

Using AI Safely and Responsibly: Privacy, Bias, Honesty and AI Scams

Protect your data when using AI, check what it tells you, notice bias, use AI honestly at study and work, recognise deepfakes and AI-powered scams, and apply a simple personal checklist.

20 min read · Content last verified 8 October 2026 · Reviewed by Shravya M

Free awareness guide. Not legal or financial advice. <https://www.shravyadigital.com/guides/using-ai-responsibly>

CONTENTS

1. Before you start
2. Chapter 1: Your data and AI
3. Chapter 2: Checking what AI tells you
4. Chapter 3: Bias and fairness
5. Chapter 4: Honesty: AI at school and work
6. Chapter 5: Deepfakes and AI-made content
7. Chapter 6: AI-powered scams
8. Chapter 7: A personal AI checklist
9. Key words

Before you start

This guide is the reading for **Course 5**, the last course of the AI Foundations Pathway. It takes about 20 minutes.

You now know how AI models, LLMs, RAG and agents work. This course is about using them well: protecting your privacy, checking answers, spotting bias, being honest about AI use at study and work, and recognising the AI-powered scams aimed at Indian families.

All names and examples here are made up.

Chapter 1: Your data and AI

Every time you use an online AI tool, you share data with it: what you type, files you upload, your voice, sometimes your photos and location.

Before you use one, ask:

- **Where does my data go?** To the company's servers, and possibly outside India.
- **Is it used for training?** Many free tools may learn from your chats unless you switch that off in the settings.
- **Who can see it?** Some services let staff review conversations for safety.
- **Can I delete it?** Look for history and data-deletion options.

Never share: **passwords, OTPs, PINs, Aadhaar, PAN or bank details**, medical records, or other people's personal information without their consent.

India's **Digital Personal Data Protection Act, 2023** gives you rights over your personal data, including to ask a company what it holds about you, to have it corrected or erased, and to complain if it's misused.

The rule: If you wouldn't post it on a public notice board, don't paste it into a chatbot.

Chapter 2: Checking what AI tells you

AI tools can be confidently wrong. Build a checking habit:

1. **Ask for sources**, then open them and check they exist and say what was claimed.
2. **Cross-check** important facts with an official or trusted source: a government website, a textbook, a qualified professional.
3. **Be extra careful** with numbers, dates, names, laws, medicines and anything very recent or very local.
4. **Notice overconfidence**. A smooth, certain tone is not evidence.
5. **For health, legal and money decisions**, use AI to understand and prepare, then decide with a qualified person.

Chapter 3: Bias and fairness

AI learns from human data, so it can repeat human unfairness:

- Assuming a doctor is "he" and a nurse is "she".
- Working less well for certain accents, languages, skin tones or regions.
- Scoring people from some backgrounds unfairly in hiring, lending or admissions.

What you can do:

- **Notice it.** Ask yourself whether an answer would change if the person were from a different group.
- **Don't let AI make important decisions about people alone.** Keep a human who can explain and change the decision.
- **Report it** to the tool's maker when you see it.

India's NITI Aayog has set out principles for **responsible AI**, including safety, equality, inclusivity, privacy, transparency and accountability. They're a useful checklist for any AI system that affects people.

Chapter 4: Honesty: AI at school and work

Using AI is not cheating in itself. Hiding it where honesty is expected can be.

- **Follow the rules** of your school, university, exam body or employer on AI use. They differ, and they change.
- **Say when you used AI** if you're asked, or if the reader would reasonably assume the work is entirely yours.
- **You're responsible for the result.** "The AI wrote it" is no excuse for errors, copied text or made-up references.
- **Respect others' work.** AI can reproduce text and images closely enough to copy someone's work. Check before you publish.
- **At work, use approved tools only** for company data, and never paste client or confidential information into a personal AI account.

Chapter 5: Deepfakes and AI-made content

A **deepfake** is audio, video or an image created or altered with AI to look or sound like a real person saying or doing something they never did. The government calls this **synthetically generated information**.

On **10 February 2026**, the government amended the IT Rules, 2021 to deal with it. Platforms must ensure **clear labelling** of permissible AI-generated content, so users can identify it, and must act against unlawful content such as impersonation.

But **a fraudster's video won't carry a label**. Clues that something may be AI-made:

- Lips slightly out of sync, odd blinking, or a face that blurs at the edges
- A voice that is flat, or has strange pauses or pronunciation
- Hands, text or backgrounds that look distorted
- Above all, an **unusual request**: money, secrecy, urgency

The best test is not your eyes. It's **checking through a different channel**: call the person back on a number you already have, or check the official website.

The rule: Seeing or hearing someone is no longer proof it's them.

Chapter 6: AI-powered scams

Criminals use AI to make old scams more convincing:

- **Voice-clone calls:** a "family member" in trouble asking for money urgently. Agree a family code word, and always call back on the number you know.
- **Deepfake endorsements:** a famous person "recommending" a trading app or scheme.
- **"AI trading bots" promising guaranteed returns:** no tool can guarantee profits. Check that any adviser is registered with SEBI before you invest.
- **Fake AI apps:** lookalike "ChatGPT" or "AI photo" apps that steal logins or money. Download only from official stores, and check the developer's name.
- **Perfect-English phishing:** AI writes flawless scam messages, so good grammar is no longer a sign of a real message.

If you've been cheated, call **1930** or report at **cybercrime.gov.in** as soon as possible. Our [Deepfake & Investment Scams guide](#) covers these in depth.

Chapter 7: A personal AI checklist

Before you rely on an AI tool, run through this list:

1. **Is it the real app or website** from the company it claims to be?
2. **What data am I giving it**, and have I checked the privacy settings?
3. **Have I checked the important facts** against a reliable source?
4. **Could the answer be biased**, and does a human need to make this decision?
5. **Am I allowed to use AI here** (school, exam, work), and should I say that I did?
6. **If it can take actions**, have I limited what it can do and kept approval for anything important?
7. **Does this request involve money, urgency or secrecy?** Then stop and verify through another channel.

AI is one of the most useful tools of our time. Used with these habits, it makes you faster and better informed, without making you easier to fool.

Key words

- **Personal data:** information that identifies you, such as your name, number, ID or photos.
- **DPDP Act, 2023:** India's law on how personal data must be handled.
- **Bias:** unfair results that favour or disadvantage some groups.
- **Responsible AI:** building and using AI safely, fairly, transparently and accountably.
- **Deepfake / synthetically generated information:** AI-made audio, video or images of real people.
- **Voice clone:** an AI copy of someone's voice.
- **Verification through another channel:** checking a request by contacting the person or organisation in a separate, trusted way.

Take the course for what you just read

Take the free 10-question assessment and pass with 8 or more. Every course you pass counts towards a Pathway Certificate issued by Shravya Digital, earned when you pass all five courses in the pathway: <https://digital.shravyaworld.org/courses/using-ai-responsibly>

Sources

Checked against these primary sources on 8 October 2026. Rules change: if something here looks out of date, trust the source.

- MeitY: Data protection framework (DPDP Act, 2023) (<https://www.meity.gov.in/data-protection-framework>)
- NITI Aayog: Responsible AI for All (<https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>)
- MeitY, "Government Strengthens Regulatory Framework to Address AI-Generated Deepfakes" (PIB, 6 Aug 2026) (<https://www.pib.gov.in/PressReleasePage.aspx?PRID=2295500>)
- SEBI registered intermediaries (<https://www.sebi.gov.in/intermediaries.html>)
- National Cyber Crime Reporting Portal and helpline 1930 (<https://cybercrime.gov.in/>)