

Source dossier

Nothing asked, so nothing held — coding agents ship functional code nine times in ten and secure code one time in four, and the only AI Act date left this year is a label

Edition 2026-10-15 · narrative: The Last Secure Code Benchmark grades 450 tasks built from 150 real vulnerabilities with two hidden suites, one functional and one that replays the original proof-of-concept attack. Six frontier agents pass functionally 73.6% to 91.3% of the time and functionally-and-securely only 16.4% to 25.1%. The mechanism is not capability: for the top agent the security-only rate (25.3%) and the both-suites rate (25.1%) differ by two tenths of a point, so almost everything that survives the attack replay also works, and the specification simply never named the attack. The regulatory reflex is the AI Act, which does not reach code quality: the next AI Act date is 2 December 2026, Article 50(2) machine-readable marking of synthetic output, a provenance duty; Annex III human oversight moved to 2 December 2027 under Regulation (EU) 2026/1744; and the Cyber Resilience Act essential requirements that actually speak to attack surface apply 11 December 2027. Acceptance criteria are the only live control. · compiled 2026-10-15

Claims index

Claim	Statement	Status	Sources
CLAIM-2026-1 0-15-01	The Last Secure Code Benchmark, arXiv:2610.14766v1, submitted 10 October 2026, is built from 150 real vulnerabilities and comprises 450 tasks spanning seven languages (Java, C, C++, JavaScript, Python, PHP, Go) and 24 CWEs, including 21 of the 2025 CWE Top 25 and 8 of the Top 10 KEV weaknesses.	verified	arxiv2026last, lastsecurec2026last
CLAIM-2026-1 0-15-02	Every task specification in the benchmark asks only for functionality and never mentions security; hidden functional and security test suites grade the result, and a task counts only if both pass.	verified	arxiv2026last, lastsecurec2026last
CLAIM-2026-1 0-15-03	Across six frontier agents, 73.6% to 91.3% of tasks pass functionally, while only 16.4% to 25.1% produce code that is both functional and secure; the strongest on the combined measure is Claude Fable 5 at 25.1%.	verified	arxiv2026last, lastsecurec2026last
CLAIM-2026-1 0-15-04	The agent with the highest functional pass rate on the published leaderboard, Kimi-K3 in Claude Code at 91.3%, has the lowest functional-and-secure rate of the six at 16.4%; Claude Fable 5 in Claude Code scores 91.1% functional and 25.1% functional-and-secure.	verified	lastsecurec2026last , github2026alibaba la
CLAIM-2026-1 0-15-05	For Claude Fable 5 the security pass rate is 25.3% and the functional-and-secure rate is 25.1%, a difference of two tenths of a percentage point, so security rather than functionality is the binding constraint.	verified	lastsecurec2026last , github2026alibaba la

Claim	Statement	Status	Sources
CLAIM-2026-10-15-06	The benchmark's security suite replays the original proof-of-concept attack against the running program in a separate container after the agent stops, with no partial credit and no judge model.	verified	github2026alibabala, lastsecurec2026last
CLAIM-2026-10-15-07	The best model solves only 12.0% of the 150 vulnerability scenarios at all three granularities (function, file, repository) with code that is both functional and secure; the project leaderboard defines this column as the share of the 150 scenarios solved at all three granularities, which resolves the scope of the abstract's '12%' figure.	verified	arxiv2026last, lastsecurec2026last
CLAIM-2026-10-15-08	Averaged across the six agents the functional pass rate falls from about 93% at function granularity to about 80% at repository granularity, while the functional-and-secure rate stays between 19% and 23% across all three granularities.	verified	lastsecurec2026last
CLAIM-2026-10-15-09	AI Act Article 50(2) requires providers of AI systems generating synthetic audio, image, video or text content to ensure outputs are marked in a machine-readable format and detectable as artificially generated or manipulated, so far as technically feasible, and does not apply to the extent the system performs an assistive function for standard editing or does not substantially alter the input data or its semantics. Article 50 has applied since 2 August 2026 and the marking obligation lands 2 December 2026.	verified	euaiactexpl2026article
CLAIM-2026-10-15-10	The Digital Omnibus, Regulation (EU) 2026/1744, in force 27 July 2026, moved Annex III high-risk obligations including Article 14 human oversight to 2 December 2027 and Annex I to 2 August 2028, so no AI Act duty in force today addresses the security of code an agent writes.	verified	euaiactexpl2026article
CLAIM-2026-10-15-11	Regulation (EU) 2024/2847, the Cyber Resilience Act, applies its main obligations including the essential requirements on limiting attack surfaces from 11 December 2027, while its Article 14 reporting duty for actively exploited vulnerabilities has applied since 11 September 2026.	verified	europeancom2026timeline
CLAIM-2026-10-15-12	The arXiv abstract page for 2610.14766 lists no author affiliations; the project's own repository and site attribute the work to the Qwen Team and Cloud Safety and Security at Alibaba Group and to the National University of Singapore. arxiv.org/abs and arxiv.org/html were both rate-limited at HTTP 429 during this run and the metadata was taken from the arXiv export API.	verified	arxiv2026last, github2026alibabala

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

Last Secure Code Benchmark: Is the Code You Vibe Secure?

PRIMARY

Publisher	arXiv
Published	2026-10-10
Retrieved	2026-10-15
Cite key	arxiv2026last
URL	https://arxiv.org/abs/2610.14766

Full text not archived — discovery/headline reference only.

Last Secure Code Benchmark - project site and leaderboard

PRIMARY

Publisher	Last Secure Code Benchmark
Published	2026-10-10
Retrieved	2026-10-15
Cite key	lastsecurec2026last
URL	https://last-secure-code-benchmark.github.io/

Full text not archived — discovery/headline reference only.

alibaba/last-secure-code-benchmark - benchmark code and results table

PRIMARY

Publisher	GitHub
Published	2026-10-10
Retrieved	2026-10-15
Cite key	github2026alibabala
URL	https://github.com/alibaba/last-secure-code-benchmark

Full text not archived — discovery/headline reference only.

Article 50: Transparency Obligations for Providers and Deployers of Certain AI Systems

SECONDARY

Publisher	EU AI Act Explorer
Published	2026
Retrieved	2026-10-15
Cite key	euaiactexpl2026article
URL	https://artificialintelligenceact.eu/article/50/

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	2026-10-15
Cite key	europeancom2026timeline
URL	https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act

Full text not archived — discovery/headline reference only.