

“Bee and I Need Diversity!” Break Filter Bubbles in Recommendation Systems through Embodied AI Learning

Xiaofei Zhou
xzhou50@ur.rochester.edu
University of Rochester
Rochester, USA

Zhenyao Cai
zhenyao@uci.edu
University of California Irvine
Irvine, USA

Yushan Zhou
yzhou101@ur.rochester.edu
University of Rochester
Rochester, USA

Annie Qiu
aqiu2@u.rochester.edu
University of Rochester
Rochester, USA

Yunfan Gong
ygong18@u.rochester.edu
University of Rochester
Rochester, USA

Qinqin Xiao
qxiao5@u.rochester.edu
University of Rochester
Rochester, USA

Alissa N. Antle
aantle@sfu.ca
Simon Fraser University
Vancouver, Canada

Zhen Bai
zbai@cs.rochester.edu
University of Rochester
Rochester, USA

ABSTRACT

AI recommendations influence our daily decisions. The convenience of navigating personalized content goes hand-in-hand with the notorious *filter bubble* effect, which may decrease people’s exposure to diverse options and opinions. Children are especially vulnerable to this due to their limited AI literacy and critical thinking skills. In this study, we propose a novel Augmented Reality (AR) application *BeeTrap*. It aims to not only raise children’s awareness of filter bubbles but also empower them to mitigate this ethical issue through sense-making of AI recommendation systems’ inner workings. By having children experience and break filter bubbles in a flower recommendation system, *BeeTrap* utilizes embodied metaphors (e.g., NEAR-FAR, ITERATION) and analogies (bee pollination) to bridge abstract AI concepts with sensory-motor experiences in familiar STEM contexts. To evaluate our design’s effectiveness and accessibility for a broad range of children, we introduced *BeeTrap* in a four-day summer camp for middle-school students from underrepresented backgrounds in STEM. Results from pre- and post-tests and interviews show that *BeeTrap* developed students’ technical understanding of AI recommendations, empowered them to break filter bubbles, and helped them foster new personal and societal perspectives around AI technologies.

CCS CONCEPTS

- Applied computing → Interactive learning environments;
- Social and professional topics → Computing education; • Human-centered computing → Empirical studies in HCI.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
IDC ’24, June 17–20, 2024, Delft, Netherlands
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0442-0/24/06
<https://doi.org/10.1145/3628516.3655802>

KEYWORDS

Embodied learning; analogy-based learning; Augmented Reality; AI literacy; K-12 STEM education; filter bubble

ACM Reference Format:

Xiaofei Zhou, Yushan Zhou, Yunfan Gong, Zhenyao Cai, Annie Qiu, Qinqin Xiao, Alissa N. Antle, and Zhen Bai. 2024. “Bee and I Need Diversity!” Break Filter Bubbles in Recommendation Systems through Embodied AI Learning. In *Interaction Design and Children (IDC ’24)*, June 17–20, 2024, Delft, Netherlands. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3628516.3655802>

1 INTRODUCTION

Personalized recommendations shape people’s daily information consumption and decision-making by reducing information overload and decision fatigue. The convenience comes at a profound cost, especially to the young generation. According to a recent report from the Center for Countering Digital Hate [19], within 2.6 minutes after a teenager’s account was created, TikTok started recommending suicide information, and within eight minutes, contents related to eating disorders were presented. This extreme case highlights the negative impact of personalized recommendation—the so-called *filter bubble* effect that isolates people from a diverse range of contents and viewpoints [35, 86]. Filter bubbles appear when algorithms narrow recommendations to solely match the interests and opinions of individuals and their social circles [33]. This self-reinforcing loop may lead to critical societal effects such as echo chambers [22] and opinion polarization [48].

Filter bubbles could be detrimental to developing curiosity, creativity, and critical thinking since the children are deprived of being exposed to diverse or different information and opinions [15, 86]. Furthermore, filter bubbles could intensify gender and racial biases, leading to unfair or unequal information exposure for people from minority groups and underrepresented communities [15, 17, 33]. Children are particularly susceptible to the impact of filter bubbles due to their underdeveloped critical thinking abilities and impulse behavior [91]. Research efforts have explored engaging youth to

reflect on AI's ethical implications through informal art exhibitions [68] and design workshops [4, 28, 120].

With some early-stage investigations instilling critical thinking around AI in children [4, 39, 67, 92], a research gap exists in empowering children to break filter bubbles by understanding recommendation system mechanisms [6]. Prior research emphasizes algorithmic transparency and user control in promoting children's digital autonomy [119, 120] and reducing the so-called algorithmic anxiety [55]. Both studies identified a lack of understanding among children regarding recommendation systems' inner workings, including inferences of user preferences [119, 120] and ranking mechanisms [55]. Understanding the actual technological mechanisms plays a fundamental role in promoting students' critical inquiry into how algorithms may cause ethical issues in diversity and equity [57, 116].

Existing K-12 AI learning experiences often apply the "low floor" principle for CS education [94]. This approach aims to engage novice learners in understanding AI concepts (e.g., data training and evaluation) while keeping the underlying algorithms as black-boxes to prevent overwhelming the students [49, 121, 127]. However, studies have shown that students often hold misconceptions about AI's inner workings [49, 59, 64]. To address this, we must thoughtfully introduce AI building blocks [49, 68, 99], ensuring accuracy and preventing the persistence of misconceptions [50].

Understanding the inner workings and mitigation of filter bubbles could be challenging for children with limited computational literacy (i.e., the ability to use computers for communication, problem-solving, and creation [30]), especially those from underrepresented backgrounds in STEM [32, 58]. First, students lacking math literacy require significant cognitive effort [21, 117] to grasp abstract AI concepts and the intricate context involved [120]. For example, comprehending how AI makes data-driven inferences and deciding user actions can be difficult for children [119, 120]; multidimensional data—a fundamental AI concept—requires advanced data literacy [21]. Second, students' experiences with AI technologies vary based on factors like access to household digital devices and socioeconomic & cultural backgrounds. The filter bubble effect may remain abstract [68, 81] and inconspicuous [37] for students who rarely use online services. Lastly, students with limited prior knowledge often exhibit reduced interest and motivation when confronted with cognitive challenges during the learning process [9, 104].

Our research explores the design of an Augmented Reality (AR) application that introduces AI concepts to students through embodied and analogical learning theories. First, embodied learning, which involves tangible manipulation, spatial interaction, and expressive representation [51], aids in simplifying complex computational concepts for children [77, 113, 125], enhancing their engagement and motivation [98]. In AI education, embodied learning supports multidimensional data analysis [13, 21, 34], learning AI training steps [18, 59, 127], and debugging algorithms [72]. Second, analogical learning could bridge abstract AI concepts with learners' concrete prior knowledge [23, 44]. The deep relational structure shared between the source and the target domain can represent inter-relations between key AI components, and thus scaffold sense-making [24, 42, 45]. Moreover, embodied metaphors and analogies in K-12 science contexts can offer a common ground that is accessible and inviting for children with diverse AI experiences [25],

enhancing learning by resonating with how novice users perceive AI technologies [6, 64].

Through iterative design and evaluation with K-12 science teachers and students, we developed a novel AR education application called BeeTrap. First, we use the image schema NEAR-FAR [52] to embody the data similarity. We map BIG-SMALL [53] with the range for similarity-based ranking. These image schemas are chosen as major embodied metaphors because physical distance and walking steps can enhance children's mathematical thinking [110]. Second, we design analogies of bee pollination, a common K-12 STEM learning topic involving science modeling [106], and has been effectively supported through embodied learning and roleplay in existing design [26, 88, 112]. In BeeTrap, the flower-pollination process is an analogy for recommendation algorithmic steps. Kids can roleplay either a bee—a recommendation system user, or an environmental scientist—an AI engineer who breaks filter bubbles. The concepts around biodiversity evoke students' reflections on how the loss of flower diversity impacts bees, and in turn a critical view of the filter bubble effect. We evaluated BeeTrap with nine middle-school students from underrepresented backgrounds in STEM. To investigate how BeeTrap supports students to understand filter bubbles, AI recommendation's inner workings, and mitigation strategy, our research questions include

RQ1 Can BeeTrap support students in learning the target AI concepts (RQ1.1) and motivating conceptual and potential behavioral changes regarding filter bubbles (RQ1.2)?

RQ2 To what extent can image schemas embodying AI concepts support students' learning?

RQ3 To what extent can analogies support AI learning in K-12 science contexts?

This work's major contributions are three-fold: (1) an AR learning application that teaches students about filter bubbles, AI recommendations' inner workings, and a diversification algorithm; (2) implications for designing interaction supported by embodied metaphors for children to learn abstract AI concepts; (3) a list of design heuristics of how analogies can integrate AI learning experience into K-12 science contexts.

2 RELATED WORK

2.1 Teaching Children about AI Recommendation Systems

A few research projects investigate how to teach children about AI recommendation systems. To explore how informal learning can support learners' critical thinking over ethical aspects of AI, researchers designed art exhibitions to provide youth with artistic first-person experiences of positive and negative impacts related to AI recommendations [68]. Workshops are also developed for children to redesign YouTube's recommendation system by identifying different stakeholders [4, 29], for teenage girls to consider how existing and future AI recommendations can impact challenges in their lives [103], and for children to learn about online datafication and coping mechanisms. Researchers have also created structured classroom curricula to deliver AI-related ethical knowledge [40]. Existing research recommends teaching kids about AI by (1) guiding them to reflect upon real-world ethical dilemmas between the

convenient personalized experience and the loss of control and diversity in AI recommendation systems and (2) empowering them to take action to tackle filter bubbles [99]. To fill the gap in unveiling the black-box underlying filter bubbles, we aim to design an educational application to teach children about AI concepts centered around filter bubbles, AI recommendation systems’ inner workings, and diversification as a mitigation strategy.

2.2 Embodied Learning for AI Education

Empirical evidence suggests that, by incorporating gestures or movements, embodied learning develops a deeper comprehension of the material [76] by promoting cognitive functions such as attention, memory, and problem-solving [61, 101]. Existing embodied learning research investigated supporting high-dimensional data analytical processes by inviting students to position in a physical space representing a 2D projection [13, 21] and by distributing the demanding cognitive load in 3D spatial environments for non-data experts to make sense of complex data more intuitively [34]. Researchers also explored concrete 3D models to reify multidimensional data [62] and body gestures to support learning the AI training pipeline [18, 59, 73, 127]. Tangible interfaces and spatial metaphors are designed to embody neural networks [27], semantic networks, feature-based machine learning algorithms [74], and debugging algorithms [72]. Empirical results show that embodied interaction and tangible representations help students perceive AI concepts more effectively compared with graphical user interfaces [27]. Combining co-located spaces and embodied interactions with roleplay, participatory simulation can support children in understanding dynamic systems, social interactions, and complex decision-making processes [20].

We aim to design an AR technology to apply image schemas—the recurrent patterns in bodily experiences [10, 52, 56]—to create embodied metaphors for abstract concepts centered around the impact, inner workings, and mitigation strategy of filter bubbles. Image schemas related to optical phenomena have been applied for children to manipulate and understand the dot product operation in AI education [125]. Existing research in math education shows that physical distance and walking steps can enhance children’s mathematical thinking [110], which is one of the major learning barriers for children to develop AI literacy [32, 58, 124]. Novice learners find it intuitive to perceive the experience of being trapped by similar data and require designs for recommendation systems to demonstrate how they are built over iterations [6]. Therefore, we choose image schemas NEAR-FAR, BIG-SMALL, and ITERATION as the major embodied metaphors underlying the interface design.

2.3 Analogical Learning for AI Education

Analogical learning is a cognitive process in which people connect concepts in a familiar source domain and concepts in an unfamiliar target domain for learning [23, 44]. An existing work used water flow as an analogy for electricity to create a more accessible experiment space with electricity building blocks augmented by virtual water flow [65]. Analogies have been commonly used to facilitate students’ conceptual transformation and may increase student engagement and motivation in activities through students’ closeness

with source domains [98, 109]. Furthermore, analogies can be enhanced through embodied interaction with the physical world and concrete representations [25] in an AR learning environment.

More research efforts are needed to investigate analogies to forge children’s understanding of abstract AI concepts. Existing work proposed using human intelligence as an analogy for machine intelligence to teach AI through roleplay and embodied cognition [25, 32, 124]. However, conflating human thinking with computer processing has been identified as one of the major barriers faced by students [107]. Therefore, we explore how analogies in K-12 science domains can support AI learning. K-12 science topics can be a common ground for analogical learning to support diverse learners [11, 68, 78]. Students have the instinct to create analogies or metaphors in scientific discourse [60]. We choose bee pollination, a common K-12 science curriculum topic [106], as the base domain for creating analogies in the BeeTrap activities.

3 THE DESIGN OF BEETRAP

3.1 Target Learning Objectives

Our educational application targets three key learning objectives (Fig. 1): (1) understanding the filter bubble effect, (2) grasping the inner workings causing filter bubbles in AI recommendations, and (3) applying a diversification algorithm to break filter bubbles. The content is centered around content-based recommendation systems, which recommend items that have features aligned with user profiles [2, 87, 95].

First, the filter bubble, a key ethical issue, is a scenario in which users receive content increasingly similar to their past choices [37]. This leads to reducing content diversity and user selection options in recommendation systems.

Second, the steps of how content-based recommendation systems form filter bubbles [3, 54, 75, 87] are introduced, including recording user choices, comparing user profiles with available items, ranking items based on their similarity to the user profile, and recommending the top-ranked items.

Third, a diversification algorithm breaks filter bubbles [66, 126] by enlarging the item list for ranking and re-ranking items by diversity, calculated as the mean distance between item pairs [89], thereby producing a more varied set of recommendations.

3.2 Iterative Design Process

Our design methodology adheres to design thinking principles [69], emphasizing iterative development and assessment. The initial phase involved collaboratively brainstorming with Human-Computer Interaction (HCI) researchers and AI experts, leading to the conceptualization of BeeTrap v1. This version features analogies of a garden, flowers, bees, and an environmental scientist, alongside the NEAR-FAR [52] embodied metaphor.

After a proof-of-concept evaluation of BeeTrap v1 with middle-school students, we co-designed with K-12 science educators and students. This phase commenced with participants interacting with BeeTrap v1 to grasp analogies and metaphors. The co-design process was facilitated using methods including storyboarding [111], layered elaboration [118], and comicboarding [79]. Insights gained from this stage directed enhancements in BeeTrap v2, notably the

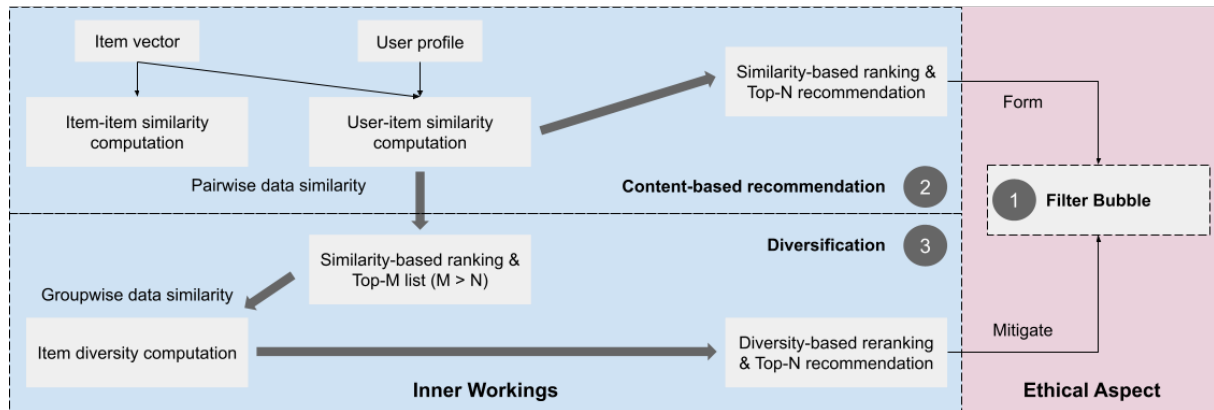


Figure 1: AI concepts related to (1) the filter bubble, (2) inner workings of AI recommendation systems, (3) diversification.

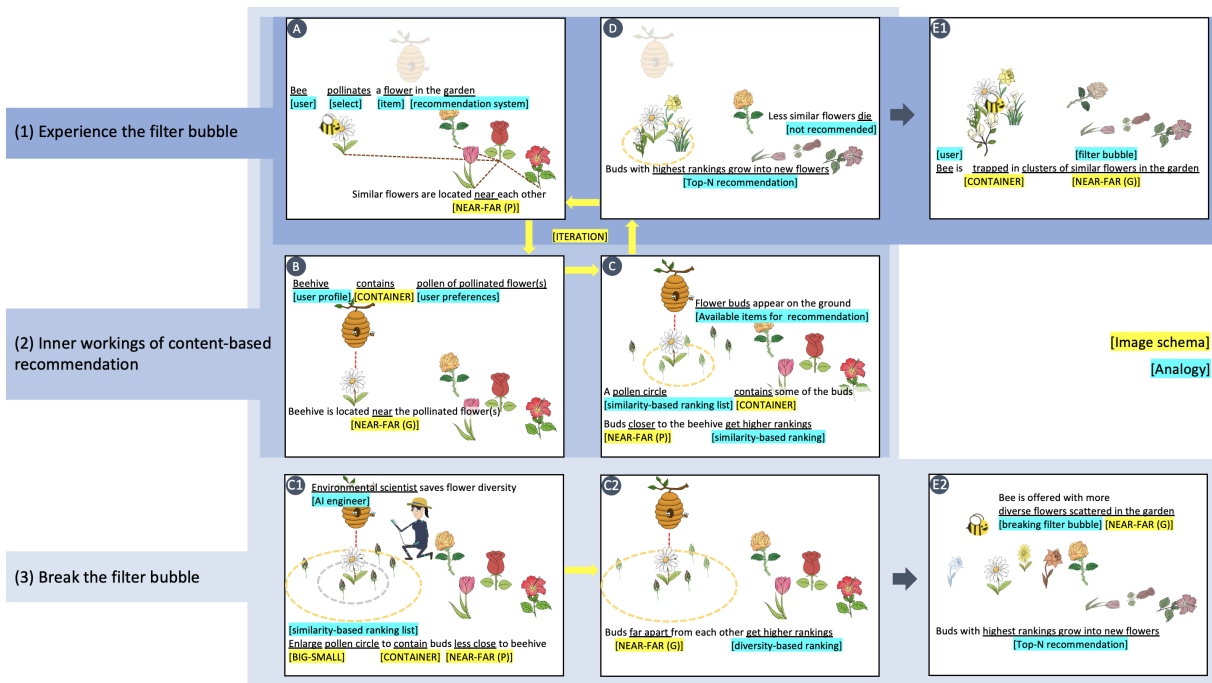


Figure 2: How different image schemas and analogies are designed in three BeeTrap activities: (1) experience filter bubbles; (2) inner workings of a content-based recommendation system; (3) diversification to break filter bubbles.

integration of additional analogies such as a beehive, flower buds, and pollen, as well as image schemas BIG-SMALL and ITERATION.

3.3 Final Design of BeeTrap Activities

BeeTrap contains three activities for three learning objectives (Fig. 2). First, students roleplay as bees, learning about filter bubbles by observing how their choices of flowers affect the flower growing and withering in the garden (Fig. 2.1). In the BeeTrap garden, the distance between flowers embodies the similarity between data that describes the flowers (Fig. 2.A). After the bee pollinates a flower, similar flowers grow and dissimilar flowers farthest away wither

(Fig. 2.D). Eventually, the bee can only interact with similar flowers clustered in small areas in the garden, indicating the filter bubble effect (Fig. 2.E). This immersive first-person experience demonstrates the entrapment of a user within filter bubbles. The inner workings of the recommendation system are visible in this activity, which aims to keep the visual representations in the BeeTrap garden consistent across different activities but risks overwhelming learners with too many visual elements.

Second, to delve into a content-based recommendation system's inner workings (Fig. 2.2), students follow a series of embodied and analogical events reflecting the algorithmic steps:

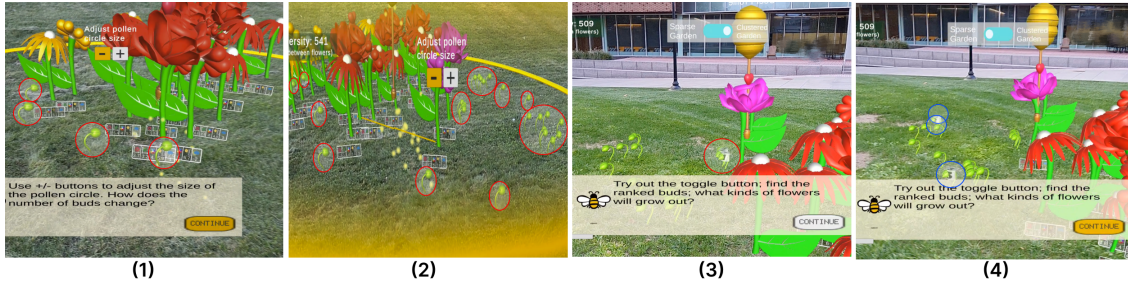


Figure 3: BeeTrap interfaces demonstrating the diversification algorithm: (1) the pollen circle with a normal size contains four to five flower buds (in circles); (2) enlarging pollen circle to contain more than 20 buds; buds less close to the beehive (i.e., less similar to the user profile) (in circles) are included for future ranking; (3) a basic content-based recommendation system assign higher rankings to the buds closer to the beehive (in circles); (4) applying a diversification algorithm on the content-based recommendation system to assign higher rankings to the buds farther away from the beehive (in circles).

- (1) a beehive moving towards the pollinated flower and collects pollen from the pollinated flowers (Fig. 2.B), which represents the update of the user profile by aggregating items accepted by the user;
- (2) six flower buds appearing on the ground as available items for recommendation and a pollen circle containing three buds closest to the beehive (Fig. 2.C), which represent item ranking based on proximity to the user profile;
- (3) the three top-ranked buds growing into new flowers (Fig. 2.D).

This process of pollination and observation unfolds over four rounds, enhancing students’ understanding of the system’s dynamics.

Third, students roleplay an environmental scientist to increase the flower diversity in the garden by acting out a two-step diversification algorithm (Fig. 2.3). To adjust the size of the ranked item list, students manipulate the pollen circle size and observe different numbers of buds contained by the pollen circle (Fig. 2.C1). By altering similarity-based and diversity-based ranking, students decide which buds to grow for the bee (Fig. 2.C2). This hands-on process actively engages students to see the real-time impact of their attempts to counteract the homogenizing effects of a basic recommendation algorithm.

3.3.1 Embodied metaphors in BeeTrap.

NEAR-FAR for data similarity/diversity. The NEAR-FAR image schema describes the spatiotemporal relationship between two points in space [12, 108]. Research shows that novice learners effectively grasp multidimensional data and complex analytical methods by positioning themselves in a physical space to represent a 2D projection of multidimensional data points [13, 21]. Moreover, walking steps enhance children’s reasoning with math concepts and methods [110]. In BeeTrap, we use the pairwise NEAR-FAR (**NEAR-FAR (P)**) as an embodied metaphor, representing the Euclidean distance between two multidimensional data points (Fig. 2.A). The groupwise distance encompassing flowers, buds, and the beehive (**NEAR-FAR (G)**) symbolizes the data diversity. This method, which promotes high bodily engagement including movement and locomotion [102], helps students better understand pairwise data similarity and groupwise data diversity.

BIG-SMALL for the range for similarity-based ranking. It facilitates tangible interaction through objects of varying sizes [12]. In the BeeTrap garden, the dynamic **BIG-SMALL** attribute of the pollen circle embodies the range of flower buds ranked based on their similarity to the beehive. During the diversification activity, students adjust the pollen circle size (Fig. 3.1). Enlarging it widens the range of items for ranking, leading to more diverse flower bud recommendations (Fig. 2.C1 & Fig. 2.C2). This interactive manipulation transforms the pollen circle into a digital manipulative [93], supporting children in exploring abstract computational concepts [85, 94] through expressive activities with concrete operations [14].

ITERATION for the formation of filter bubbles. Research suggests that users want AI recommendation systems to demonstrate how they evolve [6]. We use **ITERATION**, an image schema for the repeated cycle of events [41], to embody the progressive formation and mitigation of filter bubbles in a content-based recommendation system (i.e., garden) (Fig. 2.1 & Fig. 2.2). Each repetition of the system’s steps results in new recommendations based on the user’s choices. With time-travel buttons, students can navigate between different temporal states of the garden. This way, they can easily see and compare how the recommendations change over time (Fig. 4).

3.3.2 Analogy and roleplay in BeeTrap. To lower the entry barrier for children in learning AI, we designed analogy-based learning [43] to map AI terminologies to bee pollination context, a K-12 science learning topic that students are more familiar with [106]. Technology-enhanced play environments have been designed for students to roleplay as bees to learn about scientific phenomena and modeling [88, 112]. Furthermore, research shows that the interfloral distance is a contributing factor to a bee’s flower preference; this forms a feedback loop in which a bee pollinates a specific kind of flower causing the flowers to grow denser, which further encourages the bee to only pollinate that kind of flower [46]. This means that the bee pollination context aligns with the embodied learning experience of (1) NEAR-FAR that similar kinds of flowers grow closer, and (2) ITERATION of the filter bubble formation.

Garden as a content-based recommendation system, bee pollination as user selection. In Beetrapp, the garden represents a content-based

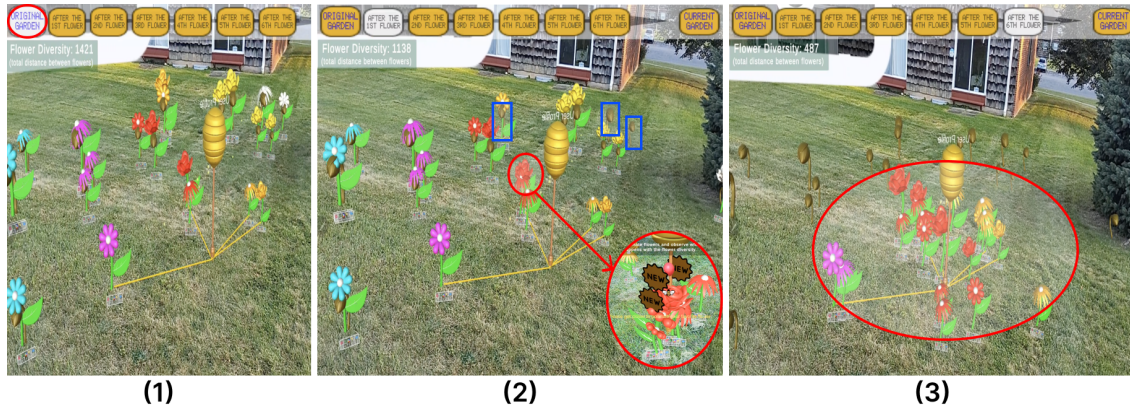


Figure 4: After being trapped in filter bubbles, students compare gardens in different iterations and reflect on the formation process of filter bubbles by using time-travel buttons: (1) students use the time-travel button (in the circle) to revisit the initial garden; (2) in the first iteration, a red rose is pollinated; three most distant (i.e., most dissimilar) flowers die (in rectangles); three closest flowers grow up (in circles); (3) in the last iteration, students roleplaying the bee are trapped in a cluster of similar flowers (in the circle); the filter bubble is formed.

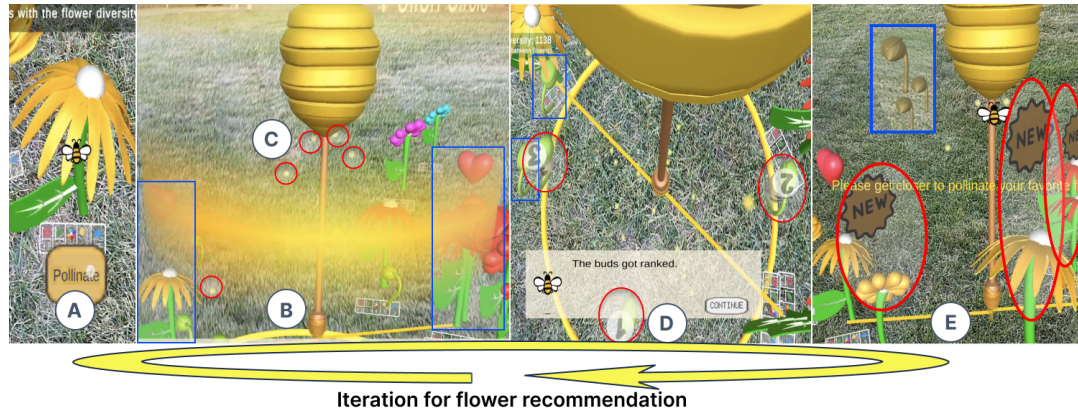


Figure 5: BeeTrap interfaces demonstrating AI recommendation inner workings: (A) bee pollinates a flower; (B) beehive moves towards the pollinated flower (in rectangles); (C) pollen (in circles) from two pollinated flowers (in rectangles) go into the beehive; (D) flower buds appear on the ground; pollen circle contains three buds inside (in circles) and some buds outside (in rectangles); buds closer to the beehive get higher rankings (in circles); (E) similar flowers with highest rankings grow (in circles); dissimilar flowers farther away die (in rectangles).

recommendation system. The bee serves as an analogy for the system's user, and the act of bee pollination mirrors a user selecting a recommended item (Fig. 2.A & Fig. 5.A). New flower growing represents newly recommended items, while the withering of flowers indicates removing items from the recommendation list (Fig. 2.D). Through roleplaying in a co-located space with flowers, students get to 'be' the bee, experiencing the first-person view of how user choices lead to the formation of filter bubbles (Fig. 2.E & Fig. 4). Such biodiversity concepts may shift students' perspectives from passive users benefiting from AI recommendations' convenience to critical thinkers who consider biodiversity loss' negative impacts.

Beehive as the user profile. A recommendation system creates a user profile as a structured data representation capturing a user's preferences [95]. BeeTrap employs a beehive as an analogy for

the user profile (Fig. 2.B). A basic approach to representing user preferences is through an aggregated data representation, essentially a weighted average of previously selected items by the user. Therefore, the user profile update is visually demonstrated by the beehive moving to a location close to pollinated flowers, which reflects the weighted average of items as user preference (Fig. 5.B). Pollen from pollinated flowers going into the beehive embodies the data collection in recommendation systems.

Flower buds as available items for recommendation, environmental scientist as an AI engineer. In BeeTrap, flower buds on the ground are available items for ranking and recommendation (Fig. 2.C). The highest-ranking buds within the pollen circle grow into new flowers and flowers farthest away die (Fig. 5.D). During diversification, students roleplay as an environmental scientist whose goal is to

Table 1: Basic demographic information of students who participated in the evaluation study

PID	Gender	Grade	Race
P1	Female	8th	Black or African American
P2	Male	7th	Black or African American
P3	Male	7th	Black or African American
P4	Female	8th	Black or African American
P5	Female	8th	White+Black
P6	Female	6th	Asian+Black
P7	Male	8th	White+Black
P8	Male	10th	Black or African American
P9	Male	10th	Black or African American

save flower diversity. This aims to motivate students to break filter bubbles as an AI engineer developing a diversification algorithm for a content-based recommendation system.

3.4 Implementation

3.4.1 Dataset construction for generating flowers. Our dataset comprises 210 flowers, each characterized by five attributes: petal color (six categories: white, yellow, orange, red, purple, blue, coded as integers 1–6), petal shape (three types: rose, sunflower, daisy, coded as 1, 2, 3), petal size, leaf size, and height. The latter three features are quantified with numerical values ranging from 1 to 6.

3.4.2 System development. We created 3D flower models based on feature values in Blender and developed BeeTrap in Unity, used a Classical Metric Multidimensional Scaling algorithm [96] to calculate 2D coordinates for flowers, used ARFoundation library [115] for plane detection and 3D virtual objects manipulation. Three BeeTrap activities are deployed to Samsung Tablet S8 devices (Fig. 4, 5, 3).

4 EVALUATION WITH STUDENTS FROM UNDERREPRESENTED BACKGROUNDS

4.1 Participant

Participants consisted of nine students from a summer camp in an ethnically and economically diverse urban school district in Northern New York (Table 1). Most participants identified as Black or African American, and some reported being of biracial background. Gender representation was almost equally distributed.

4.2 Study Procedure

The study was conducted on-site during a summer camp. Three researchers resided in two rooms throughout the entire study session to assist students with the study procedure when requested. For each BeeTrap activity, two students collaborated, or one student completed it independently if no other students were available for pairing (Fig. 6). The entire user study lasted 2 to 2.5 hours per student and was carried out on four consecutive days. The groupings of students were decided based on each day’s attendance. In the end, all students finished the three BeeTrap activities using the same procedure (Table 2).

Table 2: The overview of the study procedure.

Session	Activities
Pre-survey	(1) demographic information; (2) pre-test on AI recommendation systems.
Warm-up	Introduction to (1) AI and its everyday application; (2) bees and pollination.
1st session	(1) Filter bubble experience; (2) post-test; (3) post-activity interview.
2nd session	(1) Recommendation system mechanism; (2) post-test; (3) post-activity interview.
3rd session	(1) Diversification; (2) post-test; (3) post-activity and post-study interviews.

4.3 Data Collection

4.3.1 Pre- and post-tests. Before and after each experiment session, participants were administered the pre- and post-test, with six questions assessing AI knowledge (Table 3 “Assessment Question”). The questions were adapted from previous literature [1–3, 38, 66, 82, 89] and evaluated by two AI experts to ensure their measurement validity. Each pair of pre- and post-tests is aligned with the target learning objectives of the corresponding BeeTrap activity (Table 3 “Target concept”).

4.3.2 Interviews. After each activity, participants engaged in semi-structured interviews. This helped researchers gain a deeper understanding of participants’ perceptions of individual embodied metaphors, analogies, and the AI concepts underlying BeeTrap. For example, we explored their comprehension of user profile updates: “What does the beehive represent in the flower recommendation system?”, “Did you notice the beehive moving in the garden?”, “How did the beehive move?” In the end, we conducted a post-study interview to inquire about students’ overall experiences, their perceptions of recommendation systems, and their future use of AI. Questions are mainly adapted from existing studies in transformative informal science learning [71, 90]. All interviews were audio-recorded under consent.

4.3.3 Log data. Three types of log data were collected. First, **the movement log** records the 3D position of the device whenever the user interacts with the game. Second, **the interaction log** records (1) learner behaviors of selecting a flower to pollinate, manipulating the pollen circle size, and switching the ranking mechanism; (2) flowers pollinated, buds, newly-grown flowers, and dead flowers by their ID; (3) the user profile vector; (4) flower diversity at each time point; and (4) time stamps for each in-app interaction.

Third, we collected **in-app assessment logs** in the recommendation system mechanism activity. To evaluate if the BeeTrap analogies can be quickly connected with the underlying AI concepts by children during the activity, five in-app multiple-choice questions were inserted into the last two iterations. In the first two iterations, students are guided to observe and vocalize their understanding of various design metaphors. In the last two iterations, students answered five multiple-choice questions that probed their understanding of the recommendation mechanism: (1) What goes into the beehive (user profile)? (2) What do flower buds on the ground represent? (3) What do numbers above the flower buds represent?

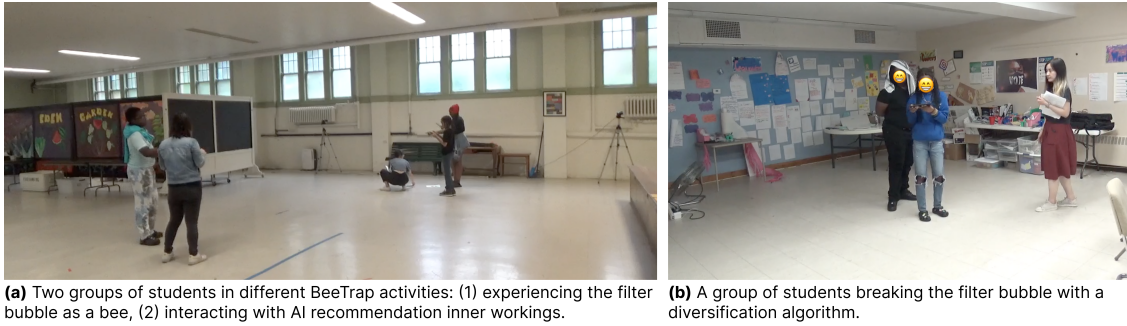


Figure 6: Students engaged in BeeTrap activities with researchers' facilitation when needed.

(4) What does the pollen circle represent? (5) How are flowers located in the garden? The questions are designed to appear twice, encouraging students' reflection.

In-app questions contribute to forming a just-in-time telling experience [84, 100], where students first explore the BeeTrap garden as bees to develop prior knowledge and then go through just-in-time telling to confront the key AI concepts.

4.4 Measures and Data Analysis

RQ1. Can BeeTrap support students in (RQ1.1) learning the target AI concepts and (RQ1.2) motivating conceptual and potential behavioral changes regarding filter bubbles? We measured students' learning gains with pre- and post-tests and identified students' conceptual and behavioral changes from post-study interviews.

RQ1.1: Learning gains in AI literacy. The accuracy of two multiple-choice questions can be analyzed quantitatively directly. For open-ended questions, two researchers separately graded students' responses. The process's reliability was affirmed with Cohen's kappa scores of 0.78 and 0.71, indicating substantial agreement [36].

RQ1.2: Self-reported conceptual and potential behavioral change. Integrating ethics into AI education equips students to critically engage with and address ethical dilemmas in AI, both as informed users and potential creators [123]. One of our primary goals is to prepare students to be aware of filter bubbles' impact on their daily lives. Therefore, we explored their conceptual and prospective behavioral change as consumers of AI recommendations as an exit ticket concluding the BeeTrap learning experience.

With the constructivist approach adopted in BeeTrap, conceptual change is identified if learners develop new perspectives regarding AI recommendation systems [31]. We measured students' potential future actions grounded in the essence of ethics, which examines moral principles guiding individual behavior.

Conceptual and potential behavioral changes are measured from the post-study interviews. The first author transcribed all audio recordings through Rev¹ and manually corrected the transcripts. Three researchers initiated an inductive open coding process [105]. Through line-by-line coding, changes were identified in students' perceptions and potential future actions regarding filter bubbles and AI recommendations. Any surprising aspects were highlighted. Through regular meetings between three researchers, the codes

¹<https://www.rev.com/>

accumulated, and categories emerged. Then an axial coding process was employed to uncover the relationships between these categories and group them into cohesive themes [97]. These themes revolved around learning outcomes, how design elements supported learning, and learners' misconceptions. Quotes were taken to illustrate and support the major themes.

RQ2. To what extent can image schemas embodying AI concepts support students' learning? To measure the learning effects of image schemas NEAR-FAR, BIG-SMALL, and ITERATION in embodying specific AI concepts for children, we analyzed a combination of interaction and movement logs for embodied learning behaviors, post-activity interviews for self-reported reflection, and pre- and post-tests for learning gains.

To understand embodied learning behaviors through NEAR-FAR, we plotted students' walking paths and areas to illustrate their body movements, distances between pollinated flowers, and the flower-growing areas for each round of bee pollination. For ITERATION, we conducted a descriptive analysis of the frequency and time to observe gardens at different time points. For BIG-SMALL, we analyzed students' interaction with pollen circle size and bud re-ranking, along with flower diversity changes.

With post-activity interviews, we identified themes around learning content and embodied metaphors through the same analysis procedures as in Section 4.4.

RQ3. To what extent can analogies support AI learning in K-12 science contexts? We first analyzed in-app assessment logs to measure analogies' effectiveness for students to map science concepts with AI concepts. Then we analyzed post-activity interviews to obtain the learning effects of individual analogies and their inter-relations for students to understand specific AI concepts.

5 RESULTS

5.1 RQ1. Can BeeTrap support students in learning the target AI concepts and motivating conceptual and potential behavioral changes regarding filter bubbles?

5.1.1 RQ1.1: Learning gains from pre- and post-test. The data was not normally distributed based on the Shapiro-Wilk test. Thus, we adopted a non-parametric test, the Wilcoxon signed-rank test on the pre- and post-test scores for each question. We calculated the

rank-biserial correlation as the effect size, which indicates a large effect with a value higher than 0.7 for all the questions.

Results of the pre-post test (Table. 3) indicate that the BeeTrap system supported statistically significant increases ($p < 0.05$) in participants' conceptual understanding of all the target key concepts about AI recommendation systems, except the basic algorithmic steps for diversification. By looking into individuals' learning gains between pre- and post-tests, we can tell that P3, P4, P7, and P9 have the highest learning performance among all the students, while P6 and P8 have the lowest performance.

5.1.2 RQ1.2: Students are empowered to develop new personal and societal perspectives around filter bubbles. We found that different students developed different perspectives around filter bubbles. First, seven students began to acknowledge the potential of recommendation systems contributing to the filter bubble formation. Previously, their exposure to AI concepts was limited, but after BeeTrap learning activities, they recognized that filter bubbles could constrain them from accessing diverse information, which potentially limited their exposure to varied perspectives: "I need diverse information.", "(Filter bubble) limits my imagination." Furthermore, eight students were able to connect these concepts to real-world situations. One student related this to their personal experience, noting, "I don't want to be a basketball player, but the social media keeps recommending me basketball content." This connection highlights that they could see the tangible effects of recommendation systems and filter bubbles in shaping their interests and choices.

Second, BeeTrap empowers all students to take action to break filter bubbles in their real lives for personal and societal changes. One participant mentioned he would want to explain the filter bubble effect to his three-year-old cousin. It would be challenging to explain this abstract phenomenon to youngsters, but he said he would use his cousin's favorite video "Cocomelon" to explain that:

"That's challenging to explain to a three-year-old. But I am thinking I am going to tell him that YouTube knows that you like Coco Melon and then it gives you a lot of Coco Melon. And you might not see other videos very often."

The other participant thought filter bubbles might impact people's judgment and mislead decision-making. She thought it's important to have independent thought and be mindful of what AI recommends to us. She mentioned she would use purchasing from Amazon as an example to teach her parents:

"When you shop on Amazon, before you decide to buy, especially what it tries to get you to buy, you gotta check the rating and read the comments to figure if it's actually good enough."

Another participant mentioned that not everyone in the community knows about filter bubbles, how the recommendation system works, how to break filter bubbles, and how to approach recommendations more critically. She wanted to make a slide show to explain these to community members:

"People might say: 'I will just take whatever my Instagram shows me, or I would just watch whatever my YouTube gave to me.' I will tell them that AI would recommend the stuff that you would usually watch. You

should try something new, so you won't be watching the same thing over and over. It will be more diverse. And you will hear more diverse opinions."

These students are motivated to take action, make changes, and deliver impact.

Third, we also observed two students (P1, P7) still need more support in developing more accurate and critical perspectives on the impact of filter bubbles. One student expressed their knowledge of what the filter bubble is, but they don't think it's a bad idea to stay in filter bubbles as the flowers inside of filter bubbles are exactly what they like. Another student mentioned that filter bubbles could be good as they filter out the things that children shouldn't watch.

5.2 RQ2. To what extent can image schemas embodying AI concepts support students' learning?

5.2.1 Understanding the definition, formation, and impact of filter bubbles. In this section, we report how image schemas NEAR-FAR (P) and (G) and ITERATION support students' understanding of filter bubbles (Fig. 1.1).

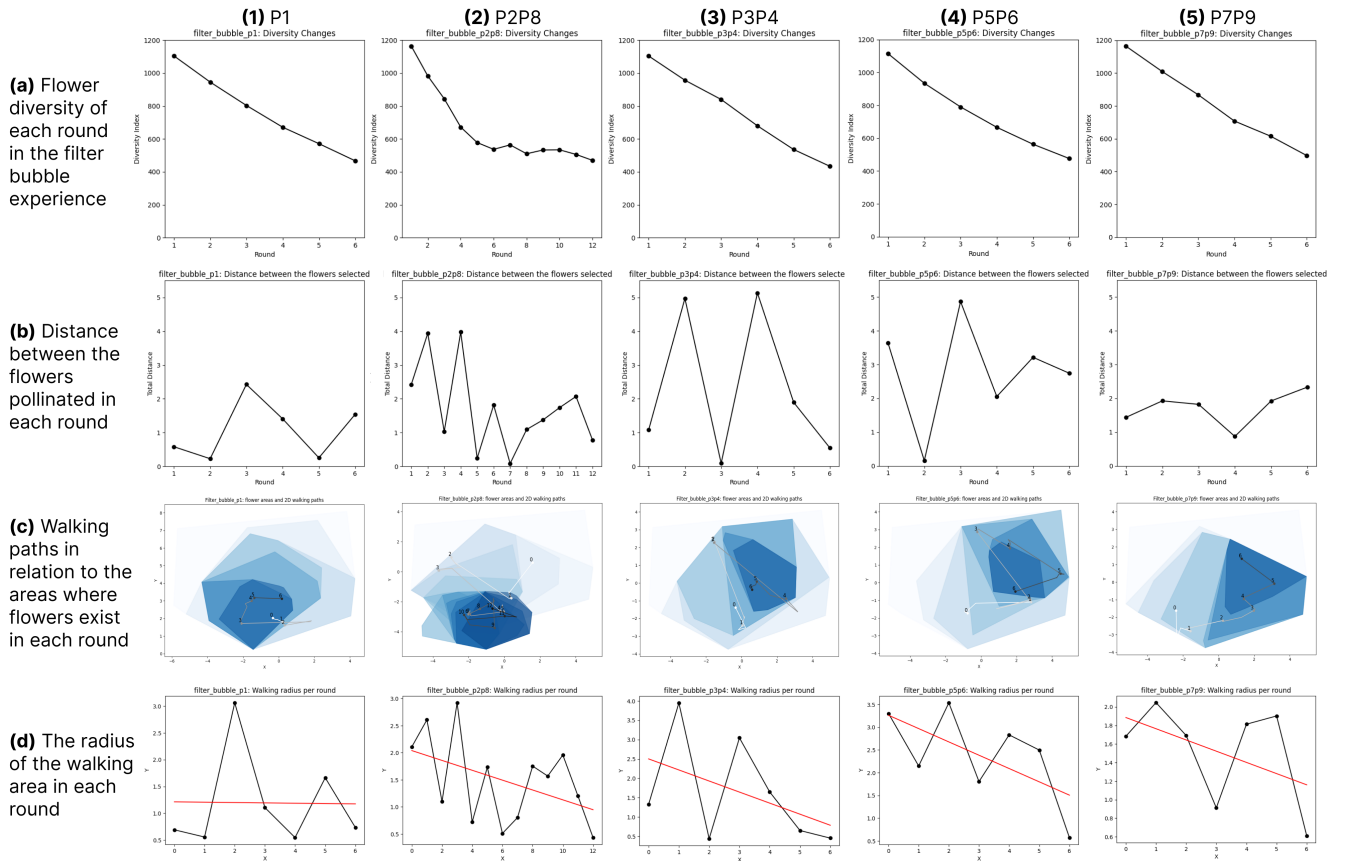
Unexpected body movements while experiencing filter bubbles. Four groups of students (P1, P3&P4, P5&P6, P7&P9) decreased the flower diversity (NEAR-FAR (G)) to the target value with six rounds of flower pollination; however, P2 and P8 took 12 rounds and the diversity increased at one point (Fig. 7.a). Walking paths overlaid on flower areas show that P2 and P8 explored pollinating flowers from the edge of where flowers exist (i.e., the most dissimilar flowers from the beehive in the current garden) (Fig. 7.c2), which caused the slow formation of filter bubbles.

The distances between pollinated flowers (Fig. 7.b) defied our expectation of students pollinating flowers in progressively closer proximity. We observed that some groups (P1, P3&P4, P7&P9) were less physically active initially and started with a very short distance between the second pollinated flowers and the first. This suggests a further design for students to explore the garden more extensively at the beginning. We did observe a trend of decreasing walking areas of four groups (P2&P8, P3&P4, P5&P6, P7&P9) (Fig. 7.d), which indicates their experience of being trapped by the filter bubble in the garden. However, most groups (P2&P8, P3&P4, P5&P6, P7&P9) still walked a lot in some lower-diversity rounds, because students got distracted by the withered flowers far away (Fig. 7.c).

NEAR-FAR helps students recognize and reflect on filter bubbles. First, groupwise NEAR-FAR (G) makes the concept of groupwise data similarity/diversity visible and comparable to children. All students described that similar flowers being clustered together made it visually striking for them to connect with the diversity loss and stimulated their reflection on why filter bubbles were formed and why diversity is needed. For example, P1: "Like it's just weird for them to like, to grow together. So it's like (growing in) one section"; P3 described their negative experience of walking in a space with fewer and fewer areas with flowers: "In the beginning, it was like spaced out... You had to walk more to see. You have more space to explore... Cause if a bee just likes one flower, it'd be all clustered up. You're not gonna go to like all flowers clustered up. (It's) weird... don't stay on one thing for too long."

Table 3: Assessment questions and mean (SD) of pre-, post-tests, and learning gains on AI concepts.

Target concept	Assessment Question (Scores range 0-3)	Mean_Pre (SD)	Mean_Post (SD)	Mean_Diff (SD)	P value
Recognizing AI recommendation systems	What AI-based recommendation systems have you used?	0.44 (0.88)	2.78 (0.44)	2.33 (1.12)	0.008
Definition and impact of filter bubbles	What's the filter bubble? What could be the negative impact of filter bubbles?	0.00 (0.00)	2 (0.87)	2 (0.87)	0.008
User profile	In a recommendation system, what could the user profile use to describe a user?	0 (0)	2.11 (0.96)	2.11 (0.96)	0.004
Pairwise data similarity computation	Which multidimensional data object below is more similar to the target data object (2,4,2,4,2)? Option 1: (2,2,4,2,2); Option 2: (4,4,2,4,4.)	0.78 (1.30)	2.33 (1.32)	1.56 (1.51)	0.033
Basic algorithmic steps in an AI recommendation system	Think of a recommendation system that you have used. How does it decide what to recommend to you? Order the major steps: (A) Update the user profile with the selected item. (B) User selects an item (C) Recommend the most similar items. (D) Rank all available items based on their distance from the user profile.	0.22 (0.67)	2.44 (0.73)	2.22 (1.09)	0.009
Basic algorithmic steps for diversification	As an AI engineer, which option below can you use to make the recommendations more diverse?	1 (0.87)	1.89 (1.36)	0.89 (1.69)	0.114
Average		0.41 (0.43)	2.26 (0.34)	1.85 (0.59)	< 0.001

**Figure 7: Students' interaction with the image schema NEAR-FAR in the filter bubble experience.**

Second, NEAR-FAR (G) inspires students to reflect on filter bubbles further in real-life scenarios. P2 and P8 transferred the connection between NEAR-FAR (G) and diversity-similarity into their daily experience: “It’s gonna decrease our chances of getting new content to see (in YouTube)”; P8 shared their in-depth reflection

about the walking experience in filter bubbles: “I am watching what I want to do. I am imagining what I want to do. Like some kids wanna be a football or basketball player. But it limits my thinking, and my imagination. Maybe I want to do other things.” Furthermore, groupwise NEAR-FAR empowers children to break filter bubbles

from the perspective of a recommendation consumer: “I should pollinate more diverse flowers to make the filter bubble bigger.”

ITERATION helps students compare and understand the process of filter bubble formation. After the formation of filter bubbles, different groups utilized the time travel buttons in distinct manners, facilitating their exploration of gardens at various iterations of the flower pollination (Fig. 8). First, two groups (P3&P4, P7&P9) used time travel buttons to jump back and forth among all the time points multiple times and spent most time observing the gardens in the later iterations. Second, one group (P2&P8) jumped between different gardens repetitively and spent most time observing the gardens in the early iterations. Third, two groups (P1, P5&P6) traveled through all the time points in the temporal order for one round.

Regarding pre- and post-test outcomes, students with higher learning performance (P3, P4, P7, P9) actively engaged with the image schema *ITERATION* by exploring all iterations of gardens multiple times with a focus on the later iterations. In contrast, students with lower levels of conceptual understanding typically revisited the iterations only once (P1, P5, P6) or concentrated on the earlier iterations (P2&P8). This suggests the learning benefit for students to intensively analyze and reflect on the various iterations, especially the later ones, as these stages more clearly and strikingly showcase the impact of the filter bubble.

Students reported that time travel with the image schema *ITERATION* supports them in understanding the changes in the garden during the formation process of filter bubbles. For instance, P3 shared: “(Filter bubble made) you see the same thing over and over again, you get bored”; P2: “Here, (this iteration has) more dying flowers. And here, (this iteration has) less different and diverse flowers”; P5: “From the first garden, there’s more like more alive flowers and more diversity. And by the end, it’s less diversity and more dead flowers.”

5.2.2 Learning AI recommendations’ inner workings. In this section, we report how the image schemas *NEAR-FAR* (P) and (G) and *ITERATION* support students in learning about the inner workings of a content-based recommendation system, including user profile, item-item similarity computation, user profile-item similarity computation, and top-N recommendation (Fig. 1.2).

NEAR-FAR (P) between flowers helps students understand item-item similarity. All groups mentioned that the physical distance supports their reasoning with data similarity between flowers, connecting the distance with the data similarities between flowers. For instance, P2 and P8 demonstrated that “I think that the more similar they are, the closer they are... But the more different they are, the farther they are”; P1: “And the distance they like bit farther apart is because like they’re, different type of flowers. they’re separated from each other.” Eight students understood that the data similarity is determined by all the features in the data vector with *NEAR-FAR* (G). For example, P5 and P6 explained: “Numbers represent like their size, the petal size, their height and the type of details about the flower that basically, giving us the rankings of the distance.”

NEAR-FAR (G) is less effective for learning how user profile is updated. Without an overhead view of objects’ locations (e.g., flowers, the beehive, buds), it could be challenging for students to observe the beehive moving towards the pollinated flower(s) (*NEAR-FAR*

(G)). Only three students (P2, P7, P9) noticed that the beehive moved towards the flower(s) they pollinated. For instance, P2: “I noticed that (the beehive moved based on the pollinated flowers) because it’s choosing the content that like.” Three students (P1, P5, P6) only noticed the beehive moved, without reasoning how it moved. Three students (P8, P3, P4) didn’t notice the beehive movement.

As a comparison, the image schemas *CONTAINER* or *IN-OUT* could be more helpful embodied metaphors for the user profile collecting data from user-selected items. Six out of nine students mentioned the effectiveness of observing pollen flying from all pollinated flowers into the beehive (Fig. 5.3). For instance, P3 and P4 explained: “(beehive) is the data that he gathered; the pollen being transferred into like beehive.”

5.2.3 Empower students to break filter bubbles. In this section, we report how the image schemas *BIG-SMALL*, *NEAR-FAR*, and *ITERATION* together empower students to break filter bubbles by a diversification algorithm. Target AI concepts include groupwise data similarity/diversity, enlarging the range for similarity-based ranking, diversity-based ranking, and diversification outcome (Fig. 1.3).

BIG-SMALL of pollen circle, NEAR-FAR (G), and ITERATION support students to break filter bubbles through exploratory trial & error. In the diversification activity, all groups successfully increased the flower diversity of the garden to the target value (Fig. 9.a). Students demonstrated four different types of exploration behaviors with pollen circle size (*BIG-SMALL*) and diversity-based ranking of buds (*NEAR-FAR* (G)) through *ITERATION* (Fig. 9.b). First, four groups (P1, P2, P3, P7) kept making small changes in pollen circle size for experimentation. All of them finished individual iterations more quickly than the other groups. Second, two groups (P9, P5&P6) initially made only mild and slow enlargements to the pollen circle, which resulted in a minor decrease in flower diversity. Upon observing these changes in *BeeTrap*, these groups adapted their approach, opting for enlarging the pollen circle more substantially, but this action alone didn’t yield a significant increase in diversity. Students recognized the need to re-rank the buds based on diversity. They iteratively adjusted the pollen circle size with deliberate precision, successfully achieving their diversification goal. Third, one group (P4&P8) started by significantly enlarging the pollen circle but failed to increase the diversity. During the second iteration, they experimented with both decreasing and increasing the pollen circle, through which they discerned the necessity of re-ranking based on diversity, and succeeded in the third iteration. Furthermore, we observed that five groups (P2, P4&P8, P5&P6, P7, P9) didn’t initially know to re-rank buds based on diversity. They figured it out after two to eight iterations of trial and error.

With pre- and post-test outcomes, we found that students with higher learning gains (P3, P9, P7, P4, P8) experimented with both decreasing and increasing the pollen circle size (*BIG-SMALL*).

In post-activity interviews, *BIG-SMALL* (i.e., directly manipulating the pollen circle size) supported seven students (P1, P2, P3, P5, P6, P7, P9) in accurately verbalizing the detailed algorithmic steps for diversification. For example, P2 described: “I used the tool by expanding the circle. Enlarge the (range of) options to recommend and then recommend the most diverse to one bee”; P7 and P9 said: “It enlarges the group... it makes the garden bigger for the bees to control and eat things they like.”

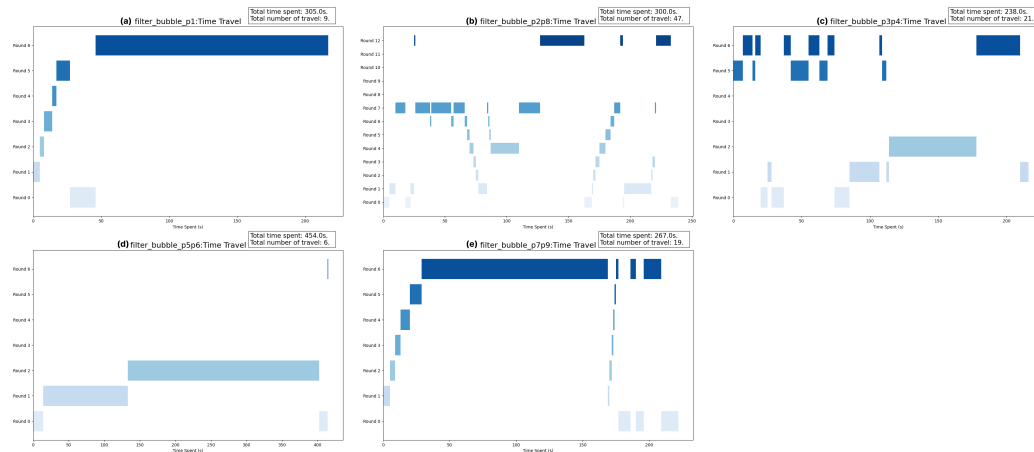


Figure 8: The sequence and time duration of using time travel buttons to jump to different gardens at different time points in the filter bubble experience; graph (b), (c), and (e): students compare all the iterations back and forth for multiple rounds; graph (a) and (d): students only compare all the iterations for one round with little back and forth.

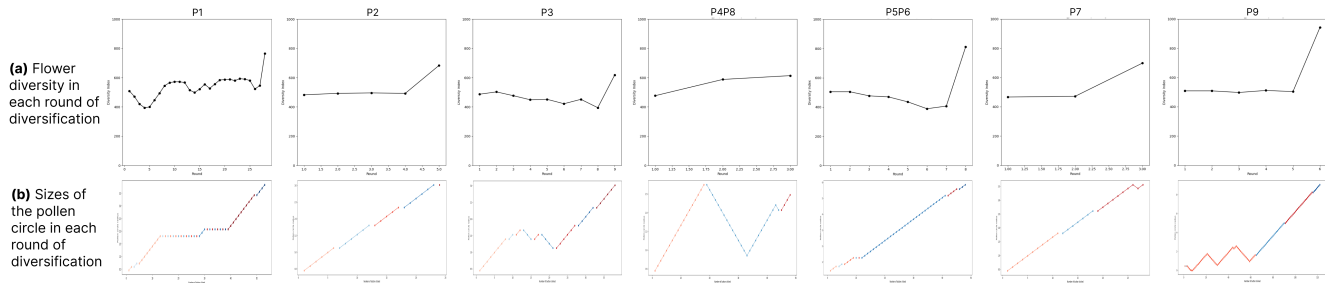


Figure 9: (a) The change in flower diversity in each iteration, (b) sizes of the pollen circle.

As for the distances among flower buds (NEAR-FAR (G)), all students except P2 understood how growing flowers far apart impacts flower diversity, compared to growing flowers clustered together. For instance, P1 mentioned: “cause when I, when I change it to a sparse (group of buds), my score increased”; P7 and P9 explained: “With clustered flowers, (diversity) just goes down.” However, P2 didn’t notice that the rankings changed after re-ranking by diversity, because it was more challenging to locate all the ranked buds that were farther apart (Fig. 3.2).

NEAR-FAR (G) is less effective in developing students’ mathematical understanding of data diversity. Although students grasped the knowledge of how to break filter bubbles with a diversification algorithm, they still missed the mathematical definition of diversity (i.e., the average dissimilarity between all pairs of items in the result set [66]). With NEAR-FAR (G) among existing flowers representing the flower diversity, only one student (P2) understood that the diversity is calculated by the distances between flowers: “the flower diversity means the differences between the flowers”; five students (P5, P6, P7, P9, P3) developed an embodied understanding that flowers located dispersedly in a larger area represent more diverse data; P5 and P6 mentioned: “Increase (pollen circle), so there can be more diverse options”; three students (P1, P4, P8) misunderstood flower diversity as the number of flowers.

5.3 RQ3. To what extent can analogies support AI learning in K-12 science contexts?

5.3.1 Students’ in-app understanding of the mapping between analogies and AI concepts. Within the two iterations of the in-app assessments, all groups correctly identified AI concepts behind the bee pollination analogies, including a beehive, flower buds, bud rankings, pollen circle, flowers, and a garden. This shows that the major BeeTrap analogies are intuitive and effective enough for students to connect with the underlying AI concepts by interacting with BeeTrap shortly.

5.3.2 Biodiversity as an accessible analogy for students with diverse backgrounds to understand information diversity and filter bubbles. We found that familiar ethical issues in STEM fields, such as bee pollination and biodiversity, may offer a unique opportunity to foster children’s understanding of AI ethical issues.

Bee pollination analogies, familiar to K-12 students, can serve as common ground. This makes the experience and implication of filter bubbles accessible for students with different levels of digital literacy. For instance, P9 rarely engages with social media or other AI-based recommendation systems. It’s challenging for them to imagine how the items recommended to them may impact their life. Through roleplaying a bee consuming flowers, P9 associated

with their prior knowledge of bees: “Bees are keeping flowers and plants grow if we need to plant vegetables... and fruit to cook and to provide for our family... so I feel like bees are impacting this world very well. Bees are important to us. (We) need to eat food.” Then they understood the negative impact of filter bubbles: “It’s useless to other plants if they don’t get enough ‘energy’ to grow. I feel like it’s not fair. Because certain bees can only pick flowers that have the most ‘energy’. But if the flowers have the least ‘energy’, the bee most likely not gonna pick the flower. But if the flower has (‘energy’), it grows every time.”

The first-person view of roleplay creates an immersive experience in the biodiversity context, which is hard to form and observe in real life. Eight students (P1, P3, P5, P6, P7, P9, P2, P8) reported that they learned about the negative impact of both the thriving flowers close together and the dead flowers farther around: “Wow, these flowers died!” Such familiar phenomena immediately invited them to connect the loss of biodiversity in flowers with bees’ health. For example, P5 and P6: “At the beginning, (there is a) more live garden. The bee will then get ill with the dying flowers”; P1: “Oh, um, it needs more diverse flowers because it can’t live off just one specific type of flower. It needs a variety”; P9: “When the bee only had a few selections of flowers, it started to become unhealthy, cause it needs multiple different sets of flowers, not only two...” Such narrative supported them to connect the necessity of biodiversity back to the need for diverse content and opinions in real life. For instance, P3: “Because you’re not trying to see the same one thing over and over again. You get bored.”

5.3.3 Analogies and their inter-relations support students in explaining complex algorithms through intuitive and imaginative storytelling.

Analogies and their inter-relations support illustrating the algorithmic steps. With the concepts from a familiar domain and the structural relationships among them, it’s easier for students to recall the detailed algorithmic steps for content-based recommendations and diversification. Eight students were able to illustrate how the content-based recommendation works in the context of bee pollination. For example, P2 and P8: “Because when you choose the flowers, that beehive chooses (that flower), that beehive decides to make that that specific flower grow more. So you can pollinate specifically that type of flowers”; P2: “I roleplayed as a bee. I know I probably pollinated more diverse flowers. the more different flowers I pollinated, the bigger the filter bubble got.” Six students (P1, P2, P3, P4, P7, P9) demonstrated algorithmic steps for diversification. For example, P3: “You enlarge it, give it more space, then you give more pollen into different plants.”

Furthermore, three students (P2, P7, P9) understood the division of labor between AI technology, user, and AI engineer by connecting them with a garden, a bee, and an environmental scientist. For example, P2 explained how to break filter bubbles as a user: “I know I should pollinate more diverse flowers. The more different flowers I pollinated, the bigger the filter bubble got”; P7 and P9 explained the difference between the AI engineer and the user: “A little bit different because we’re both trying to build a garden, but he is a human, so he has to use the tools to use the setting. And I just pollinated for him to use his tools for it to grow.” However, from the interview and post-test, we noticed a challenge in transferring specific knowledge

of analogies in the garden scenario to the target AI terminologies. Seven students described AI phenomena with their own language without referring to the AI terminologies. Two students were also confused by too many analogies presented together and got lost.

Student-created imaginary analogies to fill in the design gaps. We observed that students get inspired by existing analogies and create their own imaginary analogy bridging the gaps in a more detailed breakdown of AI concepts without any BeeTrap analogy connected. Such creation further develops students’ AI literacy. For example, with no analogy for ranking numbers placed above the buds, P7 and P9 created a new analogy “energy”, aligned with the garden context, to concretize bud rankings: “The flowers close to the beehive are fully grown and there were a couple of other flowers that were like just dead. Cause the bees, we weren’t giving them enough energy.” This demonstrates their further understanding of the complex algorithmic step that ranks items based on user profile-item similarity. They then utilized “energy” they derived from content-based recommendation systems’ inner workings to effectively understand the diversification algorithm: “After enlarging the pollen circle size, bees can get more energy to more diverse flowers far away.”

5.3.4 Beehive as a less intuitive analogy for user profile. Based on the post-interviews, six out of nine students learned the definition of a user profile by observing the beehive. P3 and P4 explained in both the bee pollination context and the real-life recommendation system context: “It’s the home of the bee and the data that he gathered, the information The thing that we like to pollinate. It’s like selecting the things that you like (in YouTube)”; P5 and P6 achieved comparable understanding: “It represents (what) we select in a recommendation system. It depends on what you watched.”

However, after the learning activity, three students (P1, P2, P8) were not able to transfer their experience with the beehive to AI recommendation systems. This is not aligned with students’ high in-app performance in that they correctly understand a user profile collects data on user-selected items. This indicates the challenges for students to memorize the connection between the beehive (i.e., source knowledge) and the user profile (i.e., target knowledge), even when they understand the target knowledge during the analogical learning experience. This may be due to the gap between a beehive moving to pollinated flowers and the real science context.

6 DISCUSSION

6.1 Embodied Metaphors for AI Concepts

We found NEAR-FAR effective in supporting students to explore and reflect on filter bubbles, although students didn’t demonstrate all expected body movements in the filter bubble activity. Combined with ITERATION, students were able to take a closer look at the gradual formation of filter bubbles, compare the new and withered flowers between iterations, and reason about filter bubbles’ negative impact and inner workings. Findings indicate benefits for students to fully explore all the iterations back and forth, spend more time observing more prominent filter bubbles, and walk inside and outside of the pollen circle to observe the garden from different perspectives (CONTAINER). This is aligned with the design guideline for tangible learning interfaces that spatial, physical, and temporal properties can trigger learners’ reflection [8].

For learning the inner workings, some students failed to notice (1) the beehive moving toward the pollinated flowers (**NEAR-FAR (G)**), and (2) ranking based on the bud-beehive distance (**NEAR-FAR (P)**). One potential reason is that these spatial relationships are less visually accessible in a first-person view of roleplaying the bee. Thus, a global overhead view could be provided for a more comprehensive and accurate understanding of the spatial organization of data [21].

For diversification, findings show that **BIG-SMALL** and **NEAR-FAR (G)** support direct manipulation and trial & error with pollen circle and bud re-ranking. Over time, such hands-on experience with immediate visual feedback helps students develop metaphorical connections between image schemas and AI concepts [7, 70, 114, 122]. However, students didn't demonstrate significant improvement in their conceptual understanding of the diversification algorithm from pre- to post-tests. Post-interviews indicate that, although image schemas make the system intuitive and engaging for students to interact with, some students didn't fully transfer their in-app understanding into post-tests. This aligned with findings from a tangible interface teaching machine learning [59].

Furthermore, students didn't develop a deeper understanding of data diversity's mathematical definition with **NEAR-FAR (G)**—the average of distances between all pairs of data objects. Therefore, **NEAR-FAR (G)** can serve as an entry point using students' real-world spatial awareness capabilities [21, 63] to interpret relationships among data; meanwhile, more visual and tangible scaffolding is needed. For example, in the garden context, a spider web can be created to connect all pairs of flowers and the total length of the thread embodies the calculation of data diversity.

6.2 Analogies for Learning AI Literacy in Science Contexts

K-12 science topics as effective analogies for AI concepts. As an extension of prior work [43, 83], we found that BeeTrap analogies have notably enabled students to connect complex concepts of a recommendation system to examples in their real lives. For example, a majority of students were able to use the **beehive** analogy to describe the nature of the user profile. One participant demonstrated his understanding: "It's the home of the bee and the data that it gathered; it's similar to selecting things you prefer, as you would do on YouTube." The process of reasoning aligns with the stages in analogical learning in which the learner recalls the familiar example, identifies the parallels between the known and the unknown, and evaluates the analogy to see if it holds [44].

Familiar ethical concepts in K-12 STEM subjects could be a common ground for students who have limited access to AI technologies to understand AI ethical issues. For example, **biodiversity**, emphasizing the necessity for bees to have diverse flowers for maintaining their health overall, served as a powerful analogy to illustrate the importance of diversification in recommendation systems. This concept fostered students' critical thinking that, much like a bee's need for diverse flowers, we as consumers of recommendation systems also need diverse content. More K-12 topics could be identified as effective common ground, such as nutrition science (the analogy between information diversity and nutrition diversity).

Powerful analogy-supported narrative with challenges in transfer learning. We found that students recalled the detailed steps of content-based recommendations and diversification better when these concepts were anchored to familiar scientific context and their structural relationships, providing an intuitive narrative for recalling the complex knowledge. Students successfully illustrated the workings of content-based recommendation systems using the analogy of bee pollination, thereby highlighting an understanding of diversification in scientific discourse. Analogies' learning effects align with existing theory that emphasizes that schema abstraction and generalization from one's prior knowledge foster the formation of abstract rules in new learning scenarios [44], especially for perceiving algorithms [6, 16].

However, we observed a significant challenge in transferring specific concepts from garden analogies to the targeted AI terminologies, though all the groups were able to correctly recognize the AI concepts behind the major analogies while using BeeTrap. From the post-interviews, students could describe certain AI phenomena in their own words but struggled to use the correct AI terms, and the introduction of too many analogies at once tended to confuse them, detracting from the intended learning outcome. This is a known challenge in analogical learning [47] due to its context sensitivity and student expertise gap.

Metaphoric representations bridge the gaps in analogies. The analogies will eventually get disrupted because the source and target domains are essentially different. In BeeTrap, gaps exist between details in bee pollination and the underlying mechanism of a content-based recommendation, including the beehive moving based on the pollinated flowers, and artificial rankings of buds based on the physical distance. These imaginative narrations break scientific knowledge and may confuse students. However, on the other side, such metaphoric representations enabled by these artificial objects (e.g., moving beehive, pollen circle, bud rankings) are also inviting and inspiring for students to come up with new analogies for further learning. For example, one group interpreted pollen circle with their self-created concept of "energy", which supports buds closer to the beehive to grow, can be collected by the bee from flowers, and requires the bee to consume more diverse flowers. Students accurately explained the full algorithmic steps behind AI recommendation and diversification by using their design of "energy". This extends the finding from existing research that novice adults can intuitively understand recommendation algorithms through metaphoric representations [6]. Extensive research supports the learning benefits of storytelling [5], where connecting with prior experience and knowledge effectively enhances learning [80].

Furthermore, using a beehive to represent the user profile is designed by teachers during the iterative design process. The lower effectiveness of the beehive indicates a steep learning curve for novices to create effective ideas for learning AI concepts. Such a challenge may request new codesign methods for novices to contribute to embodied and analogical AI learning activities.

7 LIMITATIONS AND FUTURE WORK

There are a few major limitations of this work, including the analogy design in BeeTrap, the limited number of participants, and the unexpected study setup changes caused by the summer camp's

nature of flexibility. First, as discussed in Section 6.2, gaps exist between real-world bee pollination and AI recommendations. To bridge the gap, our design involves a few artificial objects (e.g., the beehive that moves to the pollinated flowers) which break the scientific knowledge. Although these analogies have been confirmed or designed by a few in-service K-12 science teachers, and findings don't indicate their harm to learning, we still consider them as a limitation in our design and plan to improve them in future design iterations. Second, with nine students from a summer camp voluntarily signed up to participate in the study, our sample size is too limited to conduct more statistical analysis and restricts the generalizability of the findings. Increasing the sample size to include more participants with diverse levels of prior knowledge and backgrounds could enhance the results' robustness. Third, student attendance of summer camps is relatively less controllable. Thus, we have to adjust the grouping for three activities based on who showed up on the day, and this is why the grouping for the diversification activity is different from the first two activities. Moreover, an unexpected occupation of the study site made P1 wait for a long time before she could play with the diversification. Although a researcher kept her accompany, such unavoidable disruption and waiting, however, may influence P1's interest and engagement with the learning experience.

Beyond our current scope, we haven't explored the unique learning benefits of different design dimensions: (1) AR versus non-AR, (2) embodied versus non-embodied, and (3) analogical versus non-analogical approaches. In the future, we could investigate the long-term effects of embodied metaphors on students' retention and transfer learning of AI concepts. Additionally, we can explore better interaction designs using existing image schemas and new embodied metaphors for more advanced AI concepts, including the mathematical definition of data diversity. Codesigning with AI experts, K-12 teachers, and children would allow us to refine and validate new designs.

8 SELECTION AND PARTICIPATION OF CHILDREN

We recruited middle school students from an ethnically and economically diverse urban school district during a summer camp. Before enrollment, we provided comprehensive information about the study to parents and youths through consent forms. We clarified these details in person and addressed any questions. Written consent for participation and video recording was obtained from the youths. We selected all youths who volunteered to participate, following protocols approved by our Institutional Review Board.

ACKNOWLEDGMENTS

We would like to thank the support from the National Science Foundation (NSF) RETTL program (Award No. 2238675), and the industry fellowship through the NSF NRT program (Award No. 1922591). We also appreciate the generous support from Mr. Jeremy Smith, Ms. Danielle Daniels, and Dr. April Luehmann to collaborate with the Freedom Scholars Learning Center, the David T. Kearns Center at the University of Rochester, and the Sodus School Summer Camp.

REFERENCES

- [1] Shivam Agarwal. 2013. Data mining: Data mining concepts and techniques. In *2013 international conference on machine intelligence and research advancement*. IEEE, 203–207.
- [2] Charu C Aggarwal. 2016. Content-Based Recommender Systems. In *Recommender Systems: The Textbook*, Charu C Aggarwal (Ed.). Springer International Publishing, Cham, 139–166.
- [3] Charu C Aggarwal. 2016. *Recommender Systems: The Textbook*. Springer.
- [4] Safinah Ali, Blakeley H Payne, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. In *International workshop on education in artificial intelligence k-12 (eduai'19)*, Vol. 2. 1–4.
- [5] Maxine Alterio and Janice McDrury. 2003. *Learning through storytelling in higher education: Using reflection and experience to improve learning*. Routledge.
- [6] Oscar Alvarado, Vero Vanden Abeele, David Geerts, Francisco Gutiérrez, and Katrien Verbert. 2021. Exploring tangible algorithmic imaginaries in movie recommendations. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–12.
- [7] Alissa N Antle, Greg Corness, and Milena Droumeva. 2009. What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers* 21, 1-2 (2009), 66–75.
- [8] Alissa N Antle and Alyssa F Wise. 2013. Getting down to details: Using theories of cognition and learning to inform tangible user interface design. *Interacting with Computers* 25, 1 (2013), 1–20.
- [9] Noraidah Sahari Ashaari, Hairulliza Mohamad Judi, Hazura Mohamed, Meriam Tengku Wook, et al. 2011. Student's attitude towards statistics course. *Procedia-Social and Behavioral Sciences* 18 (2011), 287–294.
- [10] Saskia Bakker, Alissa N Antle, and Elise Van Den Hoven. 2012. Embodied metaphors in tangible interaction design. *Personal and Ubiquitous Computing* 16 (2012), 433–449.
- [11] Joan Boykoff Baron and Robert J Sternberg (Eds.). 1987. Teaching thinking skills: Theory and practice. *Series of books in psychology*. 275 (1987).
- [12] Cordula Baur, Carolin Wienrich, and Jörn Hurtienne. 2022. Form Follows Mental Models: Finding Instantiations of Image Schemas using a Design Research Approach. In *Designing Interactive Systems Conference (Virtual Event, Australia) (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 586–598.
- [13] Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Mille Skovhus Lunding, Marie-Monique Schaper, Maarten Van Mechelen, Mariana Aki Tamashiro, Rachel Charlotte Smith, Ole Sejer Iversen, and Marianne Graves Petersen. 2022. Supporting Critical Data Literacy in K-9 Education: Three Principles for Enriching Pupils' Relationship to Data. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference (Braga, Portugal) (IDC '22)*. Association for Computing Machinery, New York, NY, USA, 225–236.
- [14] Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Mille Skovhus Lunding, Marie-Monique Schaper, Maarten Van Mechelen, Mariana Aki Tamashiro, Rachel Charlotte Smith, Ole Sejer Iversen, and Marianne Graves Petersen. 2022. Supporting critical data literacy in K-9 education: three principles for enriching pupils' relationship to data. In *Interaction Design and Children*. 225–236.
- [15] Engin Bozdog and Jeroen Van Den Hoven. 2015. Breaking the filter bubble: democracy and design. *Ethics and information technology* 17 (2015), 249–265.
- [16] Taina Bucher. 2019. The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. In *The Social Power of Algorithms*. Routledge, 30–44.
- [17] Rafael A Calvo, Dorian Peters, Karina Vold, and Richard M Ryan. 2020. Supporting human autonomy in AI systems: A framework for ethical enquiry. *Ethics of digital well-being: A multidisciplinary approach* (2020), 31–54.
- [18] 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* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–8.
- [19] Center for Countering Digital Hate. 2022. Deadly by Design: TikTok Pushes Harmful Content Prompting Eating Disorders and Self-Harm into Users' Feeds. [URLofthereportifavailable]. Accessed: [date of access].
- [20] John Chen, Mike Horn, and Uri Wilensky. 2023. NetLogo AR: Bringing Room-Scale Real-World Environments Into Computational Modeling for Children. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. 736–739.
- [21] Xin Chen, Jessica Zeitz Self, Leanna House, John Wenskovich, Maoyuan Sun, Nathan Wycoff, Jane Robertson Evia, Scotland Leman, and Chris North. 2018. Be the Data: Embodied Visual Analytics. *IEEE Trans. Learn. Technol.* 11, 1 (Jan. 2018), 81–95.
- [22] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* 118, 9 (2021), e2023301118.

- [23] J Clement. 2013. Roles for explanatory models and analogies in conceptual change. *International handbook of research on conceptual change* 2 (2013), 412–446.
- [24] Richard K Coll, Bev France, and Ian Taylor. 2005. The role of models/and analogies in science education: implications from research. *Int. J. Sci. Educ.* 27, 2 (Jan. 2005), 183–198.
- [25] Yun Dai, Ziyang Lin, Ang Liu, and Wenlan Wang. 2023. An embodied, analogical and disruptive approach of AI pedagogy in upper elementary education: An experimental study. *Br. J. Educ. Technol.* (Aug. 2023).
- [26] Joshua A Danish, Kylie Peppler, and David Phelps. 2010. BeeSign: Designing to support mediated group inquiry of complex science by early elementary students. In *Proceedings of the 9th International Conference on Interaction Design and Children*. 182–185.
- [27] Clifford De Raffaele, Serengul Smith, and Orhan Gemikonakli. 2018. An Active Tangible User Interface Framework for Teaching and Learning Artificial Intelligence. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) (IUI '18). Association for Computing Machinery, New York, NY, USA, 535–546.
- [28] Daniella DiPaola, Blakeley H Payne, and Cynthia Breazeal. 2020. Decoding design agendas: an ethical design activity for middle school students. In *Proceedings of the interaction design and children conference*. 1–10.
- [29] Daniella DiPaola, Blakeley H Payne, and Cynthia Breazeal. 2020. Decoding design agendas: an ethical design activity for middle school students. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) (IDC '20). Association for Computing Machinery, New York, NY, USA, 1–10.
- [30] Andrea A DiSessa. 2000. *Changing minds: Computers, learning, and literacy*. MIT Press.
- [31] Janice A Dole and Gale M Sinatra. 1998. Reconceptualizing change in the cognitive construction of knowledge. *Educational psychologist* 33, 2-3 (1998), 109–128.
- [32] Stefania Druga, Sarah T Vu, Eesh Likhith, and Tammy Qiu. 2019. Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019* (New York, NY, USA) (FL2019). Association for Computing Machinery, New York, NY, USA, 104–111.
- [33] Mehdi Elahi, Dietmar Jannach, Lars Skjærven, Erik Knudsen, Helle Sjøvaag, Kristian Tolonen, Øyvind Holmstad, Igor Pipkin, Eivind Throndsen, Agnes Stenbom, et al. 2022. Towards responsible media recommendation. *AI and Ethics* (2022), 1–12.
- [34] Barrett Ens, Maxime Cordeil, Chris North, Tim Dwyer, Lonni Besançon, Arnaud Prouzeau, Jiazhou Liu, Andrew Cunningham, Adam Drogemuller, Kadek Ananta Satriadi, and Bruce H Thomas. 2022. Immersive Analytics 2.0: Spatial and Embodied Sensemaking. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI EA '22, Article 91). Association for Computing Machinery, New York, NY, USA, 1–7.
- [35] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opin. Q.* 80, S1 (March 2016), 298–320.
- [36] Joseph L Fleiss, Bruce Levin, Myunghee Cho Paik, et al. 1981. The measurement of interrater agreement. *Statistical methods for rates and proportions* 2, 212–236 (1981), 22–23.
- [37] Zhaolin Gao, Tianshu Shen, Zheda Mai, Mohamed Reda Bouadjenek, Isaac Waller, Ashton Anderson, Ron Bodkin, and Scott Sanner. 2022. Mitigating the filter bubble while maintaining relevance: Targeted diversification with VAE-based recommender systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2524–2531.
- [38] Zhaolin Gao, Tianshu Shen, Zheda Mai, Mohamed Reda Bouadjenek, Isaac Waller, Ashton Anderson, Ron Bodkin, and Scott Sanner. 2022. Mitigating the Filter Bubble While Maintaining Relevance: Targeted Diversification with VAE-based Recommender Systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) (SIGIR '22). Association for Computing Machinery, New York, NY, USA, 2524–2531.
- [39] 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.
- [40] 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* (New York, NY, USA) (AI/ES '20). Association for Computing Machinery, New York, NY, USA, 272–278.
- [41] Dirk Geeraerts and Hubert Cuyckens. 2007. *The Oxford handbook of cognitive linguistics*. OUP USA.
- [42] Dedre Gentner, Sarah Brem, Ronald W Ferguson, Arthur B Markman, Bjorn B Levold, Phillip Wolff, and Kenneth D Forbus. 1997. Analogical Reasoning and Conceptual Change: A Case Study of Johannes Kepler. *Journal of the Learning Sciences* 6, 1 (Jan. 1997), 3–40.
- [43] Dedre Gentner, Jeffrey Loewenstein, and Leigh Thompson. 2004. Analogical encoding: Facilitating knowledge transfer and integration. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 26.
- [44] Dedre Gentner and Linsey A Smith. 2013. *Analogical Learning and Reasoning*. Oxford University Press.
- [45] Mary L Gick and Keith J Holyoak. 1983. Schema induction and analogical transfer. *Cogn. Psychol.* 15, 1 (Jan. 1983), 1–38.
- [46] Christoph Grüter and Francis LW Ratnieks. 2011. Flower constancy in insect pollinators: Adaptive foraging behaviour or cognitive limitation? *Communicative & integrative biology* 4, 6 (2011), 633–636.
- [47] Shiva Hajian. 2018. The Benefits and Challenges of Analogical Comparison in Learning and Transfer: Can Self-Explanation Scaffold Analogy in the Process of Learning? *SFU Educational Review* 11, 1 (2018).
- [48] Natali Helberger, Kari Karppinen, and Lucia D'acunto. 2018. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society* 21, 2 (2018), 191–207.
- [49] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. 2019. Can children understand machine learning concepts? The effect of uncovering black boxes. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–11.
- [50] Cindy E Hmelo, Douglas L Holton, and Janet L Kolodner. 2014. Designing to learn about complex systems. In *Design Education*. Routledge, 247–298.
- [51] 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.
- [52] Jörn Hurtienne and Johann Habakuk Israel. 2007. Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction. In *Proceedings of the 1st international conference on Tangible and embedded interaction* (Baton Rouge, Louisiana) (TEI '07). Association for Computing Machinery, New York, NY, USA, 127–134.
- [53] Jörn Hurtienne, Christian Stöbel, and Katharina Weber. 2009. Sad is heavy and happy is light: Population stereotypes of tangible object attributes. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*. 61–68.
- [54] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. *Recommender Systems: An Introduction*. Cambridge University Press.
- [55] Shagun Jhaver, Yoni Karpfen, and Judd Antin. 2018. Algorithmic anxiety and coping strategies of Airbnb hosts. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
- [56] Mark Johnson. 2013. *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago press.
- [57] Yasmin Kafai, Chris Proctor, and Debora Lui. 2020. From theory bias to theory dialogue: embracing cognitive, situated, and critical framings of computational thinking in K-12 CS education. *ACM Inroads* 11, 1 (2020), 44–53.
- [58] K Megasari Kahn, Rani Megasari, Erna Piantari, and Enjun Junaeti. 2018. AI programming by children using snap! block programming in a developing country. (2018).
- [59] Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, and Marianne Graves Petersen. 2021. The Machine Learning Machine: A Tangible User Interface for Teaching Machine Learning. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Salzburg, Austria) (TEI '21, Article 19). Association for Computing Machinery, New York, NY, USA, 1–12.
- [60] Yael Kesner Baruch, Ornit Spektor-Levy, and Nira Mashal. 2016. Pre-schoolers' verbal and behavioral responses as indicators of attitudes and scientific curiosity. *Internat. J. Math. Ed. Sci. Tech.* 14, 1 (Feb. 2016), 125–148.
- [61] Markus Kiefer and Natalie M Trumpp. 2012. Embodiment theory and education: The foundations of cognition in perception and action. *Trends in Neuroscience and Education* 1, 1 (Dec. 2012), 15–20.
- [62] Jeongah Kim and Jaekwoun Shim. 2022. Development of an AR-Based AI Education App for Non-Majors. *IEEE Access* 10 (2022), 14149–14156.
- [63] Scott R Klemmer, Björn Hartmann, and Leila Takayama. 2006. How bodies matter: five themes for interaction design. In *Proceedings of the 6th conference on Designing Interactive systems*. 140–149.
- [64] Christie Kodama, Beth St. Jean, Mega Subramaniam, and Natalie Greene Taylor. 2017. There's a creepy guy on the other end at Google!: engaging middle school students in a drawing activity to elicit their mental models of Google. *Information Retrieval Journal* 20 (2017), 403–432.
- [65] Tobias Kreienbühl, Richard Wetzel, Naomi Burgess, Andrea Maria Schmid, and Dorothee Brovelli. 2020. AR circuit constructor: combining electricity building blocks and augmented reality for analogy-driven learning and experimentation. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 13–18.
- [66] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems – A survey. *Knowl. Based Syst.* 123 (May 2017), 154–162.
- [67] Susan Lechelt, Yvonne Rogers, and Nicolai Marquardt. 2020. Coming to your senses: promoting critical thinking about sensors through playful interaction in classrooms. In *Proceedings of the interaction design and children conference*. 11–22.
- [68] 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*.

- 1–22.
- [69] Jeanne Liedtka. 2018. Why design thinking works. *Harvard Business Review* 96, 5 (2018), 72–79.
- [70] Robb Lindgren and Mina Johnson-Glenberg. 2013. Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational researcher* 42, 8 (2013), 445–452.
- [71] Megan K Littrell, Anne U Gold, Kristin LK Koskey, Toni A May, Erin Leckey, and Christine Okochi. 2022. Transformative experience in an informal science learning program about climate change. *Journal of Research in Science Teaching* 59, 6 (2022), 1010–1034.
- [72] Breanne K Litts, Apoorva Chauhan, Chase K Mortensen, and Kamaehu Matthias. 2019. I’m drowning in squirrels! How children embody and debug computational algorithms through designing mixed reality games. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 267–273.
- [73] Duri Long, Takeria Blunt, and Brian Magerko. 2021. Co-Designing AI Literacy Exhibits for Informal Learning Spaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 1–35.
- [74] Duri Long, Aadarsh Padiyath, Anthony Teachey, and Brian Magerko. 2021. The Role of Collaboration, Creativity, and Embodiment in AI Learning Experiences. In *Proceedings of the 13th Conference on Creativity and Cognition* (Virtual Event, Italy) (C&C ’21, Article 28). Association for Computing Machinery, New York, NY, USA, 1–10.
- [75] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. 2015. Recommender system application developments: a survey. *Decision support systems* 74 (2015), 12–32.
- [76] Manuela Macedonia. 2019. Embodied Learning: Why at School the Mind Needs the Body. *Front. Psychol.* 10 (Oct. 2019), 2098.
- [77] Laura Malinverni and Narcis Pares. 2014. Learning of abstract concepts through full-body interaction: A systematic review. *Journal of Educational Technology & Society* 17, 4 (2014), 100–116.
- [78] Machdel Matthee and Marita Turpin. 2019. Invited Paper: Teaching Critical Thinking, Problem Solving, and Design Thinking: Preparing IS Students for the Future. *Journal of Information Systems Education* 30, 4 (2019), 242–252.
- [79] Neema Moraveji, Jason Li, Jiarong Ding, Patrick O’Kelley, and Suze Woolf. 2007. Comicboarding: using comics as proxies for participatory design with children. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1371–1374.
- [80] Paul Henry Mussen, Leonard Carmichael, et al. 1983. *Handbook of child psychology*. Wiley.
- [81] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2 (Jan. 2021), 100041.
- [82] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web* (Seoul, Korea) (WWW ’14). Association for Computing Machinery, New York, NY, USA, 677–686.
- [83] T J Nokes and D M Belenky. 2011. Incorporating motivation into a theoretical framework for knowledge transfer. *Psychol. Learn. Motiv.* (2011).
- [84] Gregor M Novak. 2011. Just-in-time teaching. *New directions for teaching and learning* 2011, 128 (2011), 63–73.
- [85] Seymour A Papert. 2020. *Mindstorms: Children, computers, and powerful ideas*. Basic books.
- [86] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- [87] Michael J Pazzani and Daniel Billsus. 2007. Content-Based Recommendation Systems. In *The Adaptive Web: Methods and Strategies of Web Personalization*, Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 325–341.
- [88] Kylie Pepppler, Joshua Danish, Benjamin Zaitlen, Diane Glosson, Alexander Jacobs, and David Phelps. 2010. BeeSim: Leveraging wearable computers in participatory simulations with young children. In *Proceedings of the 9th international conference on interaction design and children*. 246–249.
- [89] Wichian Premchaiswadi, Pitaya Poompuang, Nipat Jongswat, and Nucharee Premchaiswadi. 2013. Enhancing Diversity-Accuracy Technique on User-Based Top-N Recommendation Algorithms. In *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*. 403–408.
- [90] Kevin J Pugh, Lisa Linnenbrink-Garcia, Kristin LK Koskey, Victoria C Stewart, and Christine Manzey. 2010. Motivation, learning, and transformative experience: A study of deep engagement in science. *Science Education* 94, 1 (2010), 1–28.
- [91] Jenny Radesky, Yolanda Linda Reid Chassiakos, Nusheen Ameenuddin, Dipesh Navsaria, et al. 2020. Digital advertising to children. *Pediatrics* 146, 1 (2020).
- [92] Yim Register and Amy J Ko. 2020. Learning machine learning with personal data helps stakeholders ground advocacy arguments in model mechanics. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*. 67–78.
- [93] Mitchel Resnick, Fred Martin, Robert Berg, Rick Borovoy, Vanessa Colella, Kwin Kramer, and Brian Silverman. 1998. Digital manipulatives: new toys to think with. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 281–287.
- [94] Mitchel Resnick and Brian Silverman. 2005. Some reflections on designing construction kits for kids. In *Proceedings of the 2005 conference on Interaction design and children*. 117–122.
- [95] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. *Recommender Systems Handbook*. Springer.
- [96] Marcus Rottschäfer. 2019. *classical-mds*. <https://github.com/shukali/classical-mds>
- [97] Johnny Saldaña. 2021. *The coding manual for qualitative researchers*. sage.
- [98] Marie-Monique Schaper, Rachel Charlotte Smith, Mariana Aki Tamashiro, Maarten Van Mechelen, Mille Skovhus Lunding, Karl-Emil Kjaer Bilstrup, Magnus Höholt Kaspersen, Kasper Løvborg Jensen, Marianne Graves Petersen, and Ole Sejer Iversen. 2022. Computational empowerment in practice: Scaffolding teenagers’ learning about emerging technologies and their ethical and societal impact. *International Journal of Child-Computer Interaction* (2022), 100537.
- [99] 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*. 298–309.
- [100] Daniel I Schwartz, Jessica M Tsang, and Kristen P Blair. 2016. *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them*. WW Norton & Company.
- [101] Lawrence Shapiro. 2014. *The Routledge Handbook of Embodied Cognition*. Routledge.
- [102] Alexander Skulmowski and Günter Daniel Rey. 2018. Embodied learning: introducing a taxonomy based on bodily engagement and task integration. *Cognitive research: principles and implications* 3, 1 (2018), 1–10.
- [103] Jaemarie Solyst, Alexis Axon, Angela Stewart, Motahhare Eslami, and Amy Ogan. 2022. Investigating Girls’ Perspectives and Knowledge Gaps on Ethics and Fairness in Artificial Intelligence in a Lightweight Workshop. In *Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022*, pp. 807–814. International Society of the Learning Sciences.
- [104] Ji Soo Yi, Rachel Melton, John Stasko, and Julie A Jacko. 2005. Dust & magnet: multivariate information visualization using a magnet metaphor. *Information visualization* 4, 4 (2005), 239–256.
- [105] Maria-José Sosa-Díaz and Jesus Valverde-Berrococo. 2022. Grounded theory as a research methodology in educational technology. *International Journal of Qualitative Methods* 21 (2022), 16094069221133228.
- [106] Ngss Lead States. 2013. *Next Generation Science Standards: For States, By States*. National Academies Press.
- [107] 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.
- [108] Aleksander Szvedek. 2019. The image schema: A definition. *Styles of communication* 11, 1 (2019), 9–30.
- [109] Rodney B Thiele and David F Treagust. 1994. An interpretive examination of high school chemistry teachers’ analogical explanations. *Journal of Research in Science Teaching* 31, 3 (1994), 227–242.
- [110] Cathy Tran, Brandon Smith, and Martin Buschkuehl. 2017. Support of mathematical thinking through embodied cognition: Nondigital and digital approaches. *Cogn Res Princ Implic* 2, 1 (Feb. 2017), 16.
- [111] Khai N Truong, Gillian R Hayes, and Gregory D Abowd. 2006. Storyboarding: an empirical determination of best practices and effective guidelines. In *Proceedings of the 6th conference on Designing Interactive systems*. 12–21.
- [112] Xintian Tu and Joshua Danish. 2023. Designing a Technology-enhanced Play Environment for Young Children’s Science Modeling Practice. In *Proceedings of the 2023 Symposium on Learning, Design and Technology* (Evanston, IL, USA) (LDT ’23). Association for Computing Machinery, New York, NY, USA, 60–69.
- [113] Xintian Tu, Joshua Danish, Chris Georgen, Megan Humburg, Bria Davis, and Noel Enyedy. 2019. Examining how scientific modeling emerges through collective embodied play. (2019).
- [114] Xintian Tu, Chris Georgen, Joshua Danish, and Noel Enyedy. 2020. Extended embodiment: Physical and conceptual tools in a mixed-reality learning environment as supports for young learners’ exploration of science concepts. (2020).
- [115] Unity. 2022. *About AR Foundation*. <https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@4.2/manual/index.html>
- [116] Sepehr Vakil. 2018. Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education. *Harvard educational review* 88, 1 (2018), 26–52.
- [117] Pedro M Valero-Mora and Rubén D Ledesma. 2011. Using interactive graphics to teach multivariate data analysis to psychology students. *Journal of Statistics Education* 19, 1 (2011).
- [118] Greg Walsh, Alison Druin, Mona Leigh Guha, Elizabeth Foss, Evan Golub, Leshell Hatley, Elizabeth Bonsignore, and Sonia Francel. 2010. Layered elaboration:

- a new technique for co-design with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1237–1240.
- [119] Ge Wang, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2022. 'Don't make assumptions about me!': Understanding Children's Perception of Datafication Online. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.
 - [120] Ge Wang, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2023. "Treat me as your friend, not a number in your database": Co-designing with Children to Cope with Datafication Online. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.
 - [121] Randi Williams, Hae Won Park, Lauren Oh, and Cynthia Breazeal. 2019. Popbots: Designing an artificial intelligence curriculum for early childhood education. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 9729–9736.
 - [122] Susan A Yoon, Karen Elinich, Joyce Wang, Christopher Steinmeier, and Sean Tucker. 2012. Using augmented reality and knowledge-building scaffolds to improve learning in a science museum. *International Journal of Computer-Supported Collaborative Learning* 7 (2012), 519–541.
 - [123] Helen Zhang, Irene Lee, Safinah Ali, Daniella DiPaola, Yihong Cheng, and Cynthia Breazeal. 2023. Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education* 33, 2 (2023), 290–324.
 - [124] Xiaofei Zhou, Jessica Van Brummelen, and Phoebe Lin. 2020. Designing AI Learning Experiences for K-12: Emerging Works, Future Opportunities and a Design Framework. (Sept. 2020). arXiv:2009.10228 [cs.CY]
 - [125] Xiaofei Zhou, Pei Xiong, Qinqin Xiao, and Zhen Bai. in press. OptiDot: An Optical Interface for Children to Explore Dot Product and AI Recommendation. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*.
 - [126] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web* (Chiba, Japan) (WWW '05). Association for Computing Machinery, New York, NY, USA, 22–32.
 - [127] 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* (Boise, ID, USA) (IDC '19). Association for Computing Machinery, New York, NY, USA, 121–132.