.ai, 2021. URL https://wit.ai/.

[4] Dialogflow, 2022. URL https://dialogflow.cloud.google.com/. (Accessed April, 2022).

[5] Ibm watson assistant — ibm, 2022. URL https://www.ibm.com/products/watson-assistant/integrations.

[6] Mehdi Alaimi, Edith Law, Kevin Daniel Pantasdo, Pierre-Yves Oudeyer, and Hélène Sauzeon. Pedagogical agents for fostering question-asking skills in children. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[7] Vasudeva Rao Aravind and Marcella Kay McConnell. A computer-based tutor for learning energy and power. *World Journal on Educational Technology: Current Issues*, 10(3):174–185, 2018.

[8] Merve Arık and Mustafa Sami Topçu. Implementation of engineering design process in the k-12 science classrooms: Trends and issues. *Research in Science Education*, pages 21–43, 2020.

[9] Hilary Arksey and Lisa O'Malley. Scoping studies: towards a methodological framework. *International journal of social research methodology*, 8(1):19–32, 2005.

[10] Saptarashmi Bandyopadhyay, Jason Xu, Neel Pawar, and David Touretzky. Interactive visualizations of word embeddings for k-12 students. In *EAAI-22: The 12th Symposium on Educational Advances in Artificial Intelligence*, 2022.

[11] Amy L Baylor and Jeeheon Ryu. The effects of image and animation in enhancing pedagogical agent persona. *Journal of Educational Computing Research*, 28(4): 373–394, 2003.

[12] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. Communication breakdowns between families and alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[13] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A Kientz, Jason Yip, and Alexis Hiniker. Parenting with alexa: Exploring the introduction of smart speakers on family dynamics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[14] Dania Bilal. Children's use of the yahooligans! web search engine: I. cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for information Science*, 51(7):646–665, 2000.

[15] Julia Cambre and Chinmay Kulkarni. Methods and tools for prototyping voice interfaces. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, pages 1–4, 2020.

[16] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. Teachable machine: Approachable web-based tool for exploring machine learning classification. In *CHI EA '20: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2020.

[17] Fabio Catania, Micol Spitale, Giulia Cosentino, and Franca Garzotto. What is the best action for children to "wake up" and "put to sleep" a conversational agent? a multi-criteria decision analysis approach. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, pages 1–10, 2020.

[18] Veronica Cateté, Nicholas Lytle, Yihuan Dong, Danielle Boulden, Bita Akram, Jennifer Houchins, Tiffany Barnes, Eric Wiebe, James Lester, Bradford Mott, et al. Infusing computational thinking into middle grade science classrooms: lessons learned. In *Proceedings of the 13th Workshop in Primary and Secondary Computing Education*, pages 1–6, 2018.

[19] Mehmet Celepkolu, Erin O'Halloran, and Kristy Elizabeth Boyer. Upper elementary and middle grade teachers' perceptions, concerns, and goals for integrating cs into classrooms. In *Proceedings of the 51st ACM technical symposium on computer science education*, pages 965–970, 2020.

[20] Vanessa Cesário and Valentina Nisi. Designing with teenagers: A teenage perspective on enhancing mobile museum experiences. *International Journal of Child-Computer Interaction*, 33:100454, 2022.

[21] Sharon Lynn Chu, Francis Quek, Sourabh Bhangaonkar, Amy Boettcher Ging, and Kumar Sridharamurthy. Making the maker: A means-to-an-ends approach to nurturing the maker mindset in elementary-aged children. *International Journal of Child-Computer Interaction*, 5:11–19, 2015.

[22] Oswald Comber, Renate Motschnig, Barbara Göbl, Hubert Mayer, and Esra Ceylan. Exploring students' stereotypes regarding computer science and stimulating reflection on roles of women in it. In *2021 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE, 2021.

[23] Stefania Druga. Growing up with ai: Cognimates: from coding to teaching machines. Master's thesis, Massachusetts Institute of Technology, 2018.

[24] Stefania Druga and Amy J Ko. How do children's perceptions of machine intelligence change when training and coding smart programs? In *Interaction design and children*, pages 49–61, 2021.

[25] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. " hey google is it ok if i eat you?" initial explorations in child-agent interaction. In *Proceedings of the 2017 conference on interaction design and children*, pages 595–600, 2017.

[26] Jennifer Fereday and Eimear Muir-Cochrane. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods*, 5(1):80–92, 2006. doi: 10.1177/160940690600500107.

[27] Jamie Fraser, Ioannis Papaioannou, and Oliver Lemon. Spoken conversational ai in video games: Emotional dialogue management increases user engagement. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 179–184, 2018.

[28] Juan David Rodríguez García, Jesús Moreno-León, Marcos Román-González, and Gregorio Robles. Learningml: a tool to foster computational thinking skills through practical artificial intelligence projects. *Revista de Educación a Distancia (RED)*, 20 (63), 2020.

[29] Radhika Garg and Subhasree Sengupta. Conversational technologies for in-home learning: using co-design to understand children's and parents' perspectives. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[30] Radhika Garg, Hua Cui, Spencer Seligson, Bo Zhang, Martin Porcheron, Leigh Clark, Benjamin R Cowan, and Erin Beneteau. The last decade of hci research on children and voice-based conversational agents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2022.

[31] Bhavya Ghai, Md Naimul Hoque, and Klaus Mueller. WordBias: An Interactive Visual Tool for Discovering Intersectional Biases Encoded in Word Embeddings. *arXiv:2103.03598 [cs]*, September 2021. URL http://arxiv.org/abs/2103.03598. arXiv: 2103.03598.

[32] Lauren N Girouard-Hallam and Judith H Danovitch. Children's trust in and learning from voice assistants. *Developmental Psychology*, 58(4):646, 2022.

[33] David Antonio Gómez Jáuregui, Léonor Philip, Céline Clavel, Stéphane Padovani, Mahin Bailly, and Jean-Claude Martin. Video analysis of approach-avoidance behaviors of teenagers speaking with virtual agents. In *Proceedings of the 15th ACM International Conference on Multimodal Interaction*, pages 189–196, 2013.

[34] Christiane Gresse von Wangenheim, Jean C. R. Hauck, Fernando S. Pacheco, and Matheus F. Bertonceli Bueno. Visual tools for teaching machine learning in K-12: A ten-year systematic mapping. *Education and Information Technologies*, 26(5): 5733–5778, September 2021. ISSN 1360-2357, 1573-7608. doi: 10.1007/s10639-021-10570-8. URL https://link.springer.com/10.1007/s10639-021-10570-8.

[35] David Griol and Zoraida Callejas. Mobile conversational agents for context-aware care applications. *Cognitive Computation*, 8(2):336–356, 2016.

[36] Shuchi Grover, Roy Pea, and Stephen Cooper. Remedying misperceptions of computer science among middle school students. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, SIGCSE '14, page 343–348, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450326056. doi: 10.1145/2538862.2538934. URL https://doi.org/10.1145/2538862.2538934.

[37] Julia D Harbeck and Thomas M Sherman. Seven principles for designing developmentally appropriate web sites for young children. *Educational Technology*, 39 (4):39–44, 1999.

[38] Brian Harvey and Jens Mönig. Bringing "no ceiling" to scratch: Can one language serve kids and computer scientists. *Proc. Constructionism*, pages 1–10, 2010.

[39] Khe Foon Hew and Wing Sum Cheung. Use of three-dimensional (3-d) immersive virtual worlds in k-12 and higher education settings: A review of the research. *British Journal of Educational Technology*, 41(1):33–55, 2010.

[40] Suzanne Hidi and K Ann Renninger. The four-phase model of interest development. *Educational psychologist*, 41(2):111–127, 2006.

[41] Arthur Hjorth. Naturallanguageprocesing4all: -a constructionist nlp tool for scaffolding students' exploration of text. In *Proceedings of the 17th ACM Conference on International Computing Education Research*, pages 347–354, 2021.

[42] Arthur Hjorth. Naturallanguageprocesing4all. In *Proceedings of the 17th ACM Conference on International Computing Education Research*, pages 28–33, 2021.

[43] Anna Hoffman, Diana Owen, and Sandra L Calvert. Parent reports of children's parasocial relationships with conversational agents: Trusted voices in children's lives. *Human Behavior and Emerging Technologies*, 3(4):606–617, 2021.

[44] Hsiu-Fang Hsieh and Sarah E Shannon. Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288, 2005.

[45] W Lewis Johnson, Jeff W Rickel, James C Lester, et al. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11(1):47–78, 2000.

[46] Daniel Jurafsky and James H. Martin. *Chapter 24: Chatbots and Dialogue Systems*. 3rd edition, 2021.

[47] Juho Kahila, Jaana Viljaranta, Sanni Kahila, Satu Piispa-Hakala, and Henriikka Vartiainen. Gamer rage—children's perspective on issues impacting losing one's temper while playing digital games. *International Journal of Child-Computer Interaction*, 33:100513, 2022.

[48] K Megasari Kahn, Rani Megasari, Erna Piantari, and Enjun Junaeti. Ai programming by children using snap! block programming in a developing country. 2018.

[49] Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, Maarten Van Mechelen, Arthur Hjort, Niels Olof Bouvin, and Marianne Graves Petersen. High school students exploring machine learning and its societal implications: Opportunities and challenges. *International Journal of Child-Computer Interaction*, page 100539, 2022.

[50] Gloria Ashiya Katuka, Yvonika Auguste, Yukyeong Song, Xiaoyi Tian, Amit Kumar, Mehmet Celepkolu, Kristy Elizabeth Boyer, Joanne Barrett, Maya Israel, and Tom McKlin. A summer camp experience to engage middle school learners in ai through conversational app development. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, pages 813–819, 2023.

[51] Taylor M Kessner and Lauren McArthur Harris. Opportunities to practice historical thinking and reasoning in a made-for-school history-oriented videogame. *International Journal of Child-Computer Interaction*, 34:100545, 2022.

[52] Hankyung Kim, Dong Yoon Koh, Gaeun Lee, Jung-Mi Park, and Youn-kyung Lim. Designing personalities of conversational agents. In *CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2019.

[53] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th international conference on intelligent user interfaces*, pages 126–137, 2015.

[54] Amit Kumar, Xiaoyi Tian, Mehmet Celepkolu, Maya Israel, and Kristy Elizabeth Boyer. Early design of a conversational ai development platform for middle schoolers. In *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 1–3. IEEE Computer Society, 2022.

[55] Dale Lane. Machine learning for kids, 2018. URL https://machinelearningforkids.co.uk/.

[56] JA Large and Jamshid Beheshti. Interface design, web portals, and children. *Library Trends*, 54(2):318–342, 2005.

[57] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research Methods in Human-Computer Interaction*. Morgan Kaufmann, Cambridge, MA, USA, 2017.

[58] Irene Lee, Fred Martin, Jill Denner, Bob Coulter, Walter Allan, Jeri Erickson, Joyce Malyn-Smith, and Linda Werner. Computational thinking for youth in practice. *Acm Inroads*, 2(1):32–37, 2011.

[59] Irene Lee, Safinah Ali, Helen Zhang, Daniella DiPaola, and Cynthia Breazeal. Developing middle school students' ai literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 191–197, 2021.

[60] Jamy Li, René Kizilcec, Jeremy Bailenson, and Wendy Ju. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior*, 55: 1222–1230, 2016.

[61] Lizi Liao, Yunshan Ma, Xiangnan He, Richang Hong, and Tat-seng Chua. Knowledge-aware multimodal dialogue systems. In *Proceedings of the 26th ACM International Conference on Multimedia*, pages 801–809, 2018.

[62] Phoebe Lin, Jessica Van Brummelen, Galit Lukin, Randi Williams, and Cynthia Breazeal. Zhorai: Designing a conversational agent for children to explore machine learning concepts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13381–13388, 2020.

[63] Annabel Lindner, Stefan Seegerer, and Ralf Romeike. Unplugged activities in the context of ai. In *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*, pages 123–135. Springer, 2019.

[64] Duri Long and Brian Magerko. What is ai literacy? competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2020.

[65] Silvia B Lovato and Anne Marie Piper. Young children and voice search: What we know from human-computer interaction research. *Frontiers in psychology*, 10:8, 2019.

[66] Silvia B Lovato, Anne Marie Piper, and Ellen A Wartella. Hey google, do unicorns exist?: Conversational agents as a path to answers to children's questions. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, pages 301–313, 2019.

[67] Thomas Mahatody, Mouldi Sagar, and Christophe Kolski. State of the art on the cognitive walkthrough method, its variants and evolutions. *International Journal of Human–Computer Interaction*, 26(8):741–785, 2010.

[68] James Mylet. Amazon lex, 2012. URL https://aws.amazon.com/lex/.

[69] Zuzana Nevěřilová and Adam Rambousek. How to Present NLP Topics to Children? *Tenth Workshop on Recent Advances in Slavonic Natural Languages Processing*, page 8, 2016.

[70] Tuan D Nguyen, Chanh B Lam, and Paul Bruno. Is there a national teacher shortage? a systematic examination of reports of teacher shortages in the united states. *Annenberg Institute at Brown University*, 2022.

[71] Cansu Oranç and Azzurra Ruggeri. "alexa, let me ask you something different" children's adaptive information search with voice assistants. *Human Behavior and Emerging Technologies*, 3(4):595–605, 2021.

[72] Outlier Research & Evaluation. Basics study ecs teacher implementation and contextual factor questionnaire measures [measurement scales]. http://outlier.uchicago.edu/basics/resources/MeasuresTeacherImplementation/, September 2017.

[73] Jaehyun Park, Sung H Han, Hyun K Kim, Seunghwan Oh, and Heekyung Moon. Modeling user experience: A case study on a mobile device. *International Journal of Industrial Ergonomics*, 43(2):187–196, 2013.

[74] Kyungjin Park, Hyunwoo Sohn, Wookhee Min, Bradford Mott, Krista Glazewski, C Hmelo-Silver, and James Lester. Disruptive talk detection in multi-party dialogue within collaborative learning environments with a regularized user-aware network. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2022.

[75] Kate Pearce, Sharifa Alghowinem, and Cynthia Breazeal. Build-a-bot: Teaching conversational ai using a transformer-based intent recognition and question answering architecture. *arXiv preprint arXiv:2212.07542*, 2022.

[76] Jan Piasecki, Marcin Waligora, and Vilius Dranseika. Google search as an additional source in systematic reviews. *Science and engineering ethics*, 24:809–810, 2018.

[77] Yufeng Qian. Learning in 3-d virtual worlds: Rethinking media literacy. *Educational Technology*, pages 38–41, 2008.

[78] Mary Elizabeth Raven and Alicia Flanders. Using contextual inquiry to learn about your audiences. *ACM SIGDOC Asterisk Journal of Computer Documentation*, 20(1): 1–13, 1996.

[79] Jeba Rezwana, Mary Lou Maher, and Nicholas Davis. Creative penpal: A virtual embodied conversational ai agent to improve user engagement and collaborative experience in human-ai co-creative design ideation. In *Joint Proceedings of the ACM IUI 2021 Workshops co-located with ACM Conference on Intelligent User Interfaces (ACM IUI 2021)*, 2021.

[80] Juan David Rodríguez-García, Jesús Moreno-León, Marcos Román-González, and Gregorio Robles. Evaluation of an online intervention to teach artificial intelligence with learningml to 10-16-year-old students. In *Proceedings of the 52nd ACM technical symposium on computer science education*, pages 177–183, 2021.

[81] Daniel Rough and Benjamin Cowan. Don't believe the hype! white lies of conversational user interface creation tools. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, pages 1–3, 2020.

[82] Elisa Rubegni and Monica Landoni. Fiabot! design and evaluation of a mobile storytelling application for schools. In *Proceedings of the 2014 conference on Interaction design and children*, pages 165–174, 2014.

[83] Amrita Saha, Mitesh Khapra, and Karthik Sankaranarayanan. Towards building large scale multimodal domain-aware conversation systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[84] Theresa Schachner, Roman Keller, Florian Von Wangenheim, et al. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. *Journal of medical Internet research*, 22(9):e20701, 2020.

[85] Stefan Schaffer and Norbert Reithinger. Conversation is multimodal: thus conversational user interfaces should be as well. In *Proceedings of the 1st International Conference on Conversational User Interfaces*, pages 1–3, 2019.

[86] Marie-Monique Schaper, Rachel Charlotte Smith, Mariana Aki Tamashiro, Maarten Van Mechelen, Mille Skovhus Lunding, Karl-Emil Kjæer Bilstrup, Magnus Høholt Kaspersen, Kasper Løvborg Jensen, Marianne Graves Petersen, and Ole Sejer Iversen. Computational empowerment in practice: Scaffolding teenagers' learning about emerging technologies and their ethical and societal impact. *International Journal of Child-Computer Interaction*, page 100537, 2022.

[87] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. "hey alexa, what's up?" a mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 Designing Interactive Systems Conference*, pages 857–868, 2018.

[88] David Williamson Shaffer and Mitchel Resnick. " thick" authenticity: New media and authentic learning. *Journal of interactive learning research*, 10(2):195–216, 1999.

[89] Yukyeong Song, Gloria Ashiya Katuka, Joanne Barrett, Xiaoyi Tian, Amit Kumar, Tom McKlin, Mehmet Celepkolu, Kristy Elizabeth Boyer, and Maya Israel. Ai made by youth: A conversational ai curriculum for middle school summer camps. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Innovative Applications of Artificial Intelligence Conference and Thirteenth AAAI Symposium on Educational Advances in Artificial Intelligence*, 2023.

[90] Qingfeng Sun, Yujing Wang, Can Xu, Kai Zheng, Yaming Yang, Huang Hu, Fei Xu, Jessica Zhang, Xiubo Geng, and Daxin Jiang. Multimodal dialogue response generation. *arXiv preprint arXiv:2110.08515*, 2021.

[91] Jamaliah Taslim, Wan Adilah Wan Adnan, and Noor Azyanti Abu Bakar. Investigating children preferences of a user interface design. In *International Conference on Human-Computer Interaction*, pages 510–513. Springer, 2009.

[92] Cansu Tatar and Deniz Eseryel. A literature review: Fostering computational thinking through game-based learning in k-12. *Association for Educational Communications and Technology*, pages 288–297, 2019.

[93] Katja Thoring, Roland M Müller, et al. Understanding design thinking: A process model based on method engineering. In *DS 69: Proceedings of E&PDE 2011, the 13th International Conference on Engineering and Product Design Education, London, UK, 08.-09.09. 2011*, pages 493–498, 2011.

[94] Xiaoyi Tian, Zak Risha, Ishrat Ahmed, Arun Balajiee Lekshmi Narayanan, and Jacob Biehl. Let's talk it out: A chatbot for effective study habit behavioral change. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–32, 2021.

[95] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. Envisioning ai for k-12: What should every child know about ai? In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 9795–9799, 2019.

[96] David S Touretzky and Christina Gardner-McCune. Artificial intelligence thinking in k-12. pages 153–180.

[97] G Tsayang and DM Totev. Creativity in primary education: The role of multimedia. *International Journal of Internet Education*, 19(2):28–35, 2020.

[98] Jessica Van Brummelen. Tools to create and democratize conversational artificial intelligence. Master's thesis, Massachusetts Institute of Technology, 2019.

[99] Jessica Van Brummelen, Tommy Heng, and Viktoriya Tabunshchyk. Teaching tech to talk: K-12 conversational artificial intelligence literacy curriculum and development tools. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2021.

[100] Jessica Van Brummelen, Viktoriya Tabunshchyk, and Tommy Heng. "alexa, can i program you?": Student perceptions of conversational artificial intelligence before and after programming alexa. In *Interaction Design and Children*, pages 305–313, 2021.

[101] Jessica Van Brummelen, Mingyan Claire Tian, Maura Kelleher, and Nghi Hoang Nguyen. Learning affects trust: Design recommendations and concepts for teaching children–and nearly anyone–about conversational agents. *arXiv preprint arXiv:2209.05063*, 2022.

[102] Johanna Viitanen. Contextual inquiry method for user-centred clinical it system design. In *Studies in Health Technology and Informatics, Volume 169: User Centred Networked Health Care*, pages 965–969. IOS Press, 2011.

[103] Jeremy A. Magruder Waisome, Dennis R. Jr. Parnell, Pasha Antonenko, Brian Abramowitz, and Victor Perez. Shark ai: Teaching middle school students ai fundamentals using fossil shark teeth. In *2023 ASEE Annual Conference & Exposition*, University of Florida, 2023. American Society for Engineering Education. URL https://nemo.asee.org/public/conferences/327/papers/40095/view. Presented at NSF Grantees Poster Session.

[104] Marilyn Walker, Candace Kamm, and Diane Litman. Towards developing general models of usability with paradise. *Natural Language Engineering*, 6(3-4):363–377, 2000.

[105] Isaac Wang and Jaime Ruiz. Examining the use of nonverbal communication in virtual agents. *International Journal of Human–Computer Interaction*, 37(17):1648–1673, 2021.

[106] Randi Williams, Safinah Ali, Nisha Devasia, Daniella DiPaola, Jenna Hong, Stephen P Kaputsos, Brian Jordan, and Cynthia Breazeal. Ai+ ethics curricula for middle school youth: Lessons learned from three project-based curricula. *International Journal of Artificial Intelligence in Education*, 33(2):325–383, 2023.

[107] Ko-chiu Wu, Yun-meng Tang, and Cheng-yu Tsai. Graphical interface design for children seeking information in a digital library. *Visualization in Engineering*, 2(1): 1–14, 2014.

[108] Ying Xu and Mark Warschauer. " elinor is talking to me on the screen!" integrating conversational agents into children's television programming. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2020.

[109] Ying Xu and Mark Warschauer. Exploring young children's engagement in joint reading with a conversational agent. In *Interaction Design and Children*, pages 216–228, 2020.

[110] Ying Xu, Stacy Branham, Xinwei Deng, Penelope Collins, and Mark Warschauer. Are current voice interfaces designed to support children's language development? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2021.

[111] Ying Xu, Joseph Aubele, Valery Vigil, Andres S Bustamante, Young-Suk Kim, and Mark Warschauer. Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Development*, 93(2):e149–e167, 2022.

[112] L. Yan, L. Sha, L. Zhao, Y. Li, R. Martinez-Maldonado, G. Chen, X. Li, Y. Jin, and D. Gašević. Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 00:1–23, 2023. doi: https://doi.org/10.1111/bjet.13370.

[113] Jessica Zhu. Creating your own conversational artificial intelligence agents using convo, a conversational programming system. Master's thesis, Massachusetts Institute of Technology, 2021.

[114] Jessica Zhu and Jessica Van Brummelen. Teaching students about conversational ai using convo, a conversational programming agent. In *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 1–5, 2021.

[115] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K Kane, and R Benjamin Shapiro. Youth learning machine learning through building models of athletic moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, pages 121–132, 2019.

BIOGRAPHICAL SKETCH

Xiaoyi Tian grew up in Songyuan, a city in Jilin province, located in northeastern China. Her passion for education first took root in middle school, inspired by the lasting influence of her teacher Zhao Zhenjie. In 2018, she completed her Bachelor's degree in Management Science at Anhui University. During her sophomore year, she conducted on her very first independent research project with Dr. Jing Li. This experience has further confirmed her desire to pursue a career at the intersection of education and computing.

After her undergraduate studies, Xiaoyi came to Pittsburgh, United States to pursue a Master's degree in Information Science at the University of Pittsburgh, with a specialization in human-centered computing. During this period, she was involved in several research projects and collaborated with experts in the fields of AI in Education and Human-Computer Interaction, including Dr. Erin Walker, Dr. Rosta Farzan, Dr. Amy Ogan, and Dr. Michael Madaio.

In 2020, she joined the PhD program in Human-Centered Computing at the University of Florida and became a member of Dr. Kristy Boyer's LearnDialogue group to further her research in HCI and education. Her focus is now on introducing Artificial Intelligence to K-12 students and using computational methods to explore linguistic patterns during collaborative learning. She employs a learner-centric approach to understand how students engage with technology in learning environments and aims to design educational technology that support authentic and engaging learning experiences.