The Role of Collaboration, Creativity, and Embodiment in AI Learning Experiences

DURI LONG

Expressive Machinery Lab, Georgia Institute of Technology, duri@gatech.edu

AADARSH PADIYATH

Expressive Machinery Lab, Georgia Institute of Technology, aspadiyath@gatech.edu

ANTHONY TEACHEY, BRIAN MAGERKO

Expressive Machinery Lab, Georgia Institute of Technology, {ateachey3, magerko}@gatech.edu

Fostering public *AI literacy* (i.e. a high-level understanding of artificial intelligence (AI) that allows individuals to critically and effectively use AI technologies) is increasingly important as AI is integrated into individuals' everyday lives and as concerns about AI grow. This paper investigates how to design collaborative, creative, and embodied interactions that foster AI learning and interest development. We designed three prototypes of collaborative, creative, and/or embodied learning experiences that aim to communicate AI literacy competencies. We present the design of these prototypes as well as the results from a user study that we conducted with 14 family groups (38 participants). Our data analysis explores how collaboration, creativity, and embodiment contributed to AI learning and interest development across the three prototypes. The main contributions of this paper are: 1) three designs of AI literacy learning activities and 2) insights into the role creativity, collaboration, and embodiment play in AI learning experiences.

CCS CONCEPTS • Social and professional topics~Professional topics~Computing education~Informal education • Human-centered computing~Interaction design~Interaction design process and methods~Interface design prototyping • Human-centered computing~Human computer interaction (HCI)~Empirical studies in HCI

Additional Keywords and Phrases: AI literacy, AI education, family learning, informal learning, co-creative, collaboration, creativity, embodiment

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages.

1 INTRODUCTION

Broadening access to learning experiences about artificial intelligence (AI) is increasingly important as AI becomes more integrated into our everyday lives. Individuals who have little knowledge of AI or how it works are engaging with an increasing number of commercially available AI devices and technologies. Growing concerns about AI's role in misinformation [3,37], data privacy breaches [43], and bias/discrimination [7] suggest that technology users need new skills to be able to engage with AI critically and thoughtfully. This skillset has been referred to in the literature as *AI literacy* (i.e. "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" [34]).

There has been a recent surge in AI learning interventions for individuals without a computing background, with a focus on K-12 audiences (c.f. [34,53]). However, few efforts to date have focused on museums as venues for AI education, even though museums and science centers have historically played an important role in public science education [45]. We are investigating how to provide learning experiences in museums that can foster public AI literacy, both by developing novel AI education exhibits and by adapting existing AI research projects into educational experiences through which learners have the opportunity to interact with authentic cutting-edge research in the field.

Prior work suggests that certain design features—*collaboration*, *creativity,* and *embodied interaction*—can help facilitate effective learning experiences in museums. *Embodied* interaction is an intuitive way to engage with exhibits and can aid in concretizing abstract concepts [23–25,31,41,44]. We explore designs that utilize full-body interaction, tangible user interfaces, and spatial metaphors. Most visitors come to museums in groups [21], making *collaboration* an important part of the museum experience that contributes to learning and motivation [13,25,26]. We define collaboration in this paper as encompassing both shared dialogue and working together to achieve a shared goal. Finally, *creative* interactions have been shown to contribute to prolonged engagement at exhibits and can lead to personally-relevant meaning-making [4,18,26]. In this paper, we use the term *creativity* to refer to designs that encourage learners to generate personally creative (i.e. P-creative, or novel to the individual [5]) ideas by expressing themselves through activities like dance or generating novel artifacts and combinations of ideas. Research suggests that embodied interaction, collaboration, and creativity may also be effective at facilitating learning about computing [10,15,20,35,46,54].

Our hypothesis is that interactions with embodied, collaborative, and/or creative AI learning activities in informal learning spaces lead to interest development in AI and improved understanding of AI. In this paper, we test this hypothesis by exploring two core research questions: **1) How can embodiment, collaboration, and creativity be used in museum exhibits to encourage interest development in and learning about AI?** and **2) What design features contribute to engagement with activities that increase interest in and improve understanding of AI in informal learning spaces?** To address the first question, we designed three AI literacy exhibits—*Knowledge Net, Creature Features,* and *LuminAI*—that each incorporate collaboration, creativity, and/or embodied interaction to varying degrees. We explore how to develop AI literacy exhibits "from scratch" with *Knowledge Net* and *Creature Features.* The third exhibit *(LuminAI)* explores how to adapt and augment an existing AI research project to facilitate an educational experience about/with authentic AI technology. To investigate the second research question, we conducted remote user studies with 14 family groups (38 participants) with two study sessions. The first session of users engaged with an early iteration of our prototypes, and the second session interacted with a later iteration of the same prototypes. We present results from these studies, focusing primarily on an analysis of participant surveys (supplemented with qualitative observations when relevant). Our analysis assesses the degree to which the exhibits supported collaboration, creativity, and embodied interaction, and explores the relationship of these design features to AI learning and interest development.

2 RELATED WORK

There is a growing body of research investigating how to design AI-related learning experiences for novice audiences. Researchers are developing curricula for both K-12 audiences [2,48,50] and non-CS majors at universities [6,19,46]. Others are developing courses, interactive online tools, and programming platforms that can engage novice audiences in learning about AI (e.g. [1,15,30,54]). The exhibit designs presented in this paper are grounded in two recently published frameworks related to AI literacy. The first framework presents five "big ideas" that define areas of AI that are important for K-12 audiences to understand: 1) perception; 2) representation and reasoning; 3) learning; 4) natural interaction; and 5) societal impact [48]. The second framework is a set of AI literacy competencies and design considerations we developed based on a review of AI education literature [34]. The competencies are high-level ideas about AI intended for novice audiences, and the design considerations are intended to guide the development of AI literacy learning interventions. We used both of these frameworks to guide the design of the prototypes presented in this paper.

In the remainder of this section, we review several AI education projects that emphasize our key design considerations—collaboration, creativity, and embodied interaction. There are numerous existing platforms that are designed to engage learners in **creative** programming activities involving AI. *Cognimates* is an add-on for the Scratch programming environment that allows learners to incorporate AI technologies like image or voice recognition in their Scratch programs [14]. Similar tools exist for other coding platforms (e.g. [1,28,51]), allowing learners to incorporate AI in their creative multimedia projects. Others have developed activities to engage learners in creatively imagining alternative AI futures—like an AI ethics activity that engages middle school students in redesigning YouTube [2]. A recent paper outlined a set of design principles for introducing co-creative AI research projects in public spaces—while not explicitly focused on AI education, we draw on several of these principles in our work [33]. Research also suggests that having learners enact **embodied** simulations of algorithms (either on their own or by programming an embodied AI device [15,49]) can help them to concretize abstract concepts [15,46]. Other platforms engage learners in building machine learning (ML) models of physical gestures like dance or sports moves [8,54]. There are fewer existing projects that are focused on **collaboration**. However, recent papers suggest that facilitating social dialogue, particularly between adults and children, is important in AI learning contexts [16,34,50]. AI plugins on platforms like Scratch also facilitate social learning by allowing learners to share their work with a wide audience and "remix" others' projects [40].

3 AI LITERACY PROTOTYPES

In this section, we describe the iterative prototyping of three exhibits—*Knowledge Net*, *Creature Features*, and *LuminAI*. We provide a description of each exhibit, followed by a reflection on issues and successes with each prototype iteration. Exhibits were developed as "box-sized" versions of a real museum exhibit so they could be easily delivered to families' homes for user-testing during COVID-19[1]. We focused on designing to communicate AI competencies related to the third "big idea" of AI: "Agents maintain models/representations of the world and use them for reasoning," because we found that this "big idea" was one of the most under-explored in existing work [48]. We were interested in particular in how concepts related to AI representations/reasoning could be communicated without requiring prerequisite coding knowledge, which can be a barrier to entry. We included perspectives on both machine learning and knowledge-based AI. Throughout this section, we refer to the AI literacy competencies (hereafter, C) and design considerations (hereafter, DC) from [34] to ground our design research in a theoretical framework. We mention results from our user studies in this section to explain the iterative design process, but discuss the majority of our findings in Results.

3.1 Knowledge Net

Knowledge Net (Figure 1, left) is an exhibit prototype in which learners can use a tangible interface consisting of tiles and arrows to build *semantic networks* (a type of AI knowledge representation that is used to represent relationships

[1] We intended to evaluate exhibits by installing them as pop-up installations at the Museum of Science and Industry, Chicago (MSI). In March 2020, it became known that COVID-19 was circulating widely in the US. MSI was shut down and safety concerns arose regarding researcher travel and in-person user studies. We pivoted to designing "exhibits-in-a-box": at-home learning experiences delivered to families' doorsteps for them to interact with. We chose this method because embodiment and collaboration were central design considerations that were not easy to transfer to a virtual experience.

between objects and ideas) about topics of interest to them (e.g. family, animals, music). Learners connect object tiles (e.g. dog, cat, whiskers) with relationship arrows (e.g. is, has, likes dislikes) on a playmat to create a network (e.g. dog-HAS-whiskers, cat-LIKES-dog). Once learners build their network, they can take a picture of their playmat, upload it to our website, and interact with an AI chatbot that uses their network as its knowledge base. Our algorithm parses the image by matching user placed tiles and cards with template images. The template with the highest similarity to a user-placed tile is assumed to be the correct tile. After parsing the board, our program then input that data into a semantic network representation (adapted from [42]). Users can proceed to ask the network questions about the relationships between the tiles (e.g. "What does a dog have?). This input is matched to simple question templates (e.g. "What does a ____have?") to tell which relationship a user is querying. If a user's query is answerable, an answer template was used to respond (e.g. "A ___ has ___"). Learners interacting with *Knowledge Net* can iteratively explore and test ideas related to understanding the strengths and limitations of knowledge representations (C5, C7), recognizing the role that humans play in programming AI (C10), and understanding computer reasoning processes (C8). *Knowledge Net* incorporates interaction with a tangible interface (DC2) and open-ended creative, collaborative interactions (DC11), as well as other AI literacy design considerations such as making algorithms explainable (DC1), providing opportunities for individuals to program/teach AI (DC6), leveraging learners' interests (DC12), and facilitating a low barrier of entry (DC15).

When we user-tested the first iteration of the *Knowledge Net* exhibit prototype, we observed that the exhibit was successful at engaging learners of all ages and varying levels of prior experience with AI (DC15). The semantic network representation was intuitive for participants to understand, and learners enjoyed customizing the networks to describe topics of interest to them (e.g. their family) (DC12). The exhibit was also successful at facilitating collaborative dialogue and interaction between group members (DC11). However, the first iteration of the design suffered from several issues. Our image recognition algorithm failed or worked poorly in non-ideal conditions (e.g. bad lighting, poorly cropped picture), requiring users to manually input almost all of the network information and preventing most from interacting with the chatbot. We instructed learners who were unable to interact with the chatbot to engage in a role-playing activity where they simulated a conversation between an AI chatbot and a human user. We also asked learners to photograph the playmat with their phones, which resulted in a cumbersome uploading/cropping process. These issues collectively prevented learners from engaging in an iterative exploration/testing cycle and making connections between the tangible interface and the virtual chatbot. We aimed to resolve these issues in the second iteration of the prototype by changing the material design of the tangible interface to aid in image recognition and replicability and introducing an Osmo device and iPad for photographing the playmat to aid in image recognition and an iterative process.

3.2 Creature Features

Creature Features (Figure 1, center) is an exhibit in which learners can use a card deck and "weight tokens" to provide training data to a feature-based machine learning algorithm that classifies animals as birds. Each card depicts a creature (e.g. bluebird, flying fish) and includes a list of features describing that creature (e.g. swims, has feathers). Learners are encouraged to look at the features for each creature on the playmat and consider how to place their weights to train an algorithm that can correctly recognize many different types of birds. The more weights tokens placed on a card, the more examples of that bird will be included in the dataset. Learners can take a picture of their playmat and upload it to a website. Once the photo is uploaded, our algorithm creates a training set by including each of the cards the amount of times the learner indicated with the respective weight tokens. The attributes of each element in the training set are averaged together to form a linear classifier, and all the possible cards are classified. Results are then shown to the learners, who are encouraged to iterate on their dataset. This exhibit aims to help learners understand some of the steps and practices of machine learning (C9), explore different ways that agents represent knowledge (C7) and make decisions (C8), and engage in data curation and interpretation (C11, C12). It incorporates embodied interaction and metaphors (DC2) and facilitates collaborative discussion between group members (DC11), as well as other design considerations such as providing opportunities to program or teach AI (DC6) and creating explainable algorithms (DC1).

User testing of the first iteration of *Creature Features* indicated that the exhibit was easy for novice users to understand (DC15) and engaged learners in collaborative dialogue (DC11), although it was not as engaging for our youngest users (ages 6-8). Some learners indicated confusion over how the weight tokens affected the algorithm and wanted more explanation (DC1) of both the tokens and the algorithm's results (so they could better iterate on their dataset). The tangible card-based interface (DC2) worked well and learners were able to more easily connect it with the algorithmic output than they were in *Knowledge Net* due to fewer image recognition issues. However, we placed the creatures' features on the back of the cards, and learners tended to not turn over the cards, limiting discussion of the features that the algorithm was using to make decisions. In addition, the process of user phone photography was cumbersome (as in *Knowledge Net)*. In the second iteration, we added an Osmo device and iPad for photo capturing and experimented with material design to improve image recognition (as in *Knowledge Net*). We also moved the features to the front of the cards to make them more readily apparent. To make room for this, we moved the weight token spots to the gameboard and added writing to the board to better explain the purpose of the tokens and emphasize that learners were constructing a dataset. We also created space for both a positive training dataset (birds) and a negative training dataset (non-birds) to provide a more authentic experience of how feature-based machine learning algorithms work.

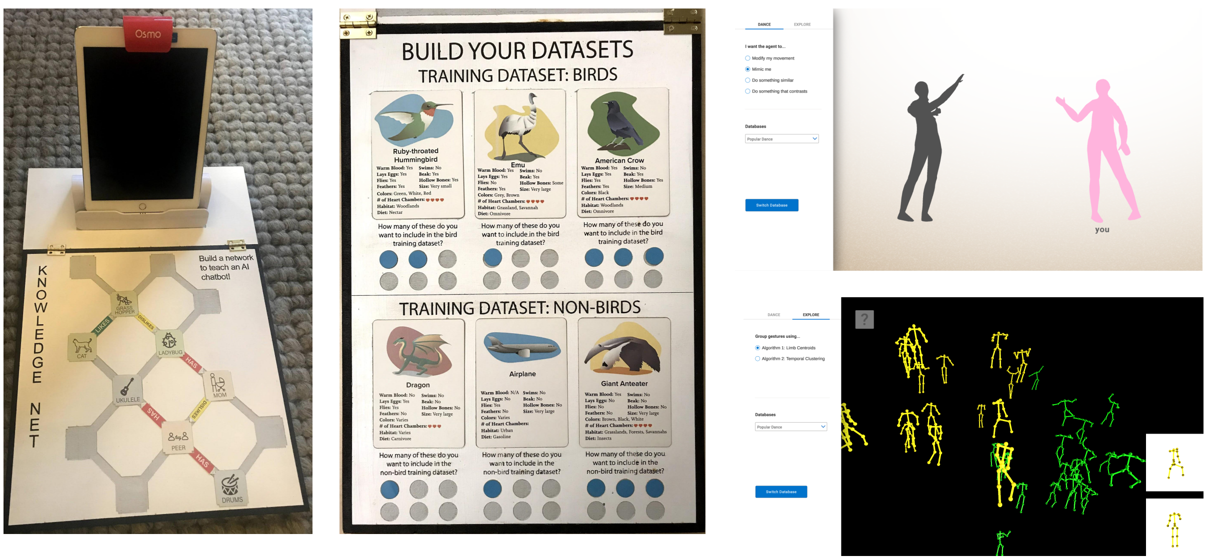

Figure 1: Final prototype designs for (from left to right): *Knowledge Net*, *Creature Features*, and *LuminAI* (dance interface (top) and 3D visualization of agent's clustered gesture memory (bottom))

3.3 LuminAI

The *LuminAI* exhibit builds on an existing AI installation in which participants can improvise movement together with an AI dance partner that is projected onto a screen [33]. In the expanded version of *LuminAI* we developed for this educational context (Figure 1, right), learners can engage with an interactive visual interface to explore different aspects of the dancer's decision-making processes and memory, such as manipulating the dancer's response modes (mimicry, transforming a gesture, performing a gesture from memory that is similar or contrasting to the observed gesture), switching between different databases of dance gestures (e.g. ballet, popular dance), and exploring a 3D visual representation of the way the dancer uses unsupervised learning to cluster (i.e. group) gestures in memory (Figure 1, right bottom). In the "boxed" version of the exhibit, learners could engage with this interface using a laptop and Microsoft Kinect motion sensor set up on a tripod. The exhibit aims to communicate AI-related competencies such as ways of representing knowledge (C7), how agents learn and make decisions (C12, C8), and exposing learners to aspects of machine learning (C9). It utilizes embodied interaction, engages learners in open-ended creative interaction, and can facilitate varying degrees of collaboration depending on the installation setup [32]. *LuminAI* also incorporates additional design considerations such as creating explainable algorithms (DC1), providing opportunities to program or teach AI (DC6), incorporating learner interests, and engaging with lesser-known forms of AI (DC14) [34].

We made only minor adjustments to *LuminAI*'s design between iterations because we wanted to gather additional data on participant interaction without drastically changing the prototype (our first user session for *LuminAI* was small). In addition, many of the changes we were considering would be vastly simpler to implement in a museum setting than an at-home environment. For instance, the small laptop screen made it difficult for learners to collaboratively dance and view the screen together (DC11). Figuring out a larger visualization option would be resolved simply in a museum space by projecting visualizations onto walls. We did make some minor adjustments to the interface. In the first iteration, the 3D visual representation displayed spheres which turned into moving gestures as the user moved towards them in 3D space. This feature was originally intended to improve the performance of the gesture display and reduce cognitive load by allowing viewers to focus only on the gestures they were nearest to. However, we observed that numerous users were confused by the spheres and did not recognize that they turned into gestures. As a result, we disabled the spheres. We also made some minor changes to the wording of the text in the UIs, including rephrasing "User dances" as "Your dances" when we observed that several users did not recognize that database as consisting of their own moves.

4 METHODS

We conducted a study session for each prototype iteration. We used several methods to recruit family groups with children ages 6 and up in the Atlanta metro area to ensure participant diversity, including posting on NextDoor, social media, coordinating with Georgia Tech's education outreach program, and reaching out to Girls Who Code chapters. All studies were conducted remotely due to COVID-19. In order to facilitate interaction with physical materials, each family was given a set of three boxes—one containing materials for data collection and two containing prototypes (each family was only given two exhibits in order to keep the total study time to ~two hours). Boxes were disinfected between participant groups per Georgia Tech's Department of Environmental Health and Safety standards.

Adult members of the family completed consent forms, and a researcher called each family at a scheduled time to ask participating children for assent. Participants were given the option to have the researcher stay on the call or for the researcher to hang up and be available as-needed. On-call, the researcher took on the role of an observer, watching quietly and only answering questions when asked so as not to unduly influence the interaction. Participants followed a detailed written instruction packet (included in the supplemental materials). We provided instructions on how to set up and use the exhibit (something that visitors might observe in a museum environment), but we kept explanatory content-related text to a minimum (i.e. the amount that you might find on a sign next to the exhibit). Participants recorded their own data using a provided audio and video recorder. We were able to collect audio/video data for all groups except one (where a device malfunctioned), and all participants over the age of 7 completed surveys after each activity. Survey questions and format were adapted to be age-appropriate (using techniques from [39], see supplemental materials).

In order to test our hypothesis "Interactions with embodied, collaborative, and/or creative AI learning activities in informal learning spaces lead to interest development in AI and improved understanding of AI," we assessed these evaluation questions (EQs): *EQ1) Usability:* Are the exhibits usable?; *EQ2) Creativity:* Do participants engage creatively with the exhibits?; *EQ3) Collaboration:* Do participants collaborate?; *EQ4) Interest:* Do participants demonstrate interest formation [22] in AI?; *EQ5) Learning*: Do participants demonstrate learning of AI literacy competencies?; and *EQ6) Cross-Installation Comparison:* Do any particular exhibits lead to greater interest development, learning gains, or creative engagement? One adult per group was asked for demographic information. We adapted items from various instruments to assess EQs1-4, including Wiebe's user engagement scale [52] (EQ1), Carroll et al.'s Creativity Support Index [9] (EQ2, EQ3), Maltese et al.'s instrument for assessing creativity in makerspaces [36] (EQ2), and Chen et al.'s instrument for assessing situational interest development [11] (EQ4) (**Error! Reference source not found.**). We supplemented these Likert-type questions with free response questions to provide qualitative insight into user engagement and interest development. Results from all Likert-type survey items are reported using appropriate statistics for ordinal data (e.g. median (*Mdn*), quartiles (*Q1, Q3*), interquartile range (*IQR*)). Some items were left blank by participants and certain items were only asked to participants ages 10+, so we specify the number of respondents for all results. Certain items assessed the same construct but were phrased slightly differently for different ages—we group results from these items together when reporting.

We assessed learning (EQ5) using two different metrics. First, we used a retrospective pre/post survey [29,38] to ask participants to reflect on how their understanding of particular concepts changed before and after their interaction with the museum installation. These pre/post questions assessed learners' self-efficacy changes (i.e. how much do they *think* they know about a topic before vs. after the activity). We also asked participants content-related post-interaction questions to assess their knowledge of certain concepts after they completed the activity. Multiple choice questions were scored by one grader, and free response questions were each scored by two graders (using a rubric, see supplemental material). A third analyst resolved discrepancies. We conducted a cross-installation comparison of survey results to determine whether any of the installation designs support learning more effectively than others (EQ6).

5 RESULTS

We recruited a total of 14 family groups (38 participants; 21 children ages 6 and up and 17 adults) to interact with the exhibit prototypes. Eight groups (n=22) interacted with Iteration 1, and six groups (n=16) interacted with Iteration 2. Among the 14 adults who answered the demographic questions, nine identified as White/Caucasian, four as African American, two as Asian American, and one as other Latin American (two participants were biracial). Most adults reported having at least a 4-year degree (79%). Among the children, 10% were 6 years old, 30% were 7-9 years old, 50% were 10-14, and 10% were 15. 60% of children identified as female and 40% as male. We also asked families about their prior experience with computing. Most adults considered their children to have some prior experience with computers (70%) and AI (60%). Most adults also reported that they worked with computers a lot or sometimes (79%) and had some prior experience interacting with AI (79%) but did not write code or program AI (93%). Unless noted, all results reported in this section are for all participants from both study sessions. This section is organized by evaluation question.

5.1 EQ1: Usability

All exhibits had an average *hold time* (i.e. time spent interacting with the exhibit) of 27-30 minutes. This is likely not indicative of the hold time in an actual museum environment, which has many more distractions. However, the length of the at-home hold time indicates that these activities have the potential to foster prolonged engagement and have a "high ceiling" for interaction. Hold times were similar across iterations, except for *Creature Features*, which saw an increase in average hold time from 19:06 to 39:05 from iteration 1 to iteration 2. We hypothesized that this was due to the changes that made it easier for participants to iterate on their dataset design. However, we found that a similar number of groups iterated on their datasets (3/5 with Iteration 1 vs. 4/5 with Iteration 2). Instead, this increase in hold time occurred because participants spent more time discussing the creatures' features once we moved them to the front side of the cards. The survey questions we asked to assess user engagement were drawn from Wiebe's User Engagement Scale [52] (**Error! Reference source not found.**). Learners generally indicated that they found all three activities enjoyable. Participants indicated that they found the activities easy to understand, but that they could not do some of the things they wanted to do.

Table 1: Participant median scores for survey items. We used a 5-point Likert scale for participants age 15+ and a 3-point scale for ages 10-14 per guidance in [39]. We grouped all scores together by mapping the 3-point scale to scores 1, 3, and 5 on a 5-point scale.

Item	EQs	*Knowledge Net* (n=15)	*Creature Features* (n=12)	*LuminAI* (n=15)
It was easy to understand how to use the activity.	1	*Mdn*=5, *IQR*=1	*Mdn=4.5, IQR=1*	*Mdn=4, IQR=2*
It was easy for me to explore different ideas, outcomes, options or designs when doing the activity.	2	*Mdn=4, IQR=2*	*Mdn=4.5, IQR=1*	*Mdn=3, IQR=2*
What I was able to produce was worth the effort I had to exert to produce it.	2	*Mdn=4, IQR=2*	*Mdn=4, IQR=1*	*Mdn=4, IQR=2*
I was able to be creative while doing the activity.	2	*Mdn=5, IQR=2*	*Mdn=5, IQR=2*	*Mdn=5, IQR=1*
I feel like I created something personally meaningful.	2	*Mdn=3, IQR=3*	*Mdn=3, IQR=1*	*Mdn=3, IQR=0*
I feel like I created something important.	2	*Mdn=3, IQR=3*	*Mdn=3, IQR=2*	*Mdn=3, IQR=1*
I was able to easily work with other people when doing the activity.	2, 3	*Mdn=5, IQR=2*	*Mdn=5, IQR=1*	*Mdn=4, IQR=2*
The activity was interesting.	4	*Mdn=5, IQR=2*	*Mdn=5, IQR=0*	*Mdn=5, IQR=1*
The topic of the activity was new to me.	4	*Mdn=4, IQR=3*	*Mdn=5, IQR=2*	*Mdn=4, IQR=2*
I was focused on the activity.	4	*Mdn=5, IQR=1*	*Mdn=5, IQR=1*	*Mdn=5, IQR=1*
I was so involved in doing the activity that I lost track of time.	4, 2	*Mdn*=3, *IQR*=2.5	*Mdn*=3, *IQR*=1	*Mdn*=3, *IQR*=1
Doing the activity was challenging in a good way	4, 1, 2	*Mdn*=4, *IQR*=2	*Mdn*=5, *IQR*=1	*Mdn*=4, *IQR*=1
The activity was enjoyable.	1, 4	*Mdn*=4, *IQR*=2	*Mdn*=5, *IQR*=1	*Mdn*=5, *IQR*=.5
I could not do some of the things I wanted to do when completing the activity.	1	*Mdn*=3, *IQR*=3	*Mdn*=2.5, *IQR*=2	*Mdn*=3, *IQR*=2
This activity made me think more about AI.	4	*Mdn*=4, *IQR*=2	*Mdn*=4.5, *IQR*=1	*Mdn*=5, *IQR*=0
I would like to do more activities like this one in the future.	4	*Mdn*=5, *IQR*=2	*Mdn*=5, *IQR*=1	*Mdn*=4, *IQR*=2

5.2 EQ2: Creativity

We drew on constructs such as exploration, effort/reward tradeoff, and expressiveness from Caroll et al.'s Creativity Support Index [9] when evaluating creativity (**Error! Reference source not found.**). Participants indicated across all three exhibits that there was an appropriate effort/reward tradeoff in the interaction, and that they felt they were able to be creative. Learners also indicated that they were able to explore a variety of ideas/outcomes during their interactions with *Knowledge Net* and *Creature Features*. Responses were more neutral when it came to exploring a variety of ideas/outcomes with *LuminAI*. We also asked learners if they felt that they had created something important or personally meaningful—responses were neutral across all exhibits. Finally, we asked learners to reflect on their perceptions of AI as creative before vs. after the activity. Learners reported positive changes in median agreement with the statement "I can create things with AI" for all activities (*KN:* Δ*Mdn*=+2, *CF:* Δ*Mdn*=+1.5, *L:* Δ*Mdn*=+2). Despite being situated in a more expressive domain, *LuminAI* did not generate notably larger gains in perceptions of AI as creative when compared to the other two exhibits.

5.3 EQ3: Collaboration

Most participants indicated that they were able to easily collaborate with others during all three activities (**Error! Reference source not found.**). Many participants said in their free-response answers that the aspect of the activity they liked the most was being able to do it together with their family (e.g. "What I like about this activity is the family engagement and my girls view on different things"; "I like that I got to do it with my mom"). We observed that *Knowledge Net* most consistently supported collaboration amongst group members of all ages. Younger participants (6-8 years old) did not appear to be as engaged with *Creature Features* as older group members were ("My child was less interested, I would have been more interested in this on my own"). However, we observed that the exhibit was quite popular with slightly older children (10-11). Since only one person is tracked by the Kinect/AI dancer at one time, learners often engaged with turn-taking with *LuminAI*, which sometimes caused conflict amongst kids who wanted to engage with the dancer simultaneously. Several participants commented that they would have enjoyed the opportunity to dance as a group with the AI. We have explored more social versions of *LuminAI* in larger installation spaces in the past [32]. Learners indicated that they felt they could "collaborate with AI" more after the activities than before. The biggest positive change in scores for this question was for *LuminAI (ΔMdn*=+2, somewhat disagree→somewhat agree*)*, which was expected, since learners were actively co-creating with the AI dancer in *LuminAI* and the other exhibits did not involve human-AI co-creation.

5.4 EQ4: Interest Development

We drew on constructs from Chen et al.'s survey for situational interest [11] (**Error! Reference source not found.**). Participants generally agreed that the exhibits were interesting. Participants found *Creature Features* to be the most novel activity, followed by *LuminAI* and *Knowledge Net.* We had anticipated *LuminAI* would be the most novel activity for participants, but several participants were familiar with the Microsoft Kinect, which may have led them to feel the exhibit as a whole was less novel. Most participants indicated agreement with the statement "The activity was challenging in a good way," suggesting that although it was easy to understand how to interact with the exhibit (EQ1), the activities provided enough challenge to be engaging. Participants generally indicated that they were focused on the activities, though responses were more neutral when we asked them to rate their agreement with the statement "I was so involved in doing the activity that I lost track of time." This may indicate that the exhibits were interesting enough to keep participants' attention, but did not go so far as to absorb them in a state of creative "flow" in most cases [12]. Participants indicated that they would like to do similar activities in the future (*Mdn*=[4,5] for all three activities) and that the activities made them think more about AI (*Mdn*=[4,5] for all three activities). Retrospective questions indicated that learners' interest in wanting to "find out more about AI" (*KN:* Δ*Mdn*=+.5, *CF:* Δ*Mdn*=+1.5, *L:* Δ*Mdn*=0) and "learn to build or program AI" (*KN:* Δ*Mdn*=+2, *CF:* Δ*Mdn*=+2, *L:* Δ*Mdn*=+1) was either high to begin with or increased after the activities, with most substantial increases after *Creature Features.* Across the board, most learners reported interest in AI both before and after the activity (*Mdn*=[4,5] for all three activities), which may be why they self-selected to participate in the study.

5.5 EQ5: Learning

A summary of median knowledge changes according to learners' retrospective pre/post self-reports for each construct is shown in Table 2. There were no constructs with a negative change. The most positive gains were seen for *LuminAI* and *Creature Features*. Particularly notable gains are shown in dark green and bolded in the table. These positive jumps occurred for "types of AI and the differences between them" (*LuminAI*) and "how AI reasons and makes decisions" (*Creature Features*). These results indicate that learners generally felt that they had learned from the activities, although perhaps only moderately in most cases. Of the three activities, the median scores suggest that *Knowledge Net* was the least successful at leading to substantial self-reported learning gains, although the Q1 quartile change indicates that learners with less prior knowledge did report some increase. We saw larger gains for some of the constructs in the Iteration 2 study than Iteration 1. This could have been due to the modifications to the exhibit designs. It could also be because the Iteration 2 participants self-reported on average that they had less prior knowledge of AI than the Iteration 1 participants, leaving them more room to learn. This would be a positive result, indicating that the activities had a low barrier of entry (DC15) and were able to communicate AI concepts to learners with little prior knowledge.

Table 2: Summary of median (*Mdn*) change in self-reported knowledge for participants ages 10 and up. Quartile change (ΔQ1, ΔQ3) is shown where there is no median change. Cells are shaded to indicate the size of the shift, with larger shifts shaded darker green.

Construct	*Knowledge Net* (n=15)	*Creature Features* (n=12)	*LuminAI* (n=15)
Similarities and differences between human and machine intelligence	Δ*Mdn*=0 (Moderate) Δ*Q1*=+1; ΔQ3=+.5	Δ*Mdn* = 0 (Moderate) Δ*Q1*=+1; ΔQ3=+1	Δ*Mdn* = +1 (Low → Moderate)
Types of AI and the differences between them	Δ*Mdn* = +1 (Low → Moderate)	ΔMdn = +1 (Low → Moderate)	**ΔMdn = +1.5 (None-Low → Moderate)**
How computers store and represent knowledge	ΔMdn = 0 (Moderate) Δ*Q1*=+1; ΔQ3=0	N/A	ΔMdn = 0 (Moderate) Δ*Q1*=+1.5; ΔQ3=0
How AI reasons and makes decisions	ΔMdn = 0 (Moderate) Δ*Q1*=+2; ΔQ3=+.5	***ΔMdn* = +1.5 (Low-Moderate → High)**	ΔMdn = +1 (Low → Moderate)
The role that humans play in programming and fine-tuning AI	ΔMdn = 0 (Moderate) Δ*Q1*=+1; ΔQ3=+1	ΔMdn = +1 (Moderate → High)	ΔMdn = +1 (Low → Moderate)
Ethical concerns about AI	ΔMdn = +1 (Low → Moderate)	ΔMdn = +1 (Low → Moderate)	ΔMdn = +1 (Low → Moderate)

Most content knowledge questions were scored on a scale of 0-3—0-inadequate, 1-partial, 2-adequate, 3-excellent (a few questions did not have a "partial" option and two were multiple choice) (see supplemental materials for the rubric for each question). An adequate score indicated that the learner demonstrated the expected level of knowledge. Excellent scores were "above and beyond". Table 3 summarizes participants' scores for each question. Median scores for all of the questions were adequate; participants responses to the *LuminAI* questions were most consistently strong. For *Creature Features*, this indicates that learners were for the most part able to predict what would happen if they placed a lot of weight on a particular card (CFQ1), list items they would include in a training dataset for another context (e.g. self-driving car) (CFQ2), and predict what would happen if they placed weight tokens on a card in the alternative context (CFQ3). For *LuminAI*, this indicates that learners were able to decide what cluster a gesture would be placed in (LQ2), consider the strengths/limitations of *LuminAI*'s representation of the human body (LQ3, LQ4), and transfer their knowledge of the agent's clustering to a new domain (LQ5). For *Knowledge Net*, this indicates that learners were able to consider similarities/differences between human intelligence and the network (KNQ1), explain how the computer would use the network to answer a question (KNQ2), transfer what they learned at the exhibit to a new domain (e.g. grocery shopping) (KNQ3), and consider how to gather data to input into the network (KNQ4, KNQ5).

LuminAI Q1 asked learners to circle all of the things that they noticed the agent doing. Everyone that completed the survey (*n=15*) noticed that *LuminAI* mimicked them, and most people noticed the other response modes (e.g. performing a similar (*n=13*) or contrasting *(n=13*) movement, modifying their movement (*n=11*) and recognizing that the AI learned from them (*n=11*)). This is a significant improvement from earlier studies we have conducted with the non-educational version of *LuminAI* in which participants are just able to dance with the installation without any UI controls over the agent's reasoning capabilities. Participants at previous versions of the installations have often not noticed that the agent is doing anything more novel than mimicking them despite complex reasoning on the backend [27,32].

Table 3: Frequency table of participant scores for content-knowledge questions. Exhibit names are abbreviated as *Knowledge Net* (KN), *Creature Features* (CF), *LuminAI* (L). Cells representing the median score are shaded and the contents are bolded.

Score	KNQ1	KNQ2	KNQ3	KNQ4	KNQ5	CFQ1	CFQ2	CFQ3	LQ2	LQ3	LQ4	LQ5
0 - Inadequate	3	4	8	4	5	6	1	2	0	2	2	3
1 - Partial	6	7	*N/A*	6	1	3	4	6	N/A	N/A	N/A	N/A
2 - Adequate	**15**	**14**	**8**	**11**	**13**	**11**	**10**	**7**	**18**	**15**	**12**	**14**
3 - Excellent	2	1	10	0	2	4	9	3	N/A	1	4	1

6 DISCUSSION

This section examines several key implications of our results as they relate to the use of collaboration, creativity, and embodied interaction in AI literacy learning interventions. Although overall our exhibits were successful at facilitating **collaborative** learning, we noted two main inhibitors to collaborative interaction in AI literacy exhibits—*intimidation* and *age*. AI and computer science more broadly can be intimidating topics to learn about, particularly if learners have negative preconceptions about their technological literacy or whether or not they "belong" in computing [47]. We saw some of this intimidation surface in parental interactions with *LuminAI*. Some users felt overwhelmed by the 3D interactive visualization we had developed to make the AI more "explainable," commenting that "We weren't sure what we were supposed to be learning." Despite these concerns, most participants scored well on the content knowledge questions related to *LuminAI*. This suggests that although they can lead to learning, visual interfaces provided to explain AI algorithms may be intimidating for novice users and require additional scaffolding or a more guided, less exploratory experience to begin with. The response mode interface was less intimidating, indicating making components customizable could be a good approach for making AI explainable.

The other factor that influenced collaboration at exhibits was *age*. *Creature Features* was not particularly engaging for our youngest participants (6-7) but was more engaging for slightly older kids. This suggests that in addition to age-appropriate learning outcomes [48], researchers may consider developing AI literacy design principles specific to particular age bands. We plan to pursue additional user testing to define appropriate age groups for our exhibits. *Knowledge Net* was most successful at facilitating collaboration amongst group members of all ages and levels of experience with AI. We hypothesize that the ability to use prior knowledge during the activity made learning about AI less intimidating—one learner commented that they liked *Knowledge Net* because they could "make relationships between things they already knew." Bolstering learners' confidence by allowing them to succeed at pulling in prior knowledge could enable them to explore activities in which they may otherwise feel less confident or discouraged [17].

We had hypothesized that *LuminAI*, which involved **creativity** to a greater extent than the other two exhibits, would be the most successful at fostering interest development and self-efficacy gains, but there was not a notable difference. We posit that the *lack of a lasting/permanent creative artifact* and the *lack of an iterative exploration/testing/revision cycle* may have played a role in limiting the impact of *LuminAI* on learner interest development/self-efficacy. Prior work has shown that learners often have increased interest/self-efficacy in computing after engaging in creative activities where they create a personally meaningful artifact (e.g. multimedia creation [20], music composition [35]). Learners interacting with *LuminAI* indicated that they did not feel they had created anything important or personally meaningful. Although *LuminAI* allowed learners to creatively express themselves, the ephemeral nature of dance means learners did not generate a lasting artifact. The lack of an artifact may also explain why *LuminAI* (along with the other exhibits) did not go so far as to absorb participants in a state of creative "flow."

We designed *LuminAI* to facilitate open-ended exploration, but learners had neutral feelings about their ability to explore a variety of ideas/outcomes when interacting with *LuminAI*. This could result from the lack of an iterative testing/reflection/revision cycle that was present at the other two exhibits. Recent research on AI education suggests that immediate feedback and opportunities for metacognition play an important role in AI learning experiences [15,53]. *Creature Features* and *Knowledge Net* had an iterative cycle built into the interaction, and learners were encouraged to reflect on their dataset/network after testing it on the computer. *LuminAI* did not have a similar cycle, and incorporating opportunities for metacognition could be an interesting direction for inspiring greater creativity and learning.

Our findings indicate that learners enjoyed the **embodied** nature of the exhibits. Learners commented, "It was a great blend of hands-on and screen time," and "I liked the tactile manipulation of the tiles representing abstract info." Learners also liked seeing their personal movements captured in *LuminAI* and observing how the agent replayed them, modified them, or responded with familiar dance moves. Our findings indicate that the embodied interfaces facilitated a low barrier to entry for AI novices. One participant commented that they "liked how the activity wasn't too complicated, and anyone could do it (as opposed to if we had to code the AI ourselves)." We noted that interaction and discussion time was skewed towards the embodied component of the activity. For instance, in *Knowledge Net,* participants focused on selecting tiles and relationships to construct the network, with less time spent interacting with and discussing the chatbot. Similarly, participants spent significantly more time dancing with *LuminAI* than they did exploring the interactive visualization. This imbalance points to the engaging nature of the embodied interactions but raises additional research questions about how to foster AI learning experiences that span both physical and digital interfaces.

7 LIMITATIONS & THREATS TO VALIDITY

The results reported in this section are not statistically significant due to the small sample size of our population. Study size was limited due to COVID-19 precautions. Participants may have self-selected for the study due to an existing interest in AI/technology. Further studies will need to be conducted to determine whether these findings generalize to a larger population. This study was also not able to probe the *individual* effect of creativity, collaboration, or embodiment on learning/interest development. For example—does collaboration play a larger role than creativity in supporting learning? This would require controlled studies with similar versions of each exhibit where (for example) one version incorporated embodied interaction and one did not.

8 CONCLUSION

In this paper, we present the design of three museum exhibits that aim to foster public AI literacy. These exhibits incorporate embodied interaction, collaboration, and/or creativity as key design features, and they also draw on design considerations outlined in [34]. We present the results from user studies in which we examined family groups' interactions with the exhibits, with a particular focus on understanding learning and interest development in relation to embodiment, creativity, and collaboration. Our findings suggest new considerations for designing collaborative, creative, and/or embodied AI literacy learning interventions.

Acknowledgements

We would like to thank Jonathan Moon, Lucas Liu, Cassandra Naiomi, and Rhea Laroya for their contributions to prototype development; all of our study participants for participating in this study during a challenging time, and our collaborators at the Museum of Science and Industry, Chicago—especially Dr. Aaron Price and Natalie Harris—for their ongoing support of this research. This work was funded by the National Science Foundation (DRL #1612644).

REFERENCES

[1] Adam Agassi, Hadas Erel, Iddo Yehoshua Wald, and Oren Zuckerman. 2019. Scratch Nodes ML: A Playful System for Children if to Create Gesture Recognition Classifiers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.*

[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. In *Proceedings of IJCAI 2019.*

[3] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2: 211–36.

[4] Leslie Bedford. 2014. *The Art of Museum Exhibitions: How story and imagination create aesthetic experiences.* Left Coast Press.

[5] Margaret A Boden. 2004. *The Creative Mind: Myths and Mechanisms.* Routledge, New York, NY, USA.

[6] Jason Borenstein and Ayanna Howard. 2020. Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*: 1–5.

[7] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, 77–91.

[8] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. 2020. Teachable machine: Approachable Web-based tool for exploring machine learning classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8.

[9] Erin A Carroll, Celine Latulipe, Richard Fung, and Michael Terry. 2009. Creativity factor evaluation: towards a standardized survey metric for creativity support. In *Proceedings of the seventh ACM conference on Creativity and cognition*, 127–136.

[10] Patricia Charlton and Stefan Poslad. 2019. Engaging with computer science when solving tangible problems. In *Proceedings of the 3rd Conference on Computing Education Practice*, 1–4.

[11] Ang Chen, Paul W Darst, and Robert P Pangrazi. 1999. What constitutes situational interest? Validating a construct in physical education. *Measurement in Physical Education and Exercise Science* 3, 3: 157–XXX.

[12] M. Czikszentmihalyi. 1996. Flow: The psychology of optimal experience. *Praha: Lidové Noviny.*

[13] Willem Doise, Gabriel Mugny, and Juan-Antonio Pérez. 1998. The social construction of knowledge: Social marking and socio-cognitive conflict. *The psychology of the social* 77.

[14] Stefania Druga. 2018. Growing Up With AI - Cognimates: from coding to teaching machines. MIT.

[15] Stefania Druga, Sarah T.Vu, Eesh Likhith, and Tammy Qiu. 2019. Inclusive AI literacy for kids around the world.

[16] 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.

[17] Carol S Dweck. 2000. *Self-theories: Their Role in Motivation, Personality, and Development.* Psychology Press, Lillington, NC, USA.

[18] John H Falk and Lynn D Dierking. 2000. *Learning from museums: Visitor experiences and the making of meaning.* Altamira Press.

[19] 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*, 272–278.

[20] Mark Guzdial. 2013. Exploring Hypotheses about Media Computation. In *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research*, 19–26.

[21] Christian Heath and Dirk vom Lehn. 2008. Configuring "Interactivity" Enhancing Engagement in Science Centres and Museums. *Social Studies of Science* 38, 1: 63–91.

[22] Suzanne Hidi and K Ann Renninger. 2006. The four-phase model of interest development. *Educational psychologist* 41, 2: 111–127. https://doi.org/10.1207/s15326985ep4102_4

[23] Michael S. Horn, Erin Treacy Solovey, and Robert JK 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*, 194–201. Retrieved October 31, 2015 from http://dl.acm.org/citation.cfm?id=1463756

[24] Eva Hornecker. 2005. A design theme for tangible interaction: embodied facilitation. In *ECSCW 2005*, 23–43.

[25] 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*, 437–446.

[26] Thomas Humphrey, Joshua Gutwill, and The Exploratorium APE Team. 2005. *Fostering Active Prolonged Engagement: The Art of Creating APE Exhibits.* Routledge, Abingdon, UK.

[27] Mikhail Jacob, Gaëtan Coisne, Akshay Gupta, Ivan Sysoev, Gaurav Verma, and Brian Magerko. 2013. Viewpoints AI. In *Proceedings of the Ninth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE '13).*

[28] Ken Kahn and Niall Winters. 2017. Child-Friendly Programming Interfaces to AI Cloud Services. In *Data Driven Approaches in Digital Education* (Lecture Notes in Computer Science), 566–570.

[29] Tony CM Lam and Priscilla Bengo. 2003. A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation* 24, 1: 65–80.

[30] Phoebe Lin, Jessica Van Brummelen, Galit Lukin, Randi Williams, and Cynthia Breazeal. 2020. Zhorai: Designing a Conversational Agent for Children to Explore Machine Learning Concepts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 13381–13388.

[31] Robb Lindgren and J Michael Moshell. 2011. Supporting children's learning with body-based metaphors in a mixed reality environment. In *Proceedings of the 10th International Conference on Interaction Design and Children*, 177–180.

[32] Duri Long, Mikhail Jacob, Nicholas Davis, and Brian Magerko. 2017. Designing for Socially Interactive Systems. In *Proceedings of the 11th Conference on Creativity and Cognition.*

[33] Duri Long, Mikhail Jacob, and Brian Magerko. 2019. Designing Co-Creative AI for Public Spaces.

[34] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 ACM Conference on Human Factors in Computing Systems (CHI 2020).* https://doi.org/10.1145/3313831.3376727

[35] 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: 14. https://doi.org/10.1145/2886418

[36] Adam Maltese. 2018. A Suite of Tools for Educators and Students: Documenting and Evaluating Creativity in Making. Retrieved from http://www.adammaltese.com/content/wp-content/uploads/2018/04/Creativity-4.11.18.pdf

[37] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think.* Penguin.

[38] Clara C Pratt, William M McGuigan, and Aphra R Katzev. 2000. Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation* 21, 3: 341–349.

[39] Janet C Read and Stuart MacFarlane. 2006. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 conference on Interaction design and children*, 81–88.

[40] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, and Kafai, Yasmin. 2009. Scratch: Programming for All. *Communications of the ACM* 52, 11: 60–67. https://doi.org/10.1145/1592761.1592779

[41] Jessica Roberts and Leilah Lyons. 2017. Scoring Qualitative Informal Learning Dialogue: The SQuILD Method for Measuring Museum Learning Talk. . Philadelphia, PA: International Society of the Learning Sciences.

[42] Paul Scott. 2012. *semnet: a Python based semantic network demo.* Retrieved from https://github.com/paulscott56/semnet

[43] Matthew Smith, Christian Szongott, Benjamin Henne, and Gabriele Von Voigt. 2012. Big data privacy issues in public social media. In *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, 1–6.

[44] Scott S Snibbe and Hayes S Raffle. 2009. Social immersive media: pursuing best practices for multi-user interactive camera/projector exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1447–1456.

[45] Karen Sullenger. 2006. Beyond School Walls: Informal Education and the Culture of Science. *Education Canada* 46, 3: 15–18.

[46] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R Cooperstock. 2019. Can You Teach Me To Machine Learn? In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 948–954.

[47] Una Tellhed, Martin Bäckström, and Fredrik Björklund. 2017. Will I fit in and do well? The importance of social belongingness and self-efficacy for explaining gender differences in interest in STEM and HEED majors. *Sex roles* 77, 1: 86–96.

[48] David Touretzky and Christina Gardner-McCune. 2022. Artificial Intelligence Thinking in K-12. In *Computational Thinking in K-12: Artificial Intelligence Literacy and Physical Computing.* MIT Press.

[49] David S Touretzky. 2017. Computational thinking and mental models: From kodu to calypso. In *Blocks and Beyond Workshop (B&B), 2017 IEEE*, 71–78.

[50] Jessica Van Brummelen and Phoebe Lin. 2020. Engaging Teachers to Co-Design Integrated AI Curriculum for K-12 Classrooms. *arXiv preprint arXiv:2009.11100.*

[51] Jessica Van Brummelen, Judy Hanwen Shen, and Evan W Patton. 2019. The Popstar, the Poet, and the Grinch: Relating Artificial Intelligence to the Computational Thinking Framework with Block-based Coding. In *Proceedings of International Conference on Computational Thinking Education*, 160–161.

[52] Eric N Wiebe, Allison Lamb, Megan Hardy, and David Sharek. 2014. Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior* 32: 123–132.

[53] Xiaofei Zhou, Jessica Van Brummelen, and Phoebe Lin. 2020. Designing AI Learning Experiences for K-12: Emerging Works, Future Opportunities and a Design Framework. *arXiv preprint arXiv:2009.10228.*

[54] 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*, 121–132.