AGENTIC AI

MORE CAPABLE UNDER DUE CONTROL

GovCERT.HK Government Computer Emergency Response Team Hong Kong

HKCERT Hong Kong Computer Emergency Response Team Coordination Centre 香港網絡安全事故協調中心

Cyber Defender 守網者

WHAT IS AGENTIC AI?

Traditional AI responds to prompts; agentic AI can autonomously plan and execute goals.

CHARACTERISTICS

- **Breaks a goal into steps and executes them**
- **Connects to tools, systems, or external services**
- **Adjusts its next move based on results**

WHY ARE THE RISKS HIGHER?

Without safeguards, agentic AI turns wrong answers into wrong actions.

1. Excessive permissions

If it has access to too many systems, accounts, or data sources, its misuse can have a much wider impact.

2. Misleading instructions or inputs

Malicious content, fake data, or untrusted inputs may cause it to make poor decisions or take unsafe actions.

3. Lack of oversight

If no one reviews what it does and produces, small issues can grow quickly and go unnoticed.

3 COMMON RISK ENTRY POINTS

1. **Automated information handling**
   Risk: incorrect references, untrusted sources, or misleading content.

2. **Automated sending or submission**
   Risk: sending wrong content, sending to wrong recipients, or acting without review.

3. **Access to internal tools or systems**
   Risk: excessive access, data leakage, or unintended actions on internal resources.

5 TIPS FOR SAFE USE OF AGENTIC AI

1. Set permissions before connecting tools
2. Keep human review for high-risk actions
3. Verify the source of inputs
4. Handle data by sensitivity level
5. Maintain records for traceability

WISE AI USE STARTS WITH PROPER GOVERNANCE

3 KEY REMINDERS

1. Verify before you trust
2. Restrict before you deploy
3. Approve before it acts

TRUST ITS PRODUCTIVITY GOVERN ITS AUTONOMY