AI Field Guide

CHAPTER FOUR

The Tools, Not the Models

Claude Code, Google Antigravity, Codex, Cursor and the rest — where the real competition is happening, and whether you should be looking over the fence.

Compiled 14 July 2026 · this is the chapter that matters most to how you actually work

CHAPTER FOUR

First, the correction

> ANTIGRAVITY IS NOT A MODEL
>
> You listed it alongside Claude, Kimi and DeepSeek. It doesn't belong there. **Google Antigravity is an agent-first *development platform*** — an IDE, a CLI, and an SDK for orchestrating coding agents. It is powered *by* Gemini models (and, notably, by Claude models too).
>
> So it doesn't compete with Claude-the-model. **It competes with Claude Code.** And that turns out to be a far more interesting comparison for you, because it's a comparison between the tool you live in all day and its most credible challenger.

The mistake is worth making, because correcting it reframes the question. You don't really want to know "which model is best" — you have that from Chapters Two and Three. You want to know **whether the thing you spend eight hours a day inside is still the right thing.** That's this chapter.

CHAPTER FOUR

What changed in Claude Code while you weren't looking

Before comparing outward, it's worth noting what shipped in the tool you already pay for — because based on your usage history, **you are not using most of it.**

> DYNAMIC WORKFLOWS — THE BIG ONE, SHIPPED MAY 2026
>
> Claude now *writes a JavaScript orchestration script* that fans out subagents: **up to 16 concurrent, 1,000 per run**, executing in a background runtime while your session stays responsive, with a grader loop that sends subagents back to revise until the output meets a rubric.
>
> The proof point is genuinely startling. Jarred Sumner used it to port **Bun from Zig to Rust — roughly 750,000 lines, 99.8% of the test suite passing, in eleven days.**
>
> Intermediate results live in *script variables, not your context window* — which is why it can go so wide without drowning. You trigger it with the keyword `ultracode:` in a prompt, or `/effort ultracode` for a whole session.

Also new since you'd have last looked: **Routines** (scheduled agents on Anthropic's infra that survive your laptop being closed — this is what your AI Brief cron actually wants to be), **Agent Teams** (3–5 peer agents that can argue with each other to disprove one another's theories), **Channels** (push events from Telegram/Discord/webhooks into a session), and `/goal` (a separate evaluator model re-checks a completion condition after every turn and restarts Claude until it holds).

CHAPTER FOUR

The master comparison

TOOL	SURFACES	MODELS	ORCHESTRATION	PRICE	STANDOUT	WEAKNESS
Claude Code Anthropic	CLI, VS Code, JetBrains, **Desktop**, Web, iOS, Slack, CI	Sonnet 5, Opus 4.8, Fable 5 (Anthropic only)	**Best in class.** 1,000 subagents/run, grader loops, 3-level hierarchy	$20 / $100 / $200 no free tier	Orchestration depth + richest extensibility (Skills, Hooks, MCP, SDK)	Most expensive floor; opaque limits; single-vendor models
Antigravity 2.0 Google	Desktop app, Go CLI, SDK, Managed Agents API	Gemini 3.5 Flash (default), Gemini 3.1 Pro, **Claude Sonnet 4.6, Claude Opus 4.6**, GPT-OSS	Manager Surface, ~5 parallel agents, **per-subagent model choice**, scheduled tasks	**Free tier** $20 / $100	**Artifacts + browser surface** — the agent runs your app, clicks it, hands back a recording	Two months old; free tier dies fast; Google killed Gemini CLI with 30 days' notice
Codex OpenAI	Web, CLI, IDE, iOS, Slack, GitHub	GPT-5.6 Sol / Terra / Luna	Local + parallel cloud tasks	Free · **Go $8** · $20 · $100 / $200	Cheapest serious on-ramp; ~4× fewer tokens than Claude Code	Thinnest extensibility of the big three
Cursor Anysphere	AI-first editor (VS Code fork) + CLI	Composer 2.5 + **40+ third-party**	Agent mode; shallower	$20 / $60 / $200	Best editor ergonomics; Composer 2.5 is 3rd on the index at 10–60× lower cost	Pricing restructured repeatedly; you must live in *their* editor

TOOL	SURFACES	MODELS	ORCHESTRATION	PRICE	STANDOUT	WEAKNESS
Copilot GitHub	VS Code, JetBrains, github.com, CLI	Multi-vendor via **Agent HQ**	Control plane over *other people's* agents	Free · **$10** · $39 · $100	**Inline completions are free and burn zero credits** on all paid plans	Not the leader on any single axis
Amp Sourcegraph	Terminal + every major IDE	Multi	Named subagents (Oracle, Librarian, Painter)	**Free, ad-supported** $10/day API cap, zero markup	Most transparent pricing in the industry	Ad-supported; smaller ecosystem

CHAPTER FOUR

Claude Code vs Antigravity, head to head

The framing the neutral reviewers keep landing on, and it's a good one:

> *Claude Code is optimised for controlled collaboration. Antigravity is optimised for autonomous execution. Claude Code feels like an AI developer; Antigravity feels like an AI engineering department.*

	CLAUDE CODE	ANTIGRAVITY 2.0
Shape	A CLI that dissolves into your existing setup	A desktop app you adopt *as* your workspace
Orchestration ceiling	**1,000 subagents**, grader loops, 3-deep hierarchy	~5 parallel agents, but **per-subagent model choice**
Verification	You review diffs. Chrome surface for debugging	**Artifacts** — the agent runs the app in a browser and returns screenshots and recordings as evidence
Models	Anthropic only	**Multi-vendor** — including Claude Sonnet 4.6 and Opus 4.6
Extensibility	Skills, Hooks, MCP, plugins, Agent SDK	AGENTS.md, Skills, Hooks, MCP, plugins — near-parity on paper, thinner in practice
Cost to try	$20/mo minimum	**$0**
Maturity	341 releases	Two months old as a standalone app

> **DO NOT DECIDE ON THE BENCHMARKS YOU'LL SEE QUOTED**
>
> The most-cited head-to-head figures (Claude Code at 87.6% SWE-bench Verified vs Gemini 3.5 Flash at 76.2% Terminal-Bench) are **not comparable** — two different benchmarks, and the Claude number uses an outdated model. Whoever quotes those at you hasn't read them.

My honest recommendation for you

Don't migrate. Your entire operating system is built out of Claude Code primitives — the `erp-builder` and `core-2.0-dashboard` skills, your MCP servers, your `launch.json` setups, your memory files, your deploy-to-Pages habits. Antigravity has analogues for most of it, but you'd be re-authoring all of it, and its orchestration ceiling (~5 parallel agents) is well below what Dynamic Workflows already gives you today for free.

> **BUT DO SPEND ONE EVENING ON IT — AND HERE'S THE SPECIFIC REASON**
>
> Antigravity's genuinely differentiated capability is **Artifacts and the browser surface**: the agent runs your app, clicks through the UI, and hands back a screenshot or a screen recording as *evidence it worked*.
>
> Look at your own tool usage. You called `preview_screenshot` **567 times** and `preview_eval` **1,434 times** in two weeks. Your memory file records that "*framer mount-anims get stuck here so keep static.*" **Visual verification of React UIs is, empirically, one of your biggest time sinks.** Antigravity is the only mainstream tool that has built a first-class answer to exactly that problem.
>
> It costs nothing and one evening. Run *one* self-contained job on it — a Sandow or Elite Mobile screen where "does the UI actually work" is the hard part — and see whether the Artifacts loop beats your current screenshot-and-squint cycle.

Two things to watch before you'd ever take it seriously

1. **Whether the free tier survives a real day.** Reported consensus is that "a handful of consecutive agent requests is often enough to end the day's allowance." Fine for evaluation, not for work.
2. **Whether Google does it again.** They killed Gemini CLI with roughly 30 days' notice in June 2026 and **broke people's CI pipelines.** You have production deploys on Cloudflare. A vendor that retires a CLI mid-flight is a vendor you keep at arm's length from anything load-bearing.

THE REALISTIC END STATE

Both, not one. **Claude Code stays your primary engine** — orchestration depth, code quality, and the skill library you've already built. **Antigravity, if it earns it, becomes the thing you reach for on UI-heavy prototyping where visual verification is the bottleneck.** That's exactly the split the neutral comparisons land on, and it happens to map cleanly onto the split in your own work.

CHAPTER FOUR

The rest of the field, briefly

- **Codex Go at $8/month** is the cheapest credible agent on the market, and Codex is *in the free ChatGPT tier* — Claude Code is not. If you ever want a second opinion on a diff for near-nothing, this is how.
- **GitHub Copilot Pro at $10** gives you unlimited inline completions that **burn zero credits**. As a pure autocomplete layer underneath Claude Code, that's close to free money.
- **Amp** is free and ad-supported with a $10/day API cap and *zero markup* — the most honest pricing in the category.
- **Cursor's Composer 2.5** ranks third on the Coding Agent Index at 10–60× lower cost than its rivals. If you ever feel the token bill, this is the escape hatch.
- **Kiro (AWS)** — spec-driven with agent hooks. Irrelevant to you; you're on Cloudflare.
- **Devin** — best at "delegate a well-specified ticket and walk away," bad at exploratory work.

CHAPTER FOUR

The uncomfortable observation

The genuinely important finding of this chapter isn't about Antigravity at all.

YOU ARE PAYING FOR A FERRARI AND DRIVING IT IN FIRST GEAR

Claude Code shipped **Dynamic Workflows** (1,000 orchestrated subagents), **Routines** (scheduled cloud agents), **Agent Teams**, `/goal`, **hooks**, and **worktrees**. In 1,318 prompts across two weeks, you have used essentially *none* of them.

The gap between Claude Code and Antigravity is real but modest. **The gap between the Claude Code you're using and the Claude Code that exists is enormous.** Closing that is worth far more to you than switching tools — and it's what Chapters Five and Seven are about.

Sources. Claude Code docs (code.claude.com) & changelog · Anthropic Dynamic Workflows announcement · Google I/O 2026 developer keynote · Antigravity docs & the Gemini CLI transition notice · OpenAI Codex pricing docs & rate card · Cursor docs · GitHub Copilot billing changes · Artificial Analysis · InfoQ · MarkTechPost · MindStudio. Antigravity's top pricing tier is secondary-source only and unverified against a Google page.