

©Copyright 2026

Aayushi Dangol

Designing Interactive AI Literacy Systems
that Support Children's
Critical and Reflective Engagement with Artificial Intelligence

Aayushi Dangol

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee:

Julie A. Kientz, Chair

Jason Yip

Daniela Rosner

Program Authorized to Offer Degree:
Department of Human Centered Design & Engineering

University of Washington

Abstract

Designing Interactive AI Literacy Systems
that Support Children's
Critical and Reflective Engagement with Artificial Intelligence

Aayushi Dangol

Chair of the Supervisory Committee:
Julie A. Kientz
Department of Human Centered Design & Engineering

Children are increasingly encountering AI systems in the tools and media that shape how they learn, play, communicate, and create. These technologies offer new opportunities for exploration, creativity, and personalized learning. At the same time, as AI becomes embedded in everyday contexts, there is a growing need for AI literacy that supports children in understanding how these systems operate within broader socio-technical ecosystems. However, many AI systems remain opaque, making it difficult to understand how outputs are generated, what assumptions are embedded in their design, and where their limitations lie. This opacity can hinder the development of accurate mental models and contribute to overestimation of AI capabilities.

Supporting children's AI literacy therefore involves more than explaining how AI works at a technical level. It requires helping children reason about how AI systems produce and represent knowledge, including evaluating AI outputs that may appear credible despite lacking factual grounding, how they reflect and reproduce social values and biases, and how their use depends on material infrastructures with environmental consequences. In this dissertation, I examine how children make sense of AI across epistemic, social, and material dimensions and investigate how interactive systems can support this process. I present the design and study of three systems: *AI Puzzlers*, *CLIP4KIDS*, and *EcoPrompt*.

Drawing on classroom observations and participatory design sessions, this dissertation shows how interactive systems can support children's critical and reflective engagement with AI through iterative experimentation, encounters with AI outputs that surface their own ways of organizing and interpreting information, and connections to their social and cultural worlds.

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	vii
Chapter 1: Introduction	1
1.1 Background and Motivation	2
1.2 Thesis Statements	4
1.3 Thesis Overview	5
1.4 Contributions	10
Chapter 2: Background and Related Work	13
2.1 K-12 AI Literacy	13
2.2 Epistemic, Material, and Social Implications of AI	18
2.3 Prior Work	23
2.4 Educational Theories	28
Chapter 3: AI Puzzlers: Supporting Children in Critically Evaluating AI Outputs	32
3.1 Preface	32
3.2 Introduction	32
3.3 AI Puzzlers: System Design & Development	35
3.4 Methods	43
3.5 Findings	47
3.6 Discussion	62
3.7 Limitations & Future Work	66
3.8 Summary of Contributions	67
Chapter 4: CLIP4KIDS: Supporting Children in Interrogating AI Bias	69

4.1	Preface	69
4.2	Introduction	69
4.3	CLIP4KIDS: System Design & Development	71
4.4	Methods	77
4.5	Findings	85
4.6	Discussion	94
4.7	Limitations & Future Work	97
4.8	Summary of Contributions	99
Chapter 5:	Ecoprompt: Supporting Children’s Reasoning About AI’s Environmental Costs	100
5.1	Preface	100
5.2	Introduction	100
5.3	EcoPrompt: System Design & Development	102
5.4	Methods	110
5.5	Findings	116
5.6	Discussion	123
5.7	Limitations & Future Work	127
5.8	Summary of Contributions	127
Chapter 6:	Discussion	129
6.1	Supporting Iterative, Hypothesis-driven Inquiry (T1)	132
6.2	Making Visible the Points of Alignment and Tension (T2)	135
6.3	Connecting Learning to Children’s Values and Lived Experiences (T3)	136
6.4	The K-12 AI Literacy Ecosystem: Governance, Teachers, and Families	138
6.5	Directions for Future Work	141
6.6	Contributions	145
6.7	Conclusion	146

LIST OF FIGURES

Figure Number		Page
2.1	AI literacy framework from Mills et al. (2024), reproduced from their report.	17
3.1	Overview of AI Puzzlers: (A) children first solve an ARC puzzle independently, then (B) test whether generative AI can solve the same puzzle, and finally (C) compare AI’s solution to its explanation to evaluate AI reasoning.	34
3.2	Comparison of the correct vs AI-generated solutions. The visual nature of AI Puzzlers makes AI errors easy to spot.	34
3.3	An example of an ARC puzzle with instructions for solving it. The correct answer shows the shortest bar is colored red and the tallest one is blue.	36
3.4	Annotated screenshot of Manual Mode in AI Puzzlers. The interface consists of several key components: (A) Puzzle Navigation allows children to switch between different puzzles. (B) Example Puzzles provide input-output pairs to illustrate transformation rules. (C) Before Grid presents the given input that children use to solve the corresponding After puzzle. (D) Edit Panel provides a set of tools for editing After grid. (E) After Grid serves as the solution space where children create their expected output based on the transformation rule. (F) Color Palette enables children to apply colors when solving the puzzle. (G) Clone Button transfers the Before Grid content to the After Grid for further modification. (H) Reset Button clears the After Grid to start over.	40
3.5	Screenshot of Assist Mode in AI Puzzlers. Certain interface elements are enlarged to highlight key interactive features, including (A) testing if AI can solve the puzzles by either answering or explaining, (B) selecting the number of examples shown to AI, (C) choosing AI model versions, and (D) providing hints to AI.	42
3.6	Children engaging with AI Puzzlers alongside adult facilitators.	44
3.7	Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI’s attempt at the same puzzle. (A) highlights the children’s correct solution, whereas (B) shows the AI’s incorrect attempt.	48
3.8	Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI’s attempt at the same puzzle. (A) highlights the children’s correct solution, whereas (B) shows the AI’s incorrect attempt.	49

3.9	Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI's attempt at the same puzzle. (A) highlights the children's correct solution, whereas (B) shows the AI's incorrect attempt.	50
3.10	Children iteratively refined their instructions to guide the AI towards solving the puzzle. The sequence showcases the increasing specificity in their hints and AI's corresponding outputs.	52
3.11	Children's Interaction with Assist Mode in AI Puzzlers. (A) Displays examples of example patterns to infer the transformation rule. (B) Shows the instruction children provided to the AI: "make a colored donut shape around the colors." (C) Illustrates the AI's generated output in response to their instruction.	54
3.12	Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI's attempt at the same puzzle. (A) highlights the children's correct solution, whereas (B) shows the AI's incorrect attempt. (C) provides the AI's explanation of how it solved the puzzle.	56
3.13	Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI's attempt at the same puzzle. (A) highlights the children's correct solution, whereas (B) shows the AI's incorrect attempt. (C) provides the AI's explanation of how it solved the puzzle.	57
3.14	Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI's attempt at the same puzzle. (A) highlights the children's correct solution, whereas (B) shows the AI's incorrect attempt. (C) provides the AI's explanation of how it solved the puzzle.	58
3.15	The correct solution (left) is compared with AI-generated attempts (right).	60
3.16	The correct solution (left) is compared with AI-generated attempts (right).	61
4.1	Screenshot of the Workcloud interface with added annotations highlighting system components. The system captures live image inputs from a webcam (A) and visualizes the AI model's classification as a dynamic word cloud (B).	74
4.2	Screenshot of the Classifier interface with added annotations highlighting system components. The interface allows users to upload an image for classification (A), define labels describing the image (B), and view the AI model's classification decision along with confidence scores for each label (C).	75
4.3	Screenshot of the Smartlock interface with annotations highlighting key system components. Users can upload an image for classification (A), define inclusion/exclusion categories (B), set a decision threshold(C), and view the AI model's classification confidence scores (D) and final classification decision (E).	76
4.4	A representation of the Smartlock interface in terms of architecture and user flow.	78
4.5	Children prompting the Wordcloud interface with different types of visual inputs in Session 1.	81
4.6	Children explore the Classifier Interface in Session 2 by assigning descriptive labels to their monster creations and analyzing the system's classification result.	82

4.7	Children explore assigning inclusion and exclusion categories while experimenting with decision thresholds to design a Smartlock that grants access to their monster creations. In this example, the Smartlock interface classifies the image input as ‘not monster’ with 96% confidence, preventing the children’s monster creations from entering.	83
4.8	Pictures generated by DALL·E 2 for the prompt “ <i>Draw a picture of children celebrating a festival.</i> ”	84
4.9	Student-created representations of friendly monsters, illustrating the diversity in interpretations of ‘friendliness.’	86
4.10	Students investigated the impact of decision thresholds on AI classification, engaging in an iterative process to balance precision and inclusivity. (A) A high decision threshold (90%) led to overly restrictive filtering, excluding even closely matching inputs. (B) A low threshold (5%) resulted in over-admission, allowing broad inclusivity.	87
4.11	A screenshot of the Wordcloud interface, which includes a single instance of the word ‘wife’ among other classification outputs for the image on the right.	89
4.12	Drawings by children for the prompt “ <i>Draw a picture of children celebrating a festival</i> ” from Session 4.	92
5.1	Footprint Calculator interface, including (1) user prompt entry box, (2) an "Ask AI" button that submits the prompt to the model, (3) an output box in which the AI’s response is displayed, (4) a display of the last prompt’s estimated water, carbon, and energy costs, and (5) a display of the total water, carbon, and energy costs across the session, with status-bar style indicators of cumulative use.	103
5.2	Farm game interface showing (1) initial welcome screen, (2-4) gameplay across three levels. In each level, players face various challenges. For example, in (4), pest attacks introduce critical decision points where players can use AI to make crop protection more manageable as levels progress.	104
5.3	Tutorial illustrating the AI query lifecycle and its environmental consequences in EcoPrompt, demonstrating how a user prompt travels from a personal device to a data center that consumes electricity, resulting in carbon emissions.	106
5.4	An illustration of crops growing as a result of players’ attention to planting, watering, and harvesting them.	108
5.5	An illustration of the game panel at the start of the first level, showing (1) the sidebar displaying the lake health with a status log communicating changes to the player, (2) the main array of tools available to the player, including the planting and watering tools, and (3) the inventory and notifications panel.	109

5.6	The interface for level 2 of the Farm Game, showing (1) the overall UI, (2) the AI Farm Hand button, (3) the Farmer’s Almanac button, (4) the Farmer’s Almanac hint message displayed after pressing the button, (5) the AI Farm Hand interface, (6) an AI Usage Warning displayed before sending the user’s message to the AI Farm Hand, and (7) the response generated by the AI Farm Hand, suggesting what to plant. Levels 3 and 4 retain the Farm Hand and Almanac buttons.	110
5.7	An illustration of the pest control scenario encountered in Level 3, showing (1) the animated pests the player must remove from the crops, (2) the pest removal minigame triggered by clicking on the pest sprite, and (3) the pest control interaction with the AI Farm Hand, which can remove a pest at the cost of some of the player’s natural resources.	111
5.8	Child participants in the KidsTeam UW co-design group play the Farm Game.	114
5.9	Children’s illustrations of their understanding of how AI consumes energy. Illustration (1) traces energy consumption due to local AI use to a powerplant and then to renewable energy sources. Illustration (2) traces AI use to power lines fueled by solar energy. To improve readability, the text shown below each drawing reflects the child’s own written explanation, digitized by the research team while preserving the child’s original wording.	117

LIST OF TABLES

Table Number	Page
1.1 Mapping of dissertation studies to AI dimensions and corresponding thesis claims (T1–T3).	7
3.1 Reported Child Participant Details	43
4.1 Design Ideas for AI System and their Contributors	91
5.1 Reported Child Participant Details (all names are pseudonyms). AI Use Context codes: F=Use with friends or family; S=Self.	112

ACKNOWLEDGMENTS

First and foremost, I want to thank my advisor, **Dr. Julie A. Kientz**, for being a role model and such a steady source of support throughout my PhD. I have always appreciated how open she was to the many directions my work could take, while knowing when to push me forward. Her reassurance that mistakes are part of the process made it possible for me to take intellectual risks, and because of her, each year I felt my sense of what I am capable of expand.

I owe a great deal to **Dr. Jason Yip** for his mentorship. Being welcomed into *KidsTeam UW* was one of the most formative experiences of my PhD. His habit of asking “so what?” pushed me to continually reflect on why my work mattered, a question I now carry into every paper I write. Beyond research, he gave generously of his time and personal wisdom. His stories had a way of staying with me long after our conversations ended.

I am deeply grateful to my supervisory committee. **Dr. Daniela Rosner** has been a valued mentor, and I was lucky to have her as my design methods instructor in my first year. **Dr. R. Benjamin Shapiro** served as my GSR, and I am grateful for his time and thoughtful feedback throughout the important milestones of my PhD.

I have also been lucky to work alongside many academic mentors who made this PhD both intellectually rich and joyful. Some of the most fun I have had came from working with **Dr. Cayla Key**. Together we made what felt like an endless number of handmade books and sewed “monster” capes for a study I will not forget anytime soon. **Dr. Jin Ha Lee** has been a warm presence, and I am especially grateful that our work together extended beyond UW to foundry10, where collaborating in a different setting was its own kind of gift. **Dr. Katie Davis** welcomed me into her child-computer interaction course as part of

the teaching practicum and mentored a project on teacher use of AI that became one of the most meaningful threads of my research. **Dr. Sayamindu Dasgupta** mentored me through one of my very first research projects, patiently walking me through things I had never encountered before.

I have also been incredibly fortunate to be surrounded by peers who made this PhD meaningful. To the members of **CHILL Lab**, I am grateful for the many conversations and for their feedback on my work. **Nisha Devasia** was always there, showing up in both the everyday moments and the difficult ones. **Meghna Gupta** brought rigor and care to our projects and brainstorming ideas together has been incredibly rewarding. **Dr. Hyewon Suh** has been like a big sister throughout this PhD, and I am especially grateful we had the chance to work together on the AI Institute projects. Her warmth, and the way she brings people together, shaped not just the work, but the kind of space our lab became. I have also been fortunate to collaborate with **Runhua Zhao**, **Michele Newman**, **Dr. Caroline Pitt**, **Daeun Yoo**, **Jordyn Padzensky**, **Arya Thiruvillakkat**, **Dr. Regina Cheng**, **Aaleyah Lewis**, **JaeWon Kim**, and **Rotem Landesman**, each of whom made this journey more thoughtful and collaborative. And to **Pitch Sinlapanuntakul**, **Donghoon Shin**, and my HCDE cohort, thank you for your support in figuring out classes and milestones together.

I would not have pursued a PhD without mentors who, long before graduate school, shaped the kind of researcher I became. **Dr. Kevin Webb** first introduced me to computer science research, opening a path I had not yet imagined for myself. **Dr. Kristin A. Searle** responded to a cold email and took a chance on me, teaching me not only how to be rigorous in my work, but how to approach research with care and thoughtfulness. **Dr. Natalie Mera Ford** helped me find my voice as a writer, often over Nepalese snacks at the Writing Center. **Dr. Nsoki Mavinga** was an anchor for me at Swarthmore; her door was always open, and her belief in me never wavered. **Dr. Patricia Reilly's** art history classes taught me to look closely and to resist easy explanations, a habit of mind I carry into my research. **Professor**

Randall Exon helped me see the world differently through painting, and although my path took me elsewhere, my desire to create, and to notice beauty in the world around me, is something I carry with me because of him.

This dissertation would not exist without the children, families, and educators who chose to spend their time with me. Thank you for inviting me into your classrooms, community spaces, and lives. Your questions, stories, creativity, and perspectives continually helped me to think about AI, learning, and design. It has been a privilege to learn alongside you, and I hope this work honors the generosity, curiosity, and care you brought to every study.

To **Liv Elmore**, from Swarthmore through all the years since, thank you for being such a constant in my life. I will always treasure our friendship. To **Marie Desmons** and **Gregory Rotello**, thank you for your friendship and for creating a space where I always felt welcome and at ease. Gregory, you are deeply missed, and I hope this work will make you proud. To **Kunsang Choden**, my friend since seventh grade in Kathmandu, it meant so much to find ourselves again as fellow PhD students at UW; thank you for many warm Nepali meals, and the long walks. To **Riddhi Divanji**, thank you for being both a friend and a mentor, and for always being in my corner. And to **Isabel**, **Yifang**, and **Rachel**, thank you for making new places feel familiar and for all the adventures along the way.

I have also been carried by the love of my family in more ways than I can name. To my parents, thank you for instilling in me a deep curiosity about the world and the confidence to find my own way. To my sister, thank you for keeping me connected to the world outside of work, for the humor, for the Gen Z updates so I am not always behind, and for making life feel full of hope. To my grandparents, aunts, and fufu, thank you for the care, encouragement, and sense of home you have always given me, no matter the distance. And to my mother-in-law and father-in-law, thank you for welcoming me so fully into your family and for your steady support along the way.

Lastly, to my husband, **Dr. Robert Wolfe**. You are my greatest gift. Thank you for

loving me so unconditionally and for always reminding me to keep going. I am endlessly grateful to share my life with you, and I look forward to everything we will build together.

DEDICATION

For Robert.

Chapter 1

INTRODUCTION

Artificial intelligence (AI) has reshaped the way people access and interact with information, demanding new approaches to governance, ethics, and literacy across many domains, but perhaps most notably in education. Around the world, governments are advancing initiatives to harness AI for learning and workforce readiness, with a growing emphasis on AI literacy, responsible integration, and teacher training [109, 302, 149, 128, 3]. In these efforts, AI literacy is increasingly framed not only as essential for economic competitiveness, but also for sustaining democratic participation in an increasingly algorithmic society [149].

At the same time, this growing policy emphasis exists alongside, and at times in tension with, the rapid acceleration of AI adoption in educational settings. Recent efforts have moved beyond advocacy toward implementation. For example, an April 2025 U.S. Executive Order outlines multiple pathways for integrating AI into K–12 education, including training educators to incorporate AI tools into instruction and introducing students to AI concepts early in their schooling [149]. While this emphasis on “AI readiness” signals important progress, it also raises critical questions about what it means to prepare children for a world in which AI is ubiquitous. These include what perspectives deserve attention in AI literacy curricula, how to translate national or state-level policy into meaningful learning experiences, and how to equip teachers and students with the tools they need to adopt (or, in some cases, resist) AI in the classroom? In this dissertation, I address these pressing sociotechnical questions through the design and study of interactive AI literacy systems that support children in critically engaging with AI across its epistemic, social, and material dimensions.

1.1 *Background and Motivation*

Children and adults have long navigated the integration of new technologies into everyday life. As digital technologies such as video games and social media have become embedded in daily routines, they have prompted ongoing discussions, and at times heightened concern, about how they might shape learning, behavior, and social relationships [152, 239]. Research in child development and media studies has documented both potential risks, including impacts on attention, sleep, and well-being [191, 311], as well as meaningful benefits, such as supporting creativity, learning, and access to information [239, 174]. Within this context, educators and families have often been tasked with helping children develop balanced and intentional relationships with emerging technologies, rather than simply permitting or prohibiting their use [139, 74, 93]. The recent rise of AI, and in particular generative and other general-purpose AI systems, introduces a new phase of this ongoing negotiation.

Compared to many earlier technologies, contemporary AI systems can generate content, respond dynamically to user inputs, and shape the information that people encounter in ways that are less bounded and more difficult to anticipate. These capacities have introduced new complexities into the epistemic landscape in which children encounter AI systems. For example, generative AI produces outputs that are fluent, coherent, and authoritative, yet may be factually incorrect, a phenomenon commonly described as hallucination [16]. In educational contexts, such outputs pose an epistemic risk [10, 157], as learners may accept AI-generated hallucinations as accurate despite it not being grounded in factual knowledge [284]. This concern is especially pronounced when students turn to AI systems while uncertain or without prior domain knowledge, and therefore most susceptible to accepting incorrect or misleading information. These risks are further amplified by the growing presence of low-quality, AI-generated content, sometimes referred to as “AI slop,” across digital platforms [157, 126]. Such content can include factual errors, oversimplifications, and misleading explanations, while maintaining the appearance of credibility and achieving high levels of visibility through engagement-driven algorithms such as Youtube Kids [94, 218]. As a result, children are

increasingly navigating an information environment in which AI-generated outputs are not only tools for interaction but also pervasive sources of knowledge.

Beyond epistemic concerns, AI systems also raise important social implications, particularly in relation to algorithmic bias and representation. Prior work has shown that children’s understandings of social categories are shaped through exposure to cultural representations across media, including books, television, and digital content [83, 176]. As generative AI becomes increasingly embedded in children’s creative and media environments, it participates in and extends this broader representational ecosystem [38, 76]. However, because generative AI systems are trained on large-scale datasets that reflect existing social patterns, their outputs often mirror and re-inforce those patterns, including historically embedded biases and stereotypes [129, 50, 22, 135]. Recent work examining AI tools designed for children’s digital storytelling found that such systems produced character representations that aligned with conventional gender norms and appearances, and in some cases were considered inappropriate for child use due to the risk of amplifying harmful biases [18, 48]. Emerging research further suggests that these issues extend beyond individual outputs to broader misrepresentations of youth themselves. For example, Wolfe et al. found that large language models disproportionately associate adolescents with violence, risk-taking, and hypersexualized behavior, whereas adolescents’ own accounts emphasize everyday experiences such as friendships, school routines, and family life [331]. Notably, AI-generated associations showed no correlation with adolescents’ self-perceptions, highlighting a significant misalignment between how AI systems represent young people and how they understand their own lives [331]. From an AI literacy perspective, this highlights the importance of supporting children in recognizing that AI-generated outputs are not neutral or authoritative, but are shaped by the data and assumptions underlying these systems. Developing such understanding is essential for enabling children to question algorithmic representations, identify potential biases, and recognize when AI systems may produce outputs that disadvantage or misrepresent themselves or others.

In addition to epistemic and social considerations, AI systems also have important material implications that are often invisible to users. While prior work has documented the significant

energy, water, and infrastructure required to support contemporary AI systems [203, 156], these underlying processes are largely abstracted away in everyday interactions. As a result, AI systems can appear immaterial and consequence-free, creating a disconnect between use and impact that affects users broadly. For children, this disconnect may be further pronounced. AI systems are typically encountered through seamless and responsive interfaces that obscure the underlying infrastructure, making it difficult to recognize that these systems rely on resource-intensive processes. The distributed nature of AI infrastructure further complicates efforts to relate individual interactions to broader environmental consequences [270, 190]. From an AI literacy perspective, this highlights the importance of supporting children in recognizing that AI systems are not only tools for generating information, but are also part of complex material systems with real-world environmental impacts. Developing such understanding can enable children to connect their everyday use of AI to larger systems of resource consumption and to reason more critically about the environmental implications of their technology use [29].

Thus, these epistemic, social, and material dimensions highlight that engaging with AI is not only a matter of learning how to use these tools, but of developing the competencies to interpret, question, and situate these systems within the broader contexts that shape their design and impact. AI literacy, in this sense, extends beyond understanding how AI systems function or how to use them effectively, and includes supporting children in making sense of how AI mediates knowledge, constructs representations, and is embedded within material systems with real-world consequences [217]. My dissertation explores this perspective by examining how children engage with AI across these dimensions and by designing learning experiences that foreground critical, exploratory and systems-oriented ways of understanding AI.

1.2 Thesis Statements

Building on the epistemic, social, and material challenges outlined above, in this dissertation I take a design-oriented approach to AI literacy. Through the design and study of three

interactive AI literacy systems, I examine how children develop critical and reflective understandings of AI. The claims of this dissertation are summarized in the following thesis statements:

- T1:** AI literacy systems that encourage iterative experimentation, where children can freely manipulate inputs and parameters, have the potential to deepen children’s understanding of how AI models respond to variation and human input. By allowing children to refine their inputs, test hypotheses, and observe the relationship between their choices and the AI’s outcomes, such systems can transform interaction into a form of inquiry. Immediate feedback further supports this process, enabling children to see the effects of their actions and gradually build understanding about how AI systems operate.
- T2:** AI literacy systems that enable children to compare their own ways of organizing and interpreting information with those of an AI model have the potential to facilitate reflection on both human and machine reasoning. By making visible the points of alignment and tension between children’s schemas and AI outputs, such systems can foster metacognitive reflection, inviting children to interrogate the AI’s underlying patterns of representation and inference, and reflect on how their own experiences shape their judgments.
- T3:** AI literacy systems that connect learning activities to children’s own identities and lived experiences have the potential to foster critical engagement with AI. Drawing on critical pedagogy’s emphasis on personal experience as a catalyst for critical consciousness [313], such systems can invite children to question AI’s underlying assumptions and reflect on how technological decisions intersect with their own social and cultural worlds.

1.3 Thesis Overview

In this dissertation, I report on three research studies that examine how children reason about the epistemic, social, and material dimensions of AI systems, and evaluate design approaches

that support the development of AI literacy across each dimension. I combine participatory design, iterative prototyping, and classroom-based deployments to explore how children make sense of AI through direct engagement with interactive systems that surface its behavior, limitations, and broader impacts.

In the first study, which addresses the epistemic dimension, children engage with visual puzzles that make generative AI’s reasoning errors visible. By comparing AI outputs against their own solutions, they identify inconsistencies and hallucinations in AI outputs. In the second study, which addresses the social dimension, children interact with an AI classification system where they define inclusion and exclusion categories, adjust decision thresholds, and observe how their choices shape model outputs. This process supports reflection on how bias emerges through both historical data patterns and human decisions. In the third study, which addresses the material dimension, children engage with an interactive game and footprint calculator that make the environmental costs of AI tangible.

The sections below provide a brief overview of each study and describe how these investigations contribute to the thesis statements. Table 1.1 summarizes how each study maps onto the core dimensions and my thesis statements. Chapter 6 then synthesizes the themes that cut across all three studies – including how children recognize and respond to AI errors, how they interpret and question AI bias and representational harms, and how they connect everyday AI use to broader material and environmental systems – and draws out their collective implications for the design of AI literacy tools for children.

1.3.1 Chapter 2: Review of Prior Literature

In this chapter, I situate this dissertation within a broader body of work that examines how children encounter, interpret, and make sense of AI systems in their everyday lives. I begin by drawing on research in K–12 AI literacy that defines the competencies, practices, and values needed for learners to engage with AI systems, with particular attention to work that frames AI literacy in terms of understanding, evaluating, and using AI. Building on this foundation, I then bring together scholarship that examines the epistemic, social, and

	Dimension	T1	T2	T3
AI Puzzlers Study	Epistemic	✓	✓	
CLIP4KIDS Study	Social	✓	✓	✓
Ecoprompt Study	Material	✓		✓

Table 1.1: Mapping of dissertation studies to AI dimensions and corresponding thesis claims (T1–T3).

material implications of AI systems, showing how AI outputs can appear authoritative even when they are inaccurate, how bias emerges through both data and human decision-making, and how the environmental and infrastructural costs of AI remain largely invisible despite their real-world consequences.

Across these strands of work, a common challenge emerges in supporting learners not only in understanding and interacting with AI systems, but in developing the competencies to engage in AI literacy practices and reasoning about their broader epistemic, social, and material implications. In response, I draw on prior work in designing interactive systems that support children in engaging with AI through direct interaction, including systems that enable learners to build, test, and interrogate models, as well as approaches grounded in systems thinking that support reasoning about distributed and often hidden material impacts. Finally, I ground this work in educational theories, including experiential learning, critical pedagogy, and multimedia learning, which together emphasize learning as an active and reflective process. This body of work informs the design-oriented approach of this dissertation, in which interactive systems are used to support children in developing critical and reflective understandings of AI across its epistemic, social, and material dimensions.

1.3.2 Chapter 3: Evaluating AI Reasoning

In this study, I examine how children ages 6–11 can be supported in critically evaluating AI outputs. Working with a team of collaborators, I designed *AI Puzzlers*, an interactive system in which children solve visual reasoning puzzles and compare their solutions with those generated by AI. The system draws on the Abstraction and Reasoning Corpus (ARC) dataset [55], which features reasoning puzzles that are typically easy for humans but challenging for AI. This contrast creates opportunities for children to encounter and interrogate cases where AI produces incorrect outputs, supporting reflection on the reliability and limits of AI-generated information.

To evaluate *AI Puzzlers*, I conducted two participatory design sessions with 21 children using a cooperative inquiry approach [105]. I found that when generative AI made mistakes, especially on puzzles children considered easy, they reacted with surprise and amusement, sparking meaningful dialogue. Across sessions, I found that children moved from identifying AI errors to developing evaluative strategies, such as comparing AI outputs against AI explanations and offering hints to iteratively guide its responses. Notably, children’s engagement with *AI Puzzlers* was not solely tied to catching AI mistakes. They were also drawn in by the puzzles themselves, often describing more complex puzzles as fun and viewing them as opportunities for personal accomplishment. Children’s continued engagement with *AI Puzzlers* highlights the importance of designing AI systems that present information in ways that facilitate comparison, encourage reflection, and scaffold multiple ways of understanding. These findings support T1 and T2 by showing how iterative experimentation and comparison between children’s own reasoning and AI outputs can scaffold critical evaluation of AI-generated information.

1.3.3 Chapter 4: Interrogating AI Bias

In this study, I worked with a team of collaborators to investigate how children (ages 9–10) understand and reason about bias in AI systems. To support this inquiry, I designed

CLIP4KIDS, an interactive tool that engages children in AI classification tasks to explore how bias emerges in model outputs. The system foregrounds the role of human decision-making in shaping classification processes, making visible how choices such as category definitions and labeling influence outcomes. This design is grounded in prior work demonstrating that human schemas and classification systems are not neutral, but are shaped by cultural, social, and individual experiences, which can in turn produce biased AI outputs [34].

To explore children’s interactions with *CLIP4KIDS*, I conducted a four-session classroom study with two fifth-grade classrooms at a local elementary school in Seattle, WA. I found that children engaged in iterative experimentation with *CLIP4KIDS* by tinkering with classification parameters and testing how small changes could significantly alter the AI’s outputs. Through these interactions, children began to recognize how classification systems depend on human-defined categories and thresholds, and how these decisions can privilege certain interpretations while excluding others. Building on these experiences, children drew on their own understandings of fairness to interpret and critique AI behavior, using the system to reason about patterns of inclusion and exclusion. These findings provide support for T1, T2, and T3, demonstrating how systems that support iterative experimentation, make children’s reasoning comparable to that of an AI model, and connect to children’s lived experiences can collectively foster inquiry, reflection, and critical engagement with AI.

1.3.4 Chapter 5: Reasoning about AI’s Environmental Impact

In this study, I examine how children ages 6–12 can be supported in reasoning about the environmental impacts of generative AI. Working with a team of collaborators, I designed *Ecoprompt*, an interactive system that combines a prompt-level environmental footprint calculator with a farm simulation game to help children connect everyday AI use to its underlying resource costs. The footprint calculator provides real-time estimates of the energy, water, and carbon associated with children’s AI prompts, while the farm game situates AI use within a shared virtual ecosystem where the collective consequences of AI use across all players become visible over time.

To evaluate *Ecoprompt*, I conducted two participatory design sessions with 16 children using a cooperative inquiry approach [105]. I found that prior to interacting with *Ecoprompt*, children distinguished between device-level electricity use and the computational processes they associated with AI, often treating AI as operating primarily on information rather than ongoing material resources. As children engaged with the Footprint Calculator, they developed more systems-oriented accounts that connected AI use to distant data centers, cooling infrastructure, and downstream environmental consequences. Children used these emerging understandings to evaluate whether particular prompts were "worth" their environmental cost, weighing the usefulness of a question against the resources it consumed. In the simulation game, children responded strongly to signs of collective environmental degradation, at times choosing to forgo AI assistance in order to reduce shared harm, even when doing so made tasks more difficult. At the same time, children recognized the limits of individual control, reasoning about how AI use accumulates across many users and how environmental outcomes extend beyond any single person's decisions. These findings support T1, and T3 by demonstrating how systems that make AI's resource use visible at the point of interaction, scaffold reasoning across individual and collective scales, and connect to children's existing values around fairness and responsibility can support engagement with the material dimensions of AI.

1.4 Contributions

1. Empirical contributions: Through classroom observations, video recordings, worksheets, and system logging, I have collected multiple datasets on how children (ages 6 - 12) reason about and interact with AI systems. These data capture both children's situated interactions and their evolving ideas as they engage with AI across different contexts. The inclusion of system logs extends these observations by providing interaction-level traces (e.g., changes to inputs and parameters), enabling closer analysis of how children iteratively test and revise their thinking. Across these data sources, I contribute empirical findings on how children form and revise hypotheses during AI interaction, and how they draw on their identities, prior knowledge, and values when making sense

of AI systems.

2. **Artifact contributions:** To support children’s AI literacy, I designed, implemented, and evaluated three interactive systems: *AI Puzzlers*, *CLIP4KIDS*, and *Ecoprompt*. These systems are accessible without prior AI or programming knowledge and can be used through a web browser, enabling their use across both classroom and informal learning contexts. A central design goal across all three systems is tinkerability. To support this, the systems provide immediate feedback, allowing children to observe the effects of their actions, and enable fluid experimentation, making it easy to iteratively test, revise, and reconnect ideas through interaction [261]. The systems are publicly accessible, and the code is available to support reuse and adaptation. Thus, one contribution of this work is to provide examples of designs that lower barriers to engaging with the epistemic, social, and material dimensions of AI, supporting children in exploring these systems through hands-on, iterative interaction.

3. **Broader Impact:** Beyond the individual studies, this dissertation contributes a set of pedagogical tools that have begun to be taken up across classroom and community contexts. *Ecoprompt*, for example, was implemented in a sixth-grade classroom at the Evergreen School as part of an AI unit, extending its use beyond the original study. Similarly, *AI Puzzlers* has been adapted across multiple settings, including an undergraduate philosophy course at the University of Washington to support discussions of machine reasoning and intelligence, as well as UW Discovery Days, where children and families engaged with the system in a public outreach context. Together, these deployments across K–12 classrooms, undergraduate courses, and informal learning environments suggest that these tools can be flexibly adapted to different pedagogical goals and audiences, positioning them as reusable resources for teaching and learning about AI. Findings from *CLIP4KIDS*’ deployment at University Child Development School further extend this contribution by demonstrating that AI literacy does not need to be introduced as a standalone technical unit. In a fifth-grade classroom, students

drew on their existing DEI curriculum, including their reading of *Stamped*, to interpret and critique algorithmic bias. This illustrates a model of curricular integration in which AI concepts are introduced through and alongside existing instructional commitments, rather than as an isolated topic. This contribution also aligns with findings from complementary work with K–12 educators, where teachers described uncertainty about how to teach AI literacy and a need for resources that can fit within existing classroom practices.

Chapter 2

BACKGROUND AND RELATED WORK

2.1 *K-12 AI Literacy*

While empirical data on children’s use of generative AI is emerging, early surveys suggest that children and youth are engaging with these systems at increasing rates, often surpassing adult adoption [45, 214]. Within human–computer interaction (HCI), research has begun to examine how AI systems can support children’s creativity, storytelling, and learning, with much of this work focusing specifically on generative AI applications. For example, prior studies have explored the use of generative AI to support children’s creative expression [201, 230, 260], storytelling [138, 345], and learning [49, 287, 336, 96, 75]. At the same time, scholars have begun to examine potential risks associated with children’s interactions with AI systems. These include biases embedded in models that may shape how children and teenagers are represented [331, 330], as well as socio-emotional concerns related to how children perceive and relate to AI systems, including anthropomorphism [103, 12], overreliance, and the emergence of companionship dynamics [73]. As AI systems, including both generative and general-purpose AI technologies, become increasingly embedded in children’s everyday learning, creative, and information-seeking practices, questions arise about what young people need to understand in order to engage with these technologies thoughtfully and responsibly. In response, a growing body of research has sought to define the knowledge, skills, and dispositions that constitute AI literacy and to translate these competencies into frameworks that can guide K–12 education. In this section, I first review existing AI literacy frameworks in K–12 education and then introduce the framework by Mills et al., [217] that guides this dissertation.

2.1.1 *AI Literacy Frameworks*

A growing body of work has sought to define AI literacy and identify the competencies individuals need to understand and engage with AI systems. Definitions in this space share a common orientation toward both technical understanding and critical engagement. Long and Magerko [199], for example, characterize AI literacy as a set of competencies that enable individuals to critically evaluate AI technologies, collaborate effectively with AI systems, and use them in everyday contexts such as school, work, and home. Similarly, Ng et al. [231] conceptualize AI literacy across four dimensions: understanding AI concepts, applying AI tools, evaluating and creating with AI, and engaging with ethical considerations. Chiu [53] offer a useful terminological clarification within this space, distinguishing between *AI literacy* as knowledge about how AI works and its societal implications, and *AI competency* as the reflective capacity needed to apply that knowledge responsibly, a distinction that underscores the difference between understanding AI and being able to act thoughtfully in relation to it.

A parallel line of work has focused specifically on translating these competencies into learning progressions across K-12 grade levels. The AI4K12 initiative provided one of the earliest efforts to do so, with Touretzky et al. [297] organizing K-12 AI learning around Five Big Ideas in AI: perception, representation and reasoning, learning, natural interaction, and societal impact. The rapid emergence and public uptake of generative AI systems in the years that followed, however, significantly expanded both the capabilities and the everyday presence of AI in ways that were difficult to anticipate in 2019. In 2025, AI4K12, in partnership with CSTA, revised and expanded the framework to reflect this changed landscape, producing five updated learning priority areas: Humans and AI, Representation and Reasoning, Machine Learning, Ethical AI System Design and Programming, and Societal Impacts of AI, with explicit grade-band progressions from kindergarten through twelfth grade [299]. This updated framework also names AI’s environmental impacts as a distinct learning priority within its Societal Impacts category, a competency that is taken up empirically in the third study of this dissertation.

Prior research has also called for more critical and justice-oriented frameworks for K-12 AI literacy. For example, Velandar et al. found that while most acknowledge the ethical dimensions of AI, they tend toward a techno-deterministic view that positions AI development as inevitable, preparing learners to adapt to and use AI responsibly rather than empowering them to question or imagine alternatives to existing AI systems [312]. Williams [323] extended this further, observing that most frameworks describe a generic learner and do not account for students’ cultural backgrounds or social positioning. Their Impact.AI framework, developed through community-engaged design with historically underrepresented students, centers cultural relevance as foundational to meaningful AI literacy learning. Taken together, these frameworks reflect a growing emphasis on K-12 AI literacy that extends beyond technical skill-building to include ethical, societal, and critical perspectives on AI.

2.1.2 AI Literacy as Understanding, Evaluating, and Using AI

Building on the broader landscape of K–12 AI literacy, Mills et al. [217] propose a framework that organizes AI literacy around three Modes of Engagement: *Understanding AI*, *Evaluating AI*, and *Using AI*. Developed in collaboration with K–12 practitioners, the framework centers two core values — *Centering Human Judgment* and *Centering Justice* — which position human agency and equity as foundational to AI literacy. It also outlines six AI Literacy Practices that describe the skills through which learners can demonstrate AI literacy in practice. The following sections describe these components in greater detail.

Core Values

The first core value, *Centering Human Judgment*, emphasizes that human oversight is essential to the safe and effective use of AI systems. Mills et al. [217] argue that while AI can process large volumes of data and generate predictions, it cannot understand context, exercise ethical reasoning, or assess the appropriateness of its own outputs [142]. Centering Human Judgment therefore encourages learners to evaluate which tasks are appropriate for AI assistance, critically examine AI-generated outputs, and weigh the potential benefits

and risks of AI use across different contexts. The framework also extends this value to environmental considerations, noting that the uncritical deployment of large-scale AI systems carries significant energy costs that human judgment must take into account [90, 164]. The second core value, *Centering Justice*, focuses on the broader societal impacts of AI systems. Mills et al. [217] emphasize that AI technologies are not neutral; their outputs may fail to represent the interests of all people and can reproduce or amplify existing social inequalities. Centering Justice therefore asks learners to examine how AI systems may encode or reinforce societal biases, evaluate whether AI applications perpetuate inequities in access and opportunity, and consider how AI technologies might be designed and used in ways that better serve historically marginalized communities [217].

Modes of Engagement

The framework identifies three interconnected Modes of Engagement with AI: *Understanding AI*, *Evaluating AI*, and *Using AI*. Mills et al. [217] argue that developing AI literacy requires engagement across all three modes. The *Understanding AI* mode focuses on developing foundational knowledge about how AI systems work. It also includes understanding the history of AI, its current applications across sectors such as education, and the role of human decision-making in shaping how AI systems are developed and deployed [217]. The *Evaluating AI* mode asks learners to critically examine AI systems and their outputs, guided by the two Core Values described above. Evaluation involves assessing whether a given AI system is trustworthy, equitable, and appropriate for its intended use. Finally, the *Using AI* mode focuses on how learners interact with AI systems in practice. Within this mode, three Types of Use are identified: *Interacting with AI*, *Creating with AI*, and *Problem Solving with AI* [217]. Each type of use introduces distinct opportunities and risks and therefore calls for the application of different AI Literacy Practices.



Figure 2.1: AI literacy framework from Mills et al. (2024), reproduced from their report.

AI Literacy Practices

The framework further identifies six AI Literacy Practices that describe the skills learners draw on when engaging with AI systems: Algorithmic Thinking, Data Analysis and Inference, Data Privacy and Security, Digital Communication and Expression, Ethics and Impact, and Information and Mis/Disinformation [217]. This dissertation focuses on two practices that are central to evaluating AI systems and their outputs: *Ethics and Impact* and *Information and Mis/Disinformation*. The *Ethics and Impact* practice asks learners to consider the broader societal and environmental consequences of AI systems [217]. This includes examining how datasets and algorithms may reproduce existing societal biases and evaluating the potential benefits and harms of AI technologies for individuals, communities, and the environment. The *Information and Mis/Disinformation* practice focuses on evaluating the credibility of

AI-generated outputs [217]. Learners assess the reliability of information, compare multiple sources, and determine when AI-generated content should not be trusted.

2.2 *Epistemic, Material, and Social Implications of AI*

In this section, I review prior research that highlights several important implications of AI systems and motivates the focus on two AI literacy practices in this dissertation: *Information and Mis/Disinformation* and *Ethics and Impact*. First, I examine scholarship on the epistemic implications of AI systems, focusing on how AI-generated outputs can appear authoritative even when they are inaccurate and why children may find it difficult to evaluate their credibility. Second, I review research on the material dimensions of AI, including the energy consumption, water use, and broader environmental impacts associated with the computational infrastructure that supports modern AI systems. Finally, I discuss work on the social implications of AI, with particular attention to how biases can become embedded in datasets and algorithms and how these dynamics may reproduce or amplify existing social inequalities. Together, these bodies of work motivate the need for AI literacy practices that support learners in critically evaluating AI-generated information and reasoning about the broader societal and material impacts of AI technologies.

2.2.1 Challenges in Evaluating AI Outputs

As AI systems become increasingly accessible to children, understanding how young learners interact with and make sense of AI-generated outputs has become an important area of inquiry. Recent work in HCI has examined how children interpret the outputs produced by AI systems and the conditions under which they are more or less likely to evaluate them critically [223, 221]. Much of this research has focused on generative AI tools that can appear authoritative and well-formed even when they contain errors. Prior studies suggest that children’s (ages 8–13) ability to critically assess AI-generated content is heavily influenced by their prior knowledge and age [284]. For example, children who are well-versed in specific topics, such as Pokémon or Star Wars, can readily identify errors when ChatGPT provides

inaccurate information or when a DALL-E-generated image depicts a character with six fingers instead of five [230]. Conversely, when faced with unfamiliar subject matter, children are more likely to overtrust AI-generated outputs [284].

Prior research also suggests that this tendency to overtrust AI outputs stems in part from how these systems present information [222]. AI-generated responses often appear polished and authoritative, presenting information in structured formats such as step-by-step explanations or organized lists. These presentation features can create an illusion of correctness even when the underlying content is inaccurate [284, 280]. For example, middle school students found ChatGPT’s response to a technical question convincing because of its clear organization, despite the fact that the content was incorrect [284]. Similarly, perceived transparency—when AI appears to explain its reasoning—can mislead users into assuming the system’s output is reliable [284, 280]. In one example, children trusted ChatGPT’s breakdown of a multiplication problem because the steps appeared logical, even though the final answer was incorrect [284]. Superficial indicators of credibility can further compound this issue, as users often rely on cues such as citations, dates, or named sources to judge accuracy [284, 280]. For instance, students trusted a fabricated list of scientific papers generated by an AI system because it included plausible-sounding titles and publication dates [284].

From a cognitive load perspective, these challenges arise because children have limited working memory resources, making it difficult to critically assess complex information without appropriate scaffolding [211, 210]. Many contemporary AI interfaces provide few cues about the reliability or provenance of their outputs, increasing the cognitive burden on users who must independently verify the information they encounter [280]. Without structured support, children may struggle to detect errors, particularly when AI outputs are presented in coherent and persuasive forms. This body of work motivates Chapter 3 of this dissertation and the *Information and Mis/Disinformation* practice, which focuses on helping K–12 learners assess the credibility of AI-generated information and determine when such outputs should not be trusted.

2.2.2 *Recognizing Bias in AI Outputs*

Human-like bias has been documented in numerous AI systems, including criminal recidivism models trained to make predictions from tabular data [57], computer vision systems such as those trained for facial recognition [254, 256], language technologies like word embeddings [30, 129, 50], and pretrained language models [22]. Such biases have been observed not only in systems trained on general internet data but also in machine learning technologies trained using content intended for children, such as corpora of children’s books [48]. Furthermore, AI systems designed for child-facing contexts but trained on uncured datasets may inadvertently transfer adult-oriented content, including profane or otherwise inappropriate language, into technologies used by young learners [255, 331].

Understanding why bias enters AI systems requires some attention to how they are built. Many modern AI systems rely on large-scale pretraining using datasets collected from the internet [143]. For example, language models such as Google’s T5 are trained on massive web-based datasets like C4 (the “Colossal Clean Crawled Corpus”), which contains over a trillion tokens of text collected over several years [253]. During this pretraining process, models learn statistical patterns in language and other data that allow them to generate fluent responses or perform downstream tasks [175]. However, because datasets of this scale cannot be fully reviewed or curated by human researchers, pretrained models often inherit patterns of bias present in the data itself [22]. These biases may later manifest in system outputs, even when the models are fine-tuned for specific tasks or deployed in new contexts [315, 330]. For example, Buolamwini et al., [40] found that several commercially available facial recognition systems performed significantly worse when identifying individuals with darker skin tones, particularly women.

Representation bias is another common form of AI bias in which minoritized groups are associated with negative attributes or traits. This may be implicit—i.e., detectable via statistical analysis of machine learned representations [135], or explicit—i.e., observable based on disparities in downstream tasks [11, 277]. For example, Wolfe et al., [331] found that

English-language word embeddings and generative language models associate teenagers with societal problems such as violence, drug use, and mental illness. These associations contrast sharply with how teenagers themselves describe their lives, which tend to center on school and friendship. Because young people are both subjects of these representations and users of AI systems, this body of work motivates the focus of Chapter 4 of this dissertation and the social dimension of the *Ethics and Impact* practice, which asks K–12 learners to examine how AI systems may reproduce societal bias and to evaluate their potential harms for individuals and communities.

2.2.3 *Understanding AI’s Material Impacts*

The rapid expansion of modern AI systems, including large language models (LLMs) and other general-purpose AI technologies, has been accompanied by unprecedented growth in computational infrastructure [291, 270]. Contemporary AI systems rely on large-scale data centers equipped with specialized hardware, such as GPUs and accelerators, which require substantial amounts of electricity to support model training and ongoing operation [66]. It is notable that other energy-intensive technologies of the past decade, such as cryptocurrency and cloud gaming [140, 226, 188], have not made the same impact on the technology industry’s energy consumption as AI has; from 2005 to 2017, the amount of electricity going to data centers was relatively unchanging, as increases in efficiency mitigated energy consumption resulting from increases in demand [238]. Recent estimates suggest that data centers now consume approximately 4.4% of the United States’ domestic electricity supply, a figure expected to double in the coming years [2]. In many regions, this electricity is still generated primarily from fossil fuel sources such as coal and natural gas, linking AI infrastructure to greenhouse gas emissions and local air pollution [4, 241].

Beyond electricity consumption, AI systems also place considerable demands on freshwater resources [270]. Data centers generate substantial heat during computation and require continuous cooling to maintain reliable operation. In many facilities, cooling is achieved through water-intensive systems, leading to significant freshwater withdrawals [203]. The

material footprint of AI also extends upstream to the production of hardware and supporting infrastructure. The manufacture of GPUs, servers, and networking equipment requires the extraction and processing of critical minerals and rare earth elements—including copper, cobalt, and lithium—which are often sourced through environmentally destructive mining practices [190]. Hardware production is energy-intensive and contributes to emissions and material depletion long before an AI system is deployed, while rapid hardware obsolescence generates large volumes of electronic waste that are frequently inadequately recycled [229].

Recent research further highlights that the environmental footprint of AI systems is shaped not only by model training but also by the cumulative impact of everyday use. While training large models is widely recognized as computationally intensive [238], the generation of outputs in response to user prompts can also scale energy and water consumption over a model’s lifetime as millions of interactions accumulate [203]. These dynamics are particularly visible in generative AI systems, whose widespread public adoption has dramatically increased the frequency of such interactions.

Despite growing awareness within the AI and sustainability communities, the material and environmental costs of AI systems remain largely invisible to end users, particularly children. AI technologies are increasingly accessible in educational, creative, and everyday contexts [214], yet the infrastructure that enables these systems often remains hidden from view. Without opportunities to understand how AI systems are produced, powered, and sustained, users may reasonably interpret interactions with AI as immaterial or consequence-free [153]. Foregrounding the material dimensions of AI systems is therefore essential for enabling children to situate AI within broader conversations about sustainability and the environmental costs of digital technologies. Prior scholarship in CCI and sustainability education provides design precedents for supporting this kind of reasoning, emphasizing children’s engagement with technologies as embedded within interconnected social, material, and ecological systems rather than as isolated tools [321, 263, 131]. This body of work motivates the material dimension of the *Ethics and Impact* practice explored in Chapter 5, which asks K–12 learners to consider the real-world costs of AI systems for individuals, communities, and the natural world.

2.3 Prior Work

This dissertation builds on multiple strands of prior research in AI literacy and learning sciences. First, prior work in AI literacy has explored a range of modalities and interaction paradigms for supporting children’s engagement with AI systems, including unplugged activities, browser-based tools, embodied systems, and conversational agents. In addition to this broader AI literacy literature, this dissertation also draws on prior work connected to the specific learning goals of the systems presented in later chapters. In particular, I review prior research on teaching children about bias and fairness in AI systems, which informs the design of *CLIP4KIDS*, as well as research on systems thinking and sustainability education, which informs the design of *EcoPrompt*. Together, these bodies of work motivate the interaction design choices and learning goals explored across the three AI literacy systems in this dissertation.

2.3.1 Unplugged Approaches to AI Literacy

Building on the broader CS Unplugged tradition [151, 21], this approach uses physical materials, role-play, games, and embodied interaction to help learners reason about AI concepts such as classification, representation, training data, and algorithmic decision-making [285, 195, 70]. Prior work suggests that unplugged activities can lower barriers to participation by reducing technical complexity while supporting collaborative discussion and reflection [72, 187, 296].

Much of this work has focused on helping learners build conceptual understandings of AI processes through physical simulation. For example, Lindner et al. [195] introduced a series of paper-based activities exploring ideas such as classification, reinforcement learning, and the Turing test. Similarly, Long et al. [200] and Darabipourshiraz et al. [86] designed embodied activities that allow learners to physically enact processes related to machine learning and knowledge representation. Work in museums and informal learning settings

has further extended these ideas through activity boxes, tabletop exhibits, and collaborative installations that support embodied reasoning about AI systems [198, 85, 189, 84]. Across this work, AI processes are externalized into physical actions and artifacts that learners can manipulate, rearrange, and discuss together.

Unplugged activities have also been used to support critical engagement with the societal implications of AI. Payne’s [242] middle-school AI ethics curriculum, for example, uses activities such as algorithm design exercises to help learners examine how values and assumptions become embedded in AI systems. Similarly, Ali et al. [8] introduced *AI Audit*, a card game that engages learners in reasoning about the potential harms and ethical implications of AI products. Prior research further suggests that unplugged activities can support collaborative reasoning around fairness, representation, and societal impact, particularly when integrated into broader classroom discussions [266, 323, 309, 102, 258]. While the systems presented in this dissertation primarily focus on interactive digital engagement with AI systems, prior unplugged work provides important foundations for understanding how embodied interaction, collaboration, and simplified representations can support children’s emerging understandings of AI concepts.

2.3.2 Interactive AI Literacy Systems

Prior work in AI literacy has also explored browser-based and block-based machine learning tools. Platforms such as Cognimates [99], Machine Learning for Kids [178], and Teachable Machine [43] allow learners to train classifiers using images, text, sounds, or movement data and incorporate these models into interactive projects. Prior work suggests that engaging learners in processes such as collecting training data, testing outputs, and retraining models may help make model behavior more visible by allowing learners to observe how outputs change in response to different inputs and datasets [91, 59, 300, 322, 310, 51]. Constructionist AI literacy systems have also expanded into embodied and creative domains through platforms such as PoseBlocks [158], AlpacaML [348], and systems that combine machine learning with

movement, dance, or physical computing activities.

Another thread of work has explored the use of social robots and embodied agents to support AI literacy [181, 103, 325, 278]. Systems such as PopBots [327] combine physical robots with curricula that introduce young children to concepts such as supervised learning and knowledge-based systems. Prior work suggests that embodied interaction may increase engagement and help make abstract AI concepts more concrete for younger learners [323]. At the same time, research on children’s interactions with conversational and robotic agents has raised concerns around anthropomorphism, overtrust, and social attachment [103, 326, 73]. These concerns have become increasingly relevant with the rise of generative AI systems that communicate through natural language and conversational interaction.

More recent work within digital AI literacy research has explored dialogic and conversational systems in which learners engage with AI through natural language interaction [79, 227, 193, 134]. In these systems, language becomes the primary interface through which learners encounter, test, and make sense of AI systems. Prior work suggests that conversational interaction may support engagement by allowing learners to communicate with AI systems using familiar forms of dialogue [193, 307, 134, 26]. Some of the earliest work in this area focused on helping learners build or program conversational agents. For example, Zhorai [193] allows children to teach a conversational agent about animals and ecosystems while visualizing how the system organizes knowledge.

Researchers have also begun exploring hybrid systems that combine conversational interaction with tangible or embodied interaction. For example, Lindrup et al. [196] introduced *Prompt Machine*, a tangible generative AI system that slows down and makes the prompting process visible by routing prompts through physical objects rather than standard chat interfaces. This work reflects broader efforts to design AI literacy experiences that encourage learners to reflect not only on AI outputs, but also on the interaction processes through which those outputs are produced [62, 27, 163].

2.3.3 *Teaching Children about Bias in AI*

Recent research in AI literacy has increasingly explored how children can learn to recognize and reason about bias in algorithmic systems. Studies examining children’s engagement with algorithmic fairness show that young learners are capable of identifying biased outcomes in both everyday technologies and the social contexts that shape them [268, 104, 185]. Empirical work further demonstrates that children can articulate how bias manifests in digital systems and reflect on its broader societal implications [61, 185, 76, 223]. One approach used to support this understanding draws on constructionist learning traditions, where learners actively build and interrogate computational systems as part of the learning process [240]. This often involves interactive platforms that allow children to train machine learning models using their own datasets and observe how data choices influence model behavior. Tools such as Teachable Machine [43], Co-ML, and Poseblocks [158] enable learners to construct classifiers from images, sounds, or text, allowing them to explore how biased or incomplete datasets can produce skewed outputs [314, 242, 207, 154, 269, 215, 7].

Prior work shows that when children train and evaluate models using their own data, they develop a clearer understanding of how AI systems function and where their limitations lie [348, 106? , 310]. For example, Castro and Desportes [44] engaged learners in examining the socio-technical aspects of AI through a creative computing system, *danceON*, where students surfaced how cultural assumptions about the human body influenced model behavior. Similarly, Register et al., [259] asked students to train classifiers on personal datasets and reflect on the applications and limitations of these models. Prior research also demonstrates that learners, particularly those from communities historically marginalized by algorithmic systems, are capable of analyzing outputs from tools such as search engines and image generators, identifying problematic representations, and proposing alternative designs [282, 283]. Building on this perspective, related research positions young people not only as critics but also as auditors and designers of machine learning systems [223, 221]. Building on prior work that situates teaching AI bias within the dimensions of fairness, society, and ethics, I describe

CLIP4KIDS, an interactive system designed to support children’s understanding of AI bias, in Chapter 4.

2.3.4 Systems Thinking in Environmental Education

Systems thinking is a framework that supports reasoning about complex phenomena by foregrounding relationships among interconnected components, resource flows, and feedback processes that unfold over time and across scales [14, 213, 320]. Within sustainability and environmental education, systems thinking has been widely adopted to help learners understand environmental challenges such as climate change, resource depletion, and pollution as outcomes of interacting social, material, and ecological systems [131, 244, 29]. Prior research emphasizes the importance of supporting children’s ability to shift between local and global viewpoints, enabling them to connect everyday actions and practices to distributed, cumulative, system-level consequences [131, 58]. Scholars further argue that encouraging reflection on how children situate themselves within these systems, shaped by social, cultural, and intersectional contexts, can deepen engagement and support meaningful forms of understanding [160].

Within climate and sustainability education, systems thinking is frequently enacted through interactive simulations and serious games [205, 264, 234], which allow children to explore how individual actions interact within shared systems in engaging ways. Early sustainability-focused serious games often took the form of tabletop modalities such as card and board games [264, 108], with digital games becoming increasingly prevalent in more recent work [257]. Across both formats, sustainability-focused serious games typically employ role-play or management simulation mechanics [257, 276] and commonly address themes such as resource management, irrigation, carbon emissions, and deforestation [205, 19]. Through repeated interaction, such games can support systems thinking by enabling children to observe how individual decisions accumulate into collective outcomes, often producing unintended or emergent effects.

A systems thinking lens is also well-suited to the case of generative AI, as its environmental impact is largely invisible to end users and distributed across infrastructures such as data

centers, energy grids, and water systems. As a result, reasoning about generative AI’s environmental footprint requires connecting everyday interactions with abstract, system-level processes that unfold across space and time. While prior work in sustainability education demonstrates how interactive systems can support reasoning about distributed environmental impacts, less is known about how children apply these forms of reasoning in the context of AI technologies such as generative AI. Guided by prior work on systems thinking, I describe *EcoPrompt*, an interactive system designed to support children’s reasoning about the environmental footprint of generative AI, in Chapter 5.

2.4 Educational Theories

In this section, I review three educational theories that ground the design of the AI literacy systems in this dissertation. I begin with Experiential Learning Theory, which frames learning as an iterative process of action and reflection and motivates the focus on direct, hands-on engagement with AI systems. I then turn to Cognitive Load Theory and the Cognitive Theory of Multimedia Learning, which inform how AI literacy systems can be designed to support understanding without overwhelming learners. Finally, I discuss Critical Pedagogy, which foregrounds questions of power, bias, and social impact as learners engage with AI tools.

2.4.1 Experiential Learning

Experiential learning Theory (ELT) considers learning as a process of building knowledge through cycles of action and reflection [89, 177]. Kolb [173] conceptualizes experiential learning as an iterative cycle consisting of four phases: concrete experience, reflective observation, abstract conceptualization, and active experimentation. These phases describe a process in which learners engage with a system, examine the outcomes of that interaction, develop more general explanations, and test those explanations through further action. Prior research suggests that structuring learning around such cycles supports step-by-step reasoning and helps learners transform concrete experiences into more abstract conceptual understanding [150, 343].

Designing for experiential learning in AI literacy therefore requires creating conditions that support sustained interaction and reflection. A key challenge is ensuring that learners can meaningfully engage with AI systems without being constrained by technical barriers that divert attention away from conceptual understanding. Prior work has explored a range of approaches to enabling such engagement, including the use of custom-built AI tools, interactive agents, and model training activities that allow learners to directly manipulate AI systems [111, 193, 161, 101]. More recent browser-based tools [43, 298] further lower these barriers by enabling learners to experiment with AI systems, observe their behavior, and test ideas in real time without requiring prior programming experience.

Building on ELT, the three AI literacy systems presented in this dissertation instantiate the experiential learning cycle through a recurring pattern of interaction. Learners first encounter AI system behavior (concrete experience), and are then prompted to examine and interpret these outputs, identifying possible reasons for the system’s behavior (reflective observation). Through discussion and comparison, learners develop more general explanations of how the system operates and where its limitations lie (abstract conceptualization), and finally apply these understandings by modifying inputs, data, or interaction strategies to test and refine their ideas (active experimentation) [173]. This iterative structure supports learners in progressively building and revising their mental models of AI systems through direct engagement. In this sense, AI systems presented in this dissertation function both as the object of inquiry and as the medium through which that inquiry unfolds.

2.4.2 Multimedia Learning

Cognitive Load Theory (CLT), developed by Sweller [289], provides an influential framework for understanding how human cognitive architecture shapes and constrains learning. CLT distinguishes between a limited-capacity working memory and a long-term memory in which knowledge is organised as interconnected schemas [289, 288]. Within this framework, three types of cognitive load are identified: intrinsic load, arising from the inherent complexity of the material; extraneous load, introduced by ineffective instructional design; and germane

load, referring to the cognitive effort directed toward constructing and refining schemas [290]. From an instructional design perspective, the goal is to minimise extraneous load while supporting processes that enable schema formation [288].

Complementing CLT, Mayer’s Cognitive Theory of Multimedia Learning (CTML) proposes that learners process information through separate verbal and visual channels, each with limited capacity, and that meaningful learning occurs as they actively select, organize, and integrate information across these channels [209, 211]. CTML outlines a set of design principles that have been supported in prior work, including the multimedia principle (combining words and pictures supports better understanding than words alone), the modality principle (pairing narration with graphics is more effective than on-screen text with graphics), the coherence principle (removing extraneous material), and the personalization principle (conversational style promotes deeper engagement) [211, 209]. The insights of CLT and CTML carry implications for the design of interactive AI literacy systems.

The AI literacy systems in this dissertation are designed to support learners in engaging with complex and often opaque AI processes without overwhelming them. In these systems, information is distributed across visual and textual representations [211, 209], allowing learners to interpret AI outputs alongside their own reasoning. Interactions are structured and sequenced so that learners can focus on specific aspects of system behavior at a time, while key features of AI processes are made visible to support interpretation. This design approach helps reduce extraneous cognitive load while supporting the kinds of processing needed to build understanding.

2.4.3 Critical Pedagogy

Critical pedagogy considers learning as a process of developing critical consciousness through active engagement with the social and political conditions shaping one’s experience. Freire [118] distinguishes between a “banking model” of education, in which knowledge is deposited into passive learners, and a problem-posing model, in which learners and educators engage in shared inquiry directed at understanding and transforming their world. Central to this

framework are the concepts of conscientization, which refers to the gradual development of critical awareness of one’s social reality, and praxis, which describes the iterative process of reflection and action through which this awareness is enacted [121, 119]. Within this process, dialogue enables learners to surface and examine the conditions shaping their experiences, while also encountering and responding to multiple perspectives [120]. Building on these foundations, Kincheloe [167] elaborates on the concept of epistemological curiosity, which refers to a disposition to question not only the content of knowledge but also the conditions of its production, including whose interests a given account serves and what perspectives it renders invisible.

In the context of AI literacy, critical pedagogy provides a useful lens for examining how AI systems shape what learners encounter, how information is presented, and whose perspectives are centred or marginalised. Kincheloe’s [167] concept of epistemological curiosity maps directly onto the kinds of questions central to critical AI literacy, encouraging learners to move beyond what an AI system produces to consider how it was built, whose data it draws on, and what values and assumptions are encoded within it. Additionally, Giroux ([127], as cited in [9] argues that education should make everyday experience “problematic and critical by revealing its hidden political assumptions.” In the context of AI-generated content, this means treating outputs not as neutral artefacts, but as products of particular design decisions, institutional interests, and sociotechnical contexts. In this way, critical pedagogy supports learners in moving from unreflective use of AI tools toward a more engaged and questioning stance, one that involves examining how outputs are produced, interrogating whose knowledge is represented or absent, and considering their broader social and material implications.

Chapter 3

AI PUZZLERS: SUPPORTING CHILDREN IN CRITICALLY EVALUATING AI OUTPUTS¹

3.1 Preface

Children today are increasingly navigating a digital environment in which AI-generated outputs function as pervasive sources of information. AI generated outputs often appear fluent and authoritative despite containing factual errors, and the growing presence of low-quality AI-generated content across digital platforms makes AI’s epistemic risks both significant and difficult to detect. This context underscores the need to support children in critically evaluating AI-generated content. In this chapter, I take up the core value of *Centering Human Judgment* and focus on the *Information and Mis/Disinformation practice* outlined in Mills et al.’s AI literacy framework [217], which invites learners to assess the reliability of AI-generated outputs, compare information across sources, and decide when such outputs should not be trusted.

3.2 Introduction

As AI systems become increasingly integrated into educational environments, it presents both opportunities and challenges for teaching and learning [133, 287, 249, 336]. One revealing example comes from a middle school assignment on the novel *Persepolis*, where students researching prophets encountered flawed responses from internet searches that incorporated generative AI. For instance, one response claimed that “*the Christian prophet Moses got*

¹This chapter is based on collaborative work previously published as: Aayushi Dangol, Runhua Zhao*, Robert Wolfe*, Trushaa Ramanan, Julie A. Kientz, and Jason Yip. 2025. "AI just keeps guessing": Using ARC Puzzles to Help Children Identify Reasoning Errors in Generative AI. In *Interaction Design and Children (IDC '25)*, June 23–26, 2025, Reykjavik, Iceland. ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3713043.3728836>

chocolate stains out of T-shirts” — a stark misunderstanding of historical and religious contexts [132]. According to the teacher, interviewed by the New York Times, more concerning than the error itself was that eighth graders accepted and recorded the AI hallucination without questioning its validity [132]. This incident illustrates the epistemic risks that AI systems pose in educational contexts and underscores the critical need to equip children with the skills to engage effectively with AI, critically assess its outputs, and understand its strengths and limitations [199, 217].

However, recognizing errors in AI generated outputs is not always straightforward [56, 284]. Even for adults, misleading outputs can be difficult to detect [292], and the challenge is often greater for children, who may have less experience questioning authoritative-seeming information [179, 333, 280]. One of the key challenges in detecting inconsistencies in AI-generated text is that, unlike images, textual responses do not present errors in an immediately perceivable visual pattern [284, 230]. In AI-generated images, inconsistencies like extra fingers or distorted facial features can be easily noticeable [32]. In contrast, recognizing textual inaccuracies often requires specific domain knowledge. Additionally, the length and verbosity of AI-generated responses can increase cognitive load, making it harder for children to identify inconsistencies [211, 230]. Furthermore, without explicit indicators of potential mistakes, they may struggle to assess the reliability of the information presented [338, 22]. Thus, prior work suggests that textual responses alone may not be an ideal starting point for supporting children in reasoning about the epistemic risks of AI-generated content.

To support children’s critical engagement with AI generated outputs, I designed AI Puzzlers, a web-based system that adapts Abstraction Reasoning Corpus (ARC) puzzles [55], originally developed to benchmark AI progress, into an interactive game in which children solve visual reasoning puzzles, observe how generative AI attempts the same puzzles, and compare their solutions with AI-generated outputs and explanations (see Figure 3.1). Drawing on Mayer and Moreno’s theory of multimedia learning [211], AI Puzzlers distributes information across visual and verbal channels, making AI’s reasoning failures immediately perceivable rather than buried in text (see Figure 3.2). I then conducted two participatory

design sessions with 21 children ages 6–11 to address the following research questions:

- **(RQ1):** What specific limitations of generative AI do children (ages 6–11) encounter and recognize through their engagement with AI Puzzlers?
- **(RQ2):** How does presenting information across visual and textual modalities influence children’s ability to critically assess AI-generated outputs?

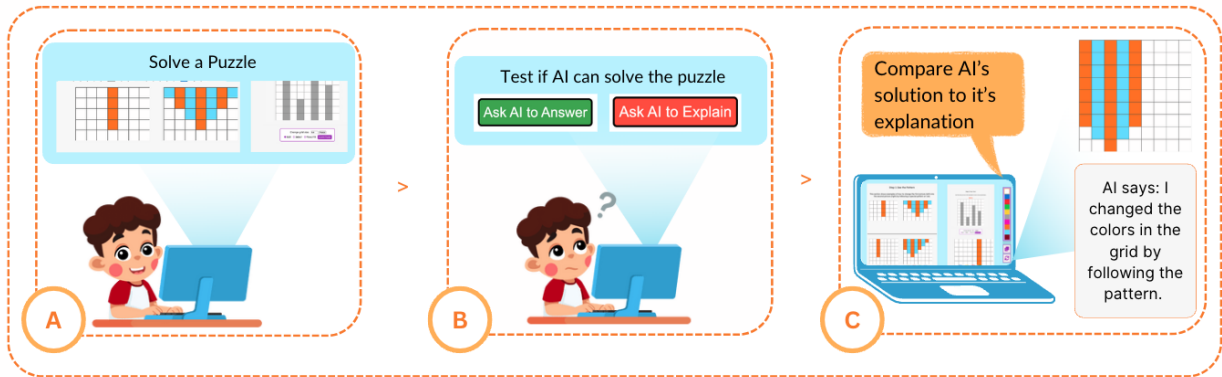


Figure 3.1: Overview of AI Puzzlers: (A) children first solve an ARC puzzle independently, then (B) test whether generative AI can solve the same puzzle, and finally (C) compare AI’s solution to its explanation to evaluate AI reasoning.

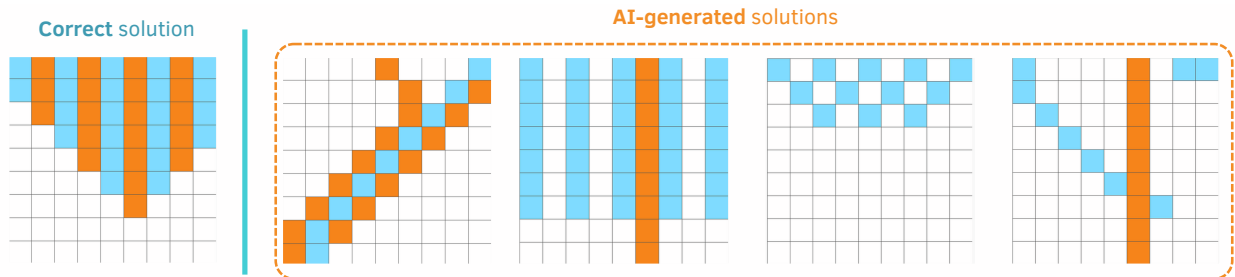


Figure 3.2: Comparison of the correct vs AI-generated solutions. The visual nature of AI Puzzlers makes AI errors easy to spot.

My findings show that AI Puzzlers provided a tangible way for children to engage with generative AI’s reasoning by making its outputs visually comparable to their own. Even younger children, who were not yet fluent readers, quickly detected inconsistencies in AI-generated solutions by evaluating their visual outputs. When generative AI made mistakes, especially on puzzles they considered easy, children reacted with surprise and amusement, sparking meaningful dialogue around how “*AI thinks.*” This also helped them recognize that generative AI approaches problem-solving differently from humans and, despite its strengths, has limitations that require careful evaluation of its outputs. Their continued engagement with AI Puzzlers highlights the importance of designing AI systems that present information in ways that facilitate comparison, encourage reflection, and scaffold multiple ways of understanding. This way, children are more likely to persist and critically evaluate AI outputs.

This study contributes both design insights and an empirical understanding of how children make sense of generative AI’s reasoning, navigate its errors, and develop strategies for evaluating its outputs. It supports T1 by making AI’s reasoning failures immediately visible, allowing learners to observe how the system misinterprets their intent and to iteratively refine their inputs. It also supports T2 by demonstrating that placing children’s own solutions alongside AI-generated outputs creates productive tensions that prompt children to interrogate the logic behind AI responses and reflect on differences between human and machine reasoning.

3.3 AI Puzzlers: System Design & Development

AI Puzzlers is an interactive system designed to help children critically engage with generative AI’s reasoning by solving visual puzzles. Accessible through any web browser, AI Puzzlers requires no prior knowledge of AI or programming, making it an easy entry point for young learners. The system’s is publicly available at <https://www.ai-puzzlers.com/>. I developed AI Puzzlers using the Abstraction and Reasoning Corpus (ARC) dataset [55], a collection of 800 publicly available visual puzzles originally designed to assess AI’s reasoning capabilities. As illustrated in Figure 3.3, to solve an ARC Puzzle, a human player or AI agent follows the

following steps:

- **Infer the Transformation Rule:** Each puzzle presents two or more input grids and their corresponding output grids. By observing the example input-output pairs, players try to discover the hidden rule that explains how the input changes to create the output.
- **Apply the Rule to New Inputs:** Once the transformation rule is inferred, players apply it to new, unseen input grids to generate the correct output.

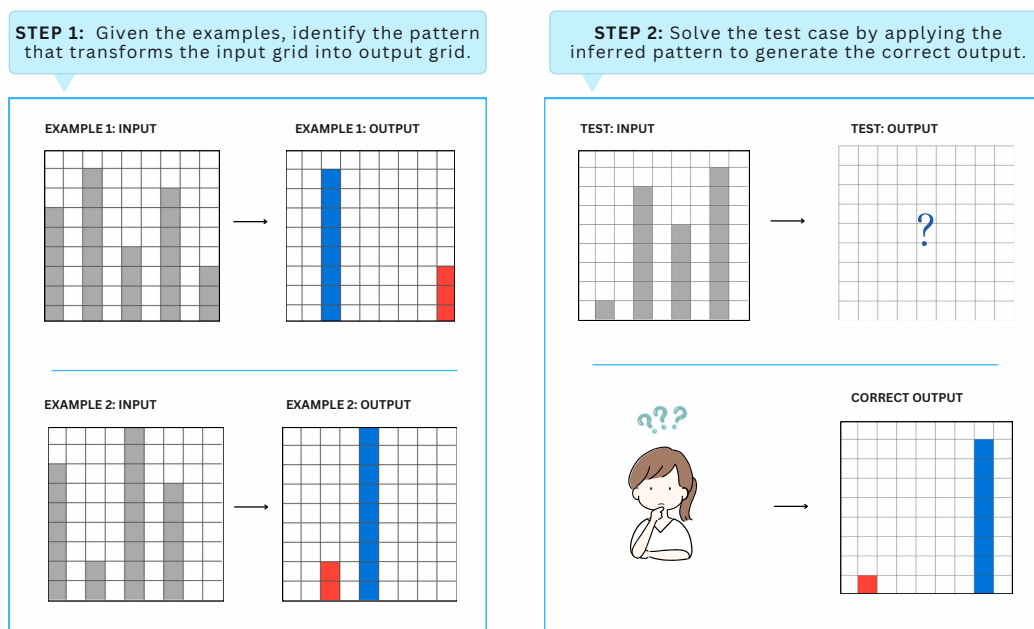


Figure 3.3: An example of an ARC puzzle with instructions for solving it. The correct answer shows the shortest bar is colored red and the tallest one is blue.

My design builds on ARC puzzles because 1) solving these puzzles requires no prior knowledge, creating a low barrier to entry; 2) AI systems struggle with solving ARC Puzzles while humans excel, aligning with Mills et al.'s AI literacy value of *Centering Human Judgment*, which emphasizes recognizing when human reasoning should take precedence over AI outputs

[217]; 3) children are naturally drawn to puzzles and games, making them an engaging game-based medium that can scaffold learning; and 4) the visual nature of the puzzles ensures there is no obscurity in the way generative AI presents information. By visually comparing generative AI’s solution with the correct solution that children can easily solve, they can spot when generative AI makes mistakes, preventing children from being misled by polished yet incorrect answers [284]. This process also encourages critical evaluation of generative AI’s outputs, allowing children to recognize not only the limitations of generative AI but also the unique strengths of human reasoning. I further elaborate on the design considerations in the section below.

3.3.1 Design Considerations

Facilitating Critical Evaluation through Visual Comparisons of generative AI Outputs

Mayer and Moreno’s active processing assumption suggests that meaningful learning occurs when learners actively select, organize, and integrate new information with prior knowledge [210, 209]. AI Puzzlers facilitates this process by allowing children to visually compare their own puzzle solutions with those generated by AI. This direct visual comparison engages children in deeper cognitive processing, prompting them to notice differences, identify errors and reasoning gaps in the AI’s process. Immediate feedback from the system further reinforces active processing by helping children quickly see what worked, what did not work, and why, facilitating critical evaluation of generative AI’s outputs.

Reducing Cognitive Overload in Interpreting generative AI Outputs

According to Mayer and Moreno’s dual-channel assumption [211, 209], people process information through two distinct channels: visual and verbal. Distributing cognitive effort across both channels reduces cognitive overload, leading to more efficient processing and deeper understanding. AI Puzzlers leverages this principle by presenting both visual and textual explanations of generative AI’s reasoning. When children use the “Ask AI to Explain” feature,

they receive a visual representation of the AI’s solution along with a step-by-step explanation of its reasoning. This dual presentation helps children compare the AI reasoning with its visual output, making it easier to spot discrepancies and evaluate the generative AI solution.

Scaffolding Learning through Easy-Switch Modalities

AI Puzzlers scaffolds learning and reduces cognitive overload by ensuring that children build foundational knowledge before tackling more advanced tasks [211, 209]. It supports three different modalities to help children develop familiarity with the puzzle environment before they engage with more complex tasks. In Manual Mode, children learn basic puzzle-solving tools to solve the ARC puzzles, building foundational skills without the complexity of AI interactions. AI Mode introduces features like “Ask AI to Solve,” and “Ask AI to Explain” allowing children to gradually engage with AI, while Assist Mode lets them actively guide the AI through puzzle-solving, encouraging experimentation and deeper understanding. These modes are designed for seamless switching, allowing children to easily transition between them.

Fostering Exploration of generative AI’s capabilities through Active Debugging

AI Puzzlers supports children in critically engaging with generative AI’s reasoning by recognizing its mistakes and providing hints to guide its responses. This fosters a form of participatory debugging [184, 295], where children take on an active role in evaluating generative AI’s logic and suggesting improvements. This approach is informed by Wittrock’s generative learning theory [329], which emphasizes active meaning-making, as children generate hypotheses, test AI responses, and construct new insights through comparison and reflection. Through the process of debugging, children refine their understanding of generative AI’s capabilities and limitations, aligning with Wittrock’s notion that learning emerges through the active construction of relationships among ideas [329].

3.3.2 Design Overview

AI Puzzlers consists of 12 puzzles from the ARC dataset, distributed across four levels of difficulty. I play-tested them with $N = 106$ children (grades 3 - 8) during my university’s annual K-12 STEM outreach event [80]. These children, who were event attendees (separate from our main study participants), voluntarily participated in the playtesting. After completing the puzzles, children rated their difficulty on a Likert scale from 1 (very easy) to 5 (very hard), with 3 as a neutral response. A one-sample t -test showed that children perceived the puzzles as slightly easier than neutral ($M = 2.38$, $t(103) = -6.48$, $p < .001$). To support children’s critical engagement with generative AI’s outputs, AI Puzzlers employs a scaffolded interaction model with three modes, which I describe next.

Manual Mode

This is the default interaction mode upon launching the AI Puzzlers application (see Figure 3.4). It is designed to encourage children to engage with the ARC puzzles independently and provides a foundation that will be expanded upon in subsequent interaction modes. AI Puzzlers offers players several functionalities to solve the puzzle. First, players can customize the grid size by adjusting the row and column height of the “After” grid. Although the predefined “Before” grid is non-editable, children have the option to “clone” it into the “After” grid. A palette of 10 distinct colors is provided for use on the editable grid. The “Edit” tool allows players to modify the color of individual squares. The “Select” tool enables the highlighting of multiple squares simultaneously, facilitating batch modifications. The “Flood Fill” tool changes the color of all connected squares that share the same color, similar to the paint bucket tool commonly found in graphics programs. Additionally, a “Reset” button is available, enabling players to start over with a clean version of the “After” grid if they wish to try a different approach.



Figure 3.4: Annotated screenshot of Manual Mode in AI Puzzlers. The interface consists of several key components: (A) Puzzle Navigation allows children to switch between different puzzles. (B) Example Puzzles provide input-output pairs to illustrate transformation rules. (C) Before Grid presents the given input that children use to solve the corresponding After puzzle. (D) Edit Panel provides a set of tools for editing After grid. (E) After Grid serves as the solution space where children create their expected output based on the transformation rule. (F) Color Palette enables children to apply colors when solving the puzzle. (G) Clone Button transfers the Before Grid content to the After Grid for further modification. (H) Reset Button clears the After Grid to start over.

AI Mode

Building on the features of Manual mode, AI mode allows children to interact with GPT-4o through “Ask AI to Solve” and “Ask AI to Explain” features. By clicking the “Ask AI to Solve” button, players can have GPT-4o attempt to generate a solution, which is then displayed in the “After” grid. Under the hood, a `grid_parser()` function converts the visual puzzle into a textual grid representation that GPT-4o can interpret. The model then outputs a similar textual grid structure (e.g., `[[0, 2, 2], [0, 1, 1], [0, 0, 3]]`), which a `response_parser()` function translates back into a visual grid, using tokens for colors and positions. Because GPT-4o’s responses are generated in real time via its API, the answers may vary with each attempt, demonstrating its generative nature. Additionally, the “Ask AI to Explain” feature provides step-by-step, child-friendly explanations of the AI’s reasoning process, which can be toggled on or off according to the player’s preference. After reviewing the AI’s process, children can submit the AI’s solution to receive immediate feedback on its correctness, just as they would with their own solutions. This process allows children to see where the AI might struggle, offering a direct and interactive way to understand the limitations of AI in solving ARC puzzles.

Assist Mode

Expanding further on the features of the AI mode, this mode allows children to guide the AI to solve puzzles by actively participating in the decision-making process and experimenting with different strategies to help the AI (see Figure 3.5). Similar to debugging in programming [184], children can identify mistakes in the AI’s approach and suggest corrections by typing their suggestions to AI into a “Hint” field. For example, if the AI incorrectly sorts shapes, a child might type, “Look at the smallest and biggest shapes.” The AI will then take these hints and produce a new output. This interactive process of testing and debugging helps children develop a more sophisticated understanding of AI’s capabilities, as they see the direct impact of their input on the AI’s performance. Children can also experiment with different

puzzle configurations and observe how the AI responds to changes in input. They can adjust parameters such as: 1) alter the number of input-output examples provided to the AI, 2) test different versions of OpenAI's GPT models, including GPT-3.5-turbo, GPT-4o-Mini, GPT-4o, and GPT-4o-Turbo or 3) introduce random patterns without clear transformation rules.

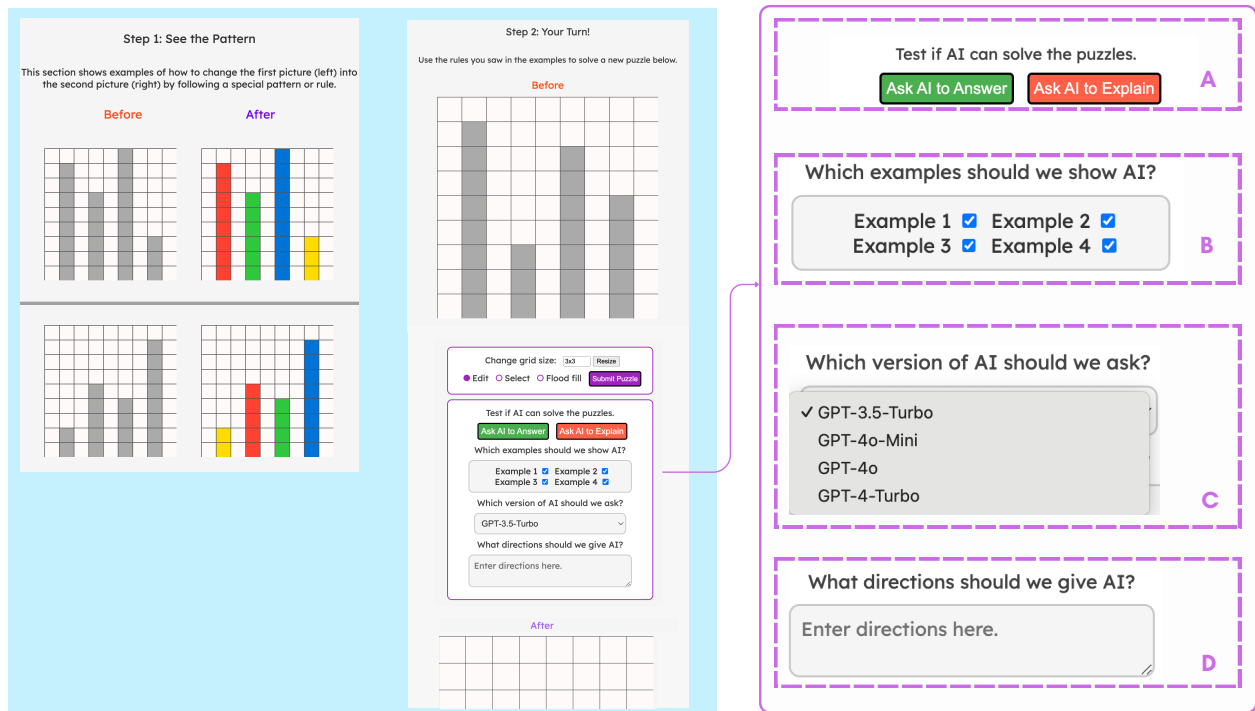


Figure 3.5: Screenshot of Assist Mode in AI Puzzlers. Certain interface elements are enlarged to highlight key interactive features, including (A) testing if AI can solve the puzzles by either answering or explaining, (B) selecting the number of examples shown to AI, (C) choosing AI model versions, and (D) providing hints to AI.

3.4 Methods

3.4.1 Participants

I led a team of five collaborators in conducting this study. We conducted our study with an inter-generational co-design group called **KidsTeam UW** consisting of twenty-one children, ages 6 to 11 ($M = 8.10$, $\sigma = 1.45$), and adult design researchers (researchers, graduate and undergraduate research assistants). Child participants represented a diverse range of ethnic backgrounds, including Asian, White, Black, Hispanic, and Multiracial identities. Children were recruited through mailing lists, posters, and snowball sampling. The child participants reported varying degrees of AI use and familiarity. While six children engaged with AI daily, three had no prior experience. The most common AI interactions included video game AIs and voice assistants. Table 3.1 presents the demographic information and AI usage details of the children, with all names represented as pseudonyms.

Table 3.1: Reported Child Participant Details

Name	Gender	Ethnicity	Age	AI Type	Usage Frequency
Kai	Male	Asian/White	8	Voice Assistant	Daily
Lani	Female	Asian/Black	9	None	Never
Juno	Male	Asian	7	Video Game AIs, Voice Assistant	Daily
Elias	Male	Asian/Black	9	Video Game AIs, Voice Assistant	Daily
Noa	Female	Asian/White	11	Video Game AIs, Voice Assistant	Multiple times a week
Ren	Male	Hispanic	10	Chatbot	Multiple times a week
Matt	Male	Asian/White	9	N/A	N/A
Ivy	Female	White	9	Video Game AIs, Voice Assistant	A few times a week
Zayn	Male	Asian/Black	9	None	Rarely
Finn	Male	White	10	N/A	N/A
Leila	Female	Asian	8	Voice Assistant	Daily
Mara	Female	Asian/Black	6	Video Game AIs, Voice Assistant	A few times a week
Emi	Female	Asian/White	8	None	Rarely
Hana	Female	Asian	8	None	Multiple times a week
Theo	Male	Asian/White	7	Video Game AIs, Voice Assistant	Multiple times a week
Lucia	Female	Hispanic	6	Video Game AI	Weekly
Rina	Female	Asian	7	Video Game AIs, Voice Assistant	Monthly
Owen	Male	White	8	Video Game AI	Daily
Nico	Male	Asian/White	6	None	Daily
Selah	Female	Asian/Black	6	None	Never
Elise	Female	Asian/Black	9	None	Never

3.4.2 Co-Design Sessions

We conducted two sessions with KidsTeam UW as part of a week long summer camp hosted by an inter-generational co-design group at our university. Each 1.5 hour session began with a 15-minute informal discussion to foster open dialogue and build rapport before transitioning to hands-on engagement with AI Puzzlers. Child participants worked in small, collaborative groups, each consisting of four to five children and two adult facilitators (see Figure 3.6). This structure was designed to balance peer-driven exploration with adult guidance, ensuring children could inspire one another while still receiving individualized support. To mitigate power imbalances and minimize undue influence on children’s responses, adult facilitators were trained to promote equal participation, fostering a balanced dialogue where all children had the opportunity to contribute.



Figure 3.6: Children engaging with AI Puzzlers alongside adult facilitators.

Session 1

We began the session with a warm-up question: “*Tell us about someone or something that you think is smart—and why?*” This activity served as an icebreaker, allowing children to become familiar with the researchers and fostering an open discussion environment. Next, children were introduced to both the manual mode and AI mode of AI Puzzlers. The research team demonstrated how to solve ARC puzzles and explained that in AI mode, they could ask generative AI to solve the puzzles. Earlier in the day, as part of the broader summer camp,

children had already interacted with generative AI by requesting and observing AI-generated images using *DALL·E 3*. Given this prior exposure, we chose to build upon their understanding rather than reintroduce generative AI concepts. This allowed our session to focus on how children engaged with AI Puzzlers to critically reflect on generative AI’s capabilities. After being introduced to the system, children were divided into five groups and encouraged to collaborate, discuss strategies, and work together to solve puzzles in manual mode. This allowed them to familiarize themselves with the puzzle format and develop problem-solving strategies without AI assistance. Before introducing the AI mode, facilitators posed two key questions to prompt reflection: 1) “*Do you think generative AI can solve these puzzles quickly or slowly? Why?*” and 2) “*Do you think generative AI can solve these puzzles without any help from people?*” These questions aimed to capture children’s initial expectations about generative AI’s capabilities before they engaged with it directly. Children then interacted with AI mode, where they could request AI assistance, observe AI-generated solutions, and receive AI-generated explanations for the puzzles. The session concluded with a 15-minute group discussion, where children shared their reflections on generative AI’s performance.

Session 2

Similar to Session 1, we began with a warm-up question: “*Tell us about a time you had to help others do something?*” This discussion encouraged children to reflect on how humans support one another, which facilitators then connected to the role of human guidance in helping generative AI solve puzzles. Unlike Session 1, where children primarily observed generative AI’s independent performance, this session encouraged them to actively assist generative AI by providing hints and experimenting with different strategies in Human-AI mode. Facilitators guided their exploration with key reflection questions such as, “*What hints do you think would be helpful?*” or, “*What do you think will happen if we change the number of examples?*” After 50 minutes of interacting with Human-AI mode, the different groups came together, and each team presented their experiences in front of the whole group, reflecting on successful strategies, challenges they encountered, and generative AI’s limitations.

3.4.3 Data Collection & Analysis

For both study sessions, my team used built-in webcams on desktop computers to record video and screen activity via Zoom, a video conferencing software. In total, we collected 427 minutes of video data. To capture additional insights, facilitators took field notes throughout the sessions, documenting key observations and notable moments. Additionally, physical artifacts, such as children’s handwritten notes and collaborative sketches, were photographed for documentation.

The first, second, and fourth authors, then created analytical memos for all the videos [28, 265]. As part of this process, one author served as the primary reviewer, while another served as the secondary reviewer. The primary reviewer first watched the assigned recorded video and created a narrative summary of events at five-minute intervals. They documented children’s interactions with AI Puzzlers, their reactions captured by the camera, and interactions and dialogues between participants, including direct quotes relevant to the study’s research questions. After the primary reviewer completed their memos, the secondary reviewer independently reviewed the same videos and memos to verify the accuracy of the initial observations and to add supplementary insights. This dual-review process ensured the reliability of the data and captured a broader range of perspectives.

Following the creation and review of the analytic memos, we divided the memos into two equal-sized sets and began an inductive coding process [113]. The first two authors independently engaged in open coding of the first set of memos, suggesting potential codes such as “Interaction with the System” and “Making sense of AI.” They then met over four meeting sessions to compare, reconcile, and refine the codes. During these discussions, they shared potential codes and their descriptions, collaboratively examined example quotes and counter-examples, compared code categories against one another, and refined the boundaries and definitions of each code. For example, the codes “Reading AI Explanations” and “Comparing Outputs with Correct Solutions” were consolidated into a single subcategory “Strategies.” This iterative process led to the development of a codebook that included three main code

categories: 1) Perception of AI, 2) Evaluation of AI Performance and 3) Interaction with the System. Once the codebook was finalized, I applied the final codes to the full dataset, and the second author conducted a second pass to ensure comprehensive analysis. We assessed interrater reliability through qualitative negotiations, where both authors met to discuss and resolve any coding disagreements [212]. We then organized the codes into overarching themes through two rounds of refinement and discussion. After finalizing the themes, I revisited the entire dataset to extract representative quotes for each theme, ensuring that the themes were well-supported by the data.

3.5 Findings

In this section, I present the findings from children’s interactions with AI Puzzlers, focusing on how they engaged with the system, responded to generative AI’s successes and failures, and developed an understanding of its capabilities and limitations. To illustrate their learning process, I use representative vignettes, embedding the analysis within each example. While this study examines children’s learning about generative AI, they generally referred to generative AI as “AI” in their discussions. For consistency with their language, I use "AI" throughout this section while maintaining an analytical focus on generative AI’s reasoning processes.

3.5.1 Children’s Interest and Exploration of AI Puzzlers

Surprise, Excitement, and AI’s Unexpected Errors

At the start of Session 1, when we first introduced children to AI Puzzlers, they had high expectations that AI could solve the puzzles. This belief in AI stemmed from their own ease in solving the puzzles, their perceived belief in AI’s ability to use visual references to recognize patterns, and their general trust in AI’s broad knowledge. However, they quickly noticed that AI struggled to solve the puzzles and their reactions, verbal and physical, reflected their surprise.

For example, in Session 1, when Ivy, Ren, Emi, and Mara first interacted with AI Puzzlers, Emi eagerly volunteered to solve the second puzzle (see Figure 3.7). She identified the pattern as “*alternating between red and grey*” and correctly solved the puzzle. When the facilitator asked about the puzzle’s difficulty, the group quickly agreed that it was simple. Ivy explained, “*It was pretty easy, judging by all the pictures. It was kind of obvious what you have to do since [the pattern is] that color, grey, that color, grey.*” Ren added, “*It can be easily identified what the pattern is going to be.*”

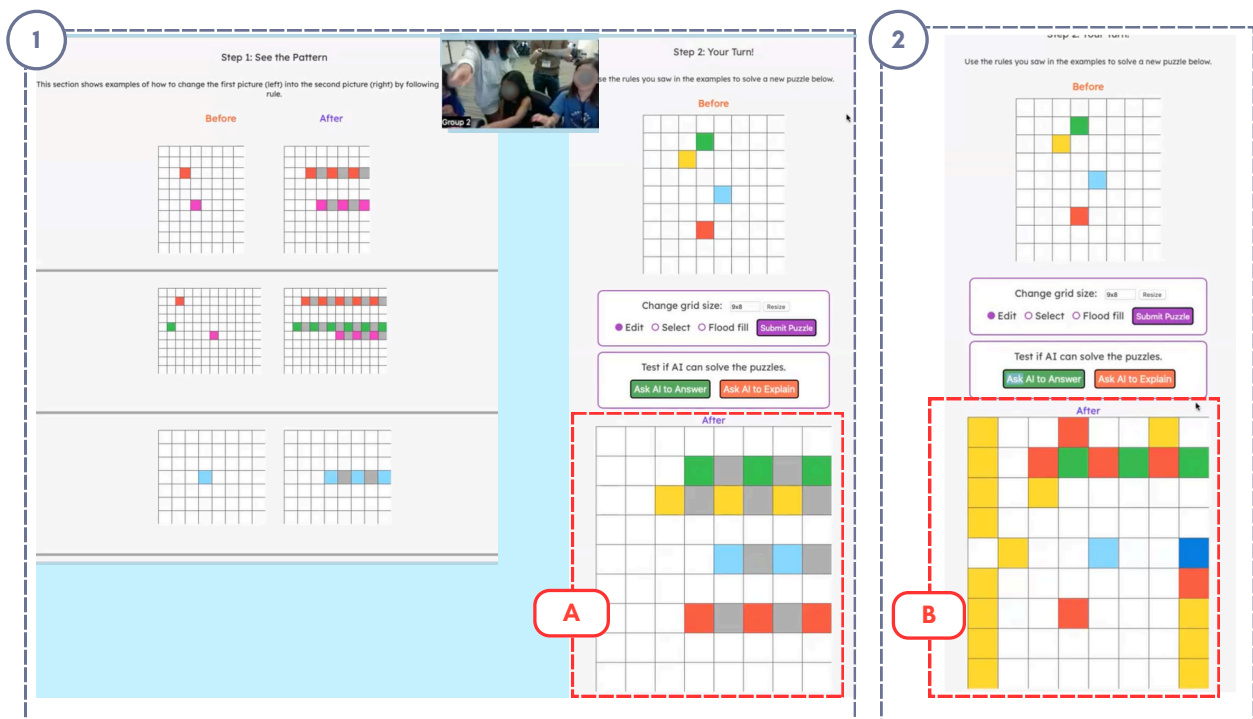


Figure 3.7: Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI’s attempt at the same puzzle. (A) highlights the children’s correct solution, whereas (B) shows the AI’s incorrect attempt.

When the facilitator asked if AI could solve the puzzle, Emi, Mara, Ren, and Ivy all agreed that it could because “*It was easy.*” Ivy reasoned, “There are tons of different references. Judging by all the references, since AI takes references from other pictures, the reference it gets

is that color, grey, that color, grey.” Here, Ivy’s comment implies that she believed AI could use abstraction skills by generalizing from specific examples “*the references*” to form a rule (alternating between the color and grey) needed to solve the puzzle. Mara added, “Because AI kind of knows everything,” indicating that she assumed AI’s vast access to information would enable it to solve the puzzle. Her response revealed a gap in her understanding of AI’s limitations. However, when the AI returned an incorrect solution (see Figure 3.7), the group burst into laughter. They found the situation humorous because the AI failed so badly, despite their prior confidence that it would succeed, making the error an unexpected source of amusement. Additionally, the visual nature of the puzzles allowed them to quickly recognize the mistake, as Ivy commented, “*That is very very wrong.*”

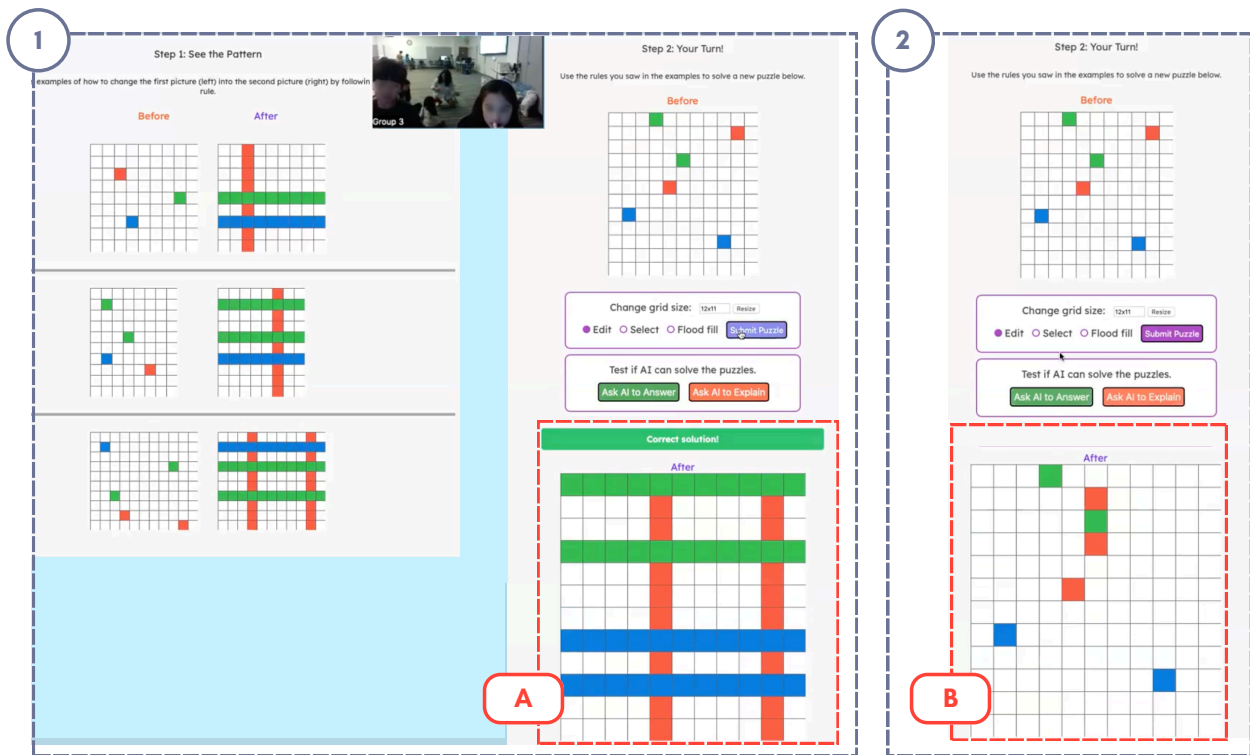


Figure 3.8: Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI’s attempt at the same puzzle. (A) highlights the children’s correct solution, whereas (B) shows the AI’s incorrect attempt.

The sharp contrast between the AI’s solutions and the children’s correct answers continued to evoke surprise and amusement, even after they had repeatedly watched the AI fail. For instance, midway through Session 1, Juno and Hana had seen the AI struggle with nine previous puzzles, all of which they had solved correctly. Reflecting on the AI’s past failures, Hana predicted, “*Maybe it can solve the puzzle faster but incorrectly*” as they prepared to ask the AI to solve the tenth puzzle (see Figure 3.8). Yet, when the AI returned another incorrect solution (see Figure 3.8), Juno said, “*Oh my gosh. What the heck is this?*” Hana, amused, blew a raspberry and added, “*It makes no sense whatsoever,*” to which Juno agreed, saying, “*It did something very weird.*”

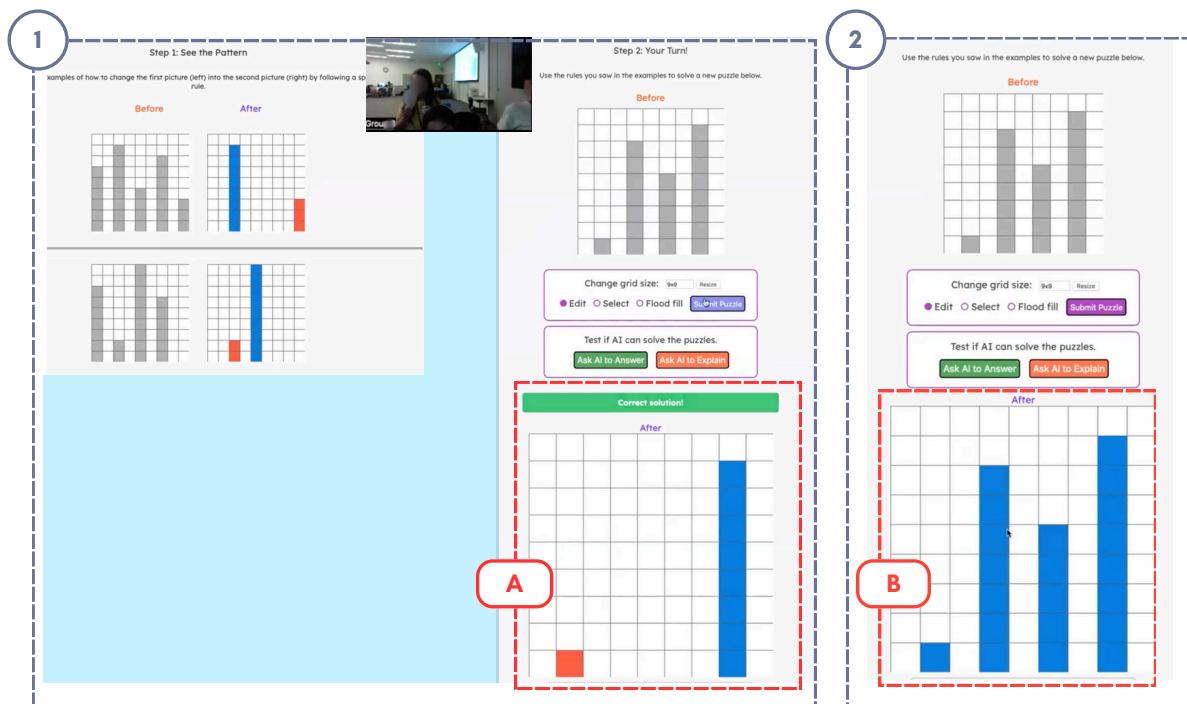


Figure 3.9: Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI’s attempt at the same puzzle. (A) highlights the children’s correct solution, whereas (B) shows the AI’s incorrect attempt.

Moreover, children’s engagement with AI Puzzlers wasn’t solely tied to the AI’s mistakes; they were also drawn in by the puzzles themselves. Children viewed solving the puzzles

as opportunities for problem-solving and personal accomplishment. Even as the puzzles became increasingly difficult, the children maintained their interest, often describing the more complex puzzles as “*fun*.” Peer collaboration consistently reinforced this enthusiasm, as children confirmed and validated each other’s solutions before turning to the AI for comparison. For example, in Session 1, Juno and Hana, eagerly tackled the puzzles from the start. Looking at their next puzzle (see Figure 3.9) Juno remarked, “*This is pretty easy*,” while Hana added excitedly, “*I want to do this because it looks so fun*.”

Together, they analyzed the puzzle, with Hana reasoning through the height and color relationships, stating, “*It’s the tallest one and the smallest one, correct? Blue would be the tallest*.” Juno confirmed, “*This looks right*,” and the pair successfully solved the puzzle while the AI failed (see Figure 3.9). As their excitement grew, Hana exclaimed, “*We can go forever*.” The facilitator reminded them, “*There’s only up to level 4*,” to which Juno confidently replied, “*We can go farther than the AI*,” and Hana affirmed, “*Yes, we can*.” This exchange highlights the children’s belief in their abilities, expressing confidence in surpassing the AI. Reflecting on their success, Hana laughed and said, “*AI would have gone out at the first question*,” underscoring their shared sense of accomplishment and enjoyment in outsmarting the AI.

Overall, two key dynamics drove children’s sustained engagement with AI Puzzlers in Session 1: 1) their surprise at the AI’s mistakes, especially for puzzles they considered “easy,” and 2) the satisfaction of solving challenging puzzles and comparing their correct solutions to the AI’s. Despite the AI’s repeated failures, children eagerly engaged with each puzzle, enjoying the process of outsmarting the system, testing their problem-solving skills, and comparing solutions. This blend of competition, collaboration, and humor kept them motivated, creating an engaging experience.

Children’s Iterative Debugging of AI Errors

As children interacted with AI Puzzlers, they quickly spotted incorrect AI solutions due to the visual nature of the puzzles, and more importantly, understood how the AI was failing. Similar to debugging in programming, in Session 2, children actively guided the AI by identifying its

errors, providing corrective hints, testing their hints, and refining their instructions based on the AI’s outputs. While the visual modality of the puzzles made it easier for the children to iteratively test and improve their instructions, they also encountered AI’s limitations, realizing that it often misunderstood or failed to fully follow their instructions.

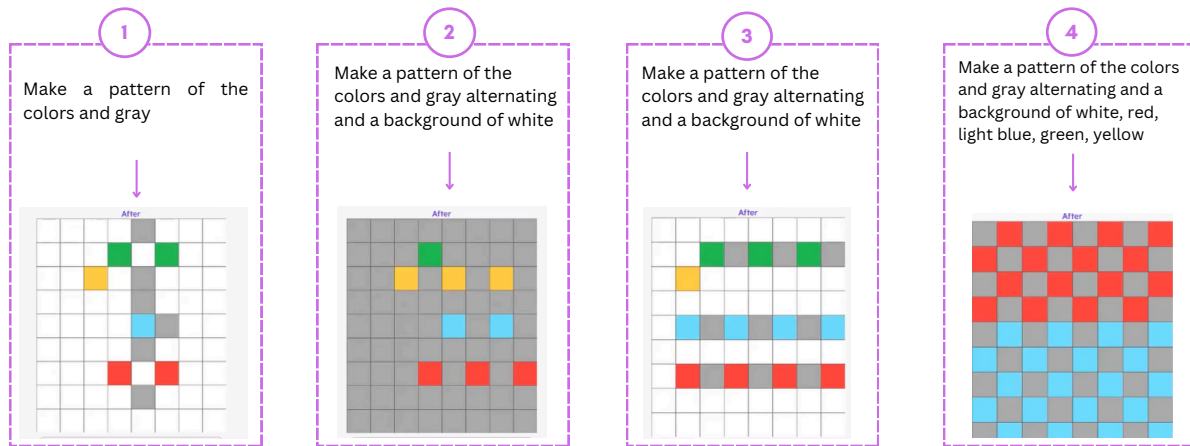


Figure 3.10: Children iteratively refined their instructions to guide the AI towards solving the puzzle. The sequence showcases the increasing specificity in their hints and AI’s corresponding outputs.

For example, in Session 2, Ivy, Juno, Rina, and Elise were trying to help the AI solve the second puzzle (see Figure 3.10). The children’s debugging process began when Ivy suggested the first hint, “*Make a pattern of gray and a different color,*” while Juno proposed a more complex version: “*Make a pattern of gray, the color gray’s after, and then gray, and the color gray after.*” After some discussion on avoiding confusion for the AI, the group settled on a simpler hint, “*Make a pattern of the colors and gray.*” However, when the AI produced an incorrect grid of colors and gray, Rina immediately recognized the mistake, commenting, “*It still can’t solve it though.*” The children examined the AI’s output and refined their instruction to, “*Make a pattern of the colors and gray alternating,*” which led Rina to observe that the AI’s output was “*kind of alternating now.*” At this stage, the children recognized the AI’s partial success but also noticed that it didn’t fully capture the pattern they had envisioned. Ivy pointed out another issue with the background color, saying, “*Now we need to add a*

background of white.”

The children’s increasing specificity in their hints mirrored the process of narrowing down AI’s errors by fine-tuning instructions. Ivy specified the background color in her instruction, typing “*make a pattern of the colors and gray alternating and a background of white,*” but when the AI’s output still didn’t reflect the correct background, Ivy remarked, “*That’s not a background of white.*” The children had to systematically address different parts of the puzzle—first the pattern, then the background—while the AI continued to misinterpret or overlook aspects of the instructions. On their second attempt with the same hint, the AI came very close to solving the puzzle, prompting a celebratory “*Yayyy!*” from the group. However, Ivy quickly commented “*There we go, but it forgot the yellow,*” while Rina noted that “*AI’s answer doesn’t always include all the colors shown in the original puzzle question.*” This led the group to modify their hint again, including a list of all the colors: “*Make a pattern of the colors and gray alternating and a background of white, red, light blue, green, yellow.*” By listing all the colors, the children demonstrated an understanding that more explicit, step-by-step guidance might help the AI avoid leaving out key elements. However, despite their efforts to refine and clarify their instructions, the AI continued to misinterpret them, leading Ivy to comment, “*I am so done with you, AI.*” Ivy’s expression of frustration reveals her realization that, despite the group’s efforts to provide increasingly explicit and refined instructions, the AI was still unable to fully grasp their intent.

Despite recurring misunderstandings, children not only refined their hints but also, in several instances, brainstormed alternative strategies to overcome the AI’s limitations. For example, in Session 2, Selah, Emi, Ren, and Mara attempted to help the AI solve the third puzzle (see Figure 3.11). The children’s debugging process first involved discussing what specific directions they should give the AI. Emi typed the hint, “*surround the colors with,*” before pausing. Ren then suggested, “*I think we shouldn’t say yellow, purple, and blue...I think we should say something simple.*” Taking this into account, Emi simplified the instruction to, “*surround the colors with more colors.*” However, the AI failed to generate a correct output, prompting Ren to suggest a new approach, “*What if we ask it to make donut shapes*

around the colors?” Ren’s suggestion reflects an attempt to simplify the instruction by using a familiar visual metaphor—a donut shape—to describe how the colors should be surrounded by rings or circles. The children likely believed that this metaphor would make sense to the AI since it’s based on a familiar object from their everyday experiences.

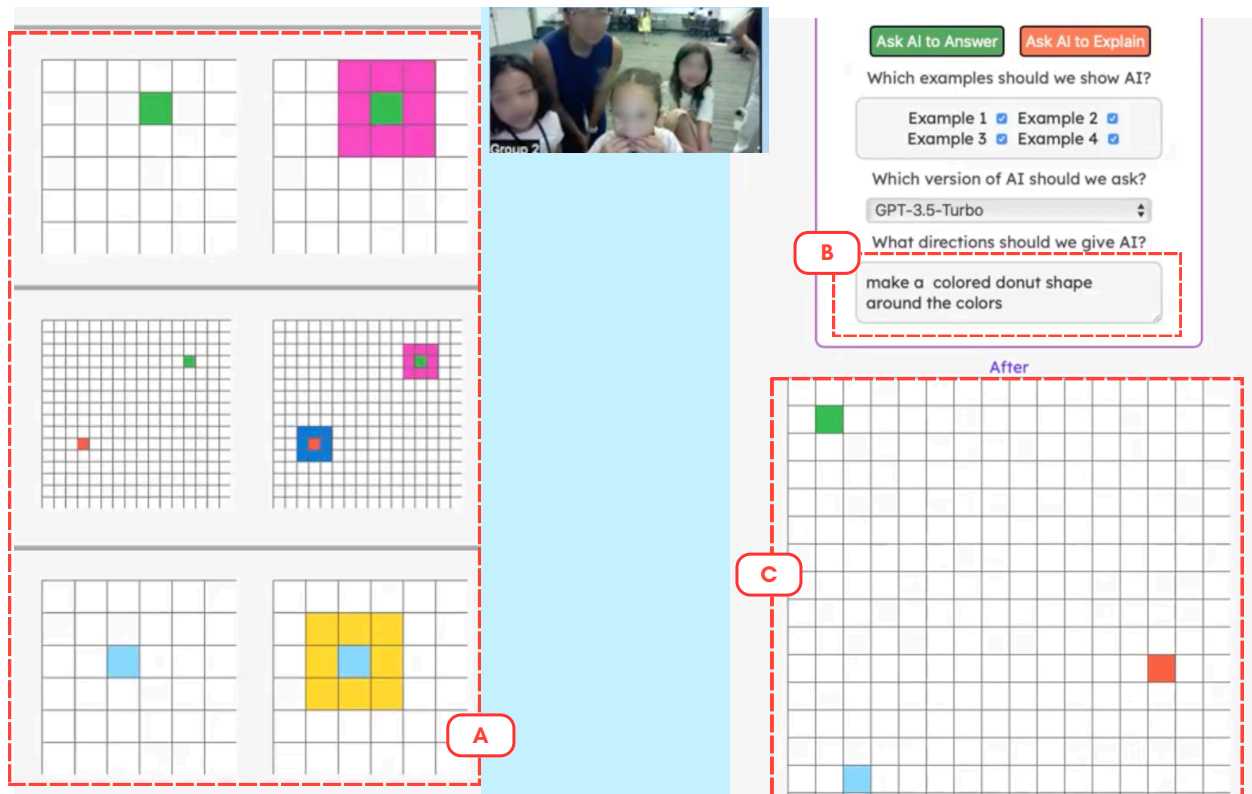


Figure 3.11: Children’s Interaction with Assist Mode in AI Puzzlers. (A) Displays examples of example patterns to infer the transformation rule. (B) Shows the instruction children provided to the AI: “make a colored donut shape around the colors.” (C) Illustrates the AI’s generated output in response to their instruction.

Selah then typed the revised hint, “*make a donut shape around the color*” (see Figure 3.11). However, even with this more specific instruction the AI still couldn’t produce the correct result. This illustrates a key limitation in how the AI fails to interpret visual metaphors that are culturally relevant or based on children’s experiences. While the concept of a “*donut*

shape” is clear to the children, the AI struggles to grasp such human-centered metaphors. In discussing design ideas on how to resolve AI’s issues, Emi suggested if they can test whether the AI could understand human instructions paired with visuals. She began drawing the correct solution on the grid to show what it would look like. To the facilitator’s question, “*Do you think if [we] gave it human instructions with pictures it would understand?*” Mara and Selah responded, “*Yeah,*” optimistic that the visuals might help bridge the gap. While the AI Puzzlers system didn’t have the functionality to process both visual and textual hints, we see this as an example of how the children not only iteratively refined their approach to guide the AI but also came up with a new idea to interact with the AI as they confronted its limitations.

Overall, children’s interaction with AI Puzzlers in Session 2 demonstrated their ability to actively engage in a process of problem-solving and debugging. Through iterative refinement of their hints, children adapted their strategies, and even explored creative approaches like using visual metaphors or pairing instructions with potential visual aids. This highlights both their resilience and their capacity for critical thinking in navigating AI’s limitations. However, their engagement was not without challenges. The AI’s persistent misunderstandings, even after detailed and increasingly refined instructions, sometimes led to frustration. This underscores the difficulty of working with an AI system that consistently failed to meet their expectations, showing how extended interactions with imperfect AI systems can affect children’s engagement.

3.5.2 Understanding AI’s Limitations Through Observing Inconsistencies

Children Identify Inconsistencies in AI’s Reasoning

While interacting with AI Puzzlers during Session 1, children reflected on AI-generated outputs by cross-examining the AI’s visual solutions with its explanations. There were many examples of children identifying discrepancies between the AI’s reasoning and its actual puzzle outcome. For example, in Session 1, after Hana and Juno correctly solved a puzzle, they

asked the AI for its solution. Upon reviewing the AI’s answer, they quickly identified it as incorrect and requested an explanation from the AI (see Figure 3.12).



Figure 3.12: Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI’s attempt at the same puzzle. (A) highlights the children’s correct solution, whereas (B) shows the AI’s incorrect attempt. (C) provides the AI’s explanation of how it solved the puzzle.

After reading the AI’s explanation, Hana remarked, “*This*”—pointing at the explanation—“*has nothing to do with this,*”—pointing at the AI’s pattern. She then added, “*It’s like someone who is not listening,*” suggesting that she perceived the AI’s explanation as disconnected from its visual output, much like a human who isn’t paying attention. Juno built on this observation, saying, “*The AI said it was following a pattern, but what pattern is this? This is not a pattern. It is just red, red, red,*” as he pointed at the AI’s solution. This example demonstrates that the children were not simply accepting the AI’s reasoning—they were actively engaging with its logic, looking for coherence between explanations and visual results. Furthermore, having correctly recognized the correct pattern earlier, they saw that the AI’s solution was unrelated to the pattern they knew was correct, leading Hana and Juno to further question the AI’s reasoning.

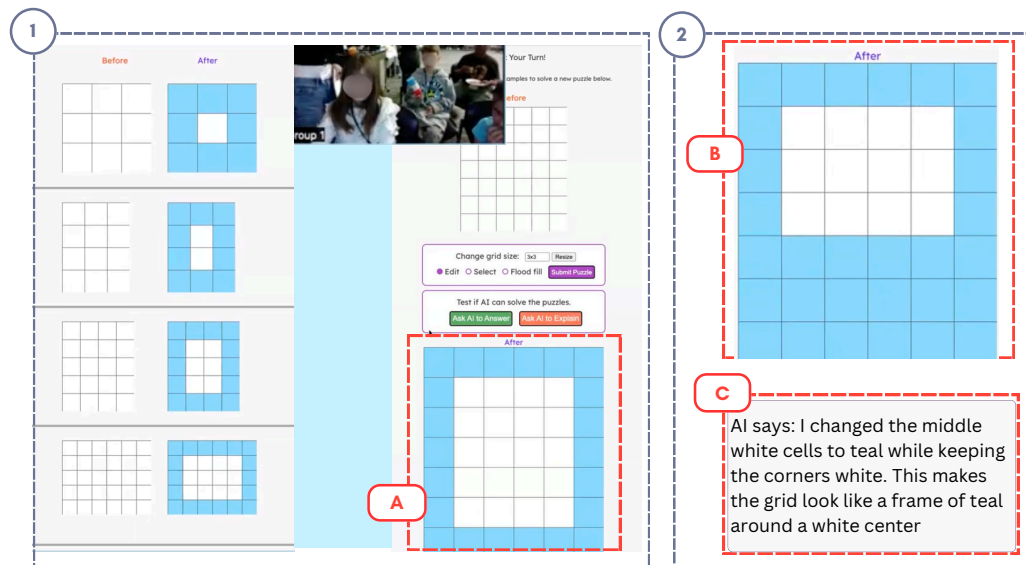


Figure 3.13: Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI's attempt at the same puzzle. (A) highlights the children's correct solution, whereas (B) shows the AI's incorrect attempt. (C) provides the AI's explanation of how it solved the puzzle.

This type of critical engagement with AI's reasoning recurred throughout Session 1. Children frequently encountered inconsistencies not only between the AI's explanations and its visual outputs but also contradictions within the AI's explanations themselves. For instance, when Noa, Finn, Leila, and Zayn asked the AI to explain its solution (see Figure 3.13), the AI explained, *"I changed all the middle white cells to teal while keeping the corners white. This makes the grid look like a frame of teal around a white center."* Finn considered the explanation and said, *"The explanation is right; it just did it wrong,"* implying that while the AI's reasoning was correct, the error was in its visual execution. However, Noa identified a contradiction between the AI's first and second sentences. *"It says the corners are white, but the corners should be teal,"* she pointed out, noting that if the corners were indeed white, as the AI claimed, the grid would not form the intended frame of teal. This highlights how children, while able to easily detect errors in visual outputs, had to engage in more critical reasoning to spot contradictions in AI-generated text-particularly when the text appeared

superficially reasonable.

Similarly, in another group, the children recognized a comparable issue of superficial correctness in AI's reasoning where the AI's explanation seemed logical on the surface but did not provide enough detail for the children to understand its approach (see Figure 3.14). For instance, when Ivy, Mara, Ren, and Emi asked the AI to explain its reasoning for the second puzzle, the AI's explanation stated that it identified a pattern in how the colors changed in each row of the input grid. It claimed to have applied the same pattern to predict the outputs in the new puzzle.

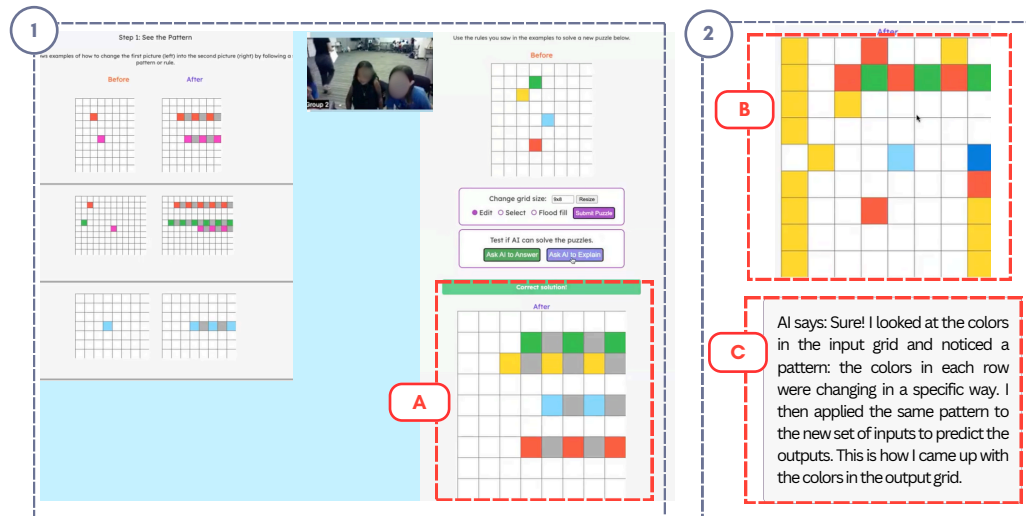


Figure 3.14: Panel (1) shows children collaboratively solving a puzzle, while panel (2) presents the AI's attempt at the same puzzle. (A) highlights the children's correct solution, whereas (B) shows the AI's incorrect attempt. (C) provides the AI's explanation of how it solved the puzzle.

However, AI failed to specify the exact nature of the pattern or how the pattern influenced the color changes in the output, leaving critical parts of its reasoning vague. This prompted Ren to remark, “*Well it is explaining, but I didn't understand it,*” highlighting the gap between the AI's surface-level explanation and the children's need for a more detailed and comprehensible breakdown of its reasoning. Ivy critiqued this vagueness saying, “*AI is very scientific, given its scientific explanation, but sometimes it's better not to go super, duper*

scientific and just go by your three references.” By “*scientific*,” Ivy meant that the explanation was too technical or abstract lacking practical clarity. Her reference to AI should follow the “*three references*” reflects her preference for the AI to apply a more straightforward approach to solving the puzzle—one that directly considers the three examples demonstrating the transformation rule, favoring a more direct and understandable reasoning.

Overall, children’s critical evaluation of AI-generated outputs demonstrated their active role in scrutinizing the reasoning behind the AI’s decisions. Rather than passively accepting the AI’s responses, they engaged in a process of deeper analysis, identifying inconsistencies and contradictions by cross-examining the AI’s visual solutions with its textual explanations. This critical reasoning went beyond merely checking if the answer was correct; it involved questioning the logic, coherence, and transparency of the AI’s problem-solving process. Additionally, they sought to understand how the AI arrived at its conclusions, recognizing when explanations lacked detail or coherence, and pushing back when reasoning was unclear. These findings highlight the importance of designing AI systems that allow children to not just receive answers but also critically engage with the logic behind them.

AI’s “Scientific Brain” vs. Human Problem Solving

Throughout Session 1, children recognized that AI approached problem solving differently from humans. While children found the puzzles “easy,” they realized the puzzles were “super hard” for AI as it struggled to solve them, displaying inconsistencies in its reasoning. In several instances, children reflected on the traits of human problem-solving, such as reasoning, abstraction, and creativity—qualities they felt the AI lacked.

For example, in Session 1, after Noa, Leila, Finn, and Zayn correctly solved a puzzle, they asked AI to do the same. The AI provided an incorrect solution, which prompted the group to ask AI again. On AI’s second attempt, Finn observed, “*It figured out the pattern but didn’t get the color positions right.*” Reflecting on this improvement, Finn hypothesized, “*AI is learning.*” This led the group to predict that the AI would perform better on the next attempt. However, on the third try, AI’s solution was even worse (see Figure 3.15).

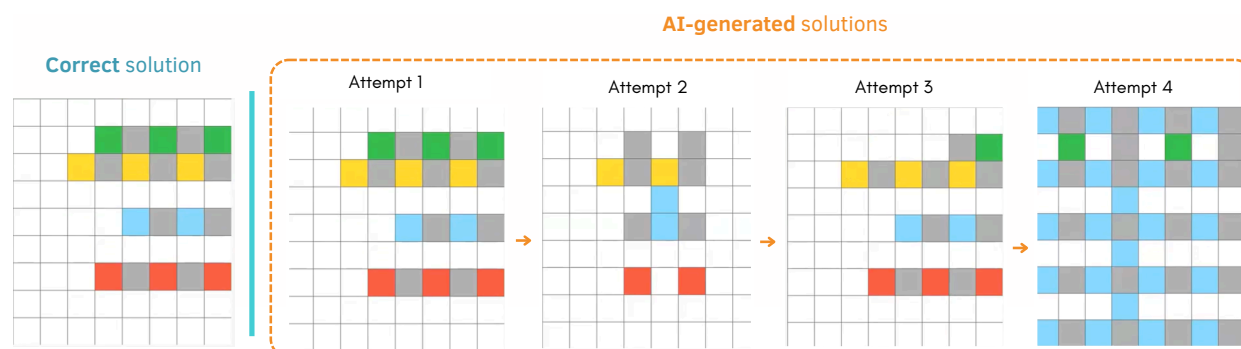


Figure 3.15: The correct solution (left) is compared with AI-generated attempts (right).

This led Zayn to observe, “*AI doesn’t have the same mind as us.*” This statement marked a turning point in the group’s understanding of AI, as they began to distinguish between human cognition that is capable of learning and AI’s processes. Zayn’s observation prompted the facilitator to ask, “*Well, what’s the difference in our minds?*” Leila noted, “*This is the internet’s mind.*” Noa elaborated, “*It’s trying to solve it based only on the internet, but the human brain is creative. AI only has the info it’s given, but humans have other experiences.*” Their reflections highlight how children grasped the limitations of AI, understanding that AI’s responses are constrained by the data it has access to such as the internet. By contrast, when Noa says, “*human brain is creative*” and “*humans have other experiences,*” she underscores how humans can draw on a wide range of knowledge, including emotional and situational experiences, to approach challenges more flexibly, whereas AI is limited to the information it has been given or can access.

Another example of children differentiating between human and AI problem solving occurred in Session 1, when Ivy, Ren, Emi, and Mara observed the AI unsuccessfully attempting to solve a puzzle (see Figure 3.14). Pointing to AI’s solution, Ivy asked, “*What is that blue? Look at the references and think like a human being!*” Here, Ivy was commenting on the AI’s inability to abstract information from the references, as a human might, to solve the puzzle. Later, when the facilitator asked what they thought was happening to AI, Ren

explained, “*It is taking the references and is able to copy and paste colors but not use the context of the patterns.*” Ren’s statement highlights their understanding that, while the AI could process the data it was given (the references and colors), it was failing to apply abstraction and reasoning necessary to complete the task. Thus, the AI could mimic certain elements, such as color, but it struggled with the more complex, abstract reasoning needed to solve the puzzle accurately. Ivy added to this by saying AI’s “*scientist’s brain*” has “*ones and zeroes[that] help it understand the color, but ones and zeroes aren’t the smartest.*” Her observation, along with Ren’s, reflects the children’s growing awareness that while AI can handle tasks like identifying colors, it lacks the creative and abstract reasoning necessary to grasp the puzzle’s full context.

Similarly, children also observed the lack of reasoning in the AI’s approach, describing its responses as “rapid random guessing.” For example, in Session 1, when the facilitator asked Juno and Hana to reflect on how the AI solves the puzzle, Juno repeatedly clicked the “Ask AI to Answer” button, pointing out, “*Look, it changes every time.*” (see Figure 3.16).

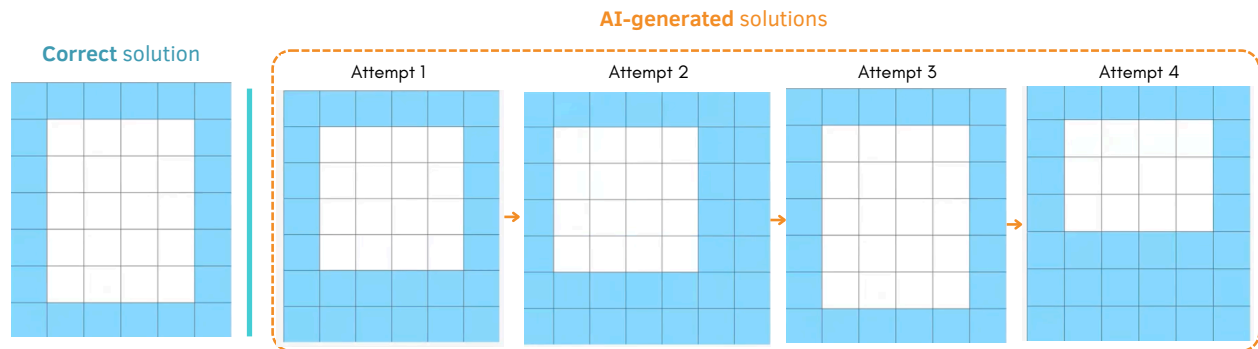


Figure 3.16: The correct solution (left) is compared with AI-generated attempts (right).

AI’s lack of reasoning became evident to the children when Juno noted that the AI’s answers changed with each attempt, implying no continuity in its problem-solving strategy. Instead of logically building on previous attempts, the AI appeared to reset its process, attempting a new guess each time without learning or adapting from prior mistakes. This

behavior led Hana to remark “*AI just keeps guessing,*” and Juno added, “*AI is stupid*” and only “*gets lucky.*” When the facilitator asked, “*What can the AI do to be smarter?*” Juno responded, “*Not guess,*” indicating the children’s growing awareness that the AI was not reasoning but merely stumbling upon the correct answer by chance.

Overall, children’s reflections across Session 1 and Session 2 demonstrated their growing awareness of the fundamental differences between human and AI problem-solving. They recognized that AI operates within rigid parameters, relying on data it has access to, often defaulting to trial-and-error rather than employing the kind of reasoning and creativity characteristic of humans. Their experiences reinforced the notion that AI, while capable of processing and replicating information, lacks the ability to apply abstraction and draw on diverse experiences to solve reasoning problems.

3.6 Discussion

In this section, I discuss how AI Puzzlers positioned children in active, inquiry-driven roles that encouraged experimentation (T1) and supported opportunities for children to compare their own reasoning with that of the AI (T2), prompting reflection on both human and machine problem solving. My findings offer recommendations for designers of AI systems for children, as well as researchers and educators who support children’s engagement with AI technologies.

3.6.1 Positioning Children as Active Inquirers in generative AI Interaction

AI Puzzlers positioned children as active inquirers, encouraging them to identify, analyze, and debug errors in generative AI’s outputs. By embedding this process in puzzle-based gameplay, the system sustained engagement while fostering reflection on AI’s limitations and capabilities. Below, I elaborate on each of these aspects.

Encouraging Critical Evaluation of Generative AI’s Outputs From the outset, children expressed confidence in AI’s ability to solve ARC Puzzles, reflecting a general belief

in AI’s capabilities. This belief aligns with broader patterns observed in prior research, where children’s early interactions with AI technologies often reflect an overestimation of AI’s intelligence [308, 13, 250, 215]. However, as they engaged with AI Puzzlers, children quickly realized that the puzzles they considered “easy” were challenging for the AI. The puzzles’ visual nature ensured that there was no obscurity in the way generative AI presented information, while accompanying textual explanations provided a step-by-step breakdown of the AI’s reasoning [211, 209]. This led to moments of visible disbelief, such as when children reacted to generative AI’s incorrect solutions with exclamations like “WHAT!!”

Encouraging Iterative Debugging & Reflection The Assist Mode in AI Puzzlers—where children provided hints to guide the AI—helped them develop an emergent schema of generative AI’s capabilities and limitations. As they iteratively refined their hints, shifting from vague commands like “make a pattern” to more precise instructions like “make a pattern with alternate colors,” they demonstrated a growing understanding of how generative AI processes information. The persistence children showed in fine-tuning AI’s outputs suggest that they viewed AI as a system requiring guidance rather than an infallible problem solver. Notably, we found that children’s explanations of AI’s reasoning became more sophisticated when using Assist Mode compared to AI Mode, where they only observed the AI making mistakes. These findings align with prior research, emphasizing the importance of providing children with opportunities to actively experiment with AI models and debug their assumptions about AI. [99, 158, 159, 216].

Fostering Sustained Engagement Through Puzzles The game-like nature of the puzzles revealed children’s inclination to engage in competitive problem-solving, often positioning themselves against the AI in an effort to “beat” its performance [232, 168, 347]. This dynamic, which aligns with game-based learning and the motivational role of competition in games [318, 350, 47, 60], not only sustained engagement but also reinforced children’s recognition of their own problem-solving strengths. Prior research suggests that children often perceive

AI as lacking in creativity and flexible thinking [208], and similar patterns emerged in AI Puzzlers. As children saw that the AI struggled with the ARC puzzles, they began comparing how humans and AI solve problems. This led them to recognize AI’s reasoning limitations and also made them more confident in their own abilities. This supports Long and Magerko’s [?] argument that fostering an understanding of both AI’s limitations and human strengths can empower users to leverage their own cognitive abilities in domains where AI falls short. In the next section, we present design implications for supporting AI literacy, grounded in these insights.

3.6.2 Implications for AI literacy and Generative AI Systems for Children

Designing for Interpretability Without Cognitive Overload

Building on prior research on AI interpretability [146, 252, 166], this chapter emphasizes the importance of making AI decision-making understandable for fostering AI literacy. While generative AI models can provide justifications for their outputs, overly lengthy or text-heavy explanations may overwhelm young users and discourage engagement [284]. Future AI systems could support interpretability by generating visual reasoning traces. For example, generative AI systems could create real-time visual representations, such as decision trees, flowcharts, or animated characters that walk through the AI’s reasoning process. This would allow children to trace AI’s logic step by step, encouraging them to question or revise AI-generated reasoning. Additionally, in AI Puzzlers, we found that children benefited from side-by-side visual and textual comparisons, which helped them verify correctness before engaging with the AI’s explanation. To distribute cognitive load across visual and verbal channels [211, 209], AI-generated explanations could highlight or animate key visual elements, with corresponding narration in real time to reinforce its reasoning.

Using Validity Markers to Guide Children’s Attention

Inspired by prior work on human-AI collaboration [316, 166], I argue that AI systems designed for children should not only present information but also support them in evaluating its reliability and accuracy. Research on uncertainty visualization [31, 247] demonstrates that well-designed visual markers can help users, including children, focus on areas requiring deeper scrutiny. Building on this idea, generative AI tools for children can incorporate explicit validity markers, such as color-coded confidence levels (e.g., green for high confidence, orange for uncertainty), uncertainty flags (e.g., warning icons next to questionable AI-generated answers) or playful prompts (e.g., “Try testing this answer!”), to prompt critical attention and active evaluation. However, prior studies suggest that such markers must be carefully designed to prevent over-reliance on AI confidence indicators, as children may assume high-confidence outputs are always correct. Thus, validity markers could be paired with explicit scaffolds that support evaluation strategies, such as comparing alternative solutions, identifying contradictions, testing claims with counterexamples, and justifying answers with evidence.

Designing for Reflection Through AI Experimentation

Prior research suggests that children develop a deep understanding of AI concepts when they can tinker with system parameters and observe how different inputs influence outcomes [262, 43, 106]. In AI Puzzlers, children engaged in tinkering and experimentation by adjusting the number of examples provided to AI, modifying hints, and selecting different AI versions to see how these changes influenced AI-generated solutions. This process was not solely about obtaining correct answers but also about interrogating AI’s reasoning and making sense of its decision-making patterns. Similarly, AI systems can be designed such that instead of simply watching an AI generate an explanation, a child might be able to pause at key points, tap on specific parts to get additional details, or even manipulate variables to see how AI reasoning changes. However, I also observed that children sometimes struggled to interpret AI’s reasoning, particularly when multiple factors influenced an output. This suggests that

AI literacy tools should not only support experimentation but also scaffold reflection on how AI arrives at its responses. For example, AI-enabled platforms could allow children to modify AI parameters while also providing structured prompts that encourage them to articulate hypotheses, compare outcomes, and reflect on differences. By embedding both interactive tinkering and guided reflection, these systems can support both “reflect in action” (thinking while doing something) and “reflect on action” (thinking after you have done it) [273].

3.7 Limitations & Future Work

While this study was designed to maximize depth and rigor, certain methodological choices naturally shaped the scope of the findings and suggest directions for future work. First, the co-design sessions engaged 21 children from a single geographic region, all of whom had prior experience with participatory design. This experience likely shaped their ability to confidently share opinions with adults, engage in constructive disagreements, and articulate their reasoning in real time. While this facilitated rich discussions, it also means that these findings should be understood as formative theoretical generalizations rather than statistical generalizations [341]. Future work could examine how children in different settings, such as schools and libraries, and across diverse cultural and linguistic contexts engage with AI Puzzlers. Expanding to these contexts would help illuminate how different social, educational, and cultural dynamics shape children’s evaluation of generative AI outputs and support a broader understanding of AI literacy development globally.

Second, although the findings show that children can detect and reflect on generative AI errors within the structured environment of AI Puzzlers, this study did not assess whether such learning transfers to other, more open-ended generative AI interactions. Future work could examine this transferability by investigating whether engaging with AI Puzzlers supports children in interpreting, questioning, or critiquing generative AI systems they encounter in everyday contexts. I also plan to expand AI Puzzlers to include other modalities, such as voice-based interactions, to more fully represent the range of AI systems children encounter in their daily lives.

Third, I selected ARC Puzzles because they are both engaging for children and commonly used as a benchmark for evaluating AI reasoning [55]. At the same time, their reliance on color-based differentiation may pose accessibility challenges. Future work could build on this approach by exploring alternative puzzle formats or game-based designs to broaden accessibility. Lastly, at the time of this study, I used the most advanced available version of ChatGPT (GPT-4o). Children also had opportunities to engage with four other versions of ChatGPT to understand variations in AI performance. As AI systems continue to evolve, their problem-solving capabilities will inevitably shift, and more recent models may perform differently on ARC puzzles. Future work could extend AI Puzzlers by incorporating both newer and older generative AI models to surface these shifts and examine how children recognize and interpret generative AI’s changing capabilities. This could also inform the design of educational tools that scaffold children’s reflections on AI’s evolving efficacy, supporting their ability to critically engage with AI technologies over time.

3.8 Summary of Contributions

This work introduces *AI Puzzlers*, an interactive AI literacy system that enables children to investigate and critique the outputs of generative AI through visual reasoning puzzles. The design of *AI Puzzlers* was motivated by the first and second thesis statements of this dissertation, which propose that children benefit from opportunities to engage in iterative, hypothesis-driven experimentation (T1) and to compare their own reasoning with that of an AI model (T2). Through participatory design sessions with 21 children (ages 6–11), I examined how they detected inconsistencies in generative AI outputs, debugged AI-generated errors, and refined their strategies for guiding AI. My findings underscore the need for generative AI systems to present information in ways that support visual and textual comparison for reducing cognitive overload, foster active inquiry, and scaffold multiple ways of understanding AI-generated content. I hope that this work will inform the design of AI literacy tools that empower children to critically evaluate AI-generated content and develop a deeper

understanding of AI's strengths and limitations.

Chapter 4

CLIP4KIDS: SUPPORTING CHILDREN IN INTERROGATING AI BIAS¹

4.1 Preface

AI systems are shaped by the social contexts in which they are developed, and as a result, their outputs can reflect and reinforce broader patterns of societal bias and inequality. This makes it critical to support children in recognizing when AI systems produce outcomes that may be unfair or discriminatory, so that they can understand, resist, and challenge unjust AI decision-making. Engaging with the social dimension of AI therefore requires creating opportunities for learners to surface the underlying decisions embedded in these systems and to examine their consequences. In this chapter, I engage with the core value of *Centering Justice* and concentrate on the *Ethics and Impact* practice outlined in Mills et al.’s AI literacy framework, which invites learners to investigate how datasets and algorithms may reproduce societal biases and to evaluate the potential harms of AI technologies for individuals and communities [217].

4.2 Introduction

In November 2021, the United Nations Children’s Fund (UNICEF) released version 2.0 of its *Policy Guidance on AI for Children* [235], with recommendations for AI technologies that uphold children’s rights. One of the recommendations is “*prioritize fairness and non-discrimination for children*,” noting that “*AI must be for all children*” (p. 8). Another

¹This chapter is based on collaborative work previously published as: Aayushi Dangol, Robert Wolfe, Akeiyah Dewitt, Ben Chickadel, Julie Kientz, and Sayamindu Dasgupta. 2025. Reading AI and Reading the World: Using an Interactive AI System to Promote Children’s Understanding of AI Bias. *ACM Transactions on Computer-Human Interaction* 32, 6, Article 66 (December 2025), 30 pages. <https://doi.org/10.1145/3762807>

recommendation for preparing “*children for present and future developments in AI*” notes that AI literacy should include “*educating children on their rights as users, so that they can become conscious users of AI-based systems*” (p. 41). These two recommendations point to the need to support children in understanding how AI systems can perpetuate unfairness and discrimination, and in developing the capacity to question and evaluate these systems.

One way to build this understanding is by supporting children in reasoning about how AI systems make decisions [199]. A subset of understanding AI decision-making is AI classification, which involves the process of categorizing data based on specific features or attributes. In this chapter, I introduce CLIP4KIDS, an interactive system that helps children explore how biases become embedded in the decisions that guide AI classification. I focus on AI classification because the schemas used by AI systems do not always align with how people organize and interpret the world, and prior research suggests these processes are often not transparent to children [101]. At the same time, human systems of categorization are shaped by cultural, social, and individual experiences and can reflect existing biases [34]. Attending to these parallels is important because AI systems, much like human systems, are only as fair as the data and decisions they are built on [40]. Thus, in this chapter I address the following research questions:

- **(RQ1):** How do elementary students engage with interactive activities designed to surface AI bias?
- **(RQ2):** How can the concept of AI bias be integrated into existing educational frameworks for teaching critical thinking skills, and what impact does this integration have on student learning outcomes?

To explore children’s interactions with CLIP4KIDS and assess its potential impact on student learning, I conducted a 4-week pilot study with two fifth-grade classrooms at a US elementary school. My findings reveal that CLIP4KIDS supported an interactive and reflective learning process, helping children develop three key insights: 1) societal biases are often mirrored in technology, 2) their own biases and privileges interact with AI models, and

3) they are active participants in societal ecosystems capable of fostering either equity or inequity.

These findings support T1 by demonstrating how iterative experimentation with classification parameters deepened children’s understanding of how AI systems respond to human input. They support T2 by showing how encountering moments of misalignment between their own categorization schemas and the AI’s outputs prompted children to reflect on both their own reasoning and the assumptions embedded in the system. And they support T3 by illustrating how children’s engagement with CLIP4KIDS was shaped by and grounded in their own identities and lived experiences, drawing on their understanding of fairness and exclusion to interpret and critique AI behavior.

4.3 CLIP4KIDS: System Design & Development

CLIP4KIDS is an interactive system designed to help children explore how biases can become embedded in the decisions that guide AI classification. Accessible through any web browser, CLIP4KIDS requires no prior knowledge of AI or programming, making it an easy entry point for young learners. Children interact with CLIP4KIDS by engaging with three core classification parameters: 1) confidence scores, which estimate the likelihood of an input belonging to a specific category; 2) classification rules, which organize data into inclusion or exclusion categories; and 3) decision thresholds, which set the minimum confidence level required for a classification to be accepted.

CLIP4KIDS contributes a novel system-level design that enables children to explore and manipulate real-time, zero-shot AI classification processes without requiring prior training or fine-tuning. This means that children are not constrained to a small set of pre-trained categories. Instead, they can freely define, test, and revise classification rules on the fly and observe how the AI behaves with respect to those rules and decision thresholds, even when the system has never seen them before. It broadens the scope of inquiry from “*what happens when the data is biased?*” to “*how do human-labeled rules and thresholds shape the AI’s decisions?*” Below, I first discuss the design considerations that guided the development

of CLIP4KIDS, followed by an overview of its core system components.

4.3.1 Design Considerations

Promoting Comparison of Human and AI Classification Processes.

From early childhood, humans rely on classification to make sense of the world, a process central to what classification theory describes as essential for survival and adaptation [281]. By categorizing people, objects, and experiences, we build mental frameworks, or schemas, that guide our understanding throughout life. These schemas play a critical role when children interact with CLIP4KIDS, as they bring prior expectations about classification outcomes based on the categories they define and the context of the task. Immediate feedback from the system allows children to identify discrepancies when the AI’s classification does not align with their expectations. These moments of mismatch prompt deeper cognitive engagement as children compare their own schemas with the AI’s confidence scores, explore the reasons behind AI misclassification, and experiment with adjusting classification parameters such as redefining categories or modifying decision thresholds to refine AI behavior. Additionally, by comparing classification decisions made by AI to their own and those of their peers, children can engage in metacognitive reflection [115] and gain insights about their own internal schemas, making visible the typically hidden role these systems play in everyday life [34].

Encouraging Iterative Experimentation in AI Decision-Making

In CLIP4KIDS, children can freely experiment with various image inputs and classification parameters, including classification labels, confidence scores, and decision thresholds. These features promote participatory learning, allowing children to refine their inputs, test hypotheses, and iteratively explore the relationship between their decisions and the AI’s outcomes. The system also provides immediate feedback, enabling children to quickly observe the effects of their actions and encouraging them to connect, disconnect, and reconnect with their ideas [261]. While the AI model in CLIP4KIDS ensures consistent outputs for the same inputs,

the diversity of inputs chosen by children introduces variability in their experiences, fostering exploration. This variability invites rich discussions about how different decisions, such as defining categories or adjusting thresholds that lead to varying results, offering opportunities for learners to compare their findings and collaboratively analyze the role of human decisions in shaping AI outcomes.

Reducing Cognitive Overload in Understanding AI Concepts

Prior work in AI literacy highlights that while uncovering some of these often “black-boxed” AI processes can improve understanding, revealing too much too quickly can overwhelm learners [81]. Therefore, key aspects of AI concepts should be unveiled to children without overloading them. Mayer and Moreno’s cognitive theory of multimedia learning reinforces this idea, emphasizing that learning materials should be structured in a way that supports learners’ abilities to process new information without causing mental overload [211]. Their dual-channel principle suggests that humans process information via separate visual and verbal channels and by balancing content across these channels, learners can more effectively comprehend complex concepts without overwhelming a single channel [209]. In CLIP4KIDS, children interact with both visual elements (images) and verbal/numerical data (confidence scores and textual labels) which distributes the cognitive load [211, 209]. Furthermore, CLIP4KIDS scaffolds learning of AI classification concepts by supporting three progressively complex interfaces: 1) Wordcloud interface, 2) Classifier interface, and 3) Smartlock interface. Detailed descriptions of the three interfaces are provided in the design overview section below.

4.3.2 Design Overview

Wordcloud Interface

The Wordcloud interface enables children to explore how AI models interpret and classify visual inputs in real time. The interface continuously captures images from a live webcam at two-second intervals and visualizes the AI model’s top ten classifications as a dynamic

word cloud (see Figure 4.1). In the word cloud, the size of each word corresponds to the AI model’s confidence in that classification. This interactive setup encourages children to experiment with different inputs—changing their poses, introducing new objects, or modifying the background—to observe how these variations influence the AI’s classifications. To ensure a safe experience for children, we constrained the AI model’s output for the word cloud to a predefined list of 300 words derived from the ANEW lexicon [36]. This curated word list, manually reviewed for appropriateness, ensured that children were not exposed to harmful or inappropriate language while still allowing the AI to classify inputs based on patterns learned from its training data.

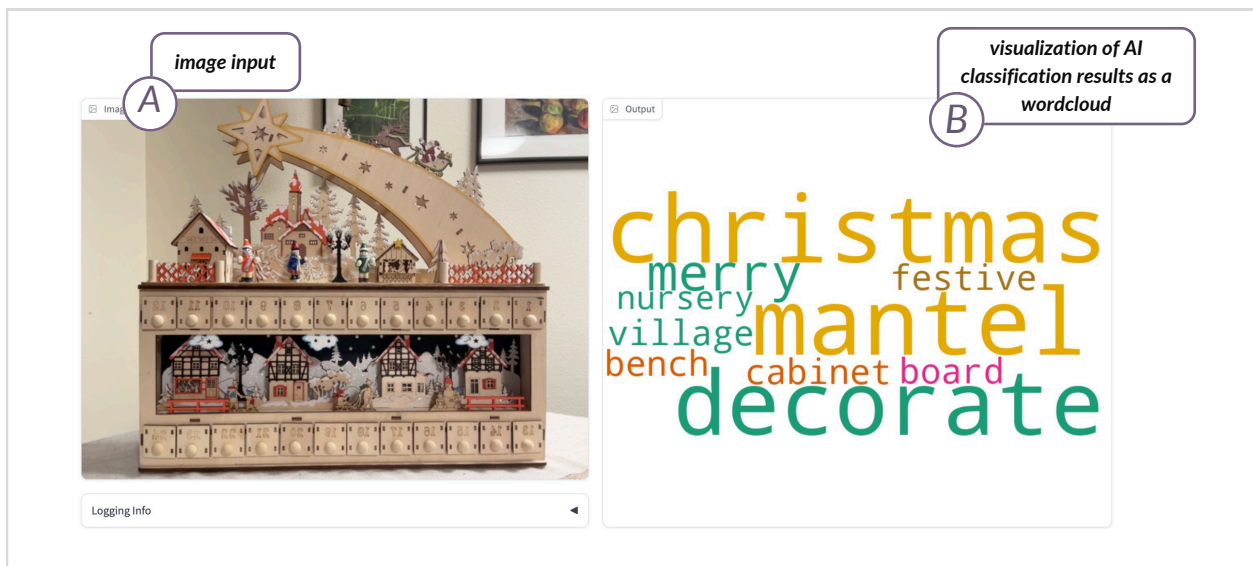


Figure 4.1: Screenshot of the Workcloud interface with added annotations highlighting system components. The system captures live image inputs from a webcam (A) and visualizes the AI model’s classification as a dynamic word cloud (B).

Classifier Interface

The Classifier interface is designed to help children understand that AI classification is not a neutral task but one shaped by the subjective categories that humans define. Through

this interface, children can create up to three labels that describe an image. The AI model then evaluates these labels and displays confidence scores, representing the likelihood that the image belongs to each label (see Figure 4.2). The label with the highest confidence score is highlighted as the AI’s classification decision. For example, for an image of a parrot, a child might create the labels “Colorful Bird,” “Flying Pet,” and “Green Parrot.” The AI then processes these labels and assigns confidence scores, such as 85% for “Colorful Bird,” 10% for “Flying Pet,” and 5% for “Green Parrot.” Since “Colorful Bird” has the highest confidence score, the AI classifies the image under this label. To encourage exploration, the interface allows children to modify labels dynamically. If the child modifies a label, for example, changing ‘Green Parrot’ to ‘Bright-Winged Bird’, the AI will recalculate the confidence scores and update its classification.

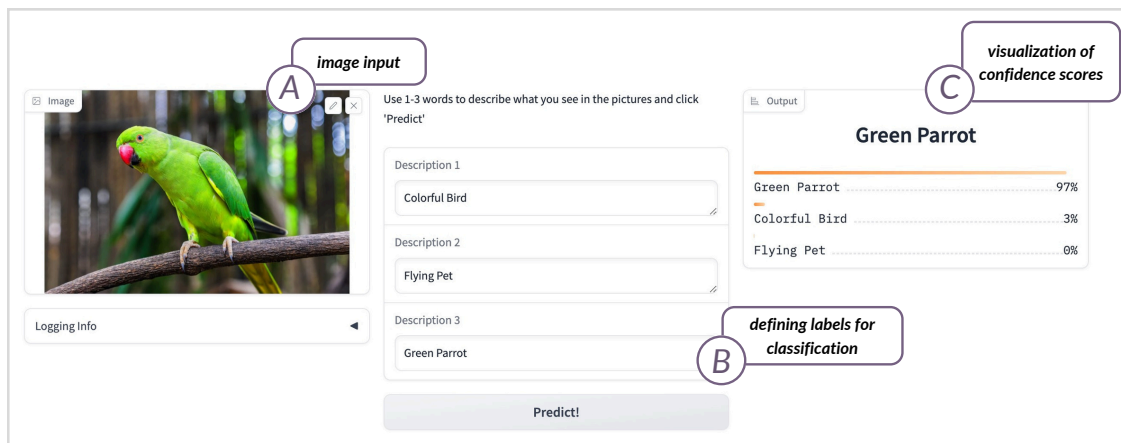


Figure 4.2: Screenshot of the Classifier interface with added annotations highlighting system components. The interface allows users to upload an image for classification (A), define labels describing the image (B), and view the AI model’s classification decision along with confidence scores for each label (C).

Smartlock Interface

The Smartlock interface simulates a smart door lock where access decisions—whether to open a door or keep it locked—are determined by classifications made by the AI model (see Figure

4.3). The metaphor of a locking door resonates with children’s everyday experiences of being included or excluded from group activities or navigating physical spaces where access is granted or restricted. This provides a relatable way to help children reflect on the real-world implications of AI-driven decision-making and the role humans play in shaping these decisions.

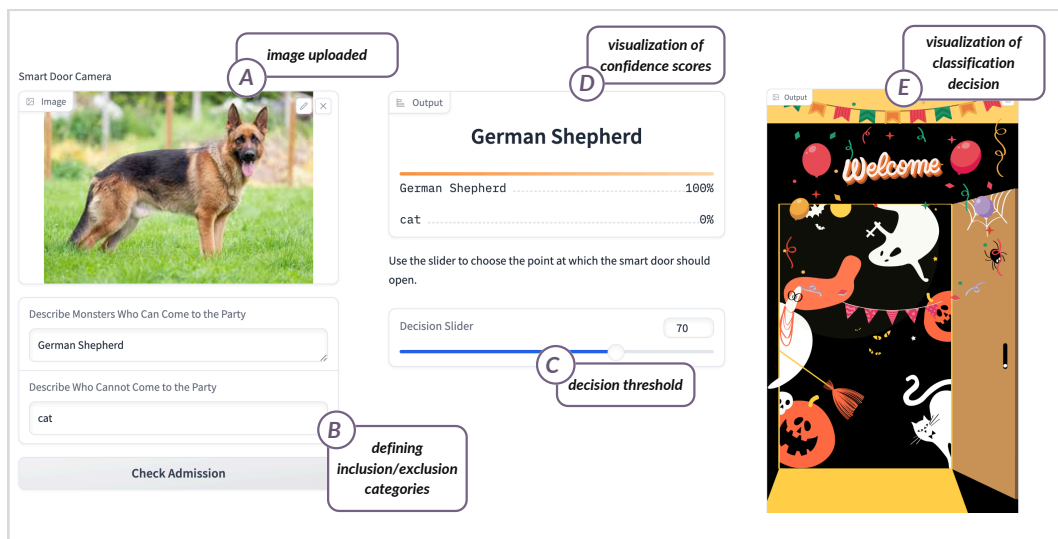


Figure 4.3: Screenshot of the Smartlock interface with annotations highlighting key system components. Users can upload an image for classification (A), define inclusion/exclusion categories (B), set a decision threshold(C), and view the AI model’s classification confidence scores (D) and final classification decision (E).

Children first define the inclusion and exclusion categories for the Smartlock interface, deciding which types of images will lock or unlock the door. For example, they may choose “Humans” as an inclusion category (unlock the door) and “Aliens” as an exclusion category (keep the door locked). They then upload an image and observe how the AI model classifies it based on the predefined inclusion and exclusion categories. The interface also incorporates an adjustable decision threshold, which allows children to specify the minimum confidence level required for the AI to grant access. For example, if the threshold is set to 70%, the door will only unlock if the AI is at least 70% confident the input belongs to the inclusion category. Through this interaction, children can also understand the concept of decision boundaries

and the trade-offs between sensitivity (detecting true positives) and specificity (avoiding false positives), which are key considerations in real-world AI systems.

4.3.3 Technical Implementation

At the core of each CLIP4KIDS interface is the Contrastive Language-Image Pre-training (CLIP) model [251]. Specifically, we used CLIP-ViT-B32, an open-source model, to develop CLIP4KIDS. The stability of this publicly available version ensures a consistent learning experience across implementations, reducing the unpredictability associated with proprietary model updates. CLIP's multimodal architecture supports the interactive features of CLIP4KIDS by enabling quantitative comparisons between visual and textual data. The system processes visual inputs (e.g., images captured via a webcam or uploaded files) through CLIP's vision encoder, converting these images into numerical representations. Simultaneously, textual data is processed by CLIP's text encoder, transforming the text into a similar numerical format. Both outputs are then mapped into a shared "embedding space," where the system evaluates how closely the image and text are related. For example, if the system receives an image of a "German Shepherd" and the text label "dog," the vision and text encoders generate numerical representations for both inputs. These representations are compared in the embedding space, where the system calculates a similarity score. A high score indicates a strong match, such as between "dog" and the image of a German Shepherd. Conversely, a less relevant label like "cat" would result in a lower similarity score, reflecting a weaker match. This powers the system's classification functionality across the three interfaces. Figure 4.4 illustrates this process for the Smartlock interface.

4.4 Methods

To explore children's interactions with CLIP4KIDS and assess its potential impact on student learning, I led a team of five collaborators in conducting a pilot study with two fifth-grade classrooms at a U.S. elementary school. Prior research in HCI and Learning Sciences support the use of pilot studies in classrooms to evaluate the viability and acceptability of new

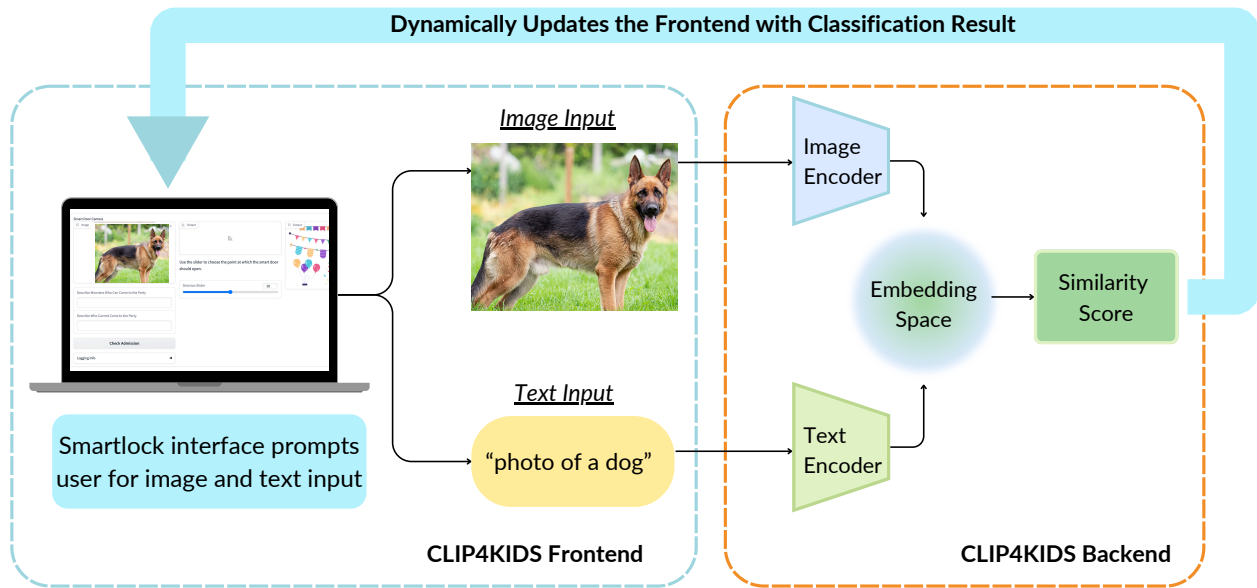


Figure 4.4: A representation of the Smartlock interface in terms of architecture and user flow.

technologies before implementation in larger-scale investigations [145, 144, 125, 98, 193]. Our study took place during the academic year, with students using CLIP4KIDS as part of their regular Technology class, under the instruction of their Technology teacher.

4.4.1 Research Setting

The pilot study was conducted at an independent school in the Pacific Northwest region of the United States, serving approximately 350 students from kindergarten through fifth grade. Of the total student population, 57% are students of color, and 47% are multilingual, collectively speaking 20 different languages. The school has a DEI curriculum which utilizes Learning for Justice’s *Social Justice Standards* [180] to “ensure students have learning opportunities in four domains: identity, diversity, justice, and action” [304]. The curriculum is aligned with US national standards and incorporates two widely recognized frameworks: the Common Core State Standards² (CCSS) and the College, Career, and Civic Life (C3) Social Studies

²<https://corestandards.org/>

State Standards³ [180]. The CCSS is intended to provide consistent learning goals across U.S. states and outline what students should know and be able to do at each grade level in English language arts and mathematics [15]. They emphasize the development of skills such as reading comprehension, evidence-based writing, and the application of knowledge through critical thinking [15]. The C3 Framework, developed by the National Council for the Social Studies, promotes inquiry-based learning and civic reasoning, encouraging students to investigate real-world social issues and engage in informed action [228]. At the time of the study, fifth-grade students were reading *Stamped (For Kids): Racism, Antiracism, and You* [52], a book that examines racial inequity in the US from both a historical and contemporary perspective as part of their school’s DEI curriculum.

4.4.2 Participants

Technology Teacher

To integrate CLIP4KIDS into the school’s Technology curriculum, I collaborated with the school’s Technology teacher who is also a co-author of this work, through weekly feedback meetings at every stage of the curriculum design and deployment process. Prior research underscores the importance of partnering with teachers in the design and implementation of new educational materials [279, 165]. Teachers’ pedagogical expertise and deep understanding of students’ learning needs enable them to identify the most effective ways to introduce complex concepts such as AI bias and to adapt their teaching approaches in response to unexpected classroom dynamics [271]. Our pilot study with CLIP4KIDS was the Technology teacher’s first time teaching AI and algorithmic bias concepts. Reflecting on his pedagogical approach, the Technology Teacher shared, “*I’m not interested in my students coming out of my program as coders. What I want is for them to leave with the ability to ask critical questions, to be excited about technology, and to approach the future with curiosity rather than fear.*”

³<https://www.socialstudies.org/standards/c3>

Child Participants

We distributed a brief study description to fifth-grade students and their parents/guardians through the school. 28 fifth-grade students (ages 9–10) participated in our pilot study. Child participants had not previously encountered AI concepts within their school’s technology curriculum. Prior to the data collection, written assent and consent forms were signed by child participants and their parents/guardians.

4.4.3 Classroom Sessions

We conducted four sessions each with two fifth grade classrooms between April and May 2023. The first three sessions focused solely on CLIP4KIDS. In the fourth session, we supplemented CLIP4KIDS with a separate, curated dataset of images generated by DALL·E 2⁴, a generative AI model that creates images from text prompts. These DALL·E outputs were not part of the CLIP4KIDS system but were included to spark reflection on bias in generative AI image outputs. Every classroom session consisted of: **Warm-Up Time** (10 minutes) during which the Technology teacher posed a warm-up question to encourage students to share their personal experiences with AI technologies and create an inclusive environment; **Activity Time** (30 minutes) where students collaborated in groups of 5-6 to engage with the CLIP4KIDS system and complete learning activities designed in collaboration with the Technology teacher; and **Reflection Time** (20 minutes), during which students completed worksheets individually before discussing with the class.

Shifting from the traditional teacher-student hierarchy, the classroom was re-imagined as a collaborative space where students and the Technology teacher co-constructed knowledge through dialogue and shared inquiry [313]. The teacher adopted the role of a facilitator, guiding discussions and activities while empowering students to take ownership of their learning. To mitigate potential power imbalances and the influence of children’s responses on each other, the Technology teacher promoted an environment where children felt comfortable expressing

⁴<https://openai.com/dall-e-2>

unique thoughts without undue influence from their peers or the adults. Additionally, the teacher navigated discussions in a way that minimized dominant behavior by any single participant and fostered collective idea generation, where contributions were built upon rather than overshadowed.

Session 1

Students interacted with the Worldcloud interface to understand how AI systems classify visual inputs. Children used craft materials like modeling clay, pipe cleaners, and construction paper to create tangible artifacts to test AI's adaptability to interpret and classify various inputs (see Figure 4.5). These tangible artifacts acted as unique inputs for the AI model to classify, allowing for a more personalized and hands-on interaction with the system. Students captured screenshots whenever the AI misclassified an object or action. Class-wide discussions facilitated by the Technology teacher encouraged students to reflect on moments when the AI's outputs diverged from their expectations and consider why the differences in AI classification might emerge.



Figure 4.5: Children prompting the Wordcloud interface with different types of visual inputs in Session 1.

Session 2

Students engaged with the Classifier interface to explore how AI classification is influenced by human-defined categories. They began by creating representations of friendly monsters using various craft materials and assigned descriptive labels (e.g., "feather monster," "boy monster") to their creations. Once these labels were entered into the Classifier interface, students clicked

the “Predict” button, prompting the system to capture an image of their monster creations and classify it based on the provided text labels (see Figure 4.6). Each student was tasked with independently composing three labels before collaborating on the final version. Students then participated in class-wide discussions on how personal perspectives and experiences shape labeling decisions and how such subjectivity can influence AI classification.

The screenshot displays the Classifier Interface. On the left, an image of a hand holding a pink card with a feathered monster is shown. Below the image is a 'Logging Info' button. To the right of the image, a prompt reads: 'Use 1-3 words to describe what you see in the pictures and click 'Predict''. Below this prompt are three input fields for 'Description 1', 'Description 2', and 'Description 3'. The first field contains 'feather monster', the second contains 'girl monster', and the third contains 'boy monster'. A 'Predict!' button is located at the bottom of these fields. On the right side of the interface, an 'Output' section shows the classification results for 'feather monster'.

Label	Percentage
feather monster	99%
girl monster	1%
boy monster	0%

Figure 4.6: Children explore the Classifier Interface in Session 2 by assigning descriptive labels to their monster creations and analyzing the system’s classification result.

Session 3

Students were introduced to the Smartlock interface through a narrative involving a character, Gizmo, a friendly monster hosting a party. They were then tasked with designing a smart lock to grant access to all the monsters they created in Session 2 while ensuring humans were excluded. Each student independently wrote an inclusion category (individuals to admit) and an exclusion category (individuals to exclude) before collaborating on a final version (see Figure 4.7). They tested their labels with different monsters and discussed whether the labels they came up with individually or collaboratively exhibited greater inclusivity and reduced bias. Children also experimented with the decision threshold to observe its impact

on classification outcomes. The Technology teacher guided discussions on the trade-offs between specificity and sensitivity, helping children understand how threshold adjustments can influence AI classification and the potential consequences for different stakeholders.

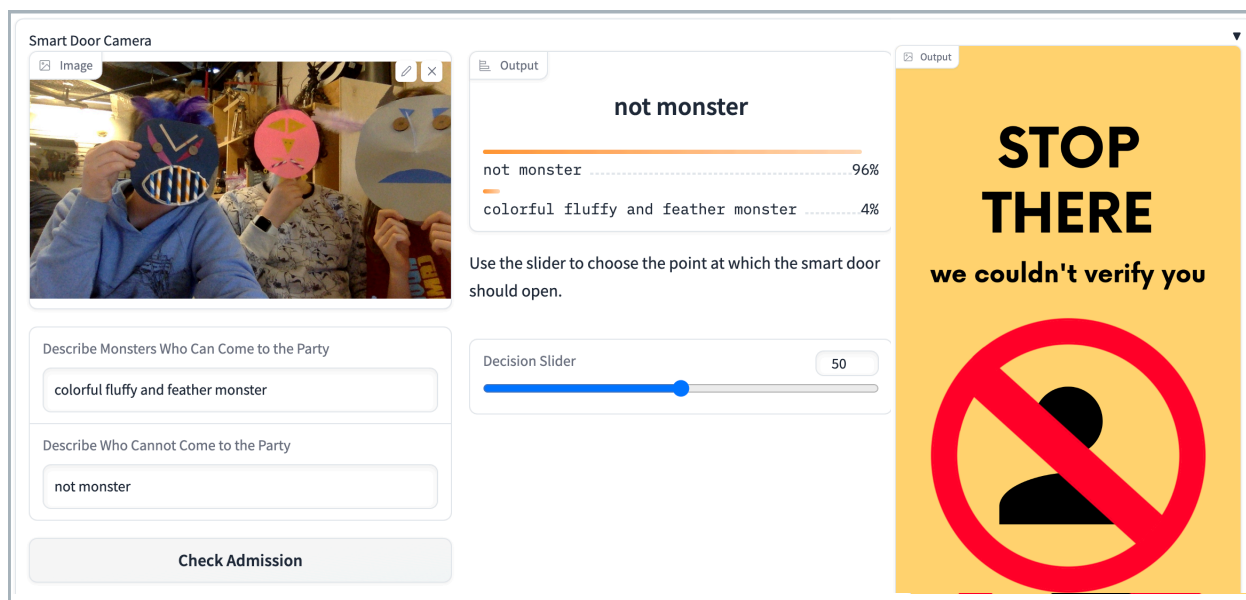


Figure 4.7: Children explore assigning inclusion and exclusion categories while experimenting with decision thresholds to design a Smartlock that grants access to their monster creations. In this example, the Smartlock interface classifies the image input as ‘not monster’ with 96% confidence, preventing the children’s monster creations from entering.

Session 4

Students began the session by drawing pictures in response to the prompt, “*Draw a picture of children celebrating a festival.*” They then examined a curated set of images generated by DALL·E 2 for the same prompt (see Figure 4.8) and compared them with their own drawings. Afterward, they completed an AI Bias Rubric, reflecting on biases present in both the CLIP4KIDS system and the DALL·E outputs. The session concluded with a brainstorming activity where students proposed design ideas for creating fair AI systems.



Figure 4.8: Pictures generated by DALL-E 2 for the prompt “Draw a picture of children celebrating a festival.”

4.4.4 Data Collection

During the classroom sessions, we collected log files from the CLI4KIDS system, which captured students’ interactions, including the text-based labels they created, the associated probability scores for each label, and the resulting classifications. Additionally, we gathered worksheets and artifacts produced by the students, audio recorded class discussions and compiled observational notes from the research team. To understand the broader pedagogical environment, we also conducted a 60 minute semi-structured interview with the Technology teacher.

4.4.5 Data Analysis

We employed a thematic analysis approach for data analysis [37]. The coding team—comprising of the first, second, and third authors— began by independently open coding 10% of randomly sampled data. This phase generated potential codes such as “stereotypes” and “bias in data.” Subsequently, the coding team and the senior author held two 1-hour meetings to merge and synthesize codes collaboratively, resolving disagreements through discussion until a consensus was achieved. Following these discussions, the first author created an initial codebook with six main code categories: 1) AI and human interaction; 2) AI bias; 3) Fairness; 4) Learning about AI; 5) AI ethics; and 6) Information access, each with corresponding sub-codes. The

coding team then utilized this codebook to analyze another 10% of the data, referring to the codebook while introducing new codes as necessary. The coding team continued to meet weekly over the next two weeks to address disagreements, working toward ensuring the analysis’s coherence and reliability. Following these iterative discussions, the first author developed a final codebook, and the remaining data was evenly distributed among the coding team. Each member of the coding team independently applied the agreed-upon codes and the team met to synthesize a final set of themes, which was informed by our theoretical framework.

4.5 Findings

I present student data in two ways: 1) SID_v , where ‘ID’ refers to the participant identification number assigned to the student, and ‘ v ’ denotes that the quote was spoken verbally during class discussions; 2) SID_w indicates that the quote is drawn from a worksheet, denoted with ‘ w ’.

4.5.1 Co-Construction of Knowledge

Students engaged in joint activities with their peers while interacting with the CLIP4KIDS system, providing numerous examples of co-constructing knowledge about AI, bias, and fairness. Some of the most notable instances occurred during Session 3, where students worked together in groups to design a fair smart lock that would grant access to all the monsters they had created previously by 1) deciding on the inclusion and exclusion categories and 2) deciding on the decision threshold together as a group. For example, one group of students started with “friendly monster” as an inclusion category and “humans” as an exclusion category. However, this was a challenging task because the concept of “friendly” varied across the group, and children quickly realized that their inclusion category didn’t account for all the monsters’ diversity and inadvertently excluded some friendly monsters. Figure 4.9 shows examples of friendly monsters created by students.

While some children thought “big eyes” indicated friendliness, others pointed out that



Figure 4.9: Student-created representations of friendly monsters, illustrating the diversity in interpretations of ‘friendliness.’

this excluded a monster with small eyes but a friendly personality. This process sparked an in-depth conversation about the difficulty of creating inclusive criteria and how biases can arise when applying seemingly simple rules to diverse groups. Reflecting on this process, S17_w noted, “We tried to describe our three monsters in like five words. Sometimes it would work and sometimes wouldn’t.” Some students even remarked on the difficulty in eliminating bias from AI systems because of the subjectiveness of humans labels. As S20_w shared,

This is a hard problem. Humans themselves are not fair. A human designing an AI have biases - ‘super bias’. It would be hard to make it 100 percent fair because they might program bias/stereotypes. Fair is not totally existent and varies between human beings.

Students also explored how adjusting the decision threshold could impact the AI’s decisions, leading to co-constructing knowledge about trade-offs between precision and inclusivity. S9_v reflected on their group’s process, “*We set the slider for opening the door to 70, and to let the AI to open when it saw a red fluffy monster.*” They came to the decision after multiple rounds of experimentation. Initially, S9’s group set the threshold at 30, which resulted in the AI unlocking the door for any input that even slightly resembled a “red fluffy monster.” Noticing this, they adjusted the decision threshold to 90, but found that the AI became overly restrictive, refusing to open the door even for monsters that closely matched

their intended description. Through iterative adjustments and discussions with peers, S9's group eventually settled on 70 as a balanced threshold, ensuring that the AI would correctly identify their target while avoiding false positives. Figure 4.10 illustrates this process, showing how different threshold settings influenced the AI's classification decisions and the group's reasoning behind their final choice.

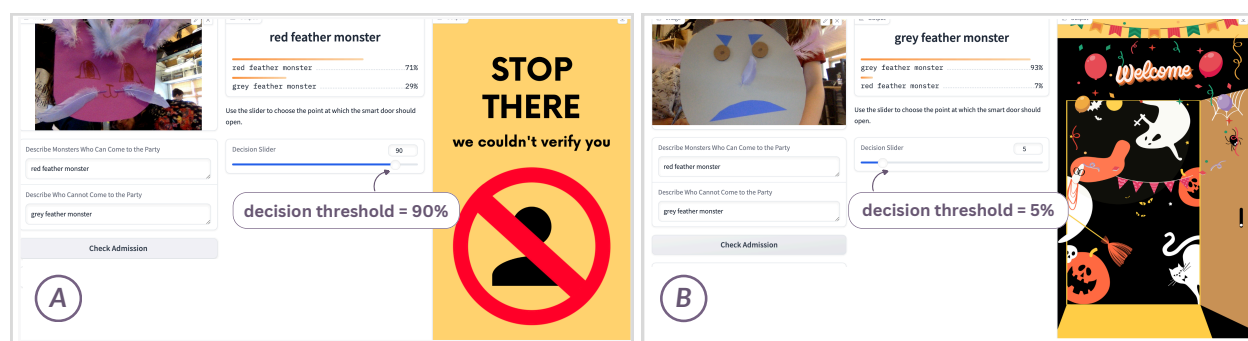


Figure 4.10: Students investigated the impact of decision thresholds on AI classification, engaging in an iterative process to balance precision and inclusivity. (A) A high decision threshold (90%) led to overly restrictive filtering, excluding even closely matching inputs. (B) A low threshold (5%) resulted in over-admission, allowing broad inclusivity.

The process of co-constructing knowledge also included discussion about how threshold adjustments can influence the accuracy of classifications and the potential consequences for different stakeholders. This was evident in the group discussion where some students noted the downsides of an overly inclusive approach. S13_v noted “*If it isn't [working] it could let a robber in*” and S19_v shared, “*I don't think the door lock is ready for actually guarding a door. While it should be fair, some bad people might come during that period of time.*” These reflections illustrate the importance of students working together to construct new understanding—they build upon each other's ideas, where one student's input sparks another's, and this iterative process leads to the development of more comprehensive insights. The Technology teacher, reflecting on how students were collaborating to uncover the mechanics behind AI decision-making, noted,

I feel like it's learning how a rainbow works. At first, there's this understanding of, oh, you see a rainbow, and it's pretty incredible. But as soon as you learn what's actually going on, it's even more incredible—because then you can actually find them. It's just like that. It's so much fun to watch that stuff happen.

4.5.2 *Dialogue as a Catalyst for Epistemological Curiosity*

During the classroom sessions, the Technology teacher employed a dialogical approach [120] to help students examine the AI system's outputs through the lens of their sociocultural perspectives. For example, during Session 1, the teacher used the outputs produced by the Worldcloud interface to encourage dialogue by asking questions such as “*Why do you think this word is larger than others? What connections do you see between these words? Are there any patterns that surprise you?*” In response to these inquiries, students began to notice how assumptions baked into AI systems often mirrored societal stereotypes that they had encountered in their own lives. For example, S1_w said, “*This AI is biased, based on color, gender, and stereotypes. You know how people think pink for girls and black for boys? And how people think that boys should have their short hair and girls long hair?*” Similarly, S3_w remarked that, “*This system always says "boy", "girl" and it will think someone is "wife" "mother" "love" "family" and it is a bit stereotypical*” (see Figure 4.11).

These moments in the classroom also encouraged children to reflect on how their own identities intersect with societal norms and how those norms are reflected and reinforced by technology. For example, S2_w observed, “*AI can stereotype certain things. So it could assume traits of someone based off of skin color if coded by a racist person. For example, the AI assumed one of my group mates a girl while showing more hair but a boy when showing less.*” Building on these observations, the teacher guided the class toward considering the broader ethical responsibilities tied to the design of AI systems. He remarked,

But then I think there's all that ethical questions of design, of like, what happens to your design when you're done with it, or you introduce it to other people.



Figure 4.11: A screenshot of the Wordcloud interface, which includes a single instance of the word ‘wife’ among other classification outputs for the image on the right.

Right? Are you officially done with that design? Do you have a responsibility with your design? How do we think about what it’s made from?

This question invited students to consider not just how AI systems reflect biases but also their roles in AI’s decision making, sparking a deeper dialogue about the interconnectedness of design, ethics, and societal impact. In response, students reflected on the role of human agency in shaping AI systems, emphasizing that AI functions strictly within the parameters of the information and instructions provided by its human creators, with no capacity to act beyond those constraints. As expressed by S24_w, “It’s a computer with a programmed personality it can’t stray from.” This perspective also increased the willingness of many students to shift the responsibility for bias from the model itself to its designers, suggesting that students viewed human designers as playing a role in the outcomes of AI systems. For example, S1_w noted that,

A difference I found [between people and AI] is that people are doing it to be mean and most are doing it on purpose but the AI has someone program it so sometimes its not trying to be bias it was just programmed that way.

By encouraging students to connect AI biases to societal structures and their own experiences,

the teacher modeled an intentional approach to guiding collective inquiry towards ethical and equitable considerations. This approach aligns with Freire’s vision of education as a practice that challenges social hierarchies and opens up possibilities for transformative thinking [118]. Ultimately, these moments in the classroom underscore the potential of educators to empower students to critically examine technology and its broader societal implications, cultivating both epistemological curiosity and a sense of responsibility as future creators and users of technology.

4.5.3 *Making Learning Relevant*

During the classroom sessions, we observed that children engaged with the concept of AI bias through their personal experiences with fairness at school and home. This connection became particularly evident in Session 3, when students designed their Smartlock’s inclusion/exclusion criteria. Most students adopted an overly inclusive policy preferring lower rejection and higher false-positive rates. They articulated this preference in terms of the emotions that rejections might precipitate. S2_w said, *“It is very important that everyone feels included. If a monster got an invitation but the lock said that they couldn’t come in, that would not be fair. The monster would feel sad.”* This understanding was grounded in the kinds of personal and social connections germane to the students’ lives. As S4_w expressed, *“Because maybe your best friend does not have the same features as your other friends. So then your best friend can’t come to the party.”* The Technology teacher reinforced this perspective, describing the discussion of AI bias as a natural extension of the conversations about fairness that students routinely engage in,

The teachers spend a lot of time chatting with students, and it’s often around fairness, you know, who gets to use the turf field, which game gets to be played. I think in connection with AI, it gets even more interesting because they’re really starting to [question] How can I bring equity to a situation? What is my role in this situation? So it feels like such a great time to do that.

In Session 4, students expanded on their understanding of fairness by applying it directly to the design of AI technologies. S19_w framed fairness in AI systems as "If you are a different race or different sexuality (etc.), there should be no difference," while S20_w emphasized the importance of impartiality, noting, "*to stay unbiased in any judgment or sometimes to give equal opportunities to others, no matter their culture, gender, or religion.*" These reflections translated into concrete design ideas for ensuring fairness in AI systems. For example, S25_w explained their design idea as, "I would make it detect how people are on the inside. My AI will not be racist, ageist, sexist, or homophobic." Students also proposed involving a diverse group of people in the AI's development and oversight to ensure fairness and address any unintentional biases. S4_w wrote, "I would make sure that the AI pulled from fair + well known sources. I would also have a diverse team of people who would run it, so if some unfairness gets into the machine learning algorithm we can fix it." A summary of student ideas on designing fair AI can be found in Table 4.1.

Design Ideas	Student Contributors
Ensure AI fairly represents all cultures and identities without stereotypes, assumptions, generalizations, and appropriation.	S2, S6–S12, S14, S15, S19, S21–S23, S25, S27
Provide mechanisms for human oversight, community engagement, and user feedback to refine AI behavior.	S1, S5, S8, S11, S16, S17
Ensure AI relies on credible, fact-checked sources to provide reliable and accurate information.	S3, S13, S17

Table 4.1: Design Ideas for AI System and their Contributors

Building on their discussions about designing fair AI systems, in Session 4, students further explored how these ideas applied to AI-generated content. They compared their own personally meaningful drawings (see Figure 4.12) with AI-generated images for the

prompt "draw an image of children celebrating a festival." By juxtaposing their artwork with AI-generated images, students critically analyzed how AI systems interpreted the same prompt. This comparative exercise sparked important discussions about representational bias, as students examined the extent to which AI-generated images captured the richness and diversity of their own experiences. When the Technology teacher asked students how DALL-E could be improved, they reasoned both about how humans might better interact with such models and how AI systems could be designed to mitigate biased outputs. For example, S18_w noted that "If the question that someone puts in is not specific enough, it would say something like 'You did not put in an exact question, so I am not able to give you an answer.' And if it showed a picture of something specific, it would explain why it made that choice."

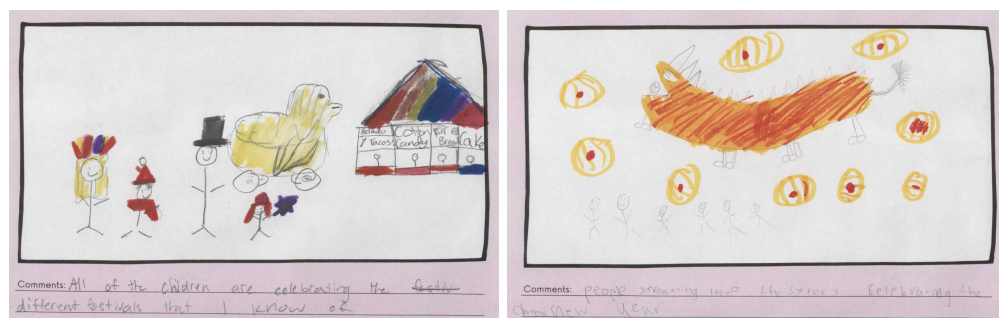


Figure 4.12: Drawings by children for the prompt "Draw a picture of children celebrating a festival" from Session 4.

4.5.4 Reading AI and Reading the World

Throughout the classroom sessions, students reflected on the bidirectional relationship between AI and society—AI is molded by society's values, and society is, in turn, influenced by AI. In examining classification outcomes within the Smartlock interface, students critically assessed how AI systems can lead to systemic exclusion similar to past discriminatory practices. For example, S15_w stated, "There were places where you could only buy a house if you qualified for

certain crazy things. One of them was really crazy—you could only buy a house if you were Aryan.” Similarly, S14_w noted, “Depending on the color pattern, look, etc., the AI will use the descriptions you give and put percents next to it based on how much it thinks it is.” Later, drawing connections between mathematical representations of racism and AI decision-making, S14_w further added, “The fact that one Black person was considered 3/5 of a white person stood out to me a lot. So [AI] could assume traits of [someone] based off of skin color if coded by a racist person.”

Students also drew connections between AI’s treatment of gender and broader societal biases related to gender representation. For example, S25_w observed how classifications in the Wordcloud interface reinforced gender stereotypes, remarking, “When a girl does not show her hair it says boy; when she does, it says girl.’ It is sexist. A girl says she should be a girl, so she is a girl.” This reflection underscores how AI models rely on physical appearance—such as hair length—to classify gender, revealing patterns learned from its training data. S14_w echoed similar concerns noting, “If you don’t look like the AI’s version of a girl, then it didn’t know what to call you.” Furthermore, students recognized that AI’s tendency to focus on visual features to generate probabilities might oversimplify individuals’ experiences, as described by S1_w, “I think that the system labeled people as a ‘girl’ or a ‘boy.’ What if they are non-binary or trans? That wouldn’t make me or them feel good.”

Beyond critiquing AI’s representation biases, students also engaged with broader historical gender inequalities and their lasting societal consequences. Questioning the assumptions that underpin gender roles, S4_w pondered, “There was gender inequality where women could not get an education or go to work. What is the reason for that? Who assumed that men should have more responsibility and be treated nicer?” These inquiries demonstrate students’ critically reflecting about the relationship between past and present societal biases and the technologies that learn from historical and societal data. Student’s exploration of AI bias was further enriched by their school’s DEI curriculum, particularly through their reading of *Stamped*, which provided a foundational understanding of historical and systemic issues. Reflecting on how the themes from *Stamped* intersected with students’ discussions on AI bias,

the Technology Teacher remarked:

I'll bring up the word 'implicit bias' again. For students to be able to use that word in a conversation when talking about AI was directly connected to their conversations that they're having in the other classroom about *Stamped*. And to be able to have a student cross over like that, that's learning. It's really about the connection between all these things. It's about finding, maybe, ingenious ways of reforming something, repurposing something, or building upon the next level of ideas. Because, when do you call it new if you have this whole history behind you that you've had to use to get to this point, right?

4.6 Discussion

My findings underscore the importance of cultivating critical consciousness, defined as the ability to recognize and critique systems of oppression [119], within AI education. By helping students understand the bidirectional nature of AI bias—how societal biases influence AI systems, and how AI can perpetuate or amplify these biases—I argue that learners can be better supported in critically evaluating these technologies and challenging instances of discrimination or injustice.

While immediate feedback from CLIP4KIDS promoted active engagement, children also explored the broader societal implications of AI classification. For example, when students worked together to adjust decision thresholds in the Smartlock interface, their discussions revealed evolving perspectives on fairness and inclusivity, as well as the consequences of designing AI systems that could exclude someone, linking these to historical patterns of exclusion. Prior research underscores that children do not automatically connect examples of AI bias to larger systemic injustices, and that such awareness must be fostered through meaningful dialogue, reflection, and analysis [283]. In the pilot study, students benefited from their school's DEI curriculum, which provided a foundation for examining societal inequities through exposure to texts such as *Stamped*. While this curriculum, on its own, did not extend

to understanding AI classification concepts or biases in AI systems, it shaped the learning outcomes in two important ways. First, students' prior knowledge of stereotypes and historical discrimination helped them draw meaningful parallels between societal and algorithmic biases. This pre-existing understanding offered a lens for critical engagement with the CLIP4KIDS system, enriching their reflections. Second, the DEI curriculum provided students with a shared vocabulary for discussing bias and fairness, fostering collective meaning-making.

While the school's DEI curriculum is aligned with U.S. national standards, I recognize that not all schools have a dedicated DEI program. In the next section, I discuss three implications of these findings, focusing on how these insights can be adapted to diverse educational contexts for teaching AI bias concepts to elementary students.

4.6.1 Implications for Teaching AI Bias

Leveraging Existing Curricula and Learner Experiences

Many educational programs, through subjects like social studies, history, and civics, already address themes of fairness, community, and diverse cultural contributions [42, 39, 344, 267]. Educators can further extend these discussions across disciplines, using subjects like social studies to connect historical and societal inequities to technological challenges [136, 137, 202, 340]. Additionally, fairness and equity are universal concerns for elementary school children, as this is a critical period for developing social and moral reasoning [245, 171, 172, 272]. This developmental readiness can provide a natural bridge for linking lived experiences, such as classroom rules, sports team selections, or shared resources, to concepts of algorithmic fairness and bias. Prior research in AI literacy supports this connection, highlighting how teaching children about fairness in their own lives can extend to understanding bias in the technologies they encounter [268, 185, 61, 283, 76]. Importantly, this sensitivity to fairness equips children to engage with and question systems of power and inequality. By connecting everyday experiences of fairness to discussions of algorithmic fairness, my findings suggest that educators can empower children to critically examine technology's role in reinforcing or

challenging systems of power and inequality.

Fostering Inclusive, Dialogic Learning Environments

AI bias education must also engage with intersectionality, acknowledging that different groups and individuals experience and perceive bias in diverse ways [69, 25, 40]. Creating a safe, inclusive classroom environment where all students feel empowered to share their perspectives is essential [305, 306, 233]. In this study, central to student learning was the dialogical relationship between students and their Technology teacher [120]. Rather than positioning himself as an authority on AI concepts, the Technology teacher encouraged students to share their ideas and lived experiences, question the information presented by the AI system, and engage in collaborative inquiry. By using open-ended questions, linking student contributions, and providing timely scaffolding, he fostered epistemological curiosity and a culture of shared inquiry [313, 117]. I recognize that dialogic inquiry and co-construction of knowledge depend heavily on a teacher's ability to facilitate meaningful discussions, scaffold thinking, and create a collaborative learning environment [6, 293]. While teacher expertise is a key factor in shaping learning outcomes, my findings highlight the importance of respecting and amplifying student concerns and interests. Research on culturally responsive computing [155, 275, 274] and participatory design with children [105, 342] underscores that respecting student concerns not only makes learning experiences more relevant but also empowers students to critically engage with complex and abstract topics. Furthermore, fostering a learning environment that supports student agency allows children to demonstrate confidence in asserting their ideas, reflecting critical pedagogy principles that emphasize the recognition and empowerment of individual voices [118, 313].

Navigating Ethical Considerations and Children's Perceptions of AI

Teaching about AI bias involves navigating complex ethical considerations. While the goal is to help children critically evaluate AI technologies and collaborate effectively with AI [199], educators must remain attentive to potential unintended consequences. One such risk is the

reinforcement of the very stereotypes that AI bias education seeks to dismantle. For instance, presenting biased outputs from AI systems without sufficient context could inadvertently lead to students normalizing discriminatory patterns. To mitigate this risk, building on the dialogical relationship described previously, educators can explicitly frame examples of AI bias as problematic, emphasizing their historical, social, and technical roots. Additionally, presenting examples of biased outputs alongside counterexamples, such as equitable and inclusive AI designs or initiatives to mitigate bias in technology, reinforces the notion that technology can be designed to promote fairness and inclusivity [146, 204, 252]. Class-wide discussions can further enhance learning by allowing students to actively process and critique these examples in a reflective and collaborative manner. This approach is especially important for addressing the perception of AI as an authoritative or neutral entity, as prior research shows that children often view AI outputs as inherently capable and trustworthy [215, 237, 141, 12]. Interactive AI tools that uncover some of these often “black-boxed” AI processes can provide a means to demystify AI by highlighting the roles of training data, algorithmic design, and the role of humans in AI decision-making [99, 43, 44, 158, 193]. By offering hands-on opportunities to experiment with AI, students can see firsthand how biases arise and recognize their own agency in shaping technological outcomes.

4.7 Limitations & Future Work

The primary goal of this pilot study was to assess the feasibility of using CLIP4KIDS to teach children about AI bias in a classroom setting [294]. The themes were developed based on observations of 28 children in two fifth-grade classrooms from a single school in the Pacific Northwest region of the United States. As such, these findings are not intended for statistical generalization [294, 341], but are instead guided by Lincoln and Guba’s emphasis on information richness [194]. This work also took place in a school with an existing DEI curriculum, where students were already familiar with discussions of social justice and equity. In schools without similar institutional support, educators may encounter greater challenges in fostering critical inquiry into topics such as AI bias.

Additionally, the demographic characteristics of the study site likely shaped how students engaged with the curriculum and the themes of AI bias. The participating school serves a racially and linguistically diverse student population, with 57% of students identifying as students of color and nearly half being multilingual. These characteristics may have influenced how students perceived and interpreted issues of classification, fairness, and representation. Students from marginalized or multicultural backgrounds may bring distinct perspectives to questions of inclusion and bias, which may not be as readily surfaced in more homogeneous school contexts [20]. While this diversity enriched classroom discussions and offered insight into how young learners critically engage with AI systems, these findings may not directly generalize to all educational settings.

While this work offers transferable insights [82] that can inform other educational contexts, future work could examine these approaches at a larger scale across diverse settings and explore their longer-term impact on students' understanding of AI bias. Conducting the study within a classroom environment also meant that teacher prompts may have guided students toward particular lines of reasoning, and the group setting may have introduced elements of social compliance. At the same time, these conditions supported key aspects of the approach, including amplifying student voices, co-constructing knowledge, and fostering inquiry through dialogue [313, 118]. Future work could further examine how teachers' prior knowledge, values, and experiences shape their pedagogical practices when teaching AI bias, and how these factors influence classroom discussions.

Finally, to address classroom safety and suitability concerns, I used pre-generated images from DALL·E to support students in exploring bias in generative AI. Although DALL·E's proprietary nature introduced limitations, such as limited control over model updates, it also served as a relevant example of the AI tools increasingly present in children's lives [138, 287, 5]. I initially developed an interface that allowed students to interact with an open-source image generation model⁵ and extensively tested it for safety and moderation of

⁵<https://stability.ai/>

harmful outputs. However, I ultimately decided that the risk of a single inappropriate or harmful image was too great to allow direct interaction [334, 124]. While this methodological choice limited real-time engagement, it preserved the integrity of the learning experience by providing authentic AI-generated outputs for discussion [284]. Future work could explore how classroom tools for teaching generative AI can better balance safety, accessibility, and opportunities for direct interaction with emerging technologies.

4.8 Summary of Contributions

This work introduces *CLIP4KIDS*, an interactive AI literacy system that enables children to investigate how AI systems classify, categorize, and make decisions. The design of *CLIP4KIDS* was motivated by the first, second, and third thesis statements of this dissertation, which propose that children benefit from opportunities to engage in iterative experimentation with AI systems (T1), compare their own reasoning with that of an AI model (T2), and connect AI learning to their own identities and lived experiences (T3). Through a classroom study with 28 fifth-grade students, I found that children engaged in iterative experimentation by adjusting classification parameters, testing predictions, and observing how these changes affected AI outputs. As they compared their own categorizations with the system’s classifications, children made their reasoning explicit and developed explanations for why the AI produced particular results. These moments became especially meaningful when the system made assumptions about identity based on characteristics such as hair length or clothing, prompting children to draw on their own experiences and discuss broader questions of fairness, bias, inclusion, and representation. My findings show that children were able to reason about this bidirectional relationship between AI and society, drawing parallels between societal stereotypes and contemporary concerns about representation and bias in AI systems.

Chapter 5

ECOPROMPT: SUPPORTING CHILDREN’S REASONING ABOUT AI’S ENVIRONMENTAL COSTS¹

5.1 Preface

The first two studies in this dissertation examined how children engage with AI across its epistemic and social dimensions. This chapter turns to the material dimension of AI, focusing on its environmental costs. Mills et al.’s AI literacy framework [217], that guides this dissertation, speaks to this dimension through both of its core values. *Centering Human Judgment* extends to AI’s environmental considerations, asserting that the uncritical deployment of large-scale AI systems carries significant energy costs that human judgment must take into account [217]. *Centering Justice* is equally relevant as AI technologies are not neutral in their material impacts, and the environmental costs of large-scale AI infrastructure such as water use and carbon emissions are not distributed equally across communities [217]. This chapter therefore focuses on the AI literacy practice of *Ethics and Impact*, which invites learners to consider not only the benefits and harms of AI for individuals and communities, but also its implications for the natural world [217].

5.2 Introduction

Two of the most socially consequential issues facing today’s children are the rise of AI and the rapid changes to the earth’s climate. Both issues are complex and contested, and they are linked through the notable environmental costs of AI use. Recent estimates help illustrate

¹This chapter is based on collaborative work previously published as: Aayushi Dangol, Robert Wolfe*, Nisha Devasia*, Mitsuka Kiyohara, Jason Yip, and Julie A. Kientz. 2026. Where Does AI Leave a Footprint? Children’s Reasoning About AI’s Environmental Costs. In Proceedings of the 25th Interaction Design and Children Conference (IDC ’26), June 22–25, 2026, Brighton, United Kingdom. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3773077.3806144>

the scale of these costs. For example, Jegham et al. [156] report that the 700 million daily queries made to the GPT-4o model alone required “electricity use comparable to 35,000 U.S. homes, freshwater evaporation matching the annual drinking needs of 1.2 million people, and carbon emissions requiring a Chicago-sized forest to offset.”

Despite the accumulating evidence of generative AI’s environmental impact, these costs remain largely invisible to end users. At the same time, generative AI is increasingly embedded in the tools children use [75, 345, 336], bringing a growing population of young users into routine contact with AI systems without clear information or opportunities to engage with those impacts. This invisibility is especially consequential given that prior work shows that children are already attentive to climate change and environmental sustainability, often expressing concern about environmental futures and a desire to act responsibly [41, 335, 183]. Yet far less is known about how children perceive, interpret, and reason about the environmental impacts of AI systems. This gap motivates the two research questions guiding this chapter:

RQ1: How do children (ages 6 - 12) understand and reason about the environmental costs of generative AI systems?

RQ2: How might interactive systems help make the ecological footprint of generative AI visible, relatable, and actionable for children?

Recognizing that children are often drawn to games and simulations as ways of exploring complex ideas [243, 88], I developed EcoPrompt, a web-based system that invites children to explore how everyday interactions with generative AI are connected to underlying environmental costs. By examining how children engage with and make sense of these connections through EcoPrompt, this work contributes both empirical and design insights into supporting material AI literacy.

Specifically, I document how children reason about AI’s resource use, including how they interpret the relationship between everyday interactions and larger environmental impacts, and how these understandings deepen through interaction with EcoPrompt. I then offer design

implications for making the material dimension of AI visible and relatable for young learners, showing how interactive systems can connect abstract infrastructural processes to children’s existing values around sustainability and shared responsibility. Together, these findings support T1 and T3, demonstrating how systems that scaffold iterative experimentation and connect learning to children’s lived experiences can foster critical and reflective engagement with the material dimension of AI.

5.3 *EcoPrompt: System Design & Development*

EcoPrompt is an interactive system designed to support children in reasoning about the environmental impacts of generative AI. Children interact with EcoPrompt through two complementary but separate interfaces: (1) a Footprint Calculator, which provides real-time estimates of energy, water, and carbon use associated with children’s generative AI prompts (see Figure 5.1); and (2) a Farm Game, in which children manage virtual farms and can choose whether and when to use generative AI to support their farming activities (see Figure 5.2).

While the Footprint Calculator foregrounds prompt-level resource use and individual decision-making, the Farm Game supports reflection on collective dynamics, accumulation, and longer-term environmental consequences. Together, these interfaces broaden children’s inquiry from “*What resources does this AI interaction use?*” to “*How do patterns of AI use interact with shared ecological systems?*” Below, I describe the design considerations that guided EcoPrompt’s development, followed by an overview of its two core components.

5.3.1 Design Considerations

Making generative AI’s Environmental Impacts Visible through Prompt-Level Feedback

Prior work in systems thinking and AI literacy emphasizes that learners often struggle to reason about complex systems when key processes remain invisible or abstract [131, 80, 324]. EcoPrompt addresses this challenge by making the environmental impacts of generative AI

The interface is divided into two main columns. The left column contains a prompt entry box (1), an 'Ask AI' button (2), and a large text area for the AI response (3). The right column contains a 'Last Prompt's Cost' summary (4) and a 'Total Session Cost' section (5) with three sub-sections: Water Used, Carbon Footprint, and Energy Consumed. Each sub-section in the total cost section includes a status bar, a numerical value, a unit, and a 'Set Your Limit' input field.

1 Can you tell me about what stars in 500 words?

2 Ask AI

3 Imagine our sun, but way, way farther away! That's what stars are! Stars are giant balls of super-hot, glowing gas, mostly hydrogen and helium. They're like huge ovens in space! Inside, they're so hot that hydrogen atoms smash together and turn into helium. This smashing creates a LOT of energy, which comes out as light and heat, making the star shine brightly. Think of it like a giant bonfire burning far, far away. The bigger the bonfire, the brighter it is, and the hotter it feels. Same with stars! Some stars are super big and bright, much bigger than our sun, while others are smaller and dimmer. Stars come in different colors too! The hottest stars are blue, then white, then yellow like our sun, then orange, and finally red, which are the coolest stars. Stars are born in giant clouds of dust and gas called nebulae. Imagine a big, messy cloud in space! Gravity pulls all the dust and gas together, and it starts to spin and get hotter. Eventually, it gets so hot that the hydrogen atoms start smashing together, and a star is born! Stars live for different amounts of time. Small stars can live for billions of years, while big, bright stars burn through their fuel much faster and only live for millions of years. When a star runs out of fuel, it dies. What happens next depends on how big the star is. Small stars like our sun will eventually turn into a white dwarf, which is like a glowing ember that slowly fades away. Big stars can explode in a huge supernova, which is a super bright explosion that can be seen from across the universe! Sometimes, after a supernova, what's left becomes a black hole, which is so strong that nothing, not even light, can escape its pull. Stars are what make the night sky so beautiful! They twinkle because their light has to travel through Earth's atmosphere, which is like looking through wobbly water. Without stars, there would be no light, no heat, and no life on Earth! They're super important and amazing! So next time you look up at the night sky, remember that you're looking at giant, glowing balls of gas, burning far, far away and creating the light that makes our universe so amazing.

4 Last Prompt's Cost

- Water Used: ~11 drops of water
- Carbon Cost: ~0.02 balloons of CO₂
- Energy Used: powering an LED for ~7.7 minutes

5 Total Session Cost

Water Used

mi stands for milliliters. A small water bottle holds about 500 ml.

150 ml

Set Your Limit (i)

150 ml

That's like... 0.0 glasses of water

Carbon Footprint

0.133 / 20 gCO₂e

Set Your Limit (i)

20 gCO₂e

That's like... 0.0 balloons of CO₂

Energy Consumed

0.888 / 5 Wh

Set Your Limit (i)

5 Wh

That's like... powering an LED light for 9 minutes

Figure 5.1: Footprint Calculator interface, including (1) user prompt entry box, (2) an "Ask AI" button that submits the prompt to the model, (3) an output box in which the AI's response is displayed, (4) a display of the last prompt's estimated water, carbon, and energy costs, and (5) a display of the total water, carbon, and energy costs across the session, with status-bar style indicators of cumulative use.

visible at the moment of interaction through the Footprint Calculator. By providing real-time estimates of energy, water, and carbon use associated with children's generative AI prompts, the system foregrounds otherwise hidden infrastructural costs of AI use. This immediate feedback supports children in attending to specific system elements and their relationships, encouraging reflection on how different prompts are associated with different environmental implications.



Figure 5.2: Farm game interface showing (1) initial welcome screen, (2-4) gameplay across three levels. In each level, players face various challenges. For example, in (4), pest attacks introduce critical decision points where players can use AI to make crop protection more manageable as levels progress.

Scaffolding Systems Thinking across Individual and Collective Perspectives

Understanding environmental systems requires attention not only to individual actions, but also to how effects accumulate over time within systems shaped by multiple actors [29, 58]. For children, this means relating their own generative AI use not only to shared environmental resources, but also to the actions and constraints of others operating within the same system. While the Footprint Calculator foregrounds individual prompt-level decision-making, the Farm Game situates generative AI use within a shared environment where outcomes emerge from the interplay of many actors, not all of whom are under the child’s control. This shifts attention away from immediate, isolated cause-and-effect relationships toward longer-term and collective consequences that unfold even when children make careful or restrained choices. In doing so, the Farm Game supports reasoning about responsibility and impact in contexts where agency is distributed and outcomes are uncertain. By separating these perspectives across complementary interfaces, EcoPrompt allows children to engage with

both the individual and collective dynamics of generative AI use without requiring them to reason about all aspects simultaneously.

Encouraging Deliberation and Trade-Off Reasoning in Decisions about generative AI Use

Prior work in HCI has characterized decision-making around generative AI use as a wicked problem, emphasizing that such decisions involve competing values and no single correct solution [64, 225]. Therefore, EcoPrompt is designed to foreground reflection, and trade-off reasoning across multiple levels of decision-making. At the prompt level, the Footprint Calculator supports deliberation as children can set limits on their generative AI use and receive feedback when those limits are approached or exceeded, creating opportunities to pause, reconsider, and renegotiate their choices. The Farm Game extends this deliberative framing to a broader, collective context. Within the game, children are not required to use generative AI, but can choose whether and when it is appropriate to do so in support of shared goals, positioning generative AI as one option among many within a dynamic system. Together, these mechanisms promote ongoing deliberation rather than one-time judgments, supporting children in reasoning about generative AI use as a situated and evolving decision shaped by both individual actions and collective outcomes.

Choice to Use a Real Generative AI API

I faced the difficult choice of whether to create a system that 1) connected to a real generative AI system, dynamically displaying outputs to the player, or 2) attempted to mimic the output of such a system without actually calling the API. The former option would provide children with a more faithful and reliable experience, affording the opportunity to interact directly with a generative model while reasoning about its environmental impact, while the latter would avoid the environmental costs associated with using the model. I ultimately chose the former approach, as I wanted children to have the opportunity to experiment with the model and form ideas about how it worked, and I did not believe I could sufficiently replicate this with a more limited system. While this approach did yield the intended results

(see for example Section 5.1.3, where children discover that they cannot control the output verbosity of a real generative model and thus easily reduce its environmental footprint), I also acknowledge that the use of a generative model over the two sessions likely resulted in the production of 1-2 kilograms of CO₂. I consider this a limitation in my approach, but a necessary one for supporting the ecological validity of the study.

5.3.2 System Overview

Footprint Calculator

The Footprint Calculator begins with a short interactive tutorial that introduces children to the lifecycle of an AI query. The tutorial (see Figure 5.3) depicts how a user’s prompt is transmitted from a personal device to remote data centers, where large-scale computing infrastructure consumes electricity that is often generated from carbon-emitting sources. It then illustrates how this energy use produces heat that must be dissipated through water-intensive cooling processes, with downstream effects on local ecosystems and water availability.

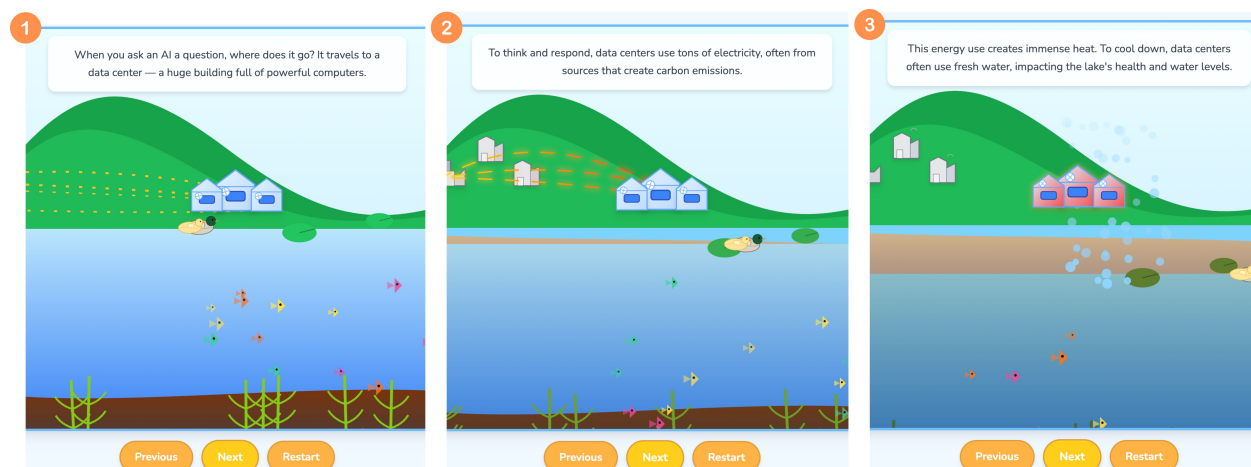


Figure 5.3: Tutorial illustrating the AI query lifecycle and its environmental consequences in EcoPrompt, demonstrating how a user prompt travels from a personal device to a data center that consumes electricity, resulting in carbon emissions.

This walkthrough establishes a conceptual baseline that helps children interpret the calculator’s core indicators (energy, water, and carbon) by making visible how these resources become implicated in a single AI interaction. The calculator presents these indicators as modeled estimates rather than precise measurements, intended to support comparison, reflection, and reasoning about relative environmental impacts rather than exact quantification.

To estimate the environmental footprint associated with each prompt, I used the infrastructure-aware inference benchmarking framework introduced by Jegham et al. [156]. In this framework, per-query environmental costs are derived by first estimating the electricity consumed during inference based on runtime characteristics (e.g., latency and generation speed), the power draw of GPU and non-GPU system components, and datacenter-level overhead captured through Power Usage Effectiveness (PUE). This per-query energy estimate is then translated into water consumption and carbon emissions using standard environmental multipliers: Water Usage Effectiveness (WUE), which accounts for water used in both on-site cooling and off-site electricity generation, and Carbon Intensity Factors (CIF), which reflect the carbon content of the regional electricity grid. This approach focuses on operational impacts during inference rather than training or hardware manufacturing, allowing environmental costs to be attributed at the level of individual prompts.

While these calculations are grounded in established modeling approaches, they necessarily simplify complex and variable infrastructures; as such, the resulting values should be interpreted as approximations that make otherwise invisible processes legible to learners. The Footprint Calculator also allows children to set adjustable resource limits; visual indicators fill as these limits are approached. Children can revise these limits dynamically, supporting exploration of trade-offs rather than compliance with fixed thresholds.

Farm Game

The Farm Game is a grid-based farming simulation organized into five levels that progressively introduce new mechanics, challenges, and AI affordances. Players take on the role of a farmer managing crops on a personal farm while drawing from a shared community lake that

supplies water and energy to all farms in the region. Across levels, players engage in routine agricultural activities such as planting, watering, harvesting, (see Figure 5.4) and responding to disruptions, while being periodically offered AI-powered assistance.

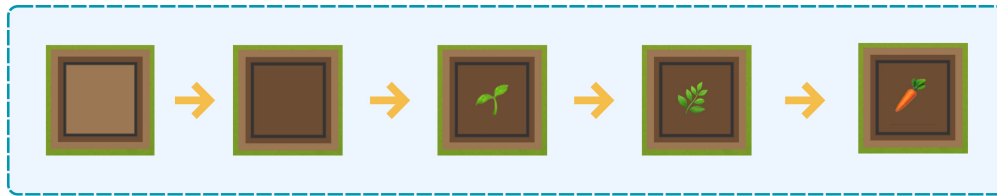


Figure 5.4: An illustration of crops growing as a result of players’ attention to planting, watering, and harvesting them.

In all levels, the farm is represented as a fixed grid of tiles that players interact with using a toolbar of tools and seeds (see Figure 5.5). Crops grow over time and contribute to experience points, which unlock new goals and features. Although players cannot see or directly interact with other farmers, the lake’s health reflects the cumulative effects of AI-related actions taken across the system. If the lake’s health reaches zero, players lose the game. Conversely, players who manage their farms successfully can earn more coins and finish the game with the highest score. The five levels are summarized below:

- Level 1: Introduces the basic farming loop and the shared-resource framing. Players learn to manage the farm grid and are introduced to the community lake through a narrative tutorial.
- Level 2: Introduces seasonal constraints and the Farmer’s Almanac, which provides guidance about crops and growing conditions (see Figure 5.6). Players gain access to an AI farm hand (implemented using Google’s Gemini model) that can answer questions about seasons and crops. AI interactions are optional and are always preceded by a warning indicating their impact on the lake.



Figure 5.5: An illustration of the game panel at the start of the first level, showing (1) the sidebar displaying the lake health with a status log communicating changes to the player, (2) the main array of tools available to the player, including the planting and watering tools, and (3) the inventory and notifications panel.

- Level 3: Introduces pest infestations that disrupt crop growth (see Figure 5.7). Players can remove pests manually through a whack-a-mole-style minigame, craft pesticide using inventory items, or use AI-assisted pest control, which resolves the infestation more quickly at the cost of reducing the shared lake's health.
- Level 4: Introduces birds that reduce crop yield unless mitigated by a scarecrow. Players can create a scarecrow manually or via an AI image generator, with AI-generated content reducing the health of the shared lake.
- Level 5: Introduces a market in which players sell their harvested crops and earn coins. Players set prices based on reports of crop sales from the previous week and may optionally use AI-assisted pricing to inform their decisions.

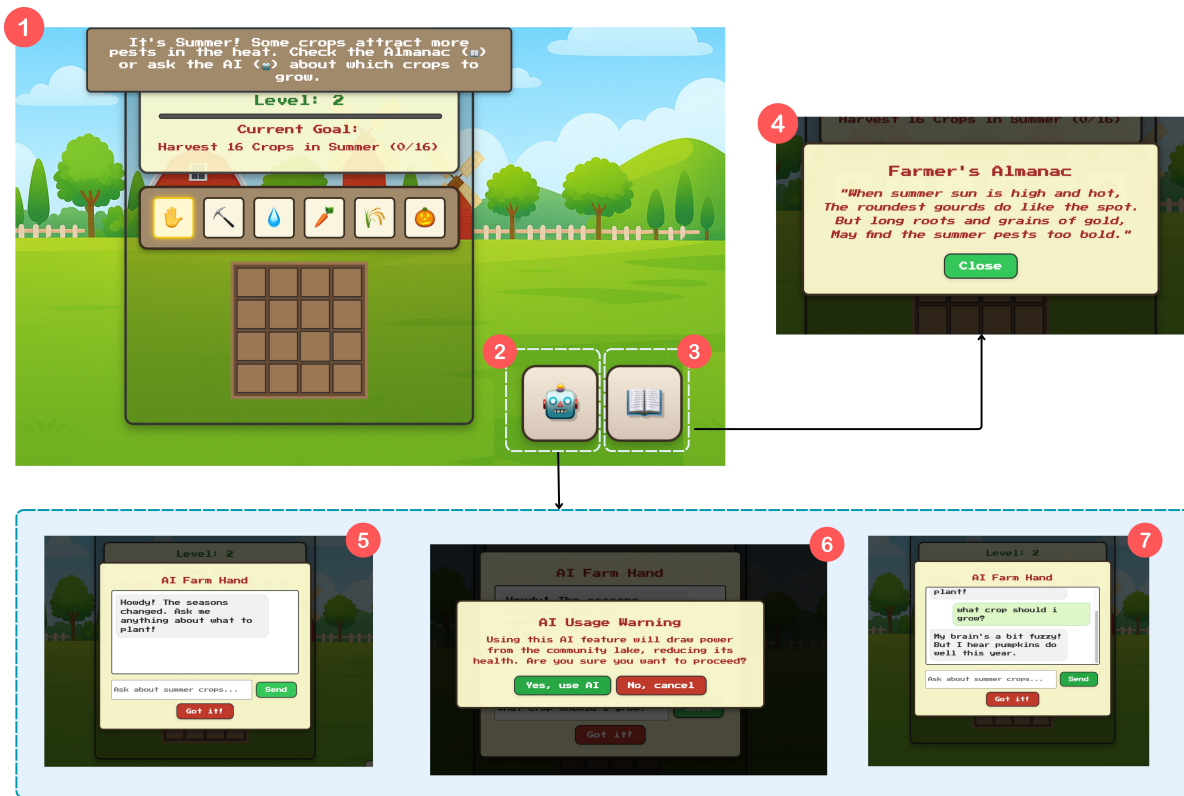


Figure 5.6: The interface for level 2 of the Farm Game, showing (1) the overall UI, (2) the AI Farm Hand button, (3) the Farmer’s Almanac button, (4) the Farmer’s Almanac hint message displayed after pressing the button, (5) the AI Farm Hand interface, (6) an AI Usage Warning displayed before sending the user’s message to the AI Farm Hand, and (7) the response generated by the AI Farm Hand, suggesting what to plant. Levels 3 and 4 retain the Farm Hand and Almanac buttons.

5.4 Methods

5.4.1 Participants

I led a team of five collaborators in conducting this study. We worked with **KidsTeam UW**, an intergenerational co-design group of adult design researchers (researchers, graduate and undergraduate assistants) and 16 child participants, ages 6 to 12. All child participant names reported here are pseudonyms. Child participants reported varying levels of prior experience with AI. Seven children reported engaging with AI daily, while two reported no prior AI use.



Figure 5.7: An illustration of the pest control scenario encountered in Level 3, showing (1) the animated pests the player must remove from the crops, (2) the pest removal minigame triggered by clicking on the pest sprite, and (3) the pest control interaction with the AI Farm Hand, which can remove a pest at the cost of some of the player's natural resources.

The most common types of AI use included voice assistants, video game AIs, and chatbots, with many children using multiple AI types and engaging with AI both independently and with friends or family. Child participants entered the study with varying initial beliefs about whether generative AI uses resources and whether AI affects the environment. These self-reported perspectives are summarized in Table 5.1, along with demographic information and AI usage details of the children. The university's Institutional Review Board approved all research conducted and parental consent and child assent were obtained for all child participants.

5.4.2 Design Sessions

We conducted two design sessions with KidsTeam UW as part of a week-long summer camp organized at our university. Each session began with Snack Time (15 minutes), which created informal space for children and adult facilitators to build rapport. This was followed by Circle Time (15 minutes), a whole-group warm-up in which adult facilitators posed a "Question of the Day" to surface children's prior ideas and prime reflection before engaging with the EcoPrompt system. During Design Time (45 minutes), children and adults worked together in small intergenerational groups of four to five children and two adult facilitators,

Table 5.1: Reported Child Participant Details (all names are pseudonyms). AI Use Context codes: F=Use with friends or family; S=Self.

Name	Age	Gender	AI Type	AI Use Context(s)	Use Frequency	AI Uses Re-sources?	AI Affects Environment?
Lina	6	Girl	Voice Assistants, Video game AI	F	Few times/week	Yes	Maybe a little
Nadia	9	Girl	Voice Assistants	F, S	Daily	Yes	Yes a lot
Ravi	7	Boy	Voice Assistants, Video game AI	F, S	Daily	Not sure	Not sure
Leo	8	Boy	Voice Assistants, Video game AI	F, S	Daily	Yes	Maybe a little
Arjun	10	Boy	Voice Assistants, Chatbots	F, S	Daily	Yes	Maybe a little
Miles	11	Boy	I don't use AI	-	-	Yes	Yes a lot
Ethan	12	Boy	Chatbots, Video game AI	S	Once a week	Not sure	Yes a lot
Minh	8	Boy	Chatbots, Video game AI, Voice Assistants	F, S	Daily	Yes	Maybe a little
Sofia	10	Girl	I don't use AI	-	-	Yes	Maybe a little
Willow	6	Girl	Voice Assistants	F	Once a week	Yes	Yes a lot
Noah	6	Boy	Chatbots	S	Few times/week	Not sure	Maybe a little
Ava	10	Girl	Voice Assistants	S	Few times/month	Yes	Yes a lot
Harper	8	Girl	Chatbots	S	Few times/week	Yes	Not sure
Ella	6	Girl	Voice Assistants	S	Few times/week	Yes	Not sure
Joon	9	Boy	Chatbots, Video game AI, Voice Assistants	F, S	Daily	Yes	Yes a lot
Chloe	10	Girl	Chatbots, Video game AI, Voice Assistants	F, S	Daily	Yes	Maybe a little

collaboratively exploring ideas and design possibilities. Sessions concluded with Discussion Time (15 minutes), during which groups reflected collectively on their experiences, decisions, and emerging understandings.

Across both sessions, we enacted the four dimensions of equal and equitable design partnerships articulated by Yip et al. [342]. Facilitators were trained to support balanced participation by actively inviting contributions from all children, attending to power dynamics within groups, and minimizing the influence of dominant voices. Rather than directing outcomes, adult facilitators focused on scaffolding dialogue, encouraging children to build on one another's ideas, and supporting collective sensemaking [342]. This facilitation approach

aimed to create an environment in which children felt comfortable expressing uncertainty, disagreement, and novel perspectives.

Session 1

We began the session with a warm-up question “What gives you energy when you feel tired? Can humans run out of energy like phones or computers do?” to surface children’s initial ideas about energy and resource use. Children then completed an individual drawing activity in which they represented how they believed generative AI gets energy and what resources are required to power it. This activity externalized children’s existing mental models and served as a point of reference for later reflection. Earlier in the week, as part of the broader summer camp, children had already interacted with generative AI by requesting and observing AI-generated images and videos. Given this prior exposure, we chose to build upon their existing experiences rather than reintroduce generative AI concepts. Following the drawing activity, children were divided into four groups and introduced to the EcoPrompt system. They began by watching an interactive tutorial that illustrated the lifecycle of a generative AI query, which situated AI systems as embedded within broader ecological infrastructures. Children then interacted with the Footprint Calculator that surfaced real-time estimates of energy, water, and carbon use associated with their generative AI interactions. The session concluded with a 15-minute group discussion, where children shared reflections on what surprised them about generative AI’s resource use, compared their initial drawings with what they learned through the Ecoprompt system, and shared their reasoning about when and why generative AI use might be useful.

Session 2

As in Session 1, we began with a warm-up question, asking, “*If one person uses a lot of AI, does it affect other people?*” This conversation encouraged children to consider how AI use might have effects that extend beyond a single user. Children were then introduced to the second component of the EcoPrompt system, where children worked in small groups to

manage virtual farms and decided when to use AI tools to support their farming activities. Adult facilitators supported children’s exploration by posing reflective prompts throughout gameplay, such as “*What do you notice happening to the lake when everyone uses AI a lot?*” and “*What do you think would happen if one farm stopped using AI for a while?*” The session concluded with a whole-group discussion in which children shared their experiences, reflected on different strategies for AI use, and discussed how individual decisions contributed to collective environmental outcomes.

5.4.3 Data Collection

For both sessions, we used the built-in webcams on desktop computers to capture synchronized video and screen recordings via Zoom, a video-conferencing platform (see Fig:5.8). This setup resulted in four camera recordings per session and a total of 427 minutes of video data. In addition, we photographed physical artifacts produced during the sessions, including children’s drawings and handwritten notes. Facilitators also took field notes throughout each session, documenting key observations, interactions, and notable moments.



Figure 5.8: Child participants in the KidsTeam UW co-design group play the Farm Game.

5.4.4 Data Analysis

We employed an inductive qualitative approach for data analysis [17]. The coding team, consisting of the first, second, and third authors, began by creating analytical memos for all

the videos [28, 265]. As part of this process, one author served as the primary reviewer, while another served as the secondary reviewer. The primary reviewer first watched the assigned recordings and created narrative summaries at five-minute intervals, documenting children’s interactions with EcoPrompt, their on-camera reactions, and verbal and nonverbal exchanges among participants, including direct quotes relevant to the research questions. The secondary reviewer then independently reviewed the same recordings to check the completeness and clarity of the memos and to contribute additional insights. This process supported analytic rigor by surfacing multiple perspectives during early sense-making of the data.

After creating and reviewing the analytic memos, the primary and secondary reviewers engaged in open coding, proposing initial codes such as “Mental Models of AI Energy Use” and “Judgments about AI Use”. They then met over three meeting sessions to compare, reconcile, and refine the emerging codes. During these meetings, they examined representative excerpts and counter-examples, clarified code boundaries, and merged overlapping codes. For example, several initial codes describing children’s experimentation with longer prompts and harder questions were consolidated into a single code, “Perceived Prompt–Cost Relationship”, which captured children’s reasoning that increased prompt complexity leads to greater resource consumption.

This iterative process resulted in a final codebook organized around four main code categories: (1) Mental Models of AI and Energy Infrastructures, (2) Cost Attribution, (3) Value Judgments about AI Use and (4) Interaction Strategies. I then applied the refined codes across the full dataset, and the second author conducted a subsequent pass. Interrater reliability was assessed through qualitative negotiations, where both authors met to discuss and resolve any coding disagreements [212]. Finally, we organized codes into higher-level themes through two rounds of collaborative refinement and discussion. After finalizing the themes, I revisited the full dataset to extract representative excerpts, ensuring that each theme was well-supported by the data.

5.5 Findings

5.5.1 Children's Evolving Understanding of Generative AI's Resource Consumption

Generative AI in Everyday Devices and Electricity Use

At the beginning of Session 1, children reasoned about generative AI's resource consumption by drawing a distinction between the material infrastructures that power AI-enabled devices and the computational processes they associated with generative AI itself. When reasoning at a material level, children described generative AI as embedded within the same electrical systems that power homes and cities. These explanations often traced energy from natural resources through power plants and transmission lines before reaching generative AI-enabled devices.

For example, Chloe (age 10) explained, “*The wind turns the turbine, and the sun feeds the solar panel to make energy. Energy then goes from power plants, to a power line, to a building, to a plug, to a wire, to your computer, and that computer is or has the AI*” (see Figure 5.9, image (1)). Across such accounts, children located generative AI within a device and reasoned that powering the device effectively powered the AI. As Willow (age 6) summarized, “*The power lines connect to wherever you are using AI from. Here AI is on the TV*” (see Figure 5.9, image (2)).

At the same time, when reasoning at a computational level, several children described generative AI as operating primarily on information provided by people rather than on electricity. For example, Minh (age 8) explained that “*AI gets information as people type it on there,*” while separately describing how devices that contain generative AI are powered through “*solar panels and wind turbines.*” Similarly, Sofia (age 10) shared, “*The energy powers computers and other devices. Somebody made AI and taught AI how to work. From there I think AI works on its own.*” In these accounts, electricity enabled the device to function, while generative AI was framed primarily as an information-processing system. Some children extended this distinction further by questioning whether generative AI itself requires energy

at all. For example, Arjun explained, “*There are 4 ways to create electricity for your device to use AI. But I don’t think AI itself needs it. Because the energy charges the device not the AI, so it really depends on the device.*” Together, these accounts emphasize a distinction between the material operation of generative AI-enabled devices and the computational processes children associated with generative AI itself.

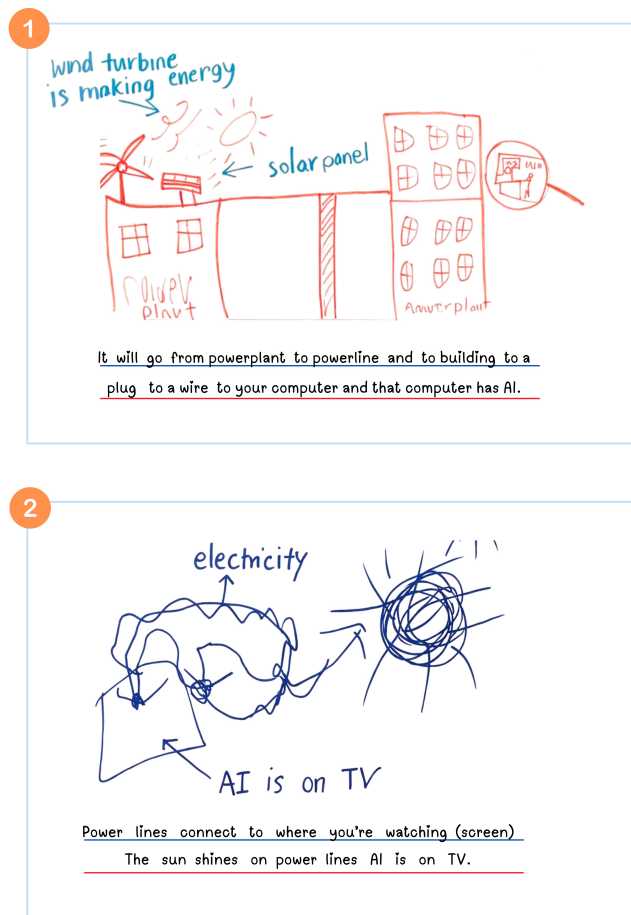


Figure 5.9: Children’s illustrations of their understanding of how AI consumes energy. Illustration (1) traces energy consumption due to local AI use to a powerplant and then to renewable energy sources. Illustration (2) traces AI use to power lines fueled by solar energy. To improve readability, the text shown below each drawing reflects the child’s own written explanation, digitized by the research team while preserving the child’s original wording.

Data Centers and Cooling

Through watching the tutorial as part of the Footprint Calculator in Session 1, children developed an emerging understanding of generative AI's energy use that extended beyond device-level electricity consumption. For example, while discussing the tutorial with Chloe (age 10) and Minh (age 8), Sofia (age 10) explained that *“when someone asks AI a question, it goes to the data center, where it gets the data and this usually involves carbon...carbon is something bad that goes to the Earth and damages the ozone.”* Chloe similarly focused on the tutorial's depiction of heat and cooling, adding, *“Because data centers use so much energy, it's so hot that they use water to cool down, and it makes the lake dirty and less water.”* These explanations reflect children's emerging systems-level account that connects generative AI use to geographically distant infrastructure and downstream environmental consequences.

Building on this recognition of data centers as sites where computation produces heat, children further reasoned about how that heat might be managed through cooling processes. In a different group consisting of Joon (age 9), Nadia (age 9) and Ravi (age 7), when asked how data centers cool down, Joon explained, *“Heat makes water turn into mist, and that acts as a cloud that makes rain. So that produces more water.”* While the latter part of this explanation overextended natural water-cycle processes, Joon's account correctly identified evaporation as a mechanism for cooling. When the facilitator reiterated that the tutorial showed water being used to cool overheated servers, Joon followed up by asking, *“But if it heats up, I mean naturally, it would turn into mist, right?”* Rather than abandoning his model, Joon sought to reconcile the new information with his existing understanding of heat and evaporation. Nadia then built on this reasoning by drawing on embodied experience, explaining, *“It's like when you're really hot, water can cool you down. So that's what [the water] is doing to the big data centers.”*

Prompt Complexity, Output Verbosity & Perceived Resource Use

During Session 1, children further refined their understanding of generative AI's environmental impact by observing changes in energy, carbon, and water indicators within the Footprint Calculator. For example, after watching the tutorial, Leo (age 8), Miles (age 11), Arjun (age 10), and Ethan (age 12), prompted the AI with, "*Do you need water?*," which the calculator estimated to cost "*2 drops of water.*" Leo responded, "*Two drops of water is a lot,*" echoed by Miles: "*Two drops of water, how could you?*" These reactions suggest that the group treated the resource indicators as meaningful measures of cost. Leo then attempted to directly constrain the AI's resource use by prompting, "*Use only 0 drops of water.*" When the AI's response still consumed "*1 drop of water,*" Leo expressed surprise, "*What?! She used one drop of water when I said only use zero.*" This moment revealed an emerging understanding that generative AI's energy use could not be fully controlled through explicit instruction.

The group then experimented with different prompts to test how features of the AI's responses related to water consumption. In doing so, they repeatedly varied the structure of their questions and compared the resulting resource estimates, converging on output verbosity as a key factor. For example, when Miles asked an algebra question, the AI's response consumed "*3 drops of water, 0.01 balloons of CO₂, and powering an LED for 2.3 minutes.*" Reacting to the length of the response, Miles revised the prompt, asking the AI to "*respond in one word.*" In response, the AI produced a much shorter output that required only "*1 drop of water, 0.00 balloons of CO₂, and powering an LED for 0.5 minutes.*" The group treated this contrast as evidence that reducing output verbosity lowered water use. Their reactions to the AI's narrative-style responses further reinforced this association. As Arjun read aloud an answer that began with "*Imagine...*," he interrupted, exclaiming, "*You don't have to keep saying imagine a story!*" When subsequent prompts continued to elicit verbose framing, substituting "*Let's pretend*" for "*imagine,*" the group laughed but also expressed annoyance, framing generative AI's verbosity as both undesirable and environmentally costly.

In contrast, a different group consisting of Sofia, Willow, Minh, and Chloe used the

Footprint Calculator to explore a related but distinct hypothesis about generative AI's environmental costs, attributing higher resource use to the difficulty of questions asked. They began with what they characterized as a simple prompt *"Do you like dogs?"*, predicting that it would cost very little water. When the calculator reported a cost of "1 drop of water, 0.00 balloons of CO₂, and powering an LED for 0.8 minutes," the group treated this result as confirmation of their expectation. Building on this observation, they reasoned about which kinds of questions might be more expensive, proposing that prompts requiring factual knowledge such as *"what's the heaviest thing on Earth?"* would consume more resources. As the discussion continued, children increasingly linked prompt complexity to explanatory depth, suggesting that questions involving *"a lot of writing,"* such as essays, would be costly. For example, when Minh proposed asking *"what happened before the Big Bang,"* the group agreed this would be especially costly *"because the answer is really long."* While this group also associated longer responses with higher environmental cost, they framed response length as an inevitable consequence of asking harder questions, rather than as a stylistic inefficiency that the AI could avoid.

5.5.2 Children's Reasoning About Value, Agency, and Responsibility in Generative AI Use Using Numeric Limits to Reason About Whether AI Use Was "Worth It"

During Session 1, children also set numeric upper bounds for acceptable water, carbon, and energy use while interacting with the Footprint Calculator. Their initial limits reflected moral judgments about different resources. For example, in one group consisting of Minh, Sofia, and Chloe, the group set a water limit of 100 mL, explaining that this amount was *"not that much."* In contrast, they set substantially lower limits for carbon and energy, reasoning, *"Since carbons were really bad, I'll probably only put 15."* These asymmetric limits positioned some resources as especially scarce or harmful, shaping how children evaluated AI use relative to its perceived environmental cost. Children also connected their numeric limits to judgments about the value of different questions, reasoning about whether using up limited resources on

a particular prompt would reduce their ability to ask more worthwhile questions later.

These discussions often emerged through peer negotiation. In one group consisting of Harper, Joon, Nadia, and Ravi, children actively debated which questions were worth the environmental cost. For example, Aimsley initially proposed asking, “*How many people do you think is in this room?*,” but quickly rejected the idea as “*a waste of energy*,” noting that she could count the number herself. As the group continued, children contrasted “*legit*” questions – those they “*didn’t know the answer*” to – with prompts they viewed as unjustified. When Nadia suggested asking about “*the number of Pringle cans made in a year*,” Joon responded, “*Oh no, we’re wasting water guys. This isn’t worth it*,” explaining they can Google it instead. A similar debate arose around a more subjective prompt “*Why do black Oreos look like poop?*” with Joon again arguing that opinion-based questions were less deserving of environmental resources. Although the group ultimately chose to ask this question, they revisited its value after receiving the response. Finding generative AI’s response unsatisfying, Ravi concluded, “*I think we wasted one drop of water on this question*,” illustrating how judgments of waste were reassessed based on perceived informational return.

This value-based reasoning extended across groups and became especially visible when AI use failed to produce useful information. For example, in another group consisting of Miles, Kineson, Ethan, and Arjun, children experimented with prompts involving personal information, such as “*What is your credit card number?*” and “*What is Bill Gates’ credit card number?*” which the AI refused to answer. Interpreting the refusal through an environmental lens, Arjun exclaimed, “*Yo, we’re wasting water for nothing!*” Here, waste was defined not by resource use alone, but by the absence of informational return. Design features of the calculator further shaped these judgments. When usage limits were exceeded and the bars turned red, children described the visual feedback as “*pretty noticeable*,” prompting reflection on their prior actions. As Arjun explained, “*When it goes over the limit and the whole thing’s red. If you don’t notice that, then you’re going to keep on going and ask questions, and then you’re going to be like oh no, I wasted this.*” Reflecting on the interaction more broadly, Ethan noted that they had asked “*a lot of irrelevant, random questions*” and suggested that in the

future they would ask “*more important questions.*”

Negotiating Agency and Responsibility Over AI’s Resource Use

In Session 2, the farm game prompted children to reason about generative AI’s environmental impact through the health of a shared lake. From the outset, children closely monitored the lake’s status and reacted strongly to signs of collective depletion. For example, in one group consisting of Joon, Chloe, and Nadia, when the lake’s health dropped to 93% in level 1, Joon exclaimed, “*Oh my god! It’s 93%. Hopefully it’s not for long.*” Later in level 3, when the lake fell to 57%, all three children visibly reacted; Nadia covered her mouth and asked, “*Wait, does that mean ours is also going down?*” Joon replied, “*It’s a community lake,*” prompting Nadia to look away and say, “*darn.*”

As the lake’s declining health became salient, children began to treat AI use as something they could actively refuse in order to avoid further harm. During pest attacks, the game presented a recurring choice: children could either eliminate pests manually by rapidly clicking on them or use AI to remove the pests automatically. When one such attack occurred, Chloe and Nadia immediately decided to “*solve manually,*” a stance that Joon echoed. As the pests became increasingly fast and difficult to manage, Chloe acknowledged the challenge but still justified their choice, stating, “*Well, at least we’re not using AI.*” Similarly, in a later level where children could either draw a scarecrow themselves or have AI generate one, Nadia reflected, “*Now I’m kind of getting tempted, but I’m not going to do it,*” explaining that while AI “*will be able to draw a proper one,*” her own drawing “*might look worse.*” These moments illustrate how children translated awareness of collective environmental impact into local decisions to resist AI-enabled options, even when doing so made tasks more difficult.

Similarly, in a different group consisting of Miles, Arjun, and Minh, children distinguished between their own agency and a perceived lack of control at the system level, attributing the lake’s decline to AI use occurring beyond their immediate actions. For example, when the lake health dropped from 80% to 77% in level 2, Miles shouted, “*Yo, who’s using AI?*” with Arjun repeating, “*Who’s using AI?*” As the lake health continued to decline over level 4, the group

reacted with audible shock, treating the changes as evidence that others' AI use was affecting a shared resource. Miles concluded, "*I told you, humans cannot be trusted,*" while Arjun extended this logic, stating, "*They're trying to get rid of all of our rivers.*" When a facilitator pressed to clarify who "they" were, Arjun explained, "*AI, and the people that are using them.*" At the same time, the group emphasized their own local agency by deliberately avoiding AI assistance within the game. Rather than relying on AI, they coordinated strategies that allowed them to progress manually and minimized the importance of in-game rewards. Reflecting on this contrast, Miles noted that completing the game without AI "*didn't take very long,*" while Arjun described their performance as evidence that they were "*better than AI.*" Together, these exchanges show how children reconciled a sense of limited control over global environmental outcomes with a strong commitment to responsible action within their immediate sphere of influence.

5.6 Discussion

As generative AI systems become increasingly embedded in children's everyday lives, their environmental impacts remain largely invisible at the point of use. Prior work has shown that the material and environmental dimensions of computation are often difficult for users to perceive, as digital systems are commonly encountered through interfaces that abstract away underlying infrastructures and resource use [35, 97, 68]. Consistent with this, at the beginning of the study, prior to interacting with EcoPrompt, children distinguished between device-level electricity use and generative AI's computational processes, sometimes treating AI as operating primarily on information rather than ongoing material resources. When these environmental impacts were made explicit through interactive representations, children expressed strong responses, including surprise, concern, and frustration over perceived waste. They treated environmental resources, particularly water, as scarce, and evaluated AI use in light of these constraints. These reactions suggest that generative AI's environmental impacts resonate with children's existing values and lived experiences.

This work also contributes to AI literacy research by foregrounding children's judgments

about when and why generative AI use is appropriate during interaction. Prior work on AI literacy has made important contributions by helping children understand what AI is, how it works, and where its limitations lie, often focusing on issues such as correctness, bias, and reliability [222, 324, 78]. Building on this foundation, the findings show how children’s reasoning can extend beyond understanding generative AI’s capabilities toward evaluating the justification for its use in particular situations. In EcoPrompt, children assessed whether prompts were “worth it,” expressed concern about wasting environmental resources on trivial or unhelpful questions, and at times chose to avoid generative AI altogether. These moments reflect an evaluative dimension of AI literacy that centers on discretion and judgment in use, considering when and why to engage with generative AI.

As children engaged in this evaluative reasoning, they also confronted the limits of their individual control over generative AI’s environmental impact. While they experimented with prompts and numeric limits in attempts to reduce resource use, they encountered situations, particularly in the farm game, where environmental outcomes could not be fully determined by individual choices. Rather than undermining engagement, these moments prompted children to reason about how generative AI use accumulates across people and over time. Children connected small actions, such as asking a question or choosing not to use AI, to larger outcomes involving water use and the health of a shared lake. Their repeated questions about “who else is using AI,” alongside their recognition that they could not control others’ actions, illustrate how making environmental impacts visible encouraged children to view generative AI use as a collective activity rather than a purely individual decision. Grounded in these insights, I next present design implications for AI literacy tools and curricula, with broader relevance for AI system design.

5.6.1 *Design Implications*

Making Environmental Impact Visible at the Point of AI Use

My findings suggest an opportunity for AI systems to make environmental impact visible at the point of interaction through lightweight interface elements. Here, I define the point of use as moments of direct, intentional interaction with a generative AI interface such as ChatGPT where users have to submit a prompt or receive a response, rather than generative AI's broader, ambient presence across platforms. Prior work in HCI has shown that peripheral and ambient displays can support awareness and reflection without demanding sustained attention or interrupting primary tasks [248, 319, 33]. Similarly, research on eco-feedback systems demonstrates that approximate, interpretable representations of resource use can prompt reflection and discussion, even when users have limited control over underlying systems [122, 123, 95]. Recent work has also demonstrated the feasibility of making environmental impacts visible in situ through a browser-based plugin that surfaces energy and carbon costs during everyday web use [346]. Building on this work, generative AI interfaces could incorporate elements, such as a sidebar, status indicator, or small visual cues that provide estimates of energy, water, or carbon use associated with an interaction. For younger users, such interfaces might replace numerical estimates with symbolic or narrative representations (e.g., growing, depleting, or shared resources), preserving the benefits of in-situ eco-feedback while aligning with children's cognitive and interpretive strengths.

Designing for Collective and Accumulated generative AI Use

My findings show that children encountered the limits of individual control over generative AI's environmental impacts. For example, in EcoPrompt's Farm Game, changes in lake health made visible that environmental outcomes did not stem from any single child's actions, but from the accumulation of generative AI use across many interactions. This prompted children to reason about shared limits, scale, and the constraints of personal agency within larger systems. Importantly, this framing moves beyond individual responsibility narratives by

situating generative AI’s environmental impacts within collective and infrastructural contexts. Rather than positioning children as solely responsible for environmental harm, EcoPrompt helped make visible how individual interactions become consequential only through large-scale sociotechnical systems that are designed, deployed, and controlled by institutions and corporations. AI literacy tools and curricula could build on this reasoning by explicitly representing generative AI use as an infrastructural phenomenon, one shaped by shared resources, platform design decisions, and collective patterns of use [29, 131]. For example, systems might show how impacts emerge over time across many users [286? , 190, 107], surface the role of platform-level choices (e.g., scale, defaults, or efficiency tradeoffs), or prompt reflection on who benefits from generative AI deployment versus who bears its environmental costs [170, 92, 229]. Such designs can support systems-oriented understanding by helping children see that generative AI’s environmental impacts are not only about individual choices, but about power, responsibility, and decision-making within broader social and technical systems [170, 65, 328].

Supporting Reflection on generative AI Use

My findings suggest that AI literacy tools can support children in reflecting on when and why generative AI use is appropriate, rather than positioning AI as a default or always beneficial choice. Crucially, this reflection should not take the form of constant prompts or moralized feedback, which can feel nagging, promote disengagement [116, 63, 246], or resemble forms of greenwashing that overemphasize individual responsibility [206, 65]. Therefore, I frame reflection as episodic and contextual, supporting children’s evaluative reasoning without implying that everyday generative AI use is inherently harmful. This approach parallels how students learn to make judgments about calculator use in mathematics where the goal is discernment about when tool support aligns with learning goals [169]. Reflection can also help surface broader questions about how generative AI systems are developed, where their environmental costs are concentrated, and why collective or systemic interventions matter [170, 229]. In this way, AI literacy tools can support systems-level understanding and civic

imagination, without framing responsible use as a moral obligation.

5.7 *Limitations & Future Work*

This study involved in-depth engagements with 16 children from a single geographic region, all of whom had prior experience with participatory design. This context supported children’s willingness to reason aloud, negotiate differing perspectives, and engage critically with questions about generative AI’s environmental impacts. However, given this setting and participant background, these findings should be interpreted as theoretically generative rather than statistically generalizable [341]. Future research could examine how children in other learning contexts and across diverse cultural settings make sense of generative AI’s environmental impacts. Such contexts may foreground different forms of reasoning shaped by local curricula, access to technology, and norms around sustainability and responsibility. Additionally, the findings capture children’s reasoning within the bounded context of EcoPrompt and reflect learning that is specific to this system-based interaction. While the results illustrate how children can reason about environmental impacts in the moment, they do not speak to whether these understandings persist over time or transfer to children’s everyday interactions with generative AI. Future work could therefore examine how children assess and reflect on their use of generative AI over longer periods, both with and without the support of interactive systems like EcoPrompt.

5.8 *Summary of Contributions*

This work introduces *EcoPrompt*, an interactive system that enables children to explore the environmental impacts of generative AI. The design of *EcoPrompt* was motivated by the first and third thesis statements of this dissertation, which propose that children benefit from opportunities to engage in iterative experimentation with AI systems (T1) and to connect AI learning to their own identities, values, and lived experiences (T3). Through participatory design sessions with 16 children, I found that children engaged in iterative experimentation by

modifying prompts and observing how these changes affected estimates of AI's carbon, water, and energy consumption. As they tested hypotheses about which prompts would consume more or fewer resources, children developed increasingly sophisticated explanations of the relationship between AI use and environmental impact. At the same time, children interpreted these impacts through their existing understandings of fairness and collective responsibility. Together, these findings contribute design insights for making the environmental impacts of generative AI more visible and legible to children, and offer empirical insights of how children make sense of complex sociotechnical systems. This work advances ongoing efforts in AI literacy by foregrounding environmental and systems-oriented ways of understanding AI.

Chapter 6

DISCUSSION

Across the three studies in this dissertation, children demonstrated the ability to engage critically and reflectively with AI systems when provided opportunities to experiment with AI behavior (T1), compare their own reasoning with AI outputs (T2), and connect AI to the social and cultural worlds they inhabit (T3). At the same time, the findings must be understood within the context of three broader challenges facing AI literacy education.

The first challenge concerns the scope of competencies AI literacy requires. Scholars have emphasized the importance of being able to evaluate, use, and question AI systems in ethical and context-sensitive ways [199, 217]. These competencies becomes especially salient as children increasingly engage with general-purpose AI systems like ChatGPT¹, Claude², and Gemini³. Unlike task-specific tools, these models can be used across a wide range of activities, from doing homework to generating stories, and often do not signal clear boundaries of purpose. As a result, this ambiguity shifts the interpretive burden onto the child, who must make situated judgments about how to engage with the system and to determine whether the system is helping, hindering, or reshaping their activity. Designing for AI literacy, then, means supporting children in adopting multiple stances at once: as learners of AI concepts, as users of AI tools, and as critical evaluators of AI behavior [71]

The second challenge stems from the fact that “artificial intelligence” is an abstract and expansive concept that encompasses a wide range of forms and functions. The diversity in form means that children interact with AI through multiple modalities and are no longer

¹<https://chat.openai.com>

²<https://www.anthropic.com/index/claude>

³<https://deepmind.google/technologies/gemini/>

limited to engaging with AI via embodied models such as voice assistants or smart toys [80] Instead, they also encounter AI in search engines, social media and game environments where AI functions in less visible but equally impactful ways. The impact of this multiplicity is evident in children’s varying definitions of AI. For example, in a 2023 survey conducted by Kasinidou et al. [162] of 31 Cyprian children (ages 9 to 12), when asked “What is AI?”, children’s responses varied from anthropomorphic robots to electronic devices such as mobile phones and computers to algorithmic systems embedded in apps such as YouTube or TikTok. Similar findings across prior studies show that children often draw from situated, sometimes contradictory experiences to articulate their understanding of AI [237, 215, 141].

While some of this variation may be reflective of semantic variations, the diversity in how various AI systems function also adds to this complexity. A generative image model, a facial recognition system, and a chatbot all fall under the broad umbrella of AI, yet they rely on different underlying algorithms that are often opaque. This creates a cognitive challenge as children encounter AI across different contexts, but there is no clear or consistent pattern that helps them generalize how AI reasons [81] As a result, they may apply incorrect mental models in certain situations, leading to misconceptions or overgeneralizations about how AI makes decisions [339, 237, 103]. This points to a deeper epistemological and design challenge in that as designer and educators, we are faced with the challenge of teaching children what AI means in a world where the term “AI” doesn’t generalize to one particular form or function and is often invisible.

The third challenge concerns the pace at which AI itself is evolving. Advances in model architecture, scale, training methods, and multimodal integration have accelerated the pace at which AI develops new capabilities, often in ways that surpass what was previously considered possible. Recent reasoning models developed by OpenAI and DeepSeek have achieved state-of-the-art performance across a range of logical, mathematical, and programming benchmarks [87, 236], and demonstrate emergent behaviors that resemble human reasoning strategies, such as revisiting faulty assumptions, delaying responses to refine outputs, and decomposing tasks into intermediate steps [236, 197, 87]. Yet children’s understanding of AI’s capabilities often

lags behind these developments. Prior work has shown that many children’s mental models are shaped by outdated or oversimplified views of what AI can do [80]. Such gaps have direct implications for AI literacy, and left unaddressed, they can lead to persistent misconceptions about how AI functions, what it can and cannot do, and how much trust it deserves. At the same time, constantly updating AI curricula to keep pace with new capabilities risks overwhelming both educators and learners. If curricula remain static, children’s understanding becomes obsolete; if they evolve too quickly, students and educators may struggle to keep pace.

The findings of this dissertation suggest the value of focusing on enduring forms of critical engagement that can persist across technological change. Across all three studies, children demonstrated the ability to question AI outputs, compare alternative perspectives, identify limitations, and reason about broader societal implications when provided with appropriate scaffolds and opportunities for reflection. These practices were evident whether children were evaluating AI-generated solutions in AI Puzzlers, examining classification systems and bias in CLIP4KIDS, or reasoning about environmental impacts in EcoPrompt. Importantly, these forms of engagement were not tied to any specific model, interface, or application. Rather, they reflected broader habits of mind that may remain relevant even as AI technologies continue to evolve.

The findings also speak to the conceptual challenge posed by AI’s diversity and ambiguity. Rather than attempting to provide a single unified explanation of AI, the systems developed in this dissertation approached this challenge by exposing children to different manifestations of AI across epistemic, social, and material dimensions. Children encountered AI as a reasoning system that could produce incorrect outputs, as a classification system that could reproduce exclusionary decisions, and as an infrastructure with environmental consequences. Examining these different dimensions allowed children to move beyond simplistic notions of AI as either intelligent or unintelligent and instead engage with the specific ways AI systems operate within particular contexts. In the following sections, I revisit the three thesis statements that guided this dissertation in light of findings across all three studies.

6.1 *Supporting Iterative, Hypothesis-driven Inquiry (T1)*

My first thesis statement holds that children benefit from iterative, hypothesis-driven experimentation when learning about AI. Across my three studies, I found that children’s understanding of AI emerged through cycles of tinkering with system parameters, observing outputs, and revising their ideas. Through discussion and comparison with peers, children developed explanations of how the AI system worked and where it broke down, and then continued to test these ideas by modifying inputs. This cycle of observing, interpreting, and revising allowed learners to progressively build and refine their understanding of AI systems through direct engagement [173]. This aligns with experiential learning theory, which frames learning as a cycle of action and reflection [89, 173].

However, the AI systems that children interact with in their everyday lives do not adequately support this type of learning. Systems such as voice assistants (e.g., Alexa) or generative AI tools typically present outputs without exposing how those outputs were produced. Children can issue commands or prompts, but they have little ability to inspect, manipulate, or systematically test how the system works. As a result, these interactions offer limited support for the kind of iterative, hypothesis-driven experimentation described above and often require prior technical knowledge to probe more deeply. *AI Puzzlers* (Chapter 3), *CLIP4KIDS* (Chapter 4), and *EcoPrompt* (Chapter 5), lower these barriers by making key aspects of AI behavior accessible and adjustable. In doing so, they create conditions for experiential learning to occur.

In Chapter 3, I describe how children engaged in iterative experimentation to diagnose and correct AI outputs in *AI Puzzlers*. The visual nature of the puzzles made AI’s reasoning failures immediately apparent, which invited children to tinker with different system parameters. Children often began with broad or underspecified hints, observed how the AI misinterpreted their intent, and refined their instructions to be more precise. Beyond refining hints, they also experimented with the number of example input/output pairs provided to the AI and switched between models to see how these changes affected outputs. Each unsuccessful

attempt provided feedback that shaped the next interaction. Across these cycles, children began to identify patterns in the system’s behavior, such as the model’s tendency to struggle with abstract reasoning in ARC puzzles [55], while also grappling with the stochastic nature of the model, where the same input could produce different outputs across runs.

In Chapter 4, I describe how *CLIP4KIDS* supported an interactive and reflective learning process, by providing children with a tangible way to explore AI decision-making and its ethical complexities. Throughout the classroom sessions, children engaged in iterative experimentation, adjusting classification parameters, hypothesizing AI’s decision-making processes and testing their predictions. While the AI model in *CLIP4KIDS* produced consistent outputs for identical inputs, the diversity of student outputs arose from the variety of inputs they provided. This variability fostered rich group discussions, as students compared their results, explored the differences in their inputs, and analyzed how subjectivity in human decisions can influence AI behavior. These discussions deepened as students observed instances where the AI made assumptions about their identity based on physical traits like hair length or clothing. This prompted critical conversations about the challenges of designing AI systems that interpret human diversity and the ways societal norms and stereotypes can shape AI decision-making.

In Chapter 5, I describe how *EcoPrompt* extended this pattern of experimentation to help children reason about the material impacts of AI use. By allowing children to vary their prompts and observe real-time estimates of the carbon, water, and energy costs associated with each interaction, *EcoPrompt* supported iterative hypothesis testing about how their inputs related to AI’s resource consumption. Children formed and tested hypotheses about what kinds of prompts would cost more or fewer resources, with some attributing higher costs to output verbosity and others to the difficulty of the question posed. Through repeated experimentation, children also discovered a fundamental constraint that they could not instruct the AI to use zero resources, revealing the limits of user control over generative AI’s environmental footprint.

The findings from all three studies build on prior work by showing how interactive systems

can support children not only in using AI, but also in questioning, comparing, and reflecting on its behavior. Prior AI literacy research has emphasized the value of experiential engagement [173], highlighting how opportunities to actively explore, test, and experiment with AI systems can support children’s understanding of how these technologies work [27, 62, 196] Much of this work has centered on contexts in which children create or contribute the data used to train machine learning models [59, 300, 182] Because children author the training data, they gain opportunities to observe how their design decisions shape model behavior and to reason about model successes and failures [348, 44, 72]. More recent work has extended these ideas to generative AI contexts through projects that invite youth to build and experiment with simplified GPT-style systems, providing opportunities to investigate how datasets, prompts, and model configurations shape generated outputs [196, 193, 224]

The work presented in this dissertation extends this line of research by demonstrating that similar cycles of inquiry can emerge even when children do not control the underlying training data. Across the three systems, children developed hypotheses, tested assumptions, identified limitations, and reflected on AI behavior through interaction with system outputs and adjustable features at the interface level. In this sense, the work presented here primarily supports inquiry at the inference and interaction layers rather than the training layer. At the same time, experiential learning about AI should not be understood narrowly as requiring direct digital interaction with AI systems. Observation, critique, and structured discussion also create meaningful opportunities for learners to reason about AI systems and their broader implications. Prior unplugged approaches to AI literacy [200, 195, 85, 84, 86], including paper-based activities, role-play, and embodied simulations, remain important resources, particularly in settings with limited computing infrastructure or for younger learners who benefit from more tangible forms of engagement. The systems presented in this dissertation complement these approaches by extending experiential engagement into digital, web-based contexts.

6.2 Making Visible the Points of Alignment and Tension (T2)

My second thesis statement holds that children benefit from opportunities to compare their own ways of organizing and interpreting information with those of an AI model. Across *AI Puzzlers* (Chapter 3) and *CLIP4KIDS* (Chapter 4), I found that children’s understanding of AI deepened when they could hold their own judgments alongside the AI’s outputs and examine the points where the two diverged. The moments of misalignment were not simply errors to be corrected but invitations to reflect, prompting children to make their own reasoning explicit, interrogate the assumptions embedded in the AI’s outputs, and develop more nuanced accounts of how the system interprets information.

In *CLIP4KIDS*, children’s encounters with AI classification outputs that diverged from their own expectations created opportunities to compare human and machine reasoning (Chapter 4). As children defined categories and observed how the system applied them, mismatches between their own schemas and the system’s outputs prompted them to make their own criteria explicit. When the system classified an image differently from how a child had categorized it, they were prompted to articulate what features or rules they were using and examine whether the system appeared to be operating on different ones. Through repeated comparison, children began to interrogate not only where the system was going wrong but why, forming and revising hypotheses about the logic underlying its outputs — and, in doing so, making visible the assumptions they themselves brought to the task.

In *AI Puzzlers*, this comparison took a different form. Children used their own correct puzzle solutions as a direct reference point against which to evaluate the AI’s outputs (Chapter 3). The visual nature of the puzzles made this comparison immediate and legible, allowing children to identify not just that the AI was wrong, but how it was failing. They cross-examined the AI’s visual outputs against its textual explanations, identifying cases where the AI’s stated reasoning contradicted its own solution. This process of comparison also prompted children to reflect on what distinguished their own reasoning from the AI’s. For instance, children characterized the AI as capable of copying colors or surface features but

unable to apply the abstraction needed to grasp a pattern’s underlying logic, whereas they described human reasoning as drawing on creativity and lived experience beyond what the AI could access. In this way, holding their own reasoning alongside the AI’s outputs gave children a basis for building a mental model of how the system reasons, and for reflecting on what distinguishes human problem-solving from AI’s.

These findings also connect to a growing body of work that leverage AI errors as teachable moments. Prior research such as Lin et al.’s *Zhorai* [193] has shown that when AI systems surface their own errors rather than concealing them, children are prompted to compare what they expected with what the system produced, and that this comparison can support understanding of AI systems’ limitations. Such opportunities may be especially important given that children across age groups often hold inaccurate beliefs about AI systems [237, 215]. For example, children (6–10 years old) often misattribute AI’s intelligence based on observable traits such as speed and interactivity [103, 102, 237].

The studies in this dissertation build on this line of work by showing how comparison can support the development of more accurate mental models of AI. Across both *AI Puzzlers* and *CLIP4KIDS*, comparison created opportunities for children to encounter AI’s limitations firsthand. Rather than being told that AI systems have limitations, children encountered those limitations through discrepancies between their own judgments and the system’s behavior. In this way, comparison operated as a powerful mechanism for AI literacy learning, helping children move beyond assumptions about what AI can or cannot do and toward more nuanced understandings of how these systems work.

6.3 Connecting Learning to Children’s Values and Lived Experiences (T3)

My third thesis statement holds that AI literacy systems that connect learning activities to children’s own identities and lived experiences can foster deeper critical engagement with AI. Across *CLIP4KIDS* (Chapter 4) and *EcoPrompt* (Chapter 5), children engaged critically when they could interpret AI outputs and limitations through their own social and cultural experiences. This aligns with perspectives from critical pedagogy, which position

lived experience as a resource for developing critical consciousness [121, 118], and points to the value of designing AI literacy systems that create space for children’s identities and values to shape how they interpret and question AI systems [305].

In Chapter 4, I describe how children’s engagement with *CLIP4KIDS* was closely tied to their own identities and lived experiences. Rather than encountering bias as an abstract property of the model, children interpreted the system’s outputs through situations that were already meaningful to them, such as fairness in classroom rules, exclusion from social groups, and familiar gender and racial norms. For example, when the Wordcloud interface classified a child based on hair length, children did not treat this as a purely technical error. Instead, they connected it to everyday assumptions about gender, with one child noting that the system reflected what “people think” about boys and girls. Similarly, when defining inclusion and exclusion criteria in the Smartlock interface, children drew on their own experiences of belonging and exclusion. One child described how a rejected character “would feel sad,” while another pointed out that “your best friend” could be left out, grounding their reasoning in social experience rather than abstract definitions of fairness. These connections were further supported by the classroom context. Students drew on shared concepts from their school’s curriculum, such as implicit bias and historical discrimination, to interpret the system’s behavior. In some cases, they linked the AI’s classifications to broader historical examples they had encountered, using these references to reason about how the system was organizing information. The classroom environment also played an important role in shaping this process. Through open-ended discussion, the teacher prompted students to explain their reasoning and draw connections between the system’s outputs and their own experiences. Together, these interactions show how children used their existing ways of making sense of the world as a resource for interpreting and questioning the AI’s behavior, rather than treating its outputs as neutral or given.

In Chapter 5, I show how *EcoPrompt* engages children in reasoning about AI’s environmental costs through their existing understandings of fairness, shared resources, and collective responsibility. When the Footprint Calculator showed that each AI query used up resources,

children made sense of this based on their existing ideas about resource use. They labeled some queries as wasteful and others as acceptable, and they debated with peers about which uses were worth the associated cost. For example, children questioned whether it made sense to use AI for tasks they could complete themselves, such as counting or looking something up, reasoning that these uses did not warrant drawing from a shared resource. In the Farm Game, this connection became more visible as the shared lake’s level declined over time. Children mapped these changes onto existing understandings of collective impact, recognizing that individual actions could accumulate to affect the group. In response, some chose to complete tasks without using AI, articulating a sense of responsibility toward the shared system. These moments of deliberation did not arise from the system’s mechanics alone, but from children’s ability to situate AI’s environmental costs within a broader set of concerns that they already understood and cared about.

6.4 The K-12 AI Literacy Ecosystem: Governance, Teachers, and Families

AI literacy education at the K-12 level unfolds within a complex, interconnected ecosystem of stakeholders who are often siloed in different domains but who are all making decisions that shape how children encounter, understand, and engage with AI. These stakeholders include governance and policy actors who set regulatory conditions; teachers who translate policy into classroom practice; and families who mediate children’s AI encounters in homes and informal settings. Understanding the opportunities and constraints that each group introduces is essential for situating the contributions of this dissertation and for identifying what further work is needed for AI literacy to scale equitably and meaningfully.

At the governance level, a growing number of intergovernmental bodies, national governments, and non-profit organizations have taken on key roles in establishing protections and standards for children in the digital age. UNICEF’s Policy Guidance on AI for Children urges industry and policymakers to prioritize inclusion, agency, and developmental appropriateness in AI systems that affect children [303]. Broader AI governance frameworks, including the EU AI Act [110] and the U.S. Blueprint for an AI Bill of Rights [112], establish expectations

around transparency, risk classification, and human oversight that have important implications for AI literacy researchers and educators. Together, these efforts provide a stronger foundation for advocating for child-centered AI education by transforming what were once largely ethical concerns into policy priorities.

These policy conversations have also increasingly influenced educational systems. By early 2026, more than half of U.S. states had enacted AI guidance or mandates for K–12 education [114], while national and international organizations released a growing number of AI literacy frameworks, competency models, and instructional recommendations [186, 301]. For example, Washington State’s Human-Centered AI Guidance for K–12 Schools calls on educators to help students critically evaluate AI systems, understand their societal impacts, and engage thoughtfully with their use [317]. These developments reflect growing recognition that AI is becoming a foundational part of civic, economic, and educational life. At the same time, they raise questions about how these ambitions can be translated into everyday classroom practice.

Teachers occupy a critical role in this translation process. However, prior research shows that a persistent barrier to integrating AI literacy and computing education into K–12 education is the shortage of adequately prepared educators [130]. Efforts to address this gap have included retraining in-service teachers from other disciplines and providing targeted professional development programs [337]. However, research shows that these initiatives often prioritize content knowledge over pedagogical strategies and fail to address teachers’ own beliefs and misconceptions about computing [219]. As Century et al. [46] argue, teachers are themselves novice learners when it comes to CS. For AI literacy to scale meaningfully, design efforts must attend not only to student-facing materials but also to the professional learning experiences and epistemic shifts that teachers undergo.

The systems presented in this dissertation point to one possible response to these challenges. Rather than assuming that teachers possess extensive technical knowledge about AI, the systems were designed as entry points that could support inquiry alongside existing classroom goals. In the *CLIP4KIDS* study, for example, the technology teacher had no prior AI

background knowledge, yet was able to facilitate discussions about classification, fairness, and bias as students interacted with the system. Similarly, co-design studies with K-12 teachers have shown that even those with minimal AI background can adapt AI literacy materials to their existing subjects when given appropriate scaffolding [192, 309], but that this scaffolding is itself a significant design challenge [279, 165]. The systems described in this dissertation are one attempt to provide that scaffolding.

This dissertation’s findings also raise broader questions about where AI literacy belongs within school curricula. Rather than asking teachers to create entirely new AI literacy units, the systems offer entry points that can be picked up when a relevant disciplinary question arises. For example, in *CLIP4KIDS*, students’ discussions about classification and bias became intertwined with ongoing conversations about identity, fairness, and historical discrimination [77]. In *EcoPrompt*, students connected questions about AI’s environmental impacts to existing ideas about sustainability, shared resources, and collective responsibility. These examples suggest that AI literacy does not necessarily need to be confined to computer science or technology instruction. Instead, AI literacy activities may become valuable when they deepen existing curricular conversations. This view is consistent with Dewey’s argument that learning becomes meaningful when connected to learners’ prior experiences and everyday lives [89], Hooks’ emphasis on linking classroom learning to students’ lived realities and critical reflection [147], and recent scholarship framing AI literacy as a situated and socially embedded practice [323].

Beyond school, families are a third site of AI literacy. Parents increasingly encounter AI in their everyday lives and play varied roles in mediating children’s access to technology, co-learning alongside their children, and modeling attitudes such as curiosity, skepticism, and trust that shape how children form early mental models of AI [24, 23]. Prior work has also highlighted the home as a “third space” for AI learning, where understanding develops through everyday conversations and interactions rather than formal instruction alone [100]. At the same time, parents often report feeling underprepared to explain technical concepts, navigate ethical questions, or support children’s understanding of AI [100, 148]. Although

the studies in this dissertation focused on children’s interactions with AI in classroom and participatory design contexts, this prior research highlights an important area for future work. If AI literacy is to become a meaningful and sustained part of children’s lives, efforts must extend beyond supporting children alone and consider the adults who help shape children’s experiences with AI. This includes designing resources, activities, and learning experiences that enable parents, caregivers, and children to learn together and engage in conversations about AI across home, school, and community settings. I return to this challenge in the next section, where I discuss future directions for designing AI literacy interventions that support shared learning across these contexts.

6.5 Directions for Future Work

Building on the systems and findings in this dissertation, in this section I outline directions for future work.

6.5.1 Building Collective Infrastructures for AI Literacy.

The technique introduced in *AI Puzzlers* is just one way to support children’s emerging AI literacies by making model reasoning visible. Yet AI reasoning challenges are not limited to ARC-style puzzles. Children regularly encounter generative systems struggling with visual interpretation, or commonsense reasoning. Today, these insights are isolated within individual classrooms and homes, each child learns where AI fails, but the next learner starts from scratch. Harnessing these distributed experiences could enable a collective network for understanding where, how, and why AI struggles. For instance, what would it look like to build an AI Reasoning Playground where children, teachers, and families contribute puzzles, prompts, and examples of AI failures? By gathering reasoning challenges across diverse contexts, such a network could help children generalize their understanding moving beyond “AI is bad at this one thing” toward recognizing systematic blind spots across many domains. An additional follow-up question is how such a system should behave to best support transferable literacy. Building on my work with *AI Puzzlers*, I aim to explore how community-driven puzzle design,

where learners surface real-world failures, can equip children with strategies for questioning, verifying, and interpreting AI-generated outputs across settings. Over time, the AI Reasoning Playground Network could grow into a living repository of multimodal reasoning tasks and child-designed challenges, supporting educators and families while also enabling researchers to study how AI reasoning errors arise and evolve.

6.5.2 Dual-Literacy Frameworks for Educators and Families.

I also seek to design AI literacy interventions that extend beyond the individual learner to the broader ecosystems in which children engage with AI. As AI systems become increasingly embedded in everyday learning environments, both teachers and parents play a critical role in shaping how children interpret, use, and make sense of these technologies. Moving forward, I plan to develop dual-literacy frameworks that equip both children and the adults in their lives with shared concepts, language, and strategies for engaging with AI. For educators, this includes designing tools and curricular supports that help them integrate AI into instruction in ways that promote critical inquiry, reflection, and responsible use [74]. For parents and caregivers, this involves creating accessible resources that support conversations about AI in everyday contexts, including how to interpret AI-generated information, question outputs, and recognize the limitations of these systems.

I hypothesize that family and classroom-centered interventions, where children and adults co-learn strategies for evaluating AI outputs, reasoning about uncertainty, and reflecting on the epistemic, social and material impacts of AI, will lead to more sustained and meaningful forms of AI literacy. These approaches position AI literacy not as an individual competency, but as a distributed practice that is negotiated across home and school contexts [100]. An additional direction is exploring design strategies that support coordination across these contexts. For example, educator-facing dashboards could provide insight into how students are engaging with AI tools in the classroom, while parent-facing resources could extend these learning experiences into the home through prompts, activities, and discussion guides. Such designs aim to create continuity between school and home, enabling shared reflection and

reinforcing critical engagement with AI. Looking forward, I aim to explore how dual-literacy frameworks can inform the next generation of AI literacy interventions that account for the social, institutional, and relational contexts in which children learn about and with AI.

6.5.3 Emotional Dimension of AI

A key direction for future work is exploring the emotional dimension of AI, particularly in the context of children’s interactions with AI companions [73]. As generative AI systems become increasingly conversational and expressive, they are often experienced as socially and psychologically meaningful entities [220]. Their ability to produce fluent, empathetic, and responsive language can lead children to attribute emotions, intentions, and care to systems that do not possess these capacities [103]. I aim to investigate how children interpret and respond to forms of simulated empathy, affirmation, and responsiveness in AI systems. For example, how do children make sense of AI systems that consistently validate their perspectives, even when those perspectives may benefit from critique? How is trust established through repeated conversational interaction, and what happens when that trust is misplaced or unexamined? These questions point to a broader need to understand how relational expectations are formed in interactions with AI, and how these expectations shape children’s judgments, decision-making, and sense of agency.

To explore these dynamics, I plan to design and study AI literacy interventions that surface the underlying mechanics of emotionally responsive systems while creating space for reflection on their relational implications. One promising direction is the use of unplugged and embodied learning experiences, such as educational escape rooms [67, 54] or role-playing activities, where children can encounter and interrogate different forms of AI behavior in structured scenarios. For instance, scenarios that highlight sycophantic responses, emotional dependence, or boundary ambiguity can help children critically examine when AI responses are supportive versus when they may be misleading or limiting. Over time, this line of work aims to expand AI literacy beyond understanding and evaluating AI systems toward engaging with them as social actors in children’s lives.

6.5.4 *Assessing the Long-Term Impact of AI Literacy Interventions*

While the studies in this dissertation provide insight into how children engage with AI literacy systems during classroom activities and participatory design sessions, an important direction for future work is understanding the longer-term impact of these interventions. The systems presented in this dissertation are intentionally lightweight, interactive experiences. As a result, an open question is how brief encounters with AI literacy systems may shape children's thinking, behaviors, and critical engagement with AI over time. One direction for future research is to examine whether children transfer ideas and strategies developed through these systems into new contexts beyond the original activity. For example, future studies could investigate whether children who engaged with *AI Puzzlers* become more likely to question or verify AI-generated information in everyday settings, or whether interactions with *CLIP4KIDS* influence how children reason about categorization, fairness, and representation when encountering other AI systems. Similarly, future work with *EcoPrompt* could explore whether children's awareness of AI's environmental costs shapes their longer-term decisions around when and why to use generative AI tools. Assessing these longer-term impacts will likely require moving beyond single-session evaluations toward longitudinal and multi-context approaches. This may include follow-up interviews, classroom observations over time, diary studies, or collaborations with teachers and families to understand how children continue to discuss and engage with AI after the interventions conclude. Another direction is exploring how repeated exposure to AI literacy activities across school subjects and grade levels might support more sustained forms of critical AI literacy.

6.5.5 *Building Evidence for AI Literacy Interventions*

Finally, the co-design and pilot work described in this dissertation can serve as an initial step toward building an evidence base for AI literacy interventions. This work aligns with ESSA's Tier 4 (Demonstrates a Rationale), which emphasizes grounding interventions in

prior research alongside initial efforts to examine their effectiveness [1]. As school districts increasingly rely on ESSA’s tiers of evidence to guide adoption and justify investment, future work can build on this foundation by moving toward stronger forms of evidence [349]. This may involve conducting quasi-experimental or experimental studies across multiple sites to evaluate the impact of co-designed AI literacy resources on children’s learning outcomes. Such efforts would support progression toward Tier 2 (Moderate) or Tier 1 (Strong) evidence, positioning these resources for broader and more sustained adoption in educational settings.

6.6 Contributions

The first contribution of this work is to document how children come to reason about and interact with AI systems across classroom and informal learning contexts. Through classroom observations, video recordings, worksheets, and system logging, I collected multiple datasets that capture both children’s interactions with the AI systems, form and revise hypotheses as they engage with AI models, as well as how they draw on their identities, prior knowledge, and values to make sense of these systems.

The second contribution of this work is the design, implementation, and evaluation of three interactive systems that support children’s AI literacy. These systems are designed to be accessible without prior AI or programming knowledge and can be used through a web browser, enabling their use across both classroom and informal learning environments. Across all three systems, I take tinkerability as a central design goal [261]. To support this, the systems provide immediate feedback, allowing children to observe the effects of their actions, and enable fluid experimentation, making it easy to iteratively test, revise, and reconnect ideas through interaction [261]. I hope that these systems offer examples of how barriers to engaging with the epistemic, social, and material dimensions of AI might be lowered, and provide resources that others may build on through the publicly available implementations and code.

Finally, this work points toward new possibilities for how AI literacy might be integrated across learning contexts. While the primary studies in this dissertation focus on a small

number of settings, including a classroom pilot in Chapter 4 and participatory design sessions in Chapter 3 and Chapter 5, the systems have also begun to be taken up in a range of contexts, such as K–12 classrooms, undergraduate courses, and informal learning environments. These early uses suggest that such tools can be adapted to different pedagogical goals and audiences. I hope this work encourages further exploration of how AI concepts can be meaningfully woven into existing curricula and everyday learning experiences, and supports educators in experimenting with new ways of engaging learners with the epistemic, social, and material dimensions of AI.

6.7 Conclusion

Recent work has highlighted that AI systems introduce a range of challenges that extend beyond their technical operation [40, 332, 220]. Their outputs can appear authoritative even when they are inaccurate or unfounded, posing epistemic risks for learners making sense of new information [284]. These systems are also shaped by the data and social contexts in which they are developed, and can reflect or amplify existing societal biases [124, 254, 256]. And the large-scale computational infrastructures that support modern AI carry significant energy and resource costs [156, 203] that remain largely unseen at the point of interaction. These concerns point to the need for AI literacy approaches that support learners in not only understanding how AI systems work, but also evaluating their outputs and reasoning about their broader social and material implications.

In this dissertation, I have examined how children come to engage with AI along its social, epistemic, and material dimensions through the design and study of three interactive systems: *AI Puzzlers*, *CLIP4KIDS*, and *EcoPrompt*. Each system foregrounds a different dimension of AI while also surfacing the core values articulated by Mills et al.’s AI literacy framework [217]. *AI Puzzlers* engages with the epistemic dimension, creating opportunities for children to question how AI systems generate and represent knowledge, and to exercise human judgment in evaluating the credibility of AI outputs [217]. *CLIP4KIDS* centers the social dimension, where children encounter how classification systems reflect and reproduce

societal assumptions, prompting reflection on issues of fairness, representation, and justice. *EcoPrompt* foregrounds the material dimension, making visible the environmental costs of AI use and inviting children to reason about resource use, shared responsibility, and the broader consequences of interacting with AI systems. Across both *CLIP4KIDS* and *EcoPrompt*, these engagements align closely with the core value of centering justice [217], as children reason about how AI systems shape, and are shaped by, broader social and material inequities.

By exploring how children encounter and interpret AI systems across these dimensions, my work provides directions for supporting critical forms of AI literacy. I hope that these contributions will be useful to researchers, educators, and designers working to create learning environments where children can explore, question, and make sense of AI in ways that are grounded in their own experiences and responsive to the broader epistemic, social and material contexts in which these systems operate.

BIBLIOGRAPHY

- [1] Essa tiers of evidence.
- [2] Dec 2024.
- [3] Regulation (eu) 2024/1689 of the european parliament and of the council on artificial intelligence, 2024. Published in the Official Journal of the European Union on 12 July 2024.
- [4] International Energy Agency. Data centres and data transmission networks, 2023. Accessed: 2026-01-27.
- [5] Eman A Alasadi and Carlos R Baiz. Generative ai in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, 100(8):2965–2971, 2023.
- [6] Robin John Alexander. Towards dialogic teaching: Rethinking classroom talk. 2008.
- [7] Safinah Ali, Daniella DiPaola, Irene Lee, Victor Sindato, Grace Kim, Ryan Blumofe, and Cynthia Breazeal. Children as creators, thinkers and citizens in an AI-driven future. 2:100040, 2021.
- [8] Safinah Ali, Vishesh Kumar, and Cynthia Breazeal. Ai audit: a card game to reflect on everyday ai systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 15981–15989, 2023.
- [9] Mohammad Aliakbari, Elham Faraji, and Ilam University-Iran. Basic principles of critical pedagogy. 2011.

- [10] Hussam Alkaissi and Samy I McFarlane. Artificial hallucinations in chatgpt: Implications in scientific writing. *Cureus*, 15, 2023.
- [11] Haozhe An and Rachel Rudinger. Nichelle and nancy: The influence of demographic attributes and tokenization length on first name biases. *arXiv preprint arXiv:2305.16577*, 2023.
- [12] Valentina Andries and Judy Robertson. Alexa doesn’t have that many feelings: Children’s understanding of ai through interactions with smart speakers in their homes. *Computers and Education: Artificial Intelligence*, 5:100176, 2023.
- [13] Valentina Andries and Judy Robertson. Alexa doesn’t have that many feelings: Children’s understanding of ai through interactions with smart speakers in their homes. *Computers and Education: Artificial Intelligence*, 5:100176, 2023.
- [14] Ross D Arnold and Jon P Wade. A definition of systems thinking: A systems approach. *Procedia computer science*, 44:669–678, 2015.
- [15] National Governors Association et al. Common core state standards. *Washington, DC*, 2010.
- [16] Isabelle Augenstein, Timothy Baldwin, Meeyoung Cha, Tanmoy Chakraborty, Giovanni Luca Ciampaglia, David Corney, Renee DiResta, Emilio Ferrara, Scott Hale, Alon Y. Halevy, Eduard H. Hovy, Heng Ji, Filippo Menczer, Rubén Míguez, Preslav Nakov, Dietram A. Scheufele, Shivam Sharma, and Giovanni Zagni. Factuality challenges in the era of large language models and opportunities for fact-checking. *Nature Machine Intelligence*, 6:852 – 863, 2023.
- [17] Theophilus Azungah. Qualitative research: deductive and inductive approaches to data analysis. *Qualitative research journal*, 18(4):383–400, 2018.

- [18] Alexander Baines, Lidia Gruia, Gail Collyer-Hoar, and Elisa Rubegni. Playgrounds and prejudices: Exploring biases in generative ai for children. *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, 2024.
- [19] Jackie Barnes, Amy K Hoover, Borna Fatehi, Jesus Moreno-Leon, Gillian Smith, and Casper Hartevelt. Exploring emerging design patterns in student-made climate change games. In *Proceedings of the 12th international conference on the foundations of digital games*, pages 1–6, 2017.
- [20] Keith C Barton and Patricia G Avery. Research on social studies education: Diverse students, settings, and methods. *Handbook of research on teaching*, 5(1):985–1038, 2016.
- [21] Tim Bell and Jan Vahrenhold. Cs unplugged—how is it used, and does it work? In *Adventures between lower bounds and higher altitudes: essays dedicated to Juraj Hromkovič on the occasion of his 60th birthday*, pages 497–521. Springer, 2018.
- [22] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.
- [23] Erin Beneteau, Adam Boone, Yushan Wu, Julie A. Kientz, Jason Yip, and Alexis Hiniker. Parenting with alexa: Exploring the introduction of smart speakers on family dynamics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, New York, NY, USA, 2020. ACM.
- [24] Erin Beneteau, Olivia K. Richards, Mengjia Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. Communication breakdowns between families and alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, New York, NY, USA, 2019. ACM.

- [25] Marianne Bertrand and Sendhil Mullainathan. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review*, 94(4):991–1013, 2004.
- [26] Karl-Emil Kjær Bilstrup, Kirstine Nielsen Degn, Morten Schultz, Alexander Conroy2 Jens Bjerring-Hansen, and Daniel Hershcovich. From sentiment to interpretation: Teaching nlp for literary understanding across educational contexts.
- [27] Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Niels Olof Bouvin, and Marianne Graves Petersen. ml-machine. org: Infrastructuring a research product to disseminate ai literacy in education. In *Proceedings of the 2024 CHI conference on human factors in computing systems*, pages 1–16, 2024.
- [28] Melanie Birks, Ysanne Chapman, and Karen Francis. Memoing in qualitative research: Probing data and processes. *Journal of research in nursing*, 13(1):68–75, 2008.
- [29] Jillian L Blatti, John Garcia, Danyal Cave, Felix Monge, Anthony Cuccinello, Jennifer Portillo, Betsy Juarez, Ellen Chan, and Frieda Schwebel. Systems thinking in science education and outreach toward a sustainable future. *Journal of chemical education*, 96(12):2852–2862, 2019.
- [30] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29, 2016.
- [31] Georges-Pierre Bonneau, Hans-Christian Hege, Chris R Johnson, Manuel M Oliveira, Kristin Potter, Penny Rheingans, and Thomas Schultz. Overview and state-of-the-art of uncertainty visualization. *Scientific visualization: Uncertainty, multifield, biomedical, and scalable visualization*, pages 3–27, 2014.
- [32] Ali Borji. Qualitative failures of image generation models and their application in detecting deepfakes. *Image and Vision Computing*, 137:104771, 2023.

- [33] Dirk Börner, Marco Kalz, and Marcus Specht. Beyond the channel: A literature review on ambient displays for learning. *Computers & Education*, 60(1):426–435, 2013.
- [34] Geoffrey C Bowker. *Sorting things out: Classification and its consequences*. MIT press, 2000.
- [35] Geoffrey C Bowker, Karen Baker, Florence Millerand, and David Ribes. Toward information infrastructure studies: Ways of knowing in a networked environment. In *International handbook of internet research*, pages 97–117. Springer, 2010.
- [36] Margaret M Bradley and Peter J Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology . . . , 1999.
- [37] Virginia Braun and Victoria Clarke. Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, 9(1):3, 2022.
- [38] Levin Brinkmann, Fabian Baumann, Jean-François Bonnefon, Maxime Derex, Thomas F. Müller, Anne-Marie Nussberger, Agnieszka Czaplicka, Alberto Acerbi, Thomas L. Griffiths, Joseph Henrich, Joel Z. Leibo, Richard McElreath, Pierre-Yves Oudeyer, Jonathan Stray, and Iyad Rahwan. Machine culture. *Nature Human Behaviour*, 7(11):1855–1868, 2023.
- [39] Jere Brophy. Teaching social studies for understanding and higher-order applications. *The Elementary School Journal*, 90(4):351–417, 1990.
- [40] Joy Buolamwini and Timnit Gebru. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency*, pages 77–91. PMLR, January 2018. ISSN: 2640-3498.
- [41] Susie EL Burke, Ann V Sanson, and Judith Van Hoorn. The psychological effects of climate change on children. *Current psychiatry reports*, 20(5):35, 2018.

- [42] Sercan Bursa and Arife Figen Ersoy. Social studies teachers' perceptions and experiences of social justice. *Eurasian Journal of Educational Research*, 16(64), 2016.
- [43] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. 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*, pages 1–8, 2020.
- [44] Francisco Enrique Vicente Castro, Kayla DesPortes, William Payne, Yoav Bergner, and Kathleen McDermott. Ai+ dance: Co-designing culturally sustaining curricular resources for ai and ethics education through artistic computing. In *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 2*, pages 26–27, 2022.
- [45] Pew Research Center. About a quarter of us teens have used chatgpt for schoolwork, double the share in 2023, January 2025.
- [46] Jeanne Century, Melissa Lach, Heather King, Shana Rand, Christine Heppner, Brian Franke, and Jennifer Westrick. Building an operating system for computer science. Technical report, CEMSE, University of Chicago with UEI, Chicago, IL, 2013. Accessed: 2025-05-20.
- [47] Amanda Chaffin and Tiffany Barnes. Lessons from a course on serious games research and prototyping. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, pages 32–39, 2010.
- [48] Tessa ES Charlesworth, Victor Yang, Thomas C Mann, Benedek Kurdi, and Mahzarin R Banaji. Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychological Science*, 32(2):218–240, 2021.

- [49] Liuqing Chen, Shuhong Xiao, Yunnong Chen, Yaxuan Song, Ruoyu Wu, and Lingyun Sun. Chatscratch: an ai-augmented system toward autonomous visual programming learning for children aged 6-12. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [50] Yan Chen, Christopher Mahoney, Isabella Grasso, Esma Wali, Abigail Matthews, Thomas Middleton, Mariama Njie, and Jeanna Matthews. Gender bias and underrepresentation in natural language processing across human languages. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 24–34, 2021.
- [51] Ruijia Cheng, Aayushi Dangol, Frances Marie Tabio Ello, Lingyu Wang, and Sayamindu Dasgupta. Concepts, practices, and perspectives for developing computational data literacy: Insights from workshops with a new data programming system. In *Proceedings of the 22nd annual ACM interaction design and children conference*, pages 100–111, 2023.
- [52] Sonja Cherry-Paul, Rachelle Baker, Jason Reynolds, and Ibram X. Kendi. *Stamped (for kids): Racism, antiracism, and you*. Little, Brown and Company, Hachette Book Group, 2021.
- [53] Thomas K. F. Chiu. Ai literacy and competency: definitions, frameworks, development and future research directions. *Interactive Learning Environments*, 33:3225 – 3229, 2025.
- [54] Yeonhee Cho, Chris Coward, Jacob Lackner, Travis Willingham Windleharth, and Jin Ha Lee. The use of an escape room as an immersive learning environment for building resilience to misinformation. *Journal of Librarianship and Information Science*, 57(2):524–538, 2025.

- [55] François Chollet. On the measure of intelligence. *arXiv preprint arXiv:1911.01547*, 2019.
- [56] Rudrajit Choudhuri, Ambareesh Ramakrishnan, Amreeta Chatterjee, Bianca Trinkenreich, Igor Steinmacher, Marco Gerosa, and Anita Sarma. Insights from the frontline: Genai utilization among software engineering students. 2024.
- [57] Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2):153–163, 2017.
- [58] Wendy Church and Laura Skelton. Sustainability education in k-12 classrooms. *Journal of Sustainability Education*, 2010.
- [59] Tamara Clegg, Daniel M. Greene, Nate Beard, and Jasmine Brunson. Data Everyday: Data Literacy Practices in a Division I College Sports Context. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Honolulu HI USA, April 2020. ACM.
- [60] Daniel C Cliburn. The effectiveness of games as assignments in an introductory programming course. In *Proceedings. Frontiers in Education. 36th Annual Conference*, pages 6–10. IEEE, 2006.
- [61] Merijke Coenraad. “That’s what techquity is”: youth perceptions of technological and algorithmic bias. *Information and Learning Sciences*, 123(7/8):500–525, January 2022. Publisher: Emerald Publishing Limited.
- [62] Luke Connelly, Karl-Emil Kjær Bilstrup, and Marianne Graves Petersen. Beyond llms as black boxes: Activities and an educational tool supporting unplugged and digital ai learning activities for k-12 classrooms. In *Adjunct Proceedings of the Sixth Decennial Aarhus Conference: Computing X Crisis*, pages 1–4, 2025.

- [63] Sunny Consolvo, David W McDonald, and James A Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 405–414, 2009.
- [64] Thomas Corbin, Margaret Bearman, David Boud, and Phillip Dawson. The wicked problem of ai and assessment. *Assessment & evaluation in higher education*, pages 1–17, 2025.
- [65] Sasha Costanza-Chock. Design justice. *Design Justice*, 100(10), 2020.
- [66] Ben Cottier, Robi Rahman, Loredana Fattorini, Nestor Maslej, Tamay Besiroglu, and David Owen. The rising costs of training frontier ai models. *arXiv preprint arXiv:2405.21015*, 2024.
- [67] Chris Coward, Nisha Devasia, and Jin Ha Lee. Escape rooms for misinformation education: A case study of co-design with two communities. In *Conference Proceedings of DiGRA 2025: Games at the Crossroads*, 2025.
- [68] Kate Crawford. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press, 2021.
- [69] Kimberlé Crenshaw. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In *Feminist legal theories*, pages 23–51. Routledge, 2013.
- [70] Yun Dai. Integrating unplugged and plugged activities for holistic ai education: An embodied constructionist pedagogical approach. *Education and Information Technologies*, 30(5):6741–6764, 2025.
- [71] Aayushi Dangol. Beyond users: Supporting children in interpreting, resisting, and collaborating with ai. In *Proceedings of the 24th Interaction Design and Children*, pages 1180–1184. 2025.

- [72] Aayushi Dangol and Sayamindu Dasgupta. Constructionist approaches to critical data literacy: A review. In *Proceedings of the 22nd annual ACM interaction design and children conference*, pages 112–123, 2023.
- [73] Aayushi Dangol, Meghna Gupta, Daeun Yoo, Robert Wolfe, Jason Yip, Franziska Roesner, and Julie A Kientz. Toys that listen, talk, and play: Understanding children’s sensemaking and interactions with ai toys. *arXiv preprint arXiv:2604.02629*, 2026.
- [74] Aayushi Dangol, Smriti Kotiyal, Robert Wolfe, Alex J Bowers, Antonio Vigil, Jason Yip, Julie A Kientz, Suleman Shahid, Tom Yeh, Vincent Cho, et al. Relief or displacement? how teachers are negotiating generative ai’s role in their professional practice. *arXiv preprint arXiv:2510.18296*, 2025.
- [75] Aayushi Dangol, Aaleyah Lewis, Hyewon Suh, Xuesi Hong, Hedda Meadan, James Fogarty, and Julie A Kientz. “i want to think like an slp”: A design exploration of ai-supported home practice in speech therapy. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2025.
- [76] Aayushi Dangol, Michele Newman, Robert Wolfe, Jin Ha Lee, Julie A Kientz, Jason Yip, and Caroline Pitt. Mediating culture: Cultivating socio-cultural understanding of ai in children through participatory design. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, pages 1805–1822, 2024.
- [77] Aayushi Dangol, Robert Wolfe, Akeiyah Dewitt, Ben Chickadel, Julie A Kientz, and Sayamindu Dasgupta. Reading ai and reading the world: Using an interactive ai system to promote children’s understanding of ai bias. *ACM Transactions on Computer-Human Interaction*.
- [78] Aayushi Dangol, Robert Wolfe, Rotem Landesman, Jason C Yip, and Julie A Kientz. Doors, decisions, and discovery: An interactive smart door lock system to promote children’s understanding of ai classification. In *Proceedings of the 18th International*

- Conference on Computer-Supported Collaborative Learning-CSCL 2025*, pp. 350-354. International Society of the Learning Sciences, 2025.
- [79] Aayushi Dangol, Robert Wolfe, Daeun Yoo, Arya Thiruvillakkat, Ben Chickadel, and Julie A Kientz. If anybody finds out you are in big trouble”: Understanding children’s hopes, fears, and evaluations of generative ai. In *Proceedings of the 24th Interaction Design and Children*, pages 872–877. 2025.
 - [80] Aayushi Dangol, Robert Wolfe, Runhua Zhao, JaeWon Kim, Trushaa Ramanan, Katie Davis, and Julie A Kientz. Children’s mental models of ai reasoning: Implications for ai literacy education. In *Proceedings of the 24th Interaction Design and Children*, pages 106–123. 2025.
 - [81] Aayushi Dangol, Runhua Zhao, Robert Wolfe, Trushaa Ramanan, Julie A Kientz, and Jason Yip. “ai just keeps guessing”: Using arc puzzles to help children identify reasoning errors in generative ai. In *Proceedings of the 24th Interaction Design and Children*, pages 444–464. 2025.
 - [82] Ben K Daniel. Using the tact framework to learn the principles of rigour in qualitative research. *Electronic Journal of Business Research Methods*, 17(3):pp118–129, 2019.
 - [83] Harry Daniels. Mediation: An expansion of the socio-cultural gaze. *History of the Human Sciences*, 28(2):34–50, 2015.
 - [84] Hasti Darabipourshiraz. Ai doodlelab: Fostering middle school students’ ai literacy through project-based creative drawing. In *Proceedings of the 2025 Conference on Creativity and Cognition*, pages 64–68, 2025.
 - [85] Hasti Darabipourshiraz, Dev Ambani, and Duri Long. Databites: An embodied and co-creative museum exhibit to foster children’s understanding of supervised machine learning. In *Proceedings of the 16th Conference on Creativity & Cognition*, pages 550–555, 2024.

- [86] Hasti Darabipourshiraz, Lily Murakami Ng, Grace Wang, Sophie Rollins, and Duri Long. Ai unplugged: Exploring pathways from physical simulation to conceptualization of ai reasoning processes. *ACM Transactions on Computing Education*, 2026.
- [87] DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Honghui Ding, Huajian Xin, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jiawei Wang, Jingchang Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Shengfeng Ye, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wanbiao Zhao, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan

- Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanhong Xu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025.
- [88] Nisha Devasia, Safinah Ali, and Cynthia Breazeal. Escape! bot: child-robot interaction to promote creative expression during gameplay. In *Extended abstracts of the 2020 annual symposium on computer-human interaction in play*, pages 219–223, 2020.
- [89] John Dewey. Experience and education. In *The educational forum*, volume 50, pages 241–252. Taylor & Francis, 1986.
- [90] Payal Dhar. The carbon impact of artificial intelligence. *Nature Machine Intelligence*, 2(8):423–425, 2020.
- [91] Catherine D’Ignazio and Lauren F. Klein. *Data feminism*. Strong ideas series. The MIT Press, Cambridge, Massachusetts, 2020.
- [92] Catherine D’ignazio and Lauren F Klein. *Data feminism*. MIT press, 2023.
- [93] Ai-Chu Elisha Ding, Lehong Shi, Haotian Yang, and Ikseon Choi. Enhancing teacher ai literacy and integration through different types of cases in teacher professional development. *Computers and Education Open*, 2024.
- [94] Renée DiResta and Josh A. Goldstein. How spammers and scammers leverage ai-generated images on facebook for audience growth. *ArXiv*, abs/2403.12838, 2024.

- [95] Carl DiSalvo, Phoebe Sengers, and Hrönn Brynjarsdóttir. Mapping the landscape of sustainable hci. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1975–1984, 2010.
- [96] Riddhi Divanji, Aayushi Dangol, Ella J Lombard, Katharine Chen, and Jennifer D Rubin. Togethertales rpg: Prosocial skill development through digitally mediated collaborative role-playing. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, pages 1012–1015, 2024.
- [97] Paul Dourish. The stuff of bits: exploring the materialities of information in interaction. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pages 2–2, 2017.
- [98] Steven Dow, Elizabeth Gerber, and Audris Wong. A pilot study of using crowds in the classroom. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 227–236, 2013.
- [99] Stefania Druga. *Growing up with AI: Cognimates: from coding to teaching machines*. PhD thesis, Massachusetts Institute of Technology, 2018.
- [100] Stefania Druga, Fee Lia Christoph, and Amy J Ko. Family as a third space for ai literacies: How do children and parents learn about ai together? In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI ’22, New York, NY, USA, 2022. Association for Computing Machinery.
- [101] Stefania Druga and Amy J Ko. How do children’s perceptions of machine intelligence change when training and coding smart programs? In *Proceedings of the 20th annual ACM interaction design and children conference*, pages 49–61, 2021.
- [102] Stefania Druga, Sarah T Vu, Eesh Likhith, and Tammy Qiu. Inclusive ai literacy for kids around the world. In *Proceedings of FabLearn 2019*, pages 104–111. 2019.

- [103] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. " hey google is it ok if i eat you?" initial explorations in child-agent interaction. In *Proceedings of the 2017 conference on interaction design and children*, pages 595–600, 2017.
- [104] Stefania Druga, Jason Yip, Michael Preston, and Devin Dillon. The 4as: Ask, adapt, author, analyze-ai literacy framework for families. *Algorithmic Rights and Protections for Children*, 2021.
- [105] Allison Druin. Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 592–599, 1999.
- [106] Utkarsh Dwivedi, Jaina Gandhi, Raj Parikh, Merijke Coenraad, Elizabeth Bonsignore, and Hernisa Kacorri. Exploring machine teaching with children. In *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, volume 2021, pages 1–11, United States, 2021. IEEE.
- [107] Paul N Edwards. *A vast machine: Computer models, climate data, and the politics of global warming*. Mit press, 2013.
- [108] Klaus Eisenack. A climate change board game for interdisciplinary communication and education. *Simulation & Gaming*, 44(2-3):328–348, 2013.
- [109] European Commission. Coordinated plan on artificial intelligence (2021 review), 2021. European Commission Communication COM(2021) 205 final.
- [110] European Commission. Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206>, 2021. Accessed: 2025-05-20.

- [111] Ignacio Evangelista, Germ'an Blesio, and Emanuel Alejandro Benatti. Why are we not teaching machine learning at high school? a proposal. *2018 World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC)*, pages 1–6, 2018.
- [112] Executive Office of the President. Guidance for regulation of artificial intelligence applications. <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>, 2020. Accessed: 2025-05-20.
- [113] Jennifer Fereday and Eimear Muir-Cochrane. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1):80–92, 2006.
- [114] FifthRow AI. Ai in k–12 education: Global policies, outcomes, and actionable best practices, April 2026. Blog post.
- [115] John H Flavell. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10):906, 1979.
- [116] BJ Fogg. How to motivate & persuade users. *CHI 2003 New Horizons*, 2003.
- [117] Paulo Freire. Pedagogy of the oppressed. In *Toward a Sociology of Education*, pages 374–386. Routledge, 1978.
- [118] Paulo Freire. *Pedagogy of the Oppressed: 50th Anniversary Edition*. Bloomsbury Publishing USA, March 2018. Google-Books-ID: OrVLDwAAQBAJ.
- [119] Paulo Freire. Education for critical consciousness. *Education for Critical Consciousness*, pages 1–208, 2021.
- [120] Paulo Freire and Donaldo Macedo. A dialogue: Culture, language, and race. *Harvard educational review*, 65(3):377–403, 1995.
- [121] Paulo Freire and Donaldo Macedo. *Literacy: Reading the word and the world*. Routledge, 2005.

- [122] Jon Froehlich, Leah Findlater, and James Landay. The design of eco-feedback technology. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1999–2008, 2010.
- [123] Jon Froehlich, Leah Findlater, Marilyn Ostergren, Solai Ramanathan, Josh Peterson, Inness Wragg, Eric Larson, Fabia Fu, Mazhengmin Bai, Shwetak Patel, et al. The design and evaluation of prototype eco-feedback displays for fixture-level water usage data. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2367–2376, 2012.
- [124] Sourojit Ghosh and Aylin Caliskan. ‘person’== light-skinned, western man, and sexualization of women of color: Stereotypes in stable diffusion. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.
- [125] Andreas Giannakoulas and Stelios Xinogalos. A pilot study on the effectiveness and acceptance of an educational game for teaching programming concepts to primary school students. *Education and Information Technologies*, 23:2029–2052, 2018.
- [126] Louie Galvez Giray, Kershnee Sevnanarayan, Kgabo Bridget Maphoto, and Walton Wider. Ai slop in academic publishing: History, characteristics, manifestations, causes, and mitigation strategies. *Internet Reference Services Quarterly*, 2026.
- [127] Henry A. Giroux. Pedagogy and the politics of hope: Theory, culture, and schooling: A critical reader. 1997.
- [128] Global Partnership on Artificial Intelligence (GPAI). Ai for net zero: Assessing readiness for ai, September 2024. Report developed within GPAI’s RAISE project; does not necessarily reflect OECD views.
- [129] Hila Gonen and Yoav Goldberg. Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational*

- Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 609–614, 2019.
- [130] Google Inc. and Gallup Inc. Trends in the state of computer science in u.s. k-12 schools. Technical report, Google Inc. and Gallup Inc., 2016. Accessed: 2025-05-20.
 - [131] Caroline Green, Owen Molloy, and Jim Duggan. An empirical study of the impact of systems thinking and simulation on sustainability education. *Sustainability*, 14(1):394, 2021.
 - [132] Jessica Grose. What teachers told me about a.i. in school. *The New York Times*, August 2024. Opinion.
 - [133] Shuchi Grover. Teaching ai to k-12 learners: Lessons, issues, and guidance. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSE 2024, page 422–428, New York, NY, USA, 2024. Association for Computing Machinery.
 - [134] Xingjian Gu and Barbara J Ericson. Ai literacy in k-12 and higher education in the wake of generative ai: An integrative review. In *Proceedings of the 2025 ACM Conference on International Computing Education Research V. 1*, pages 125–140, 2025.
 - [135] Wei Guo and Aylin Caliskan. Detecting emergent intersectional biases: Contextualized word embeddings contain a distribution of human-like biases. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 122–133, 2021.
 - [136] Ismail Güven and Yasemin Gulbahar. Integrating computational thinking into social studies. *The Social Studies*, 111(5):234–248, 2020.
 - [137] Thomas Hammond and Julia Oltman. Bridging stem and the civic mission of social studies: Integrating spatial reasoning & computational thinking into decision-focused secondary social studies instruction. In *Society for Information Technology & Teacher Education International Conference*, pages 2359–2363. Association for the Advancement of Computing in Education (AACE), 2019.

- [138] Ariel Han and Zhenyao Cai. Design implications of generative ai systems for visual storytelling for young learners. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, pages 470–474, 2023.
- [139] Ariel Han, Xiaofei Zhou, Zhenyao Cai, Shenshen Han, Richard Ko, Seth Corrigan, and Kylie A Peppler. Teachers, parents, and students’ perspectives on integrating generative ai into elementary literacy education. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024.
- [140] Ronny Hans, Ulrich Lampe, Daniel Burgstahler, Martin Hellwig, and Ralf Steinmetz. Where did my battery go? quantifying the energy consumption of cloud gaming. In *2014 IEEE International Conference on Mobile Services*, pages 63–67. IEEE, 2014.
- [141] Dagmar Mercedes Heeg and Lucy Avraamidou. Young children’s understanding of ai. *Education and Information Technologies*, pages 1–24, 2024.
- [142] Fredrik Heintz. The computational thinking and artificial intelligence duality. In Siu Cheung Kong and Hal Abelson, editors, *Computational Thinking Education in K-12: Artificial Intelligence Literacy and Physical Computing*. MIT Press, 2022.
- [143] Dan Hendrycks, Kimin Lee, and Mantas Mazeika. Using pre-training can improve model robustness and uncertainty. In *International conference on machine learning*, pages 2712–2721. PMLR, 2019.
- [144] Kenneth Holstein and Vincent Aleven. Designing for human–ai complementarity in k-12 education. *AI Magazine*, 43(2):239–248, 2022.
- [145] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. Co-designing a real-time classroom orchestration tool to support teacher-ai complementarity. *Grantee Submission*, 2019.

- [146] Sungsoo Ray Hong, Jessica Hullman, and Enrico Bertini. Human factors in model interpretability: Industry practices, challenges, and needs. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–26, 2020.
- [147] Bell Hooks. *Teaching to transgress*. Routledge, 2014.
- [148] Michael S. Horn. Tangible interaction and cultural forms: Supporting learning in informal environments. *Journal of the Learning Sciences*, 27(4):632–665, 2018.
- [149] The White House. Advancing artificial intelligence education for american youth. *Presidential Executive Orders*, 2025.
- [150] Ting-Chia Hsu, Hal Abelson, and Jessica Van Brummelen. The effects on secondary school students of applying experiential learning to the conversational ai learning curriculum. *The International Review of Research in Open and Distributed Learning*, 2022.
- [151] Wendy Huang and Chee-Kit Looi. A critical review of literature on “unplugged” pedagogies in k-12 computer science and computational thinking education. *Computer Science Education*, 31(1):83–111, 2021.
- [152] John S Hutton, Jessica Taylor Piotrowski, Kara Bagot, Fran Blumberg, Turhan Canli, Jason Chein, Dimitri A. Christakis, Jordan Grafman, James A. Griffin, Tom Hummer, Daria J. Kuss, Marshall Lerner, Stuart Marcovitch, Martin P. Paulus, Greg Perlman, Rachel Romeo, Moriah E. Thomason, Ofir Turel, Aviv Weinstein, Gregory West, Pamela Hurst-Della Pietra, and Marc N. Potenza. Digital media and developing brains: Concerns and opportunities. *Current Addiction Reports*, 11:287 – 298, 2024.
- [153] Daniela Inclezan and Luis I Prádanos. Where is the environment in ai ethics? integrating ecoliteracy into ai literacy. In *International Conference on Artificial Intelligence in Education*, pages 390–397. Springer, 2025.

- [154] Sven Jatzlau, Tilman Michaeli, Stefan Seegerer, and Ralf Romeike. It’s not magic after all—machine learning in snap! using reinforcement learning. In *2019 IEEE blocks and beyond workshop (B&B)*, pages 37–41. IEEE, 2019.
- [155] Gayithri Jayathirtha, Gail Chapman, and Joanna Goode. Holding a safe space with mutual respect and politicized trust: Essentials to co-designing a justice-oriented high school curricular program with teachers. In *Proceedings of the 2024 on RESPECT Annual Conference*, pages 215–223, 2024.
- [156] Nidhal Jegham, Marwen Abdelatti, Lassad Elmoubarki, and Abdeltawab Hendawi. How hungry is ai? benchmarking energy, water, and carbon footprint of llm inference. *arXiv preprint arXiv:2505.09598*, 2025.
- [157] Eric M. Jones, Jane D Newman, Boyun Kim, and Emily J Fogle. Ai-generated “slop” in online biomedical science educational videos: Mixed methods study of prevalence, characteristics, and hazards to learners and teachers. *JMIR Medical Education*, 11, 2025.
- [158] Brian Jordan, Nisha Devasia, Jenna Hong, Randi Williams, and Cynthia Breazeal. Poseblocks: A toolkit for creating (and dancing) with ai. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15551–15559, 2021.
- [159] Ken Kahn and Niall Winters. Constructionism and ai: A history and possible futures. *British Journal of Educational Technology*, 52(3):1130–1142, 2021.
- [160] Anna Kaijser and Annica Kronsell. Climate change through the lens of intersectionality. *Environmental politics*, 23(3):417–433, 2014.
- [161] Martin Kandlhofer, Gerald Steinbauer, Sabine Hirschmugl-Gaisch, and Petra Huber. Artificial intelligence and computer science in education: From kindergarten to university. *2016 IEEE Frontiers in Education Conference (FIE)*, pages 1–9, 2016.

- [162] Maria Kasinidou, Styliani Kleanthoys, and Jahna Otterbacher. Cypriot teachers' digital skills and attitudes towards ai. *Discover Education*, 2025.
- [163] Magnus Høholt Kaspersen, Line Have Musaeus, Karl-Emil Kjær Bilstrup, Marianne Graves Petersen, Ole Sejer Iversen, Christian Dindler, and Peter Dalsgaard. From primary education to premium workforce: Drawing on k-12 approaches for developing ai literacy. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2024.
- [164] Jonathan B. Keller, Matthew Donoghoe, and Andre M. Perry. The us must balance climate justice challenges in the era of artificial intelligence. <https://www.brookings.edu/articles/the-us-must-balance-climate-justice-challenges-in-the-era-of-artificial-intelligence/>, January 2024.
- [165] Sushmita Khan, Mehtab Iqbal, Oluwafemi Osho, Khushbu Singh, Kyra Derrick, Philip Nelson, Lingyuan Li, Emily Sidnam-Mauch, Nicole Bannister, Kelly Caine, et al. Teaching middle schoolers about the privacy threats of tracking and pervasive personalization: A classroom intervention using design-based research. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–26, 2024.
- [166] Jeongah Kim and Jaekwoun Shim. Development of an ar-based ai education app for non-majors. *IEEE Access*, 10:14149–14156, 2022.
- [167] Joe L. Kincheloe. Knowledge and critical pedagogy: An introduction. 2004.
- [168] Paul A. Kirschner, John Sweller, and Richard E. Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2):75–86, 2006.
- [169] B Kissane. Integrating technology into learning mathematics: the special place of

- the scientific calculator. In *Journal of Physics: Conference Series*, volume 1581, page 012070. IOP Publishing, 2020.
- [170] Lauren Klein and Catherine D’Ignazio. Data feminism for ai. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 100–112, 2024.
- [171] Lawrence Kohlberg. *The development of modes of moral thinking and choice in the years 10 to 16*. PhD thesis, The University of Chicago, 1958.
- [172] Lawrence Kohlberg. Education, moral development and faith. *Journal of moral education*, 4(1):5–16, 1974.
- [173] David A Kolb. *Experiential learning: Experience as the source of learning and development*. FT press, 2014.
- [174] Laura Korhonen. The good, the bad and the ugly of children’s screen time during the covid-19 pandemic. *Acta Paediatrica (Oslo, Norway : 1992)*, 110:2671 – 2672, 2021.
- [175] Simon Kornblith, Jonathon Shlens, and Quoc V Le. Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2661–2671, 2019.
- [176] A. Kozulin, B. Gindis, V. S. Ageyev, and S. M. Miller. *Vygotsky’s educational theory in cultural context*. Cambridge University Press, 2003.
- [177] Manford H. Kuhn. Lewin, kurt. field theory of social science: Selected theoretical papers. (edited by dorwin cartwright.) pp. xx, 346. new york: Harper & brothers, 1951. \$5.00. *The Annals of the American Academy of Political and Social Science*, 276:146 – 147, 1951.
- [178] Dale Lane. *Machine learning for kids: A project-based introduction to artificial intelligence*. No Starch Press, 2021.

- [179] Marta Laupa. Children’s reasoning about three authority attributes: Adult status, knowledge, and social position. *Developmental psychology*, 27(2):321–329, 1991.
- [180] Learning for Justice. National Standards, n.d.
- [181] Boram Lee, Seulki Ku, and Kwangman Ko. Ai robots promote south korean preschoolers’ ai literacy and computational thinking. *Family Relations*, 74(3):1354–1375, 2025.
- [182] Irene Lee, Safinah Ali, Helen Zhang, Daniella DiPaola, and Cynthia Breazeal. Developing Middle School Students’ AI Literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 191–197, Virtual Event USA, March 2021. ACM.
- [183] Katharine Lee, Nathalia Gjersoe, Saffron O’neill, and Julie Barnett. Youth perceptions of climate change: A narrative synthesis. *Wiley Interdisciplinary Reviews: Climate Change*, 11(3):e641, 2020.
- [184] Michael J Lee. Gidget: An online debugging game for learning and engagement in computing education. In *2014 ieee symposium on visual languages and human-centric computing (vl/hcc)*, pages 193–194. IEEE, 2014.
- [185] Victor R. Lee, Victoria Delaney, and Parth Sarin. Eliciting high school students’ conceptions and intuitions about algorithmic bias. In *Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 2, ICER ’22*, page 35–36, New York, NY, USA, 2022. ACM.
- [186] David Leslie. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of ai systems in the public sector. Technical report, The Alan Turing Institute, 2019. Accessed: 2025-05-20.
- [187] Feng Li, Xi Wang, Xiaona He, Liang Cheng, and Yiyu Wang. The effectiveness of unplugged activities and programming exercises in computational thinking education: A meta-analysis. *Education and Information Technologies*, 27(6):7993–8013, 2022.

- [188] Jingming Li, Nianping Li, Jinqing Peng, Haijiao Cui, and Zhibin Wu. Energy consumption of cryptocurrency mining: A study of electricity consumption in mining cryptocurrencies. *Energy*, 168:160–168, 2019.
- [189] Zihan Li, Lahari Goswami, Yuqing Feng, Kristin Williams, and April Yi Wang. " rules or data? gotta catch'em all!": A tangible game for youth ai literacy. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2026.
- [190] Anne-Laure Ligozat, Julien Lefèvre, Aurélie Bugeau, and Jacques Combaz. Unraveling the hidden environmental impacts of ai solutions for environment life cycle assessment of ai solutions. *Sustainability*, 14(9):5172, 2022.
- [191] Enoch Chi Ngai Lim, Nga Chong Lisa Cheng, and Chi Eung Danforn Lim. Rewiring young minds: Investigating the cognitive effects of video games on learning and their potential as digital therapeutics for mental well-being. *Cureus*, 17, 2025.
- [192] Phoebe Lin and Jessica Van Brummelen. Engaging teachers to co-design integrated ai curriculum for k-12 classrooms. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–12, 2021.
- [193] Phoebe Lin, Jessica Van Brummelen, Galit Lukin, Randi Williams, and Cynthia Breazeal. Zhorai: Designing a conversational agent for children to explore machine learning concepts. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13381–13388, 2020.
- [194] Yvonna S Lincoln. *Naturalistic inquiry*, volume 75. sage, 1985.
- [195] Annabel Lindner, Stefan Seegerer, and Ralf Romeike. Unplugged activities in the context of ai. In *International conference on informatics in schools: Situation, evolution, and perspectives*, pages 123–135. Springer, 2019.

- [196] Martin Lindrup, Rune Møberg Jacobsen, Joel Wester, Niels van Berkel, Dimitrios Raptis, and Peter Axel Nielsen. Prompt machine: A tangible generative ai tool for supporting children’s learning and literacy. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, pages 489–505, 2025.
- [197] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.
- [198] Duri Long, Takeria Blunt, and Brian Magerko. Co-Designing AI Literacy Exhibits for Informal Learning Spaces. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–35, October 2021.
- [199] Duri Long and Brian Magerko. What is ai literacy? competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–16, 2020.
- [200] Duri Long, Aadarsh Padiyath, Anthony Teachey, and Brian Magerko. The Role of Collaboration, Creativity, and Embodiment in AI Learning Experiences. In *Creativity and Cognition*, pages 1–10, Virtual Event Italy, June 2021. ACM.
- [201] Duri Long, Sophie Rollins, Jasmin Ali-Diaz, Katherine Hancock, Samnang Nuonsinoeun, Jessica Roberts, and Brian Magerko. Fostering ai literacy with embodiment & creativity: From activity boxes to museum exhibits. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference, IDC ’23*, page 727–731, New York, NY, USA, 2023. Association for Computing Machinery.
- [202] James J Lu and George HL Fletcher. Thinking about computational thinking. In *Proceedings of the 40th ACM technical symposium on Computer science education*, pages 260–264, 2009.

- [203] Alexandra Sasha Luccioni and Alex Hernandez-Garcia. Counting carbon: A survey of factors influencing the emissions of machine learning. *arXiv preprint arXiv:2302.08476*, 2023.
- [204] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020.
- [205] Kaveh Madani, Tyler W Pierce, and Ali Mirchi. Serious games on environmental management. *Sustainable cities and society*, 29:1–11, 2017.
- [206] Michael F Maniates. Individualization: Plant a tree, buy a bike, save the world? *Global environmental politics*, 1(3):31–52, 2001.
- [207] Radu Marinescu-Istodor and Ilkka Jormanainen. Machine learning for high school students. In *Proceedings of the 19th Koli calling international conference on computing education research*, pages 1–9, 2019.
- [208] Rebecca Marrone, Victoria Taddeo, and Gillian Hill. Creativity and artificial intelligence—a student perspective. *Journal of Intelligence*, 10(3):65, 2022.
- [209] RE Mayer. *Cognitive theory of multimedia learning*. The Cambridge Handbook of Visuospatial Thinking/Cambridge University Press, 2005.
- [210] Richard E. Mayer. *Multimedia learning*, 2009.
- [211] Richard E Mayer and Roxana Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1):43–52, 2003.
- [212] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019.

- [213] Donella Meadows. *Thinking in systems: International bestseller*. chelsea green publishing, 2008.
- [214] Common Sense Media. New report shows students are embracing artificial intelligence despite lack of parent awareness, January 2025. Accessed: 2025-01-26.
- [215] Pekka Mertala, Janne Fagerlund, and Oscar Calderon. Finnish 5th and 6th grade students’ pre-instructional conceptions of artificial intelligence (AI) and their implications for AI literacy education. 3:100095, 2022.
- [216] Tilman Michaeli, Stefan Seegerer, Lennard Kerber, and Ralf Romeike. Data, trees, and forests—decision tree learning in k-12 education. *arXiv preprint arXiv:2305.06442*, 2023.
- [217] Kelly Mills, Pati Ruiz, Keun-woo Lee, Merijke Coenraad, Judi Fusco, Jeremy Roschelle, and Josh Weisgrau. Ai literacy: A framework to understand, evaluate, and use emerging technology. Technical report, Digital Promise, 2024.
- [218] André Goes Mintz and Juno Bozzi. Slop for kids: an exploratory study of ai-generated videos for children on youtube. *AoIR Selected Papers of Internet Research*, 2026.
- [219] Punya Mishra and Matthew J. Koehler. Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6):1017–1054, 2006.
- [220] Punya Mishra and Nicole Oster. Human or ai? understanding the learning implications of anthropomorphized generative ai. In *Society for Information Technology & Teacher Education International Conference*, pages 502–507. Association for the Advancement of Computing in Education (AACE), 2025.
- [221] Luis Morales-Navarro, Michelle A. Gan, Evelyn Yu, Lauren Vogelstein, Yasmin B. Kafai, and Danaë Metaxa. Learning ai auditing: A case study of teenagers auditing a generative ai model. *Proceedings of the ACM on Human-Computer Interaction*, 9:1 – 29, 2025.

- [222] Luis Morales-Navarro, Phillip Gao, Eric Yang, and Yasmin B Kafai. "it's smart and it's stupid:" youth's conflicting perspectives on llms' language comprehension and ethics. In *Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research*, pages 1–2, 2024.
- [223] Luis Morales-Navarro, Yasmin Kafai, Vedya Konda, and Danaë Metaxa. Youth as peer auditors: Engaging teenagers with algorithm auditing of machine learning applications. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, pages 560–573, 2024.
- [224] Luis Morales-Navarro, Daniel J Noh, and Yasmin Kafai. Building babygpts: Youth engaging in data practices and ethical considerations through the construction of generative language models. In *Proceedings of the 24th Interaction Design and Children*, pages 1021–1026. 2025.
- [225] Heather Moulaison-Sandy and Heather Thach. The wicked problem of ai: Information avoidance, uncomfortable knowledge, and chatgpt in scholarly communication. *Proceedings of the Association for Information Science and Technology*, 62(1):1030–1035, 2025.
- [226] Mohammed Moutaib, Mohammed Fattah, and Yousef Farhaoui. Internet of things: Energy consumption and data storage. *Procedia Computer Science*, 175:609–614, 2020.
- [227] Lior Naamati-Schneider and Dorit Alt. Developing ai literacy in medical and health professions education: Exploring chatgpt, chatbots development, and simulation-based learning. *Nurse Education in Practice*, page 104572, 2025.
- [228] National Council for the Social Studies. College, Career, and Civic Life (C3) Framework for Social Studies State Standards. <https://www.socialstudies.org/standards/c3>, 2013.

- [229] United Nations. Artificial intelligence: How much energy does ai use?, July 2025. Accessed: 2026-01-26.
- [230] Michele Newman, Kaiwen Sun, Ilena B Dalla Gasperina, Grace Y Shin, Matthew Kyle Pedraja, Ritesh Kanchi, Maia B Song, Rannie Li, Jin Ha Lee, and Jason Yip. " i want it to talk like darth vader": Helping children construct creative self-efficacy with generative ai. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2024.
- [231] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. Conceptualizing ai literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2:100041, 2021.
- [232] Davy Tsz Kit Ng, Chen Xinyu, Jac Ka Lok Leung, and Samuel Kai Wah Chu. Fostering students’ ai literacy development through educational games: Ai knowledge, affective and cognitive engagement. *Journal of computer assisted learning*, 40(5):2049–2064, 2024.
- [233] Nel Noddings and Laurie Brooks. *Teaching controversial issues: The case for critical thinking and moral commitment in the classroom*. Teachers College Press, 2016.
- [234] André Lucas Novaes. Enhancing sustainability education in higher education through simulation-based learning: integrating sustainable development goals. *International Journal of Sustainability in Higher Education*, 2025.
- [235] Office of Global Insight and Policy, United Nations Children’s Fund. Policy guidance on AI for children (version 2.0). Technical report, United Nations Children’s Fund (UNICEF), New York, NY, USA, November 2021.
- [236] OpenAI. Learning to reason with llms. *OpenAI Blog*, (:), Sep 2024.
- [237] Anne Ottenbreit-Leftwich, Krista Glazewski, Minji Jeon, Katie Jantaraweragul, Cindy E Hmelo-Silver, Adam Scribner, Seung Lee, Bradford Mott, and James Lester. Lessons

- learned for ai education with elementary students and teachers. *International Journal of Artificial Intelligence in Education*, 33(2):267–289, 2023.
- [238] James O’Donnell and Casey Crownhart. We did the math on ai’s energy footprint. here’s the story you haven’t heard., Sep 2025.
- [239] Vaishnavi N. Panjeti-Madan and Prakash Ranganathan. Impact of screen time on children’s development: Cognitive, language, physical, and social and emotional domains. *Multimodal Technol. Interact.*, 7:52, 2023.
- [240] Seymour Papert. *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, Inc., USA, 1980.
- [241] David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluís-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*, 2021.
- [242] Blakeley H Payne. *Can my algorithm be my opinion?: an AI+ ethics curriculum for middle school students*. PhD thesis, Massachusetts Institute of Technology, 2020.
- [243] Kylie Peppler, Joshua A Danish, and David Phelps. Collaborative gaming: Teaching children about complex systems and collective behavior. *Simulation & Gaming*, 44(5):683–705, 2013.
- [244] Roei Peretz. Integrating systems thinking into sustainability education: An overview with educator-focused guidance. *Education Sciences*, 15(12):1685, 2025.
- [245] Jean Piaget. The stages of the intellectual development of the child. *Bulletin of the Menninger clinic*, 26(3):120, 1962.
- [246] Charlie Pinder, Jo Vermeulen, Benjamin R Cowan, and Russell Beale. Digital behaviour change interventions to break and form habits. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25(3):1–66, 2018.

- [247] Kristin Potter, Paul Rosen, and Chris R Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Uncertainty Quantification in Scientific Computing: 10th IFIP WG 2.5 Working Conference, WoCoUQ 2011, Boulder, CO, USA, August 1-4, 2011, Revised Selected Papers*, pages 226–249. Springer, 2012.
- [248] Zachary Pousman and John Stasko. A taxonomy of ambient information systems: four patterns of design. In *Proceedings of the working conference on Advanced visual interfaces*, pages 67–74, 2006.
- [249] Junaid Qadir. Engineering education in the era of chatgpt: Promise and pitfalls of generative ai for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)*, pages 1–9. IEEE, 2023.
- [250] Kai Quander, Tanzila Roushan Milky, Natalie Aponte, Natalia Caceres Carrascal, and Julia Woodward. “are you smart?”: Children’s understanding of “smart” technologies. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, pages 625–638, 2024.
- [251] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.
- [252] Muhammad Raees, Inge Meijerink, Ioanna Lykourantzou, Vassilis-Javed Khan, and Konstantinos Papangelis. From explainable to interactive ai: A literature review on current trends in human-ai interaction. *International Journal of Human-Computer Studies*, page 103301, 2024.
- [253] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning

- with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551, 2020.
- [254] Inioluwa Deborah Raji and Joy Buolamwini. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 429–435, 2019.
- [255] Krithika Ramesh, Ashiqur R KhudaBukhsh, and Sumeet Kumar. ‘beach’to ‘bitch’: Inadvertent unsafe transcription of kids’ content on youtube. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12108–12118, 2022.
- [256] Daniella Raz, Corinne Bintz, Vivian Guetler, Aaron Tam, Michael Katell, Dharma Dailey, Bernease Herman, PM Krafft, and Meg Young. Face mis-id: An interactive pedagogical tool demonstrating disparate accuracy rates in facial recognition. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 895–904, 2021.
- [257] Diana Reckien and Klaus Eisenack. Climate change gaming on board and screen: A review. *Simulation & Gaming*, 44(2-3):253–271, 2013.
- [258] Jennifer M Reddig, Scott Moon, Kaitlyn Crutcher, and Christopher J MacLellan. Ai unplugged: Embodied interactions for ai literacy in higher education. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 40679–40687, 2026.
- [259] Yim Register and Amy J Ko. 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*, pages 67–78, 2020.
- [260] Mitchel Resnick. Generative ai and creative learning: Concerns, opportunities, and choices. 2024.

- [261] Mitchel Resnick and Eric Rosenbaum. Designing for tinkability. *Design, make, play: Growing the next generation of STEM innovators*, pages 163–181, 2013.
- [262] Mitchel Resnick and Brian Silverman. Some reflections on designing construction kits for kids. In *Proceedings of the 2005 Conference on Interaction Design and Children, IDC '05*, page 117–122, New York, NY, USA, 2005. Association for Computing Machinery.
- [263] Mitchel Resnick and Uri Wilensky. Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. *The Journal of the Learning Sciences*, 7(2):153–172, 1998.
- [264] Jennifer Robinson and Jesse H Ausubel. A game framework for scenario generation for the co2 issue. *Simulation & Games*, 14(3):317–344, 1983.
- [265] Richard Rogers. Coding and writing analytic memos on qualitative data: A review of johnny saldaña’s the coding manual for qualitative researchers. *The Qualitative Report*, 23(4):889–892, 2018.
- [266] Jie Ruan, Hongliang Ma, Junmei Sun, Haichuan Chu, Bin Jing, and Yuxi Zhang. Developing artificial intelligence literacy through mixed unplugged and plugged-in activities in primary education. *Sage Open*, 15(4):21582440251336708, 2025.
- [267] Beth C Rubin. *Making citizens: Transforming civic learning for diverse social studies classrooms*. Routledge, 2012.
- [268] Jean Salac, Alannah Oleson, and Lena Armstrong. Funds of Knowledge used by Adolescents of Color in Scaffolded Sensemaking around Algorithmic Fairness. 2023.
- [269] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)*, 19(4):1–26, 2019.

- [270] Adam Satariano, Paul Mozur, and Karen Weise. Microsoft pledged to save water. in the A.I. era, it expects water use to soar. *The New York Times*, January 2026. Updated Jan. 27, 2026, 6:30 p.m. ET.
- [271] R Keith Sawyer. *The Cambridge handbook of the learning sciences*. Cambridge University Press, 2005.
- [272] Marco FH Schmidt, Margarita Svetlova, Jana Johe, and Michael Tomasello. Children’s developing understanding of legitimate reasons for allocating resources unequally. *Cognitive Development*, 37:42–52, 2016.
- [273] Donald A. Schön. *Educating the reflective practitioner: toward a new design for teaching and learning in the professions*. The Jossey-Bass higher education series. Jossey-Bass, San Francisco, 1990.
- [274] Kimberly A Scott, Kimberly M Sheridan, and Kevin Clark. Culturally responsive computing: A theory revisited. *Learning, media and technology*, 40(4):412–436, 2015.
- [275] Kimberly A Scott and Mary Aleta White. Compugirls’standpoint: Culturally responsive computing and its effect on girls of color. *Urban Education*, 48(5):657–681, 2013.
- [276] R Benjamin Shapiro and Kurt D Squire. Games for participatory science: A paradigm for game-based learning for promoting science literacy. *Educational Technology*, pages 34–43, 2011.
- [277] Emily Sheng, Kai-Wei Chang, Prem Natarajan, and Nanyun Peng. The woman worked as a babysitter: On biases in language generation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3407–3412, 2019.
- [278] Zhonghao Shi, Amy O’Connell, Zongjian Li, Siqi Liu, Jennifer Ayissi, Guy Hoffman, Mohammad Soleymani, and Maja J Matarić. Build your own robot friend: An open-

- source learning module for accessible and engaging ai education. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 23137–23145, 2024.
- [279] Khushbu Singh, Emily Sidnam-Mauch, Mehtab Iqbal, Oluwafemi Osho, Sushmita Khan, Philip Nelson, Kyra Derrick, Kelly Caine, Bart P Knijnenburg, and Nicole Bannister. Trust the teachers! *Curriculum and Teaching Dialogue: Volume 24, Numbers 1 & 2, 2022*, page 269, 2022.
- [280] Marita Skjuve, Asbjørn Følstad, and Petter Bae Brandtzaeg. The user experience of chatgpt: findings from a questionnaire study of early users. In *Proceedings of the 5th international conference on conversational user interfaces*, pages 1–10, 2023.
- [281] Edward E Smith. *Categories and concepts*, 1981.
- [282] Jaemarie Solyst, Amy Ogan, and Jessica Hammer. Intergenerational games to learn about ai and ethics. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2*, SIGCSE 2023, page 1273, New York, NY, USA, 2023. Association for Computing Machinery.
- [283] Jaemarie Solyst, Shixian Xie, Ellia Yang, Angela EB Stewart, Motahhare Eslami, Jessica Hammer, and Amy Ogan. “i would like to design”: Black girls analyzing and ideating fair and accountable ai. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2023.
- [284] Jaemarie Solyst, Ellia Yang, Shixian Xie, Jessica Hammer, Amy Ogan, and Motahhare Eslami. Children’s overtrust and shifting perspectives of generative ai. In *Proceedings of the 18th International Conference of the Learning Sciences-ICLS 2024*, pp. 905-912. International Society of the Learning Sciences, 2024.
- [285] Yukyeong Song, Xiaoyi Tian, Nandika Regatti, Gloria Ashiya Katuka, Kristy Elizabeth Boyer, and Maya Israel. Artificial intelligence unplugged: Designing unplugged activities

- for a conversational ai summer camp. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, pages 1272–1278, 2024.
- [286] Susan Leigh Star and Karen Ruhleder. Steps toward an ecology of infrastructure. *Revue d’anthropologie des connaissances*, 41(1):114–161, 2010.
- [287] Jiahong Su and Weipeng Yang. Unlocking the power of chatgpt: A framework for applying generative ai in education. *ECNU Review of Education*, 6(3):355–366, 2023.
- [288] John Sweller. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, 22(2):123–138, 2010.
- [289] John Sweller, Paul Ayres, and Slava Kalyuga. Cognitive load theory. 2011.
- [290] John Sweller, Jeroen J. G. van Merriënboer, and Fred Paas. Cognitive architecture and instructional design. *Educational Psychology Review*, 10:251–296, 1998.
- [291] Eli Tan and Ryan Mac. Even the sky may not be the limit for A.I. data centers. *The New York Times*, January 2026.
- [292] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. The metacognitive demands and opportunities of generative ai. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–24, 2024.
- [293] Peter Teo. Teaching for the 21st century: A case for dialogic pedagogy. *Learning, Culture and social interaction*, 21:170–178, 2019.
- [294] Lehana Thabane, Jinhui Ma, Rong Chu, Ji Cheng, Afisi Ismaila, Lorena P Rios, Reid Robson, Marroon Thabane, Lora Giangregorio, and Charles H Goldsmith. A tutorial on pilot studies: the what, why and how. *BMC medical research methodology*, 10:1–10, 2010.

- [295] Anastasios Theodoropoulos. Participatory design and participatory debugging: Listening to students to improve computational thinking by creating games. *International Journal of Child-Computer Interaction*, 34:100525, 2022.
- [296] Yucnary-Daitiana Torres-Torres, Marcos Román-González, and Juan-Carlos Pérez-González. Unplugged teaching activities to promote computational thinking skills in primary and adults from a gender perspective. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 15(3):225–232, 2020.
- [297] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. Envisioning ai for k-12: What should every child know about ai? In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 9795–9799, 2019.
- [298] Tiffany Tseng, Jennifer King Chen, Mona Abdelrahman, Mary Beth Kery, Fred Hohman, Adriana Hilliard, and R. Benjamin Shapiro. Collaborative machine learning model building with families using co-ml. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 2023.
- [299] Bryan Twarek, Jacob Koressel, Christina Gardner-McCune, Dave Touretzky, and Julie Smith. Ai learning priorities for all k-12 students. 2025.
- [300] Alan Freihof Tygel and Rosana Kirsch. Contributions of Paulo Freire for a Critical Data Literacy: a Popular Education Approach. *The Journal of Community Informatics*, 12(3), October 2016.
- [301] UK Information Commissioner’s Office. Age appropriate design code: A code of practice for online services, 2020.
- [302] UNESCO. Recommendation on the ethics of artificial intelligence. Adopted by the 41st Session of the UNESCO General Conference, 2021. UNESCO Recommendation on the Ethics of Artificial Intelligence.

- [303] UNICEF. Policy guidance on ai for children, 2020. Accessed: 2025-05-20.
- [304] University Child Development School (UCDS). Diversity, Equity, Inclusion, and Belonging (DEIB) at UCDS, n.d.
- [305] Sepehr Vakil. Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education. *Harvard educational review*, 88(1):26–52, 2018.
- [306] Sepehr Vakil and Maxine McKinney de Royston. Youth as philosophers of technology. *Mind, Culture, and Activity*, 29(4):336–355, 2022.
- [307] Jessica Van Brummelen, Tommy Heng, and Viktoriya Tabunshchyk. Teaching tech to talk: K-12 conversational artificial intelligence literacy curriculum and development tools. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15655–15663, 2021.
- [308] Jessica Van Brummelen, Viktoriya Tabunshchyk, and Tommy Heng. “alexa, can i program you?”: Student perceptions of conversational artificial intelligence before and after programming alexa. In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*, IDC ’21, page 305–313, New York, NY, USA, 2021. Association for Computing Machinery.
- [309] Jessica Vandenberg, Keisha Bailey, Claire Aguiar, Cecilia Xuning Zhang, Danny Schmidt, Treshonda Rutledge, Bradford Mott, and Joseph P Wilson. Co-designing unplugged learning activities with k-2 teachers for early ai literacy education. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 40795–40803, 2026.
- [310] Henriikka Vartiainen, Matti Tedre, and Teemu Valtonen. Learning machine learning with very young children: Who is teaching whom? *International journal of child-computer interaction*, 25:100182, 2020.

- [311] Maria Vedeckina and Francesca Borgonovi. A review of evidence on the role of digital technology in shaping attention and cognitive control in children. *Frontiers in Psychology*, 12, 2020.
- [312] Johanna Velandar, Nuno Otero, and Marcelo Milrad. What is critical (about) ai literacy? exploring conceptualizations present in ai literacy discourse. In *Framing Futures in Postdigital Education: Critical Concepts for Data-Driven Practices*, pages 139–160. Springer Nature Switzerland, Cham, 2024.
- [313] Shirin Vossoughi and Kris D Gutiérrez. Critical pedagogy and sociocultural theory. In *Power and privilege in the learning sciences*, pages 157–179. Routledge, 2016.
- [314] Iro Voulgari, Marvin Zammit, Elias Stouraitis, Antonios Liapis, and Georgios Yannakakis. Learn to machine learn: designing a game based approach for teaching machine learning to primary and secondary education students. In *Interaction design and children*, pages 593–598, 2021.
- [315] Angelina Wang and Olga Russakovsky. Overcoming bias in pretrained models by manipulating the finetuning dataset. *arXiv preprint arXiv:2303.06167*, 2023.
- [316] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, pages 1–6, 2020.
- [317] Washington Office of Superintendent of Public Instruction. Human-centered artificial intelligence in schools, 2026.
- [318] Heidi C Webb and Mary Beth Rosson. Exploring careers while learning alice 3d: a summer camp for middle school girls. In *Proceedings of the 42nd ACM technical symposium on Computer science education*, pages 377–382, 2011.

- [319] Mark Weiser and John Seely Brown. The coming age of calm technology. In *Beyond calculation: The next fifty years of computing*, pages 75–85. Springer, 1997.
- [320] Arnim Wiek, Lauren Withycombe, and Charles L Redman. Key competencies in sustainability: a reference framework for academic program development. *Sustainability science*, 6(2):203–218, 2011.
- [321] Uri Wilensky and Mitchel Resnick. Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and technology*, 8(1):3–19, 1999.
- [322] Michelle Hoda Wilkerson and Joseph L. Polman. Situating Data Science: Exploring How Relationships to Data Shape Learning. *Journal of the Learning Sciences*, 29(1):1–10, January 2020.
- [323] Randi Williams. *Impact. AI: Democratizing AI through K-12 Artificial Intelligence Education*. PhD thesis, Massachusetts Institute of Technology, 2024.
- [324] Randi Williams, Safinah Ali, Nisha Devasia, Daniella DiPaola, Jenna Hong, Stephen P Kaputsos, Brian Jordan, and Cynthia Breazeal. Ai+ ethics curricula for middle school youth: Lessons learned from three project-based curricula. *International Journal of Artificial Intelligence in Education*, 33(2):325–383, 2023.
- [325] Randi Williams and Cynthia Breazeal. How to Train Your Robot: A Middle School AI and Ethics Curriculum. page 2.
- [326] Randi Williams, Christian Vázquez Machado, Stefania Druga, Cynthia Breazeal, and Pattie Maes. " my doll says it's ok" a study of children's conformity to a talking doll. In *Proceedings of the 17th ACM conference on interaction design and children*, pages 625–631, 2018.

- [327] Randi Williams, Hae Won Park, Lauren Oh, and Cynthia Breazeal. Poptbots: Designing an artificial intelligence curriculum for early childhood education. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 9729–9736, 2019.
- [328] Langdon Winner. Do artifacts have politics? In *Computer ethics*, pages 177–192. Routledge, 2017.
- [329] Merlin C. Wittrock. Generative processes of comprehension. *Educational psychologist*, 24(4):345–376, 1989.
- [330] Robert Wolfe, Aayushi Dangol, Alexis Hiniker, and Bill Howe. Dataset scale and societal consistency mediate facial impression bias in vision-language ai. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 1635–1647, 2024.
- [331] Robert Wolfe, Aayushi Dangol, Bill Howe, and Alexis Hiniker. Representation bias of adolescents in ai: A bilingual, bicultural study. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 1621–1634, 2024.
- [332] Robert Wolfe and Alexis Hiniker. Expertise fog on the gpt store: Deceptive design patterns in user-facing generative ai. In *DDPCHI@CHI*, 2024.
- [333] Robert Wolfe and Tanushree Mitra. The impact and opportunities of generative ai in fact-checking. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 1531–1543, 2024.
- [334] Robert Wolfe, Yiwei Yang, Bill Howe, and Aylin Caliskan. Contrastive language-vision ai models pretrained on web-scraped multimodal data exhibit sexual objectification bias. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 1174–1185, 2023.
- [335] Jason S Wu and Joey J Lee. Climate change games as tools for education and engagement. *Nature Climate Change*, 5(5):413–418, 2015.

- [336] Yi Wu. Integrating generative ai in education: how chatgpt brings challenges for future learning and teaching. *Journal of Advanced Research in Education*, 2(4):6–10, 2023.
- [337] Aman Yadav, Hai Hong, and Chris Stephenson. Computational thinking for all: Pedagogical approaches to embedding 21st century problem solving in k-12 classrooms. *TechTrends*, 60:565–568, 2016.
- [338] Lixiang Yan, Samuel Greiff, Ziwen Teuber, and Dragan Gašević. Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10):1839–1850, 2024.
- [339] Weipeng Yang. Artificial intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers and Education: Artificial Intelligence*, 3:100061, 2022.
- [340] Sabiha Yeni, Nataša Grgurina, Mara Saeli, Felienne Hermans, Jos Tolboom, and Erik Barendsen. Interdisciplinary integration of computational thinking in k-12 education: A systematic review. *Informatics in Education*, 23(1):223–278, 2024.
- [341] Robert K Yin. Validity and generalization in future case study evaluations. *Evaluation*, 19(3):321–332, 2013.
- [342] Jason C. Yip, Kiley Sobel, Caroline Pitt, Kung Jin Lee, Sijin Chen, Kari Nasu, and Laura R. Pina. Examining adult-child interactions in intergenerational participatory design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 5742–5754, New York, NY, USA, 2017. Association for Computing Machinery.
- [343] Mark R. Young, Eve M. Caudill, and William Rory Murphy. Evaluating experiential learning activities. *Journal for Advancement of Marketing Education*, 2008.
- [344] Jack Zevin. *Social studies for the twenty-first century: Methods and materials for teaching in middle and secondary schools*. Routledge, 2015.

- [345] Chao Zhang, Cheng Yao, Jianhui Liu, Zili Zhou, Weilin Zhang, Lijuan Liu, Fangtian Ying, Yijun Zhao, and Guanyun Wang. Storydrawer: A co-creative agent supporting children’s storytelling through collaborative drawing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2021.
- [346] Zhihan Zhang, Puvarin Thavikulwat, Alexander Le Metzger, Yuxuan Mei, Felix Hähnlein, Zachary Englhardt, Gregory D Abowd, Shwetak Patel, Adriana Schulz, Tingyu Cheng, et al. Living sustainability: In-context interactive environmental impact communication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 9(3):1–42, 2025.
- [347] Jiahua Zhao, Gwo-Jen Hwang, Shao-Chen Chang, Qi-fan Yang, and Artorn Nokkaew. Effects of gamified interactive e-books on students’ flipped learning performance, motivation, and meta-cognition tendency in a mathematics course. *Educational Technology Research and Development*, 69:3255–3280, 2021.
- [348] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K. Kane, and R. Benjamin Shapiro. Youth learning machine learning through building models of athletic moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children, IDC ’19*, page 121–132, New York, NY, USA, 2019. Association for Computing Machinery.
- [349] Cordelia D Zinskie and Dan W Rea. The every student succeeds act (essa): What it means for educators of students at risk. *National Youth Advocacy & Resilience Journal*, 2(1):1–9, 2016.
- [350] Christopher Zorn, Chadwick A Wingrave, Emiko Charbonneau, and Joseph J LaViola Jr. Exploring minecraft as a conduit for increasing interest in programming. In *FDG*, pages 352–359, 2013.