2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Paper ID #51101

Building Pathways into Trustworthy AI (TAI) Technology: Experiential Learning Artificial Intelligence (ExLAI) Summer Program

Justin Zhan, University of Cincinnati
Prof. Youn Seon Lim, University of Cincinnati

Dr. Youn Seon Lim is an Assistant Professor of Psychometrics and Educational Measurement at the University of Cincinnati. Her research focuses on cognitive diagnosis modeling, Q-matrix theory, and survey, instrument, and test development, including scale development and construct validation. She also conducts interdisciplinary research on STEM career development and serves as Co-Principal Investigator on two NSF-funded projects supporting computer science pathways. She is an Associate Editor of the Journal of Classification.

Hanniel Shih, University of Cincinnati
Kate Plas, University of Cincinnati
Nicholas Jarvis, University of Cincinnati
Julia Holton, University of Cincinnati

BUILDING PATHWAYS INTO TRUSTWORTHY AI: THE EXLAI SUMMER PROGRAM

Justin Zhan, PhD, University of Cincinnati

He is the Fred A. & Juliet Esselborn Geier Chair Professor and head of the Department of Computer Science. His research spans data science, artificial intelligence, blockchain technologies, biomedical informatics, information assurance, and social computing. He serves as Principal Investigator for two NSF-funded projects focused on strengthening computer science pathways.

Youn Seon Lim, PhD, University of Cincinnati

Youn Seon Lim is an assistant professor of educational measurement and psychometrics in the Educational Studies division. Her research focuses on cognitive diagnosis modeling, Q-matrix theory, and survey, instrument, and test development, including scale development and construct validation. She also conducts interdisciplinary research on STEM career development and serves as Co-Principal Investigator on two NSF-funded projects supporting computer science pathways. She is an Associate Editor of the Journal of Classification.

***Hanniel Shih, MS, University of Cincinnati**

Hanniel Shih is a doctoral student in computer science who has taught courses in the ExLAI program.

***Kate Plas, MS, University of Cincinnati**

Kate Plas is a doctoral student in computer science who has taught courses in the ExLAI program.

***Nicholas Jarvis, BS, University of Cincinnati**

Nicholas Jarvis is a master's student in computer science who has taught courses in the ExLAI program.

Julia Holton, PhD, University of Cincinnati

Julia Holton is a Senior Research Associate at the Evaluation Services Center. She leads large-scale evaluation projects, including statewide studies of out-of-school-time programs. Trained as a school psychologist, her expertise includes qualitative and mixed-methods research focused on early literacy, mental health, and educational systems across PK–12 and higher education contexts.

Note. *Authors are listed in alphabetical order and contributed equally to this work.

BUILDING PATHWAYS INTO TRUSTWORTHY AI: THE EXLAI SUMMER PROGRAM

Abstract

Artificial intelligence (AI) increasingly shapes learning, work, health, and civic life, intensifying expectations that engineers and computing professionals develop AI systems responsibly. Persistent inequities in the AI workforce pipeline also motivate pathway programs that expand early access to authentic AI learning experiences. This paper reports short-term outcomes from a single-site implementation of the Experiential Learning Artificial Intelligence (ExLAI) summer program, an intensive eight-week introductory AI experience for early-stage learners (high school and early undergraduate students) that integrates trustworthy AI practices into the modeling workflow. ExLAI combines interactive lectures, hands-on coding demonstrations, and coached team projects in which participants define a problem, implement and evaluate an AI approach, and document trustworthiness considerations (e.g., data quality and stewardship, fairness and harmful bias, privacy, transparency/explainability, and robustness) using tools such as model cards and datasheets. Guided by self-efficacy theory and Social Cognitive Career Theory, the study uses a single-cohort pre–post design without a comparison group; findings are presented descriptively. Data sources included locally developed pre/post surveys with an embedded 11-item AI knowledge assessment, post-survey items on instructional implementation and program components, pathway awareness and intentions, and program records; end-of-program focus group input was used to provide contextual insight. Thirty-six students enrolled and 35 completed the program (97% retention). In the matched pre–post sample, AI knowledge increased from M = 6.43 to 8.87 (out of 11; n = 32). Participants also reported higher confidence across technical AI tasks and trustworthiness-relevant tasks (n = 33), including considering ethical dilemmas and reasoning about societal impacts. Post-survey ratings indicated high overall satisfaction and strong instructor communication, with more mixed ratings for pacing and course level. Participants also reported substantial awareness of AI-related educational and career pathways and strong interest in continued AI learning, while near-term opportunity-seeking behaviors were less common at the post-survey time point. Overall, the paper provides descriptive evidence from an intensive introductory program that integrates trustworthy AI practices into early technical instruction. It discusses design implications for supporting heterogeneous learners (e.g., pacing and collaboration scaffolds) and outlines future work incorporating performance-based assessment of trustworthy-AI competencies and longer-term follow-up.

Introduction

Artificial intelligence (AI) now plays a central role in products, services, and decision-making systems that shape learning, work, health, and civic life. As a result, engineering and computing programs face increasing pressure to prepare students not only to develop AI systems, but to do so in ways that are technically sound and socially responsible. In this paper, trustworthy AI (TAI) refers to AI systems and development practices that emphasize validity and reliability, safety and resilience, transparency and accountability, explainability and interpretability, protection of privacy, and the mitigation of harmful bias [1]–[4]. These principles carry clear implications for AI education: introductory learning experiences should help students connect core modeling practices to evaluation, documentation, and communication that make assumptions, limitations, and potential risks explicit.

Despite the rapid development of trustworthy AI frameworks, a gap remains in AI education research and practice. Relatively few introductory, project-based learning experiences for early-stage learners integrate trustworthy AI practices throughout the AI development workflow and provide empirical evidence of what learners gain from that integration. In many curricula, trustworthiness is presented as a stand-alone ethics topic or reserved for advanced coursework, which limits opportunities for novices to practice documentation, communicate risks, and engage in responsible decision-making alongside technical model development.

At the same time, the AI workforce pipeline continues to face persistent inequities. Broadening participation in computing and engineering is both a workforce need and an equity imperative [7]. Many learners—particularly those from groups historically underrepresented in STEM—have limited early exposure to AI, limited access to mentors, and limited opportunities to engage in authentic problem solving within supportive learning communities. Short-term, intensive programs (e.g., summer bridge and related interventions) are a common strategy for accelerating preparation and persistence, and synthesis research suggests that well-designed programs can yield measurable benefits for early academic outcomes and retention, although effects vary by design and context [8]. For early-stage learners who are still forming academic and career intentions, structured mastery experiences and supportive feedback may be especially important for confidence development and pathway exploration [6], [12].

The Experiential Learning Artificial Intelligence (ExLAI) summer program was developed as an accessible entry point into AI that integrates trustworthy AI practices into instruction and project work for heterogeneous early-stage learners. ExLAI combines interactive lectures, hands-on coding demonstrations, and team-based projects in which participants define a problem, implement an AI approach, evaluate model performance, and document trustworthiness considerations (e.g., data quality, fairness and harmful bias, privacy, transparency/explainability, and robustness) using structured artifacts such as model cards and datasheets [17], [18]. The program also incorporates demonstrations of model failure modes (e.g., distribution shift and basic adversarial examples) to emphasize stress testing and transparent reporting of limitations [19]. In this way, ExLAI treats trustworthiness as an integral component of the AI development workflow rather than as a separate ethics module. In the present study, trustworthiness is operationalized through these instructional practices and is assessed primarily through participants' knowledge and self-reported confidence in selected trustworthiness-relevant tasks, rather than through performance-based measures spanning all trustworthiness dimensions.

This paper reports findings from a single-site implementation of ExLAI delivered in an intensive eight-week format. Using matched pre–post survey measures, supplemented by end-of-program focus group input to provide contextual insight, we examine short-term changes in AI knowledge and confidence (including confidence in trustworthiness-relevant tasks), participants' ratings of instructional implementation and component helpfulness, and reported awareness of AI education and career pathways, including interest and near-term opportunity-seeking behaviors. Because the study employs a single-cohort pre–post design without a comparison condition, findings are presented descriptively and should not be interpreted as causal estimates of impact or broadly generalizable effects. The contribution is twofold: (1) a structured program model that integrates trustworthy AI practices within an introductory AI experience for heterogeneous early-

stage learners, and (2) descriptive evidence of short-term outcomes within this cohort to inform future research and program refinement.

Accordingly, this study addresses three research questions:

1. To what extent did participants demonstrate pre–post changes in AI knowledge and confidence, including confidence in trustworthiness-relevant tasks?
2. How did participants rate the learning environment and program components?
3. To what extent did participants report awareness of AI-related education and career pathways and intentions to pursue related opportunities?

Background and Related Work

This study draws on research in active and experiential learning, self-efficacy and career pathway development, and trustworthy AI and AI literacy frameworks. These perspectives provide the theoretical foundation for examining short-term changes in AI knowledge, task-specific confidence, participant perceptions of instructional implementation, and pathway awareness within an intensive introductory AI program.

Active and Experiential Learning in Introductory STEM Contexts

Research consistently shows that active learning approaches are associated with stronger academic outcomes than traditional lecture-based instruction. A meta-analysis across STEM disciplines found higher student performance and lower failure rates in active learning environments [9]. Engineering education research similarly highlights structured problem solving, guided application, and formative feedback as key mechanisms for supporting conceptual understanding and skill development [10]. In computing education, iterative practice, debugging, and scaffolded modeling tasks are particularly important for learners who enter with varied prior experience.

Project-based and inductive instructional approaches build on this foundation by introducing concepts through authentic problems rather than abstract exposition [13]. Learners apply ideas in concrete contexts and refine their understanding through guided experience. Experiential learning theory describes this process as a cycle of experience, reflection, conceptualization, and experimentation [11]. In compressed or intensive formats, such as summer bridge programs, these cycles must be deliberately structured to support rapid skill development [8]. Evidence from short-term STEM interventions suggests that knowledge gains and increases in task confidence are reasonable proximal outcomes when instruction combines structured practice, feedback, and collaboration [8].

In introductory AI education, many programs emphasize exposure to core technical topics, including supervised learning, neural networks, and model evaluation. However, empirical reports often focus on general interest, attitudes toward AI, or overall satisfaction rather than task-specific learning outcomes. Fewer studies examine changes in confidence tied to particular modeling practices, and even fewer document how responsible AI practices are integrated into early technical instruction. These gaps motivate the present study’s focus on short-term knowledge

and confidence outcomes in an AI program that embeds trustworthiness considerations within the technical workflow.

Self-Efficacy, Proximal Outcomes, and Pathway Development

The present study's focus on confidence and pathway awareness is grounded in self-efficacy theory and Social Cognitive Career Theory (SCCT) [6], [12]. Self-efficacy refers to individuals' beliefs about their ability to perform specific tasks and develops through mastery experiences, observation of others, social encouragement, and emotional responses [6]. Successfully completing structured tasks is a particularly strong source of self-efficacy. In short-term educational settings, changes in task-specific confidence can therefore serve as meaningful indicators of perceived capability.

SCCT links learning experiences to self-efficacy and outcome expectations, which influence interests, goals, and persistence in educational and career pathways [12]. For early-stage learners who are still forming academic identities and career intentions, short-term increases in confidence and awareness may precede longer-term decisions or actions. Studies of STEM enrichment and bridge programs have reported gains in self-reported confidence and pathway clarity even when longer-term outcomes require additional follow-up [8]. In AI and computing contexts, structured modeling experiences and guided participation may similarly strengthen confidence in applying technical skills and considering AI-related pathways.

Consistent with this framework, the present study examines task-specific confidence in both technical and trustworthiness-relevant AI tasks, as well as confidence in pursuing an AI career. It also assesses reported awareness of internships, jobs, and educational pathways, along with short-term intentions and opportunity-seeking behaviors. These outcomes are interpreted as proximal indicators aligned with SCCT rather than as evidence of long-term persistence or workforce entry.

Trustworthy AI and AI Literacy as Foundational Objectives

Trustworthy AI frameworks treat trustworthiness as a system-level property rather than a standalone topic. They emphasize that responsible AI development requires attention to data practices, model evaluation, documentation, and clear communication of limitations and risks [1]–[4]. Core dimensions typically include validity and reliability, fairness and mitigation of harmful bias, transparency and accountability, explainability and interpretability, robustness and safety, and protection of privacy. AI literacy frameworks similarly argue that foundational AI education should combine technical competence with the ability to reason about impacts, limitations, and appropriate use [21], [22].

In many educational settings, however, trustworthiness is presented as a discrete ethics unit separate from technical instruction. Recent work in responsible data science education instead advocates integrating documentation and evaluation tools—such as model cards and datasheets—into the modeling workflow itself [17], [18], [23]. Embedding these practices encourages learners to specify intended use, report performance characteristics, and document assumptions and limitations as part of routine development work. Despite the growing prominence of trustworthy AI

frameworks, empirical studies documenting how such practices are incorporated into introductory AI instruction remain limited. In addition, few studies examine short-term learning outcomes when trustworthiness is embedded within technical activities rather than taught as a separate topic. In this study, trustworthiness is operationalized through instructional emphasis on workflow practices and structured documentation. Outcomes are assessed through changes in knowledge and self-reported confidence in selected trustworthiness-relevant tasks, rather than through performance-based evaluation of artifacts across all trustworthiness dimensions.

Summary

Taken together, the literature motivates a focus on proximal, measurable outcomes that are plausibly responsive to an intensive introductory intervention. Active and experiential learning research supports examining short-term changes in foundational knowledge and task-specific confidence following structured practice and guided application [9]–[11], [13]. Self-efficacy theory and SCCT further justify treating task confidence, career confidence, and pathway awareness/intentions as proximal indicators that may shift before longer-term educational or workforce outcomes are observable [6], [12]. Finally, trustworthy AI and AI literacy frameworks motivate evaluating whether learners report increased confidence in tasks that explicitly involve ethical reasoning and societal implications when trustworthiness is integrated into the technical workflow [1]–[4], [17]–[19], [21]–[23].

Program Description: ExLAI Structure and Curriculum

ExLAI is an eight-week summer program that introduces early-stage learners—high school students and first-year undergraduates—to core artificial intelligence (AI) concepts while integrating trustworthy AI practices into technical instruction. Consistent with the definition presented earlier, trustworthiness is incorporated throughout the model development process, with attention to fairness (harmful bias), privacy, transparency and explainability, and robustness [1]–[4]. Instruction includes structured documentation, subgroup performance analysis, stress testing, and explicit reporting of intended use and limitations through tools such as model cards and datasheets [17], [18]. The program also incorporates demonstrations of model failure modes, including distribution shift and basic adversarial examples, to underscore the importance of robustness and transparent communication of limitations [19].
Participants engaged in approximately 20 hours per week of synchronous instruction and guided project work. Sessions combined interactive lectures, coding demonstrations, and structured project time. Students worked in teams to complete a research-oriented project and presented their work in a final presentation and poster session. Implementation details and participation records were documented throughout the eight-week program.

Table 1 summarizes the main features of the program. The curriculum covered foundational AI topics, including supervised learning, neural networks, and deep learning, while integrating trustworthiness considerations into both instruction and project activities. Guest speakers and mentoring sessions were included to connect course content to applied AI practice and to increase participants' awareness of educational and career pathways.

Table 1. ExLAI program at a glance

Feature	Description
Duration and intensity	8 weeks; approximately 20 hours per week
Primary audience	High school and early undergraduate students with heterogeneous prior programming experience
Instructional model	Interactive lectures, hands-on coding demonstrations, and coached lab/project work
Core learning activities	Guided practice, team-based research project, presentations, and poster session
TAI integration	Recurring attention to fairness, transparency/explainability, privacy, and robustness
Culminating deliverables	Team presentation and research poster with attention to documenting trustworthiness considerations

Instruction was organized into weekly modules that combined synchronous sessions, guided practice, and coached project work. The outline below summarizes the sequence of instruction and illustrates how trustworthiness considerations were incorporated throughout the program.

Week-by-Week Curriculum and TAI Integration

- Week 1: Orientation and overview of AI; introduction to trustworthiness considerations through case discussions and stakeholder and harm-mapping activities.
- Week 2: Data foundations and supervised learning; emphasis on privacy and data stewardship through structured dataset documentation (e.g., motivation, composition, recommended use).
- Week 3: Model assessment; introduction to evaluation metrics and fairness through subgroup comparisons and discussion of potential bias and mitigation strategies.
- Week 4: Neural networks and deep learning workflows; focus on transparency and explainability through documentation of features, assumptions, and limitations.
- Week 5: Robustness and security; demonstrations of performance degradation under distribution shift and basic adversarial examples to highlight stress testing and documentation of failure modes.
- Week 6: Project scoping and research planning; teams defined success criteria and developed a project-specific checklist addressing data quality, privacy, evaluation plans, and limitations.
- Week 7: Project development and revision; teams refined models with mentor feedback and incorporated documentation into technical reporting; guest sessions addressed educational and career pathways.
- Week 8: Dissemination; teams presented project results in presentations and a poster session, reporting technical findings alongside trustworthiness considerations (intended use, limitations, and potential impacts).

The week-by-week outline is provided as an example of implementation; pacing and emphasis may vary depending on participant preparation, instructional staffing, and local constraints.

Methods

Design

This study employed a single-cohort pre-post design to examine short-term learning outcomes, program experiences, and AI pathway awareness among early-stage learners participating in ExLAI. Because no comparison group was included, analyses are descriptive rather than causal. Data sources included (a) pre- and post-program online surveys administered in Qualtrics, including an embedded 11-item AI knowledge assessment; (b) end-of-program survey items on program experiences and pathway awareness; (c) an end-of-program focus group conducted using a semi-structured protocol to contextualize survey findings; and (d) program records (e.g., enrollment and completion/retention). Data sources and analytic indicators were aligned with the study research questions, as summarized in Table 2.

Table 2. Alignment of research questions, data sources, and analytic indicators

RQ	Data Sources	Analytic Indicators
1	Pre/post surveys: 11-item knowledge assessment; 7 task-confidence items; AI career-confidence item	AI knowledge total score (range 0–11); task-confidence item ratings (5-point scale; analyzed individually); AI career-confidence item rating (5-point scale)
2	Program records; post survey (helpfulness and satisfaction items); focus group	Enrollment and completion/retention; item-level helpfulness and satisfaction means; likelihood-to-recommend rating (0–10); contextual observations related to pacing and teamwork
3	Post survey (pathway awareness, interest, and opportunity-seeking items)	Awareness of internships, jobs, and educational pathways (Yes/No); interest in additional AI education (Yes/No); near-term opportunity seeking (Yes/No)

Participants

A total of 36 students enrolled in the ExLAI program and 35 completed the eight-week program (97% retention). Participants were high school students and first-year undergraduate students aged 15–19. Based on pre-survey responses, 66% identified as college students (n = 23) and 34% as high school students (n = 12). Regarding gender, 74% identified as male (n = 26) and 26% identified as female (n = 9). Participants identified as White (45%, n = 17), Asian (21%, n = 8), Black (18%, n = 7), Hispanic or Latino (8%, n = 3), and Middle Eastern or North African (5%, n = 2); two participants identified with more than one racial or ethnic group. Three participants reported being English language learners.

Surveys were administered at pre- and post-program time points using Qualtrics. Thirty-five participants completed the pre-survey and 36 completed the post-survey. Analyses examining change over time used a matched pre–post sample of 33 participants; 32 participants completed the full 11-item post-program knowledge assessment. Program records were used to describe implementation and participation patterns.

One focus group was conducted at the conclusion of the program to contextualize survey findings. Twelve students were invited to participate, and four students took part in a one-hour, in-person session. The session was audio recorded and transcribed for analysis.

Measures

Survey instruments were developed specifically for this study to align with program goals and research questions. Measures assessed four outcome areas: (1) AI knowledge, (2) task-specific confidence, (3) perceptions of the learning environment and program components, and (4) pathway awareness and intentions.

AI knowledge was measured using an 11-item multiple-choice assessment administered at pre and post. Scores were computed as the total number of correct responses (range 0–11). Task-specific confidence was assessed using seven items measuring confidence in performing AI- and trustworthy AI–related tasks (e.g., applying AI and machine-learning concepts, using deep-learning frameworks and cloud services, considering ethical dilemmas, understanding societal implications, and coding in Python). Responses were recorded on a 5-point scale (1 = Not at all confident to 5 = Extremely confident). A separate item assessed confidence in pursuing a career in AI using the same 5-point scale.

Perceptions of the learning environment and program components were assessed in the post-survey using items measuring satisfaction with instructional implementation and perceived helpfulness of program components, each on 5-point scales.

Pathway awareness and intentions were assessed in the post-survey using dichotomous (Yes/No) items addressing awareness of internships, jobs, and educational pathways; interest in pursuing such opportunities within the next 12 months; and near-term opportunity-seeking behaviors. The post-survey also included a 0–10 likelihood-to-recommend item.

With the exception of the AI knowledge total score, survey outcomes were analyzed at the item level rather than combined into multi-item scales.

Analysis

Quantitative analyses summarized survey responses descriptively. Means and standard deviations were reported for Likert-type items, and frequencies and percentages were reported for dichotomous (Yes/No) items. For matched respondents, pre–post change scores (post minus pre) were computed for the AI knowledge total score and confidence items. Descriptive summaries were also computed for selected subgroups (e.g., no prior programming experience). Item-level sample sizes are reported where missing data reduced n. Analyses were conducted using SPSS (version 29.0). Partial survey responses were retained, resulting in varying sample sizes across items.

A focus group was conducted at the conclusion of the program to provide contextual insight into participants' experiences. Observations from the discussion are used to interpret survey patterns where relevant.

Results

Results are organized by research question.

RQ1. AI Knowledge and Confidence

AI knowledge was measured using an 11-item multiple-choice assessment scored as the total number correct (range 0–11). In the matched pre/post sample, the mean knowledge score increased from 6.43 at pre to 8.87 at post (Δ = +2.44; n = 32), corresponding to an increase from 58.5% to 80.6% correct. Among participants who reported no prior programming experience at the start of the program (n = 7), the mean knowledge score increased from 5.00 to 7.29 (Δ = +2.29). These descriptive results indicate higher post-program performance on the knowledge assessment for both the overall matched sample and the subgroup entering without prior programming experience.

Task-specific confidence was assessed using seven items asking how confident participants were in their ability to perform AI- and trustworthy AI–relevant tasks, each rated on a 5-point scale (1 = Not at all confident; 5 = Extremely confident). Confidence increased across all seven tasks (n = 33). The largest gains were observed for applying basic machine-learning concepts (Δ = +1.21) and using deep-learning frameworks (Δ = +1.12). Confidence also increased for tasks aligned with trustworthiness considerations, including considering ethical dilemmas (Δ = +0.64) and understanding AI's potential for addressing societal issues (Δ = +0.52). Confidence in coding in Python increased from 2.88 to 3.58 (Δ = +0.70).

In addition, confidence in pursuing a career in AI was assessed using a separate single item (5-point scale). Career-confidence increased from 2.91 at pre to 4.42 at post (Δ = +1.51; n = 32). Taken together, these patterns indicate short-term increases in foundational AI knowledge and in self-reported capability to carry out technical and trustworthiness-relevant tasks over the duration of the program; because there was no comparison group, changes should be interpreted as within-cohort pre/post differences rather than causal effects.

Table 3. Pre/Post Changes in AI Knowledge and Self-Reported Confidence

Measure/Item	Pre M	Post M	Δ	n
AI Knowledge (11-item total score; range 0–11)				
Total score (all matched participants)	6.43	8.87	+2.44	32
Total score (no prior programming subgroup)	5.00	7.29	+2.29	7
Task-Specific and AI Confidence (8 items; 5-point scale)				
Apply basic AI concepts and applications	2.79	3.59	+0.80	33
Consider ethical dilemmas of AI	3.24	3.88	+0.64	33
Apply basic machine-learning concepts and applications	2.52	3.73	+1.21	33
Use deep-learning frameworks (e.g., TensorFlow/PyTorch)	2.00	3.12	+1.12	33
Use cloud-based AI services (e.g., AWS/Google/Azure)	2.30	2.91	+0.61	33
Understand AI's potential for addressing societal issues	3.39	3.91	+0.52	33
Coding in Python	2.88	3.58	+0.70	33
Confidence pursuing a career in AI	2.91	4.42	+1.51	32

Note. AI knowledge scores are total correct out of 11 items (range 0–11). Confidence items use a 5-point scale (1 = Not at all confident; 5 = Extremely confident). Δ = Post − Pre.

RQ2. Learning Environment and Program Components

Pre-program open-ended responses indicated that some participants anticipated challenges related to limited prior knowledge, the accelerated eight-week format, and differences in preparation (e.g., mathematics or programming background). These expectations provide context for post-program ratings of pacing and course level.

Post-survey results indicated generally positive perceptions of instructional support and implementation (Table 4a). Reported means represent item-level averages across participants (n = 33). Overall satisfaction with program implementation was M = 4.15 (SD = 0.83) on a 5-point agreement scale, and the mean likelihood of recommending the program was 8.21 on a 0–10 scale. Instructor communication received the highest rating (M = 4.30), followed by engagement of teaching methods (M = 4.00). Ratings were comparatively lower for pacing (M = 3.12) and perceived appropriateness of course level (M = 3.55), indicating more varied participant experiences on these dimensions.

Table 4a. Satisfaction with Program Implementation (Post; n = 33)

Statement	Mean	SD
Instructor communication	4.30	0.684
Overall satisfaction	4.15	0.834
Teaching methods were engaging	4.00	0.968
Material level appropriate	3.55	1.227
Pacing appropriate	3.12	1.083

Scale: 1 = Strongly disagree to 5 = Strongly agree.

Participants also rated the perceived helpfulness of specific program components (Table 5). Means reflect item-level averages across respondents. Lectures received the highest rating (M = 4.09), followed by the group research project (M = 3.88). Guest speakers and the poster session were rated similarly (M = 3.79). Standard deviations ranged from 0.90 to 1.23, suggesting variability in how participants experienced particular components.

Table 5. Perceived Helpfulness of Program Components (Post; n = 33)

Component	Mean	SD
Lectures	4.09	1.011
Group research project	3.88	1.023
Guest speakers	3.79	1.166
Poster session	3.79	1.111
Coding demonstration	3.76	1.226
In-class discussions	3.64	1.168
Group presentations	3.42	0.902

Scale: 1 = Not helpful to 5 = Extremely helpful.

Focus group findings (n = 4) provided additional context. Participants described the pace as demanding, particularly for those with less prior experience, and noted uneven participation within some teams. Guest speaker sessions were viewed as informative but primarily lecture-oriented.

These qualitative observations align with the survey findings, particularly the more moderate ratings for pacing and team-based components.

RQ3. Pathway Awareness and Intentions

Post-survey responses indicated substantial reported awareness of education and career pathways related to AI (Table 6). Percentages reflect the proportion of respondents selecting "Yes" for each item (n = 33). Fifty-five percent reported learning about AI internship opportunities (n = 18), 61% reported learning about AI job opportunities (n = 20), and 82% reported learning about AI educational pathways (n = 27). In addition, 91% indicated interest in pursuing additional AI education within the next 12 months (n = 30).

Table 6. Post-Program Pathway Awareness and Intentions (n = 33)

Indicator	n	%
Learned about AI internship opportunities	18	55%
Learned about AI job opportunities	20	61%
Learned about AI educational pathways	27	82%
Interested in additional AI education (next 12 months)	30	91%

Opportunity-seeking behaviors were relatively infrequent at the time of the post-survey (Table 7a). For example, 9.1% of participants reported applying for at least one AI internship since the start of the program (n = 3), and none reported obtaining an AI internship. A similar pattern was observed for job applications. One respondent did not answer the item regarding interest in attaining an AI job within the next 12 months (n = 32 for that item).

Table 7a. Opportunity-Seeking Behaviors (Post; n = 33 unless noted)

Indicator	Yes n (%)	No n (%)
Interested in attaining an AI internship in the next 12 months	23 (69.7%)	10 (30.3%)
Applied for at least one AI internship since the start of the program	3 (9.1%)	30 (90.9%)
Attained a new AI internship since the start of the program	0 (0.0%)	33 (100.0%)
Interested in attaining an AI job in the next 12 months (n = 32)	17 (53.1%)	15 (46.9%)
Applied for at least one AI job since the start of the program	3 (9.1%)	30 (90.9%)
Attained a new AI job since the start of the program	0 (0.0%)	33 (100.0%)
Applied for additional AI educational opportunities since the start of the program	3 (9.1%)	30 (90.9%)
Enrolled in additional AI educational opportunities since the start of the program	1 (3.0%)	32 (97.0%)

Overall, respondents reported high levels of pathway awareness and strong interest in continued AI education, whereas immediate application or enrollment behaviors were relatively uncommon at the time of data collection.

Discussion

This study examined short-term outcomes from participation in ExLAI, an intensive eight-week introductory AI program that integrates trustworthy AI practices into technical instruction and project work. Following the program, participants showed higher scores on a knowledge assessment and reported increased confidence across AI-related tasks, including those involving trustworthiness considerations. Participants also rated instructional implementation and program components positively and reported greater awareness of AI-related educational and career pathways. Because the study employed a single-cohort pre–post design without a comparison group, these results represent descriptive changes within this cohort and should not be interpreted as evidence of causal impact.

Participants' knowledge scores increased from pre to post, including among those who entered the program without prior programming experience. This pattern aligns with research on active and experiential learning, which emphasizes repeated practice, feedback, and guided application as mechanisms supporting short-term gains in foundational understanding [9]–[11]. Confidence also increased across all task-specific items, with larger changes observed for core technical skills such as machine-learning concepts and deep-learning frameworks. Increases in confidence for trustworthiness-relevant tasks—such as considering ethical dilemmas and reasoning about societal implications—are notable, particularly because introductory curricula often treat trustworthiness as a separate ethics topic rather than integrating it into technical workflows [1]–[4], [21], [22]. In ExLAI, trustworthiness was incorporated into routine development activities through documentation practices, discussion of limitations, and attention to potential harms, consistent with frameworks that view trustworthiness as a system-level property addressed through evaluation, documentation, and risk communication [1]–[4], [17], [18]. In this study, trustworthiness-related outcomes are reflected in knowledge and self-reported confidence in selected tasks, rather than in performance-based evaluation of artifacts across all trustworthiness dimensions.

Post-survey responses indicated high levels of pathway awareness and interest in continued AI education, whereas concrete opportunity-seeking actions were relatively uncommon at the time of data collection. This difference is likely related to the short interval between program completion and the post-survey, as well as the timing of internship and job application cycles. From a pathway-development perspective, increases in task-specific confidence and awareness are consistent with self-efficacy theory and Social Cognitive Career Theory (SCCT), which identify mastery experiences and contextual supports as proximal influences on interests and intentions [6], [12]. These findings suggest that an intensive introductory program may influence early indicators of pathway development, although longer-term follow-up is necessary to determine whether these shifts translate into sustained educational or career outcomes.

Overall, the study provides descriptive evidence of an instructional approach that integrates trustworthy AI practices within an introductory AI experience for learners with varied backgrounds. ExLAI demonstrates how documentation practices, including model cards and datasheets, and explicit discussion of limitations and model failure can be incorporated into early technical instruction in alignment with trustworthy AI and AI literacy frameworks [1]–[4], [17]–[19], [21]–[23]. The findings also underscore the importance of pacing and structured support for collaboration in intensive programs serving heterogeneous cohorts [8], [13]–[15].

Conclusion

This study examined short-term outcomes from a single-site implementation of ExLAI, an intensive introductory AI program that integrates trustworthy AI practices into technical instruction. Participants demonstrated increases in AI knowledge and task-specific confidence, including confidence in trustworthiness-relevant tasks, and reported positive perceptions of instructional implementation and substantial pathway awareness. Although the single-cohort pre–post design does not permit causal inference, the findings provide descriptive evidence that foundational AI concepts and workflow-based trustworthiness practices can be introduced effectively in an intensive format for heterogeneous early-stage learners. Future research should incorporate comparison conditions, performance-based assessments of trustworthiness practices, and longitudinal follow-up to evaluate longer-term educational and career outcomes.

Limitations

Several limitations should be considered when interpreting these findings. First, the study used a single-cohort pre–post design without a comparison group, which limits causal interpretation. The observed changes represent differences within this cohort over time and cannot be attributed solely to participation in the program. Second, many outcomes were based on self-reported survey responses, including confidence, satisfaction, and pathway awareness. Self-report measures may be influenced by response bias and do not necessarily reflect demonstrated performance or sustained behavioral change. In addition, several survey items were developed for this program context and were not subjected to independent validation; therefore, the findings should be interpreted as descriptive indicators rather than as validated scale scores.

Third, the AI knowledge assessment was limited to 11 items measuring foundational concepts. While appropriate for capturing short-term learning, it does not provide a comprehensive assessment of deeper conceptual understanding or applied skill. In addition, although trustworthiness practices were integrated into program activities, the study did not include performance-based evaluations of documentation quality, fairness analyses, robustness testing, or other trustworthiness dimensions. As a result, the findings reflect changes in knowledge and self-reported confidence rather than direct evidence of applied competence or artifact quality.

Fourth, data were collected immediately after the program concluded. Reported increases in pathway awareness and intentions may not translate into sustained engagement, course enrollment, internship attainment, or career outcomes. Longer-term follow-up is needed to determine whether these short-term changes persist or are associated with subsequent educational or professional decisions.

Finally, the study was conducted at a single site with a relatively small cohort. Participant demographics and levels of prior preparation may differ across institutions and contexts, which limits the generalizability of the findings. Replication with additional cohorts and in different settings would strengthen the evidence base for the program model.

References

[1] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," NIST AI 100-1, Jan. 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf. [Accessed: Feb. 22, 2026].

[2] European Commission, High-Level Expert Group on Artificial Intelligence, "Ethics guidelines for trustworthy AI," Apr. 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. [Accessed: Feb. 22, 2026].

[3] UNESCO, "Recommendation on the Ethics of Artificial Intelligence," Nov. 2021. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000380455. [Accessed: Feb. 22, 2026].

[4] Organisation for Economic Co-operation and Development (OECD), "Recommendation of the Council on Artificial Intelligence," OECD/LEGAL/0449, May 2019. [Online]. Available: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449. [Accessed: Feb. 22, 2026].

[5] S. I. Donaldson, Program Theory-Driven Evaluation Science: Strategies and Applications. Mahwah, NJ: Lawrence Erlbaum Associates, 2007.

[6] A. Bandura, Self-Efficacy: The Exercise of Control. New York, NY: W. H. Freeman, 1997.

[7] D. E. Chubin, G. S. May, and E. L. Babco, "Diversifying the engineering workforce," J. Eng. Educ., vol. 94, no. 1, pp. 73-86, Jan. 2005, doi: 10.1002/j.2168-9830.2005.tb00830.x.

[8] B. C. Bradford, M. E. Beier, and F. L. Oswald, "A meta-analysis of university STEM summer bridge program effectiveness," CBE—Life Sci. Educ., vol. 20, no. 2, Art. no. ar21, Jun. 2021, doi: 10.1187/cbe.20-03-0046.

[9] S. Freeman et al., "Active learning increases student performance in science, engineering, and mathematics," Proc. Natl. Acad. Sci. U.S.A., vol. 111, no. 23, pp. 8410-8415, Jun. 2014, doi: 10.1073/pnas.1319030111.

[10] M. Prince, "Does active learning work? A review of the research," J. Eng. Educ., vol. 93, no. 3, pp. 223-231, Jul. 2004, doi: 10.1002/j.2168-9830.2004.tb00809.x.

[11] D. A. Kolb, Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[12] R. W. Lent, S. D. Brown, and G. Hackett, "Toward a unifying social cognitive theory of career and academic interest, choice, and performance," J. Vocat. Behav., vol. 45, no. 1, pp. 79-122, 1994, doi: 10.1006/jvbe.1994.1027.

[13] M. J. Prince and R. M. Felder, "Inductive teaching and learning methods: Definitions, comparisons, and research bases," J. Eng. Educ., vol. 95, no. 2, pp. 123-138, Apr. 2006, doi: 10.1002/j.2168-9830.2006.tb00884.x.

[14] J. Lave and E. Wenger, Situated Learning: Legitimate Peripheral Participation. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[15] E. Wenger, Communities of Practice: Learning, Meaning, and Identity. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[16] A. Carpi, D. M. Ronan, H. M. Falconer, and N. H. Lents, "Cultivating minority scientists: Undergraduate research increases self-efficacy and career ambitions for underrepresented students in STEM," J. Res. Sci. Teach., vol. 54, no. 2, pp. 169-194, Feb. 2017, doi: 10.1002/tea.21341.

[17] M. Mitchell et al., "Model cards for model reporting," in Proc. Conf. Fairness, Accountability, and Transparency (FAT*), Atlanta, GA, USA, Jan. 2019, pp. 220-229, doi: 10.1145/3287560.3287596.

[18] T. Gebru et al., "Datasheets for datasets," Commun. ACM, vol. 64, no. 12, pp. 86-92, Dec. 2021, doi: 10.1145/3458723.
[19] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," arXiv:1412.6572, Dec. 2014. [Online]. Available: https://arxiv.org/abs/1412.6572. [Accessed: Feb. 22, 2026].
[20] M. Encalada, K. West, and S. Turner, "Technology for Diverse Learners into Careers in Artificial Intelligence (ExLAI) Program Evaluation Year 1," University Evaluation Services Center, Aug. 2025, unpublished.
[21] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in Proc. CHI Conf. Human Factors Comput. Syst. (CHI), Honolulu, HI, USA, Apr. 2020, pp. 1-16, doi: 10.1145/3313831.3376727.
[22] D. Touretzky, C. Gardner-McCune, F. Martin, and D. Seehorn, "Envisioning AI for K-12: What should every child know about AI?" Proc. AAAI Conf. Artif. Intell., vol. 33, no. 1, pp. 9795-9799, 2019, doi: 10.1609/aaai.v33i01.33019795.
[23] A. Lewis and J. Stoyanovich, "Teaching responsible data science: Charting new pedagogical territory," Int. J. Artif. Intell. Educ., vol. 32, no. 3, pp. 783-807, 2022, doi: 10.1007/s40593-021-00241-7.