

Research paper

## Augmented Reality Mask for Teaching Artificial Intelligence and Machine Learning in Teacher Training and K-12

Wolfgang Robinig <sup>1,2\*</sup> , Benedikt Brünner <sup>1</sup> , Harald Burgsteiner <sup>2</sup> , Johannes P. Wallner <sup>1</sup> 

<sup>1</sup> *Graz University of Technology, AUSTRIA*

<sup>2</sup> *University College of Teacher Education Styria, AUSTRIA*

\*Corresponding Author: [wolfgang.robinig@phst.at](mailto:wolfgang.robinig@phst.at)

This paper is part of the special issue: [AI and STEM: Exploring the Impact of Artificial Intelligence on Science, Technology, Engineering, and Mathematics Education](#)

**Citation:** Robinig, W., Brünner, B., Burgsteiner, H., & Wallner, J. P. (2026). Augmented reality mask for teaching artificial intelligence and machine learning in teacher training and K-12. *European Journal of STEM Education*, 11(1), Article 41. <https://doi.org/10.20897/ejsteme/18920>

**Published:** July 7, 2026

### ABSTRACT

Introducing core concepts of Artificial Intelligence and machine learning in educational contexts requires instructional approaches that are engaging, time-efficient, and accessible to learners without prior expertise. This paper presents an augmented-reality mask activity (AR mask) that guides participants through a machine learning implementation cycle within a single lesson. In a browser-based environment, participants capture and label a small dataset of four facial expression categories and train an image-classification model using transfer learning with a pre-trained convolutional neural network. Learners then deploy the resulting neural network in an interactive, Scratch-based AR application that reacts to live webcam input by overlaying expression-dependent masks. The activity is designed to connect abstract ML concepts to embodied, immediate feedback, make the role of data quality and variability immediately tangible through iteration, and surface typical limitations of ML systems (e.g., uncertainty, fragility, and dataset bias in small samples). In this exploratory study we evaluated the approach using a mixed-methods design with 32 in-service teachers (online survey) and 40 lower-secondary students (in-class questionnaire and guided reflection). Teachers perceived the AR mask as suitable for addressing neural network training processes, computer vision, reliability and error mechanisms, and dataset-related bias, and judged the cost-benefit ratio as favorable. Student responses indicated high enjoyment and showed that learners linked model performance to the quantity and diversity of training data. Overall, the AR mask appears to be a low-threshold entry point for foundational AI literacy.

**Keywords:** artificial intelligence, ai education, educational resource, teacher training, machine learning

The rapid proliferation of AI applications in recent years - encompassing generative AI, autonomous agents, and domain-specific tools such as AI-assisted molecular design - has fundamentally altered the technological landscape that today's students will inhabit and navigate. As these technologies become embedded in professional and everyday life, the ability to understand their underlying principles is no longer a specialist skill but a prerequisite for informed citizenship. This underscores the urgency of introducing AI concepts early and meaningfully within K-12 education. Over the past decade an extensive body of didactic material for Artificial Intelligence (AI) and machine learning (ML) instruction in education has emerged. Examples include comprehensive curricula such as

*Enaris* (Kandlhofer et al., 2023), *Elements of AI*<sup>1</sup> and *DAIly Curriculum*<sup>2</sup>, as well as collections of individual activities and resources such as *AI Unplugged* (Lindner et al., 2019), *Teachable Machine* (Carney et al., 2020), *AI4K12* (Touretzky et al., 2019) or a compilation maintained by TU Dresden (Marx & Chair of Didactics of Computer Science - TU Dresden, 2024). Some of these contributions provide full semester courses, whereas others focus on narrowly scoped learning objects.

Despite this diversity, no resource found simultaneously (i) covers a broad spectrum of fundamental AI concepts while representing a real-world machine learning problem, (ii) can be easily used within a single lesson, and (iii) requires only modest technical overhead - criteria that teachers identify as decisive under typical timetable constraints (Ertmer et al., 2012).

The present study introduces the *facial-expression detecting augmented-reality mask* - or simply *AR mask*. The AR mask activity guides learners through the entire machine learning pipeline:

Students take photos of themselves with four different facial expressions, label the small set of facial-expression images by assigning them to one of four possible categories, train a pre-defined convolutional neural network (Krizhevsky et al., 2012; LeCun et al., 1998), and embed the resulting model into an interactive augmented-reality application working on live webcam input.

The activity yields a functional application in about thirty minutes, and surfaces a rich set of pedagogical talking points:

- *Authenticity*. Facial expression recognition is a genuine ML application, and CNNs (convolutional neural networks) are state-of-the-art for this task.
- *Hands-on ownership*. All stages - data acquisition, labeling, model training, and application code - are executed by the learners themselves.
- *Rapid feedback*. The short development cycle lowers entry barriers and keeps cognitive load manageable.
- *Algorithmic necessity*. A rule-based implementation would be infeasible, underscoring the need for ML.
- *Model limitations*. The tiny training corpus exposes the non-deterministic, non-transparent and inexact nature of neural networks.
- *Data-performance link*. Students can immediately improve reliability by enlarging and diversifying the dataset, thus experiencing the data-centric paradigm of modern AI.
- *Resource awareness*. Scaling data volumes can lead to discussions about computational demands.
- *Bias exploration*. Imbalanced datasets quickly reveal systematic errors.
- *Robustness issues*. Unexpected inputs demonstrate failure modes.
- *Affective design*. The visual design of the mask lowers affective barriers, hopefully reducing the inhibition barrier towards content knowledge of AI.
- *Low material overhead*. Only a camera-equipped device with a browser is required, which is what most current laptops and tablets offer. No additional props are needed because learners photograph their own facial expressions.
- *Privacy preservation*. All image data remains within the local browser environment when choosing “in your web browser” as storage location, ensuring no transmission or external storage of sensitive biometric information.

The goal of the study is to empirically examine these conjectured advantages. We measured perceived suitability across sixteen AI subdomains, compared cost-benefit ratings in school and teacher-education contexts, and analyzed which components of the *Technological-Pedagogical-Content Knowledge* (TPACK) framework are reinforced. By surveying participants in three distinct contexts - teacher-education workshops, general school-teaching scenarios, and teachers' own planned classroom implementations - we explored which AI subdomains the AR mask is perceived to support effectively and where its applicability is limited. We further investigated whether the AR mask can serve as a versatile and time-efficient entry point to AI education. With two school classes we conducted a mixed-methods survey on interest gains coupled with qualitative prompts and discussion on improving the presented model - to explore the data-performance link hypothesis to see, if students can link the quality or quantity of the training images to the performance of the AR mask application.

To systematically evaluate these pedagogical conjectures and assess the AR mask's effectiveness across different educational contexts, this study addresses three primary research questions:

- **RQ1:** For which specific subdomains of AI is the instructional activity particularly effective in teacher education and K-12 classroom practice?

---

<sup>1</sup> Elements of AI. <https://www.elementsofai.com/>

<sup>2</sup> DAIly Curriculum (v2.0). <https://everyday-ai.org/resources/curriculum/daily-curriculum-v20>

- **RQ2:** How do educators evaluate the cost-benefit ratio of the AR mask activity?
- **RQ3:** To what extent does the activity stimulate critical reflection and interest on AI-specific issues among participants?

## RELATED WORK AND THEORETICAL BACKGROUND

The didactic application of the AR mask is grounded in the intersection of data-centric AI education and experiential learning. It addresses the challenge of translating abstract AI and ML concepts into tangible, learner-centered experiences.

### AI Literacy

Current frameworks for AI education, such as the "Big Five" ideas proposed by AI4K12 (Touretzky et al., 2019), the collaborative AI Literacy Framework by the European Commission, OECD and Code.org (Adams et al., 2025) or the UNESCO AI Competency Frameworks for teachers (Miao & Cukurova, 2024) and students (Miao et al., 2024), emphasize not just usage skills but a conceptual understanding of ML mechanisms. Recent scholarship emphasizes that integrating AI across all educational levels - starting as early as preschool - requires deliberate pedagogical design to nurture learner agency and critical, multimodal literacies (Katsamposaki-Hodgetts et al., 2025). An important pedagogical task lies in moving from theoretical definitions to practical understanding. While "CS Unplugged" activities (Bell et al., 2009; Lindner & Seegerer, 2019) are effective for introducing foundational logic without hardware, they often fail to capture the probabilistic, non-deterministic nature of "black box" machine learning applications based on neural networks. The instructional design of the AR mask activity responds to this data-centric approach in AI (Zha et al., 2025).

### Learning theories

To bridge the gap between theoretical AI concepts and practical comprehension, the didactic design of the AR mask activity is anchored in a framework of experiential learning and synthesizes Kolb's experiential learning theory (Kolb, 1976; Kolb et al., 2014) with principles of Embodied Cognition (Shapiro, 2019; Wilson, 2002) and Papert's concept of body syntonicity (Papert, 1980).

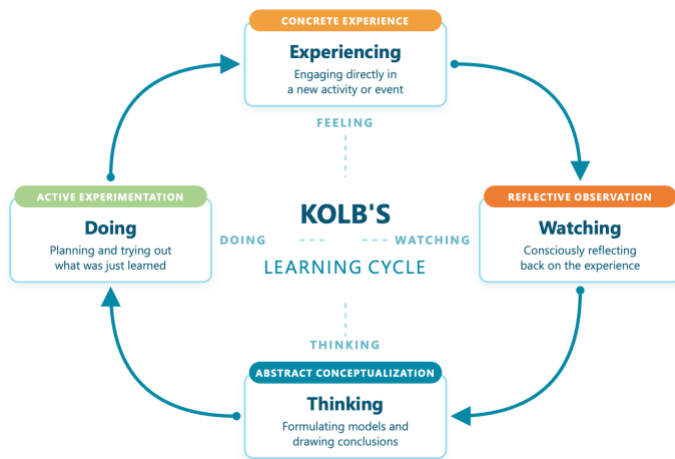
The idea of the AR mask activity is to use physical, sensorimotor engagement to demystify the probabilistic, "black box" nature of artificial neural networks. Through facial movements as the primary interaction medium, the activity leverages embodied cognition to increase cognitive receptiveness towards abstract ML concepts. Because biological brains are first and foremost the control systems for biological bodies, cognitive science grants the body a central role in shaping the mind.

While text-based coding exercises usually conceptually separate the learner from the computational object, the AR mask activity bridges this divide by utilizing the learner's own face as both the primary source of training data and the live system controller. This design specifically operationalizes Papert's concept of body syntonicity: learners understand the new mathematical construct of a classification model by relating it to their own body and sense of movement: They intuitively map the abstract mathematical boundaries of the model directly onto their own physical actions, such as the specific lip, tongue, or eye position required to trigger a targeted classification (e.g., "silly" vs. "surprised").

This embodied interaction drives a complete cycle of experiential learning (**Figure 1**) as defined by Kolb. The workflow initiates with the immediate, tangible act of capturing and labeling one's own physical variations to construct a functional training dataset (*Concrete Experience*). This transitions into observation during the live testing phase, where learners witness the model's real-time, occasionally erroneous, reactions to their physical inputs (e.g., observing the system misclassify a "laughing" expression as "surprised") (*Reflective Observation*). These observable failure modes prompt learners to conceptualize the underlying limitations of the AI system, such as probabilistic uncertainty, data variability requirements, and potential dataset bias within small samples (*Abstract Conceptualization*). This understanding is finally applied in active experimentation, where learners can return to the data acquisition stage to improve the model's performance by editing their training data (e.g., capturing images from varied angles or positions), thereby experiencing the iterative, data-centric paradigm of modern Artificial Intelligence (*Active Experimentation*).

**Figure 1**

Kolb's Experiential Learning Cycle as applied to the AR mask activity. Learners progress through a continuous loop: from the Concrete Experience of physically capturing facial data, to Reflective Observation of the model's errors, followed by Abstract Conceptualization of underlying AI limitations, and concluding with Active Experimentation to improve the dataset.



## IMPLEMENTATION AND WORKFLOW OF THE AR MASK ACTIVITY

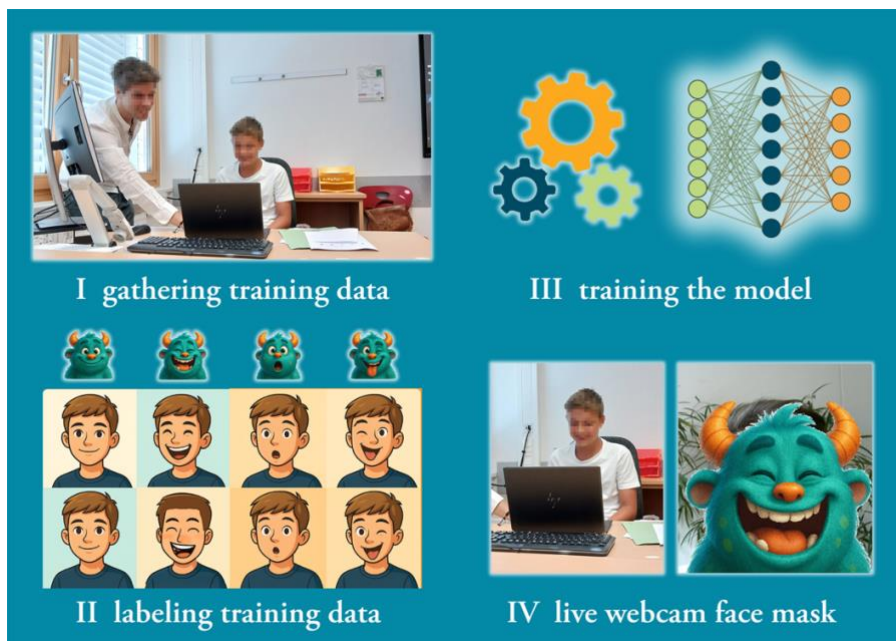
The activity operates entirely within a web browser environment, and participants are guided through four sequential steps of the Machine Learning pipeline.

### Preparation: Data collection, labeling and neural network training

Participants capture their own training dataset by photographing themselves displaying four facial expressions modeled after emoji representations as seen in **Figure 2**: happy, laughing, surprised, and silly. For optimal model performance, learners capture a minimum of 20 images per expression category. Consistent environmental conditions – stable lighting, uniform background, and seating position – enhance model performance. Following the image capture, participants manually label each photograph by assigning it to the appropriate expression category folder.

**Figure 2**

Four steps of the AR mask activity: (i) Capture photos of the learner showing approximately 20 variations of each of the four distinct facial expressions. (ii) Label each image into one of four expression categories. (iii) Train a predefined CNN on the labeled dataset. (iv) Deploy the model live: The AR mask on the webcam feed changes to match the detected expression.



The labeled dataset trains a MobileNet (Howard et al., 2017) Convolutional Neural Network from the Tensorflow.js Javascript library within the open-source framework from „Machine Learning for Kids“ locally within the browser as declared by the developer of “Machine Learning for Kids”. Training via transfer learning typically completes within 5 seconds for approximately 100 images, providing immediate feedback while maintaining engagement.

### Application development

Upon training completion, participants integrate their model into an interactive augmented reality application using a block-based programming interface in the language Scratch. The resulting application operates as a real-time webcam overlay system – as participants display different expressions, the trained CNN triggers corresponding mask images that track head movement and scale with camera distance. Participants can either customize a provided template or reconstruct the program from scratch, depending on instructional objectives. Having established the technical workflow, we now examine the pedagogical rationale behind this design.

### Didactic suitability for fundamental machine learning concepts

The proposed instructional activity is designed as an entry-level scenario for students with no prior domain-specific knowledge in Artificial Intelligence.

A central learning objective is comprehending the paradigmatic differences between classical, rule-based programming and data-driven learning. While an algorithmic definition of facial expressions via explicit “if-then” rules (e.g. geometric analysis of mouth curvature and eyebrow positions) involves high complexity and implementation effort, the activity demonstrates the implementation efficiency of neural networks in this scenario. Students experience firsthand how the preparation of labeled training data leads to the construction of an implicit knowledge base.

Simultaneously, the activity makes the limitations and inherent challenges of this technology transparent. Learners are directly confronted with the “black box” problem: in instances of misclassification, the system offers no deductive reasoning for its decision (lack of explainability). This leads to the empirical observation that model performance correlates directly with the quantity and evenly distributed variance of the training data, but some degree of uncertainty will always remain.

### Thematic context and follow up lessons

The practical procedural knowledge acquired through the AR mask activity serves as a cognitive primer, upon which further theoretical and societally relevant instructional units can be built. Following the conclusion of this unit, learners possess a foundational intuition that facilitates access to the following connected subject areas:

**Technical Foundations:** Understanding the architecture of Artificial Neural Networks (ANNs), the function of artificial neurons e.g. via the implementation of the Perceptron (Rosenblatt, 1958), and the concept of learning via weight adjustment, particularly in comparison to biological neural networks.

**Perception and Systems:** Discussing the distinctions between human and computer-aided perception (Computer Vision) e.g. in the context of Convolutional neural networks (Krizhevsky et al., 2012; LeCun et al., 1998; Robinig et al., 2025a), as well as comparing knowledge-based expert systems with data-driven approaches, weighing their respective advantages and disadvantages.

**Ethics and Society:** Analyzing the causes and types of bias (e.g., through unrepresentative or unbalanced training data) (Hellström et al., 2020; Mavrogiorgos et al., 2024; Mehrabi et al., 2021) and discussing privacy and data protection regarding the processing of biometric features. Furthermore, issues of reliability and explainability can be extended to the context of Generative AI.

**Resources and Sustainability:** The correlation between the need for large datasets for robust Machine learning and the resulting hardware and energy requirements provides an opportunity to critically examine the ecological footprint of AI systems (Pimenow et al., 2024).

Finally, the integration of the models into Scratch allows for flexible embedding within the computer science curriculum. The block-based programming interface can be utilized either as a prerequisite module for introducing algorithmic thinking or as a subsequent phase for practical application development.

### Novelty of the approach

While the availability of K-12 AI education resources has expanded significantly, a disconnect often remains between theoretical availability and practical classroom feasibility. Many existing activities, though conceptually sound, impose logistical barriers or lack algorithmic authenticity, limiting their utility as universal introductory tools.

A comparison can be drawn to similar image-classification activities, such as the "Chameleon" exercise. In this activity, a digital chameleon changes color based on the object held in front of the webcam. While engaging, this approach presents a logistical hurdle: it requires each student to have immediate access to approximately ten distinct physical objects in specific colors (e.g., red, green, and blue) to build a training set. In a typical classroom setting, ensuring every student has such a specific set of props is often unfeasible. The AR mask activity circumvents this issue entirely by utilizing the students' own faces as the data source - a resource that is intrinsically available to every learner, thereby eliminating material preparation time.

Furthermore, the "Chameleon" example highlights a discrepancy in algorithmic authenticity. From a computer vision perspective, simple color recognition is computationally inefficient when solved via Convolutional neural networks (CNNs); such tasks are traditionally and more robustly handled by deterministic algorithms like color thresholding or blob detection (Lindeberg, 1993). Using a neural network for color detection can therefore appear artificially constructed. In contrast, facial expression recognition represents a high-dimensional classification problem where classical rule-based approaches fail, and CNNs represent one genuine state-of-the-art solution. The AR mask activity thus provides a technically authentic use case, demonstrating ML in a context where it is strictly necessary rather than pedagogically forced.

Finally, while augmented reality applications that overlay masks onto webcam feeds exist within educational block-based programming environments, they typically function as technical demonstrations rather than Machine learning exercises. These existing tools generally utilize pre-trained models to detect face location, allowing students to anchor graphical elements to the head, but do not involve the students in the training process itself. The AR mask activity bridges this gap by mandating that students actively collect data and train their own models before deployment. By integrating the full ML pipeline - data collection, labeling, training, and testing - into the AR creation process, this approach transforms a standard coding activity into an experiential introduction to Machine learning.

## METHODOLOGY, ETHICS AND PARTICIPANT CHARACTERISTICS

To address these research questions, we implemented a mixed-methods design that combined an online survey for in-service teachers and a paper-based survey with post-activity discussions to gather both quantitative and qualitative data from lower-secondary students, capturing demographic profiles, perceived instructional value across AI subdomains, and reflections on model performance. The data collection took place in 2025.

### Methodology

For the teacher cohort, the online questionnaire was distributed with LimeSurvey about two weeks after the augmented-reality mask lesson had been presented.

All respondents ( $n_t = 32$ ) were teachers of the Austrian subject *Digital Basic Education*, a lower-secondary computer-science strand for grades 5-8, who were attending an in-service teacher-training course at the time of the survey.

For the student cohort, two middle school classes without prior lessons on AI-related topics – one 6<sup>th</sup>-grade and one 7<sup>th</sup>-grade group (combined  $n_s = 40$ ) – completed a printed questionnaire throughout the activity. The instrument included Likert-scale items on interest gain and open-ended questions. Written responses were followed by a brief guided discussion to capture qualitative reflections on ways to improve the model, thereby testing the data-performance hypothesis in a learner context.

### Ethical considerations

**Institutional Approval and Ethical Standards:** This study was conducted in accordance with international ethical standards for research involving human participants, including the Declaration of Helsinki, as well as European data protection regulations (General Data Protection Regulation, GDPR). Prior to the intervention, formal institutional permission and approval were obtained from the school management, which authorized the implementation of the didactic unit and the accompanying surveys.

**Informed Consent and Minor Assent:** Given the involvement of 6<sup>th</sup> and 7<sup>th</sup>-grade students, rigorous consent procedures were implemented. General consent was obtained from the parents or legal guardians regarding the capturing of recordings during classroom activities. Furthermore, assent from the minors themselves was actively secured: capturing facial data was strictly voluntary, and only students who explicitly volunteered were recorded. For the specific photographic material of the student featured in this publication, an additional, explicit written informed consent for publication was provided by the student's legal guardian.

**Data Privacy and Handling of Facial Data:** Because the activity involves capturing students' facial expressions, it was imperative that biometric information was not compromised. The application was configured to utilize the

"in your web browser" storage option. Consequently, all image data, labeling, and neural network training occurred entirely locally on the respective devices. No student images, facial data, or model data were transmitted to, processed by, or stored on external servers, thereby ensuring complete data privacy and full compliance with regional data protection regulations.

Teacher and Student Surveys: Participating teachers provided informed consent prior to engaging in the online survey. Their participation was entirely voluntary, and all collected data was strictly anonymous. Similarly, the student surveys were conducted anonymously. It was transparently communicated to all participating students prior to the survey that their responses would be evaluated anonymously and would have absolutely no impact on their academic grading or assessment.

### Participant characteristics – teachers

Almost all of the  $n_t = 32$  participating teachers opted to share demographic attributes, and the participant cohort can be described as follows:

- Gender: 56% female, 44% male
- Teaching Digital Basic Education or Computer Science: 97% reported current teaching experience in subject, 3% reported none.
- Teaching experience: 39% had taught for fewer than five years, 51% for five to fifteen years, 10% for more than fifteen years.

### Participant characteristics – students

The  $n_s = 40$  students attended a 6<sup>th</sup> and a 7<sup>th</sup> grade class of the same middle school and had no prior lessons on Artificial Intelligence.

**Gender:** 60% female, 40% male

## RESULTS

### Perceived suitability of the augmented reality mask activity for teaching AI subdomains

To measure the pedagogical use of the augmented-reality mask activity, participating teachers rated each of sixteen AI subdomains in three distinct contexts - teacher training, school teaching in general, and school teaching as they would personally implement the activity. This disambiguation was chosen to differentiate between the perceived value of the AR mask activity in teacher training at university level and school teaching. The school teaching context was split into teaching in general - how it should ideally be used and their personal teaching choice that reflects the probability of those teachers using the activity in their own lessons considering their individual barriers like time constraints, previous knowledge of their students or access to webcam-equipped hardware.

In Figure 3 and Table 1 all AI subdomains in question are ordered by the mean value of the context "teacher-training".

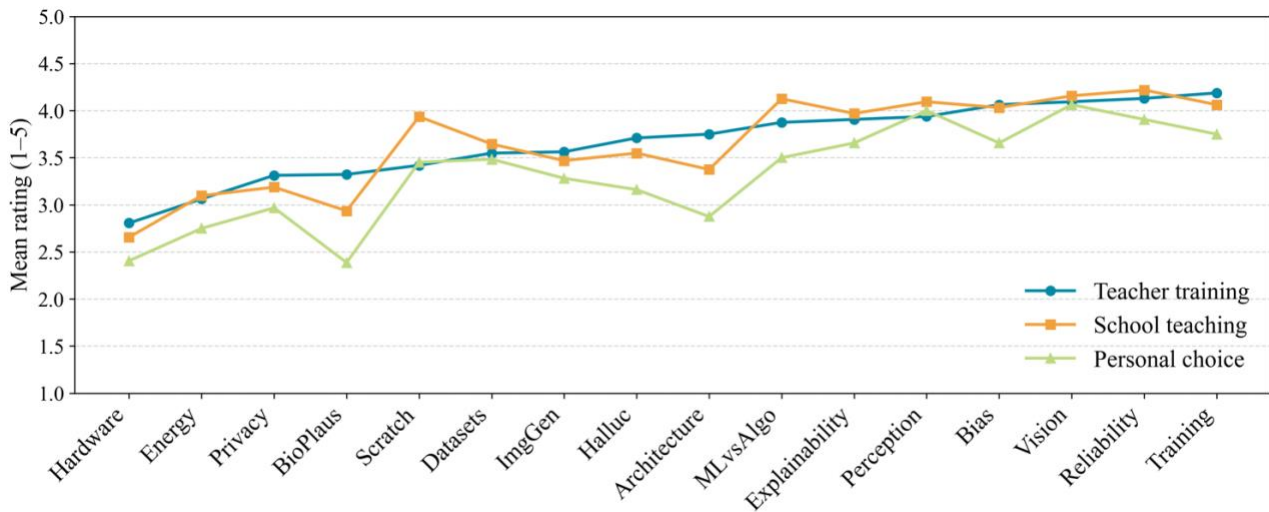
The AI subdomains were:

- Neural network training explanation (*Training*)
- Reliability of GenAI output (*Reliability*)
- How AI image recognition works (*Vision*)
- Causes of bias in AI systems (*Bias*)
- Human vs. AI perception (*Perception*)
- Explainability of AI decisions (*Explainability*)
- ML vs. classical algorithms (*ML vs. Algo*)
- Structure of neural networks (*Architecture*)
- Hallucinations in generative AI (*Halluc*)
- How AI image generation works (*ImgGen*)
- Importance of large, labeled datasets (*Datasets*)
- Block-based programming (*Scratch*)
- Biological plausibility of Neural Networks (*BioPlaus*)
- Data privacy issues (*Privacy*)
- Energy consumption of AI (*Energy*)
- AI training hardware (*Hardware*)

### Figure 3

Mean ratings on a Likert-type scale (1 = "does not apply at all", 5 = "fully applies") for the sixteen instructional targets of the perceived suitability of the augmented reality mask activity for teaching AI subdomains - collected in three contexts: Teacher training,

school teaching in general and school teaching as personally planned by the respondents. The items are ordered by ascending teacher training mean.



The selected AI subdomains are inspired by established AI literacy frameworks - particularly the AI Competency Framework for Teachers (Miao & Cukurova, 2024). Each item was chosen to broadly match the expected competence profile targeted by the planned teaching activity. The item on image generation was included because U-Nets (Ronneberger et al., 2015) (as used in diffusion models for image synthesis (Ho et al., 2020; Sohl-Dickstein et al., 2015)) share architectural similarities with convolutional neural networks (CNNs). Understanding CNN-based image classification therefore lays essential groundwork for comprehending image generation through diffusion models.

## Responses

*Teacher Training Context:* In the context of teacher training, the AR mask activity is perceived primarily as a vehicle for teaching the core mechanics of the machine learning pipeline. The mean values for the four highest-rated items - Training (M = 4.19), Reliability (M = 4.13), Vision (M = 4.09), and Bias (M = 4.06) - were all above 4.0. This indicates that pre-service teachers and seminar participants view the activity as especially effective for demystifying the "black box" processes: Illustrating how input data is processed (Vision) through iterative adjustments (Training) and how this process is susceptible to errors (Reliability) or skewing (Bias).

**Table 1**

*Mean ratings (M) and standard deviations (SD) for the sixteen instructional targets of the perceived suitability of the augmented reality mask activity for teaching AI subdomains - collected in three contexts: Teacher training, school teaching in general and school teaching as personally planned by the respondents.*

AI Sub-Domain	Teacher Training (M)	Teacher Training (SD)	School General (M)	School General (SD)	School Personal (M)	School Personal (SD)
Neural network training explanation (Training)	4.19	0.98	4.06	1.00	3.75	1.27
Reliability of GenAI output (Reliability)	4.13	1.01	4.22	0.86	3.91	1.16
How AI image recognition works (Vision)	4.09	1.26	4.16	1.03	4.06	1.12
Causes of bias in AI systems (Bias)	4.06	1.12	4.03	1.05	3.66	1.21
Human vs. AI perception (Perception)	3.94	1.12	4.09	0.95	4.00	0.97
Explainability of AI decisions (Explainability)	3.91	1.01	3.97	0.85	3.66	0.96
ML vs. classical algorithms (MLvsAlgo)	3.88	1.14	4.13	0.78	3.50	1.27
Structure of neural networks (Architecture)	3.75	0.90	3.38	0.82	2.88	1.19

**Table 1***Continue*

<i>AI Sub-Domain</i>	<i>Teacher Training (M)</i>	<i>Teacher Training (SD)</i>	<i>School General (M)</i>	<i>School General (SD)</i>	<i>School Personal (M)</i>	<i>School Personal (SD)</i>
Hallucinations in generative AI (Halluc)	3.71	1.22	3.55	1.29	3.16	1.30
How AI image generation works (ImgGen)	3.56	1.25	3.47	1.09	3.28	1.12
Importance of large, labeled datasets (Datasets)	3.55	1.34	3.65	1.26	3.48	1.41
Block-based programming (Scratch)	3.42	1.10	3.94	0.91	3.45	1.16
Biological plausibility of NNs (BioPlaus)	3.32	1.23	2.94	0.98	2.39	1.04
Data privacy issues (Privacy)	3.31	1.13	3.19	1.13	2.97	1.00
Energy consumption of AI (Energy)	3.06	1.09	3.10	1.06	2.75	1.15
AI training hardware (Hardware)	2.81	1.18	2.66	1.08	2.41	1.14

The integration of Block-based programming (Scratch) received a moderate rating ( $M = 3.42$ ) in the teacher training context. This evaluation likely reflects the curricular structure of teacher education, where block-based coding is usually covered in separate introductory modules or is assumed as prior knowledge. Consequently, within a specialized AI didactics course, participants may view Scratch as a familiar medium rather than a novel learning objective inherent to the AR activity itself.

Regarding advanced technical concepts, the item on image generation (ImgGen,  $M = 3.56$ ) received a mid-range rating. Although the activity utilizes a CNN for classification rather than generation, this rating supports the argument that understanding CNN feature extraction is a necessary scaffold for comprehending modern generative architectures. As U-Nets (Ronneberger et al., 2015) - which are fundamental to diffusion models used for image synthesis (Ho et al., 2020; Sohl-Dickstein et al., 2015; Song et al., 2020) - share architectural similarities with the classifiers used here, the activity serves as an indirect architectural introduction.

At the bottom end, topics such as Energy ( $M = 3.06$ ) and Hardware ( $M = 2.81$ ) were rated lowest. This perception is likely reinforced by the rapidity of the browser-based training process. Since the custom model trains in mere seconds, the significant computational cost typical of deep learning remains abstract. Consequently, it is not immediately apparent to learners that scaling such a system to a more extensive application would necessitate vastly greater processing power and energy consumption.

### School teaching contexts: General utility vs. personal implementation

Comparing the teacher training results with the two school-related contexts reveals distinct shifts in perceived utility. In the context of "School Teaching in General" - representing the respondents' normative view of the activity's ideal application - several items scored higher than in the training context. Notably, Reliability achieved the highest mean in the entire dataset ( $M = 4.22$ ), and Machine Learning vs. Classical Algorithms rose to  $M = 4.13$ . Furthermore, the perceived suitability of Scratch increased significantly to  $M = 3.94$ .

A divergence becomes apparent when analyzing the "School Teaching (Personal)" context, which reflects the respondents' realistic intention to implement specific aspects in their own lessons. While the "General" ratings suggest a broad acknowledgment of the activity's potential to teach complex concepts like Structure of Neural Networks ( $M = 3.38$ ) or Biological Plausibility ( $M = 2.94$ ), these theoretical topics see a sharp decline in personal implementation intent ( $M = 2.88$  and  $M = 2.39$ , respectively). Instead, respondents prioritize observable, phenomenological aspects. Items such as Vision ( $M = 4.06$ ) and Perception ( $M = 4.00$ ) remain the top priorities in personal lesson planning.

This discrepancy suggests a pragmatic filtering process: teachers appear to value the activity most for its ability to provide immediate, visual demonstrations of computer vision and the contrast to human perception. Abstract technical details (Architecture) or indirect thematic links (Energy, Hardware) appear to be considered less central to their personal teaching scenarios, likely due to time constraints or the prioritization of more basic digital literacy competencies for this age group.

### Cost-benefit ratio

Educators judged the AR mask activity to be cost-effective in both contexts - teacher training and school settings, but they assigned a somewhat higher benefit-effort balance to teacher-training courses than to regular

school lessons. In school settings, 78% of respondents classified the cost-benefit ratio as acceptable and 13% as excellent, while only 6% deemed it unacceptable. For teacher training, the share of “excellent” evaluations doubled to 28%, and 59% still rated the ratio as acceptable; again, only 6% considered it unacceptable. These distributions suggest that the preparation and technical effort required by the activity is broadly justified by its instructional value.

### **Perceived knowledge gains**

The teachers shared assessments of self-reported gains in three TPACK (Mishra et al., 2023; Mishra & Koehler, 2006) dimensions - technological knowledge (TK), content knowledge (CK), and pedagogical knowledge (PK). Among the 32 teachers, 96-97% reported at least some improvement in AI-related content (CK) and technical knowledge (TK), and 78% in pedagogical knowledge (PK). Out of these, strong gains were reported by 31% (CK), 28% (TK) and 19% (PK) - regardless of participants’ prior experience with artificial-intelligence topics. The normally distributed answer patterns, mean values and standard deviations between both groups are almost identical leading to no significant difference in a t-test for this number of participants ( $n_t=32$ ), suggesting that the perceived learning benefit was uniform across baseline expertise levels.

### **Change in AI Interest**

To gauge motivational effects, both questionnaire instruments included a single-item measure asking whether participants’ interest in Artificial Intelligence had *increased*, *remained the same*, or *decreased* after the AR mask lesson.

#### ***Teacher cohort***

Among the in-service teachers, 56% reported an *increase* in interest in AI-related topics, 34% indicated *no change*, and 3% noted a *decrease*.

#### ***Student cohort***

A comparable pattern was observed in the two student classes ( $n_s = 40$ ). Based on pre- and post-activity interest ratings, 40% showed an *increase*, 57.5% showed *no change*, and one person (2.5%) showed a *decrease*. Interestingly, the student whose interest declined nevertheless awarded the activity a maximum score on the separate overall-satisfaction item.

### ***Activity enjoyment and interest correlation.***

Correlational analyses of student responses indicated a positive relationship between pre-activity interest ( $M=3.30$ ,  $SD=0.60$ ) and post-activity interest in AI ( $M=3.68$ ,  $SD=0.73$ ), with a substantial correlation ( $r_{pre-post} = 0.61$ ). The correlation between post-activity interest and overall activity enjoyment ( $M=3.95$ ,  $SD=0.71$ ) was similarly strong ( $r_{post-enj} = 0.64$ ), and the correlation between initial interest and activity enjoyment was moderate ( $r_{pre-enj} = 0.37$ ). This coincides with the correlation between AI interest and digital learning activity motivation found in a recent study (Baez et al., 2026). These findings suggest that the AR-mask activity was effective in sustaining and enhancing AI interest, particularly among students with already elevated initial interest levels, and was generally perceived as highly enjoyable.

### **Post-application reflection**

Consistent with the trend in AI Interest change, 22 teachers (73%) stated that the example continued to prompt reflection on AI-related issues after the instructional unit, while eight (27%) disagreed and two could not be reached for this question. Taken together, the activity appears to be effective both in stimulating curiosity about AI and in fostering post-session reflection among the majority of learners.

### **Student reflections on improving model reliability**

Following the initial demonstration, students provided written suggestions on how to improve the reliability and accuracy of the AR mask application. Qualitative student feedback was independently coded by two raters into three categories: “concrete suggestions/clear understanding”, “vague suggestions/some understanding”, and “no accurate suggestions/no understanding”, achieving an inter-rater reliability of Cohen’s  $\kappa=1.0$ . Of the 40 participants, 13 provided concrete suggestions demonstrating a clear understanding of the underlying principles. These students emphasized the need for broader data collection encompassing diverse angles, varying backgrounds, multiple individuals, different lighting conditions, and varied camera distances. Nine students offered

less detailed recommendations (e.g., 'train more' or 'use more photos'), which were coded as vague suggestions with some understanding. Two students specifically suggested making the facial expressions more distinguishable.

Furthermore, four responses did not demonstrate understanding of the concepts, and 12 students provided no response. This relatively high abstention rate (30%) was likely the result of deliberate instructional choices; the trainer refrained from providing scaffolding or hints regarding model improvements prior to the data-collection stage.

In a subsequent open discussion, student participation increased once the accuracy of their earlier suggestions - such as capturing images from varied angles, backgrounds, distances, lighting conditions, and individuals - was confirmed by the trainer. The participation observed during this follow-up discussion suggested that the majority of the students had intuitively grasped fundamental principles of machine learning, effectively setting a baseline for further instructional activities on this topic.

## Limitations

While the findings from this pilot study suggest that the AR mask activity is a highly engaging and practical entry point for K-12 AI education, several methodological limitations must be acknowledged. First, the sample size was relatively small, comprising 32 in-service teachers and 40 lower-secondary students. Consequently, the findings should be interpreted as exploratory rather than strictly generalizable to broader educational contexts. The research design did not have a control group and did not employ a true pre- and post-test assessment of cognitive learning outcomes. The quantitative data collected relied heavily on self-reported perceptions of knowledge gain and perceived suitability, as well as a single-item measure for changes in AI interest.

## CONCLUSIONS, RECOMMENDATIONS AND OUTLOOK

Given the modest technical and temporal requirements, the AR mask activity seems to represent a promising instructional approach to introducing fundamental AI concepts that targets the current demand for teaching material for the understanding of AI (Robinig et al., 2025). Initial empirical data from both teacher and student cohorts indicate measurable increases in AI interest, high perceived instructional value, and effective engagement with critical AI subdomains, such as neural-network training, bias mechanisms, and the relationship between data quality and model performance. The hands-on, visually intuitive nature of the activity appears beneficial in cognitively preparing young learners to assimilate more abstract and technically complex content encountered in subsequent AI lessons.

Student reflections demonstrated that the AR mask activity fostered intuitive understanding of central machine-learning principles, like the data-performance link, bias recognition, and robustness considerations. Despite these outcomes, the activity is intended neither as definitive nor as prescriptive. Rather, educators and researchers are encouraged to adapt, extend, or modify this approach, as well as to use it as a baseline for comparative studies exploring effective introductory strategies in AI education.

## Acknowledgement

This research was done as part of the "FutureDEAL - Future of Digital Education and Learning" initiative within the doctoral program "Bildungsinnovation braucht Bildungsforschung", which is supported and partially funded by the Austrian Federal Ministry of Education, Science, and Research. Furthermore, this research was supported by the TU Graz Open Access Publishing Fund.

## Funding

This research was done as part of the "FutureDEAL - Future of Digital Education and Learning" initiative within the doctoral program "Bildungsinnovation braucht Bildungsforschung", which is supported and partially funded by the Austrian Federal Ministry of Education, Science, and Research.

The publishing fee was covered by the TU Graz Open Access Publishing Fund.

## Ethical statement

This study followed international ethical standards for research involving human participants, including the Declaration of Helsinki and GDPR. Institutional approval was obtained prior to the intervention. Given the involvement of minors (6th and 7th graders), rigorous consent procedures were applied. Parents or legal guardians provided general consent for classroom recordings; student assent was actively secured, with facial data capture strictly voluntary. For the student photograph published here, explicit written informed consent was obtained from the legal guardian.

To protect biometric data, the application was configured to process and store all image data, labeling, and model training locally on device. No facial or model data was transmitted to or stored on external servers, ensuring full GDPR compliance.

Teacher and student surveys were conducted with fully informed, voluntary participation. All survey data was collected and processed anonymously, with students explicitly informed that responses would not affect their academic assessment.

### Competing interests

No competing interests.

### Author contributions

WR: Research concept and design, Collection and/or assembly of data, Data analysis and interpretation, Writing the article, Critical revision of the article, Final approval of the article; BB: Partial research concept and design; HB: Critical feedback on the article, Mentor; JPW: Critical feedback on the article, Supervisor

### Data availability

OER Lesson material for the AR Mask activity designed after the intervention described in this paper can be found at:

<https://doi.org/10.5281/zenodo.20166352> (English version)

<https://doi.org/10.5281/zenodo.20166680> (German version)

### AI disclosure

AI image generation tools were used to produce the face mask images, the eight face illustrations, and the cogwheel icons in [Figure 2](#), as well as to generate an initial draft of Figure 1 (an SVG subsequently edited manually). AI-assisted writing tools (chatbots) were used to support proofreading, error detection, translation, writing improvement, and condensing of selected passages, as well as to refine the phrasing and formatting of questions in the teacher questionnaire. All AI-generated or AI-assisted content was reviewed, verified, and edited by the authors, who take full responsibility for the final manuscript.

### Biographical sketch

*Wolfgang Robinig* is a PhD candidate at the Institute of Software Engineering and Artificial Intelligence, Graz University of Technology, and professor of digital media education and AI at the University College of Teacher Education Styria. His research interests span AI education, audio engineering, and spatial audio. He has lectured in audio engineering at Graz University of Technology, taught digital media, mathematics, and physics, worked as a corporate audio/video technician, and collaborated as audio engineer with musicians including Jennifer Batten and Don Airey.

*Benedikt Brünner* recently completed his PhD at Graz University of Technology in the field of Educational Technology. His research focuses on self-regulated learning, AI-supported learning environments, and user experience in digital education. He has been involved in the Future of Digital Education and Learning (FutureDEAL) initiative and works closely with the Department of Educational Technology. Before his academic career, he taught computer science and biology at the secondary school level in Austria. His work bridges research and practice, with a strong emphasis on designing and evaluating innovative educational technologies in real-world contexts.

*Harald Burgsteiner*, professor for media informatics and artificial intelligence at the University of Teacher Education Styria, studied information and computer engineering at the Graz University of Technology, doctorate in neuroinformatics and artificial intelligence, teacher training in mathematics and computer science. Previous activities as university assistant at the TU Graz, professor and data protection officer at the FH Joanneum, winner of the Teaching Award 2010 at FH Joanneum. Current research areas include didactics of computer science, artificial intelligence, maker education, media informatics, educational robotics, mobile and web-based software development and its use in education.

*Johannes P. Wallner* is an associate professor at the Graz University of Technology. He received his PhD at TU Wien in 2014. Wallner's research agenda focuses on foundational aspects of knowledge representation and reasoning, including computational argumentation, inconsistency handling, and algorithmic approaches for computationally complex problems.

**Disclaimer/Publisher's Note**

The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and do not necessarily reflect the views of Lectito Publications and/or the editor(s). Lectito Publications and/or the editor(s) disclaim responsibility for any injury to persons or property resulting from any ideas, methods, instructions, or products referred to in the content.

**REFERENCES**

- Adams, C., Cachia, R., Kivinen, K., Kralj, L., Ruiz, P., South, J., Lee, V. R., Nattrass, T., & Viéville, T. (2025). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education* [Review draft]. <https://ailiteracyframework.org>
- Baez, C., Buchner, J., Ullrich, A. L., & Schallert-Vallaster, S. (2026). Exploring AI perceptions in education: Unveiling the role of student and teacher motivation and self-efficacy. *Computers and Education Open*, 10, Article 100346. <https://doi.org/10.1016/j.caeo.2026.100346>
- Bell, T., Alexander, J., Freeman, I., & Grimley, M. (2009). Computer science unplugged: School students doing real computing without computers. *New Zealand Journal of Applied Computing and Information Technology*, 13(1), 20–29.
- Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A., & Chen, A. (2020). Teachable machine: Approachable web-based tool for exploring machine learning classification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3334480.3382839>
- Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education*, 59(2), 423–435. <https://doi.org/10.1016/j.compedu.2012.02.001>
- Hellström, T., Dignum, V., & Bensch, S. (2020). Bias in machine learning — What is it good for? *CEUR Workshop Proceedings*, 2659, 3–10. <https://arxiv.org/abs/2004.00686v2>
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications*. arXiv. <https://arxiv.org/abs/1704.04861v1>
- Kandlhofer, M., Weixelbraun, P. F., Menzinger, M., Steinbauer-Wagner, G., & Kemenesi, Á. (2023). Education and awareness for artificial intelligence. *Lecture Notes in Computer Science*, 14296, 1271. <https://doi.org/10.1007/978-3-031-44900-0>
- Katsampoxaki-Hodgetts, K., Kotsidis, K., Papadakis, S., & Anastasiades, P. (2025). Reimagining preschool pedagogy: AI-enhanced teaching through multiliteracies-informed curriculum innovation. In *Proceedings of the 6th International Baltic Symposium on Science and Technology Education (BalticSTE2025)* (pp. 104–119). <https://doi.org/10.33225/>
- Kolb, D. A. (1976). Management and the learning process. *California Management Review*, 18(3), 21–31. <https://doi.org/10.2307/41164649>
- Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2014). Experiential learning theory: Previous research and new directions. In *Perspectives on thinking, learning, and cognitive styles* (pp. 227–247). Routledge. <https://doi.org/10.4324/9781410605986-9>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2323. <https://doi.org/10.1109/5.726791>
- Lindeberg, T. (1993). Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3), 283–318. <https://doi.org/10.1007/BF01469346>
- Lindner, A., Seegerer, S., & Romeike, R. (2019). Unplugged activities in the context of AI. In *Lecture Notes in Computer Science* (Vol. 11913, pp. 123–135). Springer. [https://doi.org/10.1007/978-3-030-33759-9\\_10](https://doi.org/10.1007/978-3-030-33759-9_10)
- Marx, E., & Chair of Didactics of Computer Science–TU Dresden. (2024). *Unterrichtsmaterial – Künstliche Intelligenz* [Teaching materials – Artificial intelligence]. TU Dresden. [https://tu-dresden.de/ing/informatik/smt/ddi/schulinformatik/eduinf-education\\_in\\_informatics/lehr-lern-material/kuenstliche-intelligenz/copy\\_of\\_unterrichtsmaterial](https://tu-dresden.de/ing/informatik/smt/ddi/schulinformatik/eduinf-education_in_informatics/lehr-lern-material/kuenstliche-intelligenz/copy_of_unterrichtsmaterial)
- Mavrogiorgos, K., Kiourtis, A., Mavrogiorgou, A., Menychtas, A., & Kyriazis, D. (2024). Bias in machine learning: A literature review. *Applied Sciences*, 14(19), Article 8860. <https://doi.org/10.3390/app14198860>

- Mehrabani, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6). <https://doi.org/10.1145/3457607>
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000391104>
- Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO. <https://doi.org/10.54675/JKJB9835>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108, 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education*. <https://doi.org/10.1080/21532974.2023.2247480>
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books. <https://dl.acm.org/doi/pdf/10.5555/1095592.C1104756>
- Pimenow, S., Pimenowa, O., & Prus, P. (2024). Challenges of artificial intelligence development in the context of energy consumption and impact on climate change. *Energies*, 17(23), Article 5965. <https://doi.org/10.3390/en17235965>
- Robinig, W., Burgsteiner, H., & Steinbauer-Wagner, G. (2025a). A TPACK-aligned teaching approach: Image recognition via convolutional neural networks for artificial intelligence education in teacher training. In *Proceedings of the EdMedia Conference on Educational Media and Technology* (pp. 424–429). Association for the Advancement of Computing in Education. <https://www.learntechlib.org/primary/f/226171/>
- Robinig, W., Burgsteiner, H., & Steinbauer-Wagner, G. (2025b). AI literacy and integration practices among digital education teachers: A survey study of knowledge, attitudes, and professional development needs. In *Lecture Notes in Computer Science* (pp. 141–153). Springer. [https://doi.org/10.1007/978-3-031-93567-1\\_10](https://doi.org/10.1007/978-3-031-93567-1_10)
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science* (Vol. 9351, pp. 234–241). Springer. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- Shapiro, L. (2019). *Embodied cognition*. Routledge. <https://doi.org/10.4324/9781315180380>
- Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning* (pp. 2256–2265). PMLR.
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. In *Proceedings of the 9th International Conference on Learning Representations*.
- Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 9795–9799. <https://doi.org/10.1609/aaai.v33i01.33019795>
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. <https://doi.org/10.3758/BF03196322>
- Zha, D., Bhat, Z. P., Lai, K. H., Yang, F., Jiang, Z., Zhong, S., & Hu, X. (2025). Data-centric artificial intelligence: A survey. *ACM Computing Surveys*, 57(5). <https://doi.org/10.1145/3711118>