

# Zhorai: A Conversational AI Agent for K-12 Education

Jessica Van Brummelen  
CSAIL, MIT

Phoebe Lin  
GSD, Harvard

Galit Lukin  
ORC, MIT

May 2019

We present workshop curriculum in which grade 3-5 students talk to an artificially intelligent conversational agent to learn about natural interaction with machines, how machines learn, how machines represent knowledge and reason about information, and the social impact of such agents. During the workshop, students teach the agent about animals and the agent guesses which ecosystem the animal comes from using prior knowledge. At the end of the workshop, a moderator facilitates discussion about machine perception and reasoning, as well as AI ethics.

To determine the effectiveness of the workshop, students complete a pre-workshop and post-workshop questionnaire, including questions such as "Describe how you think computers learn new things, or why you think computers can't learn new things" and "Do you think computers make fair decisions? Why or why not?". Students are also given exercises to complete individually which test their level of understanding.

To facilitate the workshop, a moderator guides the students through three modules while teaching conversational AI concepts. To aid the moderator, our website contains an interface to interact with the conversational agent, Zhorai, who takes the moderator step-by-step through the curriculum. The website also contains example discussion questions and visual representations of the agent's knowledge and reasoning for additional guidance. The website and modules are discussed in detail in the Section 2.

# 1 Motivation

Conversational AI technology is rapidly becoming commonplace in vehicles, smartphones, and even homes [3]. However, there are few resources for people without backgrounds in computer science to figure out how these devices work. We wanted to develop a tool to enable kids to explore AI through natural interaction. By talking with Zhorai, we hope students learn that conversational agents, like Siri and Alexa, are not "magic boxes", but rather tools that have been engineered to interact with and help humans.

Research has found that children ages 6-10 likely believe conversational agents, like Alexa, are smarter than them. To the authors' knowledge there are very few tools to teach young students about conversational AI. One tool for preschool through second-grade students includes *PopBots*, which are personalizable, speech-enabled robots that teach students about machine learning, generative AI, and knowledge-based systems [6]. Another tool for high school students includes MIT App Inventor's conversational AI interface, which enables students to develop their own conversational agent apps [5]. Our tool aims to teach students between these two tools' age groups about agents' knowledge representation and reasoning, natural interaction, machine learning, and social impact.

## 2 Project Design

### 2.1 Education

#### 2.1.1 Learning Objectives

The learning objectives mentioned below constitute the four Big AI themes taught in this activity:

- Learning
- Representation and Reasoning
- Natural Interaction
- Social Impact

The learning objectives of the activity are the following:

- Understand how the agent learns new information
  - The agent will visualize its knowledge and its process of understanding will be discussed throughout the activity.
- Learn how the agent represents and understands the English language
  - Students will witness how Zhorai creates the mind map and plot graph after having heard the new information.
- Learn how machines classify concepts
  - After having knowledge on both ecosystems and animals, Zhorai will classify animals to their corresponding ecosystem based on their representations. Students will visually see how Zhorai executes this task which might succeed or fail.
- Learn the capabilities and the limitations that machine learning has and why
  - Students will witness Zhorai's limitations since Zhorai does not know about all of the possible ecosystems.
  - Students will also experience mistakes in Zhorai's understanding of English
  - Lastly, students will understand that Zhorai does not not classify animals well under certain circumstances which will be discussed.
- Understand the social implications that models and data have and why they exist
  - This will be learned through a discussion held at the end of the activity.

Since students interact with the agent, they will understand its capabilities through trial and error. This aims to be a fun experience which will provide students with a good understanding of what the agent can do. The other learning objectives will be taught through visuals, discussions, and individual exercises.

### 2.1.2 Learning Assessment

Students' level of understanding will be quantified using the exercise worksheet that students will individually fill out during Module 4. The exercises will measure the understanding of all learning objectives related to AI and how Zhorai works. All learning objectives will be qualitatively assessed through the discussions held throughout the activity, and in Module 4 particularly.

Since the activity is given to a small group of students, all of the students will participate in discussions and interact with Zhorai. Thus, although the students are young, we expect them to have good understanding of the learning objectives. Given their age, it might be difficult for students to generalize the knowledge. Therefore, it is important that the moderator discuss real-world applications such as Siri and Alexa. It is important that at the end of the activity, students are able to comprehend how conversational agents learn and why they make mistakes.

### 2.1.3 The 4Ps

The activity supports the 4Ps:

- **Project:** In this activity, students learn by interacting with Zhorai, thus they learn by doing. If they teach Zhorai correctly (i.e. provide a good description of an ecosystem), Zhorai will succeed in the classifying task. However, if for example, the students do not provide enough information about an animal, Zhorai will fail at its task.
- **Passion:** Students feel like they are directing their own learning by teaching Zhorai whatever they want about the animals with no restrictions of the language.
- **Play:** The narrative is fun and engaging for the students and they enjoy sharing the opportunity to communicate with Zhorai.
- **Participation (Peers):** Students participate together in training Zhorai. Students might also learn from each other about new animals. Lastly, students engage in a discussion about Zhorai's capabilities and the importance of ethics in AI.

## 2.2 Activity

The workshop is comprised of four modules. The students will interact with Zhorai in the first three modules. In module one, they will learn about Zhorai’s current knowledge about ecosystems. In module two, students will witness how Zhorai learns new information about animals. In module three, they will learn how Zhorai infers about animals and their ecosystems. In module four, the moderator will lead a discussion summarizing the AI concepts that were brought up throughout the activity and the importance of ethics in AI. In addition, there is a warm up and cool down in which the students receive pre-learning and post-learning activity assessments. In the warm up, the students also interact with Zhorai for the first time through a casual conversation.

### 2.2.1 Activity Breakdown

The activity is expected to take about an hour and is composed of discussion, online tasks, and unplugged exercises:

- Warm-up (10 minutes; Unplugged and online)
- Module 1: Understanding Knowledge Representation in Machines (10 minutes; online and discussion)
- Module 2: Teaching New Knowledge to Machines (10 minutes; online and discussion)
- Module 3: Witnessing Machine Learning Classification (10 minutes; online and discussion)
- Module 4: AI and Ethics Discussion (15 minutes; unplugged and discussion)
- Cool down (5 minutes; Unplugged)

### 2.2.2 Example Activity Flow

A live skit of the activity will be demonstrated in the class. In this section, we describe example dialog for the activity.

The following dialog illustrates the warm-up activity, which begins with the moderator introducing Zhorai and continues with a dialog between a student and Zhorai:

**Moderator:** Today we're going to talk to an alien from another planet! It's name is Zhorai, and it's actually a computer program, or *conversational agent* that MIT students created. Would someone like to say hi to Zhorai? (Moderator chooses a student, who comes up to the computer and clicks record.)

**Student:** Hi!

**Zhorai:** Hi there! What's your name?

**Student:** My name is Erin.

**Zhorai:** Erin, what a nice name. Where are you from?

**Student:** I'm from Boston.

**Zhorai:** Interesting! I'm from planet Igbruhmmelkin. I've never heard of Boston before. Can you tell me more? (A button appears with the words, "Find out what Zhorai knows about your planet!", which takes the class to the next module.)

The following dialog illustrates the general flow of Module 1:

**Moderator:** Zhorai already knows a little bit about planet Earth. Let's ask it about what it knows. Who wants to ask Zhorai what it knows about the desert?

**Student:** What do you know about deserts?

**Zhorai:** Here's what I know about deserts. They're fascinating! (A mindmap about deserts appears, as shown in Figure 1. This mindmap was created using the sentences that appear below the mindmap, as shown in Figure 2.)

**Moderator:** Why don't we ask Zhorai what it knows about oceans?

**Student:** What is an ocean?

**Zhorai:** Oh yes, oceans are very interesting. Here's what I know. (Another mindmap and list of sentences appears.)

**Moderator:** What other ecosystem can we ask Zhorai about?

**Student:** What do you know about plains?

**Zhorai:** Hmmm, I haven't heard about that ecosystem before, but I know about rainforests.

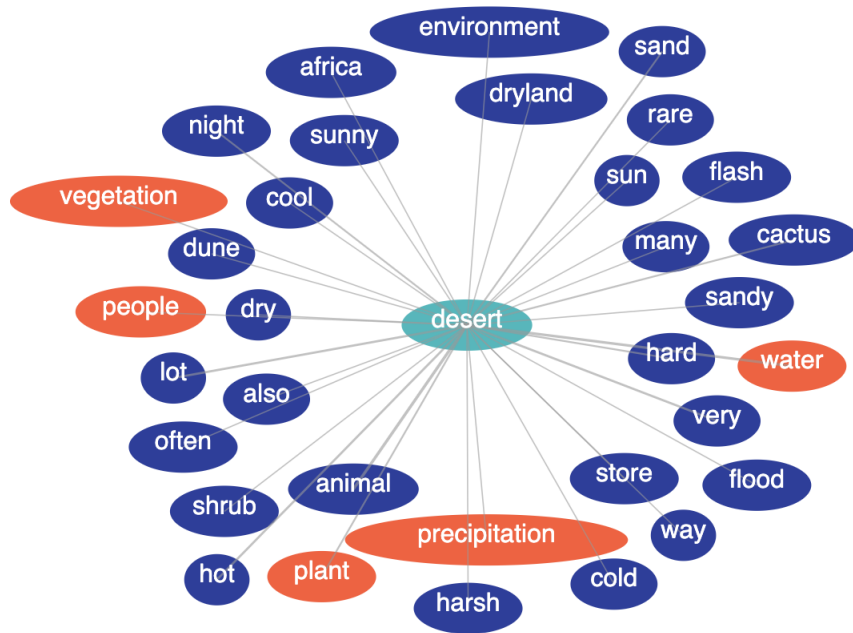


Figure 1: The desert mindmap.

Zhorai heard these things about deserts from other Earthlings:

The desert has lots of sand  
 The desert is very dry  
 Deserts have cactus  
 The desert is very hot and sunny  
 Deserts don't have much water  
 Deserts have few people  
 Deserts have flash floods

Figure 2: A few of the sentences that created the desert mindmap in Figure 1.

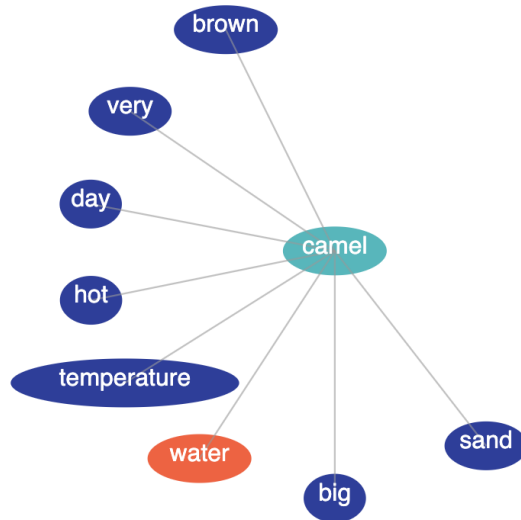


Figure 3: An example mindmap for camels.

The class discusses why Zhorai knows only about some ecosystems.

The following is a partial example dialog for Module 2:

**Moderator:** Zhorai knows a little bit about ecosystems, but doesn't know about what animals live in each ecosystem. Let's teach Zhorai about these animals. (In this module, there is a prompt on-screen that says, "Zhorai would like to know about *camels*. Could you teach it about them?".)

**Student:** Camels are big and brown. They walk on the sand all day. Camels live in hot temperatures and they drink very little water.

**Zhorai:** Wow, camels sound really interesting! Let me think for a bit and then I'll show you my thoughts. (A mindmap about camels appears on screen, as shown in Figure 3.)

A new prompt appears on the screen stating, "Zhorai would like to know about *fish*. Could you teach it about them?".

**Student:** Fish live in the water. They are very small and swim all day.

**Zhorai:** Fish sound fascinating! Now I want to visit earth and all of it's life! I'll show you what I understand after I think for a little while.

The students then discuss how well they think Zhorai understands each animal.



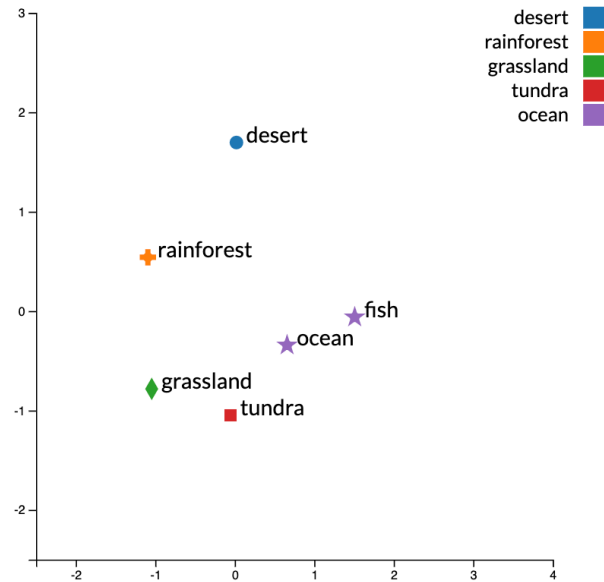


Figure 4: An example scatter-plot for fish.

The following illustrates example dialog for Module 3:

**Moderator:** Now that Zhorai knows about ecosystems and animals, let's see if it can guess which ecosystems certain animals belong to. Who would like to ask Zhorai to guess?

**Student:** Where are fish from?

**Zhorai:** Oh yeah, fish sound interesting. Let me think about where they might be from.

**Zhorai:** Based on what I know about Earth, I would guess fish live in oceans. (A scatter-plot appears on screen with fish and oceans close together, as shown in Figure 4. The mindmap about fish from Module 2 also appears on screen.)

**Moderator:** Let's think about this for a second. Zhorai is a computer program, so it uses numbers to think. This scatter plot shows what Zhorai understands about ecosystems and fish. You can see that there are numbers on the sides of the graph. Based on this graph, why do you think Zhorai thinks fish come from oceans?

**Student:** Because fish are really close to the ocean on the graph!

**Moderator:** Exactly! The distance (which can be represented by a number)

between fish and the ocean ecosystem is the smallest. Let's see if Zhorai can guess where another animal comes from.

**Student:** Guess where polar bears are from.

**Zhorai:** I'll think about polar bears for a bit and let you know!

**Zhorai:** Hmm, I'm not sure! I haven't heard much about that animal.

**Moderator:** Since we didn't teach Zhorai about polar bears, it doesn't have any idea where they are from.

Each classification will be accompanied by a discussion about how Zhorai represents knowledge and why it has succeeded or failed at the classification task. Zhorai may fail at the classification task if it is not given enough descriptors to place the animal correctly in the *embedding space* (which is the high dimensional space that is simplified and illustrated by the scatter-plot). Zhorai needs a sufficient amount of description words about animals to discern where the animals might live.

In module 4, the moderator may begin by discussing how Zhorai works. Some questions that may be discussed include:

- How does Zhorai learn and perceive the world?
- How did Zhorai learn English?
- Is it a similar way to how it learned about ecosystems?
- How did Zhorai hear us?
- What does the computer "see" when it gets information from a microphone?

The moderator may continue by discussing how Zhorai learns:

- What else could we teach Zhorai?
- Can we teach it about different cultures?
- How would that be different?
- Would it know whether what we teach it is correct?
- How would you feel if Zhorai learned something untrue about your culture?

Finally, the moderator can discuss the implications of its mistakes:

- Why were there animals whose ecosystem Zhorai could not guess?
- Is Zhorai's knowledge of the world biased?
- How can we improve this?
- Why is it important?

### **2.2.3 Activity Resources**

The following are the tools and materials needed for the activity:

- Laptop (3-4 students per device) with a working microphone and speaker
- Pencils (1 per student)
- Pre-activity assessment
- Post-activity assessment
- Worksheet (1 per student)
- Teacher guide

### **2.2.4 Teacher Guide**

The teacher guide is a written document that accompanies the teacher through the activity. The guide provides the structure of the activity as well as suggested questions and discussions that could be brought up given that the students are capable of understanding. The document takes into consideration that the teacher has little prior machine learning knowledge and provides the reader with basic information to pass on to the students.

There is no student guide because the activity is a group activity guided by the teacher, Zhorai, and the website. The students only receive an unplugged worksheet and pre- and post-assessments.

### 2.2.5 Activity Evaluation

The learning activity is designed to expose young children to interacting with machines and understanding how machines learn. Therefore, the evaluation of the learning activity will assess both whether the explicit learning goals were accomplished and the ease of interaction with the conversational agent.

In order to assess learning, students will individually complete a post-activity worksheet. This worksheet incorporates concepts that are introduced in modules 1 through 3. For example, we pose a pre-populated mindmap for a cheetah and ask students to generate sentences that would enable Zhorai to have such a mental representation of a cheetah. We also ask students about what kind of information will break Zhorai's model. Successful responses to these questions will demonstrate full comprehension of the examples given during the activity and how Zhorai works internally.

In order to assess the natural interaction between students and Zhorai, students will complete pre- and post-activity oral questionnaires. The questionnaires will be administered orally to ensure student understanding of the questions (i.e., a semi-structured interview). The questionnaires evaluate the following outcome metrics related to usability and usefulness of the system [1]. The pre-activity questionnaire is composed of only "Perception of self as an engineer" and "trust" questions, as they are not dependent on the learning activity. All items are rated on a 7-point scale.

- Learning utility: *"How useful were the activities we did today in learning about how machines learn?"*
- Visualization utility: Students answer the Likert scale questions: *"The mindmaps for each ecosystem and animal helped me think about Zhorai's brain."* and *"The scatterplot helped me understand how Zhorai made decisions."*
- Perception of self as an engineer: Students answer the Likert scale questions: *"I think I can teach a machine.:"* and *"I am curious about how a machine learns."*
- Trust: Students answer the Likert scale question on how much they would trust a similar system in the real world: *"I trust agents like Siri and Alexa."*

- Future use: Students answer the Likert scale question: *"I would continue teaching Zhorai more things and see what it can learn."*

Lastly, we acquire informal feedback from the teachers on quality of teacher resources and overall student engagement of the learning activity.

## 2.3 Technical

The activity is comprised of six technical components:

- **Speech Recognizer**  
The speech recognizer was implemented using the Web Speech API's `webkitSpeechRecognition` interface [4]. This interface enables real-time JavaScript speech recognition in the browser.
- **Speech Synthesizer**  
Zhorai's speech is synthesized through the Web Speech API's `SpeechSynthesisUtterance` interface [4]. This interface synthesizes voices to speak given text. We adjusted the API's `Google US English` voice to have a higher pitch and sound relatively gender neutral.
- **Semantic Parser**  
The semantic parser extracts meaning from the sentences outputted from the speech recognizer. It is used to create the mind map and to extract keywords such as names, ecosystems, and animals.  
The parser receives the sentences outputted from the speech recognizer. It outputs the JSON which creates the mind map and the entities that it identifies in the sentences.  
The semantic parser works as follows:
  1. The semantic parser uses a software for compositional semantics [2] to extract the structure and meaning of each sentence. This software also extracts the part of speech of each word.
  2. The parser identifies which animals or ecosystem the sentence is describing.
  3. The parser identifies which which words describe the entity.
  4. For every word found in step three, the parser identifies if the word has a positive or negative correlation with the entity described.

## 5. The parser outputs

- which name it has identified (warm up)
- which ecosystem or animal is being described (modules 1,2,3)
- the json which creates the mind map (modules 1,2)

- Conceptual Word Embedder

To aid in the detection of similarity between learned concepts (ecosystems) and new concepts (animals), we developed a word embedding model that disambiguates each concept in our set of concepts that we train and test Zhora's intelligence on. This embedding transforms the set of words in a concept to a numerical vector based on a representation that results in a small distance between similar concepts.

To evaluate our embedding model on a particular concept, we scraped the internet for sets of text that mention various ecosystems (rainforest, grassland, tundra, ocean, desert) to use as input for training the word embedding. Our test data consisted of sets of text scraped from the internet on a list of 22 animals. We additionally added a set of sentences that children might say to determine whether or not our system could correctly contextualize text with a different average structure as the data from the internet.

We utilize BERT embeddings as a starting point in which there is not much correlation between a particular animal and its corresponding ecosystem. Then we trained an embedding classification model that takes in a sequence and classifies the concept of that sequence. This model works as follows:

1. Each word sequence, where every word is represented by its BERT embedding, passes through a sequence-of-vectors generating LSTM.
2. The generated sequence is passed to an LSTM whose last hidden output is size 256
3. The 256 dimensional vector is passed through two fully connected layers
  - The first outputs embedding vectors of size 128.
  - The second generates logits used to predict the relevant concept.

- Word Embedding Comparator

To classify a word to a particular concept, the following method is used.

1. All sentences in the corpus which contain an animal are put through the network to extract the embedding from the embedding layer.
  2. Keep only the embeddings that predict the most commonly predicted ecosystem.
  3. Each of these individual embeddings are averaged, and the resulting embedding is normalized by the L2 norm of the resulting vector.
  4. The ecosystem with an embedding that is closest to this new, normalized embedding is chosen as the correct class. The embeddings of the list of potential ecosystems are derived by normalizing the column of the last fully connected layer relevant to the potential ecosystems.
- Website  
The website uses the Bootstrap library to construct the overall site framework. Each page consists of interaction for an individual modules. For modules 1 through 3, there is an explicit start recording and end recording interaction that dictates what Zhorai hears and thus what information is used. The JSON that the semantic parser outputs is then used as input for the visualizations. The mindmap and scatterplot visualizations are constructed using the D3 Javascript library.

### 3 Project Status

The project is prepared for testing on real students. This is the next step which will be done by the authors. In addition, for accessibility, the website will be hosted online. Currently, the system runs from a local machine. The details of the tasks progress can be seen in Figure 5.

The source code can be accessed from <https://github.com/jessvb/zhorai>

### 4 Future Work and Project Extensions

The possible next steps will be evaluated after the activity has been given to real students. Only then, we will know how students respond and what needs

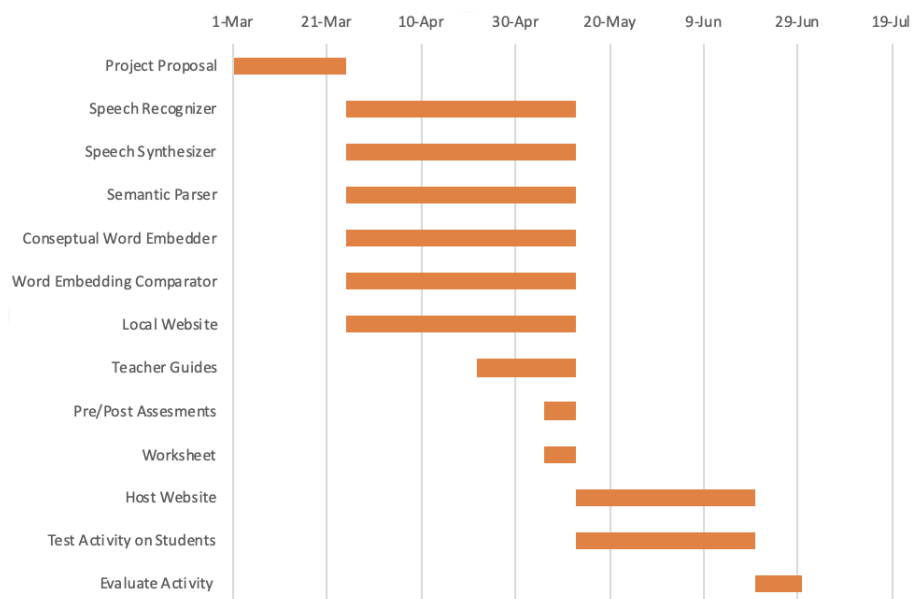


Figure 5: Zhorai's Gant Chart

to be altered, improved, or elaborated. This will be part of the evaluation shown in Figure 5.

## References

- [1] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 4:1–4:14, New York, NY, USA, 2019. ACM.
- [2] Pascual Martínez-Gómez, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki. ccg2lambda: A compositional semantics system. In *Proceedings of ACL 2016 System Demonstrations*, pages 85–90, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [3] National Public Radio. The smart audio report, winter 2018. Technical report, NPR and Edison Research, December 2018.



- [4] The Contributors to the Web Speech API Specification. The possibility of a universal social welfare function. Report, W3C, March 2019.
- [5] Jessica Van Brummelen. Tools to create and democratize conversational artificial intelligence. Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA, 2019.
- [6] Randi Williams, Hae Won Park, and Cynthia Breazeal. 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*, page 447. ACM, 2019.