

Source dossier

Green was a quarter of the check — agents close dependency tickets the rescan still fails

Edition 2026-10-18 · narrative: VulnBump scores 307 real dependency upgrades on a four-part criterion checked in a fresh container: dependencies resolve, a rescan confirms the targeted vulnerabilities are fixed, the project builds, and every previously passing test still passes. A non-agentic baseline passes 46.9%, Claude Code unaided 59.6%, the same agent handed a scanner 77.5%, and a purpose-built remediation agent 90.2%. The failure modes diverge: the general-purpose agent mostly leaves the vulnerability in place, the specialised one mostly breaks the build. The regulatory reflex points the wrong way: the Cyber Resilience Act's substantive vulnerability-handling obligations do not bind until 11 December 2027, while Article 14's duty to report actively exploited vulnerabilities has been live since 11 September 2026 — so the only CRA clock running today rewards knowing, not fixing, and an agent that marks a vulnerability resolved when it is not corrupts the input to the duty already owed. Caveat carried in the edition: the benchmark draws its task distribution from one vendor's production data and the top-scoring agent is that vendor's own; author affiliations could not be confirmed from the work's own front matter (arXiv HTML returned HTTP 429), so no institution is attributed. · compiled 2026-10-18

Claims index

Claim	Statement	Status	Sources
CLAIM-2026-10-18-01	VulnBump (arXiv:2610.17374v1, submitted 12 October 2026) comprises 307 dependency-upgrade tasks built from 2,066 fixable vulnerabilities across 43 repositories and six package managers, with the task distribution matched to one vendor's production vulnerability distributions.	verified	arxiv2026vulnbump
CLAIM-2026-10-18-02	In VulnBump an upgrade counts as passing only if, in a fresh container, dependencies resolve, a rescan shows the targeted vulnerabilities are fixed, the project builds, and all previously passing tests still pass.	verified	arxiv2026vulnbump
CLAIM-2026-10-18-03	VulnBump reports a non-agentic baseline applying the recommended version passing 46.9% of upgrades, Claude Code unaided 59.6%, Claude Code with a scanner 77.5%, and a purpose-built remediation agent 90.2%.	verified	arxiv2026vulnbump
CLAIM-2026-10-18-04	VulnBump reports that supplying scanner findings raises Claude Code's pass rate by 17.9 percentage points, and that a purpose-built harness still outperforms Claude Code with the same scanner findings by 12.7 percentage points.	verified	arxiv2026vulnbump
CLAIM-2026-10-18-05	VulnBump reports that the harnesses fail differently: the general-purpose agent most often leaves the vulnerability in place, while most failures of the purpose-built agent break the build or the tests.	verified	arxiv2026vulnbump

Claim	Statement	Status	Sources
CLAIM-2026-10-18-06	VulnBump reports that 94% of that vendor's open-source findings have a known fixed version while only about 10% are resolved in practice.	verified	arxiv2026vulnbump
CLAIM-2026-10-15-11	Regulation (EU) 2024/2847, the Cyber Resilience Act, applies its main obligations from 11 December 2027, while its Article 14 reporting duty for actively exploited vulnerabilities has applied since 11 September 2026. Reused from the edition of 15 October 2026.	verified	europeancom2026timeline, eurlex2024regulatio
CLAIM-2026-10-18-07	Author affiliations for arXiv:2610.17374 could not be confirmed from the work's own front matter during this run; the arXiv HTML rendering returned HTTP 429 and no institution is attributed in this edition.	verified	arxiv2026vulnbump

Only claims marked **verified** may carry a post. **single-source** may appear as hedged context; **disputed** must not be published at all.

VulnBump: A Benchmark for Fixing Vulnerable Dependencies in Real Projects While Keeping the Build and Tests Green

PRIMARY

Publisher	arXiv
Published	2026-10-12
Retrieved	2026-10-18
Cite key	arxiv2026vulnbump
URL	https://arxiv.org/abs/2610.17374

Assessment. Primary: the work's own abstract page. All benchmark figures in this edition are taken from it. Author affiliations are not printed on the listing page and the HTML rendering returned HTTP 429 on retrieval, so no institution is attributed. Note of interest: the benchmark's task distribution is drawn from one vendor's production data and the top-scoring system is that vendor's own remediation agent; the methodology (fresh container, rescan, build, previously-passing tests) and the honest non-agentic baseline are nonetheless strong.

Claims drawn from this source. CLAIM-2026-10-18-01, CLAIM-2026-10-18-02, CLAIM-2026-10-18-03, CLAIM-2026-10-18-04, CLAIM-2026-10-18-05, CLAIM-2026-10-18-06, CLAIM-2026-10-18-07

Full text not archived — discovery/headline reference only.

Regulation (EU) 2024-2847 — Cyber Resilience Act

PRIMARY

Publisher	EUR-Lex
Published	2024-11-20
Retrieved	2026-10-18
Cite key	eurlex2024regulatio
URL	https://eur-lex.europa.eu/eli/reg/2024/2847/oj/eng

Assessment. The instrument itself. Cite key stable since the edition of 5 August 2026.

Claims drawn from this source. CLAIM-2026-10-15-11

Full text not archived — discovery/headline reference only.

Timeline and obligations for the Cyber Resilience Act

PRIMARY

Publisher	European Commission
Published	2026-07-31
Retrieved	2026-10-18
Cite key	europeancom2026timeline
URL	https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act

Assessment. The Commission's own timeline. Separates the 11 December 2027 main obligations from the Article 14 reporting duty live since 11 September 2026 — the distinction this edition turns on.

Claims drawn from this source. CLAIM-2026-10-15-11

Full text not archived — discovery/headline reference only.

CRA reporting obligations

PRIMARY

Publisher European Commission

Published 2026-09-11

Retrieved 2026-10-18

Cite key europeancom2026cra

URL <https://digital-strategy.ec.europa.eu/en/policies/cra-reporting>

Assessment. The channel the live Article 14 duty actually runs through. Used for the reporting-path point in the first comment.

Claims drawn from this source. CLAIM-2026-10-15-11

Full text not archived — discovery/headline reference only.