When Technology Isn't Neutral: Understanding AI and Bias

Competency

Learners examine how AI systems reflect human decisions and data, identify how bias in data or design can lead to unfair or inaccurate outcomes, and apply this understanding to support students in critically analyzing AI in the classroom.

Key Method

Learners explore how artificial intelligence systems—such as chatbots, summarizers, or recommendation tools—reflect the data and design choices that underpin them. By comparing outputs from AI tools built on different information sources or perspectives, learners analyze how bias arises and shapes responses. They then design and facilitate a classroom experience that helps students recognize, question, and discuss bias in AI systems.

Method Components

AI Systems and Bias

Artificial intelligence may appear complex or mysterious, but behind every model lies a foundation of human choices: humans provide the data; humans write and program the instructions, and humans create the intended goals. By raising awareness of how human influence underlies every AI output, educators can help students deepen their understanding of AI systems.

AI models are trained on large data sets, which are human-generated collections of words, images, videos, code, and sounds. The quality of these data sets is essential. The AI's output depends on the quality of these datasets. If the data is partial, unbalanced, or prejudiced, the AI-generated response will reflect those flaws.

When educators and students alike understand how AI is trained, they can identify how and why these tools sometimes provide confident, yet incorrect or biased, answers.

Students interact with AI every day, whether in autocorrect, search suggestions, or recommended lists on streaming platforms. Educators who explore these common examples of AI use help students recognize AI is not neutral: It responds to both the information it was given and the behaviors it has learned. These insights open the door to more profound questions: How do these systems shape what we read, watch, or believe? How do they shape the ways we learn, communicate, or perceive others?

With a foundational understanding, educators can lead informed discussions about the opportunities and pitfalls of AI. Students become more critical thinkers, analysts, and users of technology when they recognize that technology is a human invention, one that can reflect and amplify bias.

Investigating Sources of Bias

Educators and students who understand that AI systems reflect human decisions recognize that bias can enter the system at multiple points. For example, bias may arise when developers or data labelers make decisions based on the datasets used to train AI models, or when users interact with the tool after deployment. Bias in AI systems is not merely a technical flaw created by "the AI," it is a social and ethical issue rooted in many human and structural decisions.

AI systems learn from datasets created by humans who naturally and unconsciously consider personal perspectives, experiences, and assumptions when producing information. If a data set overrepresents certain groups or viewpoints, the system will favor those patterns during model training. Even thoughtful design decisions, such as how to label data or adjust prompts to improve accuracy, can unintentionally reinforce imbalance. When AI is trained on examples that limit cultural perspectives or reinforce common stereotypes, the system may reproduce those patterns in its responses.

Educators can use real-life examples, such as image recognition systems that misidentify darker skin tones or text generators that reproduce stereotypes, to demonstrate how errors and bias stem from human interaction and decision-making. By creating a tangible connection, educators can help students connect technological processes to broader issues of representation, fairness, and accountability.

In the classroom, educators can explore bias safely through age-appropriate activities that focus on observable differences rather than controversial topics. For example, students may compare creative prompts, AI responses to factual questions, or historical summaries for differences in tone, accuracy, or framing. These activities foster curiosity and reflection. Analyzing these various sources of

bias may prompt students and educators to pose reflective questions about technology. By exploring bias, educators reinforce an important message: Fairness in AI requires intentional design, diverse data, and human oversight—all key components to using digital tools responsibly.

Designing and Comparing AI Systems

Educators and students can better understand AI bias when they see it firsthand. By designing a simple comparison between two AI systems—or between distinct contexts within a single tool—educators can demonstrate how datasets, design decisions, and prompts shape the AI-generated responses we receive.

To begin the comparison, educators select two AI systems or tools. They may choose two AI platforms, like NotebookLM and ChatGPT, or experiment within a single system by creating custom bots, setting distinct parameters, or adjusting the information sources each uses. AI gathers contrasting materials for each system, such as government reports versus advocacy articles, to help reveal how data sets, based on human input and decisions, influence output.

Once the setup is ready, educators pose the same prompt to both systems and compare the responses, which may differ in tone, details, or framing. Some responses may emphasize certain viewpoints, omit central perspectives, or show patterns indicative of the training data's influence. These variations become key entry points for discussion and analysis.

By comparing AI systems in this way, educators can see how algorithms mirror human priorities and data sources. This exercise shows that bias does not arise by accident but is a predictable consequence of the information and design decisions that shape each model. Educators can use this approach to safely guide students through structured inquiry into AI, supporting digital literacy, critical thinking, and ethical awareness.

Conducting Classroom Inquiry or Simulation

Once educators understand how bias can emerge within AI systems, they can guide students through their own investigations. Through inquiry-based learning, students observe, compare, and question AI-generated information in a structured, supportive environment.

In this stage, educators design lessons that allow students to lead the inquiry. The students use the same prompt to compare outputs from two AI systems—or two versions of the same system—by recording what they notice, discussing patterns, and considering what might have influenced each system's response. This approach encourages students to look beyond surface-level accuracy and think critically about how technology can frame information.

As students share their findings, educators facilitate discussion around what fairness, accuracy, and objectivity mean in the context of AI. Guiding questions may help students connect their observations to broader issues, such as how algorithms might shape what people read or believe. When discussions focus on evidence from the AI-generated output rather than personal or sensitive topics, educators maintain a reflective and respectful learning environment.

To consolidate learning, educators lead a debrief for students to reflect on how bias can appear in unexpected ways. Educators reinforce key takeaways about transparency, data diversity, and human oversight. When framing this exercise as an inquiry rather than a debate, educators foster curiosity, empathy, and critical digital literacy—essential skills for navigating a world increasingly influenced by AI.

Sharing Professional Reflections

As educators reflect on what students learned and their own discoveries, they further develop their understanding of how AI bias connects to broader issues of equity and representation. They can extend the impact beyond a single classroom by sharing those insights with colleagues, contributing to a collective understanding of responsible AI use in education.

Educators start by reviewing student reflections and work samples, looking for patterns: What did students notice about fairness or accuracy? How did the comparison exercise shift their awareness? By identifying these trends, educators can gain rich insights into how students process complex ideas about technology and bias.

Educators then reflect on how such results shape their professional practice, including new ways to choose and frame AI tools for classroom use, as well as strategies for helping students scrutinize and verify information generated by AI. Some educators may also want to collaborate with colleagues on shared lesson models, discussions of safe classroom practices, or the development of examples of AI activities that foster critical thinking.

Educators build collective capacity by sharing what they have learned with their colleagues, whether in a staff meeting or a professional learning community. By having these conversations, educators help demystify AI, moving the profession toward thoughtful and informed use rather than reactive avoidance or uncritical adoption. Recognizing bias in AI is about more than just technology. It is about collectively committing to fairness, transparency, and the ethical use of emerging AI tools in education.

Supporting Rationale and Research

Bond, Melissa, et al. “A Meta Systematic Review of Artificial Intelligence in Higher Education: A Call for Increased Ethics, Collaboration, and Rigour.” *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, 19 Jan. 2024, https://doi.org/10.1186/s41239-023-00436-z.

Ferrara, Emilio. “Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies.” *Sci*, vol. 6, no. 1, 25 Dec. 2023, p. 3, www.mdpi.com/2413-4155/6/1/3, https://doi.org/10.3390/sci6010003.

Idowu, Jamiu Adekunle. “Debiasing Education Algorithms.” *International Journal of Artificial Intelligence in Education*, vol. 34, no. 4, 4 Jan. 2024, pp. 1510–1540, https://doi.org/10.1007/s40593-023-00389-4.

Ka Yuk Chan, Cecilia. “A Comprehensive AI Policy Education Framework for University Teaching and Learning.” *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, 7 July 2023, https://doi.org/10.1186/s41239-023-00408-3.

Resources

AI Support for Educators

Microsoft Elevate for Educators

AI Glossary of Terms

Glossary

AI Systems

Video: AI, Machine Learning, Deep Learning, and Generative AI Explained

Machine Learning Foundational Courses

Artificial Intelligence (AI) K-12 Skills Progression Guide for Educators

AI Guidance for Schools Toolkit

AI Literacy 101: A Beginner-Friendly Guide for Educators

7 AI Tools That Help Teachers Work More Efficiently

AI in Education Hub

AI Considerations for Teaching and Learning

Bias in AI

Does AI Have a Bias Problem?

Thinking About Equity and Bias in AI

Navigating the Promise and Peril of AI for Students of Color

How to Combat AI Bias in the Classroom

Real or Not Quiz

Video: We Need to Talk About AI Bias

Video: Teacher Voices on AI & the Risk of Bias: What Teachers Want and What They Worry About

Mitigating Bias in Artificial Intelligence

Classroom Inquiry or Simulation

Helping Students Understand the Biases in Generative AI

Hands-On AI Project for the Classroom

Unlock Generative AI Safely and Responsibly—Classroom Toolkit

Real, Fake, or Deepfake? This Lesson Helps Students Decide

Teaching the Environmental Impact of AI Through PBL

Guiding Students to Develop AI Literacy

AI for Education: How Teachers Can Leverage AI Ethically to Enrich Their Curriculum

Professional Sharing

Meaningful Professional Development on AI

AI Teaching Strategies: Having Conversations with Students

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)

(300–400 words)
Your overview provides context for reviewers, helping them understand your instructional choices, student population, and reflections on learning. *Please do not include any information that will make you or your students identifiable to reviewers.*

Answer the following prompts.

1. Describe your teaching context, including grade level, subject area, general student demographics, and the number of students who are emergent multilingual learners, students with IEPs, or students with other specialized instructional considerations. Do not include any personally identifying information. How do these factors influence your perspective on ethical AI use?

2. Why are you exploring AI bias with your students? What is your familiarity with AI?

3. Describe your students' current use of AI. What successes and challenges have you faced with student use?

Passing:
To receive a passing score for Part 1, you must answer all questions clearly and thoughtfully and describe your teaching context without revealing identifying information.

A passing response:
- States a clear instructional purpose for exploring AI bias and your current familiarity with AI;

- Describes how your students currently use AI, noting at least one success and one challenge; and
- Provides enough context for reviewers to understand your perspective and instructional choices.

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: Lesson Plan

Task 1: Create a Lesson Plan
Create a lesson plan that helps students compare AI outputs and identify possible bias. The lesson plan should provide enough detail for a reviewer to understand how you designed and implemented the activity.

The plan must include:

- A brief focus statement (one or two sentences) that summarizes the purpose of the lesson at the top of the document;
- Title, intended grade level, topic, and subject area (for example, literacy, math, science, etc.);
- Learning objectives aligned with your content standards and the micro-credential's competency.
- AI comparison setup (for example, NotebookLM, ChatGPT, Perplexity, MagicSchool custom bots, or other tools and technologies that are available in your educational setting);
- Core prompt used for both AI systems.
  A core prompt is the single prompt you use with both AI systems to ensure a fair comparison and that the outputs relate to the same task. The core prompt should be identical (or as close as possible) across tools to ensure that any differences in the outputs reflect the AI systems themselves rather than differences in what they were asked to do.
- Implementation pathway (students use both AI systems, split-class model, or teacher-facilitated simulation);
  A split-class model means the class is divided into two groups, with each

group using a different AI tool. After working separately, students compare results, discuss differences, or share insights during the learning activity.

- Lesson sequence, with learning strategies, timing, materials, discussion prompts, and student reflection steps;
- Accommodations and modifications to your lesson for emergent multilingual learners and students with IEPs, gifted and talented, or other specific learning characteristics that require specialized instructional considerations;
- Evidence of alignment to your curriculum or standards; and
- Assessment or reflection of your students' understanding of bias tied to the competency.
  While it's ideal for students to interact directly with AI tools, this is not required. If your district restricts AI use, you may instead have students analyze AI responses you generate to meet the same learning goals.

Tip: Keep the focus on bias within AI systems. Use neutral, standards-aligned topics. Avoid themes that are political, advocacy-based, or identity-driven.

For your submission, compile all parts of your lesson plan into a single document (PDF preferred). *Submissions should not include any identifying information about you or your students.*

Task 2: Share Your Lesson Plan and Results
Share your lesson plan and the results of the classroom activity with a colleague, professional learning community (PLC), or professional team. In your results summary, include what happened, what students noticed, and what you observed. You will reflect on this professional sharing in Part 3: Reflection.

Artifact 2: Student Note Catcher/Worksheet/Outline

Create a tool for students to capture their observations and reflections during the lesson. Your design can take many forms, including, but not limited to, a worksheet, graphic organizer, or digital template.

Ensure the design clearly shows how the tool helps students analyze and reflect on AI-generated information. Examples include embedded guiding questions, checklists, or scaffolds in the tool that prompt students to examine for accuracy, bias, and/or missing perspectives. The tool must include space for students to record the prompts or questions they use when comparing AI outputs.

Artifact 3: Student Work Documentation

Submit three anonymized student work samples that demonstrate how they used the tool (created for Artifact 2) to examine differences between AI-generated responses and reflect on possible bias. Combine the student samples and your summary document into a single file (PDF preferred), with your summary as the first page.

Student work should provide evidence of one or more of the following activities:

- Identifies how AI outputs differ and why those differences matter;
- Notes possible sources of bias, missing perspectives, or inaccuracies;
- Explains how AI systems reflect human-made data and design choices; and/or
- Uses examples to discuss fairness, accuracy, or representation.

Write a summary (300–500 words) of student work samples. The summary must include:

- A brief description of the activity and what students were asked to do;
- Observed patterns or themes in the three student responses;
- Insights about how students perceived bias or fairness in the AI outputs; and
- Specific examples from the student work samples that illustrate your responses to the points above.

Part 2. Rubric

	Proficient	Basic	Developing
Artifact 1: Lesson Plan	The lesson plan is complete, well-organized, and clearly aligned to both curriculum standards and the micro-credential competency. The logical sequence of learning activities shows how students will explore and reflect on bias in AI systems. The implementation pathway and assessment plan clearly demonstrate intentional support for student understanding and reflection.	The lesson plan includes most but not all required components and shows some alignment to curriculum standards and the micro-credential competency. The sequence of learning activities addresses AI bias but lacks full clarity or cohesion. The implementation details or assessment plans are not clearly connected to student understanding or reflection.	The lesson plan is missing key components or lacks alignment to standards and the micro-credential competency. The learning activities are incomplete, unclear, or do not show how students will explore bias in AI systems. The lesson plan has limited or no mention of implementation and assessment plans.
Artifact 2: Student Tool	The tool is well-structured and clearly supports students in comparing AI outputs. The tool includes prompts or guiding questions that help students identify, document, and reflect on accuracy, bias, or missing perspectives in AI-generated responses. The tool clearly shows how the educator supports student	The tool is generally understandable but may lack clarity or organization, making it more difficult for some students to compare AI outputs. Prompts or questions may be broad or only partially guide students to examine accuracy, bias, or missing perspectives. The tool includes some instructional support, but scaffolds may be minimal,	The tool is unclear, disorganized, or does not meaningfully support the comparison of AI outputs. The prompts or questions are missing, unclear, or unrelated to the analysis of AI responses. The tool does not show meaningful instructional support or include structures that guide student analysis.

	analysis through embedded instructional moves, such as guiding questions, checklists, or reflection prompts. The tool incorporates purposeful scaffolds that foster critical thinking and inclusive discussion into the design	inconsistent, or not clearly connected to student analysis. Scaffolds may be limited or inconsistently support critical or inclusive discussion.	The scaffolds are missing, unclear, or do not support critical thinking or inclusive discussion.
Artifact 3: Student Work Samples	The submission includes a clear educator summary and three anonymized student work samples. The summary synthesizes student findings, highlights key patterns in how students recognized or discussed bias, and cites specific examples from the samples. Student work demonstrates thoughtful engagement with AI comparisons and shows evidence of learning, such as identifying differences between outputs, noticing possible bias, or explaining why those differences matter.	The submission includes an educator summary and some student information, but may lack sufficient detail or clarity. The summary identifies general themes but offers limited insight into student thinking or evidence of learning. Student samples show participation but limited analysis of AI bias.	The submission is incomplete or missing required elements (e.g., fewer than 3 samples or no educator summary). The summary lacks synthesis or evidence of student learning. Student samples show minimal engagement or understanding of AI bias.

Part 3 Reflection

(400–600 words)
Please do not include any information that will make you identifiable to your reviewers.

In your reflection, include responses to the following prompts:

1. What did you discover about how AI systems can amplify societal bias? On what do you base these discoveries?
2. Is there evidence that demonstrates student awareness of bias or fairness in AI outputs?
3. How will this experience influence your future use or discussions of AI tools in teaching and collaboration?
4. Describe how you shared your lesson plan and the results of the classroom activity with a colleague, PLC, or professional team. What feedback or insights did you receive? How will this influence your ongoing approach to teaching about AI bias?

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

Passing:
To receive a passing score, you must demonstrate that this project deepened your understanding of how AI systems can reflect and amplify human and societal bias and influence your professional practice.

A passing reflection:

- Includes specific examples of how students gained awareness of fairness, accuracy, or representation in AI output;
- Outlines clear, actionable ways this new understanding will inform future instruction, promote critical digital literacy, and foster equitable classroom discussions about AI; and
- Describes feedback and insights on shared lesson plan and classroom results.