

Agent Mode Optimization Playbook

Reduce Retries. Ship Faster. Spend Less on AI.

The Problem Most Teams Face

Agent Mode feels powerful... until:

- builds start breaking
- tasks need 3–5 retries
- context gets messy
- PRs take longer than expected

Root cause: not the model — your workflow.

The 4-Step Workflow That Fixes It

Stop prompting randomly. Use this structure every time:

1. Scan

Identify files, dependencies, and impact areas
→ Prevents editing the wrong parts

2. Plan

Define steps before execution
→ Reduces messy, inconsistent changes

3. Execute

Make scoped, controlled edits
→ Avoids over-editing and regressions

4. Validate

Run tests + check edge cases
→ Ensures merge-ready output

Model Strategy (Don't Guess)

- **Composer 1.5** → daily work, fast iterations
- **Opus 4.6** → deep tasks, large repos, audits

👉 Default to Composer
👉 Switch to Opus when complexity increases

Biggest Mistake to Avoid

❌ "Build feature X"
❌ "Refactor entire module"

✅ Break it down:

- analyze
- implement
- validate

Smaller tasks = fewer failures

How to Cut Retries by 30–50%

- define file boundaries
- mention dependencies explicitly
- always include validation step
- avoid multi-feature prompts

Example:

“Update auth middleware across API routes. Do not change DB schema. Run tests after changes.”

Handle Long Tasks Without Chaos

- split into phases
- summarize progress after each step
- don't overload context

👉 This prevents duplicate work and confusion

Cost Optimization (Simple Rule)

You don't reduce cost by switching models.
You reduce cost by **reducing retries**.

- fewer retries → fewer tokens
 - better planning → faster completion
-

What to Track (Real Performance)

- retries per task
- build success rate
- time to merge
- reviewer edits

👉 This is your real ROI from Agent Mode

Final Takeaway

Better workflow > better model

Teams that structure tasks properly:

- ship faster
- break less
- spend less

Want to fix your Agent Mode workflow?

We help teams reduce retries, improve build success, and optimize AI-assisted development.

→ [Book a Free 30-Minute Consultation](#)
