Deepfake Detection Cheat Sheet

A Practical Guide to Spotting Deepfake Red Flags, Verifying Suspicious Requests, and Preventing Fraud

1. What Is a Deepfake?

A deepfake is synthetic or manipulated media created with artificial intelligence to make it appear that a person said, did, or appeared in something that did not actually happen. Deepfakes can be highly realistic because modern AI systems can learn patterns in faces, voices, expressions, lighting, and movement, then reproduce those patterns in new media.

1.1 Deepfake Voice, Video, and Image

Deepfakes are not limited to videos. They can appear in several formats, and each type creates different detection challenges.

- **Voice deepfakes:** AI-generated or cloned speech that imitates a real person's voice, tone, accent, and cadence. Example: a fake voice call that sounds like a senior executive asking an employee to approve an urgent payment.
- **Video deepfakes:** AI-manipulated footage where a person's face, mouth movement, body motion, or entire presence is altered. Example: a fake video of a public figure appearing to make a statement they never made.
- **Image deepfakes:** AI-generated or edited still images that show people, documents, events, or scenes that are fabricated or altered. Example: a fake profile photo, synthetic ID image, or edited screenshot used in a scam.

1.2 Why Traditional "Spot the Fake" Methods Are Becoming Less Reliable

Older advice often focused on obvious errors such as frozen faces, missing blinks, strange distortions, or robotic voices. Those signs can still help, but they are no longer enough. Current AI-generated media can preserve realistic skin texture, natural-looking eye movement, smooth lip motion, and convincing voice quality.

- **Better generation tools:** Modern models can create smoother motion, more natural facial expressions, and more realistic audio.
- **Compression hides clues:** Social media platforms often compress videos and images, which can blur artifacts that might otherwise be visible.
- **Real-time deepfakes:** Some tools can alter faces or voices during live calls, reducing the time available for careful inspection.
- **Context can deceive:** A believable message from a familiar account may feel trustworthy even when the media itself is manipulated.

Practical rule: Treat visual inspection as only one layer of verification. If the content asks for money, credentials, confidential data, reputational action, or urgent decisions, confirm it through a separate trusted channel.

2. Visual Warning Signs

Visual warning signs are clues that something in a video or image does not behave like naturally captured footage. None of these signs proves a file is fake on its own, but several signs appearing together should increase suspicion.

2.1 Unnatural Blinking or Facial Movement

Natural human faces move in small, uneven, and coordinated ways. Deepfakes may struggle to reproduce subtle facial timing, especially around the eyes, cheeks, mouth corners, and forehead.

- Watch for blinking that is too rare, too frequent, too synchronized, or oddly timed.
- Look for a fixed stare, glassy eyes, or gaze direction that does not match the conversation.
- Check whether smiles, frowns, eyebrow raises, and cheek movement appear delayed or overly smooth.
- Notice whether the head turns naturally with the neck and shoulders, or whether the face seems to float separately.

Example: In a video call, a person appears to speak naturally, but their eyes barely blink for a long period and their smile does not create normal cheek movement. This may be a signal to pause and verify the call through another channel.

2.2 Lighting and Shadows That Don't Match

Deepfake systems often synthesize or swap a face separately from the rest of the scene. This can create lighting inconsistencies between the face, neck, body, background, and nearby objects.

- Compare the brightness of the face with the background and clothing.
- Check whether shadows fall in the same direction across the face, neck, and room.
- Look for mismatched skin tone between the face and the hands or neck.
- Inspect reflections in glasses, eyes, jewelry, or shiny surfaces for inconsistencies.

Example: A person's face appears brightly lit from the front, but the room and body are lit from the side. If the shadows under the chin and on the wall point in different directions, the face may have been generated or inserted.

2.3 Blurring Around Hair, Glasses, or Facial Edges

Edges are among the hardest details for AI systems to blend perfectly. Hair strands, ears, glasses, jawlines, and the boundary between the face and neck often reveal manipulation, especially when the person moves.

- Look for flickering or shimmering around the hairline, ears, jaw, chin, or neck.
- Check whether glasses frames bend, blur, disappear, or appear uneven during movement.

- Watch for hands, microphones, or objects passing in front of the face; deepfakes may blur or distort at those moments.
- Inspect whether facial edges remain sharp and consistent when the person turns their head.

Example: In a short video, the speaker turns to the side and the outline near the glasses briefly melts into the cheek. This kind of edge instability is a common warning sign in face-swap or face-reenactment media.

2.4 Audio and Lip-Sync Inconsistencies

Many deepfakes combine generated speech with generated or manipulated mouth movement. Even when both are realistic, small timing errors can appear between what is heard and how the mouth moves.

- Focus on words with hard lip closures such as “p,” “b,” and “m.” The lips should close naturally at the right moment.
- Listen for speech that sounds too smooth, too clean, monotone, metallic, or missing normal breathing sounds.
- Watch whether the mouth forms the correct shapes for vowel sounds and pauses.
- Check if background noise, room echo, or audio quality changes suddenly while the face remains visually consistent.

Example: A video appears to show a manager saying “make the payment by Monday,” but the lips do not fully close on “make” and “payment.” If the request is urgent or financially sensitive, stop and confirm through a known phone number or approved communication channel.

3. Behavioral Red Flags

Behavioral red flags are often more important than technical clues. Even if a video or voice sounds convincing, the request itself may reveal the fraud. Deepfake-enabled scams usually rely on social engineering: the attacker tries to make you act quickly, avoid normal controls, and trust the identity being presented.

3.1 Unusual Urgency

Urgency is one of the most common pressure tactics in impersonation fraud. A deepfake attacker may create a crisis scenario so the target feels there is no time to think, verify, or follow procedure.

- Be cautious of phrases such as “do this immediately,” “I need this in the next 10 minutes,” or “the deal will fail if we wait.”
- Watch for emotional pressure such as fear, embarrassment, crisis language, or claims of severe consequences.
- Pause especially when urgency is combined with money movement, credential sharing, confidential files, or policy exceptions.

Example: A finance employee receives a voice call that sounds like the CFO saying, “Wire the payment before the bank closes or we lose the acquisition.” The urgency may be designed to override normal review and approval steps.

3.2 Requests for Secrecy

Secrecy is another major warning sign. Attackers often claim the matter is confidential to prevent the target from asking a colleague, manager, legal team, finance controller, or security team for help.

- Question requests that say “do not tell anyone,” “keep this between us,” or “this is board-sensitive.”
- Be especially careful if secrecy is used to block standard approval or documentation.
- Confidential work can still follow approved procedures; secrecy should not mean bypassing controls.

Example: A fake executive video call instructs an employee to transfer funds and says, “No one else can know until the announcement.” That instruction should trigger independent verification, not automatic compliance.

3.3 Pressure to Bypass Normal Approval Processes

Deepfake scams often succeed when people make exceptions to established procedures. Fraudsters may impersonate senior leaders because employees may feel uncomfortable challenging authority or slowing down an executive request.

- Do not skip purchase approvals, payment controls, vendor verification, or legal review because the requester sounds senior.

- Be wary of instructions such as “I approve it personally,” “we can document it later,” or “don’t raise a ticket.”
- Use documented workflows for payments, vendor changes, account access, confidential data release, and contract approvals.

Example: A caller who sounds like the CEO asks procurement to approve a new vendor without onboarding checks. The correct response is to follow the vendor approval process and confirm the request through an official channel.

3.4 Unexpected Communication Channels

A sudden change in communication channel can be a red flag. Attackers may move the conversation to personal messaging apps, new phone numbers, social media accounts, or unfamiliar meeting links because those channels are harder to monitor or verify.

- Question requests sent through a new number, personal email address, unknown chat app, or unfamiliar video meeting link.
- Be careful if a person who normally uses email suddenly insists on voice calls only.
- Do not trust a link or number provided inside the suspicious message itself; use known company directories or previously saved contact details.

Example: A manager who usually communicates through corporate email suddenly contacts an employee through a personal messaging app and asks for confidential files. The change in channel should be treated as suspicious until independently verified.

3.5 Resistance to Independent Verification

A legitimate requester should not object to reasonable verification, especially for sensitive actions. Resistance to verification may indicate the interaction is being controlled by an attacker who needs you to act before the deception is exposed.

- Be cautious if the person says there is “no time” to verify or becomes frustrated when you ask to confirm.
- Watch for attempts to shame, intimidate, flatter, or pressure you into compliance.
- Independent verification is not distrustful; it is standard risk control.

Example: A caller claiming to be a senior leader says, “If you call anyone else, you’ll compromise the project.” That response should increase suspicion and trigger escalation to security or management.

4. The Verification Rule

The verification rule is simple: do not let a convincing voice, image, video, email, or message become the only proof of identity. For any sensitive request, verify through a separate trusted path before taking action.

4.1 Don't Rely on One Communication Channel

One channel can be compromised, spoofed, or imitated. A suspicious email can be paired with a fake voice call, and a fake video call can be supported by a lookalike chat account. Verification should happen outside the channel where the request originated.

- If the request arrives by video call, verify by calling a known number or using an approved internal workflow.
- If the request arrives by email, confirm through a phone number or messaging account already listed in the corporate directory.
- If the request arrives through chat, verify using an official ticket, finance system, or direct manager confirmation.

Example: A video call appears to show a department head approving a payment. Before processing it, confirm through the approved payment workflow or a separate call to the department head's known number.

4.2 Call Back Using a Known Number

For urgent or sensitive requests, call the person back using a number you already trust. Do not use a number provided in the suspicious message, email signature, chat, voicemail, or meeting invite because that may be controlled by the attacker.

- Use a number from the corporate directory, HR system, approved contact list, or previously saved contact record.
- If the person does not answer, wait or escalate through normal channels rather than acting on the suspicious request.
- For family or personal scenarios, call the person directly using the number already saved in your phone.

Example: A voice message that sounds like a colleague asks for an urgent password reset. Instead of replying to the message, call the colleague using the number listed in the official directory and confirm whether they made the request.

4.3 Verify the Request Independently

Verification should confirm both the requester's identity and the request itself. A real person may have sent a message, but the details could still be wrong, outdated, or manipulated. Independent verification means checking the request against reliable records and approved processes.

- Confirm payment requests against purchase orders, invoices, contracts, vendor master data, or finance system records.
- Confirm access requests against ticketing systems, role-based access rules, or manager approvals.
- Confirm unusual instructions with the relevant owner, such as finance, legal, HR, procurement, IT, or security.

Example: A message asks you to change a vendor's bank account details. Before making the change, verify the request using the vendor change procedure, contact the vendor through existing records, and obtain the required internal approvals.

4.4 Use a Second-Person Approval for High-Value Transactions

High-value or high-risk transactions should never depend on one person's judgment alone. A second-person approval creates a deliberate pause and reduces the chance that urgency, authority pressure, or synthetic media will lead to an irreversible mistake.

- Require dual approval for wire transfers, new vendor setup, bank detail changes, large refunds, payroll changes, and sensitive data releases.
- Make the second approver independent from the person who received the original request.

- Document the verification steps so the organization can audit the decision if needed.

Example: If an executive appears on a video call asking for a large same-day transfer, the finance team should require a second authorized approver and confirm the transaction through the approved payment system before releasing funds.

5. STOP → VERIFY → APPROVE

The safest way to respond to a suspicious request is to slow the situation down. Deepfake scams are designed to create speed, pressure, and confusion. The STOP → VERIFY → APPROVE model creates a simple decision path that anyone can follow before approving a payment, sharing information, granting access, or acting on an unusual instruction.

5.1 STOP — Don't Act Immediately

Stopping is the most important first step. It interrupts the attacker's pressure tactic and gives you time to think clearly. A legitimate request should be able to withstand a short pause for verification, especially if money, access, confidential information, or reputation is involved.

- Do not click links, open attachments, approve payments, share passwords, or release files while under pressure.
- Pause if the message contains urgency, secrecy, authority pressure, or unusual instructions.
- Write down the request exactly as received so you can verify the details without relying on memory.

Example: A video call appears to show a senior leader asking you to approve an emergency vendor payment. Instead of approving immediately, pause the transaction, capture the request details, and move to verification.

5.2 VERIFY — Confirm Through a Trusted Channel

Verification means confirming the request outside the original channel. Do not treat the suspicious call, video, message, or email as proof of identity. Use trusted records, approved workflows, and independent contacts to confirm both the person and the action being requested.

- Call back using a known phone number from the corporate directory or saved contact list.
- Confirm financial requests against purchase orders, invoices, vendor records, or payment systems.
- Use official IT, HR, finance, procurement, or security workflows when access or sensitive information is involved.
- Escalate to a manager or security team if the requester resists verification.

Example: If a voice message asks for an immediate password reset, do not respond to the message directly. Confirm the requester's identity using an approved help desk ticket, identity verification procedure, or known internal contact details.

5.3 APPROVE — Proceed Only After Verification

Approval should happen only after the request has been independently confirmed and documented. If verification fails, is incomplete, or raises doubts, do not proceed. The goal is not to block legitimate work; it is to make sure high-risk actions are completed safely.

- Proceed only when the requester's identity, request details, and approval path are confirmed.
- Document the verification method, approver names, time, and supporting records.
- If the request involves high value or high risk, require a second authorized approver.
- If anything remains unclear, escalate rather than approve.

Example: A procurement manager receives a request to onboard a new supplier urgently. Approval should happen only after vendor checks, payment details, business justification, and the required approval chain are confirmed.

6. Quick-Check: Before You Approve

Use this quick-check before approving any unusual request, especially if it involves money, credentials, sensitive data, system access, public communication, or exceptions to normal controls. If you answer "yes" to any risk question, slow down and verify before proceeding.

- **Is the request unexpected?** For example, did it arrive outside the usual process, from a person who does not normally make this type of request, or at an unusual time?
- **Is there unusual urgency or secrecy?** Be cautious if the request demands immediate action, discourages discussion, or asks you not to tell anyone.
- **Is money or sensitive information involved?** Payments, bank details, credentials, personal data, confidential files, and access changes require stronger verification.
- **Is the person asking me to bypass normal process?** Watch for instructions to skip approvals, avoid documentation, ignore system workflows, or "handle it offline."
- **Have I verified the request independently?** Confirm through a trusted channel, known number, official workflow, second approver, or reliable internal record before taking action.

Decision rule: If the request is unexpected, urgent, secretive, financially sensitive, or outside normal process, do not approve it until independent verification is complete. When in doubt, escalate to a manager, security team, finance controller, or other designated approver.

Example: A video call appears to show a senior leader asking you to approve an emergency vendor payment. Instead of approving immediately, pause the transaction, capture the request details, and move to verification.

7. What to Do If You Suspect a Deepfake

If you suspect a deepfake, treat it as a potential fraud or security incident. The goal is to stop harm first, then verify, escalate, and preserve evidence. Do not argue with the requester or try to prove the media is fake during the interaction; focus on preventing action until the request is confirmed through trusted channels.

7.1 Stop the Transaction

Stopping the transaction prevents immediate loss while the request is investigated. This applies to payments, vendor changes, access grants, data transfers, refunds, contract approvals, payroll changes, or any action that could create financial, operational, legal, or reputational impact.

- Pause the payment, approval, access change, or file release immediately.
- Do not continue the conversation inside the suspicious call, meeting, chat, or email thread.
- If the action has already started, contact the relevant operations, finance, IT, or fraud team to freeze or recall it where possible.

Example: If a voice call that sounds like a director asks for a same-day wire transfer, put the transfer on hold before doing anything else. A short delay is far safer than releasing funds based on an unverified instruction.

7.2 Contact the Person Through a Known Channel

Do not use the phone number, meeting link, email address, or chat handle provided in the suspicious message. Instead, contact the person through a channel you already trust, such as a corporate directory number, previously saved contact, official ticketing workflow, or established internal communication path.

- Call the person back using a known number, not the number supplied in the suspicious request.
- Send a separate message through the organization's approved communication platform.
- Ask simple verification questions about the request, such as the business reason, reference number, approval record, or related ticket.

Example: A video call appears to show a manager asking for confidential customer data. End the call politely, then contact the manager using their known internal number or corporate chat profile to confirm whether the request is legitimate.

7.3 Alert Your Security/Fraud Team

Suspected deepfakes should be reported quickly because they may be part of a broader attack. Security and fraud teams can help determine whether other employees received similar requests, whether accounts were compromised, and whether any financial or data exposure occurred.

- Report the incident through the approved security, fraud, IT, or compliance channel.
- Include who appeared to make the request, what action was requested, when it happened, and which channel was used.
- If money, credentials, customer data, employee data, or regulated information is involved, mark the issue as urgent.

Example: If several employees receive voice messages that sound like the same executive asking for payments or passwords, the security team can investigate the pattern and warn others before the scam spreads.

7.4 Preserve the Message, Call Details, and Other Evidence

Evidence helps investigators understand what happened and may support account recovery, payment recall, legal review, insurance claims, or law enforcement reporting. Preserve the original message and related details without editing, forwarding unnecessarily, or deleting anything.

- Save emails, chat messages, voicemail files, screenshots, caller ID details, meeting links, timestamps, and usernames.
- Record what action was requested and whether any action was taken before the suspicion was raised.

- Do not edit the media file or message, because changes may reduce its value for investigation.

Example: If a suspicious video meeting link was used, capture the meeting invitation, participant display names, chat messages, time of call, and any request made during the interaction.

8. One-Minute Deepfake Detection Checklist

Use this compact checklist whenever a request feels unusual, urgent, sensitive, or out of process. It is designed for quick use by employees, finance teams, procurement teams, IT support teams, security teams, and managers.

- **STOP:** Do not approve, pay, share, click, or grant access immediately.
- **LOOK:** Check for visual clues such as unnatural blinking, mismatched lighting, blurred edges, or lip-sync issues.
- **LISTEN:** Notice robotic tone, odd pauses, missing breathing sounds, or audio that does not match the mouth movement.
- **QUESTION:** Ask whether the request is unexpected, urgent, secretive, financially sensitive, or outside normal process.
- **VERIFY:** Contact the person through a known number, trusted directory, official workflow, or separate approved channel.
- **ESCALATE:** Alert security, fraud, IT, finance, or management if anything feels suspicious.
- **PRESERVE:** Save messages, call details, screenshots, timestamps, links, and related evidence.
- **APPROVE ONLY AFTER CONFIRMATION:** Proceed only when the requester, request details, and approval path are independently verified.

Final reminder: Deepfake detection is not just about spotting technical flaws. The safest defense is a disciplined verification habit: slow down, confirm independently, use approved processes, and escalate when something does not feel right.

Conclusion

Deepfakes are becoming harder to spot, but you don't have to identify the fake to stop the fraud. The most effective defense is a simple verification habit: pause when something feels urgent, verify through a trusted channel, and approve only after the request is confirmed.

STOP → VERIFY → APPROVE.

Use this checklist whenever a request involves money, sensitive information, or unusual instructions.

GSDC
Global Skill Development Council

CERTIFICATION IN GENERATIVE AI IN CYBERSECURITY

AGENTIC AI DEVELOPER CERTIFICATION BASED ON REAL-WORLD AGENT FRAMEWORKS TO DESIGN, BUILD, AND DEPLOY AUTONOMOUS AI SYSTEMS.

ABOUT GSDC CERTIFICATION

LIFETIME VALIDITY

GSDC Certification is an globally accreditted certification with lifetime validity.

EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.

CREATED BY EXPERTS

GSDC certifications are created and authored by world's leading experts in the field.

LEARNING MATERIALS

Get access to learning materials such as videos, ebooks, templates, and practice exams, which will help you clear the certification exam.

LEARNING OBJECTIVE

- Make an impact in the cutting-edge field of artificial intelligence.
- Validate your generative AI application skills.
- Encourage the development of generative AI technologies.

Enroll now with the code **LEARN20** To avail **20%** discount

Enroll Now

www.gsdcouncil.org