

“Alexa, Can I Program You?”: Student Perceptions of Conversational Artificial Intelligence Before and After Programming Alexa

Jessica Van Brummelen
Massachusetts Institute of Technology
Cambridge, MA, USA
jess@csail.mit.edu

Viktoriya Tabunshchyk
Massachusetts Institute of Technology
Cambridge, MA, USA
vikt@mit.edu

Tommy Heng
Massachusetts Institute of Technology
Cambridge, MA, USA
theng@mit.edu

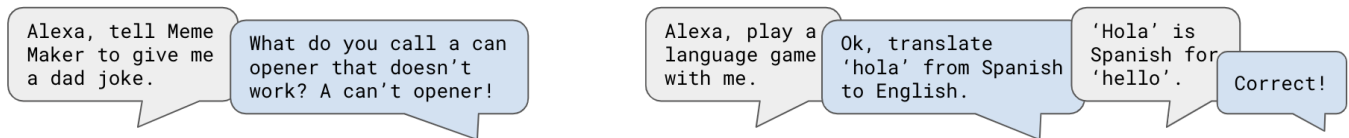


Figure 1: Example conversations from students' Alexa skill designs, including a “Meme Maker” and “Language Game”.

ABSTRACT

Growing up in an artificial intelligence-filled world, with Siri and Amazon Alexa often within arm's—or speech's—reach, could have significant impact on children. Conversational agents could influence how students anthropomorphize computer systems or develop a theory of mind. Previous research has explored how conversational agents are used and perceived by children within and outside of learning contexts. This study investigates how middle and high school students' perceptions of Alexa change through programming their own conversational agents in week-long AI education workshops. Specifically, we investigate the workshops' influence on student perceptions of Alexa's intelligence, friendliness, aliveness, safeness, trustworthiness, human-likeness, and feelings of closeness. We found that students felt Alexa was more intelligent and felt closer to Alexa after the workshops. We also found strong correlations between students' perceptions of Alexa's friendliness and trustworthiness, and safeness and trustworthiness. We recommend designers carefully consider personification, transparency, playfulness and utility when designing conversational agents for learning contexts.

CCS CONCEPTS

• **Social and professional topics** → *K-12 education; Children*; • **Human-centered computing** → **Natural language interfaces**; *User interface programming*; • **Computing methodologies** → *Intelligent agents*.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
IDC '21, June 24–30, 2021, Athens, Greece
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8452-0/21/06.
<https://doi.org/10.1145/3459990.3460730>

KEYWORDS

child-agent interaction; conversational agents; pedagogical agent design; persona; AI education

ACM Reference Format:

Jessica Van Brummelen, Viktoriya Tabunshchyk, and Tommy Heng. 2021. “Alexa, Can I Program You?”: Student Perceptions of Conversational Artificial Intelligence Before and After Programming Alexa. In *Interaction Design and Children (IDC '21)*, June 24–30, 2021, Athens, Greece. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3459990.3460730>

1 INTRODUCTION

With children asking Google to buy them more toys [7], cheating on homework with Alexa [50], and playing voice-based pranks on parents [7], conversational agents (CAs) have the potential to not only influence children's play—but also how they grow and develop [52]. For instance, researchers theorize that interacting with an agent can change people's understanding of agency concepts and their Theory of Mind [21, 29, 51]. Other research has shown engaging with CAs can change people's behavior [1, 10, 20, 46] and have positive effects on information retention [5].

Considering the impact agents have on human understanding and behavior, how prevalent these systems are becoming [40, 52], and how opaque their operations can be to humans [14, 33, 39], a growing body of research suggests it is important for people of all ages to understand AI [32, 39, 53]. Furthermore, researchers are investigating how to best teach AI literacy concepts to students, including those as young as preschoolers [58]. Studies leverage AI ethics discussions [13], interactive, collaborative learning environments [57], students familiarity with CAs [31], and gesture recognition tools [63] to engage students in learning AI. In this work, we use a constructionist approach in which students program their own CAs to teach AI concepts to 6-12th grade students [37, 56].

Another aspect of AI education research includes students' perception of AI systems themselves, including personification of such systems and how the systems work. For example, one study examines preschool- and kindergarten-aged students' perceptions of

“thinking machines” during an AI learning activity, emphasizing the importance of early childhood AI literacy development [58]. Other studies investigate children and families’ perceptions of CAs [33], how interaction modalities influence children’s perceptions of CAs [14], children’s perceptions of maze-solving agents’ intelligence [15], and whether children categorize CAs as animate objects or artifacts [62]. Yet other studies emphasize the importance of adults’ perceptions of AI, especially in decision-making and policy [22, 27, 47]. To our knowledge, few studies investigate middle and high school students’ perceptions of AI [35, 42], despite teenage years being critical in ethical perspective development [11], a key component of AI literacy [53]. Furthermore, to our knowledge, no studies investigate how middle and high school students’ perceptions of CAs change through programming CAs.

We posit that understanding students’ perceptions and feelings towards such agents can help researchers better facilitate student learning. For instance, feelings of closeness with teachers have been shown to affect students’ academic performance [2, 6, 60], which may also be the case when agents take on the teacher role. Another study indicates that the avatar used for pedagogical feedback-giving agents affect students’ emotional attachment and satisfaction with the learning process [45], alluding to the potential for students’ perception of agents to affect learning. Furthermore, research suggests understanding students’ preconceptions and mental models can improve teaching [16, 48]. By understanding students’ feelings and perceptions of agents, we expect we can create better digital learning environments.

This study investigates 6-12th grade students’ perceptions of Amazon Alexa in a learning environment described in [56]. In contrast to [56], which investigates students’ AI literacy, this study investigates how a programming and learning intervention in which students develop their own CAs affects student perceptions of AI. Our main research questions are as follows: **How does building Alexa skills and learning about conversational AI in a remote workshop affect students’ perceptions of Alexa?** and **What CA design insights can we gain from analyzing these perceptions?** By better understanding students’ perspectives on agents and how these perspectives can be changed, we contribute to ongoing research to develop more human-centered, socially useful agents—especially for K-12 education. To this end, we present four design considerations for K-12 education agents and development tools based on our findings. Specifically, we look at students’ perceptions of Alexa in terms of friendliness, human-likeness, aliveness, safeness, trustworthiness, intelligence (generally and relative to themselves), and how close they feel to Alexa.

2 RELATED WORK

2.1 Perceptions of artificial intelligence

The working definition of AI in research has changed over the years—from having a sharp focus on logical, symbolic representations of concepts and actions to a marked concentration on modelling extensive interconnected computation machines called “neural networks” [61]. In the media, AI has been depicted in many different ways—as killer robots, android caretakers, and superintelligent, disembodied voices [19]. Despite the somewhat frivolous

portrayals, people’s understanding of AI and how it works has serious implications—from policy-making to day-to-day assessments of whether a self-driving vehicle is safe to trust one’s life with [22, 34]. In this study, we focus on understanding students’ perceptions of CAs’ *persona*, which we define here as CAs’ person-like or anthropomorphized characteristics, as in [4]. Through better understanding students’ perceptions, we can likely better teach students about AI [21] and therefore better reach our research community’s goal of equipping people to live in an AI-filled world [22, 32, 53].

Children have been observed to anthropomorphize AI systems [14, 15, 62]; however, their understanding of “aliveness” of such systems is inconsistent across populations and seems to vary with age [24, 44, 58, 62]. For example, one study found 5-year-olds were less likely to attribute “brains” to robots than older children [44]. Other anthropomorphic aspects of AI systems have also been investigated for different purposes. For instance, a number of studies examine how children (3-10 years old) perceive agents’ intelligence—generally and relative to their intelligence—with the purpose of inspiring critical thinking. These studies found that 3- to 10-year-olds often perceived particular robots as smarter than themselves, and that the older children in the group tended to perceive conversational agents as smarter than themselves more often than the younger children [14, 15]. Another study investigated 5-6 year old children’s perception of CAs’ to develop CA design recommendations. The study found that more than 60% of the children agreed that the CA was friendly, alive, trustworthy, safe, funny, and intelligent [33]. Other researchers investigated similar anthropomorphic aspects, including how sociable, mutual-liking, attractive, human, close, and intelligent children (10-12 years old) perceived agents to be, finding evidence for learning benefits with anthropomorphized robots [35]. We investigate how middle and high school students’ perceive CAs in terms of related anthropomorphic aspects.

2.2 Conversational agents and education

Many studies investigate how CAs can best embody the teaching role [12, 28, 36, 38] and show that interacting with agents can positively affect learning [12, 21]. In this study, however, we take a constructionist approach, and instead of placing agents in the teaching role, we empower students to learn about AI through developing their own CAs [37, 56].

Constructionism has been shown to be effective in teaching K-12 students AI concepts. For example researchers have taught students AI ethics through constructing paper prototypes [3], machine learning (ML) concepts through developing gesture-based models [63], and AI programming concepts through creating projects with AI cloud services [23]. Our study teaches students Long and Magerko’s AI literacy competencies through developing CAs [32, 56].

Certain studies specifically investigate whether constructionist activities change student perceptions of AI agents. For example, a series of studies showed constructing a robot’s behavior shifted kindergarten students’ perspectives from technological to psychological [30] and language from anthropomorphic to technological [26]. Through an activity with the same constructionist programming environment, it was shown 5- and 7-year-olds were able to better understand robots’ behavior [51]. Another study with

programming and ML training activities showed 4–6-year-olds' perceptions of robots changed throughout the experiment [58]. In this work, we investigate whether middle and high school students' perceptions of AI and CA persona change through a constructionist CA programming activity and workshop.

3 METHODS

3.1 Participants

We conducted our workshops with 47 students in two sessions: a first group of 12 and a second group of 35 (with the first being smaller to ensure the interface worked correctly). For each group, the students' teachers (referred to as 'teachers' in this paper) acted as observers and provided feedback to the three researchers, who taught the workshops [56]. The teachers were recruited through an Amazon Future Engineers call to Title I schools. Each teacher chosen for the workshops was asked to recruit 5 or 6 of their students, with at least 50% of the students identifying as non-male. We targeted Title I schools to provide opportunities for enrichment to low-income families [49]. Students' programming experience ranged from none (9 students), to block-based (17 students), to text-based programming experience (14 students). We developed middle and high school level AI curriculum and thus targeted such students. The students had a mean age of 14.78 (range 11–18, SD=1.91) and identified as male (19), female (27) and one left unanswered. To level-set students' experience with Alexa prior to the conversational AI material, students engaged with Alexa for ~15 minutes.

3.2 Procedure

3.2.1 Programming agents. To accomplish our goal of studying student perceptions of conversational AI before and after programming Alexa, we developed a low-barrier-to-entry, block-based programming interface within MIT App Inventor for creating Alexa skills [54]. As described in [56], once a student creates a skill on the interface, the backend translates their blocks into JSON and Javascript code to be sent to Alexa's API to build and enable the skill on the student's Amazon Developer account. This allows the students to interact and have a conversation with the Alexa skill either on an Alexa-enabled device (e.g., iOS Alexa App or Amazon Echo) or an online simulated Alexa device (e.g., MIT App Inventor Alexa Testing simulator or Amazon Developer Console).

3.2.2 Workshop outline. This section provides an overview of the learning intervention, which is described in-depth in [56]. The curriculum is based on CA workshops from our previous study [54] and AI literacy literature, including [32, 53]. The intervention occurred over two sessions, both involving five consecutive days of 2.5-hour long Zoom sessions. The first day began with an introduction to the MIT App Inventor interface [59] to accustom students to block-based coding. Then the students were given ~15 minutes to interact freely with Alexa, writing down the questions they asked during the interaction. In the first week, students were provided with complimentary Echo Dots. This was not feasible for the second week of workshops due to an increased number of students, so students either used the Alexa mobile app, an online Alexa simulator (e.g., in MIT App Inventor), or Alexa devices they previously owned. Overall, 19 students used an Alexa device, 17

used the Alexa app, 10 used an online simulator, and one did not specify.

The second day involved introducing students to key AI and conversational AI concepts, discussing AI ethics, and completing a tutorial walk-through to create an Alexa skill that would respond to basic greetings. On the third day, students completed a tutorial to develop a calculator skill, in which Alexa could be asked, "What's number A multiplied by number B", or something similar. Next, we taught students about ML in more depth, including discussing the difference between a rule-based CA developed on the first day and the ML-based CAs developed on the second and third days. Finally, students engaged in an AI text generation activity.

On the fourth day, students developed a skill in which Alexa would read out text entered into MIT-App-Inventor-developed mobile apps, after Alexa was asked to "read the app's text". Students then brainstormed ideas for skills for their individual projects. Students spent the final day developing their projects and presenting them to the rest of the class, with 38 students completing presentation slides for their projects.

3.3 Questionnaires and the *Persona Questions*

Various questionnaires inspired by the perception of AI questions in [15] and [33] were given to students during the learning intervention. On the first day, students recorded their interactions with Alexa, impressions of the CA, and demographics information. At the start of the second day, students completed a questionnaire assessing their initial feelings towards Alexa on a 7-point Likert scale, which we call the *Persona Questions*. The questions stated, "Alexa is..." followed by "intelligent", "friendly", "alive", "safe", "trustworthy", "human-like", and "smarter than me". The final *Persona Question* asked how close students felt to Alexa using the *Inclusion of the Other in the Self* scale [18]. At the end of the final day, students completed the *Persona Questions* again. (Note that additional questionnaires were given, but are only discussed and analyzed in [56], which investigates students' understanding of AI and CAs).

3.4 Data Analysis

This study builds on the study presented in [56]. Thus, certain data analyzed in this study (e.g., demographics) is necessarily the same; however, this study focuses on data not analyzed in [56], including the questionnaire responses to the *Persona Questions* and students' reported interactions with Alexa.

To investigate students' reported interactions with Alexa, three researchers performed a reflexive, open-coding thematic analysis [8]. The researchers independently completed familiarization and code-generation stages, inductively coming up with codes. After several discussions, they came to a consensus on codes, which are described in Section 4.1. Codes, representative quotations, and questions from the questionnaires can be found in [55].

To measure changes in the responses to the *Persona Questions*, we employed the Wilcoxon Signed-Rank Test. Additionally, we used the Kendall Tau method to create pairwise correlation matrices. We analyzed the correlation coefficients using Cohen (2013)'s definition for correlation effect strength for behavioral and education psychology [9]. To test the validity of the strength of the coefficients, we compared Kendall Tau p-values to an alpha of 0.05.

Table 1: Types of questions asked by students to Alexa prior to the conversational AI programming intervention.

Type	Example utterances	Instances
Information updates	What time is it?, How is the weather for Wednesday?, How is the traffic?	31 (26%)
Action commands	Set a 15-minute timer, Play my Custom Spotify Playlist, Remind me that I have a meeting at 1:00 pm, What's 0 times 0?	30 (25%)
Other	Hello, Learn my voice, Are dragons real?, What are all the numbers of pi?	24 (20%)
Jokes	Tell me a joke, Can you tell me a joke?	17 (14%)
Personal questions	What's your favorite color?, When were you made?, What's your favorite video game?, How was your day?	16 (14%)

4 RESULTS

4.1 Student interactions with Alexa

To understand the types of interactions students had with Alexa prior to the intervention, we coded the phrases they reported saying to Alexa during the ~15 minute free-interaction activity. We found most of the phrases fell into one of five categories listed in Table 1. Students wrote down an average of three questions each (SD=1.19). The *Information Updates* category involved real-time events; the *Action Commands* category involved built-in Alexa applications; the *Personal Questions* category involved questions about Alexa; the *Jokes* category involved asking Alexa to say a joke; and the *Other* category involved questions and phrases that were often humorous (e.g., “Are dragons real?”) or impossible to fully answer (e.g., “What are all the numbers of pi?”), or generally fell outside of the other categories (e.g., “Hello”). Note that prior to the activity, we asked Alexa to tell us a joke, which may have contributed to a large number of students also asking Alexa for jokes.

4.2 Perceptions of Alexa pre- and post-workshop

By comparing pre- and post-survey answers to the *Persona Questions* (see Figure 2), we found significant differences in how students felt about Alexa’s intelligence and how close they felt they were to Alexa. After the intervention, students felt Alexa was more intelligent ($\bar{x} = 6.0, Mo = 6, |Z| = 2.78, p = 0.003$) and felt closer to Alexa ($\bar{x} = 3.5, Mo = 4, |Z| = 2.75, p = 0.003$). We did not find any evidence of significant differences in how students felt about Alexa being friendly, alive, safe, trustworthy, human-like or smarter than themselves before and after the intervention.

Prior to the intervention, students generally reported Alexa as being highly intelligent ($\bar{x} = 5.6, Mo = 6$), highly friendly ($\bar{x} = 6.0, Mo = 7$), not very alive ($\bar{x} = 2.9, Mo = 1$), highly safe ($\bar{x} = 5.4, Mo = 6$), moderately to highly trustworthy ($\bar{x} = 5.3, Mo = 4, 5$), and moderately human-like ($\bar{x} = 4.2, Mo = 5$). They also reported feeling Alexa was much smarter than themselves ($\bar{x} = 6.1, Mo = 7$), and feeling not particularly close to Alexa ($\bar{x} = 3.0, Mo = 2$). The results were similar after the intervention (other than the changes in intelligence and closeness described above).

4.3 Correlations between perceptions of Alexa

We found strong ($r \geq 0.5$ [9]) correlations between student reports of Alexa’s safeness and trustworthiness on both the pre- and post-test, as well as between Alexa’s friendliness and trustworthiness on the post-test. There was also a strong correlation between reports of trustworthiness on the pre-test and safeness on the post-test. Reports of friendliness and trustworthiness on the pre-test and between the pre- and post-tests were moderately ($r \geq 0.3$ [9]) correlated.

Other moderate correlations included student reports of Alexa’s intelligence and trustworthiness, friendliness and safeness, trustworthiness and feelings of closeness, human-likeness and aliveness, human-likeness and feelings of closeness, as well as aliveness and feelings of closeness. In the post-test, student reports of Alexa’s intelligence and feeling Alexa was smarter than themselves, as well as Alexa’s trustworthiness and feeling Alexa was smarter than themselves were moderately correlated. Additionally, there was a moderate correlation between students with more experience programming prior to the intervention and reports of Alexa’s human-likeness on both the pre- and post-test. Our full correlation analysis is shown in Figure 3.

5 DISCUSSION

5.1 Perceptions of Alexa’s persona

When considering our research question, **How does building Alexa skills and learning about conversational AI in a remote workshop affect students’ perceptions of Alexa?**, prior to the study, we hypothesized students would feel Alexa was *less* intelligent after learning how to program it, as they would better understand how it works; however, we found evidence that students felt Alexa was *more* intelligent after the intervention ($|Z| = 2.78, p = 0.003$). This could have been for multiple reasons. Perhaps by successfully learning fundamental AI literacy concepts [56], students realized Alexa was more complex than they initially thought and thus perceived it to be more “intelligent” (as in the Dunning-Kruger effect [17]), or perhaps their perceptions were influenced by the level of abstraction block-based coding provided. Students also generally felt Alexa was smarter than themselves (before and after the intervention). This is consistent with previous studies of students aged 3-10 [14, 33].

We also hypothesized that students would personify Alexa less after understanding the logic behind how it works, and therefore rate its “aliveness”, “human-likeness”, “friendliness”, and their feelings of closeness to it as less than prior to the intervention. However, there was no significant evidence for any change, except that they felt closer to Alexa ($|Z| = 2.75, p = 0.003$) after the intervention. Students’ increased feelings of closeness could be due to “boundary dissolution”, which is a type of closeness where two agents (usually human) no longer function completely autonomously, but rather function dependently [25]. In this case, an apparent “boundary dissolution” due to Alexa initially seeming to function independently, but seeming to function dependently on students’ programming efforts after the intervention, could have caused students’ increased feelings of closeness.

Alternatively, perhaps having programming experience fundamentally increases feelings of closeness to Alexa, seeing as we

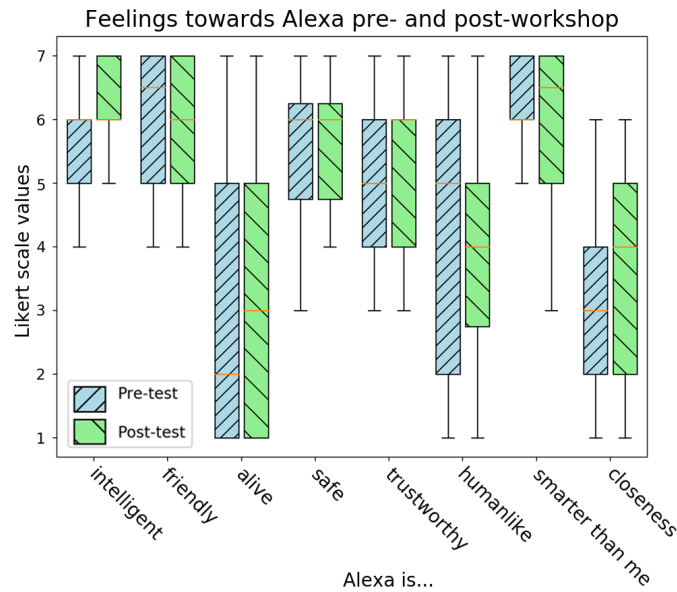


Figure 2: Students’ perceptions of Alexa prior to the workshops (in blue) and after (in green).

found evidence for moderate correlations between students’ *prior* programming experience and closeness. Furthermore, we found evidence for moderate correlations between prior programming experience and human-likeness, as well as closeness and human-likeness. One explanation could be that as students learned to program, they felt Alexa had human-like, logical reasoning, and thus felt closer to it (because of its human-like traits).

We found evidence for strong correlations between students’ perceptions of Alexa’s friendliness and trustworthiness, as well as trustworthiness and safeness, and to a lesser extent, intelligence and trustworthiness, friendliness and safeness, and closeness and trustworthiness. Although these correlations do not necessitate causation, it is important to consider the implications of potential causation when designing CAs. For instance, if a CA was purposefully designed to seem friendly and intelligent, users may associate this with trustworthiness and safeness, despite the potential for the CA to provide incorrect information (intentionally or not). Nevertheless, this could also provide positive opportunities, including how students may learn better if they feel a pedagogical agent is friendly and intelligent, and thus also trustworthy and safe. This is discussed in more depth below.

5.2 Design Considerations

We present design considerations to engage students in learning experiences with CAs, addressing the research question, **What CA design insights can we gain from analyzing student perceptions of CAs in remote workshops?**

5.2.1 Personification. As shown in Table 1, students asked Alexa many personal questions (e.g., “Alexa, do you like Siri?” and “What’s your favorite color?”), which would typically be asked of humans rather than computer systems. Alexa’s often humorous responses

(e.g., “I like ultraviolet. It glows with everything”) could have contributed to students’ perception of personified traits, like friendliness, intelligence and trustworthiness, which were all rated highly. As discussed, personified traits in CAs could play a role in effective teaching interventions [45], especially since feelings of closeness and trust can enhance human teaching and learning experiences [2, 6, 60].

We recommend pedagogical CA developers **cautiously consider personification in their designs**. Although personification could engage students in effective learning experiences, it could also increase their feelings of trust disproportionately with the actual trustworthiness of the device. For example, students could perceive the device as *always* providing unbiased, correct answers, despite AI systems often being biased [43]. Thus, we further recommend considering transparency in CA design.

5.2.2 Transparency. Students also seemed to test the limits of Alexa, asking impossible or difficult questions as encapsulated by the *Other* category in Table 1. For example, students asked Alexa to turn itself off, to tell them all the (infinite) digits of π , and to provide the answer to $\frac{0}{0}$. These behaviors could be linked to trying to understand the system’s inner workings. Thus, we recommend **developing CAs with the ability to explain themselves, and furthermore, provide transparency in terms of their abilities** (e.g., being able to explain AI bias). This is especially important when considering the correlations between CAs’ friendliness and perceived trustworthiness, and the potential for the Dunning-Kruger effect, as discussed above. This recommendation also aligns with other child-CA interaction research, which suggests designing transparent AI systems with respect to children’s level of understanding [58].

5.2.3 Playfulness. Similar to the behavior of “testing” Alexa described above, students asked Alexa playful questions like, “How

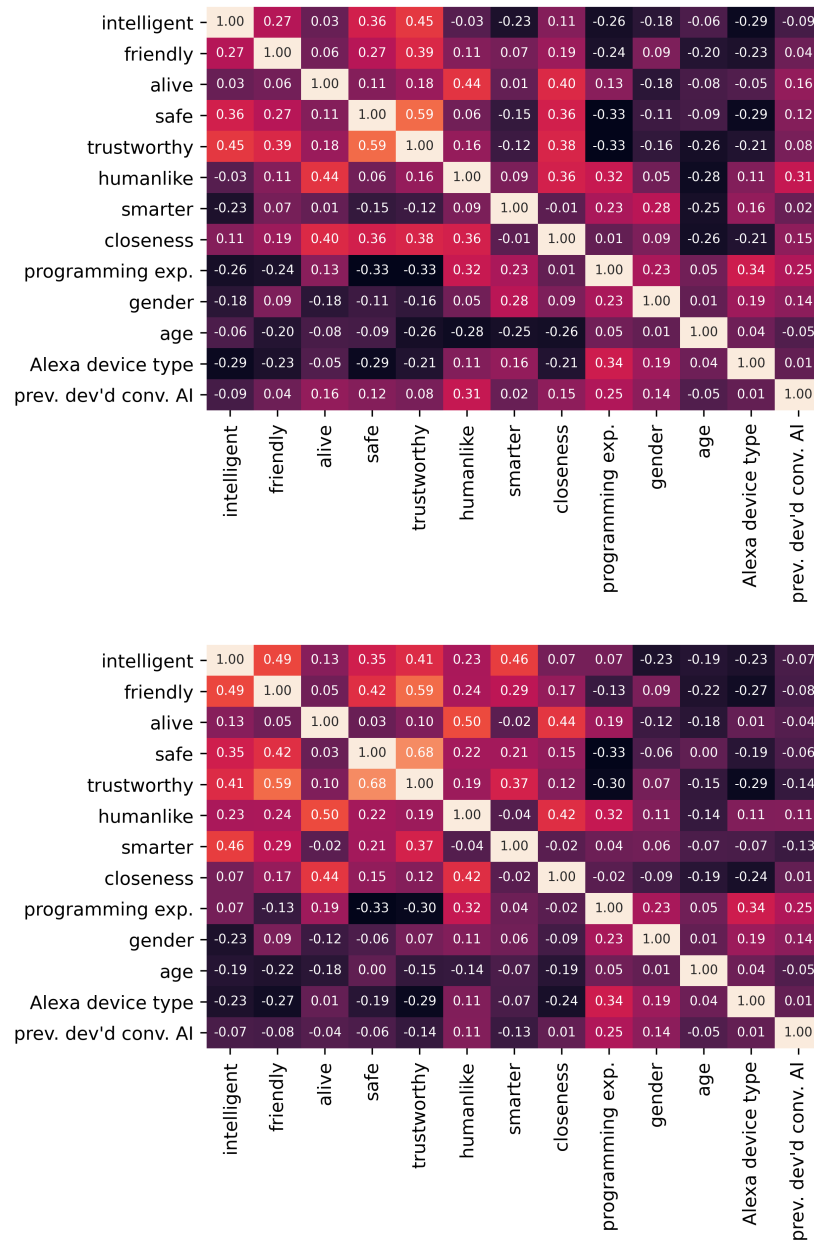


Figure 3: Correlation matrices before (top) and after (bottom) the intervention. Lighter colors correspond to higher coefficients.

much wood would a wood chuck chuck if a wood chuck would chuck wood?” and “Are dragons real?”. These questions illustrate students’—even middle and high school students’—innate desire to play. Play can be hugely beneficial in learning environments, especially from a constructionist perspective [37, 41]; thus, we recommend **considering playful learning experiences** when developing CAs. For example, in our study students had the opportunity to develop their own CA projects. Students came up with many different playful (as well as serious) ideas [56]. One very

playful idea included a CA “Meme Maker” (see Figure 1), which according to the developer, “help[ed] everyone get a quick laugh because as the old saying goes laughter is the best medicine”. This same student cited their favorite part of the workshop as “improving [their] coding ability and learning more about [CAs]”.

5.2.4 *Utility.* Many student projects’ purposes were to provide utility, with 34% being mental and physical health-, 29% being educational-, 21% being productivity- and 8% being accessibility-related CAs [56]. Utility was also reflected in students’ interactions

with Alexa, as *Information updates* and *Action commands* were the most common interactions reported and tagged. With students evidently being interested in CAs' utility, we recommend **designing CAs with useful features** to provide entry points to CA engagement and potential learning moments. For example, students might naturally engage with a CA in figuring out what the weather is like tomorrow, which would provide an opportunity to teach students about APIs and databases, and *how* CAs provide such answers.

6 LIMITATIONS AND FUTURE WORK

In this study, we engaged middle and high school students in remote workshops in which they used MIT App Inventor to program Amazon Alexa; however, the results may not generalize to other environments or grade bands. Other potential threats to validity include noise due to repeated exposure to questions and how we held workshops on two different weeks with slight differences. Thus, future work may include larger follow-up studies with students in different environments, a control group to account for noise, and additional qualitative questions to further investigate students' perceptions. Furthermore, it would be interesting to provide students with additional time to interact with Alexa after the workshops, and observe how interactions differ from the start of the workshops.

7 CONCLUSIONS

Through the programming and learning intervention, we found evidence for students' perceptions of Alexa's intelligence and closeness to Alexa changing. Based on these results, we presented four design recommendations, including considering personification, transparency, playfulness and utility when designing CAs for engaging students in learning experiences. This study contributes to AI literacy research aiming to develop students' understanding of AI to be more accurate and healthy, and CA research aiming to develop more useful, effective interactions.

8 SELECTION AND PARTICIPATION OF CHILDREN

Children aged 11-18 (Mean=14.78, SD=1.91) were selected by their teachers to participate in the middle/high school workshops. Teachers were selected from those that responded to an Amazon Future Engineers call to Title I schools and signed a consent form. Selected students of the age of 18 were given similar student consent forms to sign, and those under the age of 18 were given assent forms and consent forms to be signed by their legal guardians before participating. The university's IRB approved the study protocol and consent/assent forms, which communicated how the data would be aggregated and anonymized. Given the wide age range, teachers assigned some of their older students to be informal mentors to younger students in case they fell behind (e.g., students would ask each other for help in the Zoom chat or after class if they were struggling).

ACKNOWLEDGMENTS

We thank the teachers and students, volunteer facilitators, MIT App Inventor team, Personal Robots Group, and Amazon Future Engineer (AFE) members who made the workshops possible. Special

thanks to Hal Abelson and Hilah Barbot. This work was funded by the AFE program and Hong Kong Jockey Club Charities Trust.

REFERENCES

- [1] Rachel F. Adler, Francisco Iacobelli, and Yehuda Gutstein. 2016. Are you convinced? A Wizard of Oz study to test emotional vs. rational persuasion strategies in dialogues. *Computers in Human Behavior* 57 (2016), 75 – 81. <https://doi.org/10.1016/j.chb.2015.12.011>
- [2] Michal Al-Yagon and Mario Mikulincer. 2004. Socioemotional and Academic Adjustment Among Children with Learning Disorders: The Mediation Role of Attachment-Based Factors. *The Journal of Special Education* 38, 2 (2004), 111–123. <https://doi.org/10.1177/00224669040380020501> arXiv:<https://doi.org/10.1177/00224669040380020501>
- [3] Safinah Ali, Blakeley H Payne, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. In *International Workshop on Education in Artificial Intelligence K-12 (EDUAI'19)*.
- [4] Amy L. Baylor and Jeeheon Ryu. 2003. The Effects of Image and Animation in Enhancing Pedagogical Agent Persona. *Journal of Educational Computing Research* 28, 4 (2003), 373–394. <https://doi.org/10.2190/V0WQ-NWGN-JB54-FAT4> arXiv:<https://doi.org/10.2190/V0WQ-NWGN-JB54-FAT4>
- [5] Robbert-Jan Beun, Eveliene de Vos, and Cilia Witteman. 2003. Embodied Conversational Agents: Effects on Memory Performance and Anthropomorphisation. In *Intelligent Virtual Agents*, Thomas Rist, Ruth S. Aylett, Daniel Ballin, and Jeff Rickel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 315–319.
- [6] Sondra H Birch and Gary W Ladd. 1997. The teacher-child relationship and children's early school adjustment. *Journal of school psychology* 35, 1 (1997), 61–79.
- [7] Caroline Bologna. 2021. 40 Tweets About Parenting With Today's Technology. https://www.huffingtonpost.ca/entry/parenting-alexa-siri-tweets_1_6008e784c5b62c0057c35100. Accessed: 2021-01-27.
- [8] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. Thematic Analysis. In *Handbook of research methods in health social sciences*, Pranee Liamputtong (Ed.). Springer, Singapore.
- [9] J. Cohen. 2013. *Statistical Power Analysis for the Behavioral Sciences*. Elsevier Science. <https://books.google.ca/books?id=rEe0BQAAQBAJ>
- [10] Kevin Corti and Alex Gillespie. 2016. Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior* 58 (2016), 431 – 442. <https://doi.org/10.1016/j.chb.2015.12.039>
- [11] William Damon. 2004. What is Positive Youth Development? *The ANNALS of the American Academy of Political and Social Science* 591, 1 (2004), 13–24. <https://doi.org/10.1177/0002716203260092> arXiv:<https://doi.org/10.1177/0002716203260092>
- [12] Serkan Dincer and Ahmet Doganay. 2017. The effects of multiple-pedagogical agents on learners' academic success, motivation, and cognitive load. *Computers & Education* 111 (2017), 74 – 100. <https://doi.org/10.1016/j.compedu.2017.04.005>
- [13] Daniella DiPaola, Blakeley H. Payne, and Cynthia Breazeal. 2020. Decoding Design Agendas: An Ethical Design Activity for Middle School Students. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) (IDC '20). Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/3392063.3394396>
- [14] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 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* (Stanford, California, USA) (IDC '17). Association for Computing Machinery, New York, NY, USA, 595–600. <https://doi.org/10.1145/3078072.3084330>
- [15] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How Smart Are the Smart Toys? Children and Parents' Agent Interaction and Intelligence Attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (Trondheim, Norway) (IDC '18). Association for Computing Machinery, New York, NY, USA, 231–240. <https://doi.org/10.1145/3202185.3202741>
- [16] Reinders Duit. 2009. Bibliography: Students' and teachers' conceptions and science education. *Diakses pada tanggal* 26 (2009).
- [17] David Dunning. 2011. Chapter five - The Dunning-Kruger Effect: On Being Ignorant of One's Own Ignorance. *Advances in Experimental Social Psychology*, Vol. 44. Academic Press, 247 – 296. <https://doi.org/10.1016/B978-0-12-385522-0.00005-6>
- [18] Simon Gächter, Chris Starmer, and Fabio Tufano. 2015. Measuring the closeness of relationships: a comprehensive evaluation of the 'Inclusion of the Other in the Self' scale. *PLoS one* 10, 6 (2015), e0129478.
- [19] Andres Guadamuz. 2017. Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works. *Intellectual property quarterly* 2 (2017), 20.
- [20] Qian Hu, Yaobin Lu, Zhao Pan, Yeming Gong, and Zhiling Yang. 2021. Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*

- 56 (2021), 102250. <https://doi.org/10.1016/j.ijinfomgt.2020.102250>
- [21] Christopher Brett Jaeger, Alicia M Hymel, Daniel T Levin, Gautam Biswas, Natalie Paul, and John Kinnebrew. 2019. The interrelationship between concepts about agency and students' use of teachable-agent learning technology. *Cognitive research: principles and implications* 4, 1 (2019), 1–20.
- [22] Christopher Brett Jaeger and Daniel Levin. 2016. If Asimo thinks, does Roomba feel? The legal implications of attributing agency to technology. *Journal of Human-Robot Interaction (Symposium on Robotics Law and Policy)* 5, 3 (Dec. 2016), 23. <https://ssrn.com/abstract=3097129>
- [23] KM Kahn, R Megasari, E Piantari, and E Junaeti. 2018. AI programming by children using Snap! block programming in a developing country. 11082.
- [24] Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Nathan G Freier, Rachel L Severson, Brian T Gill, Jolina H Ruckert, and Solace Shen. 2012. "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology* 48, 2 (2012), 303.
- [25] Thomas Kreilkamp. 1984. Psychological Closeness: EXAMPLES OF CLOSENESS CONCEPTUALIZATION REFERENCES. *The American Behavioral Scientist (pre-1986)* 27, 6 (Jul 1984), 771. <https://www-proquest-com.libproxy.mit.edu/scholarly-journals/psychological-closeness/docview/194671679/se-2?accountid=12492> Copyright - Copyright SAGE PUBLICATIONS, INC. Jul/Aug 1984; Last updated - 2019-11-23; CODEN - ABHSAU; SubjectsTermNotLitGenreText - New York.
- [26] Asi Kuperman and David Mioduser. 2012. [Chais] Kindergarten Children's Perceptions of "Anthropomorphic Artifacts" with Adaptive Behavior. *Interdisciplinary Journal of E-Learning and Learning Objects* 8, 1 (January 2012), 137–147. <https://www.learnedtechlib.org/pl/44766>
- [27] David R. Large and Gary Burnett. 2019. Life on the Road: Exposing Drivers' Tendency to Anthropomorphise In-Vehicle Technology. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Sebastiano Bagnara, Riccardo Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita (Eds.). Springer International Publishing, Cham, 3–12.
- [28] Krittaya Leelawong and Gautam Biswas. 2008. Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education* 18, 3 (2008), 181–208.
- [29] D. T. Levin, J. A. Adams, M. M. Saylor, and G. Biswas. 2013. A transition model for cognitions about agency. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 373–380. <https://doi.org/10.1109/HRI.2013.6483612>
- [30] Sharona T Levy and David Mioduser. 2008. Does it "want" or "was it programmed to...?" Kindergarten children's explanations of an autonomous robot's adaptive functioning. *International Journal of Technology and Design Education* 18, 4 (2008), 337–359.
- [31] Phoebe Lin, Jessica Van Brummelen, Galit Lukin, Randi Williams, and Cynthia Breazeal. 2020. Zhorai: Designing a Conversational Agent for Children to Explore Machine Learning Concepts. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 09 (Apr. 2020), 13381–13388. <https://doi.org/10.1609/aaai.v34i09.7061>
- [32] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3313831.3376727>
- [33] Silvia B. Lovato, Anne Marie Piper, and Ellen A. Wartella. 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 (Boise, ID, USA) (IDC '19)*. Association for Computing Machinery, New York, NY, USA, 301–313. <https://doi.org/10.1145/3311927.3323150>
- [34] Alexander Meschtscherjakov, Manfred Tscheligi, Bastian Pflöging, Shadan Sadeghian Borojeni, Wendy Ju, Philippe Palanque, Andreas Riener, Bilge Mutlu, and Andrew L. Kun. 2018. Interacting with Autonomous Vehicles: Learning from Other Domains. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal QC, Canada) (CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3170427.3170614>
- [35] Joseph E. Michaelis and Bilge Mutlu. 2019. Supporting Interest in Science Learning with a Social Robot. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children (Boise, ID, USA) (IDC '19)*. Association for Computing Machinery, New York, NY, USA, 71–82. <https://doi.org/10.1145/3311927.3323154>
- [36] Elizabeth Katalina Morales-Urrutia, Jose Miguel Ocaña, and Diana Pérez-Marín. 2020. How to Integrate Emotions in Dialogues With Pedagogic Conversational Agents to Teach Programming to Children. *Innovative Perspectives on Interactive Communication Systems and Technologies* (2020), 66.
- [37] Seymour Papert and Idit Harel. 1991. Situating constructionism. *Constructionism* 36, 2 (1991), 1–11.
- [38] Diana Pérez-Marín and Ismael Pascual-Nieto. 2013. An exploratory study on how children interact with pedagogic conversational agents. *Behaviour & Information Technology* 32, 9 (2013), 955–964. <https://doi.org/10.1080/0144929X.2012.687774>
- arXiv:<https://doi.org/10.1080/0144929X.2012.687774>
- [39] Yim Register and Amy J. Ko. 2020. Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics. In *Proceedings of the 2020 ACM Conference on International Computing Education Research (Virtual Event, New Zealand) (ICER '20)*. Association for Computing Machinery, New York, NY, USA, 67–78. <https://doi.org/10.1145/3372782.3406252>
- [40] Research and Markets. 2019. Global Conversational AI Market Forecast to 2024: Integration of Advanced AI Capabilities Adding Value to the Conversational AI Offering. <https://www.researchandmarkets.com/>.
- [41] Louis Rice. 2009. Playful Learning. *Journal for Education in the Built Environment* 4, 2 (2009), 94–108. <https://doi.org/10.11120/jebe.2009.04020094> arXiv:<https://doi.org/10.11120/jebe.2009.04020094>
- [42] Juan David Rodríguez-García, Jesús Moreno-León, Marcos Román-González, and Gregorio Robles. 2021. Evaluation of an Online Intervention to Teach Artificial Intelligence With LearningML to 10-16-Year-Old Students. (2021).
- [43] Drew Roselli, Jeanna Matthews, and Nisha Talagala. 2019. Managing Bias in AI. In *Companion Proceedings of The 2019 World Wide Web Conference (San Francisco, USA) (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 539–544. <https://doi.org/10.1145/3308560.3317590>
- [44] Michael Scaife and Mike van Duuren. 1995. Do computers have brains? What children believe about intelligent artifacts. *British Journal of Developmental Psychology* 13, 4 (1995), 367–377.
- [45] Sofia Schöbel, Andreas Janson, and Abhay Mishra. 2019. A Configurational View on Avatar Design—The Role of Emotional Attachment, Satisfaction, and Cognitive Load in Digital Learning. In *Fortieth International Conference on Information Systems, Munich*.
- [46] Ryan M Schuetzler, G Mark Grimes, Justin Scott Giboney, and Jay F Nunamaker Jr. 2018. The influence of conversational agents on socially desirable responding. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. 283.
- [47] Daniel B. Shank. 2014. Impressions of computer and human agents after interaction: Computer identity weakens power but not goodness impressions. *International Journal of Human-Computer Studies* 72, 10 (2014), 747 – 756. <https://doi.org/10.1016/j.ijhcs.2014.05.002>
- [48] Bruce Sherin. 2013. A Computational Study of Commonsense Science: An Exploration in the Automated Analysis of Clinical Interview Data. *Journal of the Learning Sciences* 22, 4 (2013), 600–638. <https://doi.org/10.1080/10508406.2013.836654> arXiv:<https://doi.org/10.1080/10508406.2013.836654>
- [49] Rebecca R Skinner. 2019. The Elementary and Secondary Education Act (ESEA), as Amended by the Every Student Succeeds Act (ESSA): A Primer. CRS Report R45977, Version 2. *Congressional Research Service* (2019).
- [50] Hannah Sparks. 2019. Mom busts 9-year-old son using Alexa to cheat on homework. <https://nypost.com/2019/10/15/mom-busts-9-year-old-son-using-alexa-to-cheat-on-homework/>. Accessed: 2021-01-27.
- [51] Karen Spektor-Prezel and David Mioduser. 2015. The Influence of Constructing Robot's Behavior on the Development of Theory of Mind (ToM) and Theory of Artificial Mind (ToAM) in Young Children. In *Proceedings of the 14th International Conference on Interaction Design and Children (Boston, Massachusetts) (IDC '15)*. Association for Computing Machinery, New York, NY, USA, 311–314. <https://doi.org/10.1145/2771839.2771904>
- [52] Micol Spitale, Silvia Silleresi, Giulia Cosentino, Francesca Panzeri, and Franca Garzotto. 2020. "Whom Would You like to Talk with?": Exploring Conversational Agents for Children's Linguistic Assessment. In *Proceedings of the Interaction Design and Children Conference (London, United Kingdom) (IDC '20)*. Association for Computing Machinery, New York, NY, USA, 262–272. <https://doi.org/10.1145/3392063.3394421>
- [53] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What Should Every Child Know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (Jul. 2019), 9795–9799. <https://doi.org/10.1609/aaai.v33i01.33019795>
- [54] Jessica Van Brummelen. 2019. *Tools to Create and Democratize Conversational Artificial Intelligence*. Master's thesis. Massachusetts Institute of Technology, Cambridge, MA.
- [55] Jessica Van Brummelen, Tommy Heng, and Viktoriya Tabunshchik. 2020. <https://gist.github.com/jessvb/1cd959e32415a6ad4389761c49b54bbf>. Accessed: 2020-09-09.
- [56] Jessica Van Brummelen, Tommy Heng, and Viktoriya Tabunshchik. 2021. Teaching Tech to Talk: K-12 Conversational Artificial Intelligence Literacy Curriculum and Development Tools. In *2021 AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI)*. AAAI, Online.
- [57] Xiaoyu Wan, Xiaofei Zhou, Zaiqiao Ye, Chase K. Mortensen, and Zhen Bai. 2020. SmileyCluster: Supporting Accessible Machine Learning in K-12 Scientific Discovery. In *Proceedings of the Interaction Design and Children Conference (London, United Kingdom) (IDC '20)*. Association for Computing Machinery, New York, NY, USA, 23–35. <https://doi.org/10.1145/3392063.3394440>
- [58] Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. A is for Artificial Intelligence: The Impact of Artificial Intelligence Activities on Young Children's Perceptions of Robots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19)*. Association

- for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3290605.3300677>
- [59] David Wolber, Harold Abelson, and Mark Friedman. 2015. Democratizing computing with App Inventor. *GetMobile: Mobile Computing and Communications* 18, 4 (2015), 53–58.
- [60] Ilka Wolter, Michael Glüer, and Bettina Hannover. 2014. Gender-typicality of activity offerings and child–teacher relationship closeness in German “Kindergarten”. Influences on the development of spelling competence as an indicator of early basic literacy in boys and girls. *Learning and Individual Differences* 31 (2014), 59 – 65. <https://doi.org/10.1016/j.lindif.2013.12.008>
- [61] Michael Wooldridge. 2021. Artificial Intelligence Is a House Divided: A decades-old rivalry has riven the field. It’s time to move on. *The Chronicle* (20 Jan. 2021).
- [62] Ying Xu and Mark Warschauer. 2020. What Are You Talking To?: Understanding Children’s Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376416>
- [63] Abigail Zimmermann-Niefield, Shawn Polson, Celeste Moreno, and R. Benjamin Shapiro. 2020. Youth Making Machine Learning Models for Gesture-Controlled Interactive Media. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) (IDC '20). Association for Computing Machinery, New York, NY, USA, 63–74. <https://doi.org/10.1145/3392063.3394438>