



The minimal, *always-on* AI agent.

Only what an agent needs to run, in one Rust binary. Any model, your own keys. Full control of your agents, none of the harness complexity.

```
curl -fsSL https://gray.alignment.id/install.sh | sh
```

MIT · LINUX · MACOS · WINDOWS

13_{ms}

COLD START

37_{MiB}

INSTALLED

PROBLEM

Bloated on your disk — *or locked in their cloud.*

Self-hosted harnesses like OpenClaw and Hermes ship hundreds of megabytes of features nobody uses. The new wave — Muse, Dots, Grok Bot, Claude Duet — doesn't run on your machine at all. Either way the agent burns tokens on tools it doesn't need, or tools you don't own.

2,406 MIB OF AGENT INSTALLS — ONE BAR, SCALED BY FOOTPRINT



RUNS ON THEIR COMPUTER — NOTHING TO MEASURE

MUSE
META

DOTS
OPENAI

GROK BOT
XAI

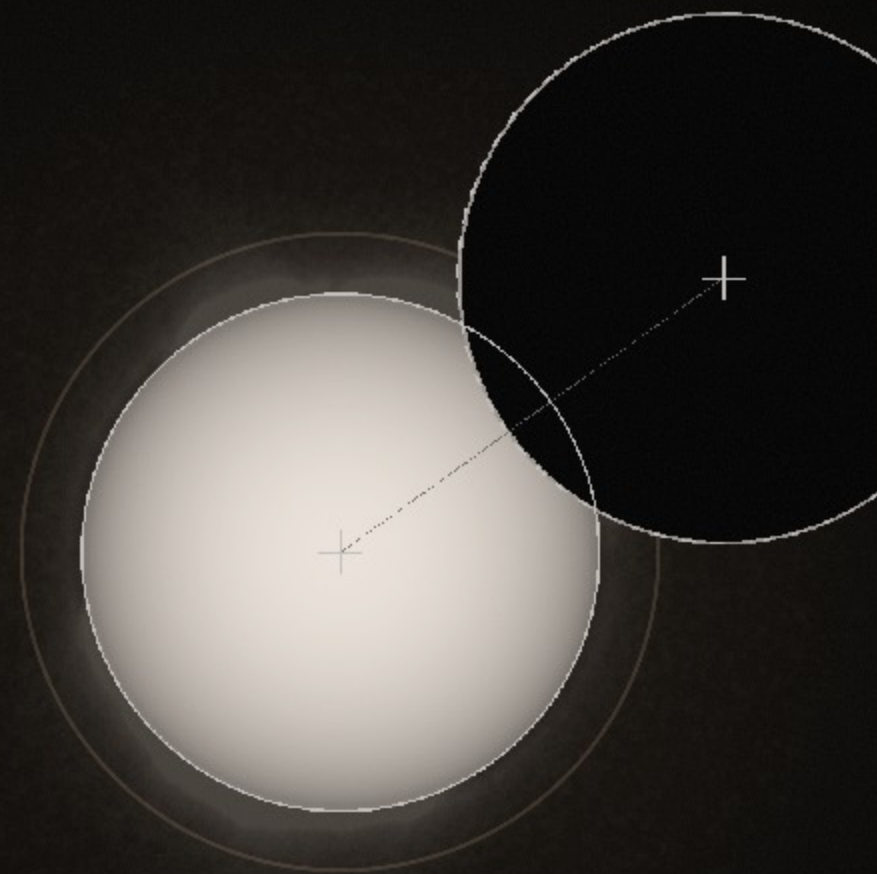
DUET
ANTHROPIC

Install footprints measured 2026-10-16 on one machine — full table on slide 04. OpenClaw figure is plugin runtime deps in `~/openclaw` alone.

PRODUCT

The bare minimum. *Everything else is a plugin.*

Less for the agent to think about means better performance at lower cost. The core has **one tool: bash**. Providers, the Discord bridge, themes and permission prompts are plugins you choose to add.



THE CORE

1

tool

\$ bash

+

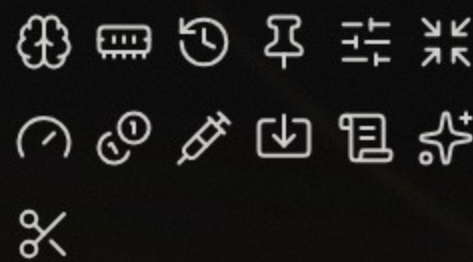
18

YOUR SUBSCRIPTIONS
ALREADY WORK

17

TALKS WHERE
YOU TALK

13

REMEMBERS
& COMPARIS

16

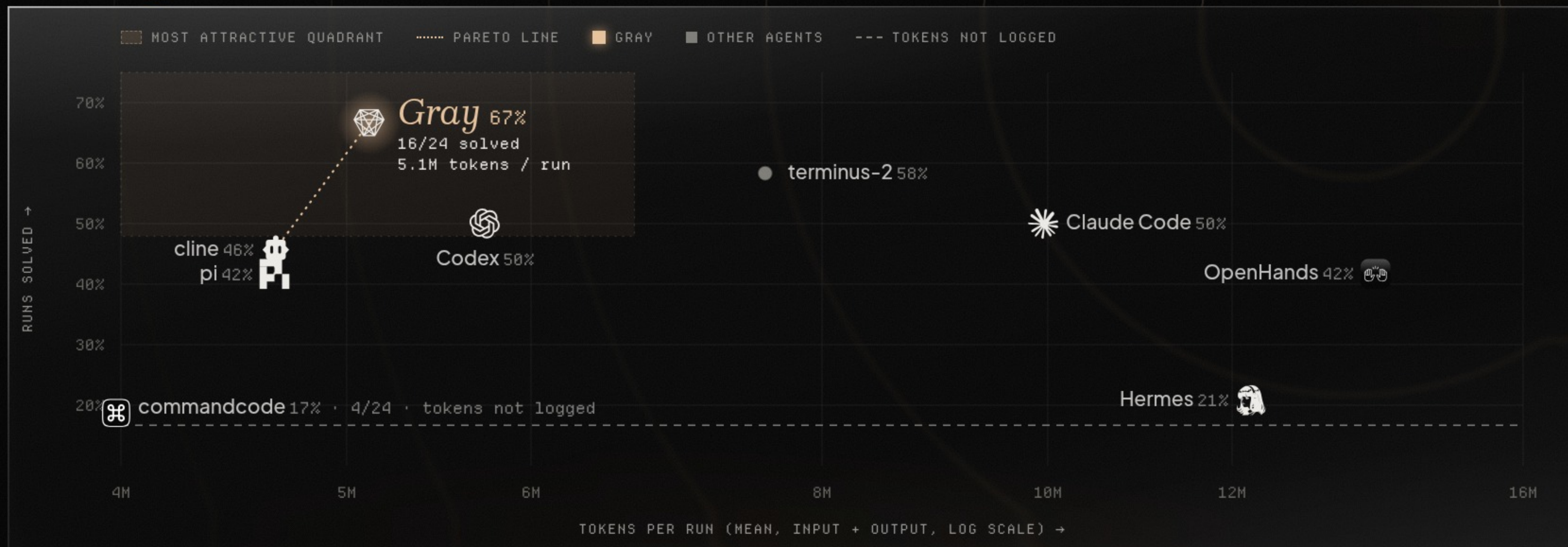
WORKERS
& WORK

15

SENSES
& OPERATES

152 official plugins on gray.alignment.id/hub — install one with `$ /plugin install <name>`. Each adds exactly the tools it needs, nothing more.

Same model, nine agents. *Gray solves the most.*



DeepSWE, 8 tasks × 3 runs per agent, all on glm-5.3-flash (z.ai, effort high), scored by each repo's tests, Oct 16-22 2026. commandcode logs no token counts, so it is drawn as a line at its solve rate; opencode excluded (config error).

PROOF · VS CLAUDE CODE · TERMINAL-BENCH 2.1

Gray used $3.4\times$ fewer input tokens than Claude Code

INPUT TOKENS / TRIAL

74k vs 249k

3.4× FEWER

OUTPUT TOKENS / TRIAL

13.4k vs 19.1k

1.4× FEWER

WALL TIME / TRIAL

361s vs 370s

9S APART · SAME PACE

TRIAL	INPUT TOKENS		RATIO	OUTPUT TOKENS		WALL TIME	
write-compressor r1	129k		1.7×	23.1k		270s	
	216k			20k		266s	
write-compressor r2	64k		2.9×	13.7k		124s	
	185k			21.3k		161s	
write-compressor r3	116k		4.7×	15.1k		132s	
	546k			39.5k		239s	
write-compressor r4	73k		3.3×	15.2k		130s	
	242k			25.1k		173s	
torch-tensor-par. r2	61k		3.0×	8.4k		578s	
	182k			7.4k		528s	
torch-tensor-par. r3	56k		3.4×	10.4k		780s	
	192k			7.8k		666s	
torch-tensor-par. r4	17k		10.8×	7.6k		515s	
	181k			12.4k		558s	

PROOF

Starts in 13 ms. *Installs in 37 MiB.*



AGENT	COLD START (--VERSION)	PEAK MEMORY	INSTALL SIZE
 Gray	13 ms 	12 MiB 	37 MiB
 Claude Code	46 ms 	37 MiB 	256 MiB
 Codex	224 ms 	46 MiB 	432 MiB
 pi	1,244 ms 	103 MiB 	128 MiB
 Hermes Agent	2,861 ms 	98 MiB 	945 MiB
 opencode	3,565 ms 	183 MiB 	242 MiB

Median of 25 runs on one x86_64 Linux machine, measured 2026-10-16. Bars scale to the slowest/largest in each column.

PROOF · VS HERMES

Half the time. *A sixth of the RAM.*

Both agents started at the same instant on the same model and endpoint. Gray finished first on 7 of 8 tasks.

 vs  Hermes

SAME MODEL, STARTED SIDE BY SIDE

4_{/8} 3_{/8}

GRAY

HERMES

RACE TASKS SOLVED

MEDIAN TIME PER TASK

Gray 32 min

Hermes 64 min

MEDIAN RAM

Gray 40 MiB

Hermes 252 MiB

INPUT TOKENS · 5 LOGGED TASKS

Gray 50.5M

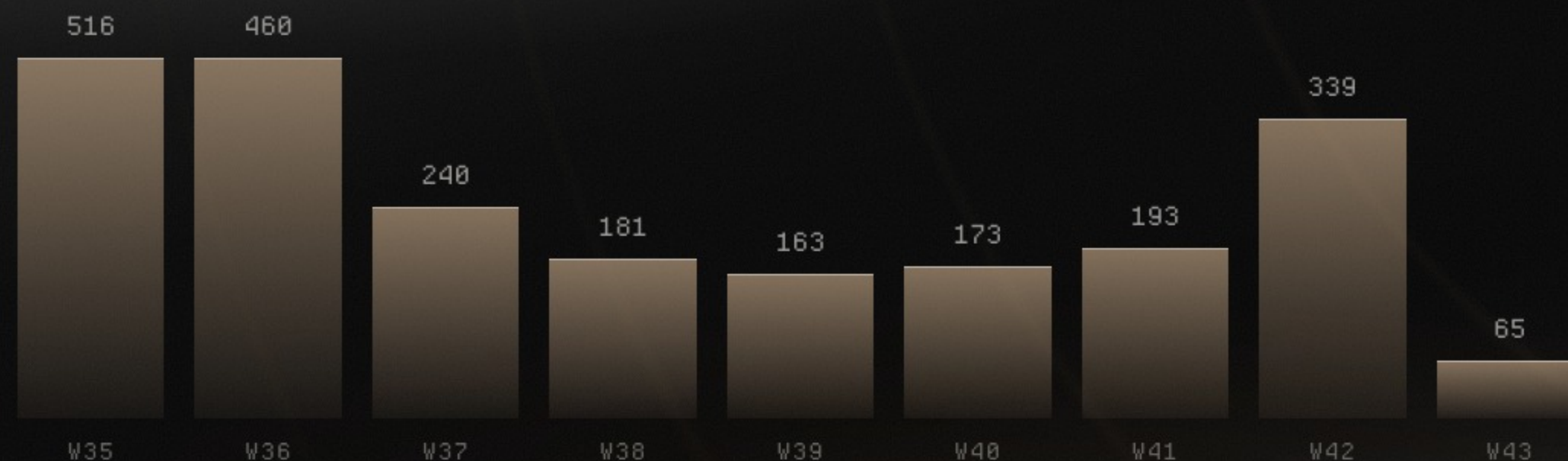
Hermes 111M

Hermes Agent 0.21.0 stock vs Gray 0.1.10, both at max effort on the same commandcode endpoint, 2026-10-06. RAM sampled every 2 s from each task container. Gray used 2.2× fewer input tokens on the 5 tasks both logged.

VELOCITY

2,330 commits in 59 days. *One founder.*

COMMITTS PER ISO WEEK · FIRST COMMIT AUG 24 · THIS WEEK IS PARTIAL



57

OF 59 DAYS
WITH A COMMIT

97%

OF COMMITTS
BY THE FOUNDER175_kLINES OF RUST
4TH ITERATION

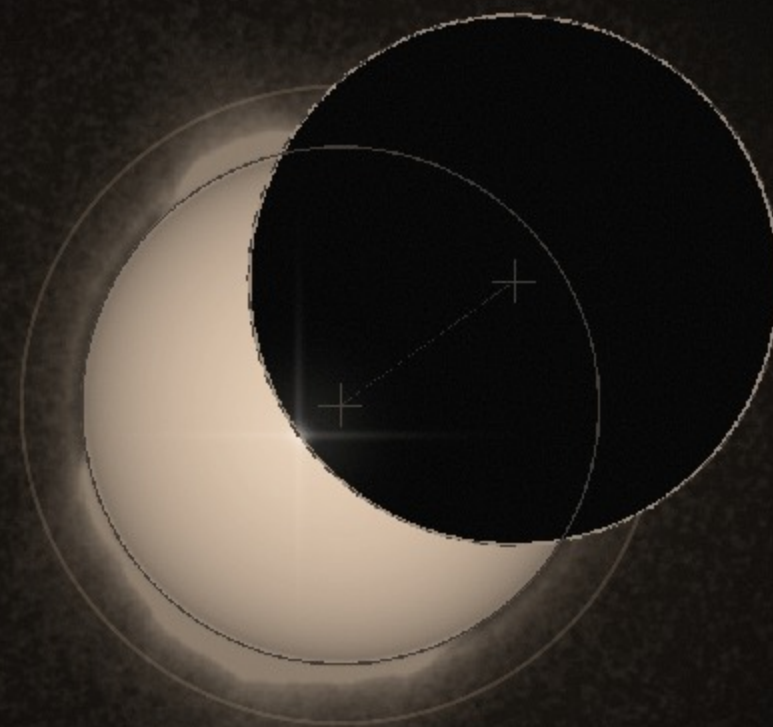
15

RELEASES
LATEST V0.2.2

TRACTION

Benchmarked *in public*.

Head-to-head runs against Hermes and Claude Code go out on X first, with every release and demo. They're the best-performing posts.



21

GITHUB STARS
6 FORKS

92

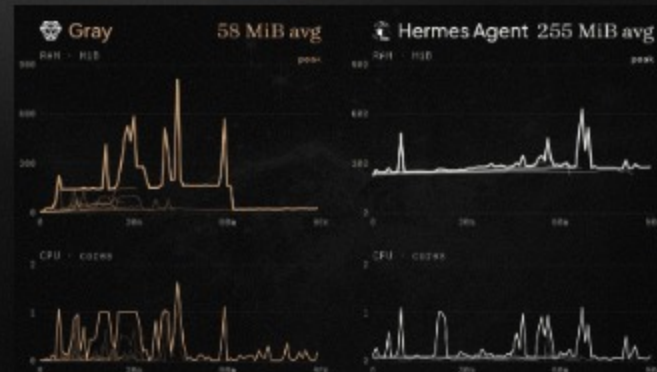
RELEASE
DOWNLOADS

7.5_k

VIEWS, LAST
100 POSTS

130

FOLLOWERS
ON X



Gray vs Hermes

	gray	claude
Score	20/20	20/20
Time	1m54s	2m36s
Requests	11	14
Cache read	117.8k	335.3k

Gray vs Claude Code
(one task on DeepSWE,
Sonnet 5.5 Med)



You can talk to Gray
anywhere using the Gray
Discord plugin!

BUSINESS MODEL

Free core. *Paid when we run it for you.*

Gray Cloud gives your agent **its own Linux machine**. It stays awake, keeps your sessions and crons, and answers while your laptop is off. Accounts, API keys and credits are live today at `gray.alignment.id/console`.

LIVE NOW · [GRAY.ALIGNMENT.ID/PLANS](https://gray.alignment.id/plans)

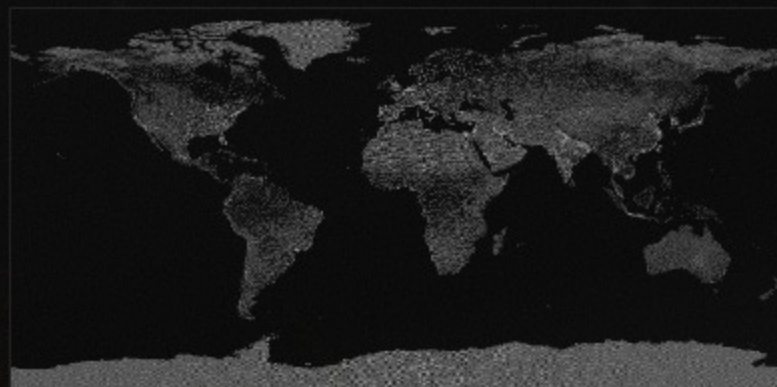
GRAYLING

\$1

ONCE

A one-time test drive. Not the subscription.

- YOUR OWN GRAY MACHINE ON GRAY CLOUD
- TALK TO IT FROM ANYWHERE WITH `gray cloud`
- \$10 OF QWEN 3.8 FLASH BY GMI CLOUD
- MINIMAX M3.1 FLASH, FREE
- CREDITS NEVER EXPIRE



MONTHLY PLANS · PAUSED UNTIL WE HAVE COMPUTE

PATHFINDER

PAUSED

\$7

/mo

Your own machine on Gray Cloud

SCOUT

PAUSED

\$27

/mo

Machine + \$30 credits

VOYAGER

PAUSED

\$77

/mo

Machine + \$90 credits

PIONEER

PAUSED

\$277

/mo

Machine + \$300 credits

COMPETITION · VS DEVIN

Devin's own model. *Gray drives it further.*

The Devin CLI and Gray both ran swe-2-high on the same 8 DeepSWE tasks. Gray finished every one.

vs Devin

DEVIN'S OWN MODEL, SWE-2-HIGH

8/8 5/8

GRAY

DEVIN

DEEPSWE TASKS SOLVED

TARGET TESTS PASSED

Gray 614/614

Devin 611/614

INPUT TOKENS, ALL TASKS

Gray 53.9M

Devin 59.3M

Devin CLI (devin -p) vs Gray driving swe-2-high through its Devin plugin, same task images and harness settings, one run per task, 2026-10-19.

FOUNDER

Vstalin Grady.

I automate everything I touch.

- Automated my own job search with AI agents: they refined my CV, then called APIs directly to send the applications.
- Built infrastructure for a clinic. When it shut down, I was asked to find everyone a new job, so I used that pipeline.
- Teaches **vibe coding**. In-person classes that take non-programmers from zero to a deployed website.



- Gray is my fourth iteration of this harness. Solo, full-time, from Indonesia.
- **Nerdiest thing**: optimized a friend's dating profile by scanning their whole photo gallery with Gemini embeddings, then set it up automatically over ADB.



THE ASK

Raising **\$100k** to keep Gray full-time and launch Gray Cloud.

USE OF FUNDS

- **Founder**, full-time
- **Inference** for development and public benchmarks
- **Hosting** for the Gray Cloud beta

WHAT IT BUYS

- **Gray Cloud** live with paying users
- **Public** DeepSWE and Terminal-Bench results
- **Autonomous research** beyond coding

My mission for Gray is to give every person the ability to own and command their own fleet of AI agents.