

Document Title	Guidelines for integrating AI Evaluation exercises into educational courses, using our approach
Project Title and acronym	PINNACLE: <i>Public Engagement in AI Evaluation for Education and System Benchmarking</i>
Pillar	III. Transformation of RTDI System
Programme	BRIDGE2HORIZON
Grant Agreement	BRIDGE2HORIZON/0823E/0203
Deliverable No.	D26
Work package No.	WP6
Work package title	AI Evaluation in Education
Authors	E. Theophilou
Contributors	M. Kasinidou, E. Christoforou
Reviewers	J. Otterbacher
Status (D: draft; RD: revised draft; F: final)	F
File Name	D26_Guidelines_Educational_Integration.docx
Date	31 July 2026

Draft Versions - History of Document				
Version	Date	Authors / contributors	e-mail address	Notes / changes
1.0	1/7/26	E. Theophilou	emily.theophilou@ouc.ac.cy	Initial version
1.1	15/7/26	M. Kasinidou	maria.kasinidou@ouc.ac.cy	Additional content
1.2	25/7/26	E. Theophilou	emily.theophilou@ouc.ac.cy	Review version
1.3	31/7/26	J. Otterbacher	jahna.otterbacher@ouc.ac.cy	Final version

Abstract	
This deliverable presents our easy-to-follow guidelines for educators, who may be interested in introducing AI evaluation exercises into their own curricula with students.	
Keyword(s):	AI education, Educational guidelines, Onboarding educators

Contents

Contents	4
1. About this Deliverable	5
1.1. Role of the deliverable	5
1.2. Relationship to other PINNACLE deliverables	5
1.3. Structure of this deliverable	5

1. About this Deliverable

1.1. Role of the deliverable

This deliverable aims to distill the PINNACLE educational approach into an appealing, easy-to-follow set of guidelines for educators. It presents an overview of the PINNACLE goals, along with the basics as to how educators can adopt our approach with their own students.

1.2. Relationship to other PINNACLE deliverables

Deliverable D26 supports our efforts to disseminate the work of PINNACLE, and to grow our community of AI Evaluators. In particular, we hope that educators will become interested in our work, and engage with us in developing an increasing range of strategies and materials to raise the AI literacy of students, teachers and parents from diverse communities.

1.3. Structure of this deliverable

The deliverable consists of the guidelines, presented in a flyer format, as to promote easy sharing and distribution to interested parties.

Guidelines for Integrating AI Evaluation into Educational Courses



PINNACLE

Public Engagement in AI
Evaluation for Education &
System Benchmarking

As AI technologies rapidly reshape every domain of modern life, AI literacy can no longer remain confined to technical or computer science disciplines. AI literacy for the public requires moving beyond passive reading materials to **authentic, hands-on learning**. Integrating structured AI evaluation exercises into course curricula can shift students from passive software consumers into critical, discerning evaluators of algorithmic systems.

Our framework introduces two educational approaches to increase AI literacy among the general public, validated empirically through a study of 243 participants.¹

Autonomous Discovery

Students independently identify, classify, and critique AI tools in their daily lives, connecting academic concepts directly to real-world experience.

Guided Mastery

Step-by-step instructions guide students through targeted evaluations of specific AI chatbots, building systematic auditing skills aligned with weekly course topics.

"The exercises and activities were the best part of the training."

"...the educational material helped in better understanding and consolidating each unit. The activities and quizzes were particularly helpful and enhanced the learning experience..."



1. Kasinidou, M., Kleanthous, S., Otterbacher, J., & Christoforou, E. (2026). Educating the Public in Artificial Intelligence: Insights from a Large-Scale Course. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(47), 40617–40625. <https://doi.org/10.1609/aaai.v40i47.41508>

Autonomous Discovery

This approach empowers students to connect academic concepts directly to their daily lives by independently identifying and evaluating **AI tools they use on a daily basis**. By analyzing applications from their own routines, they transform everyday digital interactions into critical learning opportunities.

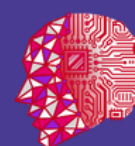


Classroom Implementation

- **Everyday AI Discovery:** Students identify and select a real-world AI application they encounter in their daily lives that uses the AI technology covered in the current unit (e.g., personalization, computer vision, recommendation systems, or generative AI).
- **Technology Classification:** Students analyze these applications, classify the underlying AI technology, and record basic observations.
- **Targeted Evaluation:** Students critically assess their selected application across key dimensions aligned with the course learning goals.
- **Reflective Synthesis:** Students share observations in class, while the instructor facilitates a guided discussion to deepen critical thinking around everyday technology use.

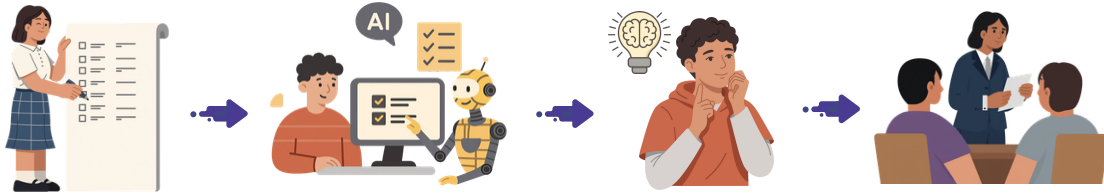
Tip!

Encourage students to look beyond chatbots toward routine digital tools (such as Spotify recommendations, email spam filters, or navigation apps). If students struggle to connect abstract weekly topics to their daily lives, scaffold the exercise by providing a curated list of example applications to spark their discovery.



Guided Mastery

This approach provides a fully scaffolded environment where students interact with **AI chatbots** using predefined prompts and built-in critical checkpoints. By pairing predefined teacher prompts with real-time critical evaluation checkpoints, students methodically test chatbots to uncover core AI behaviors, limitations, and logic gaps.



Classroom Implementation

- **Guided Setup:** Students receive step-by-step instructions with a predefined instructor prompt designed to guide them toward targeted learning outcomes.
- **Guided Interaction:** Students engage with their selected chatbot following the prescribed interaction steps.
- **In-Line Critical Evaluation:** Embedded checkpoints prompt students to critically assess the accuracy, logic, and limitations of AI outputs in real time.
- **Comparative Assessment:** Students complete a final evaluation of the chatbot to benchmark its performance against previous exercises.
- **Classroom Debrief:** The instructor facilitates a debrief to highlight critical aspects of AI behavior, discuss findings, and summarize core learning outcomes.

Tip!

Design predefined prompts that are likely to trigger subtle hallucinations or logic gaps. This ensures the built-in checkpoints actually give students something critical to catch.