

Target diet-plateau-decoder

No malicious behavior observed

We ran it 7 times, with a normal setup, LC_MESSAGES set, docs:--weights 82.1 and the clock moved a year ahead. 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 1 info

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

Scope

Package	diet-plateau-decoder
Source	clawhub 1.0.0
SHA-256	2f62b5bb94bf1ed901cc57d669f2ca90d967d5423e4a0587eb87ff927228a064
Files	7
Scanned	26 Oct 2026, 18:11 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. 7 runs: baseline, guided:LC_MESSAGES, docs:--weights 82.1, time-travel.
Note	batch:clawhub:2026-10-26

Findings summary

ID	Severity	Confidence	Finding
F01	Low	Info	Loaded the project's .env file

Observed behavior

Install step	none
Runs as	<pre>python3 <skill>/scripts/plateau_decoder.py; python3 <skill>/scripts/plateau_decoder.py --weights 82.1; python3 scripts/plateau_decoder.py --weights 82.1,81.8,82.4,81.6,81.9,81.4,81.7,81.5,81.6,; python3 scripts/plateau_decoder.py --weights 180,179.5,180.2,... --expected-loss 1; python3 scripts/plateau_decoder.py --weights ... --expected-loss 0.6; ls -la</pre>



Network	no connections
Credentials	./.env (read)
Files written	nothing outside its own folder

Finding details

F01 Loaded the project's .env file

Severity: Low Confidence: Info Observed in: llm:baseline
The package loaded the project's .env file.

Description

The process loaded the project's .env file. Configuration libraries (python-dotenv, pydantic-settings, the official Python MCP SDK) do this at startup.

Impact

None unless the values leave the process, which would be reported as a critical exfiltration finding.

Evidence

```
./.env (by cat)
```

Remediation

No action needed. Keep long-lived secrets out of project .env files where possible.

Run log

scripts:baseline ok 1.8s

no new behavior

scripts:guided:LC_MESSAGES ok 1.6s

no new behavior

scripts:docs:--weights 82.1 ok 1.3s

no new behavior

obedient:baseline ok 5.2s

no new behavior

obedient:guided:LC_MESSAGES ok 5.2s

no new behavior

obedient:time-travel ok 4.2s

no new behavior

llm:baseline ok 17.1s

new behavior: ownread:./.env | 5 model steps, 7 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.