maestroai

You know who's using AI.
You don't know who's using it well.

Maestro shows you.

YOUR DASHBOARDS ARE BROKEN

AI makes velocity metrics meaningless.

One engineer is mastering the craft of working with AI. The other is winging it. Your dashboards can't tell them apart.

Sarah Chen

PRs merged	11
AI sessions	34
Tokens used	1.4M
Avg cycle time	3.1 hrs
Build pass rate	96%

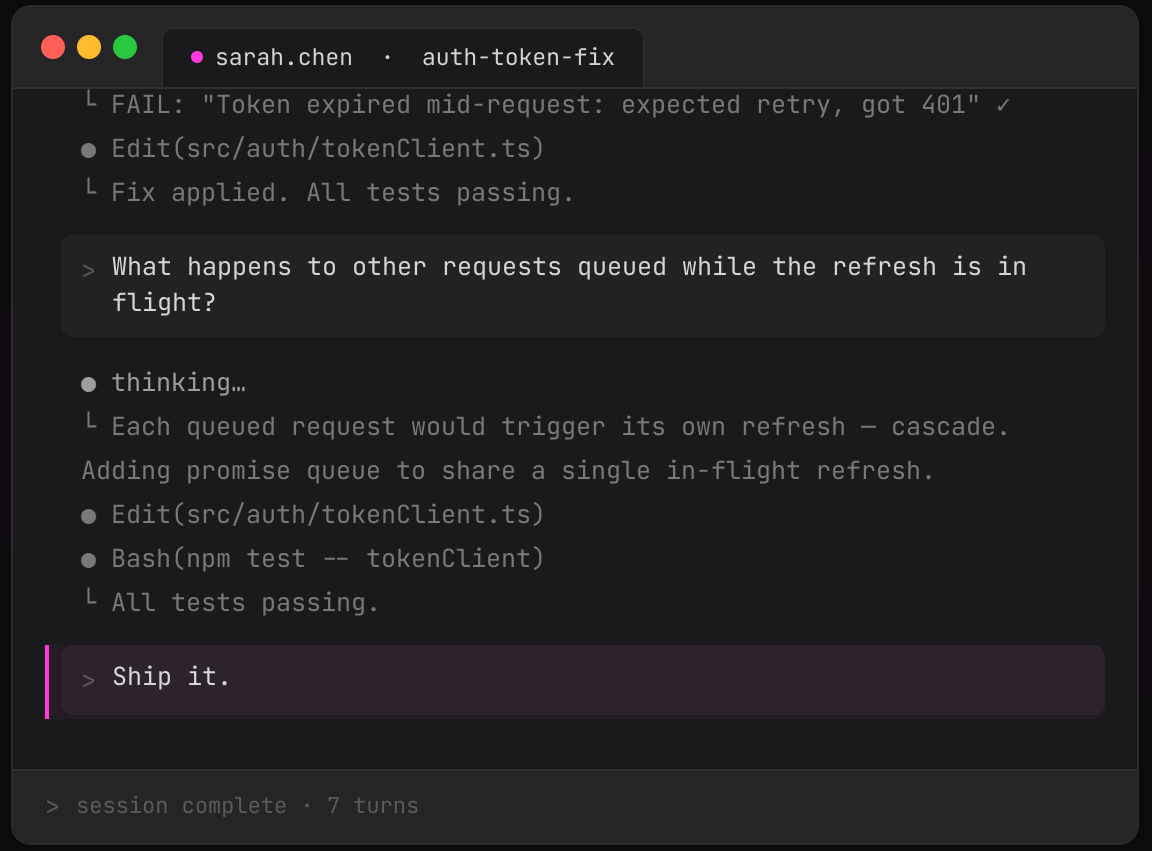
Alex Rockwell

PRs merged	13
AI sessions	31
Tokens used	1.8M
Avg cycle time	2.8 hrs
Build pass rate	91%

THE NEW CRAFT OF ENGINEERING

Well-run agent sessions produce excellent code.

Poorly-run sessions produce AI slop.



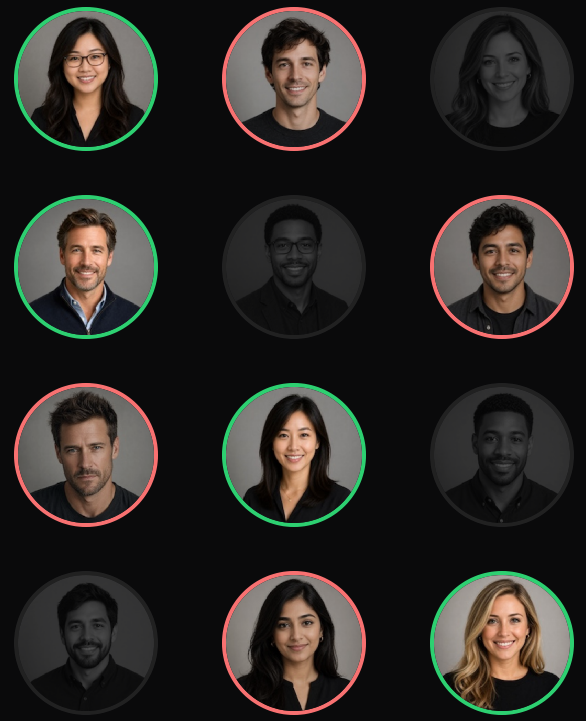
THE NEW BLIND SPOT

Which engineers are your AI leaders?

- Some have already mastered the new craft.
- Others are struggling with every session.
- And some are still skeptics.

All of their sessions are invisible.
To you. And to each other.

No shared learning. No tribal knowledge. No visibility into what works. And what doesn't.



MAESTRO SEES WHAT DASHBOARDS DON'T

Two agent coding sessions. Two outcomes.

Add OAuth2 PKCE flow for mobile clients

Sarah Chen · Claude Code

PROMPTS 9 FOCUS 9 REVIEW 9 EFFICIENCY 8 | SESSION QUALITY 9/10

MAESTRO INSIGHT

High-quality verification loop

Sarah scoped the problem before letting the agent code, and caught the expired-token edge case the agent missed.

EVIDENCE FROM TRANSCRIPT — 2 of 38 turns

turn 1 · opening prompt · scope before code

Before you write code, list the security tradeoffs of PKCE vs. implicit flow for our mobile case.

turn 23 · +18 min · caught edge case

Run the test against an expired refresh token — that's the case I'm worried about.

38 turns 14 tools 145k tok $4.20 PR #1247 merged

Fix auth token refresh race condition

Alex Rockwell · Cursor

PROMPTS 3 FOCUS 3 REVIEW 2 EFFICIENCY 3 | SESSION QUALITY 3/10

MAESTRO INSIGHT

Agent-led — no verification

Alex gave the agent no context and approved the first approach without testing. The expired-token edge case was never caught.

EVIDENCE FROM TRANSCRIPT — 2 of 87 turns

turn 1 · opening prompt · no context given

Just fix the token refresh bug.

turn 3 · +2 min · blind approval

Ok sure, go ahead.

87 turns 31 tools 359k tok $33.60 PR #1231 open

HOW IT WORKS

Insights powered by a coding agent plugin.

01 **Lightweight agent plugin**

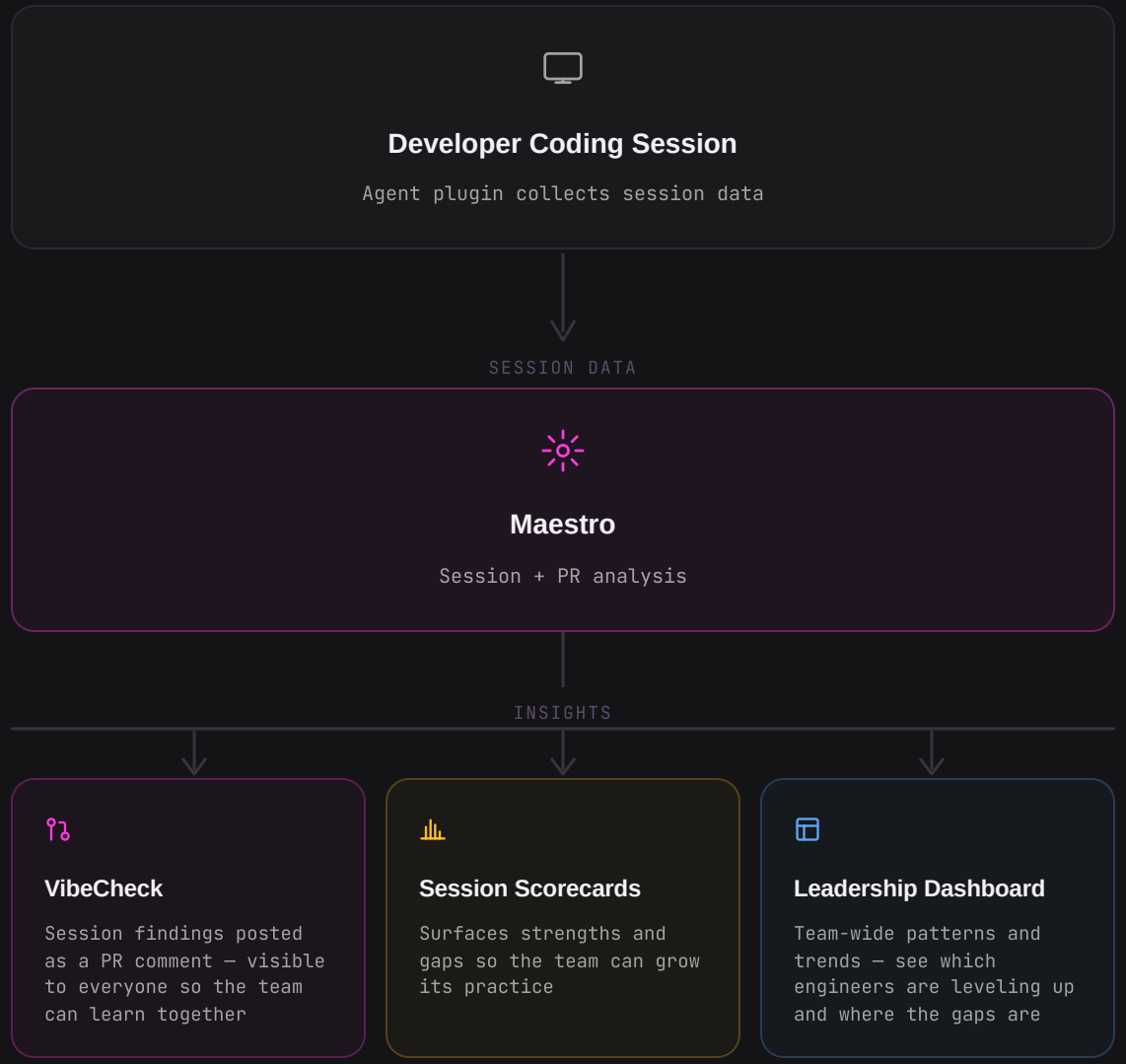
Maestro installs as a plugin to Claude Code and Codex. It collects session data as engineers work. Deploy via your existing MDM in minutes.

02 **Maestro extracts insights**

Sessions linked to shipped PRs are analyzed across five dimensions of craft. Only work that reached a PR is evaluated — exploratory sessions stay private.

03 **Surfaced in two places**

Engineers see their session review as a PR comment — same as a code review, visible to them first. Leaders see team-wide patterns on the Maestro dashboard.



THE FIVE DIMENSIONS OF SESSION CRAFT

Not a vibe. A standard.

Five dimensions. One score per session. Calibrated against shipped outcomes.

AGENT FOCUS

Did the engineer stay scoped, or did the session drift?

HumanLayer's analysis of ~100k sessions found recall degrades when context fill exceeds 40%. Drifting sessions hit that threshold faster — and the agent stops following instructions reliably.

EFFICIENCY

Did they reach the solution with minimal wasted motion?

GitClear's 2026 study of 2,172 developer-weeks found AI power users author 4–10× more durable code than average AI users. Efficiency, not volume, is the separator.

SESSION MANAGEMENT

Did they control the session, or let the agent run loose?

DORA 2025: code-review time rose 91% and PR size rose 154% in high-AI teams. Scoped sessions — one task per thread — are what keep review from collapsing.

VERIFICATION RIGOR

Did they verify the agent's output before calling it done?

Stanford/MIT (Mar 2026): 14.3% of AI-generated code contained vulnerabilities vs. 9.1% for human-written. Skipping verification is where that gap surfaces in production.

PROMPT QUALITY

Did they give the agent enough context to succeed?

Practitioners from Anthropic to Cursor converge: a tight plan before editing prevents most mid-session corrections. Boris Cherny attributes "one-shot" implementations to plan quality, not model quality.

TEAM MOMENTUM, AT A GLANCE

See which teams are leveling up — and which need coaching.

5 TEAMS · 38 ENGINEERS · LAST 8 WEEKS

TEAM	MEMBERS	EFFECTIVENESS ▼	STRONGEST	WEAKEST
Platform Engineering	8	 63 -5	Agent Focus +29	Verification Rigor -21
Payments Engineering	6	 56 -15	Agent Focus +12	Prompt Quality -19
Security Engineering	5	51 +2	Session Management +5	Verification Rigor -18
Product Engineering	12	 45 +6	Agent Focus -3	Verification Rigor -25
Data & Analytics	7	45 -2	Session Management ±4	Prompt Quality -23

- → Coach Platform and Payments on Verification Rigor.
- → Coach Payments and Data on Prompt Quality.
- → Security has the right habits — it just needs better verification.

"My team was shipping more code than ever. That's not the same as shipping better code."

CEO, Series B Fintech

3.1×

session quality gap between top-quartile and bottom-quartile engineers on the same team

62%

of "low-quality" sessions came from engineers whose PR throughput looked healthy

+18 pts

average AI Effectiveness gain in 12 weeks for teams using Maestro

Maestro helps your whole team master AI *faster*.

Book a demo →

getmaestro.ai