The background features a dark blue cityscape at night, with a suspension bridge visible in the upper right. Overlaid on this are various digital and data visualization elements, including glowing lines, grids, and abstract shapes. A person's silhouette is shown in profile on the right, holding a tablet. In the lower left, a hand in a suit sleeve is extended, palm up, as if presenting or interacting with the digital elements. The overall aesthetic is high-tech and digital.

Safe & Responsible AI Use for Cyber Resilience

Empowering Digital Citizens in the Age of AI

Agenda

- What is Cyber Resilience?
- What to trust in the age of AI?
- Everyday examples of trust zones
- Ensure Safe AI use
- Nurturing Everyday Habits
- Reporting and Safety tools



What is Cyber Resilience?

Anticipate: Know that AI scams exist before they happen.

Withstand: Use strong passwords and 2FA so an attack doesn't break you.

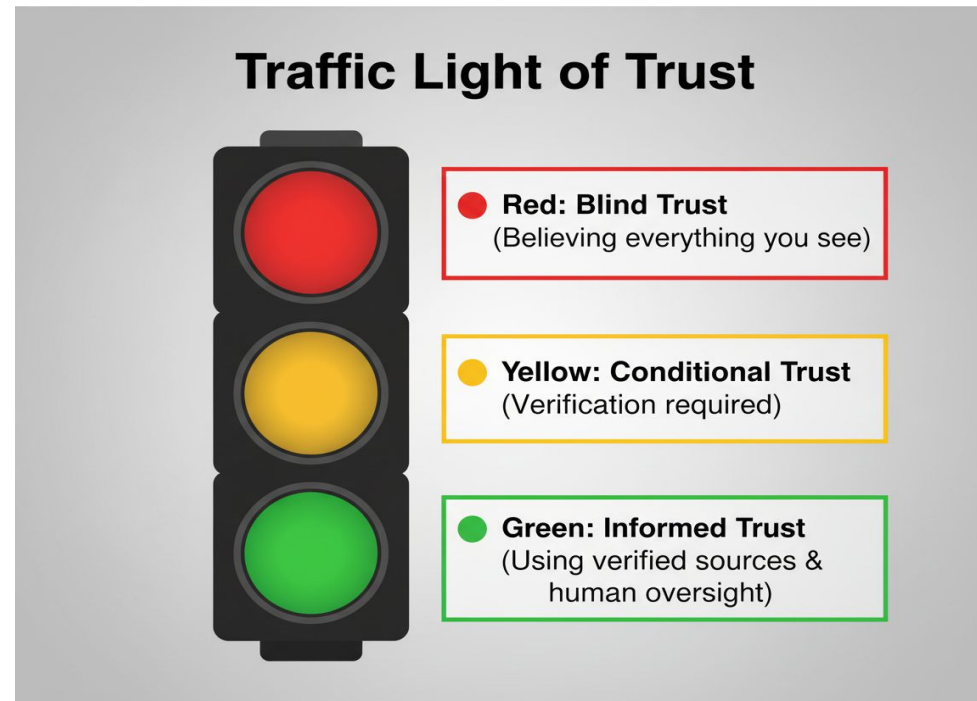
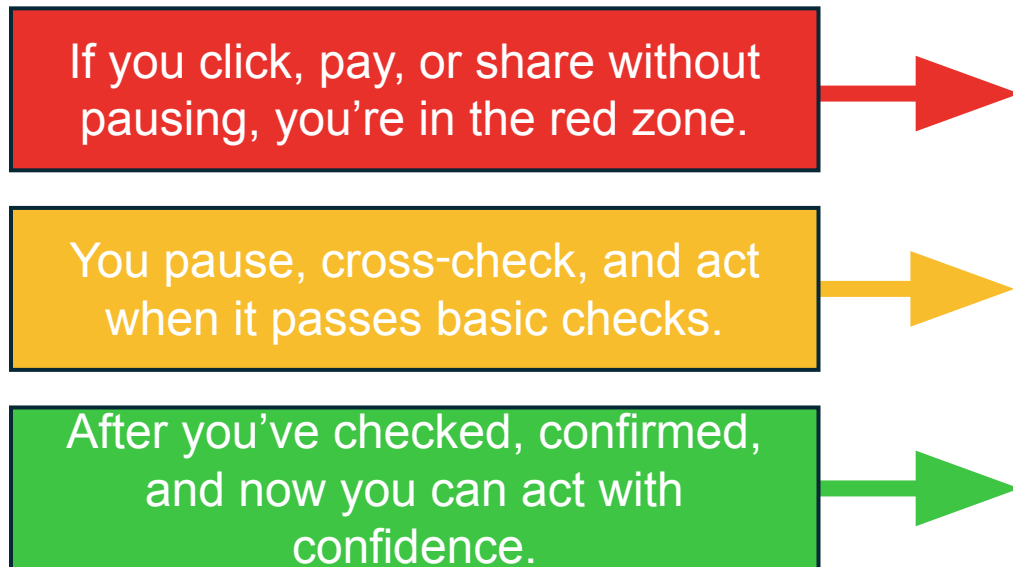
Recover: Have backups so you can get back to normal quickly.

Adapt: Learn from every "error" to be stronger next time.



The "Informed Trust" Model

Any unsolicited request for money, OTPs, or biometric data—especially if it uses a "celebrity" or "authority" or "family" voice/face.



Conversational AI (The Persona Trap)

Imagine someone chatting with an AI persona late at night. The character feels friendly, listens patiently, and even gives advice. It feels safe.

Sharing personal details because the AI persona 'feels trustworthy.'

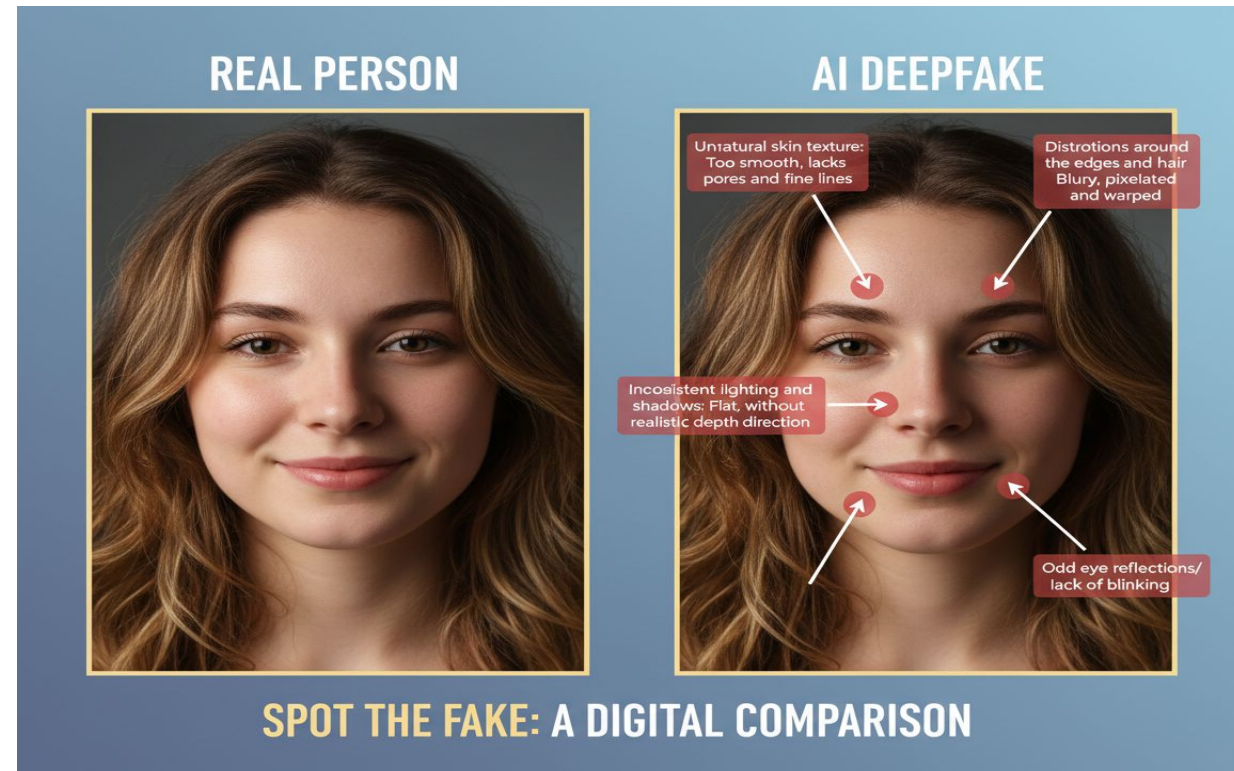
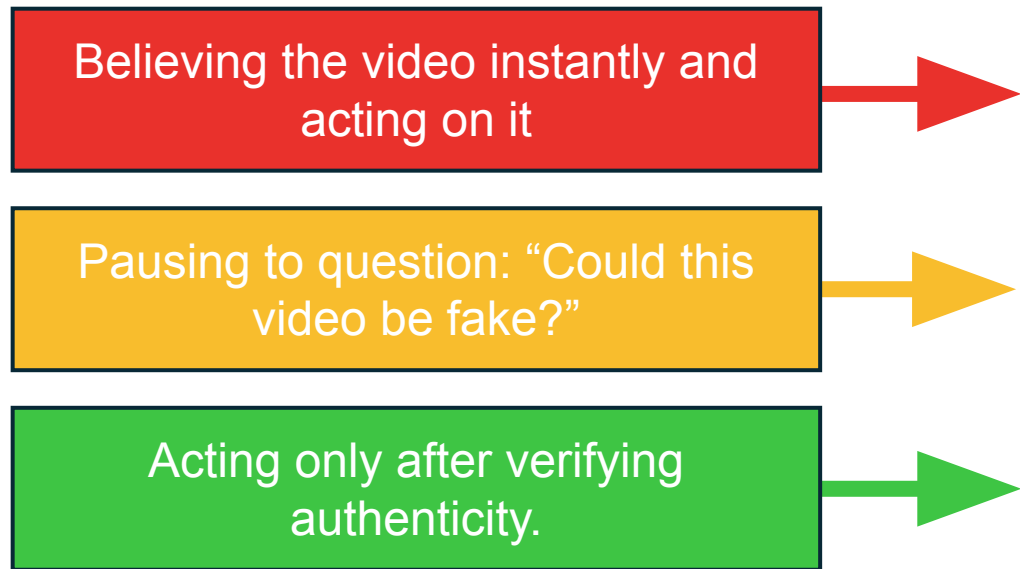
You keep chat general, avoid sensitive or emotional disclosure.

Treat AI personas as tools, not human friends. Verify advice with a trusted human source



The Deepfakes

Imagine receiving a video of a well-known business leader or even a family member asking you to urgently transfer money. The face and voice look real—but what if it's a deepfake?



AI Trends and Image Sharing

Imagine uploading a child's photo to turn it into a cute Ghibli-style portrait, or sharing family pictures online with filters. It feels fun and creative .

Posting or uploading identifiable photos publicly without caution.

Sharing photos but with some caution such as without geotags, low resolution

Reduce resolution or anonymize images before uploading to AI tools.



Archie

Biometrics & Selfies

Imagine someone posting a selfie flashing the 'V' sign. It looks harmless—but AI can reconstruct fingerprints or facial data from that photo. What feels casual can become a biometric risk.

Uploading or posting high-resolution images.

Avoid biometric gestures

Sharing screenshots



AI in the Workplace/Classroom

Imagine pasting a confidential school report into a public AI tool to get a quick summary. It feels efficient—but it's like putting the report into a shredder that tapes it back together for someone else to read later.

Copy-pasting confidential reports, student records, or exam papers directly into public AI tools.

Remove sensitive information as names, address etc.

Treat public AI as assistants for general tasks, not for sensitive data.



Ensure Safe use of AI

1. Before You Share

- 🛑 **Stop:** Don't upload high-resolution selfies, confidential documents, or sensitive personal details.
- 🟡 **Pause:** Crop, blur, or screenshot images; strip metadata; anonymize text.
- ✅ **Go:** Share only non-sensitive, low-resolution, or anonymized content.



Ensure Safe use of AI

2. Before You Trust

- 🛑 **Stop:** Don't act on unsolicited requests for money, OTPs, or biometric data—even if they look like family, authority, or celebrity.
- ⏸ **Pause:** Verify through another channel (call, official site, fact-check).
- ✅ **Go:** Act only after confirming authenticity via trusted sources.



Ensure Safe use of AI

3. Before You Paste

- 🛑 **Stop:** Don't paste confidential school/work reports into public AI tools.
- ⏸ **Pause:** Remove names, IDs, or sensitive details first.
- ✅ **Go:** Use enterprise-approved or secure AI platforms for sensitive work.



Ensure Safe use of AI

4. Before You Believe

- 🛑 **Stop:** Don't blindly trust videos or images online.
- ⏸ **Pause:** Look for inconsistencies, check fact-checking portals.
- ✅ **Go:** Trust only verified sources and official communications.



Everyday Habits

- **Limit** oversharing on social media.
- **Prefer** avatars, illustrations, or filters that obscure biometrics.
- **Keep** privacy settings updated.
- **Treat** AI personas as tools, not human friends.



Reporting & Safety Tools



Chakshu is for:

👉 Reporting **suspicious communications**

Cybercrime Portal (cybercrime.gov.in) is for:

👉 When **money is lost** or identity is stolen

National Cyber Crime Portal:

👉 Dial 1930.