

The Challenges of AI

SUBJECT	TEACHER	GRADE	DATE

OVERVIEW

This lesson explores the challenges and risks associated with generative AI systems, focusing on issues like biases, hallucinations, and the accuracy and fairness of AI-generated content. Students will engage in critical discussions, hands-on activities, and debates to examine the ethical implications of AI technologies in society, particularly in education. Through these activities, they will learn how to responsibly interact with AI systems and understand the broader societal impact of these technologies.

OBJECTIVES

By the end of this lesson, students will be able to:

- Understand biases and hallucinations in AI: Explain how biases in training data affect AI outputs and explore the phenomenon of "hallucinations" in generative AI.
- Critically evaluate AI outputs: Analyze AI-generated content for accuracy, relevance, and fairness.
- Discuss the ethical implications of AI: Identify the social, educational, and economic impacts of generative AI technologies and their potential to reinforce or challenge inequalities.
- Engage in interactive debates and policy drafting: Participate in debates on AI's role in society and draft AI usage policies that reflect principles of fairness and responsibility.

MATERIALS

- Laptops/tablets for accessing AI tools (e.g., ChatGPT, Google Bard, or another AI platform)
- Projector/whiteboard

LESSON OUTLINE

1. Introduction to AI and its Challenges (15 minutes)

- **Activity:**
 - Begin by showing slide 2 (Brainstorming).
 - Ask students to list some ways in which they can use generative AI in educational contexts. For example, help with writing, understanding texts, assistance in homework, find additional information, exploring a concept, etc.
 - Ask students if they have ever used AI tools, how often do they use them, and what their impressions or concerns are about AI-generated content.

- **Key Concepts (Slide 3):**

- Introduce the concepts of "hallucination" and "bias" as general concepts. Ask students to try and define these concepts in the context of AI.

2. Understanding Biases and Hallucinations in AI (30 minutes)

- **Activity:** Show students Slide 4. Explain that they are going to see some examples of biased AI outputs and "hallucinations." Present a selection of AI-generated content (Slide 5) and have students assess the accuracy, fairness, and potential biases in each example.
- **Discussion:** Guide students to understand how biases or other errors in training data can lead to AI outputs that perpetuate stereotypes or misinformation. Elucidate how the accuracy of the answers that AI chatbots provide depends on their training data.
- **Hands-on Exercise:** Have students use an AI tool to generate responses to the same prompt, then identify any biases or inaccuracies.
- **Small Group Work:** In groups, students critically evaluate AI's responses, identifying any factual inaccuracies or biased language.
- **Whole-Class Discussion:** Share findings and discuss strategies to address these issues in AI-generated content (e.g., improving data diversity, implementing bias detection tools).
- **Reflection:** Ask students to reflect on how this could affect real-world decision-making (e.g., hiring processes, medical diagnoses, educational tools).

3. Ethical Implications of AI in Society (25 minutes)

- **Presentation:** Provide an overview of the ethical implications of AI in society. Focus on:
 - AI in education: How can AI impact teaching and learning? What are the risks of relying too heavily on AI tools for student assessments?
 - AI and employment: How might automation and AI affect job markets? Consider both positive and negative impacts.
 - AI in decision-making: What are the dangers of AI being used for important decisions (e.g., back credit, policing)?
- **Discussion:** Facilitate a class discussion on the need for transparency and accountability in AI development. Encourage students to think about real-world examples of AI ethical challenges (e.g., biased algorithms in hiring, AI in surveillance).

4. Interactive Debates on AI Usage (20 minutes)

- **Activity:** Organize a structured debate on the topic of AI in education. Divide the class into two groups: one arguing for the use of AI in educational settings, and the other against it.
- **Preparation:** Give students 5 minutes to prepare their arguments, using the ethical considerations discussed earlier.
- **Debate:** Hold the debate, ensuring each side presents its points, followed by a rebuttal and closing arguments.

5. Drafting AI Usage Policies (15 minutes)

- **Activity:** In groups, students will draft a policy on AI usage in a specific context (e.g., AI in the classroom, AI in hiring, or AI in content creation).

- **Discussion:** Ensure that policies reflect principles of fairness, inclusivity, and transparency.
- **Presentation:** Groups will present their policies to the class.

6. Drafting AI Usage Policies (15 minutes)

- **Reflection:** Engage students in a final reflective discussion on the evolving role of AI in modern society. How can they use AI responsibly? What are their thoughts on its potential impact on the future?
- **Wrap-Up:** Summarize key points from the lesson, emphasizing the importance of critical evaluation, ethical considerations, and responsible AI usage.

UNDERSTANDING

Students will gain an understanding of the ethical, social, and technical challenges posed by generative AI systems. They will be able to critically assess AI-generated content for fairness, accuracy, and biases. The class activities will encourage students to think deeply about the implications of AI technologies in various societal contexts and develop strategies to address these challenges, ensuring AI is used responsibly and ethically.

- A short quiz based on this lesson can be found [here](#).
- Topics for discussion may include:
 - Should AI tools be regulated in classrooms, and if so, how?
 - How can we address and eliminate biases in AI training data that reflect historical inequalities (e.g., race, gender, socioeconomic status)?
 - Is it possible for AI to be completely neutral, or will bias always be an inherent part of AI systems?
 - How can AI developers ensure diversity in their data sources to reduce bias?