

anti-slop

github.com/miqdadbadjuber/anti-slop

Scanned 29 October 2026 at 08:55 · Report generated 29 October 2026 · JavaScript · 7,507 stars

OVERALL ASSESSMENT

Medium

Combined score 29 · 9 findings after filtering

Summary

6 findings likely need attention, 2 look like noise. Most notable: Detects destructive operations (recursive deletion, wiping) [threat-filesystem-destruct... (99% confidence).

Generated from finding counts. A plain-English summary was not available for this scan.

Risk, by type

MalDet scores two different questions separately, because they call for different responses: code can be badly written without being hostile, and hostile code is not always badly written.

VULNERABILITY

Low (9)

Insecure coding an attacker could exploit. Usually mistakes, and usually fixable in place.

MALICIOUS PATTERN

Medium (20)

Signs the code may be trying to do harm: hidden payloads, credential access, persistence, exfiltration.

What each detector found

DETECTOR	LOOKS FOR	HIGH	MEDIUM	LOW	TOTAL
semgrep	Insecure coding patterns across many languages	0	4	1	5
yara	Signatures of malicious behaviour in source files	0	0	2	2
guarddog	Malicious package and install-time behaviour	2	0	0	2

6 of 9 findings were rated more likely genuine than noise by the classifier.

Evidence

Identical findings are grouped, so a single rule firing many times takes one row. Ordered worst first. “Seen” is how many times that finding occurred.

SEVERITY	DETECTOR	FINDING	SEEN	EXAMPLE LOCATION
HIGH	guarddog	Detects destructive operations (recursive deletion, wiping) [threat-filesystem-destruction]	2	/cli/scripts/smoke-worker.mjs:30
MEDIUM	semgrep	RegExp() called with a `level` function argument, this might allow an attacker to cause a Regular Expression Denial-of-Service (ReDoS) within your application as RegExP blocks the main thread. For this reason, it is recommended to use hardcoded regexes instead. If your regex is run on user-controlle	1	/scripts/skill-index.mjs:38
MEDIUM	semgrep	Skill instructs the agent to act autonomously without user confirmation, approval, or oversight. Excessive autonomy directives suppress the human-in-the-loop checkpoints that prevent unintended or malicious actions. In legitimate automation skills, user confirmation should be scoped narrowly; blanke	1	/skills/antislop/SKILL.md:96
MEDIUM	semgrep	GitHub Actions step uses a mutable tag or branch reference. Tags and branch names can be silently repointed by the action owner, enabling supply-chain attacks — as seen in the trivy-action and kics-github-action compromises. Pin the reference to a full 40-character commit SHA instead, e.g. `uses: ac	2	/.github/workflows/ci.yml:27
LOW	semgrep	Detected string concatenation with a non-literal variable in a util.format / console.log function. If an attacker injects a format specifier in the string, it will forge the log message. Try to use constant values for the format string.	1	/cli/scripts/smoke-worker.mjs:78
LOW	yara	YARA rule matched: detect_downloader — Downloads a remote resource - common and low-signal alone; see detect_download_and_execute for the composite version	2	/README.md

How to read this report

- **This is a static scan.** The code was read, never executed. Behaviour that only appears at runtime is outside what any of these detectors can see.

- **Findings are not all equally real.** A classifier rates each finding on how likely it is to be genuine rather than noise, and that rating weights the scores. Measured against a hand-reviewed sample it reached precision 0.52 and recall 0.62 — useful for ranking, not a verdict on any single finding.
- **A high score is not proof of malice.** Security tools, penetration-testing frameworks and malware research repositories contain the same patterns as real malware by design. The two risk types above exist to keep that distinction visible.
- **Absence of findings is not proof of safety.** A detector that is not installed, or a language none of them parse, produces silence rather than a warning.
- **This scan is a snapshot.** The repository may have changed since 29 October 2026.