Hands-on, Interactive AI Ethics Education in a STEM Summer Program: Fostering Ethical Awareness in High School Students (Evaluation)

Labiba Rahman, Purdue University – West Lafayette (College of Engineering)

Labiba Rahman is a Ph.D. student in the School of Engineering Education at Purdue University. Her research aims to promote equitable participation in STEM by examining how students learn with and about artificial intelligence in educational settings. Her work focuses on AI ethics and explainable AI (XAI), leveraging learning technologies and curriculum interventions to foster students' ethical reasoning, transparency awareness, and sense of belonging in AI-supported learning environments.

Dr. Morgan M Hynes, Purdue University – West Lafayette (College of Engineering)

Dr. Morgan Hynes is an Assistant Professor in the School of Engineering Education at Purdue University and Director of the FACE Lab research group at Purdue. In his research, Hynes explores the use of engineering to integrate academic subjects in K-12 cla

Title: Hands-on, Interactive AI Ethics Education in a STEM Summer Program: Fostering Ethical Awareness in High School Students (Evaluation)

Hands-on, Interactive AI Ethics Education in a STEM Summer Program: Fostering Ethical Awareness in High School Students (Evaluation)

Abstract

The rapid adoption of artificial intelligence (AI) in education has raised significant ethical concerns, particularly around fairness, accountability, transparency, and ethics (FATE) [1], [2]. Although AI tools may offer many educational benefits, their widespread integration into everyday life often occurs without students being aware that they are interacting with algorithms which might infringe on privacy, reinforce biases, or make opaque decisions [3]. As AI is being more integrated into education systems, it is time to examine how all these AI- driven educational tools and technologies can be developed and used in a manner that aligns with ethical standards, ensuring fairness and accountability.

In this study, we aim to understand how high school students' knowledge and perceptions of artificial intelligence evolve after participating in an AI and ethics module within the XYZ program, a two-week hands-on program for rising juniors and seniors that provides STEM activities, faculty mentorship, and team-based projects. During the module, students engaged in interactive lessons and activities, utilizing web-based tools like Google Teachable Machine [4] and Moral Machine [5] to explore more about how biases work, and analyzed real-world scenarios through Ethical Dilemma Cards [6]. To evaluate the impact, we collected pre- and post-survey data and written group responses of ethical dilemma activities from 80 (67 consented) high school students, which will be analyzed using inductive thematic analysis and descriptive statistics to examine changes in students' ethical understanding of AI. This study helps determine the effectiveness of hands-on learning techniques in supporting high school students' understanding of AI, ethical awareness, and critical engagement with AI technologies.

The findings suggest that students' understanding of AI and engagement with related activities improved after the workshop. Many students became more comfortable using new technologies and experimenting with ways to interact with AI. The workshop also appeared to increase students' awareness of ethical considerations, including concerns about bias, misinformation, privacy, transparency, and the limitations of AI. Moreover, during ethical dilemma activities, some students were able to recognize potential biases in AI and considered that developers and organizations have a role in addressing these challenges.

These findings support the need of including inquiry-based, hands-on learning into early AI education. The study supports the use of ethical dilemma activities to promote civic engagement, critical thinking, and an understanding of fairness in technology design. It also shows that high school students are capable of meaningful, critical ethical reasoning. This work provides educators with a workable paradigm for preparing the next generation to navigate an AI-shaped future in an ethical manner.

1. Introduction

The application of Artificial Intelligence (AI) is rapidly expanding in the field of education [7], [8]. AI broadly refers to the development of machines capable of performing tasks that typically require human intelligence [9]. As this technology becomes increasingly embedded in educational tools and systems, ethical considerations grow more critical. AI ethics, defined as the study of moral principles that guide the design, implementation, and use of AI systems [10], requires a thorough examination of the societal impacts and potential unintended consequences of these technologies [11].

Despite AI's widespread integration into everyday life, many individuals remain unaware that they are interacting with algorithms that may infringe on privacy, perpetuate harmful biases, or make decisions in ways that are not transparent or understandable to users [3]. This combination of public unawareness and technological opacity, where internal decision-making processes are difficult for users to see or interpret, raises serious concerns about people's ability to engage with AI systems critically and safely [12]. As AI increasingly influences high-stakes domains – such as healthcare, employment, transportation, and criminal justice, addressing issues of ethics and fairness becomes imperative.

A central framework guiding discussions in this space is FATE—Fairness, Accountability, Transparency, and Ethics, which has gained considerable traction in both AI and education research [1], [2]. Among these principles, fairness stands as a foundational pillar, underscoring the need for AI systems to operate free from bias or discrimination [10]. However, efforts to enhance fairness often involve trade-offs with model performance, necessitating a careful and context-sensitive approach [11].

1.1 Prior Work and Gap

Prior studies on AI literacy suggest that with the help of structured, project-based learning environments, youth can interact meaningfully with concepts like bias, fairness, and transparency. For instance, scaffolded ethical discussions and hands-on AI projects have been shown to help middle school students develop more awareness of algorithmic bias. In similar vein, prior studies [12] emphasize the role of experiential, inquiry-driven activities in helping students comprehend the limitations and ethical implications of AI systems. Examples of these activities include interacting with machine learning tools and analyzing real-world scenarios. Related work in science and technology education further indicates that scenario-based learning and ethical dilemma discussions can support students' moral reasoning and civic engagement when addressing complex, socio-technical issues.

Despite these contributions, existing research remains limited in several ways. Much of the empirical work on AI ethics education has been conducted in formal classroom settings or with middle school populations, with less attention to high school learners. In addition, few studies have examined immersive, informal learning environments such as STEM summer programs, where students engage in extended, hands-on exploration of emerging technologies. There is also limited empirical evidence on how pedagogical approaches such as role-playing, ethical dilemma activities, and collaborative scenario analysis shape adolescents' ethical reasoning about AI.

These limitations highlight the need for empirical research examining how high school students develop ethical reasoning about AI in immersive, hands-on learning environments beyond traditional classroom instruction.

1.2 Purpose of Study

This evaluation explores how high school students engaged with and responded to artificial intelligence and ethics module through interactive, experiential learning methods. The study was conducted as part of the XYZ summer program, a two-week immersive STEM experience designed for rising high school juniors and seniors. The XYZ program aims to broaden students' exposure to emerging technologies, including semiconductors, microelectronics, and artificial intelligence, through hands-on activities, interdisciplinary learning, and collaborative projects, while fostering awareness of the social and technical dimensions of technology development. Prior evaluations of the XYZ program have demonstrated its effectiveness in promoting student engagement and experiential learning in informal STEM settings [13].

Building on this foundation, the present study explores whether instructional approaches such as scenario-based learning, role-playing, and real-world case analysis can enhance students' ethical reasoning and awareness of AI. Pre- and post-survey data were analyzed using inductive thematic analysis and descriptive statistical techniques to examine changes in students' knowledge, perceptions, and ethical decision-making related to AI. The findings aim to inform inclusive and impactful educational practices for integrating AI ethics into informal and experiential learning environments. This evaluation was guided by the following research questions:

RQ1: *How do high school students' knowledge and perceptions of AI change after participating in an AI and ethics module that includes lessons, interactive activities, and web-based tools?*

RQ1a: *What knowledge and perceptions of AI do students bring before the module prior to participation?*

RQ1b: *What understandings of AI and AI ethics do students demonstrate following participation in the module?*

2. Research Design

This study employed a qualitative research design to explore and identify active learning and experiential strategies that can support high school students' understanding of artificial intelligence (AI), its ethical dimensions, and principles of fairness. The AI and ethics component was conducted as part of the XYZ summer program. The focus was to assess changes in students' knowledge, attitudes, and ethical reasoning through interactive, hands-on learning. Data collected for this study included pre- and post-surveys, along with written group responses from ethical dilemma activities.

Additionally, this study was conducted as a part of the XYZ summer program and received Institutional Review Board (IRB) approval from both XY University and YZ Community College. Prior to the data collection, both parental and student consent was obtained.

2.2 Participant Demographics

A total of 80 high school students participated in the XYZ program, of whom 67 students consented to participate in this research study. All participants were rising juniors or seniors entering the 2025-2026 academic year, with ages ranging from 16 to 18 years.

Students were recruited through a competitive application process and were required to reside in and/or attend high school in one of six Midwest counties: B, C, D, E, F, or G. Eligibility criteria included being a rising junior or senior, not having previously participated in the XYZ program, and completing an application that included demographic information and short essay responses focused on students' interests, motivations, and prior accomplishments. Prior experience with artificial intelligence, semiconductors, or ethics was not required, allowing students with diverse backgrounds to participate.

Participants offered a variety of educational and geographic perspectives, representing multiple high schools in both rural and suburban communities. In addition, the program emphasized collaborative, team-based learning and provided students with exposure to foundational concepts in semiconductors, AI, and ethical decision-making. When taken as a whole, these characteristics place the study within an immersive STEM summer program designed to support diverse learners and broaden involvement in emerging technology fields.

2.3 Learning Activities and Tools

Students participated in a one-hour instructional session that served as part of the XYZ summer program. The session was facilitated by the lead researcher, with support from two undergraduate teaching assistants and an additional graduate facilitator who assisted with guiding activities and addressing student questions. The program consisted of three sections of students (Section A, n = 26; Section B, n = 27; Section C, n = 27) that rotated through program activities. Instruction was delivered to each section, and students were further organized into smaller working groups of four to six students. All three sections received the same instructional content and sequence, and the learning activities were conducted sequentially without overlap.

Before beginning the interactive components, students received a brief introduction to foundational AI concepts, including what AI is, how it makes decisions, how machine learning models are trained, where bias can emerge, and why fairness and ethics are critical considerations in AI development. The learning objectives for this lesson were to (a) build basic understanding of machine learning processes, (b) recognize how data imbalance or model design choices can introduce bias, and (c) engage students in ethical reasoning through real-world scenarios.

2.4 Active Learning Approach

In the AI and Ethics module of the XYZ program, active learning was the foundational instructional strategy. Active learning is an instructional approach in which students play a central role by engaging in activities and processing information through discussion and reflection. This method builds on students' prior knowledge, making it particularly effective for introducing new and complex topics such as AI ethics [14], [15]

In this study, students learned by doing, interacting directly with AI tools, analyzing ethical scenarios, and engaging in peer discussions. Activities were completed in small groups to encourage collaboration, idea negotiation, and collective ethical reasoning. Each activity concluded with a class-wide discussion to deepen understanding and refine perspectives.

Students then participated in two primary learning activities:

1. **Google Teachable Machine (Individual Activity):**

Students individually used the web-based tool called Google Teachable Machine to create and train simple machine learning models using images, sounds, or poses. The purpose of this activity was to give students a hands-on understanding of how machine learning models are trained and how their behavior is shaped by the data they receive. Through experimentation, students observed how models responded differently when trained on limited, imbalanced, or inconsistently labeled data. These interactions supported discussions of how bias can arise from uneven data collection, labeling errors, or model interpretation, as well as how models may behave in unexpected or inaccurate ways when trained on insufficient datasets. Figure 1 illustrates an example used during the activity to demonstrate how model predictions can vary based on the quality and balance of training data. Here, the model was trained on two image classes ("Pen" and "Pencil") using a limited and uneven number of training images. When tested with a live webcam input, the model predicted the object as a *pen* with 75% confidence and *pencil* with 25% confidence.

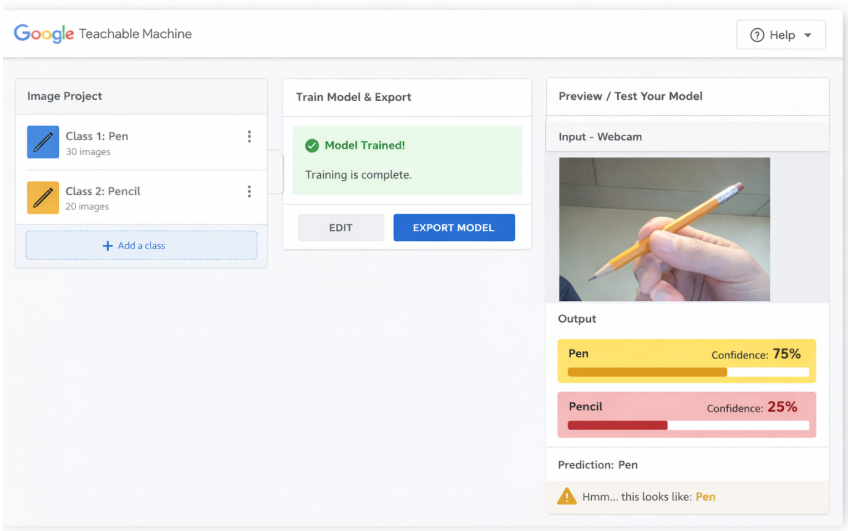

Figure 1: Example of a student-trained image classification model using Google Teachable Machine

2. **Ethical Dilemma Cards (Group Activity):**

Working in their small groups, students analyzed real-world AI scenarios, including examples such as facial recognition in schools, biased hiring algorithms, and the use of surveillance technologies. Groups discussed the potential benefits and risks of each scenario, recorded a collective decision, and documented the reasoning behind their conclusions. Each group then presented their perspectives to the class, followed by a whole-group discussion. The purpose of this activity was to help students see that real-world ethical dilemmas often require considering multiple factors before making a decision.
Both activities were intentionally designed to promote engagement through creativity, inquiry, and moral reflection, while providing opportunities for students to apply ethical concepts to practical examples of AI systems.

2.5 Data Collection and Instruments

Two primary sources of data were collected to examine changes in students' knowledge, perceptions, and ethical reasoning related to AI.

Pre- and post-surveys were administered before and after the instructional session. The surveys were developed by the lead facilitator and included a combination of Likert-scale items, multiple-choice questions, and open-ended prompts designed to assess students' familiarity with AI, comfort using AI tools, confidence in problem-solving, interest in STEM fields, and awareness of ethical considerations related to AI. Likert-scale items used a seven-point scale, with anchors such as *not at all familiar* to *very familiar*, *not comfortable* to *very comfortable*, and *very unlikely* to *very likely*. The pre-survey included 6 Likert-scale items, 6 categorical items, 4 open-ended questions, and 2 multiple-choice items, while the post-survey included 13 Likert-scale items, 6 open-ended questions, and 1 multiple-choice item. Sample survey items are provided in Appendix A.

In addition, written group responses from the Ethical Dilemma Cards activity were collected as qualitative data. Eight distinct ethical dilemma prompts were developed, and each small group was assigned one prompt. These scenarios addressed real-world applications of AI, such as school surveillance systems and smart home devices that collect user data. Students responded to guiding questions related to fairness, privacy, consent, accountability, and societal impact. Sample ethical dilemma prompts are included in Appendix B.

2.6 Data Analysis

Quantitative data from the pre- and post-surveys were analyzed using descriptive statistics, including means, percentages, and observed shifts in responses, to examine changes in students' self-reported knowledge, confidence, and perceptions of AI following participation in the module.

Qualitative data from open-ended survey responses and written group submissions were analyzed using inductive thematic analysis. The lead researcher conducted an initial open coding process by reviewing all responses and identifying recurring patterns and ideas emerging directly from the data.

Codes were iteratively refined and categorized into broader themes related to ethical awareness, perceptions of bias, privacy concerns, human oversight, and responsibility in AI systems. In addition, themes specific to students' ethical reasoning during the dilemma-based activities - such as cautious attitudes toward AI deployment and emphasis on human oversight - were identified by comparing patterns across group responses and scenarios. To enhance trustworthiness, a reliability check was conducted through a secondary review of coded data, with discrepancies discussed and resolved. Representative student quotes were selected to illustrate key themes in the findings.

3. Findings and Results

3.1 Students' Understanding and Perceptions of AI

This subsection addresses RQ1a by examining students' baseline knowledge, perceptions, and attitudes toward AI prior to participation in the instructional module.

Pre-Program Insights

Pre-Survey Analysis: Students' Perceptions of AI and Robotics

Prior Experience and Access

Most students had substantial prior exposure to STEM-related activities and digital technologies. Specifically, 52% of participants reported previous participation in a STEM club or program, while 83% indicated having access to a computer or tablet at home for learning or coding. Additionally, 66% of students reported having written or edited computer code, 85% had interacted with a robot, and an even higher 92% had used an AI chatbot.

Students most commonly reported using AI chatbots for fun or random questions **(**31%**)** and for school assignments (31%), reflecting both recreational and academic uses. This baseline exposure suggests that participants entered the program with a foundational level of technological familiarity and an emerging curiosity toward AI.

Familiarity, Interest, and Confidence

Across Likert-scale measures, students reported relatively high levels of familiarity with AI technologies, interest in STEM fields, and confidence in learning new technologies.

Table 1: *Mean pre-survey ratings of students' familiarity with AI, interest in STEM, and confidence in learning new technologies (7-point Likert scale):*

Construct	***Mean (out of 7)***
Familiarity with AI/chatbots	**5.05**
Interest in STEM fields	**5.40**
Confidence in learning new technologies	**5.39**

Interest in learning about robotics and AI	**5.59**

Qualitative responses reinforced these findings. Many students described AI as *"a tool that helps with learning," "something exciting about the future,"* and *"part of everything now."* Overall, the tone of these responses reflects optimism and eagerness to engage with emerging technologies.

Problem-Solving Approaches

To better understand students' baseline perceptions of AI as a learning and problem-solving resource (RQ1a), we examined how students reported responding to challenges prior to the instructional module. Students reported using a range of adaptive strategies when encountering challenges. A majority (87%) indicated that they had experienced being stuck on a task at some point. When this occurred, students most frequently reported the following responses:

- Asking a teacher or adult (19 mentions)
- Using AI for help (16 mentions)
- Asking a friend (14 mentions)
- Trying different solutions independently (13 mentions)
- Searching online (11 mentions)

Quantitatively, students also reported high comfort levels in seeking help from AI tools with a mean comfort rating of 5.51 (out of 7). These findings position AI as one of several human-supported and self-directed problem-solving strategies and provide insight into students' pre-existing perceptions of AI as a valid source of assistance. Together, these results help establish a baseline for examining how students' perceptions and use of AI evolve following participation in the instructional module. Changes in these problem-solving approaches are revisited in Section 3.3.

Career Aspirations

To further contextualize students' baseline perceptions of AI and its perceived relevance to their futures (RQ1a), we examined students' interest in STEM- and AI-related career pathways prior to the instructional module. Students expressed interest in pursuing future STEM-related careers, with a mean rating of 5.03 (out of 7) for interest in technology, robotics, or AI-related fields. The most frequent responses were 5, 6, and 7 indicating that more than two-thirds of participants envisioned a future involving STEM fields.

Qualitative reflections frequently referenced innovation, creativity, and future opportunities, suggesting that students already associate AI with personal growth and long-term potential. These responses provide insight into how students position AI not only as a tool for learning, but also as a meaningful component of their future academic and career aspirations.

Perceptions of AI's Role in Society

Open-ended responses revealed a nuanced understanding of both the benefits and risks associated with AI. Thematic analysis of qualitative responses identified eleven dominant themes, summarized below:

Students' open-ended responses reflect a balanced awareness of AI's benefits and risks. Thematic analysis of all qualitative questions produced eleven dominant themes:

Table 2: *Thematic analysis of students' pre-survey open-ended responses describing perceptions of AI, including theme frequency and representative student expressions:*

Theme	Frequency	Example Student Expression
AI as a Helpful Tool	**24**	"AI helps me learn faster."
Future & Evolution	**14**	"It's the future of our world."
Over-Reliance and Laziness	**11**	"People will stop thinking for themselves."
Job Displacement	**9**	"AI might take people's jobs."
Ethical & Moral Concerns	**8**	"AI doesn't know right from wrong."
Misinformation & Deepfakes	**6**	"It could spread fake information."
Control & Autonomy Risks	**5**	"AI could get out of control."
Privacy & Surveillance	**4**	"AI might track everything we do."
Environmental Impact	**3**	"Training AI uses a lot of energy."
Robots & Automation	**3**	"Self-driving cars and factory robots."
Specific Tools/Companies	**3**	"ChatGPT, Tesla robots, Boston Dynamics."

Overall, students recognized the dual nature of AI as both beneficial and potentially harmful. While many viewed AI as an essential and evolving part of modern life, they also expressed concerns about over-dependence, ethical misuse, and loss of human judgment.

Representative student quotes include:

- *"AI helps make our work easier, but we should not rely on it for everything."*
- *"It's good that AI helps with school, but people might get lazy."*
- *"AI is powerful, but it can't make moral choices."*

Overall Interpretation

These pre-survey findings indicate that students entered the program with strong prior exposure to AI and related technologies, high levels of curiosity, and generally positive attitudes toward AI and robotics. Students reported comfort using AI for learning and problem-solving while also demonstrating early awareness of ethical concerns, including bias, misinformation, and over-reliance.

These findings suggest that early engagement in STEM and robotics programs can develop not only technical familiarity but also the fundamentals of ethical awareness and critical digital literacy regarding the role of AI in society.

3.2 Post-Program Insights: Students' Understanding and Perceptions of AI

This subsection addresses RQ1b by examining students' knowledge, perceptions, and ethical reasoning related to AI following participation in the AI and ethics module.

Engagement with AI Tools and Learning Strategies

Following participation in the AI and ethics module, students reported actively using AI tools for a range of learning-related purposes. When asked how they mostly used AI throughout the activities, students most frequently said that they used AI to generate code, get ideas, and comprehend technical concepts. This suggests that students viewed AI as a flexible learning aid that supported both creative exploration and conceptual understanding. In addition, many students reported engaging in iterative interactions with AI responses, suggesting an emerging understanding of how user input influences AI outputs.

Table 3: *Post-survey means ratings of students' AI engagement and problem-solving confidence (7 – point Likert scale):*

Construct	Mean (out of 7)
Frequency of refining AI prompts	**5.08**
Confidence using AI for problem-solving	**5.14**

These responses suggest that students approached AI as an interactive tool rather than a passive source of answers.

Representative student quotes include:

- *"I realized that changing how you ask the question makes a big difference in the answer."*
- *"You can't just trust the first response - you have to think about it."*

Confidence in Learning and Using Technology

Post-survey responses indicate that students felt increasingly comfortable learning new technologies following the module. Students' self-reported confidence in learning new technology or tools was moderate to high levels ($M \approx 5.53$), reflecting a generally positive perception of their ability to adapt to unfamiliar technological systems.

Similarly, students reported increased confidence in using AI for problem-solving tasks ($M \approx 5.14$). Open-ended responses suggest that hands-on interaction with AI tools contributed to this confidence, as students described becoming more familiar with AI capabilities and limitations through direct experimentation.

Representative student quotes include:

- *"Trying it myself made it less intimidating."*
- *"I feel more confident learning new tools now because I know I can figure them out."*

Interest in AI and Robotics

Students expressed sustained interest in AI and robotics following the instructional session. Ratings for likelihood of exploring AI or robotics further were moderate to high levels ($M \approx 5.26$), indicating continued curiosity and motivation to engage with AI-related topics beyond the program. These responses suggest that participation in the module supported students' interest in continued exploration of emerging technologies.

Ethical Awareness and Reflection

Post-survey qualitative responses reveal increased attention to ethical dimensions of AI. Students frequently referenced concerns related to privacy, bias, misinformation, over-reliance on AI systems, and the importance of human oversight. Rather than viewing AI as inherently good or bad, many students articulated conditional perspectives emphasizing responsible use.

Table 4: *Thematic analysis of students' post-survey ethical themes identified from students' open-ended responses:*

Ethical Theme	Frequency	Example Student Expression
Bias and Fairness	**18**	"AI can be unfair depending on the data."
Privacy and surveillance	**15**	"It can collect too much personal information."
Over-reliance on AI	**13**	"People might depend on it too much."
Misinformation	**10**	"AI can give wrong or misleading answers."
Human oversight	**9**	"Humans should still make final decisions."

Representative student quotes include:

- *"AI is helpful, but it shouldn't replace human judgment."*
- *"We need rules, so AI doesn't hurt certain groups."*
- *"Just because AI says something doesn't mean it's right."*

Overall Interpretation

Post-program findings suggest that students demonstrated:

- Intentional engagement with AI tools and learning strategies
- Increased confidence in learning and using new technologies
- Sustained interest in AI and robotics

- Developing ethical awareness related to bias, privacy, and responsible AI use

These findings provide a foundation for examining how students' knowledge, perceptions, and ethical reasoning about AI shifted from pre- to post-participation in the module.

3.3 Changes in Students' Knowledge, Perceptions, and Ethical Reasoning

This subsection addresses RQ1 by examining changes in students' knowledge, perceptions, and ethical reasoning related to AI from pre- to post-participation in the AI and ethics module.

Shifts in Knowledge, Confidence, and Engagement with AI

Comparisons between pre- and post-survey responses indicate observable patterns in how students engaged with and perceived AI tools over the course of the module. While many students entered the program with prior exposure to AI and generally positive attitudes, post-survey responses reflect greater intentionality in how students interacted with AI systems. For instance, students reported more frequent refinement of AI prompts following the module, suggesting increased awareness of how user input shapes AI-generated responses. This pattern suggests a shift from more casual or exploratory use of AI toward more deliberate and reflective engagement.

Students' post-survey responses also reflected higher self-reported confidence related to using AI for problem-solving and learning new technologies. Although confidence levels were already moderate to high prior to participation, descriptive comparisons suggest a strengthening of students perceived ability to navigate unfamiliar tools and apply AI thoughtfully in learning contexts. Because these findings are based on descriptive statistics, they are interpreted as observed patterns rather than statistically significant changes.

These patterns are consistent with prior research on AI literacy and experiential learning, which suggests that meaningful engagement with AI emerges through opportunities for hands-on interaction and reflection rather than passive exposure alone [12]. Similarly, [3] found that project-based and interactive AI activities support learners in developing more intentional and critical approaches to using AI tools. Thus, structured experiential learning may play a role in supporting reflective engagement with AI technology, as seen by the observed increase in student's perceived confidence.

Changes in Problem-Solving Approaches

Pre-survey findings showed that students relied on a combination of peers, teachers, online searches, and AI tools when encountering challenges. Post-survey responses suggest a subtle shift in how students evaluated AI as a supporting tool. While AI continued to be viewed as a useful resource, students increasingly emphasized the importance of questioning AI outputs, refining prompts, and balancing AI assistance with human judgment. This change reflects a move toward more critical engagement with AI as part of students' broader problem-solving strategies. Thus, this shift toward more critical evaluation of AI assistance aligns with prior findings that emphasize the importance of fostering AI literacy skills that include questioning, verification, and balanced reliance on AI outputs [12].

Development of Ethical Awareness and Reasoning

One of the changes observed between pre- and post-survey responses relates to students' ethical awareness. Prior to the module, students' reflections on AI ethics often focused on general benefits and broad concerns, such as job loss or over-reliance. Following participation, students more frequently articulated specific ethical considerations, including algorithmic bias, privacy risks, misinformation, and the need for human oversight.

Qualitative comparisons suggest that students' ethical reasoning became more nuanced after engaging in scenario-based discussions and ethical dilemma activities. Students articulated more conditional perspectives, acknowledging trade-offs and the importance of context, regulation, and accountability in AI deployment, rather than viewing AI as uniformly beneficial or harmful. This pattern aligns with prior studies indicating that scenario-based and discussion-driven activities can support ethical reasoning about AI by encouraging learners to grapple with societal impact and competing values [3], [11].

3.4 Ethical Dilemma

The ethical dilemma cards provided a structured, thought-provoking opportunity for students to critically engage with the ethical dimensions of AI, particularly in socially sensitive domains such as mental health, hiring, and policing. Each scenario required students to evaluate potential benefits, risks, and societal implications of AI systems and to justify their decisions through group discussions. When combined with survey data, the written responses from these activities added rich qualitative depth to the findings.

Cautious Attitudes Toward AI Deployment

One prominent theme that emerged from the ethical dilemma responses was students' cautious approach to deploying AI systems in high-stakes contexts, particularly those involving mental health support and public safety.

In the mental health chatbot scenario, students emphasized the need for transparency and safeguards before releasing an AI system that provides emotional or psychological advice. As one group noted, *"Users should know it's not a real therapist,"* while another cautioned that deploying such a system prematurely *"could cause serious issues if it gives harmful advice."* As a result, several students recommended delaying deployment until additional safety checks and oversight mechanisms were in place. In contrast, responses to the predictive policing scenario revealed both initial acceptance and subsequent critical reflection. One group initially argued that *"the AI only uses criminal records, so there is no bias,"* asserting that increased policing would simply target areas with more crime. However, during whole-class discussion, many students challenged this view, questioning whether arrest data could reflect existing social biases and unfairly impact certain communities. This shift in perspective illustrates how engaging with contrasting viewpoints prompted students to reconsider assumptions about objectivity and risk in AI-driven decision-making.

These discussions reflect a growing recognition that AI deployment in sensitive domains warrants caution and careful ethical evaluation. While students did not always reach consensus, the process of debating real-world scenarios appeared to foster more cautious, reflective perspectives on the use of AI in contexts where errors or bias could have serious human impacts. This cautious stance aligns with prior research emphasizing the need for careful oversight when deploying AI in high-stakes domains such as mental health and public safety [2], [11].

Emphasis on Privacy and Transparency

Data privacy and transparency emerged as main ethical concerns across scenarios involving personal data collection and passive monitoring. In the data-tracking app scenario, students overwhelmingly supported informed consent, with one group stating that *"users should be told exactly how their data is being used."* While some students viewed anonymized data sharing as acceptable due to its potential benefits, such as *"improving the app for everyone"*—others raised concerns about user trust and unintended consequences, noting that *"people value their privacy and might not want their information shared."* These responses highlight students' ability to weigh innovation against individual rights.

Similarly, in the smart home device scenario, students expressed discomfort with passive audio recording without clear notification. One group stated, *"Companies shouldn't listen to conversations without a warning,"* while another acknowledged the trade-off by noting that improved accuracy *"comes at the cost of knowing too much personal information."* Across both scenarios, students consistently emphasized transparency, consent, and user awareness, reinforcing the ethical stance that privacy protections should remain central even when AI systems promise enhanced functionality.

Recognition of Bias and Community Impact

Students repeatedly identified algorithmic bias as a critical issue, particularly in the AI-assisted hiring scenario. Groups noted that keyword-based screening could exclude strong candidates, with one student explaining that *"someone could have good experience and recommendations, but if they don't use the right keywords, they get ignored."* This concern reflects students' understanding that seemingly neutral design choices can produce unfair outcomes.

Importantly, students did not place responsibility solely on the AI system. Instead, they emphasized accountability among those who design and implement such tools. As one group stated, *"The programmers and the people who decided to use the AI should be responsible if good applicants are filtered out."* While students recognized that automation could improve efficiency, several acknowledged the limitations of removing humans entirely from the process, noting that *"human review would be necessary for fair results, even if it slows things down."* These responses suggest a developing understanding of AI as a socio-technical system, where community-level impacts stem from human decisions embedded in data selection and system design.

Implications for Ethical Education

The high level of engagement with the ethical dilemma cards, along with the depth of student responses, demonstrates that high school students are capable of thoughtful and critical ethical reasoning about AI. While many students entered the program with some prior exposure to AI and ethical concepts, the structured, scenario-based discussions appeared to deepen their understanding and support the application of ethical principles to real-world contexts.

These findings support the integration of ethical dilemma activities into STEM education as a means of:

- Fostering critical thinking and civic responsibility,
- Promoting awareness of fairness, transparency, and accountability in technology, and
- Encouraging students to view themselves as future contributors to ethical technology design and deployment.

Implications for Practice

The findings of this study offer practical insights for educators, curriculum developers, and informal learning programs seeking to introduce artificial intelligence (AI) and ethics to high school students. The use of interactive tools like Google Teachable Machine helped make abstract AI concepts more accessible and engaging by allowing students to actively construct and test machine learning models. This hands-on approach supported experiential learning and strengthened conceptual understanding.

In addition, the ethical dilemma scenarios provided opportunities for students to apply ethical reasoning in structured yet open-ended ways. Through engagement with real-world examples, students demonstrated the ability to consider multiple perspectives, evaluate trade-offs, and reflect on the societal implications of AI systems, including issues related to bias, surveillance, and fairness. These activities supported students' development of ethical reasoning and critical thinking without requiring advanced technical expertise.

The active learning format of the AI and ethics module also appeared to promote student confidence and engagement. Learning through creating, discussing, and reflecting fostered curiosity and a sense of ownership over the learning process. Collaborative group work further supported this dynamic by exposing students to diverse viewpoints and encouraging peer-to-peer dialogue and negotiation.

Importantly, these findings suggest that introducing AI ethics at the high school level is not only feasible but beneficial, even for students with limited prior experience. Discussions of bias, privacy, and fairness were accessible and engaging, indicating that similar modules could be successfully implemented in both formal and informal learning environments.

Finally, the module offers a flexible and scalable model for future initiatives. Its emphasis on structured dialogue, low-barrier technologies, and ethical reflection provides a foundation for broader educational efforts aimed at preparing students to navigate and shape an increasingly AI-driven society. These findings provide practical insights for educators seeking to implement AI ethics modules in informal STEM settings.

4. Limitations

Although this study has certain limitations, it provides insightful information about teaching AI and ethics through active learning. First, only students enrolled in one summer program (XYZ) were included in the participant pool, which might not accurately represent the diversity of high school students more generally.

Consequently, the results may not be transferable to other educational or geographic contexts. Second, students' comprehension and long-term recollection of ethical reasoning may have been limited by the short duration of the AI and ethics module. Deeper examination of the technical and ethical content might have been possible with a lengthier intervention.

Third, a large portion of the data, especially from the questionnaires, depended on the knowledge, attitudes, and confidence that students self-reported. Social desirability bias or different interpretations of the survey items might have affected these answers.
Finally, although the written answers to the ethical dilemma cards provided valuable qualitative insights, the study's lack of a uniform framework to evaluate ethical reasoning proficiency restricts the analysis's consistency and comparability.

5. Conclusion

This study explored the design and implementation of active learning strategies to introduce high school students to foundational concepts in artificial intelligence (AI), with a specific focus on ethics and fairness. Through the XYZ summer program, students engaged with tools like Google Teachable Machine and participated in ethical dilemma discussions, allowing them to experience both the technical and moral dimensions of AI in an age-appropriate, accessible way.

Findings suggest that students were not only able to grasp basic AI concepts but also engage in thoughtful reflection on ethical challenges such as bias, fairness, and surveillance. The combination of hands-on experimentation and group-based ethical reasoning fostered critical thinking, increased confidence, and deeper engagement. These results affirm the value of integrating experiential, inquiry-based approaches into early AI education.

As AI continues to shape the world around us, introducing ethical awareness and responsible thinking at an early age is crucial. This study contributes a practical model for educators and program designers seeking to prepare the next generation of learners to understand, question, and ethically navigate the evolving landscape of AI.

6. Recommendations for Future Research

Future studies should look into how to scale and modify comparable AI ethics education courses for use in a variety of educational contexts, such as underfunded, urban, and rural schools. Increasing the demographic reach would enhance generalizability and show how students interact with AI ethics in various circumstances.

To evaluate long-term effects, particularly if students maintain their comprehension of prejudice, fairness, and ethical reasoning over time, longitudinal studies are also advised. Monitoring students after they graduate from the program can provide information about how early exposure to AI ethics influences their future academic or professional goals.

Furthermore, if future modules are to be incorporated into formal classroom settings, it may be beneficial to incorporate teacher viewpoints and implementation issues into their design. Students' deeper observations and emotional reactions that might not come through in written surveys could potentially be captured by including focus groups or interviews.

Lastly, to investigate how students react to more sophisticated topics like generative AI, data privacy, and algorithmic decision-making, future research may present more intricate ethical scenarios or use AI tools other than Google Teachable Machine.

References

[1] I. Inuwa-Dutse, "FATE in AI: Towards Algorithmic Inclusivity and Accessibility," in *Equity and Access in Algorithms, Mechanisms, and Optimization*, Boston MA USA: ACM, Oct. 2023, pp. 1–14. doi: 10.1145/3617694.3623233.

[2] B. Woolf, "Introduction to IJAIED Special Issue, FATE in AIED," *Int. J. Artif. Intell. Educ.*, vol. 32, no. 3, pp. 501–503, Sep. 2022, doi: 10.1007/s40593-022-00299-x.

[3] R. Williams *et al.*, "AI + Ethics Curricula for Middle School Youth: Lessons Learned from Three Project-Based Curricula," *Int. J. Artif. Intell. Educ.*, vol. 33, no. 2, pp. 325–383, Jun. 2023, doi: 10.1007/s40593-022-00298-y.

[4] "Teachable Machine." Accessed: Jan. 16, 2026. [Online]. Available: https://teachablemachine.withgoogle.com/

[5] "Moral Machine," Moral Machine. Accessed: Jan. 16, 2026. [Online]. Available: http://moralmachine.mit.edu

[6] "Moral Dilemma cards," Education. Accessed: Jan. 16, 2026. [Online]. Available: https://staff.ncsy.org/education/material/NU0qsba68B/moral-dilemma-cards

[7] M. Bearman, J. Ryan, and R. Ajjawi, "Discourses of artificial intelligence in higher education: a critical literature review," *High. Educ.*, vol. 86, no. 2, pp. 369–385, Aug. 2023, doi: 10.1007/s10734-022-00937-2.

[8] I. Roll and R. Wylie, "Evolution and Revolution in Artificial Intelligence in Education," *Int. J. Artif. Intell. Educ.*, vol. 26, no. 2, pp. 582–599, Jun. 2016, doi: 10.1007/s40593-016-0110-3.

[9] Association for the Advancement of Artificial Intelligence, Ed., *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media: ICWSM: 8-11 June 2020, held virtually*. Palo Alto, California: AAAI Press, 2020.

[10] Tejaskumar B. Modi, "Artificial Intelligence Ethics and Fairness: A study to address bias and fairness issues in AI systems, and the ethical implications of AI applications," *Rev. Rev. Index J. Multidiscip.*, vol. 3, no. 2, pp. 24–35, Jun. 2023, doi: 10.31305/rrijm2023.v03.n02.004.

[11] L. Floridi and J. Cowls, "A Unified Framework of Five Principles for AI in Society," *Harv. Data Sci. Rev.*, Jun. 2019, doi: 10.1162/99608f92.8cd550d1.

[12] D. Long and B. Magerko, "What is AI Literacy? Competencies and Design Considerations," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–16. doi: 10.1145/3313831.3376727.
[13] Y. Garje, S. Adams, B. Wellman, and M. Hynes, "Title: Building Curiosity and Competency: Designing and Evaluating Activities for Microelectronics Education (Evaluation of Program/Curricula)," in *2025 ASEE Annual Conference & Exposition Proceedings*, Montreal, Quebec, Canada: ASEE Conferences, Jun. 2025, p. 57290. doi: 10.18260/1-2--57290.
[14] J. J. McConnell, "Active learning and its use in computer science," *ACM SIGCUE Outlook*, vol. 24, no. 1–3, pp. 52–54, Jan. 1996, doi: 10.1145/1013718.237526.
[15] J. Michael and H. I. Modell, *Active Learning in Secondary and College Science Classrooms: A Working Model for Helping the Learner To Learn*, 0 ed. Routledge, 2003. doi: 10.4324/9781410609212.
[16] C. C. Bonwell and J. A. Eison, *Active learning: creating excitement in the classroom*. in ASHE-ERIC higher education report, no. 1, 1991. Washington, DC: School of Education and Human Development, George Washington University, 1991.

Appendix A: Survey Instruments

This appendix provides an overview of the pre- and post-survey instruments used to assess students' knowledge, perceptions, confidence, and ethical awareness related to artificial intelligence (AI). The surveys were administered before and after participation in the AI and ethics module as part of the summer program.

Pre-Survey Instrument

The pre-survey was designed to capture students' baseline familiarity with AI, comfort using AI tools, interest in STEM, and initial ethical perceptions. The instrument included a combination of multiple-choice, Likert-scale, and open-ended questions.

Question Types and Structure:

- 6 Likert-scale items (7-point scale: *Not at all familiar* to *Very familiar*)
- 1 Likert-scale anchor item (7-point scale: *Not comfortable* to *Very comfortable*)
- 6 Yes/No/Not sure questions
- 2 multiple-choice questions
- 4 open-ended response questions

Sample Pre-Survey Items:

- *How familiar are you with AI and chatbots?* (1–7)
- *How comfortable are you asking an AI chatbot for help with problem-solving?* (1–7)
- *Do you see yourself pursuing a career in AI, robotics, or a STEM-related field?*
- *What comes to mind when you think about AI and robotics?*
- *What concerns, if any, do you associate with AI?*

Post-Survey Instrument

The post-survey examined changes in students' understanding, confidence, and ethical awareness following participation in the module. It included similar constructs to the pre-survey, with additional emphasis on reflection and perceived learning outcomes.

Question Types and Structure:

- 13 Likert-scale items (7-point scale: *Very unlikely* to *Very likely*)
- 1 multiple-choice question
- 6 open-ended response questions

Sample Post-Survey Items:

- *How confident do you feel using AI for problem-solving after this experience?* (1–7)
- *How often did you refine your AI prompts to get better responses?* (1–7)
- *After this experience, how confident do you feel about learning new technologies or tools?* (1–7)
- *What are your thoughts on AI's role in society now?*
- *What ethical concerns about AI do you now associate with its use that you did not before?*

Responses from both surveys were used for descriptive statistical analysis and qualitative thematic analysis, as described in the Methods section.

Appendix B: Ethical Dilemma Card Scenarios

This appendix presents sample ethical dilemma scenarios used during the AI and ethics module. These scenarios were designed to prompt discussion, ethical reasoning, and collaborative decision-making among students. Each group received one scenario and responded in writing to guiding questions.

Sample Ethical Dilemma Scenarios

1. **Mental Health Chatbot**

Students evaluated an AI chatbot designed to provide mental health advice but not operated by licensed professionals.

Guiding Questions:

- Should users be informed that the chatbot is not a licensed therapist?
- What are the risks if the chatbot provides harmful advice?
- Should the system be launched immediately or delayed for additional safety checks?

2. **Predictive Policing Tool**

Students considered an AI system that predicts crime hotspots using historical arrest data.

Guiding Questions:

- Can past arrest data reflect bias?
- Could the system unfairly impact specific communities?
- Should AI be used to guide real-world policing decisions?

3. **AI for Hiring Decisions**

Students analyzed an AI-based resume screening tool prioritizing keywords and educational background.

Guiding Questions:

- Could this system overlook qualified candidates?
- Who should be accountable for biased or unfair hiring outcomes?
- Should human review be required alongside AI recommendations?

4. **Smart Home Device That Listens**

Students assessed a smart home assistant that records background conversations to improve accuracy.

Guiding Questions:

- Is it ethical to collect audio data without explicit user consent?
- Should users be notified when recording occurs?
- Who owns the collected audio data?

5. **Launching a Data-Tracking App**

Students evaluated a fitness app that collects biometric and location data for AI-driven predictions.

Guiding Questions:

- Should users be fully informed about how their data is used?
- Is anonymized data truly private?
- How should innovation be balanced with user privacy?

Written group responses to these scenarios were analyzed using inductive thematic analysis to identify patterns related to ethical awareness, bias recognition, privacy concerns, accountability, and human oversight.