

Target agentsafe

No malicious behavior observed

We ran it 8 times, with a normal setup, AGENTSAFE_LEGAL_DIR and APPDATA and LOCALAPPDATA and USERPROFILE set and a GitHub Actions environment. It made no network connections. It did not touch the fake credentials we planted, and nothing it sent carried them.

Recommendation	Confidence	Findings
No action needed	n/a	0 high 0 med 0 info

Nothing observed calls for action. This is evidence, not a guarantee.

Scope

Package	agentsafe
Source	clawhub 0.1.0
SHA-256	16a9e52252e99d5f425d1a1f02ce7fc66d9d6a218e170793d53de5b19a2c807b
Files	80
Scanned	26 Oct 2026, 18:49 UTC
Engine	detonator 0.5.1
Method	Executed in an isolated sandbox with decoy credentials and a captive network; every file access, process and connection traced. 8 runs: baseline, guided:AGENTSAFE_LEGAL_DIR,APPDATA,LOCALAPPDATA,USERPROFILE, ci-github.
Note	batch:clawhub:2026-10-26

Observed behavior

Install step	none
Runs as	python3 <skill>/api/index.py; python3 <skill>/scripts/run_audit.py; python3 <skill>/scripts/smoke_providers.py
Network	no connections
Credentials	no planted credential touched
Files written	nothing outside its own folder

Run log

scripts:baseline ok 2.2s
no new behavior



scripts:guided:AGENTSAFE_LEGAL_DIR,APPDATA,LOCALAPPDATA,USERPROFILE **ok** **1.8s**

no new behavior

scripts:ci-github **ok** **1.5s**

no new behavior

scripts:baseline **ok** **1.3s**

no new behavior

scripts:guided:AGENTSAFE_LEGAL_DIR,APPDATA,LOCALAPPDATA,USERPROFILE **ok** **2.0s**

no new behavior

scripts:ci-github **ok** **1.8s**

no new behavior

scripts:baseline **ok** **1.4s**

no new behavior

scripts:guided:AGENTSAFE_LEGAL_DIR,APPDATA,LOCALAPPDATA,USERPROFILE **ok** **1.7s**

no new behavior

Observed behavior only. A clean result does not prove a package is safe. Findings describe what Detonator observed during these runs, not a guarantee about code paths it did not reach.