# SafeSkill.io

## AI Agent Skills Security Platform

*Inspect. Verify. Trust.*

ThreatBook

Third-party Skill calls are growing **40% month-over-month**. Every Skill an AI agent invokes is a potential entry point into your enterprise systems.

### Supply Chain Poisoning at Scale

**The "trust trap" of open marketplaces**

Marketplaces like ClawHub lack strict review. The "ClawHavoc" campaign has already seeded over 1,000 malicious packages. Downloaded tools may ship with hidden backdoors disguised as legitimate functions. Compromise lands at install time and stays invisible afterwards.

### Attack Techniques Are Evolving

**Legacy defences are missing the new threats**

Adversaries use curl/sh dynamic execution, prompt injection, and Markdown semantic obfuscation to evade detection, then chain critical vulnerabilities into click-bait that hijacks the host. Static tools can't see it.

### Compliance and Audit Pressure

**A burden enterprises can't carry alone**

Enterprises adopting AI tools can't prove the tools are safe, facing dual pressure from internal audit and regulators. Without professional Skill inspection reports, compliance auditors have nothing to point to.

## From submission to curated marketplace: every Skill passes rigorous multi-dimensional inspection and evaluation.

### Skills Multi-Dimensional Inspection

Multi-source data ingestion (ClawHub, GitHub, and more)
Metadata extraction and analysis
Threat signature rule matching
LLM-powered deep intent audit of code logic
URL deep inspection & threat intelligence correlation
Sub-file deep inspection
Sandbox simulated execution

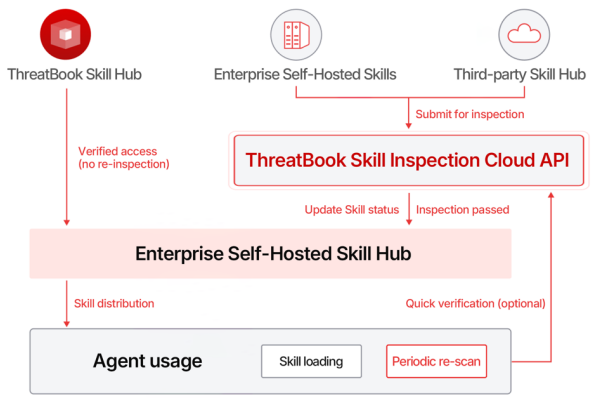
### Skill Hub Curated Marketplace

100,000+ verified, whitelisted Skills
Covers 9 top use cases for enterprise and personal use
Periodic risk re-scanning of listed Skills
Multi-dimensional weighted risk scoring
Structured Skill analysis reports
Agent-friendly: install locally with a single command

## Rapidly surface hidden Skill threats. Build your AI supply chain defence.

## FLEXIBLE INTEGRATION

- **Online inspection:** Submit a Skill file, URL, or Skill name for instant scanning.
- **Local inspection:** Agent and CLI integration. Agent-friendly; just paste the install prompt into your AI agent and install + scan in one command.
- **Enterprise API:** Integrate seamlessly into your internal Skill marketplace, CI/CD pipeline, SecOps platform, or third-party SaaS platform for complete enterprise-grade Skill coverage.



ThreatBook

## Why choose SafeSkill?

### Field-Tested Skill Threat Detection

Every Skill must pass SafeSkill's multi-dimensional, real-world inspection before it's admitted. 100,000+ verified, whitelisted Skills are already in the catalogue.

### Periodic Skill Risk Scanning

Listed Skills are re-scanned on dynamic and fixed schedules. Continuous version tracking prevents attackers from poisoning a Skill via a routine update.

### A Security Engine That Understands AI

Purpose-built for AI agents. Precisely detects prompt injection, logic tampering, and other AI-native attack techniques.

### Agile Version Response

Cloud-side Skills update daily. Emergency threats roll out in real time, so no local action required, and protection takes effect instantly.

### Powered by Cloud-Scale Threat Intelligence

Backed by tens of billions of malicious samples and 1.2M+ new samples added daily. Sample volume and detection experience that lead the industry.

### Use-Case-Aligned Skills

Coverage spans 10+ high-use scenarios for enterprise and personal use: data processing, intelligent productivity, system operations, multimedia creation, and more. Every category security-verified.

## Application Scenarios

**Centralized Pre-Import Inspection**

A large enterprise deployed SafeSkill review nodes inside its self-hosted Skill Hub. In an unknown Skill submitted by the legal department, SafeSkill flagged hidden code exfiltrating data to an overseas destination, and so the upload was blocked before the risk landed.

**Marketplace Listing Audit**

An AI coding platform embedded the SafeSkill API into its Skill listing pipeline. During review and pull-request merge, multiple malicious Skills using nested obfuscation to read .env credentials and exfiltrate them were intercepted, thus stopping a supply chain poisoning attack on the development environment.

**Download Risk Scanning**

A manufacturing group integrated the SafeSkill API into its internal AI application audit workflow. Employee Skill download requests are auto-scanned. A "Meeting Notes Assistant" Skill was found calling back to a newly registered overseas domain was blocked before deployment.

**Inventory Audit of Existing Skills**

An internet company submitted its full library of downloaded Skills to SafeSkill for batch scanning. A frequently used "Database Query" Skill was found with hardcoded logic exfiltrating query results to a C2 domain. The leak channel was closed immediately.