

RESEARCH

Open Access



Effects of AI-based and video instruction on machine learning competencies among computing and robotics education students

Christian Basil Omeh^{1*} , Valentine Joseph Owan² , Ismaila Temitayo Sanusi³ and Ilesanmi Olade⁴

*Correspondence:

Christian Basil Omeh
omehcb@unisa.ac.za

¹Department of Science and Technology Education, College of Education, University of South Africa, Pretoria, South Africa

²Department of Educational Psychology, University of Calabar, Calabar, Nigeria

³University of Eastern Finland, Joensuu, Finland

⁴Department of Education, NITE, EECS Queen's University Belfast, Belfast, Northern Ireland, UK

Abstract

The rapid expansion of artificial intelligence has increased the need for graduates who possess practical competence in machine learning (ML), particularly in developing countries where instructional resources are limited. However, many computer science education students in Nigerian universities complete their programmes without adequate hands-on ML skills. This study examined the effect of AI-based instruction, supported by ChatGPT and Copilot, compared with conventional video-based instruction on the ML competencies of computer and robotics undergraduates' education. A quasi-experimental non-equivalent pretest–posttest control design was adopted, involving 61 students enrolled in a compulsory Introduction to Machine Learning course. Participants were assigned to either an AI-based project-driven instructional group ($n=25$) or a video-based instruction group ($n=36$). Over 10 weeks, both groups covered identical ML content, but differed in instructional approach. Data were collected using validated instruments measuring data modelling, data preprocessing, data model construction, and data evaluation. MANCOVA and ANCOVA were used to test the hypotheses at a 0.05 significance level. Results showed a significant multivariate effect of instructional method on combined ML competencies (Pillai's Trace = 0.602, $p < 0.001$). Follow-up univariate tests revealed that AI-based instruction produced significantly higher achievement in data modelling, data preprocessing, and data evaluation, with large effect sizes ($\eta^2 = 0.341$ – 0.461). However, no significant difference was found in data model construction. The findings demonstrate that AI-supported instruction substantially enhances students' practical ML abilities more than video-based instruction. The study recommends integrating AI tools into ML curricula to strengthen practical competence.

Keywords Artificial Intelligence, Computer and robotics education, Data model, Machine learning

1 Introduction

Artificial intelligence (AI) is projected to add more than fifteen trillion dollars to the world economy by 2030 [1]. Yet less than twenty per cent of computer science graduates, especially in developing countries, leave school without machine learning (ML) skills to



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

meet industry needs [2]. This has positioned the ML competence as urgent for students in computer education programmes. ML is an aspect of AI, as it influences sectors like health care, banking, security and education [3, 4]. Nations that invest in AI skills gain economic and social advantages, while those with weak capacity risk widening inequality [5]. This global shift places demands on universities, especially in developing countries like Nigeria, to prepare graduates who can understand and apply machine learning skills, which form the technical core of AI. Yet many computer science education students complete their programmes without the practical competence required to work with ML systems. This problem concerns educators, industry leaders and policy makers. It also raises an important question on how Nigerian institutions can strengthen ML instruction through appropriate tools and teaching methods, especially in environments where resources, infrastructure and skilled instructors are limited. ML forms the technical base of modern AI. It requires knowledge of data handling, model development and performance assessment [6]. The progress remains slow in many developing countries. Several institutions work with poor infrastructure, outdated teaching approaches, and staff who lack adequate preparation for AI-driven instruction [7, 8]. Students also find ML difficult when they learn it without practical activities. This widens the gap between theory and practice, leaving many graduates unready for workplace demands. These concerns guide the present study.

AI-based instruction and video learning have spanned educational robotics, AI chatbots, generative AI tools, and instructional videos, consistently yielding moderate-to-large gains in computational thinking, programming self-efficacy, motivation, and academic achievement, particularly in STEM and higher education contexts [9–12]. Policy suggestions increasingly emphasise AI literacy and curricular integration across all educational levels [13]. Yet, most efforts concentrate on general computational skills or introductory AI concepts, often directed at K–12 or graduate learners rather than undergraduates in computer and robotics programs [14, 15]. Despite these advances, persistent gaps remain. Methodologically, many studies rely on AI research, not necessarily ML content; some studies rely on single interventions, single control groups, restricting generalizability and causal inference [11, 16]. Research is often discipline-specific, with little focus on undergraduate ML learning environments. No known study compares the distinct effects of AI-based versus video instruction on higher-order ML skills. Critically, empirical evidence directly addressing undergraduate ML competency development is limited. Current literature aggregates across broader STEM fields or emphasises related constructs such as computational thinking or AI literacy [12]. Addressing this gap requires rigorous, targeted research to design, implement, and evaluate instructional strategies that foster robust ML skills at the undergraduate level using AI-based and video-based instruction. However, this study examines how the integration of AI tools and video-based instruction supports the ML learning of computer science education students in a developing country context.

This study employed a quasi-experimental non-equivalent pretest–posttest control design with undergraduates enrolled in an Introduction to Machine Learning course. Participants were divided into two groups: an AI-driven, project-based instructional group and a video-instruction group. Over 10 weeks, both groups engaged with identical ML content but through distinct instructional approaches. ML competencies such as data modelling, preprocessing, model construction, and evaluation were assessed

using validated instruments and analyzed through rigorous statistical methods. Findings revealed that AI-based instruction significantly improved student performance in data modelling, preprocessing, and evaluation. The AI-supported environment offered personalized guidance, adaptive feedback, and interactive problem-solving, which proved especially beneficial for students with lower academic ability. These results align with cognitive load theory, demonstrating that intelligent feedback reduces mental overload and enhances retention. This research underscores the transformative potential of AI in machine learning. By fostering inclusivity, engagement, and practical competence, AI-based instruction can enrich ML curricula, inform educational policy, and better prepare students for success in a data-driven world.

1.1 Theoretical framework

Cognitive Load Theory (CLT), introduced by [17], provides a useful lens for examining how students learn with complex material in machine learning education. CLT emphasizes the limited capacity of working memory and the importance of instructional design in reducing unnecessary strain while fostering meaningful learning. Three types of cognitive load are central to this framework: intrinsic, extraneous, and germane. Intrinsic load arises from the inherent complexity of machine learning topics such as neural networks or optimization procedures. For learners, these concepts can quickly overwhelm working memory. This call for instructional strategies that break lessons into smaller steps and use worked examples [3, 18] help manage this load. Video tutorials, when designed according to [19] multimedia principles, can reduce intrinsic load by simplifying coding steps or algorithmic processes through clear narration and supportive visuals. On the other hand, AI tools assisted in scaffolding explanations, ensuring learners confront complexity gradually and with tailored support. Extraneous load stems from poorly designed or confusing instructional materials. Disorganised slides or lengthy, unclear explanations increase mental effort without aiding comprehension. Video tutorials often risk this problem due to their passive format; when explanations are vague or overly detailed, learners must self-regulate, which can heighten extraneous load [4]. By contrast, AI-based systems minimize extraneous load through adaptive feedback and real-time error correction, clarifying misunderstandings and presenting information in digestible steps [20, 21]. Germane load refers to the effort invested in building and strengthening knowledge structures. Activities such as self-explanation or practice enhance retention and deeper understanding [22]. AI tools actively promote germane load by encouraging learners to reason through problems, offering interactive exercises, and tracking progress to reinforce conceptual connections. Bodonhelyi et al. [23] found that students using AI tools demonstrated stronger reasoning and skill retention compared to those relying solely on videos. Video tutorials, with less interaction, can still contribute to germane load when integrated into communal learning contexts where peers discuss and apply content together. This is also supported by the African philosophy of Ubuntu, which complements CLT by emphasizing communal learning and cognitive justice. Ubuntu's principle *I am because we are*, suggests that cognitive strain can be alleviated when learners collaborate, share strategies, and co-construct knowledge [21, 24]. AI tools designed to foster collaboration align with this worldview, reducing cognitive load while promoting equity and inclusion. Video tutorials, though more static, can

also support Ubuntu when embedded in peer-learning environments where students collectively interpret and apply material.

2 Literature review

2.1 Teaching and learning of machine learning content in higher education

Artificial intelligence (AI) tools are increasingly recognized as powerful enhancers of programming education, often outperforming traditional video tutorials in higher education. Omeh [25], For instance, conducted a quasi-experimental study with undergraduate computer science students. They compared an experimental group using Google Gemini within a problem-based learning framework to a control group relying on instructional videos. The AI-assisted group demonstrated statistically significant improvements in programming skills, with a large effect size. Similarly, Yilmaz and Yilmaz [26] reported that students who used ChatGPT during programming practice achieved higher computational thinking skills, programming self-efficacy, and motivation compared to peers taught through traditional methods. Supporting this trend, Alanazi et al. [27] argued that AI-assisted learning led to notably higher student performance scores, though improvements in ease of understanding were not statistically significant. In contrast, Mellado and Cubillos [28] found that while both AI tools (Microsoft Copilot) and video tutorials improved knowledge acquisition in web programming, video tutorials produced greater gains in knowledge and curiosity at the introductory level. The authors continued that the effectiveness of AI interventions appears to depend on learners' prior experience, digital literacy, and the complexity of course content. The evidence highlights the value of integrating both approaches: using video tutorials for foundational instruction and AI tools for advanced practice. This blended strategy may optimize learning outcomes by balancing clarity with adaptability. These findings point to a growing trend of leveraging AI-driven interventions to personalize programming education in higher education. Notably, however, no study has examined the comparative of AI and video instruction specifically within machine learning course content, a gap that future research should address. Is there a significant difference in AI and video instruction, specifically within machine learning course content? Based on this position, the study proposes the following hypothesis.

H1 There is a statistically significant difference in post-intervention machine learning skill scores between students taught with artificial intelligence tools (experimental group), and those taught with video tutorials (control group), after controlling for pretest scores?

2.2 Educational process of data modelling

Machine learning (ML) education that emphasises hands-on, data-centric, and collaborative approaches can substantially strengthen students' data modelling skills, a core competency in the field [3]. Kim et al. [29] shown by developing constructivist-oriented AI datasets for classroom use; students working with context-rich materials gained deeper insight into the ML learning process and reported more meaningful data-driven experiences compared to traditional resources. Also, Galatro and Amon [30] introduced real-world battery-ageing data into undergraduate engineering courses, enabling students to explore regression, variable interactions, and exploratory analysis. The result revealed an increase in modelling-related exam scores, alongside improved critical thinking and

interpretive abilities. Zhdanovskaya et al. [31] advanced this further with a project-based curriculum focused on data collection and labelling, where iterative competitions and industry engagement helped students appreciate the pivotal role of high-quality data in model performance. Similarly, Fleischer et al. [32] found that upper secondary students using Jupyter Notebooks to document the CRISP-DM cycle developed stronger technical application skills and greater awareness of bias and context. These studies show that immersive tasks such as dataset design, feature engineering, and collaborative modelling produce measurable gains. While pre/post assessments confirm statistically significant improvements, challenges remain: advanced competencies in machine learning competencies like data modelling, require instructional method support [33]. These findings in this review underscore that constructivist, hands-on approaches reliably enhance modelling abilities compared to traditional methods. It also provides guidance for wider adoption of AI-based tools in machine learning education. Is there a significant difference in post-intervention data modelling of AI and video instruction, specifically within machine learning course content? Based on these questions, the study proposes the following hypothesis.

H2 There is a statistically significant difference in post-intervention data modelling skill scores between the experimental and control groups when pre-intervention skills are taken into account.

2.3 Educational process of data evaluation

Model evaluation relies on measures such as accuracy, precision, recall, and related indicators to assess performance [34]. In recent years, data evaluation skills have gained increasing attention in machine learning education, as educators seek to compare the effectiveness of AI-driven instruction with traditional approaches [35, 36]. Data evaluation examining model accuracy, precision, recall, and generalisability is a critical competency for computer science students [37]. Students who master these skills can interpret performance metrics, identify model weaknesses, and make informed improvements. Kim et al. [29] implemented constructivist-oriented datasets in AI education and found statistically significant improvements across all measured sub-competencies, including data evaluation. Also, Tsai [38] argued that active learning strategies, structured and unstructured data analysis assignments, led to measurable gains in data analysis and critical thinking, as reflected in improved assessment scores over time. This is supported by [39], who also reported that hands-on workshops guiding undergraduates through the full machine learning workflow (data cleaning, model training, evaluation) significantly increased students' confidence and understanding of data evaluation processes. Additional studies have shown the role of digital evaluation rubrics and project-based curricula in enhancing student engagement and improving objective measures of problem-solving and critical thinking [30, 40]. These findings suggest that interventions emphasizing real-world data handling, iterative feedback, and active participation foster a deeper understanding of data evaluation. Is there a significant difference in post-intervention data evaluation of AI and video instruction, specifically within machine learning course content? Given these differences, this study proposes the following hypothesis:

H3 There is a statistically significant difference in post-intervention data evaluation skill scores between students in the experimental and control groups, after controlling for pre-intervention data evaluation skills.

2.4 Machine learning, academic achievement and the ability level of students

The effectiveness of instructional methods in machine learning education is increasingly examined through learner characteristics, particularly academic ability. A large-scale quasi-experimental study integrating deep learning techniques for personalized learning pathways found that students using an AI-driven adaptive platform achieved an improvement in grades and engagement compared to those receiving traditional instruction [29, 41]. Subgroup analyses revealed that high-ability students benefited most, showing greater gains in complex tasks such as data modelling and evaluation, while low-ability students improved more modestly but still outperformed their control counterparts [41]. More so, studies on AI-enabled adaptive learning systems reported medium-to-large positive effects on cognitive outcomes, with student classification level a proxy for ability acting as a significant moderator. Higher-ability learners experienced larger benefits from adaptive AI interventions than lower-ability peers [42]. In another study comparing AI-generated instructional videos with traditional recorded formats, overall achievement was similar across groups; however, the AI group demonstrated higher retention scores and reduced cognitive load, suggesting that students with stronger foundational skills may leverage AI tools more effectively for deeper learning [43]. Additional [40] showed that high-achieving students tend to prefer technology-enhanced and collaborative environments, further amplifying the positive effects of AI-based instruction on advanced learning competencies [44]. For low-ability learners, improvements are evident particularly in engagement and basic competencies, but the magnitude of gains is smaller. Video-based instruction remains valuable for foundational knowledge acquisition and may be preferable for novices due to its structured format and lower cognitive demands [43, 45]. Is there a significant difference between AI and video instruction, within the machine learning course content and the academic ability group? Based on this evidence, the study proposes the following hypothesis.

H4 The effect of instructional group (AI versus video) on the combined dependent variables (data modelling, data preprocessing, data model and data evaluation) there is a significantly moderated by students' academic ability levels (low ability versus high ability).

2.5 Impact of gender in machine learning education

The role of gender in moderating the effectiveness of instructional methods has been widely examined in educational technology research. A quasi-experimental study of educational robotics for pre-service teachers found that while males initially outperformed females on computational thinking (CT) pretests, both genders reached comparable high levels after the intervention, with the gender gap closing significantly post-intervention [46]. Similarly, a large-scale AI literacy intervention in elementary education showed that although male students scored higher on pretests, no significant gender differences remained after both groups completed an AI literacy course; notably, female students outperformed males in some domains [47]. In a study of conversational AI curricula for secondary students, revealed a significant interaction between

gender and learning approach: females using project-based approaches achieved the highest CT concept scores, outperforming all other groups [48]. Research with pre-schoolers integrating CT and AI literacy found that girls significantly outperformed boys post-intervention, despite no initial differences [49]. Other studies confirm these trends across contexts. Interventions such as problem-solving instructional approaches [40, 49], STEAM–PBL–AIoT courses [47], and collaborative or unplugged activities [50, 51] consistently show that gender gaps present at baseline can be narrowed or reversed through targeted instructional design. Project-based, hands-on, and collaborative approaches sometimes benefit female students more, leading to greater gains than those of their male peers, while traditional or gamified methods may favour males unless carefully adapted [52]. Importantly, when interventions are inclusive and scaffolded (e.g., integrating real-world problems or tangible tools), both genders achieve similar proficiency levels post-intervention [53, 54]. However, affective variables such as self-efficacy, engagement, and attitudes toward technology may still differ by gender even when cognitive outcomes converge [55]. Is there a significant difference between AI and video instruction, specifically within machine learning course content with respect to gender? Accordingly, the study proposes the following hypothesis.

H5 The effect of instructional group on combined posttest scores there is a significantly moderated by student gender (male versus female).

2.6 Student age, academic ability and machine learning education

AI-driven instruction provides adaptive feedback and personalised learning pathways, which are particularly beneficial for younger and low-ability learners who require structured support to master abstract machine learning concepts [56]. Older learners, however, often use AI systems to refine higher-order skills, such as evaluating model performance, as they possess more developed metacognitive strategies [57, 58]. A meta-analysis of AI interventions in education found that age group and educational level significantly moderate the impact of AI-based instruction, with young adult learners and higher education students showing the greatest gains in learning outcomes [59]. Factors such as baseline ability and digital literacy appear to play a larger role [60]. In K–12 contexts, systematic reviews highlight that project-based and constructivist ML curricula are effective across age groups, though the depth of conceptual understanding varies. Older students, particularly those in high school, tend to grasp advanced ML concepts more readily than younger peers in elementary or middle school, especially when instructional design is tailored to the developmental stage [61–63]. Research on adaptive AI-driven platforms in non-formal education further demonstrates that the type of intervention (AI vs. static content) is the strongest predictor of post-test performance, while baseline digital literacy also contributes significantly. These effects are observed across all age groups but appear slightly stronger for mid-career adults [64]. Studies using machine learning models to predict academic performance consistently show that prior achievement and study habits outweigh demographic variables such as age or gender in forecasting success [65–67]. Taken together, these findings suggest a nuanced picture. Instructional group (AI vs. traditional/video), academic ability, and age each influence ML learning outcomes, but their effects are not simply additive. Evidence points to interaction effects, for example, older or higher-ability students benefiting

more from adaptive or project-based ML instruction, though some studies report no significant direct effect of age when controlling for other variables. Is there a significant difference between AI and video instruction, specifically within machine learning course content, with respect to gender and ability group? The study proposes the following hypothesis.

H6 The three-way interaction between instructional group, students' academic ability, and age group there is a significantly affect the combined post-intervention outcomes, including data modelling, data preprocessing, and data evaluation skills.

3 Methods

3.1 Study design and population

This study adopted a quasi-experimental, non-equivalent pretest–posttest control group design to examine competencies in machine learning among computer and robotics education students. Because intact class groups were used based on existing course registration, true randomization was not possible. This reliance on intact groups is a defining feature of quasi-experimental research and, while it limits random assignment, it reflects the authentic instructional context in which the intervention occurred.

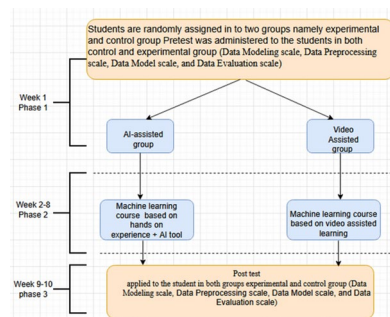
This study was conducted at a public university in a developing country, where a compulsory third-year course titled Introduction to Machine Learning is offered. This course, mandated by the National Universities Commission (NUC), provided a natural and authentic environment for implementing the instructional intervention. To ensure adequate sample, statistical power, a priori power analysis was conducted using G*Power 3.1. Assuming a medium effect size ($f = 0.25$), a significance level of 0.05, and a power of 0.80, the analysis indicated a minimum sample size of 52 participants for a $2 \times 2 \times 2$ factorial design. The final sample comprised 61 students, thereby exceeding the minimum requirement and ensuring sufficient power to detect both main and interaction effects.

Class A served as the control group (video-based instruction, $n = 36$), while Class B formed the experimental group (AI-based instruction, $n = 25$). The sample included 41 females and 20 males, aged between 16 and 25 years. All participants had completed the prerequisite course Data Science Application in Education (CRE 326). Stratification sampling technique was applied according to academic ability (low versus high) and age (≤ 20 years versus > 20 years), resulting in a $2 \times 2 \times 2$ factorial design (see Table 1). The difference in group sizes ($n = 25$ vs. $n = 36$) arose from practical constraints such as recruitment limitations and participant availability, rather than methodological oversight. Attrition and voluntary participation contributed to the imbalance, but it was not systematic or related to the variables under study, thereby minimizing the risk of bias.

Multivariate analysis of covariance was selected as the primary analytic technique because it is relatively robust to moderate inequality in group sizes, provided assumptions are met. Diagnostic checks confirmed the homogeneity of variance–covariance matrices, ensuring the validity of the results. Thus, while the imbalance is acknowledged, it reflects logistical realities of data collection and does not compromise the integrity or reliability of the analysis. This design allowed examination of both main effects and interactions among instructional strategy, academic ability, and age group. Stratification and pretest equivalence checks helped reduce potential selection bias. To further mitigate selection–regression threats inherent in quasi-experiments, pretest measures

Table 1 Participant demographics

Group	AI-based instruction	25
	video-based instruction	36
Gender	Female	41
	Male	20
Academic Ability	Low	
	High	
	AI-based instruction	11
	video-based instruction	15
	AI-based instruction	14
	video-based instruction	21
Age	≤ 20 years	
	AI-based instruction	7
	AI-based instruction	7
	video-based instruction	15
	>20 years	
	AI-based instruction	18
	video-based instruction	22

**Fig. 1** Flowchart of the instructional procedure

were incorporated to establish baseline equivalence, and statistical controls were applied through MANCOVA. Although quasi-experiments cannot eliminate all threats to internal validity, these procedures strengthen confidence in the findings by transparently addressing limitations and demonstrating that observed effects are attributable to the instructional intervention rather than pre-existing group differences.

3.2 Intervention procedure

The study intervention lasted 10 weeks and compared two instructional strategies. The experimental group received AI-based instruction supported by ChatGPT and Copilot, while the control group engaged with conventional video-based materials. Both groups covered identical curricular content aligned with the Nigerian Universities Commission (NUC) syllabus (see Fig. 1), but instructional methods were varied to isolate the effect of AI-supported project-based learning on machine learning competencies.

Each week, students participated in structured activities, including content exploration, practical exercises, and formative assessments. The experimental group interacted with AI tools to receive real-time guidance on tasks such as data preprocessing, model construction, and evaluation. AI support encouraged reflective thinking, scaffolded decision-making, and iterative refinement of solutions without replacing cognitive effort. The control group watched instructional videos that were carefully recorded by an expert and engaged in guided discussions to reinforce understanding. Weekly

assignments, machine learning logs, and code journals enabled monitoring of progress and facilitated feedback (See Online Appendix A). The intervention covered key domains including data collection, preprocessing, modelling, feature engineering, model evaluation, robotics applications, and ethical considerations. Capstone project development allowed students to integrate these skills in authentic tasks. Formative assessments ensured continuous engagement, and fidelity checks were maintained to ensure consistency across both groups. A summary of the activities is presented in Table 2 (Figs. 2 and 3).

3.2.1 Front end: user interaction

User (Student): Represents the learner engaging with the instructional platform. This layer aligns with H1–H6, where students are grouped by instructional type (AI vs. Video), gender, academic ability, and age. Their interaction with the system is the starting point for capturing pretest and posttest data across the four competency domains: Data Modelling, Data Preprocessing, and Data Model Data Evaluation.

Middle Layer: TensorFlow Platform

TensorFlow Platform: Acts as the instructional engine delivering AI-based content. **Dependencies (Other APIs):** Enable machine processing, such as: Automated feedback loops, Adaptive learning pathways, Real-time performance tracking, This layer operationalizes the AI instructional group, contrasting with the control (video-based) group. AI-based instruction here is hypothesised to outperform video instruction in post-intervention scores. TensorFlow's adaptability may benefit high-ability students more, showing moderation effects.

Back End: Artificial Tool

TensorBoard: Visualises learning data, including: Progress on each of the 40 competency items, Comparative analytics across groups, **Chat History:** Stores all student interactions, enabling: Qualitative analysis of engagement, Tracking of learning trajectories, **OpenAI Mode:** Generates instructional content, feedback, and explanations tailored to student needs. **Data Modelling OpenAI Mode** (generates schema design prompts, feature engineering tasks), **Data Preprocessing TensorFlow Platform + Dependencies** (automated pipelines, imputation, normalisation), **Data Model TensorFlow Platform** (algorithm selection, training, optimisation), **Data Evaluation, TensorBoard + Chat History** (metric visualisation, bias analysis, stakeholder feedback) (Fig. 4).

3.3 Instrument for data collection

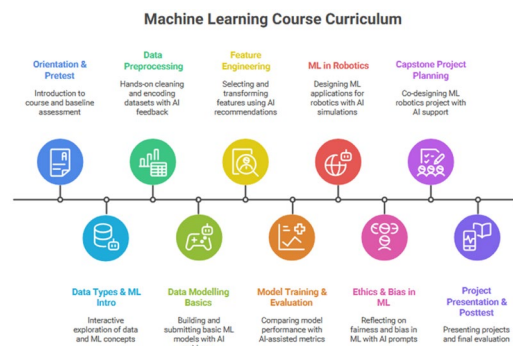
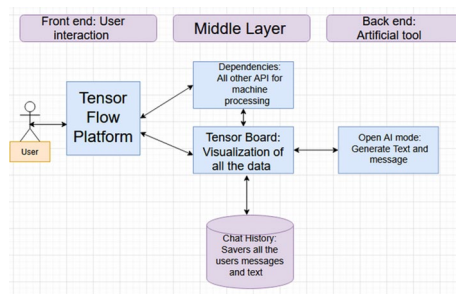
Data collection focused on five constructs: academic ability, data modelling, data preprocessing, data model construction, and data evaluation. Instruments were adapted from existing literature and validated by experts in science education and educational measurement. The Data Model comprised 10 items adapted from validated instruments on schema design, feature engineering, and scalability [68–72]. These items assessed students' competencies in structuring datasets for machine learning tasks, ensuring clarity, alignment with objectives, and proper handling of categorical and numerical variables. Sample items included: "I design schemas that capture complete relationships among data entities," and "I ensure scalability and flexibility of models to accommodate new sources while maintaining compliance with privacy standards." Data Preprocessing comprised 10 items adapted from validated instruments on data cleaning, feature selection,

Table 2 Summary of weekly activities for control and treatment groups in the machine learning instructional experiment

Week	Class activities	Control group (Video-based instruction)	Treatment group (AI-based instruction)	Assessment
1	Joint orientation session for all participants and administer a pretest to all participants	Watch academic video tutorials on ML fundamentals. The instructor-led overview of syllabus and expectations. Group discussion on prior ML knowledge. High-quality academic video tutorials (professionally produced, interactive segments, animations, and embedded quizzes) on ML fundamentals; instructor overview of syllabus; group discussion on prior ML knowledge	Structured introduction to ChatGPT as an AI-powered assistant. Guided demo on structuring raw inputs into meaningful formats. Interactive Q&A with AI on ML basics.	Quiz on ML fundamentals
2	Data Types and Collection, Introduction to Machine Learning	Interactive video modules explaining ML concepts, data types, and sources; instructor-led clarification; manual note-taking	AI chatbot introduces ML interactively Students explore applications via guided prompts AI generates examples of ML use cases for different industries	Data classification task and short reflective write-up.
3	Data Preprocessing	Tutorial Videos on cleaning, normalization, encoding Instructor-led walkthroughs with sample datasets Manual practice on preprocessing tasks.	Students clean and encode datasets with AI feedback AI suggests preprocessing strategies (e.g., imputation, scaling) Real-time error correction and explanation	Submit a cleaned dataset with a rationale for preprocessing choices.
4	Data Modelling Basics	Videos on regression, classification, clustering Instructor demonstrates manual model building in Python Students replicate models step-by-step	AI guides model selection and building Students receive real-time suggestions on algorithms AI explains trade-offs between models	Build and submit a basic ML model with annotated code.
5	Feature Engineering	Videos explain feature selection and transformation Manual application of techniques Instructor provides case study examples	AI recommends features and transformations based on dataset analysis Students test AI-suggested features AI explains impact of feature engineering on performance	Feature selection justification and model performance comparison.
6	Model Training and Evaluation	Videos on training/testing splits, accuracy, precision, recall Manual evaluation of models Instructor-led metric interpretation	AI assists with training/testing AI interprets metrics in real-time Students compare multiple models with AI guidance	Submit an evaluation report comparing two models.
7.	ML in Robotics	Videos on ML applications in robotics (vision, navigation) Case study analysis Instructor-led group discussion	AI simulates ML use cases in robotics Students design simple robotics applications with AI support AI provides feedback on feasibility	Design document for ML-enhanced robotics solution
8	Ethics and Bias in ML	Videos on fairness, bias, responsible AI Group debate on ethical dilemmas Instructor moderates discussion	AI prompts bias detection in datasets/models Students reflect on ethical challenges with AI guidance AI generates scenarios for ethical consideration	Written reflection on ethical challenges in ML.

Table 2 (continued)

Week	Class activities	Control group (Video-based instruction)	Treatment group (AI-based instruction)	Assessment
9.	Capstone Project Development	Plan ML-based robotics project using video references Instructor feedback on project scope Peer collaboration sessions	Co-design ML project with AI support AI refines data strategy and model selection AI generates draft project time-lines and objectives	Project proposal submission with time-line and objectives.
10.	Project Presentation and Review and Posttest	Present projects to peers and instructors Peer and instructor evaluation Posttest on ML concepts	Present projects with AI-generated visualizations and summaries AI assists in preparing presentation slides and reports AI provides feedback on oral defense preparation	Final project submission and oral defense

**Fig. 2** Course content for machine learning**Fig. 3** System architecture of the AI-assisted for machine learning courses proposed by this study

and automation [73–77]. These items assessed students' ability to handle missing values, detect outliers, normalize features, and balance datasets. Sample items included: "I apply imputation strategies to address missing data," and "I validate preprocessing outputs to ensure reproducibility and integrity of machine learning pipelines." Data Model comprised 10 items adapted from validated instruments on algorithm selection, hyperparameter tuning, and ensemble methods [78–82]. These items assessed students' competencies in selecting appropriate algorithms, preventing overfitting, and ensuring interpretability. Sample items included: "I optimise hyperparameters to improve model performance," and "I document model architecture and rationale to support transparency and reproducibility in machine learning applications." Data Evaluation (10 items).

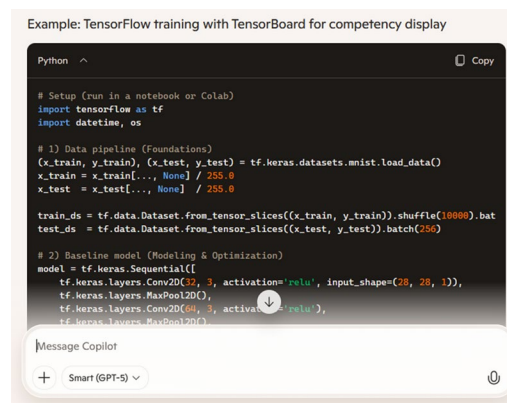


Fig. 4 Example: GPT API call parameters below:

Data Evaluation comprised 10 items adapted from validated instruments on performance metrics, fairness, and reproducibility [82–86]. These items assessed students' ability to evaluate models using multiple datasets, monitor drift, and analyse bias. Sample items included: "I use statistical significance testing to validate model performance," and "I ensure continuous evaluation post-deployment to maintain fairness, accountability, and reproducibility in AI systems."

Academic ability was assessed using an sixteen-item multiple-choice test, designed to measure logical reasoning and practical decision-making in machine learning (see Online Appendix B). Scenario-based questions simulated authentic ML challenges. A 50th percentile cut-off distinguished high- and low-ability students. The instrument demonstrated satisfactory internal consistency (KR-20=0.84). The academic ability test comprised 16 items adapted from validated instruments on conceptual learning and automated question generation [56, 87–89]. These items assessed students' applied knowledge of schema design, preprocessing, modelling, and evaluation. Sample items included: "Which element ensures referential integrity in schema design?" and "Which practice improves model interpretability for stakeholders?"

Data modelling, data preprocessing, data model construction, and data evaluation were measured using 10-item scales based on critical discourse indicators. Items were rated on a five-point scale from 1 (Not implemented) to 5 (Very good) (see Online Appendix C). Expert review and content validation using Lawshe's method ensured content relevance, with Content Validity Ratios ranging from 0.72 to 0.86 and overall CVIs between 0.74 and 0.84. Pilot testing confirmed internal consistency, with Cronbach's alpha values between 0.82 and 0.88. These instruments enabled the assessment of both theoretical understanding and practical application of machine learning concepts. For future studies, additional analyses of factorial and criterion validity could enhance psychometric rigour.

3.4 Data collection procedure

Data collection for this study was conducted over a structured ten-week period, organised into three distinct stages: the preparation phase, the intervention phase, and the assessment phase. Each phase is.

3.4.1 Preparation phase

The process began with an orientation session for all participants. During this session, the research objectives were clearly communicated, participant roles were defined, and expectations for the duration of the study were outlined. Ethical standards were upheld by obtaining informed consent from each participant, ensuring voluntary involvement and the confidentiality of all responses. Following the consent process, pre-test instruments were administered to participants in both the experimental and control groups. These instruments included the Data Modelling Rating Scale (DMRS), the Data Preprocessing Scale (DPS), the Data Model Scale (DMS), and the Data Evaluation Skill (DES). The pre-test served as a baseline measure of students' competencies before the instructional intervention. Participants were also briefed on the instructional delivery method assigned to their group—either the AI-based instruction or the video-based instruction. This orientation ensured that participants had a clear understanding of the instructional approach and helped align their expectations with the study's learning objectives.

3.4.2 Intervention phase

The intervention phase extended over ten consecutive weeks, during which students participated in structured weekly instructional sessions tailored to their assigned treatment condition. The experimental group engaged in project-based learning supported by Copilot and ChatGPT (artificial intelligence tools) designed to provide real-time assistance in machine learning tasks, including data modelling, data preprocessing, data model construction, and data evaluation. Each session promoted active exploration, allowing students to interact with the AI, reflect on its suggestions, and iteratively refine their code.

The role of Copilot was not to replace cognitive effort but to scaffold student thinking, simulate responses, and stimulate metacognitive reflection through prompt-driven engagement. In contrast, the control group received curated video-based instruction covering equivalent machine learning concepts. These instructional videos were recorded by expert in ML, offering step-by-step demonstrations of key constructs such as data modeling, data preprocessing, data model development, and data evaluation. Following each viewing session, students engaged in guided discussions with peers and instructors to deepen their understanding and apply the concepts to classroom-based machine learning tasks. Both groups completed structured weekly assignments aligned with the targeted machine learning domains. Throughout the intervention, all students maintained machine learning logs and code journals to document their progress. These artefacts were submitted weekly and used to provide formative feedback, supporting continuous improvement and reflective learning.

3.4.3 Post-intervention phase

The final week of the study was devoted to post-intervention assessment. We re-administered the same set of instruments used during the preparation phase to measure changes in students' competencies in data modeling, data preprocessing, data model construction, and data evaluation. These post-tests were conducted under standardised conditions to minimise measurement error and ensure consistency across both instructional groups. In addition to the post-test assessments, all participants submitted a culminating machine learning project that synthesised the various learning domains addressed

throughout the intervention. These projects were evaluated using the Data Modeling Rating Scale (DMRS), enabling us to assess both cognitive development and the practical application of machine learning skills in authentic scenarios. Evaluation criteria included co-design of machine learning projects with AI support, refinement of data strategy and model architecture, and written reflections on ethical considerations in machine learning. Throughout the data collection period, we closely monitored student participation, engagement, and attendance to ensure equitable access to instructional resources. The research team responded promptly to any technical or instructional challenges to minimise disruptions. This comprehensive data collection approach enabled us to capture both the immediate and sustained effects of the AI-based and video-based instructional strategies on students' learning outcomes and performance trajectories.

Fidelity of implementation was carefully monitored to ensure consistency across both instructional modes throughout the 10 weeks. For the control group, recorded video tutorials were standardised: all students accessed the same professionally produced modules, which included embedded quizzes and animations, and instructors followed a uniform teaching script to avoid variation. For the treatment group, AI interactions were structured rather than free-form. Students engaged with ChatGPT and copilot through guided prompts, weekly task sheets, and instructor-supervised sessions, ensuring comparable exposure and minimizing uncontrolled exploration. Fidelity checks included weekly observation logs, instructor debriefs, and student activity records to confirm adherence to planned procedures. These measures ensured that both groups received consistent, high-quality instruction aligned with the study design, thereby reducing the risk of confounding variables and strengthening the validity of the findings.

4 Results

4.1 Descriptive statistics

Table 3 presents the mean and standard deviation scores for the experimental group (AI-based instruction) and the control group (Video-based instruction) across the four post-intervention machine learning skills. On average, students who received AI-based instruction achieved higher mean scores in data modelling, Data Preprocessing, and data evaluation than those taught using video-based instruction. For data model skill, the mean scores for both groups were relatively close (Figs. 5, 6 and 7).

4.2 Hypothesis 1: Effect of instructional method on combined machine learning skills

A multivariate analysis of covariance (MANCOVA) was conducted with four dependent variables (data modelling, Data Preprocessing, data model, and data evaluation) and four corresponding pretest scores as covariates. The assumptions of MANCOVA were

Table 3 Descriptive statistics of post-intervention skill scores by treatment group ($N=61$)

Skill area	Treatment	Mean	SD	<i>n</i>
Data modelling	AI-based	44.2	4.66	25
	Video-based	36.39	5.49	36
Data preprocessing	AI-based	46.12	3.59	25
	Video-based	39.03	6.59	36
Data model	AI-based	37.64	6.63	25
	Video-based	36.06	6.31	36
Data evaluation	AI-based	43.4	3.11	25
	Video-based	34.64	5.31	36

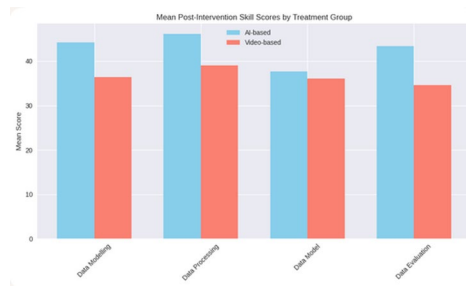


Fig. 5 Post-intervention skill scores by treatment group

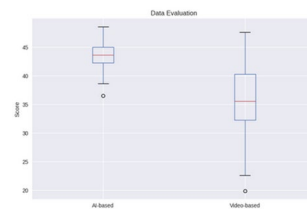


Fig. 6 Data evaluation at post test between AI-based and video based



Fig. 7 Data model at post test between AI-based and video based

Table 4 Multivariate tests for the effect of instructional treatment on combined posttest skill scores (Covariates: Pretest Scores)

Test statistic	Value	F	Hypothesis df	Error df	p	η_p^2
Pillai's trace	0.602	19.68	4	52	< 0.001	0.602

examined prior to analysis. Box's test of equality of covariance matrices was marginally significant, *Box's M* = 20.04, $F(10, 12480.99) = 1.85$, $p = 0.047$, suggesting a slight departure from equality. However, Pillai's Trace was adopted for interpretation because it is considered more robust to such minor violations [90, 91]. Levene's tests for homogeneity of variance were not significant for any of the dependent variables: data modeling ($F(1, 59) = 0.01$, $p = 0.940$), Data Preprocessing ($F(1, 59) = 1.68$, $p = 0.200$), data model ($F(1, 59) = 0.05$, $p = 0.833$), and data evaluation ($F(1, 59) = 0.18$, $p = 0.672$). These results confirm that the assumption of equal variances across treatment groups was met. The multivariate result showed a statistically significant effect of treatment on the combined dependent variables, *Pillai's Trace* = 0.602, $F(4, 52) = 19.68$, $p < 0.001$, partial $\eta^2 = 0.602$ (See Table 4). This finding means that the overall post-intervention skill performance of

Table 5 Univariate tests of between-subjects effects for treatment

Dependent variable	F	df	p	Partial η^2
Data modelling	31.62	1, 55	< 0.001	0.365
Data preprocessing	28.48	1, 55	< 0.001	0.341
Data model	1.31	1, 55	0.258	0.023
Data evaluation	46.99	1, 55	< 0.001	0.461

Covariates included pretest scores for each respective skill

Table 6 Estimated marginal means of post-test scores by instructional group (Controlling for pretests)

Skill area	AI-based instruction	Video-based instruction	Mean difference	p
Data modelling	45.8	35.28	10.53	< 0.001
Data preprocessing	48.48	37.39	11.1	< 0.001
Data model	35.14	37.79	-2.65	0.258
Data evaluation	44.92	33.58	11.34	< 0.001

Covariates were evaluated at mean pretest scores

students significantly differed based on the instructional method used, after adjusting for their pretest scores.

Follow-up univariate analyses of covariance were conducted to determine the effect of instructional treatment on each dependent variable while controlling for the corresponding pretest scores. Table 5 presents the results of these analyses. As shown in Table 5, there were significant treatment effects on data modelling, data preprocessing, and data evaluation posttest scores after controlling for pretest differences. The partial eta squared values ranged from 0.341 to 0.461, indicating large effect sizes according to Cohen's (1988) guidelines. However, the effect of treatment on data model skill was not statistically significant, $F(1, 55) = 1.31$, $p = 0.26$, $\eta^2 = 0.023$. These findings indicate that students exposed to AI-based instruction performed significantly better in data modelling, Data Preprocessing, and data evaluation than those taught through video-based instruction.

The adjusted post-test means in Table 6 further illustrate this pattern. Students in the AI-based learning condition achieved higher mean scores in data modelling, Data Preprocessing, and data evaluation, compared with those in the video-based condition. Although the mean score for data model was slightly higher in the video-based group than in the AI-based group, the difference was not statistically significant. The difference in data model skill was not statistically significant. These results provide strong evidence that AI-based instruction significantly enhances learners' overall performance in machine learning-related skills when compared to traditional video-based instruction. Thus, Hypothesis 1 was supported.

4.3 Hypothesis 2 and 3: Effect of instructional method on data modelling skills

A multivariate analysis of covariance (MANCOVA) was conducted to examine whether there were statistically significant differences in students' post-intervention data modelling-related skills between those taught using an AI-based approach and those taught using video-based instruction, after controlling for pretest scores in data modelling, Data Preprocessing, and data model skills.

Preliminary analyses were carried out to verify that the assumptions of MANCOVA were not violated. Box's test of equality of covariance matrices was not significant, Box's

Table 7 Tests of between-subjects effects for data modelling-related skills (Controlling for pretests)

Dependent variable	F(1, 56)	p	Partial η^2
Data modelling	32.19	< 0.001	0.365
Data preprocessing	22.58	< 0.001	0.287
Data model	0.40	0.529	0.007

Covariates included pretest scores for each dependent variable

Table 8 Estimated marginal means of post-test scores by instructional method (Controlling for pretests)

Dependent variable	AI-based instruction	Video-based instruction	Mean difference	95% CI	p
Data modelling	44.02	36.51	7.51	[4.86, 10.17]	< 0.001
Data preprocessing	46.13	39.02	7.11	[4.11, 10.10]	< 0.001
Data model	37.32	36.28	1.04	[-2.25, 4.33]	0.529

Covariates were evaluated at mean pretest scores

$M = 12.10$, $F(6, 18,076.80) = 1.90$, $p = 0.077$, indicating that the assumption of homogeneity of covariance matrices was satisfied. Levene's test of equality of error variances was also not significant for data modelling, $F(1, 59) = 0.42$, $p = 0.518$, and for data model, $F(1, 59) = 0.05$, $p = 0.829$. Although the test for Data Preprocessing was marginally significant, $F(1, 59) = 4.34$, $p = 0.042$, the violation was minor, and MANCOVA is generally robust against such mild deviations (Tabachnick & Fidell, 2019). Therefore, the assumption of equal variances was considered acceptable.

The multivariate test using Pillai's Trace indicated a significant main effect of treatment on the combined dependent variables, *Pillai's Trace* = 0.402, $F(3, 54) = 12.10$, $p < 0.001$, $\eta^2 = 0.402$. This finding suggests that the instructional method (AI-based versus video-based) had a statistically significant multivariate effect on students' post-intervention performance in the three data modelling-related skills, after controlling for pretest differences. Follow-up univariate ANCOVAs were conducted to examine the individual dependent variables contributing to the significant multivariate effect (Table 7). The results revealed that the treatment effect was significant for data modelling, $F(1, 56) = 32.19$, $p < 0.001$, $\eta^2 = 0.365$, and for Data Preprocessing, $F(1, 56) = 22.58$, $p < 0.001$, $\eta^2 = 0.287$. However, the effect on data model skill was not statistically significant, $F(1, 56) = 0.40$, $p = 0.529$, $\eta^2 = 0.007$. These findings indicate that AI-based instruction led to significantly higher achievement in data modelling and Data Preprocessing skills, compared with the video-based method, but not in data model skill.

As shown in Table 8, after adjusting for pretest differences, students in the AI-based group achieved higher mean posttest scores in data modelling and Data Preprocessing compared with those in the video-based group. The adjusted mean difference for data modelling was 7.51 and for Data Preprocessing was 7.11, both favouring the AI-based group. For data model, the adjusted mean difference (1.04) was not significant. These results provide partial support for Hypothesis 2, indicating that the use of AI-based instructional strategies significantly improved students' data modelling and Data Preprocessing skills compared to video-based instruction, but did not significantly influence data model skill.

4.3.1 Effect of instructional method on data evaluation skill

A one-way analysis of covariance was conducted to determine whether instructional method influenced students' post-test performance in data evaluation after controlling for pre-test scores. Descriptive statistics indicated that students exposed to the AI-based learning method obtained higher post-test scores than those taught using the video-based method. The Levene test of equality of error variances was not significant, $F(1, 59) = 1.43$, $p = 0.236$, confirming that the assumption of homogeneity of variance was satisfied.

Table 9 indicates that the overall model was significant, $F(2, 58) = 28.58$, $p < 0.001$, $R^2 = 0.496$, suggesting that approximately 49.6 per cent of the variance in post-test scores was accounted for by the combined effects of the covariate and treatment. The pre-test covariate did not significantly predict post-test performance, $F(1, 58) = 1.73$, $p = 0.193$, partial $\eta^2 = 0.029$. However, a significant main effect of treatment was observed, $F(1, 58) = 51.24$, $p < 0.001$, partial $\eta^2 = 0.469$, representing a large effect size. After adjusting for pre-test performance, the estimated marginal mean for the AI-based group was 43.90, while that for the video-based group was 34.29. The Bonferroni-adjusted comparison indicated that the adjusted mean difference of 9.61 was statistically significant. These findings support Hypothesis 3, which stated that students taught using AI-based learning would outperform those taught using the video-based method in data-evaluation performance, controlling for pre-test scores.

4.4 Multivariate analysis of covariance for moderation effects (H4–H6)

A three-way multivariate analysis of covariance (MANCOVA) was conducted to examine the effects of instructional group (AI vs. Video), students' academic ability (Low vs. High), and gender (Male vs. Female) on the combined posttest scores in data modeling, Data Preprocessing, data model, and data evaluation, while controlling for pre-test scores. Assumptions for the analysis were met. Box's test of equality of covariance matrices was not significant, Box's $M = 25.081$, $F = 0.698$, $p = 0.890$, indicating that the observed covariance matrices were equivalent across groups. Levene's tests for homogeneity of error variances were all non-significant, supporting the assumption of equal variances across groups.

Multivariate tests indicated that instructional group significantly affected the combined posttest outcomes, Pillai's Trace = 0.396, $F(4, 47) = 7.715$, $p < 0.001$, partial $\eta^2 = 0.396$. Students' ability approached multivariate significance, Pillai's Trace = 0.170, $F(4, 47) = 2.414$, $p = 0.062$. Gender, the Treatment $\times$ Gender interaction, and the three-way interaction of Treatment $\times$ Ability $\times$ Gender were not significant (all $p > 0.46$). The Treatment $\times$ Ability interaction approached significance at the multivariate level, Pillai's

Table 9 Analysis of covariance for data evaluation post-test ($N = 61$)

Source	SS	df	F	p	η_p^2
Corrected model	1 167.87	2	28.583	< 0.001	0.496
Intercept	4 340.09	1	212.441	< 0.001	0.786
Data evaluation pre-test	35.389	1	1.732	0.193	0.029
Treatment	1 046.72	1	51.235	< 0.001	0.469
Error	1 184.92	58			
Total	91 504.00	61			
Corrected total	2 352.79	60			

$R^2 = 0.496$; adjusted $R^2 = 0.479$

Trace = 0.155, $F(4, 47) = 2.150$, $p = 0.089$, suggesting a potential moderating effect of academic ability.

The estimated marginal means for each combination of treatment and ability are presented in Table 10. The results indicate that students in the AI group consistently achieved higher posttest scores compared with those in the video group. Among low-ability students, the AI group outperformed the video group in data modelling, data preprocessing, and data evaluation. High-ability students in the AI group also scored higher than their video group counterparts in data preprocessing and data evaluation, although differences in data model were minimal.

The univariate tests of between-subjects effects are presented in Table 11. Instructional group had significant effects on data modelling, data preprocessing, and data evaluation posttests. Academic ability significantly influenced data modelling and data preprocessing, and the Treatment $\times$ Ability interaction reached significance for Data Preprocessing only. Neither Treatment $\times$ Gender nor the three-way interaction Treatment $\times$ Ability $\times$ Gender was significant across any dependent variable.

Summarily, hypothesis four (H4) proposed that students' academic ability would significantly moderate the effect of instructional group on the combined posttest scores. The results partially support this hypothesis. Specifically, the interaction between instructional group and ability reached significance for Data Preprocessing posttest scores ($F(1, 50) = 8.042$, $p = 0.007$, $\eta^2 = 0.139$), indicating that the advantage of AI over Video instruction was more pronounced for students with differing ability levels in this particular skill. However, this interaction did not achieve statistical significance for Data Modelling, Data Model, or Data Evaluation posttest scores, suggesting that the moderating effect of ability was limited to Data Preprocessing.

Also, hypothesis five (H5) posited that gender would moderate the effect of instructional group on posttest outcomes. The results did not support this hypothesis. Across all dependent variables, data modelling, Data Preprocessing, data model, and data evaluation, the interaction between treatment and gender was non-significant (all $p > 0.05$),

Table 10 Estimated marginal means for posttest scores by ability and treatment

Dependent variable	Ability	Treatment	Mean	SE	95% CI (U, L)
Data modelling posttest	Low	PBL AI	46.1	1.675	42.74, 49.47
		PBL Video	38.34	1.77	34.79, 41.9
	High	PBL AI	44.04	2.695	38.62, 49.45
		PBL Video	32.81	1.529	29.74, 35.88
Data preprocessing posttest	Low	PBL AI	48.03	1.731	44.55, 51.5
		PBL Video	42.1	1.829	38.42, 45.77
	High	PBL AI	47.88	2.785	42.28, 53.47
		PBL Video	33.93	1.58	30.75, 37.1
Data model posttest	Low	PBL AI	34.64	2.146	30.33, 38.95
		PBL Video	39.02	2.268	34.47, 43.58
	High	PBL AI	38.13	3.453	31.2, 45.07
		PBL Video	39.18	1.959	35.24, 43.11
Data evaluation posttest	Low	PBL AI	44.82	1.572	41.66, 47.98
		PBL Video	32.74	1.661	29.40, 36.08
	High	PBL AI	46.1	2.53	41.01, 51.18
		PBL Video	33.43	1.435	30.54, 36.31

Covariates evaluated at pretest means: Data Modelling = 14.97, Data Preprocessing = 14.59, Data Model = 14.67, Data Evaluation = 19.15

Table 11 Tests of between-subjects effects

Outcome	Source	F	p	η_p^2
Data modelling posttest	Treatment	10.257	0.002	0.170
	Ability	5.254	0.026	0.095
	Treatment $\times$ Ability	2.441	0.124	0.047
	Treatment $\times$ Gender	0.026	0.873	0.001
	Treatment $\times$ Ability $\times$ Gender	2.320	0.134	0.044
Data preprocessing posttest	Treatment	11.585	0.001	0.188
	Ability	7.946	0.007	0.137
	Treatment $\times$ Ability	8.042	0.007	0.139
	Treatment $\times$ Gender	0.032	0.859	0.001
	Treatment $\times$ Ability $\times$ Gender	0.617	0.436	0.012
Data model posttest	Treatment	0.067	0.798	0.001
	Ability	0.010	0.920	0.000
	Treatment $\times$ Ability	0.056	0.815	0.001
	Treatment $\times$ Gender	0.419	0.521	0.008
	Treatment $\times$ Ability $\times$ Gender	0.110	0.742	0.002
Data evaluation posttest	Treatment	21.072	0.000	0.296
	Ability	0.159	0.691	0.003
	Treatment $\times$ Ability	0.006	0.939	0.000
	Treatment $\times$ Gender	0.250	0.619	0.005
	Treatment $\times$ Ability $\times$ Gender	0.019	0.892	0.000

indicating that male and female students responded similarly to AI and Video instructional approaches.

Moreover, hypothesis six (H6) suggested a three-way interaction among instructional group, academic ability, and gender on posttest outcomes. The analysis did not support this hypothesis. The three-way interaction term was non-significant across all posttest measures (all $p > 0.05$), implying that the combined influence of ability and gender did not differentially affect the efficacy of instructional treatment. Generally, the findings indicate that the instructional group significantly influenced student performance, with academic ability acting as a moderator only for Data Preprocessing outcomes. Gender did not moderate the effect of treatment, and no evidence was found for a three-way interaction between instructional group, ability, and gender.

5 Discussion of findings

This study was conceived to examine the effects of instructional method on students' acquisition of machine learning skills within Nigerian higher education, focusing on the comparative influence of artificial intelligence (AI)-based instruction and video-based instruction. The findings indicate that instructional method influenced students' combined machine learning skills, with learners exposed to AI-based instruction performing better across several skill domains. Rather than simply indicating differences in performance, this suggests that the structure of the learning environment plays a central role in shaping how students engage with and process complex tasks. According to Cognitive Load Theory, such adaptive systems can reduce unnecessary mental strain and allow students to concentrate on essential concepts, thereby improving cognitive efficiency and retention.

These results align with earlier findings by [21, 92] that AI-supported learning environments promote deeper engagement and improved outcomes through personalised content delivery. The advantage observed for the AI-based group may therefore

be interpreted as a function of increased interaction and feedback loops rather than the mere presence of technology. While this suggests that interactive tools can help address instructional constraints such as overcrowded classrooms and limited instructor attention in Nigerian universities, this implication should be understood within the specific institutional conditions under which the study was conducted. By contrast, video-based instruction, although accessible and scalable, provides limited interactivity, which may contribute to comparatively lower performance. This indicates that access alone is insufficient; the nature of learner engagement is a critical determinant of learning outcomes. This outcome is consistent with the observations of [4, 93] who associated passive video instruction with lower levels of conceptual understanding.

The second finding revealed that AI-based instruction significantly improved data modelling and data preprocessing skills, but not data model skills. This suggests that the instructional advantage of AI-based learning may not extend uniformly across all machine-learning competencies. Several explanations may account for the non-significant finding for data model skill. First, the assessment instrument may have had limited sensitivity or construct validity for measuring this competency, which could have reduced its ability to detect group differences. Second, a ceiling effect may have occurred if the data model task was insufficiently challenging for participants, thereby compressing score variability. Third, the AI tools used in the intervention may have been less well aligned with the instructional demands of data model construction than with more procedural tasks such as preprocessing and evaluation. Finally, data model skills may require additional conceptual scaffolding or instructional strategies beyond those embedded in the present AI-based environment. Although Cognitive Load Theory may help explain why abstract representational tasks are more cognitively demanding, this interpretation remains tentative because cognitive load was not directly measured in the study. Therefore, the null finding should be interpreted cautiously and warrants further investigation. These findings agree with [41, 94] who noted that AI-enhanced interaction improves procedural learning but may not automatically extend to abstract reasoning without deliberate instructional design. From the perspective of Cognitive Load Theory, the weaker effect on data model skill could reflect the high intrinsic load associated with abstract visualisation tasks, which may not have been sufficiently supported through AI scaffolding. This implies that instructional effectiveness depends not only on the presence of AI tools but also on how well these tools are aligned with the cognitive demands of specific learning tasks. Within the Nigerian higher education environment, this outcome indicates that while AI integration can strengthen applied competencies, its effectiveness may depend on curriculum alignment and instructional design quality.

The third hypothesis, which proposed that students taught using AI-based instruction would outperform those using video-based instruction in data evaluation skills, received strong support. This finding suggests that AI-based environments may enhance learners' evaluative judgment by enabling iterative engagement with feedback and error correction processes. This result aligns with the observations of [35, 36] who found that AI-driven feedback improves learners' capacity for analytical reasoning. In practical terms, this indicates that learning is strengthened when students are able to test, revise, and refine their understanding continuously rather than passively receiving information. In contexts where opportunities for applied experimentation are limited [95], such feedback mechanisms may be particularly important. This supports [17] argument

that reducing cognitive overload through timely guidance enhances understanding and retention. However, the strength of this effect may vary depending on the availability of technological infrastructure and instructional support.

The findings further revealed that academic ability moderated the effect of instructional method on Data Preprocessing performance, confirming partial support for the fourth hypothesis. It was observed that AI-based instruction appeared to benefit low-ability learners more strongly. This suggests that adaptive systems may function as compensatory mechanisms that reduce disparities in prior knowledge by adjusting instructional support. High-ability students also performed well but showed a comparatively smaller advantage, indicating that the marginal benefit of AI may decrease as prior competence increases. This trend agrees with [10, 57] who noted that AI-enhanced instruction can reduce performance gaps. This finding highlights the role of adaptive instruction in redistributing learning gains rather than uniformly increasing performance across all groups. In the Nigerian context, where student preparedness varies considerably, such systems may help reduce inequities in learning outcomes, although these effects may differ across subject areas.

Contrary to the fifth hypothesis, gender did not significantly moderate the relationship between instructional method and post-test outcomes. Both male and female students appeared to benefit similarly from AI-based instruction. This suggests that when learning environments are structured to provide equal access and adaptive support, gender-related differences in performance may be reduced. This finding is consistent with [45, 47]. However, this should not be interpreted as evidence that gender disparities no longer exist, but rather that their influence may be context-dependent. Moreover, the study design did not account for prior exposure to technology, socio-economic background, or cultural factors that often shape gendered experiences of education. Thus, the result should not be taken as evidence that gender disparities in learning outcomes have been eliminated. Instead, it highlights the importance of considering contextual influences when evaluating inclusivity claims. In the Nigerian setting, this may indicate progress toward more equitable learning opportunities, but broader generalisation requires careful examination of alternative explanations and replication across diverse populations. The sixth hypothesis proposed a three-way interaction among instructional group, academic ability, and gender on post-intervention outcomes. The results did not support this hypothesis. This indicates that the combined influence of these variables does not necessarily produce additive or interactive effects when adaptive instructional systems are in use. Similar conclusions were reported by [96, 97], who found that learning outcomes in AI-supported environments are more strongly associated with individual ability than demographic factors. This may suggest that well-designed instructional systems can reduce the explanatory power of certain learner characteristics, although such conclusions should be interpreted within the limits of the study design.

6 Conclusion

The aim of this study was to determine whether adaptive AI learning environments improve students' cognitive and practical skills more effectively than standard video tutorials. The findings indicate that AI-based instruction improved student performance in data modelling, Data preprocessing, and data evaluation. Learners who engaged with AI systems benefited from continuous feedback, personalised guidance, and interactive

problem-solving activities, which support both comprehension and practical application. The study also found that AI instruction was particularly beneficial for students with lower academic ability, as the system adjusted support according to individual needs, reducing differences in performance. Gender had no significant effect on learning outcomes, indicating similar responses to the instructional approaches. Despite these gains, improvement in conceptual model construction was limited. The study contributes to knowledge on technology-enhanced learning by showing that adaptive AI tools can improve specific aspects of teaching and learning in higher education. It supports the relevance of Cognitive Load Theory, as structured feedback and guidance may reduce mental overload and enhance retention. The findings have implications for curriculum design and teaching practice. Higher education institutions may consider integrating AI tools into science and technology programmes, while policymakers may support this through investment in digital infrastructure and professional development. Although the study was conducted within Nigerian higher education, the findings may inform similar contexts where instructional constraints, learner diversity, and limited access to personalised support affect learning outcomes. In such settings, adaptive AI-based approaches may be considered as complementary tools to enhance engagement and skill acquisition, provided that implementation is aligned with local curriculum needs and resource availability.

The findings extend Cognitive Load Theory (CLT) by demonstrating how adaptive AI systems can actively manage cognitive demands in real-time. CLT emphasises that learning is most effective when extraneous cognitive load is minimized and germane load is optimized. AI-based instruction achieves this by tailoring content to individual learners, breaking complex material into manageable steps, and providing immediate feedback. This dynamic adjustment reduces frustration and cognitive overload, enabling students to focus on essential concepts. Importantly, the study highlights that AI tools not only reduce unnecessary effort but also foster inclusivity by accommodating diverse learning speeds and abilities. Thus, AI integration enriches CLT by showing how technology can operationalize its principles in large, heterogeneous classrooms.

For teachers and universities, the results suggest several actionable strategies. First, institutions should embed AI-assisted platforms into computer science and related curricula to strengthen students' analytical and digital competencies. Exposure to AI tools builds confidence in handling modern computing systems and prepares learners for careers in data-driven industries. Second, educators must receive targeted training to use AI effectively. Professional development should address both technical skills and pedagogical strategies, ensuring teachers can balance challenge and support, provide meaningful feedback, and encourage active engagement rather than passive memorization. Third, AI instruction should be complemented with collaborative activities such as group discussions, guided practice, and project-based learning to address areas where AI alone is less effective, such as data model construction. This blended approach ensures a more holistic learning experience. At the policy level, the study underscores the need for systemic support to scale AI-based learning in Nigerian universities. Government agencies and the National Universities Commission (NUC) should invest in infrastructure, including reliable internet access, modern computer facilities, and funding for digital innovation projects. Policies should encourage collaboration between universities, technology providers, and government bodies to ensure AI systems are tailored to local

educational contexts. Furthermore, national teacher training initiatives should incorporate AI pedagogy, equipping educators with the skills to integrate these tools effectively. By aligning infrastructure, curriculum, and professional development, policymakers can create an enabling environment where AI-based instruction thrives.

6.1 Limitations and questions for future research

This study is limited by its quasi-experimental design, small sample size, and the relatively vague description of the intervention, which constrain the generalizability of the findings. To strengthen future work, researchers should adopt true randomized experiments, expand participant pools, and employ longitudinal designs to examine whether AI-supported learning in machine learning (ML) skill development endures over time. Future research may consider, does the advantage of AI-based instruction persist in capstone projects, internships, or employment months after the intervention? Why did AI instruction fail to improve data model construction skills? Are current AI tools insufficient, or do abstract tasks require alternative pedagogical scaffolds? How do outcomes differ across AI platforms (e.g., ChatGPT, Copilot, Gemini, Claude) with varying feedback mechanisms? What specific processes, such as immediate feedback, error reduction, example generation, or metacognitive reflection, drive AI's effectiveness? Mixed-methods designs could unpack these dynamics. What infrastructural and professional development supports are needed to integrate AI instruction across diverse cultural and institutional contexts?

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s44163-026-01707-0>.

Supplementary material 1

Acknowledgements

The authors wish to thank the students who participated in the study for their commitment and also the department of Computer and Robotics Education for their support during this study.

Author contributions

C.B Omeh and V.J Owan conceptualized and analysis. T. Sanusi and I. Olade data collection and editing.

Funding

None of the authors received any funding.

Data availability

the data used to support the findings of this study are available upon request to the corresponding author.

Declarations

Ethics approval and consent to participate

The author prioritized ethical integrity throughout the study. Approval was obtained from the Faculty of Vocational and Technical Education Research Ethics Committee, University of Nigeria under approval number (Ref: UNX/PG/PD/2024/95701). All procedures involving human participants complied with institutional ethical standards and the 1964 Helsinki Declaration and its later amendments. Informed consent was obtained in written form through an online checkbox that affirmed participants' understanding of the study's purpose. Participation was entirely voluntary, and no personal identifiers were collected, ensuring full anonymity and confidentiality of responses. voluntary participation, and their right to withdraw at any time without consequence.

Consent to publish

All authors have reviewed the final manuscript and consented to its submission and publication in its current form.

Competing interests

The authors declare no competing interests.

Received: 24 January 2026 / Accepted: 23 June 2026

Published online: 12 July 2026

References

1. Rizvi SAH. The economic impact of artificial intelligence on global productivity and labour markets: a 2025 perspective. SSRN. 2025. <https://doi.org/10.2139/ssrn.5431775>.
2. Aderibigbe AO, Ohenhen PE, Nwaobia NK, Gidiagba JO, Ani EC. Artificial intelligence in developing countries: bridging the gap between potential and implementation. *Comput Sci IT Res J*. 2023;4(3):185–99. <https://doi.org/10.51594/csitrj.v4i3.629>.
3. Sweller J, Van Merriënboer JJ, Paas F. Cognitive architecture and instructional design: 20 years later. *Educ Psychol Rev*. 2019;31(2):261–92.
4. Martin NA, Martin R. Would you watch it? Creating effective and engaging video tutorials. *J Libr Inform Serv Distance Learn*. 2015;9(1–2):40–56. <https://doi.org/10.1080/1533290X.2014.946345>.
5. Basil OC. Cloud computing and undergraduate researches in universities in Enugu State: implication for skills demand. *Int J Instr Technol Educ Stud*. 2022;3(2):27–33.
6. Abisoye A, Akerele JJ, Odio PE, Collins A, Babatunde GO, Mustapha SD. Using AI and machine learning to predict and mitigate cybersecurity risks in critical infrastructure. *Int J Eng Res Dev*. 2025;21(2):205–24.
7. Banerjee A, Kabadi S, Karimov D. The transformative power of AI: projected impacts on the global economy by 2030. *Rev Artif Intell Educ*. 2023;4:e20–20. <https://doi.org/10.37497/rev.artif.intell.educ.v4i00.20>.
8. Omeh CB, Olelewe CJ. Assessing the effectiveness of innovative pedagogy and lecture method on students' academic achievement and retention in computer programming. *Educ Res Int*. 2021;2021(1):5611033. <https://doi.org/10.1155/2021/5611033>.
9. Fan O, Xu W. The effects of educational robotics in STEM education: a multilevel meta-analysis. *Int J STEM Educ*. 2024;11. <https://doi.org/10.1186/s40594-024-00469-4>.
10. Wu D, Chen M, Chen X, Liu X. Analyzing K-12 AI education: a large language model study of classroom instruction on learning theories, pedagogy, tools, and AI literacy. *Comput Educ Artif Intell*. 2024;7:100295. <https://doi.org/10.1016/j.caeai.2024.100295>.
11. Yim I, Su J. Artificial intelligence (AI) learning tools in K-12 education: a scoping review. *J Comput Educ*. 2024;1–39. <https://doi.org/10.1007/s40692-023-00304-9>.
12. Su J, Yang W, Yim I, Li H, Hu X. Early artificial intelligence education: effects of cooperative play and direct instruction on kindergarteners' computational thinking, sequencing, self-regulation and theory of mind skills. *J Comput Assist Learn*. 2024;40:2917–25. <https://doi.org/10.1111/jcal.13040>.
13. Delcker J, Heil J, Ifenthaler D, Seufert S, Spirgi L. First-year students AI-competence as a predictor for intended and de facto use of AI-tools for supporting learning processes in higher education. *Int J Educ Technol High Educ*. 2024;21. <https://doi.org/10.1186/s41239-024-00452-7>.
14. Karalekas G, Vologiannidis S, Kalomiros J. Teaching artificial intelligence and machine learning in secondary education: a robotics-based approach. *Appl Sci*. 2025. <https://doi.org/10.3390/app15084570>.
15. Ng D, Su J, Leung J, Chu S. Artificial intelligence (AI) literacy education in secondary schools: a review. *Interact Learn Environ*. 2023;32:6204–24. <https://doi.org/10.1080/10494820.2023.2255228>.
16. Chou C, Shen T, Shen T, Shen C. Influencing factors on students' learning effectiveness of AI-based technology application: mediation variable of the human-computer interaction experience. *Educ Inform Technol*. 2022;27:8723–50. <https://doi.org/10.1007/s10639-021-10866-9>.
17. Sweller J. Cognitive load theory. In: *Psychology of learning and motivation*. vol 55. Academic Press; 2011. p. 37–76. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>.
18. Kirschner PA. Cognitive load theory: implications of cognitive load theory on the design of learning. *Learn Instr*. 2002;12(1):1–10.
19. Mayer RE. Constructivism as a theory of learning versus constructivism as a prescription for instruction. In: *Constructivist instruction*. Routledge; 2009. p. 196–212.
20. Owoyemi SO, Ariyibi O, Akindoju OG, Famuyide AJ, Somade MA, Adelaja MA. Awareness, skills and knowledge of lecturers and students in machine learning in Lagos State University. *Educ Perspect*. 2024;13(1):151–8.
21. Sundari MS, Penthal HR, Nayyar A. Transforming education through AI-enhanced content creation and personalized learning experiences. In: *Impact of artificial intelligence on society*. Chapman and Hall/CRC; 2024. p. 98–118. <https://doi.org/10.1201/9781032644509-7>.
22. Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. *Educ Psychol*. 2003;38(1):1–4.
23. Bodonheli A, Thaqi E, Özdel S, Bozkir E, Kasneci E. From passive watching to active learning: empowering proactive participation in digital classrooms with AI video assistant. In: *Proceedings of the 2025 CHI conference on human factors in computing systems*; 2025. p. 1–21. <https://doi.org/10.1145/3706598.3713513>.
24. Omeh CB, Ayanwale MA, Mnguni LE, Olelewe CJ. Fostering programming skill and critical thinking through AI-assisted PBL integration. *J New Approaches Educ Res*. 2025;14(1):22. <https://doi.org/10.1007/s44322-025-00041-0>.
25. Omeh CB. Impact of artificial intelligence-assisted assessment and traditional assessment on web design and development in computing education. *Educ Sci*. 2026;16(4):501.
26. Yilmaz R, Yilmaz F. The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Comput Educ Artif Intell*. 2023;4:100147. <https://doi.org/10.1016/j.caeai.2023.100147>.
27. Alanazi M, Soh B, Samra H, Li A. The influence of artificial intelligence tools on learning outcomes in computer programming: a systematic review and meta-analysis. *Comput*. 2025;14:185. <https://doi.org/10.3390/computers14050185>.
28. Mellado R, Cubillos C. Can generative artificial intelligence outperform self-instructional learning in computer programming? Impact on motivation and knowledge acquisition. *Appl Sci*. 2025. <https://doi.org/10.3390/app15115867>.
29. Kim S, Kim T, Kim K. Development and effectiveness verification of AI education data sets based on constructivist learning principles for enhancing AI literacy. *Sci Rep*. 2025;15. <https://doi.org/10.1038/s41598-025-95802-4>.
30. Galatro D, Amon CH. A similarity-based approach for diagnosing the aging of lithium-ion batteries in second life combining time series and machine learning. *Appl Sci*. 2025;15(13):7378.
31. Zhdanovskaya A, Baidakova D, Ustalo D. (2023). Data labeling for machine learning engineers: project-based curriculum and data-centric competitions. 15886–93. <https://doi.org/10.1609/aaai.v37i13.26886>

32. Fleischer Y, Biehler R, Schulte C. Teaching and learning data-driven machine learning with educationally designed jupyter notebooks. *Stat Educ Res J*. 2022. <https://doi.org/10.52041/serj.v21i2.61>.
33. Rangone G, Montejano G, Garis A, Pizarro C, Molina W. An educational data mining model based on auto machine learning and interpretable machine learning. In: 2022 IEEE global conference on computing, power and communication technologies (GlobConPT). 2022. p. 1–6. <https://doi.org/10.1109/globconpt57482.2022.9938243>.
34. Nisbet R, Miner GD, McCormick K. Handbook of statistical analysis: AI and ML applications. Elsevier; 2024. <https://doi.org/10.1016/B978-0-443-15845-2.00032-3>.
35. Alshahrani K, Qureshi RJ. Review the prospects and obstacles of AI-enhanced learning environments: the role of ChatGPT in education. *Int J Mod Educ Comput Sci*. 2024;16(4):71–86. <https://doi.org/10.5815/ijmecs.2024.04.06>.
36. Moubayed A, Injadat M, Nassif AB, Lutfiyya H, Shami A. E-learning: challenges and research opportunities using machine learning & data analytics. *IEEE Access*. 2018;6:39117–38. <https://doi.org/10.1109/ACCESS.2018.2851790>.
37. Emmert-Streib F, Dehmer M. Evaluation of regression models: model assessment, model selection and generalization error. *Mach Learn Knowl Extr*. 2019;1(1):521–51. <https://doi.org/10.3390/make1010032>.
38. Tsai Y. Empowering students through active learning in educational big data analytics. *Smart Learn Environ*. 2024;11. <https://doi.org/10.1186/s40561-024-00300-1>.
39. Amrihesari M, Brettman B. Introducing Data-driven materials informatics into undergraduate courses through a polymer science workshop. *J Chem Educ*. 2025;102:3972–81. <https://doi.org/10.1021/acs.jchemed.5c00562>.
40. Liu X, Zhao S. Research on digital evaluation rubrics for introductory computer science courses based on machine learning and big data technology. In: 2024 6th International Workshop of Artificial Intelligence Education (WAIE). 2024. p. 134–8. <https://doi.org/10.1109/waie63876.2024.00031>.
41. Adewale M, Azeta A, Abayomi-Alli A, Sambo-Magaji A. Impact of artificial intelligence adoption on students' academic performance in open and distance learning: a systematic literature review. *Heliyon*. 2024;10. <https://doi.org/10.1016/j.heliyon.2024.e40025>.
42. García-Martínez I, Fernández-Batanero J, Fernández-Cerero J, León S. Analysing the impact of artificial intelligence and computational sciences on student performance: systematic review and meta-analysis. *J New Approaches Educ Res*. 2023;12:171–97. <https://doi.org/10.7821/naer.2023.1.1240>.
43. Naseer F, Khan M, Tahir M, Addas A, Aejaz S. Integrating deep learning techniques for personalized learning pathways in higher education. *Heliyon*. 2024;10. <https://doi.org/10.1016/j.heliyon.2024.e32628>.
44. Bauer E, Greiff S, Graesser A, Scheiter K, Sailer M. Looking beyond the hype: understanding the effects of AI on learning. *Educ Psychol Rev*. 2025;37. <https://doi.org/10.1007/s10648-025-10020-8>.
45. Zheng X, Yue Y, T, Yang X. Effects of different types of cues and self-explanation prompts in instructional videos on deep learning: evidence from multiple data analysis. *Educ Tech Research Dev*. 2023;1–25. <https://doi.org/10.1007/s11423-023-10188-2>.
46. Yu Z, Yu L. A meta-analysis of online learning outcomes and their gender differences. *Int J Distance Educ Technol*. 2021;19:33–50. <https://doi.org/10.4018/ijdet.2021070103>.
47. Cheng C, Wang J, Zhai X, Yang Y. AI literacy and gender equity in elementary education: a quasi-experimental study of a STEAM–PBL–AIoT course with questionnaire validation. *Int J STEM Educ*. 2025;12. <https://doi.org/10.1186/s40594-025-00574-y>.
48. Hsu T, Abelson H, Van Brummelen J. The effects on secondary school students of applying experiential learning to the conversational AI learning curriculum. *Int Rev Res Open Distrib Learn*. 2022. <https://doi.org/10.19173/irrodl.v22i4.5474>.
49. Zhao H, Wang M, Wan H, Cavanaugh X, T, Liu J. Promoting pupils' computational thinking skills and self-efficacy: a problem-solving instructional approach. *Educ Tech Research Dev*. 2021;69:1599–616. <https://doi.org/10.1007/s11423-021-10016-5>.
50. Li Q, Jiang Q, Liang J, Xiong W, Zhao W. Engagement predicts computational thinking skills in unplugged activity: analysis of gender differences. *Think Skills Creativity*. 2024. <https://doi.org/10.1016/j.tsc.2024.101537>.
51. Hadyaoui A, Cheniti-Belcadhi L. Exploring the effects of gender in skills acquisition in collaborative learning based on the ontological clustering model. *J Comput Assist Learn*. 2023;40:2484–95. <https://doi.org/10.1111/jcal.12852>.
52. Baker R, Richey J, Zhang J, Karumbaiah S, Andres-Bray J, Nguyen H, Andres J, McLaren B. Gaming the system mediates the relationship between gender and learning outcomes in a digital learning game. *Instr Sci*. 2024;53:203–38. <https://doi.org/10.1007/s11251-024-09679-3>.
53. Pöhl S, Bogner F. Learning with computer-based multimedia: gender effects on efficiency. *J Educ Comput Res*. 2012;47:387–407. <https://doi.org/10.2190/ec.47.4.c>.
54. Mouza C, Pan Y, Yang H, Pollock L. A multiyear investigation of student computational thinking concepts, practices, and perspectives in an after-school computing program. *J Educ Comput Res*. 2020;58:1029–56. <https://doi.org/10.1177/0735633120905605>.
55. Perez-Felkner L, Erichsen K, Li Y, Chen J, Hu S, Surmeier L, Shore C. Computing education interventions to increase gender equity from 2000 to 2020: a systematic literature review. *Rev Educ Res*. 2024;95:536–80. <https://doi.org/10.3102/00346543241241536>.
56. Cheng Z, Xu J, Jin H. TreeQuestion: assessing conceptual learning outcomes with LLM-generated multiple-choice questions. *Proc ACM Hum Comput Interact*. 2024;8(CSCW2):3686970. <https://doi.org/10.1145/3686970>.
57. Luckin R, Holmes W. Intelligence unleashed: an argument for AI in education; 2016.
58. Roll I, Wylie R. Evolution and revolution in artificial intelligence in education. *Int J Artif Intell Educ*. 2016;26(2):582–99.
59. Xiu-Yi W. AI in L2 learning: a meta-analysis of contextual, instructional, and social-emotional moderators. *System*. 2024. <https://doi.org/10.1016/j.system.2024.103498>.
60. Owusu S, Zimba J, Atta F, Darling M. Evaluating the impact of AI-personalized learning systems in higher education; Examining how they affect academic performance across different age groups at Kumasi Technical University. 2024. <https://doi.org/10.55529/jaimlnn.45.19.29>.
61. Martins R, Von Wangenheim C, Rauber M, Hauck J. Machine learning for all!—Introducing machine learning in middle and high school. *Int J Artif Intell Educ*. 2023;1–39. <https://doi.org/10.1007/s40593-022-00325-y>.
62. Sanusi I, Oyelere S, Vartiainen H, Suhonen J, Tukiainen M. A systematic review of teaching and learning machine learning in K-12 education. *Educ Inform Technol*. 2022;28:5967–97. <https://doi.org/10.1007/s10639-022-11416-7>.

63. Touretzky D, Gardner-McCune C, Seehorn D. Machine learning and the five big ideas in AI. *Int J Artif Intell Educ*. 2022;33:233–66. <https://doi.org/10.1007/s40593-022-00314-1>.
64. Abbas A, Aslam M, Shahzad S, Butt S. Leveraging artificial intelligence to enhance non-formal education: bridging learning gaps beyond the classroom. *Res J Soc Affairs*. 2025. <https://doi.org/10.71317/rjsa.003.05.0360>.
65. Yakubu M, Abubakar A. Applying machine learning approach to predict students' performance in higher educational institutions. *Kybernetes*. 2021;51:916–34. <https://doi.org/10.1108/k-12-2020-0865>.
66. Chowdhury S, Das D. Academic performance analysis of the students with feature retraction using machine learning algorithms. In: 2024 4th international conference on sustainable expert systems (ICSSES); 2024. p. 1060–4. <https://doi.org/10.1109/icses63445.2024.10763310>.
67. Ahmed W, Wani M, Pławiak P, Meshoul S, Mahmoud A, Hammad M. Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. *Sci Rep*. 2025;15. <https://doi.org/10.1038/s41598-025-12353-4>.
68. Sistla S. Data modeling essentials: techniques, best practices, and future trends. *Int Sci J Eng Manag*. 2024;9(12).
69. Myeka PKR. Data modeling best practices for AI-driven applications: architectures for scale and efficiency. *World J Adv Eng Technol Sci*. 2025;4(5).
70. Saha S, Sachdeva S. Designing document stores using feature model and application workload. In: 2024 third international conference on electrical, electronics, information and communication technologies (ICEEICT); 2024. p. 1–6.
71. Koupil P, Holubová I. A unified representation and transformation of multi-model data using category theory. *J Big Data*. 2022;9(1):61.
72. Cheruku SR, Jain S, Aggarwal A. Building scalable data warehouses: best practices and case studies. *Darpan Int Res Anal*. 2024;13(3).
73. Shahidi S, Samadzai AW, Shahbazi H. Effective data preprocessing in data science: from method selection to domain-specific optimization. *J Adv Comput Knowl Algorithms*. 2025;7(2).
74. Sharifnia AM, Kpormegbey DE, Thapa D, Cleary M. A primer of data cleaning in quantitative research: handling missing values and outliers. *J Adv Nurs*. 2025;81(3):1–10.
75. Rhyu J, Božinovski D, Dubs AB, Mohan N, Bende C, Maloney EM, Braatz AJ, R. Automated outlier detection and estimation of missing data. *Comput Chem Eng*. 2023;180:108–20.
76. Zhang Y, Wang Y, Gong D, Sun X. Clustering-guided particle swarm feature selection algorithm for high-dimensional imbalanced data with missing values. *IEEE Trans Evol Comput*. 2022;26(4):1–15.
77. Subashini NJ, Pushparaj E, Venkatesh K. Transforming raw CKD data into a fixed dataset: evaluating preprocessing methods. In: 2025 2nd international conference on research methodologies in knowledge management, artificial intelligence and telecommunication engineering (RMKMATE); 2025. p. 1–6.
78. Illembayo JA, Durodola O, Alade O, Awotunde OJ, Olanrewaju AT, Falana OB, Edu OE. Hyperparameter tuning in machine learning: a comprehensive review. *J Eng Res Rep*. 2024;24(6):1–20.
79. Hussain W, Mushtaq MF, Shahroz M, Akram U, Ghith E, Tlija M, Kim T, Ashraf I. Ensemble genetic and CNN model-based image classification by enhancing hyperparameter tuning. *Sci Rep*. 2025;15(1):1–15.
80. Khan MS, Peng T, Akhlaq H, Khan MA. Comparative analysis of automated machine learning for hyperparameter optimization and explainable artificial intelligence models. *IEEE Access*. 2024;12:1–12.
81. Kwon J, Lecué G, Lerasle M. A MOM-based ensemble method for robustness, subsampling and hyperparameter tuning. *Electron J Stat*. 2018;12(2):1203–30.
82. Rainio O, Teuvo J, Klén R. Evaluation metrics and statistical tests for machine learning. *Sci Rep*. 2024;14(1):1–20.
83. Ackerman S, Farchi E, Raz O, Toledo A. Statistical multi-metric evaluation and visualization of LLM system predictive performance; 2025. Preprint at arXiv:2501.12345.
84. Davis SE, Dorn C, Park DJ, Matheny ME. Emerging algorithmic bias: fairness drift as the next dimension of model maintenance and sustainability. *J Am Med Inform Assoc*. 2025;32(3):1–12.
85. Gallegos IO, Rossi RA, Barrow J, Tanjim MM, Kim S, Dernoncourt F, Yu T, Zhang R, Ahmed N. Bias and fairness in large language models: a survey. *Comput Linguistics*. 2023;49(3):1–45.
86. Moon D, Ahn S. Metrics and algorithms for identifying and mitigating bias in AI design: a counterfactual fairness approach. *IEEE Access*. 2024;12:1–15.
87. Bai X, Stede M. A survey of current machine learning approaches to student free-text evaluation for intelligent tutoring. *Int J Artif Intell Educ*. 2022;33:992–1030. <https://doi.org/10.1007/s40593-022-00323-0>.
88. Alruwais N, Zakariah M. Evaluating student knowledge assessment using machine learning techniques. *Sustainability*. 2023;15(7):6229. <https://doi.org/10.3390/su15076229>.
89. Lee Y, Kim S, Jo Y. Generating plausible distractors for multiple-choice questions via student choice prediction. In: Proceedings of the 63rd annual meeting of the association for computational linguistics (ACL 2025). 2025;1154–68. <https://doi.org/10.18653/v1/2025.acl-long.1154>.
90. Field J. Cognitive validity. Examining listening: Research and practice in assessing second language listening, 2013;35, 77–151.
91. Tabachnick BG, Fidell LS. Using multivariate statistics, 2019;6th ed. Boston, MA: Pearson.
92. Singh B, Kaunert C, Lal S, Arora MK. Enhancing AI-augmented classrooms: teacher-centric integration of intelligent tutoring systems and adaptive learning environments. In: *Fostering inclusive education with AI and emerging technologies*. IGI Global; 2025. p. 99–130. <https://doi.org/10.4018/979-8-3693-7255-5.ch004>.
93. Navarete E, Nehring A, Schanze S, Ewerth R, Hoppe A. A closer look into recent video-based learning research: a comprehensive review of video characteristics, tools, technologies, and learning effectiveness. *Int J Artif Intell Educ*. 2025;1–64. <https://doi.org/10.1007/s40593-025-00481-x>.
94. Adeoye MA, Wirawan KASI, Pradnyani MSS, Septiarini NI. Revolutionizing education: unleashing the power of the ADDIE model for effective teaching and learning. *JPI (Jurnal Pendidikan Indonesia)*. 2024;13(1):202–9. <https://doi.org/10.23887/jpi.iundiksha.v13i1.68624>.
95. Owan VJ, Emanghe EE, Denwigwe CP, Etudor-Eyo E, Usoro AA, Ebuara VO, Effiong C, Ogar JO, Bassey BA. Curriculum management and graduate programmes' viability: the mediation of institutional effectiveness using PLS-SEM approach. *J Curriculum Teach*. 2022;11(5):114–27. <https://doi.org/10.5430/jct.v11n5p114>.

96. Zawacki-Richter O, Cefa B, Bai JY. Towards reproducible systematic reviews in open, distance, and digital education—an umbrella mapping review. *Rev Educ.* 2025;3(1):e70031.
97. Feher K, McKelvey F, Rodgers L, Marinov RN. Strategic foresight in Ai governance: the archer model for responsible Ai futures. 2025. Available at SSRN 5316818.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.