International Journal of Child-Computer Interaction 38 (2023) 100618

Contents lists available at ScienceDirect

International Journal of Child-Computer Interaction

journal homepage: www.elsevier.com/locate/ijcci

Research paper

AMBY: A development environment for youth to create conversational agents

Xiaoyi Tian [a,*], Amit Kumar [a], Carly E Solomon [a], Kaceja D Calder [a], Gloria Ashiya Katuka [a], Yukyeong Song [b], Mehmet Celepkolu [a], Lydia Pezzullo [a], Joanne Barrett [b], Kristy Elizabeth Boyer [a], Maya Israel [b]

[a] Department of Computer and Information Science and Engineering, University of Florida, Malachowsky Hall, Gainesville, 32603, Floria, United States
[b] School of Teaching and Learning, University of Florida, 1221 SW 5th Ave, Gainesville, 32601, Floria, United States

ARTICLE INFO

Keywords:
Development environment
Artificial intelligence education
Conversational agent
AI education
Conversational AI
Middle school
Learning tool
Youth
K-12

ABSTRACT

Conversational AIs such as Alexa and ChatGPT are increasingly ubiquitous in young people's lives, but these young users are often not afforded the opportunity to learn about the inner workings of these technologies. One of the most powerful ways to foster this learning is to empower youth to create AI that is personally and socially meaningful to them. We have built a novel development environment, AMBY–"AI Made By You"–for youth to create conversational agents. AMBY was iteratively designed with and for youth aged 12–13 through contextual inquiry and usability studies. AMBY is designed to foster AI learning with features that enable users to generate training datasets and visualize conversational flow. We report on results from a two-week summer camp deployment, and contribute design implications for conversational AI authoring tools that empower AI learning for youth.

1. Introduction

Conversational AIs such as Siri, Google Assistant, and ChatGPT are increasingly ubiquitous in everyday life for users of all ages (Beneteau et al., 2020, 2019; Catania, Spitale, Cosentino, & Garzotto, 2020). Conversational AI applications include virtual agents (Wang & Ruiz, 2021), intelligent personal assistants (Beneteau et al., 2019; Sciuto, Saini, Forlizzi, & Hong, 2018), and chatbots (Tian, Risha, Ahmed, Lekshmi Narayanan, & Biehl, 2021). These applications are often fun and engaging, so they present numerous innovative use cases for young learners, such as increasing engagement in reading (Xu & Warschauer, 2020b), supporting language learning (Gómez Jáuregui et al., 2013; Xu, Branham, Deng, Collins, & Warschauer, 2021), promoting story comprehension and engagement (Xu et al., 2022), and fostering question-asking behaviors (Alaimi, Law, Pantasdo, Oudeyer, & Sauzeon, 2020; Lovato, Piper, & Wartella, 2019).

Although opportunities to *interact with* conversational AI are plentiful, opportunities for young people to *deeply understand* how these technologies work are still scarce. For young learners, developing their own personally meaningful conversational agents can provide rich learning experiences (Druga & Ko, 2021; Lin, Van Brummelen, Lukin, Williams, & Breazeal, 2020; Van Brummelen, Tabunshchyk, & Heng, 2021). However, there is a lack of developmentally appropriate tools for *learning to build conversational AI* (Garg et al., 2022).

This paper 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 without prior programming experience. AMBY supports users in generating training data and visualizing conversation flow. AMBY also allows 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 directly onto a Google Home device.

This paper first describes AMBY's iterative design process with youth ages 12–13, and then presents the results from a deployment study in a summer camp. Our study targets middle school-aged youth because this age has been identified as a key developmental period for interest and identity building (Grover, Pea, & Cooper, 2014). A positive AI learning experience during this age could significantly impact learners' interest and attitudes towards AI (Lee, Ali, Zhang, DiPaola, & Breazeal, 2021). Our process began with a summer camp-based contextual inquiry, followed by the design, implementation, testing and refinement of the novel development environment over the course of a year. We deployed the final AMBY prototype in a two-week AI summer camp in 2022, where 17 youth used AMBY to create 25 conversational AI projects. In this work, we triangulate data from interviews, focus groups, system logs, and researcher observations to investigate how learners used AMBY to create conversational agents and to characterize

* Corresponding author.
E-mail address: tianx@ufl.edu (X. Tian).

https://doi.org/10.1016/j.ijcci.2023.100618
Received 13 December 2022; Received in revised form 10 September 2023; Accepted 17 October 2023
Available online 8 November 2023

the challenges they faced. We found that learners were able to use AMBY to create conversational AI projects that were personally and socially relevant.

The paper makes two main contributions. First, we introduce AMBY, a novel platform that supports youth as they create and deploy their own conversational AI applications. It uses an innovative card-based tree design, unique among conversational AI development tools, to represent conversational flow. Second, we offer design recommendations for building development environments that support youth learners in creating conversational agents. Our findings shed light on future directions for the design and research of youth-centered AI-authoring tools.

2. Background and related work

Our work stands at the intersection of AI in K-12 education and conversational AI development. This section first presents recent efforts toward creating AI learning environments for youth and then provides an overview of existing conversational AI development tools for both youth and adults. It concludes with an overview of fundamental conversational AI concepts and terminology.

2.1. Conversational AI and learning technologies for youth

Conversational agents, or chatbots, communicate with users in natural language (text, speech, or both) (Jurafsky & Martin, 2021). 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, Williams, Breazeal, and Resnick (2017) 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, Owen, and Calvert (2021) 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 (2020) 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 (2019) 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 (2021) 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 (2022) 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 (2020a) 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. (2019) 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.

The child–computer interaction (CCI) community has a long history of designing innovative learning technologies to engage young audiences (Giannakos, Markopoulos, Hourcade, & Antle, 2022; Macrides, Miliou, & Angeli, 2021). Kaspersen et al. (2022) introduced a tool for high school students to explore the ethical implications of machine learning algorithms. Theodoropoulos and Lepouras (2021) reviewed the use of AR technologies to support CS and programming learning. Schaper et al. (2022) highlighted several design principles rooted in CCI literature for engaging teenagers in learning activities about emerging technologies. One principle is "closeness", which aims to provide authentic learning experiences that are personally meaningful (Shaffer & Resnick, 1999). Additionally, the CCI community often works directly in partnership with youth in the target age group when designing technologies for young audiences (Cesário & Nisi, 2022; Chu, Quek, Bhangaonkar, Ging, & Sridharamurthy, 2015), taking into account their needs and desires.

2.2. Youth authorship of conversational AI

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 (Druga, 2018), LearningML (Rodríguez García, Moreno-León, Román-González, & Robles, 2020; Rodríguez-García, Moreno-León, Román-González, & Robles, 2021), ML4K (Machine Learning for Kids) (Lane, 2018), Zhorai (Lin et al., 2020) and eCraft2Learn (Kahn, Prasad, & Veera, 2022). 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 (Dialogflow, 2022), Rasa (Rasa, 2021), IBM Watson (Fraser, Papaioannou, & Lemon, 2018; IBM Watson, 2022), Amazon Lex (Mylet, 2012), Azure Bot Service (Azure Bot Service, 2021), and Wit.ai (Wit.ai, 2021)) 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 (Cambre & Kulkarni, 2020; Rough & Cowan, 2020) and were not designed for fostering AI learning in a robust and authentic manner to 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 (2019) 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 (2021) and Zhu and Van Brummelen (2021), 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"

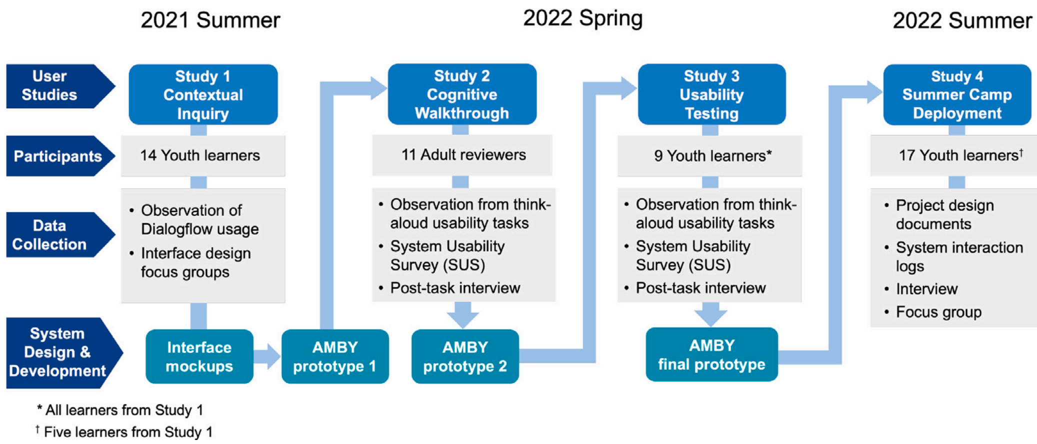

Fig. 1. Overview of the study process and system development of AMBY.

outlined by Touretzky, Gardner-McCune, Martin, and Seehorn (2019). 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 (Hew & Cheung, 2010; Johnson, Rickel, Lester, et al., 2000). 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.

2.3. 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.[1] For example, when a user says, "Can you give me a good movie title?", 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.

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

3. Iterative design studies

To develop AMBY, a novel tool that supports learners to create conversational agents, we utilized an iterative design approach, working with youth at multiple design stages (Fig. 1). This process consisted of four studies in total. Study 1, conducted in 2021, was a contextual inquiry (Section 3.1) during a summer camp with 14 youths. 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).

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 (Raven & Flanders, 1996), which can inform the design of a system that will support an improved work experience for the target users (Lazar, Feng, & Hochheiser, 2017; Viitanen, 2011). In this contextual inquiry study, our goal was (1) to investigate how youth learners use an existing conversational agent development tool, Dialogflow,[2] 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 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).

[2] https://dialogflow.cloud.google.com/.

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

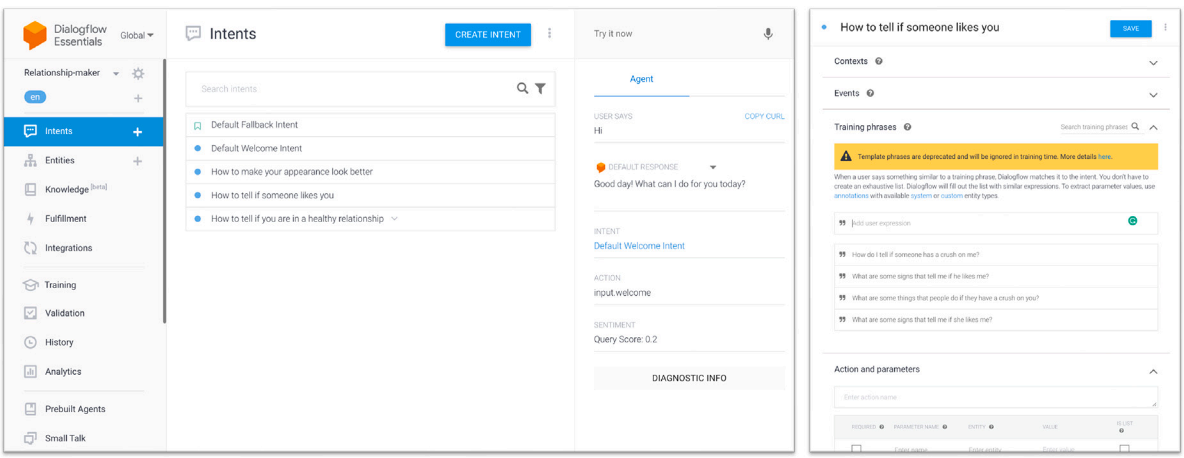

Fig. 3. Dialogflow interface; Left: main development page for intents. Right: intent editing screen.

3.1.2. Camp context with dialogflow

During the two-week summer camp, students learned about foundational principles of artificial intelligence and conversational AI (Fig. 2) (Katuka et al., 2023). 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 2.3). In the second week, learners worked in pairs to build a conversational agent using Dialogflow, with a topic or purpose of 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 (Lindner, Seegerer, & Romeike, 2019) 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 2.2). 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).
- **Overwhelming information from Dialogflow causes frustration**: Dialogflow's screens contain dense text (Fig. 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 4. The design principles were as follows:

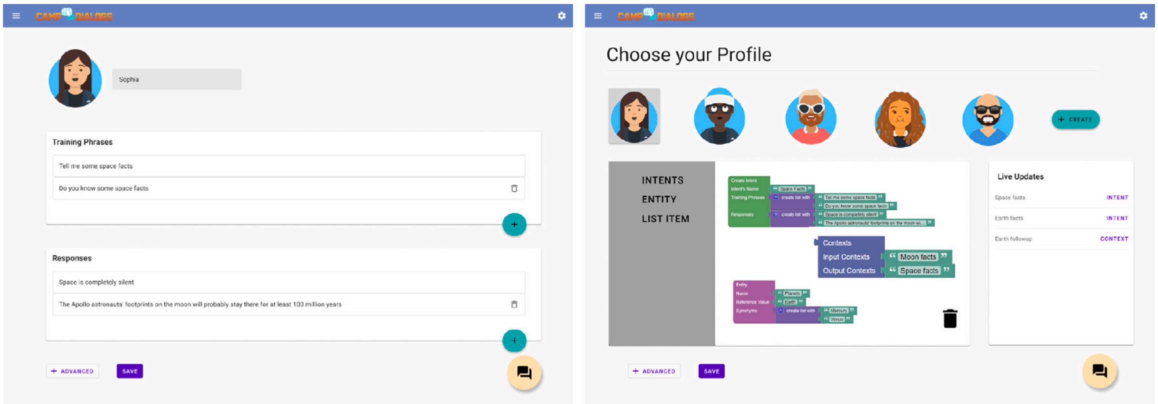

Fig. 4. The two interface mockups used during the focus group in the contextual inquiry study (Study 1).

1. **Foster an accurate conceptualization of conversational AI.** Some related work suggests strategies to introduce young learners to machine learning (Carney et al., 2020; Zimmermann-Niefield, Turner, Murphy, Kane, & Shapiro, 2019) and natural language processing concepts (Bandyopadhyay, Xu, Pawar, & Touretzky, 2022; Druga, 2018; Hjorth, 2021). 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 (Long & Magerko, 2020).
2. **Embodiment of AI agents.** Embodiment of a virtual agent can significantly improve children's engagement in a learning activity (Baylor & Ryu, 2003; Hew & Cheung, 2010; Johnson et al., 2000; Park et al., 2022). Customization of the agent's embodied characteristics, such as gender, skin tone, and voice, can enhance learner's identity (Kim, Koh, Lee, Park, & Lim, 2019) and create a better sense of belonging, thus encouraging youth to engage more with the system (Qian, 2008). However, agent customization options can also distract from the learning activity itself (Li, Kizilcec, Bailenson, & Ju, 2016). 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 (Harbeck & Sherman, 1999), simple and intuitive displays (Bilal, 2000; Taslim, Wan Adnan, & Abu Bakar, 2009; Wu, Tang, & Tsai, 2014), and concepts that are conveyed visually rather than with dense text (Large & Beheshti, 2005; Park, Han, Kim, Oh, & Moon, 2013). 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 (Griol & Callejas, 2016; Schachner, Keller, Von Wangenheim, et al., 2020). Multimodal interaction is especially beneficial for users developing conversational agents (Schaffer & Reithinger, 2019). 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 (Fig. 4). The two interface mockups pared down the information in DialogFlow and displayed it in graphical form inspired by Blockly.[3]

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 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.

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 (Fig. 4 left), many participants agreed that such a simplified interface would help them focus on creating their agents. When considering the block-based interface (Fig. 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 (Lazar et al., 2017; Mahatody, Sagar, & Kolski, 2010).

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 give 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

[3] https://developers.google.com/blockly.

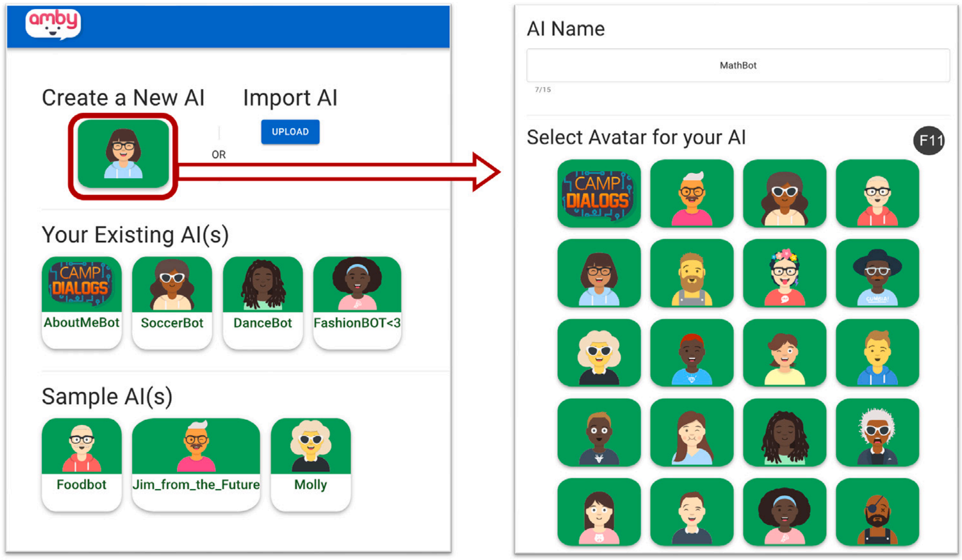

Fig. 5. Left: AMBY dashboard page. Users can create or import a new agent, select an existing agent, or tinker with sample agents. Right: The agent creation window with a collection of avatars that the learner can choose from. Based on focus group insights, avatars anchor the user's first experiences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

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 youth learners

We updated AMBY prototype 1 based on the cognitive walkthrough study (Kumar, Tian, Celepkolu, Israel, & Boyer, 2022). 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.

4. 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.

4.1. AMBY final prototype features

When users first login to AMBY, they land on the Dashboard (Fig. 5, left), where they can (1) create a new AI project, (2) import an AI project from local files, (3) open 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 (Fig. 5, right). Once the user has selected or created a new project, they are then directed to the Playground page (Fig. 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 (Baylor & Ryu, 2003). AMBY provides a menu of avatars (Fig. 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 (Fig. 6) utilizes a card-based tree design to visualize the dialogue structure (as opposed to a block-based development environment). The conversation tree begins

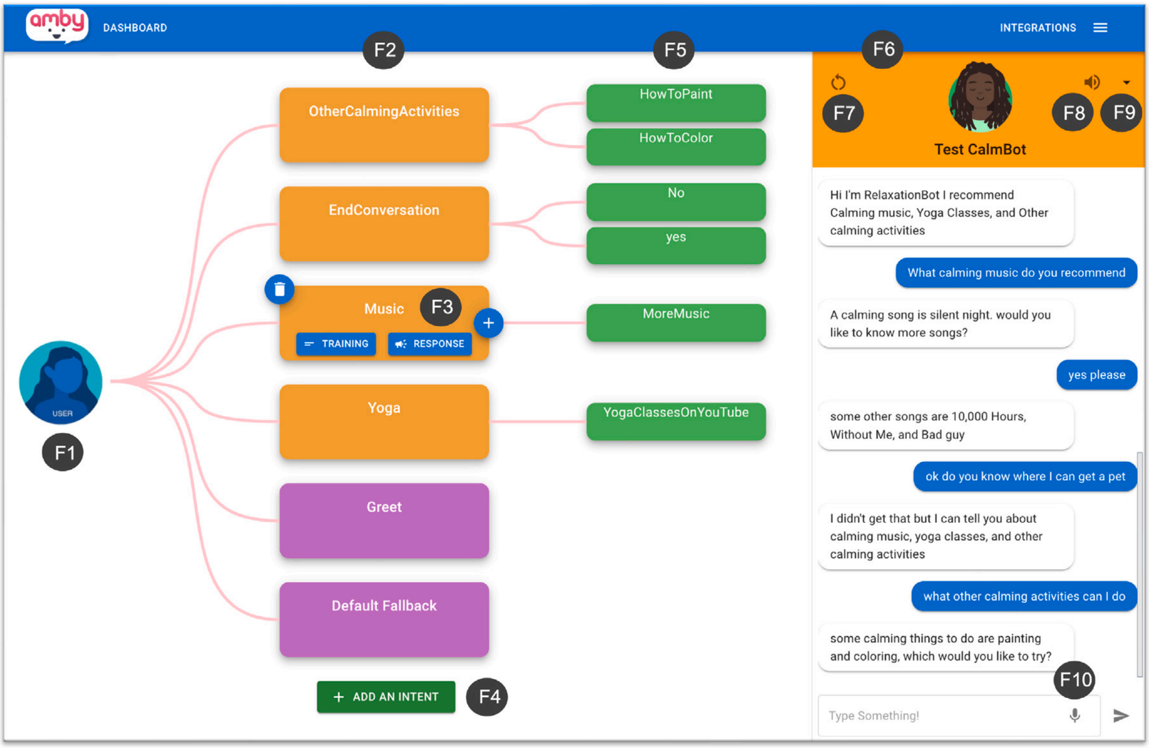

Fig. 6. AMBY playground page. F1–F10 depict specific interface elements, which are detailed in Section 4.1.

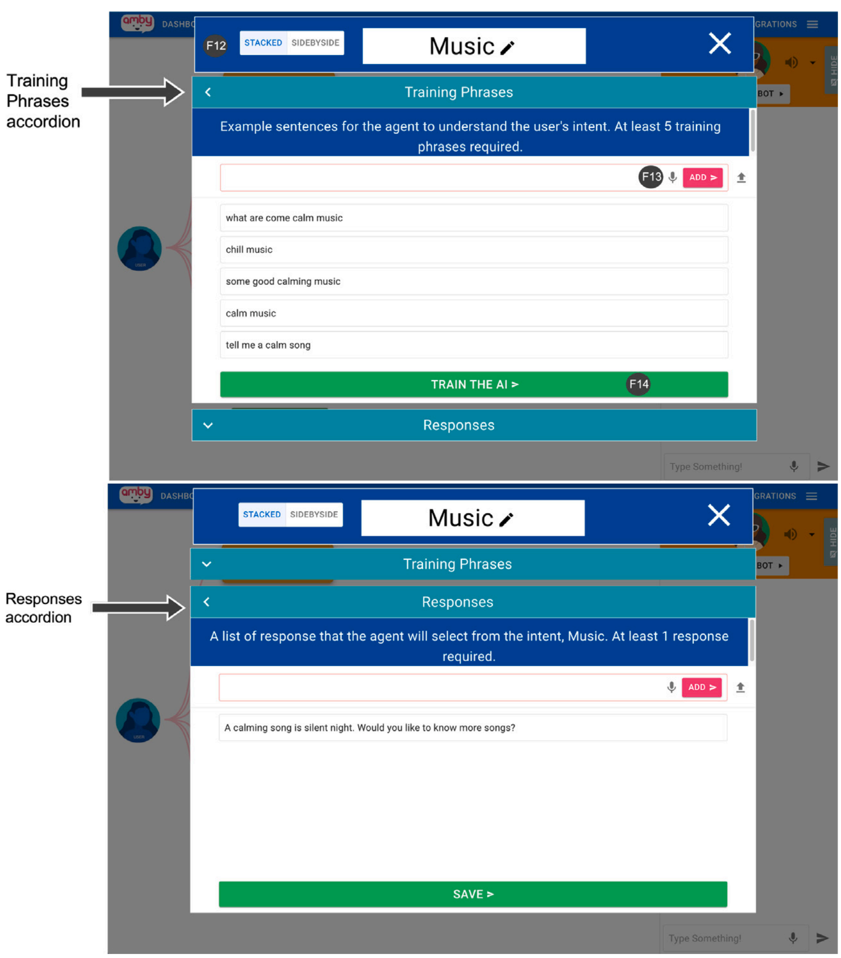

Fig. 7. Intent editing window (stacked view) for training phrases and responses.

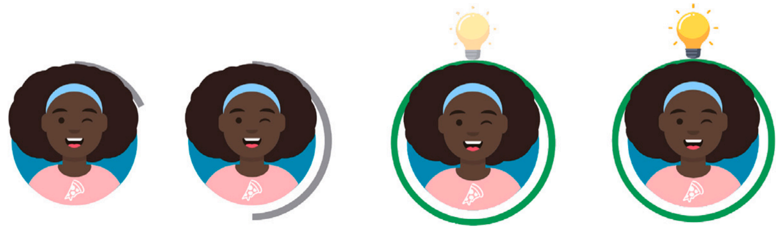

Fig. 8. The agent learning animation (triggered by the “TRAIN THE AI” button (F14) in Fig. 7).

at the end user[4] (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 (Fig. 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 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 (Fig. 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 (Fig. 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.

[4] 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.

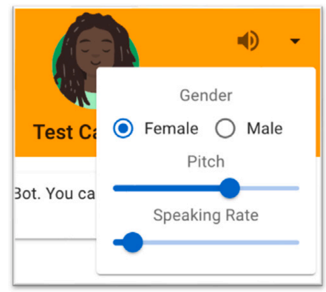

Fig. 9. Voice customization drop-down menu.

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, Fig. 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, Fig. 6 and F13, Fig. 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 (Fig. 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).

4.2. 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: client-side (front-end), Dialogflow interactions, server-side (back-end), and database (Fig. 10).

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 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.

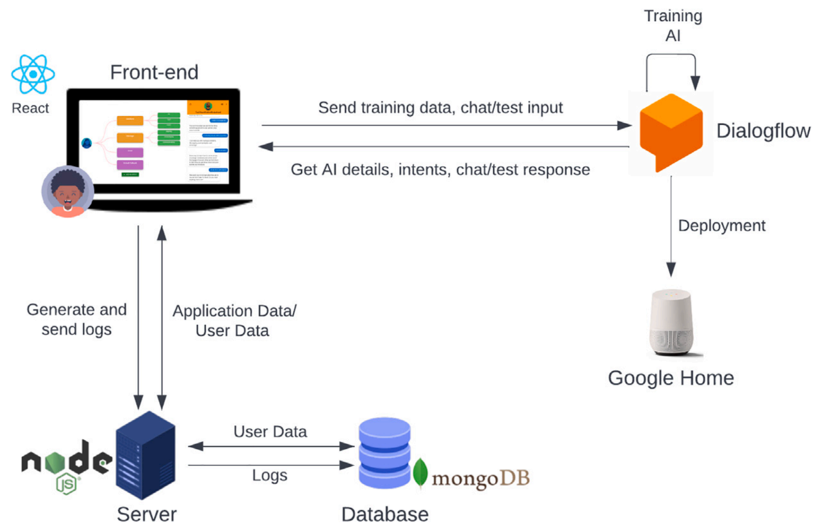

Fig. 10. Technical implementation architecture of AMBY.

5. Study 4: Summer camp 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 study was guided by the research question, *How do youth engage with a development environment designed to support them in making conversational AI?* We answer this question by analyzing several sources of data: (1) the conversational AI projects learners created using AMBY (Section 5.4.1); (2) learners' experiences using AMBY (Section 5.4.2); (3) learners' usage and perception about the features of the interface (Section 5.4.3); and (4) learners' common challenges using AMBY to develop conversational agents (Section 5.4.4).

5.1. Participants

In summer 2022, 17 youth (P1–P17) attended the summer camp. Among these 17 participants, 8 identified as female and 9 as male; 14 were Black/African American, 2 were White/Caucasian, 2 were Hispanic/Latinx, and one identified as multi-racial. The average age of the participants was 12.6 (SD = 0.7) and all participants would rising seventh or eighth-graders in the upcoming school year. Seven participants (41%) reported having no prior coding experience; ten participants (59%) 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.

5.2. Study description

AMBY learning activities spanned eight days over two weeks of the camp. Learners followed a "use-modify-create" progression approach (Lee et al., 2011) 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 (see Fig. 11). At the end of the camp, learners showcased their projects to their peers and family members on the Google Home Mini device. A detailed description of the camp curriculum is outlined in Song et al. (2023).

5.3. Data collection and analysis

During the camp, learners were introduced to design-thinking and engineering design processes (Arık & Topçu, 2020; Thoring, Müller, et al., 2011). We provided a **design log document** (Appendix 9.2) 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 who attended on day 4 when individual projects were completed. Each interview lasted about 15 min 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 (Hsieh & Shannon, 2005), specifically an inductive coding process (Fereday & Muir-Cochrane, 2006), to analyze the interview and focus group data. This method is prevalent in HCI literature (Celepkolu, O'Halloran, & Boyer, 2020; Kahila, Viljaranta, Kahila, Piispa-Hakala, & Vartiainen, 2022; Kessner & Harris, 2022). 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.

Fig. 11. Left: Learners work on their individual projects, mentored by a camp facilitator. Right: Learners work on their paired project.

5.4. Findings

5.4.1. 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 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.

5.4.2. 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"*.

5.4.3. 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, Fig. 6 and F13, Fig. 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).

5.4.4. 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.

6. 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.

6.1. Interfaces should be low-entry, but high-ceiling

Numerous studies have emphasized the importance of offering a low barrier of entry to novice learners (Harvey & Mönig, 2010; Gresse von Wangenheim, Hauck, Pacheco, & Bertonceli Bueno, 2021). 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[5] 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 (Aravind & McConnell, 2018). 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". 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.

6.2. 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 5.4.4), 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"*. This design implication maps to AI literacy competencies, specifically those regarding understanding knowledge

[5] Excluding returning participants, whose prior experience with Dialogflow would likely impact their projects' complexity.

representation and how computers reason and make decisions (Long & Magerko, 2020). Literature suggests that graphical visualizations and interactive demonstrations of models can aid a better understanding of AI (Kulesza, Burnett, Wong, & Stumpf, 2015). 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 (Bandyopadhyay et al., 2022; Ghai, Hoque, & Mueller, 2021; Hjorth, 2021), 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 (Nevěřilová & Rambousek, 2016).

6.3. 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 (Comber, Motschnig, Göbl, Mayer, & Ceylan, 2021). The projects created by the learners, as detailed in Section 5.4.1, 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 in Section 5.4.3, 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 is 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.

6.4. 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 (Tsayang & Totev, 2020) and learner engagement (Rezwana, Maher, & Davis, 2021), and there has been some research into multimodal dialogue systems (Liao, Ma, He, Hong, & Chua, 2018; Saha, Khapra, & Sankaranarayanan, 2018; Sun et al., 2021). There are existing tools such as Adaptive Cards[6] 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.

[6] https://adaptivecards.io/.

6.5. 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 2.2, 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 field notes and observations Kaspersen et al., 2022, focus groups and contextual inquiry Rubegni & Landoni, 2014) within the CCI community when an experimental study is not practical.

7. 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.

This research highlights four design recommendations to enhance interfaces for AI development, particularly for novice and youth learners. Firstly, we advocate for interfaces that are low-entry but high-ceiling, enabling learners to start easily while allowing for the creation of complex projects as their skills develop. We underscore the importance of transparency in AI systems: they should show learners how AI makes decisions and represents knowledge. The study points to the value of designing interfaces that enable learners to express their understanding and knowledge of AI in socially and personally relevant ways. Lastly, we posit the potential benefits of empowering users to incorporate multimedia in their projects, enhancing engagement and creativity. These recommendations constitute practical, user-centered insights to guide the development of more accessible, transparent, and engaging AI interfaces.

8. Selection and participation

We recruited 26 youths and 11 adults in four studies reported in this paper. In study 1, contextual inquiry (camp 2021), we recruited learners from local middle schools and a local after-school program in Gainesville, Florida in the United States. For both study 1 and study 4 (camp 2022), working in partnership with a community liaison, the project team members advertised the camp in person at local sites such as a library and a community resource hub. These efforts were also complemented with the help of digital flyer distributions on regional online platforms that host information for family-friendly events in the area. The camp was offered at no cost to families. The goal of the summer camp was to offer AI learning opportunities to middle school students who typically lack such access in their schools. We received many more applicants than the camp could accommodate, so the selection process prioritized, first, applicants from a lower-resourced geographic area of Gainesville, Florida, and then learners who identified as Black and female. Black youth have historically been excluded from STEM learning opportunities such as CS and AI (Ramsay-Jordan, 2020), and girls are often marginalized in these spaces (Solyst et al., 2022). For study 3 (usability testing), learners were recruited from the participants of study 1, in which we reached out to the parents or guardians by email about the after-school 2-hour workshop opportunity to create conversational apps using AMBY. Before any study with youth learners, including the camps, we obtained parents' consent after informing them about the study, its potential benefits, risks involved, data confidentiality, compensation, researchers' contact information, and the voluntary nature of participation. We also obtained verbal assent from youth participants after providing an age-appropriate version of the same information. For study 2, we recruited the adult participants by personally reaching out to members of an HCI research lab and researchers from the college of education at University of Florida. The adult study participants signed consent forms before participating in the study. All the research studies, procedures, flyers, and forms were approved by the institutional review board (IRB) of the researchers' university.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xiaoyi Tian, Amit Kumar, Carly E Solomon, Kaceja D Calder, Gloria Ashiya Katuka, Yukyeong Song, Mehmet Celepkolu, Lydia Pezzullo, Joanne Barret, Kristy Elizabeth Boyer, Maya Israel reports financial support was provided by National Science Foundation.

Data availability

The authors do not have permission to share data.

Acknowledgments

This research was supported by the National Science Foundation (grant# DRL-2048480). Any opinions, findings, conclusions, or recommendations expressed in this report are those of the authors, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijcci.2023.100618.

References

Alaimi, Mehdi, Law, Edith, Pantasdo, Kevin Daniel, Oudeyer, Pierre-Yves, & Sauzeon, Hélène (2020). Pedagogical agents for fostering question-asking skills in children. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–13).

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

Arık, Merve, & Topçu, Mustafa Sami (2020). Implementation of engineering design process in the K-12 science classrooms: Trends and issues. *Research in Science Education*, 21–43.

Azure bot service – conversational AI application: Microsoft azure. (2021). https://azure.microsoft.com/en-us/services/bot-services/.

Bandyopadhyay, Saptarashmi, Xu, Jason, Pawar, Neel, & Touretzky, David (2022). Interactive visualizations of word embeddings for K-12 students. In *EAAI-22: The 12th symposium on educational advances in artificial intelligence*.

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

Beneteau, Erin, Boone, Ashley, Wu, Yuxing, Kientz, Julie A, Yip, Jason, & Hiniker, Alexis (2020). 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* (pp. 1–13).

Beneteau, Erin, Richards, Olivia K, Zhang, Mingrui, Kientz, Julie A, Yip, Jason, & Hiniker, Alexis (2019). Communication breakdowns between families and alexa. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13).

Bilal, Dania (2000). 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.

Cambre, Julia, & Kulkarni, Chinmay (2020). Methods and tools for prototyping voice interfaces. In *Proceedings of the 2nd conference on conversational user interfaces* (pp. 1–4).

Carney, Michelle, Webster, Barron, Alvarado, Irene, Phillips, Kyle, Howell, Noura, Griffith, Jordan, et al. (2020). 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* (pp. 1–8).

Catania, Fabio, Spitale, Micol, Cosentino, Giulia, & Garzotto, Franca (2020). 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* (pp. 1–10).

Celepkolu, Mehmet, O'Halloran, Erin, & Boyer, Kristy Elizabeth (2020). 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* (pp. 965–970).

Cesário, Vanessa, & Nisi, Valentina (2022). Designing with teenagers: A teenage perspective on enhancing mobile museum experiences. *International Journal of Child-Computer Interaction*, *33*, Article 100454.

Chu, Sharon Lynn, Quek, Francis, Bhangaonkar, Sourabh, Ging, Amy Boettcher, & Sridharamurthy, Kumar (2015). 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.

Comber, Oswald, Motschnig, Renate, Göbl, Barbara, Mayer, Hubert, & Ceylan, Esra (2021). Exploring students' stereotypes regarding computer science and stimulating reflection on roles of women in IT. In *2021 IEEE frontiers in education conference (FIE)* (pp. 1–9). IEEE.

Dialogflow. (2022). https://dialogflow.cloud.google.com/ (Accessed April, 2022).

Druga, Stefania (2018). *Growing up with AI: Cognimates: from coding to teaching machines* (Master's thesis), Massachusetts Institute of Technology.

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

Druga, Stefania, Williams, Randi, Breazeal, Cynthia, & Resnick, Mitchel (2017). "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* (pp. 595–600).

Fereday, Jennifer, & Muir-Cochrane, Eimear (2006). 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. http://dx.doi.org/10.1177/160940690600500107.

Fraser, Jamie, Papaioannou, Ioannis, & Lemon, Oliver (2018). Spoken conversational ai in video games: Emotional dialogue management increases user engagement. In *Proceedings of the 18th international conference on intelligent virtual agents* (pp. 179–184).

Garg, Radhika, Cui, Hua, Seligson, Spencer, Zhang, Bo, Porcheron, Martin, Clark, Leigh, et al. (2022). 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* (pp. 1–19).

Garg, Radhika, & Sengupta, Subhasree (2020). 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* (pp. 1–13).

Ghai, Bhavya, Hoque, Md Naimul, & Mueller, Klaus (2021). WordBias: An interactive visual tool for discovering intersectional biases encoded in word embeddings. http://arxiv.org/abs/2103.03598.

Giannakos, Michail, Markopoulos, Panos, Hourcade, Juan Pablo, & Antle, Alissa N (2022). 'Lots done, more to do': The current state of interaction design and children research and future directions. Article 100469.

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

Gómez Jáuregui, David Antonio, Philip, Léonor, Clavel, Céline, Padovani, Stéphane, Bailly, Mahin, & Martin, Jean-Claude (2013). Video analysis of approach-avoidance behaviors of teenagers speaking with virtual agents. In *Proceedings of the 15th ACM international conference on multimodal interaction* (pp. 189–196).

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

Grover, Shuchi, Pea, Roy, & Cooper, Stephen (2014). Remedying misperceptions of computer science among middle school students. In *Proceedings of the 45th ACM technical symposium on computer science education* (pp. 343–348). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2538862.2538934.

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

Harvey, Brian, & Mönig, Jens (2010). Bringing "no ceiling" to scratch: Can one language serve kids and computer scientists. In *Proc. constructionism* (pp. 1–10).

Hew, Khe Foon, & Cheung, Wing Sum (2010). 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.

Hjorth, Arthur (2021). NaturalLanguageProcesing4All. In *Proceedings of the 17th ACM conference on international computing education research* (pp. 28–33).

Hoffman, Anna, Owen, Diana, & Calvert, Sandra L. (2021). 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.

Hsieh, Hsiu-Fang, & Shannon, Sarah E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.

IBM watson assistant | IBM. (2022). https://www.ibm.com/products/watson-assistant/integrations.

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

Jurafsky, Daniel, & Martin, James H. (2021). Chapter 24: Chatbots and dialogue systems. In *Speech and language processing* (3rd ed.).

Kahila, Juho, Viljaranta, Jaana, Kahila, Sanni, Piispa-Hakala, Satu, & Vartiainen, Henriikka (2022). Gamer rage—Children's perspective on issues impacting losing one's temper while playing digital games. *International Journal of Child-Computer Interaction*, *33*, Article 100513.

Kahn, Ken, Prasad, Ramana, & Veera, Gayathri (2022). AI snap! Blocks for speech input and output, computer vision, word embeddings, and neural net creation, training, and use. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(11), 12861. http://dx.doi.org/10.1609/aaai.v36i11.21568, https://ojs.aaai.org/index.php/AAAI/article/view/21568.

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

Katuka, Gloria Ashiya, Auguste, Yvonika, Song, Yukyeong, Tian, Xiaoyi, Kumar, Amit, Celepkolu, Mehmet, et al. (2023). 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* (pp. 813–819).

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

Kim, Hankyung, Koh, Dong Yoon, Lee, Gaeun, Park, Jung-Mi, & Lim, Youn-kyung (2019). Designing personalities of conversational agents. In *CHI EA '19: Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–6).

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

Kumar, Amit, Tian, Xiaoyi, Celepkolu, Mehmet, Israel, Maya, & Boyer, Kristy Elizabeth (2022). Early design of a conversational AI development platform for middle schoolers. In *2022 IEEE symposium on visual languages and human-centric computing (VL/HCC)* (pp. 1–3). IEEE.

Lane, Dale (2018). Machine learning for kids. https://machinelearningforkids.co.uk/.

Large, J. A., & Beheshti, Jamshid (2005). Interface design, web portals, and children. *Library Trends*, *54*(2), 318–342.

Lazar, Jonathan, Feng, Jinjuan Heidi, & Hochheiser, Harry (2017). *Research methods in human-computer interaction*. Cambridge, MA, USA: Morgan Kaufmann.

Lee, Irene, Ali, Safinah, Zhang, Helen, DiPaola, Daniella, & Breazeal, Cynthia (2021). Developing middle school students' AI literacy. In *Proceedings of the 52nd ACM technical symposium on computer science education* (pp. 191–197).

Lee, Irene, Martin, Fred, Denner, Jill, Coulter, Bob, Allan, Walter, Erickson, Jeri, et al. (2011). Computational thinking for youth in practice. *ACM Inroads*, *2*(1), 32–37.

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

Liao, Lizi, Ma, Yunshan, He, Xiangnan, Hong, Richang, & Chua, Tat-seng (2018). Knowledge-aware multimodal dialogue systems. In *Proceedings of the 26th ACM international conference on multimedia* (pp. 801–809).

Lin, Phoebe, Van Brummelen, Jessica, Lukin, Galit, Williams, Randi, & Breazeal, Cynthia (2020). Zhorai: Designing a conversational agent for children to explore machine learning concepts. In *Proceedings of the AAAI conference on artificial intelligence, Vol. 34* (pp. 13381–13388).

Lindner, Annabel, Seegerer, Stefan, & Romeike, Ralf (2019). Unplugged activities in the context of AI. In *International conference on informatics in schools: situation, evolution, and perspectives* (pp. 123–135). Springer.

Long, Duri, & Magerko, Brian (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16).

Lovato, Silvia B., & Piper, Anne Marie (2019). Young children and voice search: What we know from human-computer interaction research. *Frontiers in Psychology*, *10*, 8.

Lovato, Silvia B., Piper, Anne Marie, & Wartella, Ellen A. (2019). 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* (pp. 301–313).

Macrides, Elena, Miliou, Ourania, & Angeli, Charoula (2021). Programming in early childhood education: A systematic review. *International Journal of Child-Computer Interaction*, Article 100396.

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

Mylet, James (2012). Amazon lex. https://aws.amazon.com/lex/.

Nevěřilová, Zuzana, & Rambousek, Adam (2016). How to Present NLP Topics to Children? In *Tenth workshop on recent advances in slavonic natural languages processing* (p. 8).

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

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

Park, Kyungjin, Sohn, Hyunwoo, Min, Wookhee, Mott, Bradford, Glazewski, Krista, Hmelo-Silver, C, et al. (2022). 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*.

Qian, Yufeng (2008). Learning in 3-D virtual worlds: Rethinking media literacy. *Educational Technology*, 38–41.

Ramsay-Jordan, Natasha N. (2020). Hidden figures: How pecuniary influences help shape STEM experiences for black students in grades K-12. *Journal of Economics, Race, and Policy*, *3*(3), 180–194.

Rasa: Open source conversational AI. (2021). https://rasa.com/.

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

Rezwana, Jeba, Maher, Mary Lou, & Davis, Nicholas (2021). 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)*.

Rodríguez García, Juan David, Moreno-León, Jesús, Román-González, Marcos, & Robles, Gregorio (2020). LearningML: A tool to foster computational thinking skills through practical artificial intelligence projects: Learningml: una herramienta para fomentar las habilidades de pensamiento computacional mediante proyectos prácticos de inteligencia artificial. *Revista de Educación a Distancia (RED)*, *20*(63), http://dx.doi.org/10.6018/red.410121, https://revistas.um.es/red/article/view/410121.

Rodríguez-García, Juan David, Moreno-León, Jesús, Román-González, Marcos, & Robles, Gregorio (2021). 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* (pp. 177–183). Virtual Event USA: ACM, http://dx.doi.org/10.1145/3408877.3432393, https://dl.acm.org/doi/10.1145/3408877.3432393.

Rough, Daniel, & Cowan, Benjamin (2020). Don't believe the hype! White Lies of conversational user interface creation tools. In *Proceedings of the 2nd conference on conversational user interfaces* (pp. 1–3).

Rubegni, Elisa, & Landoni, Monica (2014). Fiabot! Design and evaluation of a mobile storytelling application for schools. In *Proceedings of the 2014 conference on interaction design and children* (pp. 165–174).

Saha, Amrita, Khapra, Mitesh, & Sankaranarayanan, Karthik (2018). Towards building large scale multimodal domain-aware conversation systems. In *Proceedings of the AAAI conference on artificial intelligence, Vol. 32*.

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

Schaffer, Stefan, & Reithinger, Norbert (2019). Conversation is multimodal: thus conversational user interfaces should be as well. In *Proceedings of the 1st international conference on conversational user interfaces* (pp. 1–3).

Schaper, Marie-Monique, Smith, Rachel Charlotte, Tamashiro, Mariana Aki, Van Mechelen, Maarten, Lunding, Mille Skovhus, Bilstrup, Karl-Emil Kjær, et al. (2022). Computational empowerment in practice: Scaffolding teenagers' learning about emerging technologies and their ethical and societal impact. *International Journal of Child-Computer Interaction*, Article 100537.

Sciuto, Alex, Saini, Arnita, Forlizzi, Jodi, & Hong, Jason I. (2018). "Hey alexa, what's up?" a mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 designing interactive systems conference* (pp. 857–868).

Shaffer, David Williamson, & Resnick, Mitchel (1999). "Thick" authenticity: New media and authentic learning. *Journal of Interactive Learning Research*, *10*(2), 195–216.

Solyst, Jaemarie, Nkrumah, Tara, Stewart, Angela, Buddemeyer, Amanda, Walker, Erin, & Ogan, Amy (2022). Running an online synchronous culturally responsive computing camp for middle school girls. In *Proceedings of the 27th ACM conference on on innovation and technology in computer science education, Vol. 1* (pp. 158–164).

Song, Yukyeong, Katuka, Gloria Ashiya, Barrett, Joanne, Tian, Xiaoyi, Kumar, Amit, McKlin, Tom, et al. (2023). 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*.

Sun, Qingfeng, Wang, Yujing, Xu, Can, Zheng, Kai, Yang, Yaming, Hu, Huang, et al. (2021). Multimodal dialogue response generation. arXiv preprint arXiv:2110.08515.

Taslim, Jamaliah, Wan Adnan, Wan Adilah, & Abu Bakar, Noor Azyanti (2009). Investigating children preferences of a user interface design. In *International conference on human-computer interaction* (pp. 510–513). Springer.

Theodoropoulos, Anastasios, & Lepouras, George (2021). Augmented reality and programming education: A systematic review. *International Journal of Child-Computer Interaction*, *30*, Article 100335.

Thoring, Katja, Müller, Roland M., et al. (2011). 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* (pp. 493–498).

Tian, Xiaoyi, Risha, Zak, Ahmed, Ishrat, Lekshmi Narayanan, Arun Balajiee, & Biehl, Jacob (2021). 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.

Touretzky, David, Gardner-McCune, Christina, Martin, Fred, & Seehorn, Deborah (2019). Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the AAAI conference on artificial intelligence, Vol. 33* (pp. 9795–9799).

Tsayang, G., & Totev, D. M. (2020). Creativity in primary education: The role of multimedia. *International Journal of Internet Education*, *19*(2), 28–35.

Van Brummelen, Jessica (2019). *Tools to create and democratize conversational artificial intelligence* (Master's thesis), Massachusetts Institute of Technology.

Van Brummelen, Jessica, Tabunshchyk, Viktoriya, & Heng, Tommy (2021). "Alexa, can I program you?": Student perceptions of conversational artificial intelligence before and after programming alexa. In *Interaction design and children* (pp. 305–313).

Viitanen, Johanna (2011). Contextual inquiry method for user-centred clinical IT system design. In *Studies in health technology and informatics, volume 169: user centred networked health care* (pp. 965–969). IOS Press.

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

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

Wit.ai. (2021). https://wit.ai/.

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

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

Xu, Ying, Branham, Stacy, Deng, Xinwei, Collins, Penelope, & Warschauer, Mark (2021). Are current voice interfaces designed to support children's language development? In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–12).

Xu, Ying, & Warschauer, Mark (2020a). "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* (pp. 1–8).

Xu, Ying, & Warschauer, Mark (2020b). Exploring young children's engagement in joint reading with a conversational agent. In *Interaction design and children* (pp. 216–228).

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

Zhu, Jessica, & Van Brummelen, Jessica (2021). Teaching students about conversational AI using convo, a conversational programming agent. In *2021 IEEE symposium on visual languages and human-centric computing (VL/HCC)* (pp. 1–5).

Zimmermann-Niefield, Abigail, Turner, Makenna, Murphy, Bridget, Kane, Shaun K, & Shapiro, R Benjamin (2019). Youth learning machine learning through building models of athletic moves. In *Proceedings of the 18th ACM international conference on interaction design and children* (pp. 121–132).