

SPOT THE FAKES

PROBLEM CONTEXT

# Deepfakes Can Amplify Scams

Cloned voices and fake media can make an urgent scam feel familiar.

**\$2.1B**

**reported lost to scams  
that started on social  
media in 2025**

FTC data show how quickly social-media scams spread. Deepfakes can make those scams feel more believable and more urgent. **28% of people reporting a scam loss said it began on social media.**

Source: FTC Data Spotlight, 2026. Broader scam total; not all involved deepfakes.

**Learn the full workflow at** [spotthefakes.org](https://spotthefakes.org). Use PAUSE before you share, pay, click, or reply.



# Use PAUSE Before You Act

Five moves for urgent media.

## Pause

Stop before you share, pay, click, or reply.

## Ask

What claim or request am I being pushed to accept?

## Use the source

Open the official site, saved contact, or original account.

## Search elsewhere

Find independent confirmation, not another repost.

## Exit pressure

Urgency, secrecy, money, or login-code demands mean verify another way. If a voice note asks for money, call the saved number.

# AI Media Can Hide Its Origin

Some major AI generators leave provider-specific signals.

22,364

AI-related complaints  
were reported to the  
FBI in 2025

When an image may come from a major provider, the provider's own checker can sometimes reveal origin evidence. That does not prove the story is true, but it can help answer **where the file came from**.

Source: FBI, 2025. AI-related complaints include more than deepfakes.

Open current checkers at [spotthefakes.org](https://spotthefakes.org). Start with the provider's own checker when you suspect DALL·E or Nano Banana.



# Check DALL·E and Nano Banana

Use provider checkers first when the image may come from OpenAI or Google AI.

## OPENAI / DALL·E

### OpenAI Verify

[openai.com/verify](https://openai.com/verify)

1. Keep the original PNG, JPG, or WEBP.
2. Upload it at OpenAI Verify.
3. Save the exact result wording.

## GOOGLE / NANO BANANA

### Gemini

[gemini.google.com](https://gemini.google.com)

1. Upload the original image in Gemini.
2. Ask: "Was this created or altered by Google AI?"
3. Save the response.

Signal found = useful origin evidence

No signal = inconclusive

Still check the caption, source, date, and claim. Origin evidence is not proof that the story is true.

# Looking Real Is Not Proof

Deepfakes can imitate faces, voices, events, and official messages.

**\$893M**

**in reported losses tied  
to AI-related  
complaints in 2025**

Deepfakes can be one part of those schemes. A detector can add evidence, but **one score cannot settle whether a file or story is authentic.**

Source: FBI 2025 IC3. AI-related complaints include more than deepfakes.

See the full guide at [spotthefakes.org](https://spotthefakes.org). Detector evidence is one clue, not a verdict.



# Use a Detector Correctly

A clean workflow prevents overconfidence.

1

## Start with the original file.

Screenshots, clips, and compression can weaken useful signals.

2

## Choose the right media type.

Image, audio, and video detectors are not interchangeable.

3

## Record the exact output.

Save the tool name, model, wording, score, or error message.

4

## Compare and continue.

If tools disagree, treat the case as unresolved. "No AI detected" is not proof of authenticity.