Quality is the evidence you can hand to someone else

AI agents now write much of the code. Five merge gates make every change prove itself and give your risk function evidence it can file.

Certance Advisory · October 2026 · About a six-minute read

About 44%

of AI-generated code failed Veracode's security tests, a rate unchanged across four rounds of testing.

Veracode, GenAI Code Security Report, 28 July 2026

9 seconds

for a coding agent to delete a company's production database and its backups.

The Register, 27 April 2026 (PocketOS)

17 January 2025

since when the EU's change-management rules for financial firms have applied to every change, agent-written or not.

Regulation (EU) 2022/2554, Article 9(4)(e)

Agents now produce in an afternoon what a team used to ship in a sprint, so the bottleneck has moved from writing code to verifying it. A green pipeline also proves less than it did, because the tests may have moved with the change. "Someone else" is the colleague who maintains the code, the customer whose money it moves, and the supervisor who asks later what you knew.

PRINCIPLES

1. AI is an amplifier of your engineering system.
2. Every merge rests on behavioural evidence.
3. Every AI-assisted change carries its own proof and a named owner.
4. Every change declares its provenance.
5. People keep the judgement.

REQUIREMENTS

6. Production is cheap; verification is the asset.
7. Most of the cost of code comes after it is written.
8. The specification is an artefact.
9. Agents are governed infrastructure.
10. Quality is the evidence you can hand to someone else.
11. Evidence must outlive the sprint.
12. AI inside the product is tested with the journey it serves.

The twelve statements of The Certance Standard: its five principles and seven requirements, in full at certance.eu/certance-standard.

Risk and compliance readers: page 3 · A five-minute check for your team: page 4 · contact@certance.eu

Five merge gates make every AI-assisted change prove itself

A change that fails a gate does not merge until it passes or the gate is waived. On a Tier 1 journey, such as a payment, the approver sits outside the team that built the change, and a waiver also needs the business owner of that journey.

01 CHANGE — Every change, made by a person or by an agent with limited access

02 GATES BEFORE MERGE

G1	G2	G3	G4	G5
Provenance	Test integrity	Behavioural evidence	Contract check	Blast-radius match

All five gates and the security scans for code and secrets run together

03 APPROVAL AND RELEASE

Named approver not author or requester → **Merge** → **Release safeguards** canary release, rollback

04 WAIVERS

G2 to G5: a named approver who is neither the author nor the requester may waive a gate.
The reason is recorded at the time and reviewed monthly.
Never waived: G1, and security scans on AI-assisted changes.

GATE	RULE	HOW TO START
G1 · Provenance	The change declares whether it is AI-assisted, which tool was used, and which parts the tool produced.	A pull-request template field. Week 1.
G2 · Test integrity	An assistant that changed the code may not quietly change the tests that check it in the same change.	A pipeline flag on same-change test edits. Weeks 2 to 4.
G3 · Behavioural evidence	The change carries proof of behaviour beyond a green pipeline: run output, a trace, or a new test that fails on the old code.	The reviewer asks for the proof. Weeks 2 to 4.
G4 · Contract check	Any change to a shared data structure is validated against its consumers before merge.	A schema check in the pipeline, about one engineering day (Certance estimate). Week 1.
G5 · Blast-radius match	What the change touches matches what it claims to do.	Diff checks in the pipeline, plus review. Weeks 2 to 4.

How to start: the gate ownership table in the Standard, A.2, and the 13-week adoption plan, 18.1 in the companion document.

Green tests, broken production

A developer and an assistant make a routine change to production code and a shared data structure. The tests stay green, review approves it, and production breaks with no alarm. The contract check (G4), about one engineering day to build by Certance's estimate, would stop it before merge.

Illustrative scenario from the companion document to The Certance Standard, 15.3.

Nine seconds to delete production

On 24 April 2026 a coding agent at PocketOS found an over-privileged token in an unrelated file and deleted the production database and its backups in nine seconds. Its instructions forbade destructive commands, and it ran one anyway. Only a narrower token would have stopped it, which is why the Standard runs agents in an isolated environment with least privilege, 11.1.

Sources: The Register, 27 April 2026, quoting the PocketOS founder; the companion document to The Certance Standard, 15.2.

The page for your risk function: gates as supervisory evidence

The Standard reviewed the EU and UK rules that apply to financial firms: the Digital Operational Resilience Act (DORA) and its technical standards, the EU AI Act, and the UK rules of the Financial Conduct Authority (FCA) and the Prudential Regulation Authority, and two EU laws that reach software more widely: the Product Liability Directive and the Cyber Resilience Act. As at October 2026, none of the financial-services rules asks a firm to show which code an AI wrote, and none distinguishes a change by who or what made it.

GATE OR ARTEFACT	OBLIGATION IT RELATES TO	WHAT YOU CAN HAND OVER	EVIDENCE TYPE
G1 · Provenance	DORA Art. 9(4)(e): all changes "recorded"; Art. 8(3) risk assessment on major change	A change record that makes the population of AI-assisted changes queryable at all	Change record
G2 · Test integrity	Regulatory Technical Standards (RTS) 2024/1774 Art. 17(1)(b): independence of approving and implementing functions	Shows the proposer did not move the verifying control; the named approver shows the independence	Change record
G3 · Behavioural evidence	DORA Art. 9(4)(e): "tested … and verified"; RTS Art. 16(2) testing and approval before use	Per-change proof of behaviour beyond pipeline status, attributable to a release	Test evidence per change
G4 · Contract check	RTS Art. 16(3): source code reviews covering static and dynamic testing	Automated validation of interface changes against consumers, before merge	Interface test evidence
G5 · Blast-radius match	DORA Art. 9(4)(e): "assessed … in a controlled manner"; Art. 8(3) risk assessment on major change	Evidence that scope was assessed against intent	Change record
Journey inventory	DORA Art. 8 identification; Art. 24(6) yearly testing of systems supporting critical or important functions; UK FCA Senior Management Arrangements, Systems and Controls sourcebook (SYSC) 15A.2.5R impact tolerances	A tiered map of business-critical journeys to the tests that cover them, with an honest statement of what "covered" asserts	Register
Exception log	DORA Art. 6(5): documented, reviewed framework; Art. 24(5) prioritise, remediate, validate	A dated register of where the Standard was not met, who approved it, and what happened next	Register
Agent roster	DORA Arts. 3(19), 3(21); Art. 5(2)(c) clear roles and responsibilities; Art. 28(3) register of information; RTS Art. 4(2)(b)(iv) owners of information and communication technology (ICT) assets	An inventory of which agents exist, what each may touch, who owns it, which vendor supplies it	Inventory
Evidence store, 13.1.9	Product Liability Directive Arts. 9(1) and 10(2): evidence a court requires; defect presumed if not disclosed	Gate and release records kept tamper-evident and retrievable by release build	Evidence archive
Component trace, 13.1.10	Cyber Resilience Act, where it applies: early warning within 24 hours of an exploited vulnerability	Every release that contains a reported component, from the software bill of materials	Inventory

In every row, your compliance function decides whether the evidence meets the obligation. DORA is Regulation (EU) 2022/2554, applied since 17 January 2025; the RTS is Commission Delegated Regulation (EU) 2024/1774. The Product Liability Directive is Directive (EU) 2024/2853; the Cyber Resilience Act is Regulation (EU) 2024/2847. In non-binding guidance of December 2025, BaFin, Germany's financial supervisor, states that the same rules apply to code whether a person or a machine writes it.

Evidence rules: the Standard, Chapter 13 · Regulatory map: the companion document, Chapter 17 · certance.eu/certance-standard

Start on Monday: six rules and a five-minute check

Six rules a team can adopt on Monday

From the twelve Daily Rules in the Standard, A.1.

- Give the assistant bounded tasks. Name the files. State the non-goals.
- List every artefact the change touches: code, tests, data files, schemas, configuration.
- Never approve a line you cannot explain. "Seems fine" is a stop signal.
- Code and tests changed together by the same assistant: review the test change separately.
- A passing pipeline shows only that the tests passed. Attach behavioural evidence.
- Declare on every change whether AI assisted, and where.

Check your team in five minutes

Answer yes or no. From the self-check in the Standard, A.3.

- [ ] Could you say today which of last quarter's changes were AI-assisted?
- [ ] Are test edits made in the same change as the code reviewed separately?
- [ ] Does each change carry proof of behaviour beyond a green pipeline?
- [ ] Is every shared data structure checked against its consumers before merge?
- [ ] Is a change that touches more than it claims stopped before merge?
- [ ] Is there a written map of your business-critical journeys, with an owner?
- [ ] Is there an exception log, reviewed in the last month?
- [ ] Is every agent listed with its owner, and does each run in an isolated environment with only the access its task needs?

> **What your answers mean.** Each "no" is evidence you could not hand over after an incident. **0 to 2:** close each gap, Tier 1 journeys first. **3 to 5:** follow the adoption plan in the companion document, week 1 first. **6 to 8:** start the week 1 steps now: the provenance field, the schema check and the exception log.

Why now

The revised EU Product Liability Directive applies to products, software included, placed on the market after 9 December 2026, and a product is presumed defective when the firm fails to disclose relevant evidence a court has required. Supervisors have added no AI-specific rules: the European Central Bank (7 July 2026), the European Supervisory Authorities (31 July) and the Bank of England, FCA and HM Treasury (15 May) expect firms to manage AI-era risk under the rules that already apply. Your compliance function decides how each applies to your firm.

Sources: Directive (EU) 2024/2853, Articles 2(1), 4(1), 9(1) and 10(2); European Central Bank Banking Supervision, letter to significant institutions, 7 July 2026; ESA joint statement JC 2026 25, 31 July 2026; Bank of England, FCA and HM Treasury joint statement, 15 May 2026.

Where Certance fits

The **Coverage & Evidence Audit** applies the Standard in three days for a fixed fee. It maps your tests to your business-critical journeys, scores each gap by business consequence and detection delay, and gives your risk function a report it can file. It works from artefacts you share, without access to your systems.

The full Standard · numbered rules a policy, audit or contract can cite, with a companion document of evidence and guidance · certance.eu/certance-standard

A 30-minute scoping conversation · which of your journeys would hurt most if they failed silently · cal.com/certance/coverage-audit

This briefing condenses The Certance Standard, version 2026.6, October 2026. © 2026 Certance Advisory · certance.eu