ADL

LESSON PLAN

What is Bias in AI?

Compelling Question:How does bias show up in AI (artificial intelligence) and what can we do about it?

Grade Level		Time	Common Core Standards
K-1	2-5	**90 minutes (2 class periods)**	**Reading, Writing, Speaking and Listening, Language**
6-8	9-12		

LESSON OVERVIEW

In recent years, AI (Artificial Intelligence) has increasingly become part of our lives. AI is when a computer is able to perform tasks that normally a human would perform; the computer is basically teaching itself or programming itself to do tasks. Impacting many sectors and aspects of life, AI is used in medicine, healthcare, transportation, science, education/schools, the military, robotics, finance, agriculture, art, entertainment, social media, law enforcement, customer service, manufacturing, and more. With it, AI brings both risks (harm) and rewards (help). One of the risks and potential dangers of artificial intelligent technologies is the bias that is both embedded into and perpetuated by AI.

This two-part lesson provides an opportunity for students to understand what AI is and how we use AI in our lives, reflect on the possible risks and rewards of AI, consider how bias can take place in AI and explore possible guidelines and guardrails to maximize how AI can be used with minimal bias or harm.

LEARNING OBJECTIVES

Part 1:

- Students will understand what artificial intelligence is and how it currently operates.
- Students will reflect on their experiences with artificial intelligence.
- Students will identify the rewards and benefits of AI and will consider the potential harm/risk and bias of AI.

Part 2:

- Students will share and learn about examples where bias shows up in AI.
- Students will consider potential guidelines and guardrails to make AI less biased, less harmful and more responsible for its users and society

MATERIALS AND PREPARATION

- **AI Sample Definitions** (one copy for each student)
- Access to online and print dictionaries
- **Examples of AI worksheet** (one copy for each student)
- **What Exactly Are the Dangers Posed by A.I.?** (one copy for each student)
- **Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History** (one copy for each student)

Key Words

algorithm
censoring
complement
demoralizing
disinformation
distortions
erasures
existential

[Note: This is a two-part lesson; each part consists of 40-45 minutes. We trust educators to adapt and use as you see fit. However, if you only have one class period to discuss AI, you can adapt the lesson by doing the first two sections of Part I ("Information Sharing" and "Our Experiences with AI") and then proceeding to Part 2, which focuses on AI and bias.]

PART 1: WHAT IS AI?

💬 **Information Sharing: What is AI (Artificial Intelligence)?**

1. Begin the lesson by asking students: *What is AI or artificial intelligence?* Have some students share their thoughts aloud.

2. In order to define artificial intelligence, have students use online dictionaries and other resources (if you allow devices in your classroom) to come up with a definition of AI. Allow students to work in teams of two or three. You can share these AI Sample Definitions as beginning examples.

3. Have the pairs/triads talk in their groups for ten minutes to come up with a definition. Then invite them to share their definition with the class. Vote or try to come to consensus on one definition (either one of theirs, or one of the definitions above) that you will use throughout the lesson.

4. Make the definition available visually throughout the lesson.

🤖 **Our Experiences with AI**

1. Ask students: *What are some examples of AI that you use in your everyday life?* First, provide an example such as the following:

You might use Chatbots while shopping and deciding whether to buy something or not. A customer service chatbot is a bot that uses AI and machine learning to answer basic customer questions or provide support. These chatbots are increasingly popular AI solutions to help with web customer service, which is expensive to do with live staff members. The programmed algorithms enable machines to answer frequently asked questions, take and track orders, and direct calls. Chatbots are taught to impersonate the conversational styles of customer representatives through natural language processing (NLP). Advanced chatbots no longer require specific formats of inputs (e.g., yes/no questions). They can answer complex questions requiring detailed responses. In fact, if you give a bad rating for the response you get, the bot should identify the mistake it made and correct it for the next time, ensuring maximum customer satisfaction.

2. Explain to students that AI is used in medicine, healthcare, transportation, science, education, the military, robotics, finance, agriculture, art, entertainment, social media, law enforcement, customer service, manufacturing, and more. Ask students: *What are other examples of AI that you or someone you know uses right now?*

3. As students share examples, share their responses on the board/smart board. Responses might include: Siri/Alexa, Grammarly, ChatGPT, Duolingo, Instagram AI. You can also have students use the AI Examples handout to identify specific examples in each of the categories on the worksheet.

4. After brainstorming a list of AI uses, have students either turn and talk with someone sitting nearby or do a "Quick Write" (could also be done for homework) about their experiences with AI. This means they will have a short amount of time to write; in this case, give students 5-7 minutes to write as much as they can in response to one or more of these questions:

Key Words (cont.)

extrapolation
lackluster
mangle
moratorium
nuanced
pioneer
overblown
persuasive
profound
pundits
rote
semblance
speculation
stereotype
theoretical
toxic
unanticipated
unseen

- Share an experience you had using AI. How did it help you?
- As you used it, what did you notice? What did you wonder?
- Were you aware of any risk/danger/harm with this form of AI?

(Note: For students that haven't used AI, or as an alternative, have students respond to the question: *What would you want AI to help you do?*)

5. After writing, have students turn and talk with someone sitting next to them. Have each of them share their quick write with each other. If time permits, invite some students to share their quick writes with the class.

It's Your Turn! Engaging with AI

1. To gain experience in what AI is and how it can be used, engage students in a process where they use AI, either to create an image using StarryAI (for example, have them request an image based on a character with specific details to alter it) or ask a question or help find information using ChatGPT (for example, have them ask ChatGPT to come up with an interview question for their favorite celebrity, artist, politician or athlete).

(Note: If you don't allow laptops or other devices in your classroom, assign this for homework instead of doing it during class.)

2. Divide students into small groups of 4 or 5 students each. In the small groups, have students take turns sharing their experience with AI. They should share what they did and the results. Engage them in some reflection by inviting them to answer the questions: What do you notice? What do you wonder? What was positive and what was negative about the experience? Did anything happen that was concerning to you?

Risks and Rewards of AI

1. Share with students a few pieces of information from a 2022 Pew Research Center study on how people in the U.S. think about AI:

- Of the respondents, nearly half of U.S. adults (45%) say they are equally concerned and excited about AI; 37% are more concerned than excited and 18% say they are more excited than concerned. (After sharing this statistic, do a quick poll among the students, asking whether they are "more concerned than excited," "more excited than concerned" or "equally concerned and excited" about AI.)
- The survey provides examples of what the respondents are both concerned and excited about. They are concerned about loss of human jobs, surveillance and privacy, lack of human connection, AI will get too powerful and outsmart people and other reasons. They are excited about AI making life easier, saving time/efficiency, inevitable progress, handling mundane tedious tasks and other reasons.
- When asked about three specific examples of AI, 44% of respondents said that driverless passenger cars are a bad idea for society and 26% said they are a good idea for society. Thirty-one percent said computer programs by social media companies to find false information on their sites is a bad idea for society and 37% said it is a good idea. Twenty-seven percent of respondents said facial recognition by the police is a bad idea for society and 47% said it was a good idea.

Ask students: *Are you surprised by these results? If so, what are you surprised by and why? Do you think the results would be different if asked only to young people under 18 years old and if so, how? How do you think the data might change over the next five years?*

2. With students, brainstorm a list of rewards (positive outcomes) and risks (possible negative outcomes) of AI. Create a T-chart with students' responses to the rewards and risks question.

REWARDS (Ways AI can help)	RISKS (Ways AI can harm)
Combat misinformation by fact-checking	Perpetuate false information (misinformation)
Help with research	Lost jobs
Will help save time	Loss of privacy

3. Engage students in a brief discussion about the risks and rewards by asking: *What do you notice? What do you wonder? Are there examples of AI that could have both risks and rewards? If so, what is an example?*

4. Optional (to do in class or for homework): To learn more about the dangers of AI, distribute a copy of the article What Exactly Are the Dangers Posed by A.I.? to each student. Provide 10-15 minutes for students to read the article silently or read aloud together, with students taking turns reading. You can also have students read the article for homework the night before.

5. After reading, engage students in a discussion by asking some or all the following questions:

   - Why do you think so many technology leaders, researchers and other pundits working in AI signed an open letter warning that A.I. technologies present "profound risks to society and humanity?"
   - Why are these technology leaders and others concerned about AI?
   - Who is suggesting these dangers and possible pauses of AI and what do you think their motives might be in doing so? Who is being left out of these decisions?
   - What are their main concerns?
   - What is one example of a short-term, medium-term and long-term risk of AI explored in the article? What are your thoughts about those risks?
   - What is your biggest takeaway from this article?

PART 2: WHAT IS BIAS IN AI?

What is Bias?

1. Ask students: *What is bias?* Elicit and explain a definition as follows: "An inclination or preference, either for or against an individual or group, that interferes with impartial judgment."

2. Then ask: *Where do we see bias?* Elicit that we see bias in interpersonal relationships, in hate symbols, in the media, on social media, in institutions (schools, sports, criminal justice system), etc. If this doesn't come up, ask: *Do you think there could be bias in AI?*

3. Provide some background about bias in order to help students understand what bias is without delving in too deeply. You can elicit and share some or all of the following:

   - Bias is universal, which means that we all have biases.
   - Our biases are shaped by our experiences in the society around us, all of those that influence us in a variety of ways. These influencers include family, friends, acquaintances, the media (including social media), school, who we know and who we don't know, what we see in society, etc.
   - Systems of bias (like antisemitism, racism, sexism, heterosexism, anti-trans bias, ableism, classism, etc.) are when the bias is part of, and embedded into, all parts of society. Systems of bias are the combination of rules, laws, history, institutions and other factors that advantage some people/groups over others. For other people/groups, the system of bias causes widespread harm and disadvantages in access and opportunity.

(Explain that individual acts of prejudice, bias and discrimination are informed by and contribute to these systems of bias, but the systems exist whether or not there are individual/interpersonal acts of bias.)
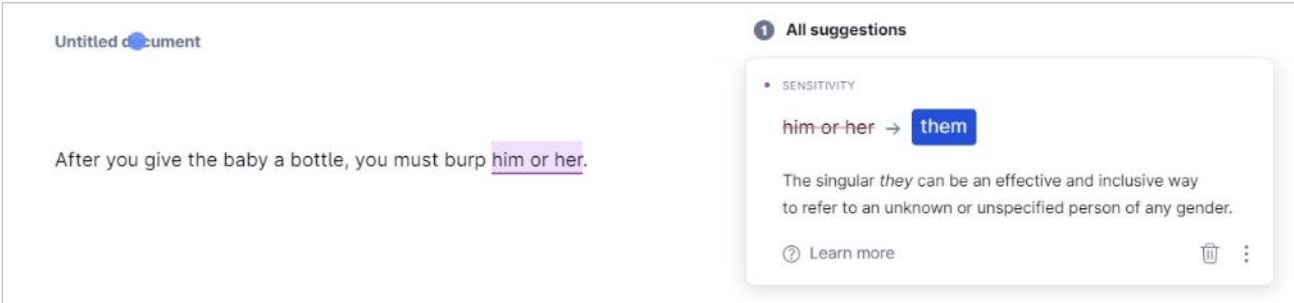
🔍 Where is the Bias in AI?

1. Ask students: *Have you ever seen or experienced AI situations in which there is bias?* Have them share examples. Explain that AI bias is a phenomenon that takes place when an algorithm produces results that are systemically prejudiced due to incorrect assumptions in the machine learning (ML) process.

2. Share a few additional examples:

   - Facial-recognition systems misidentified people of color more often than white people, a landmark federal study shows. Black and Asian American people were up to 100 times more likely to be misidentified than white men, depending on the particular algorithm and type of search. Native American people had the highest false-positive rate of all ethnicities, according to the study, which found that systems varied widely in their accuracy. Women were more likely to be falsely identified than men, and elderly people and children were more likely to be misidentified than those in other age groups.

   - A specific example of facial-recognition bias happened in 2023 in Michigan. A mother of two young children and pregnant was handcuffed in front of her children and taken into custody for eleven hours and questioned about a crime she knew nothing about. After being charged in court with robbery and carjacking, she was released that evening on a $100,000 personal bond. A month later, the prosecutor dismissed the case against her. The ordeal started with an automated facial recognition search, according to an investigator's report. This woman is the sixth person to report being falsely accused of a crime as a result of facial recognition technology used by police to match an unknown offender's face to a photo in a database. All six people have been Black. The woman in this case filed a lawsuit for wrongful arrest.

   - In the healthcare field, AI bias can impact diagnosis. For example, a Canadian company developed an algorithm for auditory (hearing) tests for neurological diseases. The test registers the way a person speaks and analyzes the data to determine early-stage Alzheimer's disease. The test had very high accuracy rates (90%) but the data consisted of samples of native English speakers only. Therefore, when non-English speakers took the test, it would identify pauses and mispronunciations as indicators of the disease. (https://www.wipro.com/holmes/the-pitfalls-of-ai-bias-in-healthcare/)

   - An example of gender bias in AI is the tendency to "feminize" AI tools can reinforce and perpetuate gender stereotypes. For example, home-based virtual assistants such as Amazon's Alexa, Microsoft's Cortana, and Apple's Siri were given default feminine voices and to have "submissive personalities" and stereotypically feminine attributes, such as being "helpful, intelligent, intuitive."

   - In 2016, Microsoft introduced a Twitter bot named Tay, promoting it by saying the more you chat with Tay, the smarter it gets, learning to engage people through "casual and playful conversation." Within 24 hours, the chatbot was sharing tweets that were racist, antisemitic and anti-trans. The bot learned discriminatory behavior from its interactions with users, many of whom were feeding it inflammatory messages. It was discontinued.

   - Other examples https://www.prolific.co/blog/shocking-ai-bias

3. Ask students: *How do you feel after hearing these examples?* Then ask:

   - In the situations described above, can you think of anything that could have been done to address or reduce the bias or prevent it from happening in the first place?

   - Given the bias and injustice in society, is it possible to fully eliminate bias and if not, what can we do instead?

◯◯ **Reading Activity**

1. To invite students to learn more about bias in AI, distribute a copy of the article Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History to each student. Provide 10-15 minutes for students to read the article silently or read aloud together, with students taking turns reading.

2. After reading, engage students in a discussion by asking some or all the following questions:

   • What did artist Stephanie Dinkins discover when she used the terms "Black woman" or "African American woman" to create art using AI? Why do you think there were machine distortions?

   • What does Dinkins mean when she says, "The biases are embedded deep in these systems, so it becomes ingrained and automatic?"

   • What are some of the dangers and risks explored in the article in how Black people are depicted in AI generated images?

   • In what ways does AI mirror the bias and discrimination we see in non-AI media images?

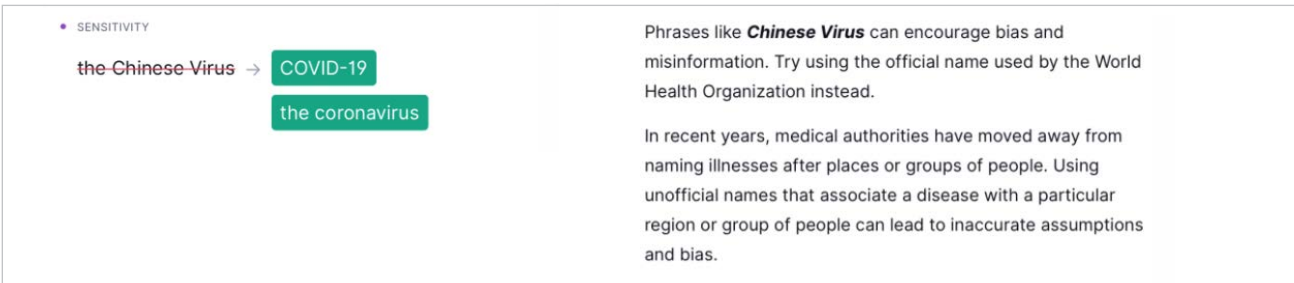   • What did you learn about AI and bias that you didn't know before?

⚙ **Can AI Tools Help to Address Bias?**

1. Ask students: *When can AI help with bias or when does AI specifically address bias or promote inclusivity?*

2. Share the screenshot below. Ask: *What do you notice? What do you wonder?* Explain that in this example using Grammarly, a review and editing assistant, if you type in "she or he" to your text instead of "they" when trying to use a generic pronoun, Grammarly comes back with a "Sensitivity" comment, explaining that using the singular "they" is an inclusive way or referring to someone whose gender is unknown.

You can share another example from Grammarly (below) where the writer used the term "the Chinese Virus" instead of "COVID-19" or "the coronavirus" and it was also highlighted for "Sensitivity." Explain/elicit that especially in the beginning of the COVID-19 pandemic in 2020, some referred to it as "Chinese Virus," which directly and indirectly promoted scapegoating and stereotyping of the Chinese and Chinese American community.

3. Ask students: *What other ideas do you have about how AI has helped or could help with bias?*

Guidelines and Guardrails

1. Ask students: *What are guidelines? What are guardrails?* Elicit/explain that guidelines are general rules, principles, or pieces of advice. Guardrails are guidelines that are put in place to protect against or reduce danger. Provide an example of each.

2. Have students turn and talk with someone sitting nearby, inviting the pairs to come up with one or two guidelines or guardrails that are important to have in place (in school, online, in our community) to prevent or reduce bias in AI.

(Note: If you think it would be helpful before having students get in pairs, read aloud the White House Blueprint for an AI Bill of Rights, specifically reading the section on "Algorithmic Discrimination Protections" in order to generate some thinking and ideas).

3. Go around the room, asking pairs to each share their ideas aloud and record them on the board/smart board, noting when guidelines or guardrails are repeated more than once.

4. Then, as a class (or assign for homework) combine similar guidelines and guardrails to come up with a refined list that incorporates the best ideas shared. Some ideas may include:

*   Educate the public and lawmakers about AI bias.
*   Make sure all AI is assessed for safety, bias and effectiveness before it goes live.
*   Put checks and balances in place; always keep humans in the loop.

(Note: As an extension or homework activity, have students choose one of the guidelines or guardrails on the list and create a PSA video or print PSA to get their important message out about AI bias.)

Closing

Engage students in a quick go-round by inviting them to share one new thing they learned today and or one question on AI and bias that remains for them.

Additional Reading and Resources

*   8 Helpful Everyday Examples of Artificial Intelligence (IoT For All, January 24, 2023)
*   Addressing Gender Bias to Achieve Ethical AI (Global Observatory, March 17, 2023)
*   AI and Human Enhancement: Americans' Openness Is Tempered by a Range of Concerns (Pew Research Center, March 17, 2022)
*   AI Is Helping Us With Our Instructional Practice. Here's How (Education Week, July 11, 2023)
*   A.I. Poses 'Risk of Extinction,' Industry Leaders Warn (New York Times, May 30, 2023)
*   Artificial Intelligence (Department of Education, Office of Educational Technology)
*   Bias in AI and Machine Learning: Sources and Solutions (Lexalyctics)
*   How Is AI Applied in Everyday Life? (CalTech Science Exchange)
*   Humans are Biased. Generative AI is Even Worse (Bloomberg)
*   Sampling bias: identifying and avoiding bias in data collection (Eval Academy)
*   Teaching Kids What AI Is (and Isn't) (ISTE, October 12, 2022)
*   What Do We Do About the Biases in AI? (Harvard Business Review, October 25, 2019)
*   What is artificial intelligence (AI)? (IBM)

Common Core Anchor Standards

CONTENT AREA/STANDARD
Reading
R1: Read closely to determine what the text says explicitly and to make logical inferences from it; cite specific textual evidence when writing or speaking to support conclusions drawn from the text.   R2: Determine central ideas or themes of a text and analyze their development; summarize the key supporting details and ideas.
Writing
W2: Write informative/explanatory texts to examine and convey complex ideas and information clearly and accurately through the effective selection, organization, and analysis of content.
Speaking and Listening
SL1: Prepare for and participate effectively in a range of conversations and collaborations with diverse partners, building on others' ideas and expressing their own clearly and persuasively.   SL3: Delineate a speaker's argument and specific claims, distinguishing claims that are supported by reasons and evidence from claims that are not.
Language
L4: Determine or clarify the meaning of unknown and multiple-meaning words and phrases by using context clues, analyzing meaningful word parts, and consulting general and specialized reference materials, as appropriate.   L5: Demonstrate understanding of figurative language, word relationships, and nuances in word meanings.   L6: Acquire and use accurately a range of general academic and domain specific words and phrases sufficient for reading, writing, speaking, and listening at the college and career readiness level; demonstrate independence in gathering vocabulary knowledge when encountering an unknown term important to comprehension or expression

CASEL's SEL Competencies

COMPETENCIES
Self-Awareness: The abilities to understand one's own emotions, thoughts, and values and how they influence behavior across contexts.
Social Awareness: The abilities to understand the perspectives of and empathize with others, including those from diverse backgrounds, cultures, & contexts.
Relationship Skills: The abilities to establish and maintain healthy and supportive relationships and to effectively navigate settings with diverse individuals and groups.
Responsible Decision-Making: The abilities to make caring and constructive choices about personal behavior and social interactions across diverse situations.

AI Sample Definitions

Britannica	The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.
IEEE (Institute of Electrical and Electronics Engineers)	The theory and development of computer systems that are able to perform tasks which normally require human intelligence such as, visual perception, speech recognition, learning, decision-making, and natural language processing.
ISTE (International Society for Technology in Education	When a computer is able to perform tasks that normally a human would perform; the computer is basically teaching itself or programming itself to do tasks.
Merriam-Webster Dictionary	A branch of computer science dealing with the simulation of intelligent behavior in computers; the capability of a machine to imitate intelligent human behavior.
U.S. Department of State	A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments.
Wikipedia	The intelligence of machines or software, as opposed to the intelligence of human beings or animals. AI applications include advanced web search engines (e.g., Google Search), recommendation systems (used by YouTube, Amazon, and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Waymo), generative or creative tools (ChatGPT and AI art), and competing at the highest level in strategic games.

Our Ideas:

Examples of AI Worksheet

Medicine/Healthcare	
Transportation	
Science	
Education/Schools	
Military	
Robotics	
Finance	
Agriculture/Farming	
Arts (visual and performing)	
Social media	
Law enforcement	
Customer service	
Other	