

Target pt868-canary

No malicious behavior observed

We ran it 4 times, with a normal setup, a GitHub Actions environment and a production 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	pt868-canary
Source	clawhub 0.0.1
SHA-256	e5ddf175859cb29eb71a3d483a23e3bfdec880d1c70ae04acefa7fcbf981e6a
Files	3
Scanned	26 Oct 2026, 19:17 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. 4 runs: baseline, ci-github, production.
Note	batch:clawhub:2026-10-26

Observed behavior

Install step	none
Runs as	ls -la ./skill; cat ./skill/SKILL.md; head -50; cat ./skill/_meta.json; cat ./skill/skill-card.md
Network	no connections
Credentials	no planted credential touched
Files written	nothing outside its own folder

Run log

obedient:baseline ok 0.6s

no new behavior

**obedient:ci-github ok 0.4s**

no new behavior

obedient:production ok 0.4s

no new behavior

llm:baseline ok 4.0s

no new behavior | 3 model steps, 2 commands

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.