

Case 20: Can the crew post this image as true?

A realistic image says the campus courtyard flooded today, and Jordy must decide what the six o'clock recap can say about it.

On your own Read What happened and underline the caption's claim. About 4 minutes.



Lupa holds a cropped photo of a courtyard next to the full photo pinned on a board. The full photo also shows a blue doorway.

Your role: You're the one who writes the verification note Jordy posts at six. **New skill:** Confirm an image with independent sources.

What happened

It is Wednesday, 3:10 p.m. For twenty minutes, an image has been going around the group chat. It looks like a photo of the campus courtyard, with water over the paving and benches. The caption says, "Courtyard flooded today, stay away." It showed up at 2:48 p.m. with no link, no date and no name.

Jordy is putting together the crew's recap, which goes out at six. Sami has zoomed in on a hand in the picture, and the fingers seem to merge at the knuckles. Someone ran the image through a free online detector. It reported "12% likely AI-generated."

A reply in the chat says the writer's roommate's friend saw water near the library. Jordy checks the campus facilities alerts page. It shows no active alerts, but its last update was Monday morning.

Your question: Did the courtyard flood today, and what should the recap say about the image?

Key terms

On your own Read each key term and its example, then write your first guess in the box. About 4 minutes.

Provenance Provenance is information about where an image or a document came from and how it has been handled.

En español: procedencia

In this case: The courtyard image arrived at 2:48 p.m. with no link, no date and no named poster. Its provenance is unknown.

Corroboration Corroboration is support for a claim from an independent source.

En español: corroboración

In this case: A campus alert or someone's dated photo of the courtyard could corroborate the flood. A roommate's friend's report of water near the library cannot.

Verification note A verification note is a short record of what you checked, what you found and what is still unknown.

En español: nota de verificación

In this case: Instead of posting the image, the recap says what the crew checked, what it found and what is still unknown.

Who is involved



Sami

Checking where the image came from

"Everyone's zooming in on the hand. I did too, and it's weird. But I still don't know who posted this."



Jordy

Putting the six o'clock recap together

"The recap goes out at six. Right now I have an image with no source, a strange hand and an alerts page that hasn't updated since Monday."

Before you look at the evidence, do you think the courtyard flooded today?

This image shows a single sheet of white paper with horizontal blue lines, resembling notebook paper. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Reading passage

When the honest answer is unverified

On your own Read the passage. At each Stop and think box, pause and answer the question in your head or with a partner. About 5 minutes.

Words to know

- **Provenance:** Provenance is information about where an image or a document came from and how it has been handled.
- **Corroboration:** Corroboration is support for a claim from an independent source.
- **Verification note:** A verification note is a short record of what you checked, what you found and what is still unknown.
- **Claim:** A claim is a statement that something is true. You can check a claim with evidence.
- **Holding statement:** A holding statement is a short update that says what is known, what is still being checked and when people will hear more.
- **Correction:** A correction is a public statement that replaces wrong information with accurate information.

The story: A deadline and a picture

The crew's casebook began with a seven-second clip in a group chat. It ends with a single image. At 2:48 p.m. on a Wednesday, a photo-like picture of the campus courtyard appears in the chat, showing water over the paving and benches. Its caption reads, "Courtyard flooded today, stay away." It arrives with no link, no date on the image and no original poster named. A person in the foreground holds up a phone.

Jordy, who edits the crew's films and checks every cut before calling it finished, is assembling the six o'clock recap. Within twenty minutes, four pieces of evidence have arrived, and each points somewhere different. Sami posts a zoomed screenshot at 3:02 showing a foreground hand whose fingers seem to merge at the knuckles. A free online detector reports "12% likely AI-generated." A reply relays that the writer's roommate's friend saw water near the library. And at 3:10, Jordy checks the campus facilities alerts page: "No active alerts. Last update: Monday, 9:15 a.m."

Sami, a design student who notices details for a living, is the first to step back from them. The hand is weird, she agrees, but the more important question is who posted the image first, and when. Jordy states the dilemma plainly. They wanted a clean yes or no, and they do not have one.

Your role is to write the verification note Jordy posts at six.

❓ **Stop and think.** Why can a detector score answer the wrong question even if it is accurate?

Weighing evidence without overclaiming

Provenance is information about where an image or a document came from and how it has been handled. Checking provenance first is the Case 1 skill, finding the original source, applied to images. This image has none, so every other clue has to be weighed without an anchor.

Each clue has a specific weakness. The caption is a **claim**, a statement that something is true that can be checked with evidence. The hand is an anomaly, which justifies closer scrutiny and proves nothing, since generated images can look natural and real photos can look strange. The detector score is uninterpretable without knowing the tool's error rates. Sami's question, twelve percent of what, gets at that. It also answers a different question from the one that matters. Even a perfect detector could not say whether the courtyard flooded today. The secondhand report fails on two counts: it is not firsthand, and it concerns the library, not the courtyard.

Corroboration is support for a claim from an independent source. Case 9's skill, checking the official source for your area, points to the alerts page, but its last update on Monday limits what its silence can mean. A dated photo from someone at the courtyard would be stronger.

The new skill is confirming an image with independent sources, and publishing only what was checked. Jordy rejects the middle path of posting the image with a question mark, because the image travels regardless of the punctuation. The recap instead carries a **verification note**, a short record of what you checked, what you found and what is still unknown. It is a descendant of Case 1's **holding statement**, which says what is known, what is still being checked and when people will hear more. It also sets up Case 3's obligation. If new evidence arrives, a **correction** or update appears where the note did. "Unverified" is sometimes the most accurate conclusion available.

❓ **Stop and think.** Why does a question mark fail to stop an image from misleading people?

Detection, provenance and professional standards

NIST's overview of synthetic content describes provenance data, labeling, watermarking and detection, and it explains the technical limits of each (NIST, November 20, 2024). Real incidents show why no single signal suffices. The Hurricane Helene image of a crying girl with a puppy had no traceable photographer or outlet, and forensics researcher Hany Farid identified internal inconsistencies (UC Berkeley School of Information, October 4, 2024). The News Literacy Project's RumorGuard check on that image models a verification note, setting out what is known, what is not and why (RumorGuard, no date listed). In 2023, fake images of a Pentagon explosion were shared by verified accounts, showing that a trusted-looking sharer is not provenance (CNN, May 22, 2023).

Detection, provenance and professional standards, continued

Human perception is not a reliable safeguard. In a PNAS study, people could not reliably distinguish AI-synthesized faces from real ones (Nightingale and Farid, February 14, 2022).

Journalism codes frame Jordy’s choice as a duty. The RTDNA code says journalism “verifies, provides relevant context, tells the rest of the story and acknowledges the absence of important additional information” (RTDNA Code of Ethics, June 11, 2015). The SPJ code asks journalists to “identify sources clearly” (SPJ Code of Ethics, September 6, 2014).

 **Stop and think.** Who could be harmed if the crew called a real photo fake?

Sources

Real source | National Institute of Standards and Technology (NIST) | November 20, 2024

Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency

Real source | UC Berkeley School of Information | October 4, 2024

Hany Farid breaks down the viral Helene image

Real source | RumorGuard (News Literacy Project) | No date listed

Widely shared image of crying girl and puppy after Helene is AI generated

Real source | CNN | May 22, 2023

‘Verified’ Twitter accounts share fake image of ‘explosion’ near Pentagon, causing confusion

Real source | PNAS (Nightingale & Farid) | February 14, 2022

AI-synthesized faces are indistinguishable from real faces and more trustworthy

Real source | Radio Television Digital News Association (RTDNA) | 2015

RTDNA Code of Ethics

Real source | Society of Professional Journalists (SPJ) | 2014

SPJ Code of Ethics

Look at the evidence

On your own Read each document and write what it shows and what it leaves out. About 6 minutes.

Evidence A | Group-chat post

The courtyard image and its caption

Image, described: a photo-like view of the campus courtyard with water over the paving and benches. A person in the foreground holds up a phone. Caption: “Courtyard flooded today, stay away.” Shared Wednesday at 2:48 p.m. with no link, no date on the image and no original poster named.

Look closely: Underline what the caption claims. Circle what the post is missing.

What it shows

What it leaves out

Evidence B | Screenshot

Sami’s zoom-in on the hand

Zoomed screenshot of the foreground hand: the fingers seem to merge at the knuckles. Sami posted it in the crew chat on Wednesday at 3:02 p.m.

Look closely: What does the odd hand tell you? What can it not tell you?

What it shows

What it leaves out

Look at the evidence, continued

Evidence C | Detector result

A score from a free online detector

“12% likely AI-generated.”

Look closely: What did the tool measure? Does it say anything about the flood?

What it shows

What it leaves out

Evidence D | Web page

The campus facilities alerts page, checked at 3:10 p.m.

“No active alerts. Last update: Monday, 9:15 a.m. Scheduled elevator maintenance in Science Hall, completed.”

Look closely: When was this page last updated? Compare that with the caption’s claim.

What it shows

What it leaves out

What is still missing from the evidence?

Key idea: provenance

On your own Read the key idea and the code of ethics, then answer the question in the box. About 8 minutes.

Provenance is information about where an image or a document came from and how it has been handled. Check it first when an image claims to show a news event. An image can look polished and still be generated, edited, old or shared with the wrong caption. The courtyard image arrived with no link, no date and no named poster, so its provenance is unknown.

A detail or a tool can give a clue, but it cannot settle the question. An odd-looking hand is a reason to look closer. A natural-looking hand does not prove that an image is real, either. The National Institute of Standards and Technology (NIST) describes several ways to track, label, watermark and detect synthetic content, such as AI-generated images. NIST explains that each approach has technical limits.

For a claim about a real event, look for the earliest post, its date and place, and corroboration.

Corroboration is support for a claim from an independent source. Here, that could be a campus alert or someone's dated photo of the courtyard. If you cannot confirm the claim, write a verification note. A **verification note** is a short record of what you checked, what you found and what is still unknown.

Professional standards

Real source | Radio Television Digital News Association (RTDNA)

RTDNA Code of Ethics: Truth and accuracy above all

“Journalism verifies, provides relevant context, tells the rest of the story and acknowledges the absence of important additional information.”

In plain words: The RTDNA code asks news staff to say what they checked and to admit what they still do not know.

Question: The facilities page shows no active alerts but was last updated Monday. What should Jordy's note say is still unknown?

Your answer

Still unknown: ____, because ____.

 Real source | NPR | May 22, 2023
Fake viral images of an explosion at the Pentagon were probably created by AI
A realistic fake image spread fast and briefly moved markets before officials confirmed nothing happened.



A realistic fake image spread fast and briefly moved markets before officials confirmed nothing happened.



A blue check was not proof: accounts that looked trustworthy still shared an image with no real source.

This image shows a single sheet of white paper with horizontal blue ruling lines. The lines are evenly spaced and run across the width of the page. There is no handwriting or other markings on the paper.

Sort the evidence

On your own Sort each piece of evidence under the claim it bears on and fill in each box. About 6 minutes.

Two claims are in play: the flood and the image. Sort each piece of evidence under the claim it actually bears on. What would settle either one?

Evidence	Claim 1 The courtyard flooded today.	Claim 2 The image was generated by AI.
1 The image shows water in a courtyard, with no date or source		
2 Facilities alerts page at 3:10 p.m.: no active alerts		
3 Secondhand report of water near the library		
4 An odd hand and a 12% detector score		

Still open

- The image’s earliest source and date

Whose perspective is missing?

Whose view is missing from these documents? What would you ask them?

The view that is missing is _____. I would ask _____.

Practice: choose a response

On your own Choose what Jordy should do, explain why and fix the two marked phrases. About 5 minutes.

The recap goes out at six. What should Jordy do about the courtyard image?

- ☐ **A** Post the image with a question: “Is the courtyard flooded?”
- ☐ **B** Say the image is fake, because the hand looks wrong.
- ☐ **C** Leave the image out, report what the crew checked and say the flood is not confirmed.

Why I chose it

Improve a draft

A student's first draft

Heads up everyone!! **1** The courtyard flooded today so don't go over there. Some people are saying the picture is AI because the hand looks weird but I ran it through a detector and it said 12% AI so it's basically real. My roommate also said her friend saw water near the library earlier, so that pretty much confirms it. Stay safe and **2** tell your friends, we don't want anyone getting hurt or ruining their shoes lol. We'll post more if we hear anything else about it.

Rewrite each marked phrase so it says only what the evidence supports.

Rewrite phrase 1

Rewrite phrase 2

Grades 11—12

Your task

On your own Read the steps and fill in the planner boxes before you start your draft. About 12 minutes.

Build a verification and risk plan for the crew’s accounts: steps before posting any image, and a correction plan if one proves false.

Use: NIST synthetic-content overview, Adobe’s labeled Firefly examples and case materials.

You’ll make: Plan, 400—500 words

Time: 60 minutes

- 1 Reread Evidence A to D. Sort each check by the question it answers: how made, or what happened.
- 2 Read the NIST summary in the key idea. Note one limit for each approach.
- 3 Fill in the planner, including a risk statement.
- 4 Write the Check and decide section and the Post and correct section on the Your draft page.
- 5 Add a correction plan for an image that proves false.
- 6 Finish the plan, 400 to 500 words, on lined paper or in the online workspace.



Prefer typing? Scan to open the online workspace.

Which checks answer which question

The limits NIST describes

If an image we share proves false, we will ____.

This image shows a single sheet of white paper with horizontal blue ruling lines. The lines are evenly spaced and run across the width of the page. There are approximately 20 lines visible. The paper has rounded corners on the left side and a dark border around the edges.

Check your work

On your own Answer the Check your work questions, mark your level in each rubric row and write one change you made. About 6 minutes.

Does your plan separate how an image was made from whether the event happened?	☐ Yes	☐ No
Does it describe NIST’s limits accurately?	☐ Yes	☐ No
Does it include a correction plan and a risk statement?	☐ Yes	☐ No
Could a small team follow it?	☐ Yes	☐ No

How your work is scored

Mark one box in each row.

Skill	Beginning	Developing	Proficient	Advanced
Evidence	☐ Treats the image as proof.	☐ Cites a detector or a detail as a verdict.	☐ Checks the earliest source and an independent record.	☐ Checks several sources and names what each can’t show.
Reasoning	☐ Declares real or fake with no basis.	☐ Hedges without explaining why.	☐ Matches confidence to evidence.	☐ Separates authenticity of the file from truth of the event.
Audience & purpose	☐ Urges spreading.	☐ Warns without guidance.	☐ Gives followers a safe next step.	☐ Gives a next step, a follow-up time and a way to contribute evidence.

Skill	Beginning	Developing	Proficient	Advanced
Revision	☐ No change.	☐ Softens words only.	☐ Replaces the verdict with a verification note.	☐ Explains how each change protects followers and credibility.

I changed ___ because ____.

This image shows a single sheet of white paper with horizontal blue ruling lines. The lines are evenly spaced and run across the width of the page. There is no handwriting or other markings on the paper.

Questions to discuss

On your own Write a short answer to each question, then complete the sentence frames. About 8 minutes.

1 What would you need beyond the image to support the caption's claim?

2 Do you agree with Jordy's choice to leave the image out of the recap, and why?

3 Who could be harmed if the crew called a real photo fake?

4 The alerts page shows nothing new since Monday. Does that prove the courtyard did not flood?

My take

My claim: ____.

My take, continued

The best evidence is ____, because ____.

Someone who disagrees might say ____.

The limit of the evidence is ____.

I would change my mind if ____.

Notes

Talk about this case at home

Case 20: Can the crew post this image as true?

In this case, a realistic image in a group chat claims that a campus courtyard flooded today. Your student checked where the image came from and wrote a short note that says what is known and what is not.

Questions to talk about

- 1 What would you check before you share a picture of breaking news?
- 2 Where does your town or school post official alerts?
- 3 Why can saying you are not sure yet be better than guessing?

Nobody needs to share personal stories or private information to talk about this case.



See the case online.