Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 6, N. 4 (2026)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Engaging Students with Computer Vision: Graphical Interfaces and Mobile Experimentation for Progressive AI Literacy in Upper Secondary Schools

Luca Zedda[1]*, Cecilia Di Ruberto[1] and Andrea Loddo[1]*

[1] Department of Mathematics and Computer Science, University of Cagliari, Italy; luca.zedda@unica.it; cecilia.dir@unica.it; andrea.loddo@unica.it
* Correspondence: luca.zedda@unica.it; andrea.loddo@unica.it

Abstract: Computer vision can make artificial intelligence observable through image labels, bounding boxes, and segmentation masks, creating opportunities for practical and critical AI literacy. This exploratory classroom study reports a progressive pathway involving 152 students in seven upper secondary classes from different schools, comprising three fourth-year and four fifth-year classes. Three-hour sessions with ten-minute breaks combined graphical interfaces, guided experimentation, and smartphone detection testing; completion required a working prototype. Reported completion rates were 82% for classification, 43% for detection, and 12% for segmentation. Among students who worked on object detection, 97% reportedly applied the model through a smartphone application. These descriptive indicators concern different activities and participant bases; they do not establish learning gains or measured increases in engagement. The low segmentation completion rate is considered in relation to available time, operational requirements, and instructional support, without isolating their effects. Classroom observations and discussions provide a formative account of interpreting predictions and model errors. The study proposes differentiated scaffolding and assessment that distinguishes practical tool use from conceptual explanation and critical evaluation. Its contribution is a situated account of designing AI education around visible outputs, authentic testing, and responsible interpretation, with methodological limitations that qualify the reported outcomes.

Keywords: AI literacy; computer vision education; student engagement

Copyright: © 2026 by the authors. Submitted for possible open-access publication under the terms and conditions of the Creative Commons Attribution (CC BY) licence (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial intelligence is encountered through everyday digital services, but access to a system does not necessarily make its operation understandable. In education, this creates a practical problem: students may use an AI tool successfully while remaining uncertain about what its output represents and what would justify trusting it. An instructional approach to AI therefore needs to connect use with explanation and critical evaluation. AI literacy encompasses competencies for understanding and evaluating AI and interacting with it appropriately, rather than treating technical operation as sufficient evidence of understanding (Long & Magerko, 2020).

Computer vision provides a concrete setting for this connection. Its outputs are visible: a category assigned to an image, a box drawn around an object, or a mask marking an image region. These representations allow teachers and students to refer to the same prediction while asking different questions about it. What has been classified? What has been localized? Which pixels or regions have been included? A

prediction can be inspected alongside its input, making disagreement and error available for classroom discussion. The educational opportunity lies in this relationship between representation and interpretation, rather than in the novelty of the software alone.

The pathway examined here introduces image classification, object detection, and image segmentation in sequence. Classification assigns a category to an image as a whole. Detection adds object localization, commonly represented through bounding boxes. Segmentation identifies regions through masks and introduces attention to boundaries and pixel-level correspondence. These tasks share a visual input but produce different types of information. Teaching them as distinct activities can help students move beyond the undifferentiated idea that AI simply recognizes images.

The study concerns 152 students across seven upper secondary classes, with three fourth-year and four fifth-year classes. Graphical interfaces and mobile experimentation served as instructional mediators. The intervention was intended to make model behavior accessible without making programming syntax the main entry requirement. Its aim was educational understanding of AI outputs, their limitations, and their relationship to real-world conditions, rather than model optimization or preparation for specialist machine learning work.

Student engagement provides a framework for interpreting this design. Fredricks et al. (2004) distinguish behavioral, emotional, and cognitive dimensions of engagement. Participating in a practical activity, expressing interest in an unexpected prediction, and reasoning about why a model failed are related but different phenomena. A high rate of successful tool use cannot represent all these dimensions. Accordingly, engagement is used here as a pedagogical lens; the available evidence does not include a validated engagement measure or a basis for estimating change in each dimension.

The main aim is to describe the pathway and interpret the reported completion pattern without overstating what it demonstrates. Two questions organize the account: how did reported completion differ across classification, detection, and segmentation, and what does smartphone application contribute to the interpretation of the detection activity? A further design question concerns the balance between time, scaffolding, and assessment when students encounter progressively richer visual outputs. The contribution is an exploratory educational case that informs instructional planning while making its evidential boundaries explicit.

2. Materials and Methods

2.1. Educational context and study scope

The intervention involved seven upper secondary classes from different schools and 152 students. Three classes were in the fourth year and four in the fifth year; class size was described as approximately 20–25 students. These contextual details locate the activity at an advanced stage of secondary education. They do not support comparisons between year groups because separate completion figures by class or year are not available. No further demographic breakdown is available for this analysis.

The study is treated as a classroom-based exploratory account. The available record comprises aggregate task completion percentages and a narrative description of classroom observation, discussion, and formative feedback. It does not provide comparison conditions, individual pre- and post-intervention scores, or the documentation needed to reconstruct a controlled evaluation. The analysis therefore describes outcomes under the reported classroom conditions. It does not estimate the intervention's causal effect on AI literacy or engagement.

The instructional sequence moved from introductory explanation to guided interaction with visual outputs and then to smartphone experimentation. The same sequence also shapes the reporting, but exposure to the tasks should not be assumed to have been equivalent in duration or support. Sessions lasted three hours and included ten-minute breaks. Completion was defined by obtaining a working prototype for the relevant task. The number of breaks and the distribution of active time across tasks are not specified, limiting interpretation of the final segmentation activity.

2.2. Computer vision activities and instructional mediation

The introductory phase compared the three types of computer vision output using concrete examples. The central distinction was between a category for the whole image, a category associated with an object location, and a region or mask within the image. The pathway sought to develop vocabulary precise enough to describe these differences. This focus is consistent with the broader aim of making core AI ideas accessible in school education, as articulated in work on K–12 AI guidelines (Touretzky et al., 2019).

In the guided phase, students interacted with graphical workflows described in the classroom account as drawing on the Ultralytics and YOLO-style ecosystem. These workflows presented images and model outputs, including labels, confidence scores, bounding boxes, and segmentation masks. The technical account is limited to the reported visual workflows; model variants and software configurations require further documentation for replication.

Classification provided an initial setting for examining the relationship between an image and an assigned label. The instructional emphasis was on observing the output, deciding whether it fitted the input, and discussing uncertainty or incorrect predictions. Detection added the interpretation of location. Students had to attend to both the predicted category and the region indicated by a bounding box. Segmentation then introduced the interpretation of masks and boundaries, requiring attention to how a region corresponded to an object or meaningful part of an image.

The mobile phase concerned object detection through a smartphone application. In the reported activity, students tested models with objects in the classroom environment and varied conditions such as distance, viewing angle, lighting, occlusion, background, and placement. These variations are described as classroom experimentation, rather than a controlled robustness test. No fixed test set, standardized acquisition protocol, or quantitative comparison across environmental conditions is available. The educational purpose was to make the dependence of predictions on the observed scene available for inspection and discussion.

2.3. Evidence and interpretation of completion

The principal quantitative evidence consists of the reported completion rates: 82% for classification, 43% for object detection, and 12% for segmentation. These values refer to obtaining a working prototype for the proposed task within the intervention. They are educational process indicators, not classifier accuracy, detection precision, or segmentation overlap metrics.

A separate indicator reports that 97% of the 43% of students who have been reported completed on object detection were able to apply the model through the smartphone application.

2.4. Ethics and availability of materials

The activity included discussion of privacy, image collection, personal data, and avoiding sensitive or identifiable visual content. Participants were informed of the activity.

The analysis is limited to aggregate outcomes and a narrative account of the intervention. Student-level records, original assessment materials, datasets, and detailed software configurations are not included in this report, which limits independent verification and replication.

3. Results

3.1. Completion of the three activities

Classification had the highest reported completion rate, at 82%. Object detection had a reported completion rate of 43%, and segmentation had the lowest reported rate, at 12%. This ordering describes completion under the available classroom conditions. It does not establish that 82% demonstrated comprehensive classification competence, that 43% mastered detection, or that the remaining students could not understand the tasks. The record does not provide item-level assessment evidence beyond the working-prototype completion criterion.

The classification activity was described as an accessible introduction to category assignment and prediction interpretation. Detection required attention to the additional spatial information in bounding boxes. Segmentation required interpreting image regions and boundaries. These differences describe the demands of the activities, but the aggregate percentages do not identify which specific demand prevented completion. A student might understand an output while needing help to complete an operational step, or complete a workflow without being able to explain its behavior independently.

The classroom account relates the low segmentation rate to the activity's greater operational requirements and the limited time available. It reports that segmentation required more guided steps and that some students did not finish within the allotted classroom time. This explanation is retained as a contextual interpretation. The available figures do not separate time limitation from task complexity, prior knowledge, order in the sequence, or differences in support. The segmentation result should therefore be reported as limited completion, with its cause remaining unresolved.

3.2. Smartphone application of object detection

The mobile phase yielded a reported application rate of 97% among students who worked on object detection. This indicates that smartphone use was widely achieved in that participant group according to the activity account. It is a distinct result from the 43% completion of the detection task. The numerical contrast is useful for considering different assessment targets, but its magnitude cannot be interpreted as an increase in success because the participant bases and criteria are not established as equivalent.

The mobile activity made detection outputs observable in the immediate classroom environment. Students could direct the camera toward objects and inspect predicted labels and bounding boxes as conditions changed. The classroom narrative describes this phase as meaningful for connecting AI explanations to practical experience. However, there are no comparative engagement scores, independent transfer tasks, or follow-up observations. The result supports a descriptive account of operational application; claims about durable conceptual transfer or increased engagement require further evidence.

3.3. Formative observations and model errors

The classroom account describes incorrect predictions, missed objects, unstable detections, and imprecise outputs as opportunities for discussion. Students were invited to consider whether an error might relate to the input image, object, background, acquisition conditions, or training data. These discussions connected practical experimentation with questions about the limits of AI. They are reported as features of the teaching activity rather than as evidence that a specified proportion of students acquired a particular explanation skill.

4. Discussion

4.1. Completion as a signal for differentiated scaffolding

The reported pattern suggests that an introductory computer vision pathway needs to allocate instructional resources by activity rather than treat all visual AI tasks as interchangeable. Classification, detection, and segmentation offer different representational demands. In this case, the completion ordering provides a practical signal for reviewing how much explanation, guided practice, and operational support each task receives. It does not supply a universal ranking of difficulty or justify removing segmentation from secondary education.

Classification can serve as an entry activity because the relationship between one image and one category is comparatively straightforward to present. Detection expands that relationship to include location, while segmentation requires attention to regions and boundaries. A future implementation could make these changes explicit by comparing outputs for the same input image. Students could explain what information is added at each step before attempting the next workflow. This is an instructional proposal arising from the case, not a procedure whose effectiveness has been tested here.

The 12% segmentation figure is especially important for planning. Allocating more time, introducing preparatory work on foreground and background, and separating mask interpretation from operational execution are reasonable responses to the reported constraints. Nevertheless, additional time should not be declared the

proven remedy. Its role would need to be examined through records of exposure, support, and completion under a revised design. The result calls for closer documentation of the pathway rather than a causal explanation based on the percentage alone.

4.2. Practical application and conceptual understanding

The smartphone finding illustrates the need to distinguish operating a tool from explaining it. Students may be able to launch an application and obtain a prediction while still needing guidance to interpret localization, uncertainty, or failure. Conversely, a student may explain the purpose of a detection output but be unable to finish a technical workflow within the session. The coexistence of different completion and application figures is therefore educationally informative even though the available record cannot link them at individual level.

An assessment approach consistent with this distinction would separate operational application, conceptual explanation, and critical evaluation. For example, obtaining a detection output would address tool use; explaining the label and box would address interpretation; testing a problematic scene and discussing alternative explanations would address evaluation. These are proposed assessment dimensions, not a reconstruction of the original rubric. Their purpose is to prevent one success indicator from standing in for the wider competence that the intervention aims to develop.

The distinction also fits the broader view of AI literacy and competency as extending beyond technical familiarity. Chiu et al. (2024) include technology, impact, ethics, collaboration, and self-reflection in their framework. The present case cannot assess that whole framework, but it makes a relevant design issue concrete: a usable interface opens participation while instructional questions determine whether students examine its outputs critically. Accessibility of the workflow and depth of understanding should be evaluated separately.

4.3. Engagement and the formative use of failure

Mobile experimentation offers a setting in which a prediction can be questioned through another attempt. Changing the viewing angle or background can give students an action to perform and an output to inspect. This creates opportunities for participation and inquiry. Yet an engaging task design is not equivalent to a demonstrated increase in engagement. Within the multidimensional framework of Fredricks et al. (2004), evidence of observable participation would still need to be distinguished from emotional involvement and sustained cognitive investment.

The reported use of errors is relevant because it redirects attention from obtaining a correct answer to examining the conditions of a prediction. A missed object can prompt comparison of scenes; an inaccurate box can prompt discussion of localization; a poor mask can prompt examination of boundaries. A teacher can use these outputs as formative material without claiming that students have diagnosed the model's internal cause. The pedagogical value lies in developing questions and testing explanations, including recognizing when available evidence does not resolve them.

Agency can likewise be treated as an instructional opportunity when students choose what to test and how to modify a scene. The current account does not measure autonomy or its development. It supports a design rationale for allowing

purposeful choices, accompanied by explanation and reflection. Future observation should record the choices students make, the reasons they offer, and whether they revise a test following feedback. Such evidence would be more informative than attributing agency to smartphone use alone.

4.4. Limitations and future development

Several limitations constrain the interpretation. The percentages lack confirmed denominators, counts, and detailed assessment descriptors. The record does not establish equivalent exposure across tasks or provide outcomes by class and school year. The absence of individual linked records prevents analysis of whether the same students completed detection and used the smartphone application. Differences in classroom context, prior experience, device access, and teacher support cannot be evaluated from the aggregate summary.

The proposed Arduino extension would connect computer vision with sensors, cameras, and tangible artifacts. A constructionist perspective provides a rationale for learning through the development and revision of meaningful computational artifacts (Papert, 1980). Here, physical computing remains future work. No Arduino prototype, sensor integration, or outcome is reported as part of the completed intervention. A future pathway should specify the feasible division of work between the sensing device and the system executing the vision model, rather than assume that all processing can run on an unspecified board.

5. Conclusions

This exploratory account describes a progressive computer vision pathway for 152 students across seven upper secondary classes. The reported completion rates were 82% for classification, 43% for detection, and 12% for segmentation; smartphone application was reported for 97% of the students who worked on detection. These indicators describe activity completion and practical use under the reported conditions. They do not establish learning gains, a causal effect on engagement, or a direct comparison between formal detection completion and mobile application.

The case supports reviewing how visual AI tasks are sequenced, scaffolded, and assessed. Classification provides an initial setting for label interpretation, detection connects category and location, and segmentation adds region and boundary reasoning. Mobile experimentation offers a concrete setting for inspecting predictions and discussing failure. The next step is to document time, participant bases, assessment criteria, and student explanations more precisely, so that practical participation can be connected to evidence of critical AI understanding.

References

Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open, 6*, 100171. https://doi.org/10.1016/j.caeo.2024.100171

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.

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*(1), 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795