AI and Truth: Seeing is not Always Believing

Competency

Educators connect AI's social and ethical impacts to its technical design, capabilities, and limitations, helping learners understand how AI's strengths and weaknesses shape its use and societal trust.

Key Method

The educator designs an inquiry-based learning experience where teachers or students grow understanding on AI-generated outputs to evaluate accuracy, bias, and misinformation (including deepfakes). Participants analyze how AI might empower and challenge truth, equity, and trust in human systems.

Method Components

As artificial intelligence gains prominence in education, educators must move beyond simply adopting a new tool to thoughtful, informed, and ethical integration. Effective AI use in the classroom requires careful selection of tools, a strong foundation in AI literacy, and plan for its intentional use. Educators play a central role in modeling responsible digital citizenship, helping students understand how AI systems function, and guiding them to analyze both the benefits and limitations of these technologies.

AI Tool Use Considerations

Selecting AI tools for a classroom requires a systematic evaluation of ease of use, learning goals, cost-effectiveness, data privacy, and scalability. Staff should prioritize the use of district-vetted AI tools and tools recommended by trusted sources.

The effective implementation of AI literacy requires high-quality, ongoing professional learning for educators. Educators must critically evaluate AI resources themselves before integrating them into the classroom. Educators should be aware of biases and inaccuracies that can arise in AI outputs and be prepared to include human verification as a critical step in AI use.

Foundational AI Literacy for Educators

Foundational AI literacy empowers educators to understand how tools actually work. At its core, AI relies on data, algorithms, and statistical inference to generate responses, meaning its outputs are shaped by the information it has been trained on and the patterns it detects. Understanding the basics of AI generative processes equips educators to support their students as they use AI. The limitations of AI must be part of their instruction. AI literacy should include information on the many applications of AI, including deepfakes. Deepfakes are synthetic images or sounds created by AI that appear realistic but do not actually exist. Trained on large datasets of images, video, and audio, these applications do not "understand" truth or intent—they merely predict what a realistic image, voice, or video should look like. Understanding how they are created helps educators guide students in their understanding of how AI can be used.

Connecting Technical Design to Social Impact

Educators can help students see that AI's technical design is directly linked to its social impact. By examining how features like speed, scale, and generative creativity emerge from how models are built and trained, learners can also understand why these same strengths also produce limitations such as bias, misinformation, and declining public trust. This phenomenon of misinformation and bias has been termed "hallucinations," though the term is increasingly falling out of use as AI appears to become more "human." These errors are more accurately attributed to poor data for modeling and to algorithmic patterns. Avoiding the word 'hallucinations' recognizes that mental health conditions are uniquely human, not possible in AI.

Through guided analysis, educators can walk students along the pathway from **Design → Capability → Limitation → Societal Impact**, showing that each technical choice—how data is selected, how algorithms operate, and how outputs are generated—shapes real-world consequences. This approach empowers learners to think critically about both the potential and the risks of AI systems. As AI grows in its use and popularity, there are many possibilities to connect the design to societal impacts, such as chatbots, grading systems, self-driving cars, and pattern recognition.

One example of how an educator may share this process with their students is:

Example-Facial Recognition with AI

The teacher starts by explaining the design in simple terms:

Design

- Facial recognition AI is trained by feeding a computer model thousands or millions of labeled facial images. Engineers choose what data to include, how the model learns patterns, and what counts as a "match."
- Class discussion surrounding who chose the training data? What kinds of faces might be over- or under-represented?

Capabilities

- Due to its training, AI can quickly identify people's faces, even in large crowds.
- Class discussion regarding why this information may be helpful
    a. Where would this be useful, such as in missing-person cases?
    b. Why is speed and computer use supportive?

Limitations

- What are some risks of this use?
- What are some limits, such as diversity or context?

Societal Impacts

- What are some errors that the system could make?
- How do these systems affect privacy or trust in institutions?
- Who benefits and who might be harmed?

Example-Deepfakes with AI

Design

- AI models are trained on massive datasets of faces, voices, and video.
- Engineers decide which data is included, how to optimize realism, and what the outputs will look like.

Capabilities

- AI generates realistic videos, images, or audio of people who were not part of the production.
- These can be created quickly and then shared widely.

Limitations

- AI does not understand truth, context, or harm.
- Accuracy is based not on realism or facts, but on output that meets criteria
- Biases in training data affect who is most easily impersonated

Societal Impacts

- Misinformation and erosion of trust.
- Individuals can be harmed through impersonations created without their knowledge or permission.
- Challenges to many fields, such as journalism, democracy, or education?

These limits on AI training can have profound impacts on society as a whole and on affected groups to a greater extent. Lessons such as this connect design to real outcomes. Educators should understand that limitations are not accidents; they are

a result of the design and data. This understanding is critical to building AI literacy skills.

Applying Critical Evaluation and Problem-Solving

Educators practice problem-solving by determining when and how to use AI responsibly. They can then promote students' responsible use of AI by modeling how to:

- Assess its **capabilities**, **limitations**, and **potential risks** rather than avoiding or overusing it
- Reflect on the use of AI as a tool for inquiry and analysis, not a substitute for human reasoning

Pedagogical Integration and Practice

Human beings must stay at the center of AI's use as a tool, and educators must be ambassadors of effective AI use in the classroom. That requires mindfully integrating AI into the curriculum, designing lessons to teach ethical reflection, and promoting discernment of digital outputs.

Using the **Human Inquiry–AI–Human Empowerment (H–AI–H)** framework is one way to support ethical and effective AI use..

- **Human Inquiry:** Begin with human curiosity and questioning.
- **AI:** Use AI tools to analyze, generate ideas, or test assumptions.
- **Human Empowerment:** End with human interpretation, critique, and ethical decision-making.

This framework keeps AI in the middle of a request, with AI generating a basic starting point that is then verified and refined by a human. According to Tasseyo De Pietro (see resource below), "The strongest results appear only when a human begins with intention, the AI expands the territory, and then the human returns to carve something real out of the noise."

Cultivating Empathy and Responsibility

Cultivating empathy and responsible AI use begins with helping educators understand how AI systems can impact different communities in different ways. By examining how training data shapes an AI's outputs, educators can more critically evaluate whether these systems reinforce existing biases or help reduce social inequities. This awareness encourages thoughtful choices about when and how AI should be used in learning environments. Equally important, educators commit to transparent and accountable AI practices—openly communicating with students, families, and colleagues about how AI tools are used, what their limitations are, and how decisions are made. Through empathy, reflection, and clear communication, educators model responsible AI use that prioritizes fairness and trust.

Limitations of AI in Education

By recognizing how an AI system's design directly shapes its social, ethical, and cultural impacts, educators model the critical inquiry needed to analyze AI-generated content with care. They apply thoughtful instructional design to help students build their own AI literacy—learning not just how to use AI tools, but how to question them. At the same time, ethical reflection remains essential, prompting ongoing attention to issues such as bias, misinformation, and responsible use. When this work is presented with clarity and professionalism, it reinforces a learning environment where AI is approached thoughtfully, transparently, and with an unwavering commitment to student understanding and well-being.

As you work through this micro-credential, keep these points in mind:

- **Integration:** As you integrate AI into your classroom, plan for ways to connect AI's technical design to its social, ethical, or cultural impacts.
- **Inquiry:** Beginning with human curiosity, your work should always focus on critical thinking skills and on identifying the purpose behind its use.
- **Application:** Appreciating the power and limitations of AI equips educators with the understanding needed to build AI literacy in their students.
- **Ethical Reflection:** Effective and ethical reflection of AI use must always include issues of bias, misinformation, and responsible AI use

Supporting Rationale and Research

Beheshti, A., & Kerridge, I. (2025). Understanding the Artificial Intelligence Revolution and its Ethical Implications. Journal of bioethical inquiry, 22(3), 497–505. https://doi.org/10.1007/s11673-025-10427-6

Gu, J., & Oelke, D. (2019). *Understanding bias in machine learning*. arXiv. https://doi.org/10.48550/arXiv.1909.01866

Melamed, C. (2025, December 8). *AI: Opportunity, risk, and a tough test for global cooperation*. United Nations Foundation. https://unfoundation.org/blog/post/ai-opportunity-risk-and-a-tough-test-for-global-cooperation/

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693/PDF/386693eng.pdf

Resources

AI Support for Educators
Microsoft Elevate for Educators

Foundational AI Literacy for Educators
AI Glossary of Terms

Common Sense Education AI Literacy Lesson for 6-12

Videos: Deepfake Detective — PBS Nova

Connecting Technical Design to Social Impact
MIT Media Lab (2021). Moral Machine Experiment.

Videos: The Social Dilemma: AI and Attention — Netflix Documentary

Applying Critical Evaluation and Problem-Solving
Helping Students Check for Bias in AI Outputs— Edutopia

SWOT Template

Watch out for false claims of deepfakes, and actual deepfakes, this election year

Pedagogical Integration and Practice
H-AI-H. The Creative Loop That Makes AI Truly... | by Tassyo De Pietro | Medium

ISTE AI Lessons

Cultivating Empathy and Responsibility
Recommendation on the Ethics of Artificial Intelligence - UNESCO

Submission Guidelines and Evaluation Criteria

To earn this micro-credential, you must receive a passing score in Parts 1 and 3, and be proficient in all components in Part 2.

Part 1. Overview Questions (Provides Context)

Answer the following contextual questions to help our assessor understand your local context.

Please do not include any information that will make you identifiable to your reviewers.

1. Describe your current role in education and how AI is used in your classroom, school, or district.
2. What grade level and subject do you teach? Include demographics and special considerations related to AI (e.g., access, prior student experience with AI tools).
3. In what way are AI tools taught about and used in your building or district? Describe your use both as a teacher and how your students are taught about and encouraged to use AI.

Criteria for a passing response:

The response provides a clear description of the educator's role within the educational system and the demographics. The educator understands how AI is currently being taught about and used by both educators and students.

Part 2. Work Examples/Artifacts/Evidence

To earn this micro-credential, please submit the following **three artifacts** as evidence of your learning. See the Rubric for the passing score. *Please do not include any information that will make you or your students identifiable to your reviewers*.

Artifact 1: SWOT analysis

Create an analysis using a SWOT diagram (Strengths, Weaknesses, Opportunities, and Threats) regarding AI. Using the template in the resource section or another SWOT format, create a diagram where you list **three to five** components in each section. After completing the diagram, write an analysis of AI using the following points.

- **Strengths**: What **benefits or advantages** do you see in AI's pattern recognition and generative modeling for learners and society?
- **Weaknesses**: What **risks or limitations** do you anticipate with these AI technologies?
- **Opportunities**: How might the ethical and safe use of AI improve teaching and learning?

- **Threats**: What potential **challenges or negative impacts** might AI pose to equity, trust, or accuracy in education?

Share your SWOT diagram and analysis with the evaluator.

Artifact 2 Lesson Plan

Create a lesson plan or adapt lessons to teach your students how to evaluate AI outputs and how these outputs affect trust in society. This lesson should explicitly teach how to **critically analyze outputs and raise awareness of AI's limitations.** Include how these limitations could influence trust in society.

Your lesson plan should include each of the following:

1. Grade level/audience, time frame, local context, and standards alignment if applicable.
2. Clear objectives and vocabulary linking AI's design (e.g., how it "works," "sees," "writes," or "creates") to societal impact (e.g., bias, misinformation, trust).
3. Include modifications or adjustments for special learners (SPed, MLs, Gifted, etc.)
4. Include an activity that engages learners in critical analysis of AI-generated outputs.

Share this lesson plan/unit with the above information clearly identified, or add an overview explaining how each of these concepts will be covered in the lesson.

Artifact 3: Analyzing AI Output — Bias, Misinformation, or Public Trust

Step 1: Select a Focus Topic
Choose one of the following topics to guide your analysis of AI output:

- Bias
- Misinformation (this may include deepfakes)
- Impact of AI on Public Trust

Step 2: Written Reflection (100–300 words)
Write a short reflection in which you:

- Explain why you chose this topic (bias, misinformation, or impact on trust)
- Describe the specific situation, tool, or product where AI output was examined
- Briefly explain what you were looking for in the AI output related to your topic

This reflection should set the context for your analysis and clearly explain what learners or educators were asked to do.

Step 3: Demonstration of Critical Analysis (Choose One Option)

Create **one** of the following artifacts that demonstrates how you, your colleagues, or your students critically analyzed AI output related to your chosen topic.

Option A: Recorded Reflection (3–5 minutes)
Submit a video, audio recording, or similar format that captures learners or teachers:

- Interacting with AI-generated output
- Discussing observations, concerns, or patterns related to the selected topic
- Explaining how they evaluated the AI's responses

Option B: Digital Portfolio (10–20 slides)
Submit a digital portfolio (e.g., photo essay, slideshow, or screenshots) that includes:

- Visual evidence of interaction with AI output
- Clear captions explaining learner actions, analysis, and reasoning
- Reflections on what was discovered or learned about the AI topic

Submission Expectations
Your submission should:

- Show **thoughtful engagement** with AI-generated content
- Provide **clear evidence of analysis**, not just description
- Explicitly **connect** observations back to your chosen topic (bias, misinformation, or public trust)

Be sure no identifying information is included (faces, names, schools, states, etc.).

Part 3. Rubric

	Proficient	**Basic**	**Developing**
Artifact 1: SWOT diagram and analysis	Includes a SWOT diagram with three - five items identified for each quadrant Analysis clearly identifies a minimum of three relevant "strengths and weaknesses" of AI's pattern recognition and generative modeling for learners and society Articulates meaningful "opportunities" focused on the ethical and safe use of AI to enhance teaching and learning Recognizes potential "threats" AI may pose to equity, trust, or accuracy in education	Includes a SWOT diagram with two - three items for each quadrant The analysis identifies one - two "strengths and weaknesses" of AI's pattern recognition and generative modeling for learners and society Implies an "opportunity" for ethical and safe use of AI, but is not explicitly described. Includes potential "threats" that AI may pose, but does not specifically refer to equity, trust, or accuracy in education	Includes a SWOT diagram with only one item in each quadrant May identify one "strength or weakness "of AI's modeling, but does not significantly analyze the impact on learners or society Does not mention an "opportunity" for ethical and safe use of AI Does not mention potential "threats" that AI may pose
Artifact 2: Lesson or Unit Plan	The lesson plan is submitted with the following either highlighted or specified in an overview document. Grade level, audience, time frame, local context, and relevant standards	The lesson plan is submitted, but includes two - three of the following points: Grade level, audience, time frame, local context, and relevant standards	The lesson plan is submitted, but the includes one or none of the following points: Grade level, audience, time frame, local context, and relevant standards

	Clear objectives for lessons and key vocabulary that connect AI's technical design to its societal impacts Includes an activity where students analyze AI-generated output Instruction includes how bias/misinformation can undermine trust in systems	Clear objectives for lessons and key vocabulary that connect AI's technical design to its societal impacts Includes an activity where students analyze AI-generated output Instruction on how bias/misinformation can undermine trust in systems	Clear objectives for lessons and key vocabulary that connect AI's technical design to its societal impacts. Includes an activity where students analyze AI-generated output Instruction on how bias/misinformation can undermine trust in systems
Artifact 3: Evidence of Practice	**Reflection** includes: Topic selected (bias, misinformation, impact of AI on public trust) Why you selected this topic: - Identify the situation, tool, or product examined - What were you looking for related to the topic **Submission** includes either a three-five-minute reflection or a 10-20 page slide presentation. The submission includes: Learners discussing or sharing insights related to the topic	**Reflection** includes: Topic selected (bias, misinformation, impact of AI on public trust) *Reflection is not clear on these points:* Why you selected this topic: - Identify the situation, tool, or product examined - What were you looking for related to the topic **Submission** includes either a three-five-minute reflection or a 10-20 page slide presentation. The submission includes: Learners discussing or sharing insights related to the topic	**Reflection does not** include: Topic selected (bias, misinformation, impact of AI on public trust) Why you selectde this topic: - Identify the situation, tool, or product examined - What were you looking for related to the topic **Or the reflection is missing** **Submission** includes a presentation of fewer than three minutes or fewer than 10 pages of slide presentation. The submission includes one or two of the following:

	Thoughtful engagement by learners Evidence of understanding of the selected topic Professional language and no identifying features shared.	Thoughtful engagement by learners is not clearly demonstrated Little evidence of understanding of the selected topic Professional language and no identifying features shared.	Learner discussing or sharing insights into learning Engagement by learners Little evidence of understanding of the impact of AI on societal trust. Professional language and no identifying features shared

Part 3 Reflection

Please do not include any information that will make you identifiable to your reviewers.

For tips on writing a good reflection, review the following resource:
How Do I Write a Good Personal Reflection?

1. What challenges and opportunities do you see about how **AI might** challenge truth, equity, and trust?

2. Looking ahead, how do you plan to sustain or expand your AI literacy and integrate new strategies that promote responsible and ethical AI use in your educational context?

Criteria for a passing response:

- Explains how adapting the lessons deepened their insights into AI's technical aspects and societal effects (bias, misinformation, trust).
- Uses specific examples from objectives, vocabulary, or activities.

Forward-Thinking & Sustainability:

- Outlines actionable next steps for continuing AI literacy initiatives.
- Describes plans to model critical, ethical AI use and expand collaborations in future practice.

Reflection is clear, specific, and connects theory to actionable practice.