Learning Affects Trust: Design Recommendations and Concepts for Teaching Children—and Nearly Anyone—about Conversational Agents

Jessica Van Brummelen[1], **Mingyan Claire Tian**[2], **Maura Kelleher**[1], **Nghi Hoang Nguyen**[1]

[1] Massachusetts Institute of Technology
[2] Wellesley College
jess@csail.mit.edu, mt1@wellesley.edu, maurakel@mit.edu, nghin@mit.edu

Abstract

Conversational agents are rapidly becoming commonplace. However, since these systems are typically blackboxed, users—including vulnerable populations, like children—often do not understand them deeply. For example, they might assume agents are overly intelligent, leading to frustration and distrust. Users may also overtrust agents, and thus overshare personal information or rely heavily on agents' advice. Despite this, little research investigates users' perceptions of conversational agents in-depth, and even less investigates how education might change these perceptions to be more healthy. We present workshops with associated educational conversational AI concepts to encourage healthier understanding of agents. Through studies with the curriculum with children and parents from various countries, we found participants' perceptions of agents—specifically their partner models and trust—changed. When participants discussed changes in trust of agents, we found they most often mentioned learning something. For example, they frequently mentioned learning where agents obtained information, what agents do with this information and how agents are programmed. Based on the results, we developed recommendations for teaching conversational agent concepts, including emphasizing the concepts students found most challenging, like training, turn-taking and terminology; supplementing agent development activities with related learning activities; fostering appropriate levels of trust towards agents; and fostering accurate partner models of agents. Through such pedagogy, students can learn to better understand conversational AI and what it means to have it in the world.

Introduction and Related Work

Reports have indicated an exponential rise in the use of voice-based agents, like Alexa and Siri (Smith 2018; Zierau et al. 2022). Researchers have taken note and begun to investigate the societal implications of agent ubiquity, like how this may affect social norms, decision-making and information spread (Seymour and Van Kleek 2021; Gaube et al. 2021). For instance, researchers have found relationships with agents form similarly to human-human relationships (Seymour and Van Kleek 2021; Straten et al. 2020). One concern researchers have is how humans may overtrust agents, especially if they have personified traits (Zhou et al. 2019; Van Brummelen, Tabunshchyk, and Heng 2021). Considering how trust is connected to relationship building, and is a key factor in misinformation spread (Xiao, Borah, and Su 2021; Seymour and Van Kleek 2021), it is important for people to be able to calibrate their levels of trust towards conversational agents (CAs), according to agents' actual trustworthiness. In this paper, we investigate the link between a conversational AI educational intervention and participants' perceptions and levels of trust of CAs, finding that participants most often mentioned learning something about CAs as reasons for changes in their trust. Furthermore, we investigate other aspects of students' learning and perceptions, including their self-efficacy and CA partner models, to determine how to best teach CA concepts.

Along similar lines, many researchers have started AI education initiatives. For instance, the AI4K12 initiative is developing tools and curriculum based on five core "Big AI Ideas"; MIT's RAISE initiative is developing vocational and K-12 tools for AI education; and Code.org is developing interactive resources for K-12 AI education (Touretzky et al. 2019; MIT 2021; Code.org 2022). Nonetheless, very few resources specifically teach about conversational AI or investigate how to best do so, despite researchers noting the importance of teaching a breadth of types of AI, a need for CA pedagogical artifacts, and the unique societal questions and challenges CAs present due to their relational nature (Long and Magerko 2020; Murad and Munteanu 2020; Seymour and Van Kleek 2021). There are especially few resources in the K-12 space, despite children potentially being more vulnerable to misinformation spread (Murad and Munteanu 2020; Van Brummelen, Heng, and Tabunshchyk 2021).

One notable K-12 resource for CA education includes DiPaola (2021)'s social robotics curriculum. In this curriculum, students aged 9-12 learn about the societal impact of social robots and how to prototype robot conversation. It focuses on social robotics and includes a portion on conversational AI, in which students learn concepts including conversational flow representation and machine learning (ML). Another resource includes Zhu and Van Brummelen (2021)'s CA curriculum, in which students aged 13-15 develop CAs through speaking with Convo, and learn concepts like training ML models and the difference between constrained and unconstrained natural language (NL). A third resource includes Van Brummelen, Heng, and Tabunshchyk (2021)'s

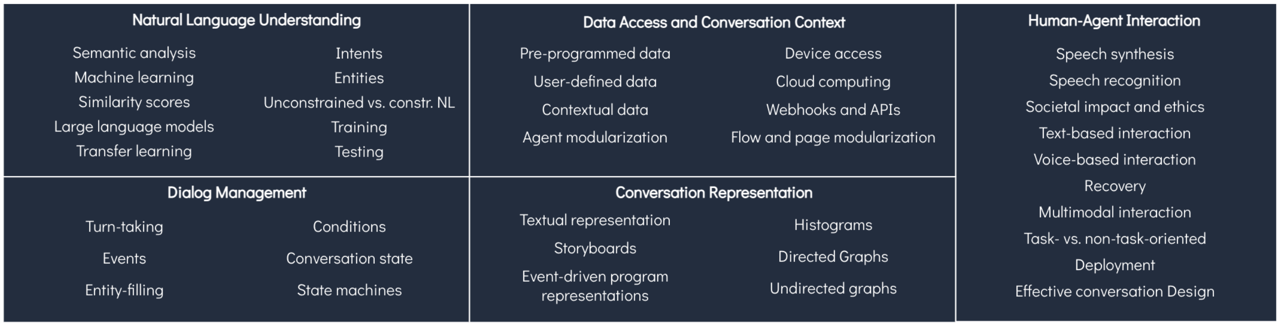


Figure 1: The framework of forty CA concepts, which are described in-depth in the Appendix (Van Brummelen 2022b).

conversational AI curriculum. It teaches students aged 13-18 how to create CAs using a block-based coding interface, and related concepts like intent-modeling and entity-filling. Since this curriculum specifically focuses on conversational AI, has a broad intended age range and a low-barrier-to-entry, open-source interface for developing CAs, we build on it to investigate our research questions. We refer to the interface as "ConvoBlocks".

Study Novelty

With this interface, Van Brummelen, Tabunshchyk, and Heng (2021) investigated changes in students' perceptions of CAs through agent-building workshops. They found correlations between perceptions of Alexa's friendliness, safeness, and trustworthiness. Trustworthiness as a concept, however, is very broad, and no one (to our knowledge) has investigated how learning to program CAs affects specific aspects of trust. Understanding this, however, could help educators better develop pedagogy to empower students to understand agents' true trustworthiness. In our study, we adopt the widely-used model of trust by McKnight and Chervany (2001). This model consists of four characteristics of trust, (1) **competence**, (2) **benevolence**, (3) **integrity** and (4) **predictability**. We also investigate people's trust of agents' correctness, as this relates directly to misinformation spread.

Another construct that could help improve agent pedagogy includes "partner models", which define how users perceive their conversational partners, or in this case, CAs. Doyle, Clark, and Cowan (2021)'s model involves three dimensions: (1) **competence and dependability**, (2) **human-likeness**, and (3) **cognitive flexibility**. By understanding users' partner models of CAs, we can better understand their expectations and reactions. For instance, if a user expects an agent to be flexible, but it is not (e.g., it only understands very specific commands), the user may become frustrated. However, by understanding users' partner models, CA designers can develop agents to foster accurate partner models (e.g., an agent which outlines the extent of its flexibility) (Cowan et al. 2017). In a similar way, by understanding students' partner models, educators can develop pedagogy to empower students to develop more accurate perceptions of agents, allowing them to better understand how they work.

There is also little literature investigating the perceptions of people from countries that are not Western, Educated, Industrialized, Rich and Democratic (WEIRD) (Sturm et al. 2015), and no literature investigating the difference between how child and parent perceptions of CAs change after programming them. Thus, we incorporate these groups in our study and aim towards developing teaching recommendations for nearly anyone to learn about CAs (although we realize there is still much work to be done in this area, and look forward to other researchers continuing in this vein).

Other Educational Interventions Affecting Trust

With other technologies, researchers have shown how educational interventions can affect trust, increase understanding and decrease the spread of misinformation (Craft, Ashley, and Maksl 2017; Vraga, Bode, and Tully 2022; Seo, Xiong, and Lee 2019; Straten et al. 2020). In terms of investigating changes in children's trust of technology, Di-Paola (2021) found children trusted robots less after engaging in societal impact curriculum. Others found children's trust decreased after learning about the programmatic nature of robots (Straten et al. 2020). To our knowledge, the only educational intervention in which researchers have investigated changes in children's trust of CAs is the ConvoBlocks study discussed above. Van Brummelen, Tabunshchyk, and Heng (2021) did not find any significant differences in students' perceptions of agents' trustworthiness through the workshops; however, they did find correlations between perceptions of trustworthiness, safeness and friendliness. Our study investigates whether there are changes in trust for specific subsets of participants not studied previously (e.g., children vs. parents, WEIRD vs. non-WEIRD).

Conversational Agent Concepts

As mentioned previously, there is a lack of CA-specific pedagogical materials, despite conversational AI's unique positioning in terms of market penetration and potential to become a primary mode of human-computer interaction (Murad and Munteanu 2020; Statista 2021; Seaborn et al. 2021). To address this need, we developed a framework of forty CA concepts, as presented in the thesis, Van Brummelen (2022a), Figure 1 and the Appendix (Van Brummelen 2022b). We developed our workshop curriculum based on teaching a number of these concepts, as described in the Procedure section.

Research Questions

To address the above literature gaps, we investigated the following research questions through our CA programming and societal impact educational intervention:

RQ1: What do various people (WEIRD, non-WEIRD, and different generations) find difficult in the intervention?

RQ2: How do various people feel in terms of self-efficacy and programming identity through the intervention?

RQ3: How do various people perceive Alexa with respect to partner models and trust through the intervention?

RQ4: How might the results from RQ1-3 inform teaching guidelines for conversational AI?

Through investigating these research questions, we developed design recommendations (DRs) for CA pedagogy. Analysis of our results also led to agent DRs, which are examined in another paper, Van Brummelen et al. (2022).

Participants

There were 99 pairs of children and parents who filled the interest forms to participate in the workshops. In total, 49 completed at least 1 of the 3 surveys. There were 27 children (age avg.=13.96, SD=1.829) and 19 parents (age avg.=46.35, SD=11.07) on the pre-survey. From the same survey, 23 participants were from WEIRD countries (age avg.=26.45, SD=19.24) and 23 were from non-WEIRD countries (age avg.=25.48, SD=15.18). We defined WEIRD countries according to Doğruyol, Alper, and Yilmaz (2019)'s method. Participants were from the US, Singapore, Canada and New Zealand (WEIRD, 57% from the US); and Indonesia, Iran, Japan, and India (non-WEIRD, 87% from Indonesia).

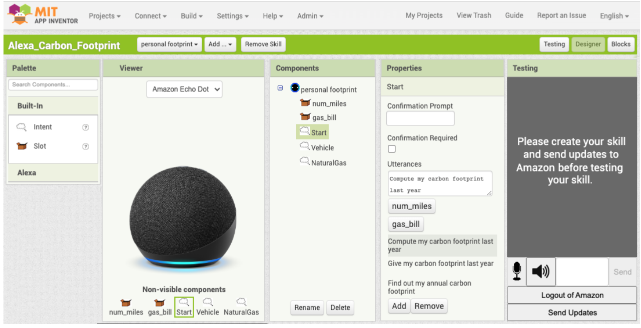
The ConvoBlocks Interface

We utilized ConvoBlocks (aka the MIT App Inventor Conversational AI Interface) to teach participants how to program CAs due to its low barrier to entry, its wide target age range, how it is open-source and how participants can create CAs to run on real smart-home devices, like Amazon Echoes (Van Brummelen, Heng, and Tabunshchyk 2021). To create such agents, participants go to a web page where they can define CA invocation names (e.g., "Carbon Footprint Calculator"), intents (e.g., "What's my carbon footprint?") and entities (e.g., miles driven), as in Figure 2. Next, they go to a page where they can connect code blocks to define how the CA responds to particular intents (e.g., by saying "Your footprint is 11.3 tons per year"), as in Figure 3. Participants can test their agent on the web page itself or any Alexa-enabled devices, like an Alexa App or Amazon Echo.

Procedure

The workshops spanned two days virtually over Zoom for approximately 3.5 hours each day. The first day introduced CA concepts through two agent coding tutorials. Participants created agents to calculate carbon footprints. We gave participants PDFs of these two tutorials, plus a third, to be completed if they finished early. After the tutorials, participants listed traits of their "ideal" CAs in an ideation session.

Prior to the tutorials, participants completed a pre-survey with questions about demographics and their perceptions and trust of CAs. At the end of Day 1, they filled out a mid-survey similar to the pre-survey. It also asked whether and why their responses changed, and which CA concepts were most challenging. On Day 1 we focused on teaching participants the CA concepts, *Training*, *Intents*, *Agent modularization*, *Entities*, *Events*, *Testing*, *Turn-taking* and associated CA terminology. These were mentioned on the mid-survey.



Compute my carbon footprint last year

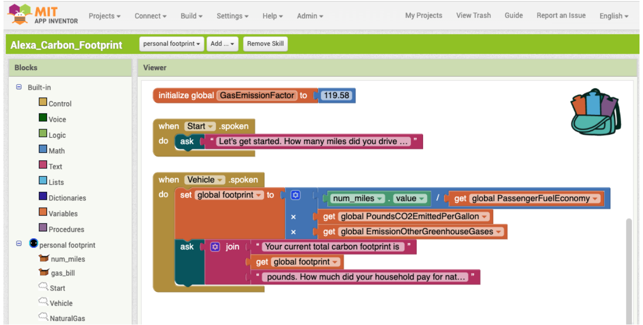
Compute my carbon footprint last year

Give my carbon footprint last year

Find out my annual carbon footprint

Please create your skill and send updates to Amazon before testing your skill.

Figure 2: The ConvoBlocks page in which users can define intents and entities for their agents to recognize.



```
initialize global GasEmissionFactor to 119.58

when Start .spoken
do  ask " Let's get started. How many miles did you drive ... "

when Vehicle .spoken
do  set global footprint to    num_miles . value / get global PassengerFuelEconomy
                             × get global PoundsCO2EmittedPerGallon
                             × get global EmissionOtherGreenhouseGases
    ask  join  " Your current total carbon footprint is "
               get global footprint
               " pounds. How much did your household pay for nat... "
```

Figure 3: The ConvoBlocks page in which users can program agents to respond to intents.

The second day focused on teaching *Societal impact and ethics* through presentations and group activities. Instructors of the workshop presented on current world challenges. Participants discussed sustainability and how technology—including conversational AI—may influence human mindsets. They ultimately created presentations on how technology could be used to combat sustainability problems. A final survey was given after the presentations, which again asked participants to reflect on how their partner models and trust of agents changed during the workshops.

Data Analysis

We analyzed the long-answer responses and ideation data using a coding reliability approach to thematic analysis, as described by Braun et al. (2019). The resulting tags are shown in the Appendix (Van Brummelen 2022b). Krippendorff's Alpha between all three researchers was $\alpha \geq .800$. The tagged data were aggregated by union between researchers, and organized with respect to the following categories: WEIRD, non-WEIRD, child and parent.

To analyze responses to Likert scale questions, we utilized independent and paired t-tests, Mann-Whitney U tests and Wilcoxon signed-rank tests, according to the sample and normality of the data. Figures show statistical significance with star symbols (i.e., *: $p \leq .05$, **: $p \leq .01$ and ***: $p \leq .001$). See Van Brummelen (2022a) for the full surveys.

Results

We describe participants' partner models, trust, difficulties learning, tutorial completion, and self-efficacy/identity here.

Overall Participants

Overall participants' feelings towards CAs shifted towards more of a friend (than a co-worker) after going through the workshops (pre/post: x̄=3.58,3.24; t(32)=2.15; p=.039). When asked in a long-answer question why they trusted or distrusted CAs to provide correct information, they generally provided answers for why they distrusted CAs, both before (72%) and after (64%) the programming activity.

About a quarter (24%) of participants' long-answer responses indicated they felt their trust of CAs changed through the programming activity. All major subsets (children, parents, non-WEIRD, WEIRD) showed an increase in trust through the workshops, although it was only significant for the WEIRD subset (pre/post: x̄=4.00,3.75; W(15)=0; p=.046) and other subsets described later. As shown in Figure 4, participants most often cited the source of the CAs' information (including human data, the internet and other sources) as the reason for their opinion changing. These answers, as well as those from the "Programmed" and "Personal learning" categories generally alluded to how participants had learned something through the workshops.

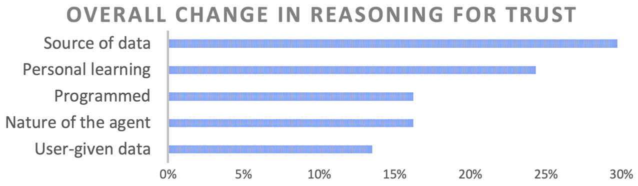


Figure 4: Participants' reasoning for changes in trust of CAs in terms of percent tag frequency. The Appendix (Van Brummelen 2022b) provides descriptions of each tag.

In terms of McKnight and Chervany's trust model, overall participants most often cited predictability and did not cite benevolence when discussing changes in trust (see Table 1).

Participants indicated *Training*, CA terminology and *Turn-taking* as the three most difficult things to learn in the workshops. In Figure 6 and 7 the concepts are ordered from most to least challenging, as chosen by overall participants.

WEIRD vs. Non-WEIRD

Participants from WEIRD countries thought Alexa was less competent after the programming activity than before (pre/mid: x̄=2.43,2.95; t(20)=-2.33; p=.030). This resulted in a significant difference between participants from WEIRD and non-WEIRD countries' feelings about Alexa's competence after the programming activity (x̄=3.00,2.11; U(38)=106.5; p=.004), as shown in Figure 5.

Subset	C[1]	I	P	B
Overall	32%	29%	**39%**	0%
Non-WEIRD	**40%**	20%	**40%**	0%
WEIRD	28%	33%	**39%**	0%
Child	13%	**47%**	40%	0%
Parent	**54%**	8%	38%	0%

[1] C: Competence, I: Integrity, P: Predictability, B: Benevolence

Table 1: Percent of long-answer responses indicating different aspects of trust when participants discussed changes in their trust of CAs through the programming activity.

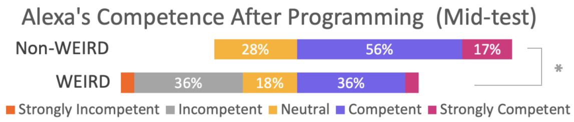


Figure 5: Distribution of responses on a 5-point Likert scale question given after the programming activity from participants from non-WEIRD and WEIRD countries when asked to rate Alexa's competence.

Those from non-WEIRD countries more often cited *Training* and *Events* as difficult concepts than those from WEIRD countries; whereas those from WEIRD countries more often cited *Testing* and *Turn-taking*, and more often described other concepts. Otherwise, the relative frequencies were quite similar, as shown in Figure 6.

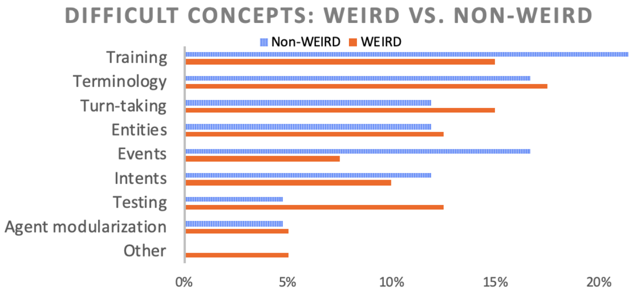


Figure 6: Relative frequency of participants from WEIRD vs. non-WEIRD countries' responses to a question asking which concepts were most difficult to learn.

Parents vs. Children

Children had more prior experience programming than parents (x̄=1.59,0.79; U(44)=144.5; p=.0026). Children from non-WEIRD countries had more prior experience learning about AI (x̄=0.62,0.13; U(19)=27; p=.017) than parents from non-WEIRD countries (although this was not so for those from WEIRD countries). Children from WEIRD countries completed more tutorials than parents from WEIRD countries (x̄=2.14,1.14; U(16)=18.5; p=.0.025).

Children more often cited *Training* and *Testing* as difficult concepts than parents did; whereas parents more often cited CA terminology and *Agent modularization*, and more often described other concepts. Otherwise, the relative frequencies were quite similar, as shown in Figure 7.

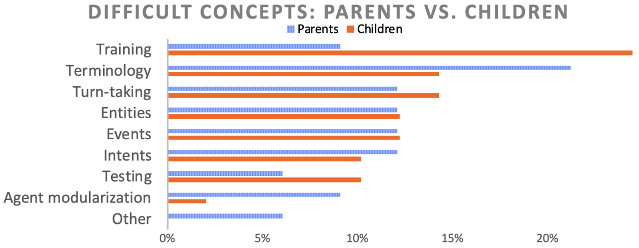


Figure 7: Relative frequency of children and parents' responses to which concepts were most difficult to learn.

Different Levels of Prior Programming Experience

According to the pre-survey, 30% of the participants had no prior programming experience, 13% had only visual/block-based programming experience, and 57% had text-based programming experience. There were no significant differences found in the number of tutorials completed based on participants' prior programming experience.

Prior to the workshops, those who had text-based programming experience thought Alexa was less competent ($\bar{x}$=2.73,2.07; W(16)=0; p=.038) than those who had no programming experience. Prior to ($\bar{x}$=3.54,1.86; U(38)=62; $p=2.50 \times 10^{-4}$), during ($\bar{x}$=3.86,2.25; U(31)=46; p=.0011), and after ($\bar{x}$=3.84,2.50; U(25)=30; p=.0065) the workshops, those who had text-based programming experience saw themselves more as programmers than those who had no experience initially. Prior to the workshops, those who had text-based programming experience also saw themselves more as programmers than those who had only visual programming experience ($\bar{x}$=3.54,2.33; U(30)=37; p=.022). After the workshops, however, there was no significant difference between how participants with text-based vs. visual programming experience felt as programmers.

Different Prior Experience Learning about AI

On the pre-survey, 46% of the participants reported having no prior experience learning about AI, whereas 54% reported they had. There were no significant differences found in the number of tutorials completed depending on whether participants had previously learned about AI or not.

Prior to the workshops, those who had learned about AI previously thought Alexa was more human-like ($\bar{x}$=2.05,2.84; U(44)=-2.87; p=.0063). Participants who had not learned AI before thought Alexa was more dependable after the programming experience (pre/mid: $\bar{x}$=3.47,3.88; W(16)=0; p=.020). Those who had learned about AI previously saw themselves more as programmers than those who had not prior to ($\bar{x}$=2.29,3.36; U(44)=-2.73; p=.0092), during ($\bar{x}$=2.82,3.50; U(37)=129; p=.047) and after ($\bar{x}$=3.83,2.87; U(31)=69; p=.0073) the workshops. They were also more confident they could design and create their own technology project than those who did not have prior AI experience prior to ($\bar{x}$=2.71,3.60; U(44)=-2.51; p=.016), during ($\bar{x}$=3.12,4.00; U(37)=92.5; p=.0026), and after ($\bar{x}$=4.28,3.13; U(31)=44.5; $p=2.7 \times 10^{-4}$) the workshops. They were also more confident they could make an impact in their community or the world using technology than those who did not have prior AI experience prior to ($\bar{x}$=3.29,4.00; U(44)=-2.52; p=.015), during ($\bar{x}$=3.18,4.18; U(37)=86.5; p=.0015), and after ($\bar{x}$=4.22,3.60; U(31)=80; p=.018) the workshops.

Those who had never learned about AI before felt CAs reported correct information more after the programming activity than before (pre/mid: $\bar{x}$=2.88,2.47; W(16)=0; p=.038). No significant difference was found for those who had learned about AI before. (See Figure 8 and 9.)

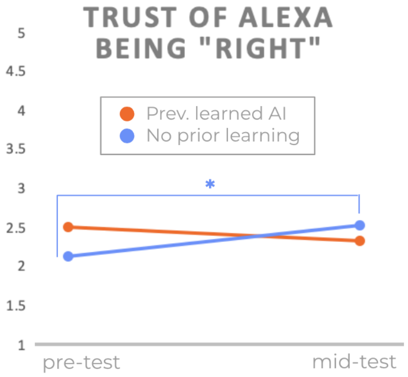


Figure 8: Mean responses to a 5-point Likert scale question about trust of Alexa's correctness given before/after the programming activity from participants with no experience and prior experience learning about AI.

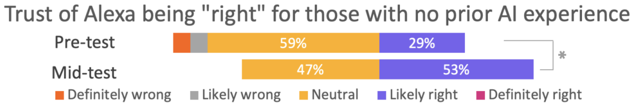


Figure 9: Distribution of responses from a 5-point Likert scale question about trust of Alexa's correctness given before and after the programming activity from participants with no prior experience learning about AI.

Different Experiences with Conversational Agents

According to the pre-survey, 52% of the participants had used more than one type of CA, 35% had used only a single type of CA, and 13% had never used a CA before. 83% reported typically using CAs in their first language and 17% reported typically using them in another language. There were no significant differences found in the number of tutorials completed depending on whether participants had used more than one type of CA or only a single CA, or on whether participants typically used CAs in their first language or not.

Those who used CAs in their first language thought Alexa was more human-like prior to ($\bar{x}$=2.61,1.88; U(44)=87.5;

p=.025), during (x̄=2.66,2.00; U(37)=67; p=.042) and after (x̄=2.82,1.80; U(31)=32; p=.025) the workshops than those who used them in another language. They also thought Alexa was more correct than those who used it in another language, prior to the workshops (x̄=4.03,3.00; U(44)=52; p=5.50×10^{-4}). As shown in Figure 10, prior to the workshops, participants who typically used CAs in their first language thought Alexa was more correct (x̄=4.03,3.00; U(44)=52; p=5.51×10^{-4}) than those who typically used it in another language. There was no significant difference after the programming activity, as shown in Figure 11.

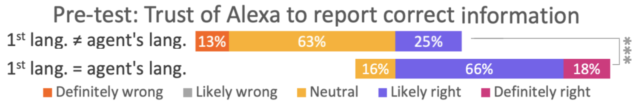


Figure 10: Distribution of responses from a 5-point Likert scale question about trust of Alexa's correctness given before the programming activity from participants who typically used CAs in their first language or not.

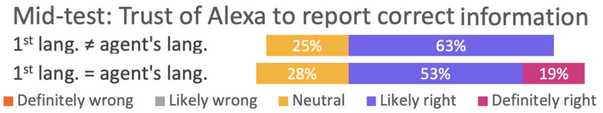


Figure 11: Distribution of responses from a 5-point Likert scale question about trust of Alexa's correctness given after the programming activity from participants who typically used CAs in their first language or not.

Discussion

In this section, we develop DRs for teaching conversational AI by discussing the results with respect to our RQs.

RQ1: Difficulties Learning about CAs

With respect to RQ1, the top three concepts most referenced as difficult in this study were *Training*, CA terminology and *Turn-taking*. Although this trend remained relatively similar for all major subsets, different subsets of the participants still found different concepts more challenging than others. For example, children cited *Training* and *Testing* more often as difficult concepts than parents did; whereas parents cited CA terminology and *Agent modularization* more often than children did. Participants from non-WEIRD countries cited *Training* and *Events* more often as difficult concepts than those from WEIRD countries did; whereas participants from WEIRD countries cited *Testing* and *Turn-taking* more often than those from non-WEIRD countries did. In other studies, students found *Constrained vs. unconstrained natural language*, *Machine learning* and *Societal impact and ethics* particularly challenging to learn (Zhu and Van Brummelen 2021; Van Brummelen, Heng, and Tabunshchyk 2021). Educators may want to focus on particularly challenging concepts for their students; thus, we propose, **DR1: Emphasize concepts that are challenging for particular audiences**.

Opportunities to implement **DR1:**

- Emphasize *Training*, *Turn-taking*, *Machine learning*, *Societal impact and ethics*, *Constrained vs. unconstrained natural language*, and CA terminology when teaching CA curricula
- With children, emphasize *Training* and *Testing*
- With parents, emphasize CA terminology and *Agent modularization*
- With those from non-WEIRD countries, emphasize *Training* and *Events*
- With those from WEIRD countries, emphasize *Testing* and *Turn-taking*

RQ2: Self-Efficacy and Identity as Programmers

In our study, those with additional experience in related topics generally had increased self-efficacy and identified more as programmers. For instance, those with previous AI experience identified more as being able to design and create their own technology projects, being programmers and being able to make impacts in their communities using technology throughout the intervention. Those with experience with more than one type of CA (as opposed to a single CA) felt similar increases to those who had learned AI. Those with text-based programming experience saw themselves more as programmers than those with no prior experience throughout the intervention. For those with initial visual programming experience, however, after engaging with the CA workshops, there was no significant difference in terms of programming identity with those who had text-based experience. Thus, creating meaningful visual programming projects and engaging in societal impact curricula may impact those with some visual programming experience more than those without any. Thus, we propose, **DR2: Supplement CA development activities with additional CA engagement, AI learning and programming activities**, as these opportunities likely significantly affect people's identity and self-efficacy as programmers. Also note, however, how these prior experiences (with additional CAs, AI learning activities and programming) may indicate differences in socioeconomic class, so the experiences themselves may not have caused the benefits seen in this study. Nonetheless, providing more opportunities for these types of activities to diverse audiences is key to democratizing technology.

Opportunities to implement **DR2:**

- Encourage activities such as learning AI, experiences with more types of CAs and programming, in addition to CA curricula activities, as the additional activities may increase self-efficacy and identification as programmers, and lead to better CA learning outcomes

RQ3: Partner Models and Trust

Through interacting more with Alexa and going through the workshops, participants overall felt Alexa was more of a friend. Such increased feelings of friendship may also increase feelings of trust long-term (Alfano 2016; Seymour and Van Kleek 2021). Furthermore, those without prior AI knowledge trusted agents more after learning to program them. Considering how trust and student-teacher relationships are important factors when learning (Schöbel, Janson, and Mishra 2019; Al-Yagon and Mikulincer 2004), it may be helpful for CAs in teaching roles to be personified. This trust-building may be encouraged by teachers, through increased interactions with CAs, or through activities in which students program CAs, depending on the student audience.

Nonetheless, this trust-building could lead to over-trust of agents, which can have serious consequences (Xiao, Borah, and Su 2021; Seymour and Van Kleek 2021). In our study, participants trusted Alexa more than their friends or parents in terms of information correctness (which may indicate over-trust, although we leave this question for future research). Thus, we propose, **DR3: Design learning activities to foster appropriate trust of agents**—which could mean either encouraging increases or decreases in trust depending on the situation. Fortunately, students' trust towards agents was not static, and about a quarter of them indicated they felt their trust changed through the programming activities. When describing how their trust changed, participants most often referenced predictability, then competence and then integrity. They also emphasized how learning about CAs, including learning about how CAs are programmed, CAs' sources of information and how CAs understand information given to them, affected their sense of trust. Educators may want to emphasize these concepts during learning activities to encourage appropriate levels of trust.

Opportunities to implement **DR3:**
- Encourage student trust of pedagogical agents through enabling students to interact with CAs more often and in ways that encourage friendship-building
- Encourage student trust of pedagogical agents through teaching students how CAs work
- Encourage student trust of pedagogical agents through using personified or friendly pedagogical agents
- Encourage student reflection on agent trustworthiness by teaching about the aspects of agents' predictability, then competence and then integrity
- Encourage reflection on agent trustworthiness by teaching about how CAs are programmed, CAs' information sources and how CAs understand information

We also found different groups of participants' partner models changed differently through the activities. For instance, after the programming activity, we found participants from WEIRD countries felt Alexa was less competent than those from non-WEIRD countries did. Before the workshops, we found participants without text-based programming experience thought Alexa was more competent than those with experience thought. Those with no prior AI experience thought Alexa was more dependable after learning how to program it. Throughout the workshops, participants who used CAs in their first language (vs. another language) thought Alexa was more human-like.

Since students' partner models could affect their understanding as well as how they interact with agents (Doyle, Clark, and Cowan 2021), it is important for students to have accurate partner models. Thus, we propose, **DR4: Foster accurate partner models through teaching related CA ideas and activities**. For example, to reinforce how CA technology is still in its infancy—and how CAs are not highly *competent* in all tasks—for a WEIRD audience, a programming activity may be appropriate, but for a non-WEIRD audience, a more direct instruction approach may be appropriate. To level-set perception of CA *competence* between those with and without text-based programming experience, a visual programming tutorial on CA development may be appropriate. To increase perceptions of CA *dependability* for those who have not learned about AI before, a programming activity may be appropriate. To increase perceptions of CA *human-likeness*, using diverse, relatable CAs and CAs in the audience's first language may be appropriate.

Opportunities to implement **DR4:**
- To reinforce how CAs are not highly competent in all tasks, for a WEIRD audience, a programming activity may be appropriate, but for a non-WEIRD audience, a more direct instruction approach may be appropriate
- To level-set perceptions of CA competence between those with and without text-based programming experience, one may use a visual programming tutorial
- To increase perceptions of CA dependability (if the CA *is* dependable) for those who have not learned about AI before, a programming activity may be appropriate
- To increase perceptions of CA human-likeness (if the CA *is* human-like), using diverse CAs and CAs in the audience's first language may be appropriate

Limitations and Future Work

While this study successfully showed various groups of people's perceptions and trust of CAs changed through learning about CAs, there were limitations. For example, there were many Indonesians in the non-WEIRD category and many Americans in the WEIRD category. Furthermore, there were more child than parent participants. Future studies could further balance subsets to verify our overall results. Future studies could also utilize various types of agents (e.g., with different voices) to verify our results. Another limitation includes how we only focused on teaching a subset of the 40 CA concepts, and how we taught the workshops virtually. Future work can investigate which of the other concepts are challenging for students, and how context (e.g., virtual vs. in-person) affects student learning. Finally, future work could also define accurate/appropriate levels of trust and partner models for various agents.

Conclusions

This paper presents results from investigating how children and parents from non-WEIRD and WEIRD countries' partner models and trust of CAs change through learning to program CAs and societal impact curriculum. It also presents pedagogical CA concepts (see Figure 1 and the appendix (Van Brummelen 2022b)), design recommendations for teaching such concepts, and opportunities to implement such design recommendations with various student audiences. With these CA concepts and educational design recommendations, educators and researchers can develop curricula to prepare students for a conversational-agent-filled world.

Acknowledgments

We would like to thank the participants, the Brave Conversations organizing committee and the MIT App Inventor team, who made this study possible.

References

Al-Yagon, M.; and Mikulincer, M. 2004. Socioemotional and Academic Adjustment Among Children with Learning Disorders: The Mediational Role of Attachment-Based Factors. *The Journal of Special Education*, 38(2): 111–123.

Alfano, M. 2016. Friendship and the structure of trust. In *From personality to virtue*, 186–206. Oxford University Press.

Braun, V.; Clarke, V.; Hayfield, N.; and Terry, G. 2019. Thematic Analysis. In Liamputtong, P., ed., *Handbook of research methods in health social sciences*. Springer, Singapore.

Code.org. 2022. Learn about Artificial Intelligence (AI). https://code.org/ai. Accessed: 2022-09-01.

Cowan, B. R.; Branigan, H. P.; Begum, H.; McKenna, L.; and Szekely, E. 2017. They Know as Much as We Do: Knowledge Estimation and Partner Modelling of Artificial Partners. In *The Annual Meeting of the Cognitive Science Society (COGSCI)*.

Craft, S.; Ashley, S.; and Maksl, A. 2017. News media literacy and conspiracy theory endorsement. *Communication and the Public*, 2(4): 388–401.

DiPaola, D. 2021. *How does my robot know who I am?: Understanding the Impact of Education on Child-Robot Relationships*. Master's thesis, Massachusetts Institute of Technology.

Doyle, P. R.; Clark, L.; and Cowan, B. R. 2021. What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. New York, NY, USA: Association for Computing Machinery. ISBN 9781450380966.

Doğruyol, B.; Alper, S.; and Yilmaz, O. 2019. The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures. *Personality and Individual Differences*, 151: 109547.

Gaube, S.; Suresh, H.; Raue, M.; Merritt, A.; Berkowitz, S. J.; Lermer, E.; Coughlin, J. F.; Guttag, J. V.; Colak, E.; and Ghassemi, M. 2021. Do as AI say: susceptibility in deployment of clinical decision-aids. *npj Digital Medicine*, 4(1): 31.

Long, D.; and Magerko, B. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.

McKnight, D. H.; and Chervany, N. L. 2001. What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology. *International Journal of Electronic Commerce*, 6(2): 35–59.

MIT. 2021. Responsible AI for Social Empowerment and Education (RAISE). https://raise.mit.edu/index.html. Accessed: 2021-09-11.

Murad, C.; and Munteanu, C. 2020. Designing Voice Interfaces: Back to the (Curriculum) Basics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.

Schöbel, S.; Janson, A.; and Mishra, A. 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*.

Seaborn, K.; Miyake, N. P.; Pennefather, P.; and Otake-Matsuura, M. 2021. Voice in Human–Agent Interaction: A Survey. *ACM Comput. Surv.*, 54(4).

Seo, H.; Xiong, A.; and Lee, D. 2019. Trust It or Not: Effects of Machine-Learning Warnings in Helping Individuals Mitigate Misinformation. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19, 265–274. New York, NY, USA: Association for Computing Machinery. ISBN 9781450362023.

Seymour, W.; and Van Kleek, M. 2021. Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2).

Smith, S. 2018. Digital voice assistants in use to triple to 8 billion by 2023, driven by smart home devices. Scientific analysis or review, Juniper Research, Hampshire, UK.

Statista. 2021. Voice commerce in the United States. https://www.statista.com/study/60607/voice-commerce-in-the-united-states/. Accessed: 2022-08-10.

Straten, C. L. v.; Peter, J.; Kühne, R.; and Barco, A. 2020. Transparency about a Robot's Lack of Human Psychological Capacities: Effects on Child-Robot Perception and Relationship Formation. *J. Hum.-Robot Interact.*, 9(2).

Sturm, C.; Oh, A.; Linxen, S.; Abdelnour Nocera, J.; Dray, S.; and Reinecke, K. 2015. How WEIRD is HCI? Extending HCI Principles to Other Countries and Cultures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, 2425–2428. New York, NY, USA: Association for Computing Machinery. ISBN 9781450331463.

Touretzky, D.; Gardner-McCune, C.; Martin, F.; and Seehorn, D. 2019. Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 9795–9799.

Van Brummelen, J. 2022a. *Empowering K-12 Students to Understand and Design Conversational Agents: Concepts, Recommendations and Development Platforms*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Van Brummelen, J. 2022b. Learning Affects Trust: Design Recommendations and Concepts for Teaching Children—and Nearly Anyone—about Conversational Agents Appendix. https://gist.github.com/jessvb/e35bc0daf859c30f73008a1ad1b37824. Accessed: 2022-09-11.

Van Brummelen, J.; Heng, T.; and Tabunshchyk, V. 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)*.

Van Brummelen, J.; Kelleher, M.; Tian, M. C.; and Nguyen, N. H. 2022. What Do WEIRD and Non-WEIRD Agent Users Want and Perceive? Towards Transparent, Trustworthy, Democratized Agents. In *Human-Computer Interaction (cs.HC)*. Ithaca, NY, USA: arXiv.

Van Brummelen, J.; Tabunshchyk, V.; and Heng, T. 2021. "Alexa, Can I Program You?": Student Perceptions of Conversational Artificial Intelligence Before and After Programming Alexa. In *Interaction Design and Children*, IDC '21, 305–313. New York, NY, USA: Association for Computing Machinery. ISBN 9781450384520.

Vraga, E. K.; Bode, L.; and Tully, M. 2022. Creating News Literacy Messages to Enhance Expert Corrections of Misinformation on Twitter. *Communication Research*, 49(2): 245–267.

Xiao, X.; Borah, P.; and Su, Y. 2021. The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs. *Public Understanding of Science*, 30(8): 977–992. PMID: 33663279.

Zhou, M. X.; Mark, G.; Li, J.; and Yang, H. 2019. Trusting Virtual Agents: The Effect of Personality. *ACM Trans. Interact. Intell. Syst.*, 9(2–3).

Zhu, J.; and Van Brummelen, J. 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)*, 1–5. IEEE.

Zierau, N.; Hildebrand, C.; Bergner, A.; Busquet, F.; Schmitt, A.; and Marco Leimeister, J. 2022. Voice bots on the frontline: Voice-based interfaces enhance flow-like consumer experiences & boost service outcomes. *Journal of the Academy of Marketing Science*.