A Best-of-Breed Chain for AI-Assisted Software Development

Why the hard part of vibe coding is verification, not generation, and how Claude Code, GitHub, Vercel or self-hosted Coolify, Next.js and Postgres close the loop we build the Nashua 360 Enterprise Platform on.

Abstract

A capable coding agent changes the economics of software development: writing code becomes fast and cost-effective, while the work around it, deciding what to build and making sure the result is correct, does not. The constraint moves from generation to verification. This paper argues that the value of an AI-assisted setup lies not in the model that writes the code but in the chain of tools around it that lets a team verify, at speed, what the agent produces and catch its mistakes before they reach users. It is a technical, opinionated account of the chain we at Nashua build and run our own product, the Nashua 360 Enterprise Platform, on: Claude Code as the agent, reachable even from a phone; GitHub as the spine, with a protected branch and required checks; a deployment platform, Vercel if you want it managed or self-hosted Coolify if you want to own it, giving every pull request its own running preview; Next.js as a typed, conventional framework whose type checker is the first gate; and PostgreSQL as durable state changed through reviewed, reversible migrations. We go into what practitioners care about, how the agent is kept in context and honest, what runs in CI, what a preview deployment really is, and how to change a live database safely, and we are candid about the costs. For AI-assisted work, best-of-breed means conventional. This is practitioner-facing advice, not a benchmark.

Keywords: AI-assisted development; vibe coding; coding agents; Claude Code; GitHub; Vercel; Coolify; self-hosted PaaS; digital sovereignty; Next.js; PostgreSQL; continuous delivery; preview deployments; code review; database migrations; Prisma ORM; biomimicry; embedded context; developer tooling; Nashua 360 Enterprise Platform

1. Introduction

For most of the history of software, the scarce resource was the writing. A developer's craft was measured in the ability to turn an intention into correct code, and the hours went into typing, looking things up, wiring libraries together, fixing the small mistakes that compile errors and failing tests revealed. A capable coding agent removes most of that. Give it a clear instruction and access to a codebase and it will produce a working change in minutes, in idioms it has seen a million times, with the boilerplate already correct. The colloquial name for working this way, popularised in early 2025, is *vibe coding*: describing what you want and letting the model produce it, steering by results rather than by writing each line. It is genuinely a step change, and it is easy to mistake it for the whole story.

It is not the whole story, because making code appear was never the hard part of software, only the laborious part. The hard parts are deciding what to build, and knowing that what was built is actually right, secure and fit to run. These do not get cheaper when generation does; if anything they get harder, because there is now far more code to understand, produced faster than a person wrote it, by an author who is confident even when wrong. The bottleneck of software development moves. It used to sit at generation, how fast a team could write correct code. With an agent it sits at *verification*: how fast a team can be sure the code is correct and let it through.

This paper is about the consequence of that shift for tooling, and it is deliberately concrete. If verification is the constraint, then the most valuable thing an AI-assisted setup can do is not generate faster, it is to help a team verify faster and more safely, catching the agent's mistakes cheaply and early. We describe the specific chain we at Nashua use to build and operate our own product, the Nashua 360 Enterprise Platform, Claude Code as the agent, GitHub as the version-controlled spine and review gate, Vercel for a running preview of every change, Next.js as the application framework, and Postgres as the database, and we go into how each piece actually works, because the detail is where the value is. Each link earns its place by tightening a loop or adding a gate, not by being individually clever. One principle runs through how we build it, borrowed from biology, *biomimicry*: the architecture and design of each module live embedded in the module, like DNA in a cell, so the agent and the people who administer it read the design where the work is.

The contribution is an argument and a worked recommendation, not a benchmark. Section 2 sets out what AI changed and what it did not, and locates the bottleneck. Section 3 presents the chain as a loop (Figure 1), makes the case that for agent-assisted work best-of-breed means conventional, and explains why independent gates compound. Sections 4 to 9 take the five pieces in turn and go into their internals, including the runtime architecture a deployment actually has (Figure 2) and the part that surprises people, that a change can now be started and reviewed from a phone. Sections 10 to 12 treat the human gate and its failure modes, the real costs, and how the chain runs in practice. The question throughout is: **once an agent can write the code, what should a team put around it so that writing fast does not mean shipping wrong?**

2. What AI changed, and what it did not

2.1. The shift

The change an agent brings is best stated precisely. It is not that the machine now understands the problem; it is that the cost of producing a plausible, often correct, implementation of a well-specified change has collapsed. A developer who would have spent an afternoon on a feature can have a first version in the time it takes to describe it, and can throw several versions away cheaply. The mode of work shifts from authoring to directing: stating intent, reading results, correcting course. This is real leverage, and a team that refuses it will be outpaced.

2.2. No silver bullet

The change is also bounded, and the bound is old. Brooks (1987) drew the distinction that still governs here: the *accidental* complexity of software, the labour imposed by our tools and languages rather than by the problem, can be attacked and reduced, but the *essential* complexity, understanding the problem and specifying a correct solution, cannot be automated away, because it is the problem. A coding agent is the most powerful attack on accidental complexity yet built: it absorbs the syntax, the boilerplate, the API lookups, the idioms. It does nothing about the essence. Knowing what the business needs, deciding the right design, and establishing that a given change actually meets the need without breaking something else, remain human work, and they are the work that was always hard.

2.3. Verification becomes the constraint

Put the two together and the conclusion follows. Generation, the accidental part, is now fast and cheap. Verification, which is where the essence lives, that this change is correct, that it does what was wanted, that it has not broken anything, that it is safe to expose, is not. In any pipeline the throughput is set by the slowest stage, and when generation accelerates past verification, verification becomes the binding constraint. A team that can generate ten changes a day but verify two will ship two, and will accumulate a backlog of unverified code that is a liability, not an asset. Worse, an agent produces confident, superficially correct work, so the cost of *under*-verifying, of merging what looks right and is not, rises exactly as the temptation to do so rises. The rest of this paper is about building a chain whose job is to raise the verification ceiling and to make under-verification hard.

3. The chain, and why conventional wins

3.1. The loop

A good AI-assisted setup is a loop, not a tool. Figure 1 draws the one we use. A developer states an intent. Claude Code, the agent, turns it into a change on a branch, not on the live code. The change becomes a GitHub pull request, where two automated gates run without anyone's attention: the type checker and the continuous-integration suite. A deployment platform, Vercel or self-hosted Coolify, builds the branch into a real, running preview deployment, a full copy of the application with the change live, reachable at its own URL. A human reviews the change, reading the diff and clicking through the preview, on a desktop or, increasingly, on a phone. Only then is it merged to the main branch, which the platform deploys to production, backed by Postgres. The outcome feeds the next intent. The point of the diagram is that the agent occupies one box; every other box is a loop closing or a gate catching, and that is where the safety and the speed come from.

3.2. Best-of-breed means conventional

The usual way to choose tools is to rank them by power and pick the most capable. For an agent-assisted chain the criterion is different, and it is the single most useful idea in this paper. An agent writes the tools it has seen most, best. Its competence on a framework or a database is a function of how much correct code using that tool was in its training, and that quantity tracks the tool's popularity, its documentation, and the stability of its conventions. A mainstream, well-documented, convention-heavy tool is therefore not just a safe human choice; it is the choice on which the agent is most capable and least likely to invent an API that does not exist. The effect compounds: popular tools get more agent-generated code, which works, which reinforces their popularity. For AI-assisted development, best-of-breed and most-conventional have converged, and every tool in Table 1 was chosen with that in mind. This is also a warning: the exotic framework a human expert could wield brilliantly is often exactly where an agent is weakest, because it has seen the least of it.

3.3. Why independent gates compound

The chain does not trust any one check. It verifies a change through several gates that look at genuinely different things: the type checker checks shapes, the continuous-integration tests check behaviour, the preview deployment shows the running result, and the human reviewer judges intent. Because the gates are largely independent, their misses multiply rather than add: a defect reaches production only if every gate fails to catch it, so four modest gates that each catch most of what they see let through only a small fraction of the product of what they miss. This is the defence-in-depth principle long known in safety engineering as the layering of imperfect barriers (Reason, 2000), and it has a sharp practical corollary. Removing or weakening one gate, skipping the preview, turning off a strict type, approving without reading, does not add a little risk; it multiplies the escape rate by the reciprocal of what that gate used to catch. The cheapest way to stay safe is to keep several different, imperfect gates honest, which is exactly what the chain is arranged to do.

4. The agent: how Claude Code fits

Claude Code is the agent: a tool that reads the codebase, writes and edits files, runs commands, executes the tests, and opens pull requests, working from an instruction and correcting itself against the results it sees. How it is run matters more than the model's raw ability, and three things about it are worth a practitioner's attention.

The first is *context*. An agent is only as good as what it knows about your codebase, and the single highest-leverage artefact is a checked-in conventions file, in our case a project `CLAUDE.md`, that tells the agent the house rules: how the project is laid out, how to run the type checker and tests, what to deploy and how, what never to touch. The agent

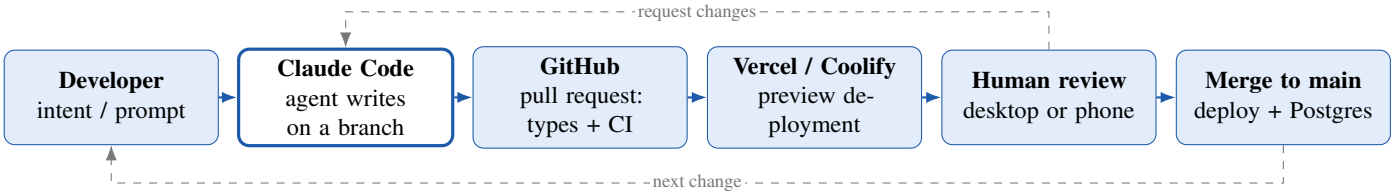

Figure 1. The development loop. The agent (Claude Code) occupies one stage; every other stage either closes a feedback loop or adds a verification gate. A change moves left to right from intent to production (solid); review sends work back to the agent, and production feeds the next intent (dashed). The type checker and continuous-integration suite gate the pull request automatically; the preview deployment lets a human see the change running; the merge to the main branch is the only path to production.

Table 1. The chain. Each link is chosen for the loop it closes or the gate it adds, and for being conventional enough that the agent handles it well. The last column names what the link mainly protects against.

Link	What we use	Loop it closes or gate it adds	What it guards against
Coding agent	Claude Code (CLI, cloud and mobile)	Turns intent into a change on a branch; runs checks and tests itself before handing over	A human doing the accidental work by hand; slow iteration
Source control and review	GitHub (branches, pull requests, Actions)	Every change is a reviewable diff; nothing reaches the main branch unreviewed; CI runs on each	Unreviewed change reaching production; lost history; no audit trail
Preview and deployment	Vercel, or self-hosted Coolify	An isolated, running preview of every pull request at its own URL; one-click deploy and rollback; managed or sovereign	Reviewing code you cannot see run; risky releases
Web framework	Next.js (TypeScript)	Types as the first automatic gate; conventional, file-based structure the agent navigates well	Whole classes of defect the type checker rejects; the agent getting lost
Database	PostgreSQL (managed)	Durable, transactional state with reviewed, versioned migrations	Data loss; silent schema drift; unrepeatable environments
Human review gate	A person, at the pull request	The judgement no gate automates: is this what we wanted, and is it safe?	Confidently wrong code that passed every automatic check

reads it on every run, so a correction written down once stops recurring. Around it sits a wider surface, reusable *skills* for repeated tasks, *hooks* that run on tool calls to enforce a rule mechanically, and *MCP* servers that give the agent typed access to outside systems, GitHub among them, so it can open and update pull requests as first-class actions rather than by screen-scraping. Context engineering, deciding what the agent should know and be able to do, is the real work of adopting one, and it is ordinary software hygiene: write the conventions down, make the checks runnable, keep the surface small.

We scale that context by a principle borrowed from biology. A single root `CLAUDE.md` states the house rules, but it cannot carry the architecture and design of a large product without becoming a document no one, agent or human, reads to the end, and the knowledge a change needs is specific to the part being changed. So rather than keep design in one central place, we embed it where the work happens, taking the lead from *biomimicry*: just as every cell carries the whole organism's DNA rather than consulting a central organ, every module of the Nashua 360 Enterprise Platform carries its own architecture and design as an in-app knowledge base alongside its code. That knowledge is dual: checked in as files the build registers and the agent reads, and surfaced inside the running product as a knowledge base in the platform's system administration module, for the people who operate it and not only the agent that changes it. An agent working in a module reads that module's design intent locally, as context that travels with the code and is versioned, reviewed and corrected with it. The right context is then always next to the change: the agent is told how this part is meant to fit the whole at the moment of editing it, which is where a wrong assumption is cheapest to catch, and design stops being tribal knowledge that erodes and becomes a checked-in artefact the next change inherits.

The second is that the agent runs *in the loop* rather than in isolation. It is expected to run the type checker and the tests itself, read what they report, and fix it, before any human looks, so that the automatic gates of the chain are already green by the time a pull request is opened. For larger tasks it can fan work out to *subagents* that each carry their own context and report back, which keeps the main thread focused and the context clean. The discipline is to let the fast, cheap feedback, types and tests, happen inside the agent's own loop, and to reserve the slow, expensive feedback, the human, for the end.

The third is *where* it runs. Beyond the terminal, Claude Code runs as a cloud session reachable from a mobile app, which changes the shape of the working day. A change can be started from a phone on the way somewhere, left to

run in the cloud, and its pull request and preview reviewed and approved from the same phone later. We are deliberate about this not because mobile work is a novelty but because it shortens the loop: the gap between noticing something and starting the change, and between a change being ready and someone approving it, collapses from hours to minutes, which in a verification-bound world is exactly the gap worth closing.

5. The spine: GitHub

GitHub is the part of the chain that makes everything else safe, and it is the least glamorous. It holds the single source of truth, the repository, and it enforces the rule on which the whole argument rests: the main branch, the one that ships, is *protected*, and can be reached only through a pull request that a human has approved and whose required status checks have passed. The agent never edits production code directly; it proposes, on a branch, and the proposal is a diff a human can read and a machine can test.

The required checks are where the automatic gates live. On every pull request, GitHub Actions runs the suite we insist on before anything merges: the type checker, the linter, the test suite, and a production build, each a required status that blocks the merge button until it is green. Because these are enforced by the platform rather than by anyone's discipline, they cannot be skipped in a hurry, and because the agent knows they will run, it satisfies them first. The pull request is the *unit* of change: small, self-contained, individually reviewable and individually revertible, which is what keeps review cheap and history legible. And because every change is a commit with an author and a message, the history is an audit trail and the undo is total, anything the agent did can be seen, attributed and reverted. A coding agent is a fast, powerful and fallible actor, and GitHub is the structure that makes it safe to let one loose, by ensuring nothing it does is unreviewed, untested or irreversible.

6. The eyes: preview deployments

Reading a diff tells you what changed; it does not tell you what the change does. A *preview deployment* closes that gap by building every pull request into a complete, isolated, running instance of the application with the change applied, at its own URL, created automatically and updated on every push. This is the gate that catches what types and tests do not, the layout that broke, the flow that got one step longer, the thing that is technically correct and wrong in use, and it is most of why the chain can move fast without moving blind. A reviewer, and the agent, click through the actual result rather than imagining it from the code.

There are two good ways to get that gate, and the choice between them is the one real fork in the chain. The *managed* way is Vercel: connect the repository and every pull request becomes a preview and every merge a production deploy, with the edge network, the runtime, the TLS certificates and the rollbacks all operated for you. It is the fastest path to the loop and the one we reach for when nothing argues against it. The *sovereign* way is Coolify, an open-source, self-hosted platform that gives the same loop, a preview deployment per pull request through its GitHub App and a one-click production deploy, on a server you own, in a

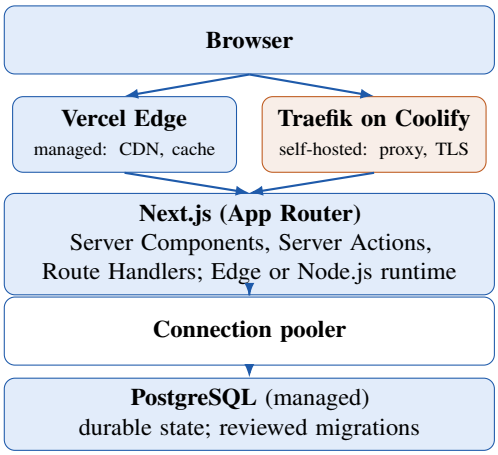

Figure 2. Anatomy of a deployment, the two ways. A request falls through the edge, Vercel's Edge Network on the managed path or Traefik on the self-hosted Coolify path, which serves static assets and cached pages and, on a miss or a dynamic route, invokes the Next.js runtime; server-side code reaches Postgres through a connection pooler. The Next.js application, the pooler and the database are identical either way; only the edge differs. Every pull request gets its own isolated copy of this entire stack at a preview URL, which is why a reviewer can click through the change running rather than imagining it from the diff. Production is the same stack promoted from a reviewed, immutable build.

country and under a law you choose. The stack runs as containers behind a reverse proxy rather than on a managed edge, but the shape the reviewer sees is identical. Which one a team picks is not a question about the chain, which is the same either way, but the managed-or-own decision taken up in Section 11: convenience and zero operations on one side, sovereignty and control on the other. We build and endorse both, and we set out the self-hosted build in full in a companion paper.

What makes the preview trustworthy is that it is a real build of the whole stack, not a mock (Figure 2): the same framework, the same runtime, its own environment variables, and, done well, its own isolated or branched database so that exercising it cannot touch production data. Deployments are *immutable*: each is a distinct, addressable build, so promoting a reviewed preview to production is a pointer move rather than a rebuild, and a bad release is an instant rollback to the previous immutable deployment rather than a frantic hotfix. On merge to the main branch the same machinery that built the preview builds production, so the thing that was reviewed is the thing that ships. In the model of this chain, preview deployments are a cheap, automatic verification gate with a high catch rate for a class of defect no other gate sees.

7. Seamless by git: why the three fit together

The reason Claude Code, GitHub and the deployment platform feel less like separate tools and more like one system is that they share a single interface, and that interface is git. Claude Code's natural output is not a deployment; it is a branch and a pull request. GitHub's job is to store those commits, run the checks and hold the review. The deployment platform watches the repository and turns every branch into a preview and the main branch into production, Vercel

natively, Coolify through its GitHub App. None of them had to be taught about the others: the agent never learns a deployment API, it opens a pull request; the platform never learns about the agent, it watches the repo. The seam between them is the git protocol all three already speak, which is why adding an agent to the chain changed nothing about how deployment worked, and why the whole path from a prompt to production has no handoffs in it.

The consequence is that deployment stops being a step anyone performs and becomes a consequence of the git action they were going to take anyway. A push is a build. A pull request is a preview environment, created and torn down on its own. A merge to the main branch is a production deploy. A revert is a rollback. Nobody runs a release; they merge, and the release happens. There is no build server to babysit, no deploy script for the agent to get wrong, no separate promotion tool, no window in which someone has written the code but not yet shipped it. For a verification-bound team this is exactly the right place to remove friction: the only deliberate, human act left in the path is the review, which is the one act that should be deliberate, and everything mechanical around it is handled by the tools watching git.

This is also why the seamlessness is safe rather than reckless, and both come from the same thing: deployment rides on git. It inherits git's guarantees: every production deploy is tied to a specific reviewed commit, with an author and a message; every preview is tied to a branch; and because deployments are immutable, rolling back is just pointing production at an earlier commit's build. The same wiring that lets a change flow from a prompt typed on a phone to a live site without anyone touching a deploy console also ensures that nothing reaches that live site except through a reviewed, attributable, revertible commit. The integration is a match not because the vendors coordinated it but because all three were built to treat git as the source of truth, and when the agent joined that convention, there was no pipeline to configure or script: the one it needed was already there.

8. The terrain: Next.js

The agent has to navigate a codebase, and the codebase's framework is the terrain. Next.js is a strong choice for two reasons that both come back to verification. The first is that it is *TypeScript*, and types are the first and cheapest gate. Run in strict mode, the type checker rejects a large class of mistake, a wrong shape, a missing field, a value that might be null, at the moment the agent writes it, which is exactly when a correction is cheapest, before the code runs and long before a human looks. We push the same idea to the program's edges, validating external input, request bodies, environment variables, database rows, against a schema so that untyped data from outside becomes typed and checked at the boundary rather than trusted. Types are not bureaucracy here; they are the gate that turns the agent's most common error, a plausible shape that is subtly wrong, into a red squiggle it fixes itself.

The second reason is that Next.js is *conventional*, and its model, though rich, is well-trodden enough that the agent writes it fluently. It is worth being concrete about that model, because it is where a lot of the engineering lives. The App Router renders on the server by default: *Server Components* run on the server and never ship their code to the browser, keeping data access and secrets server-side, while *Client Components*, marked explicitly, carry the interactivity. Mutations go through *Server Actions*, server functions a form or a button can call directly, and *Route Handlers* provide the API surface. Each route chooses how it renders, static at build time, incrementally regenerated on a schedule, or dynamically per request, and where it runs, the lightweight Edge runtime close to the user or the fuller Node.js runtime, with *middleware* running before the request to handle things like authentication and locale. These are a lot of levers, but they are public, documented, stable levers the agent has seen used correctly at scale, so it reaches for the right one, and the framework is powerful enough to build a real product while remaining the terrain on which the agent is most capable, which, by the argument of Section 3, is the combination that matters.

9. The state: Postgres

Everything above concerns code, which is safe to change because it is versioned and reversible. Data is not: a dropped column or a bad migration can destroy something no commit can restore. Postgres is the durable, transactional state at the bottom of the chain, and the discipline around it is deliberately stricter than around code. PostgreSQL earns its place the same way the rest of the chain does, it is the conventional, battle-tested, superbly documented choice, so the agent writes correct SQL and correct migrations against it, and a managed instance removes the operational toil without changing the model.

The care goes into *change*. Schema changes are migrations: ordered, versioned files checked into the repository and reviewed like any other change, run against the preview before production, and written to be reversible. A live database cannot be changed the way a fresh one can, so a risky change is made as a sequence of safe ones, the *expand and contract* pattern (Ambler & Sadalage, 2006): add the new column or table and start writing to it, backfill, move reads across, and only then, in a later deployment, remove the old shape, so that at every intermediate step the running code and the schema are compatible and the step is reversible. Postgres helps by making most schema changes transactional, so a failed migration rolls back cleanly rather than leaving the database half-changed. Two operational points matter in a serverless setting, where many short-lived function instances each want a connection: a *connection pooler* is not optional, because a database has a hard connection limit that a fleet of serverless workers will otherwise exhaust; and *row-level security* lets the database itself enforce who may see which rows, a backstop that does not depend on the application getting every query right. The rule we keep is simple and worth stating: the agent may propose a migration, but a migration is the one kind of change a human always reads with full attention, because it is the one kind the chain cannot fully undo.

An ORM makes that change fast without making it reckless. We use Prisma, a schema-first ORM whose single `schema.prisma` file is the one declaration of the data model, and editing it has two mechanical consequences. `prisma migrate` derives the SQL migration from the difference between the schema and the database, writing ex-

actly the ordered, versioned, checked-in file described above instead of one written by hand, so the expand-and-contract discipline stays intact while the boilerplate disappears. And the generated client is typed against the schema, so a renamed column or an altered relation is not a silent runtime break: every query that no longer matches surfaces as a type error, the first and cheapest gate, at the moment the agent edits the schema and long before the code runs. It is the property the whole chain is built on, applied to data, a declarative change, the type checker showing its full blast radius, the agent fixing the fallout before anyone sees it, and it is what makes a schema change both quick and hard to get subtly wrong.

10. The human gate and how it fails

Every automatic gate in this chain exists to make the human gate cheap, not to remove it. The reviewer at the pull request does the one thing no gate automates: judging whether the change is what the business actually wanted and whether it is safe to expose, questions that types, tests and previews inform but do not answer. The chain's whole purpose is to deliver that reviewer a small, running, pre-checked change so that the judgement is quick and well-founded. But the gate has characteristic failure modes, and naming them is part of using the chain well.

The first is *rubber-stamping*. When the agent is usually right, the reviewer learns to approve quickly, and the review degrades into a formality that catches nothing. The defence is to keep changes small, so that reading them properly stays cheap, and to treat the preview as mandatory, not optional: a review that did not click through the running change is not a review. The second is *over-trust*, the belief that because the code passed the type checker and the tests it must be correct; those gates catch shapes and behaviours but not intent, and the human is the only gate on intent. The third is *scale fatigue*: an agent can open more pull requests than a team can seriously review, and a backlog of half-reviewed changes is the verification-bound regime reasserting itself. The answer is not to review faster but to generate more deliberately, to let the agent produce only what the team can actually verify, because unverified change is a liability whatever its volume.

11. Costs and trade-offs

The chain is a set of choices, and honesty requires stating what they cost. The most obvious is *coupling*: Next.js and Vercel come from the same company, and building on them buys a smooth, integrated path at the price of a degree of lock-in, both to a framework that moves quickly and sometimes breaks compatibility between major versions, and to a hosting platform whose pricing and primitives you adopt. This is exactly the cost the sovereign option removes: running the chain on self-hosted Coolify keeps the framework but replaces the managed host with one you own, so the data lives where your own law reaches and there is no hosting vendor whose pricing or roadmap you must follow. It is not a free lunch; it moves a real operational responsibility, uptime, updates, backups and security, onto you or a partner who carries it, which is the mirror image of Vercel's trade: convenience bought with dependency, or sovereignty bought with operations. We run the managed path ourselves for its velocity and offer the sovereign one to organisations that need it, and on either path a team should keep its domain logic and its data portable so that the framework and host remain replaceable infrastructure rather than the product. The second cost is *the discipline the gate demands*: the chain is only as safe as the review it culminates in, and a team that lets the human gate decay to rubber-stamping has a fast way to ship confident mistakes, which is worse than a slow way to ship considered ones. The third is *the agent's access*: a tool that can edit the repository and trigger deployments is a powerful actor, and it must be scoped like one, with branch protection, with least-privilege credentials, with secrets kept out of the repository and out of the agent's reach, and with permission and infrastructure changes reviewed with particular care. The fourth is *the irreversibility of data*, treated above: the one place the chain's safety net has a hole, and the place to spend the most human attention. None of these is a reason not to adopt the chain; each is a reason to adopt it deliberately.

12. How we run the Nashua 360 Enterprise Platform on it

The chain in this paper is not a proposal we have theorised; it is the one we build and operate our own product on. The Nashua 360 Enterprise Platform is a Next.js application, deployed on Vercel, backed by Postgres, with its source on GitHub, and it is developed with Claude Code as the agent, including from the mobile app when a change is better started or approved away from a desk. For clients who need the data on infrastructure they own, we run the identical chain with the deployment link self-hosted on Coolify instead of Vercel; nothing else about it changes, which is the whole point of keeping the host replaceable. A checked-in conventions file tells the agent the house rules; every change to the platform is a branch and a pull request; every pull request gets its type check, its continuous-integration run, and its preview deployment on an isolated copy of the stack; nothing reaches the production branch without a human reading the diff and clicking through the preview; and migrations are the changes we read hardest. We did not assemble this chain to be fashionable. We assembled it because, once we accepted that an agent had made generation cheap and verification the constraint, every link followed from the question of how to verify fast and safely, and because the same reasoning that makes it right for us, conventional tools the agent handles well, wrapped in loops and gates that catch its mistakes, makes it right for most teams building web software today. That it also lets a serious product be advanced from a phone on a train is not the reason we chose it, but it is a fair sign of how far the shape of the work has changed.

13. Objections and replies

This is just continuous delivery with an agent bolted on. In large part, deliberately. The practices the chain rests on, version control, small reviewed changes, continuous integration, automated deployment with previews and rollback, are the established discipline of modern software delivery (Humble & Farley, 2010; Forsgren et al., 2018), and that

is the point: the agent does not replace that discipline, it raises the premium on it, because it is the discipline that supplies the verification the agent's speed now demands. The contribution is not a new practice but the argument for why the old practices matter more, not less, once generation is cheap.

A better model will make the tooling unnecessary. It will not, because the tooling addresses the part a better model does not touch. A more capable agent generates faster and, in a verification-bound world, generating faster does not raise what actually ships; only verifying faster does. The better model also produces more code, faster, for a human to check, which makes the gates more necessary, not less. The chain is a bet on verification, and that bet gets better, not worse, as generation improves.

You are locked into one vendor's stack. Only if you choose the managed path and stop there. The coupling of Next.js and Vercel is a real cost accepted for a real benefit, the integrated preview loop that is central to the whole argument, but the host half of it is optional: self-hosted Coolify keeps the entire chain, agent, version control, isolated preview, typed framework, reviewed database, and removes the hosting lock-in by putting the deployment on a server you own. The mitigation for the rest is the ordinary one, keep the essential logic, the domain code and the database, portable, and treat the framework and host as replaceable infrastructure, which is exactly what makes swapping Vercel for Coolify, or either for something else, a change of host rather than a rewrite.

14. Limitations

This is practitioner-facing advice, and its limits should be stated. It is a reasoned recommendation grounded in our own use, not a controlled comparison; we have not measured the chain against alternatives. The specific tools are current as of late 2026, and this is a fast-moving field: the argument, that verification is the constraint and that the chain should be read as loops and gates, is meant to outlast any particular tool, but the particular tools will change, and a reader should keep the argument and re-examine the parts. The recommendation is scoped to the kind of software we build, web applications with a database, and carries less directly to domains with different constraints, such as systems where the cost of a verification gate is itself prohibitive, or where no agent has seen enough of the stack to be trusted on it. And the chain's safety rests on a human gate that the chain makes cheap but cannot make automatic; a team unwilling to keep that gate honest should expect the speed without the safety.

15. Conclusion

A coding agent makes writing software fast, and in doing so moves the hard part into the open. Generation, the part that used to dominate the day, becomes cheap; verification, deciding what to build and being sure the result is right, does not, and becomes the constraint. The lesson for tooling is that the value of an AI-assisted setup is not in the agent but in the chain around it that lets a team verify at the speed the agent generates and catches its mistakes before users do. We build and run the Nashua 360 Enterprise Platform on such a chain: Claude Code to write, in a loop with the checks and reachable from a phone; GitHub to make every change a reviewed, reversible diff behind a protected branch; Vercel, or self-hosted Coolify, to give every change its own running preview and every release an instant rollback; Next.js as a typed, conventional terrain whose types are the first gate; and Postgres as durable state changed through reviewed, reversible migrations. Each link closes a loop or adds a gate, the gates are independent so that no one of them must be perfect, and the whole is chosen on a criterion particular to this moment: that for an agent, the conventional tool is the capable one. Write fast, verify faster, and let nothing reach production that a person has not seen running. That is the whole of it, and it is enough.

Appendix A. An adoption checklist

The practices below are what make the chain safe rather than merely fast. None is novel; the discipline is the value.

Table 2. An adoption checklist. Each practice keeps a gate honest or a loop short; the right-hand column is the failure it prevents.

Practice	What it prevents
Protect the main branch; merge only via reviewed pull request with required checks	Unreviewed or untested change reaching production
Keep a checked-in conventions file for the agent	Repeating the same corrections; context drift
Keep changes small	Review degrading into rubber-stamping
Make the agent run types and tests before handing over	A human doing the machine's verification
Treat the preview as mandatory in review	Approving code nobody saw running
Give previews isolated data	A review touching production data
Read every migration with full attention; expand then contract	The one change the chain cannot undo
Keep secrets out of the repository and the agent's reach; least privilege	A powerful actor with credentials it does not need
Generate only what you can verify	A backlog of unverified, liable change

References

[1] Ambler, S. W., & Sadalage, P. J. (2006). *Refactoring databases: Evolutionary database design*. Addison-Wesley. ISBN 978-0-321-29353-5.

[2] Bacchelli, A., & Bird, C. (2013). Expectations, outcomes, and challenges of modern code review. In *Proceedings of the 35th International Conference on Software Engineering (ICSE)* (pp. 712–721). IEEE. https://doi.org/10.1109/ICSE.2013.6606617

[3] Beck, K. (2000). *Extreme programming explained: Embrace change*. Addison-Wesley. ISBN 978-0-201-61641-5.

[4] Brooks, F. P. (1987). No silver bullet: Essence and accidents of software engineering. *Computer*, 20(4), 10–19. https://doi.org/10.1109/MC.1987.1663532

[5] Chen, M., Tworek, J., Jun, H., et al. (2021). Evaluating large language models trained on code. *arXiv preprint* arXiv:2107.03374. https://arxiv.org/abs/2107.03374

[6] Anthropic (2026). *Claude Code documentation*. https://code.claude.com/docs

[7] Coolify (2026). *Coolify documentation: self-hosted deployment platform*. https://coolify.io/docs

[8] Forsgren, N., Humble, J., & Kim, G. (2018). *Accelerate: The science of lean software and DevOps: Building and scaling high performing technology organizations*. IT Revolution Press. ISBN 978-1-942788-33-1.

[9] Fowler, M. (2006). *Continuous integration*. martinfowler.com. https://martinfowler.com/articles/continuousIntegration.html

[10] GitHub (2026). *GitHub documentation*. https://docs.github.com

[11] Humble, J., & Farley, D. (2010). *Continuous delivery: Reliable software releases through build, test, and deployment automation*. Addison-Wesley. ISBN 978-0-321-60191-9.

[12] Vercel (2026). *Next.js documentation*. https://nextjs.org/docs

[13] Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint* arXiv:2302.06590. https://arxiv.org/abs/2302.06590

[14] PostgreSQL Global Development Group (2026). *PostgreSQL documentation*. https://www.postgresql.org/docs/

[15] Reason, J. (2000). Human error: Models and management. *BMJ*, 320(7237), 768–770. https://doi.org/10.1136/bmj.320.7237.768

[16] Sadowski, C., Söderberg, E., Church, L., Sipko, M., & Bacchelli, A. (2018). Modern code review: A case study at Google. In *Proceedings of the 40th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)* (pp. 181–190). ACM. https://doi.org/10.1145/3183519.3183525

[17] Vercel (2026). *Vercel documentation*. https://vercel.com/docs