International Journal of Artificial Intelligence in Education
https://doi.org/10.1007/s40593-024-00449-3

ARTICLE

From Lecture Hall to Homeroom: Co-Designing an AI Elective with Middle School CS Teachers

William Gelder[1] · **Xueru Yu**[1] · **David Touretzky**[2] · **Christina Gardner-McCune**[3] · **Judith Uchidiuno**[1]

Accepted: 3 December 2024

Abstract
Middle School students in the United States are exposed to an unprecedented number of AI-driven consumer products. This exposure demands that educators help students develop their personal understandings of these technologies to engage with them responsibly. Designing age-appropriate AI curricula for middle school students calls for collaboration and partnership between computer and learning scientists, as well as middle school teachers. Over a 3-year period, we co-designed and successfully implemented an AI education curriculum across 9 geographically and economically diverse schools, offering it to a total of 1551 students. Drawing from our analyses of the curriculum and teacher and student experiences, we propose an effective format for teaching, assessing, and implementing fundamental AI education for middle school settings in the United States. Our research also highlights the value of empowering teachers through co-design; enriching their professional development and improving students' AI literacy.

Keywords Artificial Intelligence · AI education · K-12 · Middle school · Curriculum Format · Assessments · Co-design

✉ William Gelder
wgelder3@gatech.edu

Xueru Yu
xyu425@gatech.edu

David Touretzky
dst@cs.cmu.edu

Christina Gardner-McCune
gmccune@ufl.edu

Judith Uchidiuno
jiou3@gatech.edu

1 Georgia Institute of Technology, Atlanta, GA, USA

2 Carnegie Mellon University, Pittsburgh, PA, USA

3 University of Florida, Gainesville, FL, USA

Published online: 22 January 2025

Introduction

Middle school students in the United States engage with AI-driven products daily. Popular products such as TikTok, Instagram, Google Search, and voice-based interfaces such as Siri or Alexa all leverage AI-driven technology. Educators have an exciting opportunity to make AI instruction more accessible by helping students form connections between the products they use daily and fundamental AI concepts. This foundational understanding will better prepare students for potential careers in AI, while enriching CS education more broadly. It is crucial then for middle school students to cultivate a baseline level of AI literacy as early as possible to responsibly navigate the current and future AI-powered world.

Though national and academic initiatives have spurred the development of K-12 AI curriculum guidelines, a pedagogical support gap remains for middle school educators. Prior work has sought to create AI-oriented professional development (PD) pathways for teachers intending to transfer knowledge from AI researchers to school educators. Such programs involved teacher-researcher workshops that sought to build epistemological understandings of AI and construct lesson plans for localized use. Other studies have focused on deploying and evaluating hands-on, collaborative, and game-based teaching approaches, but few have attempted to propose a common standard for the reproducible offering of age-appropriate AI curricula for K-12 students; and fewer still for middle school students. This presents a research gap in how best to increase the transferability of such curricula, which we consider to be a sign of a larger issue in the lack of overarching PD opportunities for middle school CS educators. As extensive research on co-design has proven, the inclusion of diverse perspectives facilitates the creation of more useful and equitable tools, programs, and services. In the context of K-12 education, collaboration between Computer and Learning Scientists and middle school CS educators facilitates dialogues about which content and assessment strategies merit inclusion in an effective middle school AI curriculum.

This work took place within the context of the *Artificial Intelligence for Georgia project* (AI4GA, 2021) and was funded by the National Science Foundation's ITEST program. AI4GA is a 3-year collaboration between researchers from Carnegie Mellon University, the University of Florida, Boston College, and the Georgia Institute of Technology, plus CS specialists from the Georgia Department of Education, and middle school educators from several Georgia public school districts. An AI curriculum meant for middle school learners was co-designed by the research and educator team, evaluated by members of the project's *Education & Efficacy Evaluation* team, and implemented within 9 diverse middle schools across Georgia; collectively offered to a population of 1551 middle school students.

We seek to contribute to the growing body of research related to AI Teaching and Learning (AITL) by focusing in on AI education at the middle school elective level and by contributing our reflections on the structure of an effective co-designed AI curriculum, its implementation, and assessment strategies organized by the revised Bloom's Taxonomy (Ng et al., 2021). The revised Bloom's Taxonomy is a hierarchical model for cognitive domains and has been widely applied in CS education for designing courses and structuring assessments (Thompson et al., 2008).

To contextualize our insights, we orient our data collection and analysis to address the following research questions:

RQ1: What is an effective curriculum format for teaching AI in a middle school elective as identified by teachers and students?
RQ2: What assessment strategies do teachers employ to understand student learning outcomes?
RQ3: What aspects of the AI curriculum co-design process resonated with teachers and what do they imply for effective professional development?

Literature Review

Middle School Students & AI-Driven Products

Today, AI-driven products are well-integrated into the everyday lives of middle school students in the U.S. A 2021 survey of 1,306 U.S. children aged 8- to 18-years old was conducted by the nonprofit organization Common Sense regarding their use of screen media (Rideout et al., 2022). On average, daily use for tweens (8- to 12-year-olds) from Black and Hispanic/Latino communities were reported to be higher than for those from White communities. YouTube videos were reported to be the screen media type that tweens said they "could not live without". YouTube makes use of AI-driven algorithms to drive search results, recommend streams, adapt homepage content, and promote channel subscriptions (Geyser, 2023). Snapchat's premium feature, 'My AI', is a chatbot that allows users to customize and interact with their own conversational AI by asking it for place-based recommendations (Snapchat, 2023). TikTok's AI-driven recommendation feature (i.e., 'For You' page) surfaces content that is likely to resonate with users via an AI engine that evaluates content tags and user profiles (Chakaravarthy, 2023). Tween users are actively engaging with these types of AI-driven features daily. These usage statistics make it clear that tweens, specifically those from Black and Hispanic/Latino communities and from lower income households, are frequently exposed to AI-driven, screen-based products for extended periods of time. Our work engages with these student demographics directly and provides insight into how to effectively communicate the AI concepts that underlie these systems.

Tweens also experience AI-driven hardware in and outside of the home. According to a 2018 *Smart Audio Report* conducted by NPR and Edison Research, among the 410 smart speaker owners surveyed, 73% reported that their children use smart speakers in the home, with top requests being: playing music (55%), answering questions (44%), telling jokes (40%), playing games (28%), and helping with homework (25%) (Edison Research, 2022). Children, tweens, and teens in homes with these technologies are interacting with AI-driven digital voice assistants regardless of understanding their underlying mechanisms. Outside the home, tweens also experience AI-driven technology in cars that are equipped with AI driver assistance features such as forward collision, lane departure, rear cross traffic, blind spot warnings, and automatic braking systems (NHTS, 2021). As passengers, tweens are familiar with these types of automated systems at a basic level, but our work seeks to augment this familiarity via an AI elective curriculum that provides structure and depth to these everyday, yet opaque technologies.

AI Education Initiatives

The need to establish AI literacy among citizens has been a major driving force for the development of AITL standards. In their systematic literature review on the topic, Ng et al. mapped the current AITL landscape from 2000 to 2020, providing an overview of both government and academic research initiatives dedicated to coalescing AI education goals and pedagogical standards (Ng et al., 2023). At the global level, the United Nations Education, Scientific and Cultural Organization (UNESCO) reported in 2022 that eleven countries are developing education standards that integrate AI into existing STEM and computing curricula (UNESCO, 2022). To meet the global needs of these initiatives, working groups such as DigComp and the ISTE developed standards that incorporate examples of AI knowledge and skills that AI-literate citizens should possess, as well as example projects that educators can offer to students (Vuorikari et al., 2022; Black et al., 2023). At the strategic level, and as part of its requirements for citizens interacting with AI systems, DigComp 2.2 recommended certain knowledge, skills, and attitudes that an AI-literate citizen should possess such as the ability to "be aware of what AI systems do and what they do not" (knowledge); "use, interact and give feedback to AI systems as an end-user" (skills); and "[have] human agency and control" (attitudes) (European Commission et al., 2018). In the classroom, ISTE recommends using a student-focused approach via the implementation of guided, student-driven, online and unplugged activities that help them form associations between AI concepts and content (Black et al., 2023). From this growing body of global standards work, we further crystalize what teaching strategies are effective for middle school learners in the pursuit of a larger, AI-literate populace.

Historically, AITL frameworks and curricula have primarily been developed for higher education audiences. Published in 1995, an early textbook on the topic, *Artificial Intelligence: A Modern Approach*, sought to standardize the methods for teaching the fundamental concepts of AI to audiences at the undergraduate and graduate levels (Russell & Norvig, 1995). Discussing their motivations, authors Russell and Norvig stress that all theory should be grounded via descriptions of how AI principles apply to example agent operating environments, and that pedagogy should leverage students' existing CS knowledge. This delivery method is certainly sensible for secondary education CS students but is a challenging proposition for middle school students at varying levels of computer literacy. A more appropriate delivery method, outlined by Kumar and Meeden in 1998, proposes a more hands-on, tangible, robotics-programming approach that subsequent studies have proven to be effective in motivating learners from non-technical backgrounds across both primary and secondary education contexts (Klassner, 2002; Kumar & Meeden, 1998; Murphy, 2001). Our work seeks to further demonstrate the effectiveness of hands-on, collaborative, and activities-based approaches to AI learning but omits a robotics-programming approach for equity of implementation.

Educator focus is shifting towards a more holistic view of AI education that includes K-12 learning environments. Researchers have built on these educational foundations to further define what contemporary AI literacy entails at the K-12 level. These works articulated what competencies students should be able demonstrate after engaging with an effective AI curriculum. Long and Magerko propose an approach to teaching AI in K-12 that synthesizes existing literature into working definitions of AI literacy competencies, and scaffold a thematic framework for AI curricula organized into five themes: (1) *What is AI?*; (2) *What can AI do?;* (3) *How does AI work?;* (4) *How should AI be used?;* And (5) *How do people perceive AI?* (Long & Magerko, 2020). The answers to these themes were translated into a series of design considerations that an effective K-12 AI curriculum should incorporate, including, but not limited to: *Explainability*, or the inclusion of supplemental media or demos that help to convey how intelligent agents make decisions; and *Embodied Interactions*, or engaging students in activities that allow them to take the intelligent agent's perspective via online and unplugged activities.

An uptick in recent publications related to teaching AI to K-12 audiences optimistically suggests a movement to solidify curricula guidelines (Ng et al., 2023). One framework, developed by the AI4K12 Initiative (AI4K12.org), seeks to (1) Develop national guidelines for AI education for K-12 in the United States; (2) Provide an online AI instructional resource directory for educators; And (3) Cultivate a community of multidisciplinary experts focused on elevating the quality of the AI education ecosystem (AI4K12, 2020). This framework "defines what every student should know about AI and what they should be able to do with it" and categorizes AI concepts into what are referred to as "The Five Big Ideas in AI" (AI4K12, 2020). These ideas are defined as Big Idea 1: Perception – Computers perceive the world using sensors; Big Idea 2: Representation & Reasoning – Agents maintain representations of the world and use them for reasoning; Big Idea 3: Learning – Computers can learn from data; Big Idea 4: Natural Interaction – Intelligent agents require many kinds of knowledge to collaborate and interact naturally with humans; and Big Idea 5: Societal Impact – AI can impact society in both positive and negative ways. Along with offering this conceptual framework, this initiative has made publicly available a comprehensive directory of free online teaching resources including grade band progression charts that offer a detailed specification of what students should know about AI and what they should be able to do with it (AI4K12, 2020). This framework, among others (Black et al., 2023; Vuorikari, 2022) indicate both an exciting parallel expansion and compression of how AI can best be taught across education levels. This current work builds on these foundations by co-designing and implementing an AI elective curriculum around The Five Big Ideas in AI (Touretzky et al., 2019).

Teaching AI in Formal Learning Environments

Despite recent attempts to standardize AI education pedagogy and curricula, a unifying framework across all levels of education remains nascent. This has been attributed to the variation in learning needs between education levels, as well as learners' backgrounds and familiarity with technology. These factors necessitate pedagogical deviations in AI curricula content, structure, and implementation. For non-technical audiences (e.g., undergraduate students in medicine, business, or the humanities), AI course content is commonly delivered via lecture-based modules and make use of a variety of independent online learning resources such as AI Campus (AI Campus, 2024) or Elements of AI (University of Helsinki, 2024). These resources are meant to be completed independently by students (Laupichler et al., 2022). Laupichler et al. could not identify a universal pedagogical best practice, yet highlighted the "flip the classroom" teaching technique in which students were tasked to independently work through AI learning material and then applying newly acquired knowledge to in-class discussions and project working sessions (Bishop & Verleger, 2013; Laupichler et al., 2022). They also found that courses often provide students with programming environments in the form of dedicated lab spaces providing computers (Vazhayil, 2019) or access to digital development environments (Rodríguez-García, 2021). Of the evaluated courses, the vast majority leveraged a two-part strategy of motivating students to independently study AI topics and then prompting them to reinforce their own knowledge via hands-on activities or projects (Shih et al., 2021; Xu & Babaian, 2021). They conclude that while AI educators at these levels can successfully teach AI topics to students with non-technical backgrounds, they do not know how best to structure course content (Vazhayil, 2019) and recommend the creation of pathways that allow educators to easily evaluate AI teaching frameworks to inform their pedagogy (Laupichler et al., 2022).

Regarding the structure of AI curricula in higher education, Laupichler et al. found that most AI-oriented courses dedicated early units and lectures to establishing foundational knowledge including definitions of AI, its history, capabilities, and limitations (Laupichler et al., 2022). Most courses also covered machine- and deep-learning content. Related to format, Xu & Babaian organized content into several modules, spread across 15 weeks, categorized under: *AI foundations and intelligent agents*, *knowledge representation and probabilistic reasoning*, *problem solving*, *machine learning*, and *ethics.* These were presented via lectures and presentation slides (Xu & Babaian, 2021). Case studies were used to contextualize topics and Python programs were used to demonstrate how various algorithms such as breadth-first search work. Students participated in classroom discussions, in-class exercises, office hour visits, and group work. Finally, a final term project was deployed that asked students to form groups, submit a project proposal, collect data, implement algorithms, and submit a final term paper (Xu & Babaian, 2021). Though this work's delivery methods loosely resemble those used in higher education, drastic and fundamental differences in teaching strategies, content structure, and assessments are required when teaching AI topics in K-12.

AI education for K-12 settings is commonly offered via more hands-on and group approaches through activities and age-appropriate projects, especially for younger children (Yi, 2021). Delivery variation can also be attributed to the differences in educator motivation for learners in primary and secondary education; with the former being oriented towards professional preparation and the latter oriented towards establishing foundational knowledge about AI (Laupichler et al., 2022). In recognizing these disparities between learning contexts, Laupichler et al. have organized the current publication landscape into two general themes: works meant to inform AI education theory and definitions (Long & Magerko, 2020) and works meant to inform the ways that AI should be taught to non-experts (Lin et al., 2021). Our work situates itself within the latter by expanding upon, and recommending effective implementation and assessment strategies that educators can leverage to build the foundational AI knowledge needed by middle school learners.

Literature for K-12 AI education modalities focus on two major pedagogies: *collaborative learning activities* and *game-based learning*. These two approaches, investigated and outlined by recent studies and guidelines (Black et al., 2023; Long & Magerko, 2020; UNESCO, 2019), enable educators to align curriculum content with students' interests, and to lower learning barriers using elements of play (Ng et al., 2023). The use of *collaborative learning activities* is a common and popular curriculum delivery method that incorporates elements of robotics and game-based learning via interest-driven, iterative experimentation, and allows students to practice and develop their interpersonal skills (Mota-Valtierra et al., 2019; Ng et al., 2023). The use of *game-based learning* activities remains a popular delivery method and has continued to succeed in driving interest and engagement through the medium of student-led projects. Students can successfully learn and integrate AI concepts into the design of their games and are able to gather immediate feedback through iterative playtesting or learn by playing existing games (Sailer et al., 2017). While this work does not specifically focus on game-based learning activities, it seeks to validate social-oriented AI learning pathways for middle school students.

Studies specifically investigating AI education for underrepresented groups in K-12 are few yet rich in terms of curriculum design and implementation insights. In their experience report, Lee et al. describe their summer workshop that developed and virtually implemented a 30-h AI (DAILy) curriculum over two weeks (Lee et al., 2021). This workshop focused on AI concepts such as decision trees and neural networks and structured AI concepts into several units: (1) An Introduction to AI; (2) Logic systems; (3) Supervised learning; (4) Neural networks); And (5) Generative Adversarial Networks. They found high levels of student engagement when discussing topics of bias and ethics in AI. The virtual setting posed challenges during their implementation such as limited student engagement. However, despite these challenges, learning outcomes were successful as measured by two instruments: (1) an AI Concept Inventory related to AI general concepts, logic systems, and machine learning, and (2) an Attitudes toward AI careers survey. This past work suggests that a successful middle school AI curriculum should focus on cultivating student interest in future AI careers and include in-person activities and discussion as much as possible, particularly related to issues of bias and ethics in AI. Our work incorporates each of these successful aspects. Professional development was found to be effective in improving the quality of AI teaching strategies in K-12. In a follow-up study, Zhang et al. found that teachers who had received PD in the DAILy curriculum were effective in teaching its concepts. When teachers taught the curriculum every day, students were able to make consistent learning progress. They also found no significant difference in learning improvements between student characteristics such as grade, gender, race, and ethnicity; students of different characteristics were able to benefit similarly from the curriculum. Our work incorporates these same professional development and teaching strategies in hopes of reaping the same learning benefits for students.

Design & Assessment Frameworks

Collaboration between Computer and Learning Scientists and middle school CS educators demand a structured approach to ensure equitable partnership. Several important theoretical and implementation frameworks exist that describe how researchers and teachers should collaborate to successfully construct, implement, and assess new technological curricula. These include Design-based Implementation Research (DBIR), the Technological Pedagogical and Content Knowledge (TPACK) framework, and the Substitution, Augmentation, and Redefinition (SAMR) model, and the revised Bloom's Taxonomy.

Design-based Implementation Research (DBIR) has long been used as a framework by education researchers for conceptualizing sustainable education innovations via research and practice partnerships. The characteristics of a DBIR project include: (1) A focus on persistent problems of practice from multiple stakeholders' perspectives; (2) A commitment to iterative, collaborative design; (3) A concern with developing theory and knowledge related to both classroom learning and implementation through systematic inquiry; And (4) a concern with developing capacity for sustaining change in systems (Fishman et al., 2013). This current work certainly meets these criteria in its goals for developing national guidelines for AI education for K-12 informed by iterative co-design and implementation. We engaged in a systematic researcher-practitioner approach that designed and implemented an AI curriculum such that its output may serve as a shared resource for both researchers and teachers.

Regarding equitable collaboration practices, co-design has been proven to elevate teacher voices in the design and implementation of CS/AI curricula. Cooenraad et al. leveraged participatory design to incorporate the diverse voices of students, teachers, administrators, and parents to produce a culturally relevant CS curriculum for marginalized communities (Cooenraad et al., 2022). Lin & Van Brummelen held co-design workshops with K-12 teachers to co-create lesson plans via AI tools and incorporated AI concepts into its curricula (Lin & Van Brummelen, 2021). Grover found that curricular co-design is an increasingly popular method of professional development for teachers who are unfamiliar with AI concepts (Grover, 2021; Grover et al. 2024; Severance et al., 2018). This current work collaborated with middle school CS teachers and adopted Steen's definition of co-design as "a set of processes to engage stakeholders and collaboratively work to identify requirements, brainstorm, and prototype solutions for new products and technologies" (Steen, 2013). Our insights contribute to the growing body of evidence for co-design's efficacy in producing culturally relevant CS and AI curricula.

Regarding teaching theory, the Technological, Pedagogical, and Content Knowledge (TPACK) framework has been used to define specific knowledge domains that teachers possess. It outlines the knowledge areas needed by teachers to successfully integrate technology in the classroom (Koehler et al., 2014). This framework is broken up into three categories: (1) Content knowledge (i.e., subject matter); (2) Pedagogical knowledge (PK) (i.e., teaching strategies); and (3) Technological knowledge (i.e., how technologies can be integrated into curricula). Combined, these three teacher knowledge categories can both afford and constrain how deeply teachers implement technology within the classroom. This framework provides valuable language in understanding how teachers choose to adopt or adapt different teaching strategies during implementation of the co-designed AI curriculum.

On the topic of teacher adaptations, the Substitution, Augmentation, Modification, and Redefinition (SAMR) model exists to classify the types of transformations teachers make during implementation. It is defined as a 4-level hierarchy in which teachers select, use, and evaluate technology in the classroom and proposes that teachers moving upwards along this hierarchy will achieve higher levels of teaching and learning outcomes in the classroom (Hamilton et al., 2016; Puentedura, 2006). At the *Substitution* level, teachers replace analogue technology for digital equivalents (e.g., replacing printed test questions with digital questions). At the *Augmentation* level, teachers introduce technology that augments a task (e.g., students read *and* listen to stories via personal handheld devices, as opposed to a shared teacher-led reading session). At the *Modification* level, a classroom task must be modified fundamentally with the incorporation of technology (e.g., a lesson on light waves using interactive simulation as opposed to static diagrams). And finally, at the *Redefinition* level, the use of technology creates novel tasks (e.g., students creating their own videos instead of writing a persuasive essay). This model serves as a valuable contextualizing framework for many of the teaching strategies used during implementation of the co-designed AI curriculum.

Regarding the design of assessments, Bloom's Taxonomy is a common reference for setting learning objectives and aligning assessment questions to ensure that students develop lower to higher order skills (Bloom et al., 1956; Thompson et al., 2008). Ng et al. adapted this foundational framework to AI literacy in an attempt to address the lack of a classification system for cognitive learning related to AI knowledge. Their revised Bloom's Taxonomy represents a hierarchy of learning levels that demand greater mastery over AI concepts at each level. These levels are defined as: *Know*, *Understand*, *Apply*, *Analyze*, *Evaluate*, *and Create* (Ng et al., 2021). Prior studies have also applied this model to classify assessment questions (e.g. Chang & Chung, 2009). We also leverage this revised Bloom's Taxonomy in the classification of assessments used in the co-designed curriculum.

Methodology

This project developed a 9-week ungraded elective course via in-person and online teacher professional development and co-design sessions. Its contents incorporate

The Five Big Ideas in AI framework and the Georgia AI standards, both of which are based on the AI4K12 national guidelines for teaching AI in K-12 (AI4K12, 2020). This work is dedicated to the development of competencies, career awareness, and interest in artificial intelligence for both Georgia middle school teachers and students. It is a response to the NSF Dear Colleague Letter 20–101, Advancing Educational Innovations that Motivate and Prepare PreK-12 Learners for Computationally Intensive Industries of the Future (Martonosi & Marrongelle, 2020). The curriculum provides students with opportunities to explore how AI works, how it is designed, and how it impacts their personal lives and communities. In addition, students learn about the wide range of professions in which people design and use AI applications in their work. The middle school elective and teacher PD program were collaboratively designed by the chair and co-chair of AI4K12.org, the Georgia Department of Education, and teachers in districts that represent geographically and ethnically diverse populations in Georgia: rural, suburban, and urban districts with Black, Hispanic, and White students. Co-designing the curriculum for these diverse contexts produced materials that were student-centered and culturally relevant to each. The AI elective course was iteratively refined over multiple offerings akin to a design-based implementation research (DBIR) approach and incorporated feedback from evaluators, teachers, and students (Barab, 2022; Fishman et al., 2013).

Program Partners & Participants

The research team obtained district approval for teacher participation in the initiative and worked with school administrators to identify teacher candidates. Participation criteria asked that teachers had obtained CS certification and expressed a willingness to engage and participate in co-design sessions and implementation for any duration of the initiative. Administrators recommended teachers who met these criteria. Across all 3 years of the initiative, a total of 16 teachers had participated for any duration: 5 teachers participated in Year 1 (co-design and pilot implementation), 6 teachers participated in Year 2 (implementation-only and inclusive of 1 Year 1 teachers who had continued), and 8 teachers participated in Year 3 (implementation-only and inclusive of 1 Year 1 teacher who had continued). The total number of students who enrolled in the elective was 1551. The current work focused on 8 teachers who completed Units 1 and 2 of the co-designed curriculum spanning Years 1 and 2. The 8 teachers offered the elective to a total of 759 students. Table 1 describes each teacher's school category, program year, background, student demographics, and number of enrolled students.

Co-design Methodology & Prior Work

As outlined in our previous work (Gardner-McCune et al., 2023), during Year 1, our co-design process engaged researchers and middle school teachers in a year-long, three-phase process to design a middle school AI curriculum that teachers adapted and implemented in their classrooms. The co-design team was composed of 5 researchers, 3 curriculum & professional development specialists, 2 evaluators, and

Table 1 Georgia Middle School Teacher and Student Demographics

Teacher ID	School Category	Program Year(s)	Background	Gender	Student Demographics	Enrolled Students
A	Suburban	Year 1, Year 2	Web design/html, Javascript, CSS, game design	Female	Predominantly Black	255
B	Urban	Year 1	AI/CS curriculum development, game design/VR, web development, Javascript and Python	Female	Predominantly Black	47
C	Rural	Year 2	Block programming, Python, Javascript	Female	Predominantly Black	50
D	Urban	Year 1	No previous CS classes	Male	Predominantly Black	19
E	Urban	Year 2	Mathematics, Business Education, CS Ed certification	Female	Predominantly Black	85
F	Urban	Year 2	8 years total, CS certification since 2018	Female	Predominantly Black	85
G	Urban	Year 2, Year 3	Taught CS for 9 years	Female	Predominantly Black	30
H	Rural	Year 1, Year 2, Year 3	Taught CS classes (Raspberry Pi, Python, pressure sensors), Arduino, 3D printers, competitive programming team	Male	Predominantly White	188
					Total Students	759

5 teachers. Weekly 1-h sessions were held over 33 weeks. Activities during these sessions varied according to the current phase of the co-design process.

Phase 1 had researchers frame AI curriculum ideas to teachers who then ideated on improvements over the course of 12 weeks. The goal was to determine the scope of AI concepts needed to include in the curriculum, the nature of student activities, and resources needed for implementation. Inputs included overviews of unit topics, 30 h of teacher PD via lecture and interactive activities, and teacher interviews to learn about their respective teaching experiences, styles, students, and backgrounds. A researcher played the role of facilitator, 2 researchers scaffolded the first-iteration curriculum, 1 teacher acted as a teacher-positioning expert, and 5 teachers acted as middle school teaching experts. Example questions to teachers included: "*Do the concepts make sense?"; "Would this work in your classroom and for your students?"*; "*What resources and support do you need to teach and engage your students?"*; And "*What might other teachers who didn't engage in professional development need?"*.

Phase 2 saw teachers piloting the curriculum resources over 16 weeks and adapting them to align with their teaching styles and student needs. Co-design questions for this phase included: "*How are teachers adapting materials?*"; "*What curriculum materials are working or not working?*"; "*How are students engaging?*"; And "*What resources are needed?*". Five middle school teachers reviewed the slide and activity materials as they prepared for instruction each week or day. Researchers assumed the role of observers and coaches throughout the pilot. During co-design sessions, teachers shared how their pilot implementations were going, shared student work and new resources that they created, and acted as peer-mentors for each other. The team ideated improvements and resource modifications. Example adaptations included: increased personalization of materials such as stylizing slides, adding images, trimming and pacing lessons, creating student versions of slides, creating worksheets, and creating unplugged activities.

Phase 3 saw teachers frame new curriculum ideas and adaptations over the course of 5 weeks. Co-design goals included: re-scoping curriculum concepts, refining student activities, creating lesson plans for new teachers, refining resources for concepts, and addressing student engagement and relevance-related challenges. Teachers were positioned as experts in teaching the curriculum and as curriculum developers. During sessions, the team reviewed each unit and discussed changes to be made.

The output co-designed curriculum was evaluated for educational efficacy by The Findings Group, LLC, an Atlanta-based research and evaluation firm that provides evaluation services to K-12 public education programs (The Findings Group, 2023). This evaluation process aligns with Laupichler et al.'s recommendation that newly created AI courses should be properly inspected to ensure "accountability and quality of teaching practices" (Laupichler et al., 2022). By engaging in this multi-disciplinary and collaborative process, we attest to the credibility of the co-designed curriculum regarding lesson material, age-appropriateness of content, and educational quality standards.

The co-designed curriculum is organized into three units, each housing several sub-modules and named as follows: *Unit 1—Autonomous Robots & Vehicles*, *Unit 2—How Computers Understand Language*, and *Unit 3—Decision Making and Learning*. This curriculum was then used and adapted by Year 2 and Year 3 teachers for subsequent implementations. For the purposes of this current work, only Year 1 and Year 2 teacher implementations of Units 1 and 2 of the curriculum were analyzed. Not all Year 1 and Year 2 teachers were able to complete Unit 3 within their implementation schedules.

Data Collection & Analysis

Across all years of the initiative, data was collected from participating teachers. This work features 8 teachers who implemented the co-designed curriculum as an AI elective course in their respective schools over a 9-week period, starting in the spring of 2021. The research team followed up with teachers to collect qualitative data regarding implementation (and how it was being received by students) via 9 classroom observations, 8 teacher interviews, 2 student focus groups, and 6 student interviews. Artifact analysis was conducted on the output co-designed curriculum slides and Grounded Theory Method was used for all qualitative data analysis, first by analyzing data per data collection method, and then again across methods to derive higher order themes (Muller & Kogan, 2010).

Researchers conducted pre- and post-implementation virtual interviews with the teachers individually, centered around the pre- and post-co-designed curriculum materials to gather his or her thoughts about its content, format, and appropriateness. Teachers were then asked to reflect on specific unit modules and assessment strategies they found effective during their implementations.

During classroom observation, researchers circulated throughout classrooms to engage individual students and groups of students to inquire about their work and impressions of the curriculum. At the end of 2 observation sessions, researchers engaged in focus group discussions involving groups of 4–7 students, and posed questions related to each group's impressions of the curriculum, its projects and activities, the aspects that they found enjoyable (or not enjoyable), and topics that they would have liked to explore more deeply. After observation, researchers either debriefed together to compare field notes and summarize observations, or individually record reflections via audio recorder or written summaries. Data were captured via classroom observation notes, recorded teacher interviews, recorded student focus groups, recorded student interviews, annotated curriculum slides, and student assessment artifacts. All recorded video and audio were collected with the consent/assent of school administrators, teachers, students, and student families. Researchers interviewed a total of 6 students individually to discuss their project work and asked them questions regarding how they felt about the curriculum. Student interviews were captured via audio and video recordings and centered around produced project work.

Analysis of classroom observations was done by first aggregating all individual researcher field notes and media into summary documents and uploaded to Atlas.ti (ATLAS.ti, 2023). We used a mixture of open and axial coding techniques to categorize raw observations into higher-order themes. These were then used as thematic framework for subsequent analysis of teacher/student interviews and student focus groups (Corbin & Strauss, 1990). Open qualitative coding was then conducted to identify and organize data into categories related to: student demographics (including how many students were present, their familiarity and experience with AI and technology, race/ethnicity, and gender); teachers' and students' use of technology (including its use in classroom management, lesson delivery, and student assessments); and teaching techniques (including behaviors related to driving student engagement, assessing student understanding, and tactics for classroom management).

Following the initial open coding of all 9 observation summaries, emergent themes were sorted by frequency and further articulated into sub-categories to be used as an analysis framework during second-iteration, axial coding conducted on all 9 summaries. Analysis of the 8 teacher interviews began with uploading and transcribing all video recordings into Dovetail (Dovetail, 2023). Open qualitative coding was conducted on each interview transcript to establish a thematic framework for second-pass axial coding (Corbin & Strauss, 1990). The coding of interviews focused on identifying and organizing teacher responses related to: aspects of the curriculum that they found to be engaging for their students, rationales for adjustments to materials and activities, and rationales for using certain assessment techniques or activities during lessons. A similar analysis process was used for the 6 student interviews and the 2 student focus group sessions. Initial open coding and secondary axial coding of student interviews/focus groups were dedicated to identifying and categorizing responses related to: aspects of the curriculum that students found engaging (or not engaging), and how well they understood its material. Finally, all collected and coded data were analyzed holistically to identify patterns in the curriculum's structure, implementation, and teacher and student behaviors and reflections.

Results

Research Question 1 – What is an effective curriculum format for teaching AI in a middle school elective as identified by teachers and students?

This question seeks to propose an effective format used to organize and implement the co-designed curriculum such that teachers, curriculum developers, and education researchers may benefit by way of its example. The units of the co-designed AI curriculum took the form of online Google Slides presentations, worksheets, and interactive activities. The 2 analyzed units were titled *Unit 1—Autonomous Robots and Self-Driving Vehicles* and *Unit 2—How Computers Understand Language*. Each unit contained 6–7 modules, each with 30–50 slides, and focused on teaching one or more of The Five Big Ideas in AI. Some modules also introduced and described assessment worksheets and instructions for in-class activities. Tables 3, 4, 5 (found in the appendix) describe these modules, activities, and related Big Idea(s) within Units 1, 2, and 3 respectively. Each module's set of Big Ideas were derived by analyzing their co-designed learning objectives. This paper focuses exclusively on the co-design and the deployment of materials used in Units 1 and 2.

Materials for Unit 3 are currently in the process of active co-design and implementation. The outcome of these activities will be published with the AI4GA website curriculum materials upon completion.

Teacher Perspectives

Slide content elements were found to fall into one or more of the following categories: ***Activators***, ***Definitions***, ***Prompts***, ***Asynchronous Tools***, ***Relatable Examples***, and ***Activities***.

Activators are introductory classroom procedures that take the form of questions, activities, or videos meant to focus students' attention, and introduce or reinforce new knowledge. Examples of activators include posing the question "What do you think AI is?" or playing a video such as the *Evolution of Boston Dynamics Since 2012* (Boston Dynamics, 2019). Teacher H elaborates on what purpose activators serve:

> "*I would say that's dual purpose. It is content reinforcement, but also it's effective classroom management because it's a way that students can come in, know the expectation, know what you've set and be working on that.*"

Teacher B articulates in more detail how video-based activators in particular are effective in engaging students:

> *"Of course the activators… It's hard to go wrong with that with middle school because they're coming in from a transition. It's a good way to start the class… It's passive… also including activators sometimes just inspires them… it doesn't have to always necessarily be artificial intelligence… we are exposing them to different career opportunities and different pathways that they can take."*

In other words, teachers and students benefit from the use of activators in that a natural classroom tone shift occurs that aligns both teachers and students towards two synergistic goals: (1) teachers can offload the task of classroom management onto the activator, and (2) students focus their attention and curiosity onto the activator.

Definitions are terms and their associated meanings presented in text and verbally described by teachers. An example being *AI is a branch of computer science that studies techniques for getting computers to produce intelligent behavior*. In her interview, Teacher A critiqued the pre-co-design definition for "perception":

> *"'Perception is the extraction of meaning from sensory information using knowledge.' There's a lot of big words and these aren't terribly difficult big words, but they [students] have to learn the word 'perception', which they've probably never defined before this class. Then they have to use the word 'extraction', which some of them are going to be familiar with but they might not have pinned down what that word means... And so, in that you've got 'perception', 'extraction', 'sensory'... at least two of those words are kind of new in this setting... So I would've probably rewritten that."*

This critique of pre-co-designed curriculum vocabulary demonstrates Teacher A's ability to anticipate her students' challenges with comprehension, and, in the face of time constraints, reaches the conclusion that simplifying the definition would be the optimal solution. In this manner, definitions throughout the co-designed curriculum were found to have undergone similar condensing and simplifying revisions to improve student comprehension. In this way, Teacher A exercised her Pedagogical Knowledge to tune curriculum materials to the benefit of her students (Koehler et al., 2014).

Prompts are questions posed further within the lesson that are meant to assess students' understanding, motivate classroom discussion, or introduce new topics. Examples include posing the question "Do all robots have AI?" after the teacher presented the definitions for both robots and AI, or asking students to identify the types of sensors used in autonomous vehicles to detect traffic signs. *Module 1.2* showed 12 slides with the prompt "Robot or Not?" coupled with images of various objects like self-driving tractors, TikTok, and robot vacuums. Students were then asked to determine if each object was a robot via in-class discussion. When asked why her students enjoyed this activity, Teacher B articulated:

> *"Because it caused a debate! I mean, we had to think about different criteria to determine whether or not it was actually a robot and it was like 'I think it is', 'I think it's not'. Well let's go back to the list! … To see them kind of get those concepts and apply them, it was really fun to watch!"*

In ways like this, the consistent use of interwoven prompts across implementations sought to both reinforce new knowledge through guided yet natural classroom discussion.

Asynchronous Tools are instances of interactive polling software that can be used asynchronously to capture student responses and/or facilitate real-time collaboration (Quizizz, 2023; Slido, 2023; Nearpod, 2023a, 2023b, Pear Deck, 2023). Teacher E had fallen ill at the time of her implementation of *Module 1.6,* which asks students to watch a news report about drivers sleeping in self-driving cars. Students were then meant to discuss the video's ethical implications. Since she could not facilitate this discussion in real-time, Teacher E adapted the lesson slides into an interactive and asynchronous format via Nearpod. After independently watching the news report, students were asked to leave comments and respond to open-ended questions related to what they saw. The asynchronous nature of the Nearpod format gave her the ability to (1) maintain her lesson schedule despite experiencing a setback preventing her from teaching, and (2) asynchronously assess her students' understanding of the material by analyzing their submitted responses upon her return. This adaptation classifies as a *Substitute* application of technology standing-in for an analogue equivalent in the SAMR model (Puentedura, 2014). Teachers C, G, and F also found that the use of Nearpods or Pear Decks provided students with lesson structure and format with which they were already familiar.

Relatable Examples refer to the use of relatable, real-world anecdotes that help students associate new knowledge to their lived experience. These took the form of references to movies, social media apps, GPS and search tools, robot vacuum cleaners, etc. This type of content occurred most frequently across all units. *Module*

1.1—Living and Working With AI had the largest collection of examples across the curriculum. Teachers also stressed the importance of real-world examples in facilitating ‘aha’ moments. One such moment was when Teacher G used a co-designed example that articulated the distinction between *sensing* and *perception.* This example featured an image of a deer that had accidentally wandered into an Aldi supermarket (Fig. 1). This particular example was effective in solidifying her students’ understanding.

However, Teacher F expressed a nuance that the number of shown examples should be tuned to students’ collective understanding measured through classroom discussion:

> *“If the presentations had more than enough examples, I would just cut a few. Some students would get it from one example, some students wouldn’t…in the folders that I re-did, you’ll see discussion prompts throughout the presentation because that’s kinda how I assess them in class. We would discuss and then I would put a discussion post up and they would respond. So with that… I would cut off excessive stuff.”*

In other words, teachers appreciated having a pool of relatable examples to draw from, but ultimately relied on classroom discussion and prompts to assess student understanding, and to determine how many examples to feature. Again, an expert application of Pedagogical Knowledge (PK) (Koehler et al., 2014).

Activities refer to teacher-led, individual, or group-based exercises, assignments or projects that assess understanding through active or constructionist-based work (Ali et al., 2019). An example activity includes the *My Dream Robot* project in which students design their own personal robot as they progress through

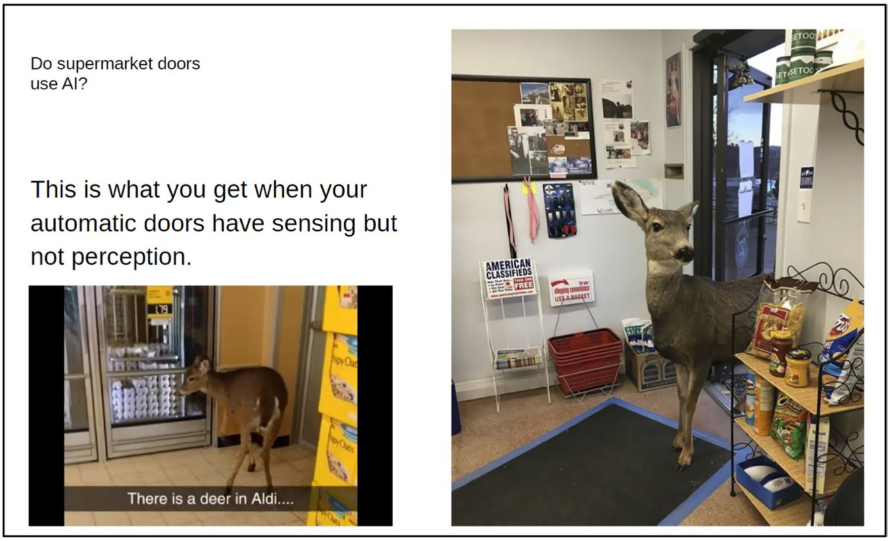


Fig. 1 Example slides featuring a deer walking into Aldi used in Module 1.1 to distinguish ‘sensing’ vs. ‘perception’

Unit 1. At the end of each module, students were presented with a “Ticket out of the Door”. This was an end-of-lesson activity in which students were asked to make increasingly lower-level design decisions about their robots while integrating new concepts. Students were allowed to design their robot via the modality of their choice; some chose written worksheets, slide presentations, or 3D models to depict their robots. Teacher G found that the former worksheet-based format was less supportive for her students who find writing to be challenging. She thus pivoted to a slide-based format:

> *“The writing was a problem… they didn’t have a problem with putting together a slide. That’s how I had to modify that, putting together a slide that showed what their robot could look like… and giving me a sentence or two to describe what it does.”*

In other words, curriculum activities in which students were able to leverage their creativity, e.g. via slides as opposed to written assignments, were found to effectively address reading inequities within the classroom.

Instances of each of the above content categories found throughout Units 1 and 2 were counted and a total of 383 slide content elements were analyzed. 52% of slide content was dedicated to *relatable examples*, 22% to *prompts*, 14% to *definitions*, 8% to *activities*, 3% to *activators*, and 1% to *asynchronous tools*. Figure 2 represents these content type frequencies and proportions for each unit and combined.

In practice, teachers adapted modules in ways that they felt would maximize their students’ engagement, such as making format transformations of curriculum content and lesson activities. To surface meaningful comparisons between all 8 teachers, we report on how each taught Unit 1 in their respective classrooms. Teachers divided lesson content across multiple days to present it in ways that were engaging, relevant, and comfortable for their students. For example, when teaching Module 1.1, Teacher A dedicated individual class periods to a single big idea of AI. This was done to mitigate student boredom in the face of text-heavy module content, which was her main concern:

> *“I think the first time I used it [slide presentation], I’m mostly certain I lost them cuz it was a lot of information. It was long and there wasn’t built-in engagement.”*

When probed further about what would be more engaging for her students, Teacher A responded:

> *“At the very least, back and forth discussion and at the… most extreme… full blown activities like taking a piece of a lesson and being like, ‘we’re only gonna learn this and this is gonna be ALL activity’.”*

Her solution to this was to create Nearpod presentations that incorporated fill-in-the-blank questions, quiz questions, and a review game, i.e. “Time to Climb”, a Nearpod feature that gamifies learning material (Nearpod, 2023a). For Teacher F, her solution to student boredom with Module 1.1 was to pull in relevant video

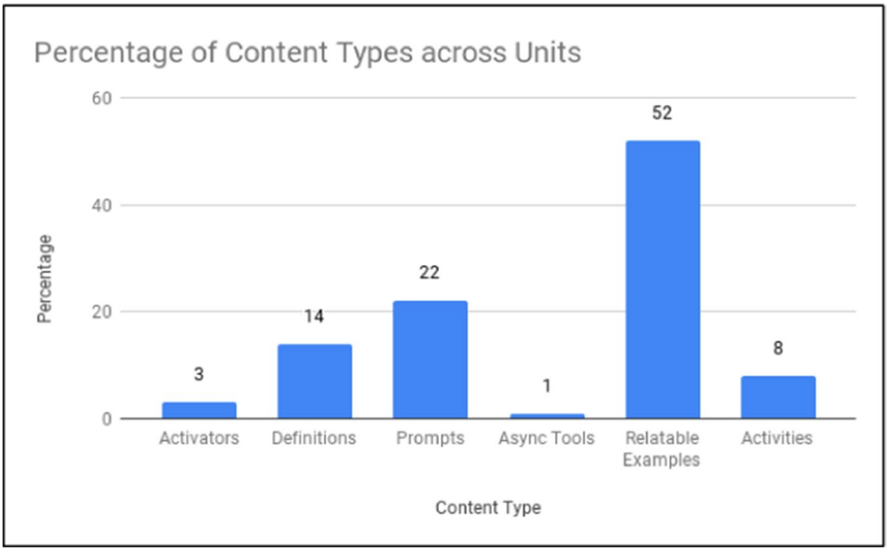


Fig. 2 Percentage of Content Types found in Co-designed Curriculum Slides (n = 383)

examples from Unit 2 as she introduced each of The Five Big Ideas. For Module 1.5, Teacher E divided her lesson across three days because she realized that students needed more time to fully absorb the concept of route finding:

> *"I discovered that all my students didn't see it as clearly as I see it. But that's okay! I still enjoyed teaching it. I didn't mind slowing down a little bit. What I did discover was I couldn't teach it in just a session; day one, day two, day three. I needed to give them some time and then come back and assess."*

Through dividing lessons into smaller, interactive, collaborative discussions and activities, teachers were able to drive student engagement and clarify AI concepts through in-class discussion.

To adapt to unexpected classroom scenarios, teachers deployed independent or self-paced activities. Such scenarios included unavoidable teacher absences, meeting the demands of teaching in joint classrooms, and mitigating student boredom. For example, at the time of his implementation of Module 2.1 and Module 2.2, Teacher H was responsible for teaching two classes of students simultaneously and so he, like Teacher E, created independent assignments for his students, which afforded him the ability to support both sets of students in the classroom while other students worked during his temporary absence. This use of independent assignments in the face of temporary teacher absences both within and outside of the classroom is noteworthy.

Student Perspectives

Analysis of the 6 student interviews, 9 classroom observations, and 2 student focus groups uncovered several major themes in student preconceptions about the difficulty of learning AI concepts, classroom social dynamics, students' keen interest in ethics and AI, student-offered curriculum improvements, and student reflections about their future careers.

Two out of the six students interviewed expressed that they had preconceptions about how difficult the AI material would be to learn, as well as its relevance to their everyday lived experience. This was best expressed by one student:

> *"At first, it almost seems like it's just like robots and you feel like you have to be a certain way to work with it, and you have to have this and that, but really **it's not that hard**. And I think it's definitely helpful no matter what I do; to be able to know what I'm using… computers and having Siri and stuff… you just know so much more about it and how it works… **I think that's just probably the best part about this… is just being able to have that knowledge.**"*

In other words, this student overcame her hesitation about learning AI concepts prior to taking the elective and was able to incorporate its material into her understanding of computers and AI. Finally, she was able to relate this knowledge to her own personal experience using AI-enabled products like Siri. This newfound understanding gave this specific student confidence in knowing how AI-enabled products operate within her everyday life.

Classroom observations noted the use of self-directed, pair, and group activities to connect AI concepts to lived experience. A pair of female students worked on their Unit 3 project that had them design and train their own AI. One student explained that she was "feeding the machine" by supplying her model with pictures of horses and cows and further shared that she works with farm animals at home daily. Students also chose to work in groups when afforded the opportunity such as in Unit 2's *Case Study – Is Alexa Safe?* assignment. This activity has students read articles related to Alexa and then engage in whole-class discussion and debate. When probed by researchers, a student within a group shared that her younger cousins used Siri at home and that they were becoming disrespectful of their parents, citing that Siri never says 'no' (Fig. 3). Another student shared that they have a Google Assistant at home which they tried to use for answering homework questions (it did not). Teacher B expressed that this type of reading and discussion-based activity is commonplace in her classroom and is one of the factors for her relatively slow pace when moving through the co-designed curriculum. This emphasis on group work, coupled with providing space for discussion has strong implications for the structure and pacing of future curriculum iterations. These examples reinforce the effectiveness of self-directed projects, working within groups, and providing opportunities to link material to lived experience.

On the topic of in-class discussion, it was observed that students greatly enjoyed engaging in debates around ethics in AI. During her interview, one student expressed that she liked having the opportunity to talk about autonomous vehicles. This then

Fig. 3 A group of students working on *Unit 2 Case Study – Is Alexa Safe?*

progressed into a discussion about the issue of AI "taking jobs away from certain people but then it gives more jobs in other directions."

When asked if they could envision future careers in AI, student responses ranged from ambivalence to keen interest. Out of the 6 interviewed students, 2 expressed no change in their intended career paths in medicine, law, genetics, and criminal justice. Two students acknowledged that AI may play a part in their future careers, but it would not be the central focus of their work. Finally, students offered their own suggestions for curriculum improvement that included: more hands-on projects and in-class activities, use of physical learning aids like robots, more opportunities to work in groups, use of mnemonic devices to help memorize content such as the different types of sensors, and more videos and less text in slides. These student suggestions provide valuable insight into the next iteration of the curriculum.

Research Question 2 – What assessment strategies do teachers employ to understand student learning outcomes?

This question seeks to highlight the ways that teachers leveraged co-designed assessments during implementation. Dedicated work that examines student perspectives and quantitative measures of their learning is currently ongoing. Defined in the co-designed curriculum's outline, each unit's set of assessments strove to meet one or more learning objectives that aligned with specific revised Bloom's taxonomy labels, e.g. module 1.1's learning objective of *Describe AI to someone* corresponds to the *comprehension* level of the Bloom's taxonomy. These assessment types and strategies were categorized into the following revised Bloom's Taxonomy levels: *knowledge*, *comprehension*, *application*, *evaluation*, and *creation* (Table 2) (Ng et al., 2021).

During analysis, frequencies and proportions of each assessment were measured. A total of 43 assessments were analyzed. 30% of assessment learning objectives aligned with the *knowledge* level, 26% with the *comprehension* level, 12% with the *application* level, 9% with the *analysis* level, 7% with the *evaluation* level, and 16% with the *creation* level. Figure 4 represents the percentage of assessments aligned with corresponding revised Bloom's taxonomy levels.

Knowledge assessment strategies focus on measuring students' memorization of terms or concepts through recollection tasks. 30% of assessments aligned with this revised Bloom's Taxonomy level. An example assessment was the *Unit 1 AI Concepts Review* assignment in which students were given 10 concepts learned throughout the unit and were tasked with (1) defining the concept, (2) explaining the concept in terms that a five-year-old could understand (ELI5), (3) finding an image that represents the concept, and (4) describing why the concept is important to robots and self-driving vehicles. Task 1 specifically required students to recall their own working definitions of AI concepts and translate that understanding into writing. The student in Fig. 5 correctly defined "Object Recognition" as the process of identifying objects from visual input.

Comprehension assessment strategies focused on measuring students' understanding of basic AI vocabulary and concepts through comprehension-based means. 26% of assessments aligned with this revised Bloom's Taxonomy level. Students should be able to demonstrate surface-level understanding of basic AI vocabulary and AI concepts. Example assessments in this category were an activator-type, warm-up activity in which students answered essential questions from prior lessons. Figure 6 shows a jam board, a digital interactive whiteboard, from Teacher F's class posing two essential questions that focused on the sensors and subsystems that made up self-driving cars and autonomous robots. It provided a quick opportunity to activate students' prior knowledge and made it easier for them to connect to the next topic, how computers process the information from sensors and perceive the world. In the SAMR model, this could be classified as an *Augmentation* technology integration; the use of a virtual whiteboard enabled whole-class collaboration, circumventing constraints related to synchronous student expression in a traditional classroom discussion (Hamilton et al., 2016; Puentedura, 2006).

Teachers sometimes structured lessons that progressed through a single activity that started in simple terms and increased in complexity over the course of the class. For example, when teaching *Module 1.5—Route Finding*, Teachers A, C, and E used a Pear Deck presentation that, in combination with a worksheet, guided students in labeling a map of Georgia cities. Teachers added nodes to cities, linked said nodes, and then introduced the breadth-first search algorithm by having students color-code search levels for various route-finding scenarios (Fig. 7). Teacher A referred to this technique as "I do. We do. You do." to help convey the progressive nature of the activity. In the SAMR model, this again demonstrates the *Augmentation* of a traditional task (i.e., a worksheet) via the Pear Deck technology (Hamilton et al., 2016; Puentedura, 2006).

Upon reflection, Teacher C noted the activity's demand of students to critically think by themselves while she roamed the classroom and supported students as they worked:

> *"It gave me the opportunity to walk around in real time and say 'looking good', 'keep working'. 'that's great', 'that's perfect', 'uh-oh almost', you know? And I was even as technical as I didn't give them an 'attaboy until they had circled the best route. So they would sit there and go 'what is she talking about?' They would have everything completed; the whole breadth-*

Table 2 Revised Bloom's Taxonomy applied to AI education

Bloom's Level	Description	Example Task
Knowledge	Memorize an AI concept	Define "Artificial Intelligence"
Comprehension	Explain an AI concept	Explain how autonomous vehicles and robots sense the world
Application	Use an AI concept in a practical context	Find routes between two locations using breadth-first search algorithm
Analysis	Connect AI concepts across contexts	Explain which sensor is malfunctioning in a given scenario
Evaluation	Justify decisions based on AI concepts	Discuss the societal implications of self-driving vehicles
Creation	Instantiate AI concepts through *making*	Create your own chatbot

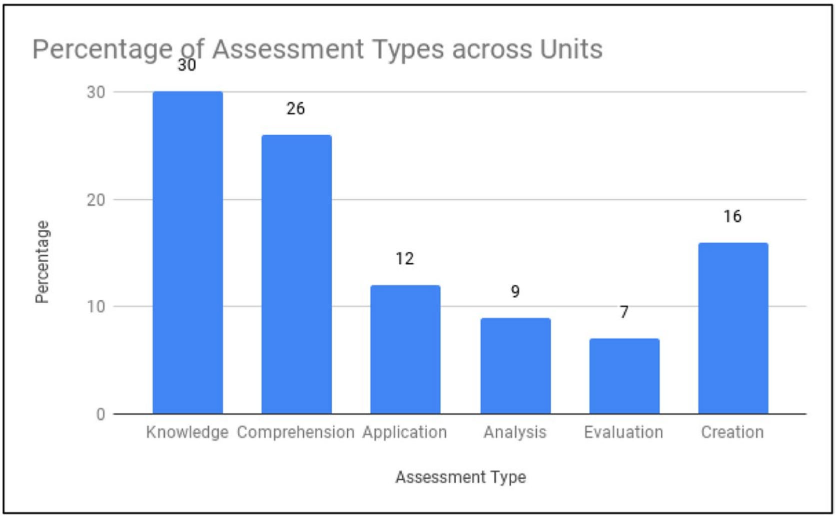


Fig. 4 Percentage of Assessment Types across Units ($n=43$)

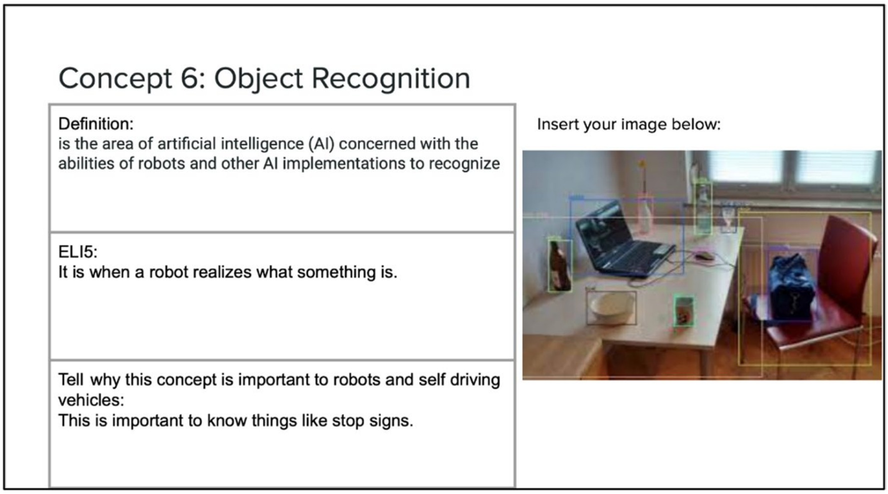


Fig. 5 Student Unit 1 AI Concepts Review assignment

> *first. But they had not done that final outline of the best route… And so I would just say 'almost' and so some of them would go, 'Oh, I got it! I think I know!'… so I appreciated the thinking that happened in that activity."*

Application assessment strategies focused on assessing students' use of AI-related learning concepts without having to demonstrate deeper-level knowledge via application-based means. 12% of assessments aligned with this revised

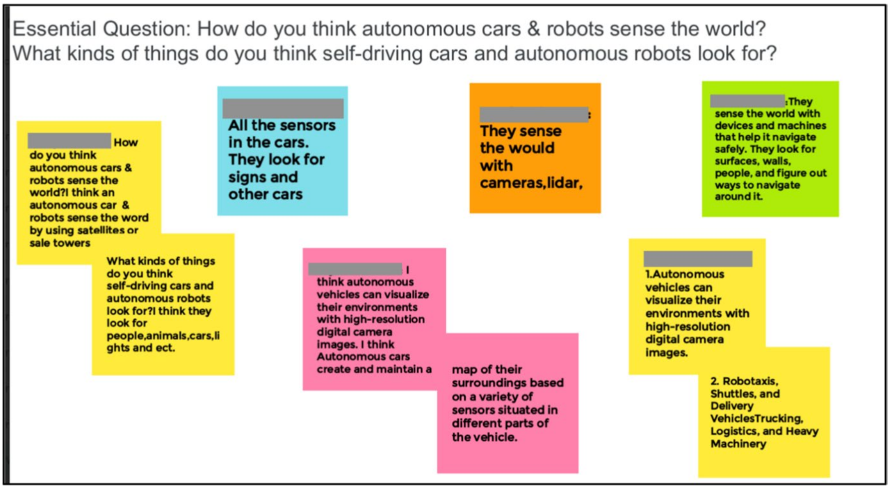
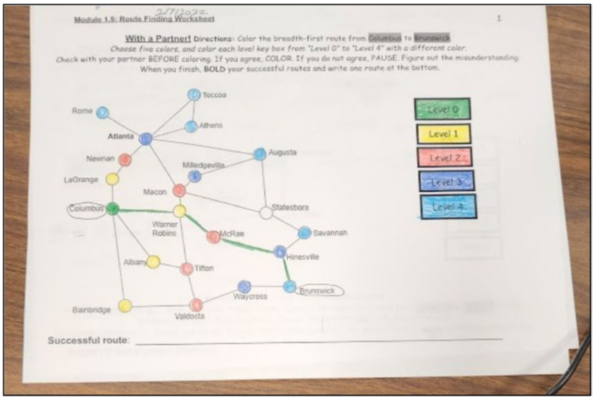


Fig. 6 A screenshot of a Jam board from Teacher F's class with students answering essential questions as a warm-up activity

Fig. 7 Student's route finding worksheet



Bloom's Taxonomy level. Students can apply their working knowledge in new and authentic situations. Unit 1's comic strip assignment asked students to create a comic that explains sensors on self-driving cars to students' family members. Teacher B commented on students' high levels of engagement:

> *"A lot of them enjoyed that. Again, they have the opportunity to express their creativity, work with partners if they like on this particular assignment, to give those different examples and to present it in something other than a Google Doc with several sentences on it."*

The comic strip assignment allowed students to exercise their creativity and humor while demonstrating their understanding of sensor limitations and mechanisms of operation in non-traditional writing formats. Figure 8 depicts a story in which children and adults teleport to different scenarios and adults explain what sensors would function well or malfunction under each scenario. This student showed a solid grasp of how RADAR, LIDAR and camera sensors work, along with their limitations.

Analysis assessment strategies focused on measuring students' synthesis of AI knowledge. 9% of assessments aligned with this revised Bloom's Taxonomy level. The strongest example of this assessment type was an unplugged activity related to word embeddings: a representation of words as points in a high-dimensional space such that nearby words have similar meanings. In this activity, Teacher A placed post-it notes on classroom walls to denote X and Y axes, e.g. X representing 'edible' vs. 'not edible' and Y representing 'animal' vs. 'not animal'. Students were then given index cards (Fig. 9) and asked to place themselves within the classroom in response to prompts like 'tomato', 'turtle', and 'human'. Students were asked to explain why they placed their card where they did and

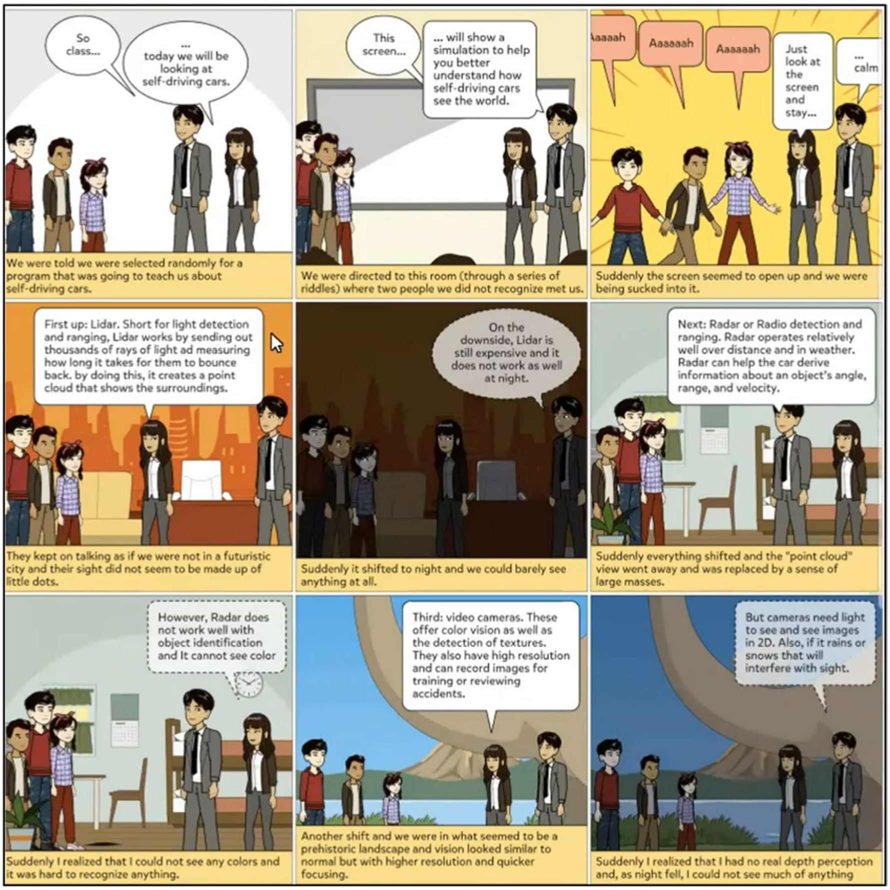


Fig. 8 Student comic strip depicting a story about how self-driving cars sense their surroundings

were encouraged to debate their opinions to support their positions. One such prompt was why some students felt that a turtle leaned more towards being an 'animal' and 'not edible' (although many students made the humorous case that turtles were quite edible). Teacher A then added Z-axis post-it notes representing 'cute' vs. 'not cute' and had students repeat the exercise, now in three dimensions. Finally, she asked students why this activity was done in a physical space to assess students' comprehension of the concept of multi-dimensional semantic feature space. After participating in this activity, students were shown an interactive visualization of word embeddings using a 300-dimensional feature space that a large language model would use. In an interview, one student commented that this embodied activity was particularly effective for her as a kinesthetic learner: "I just like that, cuz like we gotta move around, cuz I can't sit still for very long".

Evaluation assessment strategies focus on assessing students' ability to accurately critique produced work by incorporating AI knowledge into rationales. 7% of assessments aligned with this revised Bloom's Taxonomy level. Students should be able to evaluate their work or the work of others from a position of deep understanding of multiple AI-related concepts simultaneously. Unit 1's *Case Study: Sleeping in Self-Driving Cars* presented students with videos of Tesla drivers seemingly asleep while driving on highways and engaged students to debate about ethical questions related to the current state of autonomous driving technologies (e.g. should autopilot mode be banned). Teacher G pointed out that discussing real-world problems excited students and created a more dynamic learning environment than her usual CS classes:

> "Thinking about it from a moral standpoint **...** brought out good conversations **...** I just didn't expect that... that's what I appreciated the most out of this curriculum is how it, you know, allowed the students to just think and talk about things that were going on in a way that other things that I'm using as a computer science teacher hasn't done. And it's because of how current and relevant it is, that's interesting because we don't understand it and it seems taboo and, you know, that kind of stuff is attractive to students, you know?".

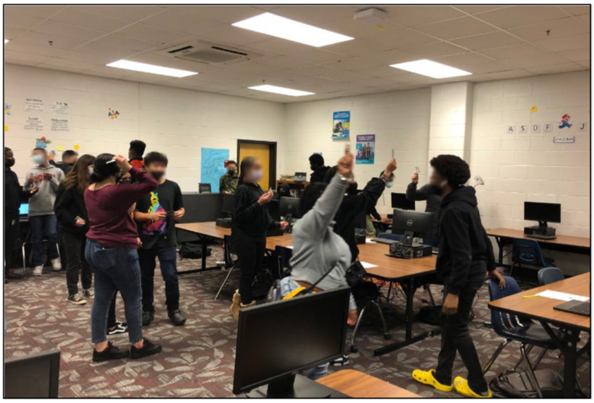


Fig. 9 Students exploring three-dimensional semantic feature space by placing an index card in the space where they perceive it belongs

Some students also connected material to their personal lives when debating about the impacts of autopilot mode. For example, one student shared that her diabetic father could benefit from autopilot in that it could enable him to take medication during emergencies: *"my dad is diabetic and you never know when your blood sugar will drop and you need to eat something."*

Creation assessment strategies focus on assessing students' ability to design or create work from a deep level of understanding through synthesis-based means. 16% of assessments aligned with this revised Bloom's Taxonomy level. Students should be able to apply their knowledge of multiple AI-related concepts to the creation of more advanced or complex work or projects. One example was Unit 1's *My Dream Robot* project in which students envision a robot for solving a personal problem that they care about. At the end of each module, students added increasingly lower-level details about their robot and how it would work based on the main concepts they learned from each module. Teachers gave students the freedom to choose whatever format they were comfortable with for their final deliverable. Student deliverables ranged from slide presentations, TinkerCAD models (Autodesk, 2023), and Hero Forge models (Hero Forge, 2024). During interviews, students expressed that they enjoyed hands-on projects like this because they had opportunities to envision a tangible robot, give it a name, think about sensors and explore its societal impacts. This project motivated students by allowing them to create meaningful AI-powered products and leaving plenty of room for customization and creativity. Figure 10 shows a student robot designed to prepare and deliver meals as well as clean up. This student used different sensors to help the robot function and critically considered how the robot can impact different stakeholders (e.g. people with obesity and people with difficulties preparing meals).

Research Question 3 – What aspects of co-design resonated with teachers while creating the AI curriculum and what do they imply for effective Professional Development?

This question seeks to understand what aspects of co-design resonated with teachers as a mechanism for professional development. Semi-structured interviews with each teacher were conducted after their respective individual implementations of the curriculum. During these interviews, teachers unanimously expressed positive feelings towards the process of co-design. They saw benefits not only in their own professional development, but in how the co-designed curriculum impacted their students. Teacher A best articulated this positive sentiment when reflecting on the impact that co-design and the resulting curriculum had on her and her students:

> *"When I came to the school, I wanted my computer science program to be something noteworthy. It didn't become what I wanted it to become, but this (program) is something I can point to and I can say… for sure…* ***I gave my students something that someone else could not have necessarily given them****… to the point that another computer science teacher wouldn't have necessarily been able to give that to them.*
> *Sure, I wanted them to do certain things with their programs or competitions or certain 'this' or 'that'. Most computer science teachers have access to that. (But) this (program) makes me feel like I can really brag on it… And I said it*

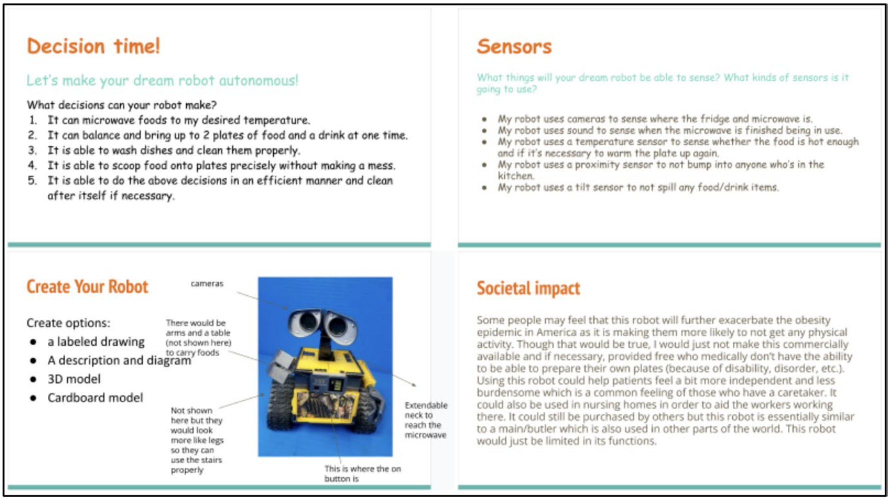


Fig. 10 A student robot designed to prepare, deliver, and clean up after meals

> *to my students: 'look y'all,* ***it's just us****. Y'all are the first people to do it. Your opinions are gonna be the first ones to change it. Nobody else has access to it.* ***This is college level material that y'all are able to handle.****' It felt very braggable. I loved being a part of this program… This definitely for me is a highlight of teaching here.*
>
> ***This was something my eighth graders engaged with well****. They had some really good opinions. They changed their opinions using the knowledge. They made some really cool looking projects. They dove into some of the assignments. It was all brand new. I loved it."*
>
> *"I've heard that when people do this (curriculum development), they don't do it the same way… the usual method is the professors or somebody makes it and they just give it to the teachers…. not taking anything away from people who have made materials and then just handed it to people, but I've certainly been on the receiving end of materials that I've been given access to and being like: 'they want me to do this, this, this, and this in a 45 minute class!? That doesn't make any sense!' I see now there's no way there was a teacher in the room (during curriculum development)… And so with these (co-designed) lessons, I see how the fact that we actually meet and discuss them and then mess with them together. I see how that makes a* ***realistic*** *lesson."*

In other words, Teacher A felt pride in being uniquely able to offer bespoke learning material for her students. She felt joy in seeing her students readily engage with lessons and shape their own understanding. She realized that the lack of teacher input in curriculum development explains her past negative experience with pre-made lessons and their unrealistic expectations for teachers.

The benefits of teacher collaboration were another strong inflection point heard throughout several interviews. Some teachers, like Teacher D, expressed that co-design, as a venue for educator discourse, also acted as a camaraderie-building mechanism in juxtaposition to her past solitary experience in education:

> *"I really enjoyed working as a team. I have been a loner ever since I began teaching. And this is the first time I can honestly say that I enjoyed working with the group (laughs)... Everyone's expertise… I appreciated their input. I appreciated the free speaking. Even if you didn't like it, you could say it… 'that's not gonna work for me'. But the respect of us, one to another… That for me… created a good bond between us."*

This sense of camaraderie stemmed not only from respectful discourse between teachers, but also from the ability to share and compare teaching tactics and practices and learning directly from AI experts who otherwise would have remained inaccessible.

Some teachers did not have the capacity to engage in the co-design and implementation process. This was due to the limited time and resources needed to devote to external projects that weren't related to immediate teaching responsibilities. Nonetheless, Teacher D acknowledged this fact and provided constructive suggestions in that an AI educational platform should be designed, developed, and made available for teachers like him who need a curriculum that is immediately implementable; thereby mitigating the time and resources needed to engage in collaboration and manual preparation, and instead put into practice a turn-key curriculum. Along similar lines regarding preparation processes, Teachers B and E expressed their appreciation towards the notes and Nearpod slides created by past teacher implementations as ways to accelerate their own learning and preparation.

Interestingly and seemingly in contrast to the previous point on ease-of-use, some teachers felt that the manual creation of materials should always occur regardless of the co-designed curriculum's maturity. Teacher H commented that the act of creating materials for the curriculum increased teachers' sense of ownership and empowerment over the material. This notion was validated by Teacher G who expressed continued interest in further developing the curriculum even after the conclusion of her implementation.

Discussion

AI literacy is conceptualized as a set of competencies, including critically knowing, building, and collaborating around AI technologies (Long & Magerko, 2020). We reflect on how our results cultivate such competencies regarding its content structure and assessments strategies. We also situate these results in the context of existing literature surrounding how AI education can be made more relatable to middle school learners. Finally, we discuss the effects that co-design has had on teachers regarding their feelings of empowerment and confidence in teaching topics in AI, and what these imply for the future of PD in K-12 AI education.

Curriculum Format & Assessment Reflections

Teachers Make AI Learning More Accessible to Diverse Students Via the Use of Specific Curriculum Formats and Engagement Techniques

Both Dai and Payne observed similar practices in that teachers framed their own epistemological understandings of AI when presenting students with real-word scenarios (Dai, 2023; Payne, 2020). Teachers appreciated the abundance of examples, but ultimately used their judgment on how many examples to feature for students based on student comprehension. Prompts were interwoven within modules, indicating that teachers emphasize providing ample opportunities for students to solidify their new knowledge, either via in-class discussion and debate, or by asynchronous means via Nearpod or Pear Deck presentations. In middle school, the use of asynchronous tools was less represented in curriculum slides themselves but were nonetheless commonly recalled in interviews and observed to be used in classroom observations across implementations. Asynchronous tools were also found to work well in higher education contexts to help students reinforce their new AI knowledge (Shih et al., 2021; Xu & Babaian, 2021). Definitions typically resided at the beginning of modules or as new concepts were introduced. This aligns with Payne's assessment that AI terms such as 'algorithm', 'robot', and 'programming' need to be explicitly clarified to avoid misconceptions (Payne, 2020). Activities were pivotal in driving student engagement within the classroom and teachers shined in their ability to balance classroom management and guide students. While activity slides themselves were largely left unmodified by teachers, some teachers chose to create activities out of what was originally static material to boost engagement. Use of activators were consistent at the beginning of all modules and teachers cited videos specifically for their ability to drive student engagement while offering students glimpses into AI-related careers. We posit that this orchestration of techniques successfully lowered students' perceived levels of difficulty for learning AI material while also reinforcing new AI knowledge by making it relatable.

Teachers Leverage Rich and Varied Strategies for Reinforcing AI Knowledge

Recognizing that each teacher, class, and learning context can vary widely, we suggest the types of teaching strategies that are effective for solidifying certain types of knowledge as characterized by the revised Bloom's Taxonomy. For solidifying learning at the *knowledge* level, effective strategies included simplifying definitions (e.g. "Perception") and memorization assessments (e.g., AI Concepts Review); techniques geared towards rote memorization. For learning at the *comprehension* level, strategies like activators and prompts (e.g., "Robot or Not"), as well as the use of asynchronous tools (e.g., Jam boards and Nearpod) can be effective; these strategies demand students to explain AI concepts in their own words. *Application*-level learning can be reinforced by the use of more hands-on activities (e.g., route-finding worksheet); these demand students to apply AI concepts towards more practical goals in real-time. *Analysis* level learning can be reinforced via activities that progressively introduce complexity while allowing for debate (e.g., word-embeddings);

these strategies allow students to synthesize their understanding of AI concepts over time through incremental guidance. *Evaluation* level learning can be reinforced via the use of relatable examples mixed with discussion-based activities (e.g., Aldi deer and case study); these strategies allow students to justify their positions using familiar contexts while incorporating new AI concepts. Finally, *creation* level learning can be reinforced with self-directed projects (e.g., "Ticket Out the Door" and My Dream Robot); these types of activities allow students to instantiate AI concepts incrementally and facilitated through the creative act of making. These strategies highlight how highly capable teachers truly are in designing effective strategies for introducing and reinforcing AI knowledge for their unique students.

Long-running Creativity-based Assessments Should be used to Increment Challenge Over Time While Reinforcing AI Knowledge

Our assessments target both lower and higher order levels of thinking. Noticeably, as students dove deeper into the curriculum, there was an increase in the proportion of assessments in the cognitive levels of creation. In Unit 1, the creation assessment, *My Dream Bot*, had students envision and design their ideal robot. This assessment was broken down into several parts and embedded at the end of lessons throughout Unit 1 to help students apply what they had learned in each lesson. Unit 1 was particularly unique from an assessment perspective in that the project acted as a unifying and incremental project. Its incremental addition of complexity seemed to contribute to students' sense of creative challenge as reflected upon by teachers and students alike. This suggests the importance of chunking, especially in intellectually challenging projects, to reduce students' cognitive load (Lah et al., 2018). Future AI units should also strive to offer unit-wide projects to reap these same learning benefits via incremental project progression. Given different assessment methods, future research should focus on identifying the limitations and strengths of each in the context of AI education. Effective remix of assessments should disclose both students' learning outcomes and thinking processes and will facilitate the design of learning environments. Finally, future work is needed to develop more AI learning assessment options geared towards the *creation*, *application*, and *evaluation* levels of the revised Bloom's Taxonomy (Ng et al., 2021).

Multiple Forms of Assessments Should be Embedded into AI Curricula to Assess Holistic Learning of AI Concepts

In describing assessments that spanned Bloom's Taxonomy levels, we argue that multiple forms of assessments, or "systems of assessment" proposed by Grover, should be embedded into AI curriculum over time to holistically evaluate student development of AI literacy (Grover, 2017). Different assessments provide different evidence for student learning. For example, Unit 1's case study assignment revealed deep levels of student understanding in their evaluation of AI technologies and their societal impacts. Interpretations of different assessments captured how well students developed expected learning outcomes and provided information on how the design of assessments, pedagogies and learning environments can be improved in future design iterations. During co-design and implementation, teachers integrated their expertise, or Technological Pedagogical Content Knowledge (TPACK), into assessments to meet the specific needs of their students (Koehler et al., 2014). Although assessments were aligned with learning objectives, there was variation in how information was presented and how students demonstrated their understanding. None of the assessment methods can provide a comprehensive picture of student learning, however teacher input is vital to the design of appropriate assessments that address the interests and priorities of students.

AI Education within Middle School Electives

Middle School AI Education is Heavily Designed to Maximize Student Engagement Rather Than Comprehensive Knowledge Coverage

Historically, AI education practices have been targeted at students in higher education using a range of independent formats and assessments meant to assess comprehensive and technical AI knowledge. These largely take the form of lecture- and self-directed project-based strategies (Russel and Norvig, 1995; Laupichler et al., 2022). For middle school learners, however, we've found that teachers go to great lengths to customize and prepare the curriculum's format and implementation based on their teaching expertise related to student interests, learning styles, and engagement needs. As found by Laupichler et al., this motivation is driven by primary educators' goal of establishing fundamental AI knowledge in students as opposed to career preparation (Laupichler et al., 2022). However, our classroom observations suggest that some teachers balance both goals. There are several means by which educators made lessons more relevant and engaging to middle school learners.

Effective Middle School AI Curricula Should Provide an Overabundance of Teacher Resources, Assuming and Encouraging Transformations

These techniques can be described as instantiations of the *Substitute*, *Augment*, *Modify*, and *Redefine* (SAMR) model (Hamilton et al., 2016; Puentedura, 2006). Depending on the resources available to individual teachers, different SAMR techniques were used. We posit that a robust AI curriculum should afford teachers the ability to make such modifications based on known resource limitations while conveying equivalent knowledge to students regardless of modality. By doing so, we also posit that teachers will have the scaffolding necessary to leverage as-is materials (e.g., assessments and slides) while building in affordances for individual teacher customization (e.g., assessments- and slides-as-templates). These materials would serve as implementation accelerators for practitioners given varied resource limitations. Such adaptations, like the use of video for prompting student discussions or the creation of embodied activities were done in efforts to drive student engagement, interest, and facilitate ease of learning abstract AI concepts.

Curricula Should Convey how AI can Augment More Commonly-known or Relatable Careers (as Opposed to Highlighting Careers in AI)

While adaptations were found to spark some student interest in possible careers in AI, some students' career plans remained unchanged even after the AI elective. These students acknowledged AI's ability to augment future work in fields like medicine or genetics, but none of the interviewed students expressed interest in pursuing an AI-focused career. This insight augments Lee et al.'s *Attitudes Toward AI Careers* survey results that demonstrated a statistically significant increase in students' *awareness* of AI-related or AI-enabled careers after intervention. Our findings suggest that more work is needed to visualize AI-focused career pathing and increase students' desire to pursue AI-focused careers. However, we hypothesize that students need even more personal and relatable examples of such pathing. Curricula should also convey how AI can augment more well-known fields of work (e.g., agriculture, medicine, etc.) or relate to students' personal wants and needs (e.g., doing chores or socializing with friends).

Engagement in Co-design Better Prepares Teachers to Facilitate Discussions Around AI Ethics, Which is of Particular Interest to Students

Students' willingness to engage in AI ethics discussions were strongly represented. This aligns with what Payne found in that students can be "conscientious consumers and ethical designers of AI" (Payne, 2020). This interest in discussions around AI ethics was also noted by Lee et al. (Lee et al., 2021). Discussions within this current work were largely motivated by students' conceptions and misconceptions surrounding AI, of which their views could be validated through socialization with peers and facilitated by teachers in the classroom. Similar to what Williams et al. report (Williams, et al., 2021), our work found that teachers felt equipped to facilitate such conversation in the classroom as a result of PD workshop engagement and co-design. This contrasts with traditional CS teaching practices in higher education, where some educators desire to leave the social impacts of AI and its ethics to other courses, prioritizing technical topics and citing a lack of time, incentive, or expertise in teaching such topics (Smith et al., 2023). Through PD and co-design, teachers felt that they were able to facilitate AI ethics discussions in a controlled fashion. This provides further motivation for establishing formal AI education PD pathways.

Teacher Training & Empowerment via Co-Design

Future Collaborations Should Position AI Experts as Sources for Theoretical AI Knowledge and Teachers as Grounded Shapers of Said Knowledge

Teachers prioritized how they would pragmatically convey AI knowledge while participating in co-design and implementation. As Dai noted, the participation of teachers in the development of their own epistemological understandings of AI in partnership with AI scientists acts as a filtering mechanism that distills AI knowledge into practical formats (Dai, 2023). Lin and Van Brummelen built on this notion via their Value-Sensitive Design approach during their collaboration with teachers to develop AI lesson plans. They found that educators emphasized the filtering of knowledge via aspects of practicality, namely: evaluation, engagement, logistics, and collaboration (Lin & Van Brummelen, 2021). We give credence to these insights by noting teacher reflections on how immediately implementable the co-designed curriculum was and the effectiveness of modifications made to the pre-co-designed curriculum. These included the breaking up lessons into multiple activities to slow down the flow of material, deconstructing and clarifying AI terminology for students, and making content more aesthetically pleasing. We provide evidence for the success of using The Five Big Ideas in AI as a co-design/PD framework that enabled teachers to build up their own TPACK models related to AI; specifically in their *Content Knowledge* (CK) and *Technological Knowledge* (TK) (Koehler et al., 2014).

Future Collaborations Should Incorporate a Community-oriented Approach When Iterating on AI Curricula and Strive for a more Streamlined Pipeline for Teacher Training, Co-design, and Implementation

Teachers felt that future collaborations, training, and implementations could be improved by several methods. Teacher D described how a future online AI learning platform, which houses all curriculum materials and activities, would make implementations more efficient. Such a platform would offer a more uniform user experience for teachers and students. Teacher B commented on how notes left by past teacher implementations within curriculum slides assisted her with her own training experience. This sentiment is echoed in Kim and Kwon's findings that teachers lack the content, technical, and pedagogical knowledge needed to feel confident in teaching AI (Kim & Kwon, 2023). Annotations left by past teacher implementations offered a means to help new teachers bridge their knowledge gaps. This dovetails with what Fishman et al. describe as being one of the core principles of Design-Based Implementation Research (DBIR) (Fishman et al., 2013); a sustainable practice whereby past implementations can inform future improvements to the AI curriculum. Through this work, we offer the above insights to help bolster the under-researched area of teacher PD as it relates to K-12 AI education as called out by Dai (Dai, 2023).

Finally, Co-design is a More Empowering Professional Development Modality Over Traditional Formats

Teacher attitudes towards co-design were overwhelmingly positive for a diverse set of reasons. Teacher H commented on how the process of co-design helped teachers develop a sense of ownership and empowerment, going so far as to recommend intentionally leaving gaps within modules to inspire creativity from future teacher implementations. Teacher G, after having implemented the curriculum, expressed a strong desire to further refine the curriculum as a developer, suggesting that the practical knowledge gleaned during her implementation inspired her to want to make improvements using her newly found and grounded insights. Finally, all middle school teachers expressed the current lack of opportunities for collaboration with professors and AI experts prior to engaging in the process of co-design. Like Lin and Van Brummelen's approach to curriculum co-design, we too found that teachers prioritize learner evaluation, engagement, logistics of implementation, and fostering collaboration between students (Lin & Van Brummelen, 2021). Throughout their partnership with us and onwards, teachers expressed feelings of gratitude, humbleness, honor, and camaraderie. Teachers felt recognized as equal team members empowered to pursue how best they saw fit to educate and inspire their students. These sentiments point towards future successes in bridging collaboration gaps between Computer and Learning Scientists and middle school CS educators. We recognize these teachers' contributions towards establishing grounded and high-quality middle school AI curricula standards, both nationally and globally.

Limitations & Future Work

This work is not without its limitations regarding the scalability of our co-design and implementation methodologies, generalizability of assessments, access to produced resources, bias of researchers and participating teachers, and systematic incorporation of student perspectives. We offer our thoughts on how a community-driven online PD platform could effectively address at least some (and certainly not all) of these limitations.

There is an inherent limitation in the scalability of our co-design and implementation process for teachers who were not part of the original co-design process. This limitation also presents an issue of teacher bias in the produced works that have embedded perspectives of both researchers and participating teachers. We acknowledge these limitations fully but posit that an online community-oriented PD platform could reflect more diverse teacher perspectives in shared curriculum resources. This current work makes all co-designed materials publicly available for use and modification by any teacher at ai4ga.org (AI4GA, 2021) and ai4k12.org (AI4K12, 2020). However, these websites in their current form lack a community engagement feature which we found to be vital for teachers during co-design and implementation.

Regarding the scalability and bias of the co-designed resources themselves (slides, activities, and assessments), our findings unsurprisingly found that teachers made modifications to nearly all curriculum materials to meet the unique and diverse needs of their students. Whether by adding or removing examples, creating new activities, or adding/removing slide content or visuals, it is clear that teachers value choice and ownership over their educational tools. Regarding the feasibility and scalability of assessments, those highlighted in this article can be thought of as examples for the types of assessments that educators can leverage for themselves while also providing a framework (i.e., the revised Bloom's taxonomy) for mapping learning objectives to these assessment types. Congruent with our insights, an online PD platform could serve as a "marketplace" for such freedom of choice where teachers can both find resource templates and upload their own adaptations for the benefit of the larger community. Such a platform could support issues of access by implementing a tagging system that affords the organic creation of curricular taxonomy (inclusive of local shorthand terms like “Tickets Out the Door”, “Bell Ringers”, “Voice and Choice”, etc.), which could broaden visibility and adoption among the larger teaching community. We plan to use our insights relating to the feasibility of activities and lesson pacing to inform our continued efforts in establishing implementation best-practices as evidenced by our teacher partners’ expertise and experience.

Not all teachers were able to fully implement the curriculum to completion. These instances were due to the resource capacities of individual teachers, but we fully acknowledge that this issue speaks to a larger need for equitable and sustainable preparation and implementation support and signals to us that further pacing improvements must be made. An online PD platform could afford direct communication support to AI education and researcher experts through a public forum for discussing tactical implementation questions and addressing challenges unique to individual educators.

Conclusion

In this work, we described the format and assessment strategies used within a co-designed AI curriculum structured around *The Five Big Ideas in AI.* It was offered to middle school learners enrolled in an AI elective course across several Georgia middle schools. Though national and academic initiatives have spurred the development of K-12 AI curriculum guidelines, pedagogical support gaps remain for middle school educators needing to make AI relevant and engaging for their students. In response, this work described the co-design and implementation of an AI curriculum for middle school learners. Our analyses of the co-designed curriculum, its assessments, and the learning strategies used by teachers inform what an effective AI curriculum should look and feel like for diverse classrooms. Finally, we reflected on how teachers’ feelings of confidence, camaraderie, and ownership over the AI curriculum improved as a result of co-design. This work contributes effective AI curriculum format and assessment strategies for further elevating the quality of K-12 AI education.

Appendix

Table 3 Module Structure for *Unit 1: Autonomous Robots and Self-Driving Vehicles*

Module	Title	Description	Related Big Idea(s) in AI
1.1	*Course Overview*	What is AI? The Five Big Ideas in AI	The Five Big Ideas in AI
1.2	*Autonomous Robots and Self-Driving Vehicles*	How autonomous robots work	(1) Perception (2) Representation & Reasoning (4) Natural Interaction
1.3	*Anatomy of a Self-Driving Car*	N/A	(1) Perception
1.4	*Robot and Computer Perception*	How face recognition works	(1) Perception (2) Representation & Reasoning
1.5	*Route Finding*	Breadth-first search	(2) Representation & Reasoning
1.6	*Case Study: Sleeping Drivers in Self-Driving Cars*	Case studies for discussion	(4) Natural Interaction (5) Societal Impact
End of Unit 1	*Mini-Project*	Robot design: My Dream Bot	(1) Perception (2) Representation & Reasoning (4) Natural Interaction (5) Societal Impact

Table 4 Module Structure for *Unit 2: How Computers Understand Language*

Module	Title	Description	Related Big Idea(s) in AI
2.1	*Unit Overview: Understanding Language*	Conversing with intelligent assistants	(1) Perception (2) Representation & Reasoning (3) Learning (4) Natural Interaction
2.2	*How Do Computers Understand Language?*	Waveforms, spectrograms, and speech recognition	(1) Perception (2) Representation & Reasoning (3) Learning (4) Natural Interaction
2.3	*How Intelligent Assistants Understand & Answer*	Syntax, semantics, and understanding search queries	(2) Representation & Reasoning (3) Learning (4) Natural Interaction
2.4	*Word Embeddings*	Semantic feature space, word embeddings	(2) Representation & Reasoning (3) Learning
2.5	*How Computers Represent and Generate Meaning*	Machine translation	(2) Representation & Reasoning (3) Learning
2.6	*Sentiment Analysis*	How computers know how people feel	(2) Representation & Reasoning (3) Learning (4) Natural Interaction
2.7	*Chatbots*	Understand the different uses for chatbots and their limitations	(2) Representation & Reasoning (3) Learning (4) Natural Interaction
End of Unit 2	*Mini-Project: Cognimates*	Programming practice in speech recognition, sentiment analysis, and visual pattern detection	(5) Societal Impact

Table 5 Module Structure for *Unit 3: Machine Learning & Automated Decision Making*

Module	Title	Description	Related Big Idea(s) in AI
3.1	*How do computers make decisions?*	Types of reasoning problems with example applications	(2) Representation & Reasoning (3) Learning
3.2	*How can a computer learn to classify objects from examples?*	Decision tree classifiers	(2) Representation & Reasoning
3.3	*How do we train a computer to make decisions?*	Feature spaces and decision tree learning algorithms	(2) Representation & Reasoning (3) Learning
3.4	*Machine Learning with Datasets*	Using machine learning with tabular datasets to classify or predict	(2) Representation & Reasoning (3) Learning
3.5	*Neural Networks*	How neural networks work	(2) Representation & Reasoning (3) Learning
3.6	*Case Study: Does AI Make Better Decisions than People?*	Societal impacts of AI decision making	(2) Representation & Reasoning (3) Learning (5) Societal Impact
End of Unit 3	*Mini-Projects: AI Design*	Using AI model development tools	(2) Representation & Reasoning (3) Learning (5) Societal Impact

Author Contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by David Touretzky, Christina Gardner-McCune, and Judith Uchidiuno. Data analysis was performed by William Gelder and Xueru Yu. The first draft of the manuscript was written by William Gelder and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was funded by the National Science Foundation, under awards DRL-2049029 and DRL-2048502. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

References

AI Campus. (2024). *Frontpage*. https://ki-campus.org/

AI4GA (2021). *Artificial Intelligence For Georgia*. https://ai4ga.org/

AI4K12. (2020). *AI4K12*. https://ai4k12.org/

Ali, S., Williams, R., Payne B., Park H., Breazeal C. (2019) *Constructionism, Ethics, and Creativity: Developing Primary and Middle School Artificial Intelligence Education*. Presented in the International Workshop on Education in Artificial Intelligence K-12 (EDUAI '19). Palo Alto, CA, USA. Proceedings of IJCAI 2019.

ATLAS.ti. (2023). *The #1 software for qualitative data analysis*. https://atlasti.com.

Autodesk. (2023). *From mind to design in minutes*. Tinkercad. https://www.tinkercad.com/

Barab, S. (2022). Design-Based Research: A Methodological Toolkit for Engineering Change. In R. Sawyer (Ed.), The Cambridge Handbook of the Learning Sciences (Cambridge Handbooks in Psychology, pp. 177–195). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108888295.012

Bishop, J., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. *Paper presented at 2013 ASEE Annual Conference & Exposition*. https://doi.org/10.18260/1-2--22585.

Black, N., Brooks-Young, S., Amalong, Poth, R.D., Puri, A., South, J., Aiwohi, L., Lowenstein, J., Ramos, Y., Waite, V., & Woodall, C. (2023). *Hands-On AI Projects for the Classroom: A Guide for Electives Teachers*. ISTE. General Motors. https://cms-live-media.iste.org/www-root/Libraries/Documents%20%26%20Files/Artificial%20Intelligence/AIGDEL_0820-red.pdf

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives Handbook 1: Cognitive domain*. Longman Group Ltd.

Chakaravarthy, E (2023). *The tiktok AI algorithm that got millions of users hooked*. The TikTok AI Algorithm that got Millions of Users Hooked. https://www.argoid.ai/blog/the-ai-algorithm-that-got-tiktok-users-hooked

Chang, W.-C., & Chung, M.-S. (2009). Automatic applying Bloom's taxonomy to classify and analysis the cognition level of English question items. *Joint Conferences on Pervasive Computing (JCPC), 2009*, 727–734. https://doi.org/10.1109/JCPC.2009.5420087

Coenraad, M., Palmer, J., Eatinger, D., Weintrop, D., & Franklin, D. (2022). Using participatory design to integrate stakeholder voices in the creation of a culturally relevant computing curriculum. *International Journal of Child-Computer Interaction, 31*, 100353. https://doi.org/10.1016/j.ijcci.2021.100353

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology, 13*, 3–21. https://doi.org/10.1007/BF00988593

Dai, Y. (2023). Negotiation of Epistemological Understandings and Teaching Practices Between Primary Teachers and Scientists about Artificial Intelligence in Professional Development. *Research in Science Education, 53*, 577–591. https://doi.org/10.1007/s11165-022-10072-8

Pear Deck. (2023). Homepage: Pear Deck. Homepage | Pear Deck. https://www.peardeck.com/

Dovetail. (2023). *Customer insights hub*. https://dovetail.com/

Boston Dynamics. (2019). *Evolution Of Boston Dynamics Since 2012*. Retrieved November 23, 2023, from https://www.youtube.com/watch?v=NR32ULxbjYc

Edison Research. (2022). *Smart audio report 2022 from NPR and Edison Research*. https://www.edisonresearch.com/smart-audio-report-2022-from-npr-and-edison-research/

Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-Based Implementation Research: An Emerging Model for Transforming the Relationship of Research and Practice. *Teachers College Record, 115*(14), 136–156. https://doi.org/10.1177/016146811311501415

Gardner-McCune, C., Touretzky, D., Cox, B., Uchidiuno, J., Jimenez, Y., Bentley, B., Hanna, W., and Jones, A. 2023. Co-Designing an AI Curriculum with University Researchers and Middle School Teachers. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2 (SIGCSE 2023). Association for Computing Machinery, New York, NY, USA, 1306. https://doi.org/10.1145/3545947.3576253

Geyser, W. (2023). *How does the YouTube algorithm work: A peek into its changes in 2024*. How Does the YouTube Algorithm Work: A Peek into its Changes in 2024. https://influencermarketinghub.com/how-does-the-youtube-algorithm-work/#:~:text=Like%20Netflix%2C%20YouTube%20uses%20AI,account%20is%20currently%20logged%20in

Grover, S. (2017). *Assessing Algorithmic and Computational Thinking in K-12: Lessons from a Middle School Classroom*. In: Rich, P., Hodges, C. (eds) Emerging Research, Practice, and Policy on Computational Thinking. Educational Communications and Technology: Issues and Innovations. Springer, Cham. https://doi.org/10.1007/978-3-319-52691-1_17

Grover, S. (2021). *Toward A Framework for Formative Assessment of Conceptual Learning in K-12 Computer Science Classrooms*. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21). Association for Computing Machinery, New York, NY, USA, 31–37. https://doi.org/10.1145/3408877.3432460

Grover, S. (2024). Teaching ai to K-12 learners: Lessons, issues, and guidance. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*. https://doi.org/10.1145/3626252.3630937

Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) model: A critical review and suggestions for its use. *TechTrends, 60*, 433–441.

Hero Forge custom miniatures. (2024). https://www.heroforge.com/

Kim, K., & Kwon, K. (2023). Exploring the AI competencies of elementary school teachers in South Korea. *Computers and Education: Artificial Intelligence, 4*, 100137.

Klassner, F. 2002. *A case study of LEGO Mindstorms'™ suitability for artificial intelligence and robotics courses at the college level*. In Proceedings of the 33rd SIGCSE technical symposium on Computer science education (SIGCSE '02). Association for Computing Machinery, New York, NY, USA, 8–12. https://doi.org/10.1145/563340.563345

Koehler, M. J., Mishra, P., Kereluik, K., Shin, T. S., & Graham, C. R. (2014). The technological pedagogical content knowledge framework. *Handbook of research on educational communications and technology*, 101–111.

Kumar D., & Meeden, L. 1998. *A robot laboratory for teaching artificial intelligence*. In Proceedings of the twenty-ninth SIGCSE technical symposium on Computer science education (SIGCSE '98). Association for Computing Machinery, New York, NY, USA, 341–344. https://doi.org/10.1145/273133.274326

Lah, N. C., Saat, Â. R. M., & Hassan, Â. R. (2018). Cognitive strategy in learning chemistry: How chunking and learning get together. *MOJES: Malaysian Online Journal of Educational Sciences, 2*(1), 9–15.

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence, 3*, 100101. https://doi.org/10.1016/j.caeai.2022.100101

Lee, I., Ali, S., Zhang, H., DiPaola, D., & Breazeal, C. (2021). Developing middle school students' ai literacy. Proceedings of the 52nd ACM Technical Symposium on Computer Science Education. https://doi.org/10.1145/3408877.3432513

Lin, C.-H., Yu, C.-C., Shih, P.-K., & Wu, L. Y. (2021). STEM based Artificial Intelligence Learning in General Education for Non-Engineering Undergraduate Students. *Educational Technology & Society, 24*(3), 224–237. https://www.jstor.org/stable/27032867.

Lin, P., & Van Brummelen, J. (2021). Engaging teachers to co-design integrated AI curriculum for K-12 Classrooms. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445377

Long, D., Magerko, B. 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, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376727

Martonosi, M., & Marrongelle, K. (2020). *Dear colleague letter: Advancing Educational Innovations that motivate and prepare prek-12 learners for computationally-intensive industries of the future*. NSF. https://www.nsf.gov/pubs/2020/nsf20101/nsf20101.jsp

Mota-Valtierra, G., Rodríguez-Reséndiz, J., & Herrera-Ruiz, G. (2019). Constructivism-Based Methodology for Teaching Artificial Intelligence Topics Focused on Sustainable Development. *Sustainability., 11*(17), 4642. https://doi.org/10.3390/su11174642

Muller, M., & Kogan, S. (2010). *Grounded Theory Method in HCI and CSCW*.

Murphy, R. R. (2001). Introduction to AI Robotics. *Industrial Robot, 28*(3), 266–267. https://doi.org/10.1108/ir.2001.28.3.266.1

National Highway Traffic Safety Administration (NHTS). (2021). *Driver assistance technologies*. NHTSA. https://www.nhtsa.gov/equipment/driver-assistance-technologies

Nearpod. (2023). *You'll wonder how you taught without it*. https://nearpod.com/

Nearpod. (2023a). *Time to climb: Game-Based Learning*. https://nearpod.com/time-to-climb

Ng, D. T. K., Lee, M., Tan, R. J. Y., et al. (2023). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies, 28*, 8445–8501. https://doi.org/10.1007/s10639-022-11491-w

Ng, D.T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). *Conceptualizing AI literacy: An exploratory review*. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Payne, B. H. (2020). *Can my algorithm be my opinion? : an AI + ethics curriculum for middle school students.* Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences. Massachusetts Institute of Technology.

Puentedura, R. (2006). *Transformation, technology, and education*. In: Presentation given August 18, 2006 as part of the Strengthening Your District Through Technology workshops, Maine, US.

Puentedura, R. (2014). Building transformation: An introduction to the SAMR model [Blog post]. Retrieved from http://www.hippasus. 440 TechTrends (2016) 60:433–441 com/rrpweblog/archives/2014/08/22/BuildingTransformation_ AnIntroductionToSAMR.pdf.

European Commission, Joint Research Centre, Punie, Y., Pujol Priego, L., Carretero, S. (2018). *DigComp into action, get inspired make it happen: a user guide to the European Digital Competence framework*. (Y.Punie, editor, S. Carretero, editor, R. Vuorikari, editor) Publications Office. https://data.europa.eu/doi/10.2760/112945

Quizizz. (2023). *Free online quizzes, lessons, activities and homework*. https://quizizz.com/

Rideout, V., Peebles, A., Mann, S., & Robb, M. B. (2022). *Common Sense census: Media use by tweens and teens, 2021*. Common Sense.

Rodríguez-García, J. D., Moreno-León, J., Román-González, M., & Robles, G. 2021. *Introducing Artificial Intelligence Fundamentals with LearningML: Artificial Intelligence made easy*. In Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'20). Association for Computing Machinery, New York, NY, USA, 18–20. https://doi.org/10.1145/3434780.3436705

Russell, S., & Norvig, P. (1995). *A modern, agent-oriented approach to introductory artificial intelligence*. SIGART Bull. 6, 2 (April 1995), 24–26. https://doi.org/10.1145/201977.201989

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. *How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction*. Computers in Human Behavior. Volume 69, 2017. Pages 371–380. ISSN 0747–5632. https://doi.org/10.1016/j.chb.2016.12.033

Severance, S., Penuel, W. R., Sumner, T., & Leary, H. (2018). Organizing for teacher agency in curricular co-design. In *Cultural-historical activity theory approaches to design-based research* (pp. 45–78). Routledge.

Shih, P.-K., Lin, C.-H., Wu, L. Y., & Yu, C.-C. (2021). Learning Ethics in AI—Teaching Non-Engineering Undergraduates through Situated Learning. *Sustainability., 13*(7), 3718. https://doi.org/10.3390/su13073718

Slido. (2023). *Audience interaction made easy*. Slido. https://www.slido.com/

Snapchat. (2023). *SPS 2023: What's next for my ai. Snap Newsroom*. https://newsroom.snap.com/sps-2023-whats-next-for-my-ai

Steen, M. (2013). Co-Design as a process of joint inquiry and Imagination. *Design Issues, 29*(2), 16–28. https://doi.org/10.1162/desi_a_00207

Smith, J. J., Payne, B. H., Klassen, S., Doyle, D. T., & Fiesler, C. (2023). Incorporating ethics in computing courses: Barriers, support, and perspectives from educators. In Proceedings of the 54th ACM technical symposium on computer science education Vol. 1 (pp. 367–373).

The Findings Group. (2023). The Findings Group, LLC. https://thefindingsgroup.org/

Thompson, E., Luxton-Reilly, A., Whalley, J. L., Hu, M., & Robbins, P. (2008). *Bloom's taxonomy for CS assessment*. In Proceedings of the tenth conference on Australasian computing education-Volume 78 (pp. 155–161).

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What Should Every Child Know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(01), 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795

UNESCO (United Nations Educational, Scientific and Cultural Organization). (2019). *Artificial Intelligence for Sustainable Development: Program*. UNESCO. Paris, France. 4–8 March 2019.

UNESCO (United Nations Educational, Scientific and Cultural Organization). (2022). *K-12 AI curricula: a mapping of government-endorsed AI curricula*. Paris: UNESCO. ED-2022/FLI-ICT/K-12 REV.

University of Helsinki. (2024). *A free online introduction to Artificial Intelligence for non-experts*. Elements of AI. https://www.elementsofai.com/

Vazhayil, A., Shetty, R., Bhavani, R. R., & Akshay, N.. *Focusing on Teacher Education to Introduce AI in Schools: Perspectives and Illustrative Findings*. 2019 IEEE Tenth International Conference on Technology for Education (T4E). Goa, India. 2019. pp. 71–77. https://doi.org/10.1109/T4E.2019.00021.

Vuorikari, R., Kluzer, S., & Punie, Y., *DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes*, EUR 31006 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978–92–76–48882–8, https://doi.org/10.2760/115376, JRC128415.

Williams, R., Kaputsos, S.P., & Breazeal, C. *Teacher Perspectives on How to Train Your Robot, A Middle School AI and Ethics Curriculum*. Proceedings of the 10th Symposium on Education Advances in Artificial Intelligence (EAAI '21). 2021.

Xu, J. J., & Babaian, T. *Artificial intelligence in business curriculum: The pedagogy and learning outcomes*. The International Journal of Management Education. Volume 19, Issue 3. 2021. 100550. ISSN 1472–8117. https://doi.org/10.1016/j.ijme.2021.100550

Yi, Y. (2021). Establishing the concept of AI literacy Focusing on competence and purpose. *Jahr European Journal of Bioethics, 12*(2), 353–368. https://doi.org/10.21860/j.12.2.8