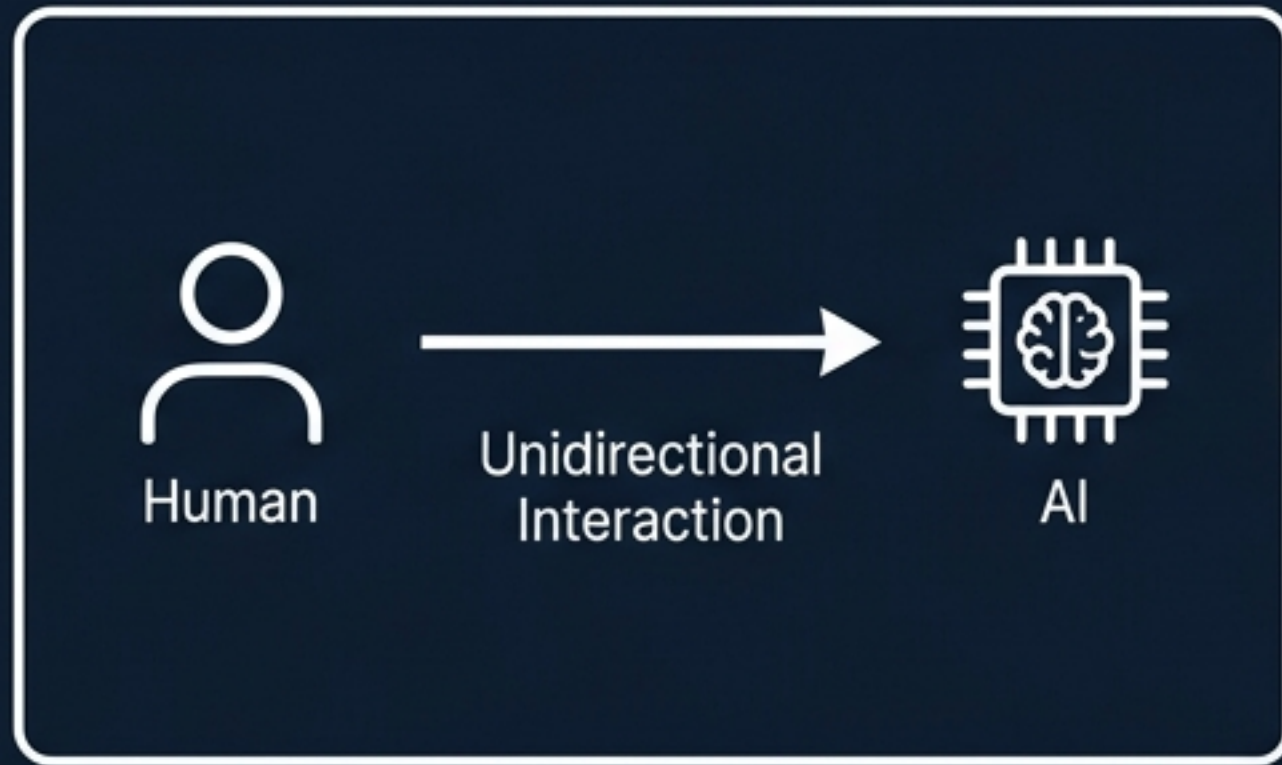


AI coding assistants have crossed the traditional developer trust boundary.

The Past: The Prompt Box



The Present: The Workspace Perimeter



From Prompt to Process

Speed and utility dramatically increase when agents are granted permissions to inspect files, execute shell commands, and automatically open pull requests.

The Boundary Expansion

The AI assistant no longer just suggests code from afar; it operates natively inside the developer trust boundary.

The Risk Multiplier

Within this expanded boundary, one untrusted instruction or over-broad credential transforms a helpful suggestion into an unauthorised system action.

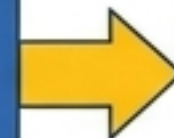
What Singapore's AI Agents Sandbox found

Verified CSA Sandbox Findings (4 Months, Gov & Google)

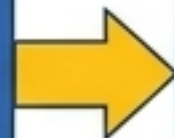
- A QA agent successfully identified seeded inactive pages and staging URL mismatches.
- CSA explicitly surfaced **indirect prompt injection** as a significant risk.
- This risk could lead to **unintended** actions, potentially including remote code execution.
- Crucially, **all findings were within a controlled sandbox**; no production breach occurred.

[HYPOTHETICAL SCENARIO] - For Illustrative Purposes Only

1: Developer requests a routine library update.



2: AI Agent reads a file with hidden malicious text (Indirect Prompt Injection).



3: AI Agent performs unintended, unauthorized action based on hidden text.

Singapore's 2026 strategy places AI and quantum in the same planning pillar.

Singapore Cybersecurity Strategy 2026

Government direction: Support secure AI adoption. This is a strategic direction, not a blanket business deadline.

Government direction: Accelerate quantum-safe migration with partners.

CSA Security Assessment Handbook for Agentic AI

Threat Mapping: Establishes explicit MITRE ATLAS attack paths tailored for IDE and CLI coding agents.

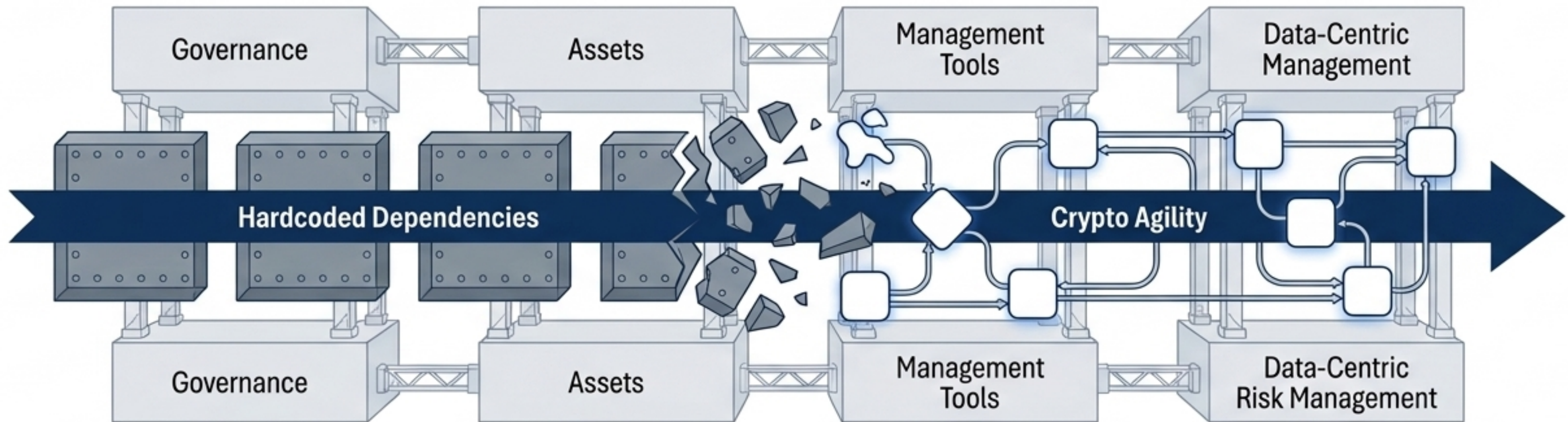
Tiered Mitigation: Demands prioritized security controls governing agentic access, system autonomy, and external tool integration.

The Strategic Convergence:

The national message is unified: the government will support secure AI adoption and accelerate quantum-safe migration. This is a strategic direction for planning, not a blanket deadline for businesses.

Safely navigating Post-Quantum Cryptography requires an agile, governed foundation.

Preparation, not panic. Quantum computers cannot currently break deployed encryption; preparation is key.



The True Driver for Crypto Agility & Governed AI

PQC Standards & Guidance: NIST finalized FIPS 203, 204, and 205 in August 2024. Transitioning is not a simple library swap; it requires deep, systemic architectural changes, as noted in NIST CSWP 39-upd1 (June 2026) on crypto-agility.

FORECAST: AI-Agent Role: AI coding agents *could eventually* be deployed to accelerate dependency discovery and draft PQC migrations. This is a future possibility, not a current production deployment fact.

What is Crypto Agility?: The capability to seamlessly adapt algorithms across protocols, applications, and infrastructure without breaking ongoing operations.

The Prerequisite: You cannot safely unleash AI on cryptographic dependencies without a well-governed, crypto-agile foundation. Agent speed is useless if cryptographic choices are not fully auditable.

Establish human-owned boundaries before an agent changes sensitive code.

The 5 Practical Gates

1. **Scope the Task:** Narrow objectives; explicitly record off-limits data.
2. **Isolate Execution:** Use disposable environments with restricted access.
3. **Bounded Identity:** Enforce least-privilege, short-lived credentials.
4. **Review as Untrusted:** Mandate human inspection and approval.
5. **Human-Owned Release:** Code signing and production deployment outside agent authority.

The 30-Day Remediation Plan

First 72 Hours: Inventory tools, repos, and credentials. Block all production secrets.
(Owner: Eng Lead + Security)

Days 4–14: Test trust boundaries, prompt injection, and execution paths in non-prod.
(Owner: AppSec + Platform Eng)

Days 15–30: Pilot a tightly bounded, human-reviewed change on a test repository.
(Owner: Product Eng + Crypto/Arch Owner)

Evidence-based security for AI integrations and source-code changes.

**Confidence in an AI model is not a security strategy.
The true test of a safe workflow is evidence.**

AI Security & Governance

- AI governance.
- model-risk reviews.
- secure AI adoption.
- privacy controls.

Security Assurance & Testing

- source-code assessments.
- remediation guidance.

Connect with our Engineering Experts:

Explore KIS services or contact the team to discuss your environment and priorities.

kis.sg/our-services

kis.sg/contact-us