AI FRAUD PREPAREDNESS / VERSION 2.0

AI SAFETY.

Everyday confidence.

AI can fake the voice, the face and the story.

A practical companion to the You AI Fool? test.
For you, your family and the people who trust you.

SIX AREAS. REAL SITUATIONS.

01 Voices & identity 02

02 Personalised scams 03

03 Fake evidence 04

04 Sharing with AI 05

05 AI access & permissions 06

06 Response & recovery 07

PAUSE. CHECK. THEN ACT.

Slow down. Check through a route you already know.
Then decide what to do.

01 / FAMILIAR VOICE.

Familiar voice. Fresh check.

A familiar voice, face or account can be copied or taken over. Check the request, not just how convincing it feels.

THE SITUATION

A message sounds like a relative who needs money urgently. Or a colleague appears on video with new bank details for a supplier. The person feels familiar. The request still needs a check.

BEFORE YOU ACT

Pause the payment. Do not send a small amount as a test.

Call a saved number. Use contact details you already had, rather than a number in the message.

Check changes separately. Confirm supplier bank changes with an established contact and follow the normal approval process.

WORDS YOU CAN USE

> **I want to help. I'll call you back on the number I already have.**

MAKE IT A FAMILY HABIT

Agree how everyone will check unusual requests for money or help. A private phrase can be an extra check, alongside a callback. Keep it out of public posts and AI chats; change it if exposed. If you cannot reach the person, check with another trusted relative before paying.

Informed by sources 1 and 2. Sample phrases are written for this guide.

02 / A STORY THAT

A story that knows you.

A message can know your job, friends or recent plans without coming from someone you can trust.

WORDS YOU CAN USE

Please send the details.
I’ll check them and come back to you.

WHY IT FEELS CONVINCING

A recruiter refers to your work and interviews you on video. A proposal names a mutual contact. An online acquaintance gradually introduces an investment.

Personal details, polished writing and realistic calls can all be copied or fabricated.

WHAT TO DO

Contact the employer or mutual contact using details you find separately. Open the service yourself instead of following a sign-in link.

Before sharing ID, installing an app or moving money, confirm who is asking, why it is needed and where it will go.

NOTICE THE PRESSURE

Urgency, secrecy, flattery or fear are reasons to slow down. Talk to someone outside the conversation before a decision you cannot easily undo.

A credible profile or HTTPS connection does not prove who is behind an offer. Sources 1 and 2.

03 / LOOKS REAL.

Looks real. Proves what?

Separate the claim from the picture. Even genuine media can be used to support a false request.

WHAT YOU SEE	WHAT TO CHECK
A familiar investor in a video	Find the original recommendation and check the firm through the relevant official register.
A genuine disaster photo in an appeal	Confirm the appeal with the charity and find its official donation page separately.
A screenshot showing a payment or profit	Check your own bank or service account. A screenshot is not confirmation that funds have arrived.

WORDS YOU CAN USE

“

I’ll check the source before I act on the story.

KEEP IN MIND

More copies of a claim are not more independent sources. Another chatbot or AI detector is not a final verdict.

Registration or a genuine endorsement alone does not make an investment safe or suitable. Sources 1, 2 and 4.

04 / ANONYMISATION:

Anonymisation: more than names.

Initials can hide a name while leaving enough context to identify the person or reveal private information.

A SAFER START

Explain this type of contract clause using a fictional example.

BEFORE YOU PASTE OR UPLOAD

Check what the content reveals and whether you may share it with this particular tool and account. For work information, follow your organisation’s rules and use approved services.

Look beyond names: job titles, dates, project details, file properties and screenshots can reveal people or confidential work.

SETTINGS ARE NOT PERMISSION

Temporary chats and training controls may reduce some exposure. They do not replace checking the service’s privacy, retention and access arrangements, or your right to share the material.

SHARE LESS, CHECK MORE

Use fictional examples or the smallest necessary extract. Keep passwords, login codes and recovery keys out of prompts. Verify important AI answers against reliable original sources before using them.

REMOVE NAMES **+ CHECK THE REMAINING CONTEXT**

Sources 3 and 4. Data controls differ between services, accounts and settings.

05 / WHEN AI

When AI can act for you.

An assistant that can send, share or delete needs firmer limits than one that only drafts.

01

LIMIT ACCESS

Only what the task needs

02

REVIEW THE ACTION

Recipient, file and purpose

03

APPROVE THE STEP

Before anything is sent

BEFORE YOU CONNECT

Read the permissions. Can the tool only read messages, or can it send and delete them too? Start with limited access or a separate test account where possible.

Use settings that require approval. Telling the assistant “ask me first” is different from a control that prevents an action without approval.

WATCH THE INSTRUCTIONS IT READS

An email or webpage may tell an assistant to upload a work document. That content is not permission to share it.

Confirm the request another way. Check the destination and the actual file before approving. Review connected apps and remove access you no longer need.

If it acts unexpectedly: **stop it, remove its access, then investigate.**

BEFORE YOU CONNECT

What can it read?
What can it change?
What still needs my approval?

Sources 5 and 6. Removing an extension may not remove its permissions in the connected account.

06 / SOMETHING HAPPENED.

Something happened. Start here.

You do not need to prove that AI was involved before asking for help. Fast, calm action matters.

01

Money sent

Contact your bank or payment provider through its official app or a known number. Ask whether the payment can be stopped or recovered. Recovery is not guaranteed.

02

Account details exposed

Use a trusted device to secure affected accounts. Change exposed or reused passwords, sign out other sessions and check recovery details. Protect email first if it resets other accounts.

03

An AI tool shared a file

Stop the tool and remove its permissions in account settings. Check what was sent and to whom. If work information is involved, tell your IT or security contact promptly.

04

Unwanted device access

Disconnect a device with suspected remote access or malicious software from the network. Stop using it for banking or other sensitive tasks and seek trusted technical help.

KEEP EVIDENCE. REPORT THE INCIDENT.

Keep messages, account names, times, links and payment references. Report through the affected service and your local police or national fraud-reporting route. Warn contacts if your account was used to approach them.

Beware a second scam: do not pay an unexpected contact who promises to recover your money.

Follow your bank’s, employer’s and service provider’s incident instructions. Sources 7 and 8.

MAKE IT A HABIT / SMALL STEPS.

Small steps. Shared confidence.

Choose three actions today. Add the others this week. You do not have to be a technical expert to make a useful start.

01 **Agree the family check**

Decide which number to call and how to check an unexpected request. Anyone can pause without embarrassment.

02 **Protect your key accounts**

Use unique passwords or passkeys where available. Turn on two-step verification, especially for email and banking.

03 **Review AI connections**

Remove access you no longer need. Require review before sending, deleting, paying or sharing private information.

04 **Keep a short response plan**

Save your bank’s official contact route and a trusted person to call. Know how to get support and remove connected-tool access.

05 / SHARE WITHOUT BLAME

If something feels wrong, call me.
Even if you have already clicked or paid.

Getting fooled does not make someone foolish. Help first; work out what happened together.

KEEP NEARBY

The 20-second safety check.

01 Was I expecting this?

02 Does it ask for money, access or private information?

03 Have I checked through a route I already know?

If you are unsure, pause the action.

SOURCES & FURTHER READING

1 **FBI / IC3**
Generative AI and financial fraud

2 **INTERPOL**
AI deepfake scams: public awareness campaign

3 **UK NCSC**
ChatGPT and large language models: the risks

4 **Canadian Centre for Cyber Security**
Generative artificial intelligence: ITSAP.00.041

5 **UK NCSC**
Managing the cyber risk of agentic AI

6 **OWASP**
AI Agent Security Cheat Sheet

7 **UK NCSC**
Top tips for staying secure online

8 **GOV.UK**
Avoid and report internet scams and phishing (UK)

Sources checked 8 November 2026. Examples are illustrative. Public guidance informs this guide; the organisations listed have not endorsed You AI Fool? or Globawise. This is not an audit, certification or guarantee of safety.

The test summarises 18 answers across six areas. Its score does not indicate the probability of being scammed. Get your result first, then optionally save it and sign up for notifications about new threats and test updates.

youaifool.com