



Deepfakes, Disinformation, and Digital Trust

Grades 9–12 Media Literacy / Social Studies | 40-Minute Lesson Plan

1. OVERVIEW

Deepfakes have moved from novelty to a tool for disinformation and fraud, appearing in politics, finance, and everyday life. Students need both technical verification skills to investigate suspicious media and frameworks for reasoning about the societal impact of manipulated content.

Important distinction: some fake videos are fully AI-generated, while others use real footage with a real person's face or voice swapped in (deepfakes in the strict sense). Detection skills overlap across these types.

This lesson's cases cover both, so students can practice evaluating and responding to each.

2. LEARNING OBJECTIVES

- Evaluate real-world deepfake incidents and their consequences.
- Use technical verification tools on suspicious media.
- Construct evidence-based arguments about how society should respond.

3. MATERIALS

- Projector
- Student devices
- Case cards (included below)
- CER graphic organizer (Claim-Evidence-Reasoning)



Teacher Prep

- Preview all materials and cases to become familiar with content and questions.
- Prepare student grouping (pairs or small groups).
- Have a sample suspicious image or key frame ready for the verification demo.



Deepfakes, Disinformation, and Digital Trust

Grades 9–12 Media Literacy / Social Studies | 40-Minute Lesson Plan

4.



LESSON FLOW

PART	FOCUS	TIME
A.	Hook	5 min
B.	Case Study Jigsaw	15 min
C.	Verification Workshop	10 min
D.	Structured Discussion	7 min
E.	Exit Ticket	3 min
		TOTAL: 40 MINUTES

5 min



A. Hook

- Quick poll: “How confident are you that you could spot a deepfake?”
- Reveal: A preregistered experiment found people cannot reliably detect deepfakes, overestimate their own ability, and neither warnings nor financial incentives improve accuracy (*Köbis et al., 2021, iScience*).

15 min



B. Case Study Jigsaw

- Groups each analyze one case card.
- Students identify: who was harmed, what made it believable, and what could have exposed it.
- Groups prepare a brief share-out (1–2 key points).

10 min



C. Verification Workshop

- Demonstrate reverse image search on video keyframes with Google Lens or TinEye and lateral reading.
- Note: Reverse search is especially effective against face-swap fakes because it often surfaces the original footage.
- Introduce provenance standards such as C2PA content credentials.
- Emphasize honestly: Detection tools are probability-based and never 100% reliable.

7 min



D. Structured Discussion

- Using a CER organizer, students respond to:
“Should platforms be legally required to label AI-generated media?”
- Provide a claim, evidence from a case, and reasoning.

3 min



E. Exit Ticket

- Students write:
 - One verification habit they will use.
 - One lingering question they still have.



Deepfakes, Disinformation, and Digital Trust

Grades 9–12 Media Literacy / Social Studies | 40-Minute Lesson Plan

5.



CASE CARDS

Analyze each case using the questions provided.



CARD 1

New Hampshire fake Biden robocall (Jan 2024)

political consultant Steve Kramer commissioned an AI-cloned voice of President Biden urging Democrats not to vote in the primary; thousands of voters received it.

Outcome: the FCC fined Kramer \$6 million (caller-ID spoofing violations) and the transmitting carrier Lingo Telecom paid a \$1 million settlement; in June 2025 a New Hampshire jury found Kramer NOT GUILTY on all criminal charges (13 felony voter-suppression counts, 13 misdemeanor impersonation counts).

As of late 2025 the FCC fine remains unpaid.

- 1 Who was harmed?
- 2 What made it believable?
- 3 What could have exposed it?



Facilitator Note

Election integrity + “technology is new, laws are old” — the acquittal shows how existing law struggles with AI cases.



CARD 2

Hong Kong Arup deepfake video call (Jan 2024)

a finance employee at engineering firm Arup received a suspicious email about a confidential transaction, but set doubts aside after a video conference where every other participant — including the “CFO” — was a deepfake.

The employee transferred HK\$200 million (about US\$25.6 million) in 15 transfers to 5 accounts.

No arrests; funds unrecovered.

- 1 Who was harmed?
- 2 What made it believable?
- 3 What could have exposed it?



Facilitator Note

The failure of “seeing is believing”; the initial suspicion was correct and the video call defeated it.



Deepfakes, Disinformation, and Digital Trust

Grades 9–12 Media Literacy / Social Studies | 40-Minute Lesson Plan

5.



CASE CARDS

CARD 3

Maryland Pikesville High School audio (Jan 2024)

Athletics director Dazhon Darien, facing contract non-renewal, used AI to clone principal Eric Eiswert's voice in a fake racist and antisemitic recording, which spread widely.

The principal was suspended and received threats before investigators traced the fake.

In April 2025 Darien entered an Alford plea to disrupting school operations and was sentenced to 4 months; the principal later settled a lawsuit with the school system.



ANALYSIS QUESTIONS

- 1 Who was harmed?
- 2 What made it believable?
- 3 What could have exposed it?



FACILITATOR NOTE

Audio deepfakes harm ordinary people, not just celebrities; the victim's perspective.

CARD 4

Fake celebrity endorsement scams

AI-generated videos of celebrities and doctors promoting fake investments, miracle cures, and giveaways circulate widely on social platforms; the FBI's IC3 received 22,364 complaints citing AI in fraud in 2025, with \$893 million in reported losses.



ANALYSIS QUESTIONS

- 1 Who was harmed?
- 2 What made it believable?
- 3 What could have exposed it?



FACILITATOR NOTE

Financial self-defense; why platforms struggle to keep up.



Deepfakes, Disinformation, and Digital Trust

Grades 9–12 Media Literacy / Social Studies | 40-Minute Lesson Plan

6.



DISCUSSION RUBRIC

Discussion Prompt: Should platforms be legally required to label AI-generated media?

Criteria	3 — Strong	2 — Developing	1 — Beginning
 Claim	Presents a clear, focused, and defensible answer that directly addresses the prompt.	Presents a general or partially focused answer that addresses the prompt but lacks specificity.	Presents an unclear, vague, or off-topic answer that does not address the prompt.
 Evidence Use	Uses relevant, accurate, and specific evidence from the case study to support the claim.	Uses some relevant evidence from the case study, but it may be limited or only partially connected to the claim.	Uses little, irrelevant, or inaccurate evidence from the case study, or no evidence.
 Reasoning Quality	Clearly explains how the evidence supports the claim with logical, coherent reasoning.	Attempts to explain how evidence supports the claim, but reasoning may be incomplete or uneven.	Provides weak, unclear, or missing reasoning that does not connect evidence to the claim.



CER Reminder

Claim: Your answer or position to the question.

Evidence: Facts, data, or examples from the case study that support your claim.

Reasoning: The explanation of how your evidence connects to and supports your claim.

Sentence Starters:

- My claim is that _____.
- This is because _____.



Deepfakes, Disinformation, and Digital Trust

Grades 9–12 Media Literacy / Social Studies | 40-Minute Lesson Plan

7.



TEACHER NOTES



SENSITIVE-CONTENT ADVISORY

This lesson may touch on sensitive topics. Teachers should preview all materials and consider student needs before instruction.

- If students raise non-consensual intimate imagery, state accurately:
The federal TAKE IT DOWN Act was signed in May 2025 — knowingly publishing such images (including AI “digital forgeries”) is now a federal crime. Since May 2026, platforms must remove reported images within 48 hours.
- Direct affected students to trusted adults and reporting channels rather than discussing case details in class.
- Prioritize a safe, supportive environment. Pause or adjust activities if students show distress.

8.



STANDARDS ALIGNMENT

Standard	How This Lesson Aligns
CCSS.ELA-Literacy.CCRA.R.1	Students evaluate claims and evidence in media, identifying credible sources and reasoning.
CCSS.ELA-Literacy.CCRA.R.7	Students integrate information from diverse texts and media to analyze topics and issues.
CCSS.ELA-Literacy.W.1	Students write arguments that introduce a claim, provide reasons, and use evidence to support conclusions.
NGSS.SEP.7 (engaging in argument from evidence)	Students construct and critique evidence-based arguments about the societal impact of manipulated content.
ISTE Digital Citizen	Students practice responsible, ethical, and safe use of technology while evaluating digital content and protecting privacy.