ALLCODE · SAMPLE SUMMARY

Autonomous Software Development Assessment

Illustrative example — not a real client engagement

At a Glance

Question	Answer
Where the company is today	**Level 2.** AI helps engineers plan and write code. A person reviews and decides on everything it produces.
Where the company wants to be	**Level 4.** An AI agent does defined work inside fixed limits; people approve the result from evidence, not by re-reading every line.
The key insight	AI can now produce a working change faster than a person can verify it's correct. The bottleneck isn't writing code anymore, it's proving the work is right.
What we recommend	A four-week pilot on one service, using 5-10 real backlog items. Continue only if human review effort drops at least 20%, with no slower delivery and no weaker proof of correctness.
What needs to be decided	Which service to pilot, who owns each approval step, and authorization to begin readiness work.
What stays the same	The existing work tracker, CI system, and hosting stay exactly as they are. People still approve every release.

Where the Company Stands Today

We rate authority across seven delivery stages, not engineering quality. A team can have excellent engineering and still be Level 0 for AI, because the rating measures how much decision-making is handed to AI, not how good the pipeline is.

Stage	Level	What we found
Plan	L2	AI drafts initiatives; product owners revise over several rounds
Code	L2	Engineers use AI for research and first drafts, then write and review by hand
Verify	L0	Automated builds and compatibility checks; all judgment calls are manual
Release	L0	A person decides when each release ships; a dedicated team validates first
Deploy	L0	Deployments are triggered and run by people
Monitor	Unverified	No run records available to rate this stage
Govern	Unverified	No run records available to rate this stage

> **Overall: Level 2.**

One Bug, Two Ways

Today (Level 2): A customer reports a shipping address isn't validating correctly for one country's postal code format. An engineer asks an AI assistant for a fix, gets a working change the same day. Automated checks pass, but none of them test the specific postal code pattern the customer hit. A reviewer reads the diff and it looks right. It ships in the next release. If nobody happens to test that exact format by hand, the bug reaches other customers in the same country.

At Level 4: A small team of role-limited AI agents handles the ticket. A research agent investigates and drafts the acceptance rule, including edge cases nobody had flagged, before any code is touched. A coding agent makes the fix. A separate testing agent, without seeing the fix, writes tests designed to break it, including property tests across hundreds of real-world postal formats. A security agent checks the change touches only what it's allowed to. Then an independent checker, built from business examples the agents have never seen, runs the final test. One hidden case fails: a format the original spec never mentioned. A person decides the correct behavior, the spec updates, and the next attempt passes. A reviewer approves from a single evidence packet, in minutes, not by reading the code.

> **What changed:** the work gets challenged from multiple independent angles before a person has to trust it by eye, and people still own the two decisions that actually matter: what "correct" means, and whether the evidence is good enough to ship.

The Four-Week Pilot

Week	What happens
0 (readiness)	Baseline measured, service and backlog items chosen, access limits agreed
1	First items run; measurements confirmed
2-3	Remaining items run; most important missing checks added
4	Evidence reviewed; company chooses: expand, invest in verification only, or stop

The deciding measure: human effort per accepted change must fall at least 20%, without slowing delivery or weakening proof the work is correct.

Why AllCode

AllCode holds AWS's Agentic Competency, is an Anthropic partner (with a Professional Solution Architect on staff), and an OpenAI partner, so we design and run this assessment on whichever model and platform stack actually fits your environment.

Want a real assessment for your own team? Visit allcode.com/autonomous-software-development-assessment to schedule yours.

© 2026 AllCode. This document is an illustrative example for prospective clients and does not describe an actual engagement.