

Executive Brief — AI-Augmented Software Delivery

Autonomous Loops, a Strong Harness, and Spec-Driven Development

A Hamilton Road Agility Technical White Paper · John Crider · v1.0 · July 2026 *One-page brief. Full paper, evidence, and citations: [ai-augmented-delivery-white-paper.md](#)*

Bottom line. AI amplifies individual engineering output on well-scoped work — and measurably *slows* experienced engineers on dense, familiar code. Whether that amplification converts into delivery velocity, quality, and ROI depends entirely on the operating model around the model: a governance **harness**, **spec-driven development (SDD)**, and an **evolved human team** — and the claim is *conditional*: the heavy architecture pays **above a threshold** (where work volume, parallelism, and spec density make human review the binding constraint); below it, a lightweight team with AI assistance captures much of the benefit for far less. The real question is *which side of that line you’re on*.

The problem — the amplification trap. Adoption is near-universal (88% of organizations) yet value is not: an MIT study finds ~95% of enterprises get *zero return* on generative AI, and only 39% report enterprise-EBIT impact. Teams speed up code *generation* without expanding the real constraint — review, testing, integration — so DORA finds AI *raises delivery instability* even where it lifts individual output.

The architecture — three pillars.

- **Autonomous loops** — planner / spec / builder / merge agents advance work through a shared, single-writer state board, with humans off the per-step path.
- **A strong harness** — validated boundaries, mechanical policy gates, adversarial review, escalation-to-human, and cost/audit accounting: what makes autonomy safe (and the security-control surface).
- **Spec-driven development** — a validated spec is the durable control surface; code is a regenerable rendering the harness verifies against.

The team of five. One product owner owns outcomes and validation (the irreducible role); four engineers shift from typing to spec authoring, escalation handling, the hard 20%, and harness stewardship. This is the *squad-level unit* — it replicates at the squad boundary for larger organizations.

What it delivers vs. what it costs.

Delivers	Costs
Throughput unbounded by headcount	A real J-curve dip before the lift
Code-quality + documentation gains, captured systematically	Permanent stewardship load + inference spend
Durable, living specifications	New failure modes: prompt injection, compounding errors, lock-in
Governance & audit by construction	Build-trap risk at machine speed
Human time redirected to judgment	High failure rate if adopted as a tool swap, not an operating-model change

Earned, not installed. Like electrification, the technology alone yields a dip, not a lift, until the organization and its people are redesigned around it. You cannot drop a high performer from a human-only team into an AI-human team and expect competence — skills, culture, and governance must co-evolve over quarters.

Decision guidance.

- **Buy vs. build:** most 50-person firms should *buy* the loop-and-SDD layer and spend scarce capacity on harness fit and people; build only where delivery is itself a competitive asset.
- **Find your threshold first:** if review-and-integration time dominates cycle time and you carry more well-specified work in parallel than you can review, you’re above the line — **adopt** (with spec/test discipline, a real PO, and patience for a quarters-long dip). Otherwise stay lightweight and **skip** the heavy path.
- **Advance by evidence, not calendar; expect the dip; treat regression as a healthy outcome.**

Next steps.

1. **Find your threshold** (full paper, §11) — below it, stay lightweight; that’s the honest answer for many teams.
2. Baseline DORA’s four keys *before* automating; run the readiness self-assessment (§10).
3. Build the harness first, autonomy second (aligned to NIST AI RMF).
4. Invest in people — reskilling toward spec, verification, and agent supervision.
5. Elevate the retrospective to a double-loop whole-system review; advance by evidence, expect the dip.

Independent — Hamilton Road Agility, not affiliated with Anthropic; Claude Code is an illustrative platform. Figures verified against primary sources; AI used in creation under human direction. Full paper carries citations and disclosures.