

1 **ACORN: Creating Interactive Educational Stories that Adapt to Learners and**  
2 **Their Environments**

3  
4 ANONYMOUS AUTHOR(S)



28 Fig. 1. *Oak Story* is a mobile app that uses the ACORN system to engage children in narrative-based, real-world learning.

29 Narrative-based education can support children’s interest in learning, but the stories featured are typically rigid and lack  
30 adaptation to learners’ preferences or offer choices in a way that breaks immersion. In pursuit of student-driven learning  
31 experiences that invite active participation in both the story and the real world, we design, build, and evaluate an Agentic  
32 Conduit for Responsive Narratives (ACORN), an architecture employing a large language model to facilitate interactive stories  
33 aligned with educational goals. Driven by learners’ choices, ACORN generates character dialogue, evolves the storyline, and  
34 launches relevant multimodal and augmented reality activities that extend the narrative to the real world. Using ACORN, we  
35 create *Oak Story*, a mobile application teaching ecology through the inhabitants of oak woodlands. In a between-participants  
36 ablation study with fourth–sixth graders ( $N = 47$ ), we find ACORN produces statistically significant increases in engagement,  
37 learning gains, and learner agency compared to traditional rule-based narrative progression.

38  
39 **ACM Reference Format:**

40 Anonymous Author(s). 2025. ACORN: Creating Interactive Educational Stories that Adapt to Learners and Their Environments.  
41 In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*.  
42 ACM, New York, NY, USA, 32 pages. <https://doi.org/XXXXXXX.XXXXXXX>

43  
44  
45 Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that  
46 copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.  
47 Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy  
48 otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from  
49 [permissions@acm.org](mailto:permissions@acm.org).

50 © 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
51 Manuscript submitted to ACM

52 Manuscript submitted to ACM

## 1 INTRODUCTION

Stories have unique power to draw us into worlds that inspire the human imagination. In *The Diamond Age: Or, A Young Lady's Illustrated Primer* [48], a science fiction novel by American writer Neal Stephenson, protagonist Nell opens a high-tech book, her Illustrated Primer, causing four toys in her bedroom to spring into life. As they guide Nell through various challenges, these characters teach her skills ranging from reading to self-defense, thereby turning her otherwise drab life into a story of adventure. Nell's Illustrated Primer is a powerful learning tool because it adapts to the ideas and activities she chooses. By maintaining contextual awareness of Nell's thoughts, feelings, intentions, and environment, the Illustrated Primer creates learning experiences for Nell at the most opportune moments, which often occur outside the confines of a classroom.

The Primer gives Nell the power to steer her own learning journey—directing it not only by selecting topics that pique her interest but also by moving to different locations. This is self-directed learning [5] at its ideal, and research has repeatedly shown the effectiveness of student empowerment, particularly in informal learning settings outside of the classroom [4]. However, without an educator's support, it can be difficult for learners to take advantage of out-of-classroom opportunities to learn, as they lack the expert vision [20] to fully understand what they are looking at and why it matters. To a typical person, an oak tree might look like any other tree. Yet, to a botanist, an oak tree is the foundation of a diverse, thriving ecosystem. How can we help learners tap into the rich stories behind the seemingly mundane things in their environment? Inspired by *The Diamond Age*, we look to support learners with independent, out-of-classroom opportunities by leveraging automated agents that can enact contextually-aware, personalized learning experiences through stories. Modern advances in artificial intelligence (AI) and augmented reality (AR) have brought this vision closer to fruition, enabling learning opportunities to be tied to captivating, context-aware narratives.

In this work, we create educational stories that are unique to each learner and their surrounding environment by using AI to generate dialogue and behaviors for the characters in the story and AR to bridge the digital and physical worlds. To do this, we designed, built, and evaluated an Agentic Conduit for Responsive Narratives (ACORN), a system that interprets educational goals and a desired storyline to situate learning goals within a conversationally interactive experience driven by the student's interests and preferences.

ACORN leverages a large language model (LLM) to support a two-way conversational interface between the user and a narrative world that includes multiple characters, scenes, and AR activities. ACORN serves two primary functions. First, it interprets natural-language user input to guide a student towards learning goals, progress the narrative, and start activities that interface with the real world. Second, it uses the current state of the conversation and world to generate character dialogue that responds to student queries while incorporating contextual information about learning goals, available options, and educational content retrieved from a predefined knowledge base.

Extending prior work on outdoor learning experiences with narratives, AI, and AR [10, 11], we use ACORN to create *Oak Story*, an educational mobile application for 4th–6th graders centered on local oak woodlands (see Figure 1). *Oak Story* immerses students in oak ecosystems through a narrative in which they meet and interact with different animal characters, each with distinct personalities (see Figure 2). These characters guide learners through various activities of the learner's choice that connect to their physical surroundings, such as helping to identify oak trees in their vicinity. This context enriches the learner's knowledge about their local environment while keeping them engaged with the story.

105 We conducted a between-participants experimental study to evaluate ACORN and *Oak Story* in order to  
106 answer three primary research questions:  
107

108 **RQ1:** How does using *Oak Story* affect learning gains, motivation toward science learning, and engagement?

109 **RQ2:** How does the full *Oak Story* experience compare to an ablated version where ACORN is replaced with  
110 traditional linear narrative progression?  
111

112 **RQ3:** How do young learners (4th–6th grade) perceive and interact with ACORN and *Oak Story*?

113 We make three contributions to the literature. First, we present the design and implementation of ACORN, a  
114 system leveraging an LLM to enable a conversational interface that manages interactions between the user and  
115 the narrative to create personalized educational experiences linked to the physical world. Second, we contribute  
116 *Oak Story*, an educational app that combines an interactive story with real-world activities, and we provide  
117 empirical evidence that deploying ACORN within *Oak Story* promoted learning gains, improved engagement,  
118 and increased learners’ sense of agency among 4th–6th grade students. Third, we discuss several implications  
119 regarding the specific role of LLM-mediated interactions in enhancing user autonomy and engagement with  
120 other modes of interaction, including augmented reality.  
121

122 The results of this research support the creation of personalized learning environments that are readily avail-  
123 able to supplement in-school learning with compelling out-of-classroom education. With greater independence,  
124 learners can become more observant, inquisitive, and engaged with their immediate environment.  
125  
126

## 127 2 RELATED WORK

128 First, we summarize prior research on narrative-based learning environments, student agency, and LLMs in  
129 education in order to situate the present work in the literature.  
130  
131

### 132 2.1 Narrative-Based Learning Environments

133 Narrative-based learning environments integrate elements of storytelling into educational technologies. In  
134 particular, narratives can provide scaffolding that engages students and supports their learning [32]. For instance,  
135 prior work has demonstrated that narrative-based learning can lead to a variety of positive outcomes, such as  
136 increasing self-efficacy [34], promoting interest in subjects [34, 41], and fostering a sense of identity [41].  
137

138 Narrative-based learning environments can take many forms, but one area of focus in prior work is on  
139 experiences where the learner can take an active role in shaping the direction of the story itself. This improves  
140 engagement and may ultimately improve learning. Past research has advanced techniques in artificial intelligence  
141 (e.g., [43, 44]) to tailor story events to the learner’s decisions, but these systems are difficult to build.  
142

143 Recent advances in large language models (LLMs) and their generative capabilities have made it easier to  
144 create interactive narratives, sparking a new wave of interest in AI-supported storytelling [35, 62]. However,  
145 LLM chatbots are notoriously difficult to design and keep on task due to the fickle nature of prompting and  
146 unpredictable outputs [61]. In our system, we strike a balance between structure (imposed by the story and  
147 educational goals) and freedom (for the learner). By defining a limited set of events in the narrative, with  
148 educational goals provided for each event. While the conflict and resolution of the story are pre-defined, our  
149 architecture, ACORN, allows for flexibility in how the narrative unfolds based on the learners’ interactions and  
150 environmental context. At the same time, it ensures that the goals for each event are met and prevents the  
151 narrative from being derailed. Furthermore, ACORN streamlines the creation of our story, removing the need  
152  
153  
154  
155  
156

157 to write and edit countless lines of dialogue. High-level descriptions of the characters, scenes, and goals are  
158 defined by an education content creator and used by ACORN as bases for the text generated by the LLM.  
159

## 160 2.2 Student Agency in Education

161  
162 Learner agency, defined as students taking ownership of their learning [53] is associated with increased  
163 engagement, motivation, and academic achievement. The pivotal role of learner agency in education is amply  
164 documented [16]. For example, Hattie and Anderman [23] identify learner agency as a critical factor for student  
165 achievement, emphasizing the importance of feedback and teacher-student relationships in its development.  
166 Bandura’s social cognitive theory [2, 3] emphasizes the role of self-efficacy—and found that individuals with a  
167 higher sense of agency, tied to self-efficacy, may be more likely to seek help when facing challenges.  
168

169 Despite the amount of attention that the concept of learner agency has received from scholars in the learning  
170 sciences, there is relatively little work exploring how to design *technology* to support learner agency, especially for  
171 out-of-classroom contexts where an educator is not available. In this work, we present an approach employing  
172 interactive narrative and an LLM-based architecture to allow learners to shape their own path through the  
173 narrative.  
174  
175  
176  
177

## 178 2.3 Large Language Models in Education

179 Language models offer new possibilities for creating adaptive and personalized educational experiences:  
180 recent studies have explored the use of LLMs in generating explanations [31], feedback [47, 54], guided  
181 practice [9, 60], and engaging in tutoring dialogues [13, 21]. These applications have shown promise in  
182 delivering personalized learning experiences at scale.  
183

184 Character chatbots powered by LLMs have been previously applied in narrative education to improve interest  
185 and engagement [59]. However, these applications lack the capability to take tangible actions beyond text,  
186 limiting their perceived impact. Our approach in Oak Story represents a departure from prior applications of  
187 LLMs in education: unlike traditional chatbots, the characters in Oak Story are not only conversational partners,  
188 but also agents that can directly influence the story and perception of the real world (through AR). In this way,  
189 the LLM bridges between different modes of interaction, unifying both conversations and actions under the  
190 broader umbrella of a narrative where the learner is in control.  
191

192 To this end, our system employs a language model as a conduit between the learner’s natural language  
193 inputs and the narrative state. This approach allows us to leverage the strengths of LLMs in natural language  
194 processing and generation, while supplementing immersive activities with grounding in an existing narrative.  
195 The characters present options for exploration to the learner, and their response influences narrative progression  
196 and immersive activities. This bidirectional communication allows for the integration of multimodal and AR  
197 interaction with interactive storytelling, strengthening the impact of both.  
198  
199  
200

## 201 3 OAK STORY APPLICATION

202  
203 *Oak Story* is an interactive narrative-based learning app built with ACORN that engages with the learner’s  
204 environment. In this section, we describe the educational objectives, narrative design process, and resulting  
205 system, story, and interactions.  
206  
207

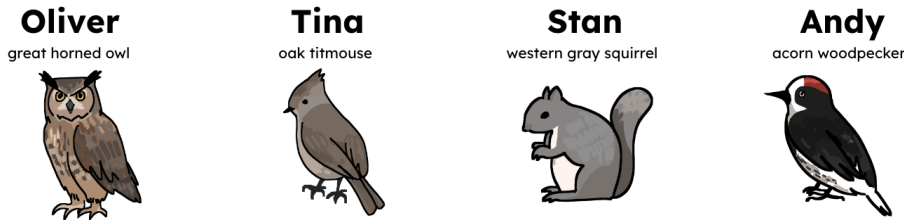


Fig. 2. The characters of *Oak Story*. From left to right: Oliver the owl, Tina the oak titmouse, Stan the squirrel, and Andy the acorn woodpecker.

### 3.1 Educational Objectives

*Oak Story* focuses on oak trees and their ecosystems. Oaks are keystone species that are environmentally, historically, and culturally significant in our local area, and they are the subject of active revitalization efforts [45]. We use oak trees and their surrounding ecosystems as an anchoring phenomenon to explore broader concepts in ecology such as seed dispersal and symbiosis. Our system targets the following Next Generation Science Standards (NGSS) evidence statements [46] for fourth grade and middle school:

- (1) Construct an argument that plants and animals have internal and external structures that function to support survival, growth, behavior, and reproduction.
- (2) Construct an argument supported by empirical evidence that changes to physical or biological components of an ecosystem affect populations.
- (3) Construct an argument supported by evidence for how increases in human population and per-capita consumption of natural resources impact Earth's systems.

Because the NGSS evidence statements are broad, we scoped them down to the domain of oak ecosystems and came up with the following list of learning objectives for *Oak Story*, which were selected to target different levels within Bloom's Taxonomy [1]:

- (1) Identify an oak tree by looking at its structures (leaves, acorns) and distinguish oak trees from other trees.
- (2) Explain why oak trees are important to both natural ecosystems and human society.
- (3) Analyze the relationships within an oak ecosystem and explain how different organisms in that ecosystem interact with each other.
- (4) Understand how an animals' structures and behaviors help them survive.
- (5) Give an example of how human activity affects natural ecosystems.

### 3.2 Narrative Design

In *Oak Story*, the learner explores different facets of the oak woodland ecosystem with the help of an animal companion, Oliver the owl. Eventually, they meet Tina the oak titmouse, whose oak tree home has been cut down. As the learner investigates different parts of the ecosystem, they find out about Tina's predicament, meet new friends in the forest, and gather clues to help Tina find a new home. To avoid introducing "seductive details" that may hinder or distract from learning [22, 49], we chose to keep the narrative grounded in real-life ecosystem relationships, such as the fact that titmice sometimes live in cavities excavated by woodpeckers [45].

261 In selecting the narrative’s primary conflict, we wanted to show users a relevant ecological issue and empower  
262 them with an opportunity within the story to take concrete steps to rectify it. We thus focused the story on  
263 finding a new home for an oak titmouse, a local bird that relies on cavities in oak trees for its habitat.  
264

265 We draw inspiration from the 5E model of instruction for inquiry learning [7, 18], an approach rooted  
266 in constructivist theories of learning that provides a recommended sequence of instructional steps (Engage,  
267 Explore, Explain, Elaborate, Evaluate) to promote effective active learning. We structure our narrative to align  
268 with the Engage, Explore, and Evaluate phases of the 5E model.  
269

270  
271 *3.2.1 Engage.* In the introduction of the story, the learner is introduced to Oliver the owl, who **engages** the  
272 learner by establishing the environmental, historical, and cultural significance of oak trees and asking what the  
273 learner wants to know more about. The learner is prompted to type or dictate their response, and the system  
274 considers their expressed preference when Oliver proposes possible next actions.  
275

276  
277 *3.2.2 Explore.* Next, Oliver brings the learner to a “central area” where the learner can start to **explore** the  
278 oak ecosystem and its various inhabitants. The following scenes are available, and which ones Oliver suggests  
279 depends on what preferences the learner expressed earlier:  
280

- 281 • Go to a magic portal that shows what the oak woodlands in the area might have looked like 200 years  
282 ago (see Figure 3, left).
- 283 • Meet Tina the oak titmouse, a dispirited bird who recently lost her home due to deforestation.
- 284 • Meet Andy the acorn woodpecker, an acorn-obsessed woodpecker who might have just the beak to peck  
285 Tina a new home if given the right incentive.
- 286 • Meet Stan the squirrel, a friendly rodent who’s happy to share his stashed acorns.
- 287 • Find and identify an oak tree in real life (which might be the perfect new home for Tina!).
- 288 • Try to find Tina a new home (this scene is available after meeting Tina but cannot be fully completed  
289 until the learner has befriended all the animals).  
290

291  
292 An overview of each scene in *Oak Story* can be found in Table 1.  
293

294 Because the **explore** phase in the 5E model is meant to be learner-directed, we wanted to give the learner  
295 multiple options within the story to investigate. Therefore, we designed *Oak Story*’s narrative as a branching  
296 story where the learner can have control over which paths to explore and when.  
297

298  
299 *3.2.3 Evaluate.* Upon completing one of the exploration scenes, the learner returns to the central area, and  
300 Oliver the owl asks the learner to **evaluate** what they have learned by asking them to reflect on what they have  
301 seen and read so far. As suggested by the 5E model, these reflections are embedded throughout the learning  
302 experience as lightweight, formative assessments.  
303

304  
305 *3.2.4 Repeat and Resolve.* After the learner is done reflecting, Oliver asks them what they want to do next,  
306 starting another exploration phase. This cycling of exploring and then evaluating repeats until all the exploration  
307 scenes are complete.

308 Then, the story transitions to the resolution, where the learner gathers all the animals in the woodland. The  
309 learner take the acorns they received from Stan the squirrel and gives them to Andy the acorn woodpecker. In  
310 exchange, Andy pecks a large AR hole in the (real-life) oak tree the learner found for Tina the oak titmouse to  
311

Scene Description	Activity	Lesson	Prerequisites
Meet Oliver the owl (Intro)	[None]	Overview of oaks and their ecological and societal importance	[None]
The “central area”	Traverse to other scenes	Reflect on the previous scene	[None]
Go to the magic portal	Portal	History of local oak woodlands	[None]
Meet Tina the oak titmouse	[None]	Causes and impact of habitat destruction	[None]
Meet Andy the acorn woodpecker	[None]	Organisms’ adaptations for survival and competition between animals	[None]
Meet Stan the squirrel	Squirrel	Symbiotic relationships	[None]
Find an oak tree in real life	Tree Identification	Species identification	[None]
Attempt to rehome Tina	[None]	[None]	Meet Tina
Find Tina a new home	Woodpecker	[None]	Meet Tina, Andy, and Stan, and identify an oak tree
Conclusion	Oak lifecycle	Lifecycle of an oak tree	All other scenes

Table 1. The scenes in *Oak Story* and their corresponding activities, lessons, and prerequisites.

live in. Finally, after Tina has successfully been rehomed, the learner participates in a final tree planting activity, which concludes the *Oak Story* experience (Figure 4, right).

### 3.3 System Description

*Oak Story* is a mobile application with a client-server architecture. The client is built with the Unity [51] game engine and deployed on iOS devices. The UI for dialogue was implemented with Yarn Spinner [30], a dialogue engine. All the real-world activities were implemented client-side, and the augmented reality (AR) features for such activities were developed using Niantic Lightship ARDK [36]. We drew all artwork in the Unity client, including character and background art.

The Unity client connects to ACORN, our LLM architecture that is described in greater detail in Section 4, over a WebSocket connection. ACORN is implemented in Python and hosted on Heroku [24], and it runs on Uvicorn [52], an Asynchronous Server Gateway Interface (ASGI) webserver. Under the hood, ACORN uses OpenAI’s GPT-4o model [40], a state-of-the-art LLM with a low-latency inference provider that supports multimodal inputs with fast inference so we can give timely outputs to the learner.

We now describe the core features of *Oak Story*.

**3.3.1 Dynamic dialogue.** To adapt learning content and narrative progression to each learner, dialogue is dynamically generated by ACORN in response to text or image input received from the user. At each timestep (one user input-system response pair), the learner is usually prompted to enter a text response. They are also able to use image input from the phone camera when the narrative calls for it, such as when the learner is guided to identify an oak tree (see Figure 4). On the frontend, the character dialogue to display next is received in the format of a screenplay and each line of dialogue is presented sentence by sentence to the learner. The text is accompanied by a gibberish chatter-like voiceover, differing for each character, akin to the speech audio



Fig. 3. Participant interactions with *Oak Story* and *ACORN*. In the left two images: a learner views an AR scene showcasing how local oak woodlands looked 200 years ago. In this portal scene, dialogue from the characters is overlaid on the AR activity. In the right two images: the learner has asked a question, and the answer is given by the characters and situated in the narrative. When not engaged in a multimodal/AR activity, the characters in the scene and their messages are displayed to the learner in front of an illustrated background.

in *Animal Crossing* [37] games with the intention of increasing sensory immersion. To address response latency, a thinking animation for the characters is displayed when the learner is waiting for the characters' response.

The responsive nature of character messages is reflective of the dialogic framework of teaching and learning [57]: characters ask and answer open-ended questions, provide relevant knowledge and feedback, and guide the learner to learn about ecological concepts in their own environment through personal exploration that matches their desired pace (see Figure 3).

**3.3.2 Real world activities.** At several points in the experience, real world activities situated in the narrative are initiated when *ACORN* sends a corresponding action call to the Unity client. (More information about this is provided in Section 4.) When an action corresponding to starting a real world activity is received, the client launches the activity once all dialogue is complete. When the learner finishes the activity, they return to a conversation with the characters and *ACORN* provides the LLM with a description of the completed activity, creating a seamless experience without interruption to narrative continuity. Some real world activities feature specific characters in the narrative, while others are general and center broader educational content. See Figure 4 for screenshots of these activities that help support learning.

**Tree identification.** In the *Tree Identification* activity, the learner is tasked with finding an oak tree in their local environment. Unlike the other activities, which are created with Niantic Lightship, the tree identification activity leverages the multimodal capabilities of GPT-4o to receive and parse images that the learner sends.



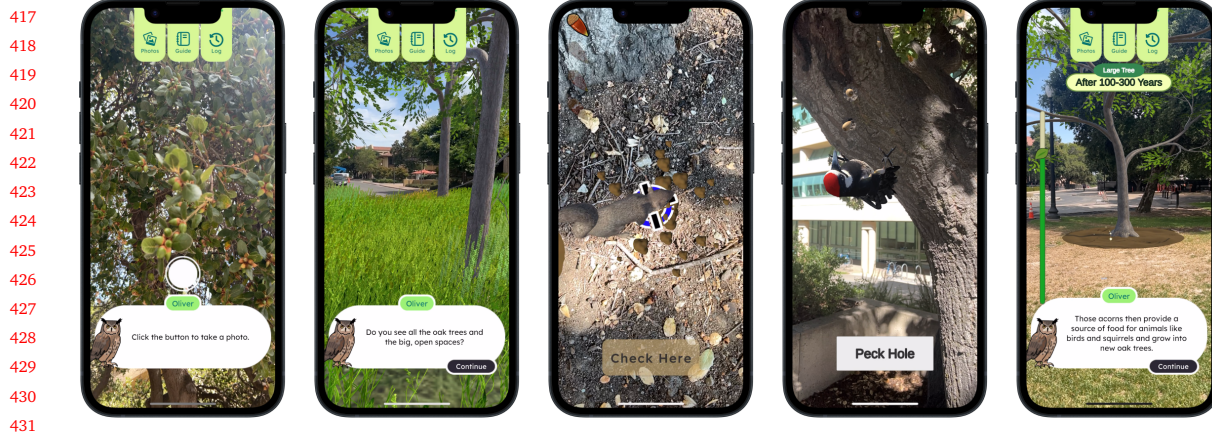


Fig. 4. Real-world activities in *Oak Story*. From left to right: Tree Identification, Portal, Squirrel, Woodpecker, Oak Lifecycle. See Table 2 for descriptions.

To begin, Oliver the owl first tells the learner to go outside and find a nearby tree. Then, Oliver provides a description of oak leaves and asks learners to describe whether the leaves on the tree match the description, as well as take a picture of the leaves. Oliver also asks the learner to look for acorns, which are unique to oak trees, and may request a picture of acorns. Finally, Oliver asks the learner to synthesize their observations and come to a judgment on whether the tree is an oak or not. If not, the learner is prompted to find another tree and try again.

This activity is purposely designed to position the learner as the decision-maker, rather than having Oliver (i.e., the AI) identify the tree's species on the learner's behalf. Oliver instead serves as a facilitator.

*Portal.* In the *Portal* activity, the learner places a AR portal in their environment and walks through it to be transported into a 3D environment of a historical oak woodland. This activity is accompanied by an explanation from Oliver on the long history of native oaks in the local area, how Indigenous peoples used and lived with the oaks, and the negative impact of colonial settlement on the oak population.

*Squirrel.* The *Squirrel* activity is launched in the scene where the learner meets Stan the squirrel. Learners use an AR crosshair in their environment to identify places where an acorn may have been buried to help Stan the squirrel locate them. An AR version of Stan the squirrel will run to the user-selected spot and attempt to dig up a virtual acorn. Whether or not an acorn is found is determined by chance, with the probability of success increasing with each attempt. This activity uses Niantic Lightship's semantic segmentation library for the detection of dirt and grass, prompting the learner to move towards natural ground if they attempt to dig elsewhere. The learning goal of this activity is to reinforce the concept that squirrels engage in caching behavior for survival.

*Woodpecker.* The *Woodpecker* activity is launched after the learner has met all of the other characters and has reached the scene for rehoming Tina the oak titmouse. In this activity, learners use an on-screen crosshair to place Andy the acorn woodpecker using AR on the real-life oak tree they previously found. Andy will peck holes to store the acorns that learners obtained from helping Stan the squirrel, culminating in pecking a larger

469 hole that Tina the oak titmouse flies into. This activity uses Niantic Lightship’s semantic segmentation library to  
 470 detect foliage and prompts the learner to choose a different location if it does not detect that the crosshair is  
 471 pointed at a tree. The learning goal of this activity is to help learners visualize how woodpeckers’ specialized  
 472 beaks help them survive.  
 473

474 *Oak Lifecycle*. In the *Oak Lifecycle* activity, the learner measures out a canopy space of at least 25 feet in their  
 475 environment, plants an virtual acorn, and watches an AR visualization of an oak tree’s growth through each  
 476 stage of the oak tree life cycle, from seed to sprout to sapling to young tree to great oak. The measurement  
 477 component of this activity uses real-world distance estimation, implemented with raycasting and iPhone Lidar  
 478 capabilities.  
 479  
 480

Activity	Description
Tree Identification	The learner receives pictures of oak leaves and acorns, and sends pictures of a tree they’re looking at to confirm if it is an oak.
Portal	Oliver opens a portal, showing the local oak woodlands 200 years ago.
Squirrel	Stan digs for acorns that he’s buried nearby after the learner points at dirt or grass.
Woodpecker	Andy pecks holes into a tree trunk for acorns and for Tina’s home
Oak Lifecycle	The learner plants an acorn and watches it grow from seed to oak tree.

481  
482  
483  
484  
485  
486  
487  
488  
489 Table 2. Descriptions of the real-world activities in *Oak Story*.

## 490 491 492 4 ACORN ARCHITECTURE

493 To enable the learner-driven features of *Oak Story*, we designed and built ACORN. ACORN uses a large language  
 494 model (LLM) to provide a conversational interface between a learner and a collection of narrative educational  
 495 materials supplied by a learning experience designer. ACORN uses character messages to provide learners  
 496 with information about lessons and available actions. In response, learners express their intentions in natural  
 497 language, and ACORN converts these intentions into system functions (see Figure 5) to advance the state of the  
 498 narrative. In this section, we elaborate on ACORN’s architectural details.  
 499  
 500  
 501

### 502 4.1 Story Scenes and State

503 To create an experience like *Oak Story*, a learning experience designer first supplies a narrative in the form of  
 504 distinct story scenes. Each scene necessarily includes a high-level description and lists of goals, actions, and  
 505 characters. Designers can also include activities that occur outside of the conversations with characters, like  
 506 an activity for planting a tree in AR. Scenes can thus be created to center particular learning objectives and  
 507 activities.  
 508  
 509

510 One example from *Oak Story* is the scene in which “the student is introduced to Stan the squirrel.” This short  
 511 description of the scene is provided to the LLM, alongside a list of goals that is updated whenever a goal is met.  
 512 The goals in this scene include “Stan the squirrel has mentioned his huge stash of acorns” (an event related  
 513 to the narrative), “Stan or Oliver describes the symbiotic relationship between oaks and squirrels” (a learning  
 514 goal), and “The student has helped Stan collect acorns” (a goal related to a real world activity). The scene also  
 515 includes the actions “COLLECT\_ACORNS”, which launches the Unity activity, and “LEAVE\_STAN”, which exits  
 516 the scene and returns to the central area. Finally, the scene contains a list of characters present, in this case  
 517 Oliver and Stan.  
 518  
 519

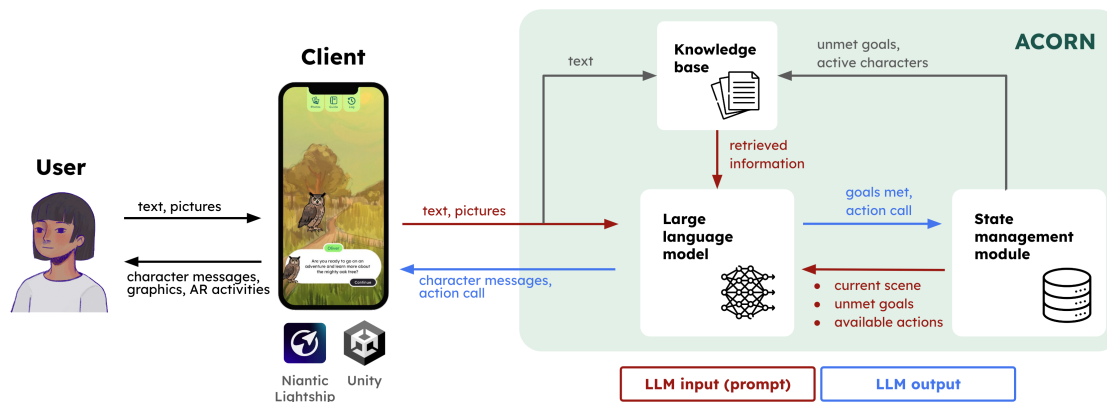


Fig. 5. ACORN employs a large language model for ‘translation’ between the user and the narrative system. The narrative state provides the LLM with information about the learner’s options, as well as some learning goals that need to be fulfilled. These are converted into messages from characters in the story, who convey educational information to complete a learning goal or inform the learner about the actions available to them. Likewise, the learner’s message is interpreted for any goals that were met or any actions that were requested, and translated into the appropriate system function to advance the narrative.

Internally, ACORN’s state management module holds information about the current state of the narrative, including the current scene, unfulfilled goals derived from the scene, available and upcoming actions, and a LLM-generated summary of prior narrative events. The state is updated by parsing the LLM response for goals met and action calls.

At the start of the aforementioned example scene, the state would contain the list of goals and update this list as goals are completed. It would also provide the LLM with “COLLECT\_ACORNS” as an action that is available at the start of the scene, and list “LEAVE\_STAN” as an action that will become available once all goals in the scene are met. If the student had completed other scenes prior to visiting this one, their actions and conversations in those scenes would be summarized and provided with other state information.

## 4.2 Goals

To produce the desired learning outcomes, ACORN prompts the language model at each timestep (a user message-character messages pair) with a list of incomplete goals and instructions to meet all goals. To invite the learner’s continued engagement, the characters’ responses should typically contain a question for the learner to answer, so goals often serve the purpose of guiding the generation of questions. For example, when the learner returns to the ‘central area’ after completing a scene, one goal is “Oliver has asked the student to reflect on what they learned from the most recent activity.”

Notably, the order in which goals are met within a state is not deterministic: the conversation history is used to select the goal that is most appropriate for each exchange. Additionally, narrative coherence requires that some goals are contingent on the completion of others. After receiving a user’s message and updated state information, the LLM identifies which goals were met, and the response is used to update the state. This enables dependencies: certain goals within a scene may be best presented in a logical order, so new goals can be added to the list of unmet goals once a prerequisite is completed.

Parameter	Purpose
Description	Information about the functionality and appropriate usage of the action
State Functions	Used by the state management module to switch scenes, add/remove actions, add/remove characters, and/or update narrative prerequisites
System Message	A message inserted into the LLM context to provide feedback that the action was successfully called, including a description of the real world activity for external actions
Example Usage	Example(s) of action usage, inserted into the prompt when the action is available and used for few-shot learning

Table 3. Components of an action in ACORN.

### 4.3 Actions

Actions in ACORN bridge the narrative world and the real world, and are the means by which the story progresses. Each action is defined by a name and a required description, which is passed to the LLM and used to ensure appropriate action calling. This description acts as a guide for the LLM to understand when and how to use each action within the narrative context. Furthermore, an action contains a set of state functions that tell the state management module which scenes the action should affect and how. Each action also includes two additional parameters that improve their integration with the narrative, shown in Table 3: a system message providing feedback to the LLM and an example of when the action should be called for use in few-shot prompting[6].

Actions can be broadly categorized into two types: story and external. Story actions progress the narrative by traversing between scenes, such as moving from the ‘central area’ to meeting a new character. They can also be used to handle prerequisites by preventing the user from accessing the story’s conclusion until a set of other scenes have been visited. For instance, in *Oak Story*, the learner is unable to rehome Tina the oak titmouse until they have met all the animals and identified an oak tree. The story actions that exit those scenes also update a global list of prerequisites. Then, in the rehoming scene, the list of prerequisites is passed to the LLM to inform the character feedback on whether the learner can successfully create a new home for Tina. External actions interact with the real world through launching activities that appear on the frontend client—in the case of *Oak Story*, this comprises AR activities and the sending and receiving of pictures. Upon completion of an external activity, ACORN inserts a educational content creator-supplied description of the activity into the LLM’s context via the system message. As a result, the learner’s real-world actions appear to be known to the characters.

Certain actions can also be both internal and external: *Oak Story*’s portal scene uses both AR and a distinct scene with goals related to history to communicate, visually and conversationally, how the local area used to look when it was a lush oak woodland. In this case, the description of the activity is provided within the scene and passed to the LLM for generating appropriate character dialogue about what the learner sees in AR.

**4.3.1 Suggesting and Taking Actions.** The ACORN architecture uses a process resembling function calling [39] to convert LLM “action calls” into the corresponding system functions. This process begins with the LLM being provided a list of valid actions as part of its input, alongside the learner’s message. Then, the learner’s intent is compared with the available actions to determine if an action should be called in the appropriate part of the LLM output. To ensure the integrity of the narrative and prevent unexpected behaviors, the output is validated against the actions that are available in the current state before being executed. Upon execution, the action is passed to both the state management module and to the Unity client. Simultaneously, the LLM generates

625 dialogue for the characters that provide feedback to the learner to inform them that the action was successfully  
626 taken.

627 In some cases, there may be many different available actions, such as when the learner is in the central area  
628 and able to access several other scenes. Combined with the open-ended text input, the available affordances  
629 may not be immediately obvious. To support user sensemaking, ACORN can have characters suggest actions  
630 to the learner based on the conversation history and the description of each action. In other cases, where an  
631 action must necessarily be taken to exit a scene, ACORN uses the context of upcoming actions to guide the  
632 conversation towards an appropriate conclusion. These upcoming actions, which become available once some  
633 or all goals in the scene are met, provide a natural progression to the narrative while maintaining the learner’s  
634 sense of agency.  
635  
636  
637  
638

#### 639 4.4 Character Messages

640 ACORN invites the learner into the narrative by incorporating contextual information about their responses and  
641 real-world interactions into character messages. State information like goals and available actions is delivered in  
642 the voices of characters. Likewise, the learner’s input is interpreted for relevant goals or actions and translated  
643 into the corresponding state updates. Character messages, as well as interface changes, then provide feedback  
644 that something has happened.  
645

646 Each character has a unique personality, background, and area of expertise, making the learning experience  
647 more immersive and memorable. In *Oak Story*, Oliver the owl is an inquisitive guide who is well-informed about  
648 history, but not much about oak trees in the present day; Andy is an irritable woodpecker frustrated with the  
649 competition he experiences against squirrels for acorns; Tina the oak titmouse is anguished after losing her  
650 home, showcasing the ecological impacts of habitat destruction. By having these characters interact with each  
651 other and the learner, complex concepts can be presented in a engaging way that emphasizes the connections  
652 between topics. This character-based approach not only makes the learning experience more enjoyable but also  
653 aids in comprehension by associating lesson concepts with distinct personalities [58].  
654  
655  
656  
657  
658

#### 659 4.5 Knowledge Base and Retrieval-Augmented Generation

660 To ensure that the system provides accurate information while maintaining the individuality of different  
661 characters, ACORN uses a derivative of retrieval-augmented generation that supplements a natural language  
662 description of each character with differing knowledge bases that reflect their limited knowledge based on their  
663 role in the ecosystem. The documents in each character’s knowledge base is curated by the experience designer.  
664 For instance, *Oak Story*’s Andy the acorn woodpecker is presented as an expert on the oak trees’ acorns. As  
665 such, whenever the learner is speaking to Andy, the LLM will have access to information about acorns, and  
666 will produce messages for Andy that reference this information. The information available at each time step is  
667 pertinent to the unmet goals and user messages, e.g., Andy’s knowledge base contains answers to questions  
668 like “How do oak trees disperse their acorns?” and “How did Native Americans use and manage acorns as a  
669 food source?” but only the most relevant items are retrieved. This method also supports multiple characters  
670 in a scene: at each step, the learner’s input and the text of unmet goals are embedded to retrieve relevant  
671 knowledge for each character, and specialized knowledge is attributed to the appropriate speaker.  
672  
673  
674  
675  
676

#### 4.6 LLM Prompting

We use a prompt template that is filled in at each time step with updated information to communicate the current state to the LLM. This includes the high-level scene description, a list of unmet learning goals, character descriptions and retrieved knowledge, few-shot learning examples, and details about available and unlockable actions.

The prompt also contains persistent information about the user, including their name and grade level, to support a personalized experience. To ensure that the language model’s generated text is appropriate for the learner, we employ a zero-shot prompting approach that instructs the language model to produce messages suitable for the user’s reading level. In our user studies, we choose to use the last completed grade level rather than the current grade when calibrating the reading level, for two reasons: first, the user studies were conducted at the beginning of the school year when students had just entered a new grade, and second, prior work finds that this method produces text at or above the target reading level [25].

Additionally, we apply two common best practices in prompting: first, we utilize chain-of-thought reasoning, which involves conditionally populating reasoning steps based on the current state. This approach guides the language model through a series of questions that promote correct behavior before generating the final output. The result is more consistent and logically coherent responses from the system. Second, we incorporate few-shot learning tailored to each scene. This provides the language model with examples of desired responses and interactions, further improving the alignment of its outputs with the designer’s intentions.

The complete prompt template can be found in the appendix.

## 5 EXPERIMENT

We conducted an IRB-approved study with fourth–sixth graders ( $N = 47$ ). In this section, we will describe the procedure we followed, how participants were assigned to a specific condition, our participants and recruitment approach, and our experimental measures.

### 5.1 Pilot Study

To prepare for the larger evaluation study, we ran a pilot study to collect preliminary data and debug the *Oak Story* app, the ACORN architecture, and the study procedure. During this pilot we made some adjustments to the system and to the study. First, we extended the estimated study duration from one hour to 1.5 hours to account for the time it took to use *Oak Story*, complete the pre- and post-study questionnaires, and conduct the post-study interview.

Next, we noticed pain points and frustrations from some pilot participants when using *Oak Story*. These frustrations seemed to arise more often with the youngest participants, since the experience required reading and writing for an extended period of time, so we raised the minimum eligible grade level for our study from 3rd grade to 4th grade.<sup>1</sup>

This was because the system expected the participant to respond with a message that directly meets the goal, and if the participant did not meet the goal, it was likely that the system would prompt them again with a similar question. This pattern could repeat over and over, leading to frustration. To address this, we

---

<sup>1</sup>We ran the pilot study and official study at the beginning of the school year, so 3rd graders had only just completed 2nd grade and were quite young.

729 changed some “Student does X” goals to “[Character] asks student to do X”, which reduced the burden  
730 on students.

731 Lastly, we explored ACORN’s potential for scaling to languages beyond English, namely German and Spanish  
732 and found that it seems to perform adequately.  
733  
734

## 735 5.2 Procedure

736 We conducted a between-participants study of our system in which participants completed reading and learning  
737 activities using our system, filled out pre- and post-study surveys and quizzes, and provided their feedback on  
738 the app and activities.  
739

740 Each participant was invited to come to our campus for a 90-minute individual session. We used two iPhones  
741 (an iPhone 13 Pro Max and an iPhone 14 Pro). One was used for running the system, and the other for recording  
742 the session and as a back-up. We used an iPad for pre- and post-study surveys and quizzes. Sessions started  
743 indoors in a building at our institution. We first obtained consent from the participants’ parents or guardians  
744 (adults) and assent from each participant. To obtain assent we asked participants two questions: 1) *You’ll be*  
745 *using our app to learn science. Some parts of the app will ask you to go outside. Is that okay with you?* and 2)  
746 *Before and after you use the app, we’ll ask you to answer some questions on the iPad. Also, we’ll ask you some*  
747 *questions about your experience with the app. Is that okay?* We also explained that they could stop the study at any  
748 point by letting a researcher know. Participants’ parents/guardians were also asked to fill out a demographics  
749 survey on behalf of the participant.  
750  
751

752 Participants then filled out a pre-study questionnaire and a pre-test (knowledge quiz) on the iPad before  
753 starting to use *Oak Story*. The knowledge quiz was filled out using the microphone (dictation) feature to input  
754 answers for consistency across participants, but they were allowed to correct any text by typing if needed. Not  
755 all participants were accustomed to typing on an iPad, and this helped ensure participants didn’t spend too long  
756 on the quiz.  
757  
758

759 We video-recorded sessions in which consent had been granted. When video-recording, we aimed to capture  
760 interactions while doing our best to avoid capturing the child’s face. We then started screen recording on  
761 the iPhone running our system, and handed it to the participant to begin their interactions. When the first  
762 request for user input popped up in our system, we let the participant know that they could use the microphone  
763 (dictation) feature to input their answer. Once it was time to do an AR activity, a researcher took the participant  
764 outside. When the app said to “go to the central area,” participants could return inside the building or find  
765 somewhere outside to sit and continue using the app. There were two to three researchers at each session to  
766 moderate, observe, and record the study.  
767  
768

769 Once participants were done using our app, we asked the participant if they wanted to take a break. They  
770 were then asked to complete the same knowledge quiz from the beginning in the same manner. Next, they  
771 completed a post-study questionnaire. Finally, participants completed a semi-structured interview that was  
772 audio-recorded.  
773

774 A follow-up survey and knowledge quiz<sup>2</sup> was sent a week after the in-person session to be completed at  
775 home, and participants were asked to complete it as soon as possible. Upon completing the follow-up survey,  
776 we sent them a \$25 Amazon gift card to thank them for their participation.  
777

778 <sup>2</sup>Not all follow-up surveys have been completed at the time of this submission, so we will include the complete analysis of that data in the  
779 revision.  
780

781 Two of the studies were conducted at a park within driving distance of our institution, as the families of those  
782 participants did not have the means to drive to our institution. We also used these sessions as an opportunity to  
783 test whether ACORN and *Oak Story* could operate in other locations, and because they worked appropriately, we  
784 include these studies in our analysis.  
785

### 786 5.3 Condition Assignment

787 Participants were assigned to one of two experimental groups: the control group ( $n = 21$ ), and the treatment  
788 group, or ACORN ( $n = 26$ ). We created a control condition to isolate the effects of learner agency afforded  
789 by ACORN. In the control, we replace the real-time dialogue generation with messages comprised of two  
790 parts: a LLM-generated message, and hardcoded dialogue lines generated in advance by ACORN. To keep the  
791 conversational interface the same between conditions, we still allow learners in the control condition to write  
792 open-ended responses, and a LLM generates a reply that responds to the learner while connecting it to the next  
793 hardcoded line of dialogue in the narrative. The narrative progresses linearly, and the learner is not given the  
794 choice to decide which story scenes to go to next. In other words, the experience is linear but the character  
795 messages remain responsive to the specific user input. This allows us to specifically examine the impact of  
796 ACORN in managing the interactive experience.  
797

800 We utilized systematic stratified randomization to maintain a roughly even distribution across grades and  
801 gender. Condition was first randomized by strata, each stratum being defined by a grade-gender pairing, and  
802 then alternated within each stratum between the control and treatment conditions.  
803  
804

### 805 5.4 Participants and Recruitment

806 We recruited a total of 51 participants, but two participants (P10, P27) did not finish the study due to  
807 irrecoverable bugs in the system, one participant (P50) declined to finish using *Oak Story*, and one participant  
808 (P34) declined to do the final surveys and interview.<sup>3</sup>  
809

810 This left us with a final total of 47 participants. These participants (30 male and 17 female) were in fourth ( $n$   
811 = 13), fifth ( $n = 15$ ), and sixth grade ( $n = 19$ ). Participants' parents or guardians identified them as East Asian  
812 ( $n = 16$ ), White ( $n = 15$ ), Hispanic or Latino ( $n = 6$ ), South Asian ( $n = 2$ ), Southeast Asian ( $n = 1$ ), and mixed  
813 ( $n = 7$ ). They were recruited through mailing lists, physical flyers, and word of mouth. Some mailing lists were  
814 pre-existing ones in our lab and others were gathered through connections with local schools and afterschool  
815 organizations.  
816  
817  
818

### 819 5.5 Measures

820 In this section we describe our quantitative measures: learning gains, pre-post measures, engagement, and  
821 perceptions of the characters and story. In addition, we describe our semi-structured interview and qualitative  
822 data analysis approach. Copies of our questionnaires and interview guide are included in the Supplementary  
823 Material.  
824

825  
826  
827 *5.5.1 Learning gains.* We used the knowledge quiz to measure learning gains. One question was a multiple-choice  
828 question that asked participants to select all the oak trees out of a variety of tree photos. For the remaining  
829 questions, participants were asked to verbally describe: 1) the importance of oak trees, 2) their role in today's  
830

831 <sup>3</sup>We use an excerpt from P34's in-app conversations to demonstrate adversarial inputs but otherwise exclude their data from analysis.



833 cities, 3) the benefits they can provide to people, and 4) how different parts of an oak tree’s ecosystem work  
834 together (e.g., what roles do oak trees, acorns, squirrels, and acorn woodpeckers play in the ecosystem). We  
835 had initially also included two “Apply” questions about knowledge regarding a sunbird and hickory trees that  
836 was not covered in *Oak Story*, so we omitted those from the analysis. The rubric for how the knowledge quiz  
837 responses were graded is included in the Supplementary Material.  
838  
839

840 *5.5.2 Pre-post measures.* In the pre and post tests we measured students’ motivation toward science learning  
841 (SMTSL) [50], Environmental (2-MEV) Scale questionnaire for children [26], and **interest levels**.

842 For the SMTSL, we selected and adapted six items on a five-point Likert scale (1 = strongly disagree, 5 =  
843 strongly agree) encompassing the following categories: self-efficacy, active learning strategies, and science  
844 learning value. We also used ten five-point Likert-type items from the 2-MEV Scale to measure children’s  
845 attitudes toward the environment. These items encompass preservation categories: 1) Intent of support, 2) Care  
846 with resources, and 3) Enjoyment of nature, and utilization categories: 1) Altering nature and 2) Dominance.  
847  
848

849 Finally, to measure interest levels, we used a five-point Likert scale (1 = Not interested at all, 5 = Extremely  
850 interested) to measure interest on seven items: 1) animals, 2) plants, 3) tress, 4) oak trees, 5) biology (the  
851 study of life), 6) ecology (the study of organisms and their environments), and 7) wildlife in our region.  
852

853 *5.5.3 Post-only measures.* In the post-completion questionnaire we additionally asked for feedback on *Oak Story*.  
854 To measure engagement, we adapted the seven items from Dietz, et al.’s Giggle Gauge [17] to a five-point Likert  
855 scale (1 = strongly disagree, 5 = strongly agree). We also developed a set of 12 items on a five-point Likert  
856 scale (1 = strongly disagree, 5 = strongly agree) to evaluate interactions with *Oak Story*. A set of eight items  
857 dealt with the conversations with the animals in the story, “*The characters were unique compared to each other*”  
858 or “*The characters spoke in a kind way.*” A set of four questions dealt with the story overall, such as “*I felt excited*  
859 *to find out what was going to happen next*” and “*My actions, interests, and messages directly influenced the story.*”  
860  
861  
862

863 *5.5.4 Interview.* Once participants had completed the last in-person questionnaire, they started a short, semi-  
864 structured interview about their experience. In this interview, we asked questions such as “*How was your overall*  
865 *experience with the app?*” or “*How did this experience compare to how you usually learn science?*” The interview  
866 allowed us to obtain more qualitative insights.  
867  
868

## 869 6 FINDINGS

870 First, we provide our findings from a technical evaluation of ACORN. Then, we present the results from our  
871 experimental study of *Oak Story* and ACORN, analyzing both the quantitative student outcomes collected from  
872 our questionnaires and the qualitative data gathered from our interviews.  
873  
874

### 875 6.1 Technical evaluation

876 Achieving our goal of building adaptive learning experiences requires that ACORN accurately bridges the learner  
877 and the narrative state. To that end, we coded the logs of user and character messages from all user studies to  
878 validate system behavior. In total, 1405 rows from the ACORN condition and 1060 rows from the control were  
879 analyzed, with each row containing a user message and the response from the characters, as well as any goals  
880 met, actions taken, and knowledge retrieved. Below, we provide both standalone and comparative analyses of  
881 ACORN. The codes used can be found in the appendix.  
882  
883  
884

885 We first focus on ACORN’s ability to make correct action calls when learners express interest in an action.  
 886 Our analysis finds that 98.03% of user-requested actions out of 203 total instances were successfully executed,  
 887 demonstrating the robustness of the system in correctly interpreting user intent from natural language and  
 888 translating it into appropriate state changes. This ensures smooth narrative progression, with the few errors  
 889 primarily occurring in edge cases where user input was ambiguous about their intended next step or the  
 890 requested action was unavailable. For instance, one student wanted to repeat a previous action:  
 891

892 [after the Squirrel AR activity, where the student searches for virtual acorns]

893 Oliver: [...] Do you have any more questions about oak trees or squirrels?

894 Participant: Let me hunt acorns again but harder

895 Stan: Wow, [name], you did such a great job finding those acorns! If you have any more  
 896 questions about acorns, squirrels, or oak trees, just let me know.  
 897

898 This error arose from the participant not being aware of the actions they can take. To minimize confusion  
 899 resulting from this, ACORN communicates options to the learner in the voices of characters when actions are  
 900 available or the learner expresses uncertainty. Across 26 participants interacting with ACORN, actions were  
 901 suggested 129 times, and learners took the suggested action 120 times. Since actions become available when  
 902 the prerequisite goals are completed, we also evaluate detection of goal completion. ACORN accurately marked  
 903 98.51% of goals as completed when users successfully achieved them, out of 875 total goals completed. This  
 904 indicates that ACORN effectively tracks user progress through the narrative and learning objectives.  
 905

906 ACORN also contains a knowledge base with curated educational materials, sections of which were delivered  
 907 to the LLM for incorporation into character responses when any sections met a threshold of cosine similarity  
 908 to the user input or unmet goals. This threshold was met, and consequently retrieval was used, in 40.00% of  
 909 LLM-generated responses. Out of these 562 instances, character messages incorporated externally validated  
 910 knowledge to provide accurate information to the learner in 64.59% of responses, demonstrating effective  
 911 integration of the knowledge base into the dialogue.  
 912

913 Next, we conducted comparative analyses between the two versions of *Oak Story*. Both versions are able to  
 914 answer student questions and provide specific responses due to the use of a LLM in the control. Our first analysis  
 915 showed that 100.00% of 39 total questions posed by learners interacting with ACORN were adequately answered,  
 916 compared to 76.47% (13 of 17) in the control, indicating a strong ability to provide relevant information in  
 917 response to user queries. Furthermore, 58.97% of answers from ACORN were supplemented with relevant  
 918 retrieved knowledge, reducing the likelihood of LLM hallucination by incorporating curated information.  
 919

920 Additionally, coders were instructed to label messages as “specific” if they directly quoted or referenced the  
 921 student’s input and “generic” if the response failed to acknowledge the student’s input at all, for instance:  
 922

923 (specific)

924 Andy: [...] What do you think squirrels do to get their acorns?

925 Participant: I think they climb the oak tree.

926 Oliver: That’s right, [name]! Squirrels are great climbers and use their strong legs and claws to  
 927 scurry up oak trees quickly [...]

928 (generic)

929 Oliver: [...] What do you think is the most interesting thing you’ve learned?

930 Participant: That animals have always lived in oak trees  
 931

937 Oliver: It’s fascinating to see what this area used to look like with so many oak trees. They  
 938 played a big role in the ecosystem [...]  
 939

940 Messages that were not flagged as ‘specific’ nor ‘generic’ comprised 79.15% of responses in the ACORN version  
 941 of *Oak Story*, and 83.21% of responses in the control. Messages that respond to short inputs from the student,  
 942 like “yes” and “okay,” comprise the majority of these neutral messages.  
 943

944 We find that ACORN enables a higher degree of narrative personalization, with 20.36% of characters’ responses  
 945 tailored specifically to the learner’s input compared to 10.85% in the control. Characters respond to the learner’s  
 946 expressed interests, prior narrative events, and real world actions nearly twice as often—similarly, the control  
 947 produced non-responsive character messages more often (5.94% vs 0.50% with ACORN).  
 948

949 These results demonstrate the effectiveness of ACORN in creating personalized, responsive learning experi-  
 950 ences. The high accuracy in action execution and goal tracking ensures that the narrative progresses coherently  
 951 based on user choices. The remaining metrics provide evidence that ACORN successfully manages the complex  
 952 task of concurrently interpreting user input, updating the narrative state, and generating responsive character  
 953 responses. In the next section, we describe how this capability enhances learning experiences and outcomes for  
 954 learners.  
 955

## 957 6.2 Student Outcomes

958 We gathered data on student outcomes from our pre-test and post-test, as well as the pre- and post-study  
 959 questionnaires. For pre-post measures, we compare whether there was a shift from pre- to post-, and if there  
 960 was, we tested for a difference between the ACORN and the control conditions. For post-only measures, we test  
 961 for differences between conditions only.  
 962  
 963  
 964

	Quiz Score Mean (SD)		
	Pre-test	Post-test	Gain (Post - Pre)
Control ( $n = 20$ )	4.23 (2.49)	5.18 (2.49)	0.95 (2.23)
ACORN ( $n = 26$ )	3.08 (1.98)	6.15 (1.94)	3.08 (2.11)
Overall ( $N = 46$ )	3.58 (2.26)	5.73 (2.22)	2.15 (2.39)

965 Table 4. Learning gains based on the knowledge quiz, separated by condition. The sample size is 46 because we remove one outlier  
 966 who scored greater than 2.5 standard deviations above the mean in the pre-test (the outlier was assigned to the control condition).  
 967  
 968  
 969  
 970  
 971  
 972  
 973  
 974

975 6.2.1 *Knowledge quiz.* The answers to the knowledge quiz were graded by two researchers who were blind to  
 976 the participants’ condition and to whether the responses were from the pre-test or post-test. They achieved high  
 977 inter-rater reliability, with an intraclass correlation coefficient (ICC) of 0.918, indicating excellent reliability [27].  
 978

979 To measure immediate learning gains after using *Oak Story*, we scored each pre-test and post-test (range:  
 980 [-2, 14]).<sup>4</sup> For our analysis, we removed an outlier who scored over 2.5 standard deviations higher than the  
 981 mean on the pre-test. Summary statistics for pre-test scores, post-test scores, and gain (difference between  
 982 pre and post) are provided in Table 4. Based on a paired t-test, we observed a significant post-test increase in  
 983 scores across both conditions,  $t(45) = 6.11$ ,  $p < 0.001$ , indicating learning gains. The effect size, using Cohen’s  $d$   
 984 calculated with change score variability [14], was  $d = 0.609$ , suggesting a medium-to-large effect size.  
 985

986 <sup>4</sup>The minimum possible score is -2 because it is possible to lose up to two points by guessing on the multiple choice question on the quiz. For  
 987 more details on scoring, see the rubric in the Supplementary Material.  
 988

Furthermore, we compared the difference in learning gains between the ACORN and control conditions. We conducted a one-way ANCOVA [38] and found a significant effect of the condition on the post-test score after controlling for the pre-test score,  $F(1, 43) = 7.25$ ,  $p = 0.010$ . The effect size, measured with the partial eta-squared statistic, was  $\eta^2 = 0.144$ , indicating a large effect [12]. This finding suggests that **ACORN led to increased learning gains** compared to the control.

**6.2.2 Engagement.** To assess whether there was a difference in engagement between the ACORN and non-ACORN groups, we sum the items in our modified Giggle Gauge scale for each participant (range: [0, 30]). Only six out of seven items in the scale are summed because one item was mistakenly excluded from the survey for early participants. We used an unpaired t-test to compare the results and found that engagement in the ACORN group ( $M = 26.1$ ,  $SD = 2.47$ ) was significantly higher than the control group ( $M = 24.2$ ,  $SD = 2.99$ ),  $t(45) = 2.41$ ,  $p = 0.020$ , suggesting that **ACORN made Oak Story more engaging** compared to the control.

Scale	Items	Range	Condition	Mean (SD)	<i>p</i>
System responsiveness	<ul style="list-style-type: none"> <li>“The characters responded to what I specifically said or asked.”</li> <li>“My actions, interests, and messages directly influenced the story.”</li> </ul>	[2, 10]	ACORN	8.92 (0.89)	0.009*
			control	7.76 (1.48)	
Characters’ distinctiveness	<ul style="list-style-type: none"> <li>“The characters were unique compared to each other.”</li> <li>“I remember a lot about each character and what I learned from them.”</li> </ul>	[2, 10]	ACORN	8.96 (1.00)	0.164
			control	8.43 (0.81)	
Feelings toward characters	<ul style="list-style-type: none"> <li>“I felt happy when I read what the characters said to me.”</li> <li>“The characters spoke in a kind way.”</li> <li>“I wanted to have more conversations with the characters.”</li> </ul>	[3, 15]	ACORN	12.62 (1.96)	0.908
			control	12.19 (1.86)	
Connection to real-life activities	<ul style="list-style-type: none"> <li>“Seeing the characters in activities (where they interact with the real world) made me more interested in talking to them.”</li> <li>“The activities (where characters interact with the real world) felt connected to the conversations I had with the characters.”</li> </ul>	[2, 10]	ACORN	9.00 (1.10)	0.908
			control	9.05 (0.97)	
Story enjoyment	<ul style="list-style-type: none"> <li>“I felt excited to find out what was going to happen next.”</li> <li>“The story made learning more fun.”</li> </ul>	[2, 10]	ACORN	9.23 (0.95)	0.088
			control	8.43 (1.36)	

Table 5. Groupings of our Likert-type items into scales. Asterisks on *p*-values indicate that learners using ACORN gave statistically significantly higher responses than learners using our control app.

**6.2.3 Conversation- and story-related survey items.** To analyze the 12 Likert-type items we developed to evaluate participants’ perceptions of the characters and story, we clustered them into groups of 2-3 similar items and summed them, resulting in five separate Likert scales with the following themes: system responsiveness,

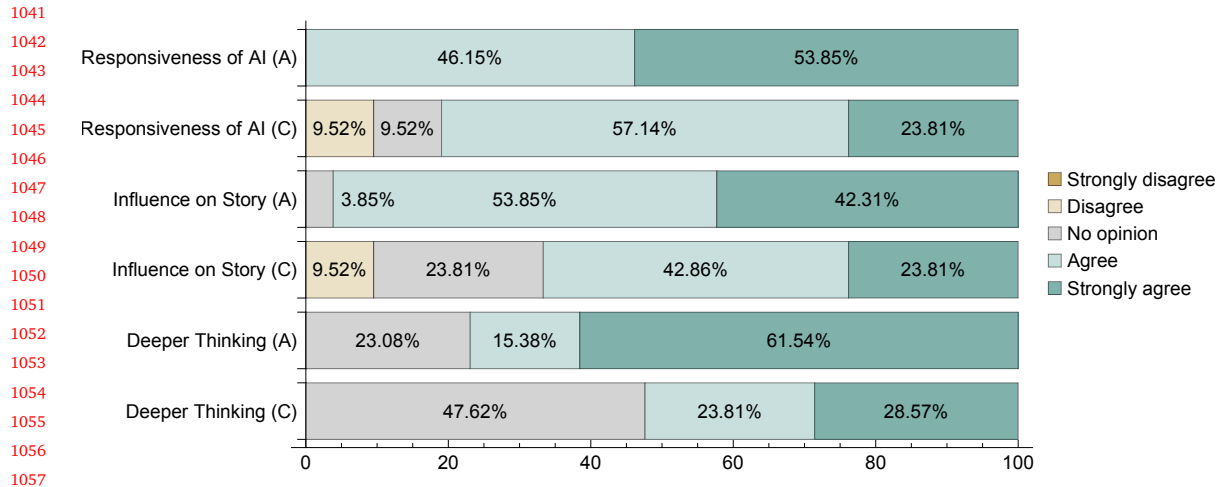


Fig. 6. The distributions of responses to 3 (out of 12) of our conversation- and story-related survey items where we observed the greatest difference between ACORN (A) and control (C).

characters’ distinctiveness, feelings toward characters, connection to real-life activities, and story enjoyment (see Table 5).

One item (“The conversations encouraged me to think deeper.”) did not fit well into any of the themes, so we omit it from hypothesis testing, as testing a single Likert item can be unreliable [8]. Instead, we simply note that the distribution of responses to this item skews much more strongly toward “Strongly agree” among those in the ACORN condition, as shown in Figure 6.

We performed an unpaired t-test on each scale and correct for multiple comparisons using the Holm-Bonferroni method. Table 5 shows the p-value for each scale after correction. We found that participants’ attitudes toward system responsiveness were significantly higher in the ACORN group compared to the group who received the control, suggesting that **characters and stories driven by ACORN feel more responsive** to learners’ inputs. The distributions for these items are also presented in Figure 6.

**6.2.4 Motivation to learn science.** We summed the items in the Student Motivation Towards Science Learning (SMTSL) scale for each participant (range: [6, 30]) and used a paired t-test to compare how participants’ scores may have changed between the pre-study survey and post-study survey. We observed a significant increase from pre ( $M = 23.3, SD = 3.28$ ) to post ( $M = 24.4, SD = 2.55$ ),  $t(46) = 3.35, p = 0.0016$ , which indicates that **Oak Story increased participants’ motivation to learn science**. The effect size, measured with Cohen’s  $d$  calculated with the change score variability [14], was  $d = 0.489$ , suggesting a medium effect size.

We did not find evidence to suggest that this change in motivation to learn science differed between the two conditions.

**6.2.5 Attitudes toward the environment.** For each participant, we summed the items in the Environmental (2-MEV) Scale questionnaire for children (range: [10, 50]) and used a paired t-test to compare for differences between pre and post. We did not find a significant difference from pre ( $M = 38.4, SD = 5.61$ ) to post ( $M = 39.0, SD = 6.43$ ).

1093 The effect size, measured with Cohen’s  $d$  calculated with the change score variability [14], was  $d = 0.161$ ,  
 1094 suggesting a small effect size.  
 1095

1096 **6.2.6 Interest in scientific topics.** Similar to the conversation- and story-related survey items, we group similar  
 1097 items together, condensing seven Likert-type items into three scales: interest in animals (animals and local  
 1098 wildlife), interest in plants (plants, trees, and oak trees), and interest in science (biology and ecology). Then, we  
 1099 performed a paired t-test on each scale to compare how scores changed from pre to post and correct for multiple  
 1100 comparisons using the Holm–Bonferroni method. We found that the scores for all three scales significantly  
 1101 increased from pre to post, indicating that **Oak Story increased learners’ interest in the topics covered.**  
 1102  
 1103

	Range	Pre	Post	$p$
Interest in animals and local wildlife	[2, 10]	7.57 (1.70)	7.98 (1.70)	$p = 0.026$
Interest in plants, trees, and oak trees	[3, 15]	10.1 (2.64)	12.0 (2.45)	$p < 0.001$
Interest in biology and ecology	[2, 10]	7.09 (1.85)	7.68 (1.87)	$p = 0.026$

1104  
 1105  
 1106  
 1107  
 1108  
 1109 Table 6. Ranges, pre-app scores, and post-app scores for participant interest in scientific topics related to *Oak Story*.

1110  
 1111  
 1112  
 1113  
 1114  
 1115  
 1116  
 1117  
 1118

We also find that the increase in interest in oak trees is significantly larger in the ACORN group ( $M = 1.12$ ,  
 $SD = 0.91$ ) compared to the non-ACORN group ( $M = 0.52$ ,  $SD = 1.08$ ) using a Mann–Whitney U test:  $Z = 163.5$ ,  
 $p = 0.014$ .

These findings suggest that not only did *Oak Story* increased participants’ interest in the topics covered, but  
 also ACORN leads to a **greater increase in interest in the topic than the control.**

1119  
 1120

### 6.3 Qualitative Observations

1121  
 1122  
 1123  
 1124

Here, we share the themes we found in the interview transcripts, including potential differences we found  
 between participants in the ACORN and control groups. We provide the condition (ACORN or Control) and grade  
 level (4th–6th) for each quoted participant. Quotes are lightly edited for clarity.

1125  
 1126  
 1127  
 1128  
 1129  
 1130

**6.3.1 ACORN supported participants’ sense of agency.** *Oak Story* allowed participants to write freeform responses  
 to the questions posed by characters in the story, allowing them more freedom to express themselves. One  
 participant explained, “I actually loved [writing responses] because **I get to type my own opinions** and stuff”  
 (P20, ACORN, 5th).

1131  
 1132  
 1133  
 1134

While participants across both groups appreciated being able to input what they wanted to say, those in the  
 ACORN group noticed that the system was giving them opportunities to make meaningful choices to shape both  
 their own learning and the direction of the story. P17 (ACORN, 5th) described their feelings as follows:

1135  
 1136  
 1137  
 1138  
 1139

Well, it asked me what I would want to do next. So it’s really generous. That helped me think  
 that **it gives me the choice and gives me the power to think about it**, and when I say what I  
 want, it really just responds with a really helpful answer, like really, it just responds exactly to  
 what I want. I felt like I had a lot of choices, and the story was really good.

1140  
 1141  
 1142  
 1143  
 1144

**6.3.2 ACORN generated dialogue that felt responsive to participants.** We noticed in the interviews that participants in  
 the ACORN group regularly emphasized that the characters responded meaningfully to their inputs. For example,  
 P14 (ACORN, 6th) remarked, “I liked how the animals took a minute to process the information, and then they

1145 actually gave me extremely useful answers, and **they knew exactly what I was asking**, and they gave me very  
1146 perfect answers. **So then I learned better.**”

1147 The precision of ACORN in responding to inputs not only entertained participants, but it also motivated  
1148 them to learn more. Another participant expressed similar feelings of increased learning due to ACORN’s strong  
1149 responses, saying, “I like that the characters are really responsive, and the responses are actually pretty nice,  
1150 like, they can easily take what I respond and transform it into like, oh, normally talking to a real person. [...] **I**  
1151 **hate science, but I can learn a lot better from this**” (P6, ACORN, 6th).

1152  
1153  
1154 Meanwhile, participants in the control group sometimes expressed reservation about whether the system  
1155 understood them: “Well, **sometimes it doesn’t understand my response very well**” (P54, Control, 4th).  
1156 This suggests that, even though the control version of *Oak Story* also used an LLM to respond to participants’  
1157 messages, its inability to meaningfully act on those messages hurt the overall experience.  
1158

1159  
1160 **6.3.3 ACORN was able to handle adversarial inputs.** While the vast majority of participants engaged earnestly with  
1161 *Oak Story*, a few participants tested the limits of the system with adversarial inputs. ACORN’s guardrails could  
1162 handle these situations, either by ignoring the input altogether (if they were non sequiturs) or by acknowledging  
1163 the participant’s input and steering the conversation back on track. We provide an example conversation where  
1164 this occurred:  
1165

1166  
1167 Oliver: Now that we’re back at the central area, we have two things we could do. We could  
1168 look for a tree and see if it’s an oak, or we could start thinking of how to help Tina find a new  
1169 home. What would you like to do next?

1170 P34 (ACORN, 5th): Tona Home gib her boba and in n out [a fast food restaurant]

1171 Oliver: That’s a fun idea, [name]! **While boba and In-N-Out might make us happy, Tina**  
1172 **needs a new home more than a snack.** We can help by gathering ideas from our animal  
1173 friends, or we can explore a nearby tree and see if it’s an oak. What do you think we should  
1174 do?  
1175  
1176

1177  
1178 **6.3.4 Participants enjoyed the outdoor, real-world activities.** Almost all participants across both conditions expressed  
1179 that they enjoyed the real-world activities. Many participants contrasted *Oak Story* with their more repetitive  
1180 experiences learning in school. P38 (Control, 6th) remarked, “I liked that you had to go outside. [...] **Because**  
1181 **that way you actually get to do something, it would help you focus more.** Like fidgeting with something, like  
1182 doing something, not just staring at a textbook.” Specifically, they appreciated that active interaction replaced  
1183 idleness; one observed, “It included a lot of physical stuff instead of just sitting around and not doing anything  
1184 and just watching it instead of interacting with it” (P39, Control, 4th)  
1185

1186 Participants also positively compared *Oak Story* to other mobile apps, noting that the outdoor portions  
1187 prevented boredom and made the app more challenging. P56 (ACORN, 6th) emphasized, “Well, [going outside]  
1188 is more interesting. It’s like, the other games, you’re usually just like sitting on your computer in the classroom.  
1189 [...] I liked how you had to walk around outside, scavenging around, trying to find things like acorns and  
1190 oak trees.” Another participant described how *Oak Story* prevented sedentary behavior, a common problem in  
1191 children when using screens: “**I feel like not enough apps really take into consideration that kids shouldn’t**  
1192 **be just glued onto the screen;** even if they’re learning something they should also be able to go outside and  
1193 interact with the real world” (P43, ACORN, 6th).  
1194  
1195  
1196

1197 6.3.5 *Participants appreciated that the narrative and characters connected to the real world.* Participants enjoyed the  
1198 narrative-based learning environment, feeling it bolstered their engagement and prevented the experience  
1199 from becoming stale. Several participants noted a sense of accomplishment after successfully rehoming the  
1200 oak titmouse, with one commenting, “I think the story definitely made it more interesting. **I probably would**  
1201 **have gotten bored if there wasn’t a story** [...] It did feel good that I saved a tiny bird” (P24, Control, 4th).  
1202 Furthermore, the presence of characters with distinct personalities made participants eager to see their responses  
1203 to the world around them. P40 (Control, 6th) noted, “It was just like fun to interact and see how the characters  
1204 would also react with, like, how everything was out here.”  
1205  
1206

1207 6.3.6 *Participants valued ACORN’s context-awareness and wanted more.* In certain scenarios, the open-ended prompt-  
1208 ing allowed participants to make requests or observations that the system was not fully equipped to handle due  
1209 to limited contextual awareness. One such example came with the final rehoming activity, where participants  
1210 would ask the acorn woodpecker to peck a new home for the titmouse in the oak tree they’d previously identified.  
1211 In certain scenarios, participants recognized holes that were already present on the tree as a potential home for  
1212 Tina. The app was not advanced enough to be aware of this and was unable to effectively respond to these  
1213 observations.  
1214

1215 While the freedom afforded to participants allowed limitations like these to be exposed, these unexpected  
1216 outcomes indicate the variety of responses among participants and the natural manner with which they  
1217 interacted with the app. In other words, ACORN encouraged them to be creative and think for themselves,  
1218 to the extent that they made unexpected discoveries. While these limitations hindered response precision in  
1219 places, they generally did not cause major issues and participants were able to proceed without much difficulty.  
1220 Occasionally, participants would extrapolate from one of the app’s context-unaware moments and assume it  
1221 was context-aware. For example, P42 (ACORN, 6th) prescribed an ecological meaning to the randomness of the  
1222 acorn-digging activity: “I liked all [the parts of the app] but finding the acorns was kind of challenging. [...]”  
1223 And then **I remembered that squirrels usually bury their acorns by oak trees**. So I went to the oak tree, and  
1224 they’re all in the same— they’re all really close to each other.”  
1225  
1226  
1227  
1228  
1229

## 1230 7 DISCUSSION

1231 Through the process of designing, developing, and evaluating both ACORN and *Oak Story*, we found evidence  
1232 that *Oak Story* not only effectively facilitates an interactive narrative that ties to the real world, but also supports  
1233 children’s learning and engagement. Here, we synthesize our findings and discuss the broader implications of  
1234 our work.  
1235  
1236

### 1237 7.1 ACORN improves learning and engagement by giving learners agency

1238 Our quantitative findings support the conclusion that ACORN improved both children’s learning and engagement  
1239 when using *Oak Story*. We believe these large increases are due to ACORN giving learners greater *agency* over  
1240 their learning experience.  
1241

1242 Our study contributes an additional perspective to the growing body of research investigating the impact of  
1243 AI (and LLMs) on learners’ agency. AI assistance can rob learner’s agency by encouraging overreliance [15], but  
1244 AI-driven systems can also support agency by empowering learners [62]. We provide an example of the latter,  
1245 where ACORN offers a way to take advantage of LLMs’ affordances to accept and make sense of open-ended  
1246



1249 input from the student to shape their learning experience. In the real world, the space of actions that the learner  
1250 can take is broad—ACORN supports agency by giving learners the same sense of open-endedness, while pairing  
1251 the learner’s input with the narrative state to translate their intent into narrative progression and characters’  
1252 reactions.  
1253

## 1254 1255 **7.2 ACORN effectively facilitates an interactive, narrative-driven educational experience** 1256

1257 When designing interactive narratives, where the user makes choices to impact the story, there is a central  
1258 tension between coherent story progression and the user’s agency [42]. When these narratives also need to be  
1259 educational, another constraint emerges: the interactive story must now support a set curriculum and allow the  
1260 learner to achieve certain learning objectives. ACORN complements recent work exploring how LLMs can enable  
1261 interactive narrative experiences (e.g., [29]) by exploring how we might make user-driven interactive stories  
1262 possible within the constraints of an educational curriculum.  
1263

1264 For *Oak Story*, we prioritized learner autonomy within a narrative about oak trees, their ecosystems, and  
1265 animal inhabitants. Our goal was to empower students to make impactful choices that shape the story while  
1266 ensuring educational objectives were met. We structured the experience around key scenes, each with specific  
1267 goals expressed in natural language. These goals are designed to be flexible, allowing for both system-driven  
1268 elements (e.g., "Oliver explains how oak trees benefit humans and cities") and learner-initiated actions  
1269 (e.g., "The student proposes a method to protect oak trees").  
1270  
1271

## 1272 1273 **7.3 ACORN enables narrative-based learning coupled with the learner’s physical environment** 1274

1275 While narrative on its own is an effective way to engage learners, we found in the post-study interviews that  
1276 participants tended to highlight the real-world augmented reality activities as their favorite. We also observed  
1277 from the interviews that some participants enjoyed the fact that events in the narrative, such as finding Tina  
1278 the oak titmouse a new home, connected directly to the real-world activities, suggesting that engagement  
1279 through narrative and engagement through physical exploration could have an amplifying effect when employed  
1280 together.  
1281

1282 Finally, we also note that the deeper the narrative is entwined with the real world, the expectations of  
1283 learners may also increase with it. For example, because finding Tina a new home requires the learner to create  
1284 a (virtual) hole in a (real) oak tree, some participants, upon seeing a real hole in a tree, hoped that they could  
1285 put Tina in that hole and were disappointed that they could not.  
1286  
1287

## 1288 1289 **7.4 ACORN takes a step toward personalized learning at scale**

1290 We designed ACORN to work in a variety of contexts and to facilitate development of new experiences similar  
1291 to *Oak Story*. We found that ACORN is scalable through successful implementation in two distinct locations,  
1292 providing evidence that the underlying models are not overfitted to a particular environment or context. We  
1293 also explored the scalability of ACORN by pilot testing it with different languages, e.g., German and Spanish,  
1294 demonstrating its ability to generate coherent and engaging narratives for linguistic contexts beyond the  
1295 experience designer’s. This adaptability can thus enable educators to create inclusive tools that serve diverse  
1296 student populations. Additionally, ACORN enables learning experience designers to generate dialogue from a set  
1297 of high-level descriptions, making it easier to iterate on educational content and modify scenes in the narrative.  
1298  
1299

## 1301 8 LIMITATIONS AND FUTURE WORK

1302 While ACORN demonstrates significant potential in enhancing learner agency in narrative-based education, we  
1303 identified limitations that open up exciting areas for future research.  
1304

1305

### 1306 8.1 Technical Limitations

1307

1308 Several of the limitations we identified are technical in nature.

1309

1310 *8.1.1 Language model constraints.* At the core of ACORN is a large language model, which, while powerful,  
1311 introduces a set of complex challenges. The use of prompting techniques like chain-of-thought reasoning [55]  
1312 and few-shot learning [6] improves the consistency of outputs but comes with tradeoffs. Chain-of-thought  
1313 reasoning, for instance, increases the coherence of the model’s responses but produces more output tokens,  
1314 potentially introducing latency that can disrupt the flow of interaction. Similarly, few-shot examples, while  
1315 helpful in guiding the model’s behavior, risk overfitting to specific scenarios, potentially limiting the system’s  
1316 flexibility.  
1317

1318

1319 *8.1.2 Multimodal and extended reality integration.* In its present form, ACORN utilizes AR activities as complemen-  
1320 tary components to the text-based narrative interactions. These AR elements offer students the opportunity  
1321 to engage with concepts introduced in the story through immersive, real-world interactions. However, the  
1322 connection between the language model driving the narrative and the AR experiences remains somewhat  
1323 disconnected. The language model receives a text description of what the student sees in AR during the activity  
1324 or after its completion, but it lacks real-time insight into the student’s actions and experiences during the  
1325 interaction.  
1326

1327

1328 Exploring the continued use of multimodal language models could enable real-time processing of visual and  
1329 interactive data from AR activities. This would allow the narrative to adapt dynamically based on the student’s  
1330 actions and discoveries in the AR environment. For instance, if a student struggles with a particular AR task,  
1331 characters could immediately offer guidance or adjust the narrative to provide additional context.  
1332

1333

1334 Developing mechanisms for bidirectional communication between the AR components and the language  
1335 model could also create a more cohesive learning experience—currently, ACORN can launch activities on the  
1336 client but does not receive any synchronous data about it. Improvements in this area might involve designing  
1337 protocols for the AR system to send real-time updates to the language model, allowing it to incorporate  
1338 immediate feedback or observations into the narrative flow. Advancing the LLM’s ability to interpret the  
1339 semantic meaning of AR interactions could enhance the educational value of these experiences. This might  
1340 involve integration of existing semantic segmentation capabilities in Niantic Lightship or methods for the  
1341 language model to influence or generate AR content in real-time.  
1342

1343

### 1344 8.2 Pedagogical Limitations

1345

1346 Our implementation also has some pedagogical limitations.

1347

1348 *8.2.1 Limitations of our study.* Our experimental study was conducted at our institution’s campus, which is located  
1349 in an relatively affluent area with a skewed population distribution relative to national (and international)  
1350 demographics. Although we made great efforts to recruit widely across nearby cities and counties with different  
1351

1352 Manuscript submitted to ACM

1353 demographic makeups, the children in our sample may not be representative of the broader population. For  
1354 example, none of our participants identified as Black.

1355 Also, lower-income families within driving distance did not always have the means to drive over to our  
1356 institution to participate in the study. While we were able to run two sessions in families' neighborhoods for this  
1357 current study, we hope to conduct future studies closer to individual families now that we have confirmed that  
1358 *Oak Story* and ACORN are flexible enough to operate in alternative locations.  
1359  
1360

1361 *8.2.2 Accuracy, safety, and transparency.* The application of language models in children's education amplifies  
1362 concerns about accuracy, bias, and safety [28, 56]. The potential for model hallucinations or inconsistencies [33,  
1363 56] becomes particularly concerning when students may not fully understand that they are interacting with  
1364 a tool prone to these errors. Interviews with participants revealed that several students believed they were  
1365 communicating with a human or a system capable of real-time web searches, highlighting the need for  
1366 additional safeguards or transparent communication about the nature of the interaction without breaking  
1367 immersion. Exploring techniques to improve the factual accuracy of language model outputs, such as improved  
1368 retrieval-augmented generation or fact-checking mechanisms, could enhance the educational reliability of the  
1369 system.  
1370  
1371  
1372

1373 *8.2.3 Understanding learner needs.* Adapting to the diverse needs of learners remains a complex challenge. While  
1374 ACORN aims to cater to varying grade levels and engagement styles, the nuanced decision-making required to  
1375 balance challenging students and allowing uncertainty is an area requiring further refinement. The system's  
1376 ability to make effective judgements in this regard impacts both the learning outcomes and the student's overall  
1377 experience.  
1378

1379 Future research might focus on advancing the adaptive capabilities of ACORN through improved learner un-  
1380 derstanding, such as capturing and storing 'memories' about the student's learning preferences and progress over  
1381 time. Additionally, investigating methods for real-time assessment of student engagement and comprehension  
1382 could accompany the existing self-assessment, further supporting learner autonomy and growth.  
1383  
1384

1385 *8.2.4 Educator involvement and authoring tools.* The current iteration of ACORN requires substantial involvement  
1386 from educational practitioners in shaping the learning experience. Moreover, the sensitivity of language models  
1387 to minor variations in prompts poses a challenge for non-expert users in eliciting desired behaviors consistently.  
1388

1389 To address these limitations, future work might involve the development of user-friendly authoring tools  
1390 that empower educators to create and customize narrative-based learning experiences with minimal technical  
1391 expertise. Drawing inspiration from platforms like Twine [19], these tools could provide intuitive interfaces for  
1392 defining character personalities and knowledge, setting educational goals, and structuring narrative flows.  
1393  
1394

## 1395 9 CONCLUSION

1396 In this work, we presented the design of ACORN, an LLM architecture that empowers students to direct their own  
1397 learning in an interactive, narrative-driven experience. ACORN facilitates and interprets learners' conversations  
1398 with story characters, uses this input to guide story progression, and seamlessly extends the narrative into the  
1399 learner's real-world environment at opportune moments.  
1400

1401 We implemented ACORN in *Oak Story*, a mobile application teaching oak woodland ecology, and performed  
1402 a user study to demonstrate that ACORN significantly enhances learner agency compared to a control. While  
1403  
1404

both versions feature LLM-generated responsive dialogue, the same overarching storyline, and identical multimodal/AR activities, ACORN's unique use of a LLM as two-way conduit between learners and the story allows for meaningful influence on the narrative.

Learners using ACORN-powered narratives not only showed increased engagement and learning gains but also reported a greater sense of ownership over their learning. They appreciated the ability to steer the narrative, ask questions that mattered to them, and explore topics aligned with their personal interests. Enhanced agency, alongside the real-world connection, appeared to deepen their connection to the material and motivate more active participation in the learning process.

Our findings underscore the potential of adaptive, learner-driven narratives in education and demonstrate how new technologies can be leveraged to nurture learners' natural curiosity about the world around them. This approach represents a step forward in supporting active, curious, and self-directed learners who are eager to explore and understand the world beyond the classroom.

## REFERENCES

- [1] Lorin W. Anderson and David R. Krathwohl (Eds.). 2001. *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives* (2 ed.). Allyn & Bacon, New York.
- [2] Albert Bandura. 2001. Social cognitive theory: An agentic perspective. *Annual review of psychology* 52, 1 (2001), 1–26.
- [3] Albert Bandura. 2006. Toward a psychology of human agency. *Perspectives on psychological science* 1, 2 (2006), 164–180.
- [4] Brigid Barron and Philip Bell. 2015. Learning environments in and out of school. In *Handbook of educational psychology*. Routledge, 337–350.
- [5] Stephen D Brookfield. 2009. Self-directed learning. In *International handbook of education for the changing world of work: Bridging academic and vocational learning*. Springer, 2615–2627.
- [6] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Ma teusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. *ArXiv abs/2005.14165* (2020). <https://api.semanticscholar.org/CorpusID:218971783>
- [7] Rodger W Bybee. 1990. Science for life & living: An elementary school science program from biological sciences curriculum study. *The American Biology Teacher* 52, 2 (1990), 92–98.
- [8] James Carifio and Rocco J Perla. 2007. Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of social sciences* 3, 3 (2007), 106–116.
- [9] John Chen, Xi Lu, Yuzhou Du, Michael Rejtig, Ruth Bagley, Mike Horn, and Uri Wilensky. 2024. Learning Agent-based Modeling with LLM Companions: Experiences of Novices and Experts Using ChatGPT & NetLogo Chat. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). <https://api.semanticscholar.org/CorpusID:267320240>
- [10] Alan Y Cheng, Meng Guo, Melissa Ran, Arpit Ranasaria, Arjun Sharma, Anthony Xie, Khuyen N Le, Bala Vinaithirthan, Shihe Luan, David Thomas Henry Wright, et al. 2024. Scientific and Fantastical: Creating Immersive, Culturally Relevant Learning Experiences with Augmented Reality and Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–23.
- [11] Alan Y Cheng, Jacob Ritchie, Niki Agrawal, Elizabeth Childs, Cyan DeVeaux, Yubin Jee, Trevor Leon, Bethanie Maples, Andrea Cuadra, and James A Landay. 2023. Designing Immersive, Narrative-Based Interfaces to Guide Outdoor Learning. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–22.
- [12] J. Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- [13] Yuhao Dan, Zhikai Lei, Yiyang Gu, Yong Li, Jia-Peng Yin, Jiaju Lin, Linhao Ye, Zhiyan Tie, Yougen Zhou, Yilei Wang, Aimin Zhou, Zeyang Zhou, Qin Chen, Jie Zhou, Liang He, and Xipeng Qiu. 2023. EduChat: A Large-Scale Language Model-based Chatbot System for Intelligent Education. *ArXiv abs/2308.02773* (2023). <https://api.semanticscholar.org/CorpusID:260681803>
- [14] Scott J Dankel and Jeremy P Loenneke. 2021. Effect sizes for paired data should use the change score variability rather than the pre-test variability. *The Journal of Strength & Conditioning Research* 35, 6 (2021), 1773–1778.
- [15] Ali Darvishi, Hassan Khosravi, Shazia Sadiq, Dragan Gašević, and George Siemens. 2024. Impact of AI assistance on student agency. *Computers & Education* 210 (2024), 104967.
- [16] Edward L Deci and Richard M Ryan. 2008. Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne* 49, 3 (2008), 182.

Manuscript submitted to ACM

- 1457 [17] Griffin Dietz, Zachary Pease, Brenna McNally, and Elizabeth Foss. 2020. Giggle gauge: a self-report instrument for evaluating children's  
1458 engagement with technology. In *Proceedings of the Interaction Design and Children Conference*. 614–623.
- 1459 [18] Lena Ballone Duran and Emilio Duran. 2004. The 5E instructional model: A learning cycle approach for inquiry-based science teaching.  
1460 *Science Education Review* 3, 2 (2004), 49–58.
- 1461 [19] Interactive Fiction Technology Foundation. 2024. Twine: An open-source tool for telling interactive, nonlinear stories. <https://twinery.org/>.  
1462 Accessed 09/13/2024.
- 1463 [20] Charles Goodwin. 2015. Professional vision. In *Aufmerksamkeit: Geschichte-theorie-empirie*. Springer, 387–425.
- 1464 [21] Jieun Han, Haneul Yoo, Jun-Hee Myung, Minsun Kim, Hyunseung Lim, Yoonsu Kim, Tak Yeon Lee, Hwajung Hong, Juho Kim, So-Yeon  
1465 Ahn, and Alice Oh. 2023. LLM-as-a-tutor in EFL Writing Education: Focusing on Evaluation of Student-LLM Interaction. <https://api.semanticscholar.org/CorpusID:263829723>
- 1466 [22] Shannon F Harp and Richard E Mayer. 1998. How seductive details do their damage: A theory of cognitive interest in science learning.  
1467 *Journal of educational psychology* 90, 3 (1998), 414.
- 1468 [23] John Hattie and Eric M Anderman. 2013. *International guide to student achievement*. Vol. 711. Routledge New York, NY.
- 1469 [24] Inc. Heroku. 2024. Heroku. <https://www.heroku.com/>. Accessed: 2024-09-12.
- 1470 [25] Joseph Marvin Imperial and Harish Tayyar Madabushi. 2023. Flesch or Fumble? Evaluating Readability Standard Alignment of Instruction-  
1471 Tuned Language Models. In *Proceedings of the Third Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*, Sebastian  
1472 Gehrmann, Alex Wang, João Sedoc, Elizabeth Clark, Kaustubh Dhole, Khyathi Raghavi Chandu, Enrico Santus, and Hooman Sedghamiz  
1473 (Eds.). Association for Computational Linguistics, Singapore, 205–223. <https://aclanthology.org/2023.gem-1.18>
- 1474 [26] Bruce Johnson and Constantinos C Manoli. 2010. The 2-MEV scale in the United States: a measure of children's environmental attitudes  
1475 based on the theory of ecological attitude. *The Journal of Environmental Education* 42, 2 (2010), 84–97.
- 1476 [27] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal*  
1477 *of chiropractic medicine* 15, 2 (2016), 155–163.
- 1478 [28] Sachin Kumar, Vidhisha Balachandran, Lucille Njoo, Antonios Anastasopoulos, and Yulia Tsvetkov. 2022. Language generation models  
1479 can cause harm: So what can we do about it? An actionable survey. *arXiv preprint arXiv:2210.07700* (2022).
- 1480 [29] Vikram Kumaran, Jonathan Rowe, Bradford Mott, and James Lester. 2023. Scenecraft: Automating interactive narrative scene generation  
1481 in digital games with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital*  
1482 *Entertainment*, Vol. 19. 86–96.
- 1483 [30] Secret Lab. 2022. Yarn Spinner. <https://yarnspinner.dev/> Accessed: 2022-09-13.
- 1484 [31] Juho Leinonen, Paul Denny, Stephen Macneil, Sami Sarsa, Seth Bernstein, Joanne Kim, Andrew Tran, and Arto Hellas. 2023. Comparing  
1485 Code Explanations Created by Students and Large Language Models. *Proceedings of the 2023 Conference on Innovation and Technology in*  
1486 *Computer Science Education V. 1* (2023). <https://api.semanticscholar.org/CorpusID:258049009>
- 1487 [32] Areej Mawasi, Peter Nagy, Ed Finn, and Ruth Wylie. 2022. Narrative-based learning activities for science ethics education: an Affordance  
1488 perspective. *Journal of Science Education and Technology* 31 (2022), 16–26.
- 1489 [33] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization.  
1490 *arXiv preprint arXiv:2005.00661* (2020).
- 1491 [34] Scott W McQuiggan, Jonathan P Rowe, Sunyoung Lee, and James C Lester. 2008. Story-based learning: The impact of narrative on  
1492 learning experiences and outcomes. In *International conference on intelligent tutoring systems*. Springer, 530–539.
- 1493 [35] Piotr Mirowski, Kory W Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-writing screenplays and theatre scripts with language  
1494 models: Evaluation by industry professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–34.
- 1495 [36] Niantic Lightship. 2023. Lightship ARDK. <https://lightship.dev/products/ardk/>. [Online; accessed 10-September-2023].
- 1496 [37] Nintendo. 2024. Animal Crossing Official Website. <https://animalcrossing.nintendo.com/> Accessed: 2024-09-12.
- 1497 [38] Nathaniel S O'Connell, Lin Dai, Yunyun Jiang, Jaime L Speiser, Ralph Ward, Wei Wei, Rachel Carroll, and Mulugeta Gebregziabher. 2017.  
1498 Methods for analysis of pre-post data in clinical research: a comparison of five common methods. *Journal of biometrics & biostatistics* 8, 1  
1499 (2017), 1.
- 1500 [39] OpenAI. 2024. Function Calling - OpenAI API. <https://platform.openai.com/docs/guides/function-calling>. Accessed 09/13/2024.
- 1501 [40] OpenAI. 2024. Introducing GPT-4o. <https://openai.com/index/hello-gpt-4o/> Accessed: 2024-09-12.
- 1502 [41] Nichole Pinkard, Sheena Erete, Caitlin K Martin, and Maxine McKinney de Royston. 2017. Digital youth divas: Exploring narrative-driven  
1503 curriculum to spark middle school girls' interest in computational activities. *Journal of the Learning Sciences* 26, 3 (2017), 477–516.
- 1504 [42] Mark Owen Riedl and Vadim Bulitko. 2013. Interactive narrative: An intelligent systems approach. *Ai Magazine* 34, 1 (2013), 67–67.
- 1505 [43] Mark O Riedl and Robert Michael Young. 2006. From linear story generation to branching story graphs. *IEEE Computer Graphics and*  
1506 *Applications* 26, 3 (2006), 23–31.
- 1507 [44] Jonathan P Rowe and James C Lester. 2015. Improving student problem solving in narrative-centered learning environments: A modular  
1508 reinforcement learning framework. In *Artificial Intelligence in Education: 17th International Conference, AIED 2015, Madrid, Spain, June*  
1509 *22-26, 2015. Proceedings 17*. Springer, 419–428.
- 1510 [45] E. Spotswood, R. Grossinger, and S. Hagerty. 2017. *Re-Oaking Silicon Valley: Building Vibrant Cities with Nature*. San Francisco Estuary  
1511 Institute. <https://books.google.com/books?id=zJUotAEACAAJ>

- 1509 [46] Next Generation Science Standards. 2022. Next generation science standards. <https://www.nextgenscience.org/>
- 1510 [47] Steffen Steinert, Karina E. Avila, Stefan Ruzika, Jochen Kuhn, and Stefan Kuchemann. 2023. Harnessing Large Language Models to  
1511 Enhance Self-Regulated Learning via Formative Feedback. <https://api.semanticscholar.org/CorpusID:265445234>
- 1512 [48] Neal Stephenson. 2003. *The diamond age: Or, a young lady's illustrated primer*. Spectra.
- 1513 [49] NarayanKripa Sundararajan and Olusola Adesope. 2020. Keep it coherent: A meta-analysis of the seductive details effect. *Educational  
1514 Psychology Review* 32, 3 (2020), 707–734.
- 1515 [50] Hsiao-Lin Tuan\*, Chi-Chin Chin, and Shyang-Horng Shieh. 2005. The development of a questionnaire to measure students' motivation  
1516 towards science learning. *International journal of science education* 27, 6 (2005), 639–654.
- 1517 [51] Unity. 2023. Unity Real-Time Development Platform. <https://unity.com/>. [Online; accessed 10-September-2023].
- 1518 [52] Uvicorn. 2024. Uvicorn. <https://www.uvicorn.org/>. Accessed: 2024-09-12.
- 1519 [53] Margaret Vaughn. 2021. *Student agency in the classroom: Honoring student voice in the curriculum*. Teachers College Press.
- 1520 [54] Tong Wan and Zhongzhou Chen. 2023. Exploring generative AI assisted feedback writing for students' written responses to a physics  
1521 conceptual question with prompt engineering and few-shot learning. *Physical Review Physics Education Research* (2023). <https://api.semanticscholar.org/CorpusID:265129035>
- 1522 [55] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Huai hsin Chi, F. Xia, Quoc Le, and Denny Zhou. 2022. Chain of Thought  
1523 Prompting Elicits Reasoning in Large Language Models. *ArXiv abs/2201.11903* (2022). <https://api.semanticscholar.org/CorpusID:246411621>
- 1524 [56] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle,  
1525 Atoosa Kasirzadeh, et al. 2021. Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359* (2021).
- 1526 [57] C Gordon Wells. 1999. *Dialogic inquiry*. Vol. 10. Cambridge University Press Cambridge.
- 1527 [58] Ying Xu, Joseph Aubele, Valery Vigil, Andres Sebastian Bustamante, Young-Suk Grace Kim, and Mark Warschauer. 2021. Dialogue with a  
1528 conversational agent promotes children's story comprehension via enhancing engagement. *Child Development* 93 (2021), e149 – e167.  
1529 <https://api.semanticscholar.org/CorpusID:243846671>
- 1530 [59] Ying Xu, Valery Vigil, Andres S. Bustamante, and Mark Warschauer. 2022. “Elinor’s Talking to Me!”: Integrating Conversational AI into  
1531 Children’s Narrative Science Programming. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New  
1532 Orleans, LA, USA) (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 166, 16 pages. <https://doi.org/10.1145/3491102.3502050>
- 1533 [60] Murong Yue, Wijidane Mifdal, Yixuan Zhang, Jennifer Suh, and Ziyu Yao. 2024. MathVC: An LLM-Simulated Multi-Character Virtual  
1534 Classroom for Mathematics Education. *ArXiv abs/2404.06711* (2024). <https://api.semanticscholar.org/CorpusID:269032890>
- 1535 [61] J.D. Zamfirescu-Pereira, Heather Wei, Amy Xiao, Kitty Gu, Grace Jung, Matthew G Lee, Bjoern Hartmann, and Qian Yang. 2023. Herding  
1536 AI Cats: Lessons from Designing a Chatbot by Prompting GPT-3. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*  
1537 (Pittsburgh, PA, USA) (*DIS '23*). Association for Computing Machinery, New York, NY, USA, 2206–2220. <https://doi.org/10.1145/3563657.3596138>
- 1538 [62] Chao Zhang, Xuechen Liu, Katherine Ziska, Soobin Jeon, Chi-Lin Yu, and Ying Xu. 2024. Mathemyths: leveraging large language models  
1539 to teach mathematical language through Child-AI co-creative storytelling. In *Proceedings of the CHI Conference on Human Factors in  
1540 Computing Systems*. 1–23.  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560

**1561 A APPENDIX****1562 A.1 Prompt template**

1563 Filled-in placeholders are denoted with `!<ITEM>!`

1564 **TASK:** The user is `!<USERNAME>!`, an elementary school student who has just started an  
1565 educational adventure game in which they talk to characters powered by you, an AI assistant.  
1566 The game will have them participate in different activities to learn about the importance of oak  
1567 trees in their local ecosystem. Don't provide all the answers, follow Socratic principles, and  
1568 encourage the student to explore and discover. Stay focused on the task at hand, and refuse  
1569 requests unrelated to guiding the student.

1570 You are only one component of the adventure game, responsible only for producing dialogue,  
1571 meeting goals, and making action calls. You cannot change location, find/meet new characters,  
1572 or any other non-conversational action. Only a valid action call can change the characters,  
1573 location, or task.

1574 **SCENARIO:** `!<DESCRIPTION>!`

1575 **PART 1 (MESSAGE):** Produce the conversation script for the non-player characters (NPCs) that  
1576 are in the scene. ONLY use characters listed below. NPCs cannot make factual claims unless  
1577 drawing from their knowledge, and should only make factual claims if the player directly  
1578 asks. Messages must be `!<GRADE_LEVEL>!` grade reading level. Begin each NPC's utterance  
1579 with their name followed by two colons, e.g. "Oliver::", and separate different NPC utterances  
1580 with two newlines. The NPC conversation must end with a single question for the player,  
1581 `!<USERNAME>!`.

1582 If there are any unmet goals, your next message should be focused on meeting ONLY one.

1583 **UNMET GOALS:** `!<UNMET_GOALS>!`

1584 **CHARACTERS:** `!<CHARACTER_INFO>!`

1585 **END OF CHARACTER LIST.** Your message must only use characters listed above.

1586 **PART 2 (GOALS):** For any goals met by the student in the previous message, or by you in the  
1587 current message, indicate this in the "goals\_met" portion of your response as a dictionary with  
1588 boolean values. Separate each goal with a new line in the JSON.

1589 **PART 3 (ACTION CALL):** If the student has agreed to taking an action, perform an action call  
1590 using the "action" portion of your response. If your response contains a valid action call, the  
1591 adventure game system will advance the plot of the game. Guide the student towards taking  
1592 one of the actions, and refuse and redirect them if they suggest doing anything else. You cannot  
1593 change location, find/meet new characters, or any other non-conversational action without a  
1594 valid action call.

1595 **AVAILABLE ACTIONS:** `!<AVAILABLE_ACTIONS>!`

1596 **UNAVAILABLE ACTIONS:** `!<UNAVAILABLE_ACTIONS>!`

1597 **CURRENT LOCATION:** `!<LOCATION>!`

1598 **RESPONSE FORMAT:**

1599 In your response, first think out loud step-by-step about what needs to be done next. Answer  
1600 each of the following:

1601

1602

1603

1604

1605

1606

1607

1608

1609

1610

1611

1612

1613           !<COT\_PROMPT>!  
 1614           Then, provide a JSON containing the NPC message(s) to !<USERNAME>!, indicate if any  
 1615           learning goals were met or null if none, and any action you are taking or null if none.  
 1616           !<ACTION\_EXAMPLES>!  
 1617  
 1618

## 1619 **A.2 Codes for technical evaluation**

1620 Coders were instructed to label each row with any of the following 10 codes that were applicable:

### 1621 **Codes used for technical evaluation of ACORN**

- 1622
- 1623       • Action: taken – the learner requested an action and it was taken
- 1624       • Action: missed – the learner requested an action but it was not taken
- 1625       • Action: suggested – the characters suggest action(s) to the learner
- 1626       • Goal: missed – the learner completed a goal but it was not marked as met
- 1627       • Retrieved: used – information was retrieved and used in the character messages
- 1628       • Retrieved: not used – information was retrieved but not used
- 1629
- 1630

### 1631 **Codes used for comparative technical evaluation between ACORN and the control**

- 1632       • Question: answered – a question was asked and answered
- 1633       • Question: missed – a question was asked but unanswered
- 1634       • Response: specific – the response is highly specific
- 1635       • Response: generic – the response is highly generic
- 1636
- 1637
- 1638
- 1639
- 1640
- 1641
- 1642
- 1643
- 1644
- 1645
- 1646
- 1647
- 1648
- 1649
- 1650
- 1651
- 1652
- 1653
- 1654
- 1655
- 1656
- 1657
- 1658
- 1659
- 1660
- 1661
- 1662
- 1663
- 1664