CERT-In Guidelines on AI Security Auditing – Agentic AI

Prelims: *Current events of National and International importance| Science & Technology*

Why in the News?

The Indian Computer Emergency Response Team (CERT–In) has introduced the AI Security Auditing Guidelines– Framework for Secure, Ethical, and Auditable Artificial Intelligence Systems to strengthen the security and accountability of autonomous AI systems.

What is Agentic AI?

- Agentic AI refers to AI systems capable of independently planning, making decisions and executing multi–step tasks with limited human intervention.
- It can access external tools, execute code, communicate with external systems and delegate tasks to other AI agents.
- **Key features**– Autonomous execution, persistent memory, tool utilisation and task delegation.

Major Security Risks Identified

- **Privilege abuse** – Accessing resources beyond authorised permissions.
- **Unauthorised tool use**– Using external tools without approval.
- **Credential misuse**– Misusing passwords and access tokens.
- **Memory manipulation**– Altering stored information.
- **Data leakage**– Sharing one user's information with another.
- **Unsafe delegation**– Uncontrolled transfer of tasks between AI agents.

Key Guidelines Prescribed by CERT-In

Area	Security requirements
Access control	Test permissions and prevent privilege abuse.
Task delegation	Detect repeated delegation and uncontrolled task creation.
Authentication	Unique identity and authenticated communication for every agent.
Audit trail	Record requests, decisions, permissions, tools and external actions.
Emergency control	Suspend individual agents/workflows and revoke credentials.
Data protection	Prevent memory manipulation and unauthorised data retention.

Multi-Agent AI Systems

- Multiple AI agents collaborate and exchange tasks. CERT–In requires authenticated, logged communication and traceable records of their activities.

Additional Requirements

- Maintain an inventory of AI models, training data, software and external tools.
- Document information gaps relating to third–party models.
- Independently verify AI–generated outputs; they cannot replace audit evidence or professional judgement.

Quick facts

CERT-In
• **Established**– 2004. • **Nodal ministry**– Ministry of Electronics and Information Technology (MeitY). • **Statutory basis**– Section 70B of the Information Technology Act, 2000. • **Headquarters**– New Delhi. • **Function**– National agency for cybersecurity incident response.

Reference

[The Hindu| CERT–In Guidelines on AI Security Auditing – Agentic AI]

SHANKAR IAS PARLIAMENT
Information is Empowering