DIS 2026
Honourable Mention
Paper Award (Top ~5%)

# Conceptualizing How to Design for AI Literacy through Game Artifacts

Joseph Tu*†
Stratford School of Interaction Design and Business
University of Waterloo
Waterloo, Ontario, Canada
joseph.tu@uwaterloo.ca

Geneva M. Smith‡
Stratford School of Interaction Design and Business
University of Waterloo
Waterloo, Ontario, Canada
g38smith@uwaterloo.ca

Simone Bassanelli
Motivational Digital Systems (MODIS)
Fondazione Bruno Kessler
Trento, Povo, Italy
simone.bassanelli@unitn.it

Annapaola Marconi
Motivational Digital Systems (MODIS)
Fondazione Bruno Kessler
Trento, Povo, Italy
marconi@fbk.eu

Lennart E. Nacke*
Stratford School of Interaction Design and Business
University of Waterloo
Waterloo, Ontario, Canada
lennart.nacke@uwaterloo.ca

## Abstract

AI literacy is an emerging research area that increasingly incorporates new forms of computational intelligence, creating opportunities to enrich human learning and interactive experiences. Our scoping review examines design interventions and discourses within AI literacy and games to identify and characterize the learning experiences that prior research has sought to support. Drawing from 45 papers, we identified and analyzed 48 unique design artifacts, including game-based learning prototypes and gamified systems. We constructed a comprehensive matrix charting each artifact's game or gamification approach, platform, AI literacy focus, game design elements, player requirements, modality, and open-source availability, providing a detailed view of how AI literacy is represented across these systems. Building on this matrix, we identify nine design suggestions that illustrate how specific design choices shape learners' knowledge, skills, and experiences with AI. Our work clarifies game-based AI literacy interventions and offers actionable suggestions for designing future systems that effectively leverage games and gamification to support AI learning.

## CCS Concepts

• **Human-centered computing** → **Interaction design theory, concepts and paradigms**.

## Keywords

scoping review, AI literacy, game, methodology, design, HCI

*The author(s) are also affiliated with the Department of Systems Design Engineering at the University of Waterloo.
†The author is affiliated with Motivational Digital Systems (MODIS) group at Fondazione Bruno Kessler.
‡The author is affiliated with the Department of Management Science and Engineering at the University of Waterloo.

*DIS '26, Singapore, Singapore*
© 2026 Copyright held by the owner/author(s).
This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Designing Interactive Systems Conference (DIS '26), June 13–17, 2026, Singapore, Singapore*, https://doi.org/10.1145/3800645.3812865.

**ACM Reference Format:**
Joseph Tu, Geneva M. Smith, Simone Bassanelli, Annapaola Marconi, and Lennart E. Nacke. 2026. Conceptualizing How to Design for AI Literacy through Game Artifacts. In *Designing Interactive Systems Conference (DIS '26), June 13–17, 2026, Singapore, Singapore.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3800645.3812865

## 1 Introduction

Artificial intelligence (AI) is becoming a part of everyday life, making the understanding of and interaction with AI systems an increasingly important skill [7]. As AI is integrated into social game simulations [54], educational game-based approaches [75, 77], and gamified creative contexts [15, 83], researchers and educators emphasize the importance of conceptualizing AI literacy—the knowledge, skills, and attitudes required to critically engage with and make informed decisions about AI-driven technologies [41, 51]. Consequently, the ability to comprehend, question, and collaborate with AI systems has become a fundamental sub-component of *digital citizenship*[1] [53].

Games offer powerful affordances for learning complex concepts, enabling players to engage with AI systems[2] in situated and embodied ways [48]. In addition to fostering understanding, game-based interventions can shape players' perceptions of AI's capabilities and limitations, prompting reflections on trust, agency, and ethics. Human-Computer Interaction (HCI) researchers have explored how people can learn about and through AI, but definitions and methods vary widely [50, 51]. Among these, *gamified systems* [26] and *game-based learning approaches* [60] are particularly notable. In this paper, we collectively refer to both as *games* when discussing interventions designed to foster AI literacy.

[1]Digital citizenship (which builds upon *digital literacy*) is the capacity to actively, continuously, and responsibly participate in communities on and offline through competent and positive engagement with digital technologies (by creating, working, sharing, socializing, investigating, playing, communicating, and learning).
[2]In this work, we use the terms AI and Machine Learning (ML) in line with common public discourse, treating ML as a subset of AI. Although these terms are technically distinct, we do not maintain a strict differentiation here, while acknowledging that more specialized literature treats them separately.

However, while many researchers have begun to leverage games to *teach* or *explore* AI literacy, the field lacks a consolidated understanding of how such interventions are being designed, evaluated, or conceptualized (primarily just preliminary, descriptive and exploratory in nature) [69]. For example, recent systematic literature reviews by Casal-Otero et al. [12] and Ng et al. [52] encourage further exploration of how games can be designed and used in conjunction with AI literacy.

This gap highlights the need for a comprehensive mapping of the literature to understand how AI literacy is being addressed in *games*, the design choices researchers make, and the types of learning or interaction they aim to support. A scoping review in this intersection allows us to uncover both the educational and experiential intentions behind such interventions, whether they aim to teach technical concepts, foster critical awareness, or promote ethical reasoning. This knowledge can help guide the design choices towards meaningful AI literacy objectives. Thus, we address this through the following research questions (RQs):

**RQ1** What aspects of AI literacy are addressed by current game- and interaction-based learning systems, and how do these aspects distribute across platforms, modalities, and design artifacts?

**RQ2** What game design approaches and broader design considerations support the development of effective, equitable, and scalable AI literacy through interactive systems?

We identified and analyzed $n = 48$ unique *design artifacts* (including game-based learning prototypes and gamified systems from $n = 45$ papers[3]) contributing to the research and design of games for AI literacy. This includes:

(1) A comprehensive matrix of 48 game and gamified systems for AI literacy interventions by analyzing how they address different AI literacy domains and identifying patterns across platforms, modalities, and interaction designs (**RQ1**).
(2) Nine design suggestions (DS1—DS9) highlighting how specific design choices shape learners' knowledge, skills, and experiences (**RQ2**).

By integrating insights from the matrix analysis with the nine design suggestions, we provide a clearer understanding of the current landscape and offer actionable guidance for developing effective AI literacy interventions through games and gamification.

## 2 Background and Related Work

We first clarify our scoping review's key terms, then review existing secondary research on AI learning, education and literacy to identify gaps that motivate our research.

### 2.1 Clarifying Relevant Terminology Pertaining to the Scoping Review

We built our scoping review around the following terms, which also determined how we sorted our charting data:

**"AI Literacy":** Several complementary frameworks have articulated the concept of AI literacy. Building on Touretzky et al. [71]'s foundational perspective, AI literacy encompasses five core ideas: *(1) Perception*, concerning how computers interpret sensory input and extract meaning through knowledge; *(2) Representation and Reasoning*, wherein intelligent agents construct and manipulate models of the world to support decision-making; *(3) Learning*, algorithms' capacity to adjust internal representations in response to data, as seen in models such as decision trees or neural networks; *(4) Natural Interaction*, which involves enabling agents to communicate and collaborate effectively with humans by leveraging linguistic, cultural, and emotional knowledge; and *(5) Societal Impact*, concerning the ethical, economic, and cultural consequences of AI systems, including issues of bias, fairness, transparency, and applications of AI for social good.

From a HCI perspective, Long and Magerko [41] define AI literacy as a *set of competencies* that enables individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI systems, and use AI tools across personal, educational, and professional contexts. This framing positions AI literacy as an extension of *digital literacy* that is not necessarily dependent on *computational* or *scientific* literacy [65]. In line with broader literacy frameworks, AI literacy can be understood as the ability to identify, interpret, create, communicate, and reason within contexts mediated by intelligent technologies.

Complementing these perspectives, Ng et al. [51] conceptualize AI literacy through progressive domains: "*knowing and understanding AI*", "*using and applying AI*", "*evaluating and creating AI*" and "*AI ethics*", which map onto Bloom et al. [9]'s taxonomy of cognitive processes and its later refinements [30]. Collectively, these frameworks position AI literacy as both a *cognitive* and *socio-technical* construct, integrating knowledge of AI concepts with the capacity to apply, critique, and design intelligent systems responsibly. In this review, we adopt this integrated understanding and use the term "AI literacy" inclusively to encompass these overlapping definitions.

**"Game" (Design Elements) in AI Literacy:** We conceptualize games in AI literacy through two complementary lenses. The first is *gamification*, which refers to the use of game design elements in non-game contexts [17]. The second is *game-based learning*, which emphasizes structured gameplay designed to facilitate specific learning experiences in digital games [58]. While the literature often uses "serious games" and game-based learning interchangeably, they are distinguishable by their intent and scope. As Becker [6] distinguishes in a brief academic letter, game-based learning integrates learning objectives directly into gameplay mechanics to make the learning process engaging and effective. In contrast, serious games are dedicated tools in this approach—games explicitly designed for educational purposes that are "shifted" towards targeted learning outcomes [20].

Gamification has been shown to enhance motivation, engagement, and learning outcomes across educational levels. Studies report positive effects on students' intrinsic motivation, learning attitudes, and emotional experiences [16, 89]. Common game elements (such as points, badges, and leaderboards) can promote engagement by providing immediate feedback and progress tracking [18, 34]. However, gamification is most often implemented in digital learning environments, such as e-learning systems, mobile platforms, and flipped classrooms [89]. For instance, Gini et al. [23] highlight the growing application of gamification in STEM and

[3]The discrepancy between the number of papers and artifacts arises because two papers contained multiple unique design artifacts.

K–12 education, as well as its relevance in online and personalized learning contexts.

Building on Plass et al. [57], we view game-based learning as an approach that "*integrates cognitive, motivational, affective, and sociocultural perspectives*". These dimensions are essential for designing and studying games that effectively support learning across physical, digital, and hybrid environments. However, despite these distinctions, debates persist regarding what constitutes a "game," what qualifies as a "non-game" [70], and whether digital games are the only valid medium for learning [89]. To address this, our review uses "gam*" as a search keyword to encompass a broad range of interpretations rather than adhering to a single definition. We therefore treat games in AI literacy as context-dependent, shaped by underlying *game design elements* that span both gamified and game-based scenarios. For example, the game content may include problem-solving challenges that foster a sense of achievement while engaging learners with AI concepts [22, 29, 60].

## 2.2 Secondary Research on AI-Related Pedagogical Approaches in Games

Several works have reviewed the intersections between AI and education, as the term AI literacy encompasses elements of AI education [13], AI ethics [35], machine learning (ML) education [64], and large language models (LLMs) in education [85], along with their respective challenges and implications for future research. We focus on the domain of games and their teaching and learning applications in AI-related contexts, as examined through secondary research (e.g., scoping and systematic reviews). For example, Hu et al. [29] mapped alignments between AI techniques and game genres for experimental purposes. However, there is a lack of research approaching AI–game integration from a *pedagogical and literacy-driven perspective*.

A few reviews examined overlaps between AI education and game-based learning platforms. For instance, Giannakos et al. [22] reviewed digital games designed to support computational thinking and programming in K–12 education, finding that "coding" remains the central goal in most environments—despite *computational thinking*'s emphasis on problem decomposition, abstraction, and algorithmic reasoning as general cognitive skills. *AI education* broadly encompasses the teaching of technical AI and ML concepts, often targeting specialist or pre-professional audiences [13, 64]. In contrast, *AI literacy* is oriented toward a general population and emphasizes critical evaluation, ethical reasoning, and the ability to interact with and apply AI tools across everyday contexts [41, 71]. Existing reviews have not consistently maintained these distinctions, leading to a fragmented picture of what game-based interventions actually address.

Meanwhile, Zhan et al. [91] observed that many researchers use puzzle games for teaching AI concepts because of their compact size, ease of design, and development efficiency. This observation aligns with findings from Abd-alrazaq et al. [1], who analyzed serious games in the health domain and additionally found that motor impairment was the most common condition targeted. In particular, they noted that most studies used serious games for rehabilitation purposes, with popular genres including role-playing, puzzle, and platform games—consistent with the trends identified by Zhan et al. [91]. Similarly, Ullah et al. [74] investigated serious games in science education, identifying country-level research trends, common evaluation methods, and both positive and negative perceptions toward their educational use. Yet a common limitation across this body of work is the tendency to treat game design as a delivery mechanism rather than as a variable that itself shapes learning outcomes. Few reviews interrogate how specific game design elements (such as mechanics, feedback structures, or narrative framing) map onto particular AI literacy competencies [29]. Furthermore, as noted by Zhai et al. [90], while AI techniques in educational contexts can demonstrate "smart" computational capabilities, they often fail to deliver meaningful added value at scale due to high implementation costs and limited domain-specific adaptability, restricting their ability to support personalized learning experiences.

# 3 Methodology

Rather than providing an exhaustive overview of existing studies, our focus lies in identifying the specific features AI literacy and games aims to enhance or introduce. We also examine how researchers seek to shape the human experience of play through design mechanisms and technological interventions. For now, we set aside broader concerns, such as the digital divide[4].

Predefined research questions and a structured search strategy informed our review, but we did not follow a formal scoping review protocol [5]. Instead, we adopted a flexible approach that allowed the first author's growing familiarity with the literature to guide the analysis. To ensure rigour and transparency, we drew on the seven stages of the INSPIRE Framework [73], which better aligns with the "search and selection" stages in formalized HCI literature review practices. We expand on this recent methodology adopted and supported by librarians within the HCI field.

## 3.1 Involvement: Inclusive Research Participation

We critically reflect on our assumptions and biases when analyzing the data and constructing findings with our librarians. The research team brings diverse perspectives from both academic and professional contexts, with representation spanning from graduate students to full professors at the time of this research.

Our team combines expertise in HCI and game user research (GUR). The first author has access to both the University of Waterloo's and Fondazione Bruno Kessler's ILL networks to ensure broad coverage of the literature, and brings extensive knowledge of how metadata is embedded in academic databases, developed in collaboration with librarians[5]. The first and second authors have conducted interdisciplinary work in game design and have strong relationships with their institutional librarian to fully use the ILL network. The third, fourth, and last author bring deep expertise in gamification, providing critical evaluation of relevant works. All authors have taught game design courses at the university level.

## 3.2 Inquiry: Learning About the Research Space

Following the first stage of the INSPIRE framework [73], we began by exploring the research space to understand how AI literacy has been approached in games (Figure 1). While AI literacy has received

[4]Unequal access to and effective use of digital technology

[5]All authors are familiar with the search and selection of papers in HCI related databases [73].




**Figure 1: A flowchart of the Inquiry Stage within the INSPIRE Framework.**

increasing attention in educational research [12, 52], little work has synthesized how the design of game-based and interactive systems addresses different AI literacy domains—a gap this review aims to fill. Thus, we focused on the central question: What do we envision and study when designing for AI literacy in games? We derived our research questions from this activity (Section 1).

To ensure our review's novelty, we also confirmed that no existing or ongoing systematic reviews on AI literacy and games were available in HCI venues using an open-registered repository such as the Open Science Framework (OSF) Registries[6].

## 3.3 Navigation: Find the Seed / Anchor Papers

The second stage of our review involved selecting anchor works and mapping the relevant research space. We began by tracing the origins of the term "AI literacy" in the literature [41, 51] and "games" like *Artbot* [77] and *TreasureIsland* [49]. These works, our *seed papers*, highlight both the need for AI literacy and the HCI community's tendency to approach it primarily through technological interventions. Building on this foundation, we expanded our search to identify additional seed papers that explicitly addressed the intersection of AI literacy and games. Figure 2 shows our initial keyword extraction.

## 3.4 Searching: Different Approaches to Finding Papers

To develop the final key string, we began by extracting key terms from titles, abstracts, CCS concepts, domain, research area, highlights and author keywords from our *seed papers* (Figure 3). Through divergent searching, we expanded these terms to include synonyms and related concepts, incorporating keystrings like AI EDUCATION, AI LEARNING, and AI TEACHING. However, we found that the Scopus, SpringerLink, and IEEE databases treat the string "AI" as a stop word[7], just as ACM DL does for "Machine Learning".

We then engaged in iterative testing across multiple databases that indexed our seed papers, refining keyword combinations and Boolean operators to maximize relevant study retrieval while minimizing irrelevant results (i.e., convergent searching). For example, we found that the wildcard term GAM* was effective, as it captured variations such as GAME, GAMES, GAMIFIED, GAMIFICATION, and GAME-BASED. However, it also retrieved irrelevant results, such as papers related to GAMMA.

[6]A permanent, transparent, easily accessible repository that enables the archiving, sharing, searching, and aggregating of funded study plans, designs, data, and outcomes. Researchers can create robust, timestamped registrations of research projects, or discover existing ones.

[7]A common word like "the," "a," or "is" that search engines often ignore to help reduce noise, improve efficiency, and prioritize more meaningful keywords, yielding more relevant results.

## 3.5 Point of Saturation

During this stage, we consulted closely with our librarian, whose extensive knowledge of indexing practices proved invaluable in refining search terms to better capture topics aligned with our research objectives. As we used the University of Waterloo and Fondazione Bruno Kessler institutional interlibrary loan (ILL) network, our librarian recommended exploring additional databases such as IEEE Xplore and SpringerLink, since several papers that frequently cited our seed papers are published there. At this stage, we determined a point of saturation, where enough studies had been identified to finalize the search strategy and corpus, as of 1908—the first available year filter on ACM Digital Library (ACM DL). We selected this year to guarantee complete search exhaustion, as ACM DL's search engine was returning inconsistent corpus sizes for the same search string at the time of the search.

Using the final keystring, we conducted our search in $n = 4$ databases for the years (1908–2025): ACM DL, IEEE Xplore, Scopus and SpringerLink (Figure 4). Although we iteratively refined search terms, we found that terms such as AI EDUCATION, AI LEARNING, and AI TEACHING produced fewer relevant results, however the term AI LITERACY consistently yielded more, particularly those related to teaching AI. Based on our librarian's advice, we focused instead on AI LITERACY, which, though a relatively new concept, proved more effective for capturing relevant studies, particularly those related to games. Accordingly, we simplified the final search string to emphasize the term AI LITERACY across all databases (Section 3.7).

## 3.6 Reporting: Document the Process

To avoid citation inflation, we included only the most recent version of each work, sourced from ACM DL ($n = 373$), IEEE Xplore ($n = 17$), Scopus ($n = 57$), SpringerLink ($n = 624$), and snowballing following full-text screening ($n = 16$), yielding a corpus of $n = 1087$ papers before de-duplication (Figure 5). We uploaded the resulting dataset of $n = 1071$ records to Covidence[8] for the screening process. We also document anomalies relevant to replicating our methodology:

- Wiley, one of our seed paper's publisher, acquired the Open Access company Hindawi in 2021. Shortly thereafter, in 2022, a large number of Hindawi's publications were retracted, leading Wiley to discontinue all associated entries from its database. Although one of our seed papers is from Wiley, we excluded this database due to these retraction issues. Fortunately, Scopus was still able to retrieve our seed paper.

[8]Covidence is a cloud-based platform designed to facilitate the screening of titles and abstracts in large-scale literature reviews.

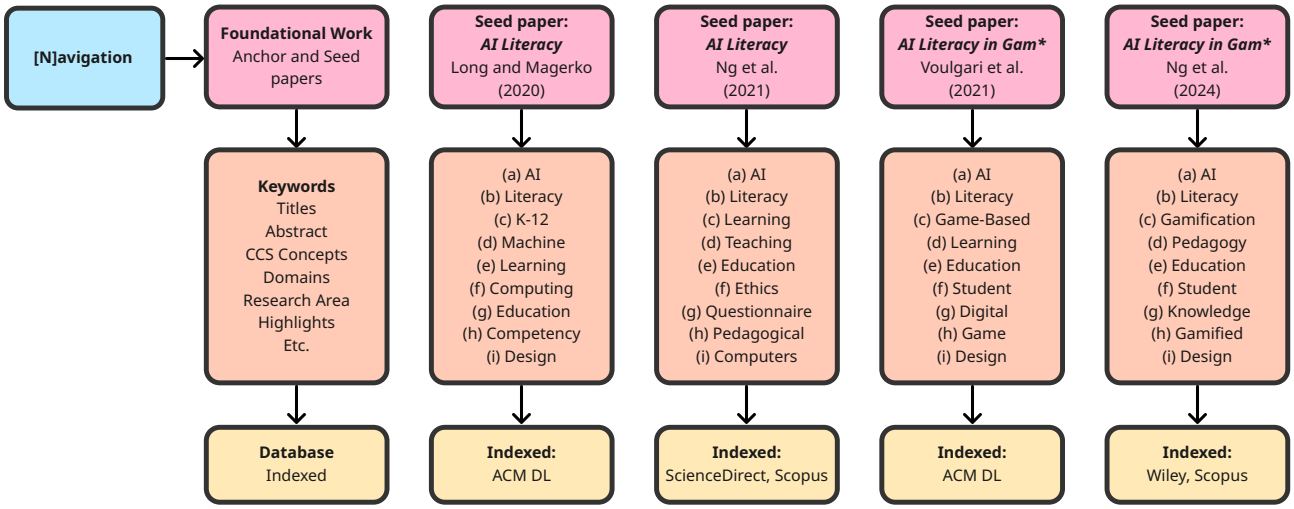



**Figure 2: A flowchart of the Navigation Stage within the INSPIRE Framework.**

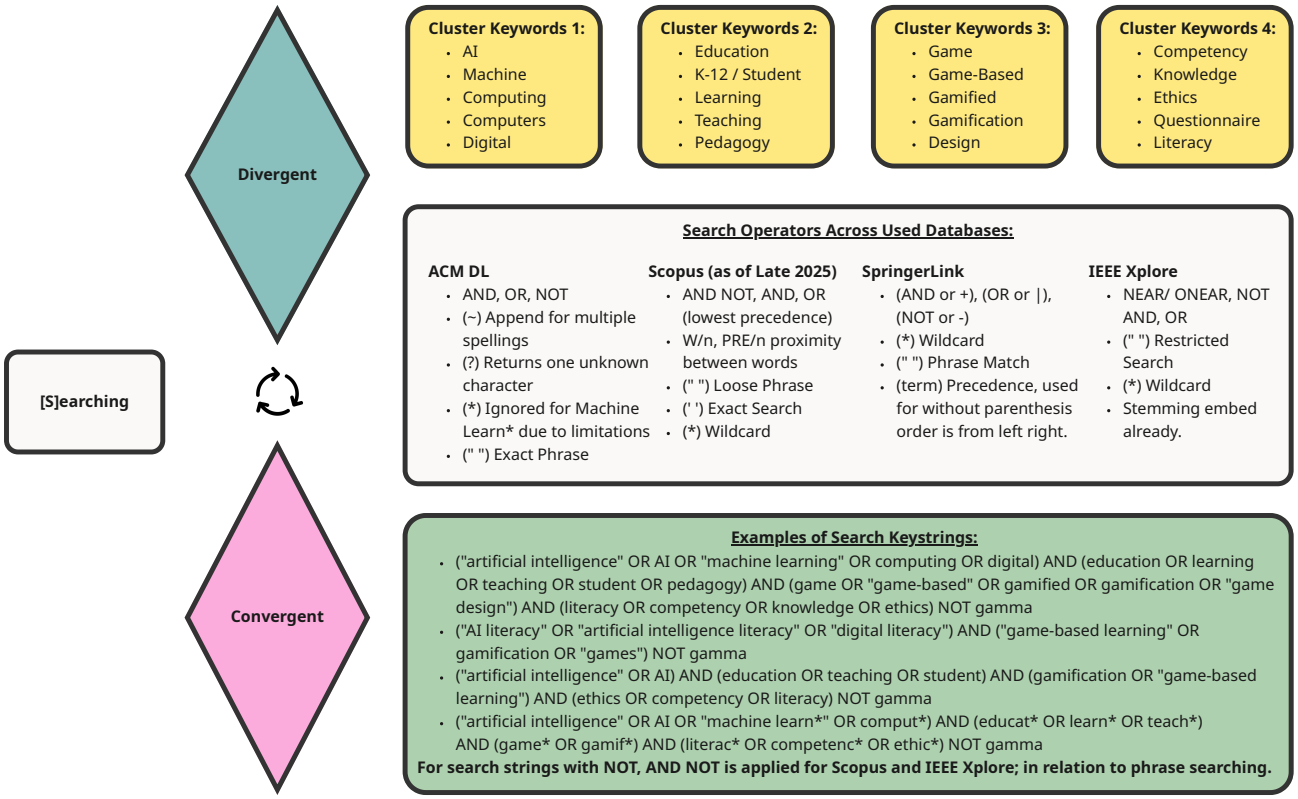



**Figure 3: Overview of divergent and convergent search strategies within the Searching Stage of the INSPIRE Framework.**

- IEEE Xplore updates its index at the end of each fiscal year, so publications from 2025 were not yet fully indexed at search time. The database itself acknowledges this limitation.
- We also note that SpringerLink, which we access through our institutional ILL, provides paid access to otherwise pay-walled publications. Therefore, those that do not have this access might find it difficult to reproduce our results.

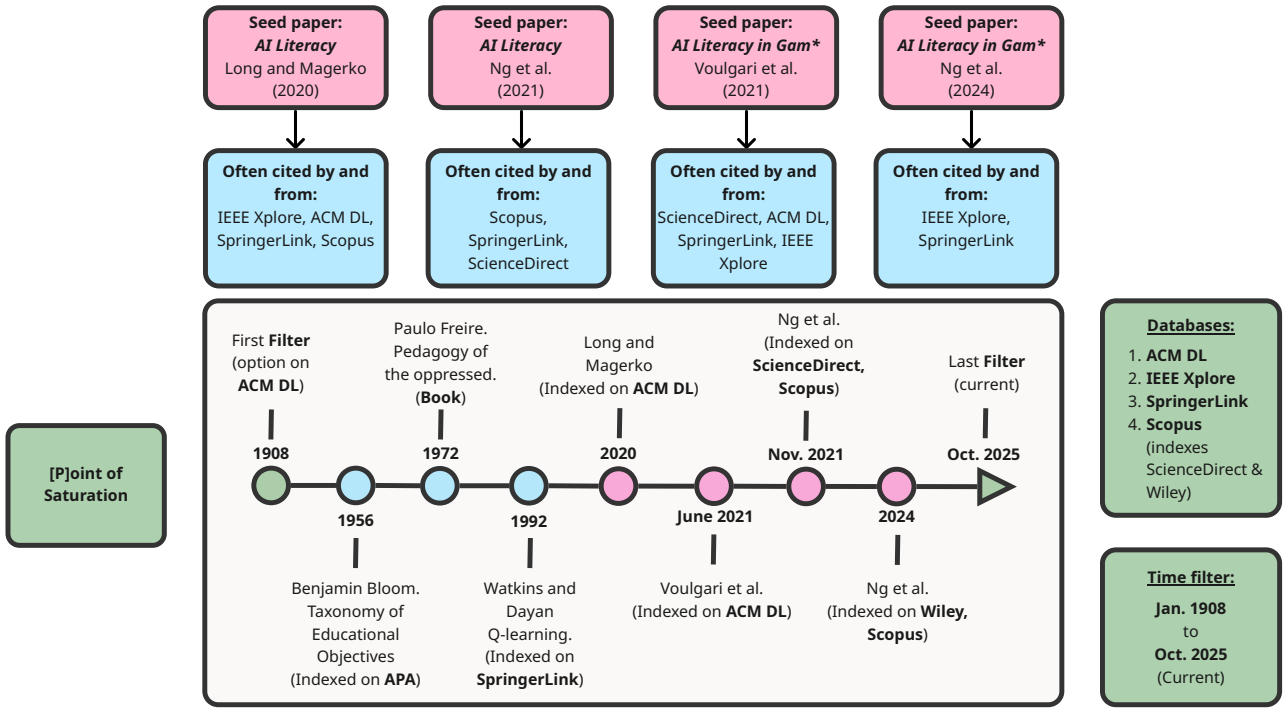



**Figure 4: A figure of the Point of Saturation Stage within the INSPIRE Framework**

- For SpringerLink, we used an online conversion tool[9] to transform exported comma-separated values (`.csv`) into bibliography (`.bib`) files format for reference management.
- Since SpringerLink does not export abstracts or DOIs, we retrieved them using a custom script developed by the first author.
- Springer has started to merge their entire branding to Springer Nature Link which now also includes content from BioMedCentral, Apress, Palgrave, and SpringerOpen[10].

*3.6.1 Duplicate Removal.* We identified potential duplicates using Covidence's automated duplication percent match (any abstracts and titles with 95% similarity), then manually checked and removed those with identical contents. After eliminating duplicates ($n = 33$), we obtained $n = 1038$ unique records[11] for screening.

*3.6.2 Screening Phase and Study Selection.* The first and second authors oversaw all screening, with the third or fourth author acting as mediators to resolve disagreements as needed. In the context of this paper, each study's quality depended on its ability to explain AI literacy design decisions in detail. This implicit measure assigns less weight and representation to papers that do not provide substantial insight into games.

For title and abstract screening, we included papers if they met the following inclusion criteria (IC):

**IC1** The item contains AI literacy related to games.
**IC2** The item is a peer-reviewed paper (i.e., no books or theses).

We derived our exclusion criteria (EC) from our ICs alongside the item's full-text availability (i.e., items that were not available to the research team in full text using their academic libraries and ILL services were excluded):

**EC1** The item's full-text must be accessible.
**EC2** The item is a formalized literature review (systematic review, scoping review, etc).
**EC3** The item is a book or a full proceedings.
**EC4** The item is written in a language other than English.

If a study's relevance was unclear from its abstract, we retained it for full-text screening. For example, we included any studies related to "games", as it was often difficult to determine from titles and abstracts whether they addressed aspects of AI education, learning, teaching, or literacy design. A total of $n = 89$ papers proceeded to the full-text screening, of which $n = 46$ formed our corpus.

During the full-text screening, we observed that several papers presented preliminary results of interest, prompting us to conduct citation-based snowballing. Specifically, we searched for follow-up work using prototype names and "cited-by" links, allowing us to identify studies where authors reference their earlier work. This yielded an additional $n = 16$ papers, resulting in a final corpus of $n = 45$ papers after a final de-duplication. Given the rapid growth

[9] A JavaScript-based tool for converting SpringerLink CSV files to BibTeX format.
[10] We acknowledge that the corpus size may change when reproducing our work, however, relevant inclusion criteria and exclusion criteria should result in a similar final corpus size
[11] We note that this does not account for lecture notes, which could be indexed separately with a preceding or subsequent DOI.

of the field and our exploratory research goals, we included peer-reviewed design artifacts. Under a strict critical appraisal, only nine of these works would meet the inclusion threshold in a traditional systematic review. However, because our approach is intended as a *scoping review*, we deliberately adopted broader inclusion criteria.

*3.6.3 Charting, Collating, and Reporting the Data.* Scoping reviews tend to address broader topics where various study designs may be applicable, rather than strictly adhering to systematic review conventions. Accordingly, we followed recommendations from Arksey and O'Malley [5] to record the following items using a "descriptive–analytical method". This involves the application of a common analytical framework to all primary research reports, collecting standardized information from each. The first author designed the initial data extraction template in alignment with our research goals (see Appendix), and any discrepancies in the resulting matrix were resolved through discussion involving at least two authors. As patterns emerged during analysis, we restructured it into the matrix format seen in Figure 7:

- **Name:** The title of the prototype or system (denoted by the author).
- **Gam*:** Specifies if the intervention uses full-fledged game-based learning or gamified elements.
- **AI Literacy Focus:** The general AI literacy theme addressed (e.g., conceptual understanding, ethics, data use) [41, 51].
- **Target Audience:** The target learner group or demographic (e.g., students, teachers, public), contextualizing which population the AI literacy competencies are designed for [71].
- **Research Question(s)/Goal(s):** The study's main aims or research objectives to clarify the intended learning outcomes or pedagogical inquiry.
- **Co-Design Process:** Captures whether learners, teachers, or stakeholders co-designed the game/intervention, reflecting participatory and socio-technical dimensions of AI literacy [41].
- **Evaluation Gam*:** Describes how the gam* intervention was evaluated (e.g., number of participants, data type, and analytical approach).
- **Study Duration:** The temporal scope of the intervention (e.g., single session, multi-week, longitudinal).
- **AI Role in the Gam*:** Specifies how AI is represented or functions in the game (e.g., simulated agent, adaptive tutor, content generator) [71].
- **Interactivity or Play:** The degree and type of interaction (e.g., collaborative play, single-player tasks), and if the gam* can be downloaded (open-source).
- **Key Findings:** The major results or implications related to AI literacy learning and game effectiveness.

*3.6.4 Collating, Summarizing and Reporting the Results.* Crucially, our approach reflects a distinctly *human practice*, without the use of computational tools in meaning interpretation [32]. First, we examined the corpus's characteristics such as the papers' publication venues. Next, we analysed the papers using reflexive thematic analysis (RTA), forming two clusters (descriptive and interpretive) based on recent recommendations from Braun and Clarke [11]. We also applied deductive coding informed by our focus on AI literacy, using work from Long and Magerko [41] (one of our seed papers). The coding process was highly iterative and completed over multiple rounds by the first author, with intermittent discussions with the full research team. This ultimately produced two overarching thematic areas: how AI literacy is conceptualized and represented across game-based systems (**RQ1**), and the design features and mechanisms that meaningfully support AI literacy in play (**RQ2**), aligning with a codebook-driven analysis [10].

Finally, the full research team refined and finalized the resulting themes over four meetings. These discussions focused on the themes' suitability and terminology, supported by the example sentences in Figure 3. A detailed overview of our themes, the full list of corpus papers, and the complete analysis process are available as supplementary materials.

## 3.7 Exhaustion: Re-Search and Update

Table 1 summarizes the final search strategy, documenting exact queries and coverage for transparency and reproducibility. To ensure the robustness and currency of our review, we conducted the final search multiple times over several months until October 1st, 2025. Our goal was to stabilize the size of the resulting corpus. By reapplying the search strategy and monitoring the results, we ensured that our corpus comprehensively reflects the literature on AI literacy and games up to the cut-off date, providing a stable foundation for analysis and synthesis.

Although new publications can appear during the selection, extraction, and analysis phases, we observed inconsistencies across a few databases. Since each database processes searches differently, we outline below the rationale for each query syntax.

ACM DL provides filters for field, month, and year. However, when restricting results to September 2025, the search unexpectedly returned fewer works. Therefore, since ACM DL does not allow filtering by exact dates, we extended the filter to include publications up to October 31st. In contrast, Scopus only supports year-based filtering. To ensure inclusion of `2025` publications, we applied the syntax filter "`before 2026`." While our last search was performed on October 1st, this approach means that some in-press or in-camera-ready publications may also be present in the dataset depending on when the paper goes live.

Our IEEE Xplore search process mirrors that of Scopus, as both rely on the same indexing system. In IEEE Xplore, the year filter is automatically applied, with the earliest indexed paper defaulting to `1884`. However, for both Scopus and SpringerLink, we did not use the query modifier NOT GAMMA, because its use did not capture one of our seed papers in the results.

On the morning of October 1st, 2025, we retrieved 624 papers from SpringerLink through University of Waterloo's ILL network. However, on the same day, Fondazione Bruno Kessler's ILL network only returned 623 papers. After verifying that the results were otherwise the same, we decided to include the search results from University of Waterloo's ILL network because it captured one additional paper. The extra paper was deemed irrelevant after screening, but we retained this search to ensure maximal coverage. Our institutional librarian supported our access through the ILL network. We provide all public search links without institution

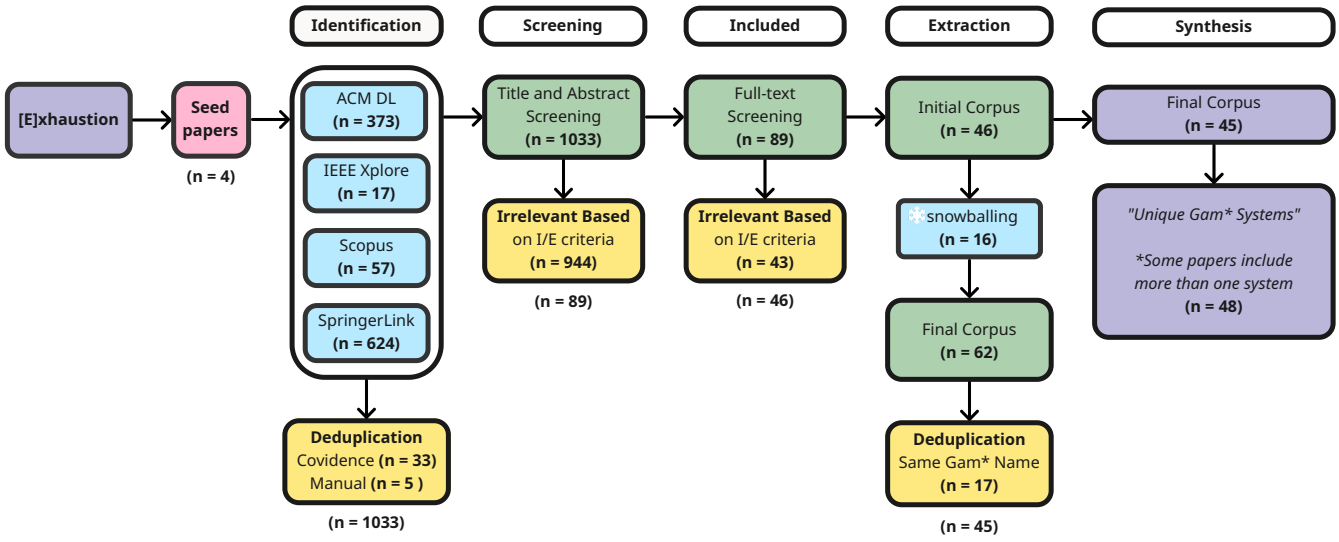



**Figure 5: A figure of the Exhaustion Stage within the INSPIRE Framework.**

| Database | Syntax Query | Corpus Size* |
|---|---|---|
| ACM DL (Link) | "query": AllField:("AI literacy") AND AllField:(gam* AND NOT gamma) "filter": E-Publication Date: (01/01/1908 TO 10/31/2025), ACM Content: DL | 373 |
| IEEE Xplore (Link) | ("All Metadata":"AI literacy") AND ("All Metadata":gam*) NOT ("All Metadata":gamma) | 17 |
| Scopus (Link) | TITLE-ABS-KEY ( "AI literacy" AND gam* ) AND PUBYEAR > 1884 AND PUBYEAR < 2026 | 57 |
| SpringerLink (Link) | "query": ("AI literacy" AND gam*) "filter": Date published, Custom dates: (Start year: 1908) (End year: 2025) | 624 |

** The corpus size was filtered up until 31-OCT-2025. The public link will display the filter until the end of 2025.*

**Table 1: This table lists the final search queries in each database's syntax. The corresponding filters are applied to the ACM DL search results to mirror the filters in the other database searches. The final search was conducted on 31-OCT-2025. We provide a public link to the timestamp of each database (without custom filters from our ILL, thus these public link only allow to only filter by year). However, due to inconsistencies of these databases as noted by Tu et al. [73], we also provide the full corpus in the supplementary materials.**

virtual private network (VPN) attachments and without the date filter (only available on ILL networks).

## 4 Findings

We organize our findings around our research questions: what AI literacy domains are addressed through existing game-based systems, revealed through our matrix diagram and charting process (**RQ1**), and which game design elements and broader design considerations support these AI literacy learning goals (**RQ2**).

### 4.1 Characteristics of the Literature

Our results indicate that AI literacy and games research has increased in popularity over the last four years (Figure 6). Note that we prioritized the latest peer-reviewed publication describing a design artifact when there are also preliminary publications to capture the most complete and accurate representation of the research. This reduces inconsistencies that can arise from earlier iterations, avoiding the fragmentation associated with *salami publishing*[12].

### 4.2 Mapping AI Literacy Coverage in Game and Gamified Systems (RQ1)

We identified 48 unique game- and interaction-based design artifacts for AI literacy in our 45-paper corpus. Given the rapid growth of the field and our exploratory research goals, we break them down by paper type: full papers presenting evidence-driven studies ($n$ = 9, 🌱); short papers focused on empirical work ($n$ = 6, 🌱); design papers ($n$ = 13, 🌱); ongoing work with preliminary findings ($n$ = 12, 🌱); and posters ($n$ = 5, including one doctoral consortium submission, 🌱).

[12]The unethical practice of splitting a single research study into multiple smaller publications to artificially increase publication output.

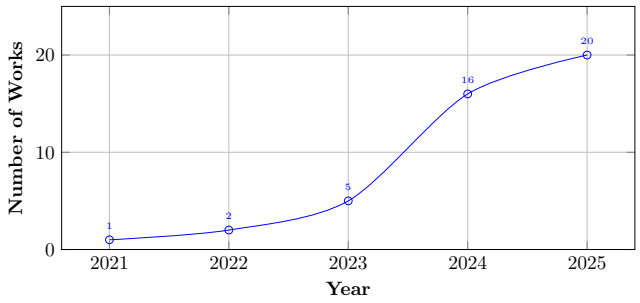



**Figure 6: Number of AI literacy and game-related works published per year.**

We charted our data into a matrix diagram (Figure 7) as described in Section 3.6.3. In the following sections, we report our findings for each matrix sub-column. Given that most of the reviewed studies are not full papers, there was often insufficient detail resulting in potentially unreliable tagging. To combat this, we also examined any supplementary material provided by the authors where available.

*4.2.1 Game and Gamified Design Artifacts.* We derived these classifications from our disciplinary expertise in game design and gamification (Section 3.1). Our analysis reveals consistent trends in how game-related studies operationalize both games and gamification. Most of the game-based papers we analyzed integrate multiple gamification elements within their designs, contributing to a conceptual blurring between games and gamified systems. This blurring stands in contrast to the foundational definition of gamification as the application of game elements in non-game contexts, indicating that researchers frequently use the term inconsistently in the literature. This issue is particularly evident in design and works-in-progress submissions (🌱, 🌱), which often use the terms "game" and "gamification" interchangeably. Specifically, we identified: nine papers reporting game-based learning approaches; two describing serious games; and two focusing on game interfaces, which we can interpret as using game design elements rather than constituting full games.

*4.2.2 Deployment Platforms.* Our data show that 32 design systems were computer-based, 6 are for mobile devices, 14 incorporated tangible components (3 of which were situated in classroom or course settings), and 1 used augmented reality (AR). We did not classify systems designed to be played entirely on a computer as involving tangibility unless physical interaction was an intentional and required part of the experience. For example, DiPaola et al. [19] presents an escape-room style system that, although delivered through a computer interface, still requires players to physically navigate and interact with the surrounding space. For this reason, we is categorized it as both computer-based and tangible. In contrast, we classified systems that rely on physical card decks or other handheld materials as tangible interaction alone.

*4.2.3 AI Literacy.* Based on the terms we use for this scoping review (Section 2.1), *Know & Understand AI* represents the largest proportion (n=30), followed by *Use & Apply AI* (n=21), *Ethics & AI* (n=14), and *Evaluate & Create AI* (n=9). We also note that all six short papers (🌱) focus on *Know & Understand AI* and *Use & Apply AI*. This suggests that current work most frequently prioritizes practical engagement with AI systems, while comparatively less attention is given to critical evaluation and creative construction of AI. This highlights a potential gap in existing approaches, pointing to an opportunity for future game design research to more deeply support higher-order AI literacy skills. We also note that only two design artifacts, *Apricot Stone City* [45] and *TreasureIsland* [49], incorporated all four AI literacy focus areas.

*4.2.4 Game Design Elements.* A majority of game-related design considerations arise from decision-making, feedback systems, levels and phases (particularly within gamification-oriented work), and narrative and story. For instance, decision-making supports active engagement and agency (e.g., Kim et al. [36]); feedback systems provide immediate reinforcement and guide player learning (e.g., Lee and Ali [39]); and levels or phases offer structured progression that can scaffold increasing complexity (e.g., Zammit et al. [88]). Narrative and story elements, meanwhile, help contextualize AI concepts within meaningful scenarios, making abstract or technical ideas more relatable (e.g., Moore et al. [45]). There are comparatively fewer instances of puzzles, role-playing, guessing or hidden object mechanics, time-based gameplay, pathfinding, simulation, badges, points, and leaderboards. This points to future research avenues exploring how to effectively use these game design elements in AI literacy games or gamified applications.

*4.2.5 Player Requirements.* Our data show that 28 of the systems support single-player interaction, while 20 support multiplayer experiences. Notably, 7 of the multiplayer systems require a moderator—an external person responsible for facilitating or running the activity. This moderator role is not embedded in the system itself, typically falling to a researcher or another individual outside the gameplay experience. This can limit the utility of such systems, as moderators are not always available and players might not want to take on the role because it often requires deep game knowledge to be effective.

*4.2.6 Modality of Interactivity and Play.* Of the 48 design artifacts, player-system interaction dominates both individual and team-based play. At the individual level, 27 artifacts have players engaging directly with the digital or physical system rather than with each other, compared to only 10 built around player-versus-player interaction. Team-based designs follow the same pattern. Nine artifacts structure play as team–system collaboration, while just three involve direct team-versus-team dynamics.

The patterns indicate that most AI-focused educational game systems centre on individual interaction with the technology itself, rather than fostering interpersonal or competitive gameplay. This suggests an opportunity for future work to explore richer social interaction models such as cooperative problem-solving, inter-group negotiation, or peer-based strategy development. These could further support AI literacy through shared inquiry and collaborative reasoning.

*4.2.7 Open Source.* Only 12 of the design systems in our review are fully open source, meaning they are playable; of those, 3 are playable but offer nothing further (no source, documentation, etc.). In these cases, the complete materials are publicly accessible, meaning that

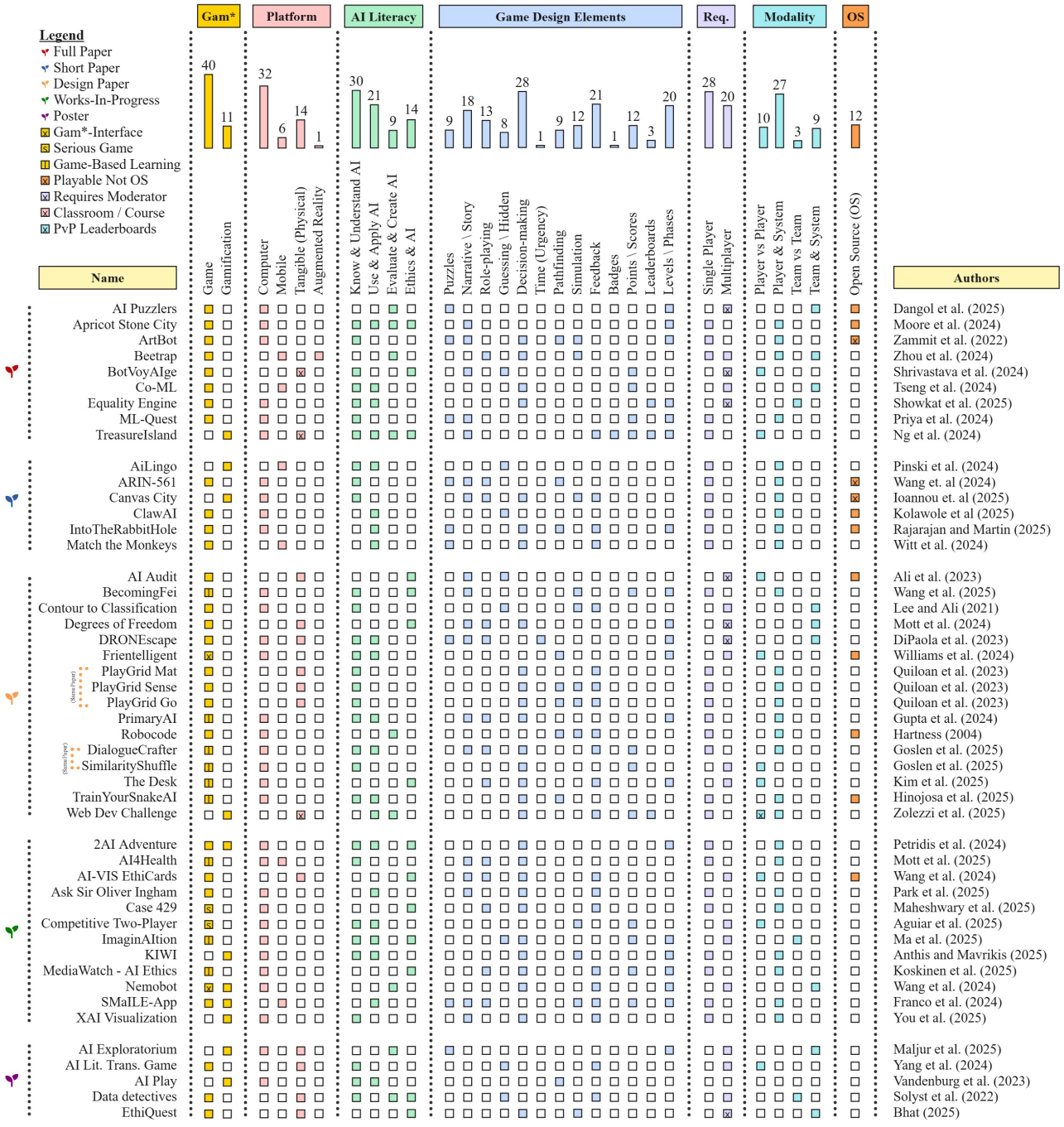



**Figure 7: A matrix diagram of our charting data. Frequency was reported because some papers addressed multiple categories. AI literacy categories were clustered based on Ng et al. [51]'s four domains.**

| Design Suggestions (DS1—DS9) | Description |
|---|---|
| **Foster Critical Engagement with AI Reasoning (DS1)** | Design activities that prompt learners to interrogate AI outputs rather than passively accepting them: test assumptions, compare AI- and human-generated solutions, identify contradictions, justify decisions with evidence, and surface confidence/provenance via validity markers and structured reflection prompts [14]. When introducing LLMs, frame them as evolving systems to encourage exploration while recognizing limitations and uncertainty [67]. |
| **Integrate AI/ML Ethics and Bias Literacy with Plain Language (DS2)** | Expose learners to different bias types and de-biasing strategies; use dilemma-based gameplay to practice fairness, accountability, and responsibility; avoid heavy ML jargon and use age-appropriate language [8, 36, 66, 68]. Collaborative discussion can broaden ethical perspective-taking [67]. |
| **Sustain Engagement with Purposeful Game Mechanics (DS3)** | Address reports of slow or monotonous play by implementing progressive difficulty, meaningful rewards, badges, leaderboards, and narrative challenges aligned with learning goals [49, 88]. Timed challenges and competitive elements can drive engagement [37, 62]. Consider richer visualizations, concept expansion, and optional multiplayer modes to deepen engagement [59]. |
| **Scaffold Problem-Solving and Concept Mastery (DS4)** | Provide embedded guidance (e.g., prompts, worked examples, embodied/virtual agents) to help learners develop systematic strategies and metacognitive habits; design scaffolds that fade over time to encourage new skill acquisition and promote independent reasoning and knowledge transfer [19, 84]. |
| **Leverage Collaboration for Shared Understanding and Equity (DS5)** | Use dyads and small teams to support strategic problem-solving and reduce knowledge gaps; recognize listening as legitimate participation and structure roles to ensure quieter learners benefit and contribute [66, 72]. Maintain high team participation with clear roles, shared artifacts, and facilitator/agent supports [19]. |
| **Use Tangible and Embodied Interfaces to Ground Abstractions (DS6)** | Incorporate physical/AR interactions (e.g., robots, tangible ML prototypes, spatial interfaces, interactive exhibitions) to make abstract AI/ML ideas concrete and memorable; connect conceptual content to embodied action and everyday contexts [44, 61, 83, 92]. Use embodied analogies judiciously to avoid confusion from too many metaphors at once [92]. |
| **Reinforce Prerequisites with Just-in-Time Mini-Lessons (DS7)** | Close gaps between in-game performance and post-test outcomes by embedding concise mini-lessons (e.g., logic, probability, data reasoning) and reflection prompts; align assessments with transfer targets to verify durable understanding [79, 92]. |
| **Promote Inclusivity and Retention Across Learner Attitudes and Identities (DS8)** | Design for low-attitude learners and underrepresented groups (e.g., girls) via supportive messaging, diverse narratives, and choice of content; track and address drop-off; equip teachers with age-appropriate resources and simplified language where needed [21, 31, 45, 49]. |
| **Ensure Accessibility, Robustness, and Scalability in Real Contexts (DS9)** | Design the system so it works well on learners' preferred devices (e.g., tablets and phones with size-responsive interfaces). Ensure classroom-ready robustness through backup plans and offline options, and provide multiple learning pathways that support varied learner needs. Plan for deployment across diverse educational and cultural contexts and for longitudinal use [4, 19, 21, 25, 43, 59, 88]. |

**Table 2: Design suggestions (DS1–DS9) for game- and interaction-based AI literacy derived from the synthesis.**

tangible systems can be printed or assembled using the provided documentation, and digital systems can be downloaded or modified through freely available source code repositories such as GitHub. This level of transparency and accessibility supports reproducibility and broader adoption, but it remains relatively uncommon across our corpus.

## 4.3 Design Suggestions Supporting AI Literacy in Games (RQ2)

From our reflexive thematic analysis of the corpus (Section 3.6.4), we identified nine similar design suggestions (DS1—DS9) for game- and interaction-based AI literacy in our corpus (Table 2). Collectively, they argue that we need to move beyond answer delivery toward *learner-facing AI reasoning*, where students actively interrogate outputs, compare alternatives, and justify decisions with evidence [14].

Systems that foreground *limitations and uncertainty* (especially for LLMs like ChatGPT) invite productive exploration without overtrust [67]. Ethical reasoning and bias literacy benefit from *dilemma-driven play* and *plain language* [36, 66], while sustained engagement hinges on purposeful mechanics (e.g., progression, rewards, narrative) [49, 88]. Collaboration structures (e.g., dyads, small teams) broaden participation and strategy formation, including for quieter learners [66, 72]. Tangible and embodied interfaces help concretize abstractions [61, 83, 92], but we note and *stress* that the gaps between in-game performance and post-test transfer motivate the embedded micro-instruction and prerequisite reinforcement reported by Wang et al. [79] and Zhou et al. [92]. Lastly, inclusive and context-aware designs should be paired with mobile accessibility and classroom robustness to support equitable uptake across settings [45, 59].

*4.3.1 Mapping AI Literacy Domains to Game Design Elements.* To understand which game design elements researchers use to support

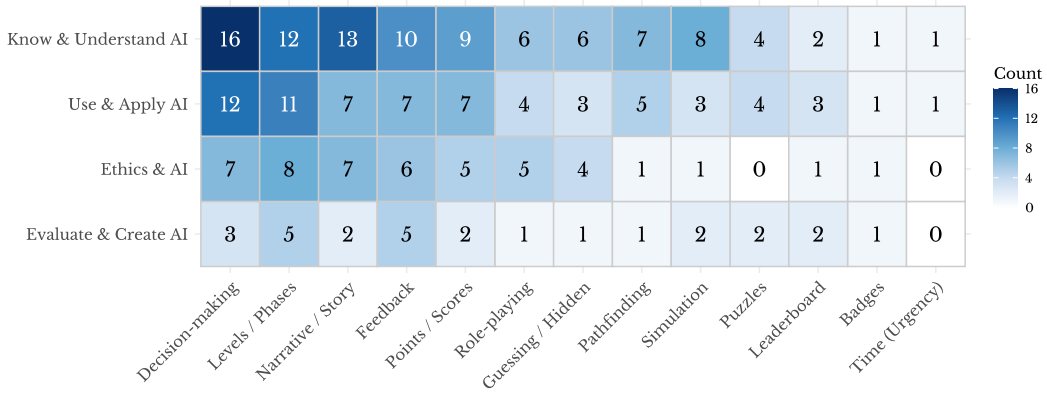



**Figure 8: Co-occurrences of AI literacy domains and game design elements for the 48 design artifacts in our corpus.**

specific AI literacy competencies, and identify potential gaps in how the corpus addresses different domains, we constructed a co-occurrence heatmap (Figure 8). The concentration of high values in the top-left quadrant reflects both the dominance of *Know & Understand AI* as a learning objective and the reliance on a small set of core mechanics, specifically decision-making, narrative, levels, and feedback, to deliver AI literacy content. *Evaluate & Create AI* has the lowest co-occurrence counts overall, in alignment with our findings suggesting that examined design artifacts less frequently target higher-order AI competencies (Section 4.2.3).

Building on findings in Section 4.2.3 and Section 4.2.4, decision-making is the most frequently paired game element across all four AI literacy domains, with particularly strong co-occurrence with *Know & Understand AI* (n=16). Narrative/Story, Levels/Phases, and Feedback also appear consistently across domains, suggesting these mechanics serve as versatile vehicles for AI literacy regardless of the targeted competency. In contrast, elements such as Time (Urgency), Badges, and Leaderboards appear infrequently, indicating they are used sparingly or in narrower contexts.

## 5 Discussion

Our scoping review findings articulate how game- and interaction-based systems conceptualize and implement AI literacy. By synthesizing research goals, target audiences, AI literacy domains [51], platforms, modalities, mechanics, and evaluations, we identify what aspects of AI literacy are addressed (**RQ1**) and how design features and mechanisms support these goals (**RQ2**). We discuss the implications of these findings in relation to delivery formats and contexts, coverage and learning processes, and design patterns for AI literacy.

We stress that the design suggestions presented here are derived from our synthesis of existing work, and should be treated as a developing rather than definitive list. Our corpus includes design papers, works-in-progress, and preliminary results alongside full empirical studies, meaning some suggestions rest on early-stage evidence that has yet to be validated through rigorous evaluation, which is consistent with the exploratory nature of scoping reviews.

### 5.1 Delivery Formats and Infrastructure for AI Literacy

Our analysis highlights that youth (primarily students) engagement with AI literacy is not confined to digital formats. Work by Maljur et al. [44] shows that analogue, tangible, and embodied interactions appear across our corpus as viable alternatives to screen-only activities for delivering AI literacy content, particularly when the goal is to make opaque computational processes observable and manipulable. Beyond AR platforms that illustrate AI reasoning [92], several artifacts demonstrate the promise of low-infrastructure tangible designs: card- and table-top formats for data ethics and bias awareness [3, 8, 68, 81], floor mats and button-pad interfaces that externalize algorithmic processes [61], and robot-mediated environments that synchronize physical and virtual states to make policy and path planning observable [83]. Tangible escape-room sequences further show how embodied puzzle structures can anchor abstract concepts such as embeddings, reinforcement learning, and human-in-the-loop coordination in physical action [19]. These approaches make model behaviour, constraints, and trade-offs more legible than on-screen widgets alone (as seen in Figure 7). Work by Rodić and Granić [63] further supports this direction, showing that tangible interfaces reduce cognitive load and support collaborative meaning-making for young children in ways screen-based interfaces cannot.

However, the same concreteness that makes tangible formats effective for AI literacy also introduces risks specific to how learners form mental models of AI systems. When multiple physical metaphors are layered without careful sequencing, learners can overgeneralize from the tangible representation to the computational process. For example, concluding that an AI recommendation system literally "traps" users the way bees are trapped in a physical game might impede learners' understanding of the underlying statistical filtering mechanism [92]. This is a distinctly AI literacy-specific concern: unlike mathematics or physics, where physical manipulatives map relatively directly onto the target concept, AI/ML processes (e.g., gradient descent, embedding spaces,

stochastic optimization) have no natural physical analogue. Therefore, every tangible metaphor is a deliberate simplification that risks being mistaken for the thing itself (👆 DS6). Explicit debriefs tying physical affordances to dataset properties, uncertainty, and model limitations can reduce this risk.

Additionally, physical and print-ready formats can help mitigate digital divide constraints in under-resourced classrooms by reducing reliance on personal devices and broadband. Card games and printable kits are low-cost and easy to localize [3, 68, 81]. This is particularly salient for AI literacy because access to AI tools and infrastructure is already unevenly distributed; analogue formats ensure that learning about AI does not itself require AI infrastructure. That said, analogue formats have their limitations: they cannot provide the characteristic real-time feedback loops of actual AI systems (e.g., a classifier updating its predictions as new data arrives, or visible agent improvements over training iterations). This means that learners may develop accurate conceptual models of how AI works in principle, while lacking experience with the temporal dynamics, scale, and unpredictability of live AI systems.

Although there is increased interest in mobile learning, most systems in our analysis remain computer-based and only a small subset target phones or tablets. These patterns highlight a need to invest in tangible-first, low-infrastructure designs that align with real classroom realities while being transparent about what the physical metaphor does and does not capture about the AI system it represents.

*5.1.1 Teacher Readiness and Facilitation.* Teachers' confidence for teaching AI literacy concepts and age-appropriate materials remain limited [31]. This gap is more acute for AI than established subjects like mathematics because teacher preparation programs rarely cover AI/ML concepts (♿ DS8). AI literacy content requires facilitators who understand AI/ML concepts themselves; unlike general game facilitation, guiding learners through algorithmic reasoning, bias identification, or model evaluation demands domain-specific expertise that limits who can run these activities (👥 DS5, 📱 DS9) [47]. Even when tools are available, teachers report a lack of necessary resources for AI content delivery [31]. For example, several systems rely on external facilitators such as a game master or an AI-supported moderator, restricting replication [47, 67]. Some design artifacts use LLMs such as ChatGPT for assistance, but this becomes fairly outdated during longitudinal studies.

Our findings also echo work by Kangas et al. [33] which notes that, depending on the learning goals and the game context, the teacher's role varies from a leader to a facilitator and from an organizer and planner of pedagogical entities to a guide and a tutor during the gameplay sessions. From a pedagogical standpoint, sometimes the most important "teachable moments" [82] happen during play, while in other cases the teacher's involvement as an organizer or co-learner becomes essential. This means teachers also operate as bridges between the game world and everyday experiences, helping students connect in-game model behaviour to the AI systems they encounter in daily life, such as recommendation algorithms, voice assistants, or content filters. In-system facilitation, such as adaptive onboarding, role prompts, and pause-and-discuss checkpoints, can shift this burden [19].

*5.1.2 Cultural Relevance and LLM Dependencies.* Localized datasets and culturally relevant scenarios further improve alignment with learners' norms [4, 43, 67]. Shrivastava et al. [67] found that ChatGPT's Western-biased outputs can misrepresent culturally specific concepts, directly affecting how learners evaluate and apply AI literacy skills in non-Western contexts. AI literacy tools that depend on live APIs (e.g., GPT-4o, real-time model training) also introduce infrastructure dependencies that general educational games do not face: network outages, API costs, rate limits, and model version changes can break the learning experience in ways that are outside the designer's control (📱 DS9) [54]. Learners' self-assessments of their own AI knowledge are further complicated by the Dunning–Kruger effect, where novices overestimate their AI understanding while experienced learners underestimate it, making self-reported AI literacy a poor basis for adaptive support or placement (♿ DS8) [45].

> These patterns suggest that low-infrastructure designs transparent about their representational limits, in-system facilitation supports that reduce reliance on teacher AI expertise, and defensive integration of LLMs against cultural bias and API instability can make games and gamified applications for AI literacy more accessible, more pedagogically sound, and more durable in real classroom settings.

## 5.2 AI Literacy Coverage and Learning Processes

The design artifacts from our corpus predominantly target the AI literacy domain of *Know & Understand AI*, followed by *Use & Apply AI*, *Ethics & AI*, and *Evaluate & Create AI* (Section 4.2.3). This ordering reflects a pragmatic sequence where foundational understanding precedes application [49]. As shown in the co-occurrence heatmap (Figure 8), *Know & Understand AI* dominates across nearly all game design elements, while *Evaluate & Create AI* shows the lowest co-occurrence counts overall, suggesting that game designs in our corpus less frequently target higher-order AI competencies.

Just-in-time mini-lessons on probability, data reasoning, and logic reinforce performance within gameplay and improve transfer to novel tasks [79]. For instance, data literacy (including dataset selection, labelling, sampling, and bias mitigation) is less frequently addressed but is crucial to understanding AI systems [24, 68].

*5.2.1 Conceptual Clarity and Misconceptions.* A major concern is conceptual clarity. Many works blur the line between games and gamification, or between AI literacy and general STEM or coding activities. While programming-heavy systems can be valuable, they do not inherently support learners' understanding of model limitations, uncertainty, or bias unless explicitly surfaced. We emphasize that researchers should clarify these boundaries when designing experiences that target AI literacy to avoid targeting adjacent competencies instead.

Learner misconceptions are also common, necessitating deliberate design. Confusion arises when too many analogies or abstract concepts are presented simultaneously [92], and exploratory trial-and-error often precedes structured strategy formation [84].

Over-trust of model outputs and misinterpretation of uncertainty remain recurring issues, emphasizing the potential value of "predict–observe–explain" cycles, contradiction highlighting via validity markers, and structured reflection prompts highlighted by Dangol et al. [14].

*5.2.2 The Gap Between Experiential and Assessable AI Knowledge.* A persistent tension across our corpus is that learners often develop functional intuitions about AI behaviour without acquiring the necessary vocabulary or conceptual precision to demonstrate their understanding on assessments. For example, in Beetrap, students could describe AI phenomena such as filter bubbles in their own words but struggled to use correct AI terminology [92]. This gap between experiential AI literacy and assessable AI knowledge complicates the simultaneous use of DS1, DS4, and DS7: fostering critical engagement with AI reasoning (DS1) may produce rich in-context understanding that scaffolds learning (DS4) but mini-lessons (DS7) fail to translate into transferable knowledge. The transfer problem appears directly in ARIN-561, where in-game tutorial performance on search algorithms was only weakly predictive of success on transfer tasks [79], and in Beetrap, where learners who successfully manipulated recommendation algorithms in-app did not consistently transfer that understanding to post-test items [92]. These findings suggest that procedural fluency with AI tasks—training a classifier, building a decision tree, executing a search algorithm—may develop independently from conceptual AI understanding.

A related challenge affects DS2: simplifying AI ethics language to make bias and fairness concepts accessible risks stripping them of technical precision. Even with plain language, learners who engaged with AI Audit struggled to distinguish between related concepts such as fairness, accountability, and transparency [3]. Moreover, reducing bias assessment to binary correct/incorrect outcomes cannot capture the contextual and contested nature of algorithmic bias, where reasonable people may disagree on what counts as biased [66]. Gameplay exposure may raise awareness of specific bias types encountered in-game while leaving learners unable to recognize novel forms of bias in unfamiliar AI systems [66]. This implies that ethical reasoning can be introduced in short sessions through dilemma-based gameplay. However, several works call for longer observation to measure retention and transfer [4, 25, 38], as one prominent game design element was "decision-making" (see Figure 7).

> This suggests that we can improve AI literacy coverage and learning processes through games and gamified applications by sharpening the conceptual focus of AI literacy (e.g., distinguishing it from general computational thinking), making concepts like uncertainty, bias, data quality, and model limitations visible within gameplay, using misconception-sensitive sequencing and reflection to reinforce these ideas, and addressing the gap between experiential understanding and assessable AI knowledge by aligning in-game scaffolds with transfer-oriented assessments.

## 5.3 Design Patterns and Game Mechanics

Critical engagement (DS1) is emphasized through activities that prompt learners to interrogate AI outputs, compare human and AI solutions, and justify decisions using evidence. In short, validity markers and structured reflection support trust calibration, while introducing LLMs as evolving tools encourages exploration without fostering overconfidence [14, 67]. Dilemma-based gameplay and collaborative discussion reinforce ethical and bias literacy (DS2), exposing learners to multiple perspectives and concrete bias types [36, 66, 68].

*5.3.1 Engagement Mechanics and AI-Specific Learning.* Not all game mechanics need to directly encode AI concepts. Leaderboards, badg-es, and narrative framing serve ludological functions: they sustain motivation, regulate pacing, and provide affective scaffolding to keep learners engaged with the experience long enough for AI-specific learning to occur [49, 88]. The design challenge is not to make every element teach AI, but to ensure that the elements which do carry conceptual weight are not drowned out by those which do not. This is an important distinction because over-instrumentalizing engagement mechanics can flatten gameplay into a compliance exercise, while leaving them entirely disconnected from AI content risks learners treating the AI material as an obstacle between rewards. However, our analysis suggests that many engagement mechanics operate at a surface level largely disconnected from AI understanding (DS3). Competitive mechanics such as timers can redirect learner attention toward gaming the AI system for points rather than understanding how it works like in ClawAI, where students optimized descriptions for scoring rather than reflecting on how text classification operates [37].

AI concepts such as training, iteration, and model refinement are inherently repetitive processes, creating tension between faithful representations of how AI works and maintaining engaging pacing [88]. Some AI processes, such as the training of reinforcement learning agents training where improvements only appear after many iterations, have inherently delayed payoffs that conflict with the need for immediate, visible feedback in game-based learning [28]. Learners may also engage with game rewards (e.g., badges, leaderboards) without connecting them to the underlying AI learning objectives, producing high game engagement but shallow AI understanding [49].

A key opportunity is the re-purposing of existing mechanics so that progression, rewards, or narrative branches respond to how learners interrogate or correct AI behaviour instead of generic performance metrics [59]. The co-occurrence heatmap (Figure 8) also reveals that elements such as Time (Urgency), Badges, and Leaderboards appear infrequently across all AI literacy domains, indicating these mechanics are used sparingly despite being commonly recommended for sustaining engagement (DS3). In contrast, decision-making and feedback are the most consistently paired elements, suggesting designers default to these mechanics regardless of the specific AI literacy competency being targeted. While agents, prompts, and guided strategies support metacognition and performance [19, 84], these supports often exist alongside gameplay rather than being intertwined with AI-specific reasoning. AI/ML concepts involve layered abstractions (data, algorithm, model) that learners may navigate through trial-and-error even when scaffolds

are available, because the system's internal logic is not directly observable [84]. Scaffolds that help learners complete AI tasks in-game (e.g., training a classifier, building a decision tree) may produce procedural fluency without supporting transfer to reasoning about AI systems in new contexts (DS4) [79]. As seen in designs that allow or encourage repeated "guessing" as a game design element, such patterns raise an important pedagogical concern: if learners rely on guessing to progress, to what extent are they actually developing AI literacy? Embedding scaffolds that direct attention explicitly toward data quality, model behaviour, and uncertainty can help strengthen the connection between gameplay actions and conceptual development [14].

*5.3.2 Single-Player vs. Collaborative Modalities.* Our analysis found that most design artifacts centre player–system interaction, with relatively few structuring collaboration around AI-specific reasoning tasks such as joint inspections of misclassified data or debating dataset composition [72]. The limited examples that do, such as Co-ML's shared dataset construction [72] and Equality Engine's dyadic bias identification [66], report that collaboration surfaced perspectives that individual learners missed, suggesting this is an under-explored design space rather than an established best practice. Our findings show a strong reliance on system-generated feedback as a primary design element. However, Li et al. [40] highlight that human-human feedback can be just as essential in shaping learning experiences.

Multiplayer configurations are not inherently superior for AI literacy. Single-player formats afford a distinct cognitive mode: uninterrupted reflection, self-paced experimentation with model parameters, and the freedom to make mistakes without social pressure, all support the kind of deliberate hypothesis testing that deepens understanding of AI system behaviour [84]. Collaborative formats, by contrast, distribute cognitive load and surface diverse perspectives on data and bias [66, 72], but can allow passive participation to go undetected [72]. Verifying whether passive participants genuinely develop AI understanding, or merely defer to more vocal group members, remains methodologically difficult (DS5). When collaboratively building ML datasets or configuring AI agents, the range of available parameters (e.g., distance metrics, policy options) can also overwhelm learners, making collaborative decision-making confusing rather than productive [83]. The choice between modalities should therefore be driven by the specific AI literacy domain being targeted: individual exploration may better serve procedural understanding of algorithms (*Know & Understand AI*), while collaborative formats may better support ethical deliberation (*Ethics & AI*) where multiple stakeholder perspectives are inherently part of the learning objective. Thus, rather than assuming multiplayer is always preferable, designers should consider structuring collaborative roles that obligate learners to contribute distinct forms of AI reasoning (e.g., one student monitors uncertainty while another evaluates input features), which could support deeper collective understanding.

*5.3.3 Evaluation Limitations.* In-game performance on AI tasks (e.g., executing search algorithms, classifying data) is only weakly predictive of learners' ability to reason about those same AI concepts on transfer assessments (DS7), suggesting that procedural AI task completion and conceptual AI understanding may develop independently [79]. Therefore, our findings indicate that evaluation practices limit the field's ability to draw strong inferences about learners' AI literacy gains. Many studies use brief interventions, convenience samples, or limited controls, which constrains claims about causal effects or durable learning [83, 92]. Although short sessions can yield measurable gains in ethics awareness [38], extended and mixed-method designs are necessary to distinguish between novelty-driven engagement and conceptual retention [25, 51]. Likewise, inconsistent measures of engagement, ethics literacy, and trust calibration make it difficult to compare findings across systems or identify reliably effective mechanics.

> This suggests that game design and evaluation for AI literacy can be advanced by integrating AI-specific reasoning directly into core mechanics rather than treating engagement elements as conceptual carriers, structuring collaborative roles around distinct analytical tasks (e.g., monitoring uncertainty vs. evaluating features), matching single-player and multiplayer modalities to the AI literacy domain being targeted, designing scaffolds that bridge procedural AI task completion and conceptual transfer, and adopting shared evaluation frameworks that distinguish novelty-driven engagement from durable learning—making games and gamified applications for AI literacy more conceptually integrated, more pedagogically intentional, and more rigorously evaluated.

## 6 Limitations & Future Work

Like all reviews, our work is subject to several limitations. First, it is possible that our search terms did not return all relevant AI and ML studies because database indexing sometimes treats the term "AI" as a stop word. Given our focus on AI literacy and games, we prioritized the capture of educational and design-oriented work over technical ML optimization papers. To accommodate this constraint, we followed best-practice scoping procedures outlined by Tu et al. [73] instead of the medical-oriented PRISMA Scoping protocol. This is a deliberate trade-off: while our search might have excluded some ML-specific papers, we determined that ML-heavy studies often centre system performance or model engineering rather than the design, interaction, and learning mechanisms that our review examines.

Another limitation concerns the inclusion of works-in-progress, posters, and early-stage design reports (typically excluded). However, with the rapid publishing rate noted in our findings (Section 4.2), we felt it necessary to include them in our scoping review. Although these sources provide important visibility into emerging directions, they typically offer limited methodological detail, abbreviated descriptions of design features, or incomplete evaluation results. Consequently, several of our categorizations (such as AI literacy coverage, game mechanics, facilitation requirements, and evaluation strength) rely on partial information. This might result in conservative tagging, underreporting system capabilities, or incomplete interpretations of how designers intended for mechanics to support learning. Finally, the concept of AI literacy is still evolving, and there is no consensus on a unified framework or best practices for evaluating learning outcomes in game-based

AI literacy interventions. This lack of shared definitions complicates cross-study comparison, highlighting the need for clearer conceptual and methodological guidance in the field.

For future work, we hope that this review provides a foundation for researchers developing frameworks and standardized instruments for assessing AI literacy outcomes in game-based or gamified learning environments. As this is a scoping review, the next logical step is a systematic review that examines the effectiveness of individual design strategies, compares short- and long-term learning gains, and evaluates which mechanics, scaffolds, and interaction modalities most consistently produce meaningful improvements in AI literacy. Such a systematic review would also enable meta-analytic insight into transfer, misconceptions, and equity outcomes, thereby refining the design space mapped here. Along with a systematic review, designing and evaluating one or more artifacts embodying extracted design suggestions (DS1–DS9) that link game elements, pedagogical practices, and learning outcomes would be useful demonstrations of these findings, informing future AI literacy-focused games and gamified system design efforts. Such artifacts would also allow us to empirically investigate different learning strategies, combinations of game elements and AI literacy dimensions, interaction modalities, and longitudinal learning outcomes. Together, the systematic review and design artifacts would create a foundation of a design framework for developing AI literacy through games and gamified applications.

## 7 Conclusion

Our scoping review examines 45 papers containing 48 design artifacts to provide a consolidated view of how AI literacy is supported through game- and interaction-based learning systems. By mapping the dimensions of AI literacy addressed across platforms, modalities, and play configurations (**RQ1**) and identifying the design elements and broader considerations that underpin these learning goals (**RQ2**), we highlight both strengths and gaps in the field. Current systems often succeed at introducing foundational concepts, sustaining engagement through familiar mechanics (e.g., decision-making, feedback, levels, and narrative). Our work is not intended as a critique of existing efforts in AI literacy and games. Instead, our collated design suggestions (DS1–DS9) highlight opportunities for deepening this emerging area by emphasizing critical engagement with AI literacy.

## Acknowledgments

Foremost, this work would not have been possible without the help of our local librarians at the University of Waterloo, mainly Timothy Ireland. We thank our colleagues at Fondazione Bruno Kessler for their contributions to European and international AI research. This study was supported by the NSERC Discovery Grant (RGPIN-2023-03705), and the CFI John R. Evans Leaders Fund (CFI JELF-41844). Software for proofreading and grammar, such as Grammarly (with AI), assisted in restructuring some of the sentences and reducing the words within the paper to help with readability. **Supplementary material** is available on the Open Science Framework at: https://osf.io/3b2p9/.

## References

[1] Alaa Abd-alrazaq, Israa Abuelezz, Asma Hassan, AlHasan AlSammarraie, Dari Alhuwail, Sara Irshaidat, Hashem Abu Serhan, Arfan Ahmed, Sadam Alabed Alrazak, and Mowafa Househ. 2022. Artificial Intelligence–Driven Serious Games in Health Care: Scoping Review. *JMIR Serious Games* 10, 4 (Nov. 2022), e39840. doi:10.2196/39840

[2] C. Aguiar, D. Carpenter, J. Vandenberg, W. Min, V. Cateté, and B. Mott. 2025. Fostering AI Literacy Through Strategic Play: A Competitive Pathfinding Game for Middle School. In *2025 IEEE conference on games (CoG)*. IEEE, Piscataway, NJ, USA, 1–4. doi:10.1109/CoG64752.2025.11114117

[3] S. Ali, V. Kumar, and C. Breazeal. 2023. AI Audit: A Card Game to Reflect on Everyday AI Systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. AAAI, Washington, DC, USA, 15981–15989. doi:10.1609/aaai.v37i13.26897

[4] Z. Anthis and M. Mavrikis. 2025. In Search of Answers: Supporting Educators to Understand AI Through Heuristic Thinking. In *Artificial Intelligence in Education*, Vol. 15880 LNAI. Springer Nature Switzerland, Cham, 119–133. doi:10.1007/978-3-031-98459-4_9

[5] Hilary Arksey and Lisa O'Malley. 2005. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 8, 1 (Feb. 2005), 19–32. doi:10.1080/1364557032000119616

[6] Katrin Becker. 2021. What's the difference between gamification, serious games, educational games, and game-based learning? 4 pages. doi:10.20935/AL209 Academic Letter.

[7] Martina Benvenuti, Angelo Cangelosi, Armin Weinberger, Elvis Mazzoni, Mariagrazia Benassi, Mattia Barbaresi, and Matteo Orsoni. 2023. Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior* 148 (Nov. 2023), 107903. doi:10.1016/j.chb.2023.107903

[8] Maalvika Bhat. 2025. Designing AI interfaces for transparent decision-making and ethical reflection. In *Companion proceedings of the 30th international conference on intelligent user interfaces (IUI '25 companion)*. Association for Computing Machinery, New York, NY, USA, 211–214. doi:10.1145/3708557.3716150

[9] Benjamin S Bloom, Max D Engelhart, Edward J Furst, Walker H Hill, and David R Krathwohl. 1956. *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* Longmans, Green, New York, NY, USA.

[10] Virginia Braun and Victoria Clarke. 2022. *Thematic analysis: a practical guide.* SAGE, Los Angeles London New Delhi Singapore Washington DC Melbourne.

[11] Virginia Braun and Victoria Clarke. 2023. Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a *knowing* researcher. *International Journal of Transgender Health* 24, 1 (Jan. 2023), 1–6. doi:10.1080/26895269.2022.2129597

[12] Lorena Casal-Otero, Alejandro Catala, Carmen Fernández-Morante, Maria Taboada, Beatriz Cebreiro, and Senén Barro. 2023. AI literacy in K-12: a systematic literature review. *International Journal of STEM Education* 10, 1 (April 2023), 29. doi:10.1186/s40594-023-00418-7

[13] Lijia Chen, Pingping Chen, and Zhijian Lin. 2020. Artificial Intelligence in Education: A Review. *IEEE Access* 8 (2020), 75264–75278. doi:10.1109/ACCESS.2020.2988510

[14] Aayushi Dangol, Runhua Zhao, Robert Wolfe, Trushaa Ramanan, Julie A. Kientz, and Jason Yip. 2025. "AI just keeps guessing": Using ARC puzzles to help children identify reasoning errors in generative AI. In *Proceedings of the 24th interaction design and children (Idc '25)*. Association for Computing Machinery, New York, NY, USA, 444–464. doi:10.1145/3713043.3728836

[15] Hasti Darabipourshiraz. 2025. AI DoodleLab: Fostering middle school students' AI literacy through project-based creative drawing. In *Proceedings of the 2025 conference on creativity and cognition (C&C '25)*. Association for Computing Machinery, New York, NY, USA, 64–68. doi:10.1145/3698061.3734417

[16] Hojjat Dehghanzadeh, Mohammadreza Farrokhnia, Hossein Dehghanzadeh, Kiumars Taghipour, and Omid Noroozi. 2024. Using gamification to support learning in K-12 education: A systematic literature review. *British Journal of Educational Technology* 55, 1 (2024), 34–70. doi:10.1111/bjet.13335

[17] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (Tampere, Finland) *(MindTrek '11)*. Association for Computing Machinery, New York, NY, USA, 9–15. doi:10.1145/2181037.2181040

[18] Darina Dicheva, Keith Irwin, and Christo Dichev. 2018. Motivational Factors in Educational Gamification. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. IEEE Computer Society, Los Alamitos, CA, USA, 408–410. doi:10.1109/ICALT.2018.00102

[19] D. DiPaola, J. Shen, R. Hu, S. Alghowinem, and C. Breazeal. 2023. DRONEscape: Designing an Educational Escape Room for Adult AI Literacy. In *2023 IEEE conference on games (CoG)*. IEEE, New York, NY, USA, 1–8. doi:10.1109/CoG57401.2023.10333148

[20] Patrick Felicia and Robert Tennyson (Eds.). 2011. *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches*. IGI Global, Hershey, PA, USA. doi:10.4018/978-1-60960-495-0

[21] S. Franco, M.G. Ballatore, S. Bernardini, G. Como, L. Damonte, A. Neacsu, and M. Vanelli. 2024. Unlocking AI Literacy: The SMaILE-App Gaming Experience. In *2024 IEEE gaming, entertainment, and media conference (GEM)*. IEEE, New York, NY, USA, 1–6. doi:10.1109/GEM61861.2024.10585394

[22] Michail Giannakos, Iro Voulgari, Sofia Papavlasopoulou, Zacharoula Papamitsiou, and Georgios Yannakakis. 2020. Games for Artificial Intelligence and Machine Learning Education: Review and Perspectives. In *Non-Formal and Informal Science Learning in the ICT Era*, Michail Giannakos (Ed.). Springer Singapore, Singapore, 117–133. doi:10.1007/978-981-15-6747-6_7

[23] Federica Gini, Simone Bassanelli, Federico Bonetti, Reza Hadi Mogavi, Antonio Bucchiarione, and Annapaola Marconi. 2025. The role and scope of gamification in education: A scientometric literature review. *Acta Psychologica* 259 (2025), 105418. doi:10.1016/j.actpsy.2025.105418

[24] Alex Goslen, Claire Aguiar, Jessica Vandenberg, Veronica Cateté, Wookhee Min, and Bradford Mott. 2025. Playful Pathways to AI Literacy: Designing NLP Activities for Middle School AI Education. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium, Blue Sky, and WideAIED*, Alexandra I. Cristea, Erin Walker, Yu Lu, Olga C. Santos, and Seiji Isotani (Eds.). Vol. 2592. Springer Nature Switzerland, Cham, 60–67. doi:10.1007/978-3-031-99267-4_8

[25] Anisha Gupta, Seung Lee, Bradford Mott, Srijita Chakraburty, Krista Glazewski, Anne Ottenbreit-Leftwich, Adam Scribner, Cindy E. Hmelo-Silver, and James Lester. 2024. Supporting Upper Elementary Students in Learning AI Concepts with Story-Driven Game-Based Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 21 (March 2024), 23092–23100. doi:10.1609/aaai.v38i21.30354

[26] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences*. IEEE, Waikoloa, HI, 3025–3034. doi:10.1109/HICSS.2014.377

[27] Ken Hartness. 2004. Robocode: using games to teach artificial intelligence. *J. Comput. Sci. Coll.* 19, 4 (April 2004), 287–291.

[28] Cesar Hinojosa, Priyanka Kumar, Pragathi Durga Rajarajan, and Fred Martin. 2025. TrainYourSnakeAI: a novel tool to teach reinforcement learning to middle school students. In *Proceedings of the 56th ACM technical symposium on computer science education V. 1 (Sigcsets 2025, Vol. 1)*. Association for Computing Machinery, Pittsburgh, PA, USA, 506–512. doi:10.1145/3641554.3701907

[29] Chengpeng Hu, Yunlong Zhao, Ziqi Wang, Haocheng Du, and Jialin Liu. 2024. Games for artificial intelligence research: A review and perspectives. *IEEE Transactions on Artificial Intelligence* 5, 12 (2024), 5949–5968. doi:10.1109/TAI.2024.3410935

[30] William Huitt. 2011. Bloom et al.'s taxonomy of the cognitive domain. *Educational psychology interactive* 22 (2011), 1–4.

[31] Nicolas Ioannou, Maria Kasinidou, Anthi Ioannou, and Styliani Kleanthous. 2025. From teachers to students: Evaluating canvas city as a path to AI literacy. In *Proceedings of the 30th ACM conference on innovation and technology in computer science education V. 1 (ITiCSE 2025)*. Association for Computing Machinery, Nijmegen, Netherlands, 16–22. doi:10.1145/3724363.3729069

[32] Tanisha Jowsey, Virginia Braun, Victoria Clarke, Deborah Lupton, and Michelle Fine. 2025. We reject the use of generative artificial intelligence for reflexive qualitative research. doi:10.2139/ssrn.5676462

[33] Marjaana Kangas, Antti Koskinen, and Leena Krokfors. 2016. A qualitative literature review of educational games in the classroom: the teacher's pedagogical activities. *Teachers and Teaching* 23, 4 (July 2016), 451–470. doi:10.1080/13540602.2016.1206523

[34] Firas Layth Khaleel, Noraidah Sahari Ashaari, and TSMT Wook. 2020. The impact of gamification on students learning engagement. *International Journal of Electrical and Computer Engineering* 10, 5 (2020), 4965–4972. doi:10.11591/ijece.v10i5.pp4965-4972

[35] Arif Ali Khan, Sher Badshah, Peng Liang, Muhammad Waseem, Bilal Khan, Aakash Ahmad, Mahdi Fahmideh, Mahmood Niazi, and Muhammad Azeem Akbar. 2022. Ethics of AI: A Systematic Literature Review of Principles and Challenges. In *The International Conference on Evaluation and Assessment in Software Engineering 2022*. ACM, Gothenburg Sweden, 383–392. doi:10.1145/3530019.3531329

[36] Sung-Eun Kim, Kyuwon Kim, Jeanhee Lee, Yeji Ko, Yue Wang, and Hyo-Jeong So. 2025. Dilemmas in AI Ethics: A Digital Game for Moral Reasoning and Collective Decision-Making. In *Artificial Intelligence in Education*, Alexandra I. Cristea, Erin Walker, Yu Lu, Olga C. Santos, and Seiji Isotani (Eds.). Vol. 15878. Springer Nature Switzerland, Cham, 434–447. doi:10.1007/978-3-031-98417-4_31

[37] Aishat Kolawole, Yeshaswini Parvatham, and Fred Martin. 2025. ClawAI: a software tool teaching text classification and AI systems to middle school students. In *2025 IEEE integrated STEM education conference (ISEC)*. IEEE, New York, NY, USA, 1–7. doi:10.1109/ISEC64801.2025.11147299

[38] A. Koskinen, A. Lindstedt, and K. Kiili. 2025. Game-Based Learning Can Enhance Students' Understanding of AI Ethics. In *Games and Learning Alliance*, Vol. 15348 LNCS. Springer Nature Switzerland, Cham, 47–56. doi:10.1007/978-3-031-78269-5_5

[39] Irene Lee and Safinah Ali. 2021. The Contour to Classification Game. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 17 (May 2021), 15583–15590. doi:10.1609/aaai.v35i17.17835

[40] Guang Li, Mohammad Amin Zarei, Goudarz Alibakhshi, and Akram Labbafi. 2024. Teachers and educators' experiences and perceptions of artificial-powered interventions for autism groups. *BMC Psychology* 12, 1 (April 2024), 199. doi:10.1186/s40359-024-01664-2

[41] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems (Chi '20)*. Association for Computing Machinery, Honolulu, HI, USA, 1–16. doi:10.1145/3313831.3376727

[42] Qianou Ma, Anika Jain, Jini Kim, Megan Chai, and Geoff Kaufman. 2025. ImaginAItion: Promoting Generative AI Literacy Through Game-Based Learning. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 331, 9 pages. doi:10.1145/3706599.3719844

[43] Pragati Maheshwary, Aditi Haiman, Emma Brown, Aarnav Sangekar, and Canwen Wang. 2025. Case 429: a murder mystery game for teaching bias in generative AI. In *Companion proceedings of the annual symposium on computer-human interaction in play (CHI PLAY companion '25)*. Association for Computing Machinery, New York, NY, USA, 316–321. doi:10.1145/3744736.3749366

[44] Tina Maljur, Jasminko Novak, Jessica Laufer, and Sarah Gnoth. 2025. AI exploratorium: a gamified approach to AI literacy through experiential learning in interactive exhibitions. In *Proceedings of the 30th ACM conference on innovation and technology in computer science education V. 2 (ITiCSE 2025, Vol. 2)*. Association for Computing Machinery, Nijmegen, Netherlands, 784. doi:10.1145/3724389.3730782

[45] Dylan Edward Moore, Sophia R. R. Moore, Bansharee Ireen, Winston P. Iskandar, Grigory Artazyan, and Elizabeth L. Murnane. 2024. Teaching artificial intelligence in extracurricular contexts through narrative-based learnersourcing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 270, 28 pages. doi:10.1145/3613904.3642198

[46] B. Mott, J. Vandenberg, C. Penilla, S. Hennigan, R. Navarro, J. Lester, and E. Ozer. 2025. Designing a Narrative-Centered Game to Promote AI Literacy and Health Career Exploration. In *2025 IEEE conference on games (CoG)*. IEEE, Piscataway, NJ, USA, 1–4. doi:10.1109/CoG64752.2025.11114233

[47] T. Mott, M. Higger, A. Bejarano, and T. Williams. 2024. Degrees of Freedom: A Storytelling Game that Supports Technology Literacy about Social Robots. In *2024 33rd IEEE international conference on robot and human interactive communication (ROMAN)*. IEEE, New York, NY, USA, 2095–2102. doi:10.1109/RO-MAN60168.2024.10731340

[48] D.T.K. Ng, J. Su, and S.K.W. Chu. 2024. Fostering Secondary School Students' AI Literacy through Making AI-Driven Recycling Bins. *Education and Information Technologies* 29, 8 (2024), 9715–9746. doi:10.1007/s10639-023-12183-9

[49] D.T.K. Ng, C. Xinyu, J.K.L. Leung, and S.K.W. Chu. 2024. Fostering students' AI literacy development through educational games: AI knowledge, affective and cognitive engagement. *Journal of Computer Assisted Learning* 40, 5 (2024), 2049–2064. doi:10.1111/jcal.13009

[50] Davy Tsz Kit Ng, Jac Ka Lok Leung, Kai Wah Samuel Chu, and Maggie Shen Qiao. 2021. AI Literacy: Definition, Teaching, Evaluation and Ethical Issues. *Proceedings of the Association for Information Science and Technology* 58, 1 (Oct. 2021), 504–509. doi:10.1002/pra2.487

[51] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2 (2021), 100041. doi:10.1016/j.caeai.2021.100041

[52] Davy Tsz Kit Ng, Jiahong Su, Jac Ka Lok Leung, and Samuel Kai Wah Chu. 2024. Artificial intelligence (AI) literacy education in secondary schools: a review. *Interactive Learning Environments* 32, 10 (Nov. 2024), 6204–6224. doi:10.1080/10494820.2023.2255228

[53] Luci Pangrazio and Julian Sefton-Green. 2021. Digital rights, digital citizenship and digital literacy: What's the difference? *Journal of New Approaches in Educational Research* 10, 1 (Jan. 2021), 15–27. doi:10.7821/naer.2021.1.616

[54] Kieun Park, Hyungwoo Song, Seungbae Seo, Junghwan Kim, and Bongwon Suh. 2025. "Ask Sir Oliver Ingham": LLM-based Social Simulations for History Education. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 2, 13 pages. doi:10.1145/3706599.3719728

[55] Panagiotis Petridis, Vladlena Benson, Mariam Garibyan, Gonçalo Meireles, Alex Carpov, Dimitra Dimitrakopoulou, and Marisa Teles. 2024. Adventure in AI Project (2AI): Promoting AI Knowledge for Kids Aged 7–12 Using Gaming. In *Smart Mobile Communication & Artificial Intelligence*, Vol. 937. Springer Nature Switzerland, Cham, 307–315. doi:10.1007/978-3-031-56075-0_29

[56] M. Pinski, M.-J. Haas, and A. Benlian. 2024. Building Metaknowledge in AI Literacy - The Effect of Gamified vs. Text-based Learning on AI Literacy Metaknowledge. In *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS)*. ScholarSpace, Honolulu, HI, USA, 5164–5173. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85199783495&partnerID=40&md5=9d537140cd2d7c4fd95e0af758ac701b

[57] Jan L. Plass, Bruce D. Homer, and Charles K. Kinzer. 2015. Foundations of Game-Based Learning. *Educational Psychologist* 50, 4 (Oct. 2015), 258–283. doi:10.1080/00461520.2015.1122533

[58] Marc Prensky. 2003. Digital game-based learning. *Computers in entertainment (CIE)* 1, 1 (2003), 21–21.

[59] Shruti Priya, Shubhankar Bhadra, Sridhar Chimalakonda, and Akhila Sri Manasa Venigalla. 2024. ML-Quest: a game for introducing machine learning concepts to K-12 students. *Interactive Learning Environments* 32, 1 (Jan. 2024), 229–244. doi:10.1080/10494820.2022.2084115

[60] Meihua Qian and Karen R. Clark. 2016. Game-based Learning and 21st century skills: A review of recent research. *Computers in Human Behavior* 63 (Oct. 2016), 50–58. doi:10.1016/j.chb.2016.05.023

[61] R. Quiloan, S.R. Moosavi, R. Tankelevich, and M. Karlin. 2023. PlayGrid: Sparking AI Literacy with an Intuitive and Accessible Learning Interface. In *2023 Fifth International Conference on Transdisciplinary AI (TransAI)*. IEEE, New York, NY, USA, 171–178. doi:10.1109/TransAI60598.2023.00045

[62] Pragathi Durga Rajarajan and Fred Martin. 2025. IntoTheRabbitHole: a web application for teaching middle school students about search algorithms. In *Proceedings of the 30th ACM conference on innovation and technology in computer science education V. 1 (ITiCSE 2025)*. Association for Computing Machinery, Nijmegen, Netherlands, 340–346. doi:10.1145/3724363.3729080

[63] Lea Dujić Rodić and Andrina Granić. 2022. Tangible interfaces in early years' education: a systematic review. *Personal and Ubiquitous Computing* 26, 1 (Feb. 2022), 39–77. doi:10.1007/s00779-021-01556-x

[64] Ismaila Temitayo Sanusi, Solomon Sunday Oyelere, Henriikka Vartiainen, Jarkko Suhonen, and Markku Tukiainen. 2023. A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies* 28, 5 (May 2023), 5967–5997. doi:10.1007/s10639-022-11416-7

[65] R. Benjamin Shapiro, Rebecca Fiebrink, and Peter Norvig. 2018. How machine learning impacts the undergraduate computing curriculum. *Commun. ACM* 61, 11 (Oct. 2018), 27–29. doi:10.1145/3277567

[66] Dilruba Showkat, Lingqing Wang, Laveda Chan, and Alexandra To. 2025. Equality Engine: Fostering Critical Machine Learning Bias Literacy Through a Transformational Game. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW068 (May 2025), 31 pages. doi:10.1145/3710966

[67] Vaishnavi Shrivastava, Sumita Sharma, Dipanjan Chakraborty, and Marianne Kinnula. 2024. Is a Sunny Day Bright and Cheerful or Hot and Uncomfortable? Young Children's Exploration of ChatGPT. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction* (Uppsala, Sweden) *(NordiCHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 61, 15 pages. doi:10.1145/3679318.3685397

[68] Jaemarie Solyst, Jennifer Kim, Amy Ogan, and Jessica Hammer. 2022. Data detectives: a tabletop card game about training data. In *Proceedings of the 27th ACM conference on on innovation and technology in computer science education vol. 2 (ITiCSE '22)*. Association for Computing Machinery, Dublin, Ireland, 632. doi:10.1145/3502717.3532128

[69] Jiahong Su, Davy Tsz Kit Ng, and Samuel Kai Wah Chu. 2023. Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. *Computers and Education: Artificial Intelligence* 4 (2023), 100124. doi:10.1016/j.caeai.2023.100124

[70] Katie Salen Tekinbas and Eric Zimmerman. 2003. *Rules of play: Game design fundamentals*. MIT press, Cambridge, MA, USA.

[71] David Touretzky, Christina Gardner-McCune, and Deborah Seehorn. 2023. Machine Learning and the Five Big Ideas in AI. *International Journal of Artificial Intelligence in Education* 33, 2 (June 2023), 233–266. doi:10.1007/s40593-022-00314-1

[72] Tiffany Tseng, Matt J. Davidson, Luis Morales-Navarro, Jennifer King Chen, Victoria Delaney, Mark Leibowitz, Jazbo Beason, and R. Benjamin Shapiro. 2024. Co-ML: Collaborative Machine Learning Model Building for Developing Dataset Design Practices. *ACM Trans. Comput. Educ.* 24, 2, Article 25 (April 2024), 37 pages. doi:10.1145/3641552

[73] Joseph Tu, Lennart Nacke, and Katja Rogers. 2025. Introducing the INSPIRE Framework: Guidelines From Expert Librarians for Search and Selection in HCI Literature. *Interacting with Computers* 38, 2 (Feb. 2025), 179–200. doi:10.1093/iwc/iwaf001

[74] Mohib Ullah, Sareer Ul Amin, Muhammad Munsif, Muhammad Mudassar Yamin, Utkurbek Safaev, Habib Khan, Salman Khan, and Habib Ullah. 2022. Serious games in science education: a systematic literature. *Virtual Reality & Intelligent Hardware* 4, 3 (June 2022), 189–209. doi:10.1016/j.vrih.2022.02.001

[75] Saniya Vahedian Movahed and Fred Martin. 2025. From play to pedagogy: Discovering the ecosystem of AI educational tools and curricula. In *Proceedings of the 30th ACM conference on innovation and technology in computer science education V. 1 (ITiCSE 2025)*. Association for Computing Machinery, Nijmegen, Netherlands, 333–339. doi:10.1145/3724363.3729068

[76] Jessica Vandenberg, Wookhee Min, Veronica Cateté, Danielle Boulden, and Bradford Mott. 2023. Promoting AI education for rural middle grades students with digital game design. In *Proceedings of the 54th ACM technical symposium on computer science education V. 2 (Sigcse 2023)*. Association for Computing Machinery, Toronto ON, Canada, 1388. doi:10.1145/3545947.3576333

[77] Iro Voulgari, Marvin Zammit, Elias Stouraitis, Antonios Liapis, and Georgios Yannakakis. 2021. Learn to machine learn: Designing a game based approach for teaching machine learning to primary and secondary education students. In *Proceedings of the 20th annual ACM interaction design and children conference (Idc '21)*. Association for Computing Machinery, Athens, Greece, 593–598. doi:10.1145/3459990.3465176

[78] Ning Wang, Boxi Fu, Betul Dincer, Omkar Masur, David Faizi, Harshul Ravindran, Julia Wang, Devashish Lai, and Chirag Merchant. 2025. Becoming Fei: An Educational Game for AI and Data Science Education for Novice Learners. In *Learning and Collaboration Technologies*. Vol. 15808. Springer Nature Switzerland, Cham, 69–79. doi:10.1007/978-3-031-93746-0_6

[79] Ning Wang, Eric Greenwald, Ryan Montgomery, and Maxyn Leitner. 2023. The Analysis of Student Errors in ARIN-561 – An Educational Game for Learning Artificial Intelligence for High School Students. In *Learning and Collaboration Technologies*. Vol. 14041. Springer Nature Switzerland, Cham, 570–581. doi:10.1007/978-3-031-34550-0_40

[80] Yuchen Wang, Shangxin Guo, Lin Ling, and Chee Wei Tan. 2024. Nemobot: Crafting strategic gaming LLM agents for K-12 AI education. In *Proceedings of the eleventh ACM conference on learning @ scale (L@S '24)*. Association for Computing Machinery, Atlanta, GA, USA, 393–397. doi:10.1145/3657604.3664671

[81] Zezhong Wang, Shan Hao, and Sheelagh Carpendale. 2024. Card-Based Approach to Engage Exploring Ethics in AI for Data Visualization. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 69, 7 pages. doi:10.1145/3613905.3650972

[82] William R. Watson, Christopher J. Mong, and Constance A. Harris. 2011. A case study of the in-class use of a video game for teaching high school history. *Computers & Education* 56, 2 (Feb. 2011), 466–474. doi:10.1016/j.compedu.2010.09.007

[83] Randi Williams, Safinah Ali, Raúl Alcantara, Tasneem Burghleh, Sharifa Alghowinem, and Cynthia Breazeal. 2024. Doodlebot: An educational robot for creativity and AI literacy. In *Proceedings of the 2024 ACM/IEEE international conference on human-robot interaction (Hri '24)*. Association for Computing Machinery, Boulder, CO, USA, 772–780. doi:10.1145/3610977.3634950

[84] Clemens Witt, Thiemo Leonhardt, Erik Marx, and Nadine Bergner. 2024. From Guesswork to Game Plan: Exploring Problem-Solving-Strategies in a Machine Learning Game. In *Informatics in Schools. Innovative Approaches to Computer Science Teaching and Learning*, Zsuzsa Pluhár and Bence Gaál (Eds.). Vol. 15228. Springer Nature Switzerland, Cham, 73–84. doi:10.1007/978-3-031-73474-8_6

[85] Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. 2024. Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology* 55, 1 (Jan. 2024), 90–112. doi:10.1111/bjet.13370

[86] Ellia Yang, Amy Ogan, Jessica Hammer, and Jaemarie Solyst. 2024. Designing an AI literacy transformational game for families. In *Proceedings of the 2024 ACM conference on international computing education research - volume 2 (Icer '24, Vol. 2)*. Association for Computing Machinery, Melbourne, VIC, Australia, 548. doi:10.1145/3632621.3671433

[87] Yuzhe You, Helen Weixu Chen, and Jian Zhao. 2025. Enhancing AI Explainability for Non-technical Users with LLM-Driven Narrative Gamification. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 221, 7 pages. doi:10.1145/3706599.3719795

[88] Marvin Zammit, Iro Voulgari, Antonios Liapis, and Georgios N. Yannakakis. 2022. Learn to Machine Learn via Games in the Classroom. *Frontiers in Education* 7 (June 2022), 913530. doi:10.3389/feduc.2022.913530

[89] Nilüfer Zeybek and Elif Saygı. 2024. Gamification in education: Why, where, when, and how?—A systematic review. *Games and Culture* 19, 2 (2024), 237–264. doi:10.1177/15554120231158625

[90] Xuesong Zhai, Xiaoyan Chu, Ching Sing Chai, Morris Siu Yung Jong, Andreja Istenic, Michael Spector, Jia-Bao Liu, Jing Yuan, and Yan Li. 2021. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity* 2021, 1 (Jan. 2021), 8812542. doi:10.1155/2021/8812542

[91] Zehui Zhan, Yao Tong, Xixin Lan, and Baichang Zhong. 2024. A systematic literature review of game-based learning in Artificial Intelligence education. *Interactive Learning Environments* 32, 3 (March 2024), 1137–1158. doi:10.1080/10494820.2022.2115077

[92] Xiaofei Zhou, Yushan Zhou, Yunfan Gong, Zhenyao Cai, Annie Qiu, Qinqin Xiao, Alissa N. Antle, and Zhen Bai. 2024. "Bee and I need diversity!" break filter bubbles in recommendation systems through embodied AI learning. In *Proceedings of the 23rd annual ACM interaction design and children conference (Idc*

*'24)*. Association for Computing Machinery, Delft, Netherlands, 44–61. doi:10.1145/3628516.3655802

[93] Daniele Zolezzi, Giovanni Adorni, Ilaria Torre, and Gianni Viardo Vercelli. 2025. The Web Dev Challenge: Designing GenAI-Supported Gamified Learning in a University Classroom. In *Methodologies and Intelligent Systems for Technology Enhanced Learning, 15th International Conference*. Vol. 1619. Springer Nature Switzerland, Cham, 184–195. doi:10.1007/978-3-032-05070-0_17

**Table 3: We acknowledge that our analysis focused on the core game mechanics of each game — specifically, whether the authors explicitly designed a given mechanic to teach AI literacy. Note that these data reflect the first author's initial coding, prior to reconciliation. Discrepancies between this initial coding and the final matrix were resolved through discussion involving at least two authors. $N/R$ = Not Reported. Artifacts with insufficient detail on their AI literacy focus or game design elements were excluded from the matrix. "Target Audience" reflects the age group intended by the authors and may differ from the actual participant pool evaluated. * denotes uncertainty with mixed numbers reported within the paper.**

| Name | Gam*, Space, Modality | AI Literacy Focus | Target Audience | Primary Objective | Evaluation Gam* | Results | Interactivity or Play | Suggestions |
|---|---|---|---|---|---|---|---|---|
| *AI Puzzlers* Dangol et al. [14] (2025) | Game<br>Puzzles<br>Computer | Evaluate & Create AI | Children (Ages 6 - 11) | Aims to help children understand and critically evaluate generative AI reasoning by providing a visual and interactive comparison between AI-generated solutions and their own, leveraging multimedia learning principles to reduce cognitive overload and support deeper engagement with AI outputs. | $N$ = 21 Children (Ages 6 - 11) 1.5 hours, Two Co-Design Sessions, Observational Video Data (927 mins) | Findings show that children viewed solving AI Puzzlers as fun problem-solving challenges, maintained interest even as puzzles became more difficult, quickly spotted AI errors visually, and learned to refine their instructions to address superficial AI reasoning. | Multiplayer (4 - 5 Players) Collaboration | Suggests that AI systems should allow children to critically engage with AI reasoning rather than just receiving answers, providing opportunities to experiment, test assumptions, and use scaffolds like validity markers to compare solutions, identify contradictions, test claims, and justify answers with evidence. |
| *Apricot Stone City* Moore et al. [45] (2024) | Game<br>Narrative<br>Computer | Know & Understand AI<br>Use & Apply AI<br>Evaluate & Create AI<br>Ethics & AI | Learners K-16 (Ages 12 - 17) | Aims to develop and evaluate a narrative-based learnersourcing platform that increases AI knowledge, promotes engagement, and supports peer-driven learning through content creation and feedback. | $N$ = 18 Learners (Ages 12 - 17) Online Workshop<br>$N$ = 31 Learners (Ages 15 - 19) In-Person Workshop<br>$N$ = 16 Learners (Ages 14 - 17) In-Person Workshop<br>$N$ = 27, 14 high schoolers (Ages 14 - 18), 11 college students (Ages 19 - 22), and 2 participants (Ages over 23).<br>1 Week, Pre- and Post-Exams, Interaction Data, Pre- and Post-Questionnaires, User Engagement Scale (UES-SF) and Unified Theory of Acceptance and Use of Technology (UTAUT2), Focus Group Interviews | Findings show that learners across experience levels benefited from the learnersourcing platform, but self-assessments were unreliable due to a Dunning–Kruger effect, highlighting the need for objective, expert-made assessments. Learners were more likely to re-engage with and expand narratives they found personally engaging, although some, particularly male learners, viewed narrative as non-essential or distracting. | Single Player | Suggests the need to reduce learner drop-off, expand testing across diverse educational and community contexts, better support low-attitude learners, and strengthen girls' interest and retention in computing given women's underrepresentation in the field. |
| *ArtBot* Zammit et al. [88] (2022) | Game<br>Puzzle Game<br>Computer and Mobile | Know & Understand AI | Students and Teachers<br>(Ages $N/R$) | Aims to introduce young learners to fundamental AI and ML concepts, specifically supervised learning and reinforcement learning, through a game-based approach. The game seeks to raise awareness of AI processes, highlight algorithmic capabilities and limitations, and foster critical thinking while engaging students in interactive, narrative-driven mini-games. | $N$ = 2222 Unique Users, Survey data for $N$ = 130 students (Mean Age = 14.3, SD=5), $n$ = 35 Teachers<br>Longitudinal Study (About 1 year), Surveys, Game Analytics | Findings indicate that players understood the principles of reinforcement learning and how human-defined settings can influence AI outcomes. While some found manual labelling less enjoyable or not impactful enough, students appreciated learning new concepts through problem-solving and critical thinking in a game environment, which teachers noted could engage students, particularly those already interested in games. | Single Player | Suggests that students found the game slow, monotonous, and time-consuming, indicating a need for improved engagement with learning content. Additional features, such as a gradual difficulty increase and a better reward system—like points, leaderboards, or achievement lists—could enhance motivation and participation. |
| *Beetrap* Zhou et al. [92] (2024) | Game<br>Augmented Reality<br>Mobile (Samsung Tablet S8) | Evaluate & Create AI | Students (6th - 10th grade) | Aims to help students understand and engage with AI recommendation systems by teaching three key concepts: (1) the filter bubble effect and its ethical implications, (2) how content-based recommendation systems create filter bubbles, and (3) how to apply a diversification algorithm to increase content diversity and mitigate filter bubbles. | $N$ = 9 students<br>Pre- and Post-Questionnaire, Interview data, Interaction Data | Findings show that using embodied analogies helped students understand filter bubbles, connect abstract AI concepts to familiar contexts, and create their own analogies, supporting recall of algorithmic steps in recommendation systems. However, students could describe certain AI phenomena in their own words but struggled to use the correct AI terms, and the introduction of too many analogies at once sometimes caused confusion, reducing the effectiveness of learning. | Single or Multiplayer (2-player paired preferred) | Suggests that tangible learning interfaces with spatial, physical, and temporal properties enhance student reflection and understanding, though some students did not fully transfer their in-app learning to post-tests. |
| *BotVoyAIge* Shrivastava et al. [67] (2024) | Game<br>Board Game<br>Tangible Elements | Know & Understand AI<br>Ethics & AI | Young Indian Children (Ages 10 - 12) | Aims to explore how young Indian children engage critically with ChatGPT-3 through the game, examining the biases and sensitivities they uncover, how they respond to prompts, and how they perceive the gameplay experience, with the goal of fostering ethical and responsible interaction with generative AI. | $N$ = 20 Children (Ages 10 - 11) across Three Workshops Gameplay about 60 - 90 minutes Pre- and Post-Questionnaire, Observational Data | Findings show that participants developed a sense of competition by identifying ChatGPT responses that were culturally or socially inappropriate. Their own responses reflected local cultural and geographical nuances, highlighting mismatches with ChatGPT's Western-biased outputs, such as the interpretation of the term "president" in the Indian political context. | Multiplayer (4 - 5 Players, Requires one game moderator and ChatGPT-3 as an assistant) | Suggests that Child-Computer Interaction (CCI) researchers should empower children to critically engage with AI technologies and understand their limitations. Suggests that ChatGPT can be introduced as a tool still in development, using a game-based approach to promote exploration while preventing incorrect assumptions. |

| Name | Gam*, Space, Modality | AI Literacy Focus | Target Audience | Primary Objective | Evaluation Gam* | Results | Interactivity or Play | Suggestions |
|---|---|---|---|---|---|---|---|---|
| *Co-ML* Tseng et al. [72] (2024) | Classification Game<br><br>Mobile Application<br><br>Tablet (iPad) | Know & Understand AI<br>Use & Apply AI | Students (Ages $N/R$) | Aims to introduce a collaborative ML model-building tool designed for tablets that enables multi-user data collection and model testing. The system centres collaboration to help learners surface diverse perspectives during dataset creation, analysis, and model evaluation. Co-ML further emphasizes iterative model testing, allowing students to monitor how model performance shifts with each dataset revision. | $N = 18$ (Ages 15 - 18)<br>2-week Summer Camp<br>Observational Data, Model Training Data, Focus Group Interviews | Findings show how a collaborative experience, specifically utilizing a shared dataset and collective discussion, can help surface differences in data collection approaches that ultimately inform and enhance ML datasets. This requires an iterative cycle where groups inspect misclassified data, diagnose causes, and coordinate dataset modification and model retraining. | Multiplayer (Cooperative 3 - 4 players) | Suggests that even learners who participate less in verbal discussion can still benefit by listening to others' concerns, enabling them to act on newly introduced ideas. |
| *Equality Engine* Showkat et al. [66] (2025) | Game<br><br>Online Boardgame<br><br>Computer | Know & Understand AI<br>Use & Apply AI | Post-secondary Students (Ages $N/R$) | Aims to explore how a transformational multiplayer game can teach post-secondary students about ML biases and debiasing concepts, providing a playful, immersive learning environment to help students critically understand algorithmic bias, its sources, and strategies for mitigation. | $N = 14$ (7 groups, two players per group, Ages 18 - 28)<br>All STEM majors<br>Gameplay Data, Pre- and Post-Questionnaire, Interview | Findings show that prior ML knowledge enhanced students' understanding of bias concepts in the game. Binary scoring encouraged attention and memorization, though the complexity of bias scenarios limited assessment via correct/incorrect outcomes. Gameplay fostered collaboration, negotiation, and collective exploration of ideas. | Multiplayer Cooperative Two-Player (or in multiple pairs) | Suggests that ML bias literacy depends on exposure to biases. Gameplay increased participants' awareness of bias types and debiasing methods, though broader recognition remains limited; however, dyadic pairing supports collaborative problem-solving and strategic decision-making, compensating for individual knowledge gaps. Future work should consider avoiding ML jargon, addressing a gap in ethical AI/ML education that often assumes prior technical knowledge. |
| *ML-Quest* Priya et al. [59] (2024) | Role-Playing Game<br><br>Computer | Know & Understand AI | Students (Ages 12 - 13) | Aims to introduce higher-secondary students to three ML concepts—Supervised Learning, Gradient Descent, and KNN Classification—through a step-by-step game. The game conveys a conceptual overview without complex math, using clues, an instruction board, and dialogue to guide intuitive task-solving, while fully explaining the ML concepts and terminology only at the end. | $N = 30$ Experimental Group, $N = 30$ Control Group (High School Students) Analyzed $N = 18$ from Experimental, and $N = 23$ from Control<br><br>Four factors out of six factors measured in Technology Acceptance Model (TAM): Perceived Ease of use (EU), Perceived Usefulness (U), Intention to Use (I) and Correctness (C) | Findings show that students who played the game outperformed those who did not, and that the gameplay, storyline, and challenges helped students build clearer mental models of the ML concepts. It also indicates that mentor-guided gameplay improved understanding for both groups, with slightly higher gains for the control group. | Single Player | Suggests that ML-Quest could be strengthened by adding new ML concepts with improved visualizations, increasing game complexity and user engagement, introducing multiplayer features, and enhancing graphics. It also indicates a need to explore mobile platforms, as many students prefer smartphone-based gameplay. |
| *Treasurels-land* Ng et al. [49] (2024) | Gamification<br><br>Classroom<br><br>Computer | Know & Understand AI<br>Use & Apply AI<br>Evaluate & Create AI<br>Ethics & AI | Students (Ages 12 - 13) | Aims to explore how 12-to-13-year-old students learn AI concepts in a formal curriculum using a gamified platform built on top of ebook elements such as videos, readings, and digital artifacts. | $N = 33$ Control Group, $N = 35$ Treatment Group, Students (Grade 7)<br><br>2 months, 10 sessions, 35 minutes<br><br>Pre- and Post-Questionnaires, Interviews | Students with lower initial achievement were motivated to improve once they saw they could move up the leaderboard. Without enough practice time, students struggled to apply AI knowledge to real-world contexts and suggested longer lessons for more hands-on learning. | Single Player | Suggests that leveraging game elements, such as peer competition, digital badges, points, and a storyline-based leaderboard, supports retention and helps students develop problem-solving skills, creativity, and collaboration. |
| *AiLingo* Pinski et al. [56] (2024) | Gamification<br><br>Application<br><br>Mobile | Know & Understand AI<br>Use & Apply AI | Non-Expert Adult | Aims to investigate whether a gamified learning experience improves AI literacy and AI metaknowledge by fostering reflection, critical self-assessment, and deeper engagement with AI concepts. | $N = 101$ 53 treatment group, 48 control group (Mean Age = 31.4, SD= $N/R$) | Findings demonstrate that a gamified learning experience is more effective than a text-based approach in enhancing objective AI literacy, subjective AI literacy, and AI metaknowledge, with the text-based group exhibiting reduced metaknowledge and lower learning gains. | Single Player | Suggests that metaknowledge is crucial for current and future human-AI collaborations, and organizations need effective methods to cultivate it. |
| *ARIN-561* Wang et al. [79] (2023) | Game<br><br>Role-Playing and Puzzles<br><br>Computer | Know & Understand AI | K-12 Students | Aims to cover classical search algorithms (breadth-first search (BFS), depth-first search (DFS), and greedy search) via tutorial and transfer modules with embedded quizzes for self-assessment by integrating in-game challenges, such as searching for spaceship parts or cracking passwords, to contextualize algorithm concepts and connect them to real-world problem solving. | $N = 764$ Students (Grade 9 - 12) for final analysis, Pre- and Post-Questionnaire, Game Log | Findings show students made more errors when learning DFS than BFS or greedy search, that performance in transfer modules positively correlates with tutorial performance, and that tutorial performance is only weakly related to success on transfer tasks. | Single Player | Suggests that integrating in-game mini lessons to reinforce math knowledge relevant to AI concepts can improve students' performance in the game. |
| *Canvas City* Ioannou et al. [31] (2025) | Gamification<br><br>Choice-Based Narrative<br><br>Computer | Know & Understand AI | Primary and Secondary Teachers (Ages $N/R$) | Aims to enhance awareness of AI decision-making processes and their societal implications by providing Canvas City, an interactive scenario-based platform for studying user perceptions and educating users about AI decision-making, explanation types, and how AI systems generate them. | $N = 90$ Teachers, Average Completion Time 19.36 mins<br>Questionnaire, Interaction Data | Findings show teachers valued the real-world examples in the scenarios. Many teachers recognized the importance of AI education, yet lacked the confidence, training, and resources to teach AI concepts or use AI tools effectively. The results also showed that attending AI training did not significantly influence teachers' views on the tool's benefits or their confidence in integrating it into their teaching. | Single Player | Suggests that additional resources are needed for those who want to learn more, while also noting that some teachers found the information overwhelming and the language too complex for younger students. |

| Name | Gam*, Space, Modality | AI Literacy Focus | Target Audience | Primary Objective | Evaluation Gam* | Results | Interactivity or Play | Suggestions |
|---|---|---|---|---|---|---|---|---|
| *ClawAI* Kolawole et al. [37] (2025) | Guessing Game<br><br>Computer | Use & Apply AI | K-12 Students | Aims to teach children how AI recognizes text through an interactive game where players describe objects for an AI-powered claw machine to identify. The tool introduces core ideas such as text classification and confidence levels across three gameplay rounds, showing students how their written descriptions influence AI recognition. | $N$ = 43 Students (Ages 7 - 14), After-school Program Exhibition, 4 hours of interaction<br>Interaction Data, Pre- and Post-Questionnaire, Screen and Audio Recordings | Findings show that students' approach to the game strongly influenced both their success and their understanding of how AI interprets data based on the datasets provided. The timer also increased competitiveness, motivating students to score points, obtain items, or outperform their peers. | Single Player | Suggests that students became competitive and applied problem-solving strategies by trying different categories and words and looking for patterns in correct and incorrect answers to refine their approach to the game. |
| *IntoTheRabbitHole* Rajarajan and Martin [62] (2025) | Puzzle Game<br><br>Computer | Use & Apply AI | K-12 Students (Ages $N/R$) | Aims to introduce middle school students to foundational AI search algorithms using a web-based learning application designed to teach Depth First Search (DFS) and Breadth First Search (BFS). Aims to evaluate whether the application effectively supports learning and helps students compare and understand the two algorithms within an accessible, age-appropriate AI literacy experience. | $N$ = 49 Students (Grade 6 - 8) interacted with the game (about 5 mins)<br>Interaction Data, Discussion and Post-Survey | Findings show that many students enjoyed the game-like nature of the timed challenge level and actively competed to achieve the highest score. In terms of learning outcomes, 23 student responses shifted from incorrect to correct. Such a shift provides strong evidence that the application supported learning of DFS and BFS concepts. | Single Player | Suggests the effectiveness of activity- and game-based methods for introducing children to search algorithms. |
| *Match the Monkeys* Witt et al. [84] (2024) | Multi-Touch Puzzle Game<br><br>Tablet | Use & Apply AI | Students (Ages 12 - 13) | Aims to investigate how students approach problem-solving when creating decision-tree classification models, examining the strategies they use, how these strategies relate to learning success, and what behavioural patterns such as persistence or strategy switching emerge during the task. | $N$ = 49 Students (Grade 8) and $N$ = 44 Students (Grade 10) in workshops.<br>$N$ = 93 but one is excluded due to loss of connection, so $N$ = 92<br>14 min and 54 s average gameplay<br><br>Recorded Gameplay Footage, Content Analysis | Findings show that learners often begin with exploratory, trial-and-error approaches to understand the task and the game environment. These findings also suggest that structured problem-solving strategies are essential for successfully completing the game, highlighting the importance of supporting learners in developing more systematic thinking. | Single Player | Suggests that to help students develop consistent and systematic solution strategies independently, ongoing support for their problem-solving processes is important. |
| *AI Audit* Ali et al. [3] (2023) | Card Game (Physical Environment)<br><br>Tangible | Ethics & AI | Middle and High Schoolers (Ages $N/R$) | Aims to encourage children to critique everyday AI systems for potential harms while simultaneously thinking of ways to mitigate them. | $N$ = 5 Adults<br><br>Playtesting of the game | Players struggled with some AI-related terminology and concepts. | Multiplayer (Up to 7 players) | Plans to test AI Audit with middle and high school students and teachers in classroom settings to evaluate accessibility, engagement, and learning outcomes. |
| *BecomingFei* Wang et al. [78] (2025) | Game-Based Learning (Digital Environment)<br><br>Computer | Know & Understand AI<br>Ethics & AI | General Public and Pre-College Students (Ages $N/R$) | Aims to teach three concepts: (1) data: how type and quality affect training and the resulting model; (2) algorithm: the advantages, limitations, and suitability for different data and problems; (3) model: evaluation metrics such as accuracy, error rate, precision, recall, and false positives/negatives. | $N/R$ | $N/R$ | Single Player | Plans to evaluate the game with participants to assess its effectiveness in promoting AI learning and developing data literacy. |
| *Contour to Classification* Lee and Ali [39] (2021) | Classification Game (Digital Environment)<br><br>Computer | Know & Understand AI | Middle School Students (Grade 6 - 8) | Aims to engage learners in activities that build mental models of supervised learning in neural networks, including the structure and function of input, hidden, and output nodes, as well as the processes of feedforward, evaluation, backpropagation, and gradient descent. | $N$ = 65 Students (Ages 10 - 16)<br><br>Three Workshops<br>Playtesting of the game | Students valued participating in the live simulation, observing network changes, and connecting them to a previous decision-tree game. Feedback on correct contours and opportunities to revise selections may enhance understanding of how channels strengthen or weaken and how nodes adjust their algorithms. | Multiplayer Requires 10 players | Plans to play-test the game with more students, allowing them to select images or themes that are personally relevant and engaging. |
| *Degrees of Freedom* Mott et al. [47] (2024) | Narrative Role-Playing Game (Physical Environment)<br><br>Tangible Elements | Ethics & AI | $N/R$ | Aims to introduce a multiplayer interactive storytelling game that fosters technology literacy about social robots. The game promotes technology literacy competencies by encouraging players to explore how values are embedded in robot designs, reflect on the risks and limitations of robots, and relate these considerations to their own lives and personal values. | $N$ = 3 Adults<br><br>Playtesting of the game | Participants' in-game decisions reinforced AI literacy by reflecting challenges faced by roboticists, such as the consequences of robots performing human-like roles and retaining user data. The dilemmas also prompted discussion on robots' potential role in law enforcement. | Multiplayer Requires a Game Master (GM) that is a researcher | Currently, the game requires a GM with HRI and roboethics expertise, making it best for guided settings. Plans to iterate include adding materials to make the game self-contained for broader use. |
| *DRONEscape* DiPaola et al. [19] (2023) | Puzzle Game<br><br>Escaperoom<br><br>Inside A Tent & Computer | Know & Understand AI<br>Use & Apply AI | Air Force Participants (Ages $N/R$) | Aims to design an educational escape room for Air Force participants as part of a broader AI literacy training program. The objective was to reinforce learning objectives by providing an embodied experience with real-world AI systems. The missions and puzzles were specifically designed to incorporate six key concepts: (1) word embeddings, (2) feature engineering (FE), (3) reinforcement learning (RL), (4) human-in-the-loop processes, (5) supervised learning (SL), and (6) optimum policy switching. | $N$ = 22, 5 teams in groups of 4-5 members, workshop 2 hours to escape<br>Questionnaire, Interviews | Findings show that while the escape room was not a primary source of new information, it excelled at aiding the retention and real-world application of existing AI concepts. Observations also highlighted positive collaborative behaviours, as teammates demonstrated mutual respect and active listening. | Multiplayer Cooperative (Groups of 4 - 5 players) | Suggests integrating embodied agents for scaffolding and ensuring high team participation, thereby maintaining strong engagement and teamwork. Additionally, designs must be technically robust, with backup plans and multiple learning pathways, and the focus should shift to creating clues that lead to new skill acquisition, incorporating a beneficial mix of evaluation and new learning. |

| Name | Gam*, Space, Modality | AI Literacy Focus | Target Audience | Primary Objective | Evaluation Gam* | Results | Interactivity or Play | Suggestions |
|---|---|---|---|---|---|---|---|---|
| *Frientelligent (interface of Doodlebot)* Williams et al. [83] (2024) | Game-based Interface<br><br>Embedded within Robot<br><br>Computer & Robot | Know & Understand AI<br>Use & Apply AI | Primary and Middle School Youth (Ages 6 - 14) | Aims to introduce students to multiagent AI concepts by enabling them to explore path planning, distance metrics, and policy making through a synchronized virtual and physical game environment with autonomous Doodlebots. | $N = 8$ children (Ages 9-11), $N = 2$ conditions, virtual and physical<br>Pre- and post-Questionnaires, Interviews, Observations | Findings show that while students were generally able to link the physical Doodlebot to AI examples from their everyday lives, two students reported that the wide range of policy and distance metric options in the interface felt confusing. Additionally, the physical setup required more effort and was constrained by the limited number of objects available for constructing the map. | Multiplayer (Team Cooperation) | Suggests collaborating with educators to design low-cost, classroom-ready educational robots, enhancing cost-effectiveness through multiple use cases across learning topics, while ensuring the robots are engaging and effective, and that they link technical concepts to real-world experiences. |
| *PlayGrid Mat, PlayGrid Sense, PlayGrid Go* Quiloan et al. [61] (2023) | Pathfinding Game<br><br>PlayGrid Mat: Floor<br><br>PlayGrid Sense and PlayGrid Go: 8x8 Physical Button Pads | Know & Understand AI | Young Children (Ages $N/R$) | Aims to introduce AI literacy to young children through a handheld learning interface, helping them understand the capabilities and limitations of AI in a safe, educational environment without teaching coding. | $N/R$ | Proposes that PlayGrid Mat, a floor-based interactive grid where users step, jump, or dance, helps develop the sequential thinking skills assumed by many coding toys and games. Due to budget constraints, iterations evolved into physical button pads: PlayGrid Sense and PlayGrid Go. | Single Player | Suggests that ML algorithms can support early AI literacy by providing young children with hands-on, tangible experiences through playful, interactive prototypes. |
| *PrimaryAI* Gupta et al. [25] (2024) | Game-Based Learning (Digital Environment)<br><br>Computer | Know & Understand AI<br>Use & Apply AI | Students (Ages $(8-11)$) | Aims to create a story-driven, game-based learning environment that engages upper elementary students in situated, AI-infused life-science problem-solving. By integrating AI and science instruction, the experience is designed to reach a broad range of students, including those who might not otherwise choose to explore AI. The classroom-friendly game is supported by a curriculum featuring "unplugged" AI learning activities that introduce key concepts prior to gameplay. | $N = 16$ Fourth Grade Class<br>Observational Data | Findings suggest that the Use-Modify-Create framework was generally successful in providing students with the knowledge they needed to progress to the next mission in the quest. | Single Player | Plans to explore how AI-driven adaptive learning methods can best support AI education by delivering tailored feedback to students. It is also important to conduct large-scale classroom studies to better understand the impact of story-driven, game-based learning on students' understanding of AI and their interest in the field. |
| *Robocode* Hartness [27] (2004) | Game (Digital Environment)<br><br>Computer | Evaluate & Create AI | Students (Ages $N/R$) | Aims to use the game to let students develop virtual robots in Java, apply class algorithms, deepen AI understanding, and build confidence. Through the game, students create autonomous agents and observe their interactions. | $N/R^*$ | Suggests students gradually learn and experiment with strategies—such as hide, chase, chaotic, or "sit still and shoot everything"—combining and modifying them iteratively. Robots can be quickly adjusted and retested if they perform unsatisfactorily. | Single Player | Plans to apply game tree search in Robocode, modelling the environment and iteratively testing robot behaviours. |
| *Dialogue-Crafter (DC), SpongeWords (SW), SimilarityShuffle (SS)* Goslen et al. [24] (2025) | Game-Based Learning (Digital Environment)<br><br>Computer | Know & Understand AI (DC,SS), Use & Apply AI (DC) | K-12 Students | Aims to propose three game design artifacts:<br>**SpongeWords:** a playful, unplugged activity demonstrating how language models generate sentences based on text and word-sequence probabilities.<br>**DialogueCrafter:** a digital game-based activity where students explore NLP concepts by creating dialogue for NPCs, learning about AI applications in farming, a topic particularly relevant to their lives.<br>**SimilarityShuffle**: an unplugged, game-based activity introducing embeddings and similarity scores. Students practice calculating distance-based similarity within a vector space. | $N = 10$ Middle School Students (Summer Camp One)<br><br>$N = 23$ Middle School Students (Summer Camp Two) | Students showed an understanding of voice assistants' ability to process natural language, but were unable to give specific examples of challenging NLP scenarios, indicating difficulty grasping the limitations of NLP after the first summer camp.<br><br>When surveyed, most students correctly identified the definition of a similarity score and named challenging scenarios illustrating language model limitations, showing improved understanding from the previous camp. | Single Player for Dialogue-Crafter<br><br>Multiplayer (4 - 8 players) for SpongeWords and SimilarityShuffle | Plans to explore integrating large language models into student-facing tools, enabling richer discussions of AI's capabilities, limitations, and societal implications. |
| *The Desk: Dilemmas in AI Ethics* Kim et al. [36] (2025) | Game-Based Learning<br><br>Decision Making Game<br><br>Computer | Ethics & AI | Students (Ages $N/R$) | Aims to design a dialogue-based game grounded in Moral Foundations Theory to teach AI ethics, using dilemmatic scenarios such as "killer robots" to foster moral reasoning, ethical discussions, and collective decision-making among learners. | $N = 9$ Students (Graduates and Undergrads, Ages 23 - 27)<br><br>60 mins gameplay<br><br>Observational Data | Students noted that the digital game effectively engaged them in moral reasoning about AI ethics, with all six Moral Foundations discussed, especially Care and Purity, and that moral conflicts through dilemmas and collective decision-making promoted diverse arguments and deeper moral engagement. | Multiplayer (3 Players, Free For All) | Suggests expanding the game with more dilemma scenarios and diverse participants, such as small sample size, cultural context, and focus on adult learners, while highlighting its potential to foster moral judgment and reasoning in AI ethics education. |
| *TrainYourSnakeAI* Hinojosa et al. [28] (2025) | Game-Based Learning (Digital Environment)<br><br>Computer | Know & Understand AI<br>Use & Apply AI | Middle School Student | Aims to use the game to teach middle school students about reinforcement learning, allowing them to train an agent to play Snake and observe how their input shapes the agent's learning, fostering AI literacy and preparing students for future AI engagement. | $N = 71$ Middle School Students (Grade 6 - 8)<br><br>Two 3-hour sessions over two days | After using the application, 21 students accurately described reinforcement learning using keywords like rewards or punishment, and 19 compared it to natural learning in biological organisms. Among children previously unfamiliar with RL, many were able to provide an accurate definition post-application. | Single Player<br>5 - 10 minutes | Plans to address the delayed learning of the Snake RL agent, which showed visible improvement only after around 75 of 100 training iterations, by adding a live visualization of the agent's state so participants can see what the AI snake "thinks" during training. |

| Name | Gam*, Space, Modality | AI Literacy Focus | Target Audience | Primary Objective | Evaluation Gam* | Results | Interactivity or Play | Suggestions |
|---|---|---|---|---|---|---|---|---|
| *Web Dev Challenge* Zolezzi et al. [93] (2025) | Gamification (Physical Environment) Course | Use & Apply AI Evaluate & Create AI | University Students (Ages $N/R$) | Aims to create a GenAI-supported gamified learning experience by pairing weekly coding tasks with an AI-driven virtual tutor and progression-based game mechanics to boost technical skills, prompt literacy, and motivation. | $N/R$ | Suggests that students frequently used the virtual tutor for syntax help, debugging, and CSS/JavaScript guidance. Context-rich prompts led to more effective responses, while vague prompts caused misinterpretations. Students valued the tutor's availability but sometimes questioned the accuracy of its suggestions. | Single Player Leaderboard System | Plans to systematically analyze student-tutor interactions for accuracy, relevance, and educational value, while tracking prompt evolution over time to identify learning patterns and optimize the tutor's instructional effectiveness. |
| *2AI Adventure* Petridis et al. [55] (2024) | Gamification* (A Game* Digital System) Computer | Know & Understand AI Use & Apply AI Ethics & AI | Students (Ages 6 - 12) | Aims to develop a gamified approach to enhance children's AI literacy and equip educators to address AI topics across formal and informal learning contexts across three levels. | $N = 68$* Students (Inconsistency reported) Usability Questionnaire on evaluating the Game Concept Idea | Findings show game difficulty decreases as participant age increases, with older players relying more on skill than on luck. | Single Player | Plans to develop an EU/UK platform for teacher and student training in AI and its societal implications, supporting the Road to AI and EU Digital Market initiatives through interactive learning tools. |
| *AI4Health* Mott et al. [46] (2025) | Narrative Game (GBL) (Digital Environment) Computer, Tablets | Know & Understand AI | Middle School Students (Ages $N/R$) | Aims to promote AI literacy and spark early interest in health-related careers by situating learning within an engaging narrative framework. | $N = 25$ Middle School Students Usability of the game | Suggests that the game enhances engagement, deepens understanding of AI models, and fosters interest in related careers in health. | Single Player | Plans to expand the system with more episodes and conduct classroom studies to assess long-term learning and engagement. |
| *AI-VIS EthiCards* Wang et al. [81] (2024) | Card Game (Physical Environment) Tangible Elements | Ethics & AI | Students (Ages $N/R$) | Aims to offer users an exploratory and open approach to discussions, providing multiple perspectives to guide ethical considerations when applying AI for visualization. | $N/R$ | $N/R$ | More than 2 players | Plans to organize workshops with participants from diverse disciplines such as ethics, data science, visualization design, and other professional domains to refine the cards and activities. |
| *Ask Sir Oliver Ingham* Park et al. [54] (2025) | Simulation Game (Digital Environment) Computer GPT-4o | Use & Apply AI | Students & Teachers (Ages $N/R$) | Aims to investigate the feasibility and potential of an LLM-based historical simulation platform designed to balance immersive student engagement with teacher-directed customization, such as AI-generated content. | $N = 15$ Elementary and middle school teachers System Usability Scale (SUS) | Teachers noted the platform's immersive potential, support for multiple perspectives, and reflective exploration of history. Concerns about AI-generated content were addressed through guided discussions. | Single Player | Plans to extend the platform to educational settings across age groups, conducting studies to evaluate learning, engagement, and historical thinking through longitudinal and comparative analyses with students. |
| *Case 429* Maheshwary et al. [43] (2025) | Serious Game (Digital Environment) Computer | Ethics & AI | Late Secondary and University Students (Ages 16 - 22) | Aims to enable players to detect and critique algorithmic bias, misrepresentation, and hallucinations in high-stakes AI outputs, such as those affecting criminal justice, fostering ethically aware and critically minded digital citizens. | $N = 12$ Students for Concept Design $N = 6$ Students for Playtesting | In the concept design phase, findings indicate that (1) race and gender stereotypes are easiest to detect, while cultural or domain-specific biases are often overlooked, and (2) AI is most frequently used for summarizing large documents or identifying trends across multiple files. | Single player | Plans to expand Case 429 as an open, modular platform by diversifying domains, broadening audiences and cultures, and enhancing assessment and social interaction, transforming it into a flexible research testbed for critical AI-literacy interventions across disciplines and educational settings. |
| *Competitive Two-Player Game* Aguiar et al. [2] (2025) | Serious Game (Digital Environment) Computer | Know & Understand AI Use & Apply AI | Students (Ages $N/R$) | Aims to provide students with a hands-on learning experience that allows them to observe in real time how search algorithms operate and efficiently find a path between two points. | $N = 40$ Middle School Students (Grade 7 - 8) Playtesting of the game | Observations indicate that some students initially struggled with level editing and boost tiles. Instructor walkthroughs clarified the mechanics, highlighting the need for clearer onboarding, a more intuitive interface, and in-game tutorials to better support learning. | 2-Player Versus | Plans to iterate on the game by integrating additional topics, such as reinforcement learning or heuristic search, while ensuring novice learners are not overwhelmed. |
| *ImaginAItion* Ma et al. [42] (2025) | Game-Based Learning (Digital Environment) Computer | Know & Understand AI Use & Apply AI Ethics & AI | Adults | Aims to understand how adult novices perceive bias in GenAI, interpret its capacities, and engage in prompting. It also seeks to examine whether ImaginAItion can support the development of GenAI literacy by improving bias awareness, strengthening mental models, and enabling more successful and efficient prompt creation and refinement. | $N/R$ | $N/R$ | Multiplayer 3 players, 3 teams | Plans to add more complex mechanics, letting the prompter customize prompts or add distractors, and supporting team or individual gameplay to reinforce turns. Future iterations may also include explicit guidance on prompt engineering based on existing training literature. |
| *KIWI (Knowledge Inferencing Web-based Interface)* Anthis and Mavrikis [4] (2025) | Gamification (Digital Environment) Computer | Know & Understand AI Use & Apply AI | In-Service Educators (Ages $N/R$) | Aims to investigate how in-service educators experience the usability and cognitive demands of KIWI, a gamified learning environment for AI literacy, while also examining the impact of non-analytical search on their problem-solving performance over time and how interactive, learning-based experiences shape their engagement and approaches to solving AI-related problems. | $N = 90$ In-Service Educators (actively teaching at primary or secondary education level) System Usability (SUS), Cognitive Load Questionnaire (CLQ), search efficiency | Novice users initially struggled with stochastic variability but gradually developed a clearer understanding of when to explore versus exploit. Over time, they calibrated their trust in AI-generated solutions more effectively, demonstrating a deeper understanding of explainability, justification, and transparency. | Single Player | Plans to study how usage and performance evolve with reflection and self-regulation through longitudinal studies to assess sustained AI literacy. Expanding to diverse educator groups could reveal cross-cultural engagement patterns. Integrating AI explainability and adaptive gamification could enhance motivation and support learning from conceptual understanding to procedural retention and trust calibration. |
| *MediaWatch* Koskinen et al. [38] (2025) | Game-Based Learning (Digital Environment) Computer | Ethics & AI | Students (Ages $N/R$) | Aims to provide empirical evidence on the influence of motivational, cognitive, and emotional factors in game-based AI ethics instruction, while informing the development of future curricula in this emerging educational domain. | $N = 40$ 9th Grade Students (Age: M=15.53, SD=.51, Male = 17, Female = 23) 7.1 minute gameplay (SD=1.53) | Suggests that even short gameplay sessions can enhance students' understanding of AI ethics, despite their prior familiarity with the topic. | Single Player | Plans to examine the long-term effects of game-based learning on students' understanding of AI ethics and their ability to apply this knowledge in real-world contexts. |

| Name | Gam*, Space, Modality | AI Literacy Focus | Target Audience | Primary Objective | Evaluation Gam* | Results | Interactivity or Play | Suggestions |
|---|---|---|---|---|---|---|---|---|
| *Nemobot* Wang et al. [80] (2024) | Gamification and Game-Interface (Digital Environment)<br><br>Computer | Evaluate & Create AI | K-12 Students | Aims to focus on the inner workings of AI, support collaborative and scalable learning, and remain accessible to K–12 students with diverse proficiency levels. The learning experience begins with students creating a two-player strategic game where players take turns acting and earn corresponding rewards. | $N = 30^*$<br><br>Proof-of-concept of Gamified Process | Findings show the efficacy of Nemobot in teaching K–12 students to use the platform effectively for AI creation in strategic games. | 2-Player | Plans to expand it beyond strategic gaming to include diverse AI tasks, such as multimedia and programming, thereby enhancing AI education efforts. |
| *SMaILE-App* Franco et al. [21] (2024) | Gamification and Mini-games<br><br>Building-Simulation and Puzzles<br><br>Mobile | Use & Apply AI | Students (Ages 10 - 18) | Aims to provide a simulation experience in the realm of city-building games, where the primary objective is to construct a thriving metropolis from the ground up, starting with a sustainable master plan to enhance livability. | $N = 80$ Children (Ages 10-13) in 5 Groups for difficulty and playability<br><br>$N = 20$ Students (Ages 11 - 12) in 2 Groups for interface and usability<br><br>$N = 40$ Students (Ages 10-18) in 2 workshops, $N = 900$ middle school students but only analyzed $N = 450$ in the field study<br><br>GUESS-13 Questionaire | Findings revealed that committed players spent more than an hour on the app on average, showing higher user satisfaction. Female players showed a stronger preference for the app's visuals, engagement, design, and balance. | Single Player | Suggests incorporating more real-world applications of engineering concepts, which students perceive as both educational and enjoyable. |
| *XAI Visualization** You et al. [87] (2025) | Gamification (Digital Environment)<br><br>Computer | Know & Understand AI | Non-Technical Users | Aims to explore whether LLM-driven narrative gamification can promote intuitive understanding of complex XAI visualizations and actively engage non-technical users in their exploration. | $N = 10$ Limited Technical Background in AI and ML (6 men, 4 women, Ages 22 - 23)<br><br>Comparison Analysis Between Gamified Prototype and baseline without gamification | Findings show potential in motivating users to engage more deeply with visualizations, fostering curiosity and proactive learning. Participants reported feeling comfortable asking questions without fear of judgment. | Single Player | Plans to explore the pros and cons of combining gamification with XAI visualizations, examine user workflows and use of gamification features to extract insights, and investigate the impact of LLM hallucinations on understanding and engagement, including mitigation strategies such as improved prompting or retrieval-augmented generation. |
| *AI Exploratorium* Maljur et al. [44] (2025) | Gamification (Physical Environment)<br><br>Computer and Tangible Elements (Interactive Exhibition) | Evaluate & Create AI | Students (Ages $N/R$) | Aims to use hands-on challenges that have participants train AI models for real-world applications like autonomous cars and deepfake detection, and encourage them to reflect on the ethical implications of AI. | $N = 7$ Male Students (Ages 12 - 14)<br><br>Pilot Data | Participants replayed the card game multiple times, improving their ability to spot deepfakes and understand AI misuse potential. Quizzes effectively reinforced learning. Participants correctly identified causes of AI misclassifications, such as visual similarities, showing understanding of pattern recognition. | Multiplayer (Pairs or small teams) | Plans to explore the challenges and opportunities of integrating interactive, puzzle-based AI learning approaches—where parts of the "black box" are revealed through acquired knowledge—along with assessment methods to evaluate its long-term impact on AI literacy. |
| *AI Literacy Transformational Game* Yang et al. [86] (2024) | Card Game (Physical Environment)<br><br>Tangible Elements | Know & Understand AI | Youth and Families (Ages $N/R$) | Aims to design a game to support youth and families in learning about ML and algorithmic bias. | $N = 7$ Children (Ages 9 - 12)<br><br>$N = 3$ Families (3 parents, 6 children (Ages 9 - 13)) Playtesting at workshops in an out-of-school context | Findings indicate that transformational games incorporating persuasive design techniques can effectively support learning about machine learning and algorithmic bias, especially in family contexts. | 2-Player Game | Future work could further investigate playful approaches that help families learn about and make sense of AI, including emerging areas such as generative AI. |
| *AI Play* Vandenberg et al. [76] (2023) | Gamified (Digital Environment)<br><br>Computer | Know & Understand AI<br>Use & Apply AI | Rural, middle grades students (Ages 11 - 14) | Aims to enable students to learn AI concepts and practices while creating gameplay experiences using AI-driven techniques such as machine-generated content, search-based pathfinding, non-player character reasoning, and natural language processing-based interactions. | $N/R$<br><br>Co-design Sessions | Findings show teachers could not define AI and noted they did not teach coding. Students' answers to "What is AI?" ranged from general computer use to examples like doorbell cameras and self-driving cars. | $N/R$ | $N/R$ |
| *Data detectives* Solyst et al. [68] (2022) | Card Game (Physical Environment)<br><br>Tangible Elements | Know & Understand AI<br>Evaluate & Create AI<br>Ethics & AI | Child-Parent (Ages $N/R$) | Aims to facilitate three main learning outcomes: players are able to (1) understand how training data influences algorithm outputs, (2) recognize that algorithmic bias can arise from training data—including its selection, labelling, and quantity, and (3) critically evaluate training data decisions in the context of supervised machine learning. | $N = 6$ Adults Playtesting of the game | $N/R$ | Multiplayer (2 - 6 players, 2 teams) | Plans to play test with youth and families. |
| *EthiQuest* Bhat [8] (2025) | Card Game (Physical Environment)<br><br>Tangible Elements | Ethics & AI | Middle School Students (Ages $N/R$) | Aims to engage middle school students in exploring nuanced ethical dilemmas surrounding AI through a generative AI– and ethics-based card game, prompting reflection on authenticity, creativity, fairness, accountability, and responsibility, and fostering foundational AI ethics understanding through collaborative, speculative play. | $N = N/R$ Students and Teachers<br><br>5 day workshop | Surveys showed clear gains in students' understanding of fairness, bias, and accountability in AI. Reflections and discussions articulated nuanced perspectives, evaluated trade-offs, and recognized ethical complexities. Teachers reported that the scenario-driven format kept students engaged and helped them internalize abstract ethical concepts. | Requires moderator, teams | Plans to include exploring how interactive tools can more explicitly communicate the ethical trade-offs in high-stakes AI applications where errors have significant societal impact. |