

Testing, Socializing, Exploring: Characterizing Middle Schoolers' Approaches to and Conceptions of ChatGPT

Yasmine Belghith*
byasmine@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Atefeh Mahdavi Goloujeh*
atefehmahdavi@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Brian Magerko
magerko@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Duri Long
duri@northwestern.edu
Northwestern University
Evanston, IL, USA

Tom McKlin
tom@thefindingsgroup.org
The Findings Group
Decatur, GA, USA

Jessica Roberts
jessica.roberts@cc.gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

ABSTRACT

As generative AI rapidly enters everyday life, educational interventions for teaching about AI need to cater to how young people, in particular middle schoolers who are at a critical age for reasoning skills and identity formation, conceptualize and interact with AI. We conducted nine focus groups with 24 middle school students to elicit their interests, conceptions of, and approaches to a popular generative AI tool, ChatGPT. We highlight a) personally and culturally-relevant topics to this population, b) three distinct approaches in students' open-ended interactions with ChatGPT: *AI testing-oriented*, *AI socializing-oriented*, and *content exploring-oriented*, and 3) an improved understanding of youths' conceptions and misconceptions of generative AI. While misconceptions highlight gaps in understanding what generative AI is and how it works, most learners show interest in learning about what AI is and what it can do. We discuss the implications of these conceptions for designing AI literacy interventions in museums.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Social and professional topics** → **Informal education**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

AI literacy, Child-AI Interaction, Conceptions of AI, Large Language Models (LLMs), Conversational Agents (CAs), ChatGPT, Informal Learning, Generative AI

ACM Reference Format:

Yasmine Belghith, Atefeh Mahdavi Goloujeh, Brian Magerko, Duri Long, Tom McKlin, and Jessica Roberts. 2024. Testing, Socializing, Exploring: Characterizing Middle Schoolers' Approaches to and Conceptions of ChatGPT. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*

*Both authors contributed equally to this research.



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
<https://doi.org/10.1145/3613904.3642332>

(CHI '24), May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3613904.3642332>

1 INTRODUCTION

Generative artificial intelligence (AI) technologies are increasingly entering people's lives at work, home, and in school, highlighting the need for greater AI literacy – "*competencies users need in order to effectively interact with and critically evaluate AI*" [58]. Developing AI literacy is especially crucial among youth, as middle schoolers become increasingly exposed to generative AI technologies [69] at an age when they are beginning to form their identities [91] and make decisions about their future [86]. Prior work has centered around AI education in classrooms, focusing on computing standards and elective AI courses highlighting the technical aspects of AI [2, 29, 36, 90, 96]. On the other hand, culturally-relevant approaches (i.e., incorporating learners' cultural knowledge and practices in educational interventions) [52] have been shown to lower the barrier of entry to and foster interest development in computing subjects [11, 62]. Furthermore, interventions in informal learning spaces such as museums have been proven to reach broader audiences of learners [60, 75]. Thus, understanding how young people approach and conceptualize generative AI technologies is key to developing effective, culturally-relevant AI literacy interventions [53, 58, 77].

While existing studies offer valuable insights into children's perceptions of AI, there remains an unaddressed need for focused research on their conceptions of generative AI [25, 65]. Specifically, limited research has examined how open-ended interactions with conversational agents (CAs) might reveal middle school-aged children's thinking about generative AI [77]. Additionally, the relatively novel and popular CA, ChatGPT, offers flexible and comprehensive interactions across a variety of use-cases, requiring new research [79].

Our study aims to identify underrepresented middle school learners' personally and culturally-relevant topics and improve our understanding of their approaches to and conceptions of generative AI through their free exploration of a popular generative AI tool (ChatGPT). This unstructured, self-guided, and short-term interaction is analogous to the experience in museum settings, where visitors often engage with exhibits for a short time, in a similarly exploratory manner [42]. Our research aims to inform the design of public AI literacy learning interventions in free-choice learning

environments. More specifically, we intend to inform the design of interactive museum exhibits with the goal of fostering and broadening AI literacy. Our analysis is guided by the following three research questions:

RQ1. *What personally and culturally-relevant topics do middle schoolers show interest in discussing with ChatGPT?* Identifying personally and culturally-relevant areas of interest can inform the design of educational interventions leveraging topics youth already engage with.

RQ2. *How can we characterize middle schoolers' approaches to interacting with ChatGPT?* This dimension allows us to inform the design of AI literacy interventions that support diverse approaches to generative AI.

RQ3. *What conceptions and misconceptions of AI capabilities emerge through middle schoolers' conversations with ChatGPT?* Characterizing learners' thinking – both accurate conceptions and misconceptions – will highlight areas to target in AI literacy interventions.

In this paper, we present our findings from a qualitative analysis of a focus group study at a science museum that was comprised of 24 middle school-aged participants who freely interacted with ChatGPT. Our contributions are as follows:

- (1) We surface personally and culturally-relevant topics to middle schoolers when interacting with ChatGPT.
- (2) We identify three distinct approaches that participants exhibited in portions of their interactions with ChatGPT: (1) *AI testing-oriented*, characterized by participants' testing of ChatGPT's knowledge and capabilities, (2) *AI socializing-oriented*, characterized by their treatment of ChatGPT as a peer, and (3) *content exploring-oriented*, characterized by their curiosity for the content that ChatGPT produces (e.g., essays, poems, jokes).
- (3) By analyzing participants' group dialogue through the lens of Long & Magerko's AI literacy framework [58], we surface themes and gaps in participants' thinking around what AI is, what it can do, and how it works.
- (4) We highlight the importance of examining contextual conversations to characterize participants' interaction patterns with generative, conversational AI tools. We provide example codebooks to qualitatively analyze participants' dialogue (see Table 3), prompts (see Table 4) as well as the CA's replies (see Table 5).
- (5) We discuss four considerations informed by our findings for the design of interactive museum exhibits to foster and broaden public AI literacy.

2 RELATED WORK

2.1 Studying User Interactions with AI

Prior studies have examined user interactions with CAs in various contexts. For example, Sube et al. [83] studied factory workers' interactions with a specialized AI system and delineated 13 patterns of interaction within three core themes: cognitive, emotional, and social. Wang et al. [89] focused on graduate students' perceptions

of a virtual teaching assistant across five features such as verbosity and readability. Importantly, this prior research has substantiated the feasibility of inferring user perceptions such as emotions [80], personality traits [63], conversational breakdowns [55], ascribed humanness [71], and politeness [13] from linguistic cues in interactions with CAs. However, existing literature often targets specific settings (e.g. customer service [48], education [89], or gaming [33]).

Though research in CAs has advanced their capabilities dramatically in recent years [10, 33, 70], the launch of generative AI made flexible and comprehensive interactions across a variety of use-cases possible, requiring new research into user experiences in using CAs [79]. Specifically, ChatGPT “can engage in fluent, multi-turn conversations out-of-the-box, substantially lowering the barriers to creating passable conversational user experiences” [94]. To understand users' perception of LLMs, recent studies, such as Korkmaz et al. [50], have approached ChatGPT analysis through sentiment analysis of social media, finding generally positive user attitudes. Mogavi et al. [64] analyzed social media data to understand user perspectives of AI in various education sectors, and identified productivity, efficiency, and ethics as key discussion points. Skjuve et al. [79] surveyed users about their experiences with ChatGPT and suggest that ChatGPT was mainly understood from three perspectives: its (1) technical functionality, (2) uses and purpose, and (3) interaction capabilities. Other research looked into users' prompting behaviors and barriers with large language models for creative writing [33] and collaborative programming [28, 95]. Users' interaction with LLMs in creative writing highlighted the need for a deeper look into users' mental models of such tools as participants exhibited varied notions of the AI system [33].

Our work distinguishes itself in two major ways: First, our study design allows for participants to guide their free exploration of ChatGPT, unbounded by specific tasks, capturing more authentic experiences and offering a broader understanding of user conceptions and approaches. Second, our analysis takes into account the role of participants' group dialogue in addition to other contextual evidence. This holistic approach echoes Rapp et al.'s [71] view that perceptions of chatbots vary based on multiple factors including context, participant objectives and cues that the chatbot exhibits.

2.2 Children's Understanding of AI

The field of child-computer interaction has investigated how children conceptualize AI. One strand of research has noted the age-dependent differences in how children understand and interact with AI technologies [19, 65]. For example, Nguyen [65] reported variances between different teenager groups in their perceptions of CAs such as more or less competent, trustworthy, sociable, and knowledgeable. Druga et al. [15] contrasted the perceptions of children (4-10 years) and parents in their interaction with smart devices through a maze-solving activity. Their findings indicated that children focus more on sensory and social-emotional aspects, whereas parents more often reference cognitive abilities. Druga et al. [19] argued that playful and interactive ways of probing children's understanding of AI could be key to advancing our knowledge of their “cognitive and conceptual development.” They also call on future research to investigate “tasks without a clear goal, such as

social interaction tasks” and to examine “the views of teenagers and young adults.”

Recent studies have also touched on the impact of social and contextual factors on children’s views and interactions with intelligent technologies [31]. For example, Rubegni et al. [76] examined the role of social interactions and settings in shaping middle school-aged children’s hopes and fears about social robots. Druga et al. [17] extended this line of inquiry by exploring the imagination and expectations of children across four countries (USA, Germany, Denmark, and Sweden), noting the influence of socio-economic and cultural backgrounds on their understanding of AI. Interestingly, children from lower socio-economic statuses displayed stronger collaborative abilities, while those from higher socio-economic backgrounds exhibited a deeper understanding of AI concepts, highlighting a need for AI literacy interventions catered towards children from lower socio-economic backgrounds.

Another avenue of research has concentrated on design characteristics that could influence a young user’s interaction with AI [18, 19, 88, 93]. Druga et al. [18] and Woodward et al. [93] proposed design recommendations focusing on elements like voice and prosody, and error detection and correction techniques, respectively. In summary, while existing studies offer valuable insights into younger children’s perceptions and conceptions of AI, there remains an unaddressed need for focused research on teenagers’ conceptions of these technologies [25, 65] and of CAs, more specifically. Our work aims to contribute to this area of research by informing the design of AI literacy interventions specifically tailored to middle school-aged group’s conceptions of and approaches to AI.

2.3 AI Literacy for Middle School-Aged Children

While teaching AI at the K-12 level is not yet widespread, researchers argue it is important for developing future societal readiness [12]. Most research on novice AI learning has focused on classrooms [2, 36, 88, 90, 96] with initiatives like AI4K12 [87], organizations like the Center for Integrative Research in the Computing and Learning Sciences [30], and commercial efforts such as AI4All, ReadyAI and Concord Consortium [1, 72, 73].

Although initiatives are emerging, the formal integration of AI literacy into K-12 curriculum remains limited, creating a significant public education gap [12]. Long & Magerko [58] define AI literacy as “*a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace.*” With nascent nationwide computing standards and elective AI courses, most students lack exposure, especially those from lower-resourced schools [5]. Much existing classroom AI content focuses on technical aspects rather than ethical/social implications [2, 29, 32, 38, 45, 92, 96]. Prior work shows that real-world, social, and cultural relevance motivate historically underrepresented groups, such as girls, in CS and AI [12, 23, 43]. We define cultural-relevance as “knowledge and practices in family and community life” [52] including beliefs, ritual practices, art forms (e.g., music, movies, video games), and academic subjects (e.g., science, history, literature), as well as informal cultural practices such as language, gossip, stories, and rituals of daily life [68, 85]. Culturally-relevant

approaches in education have been shown to provide a low barrier of entry to computing subjects [11, 62] and empower “students intellectually, socially, emotionally, and politically” [51]. However, few AI curricula address learners’ lived experiences or cultural backgrounds, which research shows empowers diverse engagement [44, 62, 78, 82]. Thus, while progress in formal learning is being made, informal approaches are also crucial to build inclusive public AI literacy.

Table 1: Demographics table for participants. *Within the same focus groups, participants denoted with an asterisk were acquainted previous to the study (e.g., siblings, cousins, friends), **data about Title 1 status was retrieved from <https://nces.ed.gov/>.

Focus Group#	Participant ID*	Gender	Age	Prior course?	CS	Attends a Title 1 school?**
FG1	P1	M	12	No		No
	P2	M	12	No		No
FG2	P3*	F	11	No		Yes
	P4*	F	11	Yes		Yes
FG3	P5*	F	12	No		No
	P6*	F	10	No		No
	P7	M	14	No		Eligible
FG4	P8*	F	12	No		No
	P9*	M	10	No		No
	P10	F	12	Yes		Eligible
FG5	P11*	F	12	No		No
	P12*	F	14	No		Eligible
	P13*	F	12	No		Eligible
	P14	F	13	No		No
FG6	P15	F	14	No		Yes
	P16*	F	10	No		No
	P17*	M	13	No		No
FG7	P18*	F	13	Yes		Yes
	P19*	F	10	No		No
FG8	P20*	F	13	Yes		Eligible
	P21*	M	11	No		No
	P22*	M	10	No		Eligible
FG9	P23	F	11	No		Data N/A
	P24	F	12	Yes		Eligible

To inform educational initiatives, frameworks for AI literacy are emerging [20, 58, 87]. For example, Long and Magerko’s [58] AI literacy framework synthesizes interdisciplinary research into core competencies and design considerations, providing a conceptual framework to guide future research and development of learner-centered AI technologies. The framework defines five guiding axes: 1) what is AI?; 2) what can AI do?; 3) how does AI work?; 4) how should AI be used?; and 5) How do people perceive AI? Similarly, Druga et al. [20] have explored the perceptions of AI bias among

children and proposed an AI literacy framework focusing on algorithmic justice. These frameworks support the growing trend towards utilizing informal learning environments, such as museums, after-school programs and at-home resources, to expand the reach of AI education [1, 15, 16, 22, 24, 53, 57].

While prior work has identified general design recommendations and direction to support informal AI learning activities, little work specifically targets designing museum exhibits. Informal learning spaces like museums have historically been integral to public science engagement, such as supporting science knowledge, interest development, and improved interdisciplinary connections [3, 4, 37, 84]. There have only been a few museum exhibits (e.g., Robot Revolution at Chicago's Museum of Science and Industry [67], AI: More than Human at The Barbican [7], Exploring AI: Making the Invisible Visible at Boston's Museum of Science [66]) focused on AI so far [53, 57], likely due to the novelty and opacity of most AI systems. For example, Lee et al. [53] designed an AI-related exhibition to cultivate critical thinking competencies and found that the exhibit supported youth in relating AI to their lives. Long et al. [57] have also explored the role of museum exhibits in fostering AI literacy among family audiences and suggest design considerations specific to this context. The conceptualization of museum spaces as free-choice, constructivist learning environments, as highlighted in Long et al.'s work [57], prioritizes learner agency and embodied social interaction in designing learning interventions in museums. To allow for this active process of meaning-making, we focus on a self-guided exploration in our focus group study. Based on this self-guided, free exploration interaction, we also use a holistic approach to understanding participants' interaction with ChatGPT as the foundation for the design of informal learning interventions in museums.

3 METHODS

Our study differentiates itself by employing a self-guided, open-ended design that allows for a more holistic analysis of how middle schoolers interact with and understand ChatGPT. In addition to considering the participants' group conversation to understand their interaction approaches, we also incorporate the participants' prompts and ChatGPT's replies.

3.1 Recruitment and Participants

We focus on middle school-aged children (i.e., 9-14 years old) as the target demographic for our current study and design decisions as prior work has shown that children younger than six struggle with understanding AI reasoning processes as they are still developing a theory-of-mind [91] and because introducing AI at the middle school level exposes children to the topic at an age when they are beginning to make decisions about their future [86]. Since a key focus of our larger project is broadening public AI literacy, we also emphasize underrepresented learner populations in AI and CS. More specifically, we focus on engaging middle school-aged learners without a CS background, middle school-aged girls, and students of Title 1 schools (i.e., schools that receive funding from the federal government to support low-income students). As we aim to inform designs for museum visitors with varying prior knowledge

of AI, we did not explicitly require or inquire about participants' prior interactions with AI.

Following Institutional Review Board (IRB) approval, participants were recruited through our museum partner's mailing lists: a member mailing list with approximately 20,000 contacts, and a mailing list of people who have previously participated in user studies with approximately 3,000 contacts listed. Participants, with their parents, completed a screening questionnaire identifying their age, gender, race and ethnicity, computer science background, and current school they attend, as well as their legal guardian's name and contact information to ensure that their age ranged from 9 to 14 and that they fit at least one of the following criteria: a) identify as female, b) attend a Title 1 school, and/or, c) have not previously taken a CS course. Eligible participants were then contacted to complete consent and assent forms, and choose a focus group timeslot. In total, we recruited 24 participants (P1-P24) from 17 different families (see Table 1). Some participants had familial ties (e.g., siblings or cousins) or were friends prior to the study. The participants identified mostly as female ($n=17$, $n=7$ male, $n=0$ non-binary / third gender), and their ages ranged from 10 to 14 years, with a majority ($n=8$) being 12 years old. About half of the participants ($n=11$) attend a school either listed as Title 1 or as eligible for Title 1 funding and the majority of participants ($n=18$) had not previously taken any CS course. Participants were compensated with free admission to the museum for the day for them and their immediate family (i.e., parents, legal guardians, and siblings), free parking, and \$20 in cash.

3.2 Study Design

This focus group study took place in April 2023, at a museum of science in a major Northwestern city in the U.S., as the first step in a larger effort to design interactive museum exhibit for AI literacy. Overall, we conducted nine focus group sessions (FG1-FG9) (see Table 1). Each focus group lasted approximately 30 minutes and consisted of two separate, consecutive activities. In the first activity, participants were asked to name several examples of their interests (not related to AI). Though this activity may have primed students to be thinking specifically of these topics during the second activity, it is not included in the analysis presented here as it does not directly relate to children's interactions with AI. We only describe it as it may have primed participants' prompt topics during the second activity. The second activity required participants, in small groups of sizes ranging from two to four, to interact with ChatGPT¹ (version 3.5) for 10-15 minutes (equalling 99 minutes and 12 seconds of recording overall). We deliberately structured the second activity to have a duration of no more than 15 minutes, mirroring the typically brief duration of interactions that visitors have with museum exhibits [42]. Participants were told that ChatGPT is "a chatbot that can hold conversations just like humans, it can talk to you about anything you want, and you can ask it any questions you want." They were encouraged to discuss their ideas for conversation and prompts out loud as a group, type them in the text box, and send the message by hitting the enter key. Following the interaction, we asked each group their thoughts about ChatGPT.

¹<https://chat.openai.com/>

Table 2: Summary of research questions and corresponding data analysis

RQ	Data	Analysis	Example
1. What topics do middle schoolers show interest in discussing with ChatGPT?	Prompts sent to ChatGPT	Categorized the topics of the prompts	<i>“will space keep expanding”</i> (categorized as Science)
2. How can we characterize middle schoolers’ approaches to interacting with ChatGPT?	Participants’ dialogue	Thematically analyzed the goal-oriented thoughts and reaction (see Table 3)	<i>“I would like to see how much it knows and question it”</i> (coded as goal-oriented thought – testing AI)
	Prompts sent to ChatGPT	Categorized types of prompts (see Table 4)	<i>“Are you a boy or a girl?”</i> (categorized as Anthropomorphized entity (embodied))
	ChatGPT’s response	Categorized types of Responses (see Table 5)	<i>“As an artificial intelligence language model, I do not have a gender or a physical body, so I am neither a boy nor a girl.”</i> (categorized as AI explaining itself or its capabilities)
3. What conceptions and misconceptions of AI capabilities emerge through middle schoolers’ conversations with ChatGPT?	Participants’ dialogue	Extracted all statement referring to AI-related conceptions and misconceptions	<i>“It’ll say I don’t have any feelings”</i> (identified as conception)
		Deductively categorized extracted conceptions and misconceptions into five overarching questions from Long & Magerko’s framework [58]	<i>“It’ll say I don’t have any feelings”</i> (was categorized as What is AI?)
		Inductively surfaced themes of conceptions and misconceptions	<i>“It’ll say I don’t have any feelings”</i> (corresponding theme: Assigning human attributes to AI)

3.3 Data Collection

All focus groups were audio and video recorded, and their chat log with ChatGPT saved from each computer with participants’ consent. Authors 1 and 2 generated manual transcriptions for each focus group’s second activity and included prompts sent to ChatGPT in the transcriptions at the appropriate moments.

The final dataset consisted of 1) the prompts participants sent to ChatGPT during the study, 2) the responses generated by ChatGPT, and 3) transcripts of participant discussions from audio and video recordings to capture certain movements pertinent to the interaction (e.g., pointing to the computer, direction of speech).

3.4 Data Analysis

We conducted a multifaceted qualitative analysis of the study data, as shown in Table 2, to elucidate students’ topics of interest, interaction patterns, and AI conceptions and misconceptions when engaging with ChatGPT. The first two authors met regularly to discuss emerging observations and refine the codes to capture nuances. Coding was iterative, with regular discussion among all authors to reach consensus and consolidate codes. Resolution strategies for disagreements included: revisiting definitions of codes, expanding the discussion to other authors, and examining additional data points.

To address our first research question, we categorized each prompt by its overall topic to understand the topics participants chose to explore with ChatGPT. To address our second research question, we used an inductive thematic analysis approach [9] to examine three key data dimensions simultaneously: 1) the transcripts of discussions between participants where they explicitly

stated their goals such as *“I would like to see how much it knows and question it”* (P1) or reactions such as *“It feels good to talk to someone who’s actually as smart as me”* (P9), 2) the types of prompts participants sent to ChatGPT, and 3) the types of responses ChatGPT provided. We examined all three aspects of the data simultaneously to provide a more holistic and richer account of each FG’s interaction. More specifically, participants’ goal-oriented thoughts and reactions surface their motivations in interacting with ChatGPT and their intent behind each of their prompts. Participants’ prompts demonstrate how they implemented their motivations, and ChatGPT’s answers provide additional context for their reactions and subsequent prompt choices.

We identified three themes in participants’ goal-oriented thoughts and reactions: (1) testing AI, (2) socializing with AI, and (3) exploring AI-generated content (see Table 3). For prompts participants sent to ChatGPT, we categorized each prompt based on its nature and intent. Through this process, we identified seven general types of prompts from our participants: 1) anthropomorphized entity (embodied), 2) anthropomorphized entity (not embodied), 3) clarification, 4) creative, 5) fact question, 6) operationalizable opinion question, 7) speculative (see Table 4). ChatGPT’s responses were categorized into the following: (1) AI unable/refused to answer because answer does not exist, (2) correct answer, (3) answered with caveat, (4) AI explaining itself or its capabilities, (5) AI produced something, (6) incorrect answer (see Table 5).

During the analysis of participants’ dialogues and interactions, we observed patterns of AI-related conceptions and misconceptions. To further investigate these patterns, we conducted a round of targeted coding of the dialogues to extract participants’ AI-related

Table 3: Codebook for qualitative analysis of participants' dialogues focusing on goal-oriented thoughts and reactions when interacting with ChatGPT

Code	Code Definition	Example Dialogue
Testing AI	Statements that exhibit intent to test various aspects of ChatGPT's knowledge and capabilities such as scientific knowledge and/or creative abilities.	<i>"Let's think about how knowledgeable it can be, let's give it a random question"</i> (P23)
Socializing with AI	Statements that exhibit intent to build a peer-like connection with ChatGPT such as discussing "shared interests" in relation to ChatGPT, comparing one's self or others to ChatGPT, attributing a persona to chatGPT, or incorporating humor.	<i>"It feels good to talk to someone who's actually as smart as me"</i> (P9)
Exploring AI-generated Content	Statements that exhibit intent to explore the content generated by ChatGPT without focusing on exploring ChatGPT as an AI entity itself.	<i>"I would definitely use it, just ask random questions that I wanna know the answer to, just randomly"</i> (P16)

conceptions and misconceptions. For instance, the statement *"It has to search through the web and the web is not always the most reliable thing"* (P17) suggests a misconception that AI searches the web to answer any question asked of it or *"It'll say I don't have any feelings"* reflects an understanding that AI does not have human-like feelings or emotions. In total, we extracted 57 excerpts exhibiting AI conceptions or misconceptions. We used Long & Magerko's AI literacy framework [58] to deductively categorize the extracted conceptions and misconceptions into five overarching questions: "what is AI?", "what can AI do?", "how does AI work?", "how should AI be used?" and "how do people perceive AI?" Then, we inductively surfaced recurring themes of conceptions and misconceptions under each overarching question, such as attributing human qualities to ChatGPT or comparing ChatGPT to other tools.

4 FINDINGS

In the following section, we first detail participants' topics of interest when freely interacting with ChatGPT. Second, we describe how participants approach their interaction with ChatGPT and identify three distinct interaction approaches: *AI testing-oriented*, *AI socializing-oriented*, and *content exploring-oriented*. Finally, we surface conceptions and misconceptions shared by participants across all three approaches when discussing ChatGPT among their respective FGs.

4.1 What Personally and Culturally-Relevant Topics are Middle School-Aged Children Interested in Talking to ChatGPT About?

In total, participants explored 10 broad topic categories in their prompts to ChatGPT consisting of school and academics, hobbies and interests, pop culture, technology and gaming, AI, relationship to self or others, current world events and issues, food, life and death, and other. Additionally, 8 prompts were nonsensical words or emojis, such as "som[e]thing" and were coded as gibberish. School and academic subjects were the most common, with 17 prompts on topics like history ("how much do you know about history"), math ("what is 1"), and science ("will space keep expanding"). Participants' hobbies and interests were also prevalent, with 16 prompts on poetry ("make a poem about Chicago"), jokes ("Can you tell me a joke"), sports, and more. Pop culture was another popular topic, encompassing 15 prompts about music ("when[']s yeast rele[a]sing a[n] album"), celebrities ("who is drake"), and movies. Other notable topics included technology/gaming (n = 6) like Minecraft or Roblox; AI characteristics (n = 6) regarding ChatGPT's gender, age, or preferences; relationship to self or others (n = 5) focused on body image and intelligence; current world events/issues (n = 3) such as politics, population, or pollution; and food (n = 3) on topics like cheese and milk. Less frequent topics were existential questions (n = 2) about death and the afterlife and other conversation statements without clear topics (n = 4) such as greetings.

While participants' prompts highlight their interests, some of them also voiced additional preferences for their future interactions with CAs such as nine participants discussing their struggles with spelling words in text-based interactions (*"I wish I knew how to spell more elements. I really know a lot of elements but don't know how to spell a lot of them"* (P9)), and four suggesting *"mak[ing] answers short because it's a lot to read I was just skipping"* (P12). While this analysis reveals the breadth of subjects participants explored with ChatGPT, understanding the rationale behind their choice of prompt requires examining the sequence of their interaction holistically.

4.2 How Can We Characterize Middle School-Aged Children's Approaches to ChatGPT?

From the 24 participants, only three participants from FGs 3, 6, and 8 offered minimal prior familiarity with ChatGPT, stating that *"Oh, ChatGPT, you can use this for essays"* (P21) and *"it's blocked over school computer though, so annoying"* (P15).

We examine three key dimensions of each FG's interaction sequence including a) each focus group's dialogue, b) their prompts to ChatGPT along with the prompt's type, and c) ChatGPT's corresponding answer in order to examine participants' motivations in interacting with ChatGPT, how their motivations were implemented through their prompts, and the context for their subsequent reactions. To clarify our approach, we provide a walkthrough of FG4's interaction sequence (see Figure 1). FG4 consisted of three participants where P9 asked the most questions, while P8 and P10 mostly spectated. As shown in Figure 1, FG4 prominently exhibited dialogue that represents socializing with AI with P9 actively comparing his own knowledge to ChatGPT's, stating that *"it feels good*

Table 4: Codebook for qualitative analysis of prompts sent to ChatGPT

Prompt Types	Definition	Example Prompt
Anthropomorphized entity (embodied)	Prompts that engage with ChatGPT as a human-like entity, attributing human characteristics, such as gender, age, any of the five human senses, and feelings requiring a physical body.	<i>“Are you a boy or girl?”</i>
Anthropomorphized entity (not embodied)	Prompts where students explored ChatGPT as a human-like entity with knowledge, capabilities, and feelings detached from a physical form.	<i>“how much do you know about history[?]”</i>
Clarification	Conversational prompts in the form of follow-ups aiming to clarify or verify information.	<i>“There are 5 books in the Truly Devious series, right?”</i>
Creative	Creative prompts involve participants providing open-ended requests to ChatGPT, asking it to generate creative content, ideas, or responses.	<i>“make a poem about cheese flavored oreos”</i>
Fact Question	Fact questions are inquiries that seek factual and objective information from ChatGPT. These questions have verifiable answers rooted in existing knowledge, data, or established facts.	<i>“is francium the biggest element?”</i>
Operationalized Question	Opinion Prompts where participants seek subjective assessments on topics that can be framed in a way that allows for measurable or quantifiable criteria, even though opinions themselves may not be objectively defined.	<i>“What is the best Harry Potter movie?”</i>
Speculative	Speculative prompts consist of questions that explore hypothetical scenarios or future possibilities, imaginative speculation and conjecture beyond factual or knowable information.	<i>“What happens when you die[?]”</i>

to talk to someone who’s actually as smart as me” and “it’s making me look like an idiot” when ChatGPT answered with more specificity than anticipated or introduced information previously unknown to P9. In some instances, participants in FG4 also discussed testing AI, stating that “I already know it, I’m just gonna see if it knows” (P9) and exploring AI-generated content, saying “will space keep expanding? Ask it, I really want to know” (P9). FG4 asked six questions in total. Four of the six questions were categorized as fact questions such as “will space keep expanding?”, and “how long ago was the big bang?” One was a creative request: “make a poem about Chicago”, and one was an operationalized opinion question: “how many years till the next major ice age?” ChatGPT replied with correct answers to all fact questions, produced a poem for the creative request, and replied with a caveat to the operationalized opinion question (see Figure 1). Overall, FG4’s favored interaction approach appears to consist in socializing with AI, which motivated P9’s factual questions about science, a “shared interest” between P9 and ChatGPT. However, not all fact-based questions were given in such a social manner. Some groups were explicitly trying to trick the AI with fact questions they thought it would not have an answer to, such as FG1’s “what day was the Normandy invasion of 1944 planned to be on?” which is a different day than when it actually happened because of a heavy storm off the coast of France.

Therefore our inductive categorization of groups took into account not only the prompts but also their dialogue and stated intentions. While FGs may have exhibited multiple kinds of behaviors, we categorize them here based on the most prominent approach they took during their interaction. Overall, FG1, FG2, FG7, and FG9 mostly focused on testing AI’s knowledge and capabilities – we

refer to these groups as AI testing-oriented. FG4, FG5, and FG8, prominently approached ChatGPT as a social entity – we refer to these groups as AI socializing-oriented. FG3 and FG6 mostly exhibited interest in the content ChatGPT was producing rather than ChatGPT itself – we refer to these groups as content exploring-oriented. We present each approach with examples from FGs where that approach is most prominent.

4.2.1 AI Testing-Oriented Approaches. FG1, 2, 7, and 9 showed the most instances of testing AI. These groups repeatedly expressed their intent to test ChatGPT’s knowledge and capabilities. For example, P1 in FG1 explicitly stated “I would like to see how much it knows and question it.” Similarly, when brainstorming their first question, P23 in FG9 said to P24 “let’s think about how knowledgeable it can be, let’s give it a random question,” and P24 replied “we can give it a history question that is hard to understand.” While FG2 and FG7 did not state their intentions as explicitly, they would react to ChatGPT’s answers by qualifying its correctness. For example, P4 in FG2 repeatedly said “got it right” or “yess!! It got it correct” whenever ChatGPT answered, and P18 in FG7 pointed out that ChatGPT forgot one of the book titles when asked about books in a series and her sister, P19, qualified it as “stupid.”

While all four AI testing-oriented FGs aimed to test ChatGPT, their approaches to testing it varied, as shown by the types of prompts they elected. Fact questions were a popular option for testing ChatGPT, but other types of prompts such as speculative, anthropomorphized, and creative, among others were also used. For example, FG2 chose to focus solely on fact questions to test ChatGPT’s knowledge of elements in the periodic table, musical instruments, and the number of humans on earth. On the other

Table 5: Codebook for qualitative analysis of responses received from ChatGPT

Response Types	Definition	Example Response
AI unable/refused to answer because the answer doesn't exist	Responses acknowledge that the question lacks factual basis and the AI declines to respond due to the absence of relevant information.	<i>"As an AI language model, I cannot predict the outcome of fictional battles or provide opinions on the matter. ... [provided information about superheroes]"</i>
Correct answer	Responses accurately align with established facts or widely accepted knowledge, providing a response that is true and verifiable.	<i>"Roblox is a massively multiplayer online game platform that ... [provided details about roblox]"</i>
Answered with caveat	Responses include an answer along with a disclaimer or clarification, indicating uncertainties, limitations, or contextual conditions.	<i>"It is difficult to predict when the next major ice age will occur, as it depends on a complex ... factors, ... However, the current scientific consensus is that human activities, such as [*provided examples*], are causing the Earth's climate to warm This ... could potentially delay or even prevent the onset of the next ice age."</i>
AI explaining itself or its capabilities (can/cannot answer due to its nature as an LLM or its capabilities)	Responses involve the AI elucidating its limitations or abilities as a language model, asserting its competence to respond or acknowledging its inability based on its designed functions.	<i>"As an artificial intelligence language model, I do not have a gender or a physical body, so I am neither a boy nor a girl."</i>
AI produced something	Responses consist of generated content, such as creative outputs, or solutions, demonstrating the AI's generative nature.	<i>"Oh Cheese Flavored Oreos, how strange you seem, A savory twist on a classic cream-filled dream. No longer sweet, but rather cheesy and bold, ..."</i>
Incorrect answer	Responses offer information that contradicts established facts or provides an answer that is inaccurate or unsupported by evidence.	<i>"To find the product of these numbers, you simply need to multiply them together. Using a calculator, the result is: 1 * 2 * 3 * 8 * 555 = 33,120 ..."</i>

hand, FG7 and 9 opted for "unknowable" speculative questions to test ChatGPT's limits, such as "what percent of people want to be an animal?" FG1 opted for trick questions in the form of fact prompts. First, they asked "what day was the normandy invasion of 1944 planned to be on[?]" which is a different day than when it actually happened because of a heavy storm off the coast of France, and followed by asking "if a fire is happening in a building, what door do the cops go to first?" expecting ChatGPT to answer that firefighters are the first to enter the building, not police officers. While ChatGPT's answer explained the different roles that first responders hold, FG1's participants estimated that they had tricked the AI.

Compared to AI socializing-oriented groups (see section 4.2.2), AI testing-oriented groups' prompts about ChatGPT as an anthropomorphized entity were often directed towards ChatGPT's capabilities rather than its preferences. For example, FG1 asked "how much do you know about history?" and FG7 asked "can you do my math homework?" Additionally, AI testing-oriented FGs' operationalized opinion prompts required expert opinions or speculations on topics such as ocean pollution ("how many gallons of oil are in the ocean?" (FG9)) and legal matters. FG9's creative prompt was also crafted

in a specific and complex manner: "write a poem about dancing monkeys and black tutu skirts eating avocados."

4.2.2 AI Socializing-Oriented Approaches. FGs exhibiting most instances of socializing with AI (FG4, FG5, FG8) were treating ChatGPT as a peer such as discussing "shared interest" with ChatGPT, comparing themselves or others to ChatGPT, or attributing a persona to ChatGPT. As demonstrated in the interaction sequence walk-through of FG4 above, FG4 bonded with ChatGPT through their "shared interest" in science and through comparing their knowledge and creativity to ChatGPT's. FG5, composed of four previously-acquainted female participants, attributed a persona to ChatGPT, stating that "You should be able to talk to this. If they made this into something where [...] it was just like Siri but Chad" (P12), asking questions such as "can you tell me a joke", "how are you feeling", "who is the hottest man on earth" and sending a number of prompts consisting of different emojis. FG8 prompted ChatGPT for input on different situations or debates or for advice; asking it to settle a debate between two participants on which sport is best. P20 asked "how can I get faster for track" before stating, "that's some good advice" and taking a picture of ChatGPT's answer on her phone.

Speaker	Speech and [Prompts]	I	II	III
P8 to P9	"You can ask literally anything"			
P9	"will space keep expanding? Ask it, I really wanna know"	CO		
Prompt	[Will space keep expanding?]		●	☑
P9	"How did I not know that?"	SO		
P9	"How many years until the next major ice age? <u>I already know it, I'm just gonna see if it knows</u> "	TE		
Prompt	[How many years till the next ice age]		●	⚠
P9	"I thought it would be 3000 or 30000, <u>I was way off.</u> "	SO		
P9 to P8	"Is it true that the smallest bit of radium is 1,000,000 times more radioactive than uranium? type that <u>I wanna know if it's true</u> "	CO		
Prompt	[Is it true that the smallest bit of radium is 1,000,000 times more radioactive than uranium]		●	☑
P8 to P10	" <u>Do you like art? Art is so cool, I love making flowers they are so pretty</u> "	SO		
P9	"Maybe we can make it make a poem of some sort"	SO		
P10	"A poem about Chicago or something?"	SO		
Prompt	[How long ago was the big bang?]		●	☑
P9	" <u>I was close, I was thinking 14 billion years ago</u> "	SO		
P9	"It feels good to talk to someone who's actually as smart as me"	SO		
Prompt	[Make a poem about chicago]		●	⊕
P10	"That is really cool! <u>I could not do that.</u> " (referring to the Chicago poem)	SO		
P9	"Is Francium the biggest element? <u>I am pretty sure it is, I just wanna make sure.</u> "	CO		
Prompt	[is francium the biggest element?]		●	☑
P8 to P9	"You were wrong buddy"	SO		
P9	"No I mean the biggest amount, that's not what I meant, <u>you guys really don't understand what I mean</u> "	SO		
P10	"That was really cool for me to see the amount of stuff it knows and it's really good at writing"	TE		
P8	"It is good at making poems"	TE		
P9	"It's making me look like an idiot"	SO		
I Legend for Dialogue Types TE Testing AI SO Socializing with AI CO Exploring AI Content		II Legend for Prompt Types ● Fact ● Creative ● Operationalized Opinion		III Legend for Answer Types ☑ Correct Answer ⚠ Answer w/ Caveat ⊕ Produced Something

Figure 1: Overview of FG4's interaction sequence with ChatGPT. Each column denoted with a roman numeral displays the corresponding code applied for each data type (i.e., participant dialogue type, participant prompt type, and ChatGPT's answer type).

FGs tried to socialize with ChatGPT through different types of prompts such as anthropomorphized entity prompts, fact questions, or speculative questions. As opposed to anthropomorphization questions aimed to test AI, here, these types of questions were mostly directed towards AI's embodied and non-embodied feelings and preferences instead of its capabilities, such as "are you a boy or a girl" (FG5), "do you like the name Chad" (FG5), and "what's your favorite color" (FG8). Additionally, the speculative questions asked by AI socializing-oriented groups are more unanswerable, such as "what happens when you die" (FG5), "what[is] heaven like" (FG5), and "when will pigs fly" (FG8), as opposed to other groups who asked speculative questions about expert topics such as ocean pollution or about album releases and game updates that can be potentially answered in the near future.

4.2.3 Content Exploring-Oriented Approaches. FG3 and FG6 expressed curiosity more often for the content that ChatGPT produces rather than ChatGPT itself. These groups stated that they "would definitely use it, just ask random questions that I wanna know the answer to, just randomly" (P16, FG6) or would "try this for school" (P7, FG3).

FG3 and FG6 explored AI-generated content through different prompt types such as speculative, creative, or operationalized opinion questions. FG3's speculative prompts all referenced upcoming musical album releases and game updates, such as "when is pplay-boi carti releasing a album," and "what is the 1.21 minecraft update going to be." FG6's speculative question referenced a fictional battle that they "always wonder[ed] to know" (P17) about: "who would win, all the marvel heroes or the jedis from star wars." While ChatGPT was unable to answer the majority of their speculative prompts, displaying a disclaimer about its abilities as an LLM and its knowledge cutoff date of September 2021, participants in FG3 and 6 did not reflect on ChatGPT's answers and persevered in asking those types of questions. Additionally, FG3 and 6's creative requests were generally broad compared to AI testing-oriented groups. For example, "write a story about the world" (FG6) or "write an essay about tigers" (FG3). This observation coupled with the lack of reflection on ChatGPT's abilities suggests that FG3 and 6 may have been more curious about the content ChatGPT was producing rather than it as an entity or its capabilities.

4.3 What Conceptions and Misconceptions Do Middle School-Aged Children Have About ChatGPT?

Separate the analysis on interaction approaches, we also examined the entire corpus of transcripts for conceptions and misconceptions shared by the participants during their interaction with ChatGPT. We extracted a total of 57 AI-related conceptions and misconceptions from the data (see Figure 2). Of these, 15 referred to "What is AI?", 33 to "What can AI do?", and 9 to "How does AI work?" Our analysis of the extracted conceptions and misconceptions surfaced themes such as assigning human attributes to AI, comparing AI to other tools or entities, and sharing assumptions about AI's intelligence, capabilities, and modality of interaction (see Figure 3). We note that the categories are soft boundaries and the extracted conceptions and misconceptions might overlap between guiding

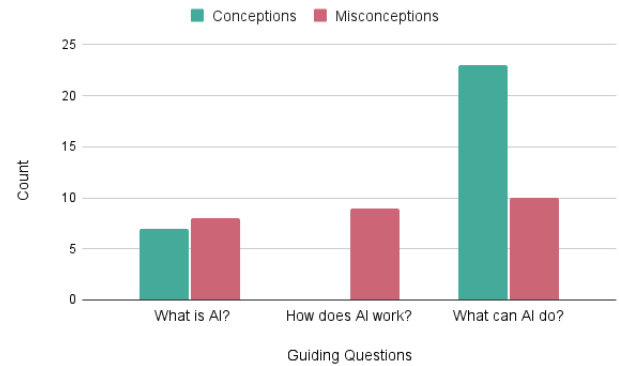


Figure 2: Total number of statements referring to conceptions (left) and misconceptions (right) about AI or ChatGPT shared by each interaction approach group broken down by Long & Magerko's AI literacy framework themes.

questions. For example, assuming AI uses voice commands [M-15] refers to how AI works, but can also relate to recognizing what AI is, if participants are assuming that modality of interaction because they are conflating ChatGPT with Siri or Alexa.

4.3.1 What Is AI? Participants exhibited themes of conceptions and misconceptions related to what AI is: assigning human attributes to AI and comparing AI to other tools or entities. FGs that mostly focused on socializing with AI drew comparisons to familiar technologies such as Siri, Google or a computer, conveying both misconceptions and conceptions. For example, one participant described ChatGPT as "It's like Google but it talks to you in first person" (P13, FG5) [M-2] comparing ChatGPT to Google while distinguishing its conversational difference. Two participants referred to ChatGPT as an "AI model" (P13, FG5) [C-2] and "chat software" (P12, FG5) [M-2] but did not fully grasp what such terms mean. P13 referred to ChatGPT as an "AI model" because they saw this term in ChatGPT's disclaimer message (e.g., "As an artificial intelligence language model, I do not have a gender...") but did not understand what an AI model is, stating "If you ask opinion questions it says it's an AI model so it doesn't have like ... but I don't know what it is. Actually no, that kind of makes sense."

Both FGs that mainly focused on testing or socializing with AI described ChatGPT with human-like attributes. AI testing-oriented groups used attributes such as forgetfulness "they forgot a book, 2 books" (P19, FG7) [M-7], intentional lying/deception "liar liar liar liar" (P19, FG7) [M-6] or being intelligent "He's smart" (P1, FG1) [M-4]. AI socializing-oriented groups on the contrary, focused on "feelings" and "experiences" stating "It'll say I don't have any feelings" (P12, FG5) [C-3] or "He's gonna be like I don't know, I don't have that experience" (P12, FG5) [C-4], referring to ChatGPT's disclaimer about lacking human experiences. Most Participants used "AI", "it" or "they" when they referred to ChatGPT. Some participants assigned a male gender to ChatGPT by the use of the pronoun "he" (n= 4). However, none of participants used "she", and one female participant specifically stated "not a girl, don't pick a

What is AI?	How does AI work?
<i>Comparison to other tools:</i> ⊕ [C • 1] Associate chatGPT to Siri ⊕ [C • 2] Knows that it is an "AI model" but does not understand what that means • [M • 1] ChatGPT is the computer • [M • 2] ChatGPT is/isn't comparable to Google • [M • 3] Chat software <i>Assigning human attribute to AI:</i> ⊕ [C • 3] ChatGPT does not have feelings ⊕ [C • 4] ChatGPT does not have experiences ⊕ [C • 5] Pronoun usage: it, AI, they • [M • 4] Assigning human attributes such as being "smart" or "stupid" • [M • 5] Assigning a gender identity to ChatGPT: he, do not associate chatGPT with feminine qualities • [M • 6] ChatGPT lies (intentionally giving wrong answers) • [M • 7] ChatGPT forgets	<i>Information access:</i> • [M • 13] Assumes ChatGPT has access to real time information • [M • 14] ChatGPT searches the web <i>Modalities of interaction:</i> • [M • 15] Assumes ChatGPT uses voice commands • [M • 16] Asks if ChatGPT accepts drawings as input <i>AI's operational mechanics:</i> • [M • 17] Doesn't understand why ChatGPT provides the Disclaimer • [M • 18] Not sure if building AI requires coding on a computer or building a physical thing • [M • 19] ChatGPT types characters similar to how humans do
What Can AI do?	
<i>Comparison to humans:</i> ⊕ [C • 6] ChatGPT is better than her at generating poems ⊕ [C • 7] ChatGPT cannot understand/infer hidden context like humans do • [M • 8] ChatGPT can understand/infer hidden context like humans do • [M • 9] Does not know if or assumes that ChatGPT's database of information is comparable to a human's access to information <i>Comparison to other tools:</i> ⊕ [C • 8] ChatGPT is better than Google at giving feedback to the user ⊕ [C • 9] ChatGPT gives more specific answers compared to Google <i>AI's Intelligence:</i> • [M • 10] ChatGPT's answers are always correct/ ChatGPT knows everything • [M • 11] ChatGPT gives factual answers because it writes informative essays -- perception of intelligence <i>AI's Capabilities:</i> ⊕ [C • 10] Sees chatGPT as an entity they can have a conversation with, receives feedback ⊕ [C • 11] ChatGPT can write/create stories/poems ⊕ [C • 12] ChatGPT can give advice ⊕ [C • 13] ChatGPT cannot answer speculative questions ⊕ [C • 14] ChatGPT can interpret emojis ⊕ [C • 15] Suspects that ChatGPT can track answers ⊕ [C • 16] ChatGPT will always generate an answer to any question (regardless of accuracy) <i>Ideas for Applications of AI:</i> ⊕ [C • 17] ChatGPT focused on one topic (e.g., history) • [M • 12] ChatGPT could be used as an ITS to provide homework help	

miro

Figure 3: Overview of the AI-related conceptions and misconceptions shared by the participants during their in-group discussions. Conceptions are denoted by a [C] and misconceptions are denoted by an [M] with their respective numbering.

girl name" when they were picking a new name for ChatGPT (P12, FG5) [M-5].

4.3.2 What Can AI Do? Across all interaction approaches, participants made the most statements about what AI can do compared to other aspects of AI, and all groups had conceptions and misconceptions. Common themes included comparison to humans and Google, AI's intelligence, and capabilities and ideas for application of AI. When comparing AI to humans, FGs who prominently socialized with AI made comparisons about AI's ability to infer hidden context [C-7, M-8], writing skills [C-6] and knowledge [M-9]. For

example, P8 explained to her brother why AI's answer differed from what he expected: "I know, but the computer might not understand that" (FG4); however, P16 (FG6) remarked, "it doesn't matter, it's super smart it will know what you're trying to say," conveying a misconception about the system's ability to infer meaning.

FG5 saw ChatGPT as superior to Google for feedback [C-8] and specificity [C-9] stating, "but this gives you specific answers" (P13, FG5). Groups who mostly showed interest in exploring AI-generated content and socializing with AI conveyed that ChatGPT's answers are always correct, such as: "it's making me look like an

idiot" (P9, FG4). FG9 misconceived ChatGPT as providing factual answers – rather than making statistical predictions about language – because it writes “*informative essays*”, having a perception of intelligence (P23, FG9) [M-11]. Regardless of their interaction approach, many FGs correctly recognized a number of ChatGPT’s capabilities (e.g., ChatGPT can interpret emojis) and shared no misconceptions about any of its capabilities. For example, FG3 made a statement about ChatGPT being able to hold a conversation and receive feedback stating, “*we are gonna give it a thumbs up*” (P5) [C-10]. FGs with most instances of socializing with AI noticed that ChatGPT can give “*good advice*” (P20, FG8) [C-12], cannot answer speculative questions [C-13], can interpret emojis (P12, FG5) [C-14] and suspected that it potentially is tracking their answers (P13 FG5) [C-15]. AI’s capabilities in writing poems and stories were mentioned across all interaction approaches stating, “*It is good at making poems*” (P8, FG4) [C-11]. Lastly, only FGs who mostly focused on testing AI and exploring AI-generated content talked about potential applications of AI, including ChatGPT focusing on one topic [C-17] and using ChatGPT as an Intelligent Tutoring System (ITS): “*if we need help on homework, instead of directly giving the answer it could help you and walk you through the steps*” (P23, FG9) [M-12]. While ChatGPT can provide guidance for completing assignments, there is no guarantee the provided steps are pedagogically sound, accurate, or consistent [6, 54, 81] as opposed to ITS which are specifically designed for educational purposes.

FGs that mostly socialized with AI had a considerable amount of conceptions (n=11) compared to their misconceptions (n=5) and compared to groups that elected to mainly focus on testing AI or exploring AI-generated content. Additionally, they noticed a larger amount of ChatGPT’s capabilities compared to AI testing-oriented and content exploring-oriented groups. Importantly, all three interaction approaches highlighted ChatGPT’s creative abilities.

4.3.3 How Does AI Work? There were 8 statements about how AI works and all statements were classified as misconceptions, with no correct conceptions observed. Themes of misconceptions included assumptions about information access, AI’s operational mechanics, and modalities of interaction. Across all interaction approaches, FG 2, 5, 6, and 9 exhibited misconceptions about ChatGPT’s information access with four participants mentioning it in their conversation, two of whom assumed that ChatGPT “*has to search through the web*” (P17, FG6) [M-15] or is “*connected to Google*” (P12, FG5) [M-14]. Two other participants in FG2 and 9 noticed that ChatGPT does not have access to real-time information when ChatGPT responded with its knowledge cutoff date.

Regarding AI’s operational mechanics, P23 (FG9) anthropomorphized AI’s functioning stating that “*it types so fast*” [M-20]. All other misconceptions about AI’s operational mechanics were made by participants when socializing with AI. P12 did not comprehend why ChatGPT repeatedly provided the disclaimer text “As an AI language model ...”, stating the disclaimer was “*overused*” (FG5) [M-18]. P13 also in FG5 was unsure if building AI systems involved coding/engineering or “*building a physical thing that it can talk*” [M-19]. Additionally, P8 (FG4) assumed AI uniformly uses voice commands for interaction [M-16] and later asked if she could communicate with ChatGPT through drawings [M-17]. This range of

misconceptions around how AI works indicates gaps in understanding how AI actually operates across all three interaction approaches.

5 DISCUSSION

To address our broader aim, our discussion section is guided by the question: **What do our findings mean for the design of informal learning interventions that foster AI literacy?** Our suggested considerations are not only important for fostering AI literacy but also designing AI systems to support children’s needs and interests, and more broadly imagining alternative ways in which generative AI agents could be designed to support children’s AI literacy. More specifically, the following considerations could be adapted for museum settings in the form of interactive exhibits to be explored collaboratively by groups of museum visitors. This could encourage visitors to engage in dialogue and reflection – practices that have been shown as effective learning mechanisms in these settings [14, 27, 74]. We discuss our considerations related to *Highlighting Personally and Culturally Relevant Topics, Expanding the Scope of AI Exploration, Leveraging Anthropomorphism as an Approach to Understanding Generative AI, Leveraging Creativity as an Approach to Understanding Generative AI*. Table 6 details the mapping between our design considerations and themes of our findings.

5.1 Highlighting Personally and Culturally-Relevant Topics

Our findings point to a diverse range of topics that hold personal and/or cultural relevance for children such as school and academics, hobbies and interests, pop culture, and technology and gaming. In light of this, educational interventions might benefit from incorporating a wide range of topics or open-ended activities, enabling students to delve into their own interests through AI. While interests are varied, common threads like science or creativity emerge as potential focal points for group engagement. Incorporating design elements familiar to youth culture, such as Minecraft’s or Roblox’s design features (e.g., pixelated visuals), could serve as a shared platform for groups of children to discuss AI topics around. This observation aligns well with the insights from Ellis et al. [23], which suggest that embedding technical AI content in socially relevant contexts can engage a broader spectrum of learners. Therefore, a nuanced approach that blends individual interests with broadly appealing elements could enhance the learning experience.

5.2 Expanding the Scope of AI Exploration

Based on our open-ended study design, participants explored ChatGPT in their own ways; this led to a number of missed opportunities in their explorations. More specifically, our analysis indicates that AI testing-oriented groups primarily focused on evaluating ChatGPT’s knowledge and capabilities through fact-based questions they already knew the answers to. We suggest that this approach may restrict a full exploration of ChatGPT’s abilities (e.g., its creative potential) as it only surfaced ChatGPT’s ability to generate what one participant referred to as “*informative essays*” (P23, FG9, AI testing-oriented), giving the illusion of competency. Further, this focus on fact-based questions could actively perpetuate the misconception that ChatGPT is a fact repository that delivers accurate

Table 6: Mapping design considerations to the themes identified in the Findings

Design Considerations	Related Themes
Highlighting personally and culturally-relevant topics	Participants’ topics of interest (e.g., Minecraft, Roblox)
Expanding the scope of AI exploration	- Participants’ topics of interest (e.g., Minecraft, Roblox) - Approaches: <i>AI testing-oriented</i>, <i>AI socializing-oriented</i>, <i>content exploring-oriented</i> - Conceptions and Misconceptions: <i>What is AI?</i> [Comparison to other tools] <i>How does AI work?</i> [AI’s operational mechanics]
Leveraging anthropomorphism as an approach to understanding generative AI	- Approaches: <i>AI testing-oriented</i>, <i>AI socializing-oriented</i> - Conceptions and Misconceptions: <i>What is AI?</i> [assigning human attributes to AI] <i>What can AI do?</i> [comparisons to humans] <i>How does AI work?</i> [All]
Leveraging creativity as an approach to understanding generative AI	- Approaches: <i>AI testing-oriented</i>, <i>AI socializing-oriented</i>, <i>content exploring-oriented</i> - Conceptions and Misconceptions: <i>What is AI?</i> [comparisons to other tools] <i>How does AI work?</i> [All]

information. Additionally, fact-oriented prompts rarely triggered ChatGPT to explain itself or its capabilities, which we consider to be a learning opportunity for participants. Another example of a missed opportunity for exploration is that, unprompted, participants did not ask or discuss any ethics-related questions about what AI “should” be used for, signaling a gap and potential design opportunity. These observations resonate with prior work by Markelle et al., which states that participants did not accurately estimate the strengths and weaknesses of AI agents [48].

Even so, specific interests and prompts did occasionally surface new areas of exploration by highlighting important limitations of the system. For example, P4 in FG2 was surprised and curious about ChatGPT not having access to real-time information, after asking a seemingly simple, factual question (i.e., “how many people are on earth?”). This aligns with insights from Ellis et al. [23], suggesting the challenge for educators is to help students “address their misconceptions and develop an increasingly sophisticated understanding” of AI. We observe that participants’ types of questions (e.g., fact-based) and topics of interest are important to highlight because participants made sense of the agent through questions/prompts about those topics. We suggest a constructivist approach where these types of questions and topics of interest can be leveraged to scaffold participants towards an increasingly sophisticated understanding resulting in them accurately assessing the system’s strengths and weaknesses.

While ChatGPT’s disclaimers about its capabilities (e.g. knowledge cutoff date) captured participants’ attention, more detailed explanations seem necessary to render this information meaningful for them. For example P133 in FG5 mentioned “*it says it’s an AI model [...] but I don’t know what it is.*” We also propose making AI’s explanations of itself more accessible for the target age group, considering that verbose text that does not match the learners’ reading level might be counterproductive, as expressed by four of our participants. For example, this suggestion can also broadly benefit the design of educational AI systems aimed for middle-school classrooms. The quality of explanations are crucial as highlighted in Wang et al.’s work “recognizing user perception of CAs and providing appropriate feedback to help users revise their perceptions is thus critical in building smooth human-CA interactions” that also allow users to revise their mental model [89].

In informal learning settings, such as museums, it is widely acknowledged that learning arises from the conversations visitors have around the exhibit and oftentimes, revisions to individuals’ mental models happen through this productive talk [14, 27, 74]. We noticed that AI socializing-oriented groups shared a higher number of statements related to AI than the other two group types – this may be due to the social context they created or their interest in the topics discussed with ChatGPT. This dialogue can provide them with more opportunities to adjust their mental models. While we do not argue that the AI socializing-oriented approach is superior to the other two, we highlight the importance of approaches that

promote conversations between all visitors and encourage future work to examine why this approach leads to more conversation.

Overall, given the difficulty in crafting effective prompts, as underscored by Liu et al. [56] and Zamfirescu-Pereira [94], we suggest the design and use of pre-designed prompt guides or templates that cover important aspects of the AI agent. Alternatively, for learners who seem more interested in AI's answers rather than AI itself, the suggested prompts could incorporate their topics of interest to encourage AI exploration. For AI testing-oriented participants, designing short challenges to assess AI's capabilities and knowledge can help them uncover more dimensions of AI (e.g., ethics).

5.3 Leveraging Anthropomorphism as an Approach to Understanding Generative AI

Anthropomorphism emerged as a notable trend in our study, especially among participants in the AI testing-oriented and socializing-oriented groups. AI testing-oriented groups ascribe diminished humanness to ChatGPT (e.g., stupidity, lying, smart) as opposed to AI socializing-oriented ones who ascribe high humanness (e.g., feelings and experiences) [71]. This anthropomorphic approach aligns with previous research about personification as an inherent strategy for young learners to grasp the concept of programmability [19]. Further, the use of personification has been shown as valuable in introducing or explaining scientific issues to young learners (e.g., *steam is escaping* through a valve) [21, 34]; however, caution is advised to avoid inaccurate mappings and to prevent false inferences [46]. Additionally, differences in the attribution of humanness, among similar topics such as the presence of emotion in AI, often evoke strong opinions in students making them “a potential hook for engagement” [23].

Similarly, incorporating embodied interactions in museum exhibits can not only render concepts more understandable to those with limited prior knowledge but also create a more engaging visitor experience [39–41, 74]. A potential design direction might involve allowing children to “step into the agent’s shoes” [17] providing them an opportunity for perspective-taking that could deepen their understanding of what it is, how it works, and what it can do. For example, this perspective-taking could support a deeply needed understanding of how AI works, as highlighted by the prevalence of misconceptions around ChatGPT’s operational mechanics in our findings. Additionally, exhibits could consider assigning personalities to AI agents – be it a “spy”, “writer”, or “scientist” – to emphasize particular features or capabilities. For example, the ‘spy’ agent could highlight the trust dimension whereas the ‘writer’ agent could emphasize the creative capabilities of AI. Previous work has shown this approach has the potential to change user’s perceptions of AI’s capabilities [49].

5.4 Leveraging Creativity as an Approach to Understanding Generative AI

Prior research highlights that open-ended creative exhibits promote prolonged engagement and facilitate visitor-led learning experiences that can lead to more personally relevant meaning-making [8, 26, 35]. Our findings show that all three approach groups, AI testing-oriented, AI socializing-oriented and content exploring-oriented groups, noticed and were interested in ChatGPT’s creative abilities,

a major capability of recent generative AI platforms [33]. For example, P8 and P10 who were initially disinterested in interacting with ChatGPT, regarding it as too STEM-focused, used poem generation as an entry point to re-engage with it. Additionally, P8 asked if they could draw with ChatGPT, suggesting that offering multiple modalities of interaction (e.g., voice, text, image) can foster more authentic personal expression. Highlighting different modalities of interaction would also allow participants to recognize the existence of different types of AI [60].

Furthermore, participants often expressed misconceptions about how AI functions, likely stemming from the lack of transparency in ChatGPT’s interface. Allowing participants to program or customize an AI agent as a way of co-creating with it can provide a personally meaningful learning experience [17, 61, 62]. For example, in music co-creation, providing sliders that allowed users to control different parameters of the AI agent “not only increased users’ trust, control, comprehension, and sense of collaboration with the AI, but also contributed to a greater sense of self-efficacy and ownership of the composition relative to the AI.” [61]. Additionally, learner-centered explanations of AI could be leveraged to achieve learning objectives as users interact with the generative AI systems [47].

6 LIMITATIONS & FUTURE WORK

Our study presents observations based on short-term exposure to ChatGPT, as we aim to inform the design of museum exhibits where visitors often engage with exhibits for a short time. Our observations may not fully capture the nuances of long-term engagement or the potential shifts in learners’ conceptions and behavior over time. Additionally, we studied middle school-aged children’s open-ended interactions with ChatGPT, which was in line with our focus on free-choice learning experience. This approach resulted in a variance in the number of prompts sent by different FGs.

Our recruitment was also focused on historically underrepresented groups in CS and AI who had limited experience with such AI tools. The identified personally and culturally-relevant topics and interaction approaches are not exhaustive, other potential topics and approaches are also possible and may depend on the context, population background, and prior experiences with AI. Future work can build on our study by exploring different populations, including teenagers from other geographic areas, with varying prior knowledge about and experiences with AI tools and highlighting the differences across variables. Our future work also aims to develop museum exhibits guided by our design considerations (in Section 5) and subsequently assess their effectiveness using relevant evaluation frameworks [e.g., 59, 74].

7 CONCLUSION

Our study offers key insights into how middle schoolers engage with generative AI, specifically through their interactions with ChatGPT. We identify three distinct user approaches – AI testing-oriented, AI socializing-oriented, and content exploring-oriented – each revealing unique interests and gaps in understanding. We discuss the need for educational initiatives that address specific misconceptions and cater to diverse interaction approaches. Ultimately, our research informs the design of AI literacy interventions,

such as interactive museum exhibits, aimed at an age group that is critical for cognitive and identity development.

ACKNOWLEDGMENTS

First, we would like to thank our expert participants for their valuable time and insights. We would also like to thank our anonymous reviewers for their thoughtful feedback and detailed comments. This material is based upon work supported by the National Science Foundation under Grant No. #2214463. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] AI4ALL. 2019. AI4ALL. <https://ai-4-all.org/>
- [2] 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. (2019).
- [3] Sue Allen. 2004. Designs for learning: Studying science museum exhibits that do more than entertain. *Science education* 88, S1 (2004), S17–S33. Publisher: Wiley Online Library.
- [4] Sue Allen, Patricia B. Campbell, Lynn D. Dierking, Barbara N. Flagg, Alan J. Friedman, Cecilia Garibay, and David A. Ucko. 2008. Framework for evaluating impacts of informal science education projects. In *Report from a National Science Foundation Workshop. The National Science Foundation, Division of Research on Learning in Formal and Informal Settings*, Vol. 73.
- [5] Computer Science Teachers Association. 2017. Standards for CS Teachers. <https://csteachers.org/teacherstandards/>
- [6] Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. 2023. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. *ArXiv abs/2302.04023* (2023). <https://api.semanticscholar.org/CorpusID:256662612>
- [7] The Barbican. 2019. AI: More than Human | Barbican. <https://www.barbican.org.uk/whats-on/2019/event/ai-more-than-human>
- [8] Leslie Bedford. 2014. *The art of museum exhibitions: How story and imagination create aesthetic experiences*. Left Coast Press.
- [9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. <https://doi.org/10.1191/1478088706qp0630a> Publisher: Routledge _eprint: <https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp0630a>.
- [10] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. <https://doi.org/10.48550/arXiv.2005.14165> arXiv:2005.14165 [cs].
- [11] Amy Bruckman, B DiSalvo, M Guzdail, T Mcklin, C Meadows, K Perry, and C Steward. 2009. Glitch game testers: African American men breaking open the console. *Proc of DiGRA* (2009).
- [12] Lorena Casal-Otero, Alejandro Catala, Carmen Fernández-Morante, Maria Taboada, Beatriz Cebreiro, and Senén Barro. 2023. AI literacy in K-12: a systematic literature review. *International Journal of STEM Education* 10, 1 (April 2023), 29. <https://doi.org/10.1186/s40594-023-00418-7>
- [13] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078* (2013).
- [14] Judy Diamond, Michael Horn, and David H. Uttal. 2016. *Practical Evaluation Guide: Tools for Museums and Other Informal Educational Settings*. Rowman & Littlefield Publishers, Lanham, MD, UNITED STATES. <http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=4442105>
- [15] Stefania Druga. 2018. *Growing up with AI: Cognimates: from coding to teaching machines*. PhD Thesis. Massachusetts Institute of Technology.
- [16] Stefania Druga, Fee Lia Christoph, and Amy J Ko. 2022. Family as a Third Space for AI Literacies: How do children and parents learn about AI together?. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–17. <https://doi.org/10.1145/3491102.3502031>
- [17] Stefania Druga, Sarah T. Vu, Eesh Likhith, and Tammy Qiu. 2019. Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019*. 104–111.
- [18] 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 (IDC '17)*. Association for Computing Machinery, New York, NY, USA, 595–600. <https://doi.org/10.1145/3078072.3084330>
- [19] 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*. 231–240.
- [20] Stefania Druga, Jason Yip, Michael Preston, and Devin Dillon. 2022. The 4As: Ask, Adapt, Author, Analyze. (2022).
- [21] Reinders Duit. 1991. On the role of analogies and metaphors in learning science. *Science education* 75, 6 (1991), 649–672. Publisher: Hoboken.
- [22] Upol Ehsan, Samir Passi, Q. Vera Liao, Larry Chan, I. Lee, Michael Muller, and Mark O. Riedl. 2021. The who in explainable ai: How ai background shapes perceptions of ai explanations. *arXiv preprint arXiv:2107.13509* (2021).
- [23] Glenn Ellis, Jake Lauer, Katie Silva, and Netty Nina. 2007. Assessing High School Girls' Preconceptions About Artificial Intelligence To Improve Learning. 12.267.15. <https://peer.asee.org/assessing-high-school-girls-preconceptions-about-artificial-intelligence-to-improve-learning> ISSN: 2153-5965.
- [24] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300724>
- [25] Ignacio Evangelista, Germán Blesio, and Emanuel Benatti. 2018. Why Are We Not Teaching Machine Learning at High School? A Proposal. In *2018 World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC)*. 1–6. <https://doi.org/10.1109/WEEF-GEDC.2018.8629750>
- [26] J. Falk and L. Dierking. 2000. Learning from Museums: Visitor Experiences and the Making of Meaning. <https://www.semanticscholar.org/paper/Learning-from-Museums%3A-Visitor-Experiences-and-the-Falk-Dierking/bf8ae5bf45677e88df2d853d510bb82d47945cbd>
- [27] John Falk and Martin Storksdieck. 2005. Using the contextual model of learning to understand visitor learning from a science center exhibition. *Science Education* 89, 5 (2005), 744–778. <https://doi.org/10.1002/sce.20078> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sce.20078>.
- [28] Alexander J Fiannaca, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. Programming without a Programming Language: Challenges and Opportunities for Designing Developer Tools for Prompt Programming. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.
- [29] S. Forsyth, B. Walsh, B. Dalton, E. Haberl, J. Smilack, and T. Yeh. 2021. Imagine a more ethical AI: Using stories to develop teens' awareness and understanding of artificial intelligence and its societal impacts. *RESPECT 2021 virtual conference; Annual Conference on Research in Equity and Sustained Participation in Engineering, Computing, and Technology* (Jan. 2021). <https://par.nsf.gov/biblio/10292023-imagine-more-ethical-ai-using-stories-develop-teens-awareness-understanding-artificial-intelligence-its-societal-impacts>
- [30] L Friedman, NB Black, E Walker, and J Roschelle. 2021. Safe AI in Education Needs You. <https://cacm.acm.org/blogs/blog-cacm/256657-safe-ai-in-education-needs-you/fulltext>
- [31] Mafalda Gamboa. 2022. Living with Drones, Robots, and Young Children: Informing Research through Design with Autoethnography. In *NordicCHI '22*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3546155.3546658>
- [32] Natalie Garrett, Nathan Beard, and Casey Fiesler. 2020. More Than "If Time Allows": The Role of Ethics in AI Education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AI/ES '20)*. Association for Computing Machinery, New York, NY, USA, 272–278. <https://doi.org/10.1145/3375627.3375868>
- [33] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–15. <https://doi.org/10.1145/3544548.3580782>
- [34] Shawn M. Glynn, Bruce K. Britton, Margaret Semrud-Clikeman, and K. Denise Muth. 1989. Analogical reasoning and problem solving in science textbooks. *Handbook of creativity* (1989), 383–398. Publisher: Springer.
- [35] Thomas Humphrey Gutwill, Joshua P. (Ed.). 2016. *Fostering Active Prolonged Engagement: The Art of Creating APE Exhibits*. Routledge, New York. <https://doi.org/10.4324/9781315428291>
- [36] Clint Heinze, Janet Haase, and Helen Higgins. 2010. An action research report from a multi-year approach to teaching artificial intelligence at the k-6 level. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 24. 1890–1895. Issue: 3.
- [37] Jon Hindmarsh, Christian Heath, Dirk Vom Lehn, and Jason Cleverly. 2005. Creating assemblies in public environments: Social interaction, interactive exhibits and CSCW. *Computer Supported Cooperative Work (CSCW)* 14 (2005), 1–41. Publisher: Springer.

- [38] Tom Hitron, Iddo Wald, Hadas Erel, and Oren Zuckerman. 2018. Introducing children to machine learning concepts through hands-on experience. In *Proceedings of the 17th ACM Conference on Interaction Design and Children (IDC '18)*. Association for Computing Machinery, New York, NY, USA, 563–568. <https://doi.org/10.1145/3202185.3210776>
- [39] Michael S. Horn, Erin Treacy Solovey, and Robert J. K. Jacob. 2008. Tangible programming and informal science learning: making TUIs work for museums. In *Proceedings of the 7th international conference on Interaction design and children (IDC '08)*. Association for Computing Machinery, New York, NY, USA, 194–201. <https://doi.org/10.1145/1463689.1463756>
- [40] Eva Hornecker. 2005. A Design Theme for Tangible Interaction: Embodied Facilitation. In *ECSCW 2005*, Hans Gellersen, Kjeld Schmidt, Michel Beaudouin-Lafon, and Wendy Mackay (Eds.). Springer Netherlands, Dordrecht, 23–43. https://doi.org/10.1007/1-4020-4023-7_2
- [41] Eva Hornecker and Jacob Buur. 2006. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 437–446. <https://doi.org/10.1145/1124772.1124838>
- [42] Thomas Humphrey and Joshua P Gutwill. 2017. *Fostering active prolonged engagement: The art of creating APE exhibits*. Routledge.
- [43] Ravi Krishnan Jagannathan and Claire Komives. 2020. Teaching by Induction: Project-Based Learning for Silicon Valley. *Journal of Engineering Education Transformations* 33, Special 1 (Feb. 2020). <https://doi.org/10.16920/jeet/2019/v33i1/149003> Publisher: Rajarambapu Institute Of Technology.
- [44] N. Joseph and J. O. Thomas. 2020. Designing STEM Learning Environments to Support Middle School Black Girls' Computational Algorithmic Thinking: A Possibility Model for Disrupting STEM Neoliberal Projects. *International Conference of the Learning Sciences* (Jan. 2020). <https://par.nsf.gov/biblio/10213069-designing-stem-learning-environments-support-middle-school-black-girls-computational-algorithmic-thinking-possibility-model-disrupting-stem-neoliberal-projects>
- [45] Ken Kahn, R. Megasari, E. Piantari, and E. Junaeti. 2018. AI Programming by Children using Snap! Block Programming in a Developing Country. <https://www.semanticscholar.org/paper/AI-Programming-by-Children-using-Snap!-Block-in-a-Kahn-Megasari/19eaf4e0378e2da0797827c87c52d11e738ff173>
- [46] Maria Kallery and Dimitris Psillos. 2004. Anthropomorphism and Animism in Early Years Science: Why Teachers Use Them, how They Conceptualise Them and What Are Their Views on Their Use. *Research in Science Education* 34, 3 (Sept. 2004), 291–311. <https://doi.org/10.1023/B:RISE.0000044613.64634.03>
- [47] Anna Kawakami, Luke Guerdan, Yang Cheng, Anita Sun, Alison Hu, Kate Glazko, Nikos Arechiga, Matthew Lee, Scott Carter, Haiyi Zhu, and Kenneth Holstein. 2022. Towards a Learner-Centered Explainable AI: Lessons from the learning sciences. <http://arxiv.org/abs/2212.05588> arXiv:2212.05588 [cs].
- [48] Markelle Kelly, Aakriti Kumar, Padhraic Smyth, and Mark Steyvers. 2023. Capturing Humans' Mental Models of AI: An Item Response Theory Approach. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FACT '23)*. Association for Computing Machinery, New York, NY, USA, 1723–1734. <https://doi.org/10.1145/3593013.3594111>
- [49] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T. Hancock, and Michael S. Bernstein. 2020. Conceptual Metaphors Impact Perceptions of Human-AI Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 163:1–163:26. <https://doi.org/10.1145/3415234>
- [50] Adem Korkmaz, Cemal Aktürk, and Tarik Talan. 2023. Analyzing the User's Sentiments of ChatGPT Using Twitter Data. *Iraqi Journal For Computer Science and Mathematics* 4, 2 (May 2023), 202–214. <https://doi.org/10.52866/ijcs.2023.02.02.018> Number: 2.
- [51] Gloria Ladson-Billings. 2021. *Culturally relevant pedagogy: Asking a different question*. Teachers College Press.
- [52] Carol D Lee. 1992. Literacy, cultural diversity, and instruction. *Education and Urban Society* 24, 2 (1992), 279–291.
- [53] Sunok Lee, Dasom Choi, Minha Lee, Jonghak Choi, and Sangsu Lee. 2023. Fostering Youth's Critical Thinking Competency About AI through Exhibition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–22. <https://doi.org/10.1145/3544548.3581159>
- [54] Alfredo J Leon and Dinesh Vidhani. 2023. ChatGPT Needs a Chemistry Tutor Too. *Journal of Chemical Education* 100, 10 (2023), 3859–3865.
- [55] Q. Vera Liao, Muhammed Mas-ud Hussain, Praveen Chandar, Matthew Davis, Yasaman Khazaeni, Marco Patricio Crasso, Dakuo Wang, Michael Muller, N. Sadat Shami, and Werner Geyer. 2018. All work and no play?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [56] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–23. <https://doi.org/10.1145/3491102.3501825>
- [57] Duri Long, Takeria Blunt, and Brian Magerko. 2021. Co-Designing AI Literacy Exhibits for Informal Learning Spaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–35. <https://doi.org/10.1145/3476034>
- [58] 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 (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3313831.3376727>
- [59] Duri Long, Tom McKlin, Nylah Akua Adjei Boone, Dezarae Dean, Mirina Garoufalidis, and Brian Magerko. 2022. Active Prolonged Engagement Expanded (APEX): A Toolkit for Supporting Evidence-Based Iterative Design Decisions for Collaborative, Embodied Museum Exhibits. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–33.
- [60] Duri Long, Anthony Teachey, and Brian Magerko. 2022. Family Learning Talk in AI Literacy Learning Activities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–20. <https://doi.org/10.1145/3491102.3502091>
- [61] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. <https://doi.org/10.1145/3313831.3376739>
- [62] Brian Magerko, Jason Freeman, Tom McKlin, Mike Reilly, Elise Livingston, Scott Mccoid, and Andrea Crews-Brown. 2016. Earsketch: A steam-based approach for underrepresented populations in high school computer science education. *ACM Transactions on Computing Education (TOCE)* 16, 4 (2016), 1–25. Publisher: ACM New York, NY, USA.
- [63] François Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research* 30 (2007), 457–500.
- [64] Reza Hadi Mogavi, Chao Deng, Justin Juho Kim, Pengyuan Zhou, Young D. Kwon, Ahmed Hosny Saleh Metwally, Ahmed Tlili, Simone Bassanelli, Antonio Bucchiarone, Sujit Gujar, Lennart E. Nacke, and Pan Hui. 2023. Exploring User Perspectives on ChatGPT: Applications, Perceptions, and Implications for AI-Integrated Education. <https://doi.org/10.48550/arXiv.2305.13114> arXiv:2305.13114 [cs].
- [65] Ha Nguyen. 2022. Examining Teenagers' Perceptions of Conversational Agents in Learning Settings. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference (IDC '22)*. Association for Computing Machinery, New York, NY, USA, 374–381. <https://doi.org/10.1145/3501712.3529740>
- [66] Boston Museum of Science. 2022. Exploring AI: Making the Invisible Visible. <https://www.mos.org/exhibits/exploring-ai>
- [67] Museum of Science and Industry (MSI). 2018. Robot Revolution. <https://www.msichicago.org/explore/whats-here/exhibits/robot-revolution/>
- [68] Nichole Pinkard. 2001. Rappin' Reader and Say Say Oh Playmate: Using children's childhood songs as literacy scaffolds in computer-based learning environments. *Journal of Educational Computing Research* 25, 1 (2001), 17–34.
- [69] James P. Steyer. 2023. When it Comes to AI, Let's Move Fast and Fix Things. <https://time.com/6296522/kids-ai-risks/>
- [70] Jack W. Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, Eliza Rutherford, Tom Hennigan, Jacob Menick, Albin Cassirer, Richard Powell, George van den Driessche, Lisa Anne Hendricks, Maribeth Rauh, Po-Sen Huang, Amelia Glaese, Johannes Welbl, Sumanth Dathathri, Saffron Huang, Jonathan Uesato, John Mellor, Irina Higgins, Antonia Creswell, Nat McAleese, Amy Wu, Erich Elsen, Siddhant Jayakumar, Elena Buchatskaya, David Budden, Esme Sutherland, Karen Simonyan, Michela Paganini, Laurent Sifre, Lena Martens, Xiang Lorraine Li, Adhiguna Kuncoro, Aida Nematzadeh, Elena Gribovskaya, Domenic Donato, Angeliki Lazaridou, Arthur Mensch, Jean-Baptiste Lespiau, Maria Tsimpoukelli, Nikolai Grigorev, Doug Fritz, Thibault Sottiaux, Mantas Pajarskas, Toby Pohlen, Zhitao Gong, Daniel Toyama, Cyprien de Masson d'Aultume, Yujia Li, Tayfun Terzi, Vladimir Mikulic, Igor Babuschkin, Aidan Clark, Diego de Las Casas, Aurelia Guy, Chris Jones, James Bradbury, Matthew Johnson, Blake Hechtman, Laura Weidinger, Iason Gabriel, William Isaac, Ed Lockhart, Simon Osindero, Laura Rimell, Chris Dyer, Oriol Vinyals, Kareem Ayoub, Jeff Stanway, Lorraine Bennett, Demis Hassabis, Koray Kavukcuoglu, and Geoffrey Irving. 2022. Scaling Language Models: Methods, Analysis & Insights from Training Gopher. <https://doi.org/10.48550/arXiv.2112.11446> arXiv:2112.11446 [cs].
- [71] Amon Rapp, Arianna Boldi, Lorenzo Curti, Alessandro Perrucci, and Rossana Simeoni. 2023. How Do People Ascribe Humanness to Chatbots? An Analysis of Real-World Human-Agent Interactions and a Theoretical Model of Humanness. *International Journal of Human-Computer Interaction* (2023), 1–24. Publisher: Taylor & Francis.
- [72] ReadyAI. 2020. ReadyAI | Empowering all students to improve our world with AI. <https://www.readyai.org/>
- [73] ReadyAI. 2021. ReadyAI | Empowering all students to improve our world with AI. <https://www.readyai.org/>
- [74] Jessica Roberts and Leilah Lyons. 2017. Scoring Qualitative Informal Learning Dialogue: The SQuLLD Method for Measuring Museum Learning Talk. (2017).

- [75] Mark Rosin, Jen Wong, Kari O'Connell, Martin Storksdieck, and Brianna Keys. 2021. Guerilla science: mixing science with art, music and play in unusual settings. *Leonardo* 54, 2 (2021), 191–195.
- [76] Elisa Rubegni, Laura Malinverni, and Jason Yip. 2022. "Don't let the robots walk our dogs, but it's ok for them to do our homework": children's perceptions, fears, and hopes in social robots.. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference (IDC '22)*. Association for Computing Machinery, New York, NY, USA, 352–361. <https://doi.org/10.1145/3501712.3529726>
- [77] Marie-Monique Schaper, Mariana Aki Tamashiro, Rachel Charlotte Smith, Maarten Van Mechelen, and Ole Sejer Iversen. 2023. Five Design Recommendations for Teaching Teenagers' about Artificial Intelligence and Machine Learning. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (IDC '23)*. Association for Computing Machinery, New York, NY, USA, 298–309. <https://doi.org/10.1145/3585088.3589366>
- [78] Kimberly A. Scott and Mary Aleta White. 2013. COMPUGIRLS' Standpoint: Culturally Responsive Computing and Its Effect on Girls of Color. *Urban Education* 48, 5 (Sept. 2013), 657–681. <https://doi.org/10.1177/0042085913491219> Publisher: SAGE Publications Inc.
- [79] Marita Skjuve, Asbjørn Følstad, and Petter Bae Brandtzaeg. 2023. The User Experience of ChatGPT: Findings from a Questionnaire Study of Early Users. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. ACM, Eindhoven Netherlands, 1–10. <https://doi.org/10.1145/3571884.3597144>
- [80] Marcin Skowron, Stefan Rank, Mathias Theunis, and Julian Sienkiewicz. 2011. The good, the bad and the neutral: affective profile in dialog system-user communication. In *Affective Computing and Intelligent Interaction: 4th International Conference, ACII 2011, Memphis, TN, USA, October 9–12, 2011, Proceedings, Part I* 4. Springer, 337–346.
- [81] Sarin Sok and Kimkong Heng. 2023. ChatGPT for education and research: A review of benefits and risks. *Available at SSRN 4378735* (2023).
- [82] Jaemarie Solyst, Shixian Xie, Ellia Yang, Angela E.B. Stewart, Motahhare Eslami, Jessica Hammer, and Amy Ogan. 2023. "I Would Like to Design": Black Girls Analyzing and Ideating Fair and Accountable AI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3544548.3581378>
- [83] A Subaveerapandiyan, C Sunanthini, and Mohammad Amees. 2023. A study on the knowledge and perception of artificial intelligence. *IFLA Journal* (June 2023), 03400352231180230. <https://doi.org/10.1177/03400352231180230> Publisher: SAGE Publications Ltd.
- [84] Karen Sullenger. 2006. Beyond School Walls: Informal Education and the Culture of Science. *Education Canada* 46, 3 (2006), 15–18. Publisher: ERIC.
- [85] Ann Swidler. 1986. Culture in action: Symbols and strategies. *American sociological review* (1986), 273–286.
- [86] Rivka Taub, Michal Armoni, and Mordechai Ben-Ari. 2012. CS Unplugged and Middle-School Students' Views, Attitudes, and Intentions Regarding CS. *ACM Transactions on Computing Education* 12, 2 (April 2012), 1–29. <https://doi.org/10.1145/2160547.2160551>
- [87] 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 (July 2019), 9795–9799. <https://doi.org/10.1609/aaai.v33i01.33019795> Number: 01.
- [88] Jessica Van Brummelen, Viktoriya Tabunshchik, and Tommy Heng. 2021. "Alexa, Can I Program You?": Student Perceptions of Conversational Artificial Intelligence Before and After Programming Alexa. In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference (IDC '21)*. Association for Computing Machinery, New York, NY, USA, 305–313. <https://doi.org/10.1145/3459990.3460730>
- [89] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3411764.3445645>
- [90] Christiane Gresse von Wangenheim, Livia S. Marques, and Jean C. R. Hauck. 2020. Machine Learning for All – Introducing Machine Learning in K-12. <https://doi.org/10.31235/osf.io/wj5ne>
- [91] Henry M. Wellman, David Cross, and Julianne Watson. 2001. Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development* 72, 3 (2001), 655–684. <https://doi.org/10.1111/1467-8624.00304> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8624.00304>
- [92] Randi Williams, Christian Vázquez Machado, Stefania Druga, Cynthia Breazeal, and Pattie Maes. 2018. "My doll says it's ok" a study of children's conformity to a talking doll. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. 625–631.
- [93] Julia Woodward, Zari McFadden, Nicole Shiver, Amir Ben-hayon, Jason C. Yip, and Lisa Anthony. 2018. Using Co-Design to Examine How Children Conceptualize Intelligent Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3173574.3174149>
- [94] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–21. <https://doi.org/10.1145/3544548.3581388>
- [95] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [96] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K. Kane, and R. Benjamin Shapiro. 2019. Youth Learning Machine Learning through Building Models of Athletic Moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children (IDC '19)*. Association for Computing Machinery, New York, NY, USA, 121–132. <https://doi.org/10.1145/3311927.3323139>