ENGINEERING AN AUTONOMOUS SDLC AT SCALE

Prompt → Prod.

From ad-hoc prompting to orchestrated agentic work.

SPEAKER

Andrew Swerdlow

ROLE

Sr. Director · Engineering Acceleration & Core Platforms · Roblox

TRACK

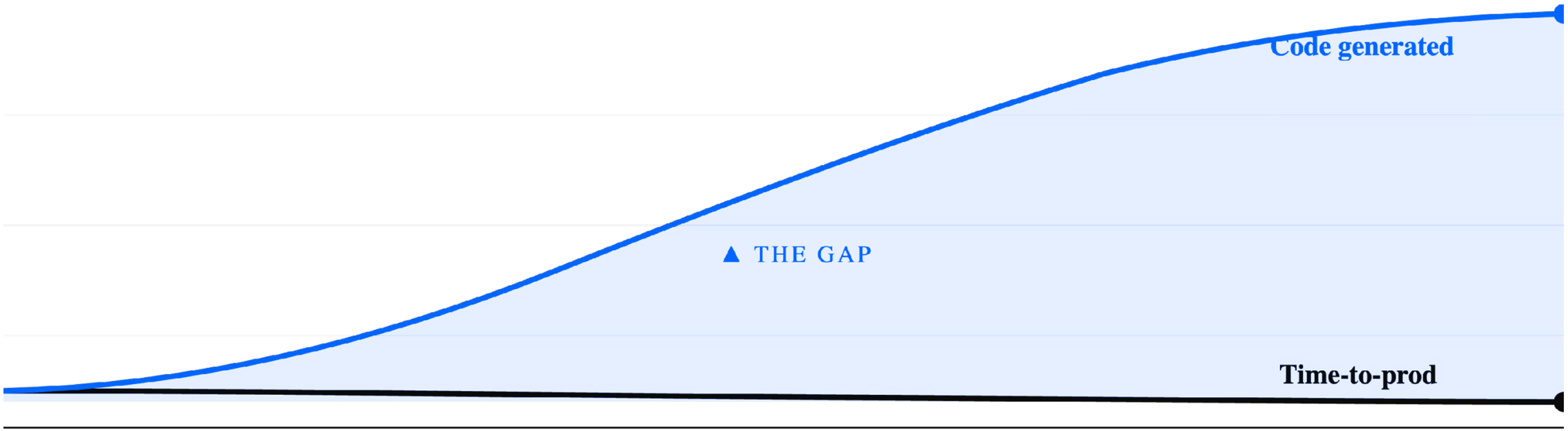
AI for Software Engineers

SLOT

Mon · 03:40 PM · Grand Ballroom

THE PRODUCTIVITY PARADOX

We solved typing. We didn't solve trust.



The bottleneck moved from writing code to reviewing, validating, and trusting it in production

The future isn't better ~~autocomplete~~.

It's Infra systems that can safely absorb autonomous change at the scale of a mature codebase.

We need to build Infrastructure for Self-Driving Systems. Everything that follows is about building it.

SHIFTING UP THE MATURITY LADDER

A reframe. Not an upgrade.

AD HOC

Individual & assistive.

IDE completions. Prompt files. Inconsistent across repos. Productivity gain is local.

ORCHESTRATED

Multi-agent & governed.

Agentic PRs. Continuous migrations. Workflows committed to Git. Productivity gain is structural.

Different cognitive posture. Different infrastructure. Different metrics.

THE TRUST GAP

Speed and autonomy without safety just creates debt faster.

01	Context & Ownership
02	Reliability & Security
03	Accountability when something breaks at 3am

We need systems that limit the blast radius of autonomy gone wrong

WHAT YOU'LL WALK OUT WITH

Three takeaways

TAKEAWAY · 01

Alignment & Guardrails

Encoding expert judgment as operational systems.

TAKEAWAY · 02

Security & Access

Architecting an agentic SDLC that securely runs on its own.

TAKEAWAY · 03

Measuring What Matters

Productivity metrics for an Agentic native organization.

TAKEAWAY · 01

Alignment & Guardrails

Encoding expert judgment as operational systems.

Mosts company's first reaction to autonomous agents is rejection.

How do we get to "Yes", without compromising prod?

A STORY BEFORE A FRAMEWORK

Security said slow down.

We had to rethink how to introduce less-controlled agentic frontier harnesses and models.

CONCERN · 01

Prompt-injection → source exfiltration

A poisoned dependency or malicious doc could turn an agent into a data-leak channel.

CONCERN · 02

Inherited user permissions, including prod access

Agents acting as developers would inherit everything developers can do — by default.

CONCERN · 03

The rogue-agent scenario

One compromised context could mean real data deletion risk against production systems.

THE TENSION

Many issues to consider, but the industry was moving fast

We did not want our engineers to fall behind.

16 / 30

THE THREE GUARDRAILS

Not features. Agent safe Infra.

GUARDRAIL · 01

Sandboxing

Blast-radius containment for every agent run. Local + cloud.

GUARDRAIL · 02

Policy gateways

MCP gateway · deep packet inspection · JIT Vault · Least privileged access. No long-lived prod secrets.

GUARDRAIL · 03

Hardening

Agent identities separate from user identities. Always auditable, by design.

Designed to treat agents as untrusted and limit any blast radius

Why autonomous coding can break in production.

And what alignment actually means when the model has read the internet — but not your company.

CONTEXT · ALIGNMENT · REPO MEMORY

THE CONTEXT GAP

The model knows the internet.

It doesn't know your company.

NOT IN THE TRAINING SET

- Company conventions
- Migration scars
- Security patterns
- SEV lessons
- Infra assumptions
- Ownership graphs

"Your engineering culture is not in the training set."

EXEMPLAR ALIGNMENT

Not prompting. Not fine-tuning. Operational learning.

WHAT IT IS

Expert judgment captured as scoped, testable exemplars

Config-as-code · versioned · auditable. Each exemplar runs as a focused sub-agent against matching file patterns.

WHY IT MATTERS

The alignment engine — not the model — is the durable asset

The model can change tomorrow. The exemplars stay. Your institutional voice is portable across frontiers.

HYBRID REPRESENTATIONS

Codebases are software and memory systems.

SYMBOLIC

ASTs · dependency graphs · ownership

What calls what. What breaks if I change this. Who's responsible.

Correctness.

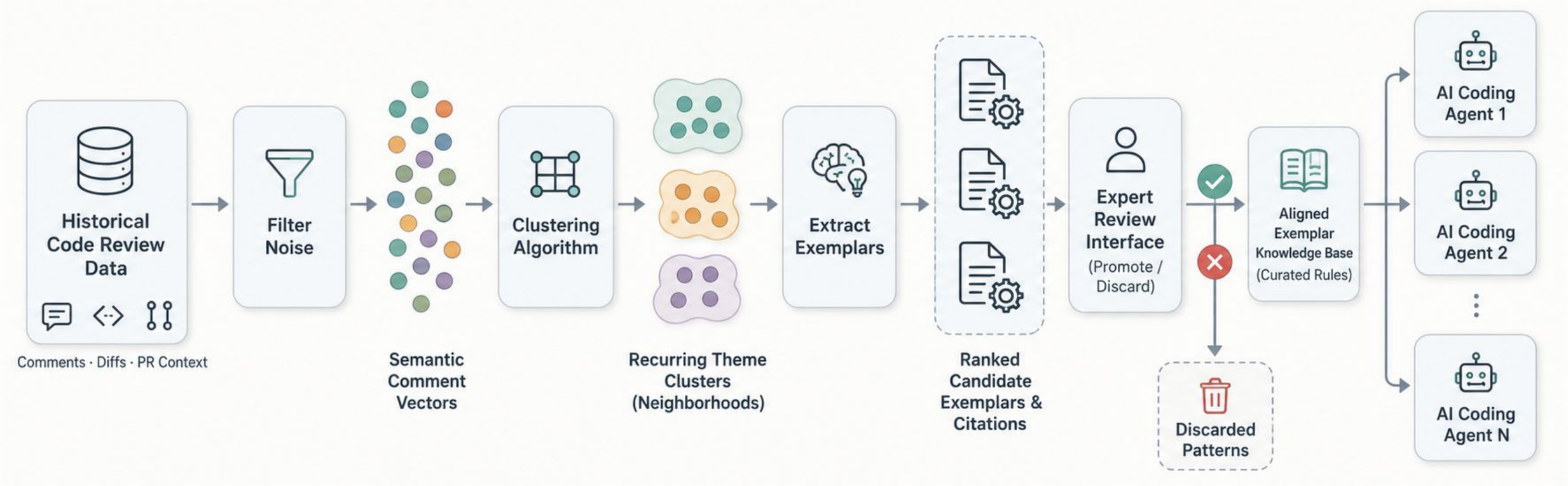
VECTOR

Embeddings over PRs · reviews · exemplars

What this code means. What's similar elsewhere. **Meaning.**

Together — agents reason like a senior engineer, not a stranger.

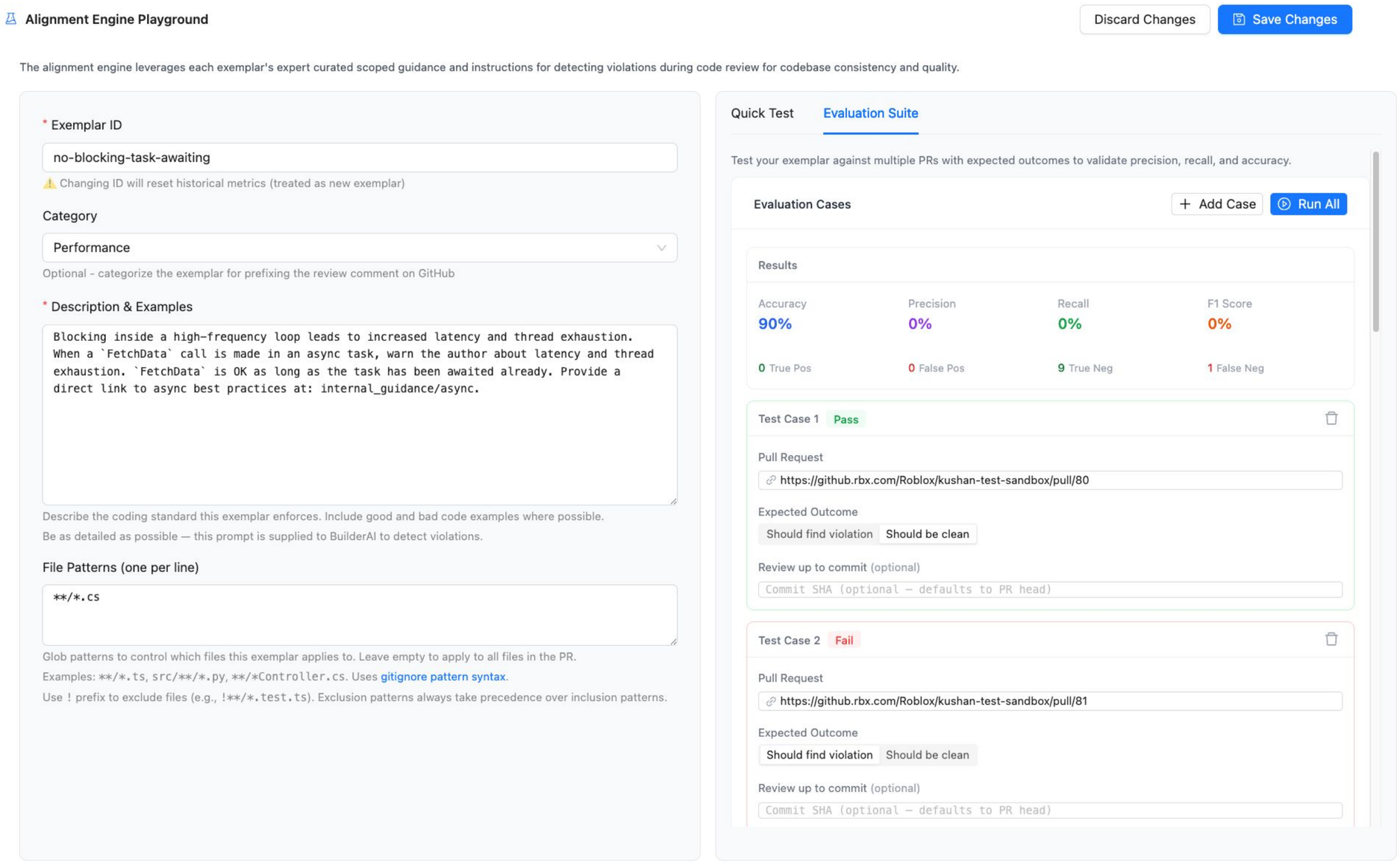
Turning 1.7M review comments into organizational memory.



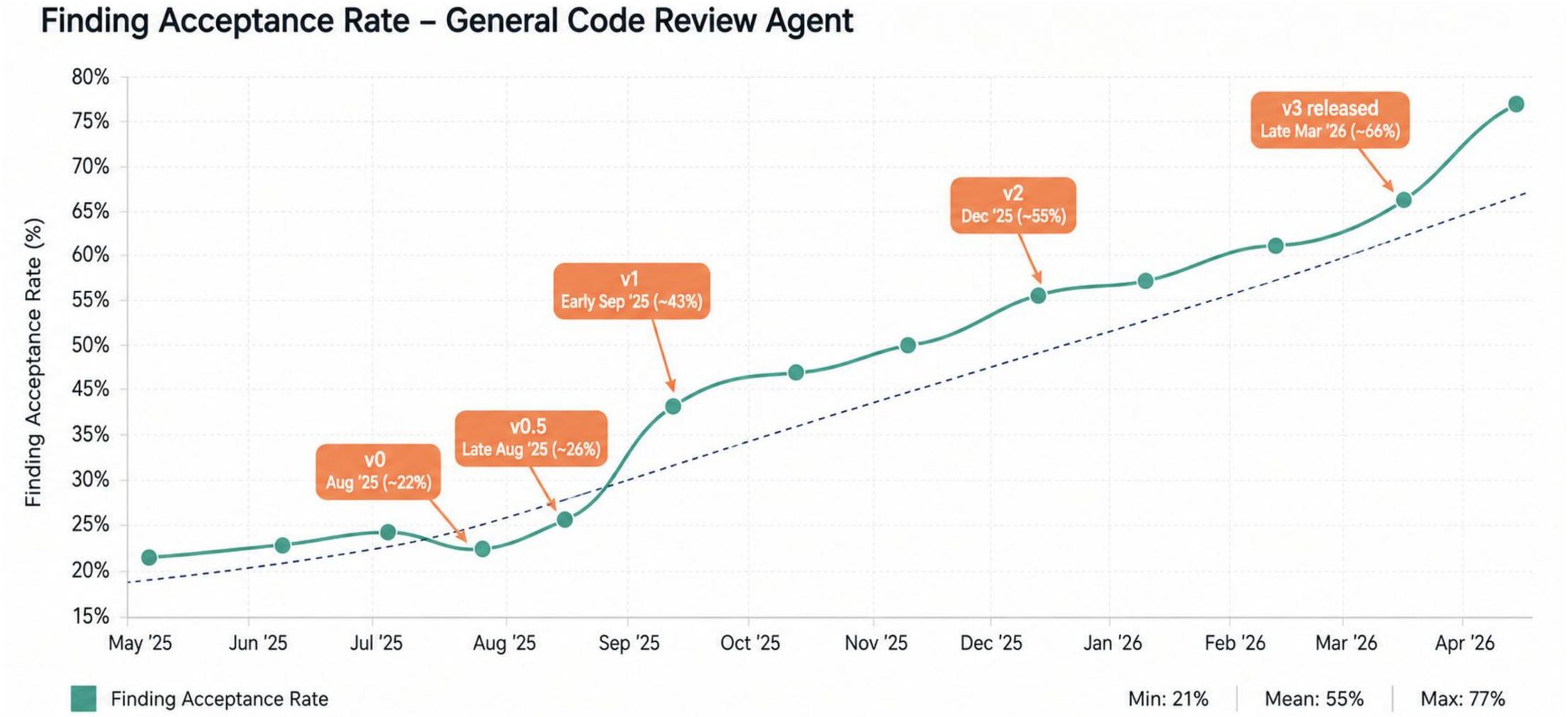
700K PRs (3 yrs). Deep research over historical reviews → exemplar candidates → expert promote / kill. Expert "no"s become training signal, not tribal knowledge.

THE INTERFACE

A systems based approach to curating deep cultural knowledge.



21% → 65%+ accepted findings, weekly.



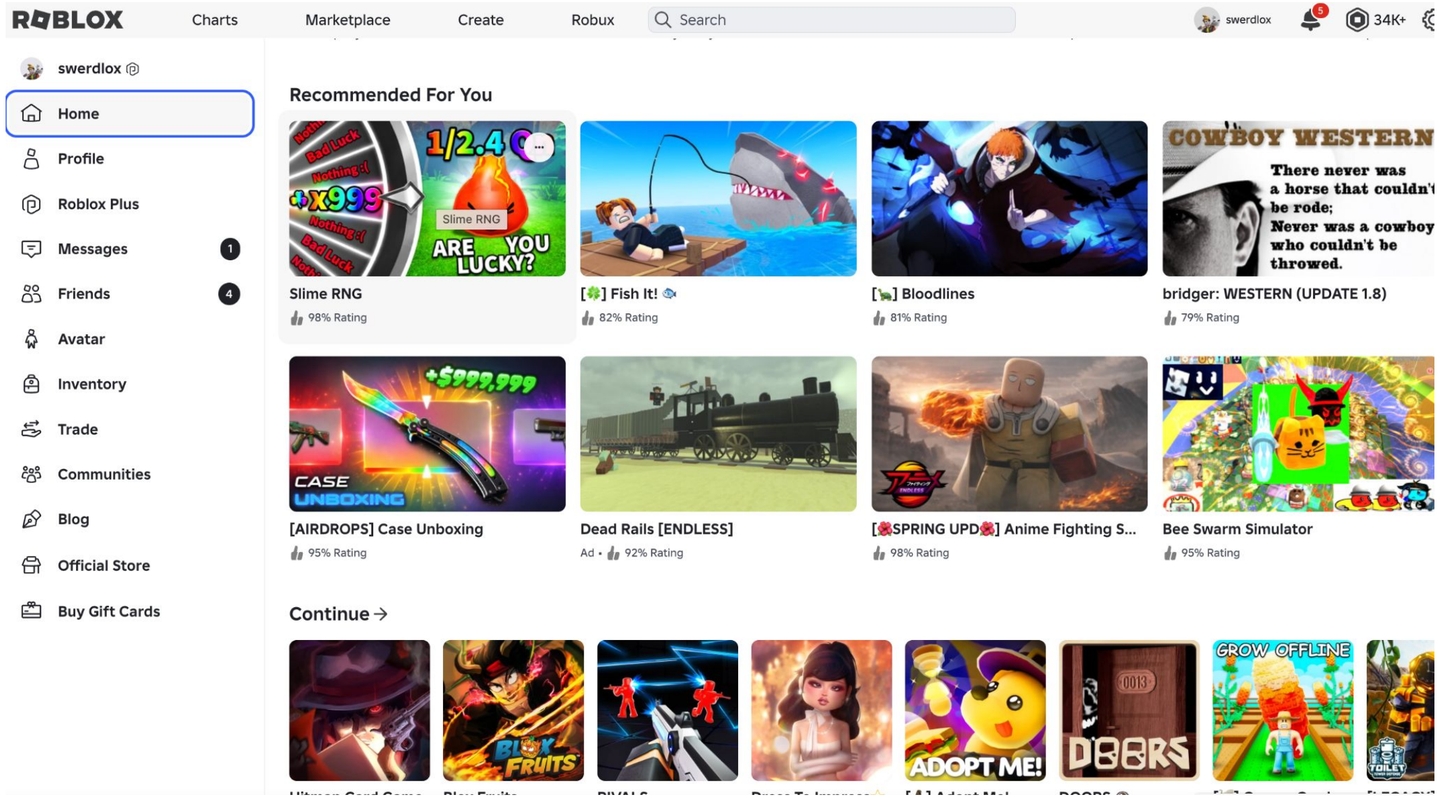
87% LGTM recall . Human review acceptance ≈ 55%... the agent now exceeds it.

TAKEAWAY · 02

Orchestration & Access

Architecting an agentic SDLC that runs on its own — prompt to production.

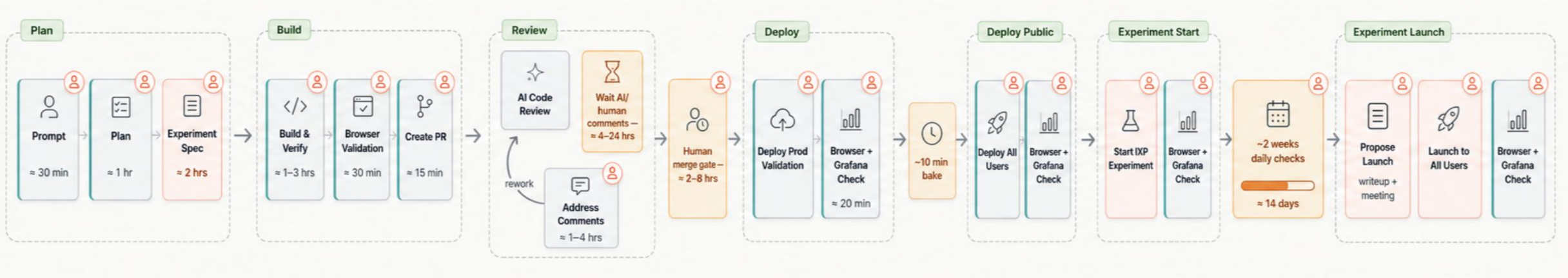
Vision: Prompt a change on the Roblox home page, all the way to production!



Roblox Homepage 144M Daily Active Users

THE AFTER STATE

Old workflow were cumbersome even for humans



TOIL TALLY

≈ 18 human touchpoints

≈ 3 weeks elapsed

≈ 1–2 days active engineer time

THE AFTER STATE

Quickly found the gaps

We needed to make sure that Agents had:

Programmatic Access

Many tools were inaccessible to Agents and needs MCPs, CLIs, or APIs.

Safe Usage

We needed safe runtimes and more validation like tests and observability used by the Agents.

Policy Enabled

We had to update policies like Agent access to mutate tools, and change production autonomously.

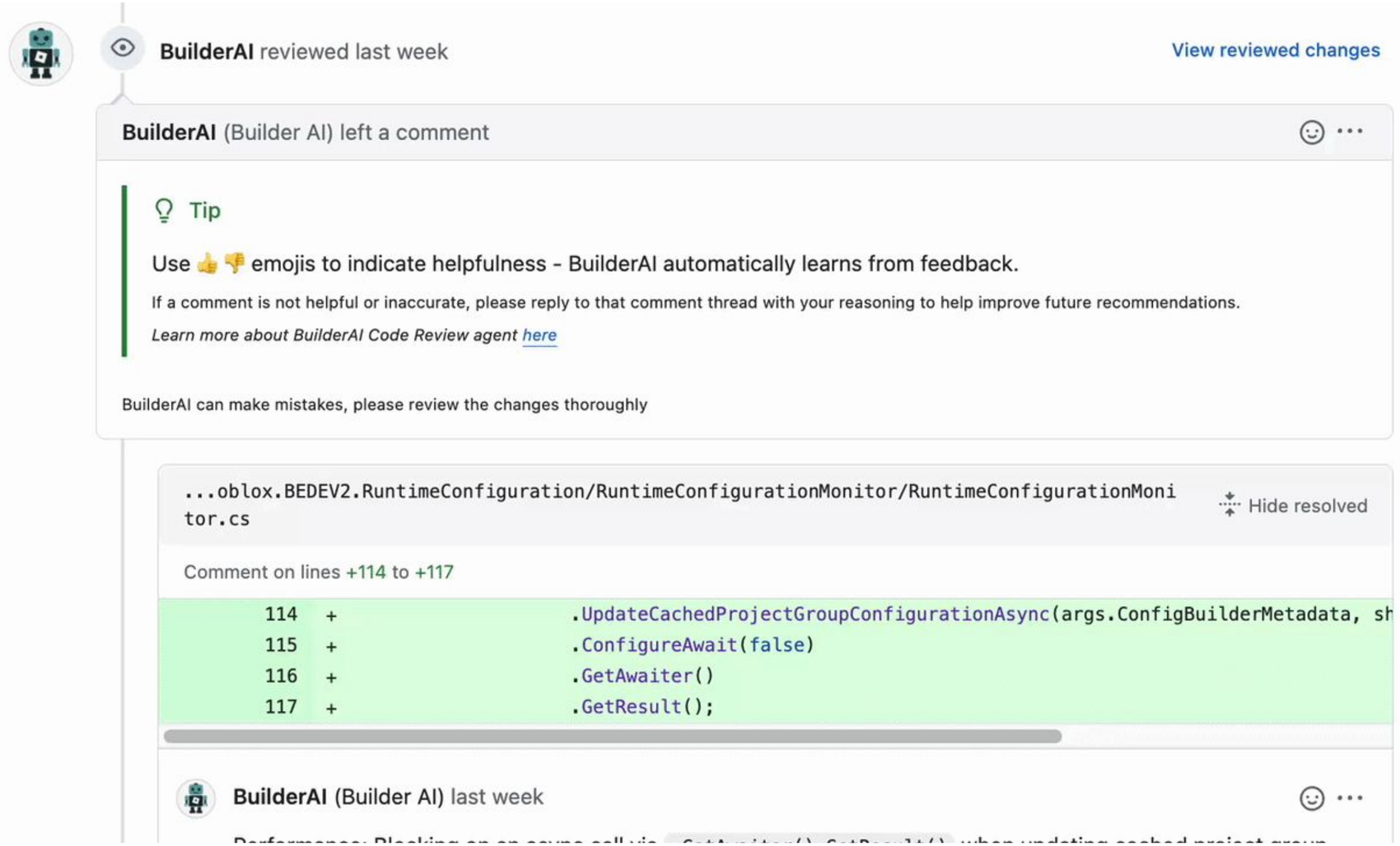
TOOLING UNLOCKED

Programmatic access to every system the workflow touches.

TOOL · 01	Experimentation CLI · start · stop · ramp · fetch results (replaces portal clicks)
TOOL · 02	Web-app deploy CLI · push frontend bundles to staging and production from an agent
TOOL · 03	Browser-automation CLI · headless UI validation inside the agent's verification loop
TOOL · 04	Telemetry CLI · query production metrics during rollout to confirm no regressions
TOOL · 05	Pull-request CLI · open PRs, attach artifacts, iterate on review-agent feedback

Agents:

- Write
- Review
- Fix
- Approve
- Deploy
- Iterate



TAKEAWAY · 03

Measuring What Matters

Productivity metrics for an Agentic native organization.

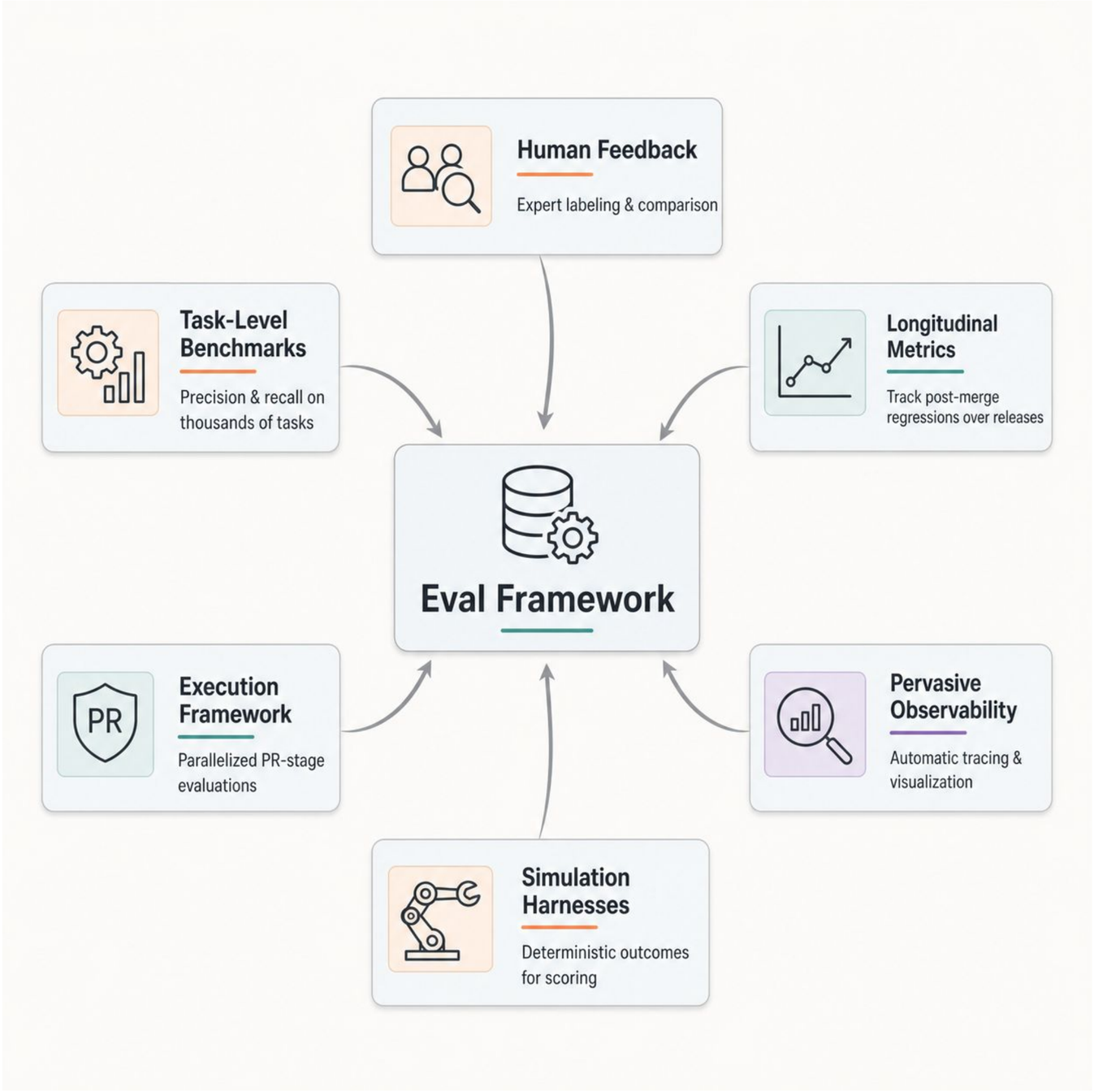
OUTPUT VS. OUTCOME

Most teams measure the wrong thing.

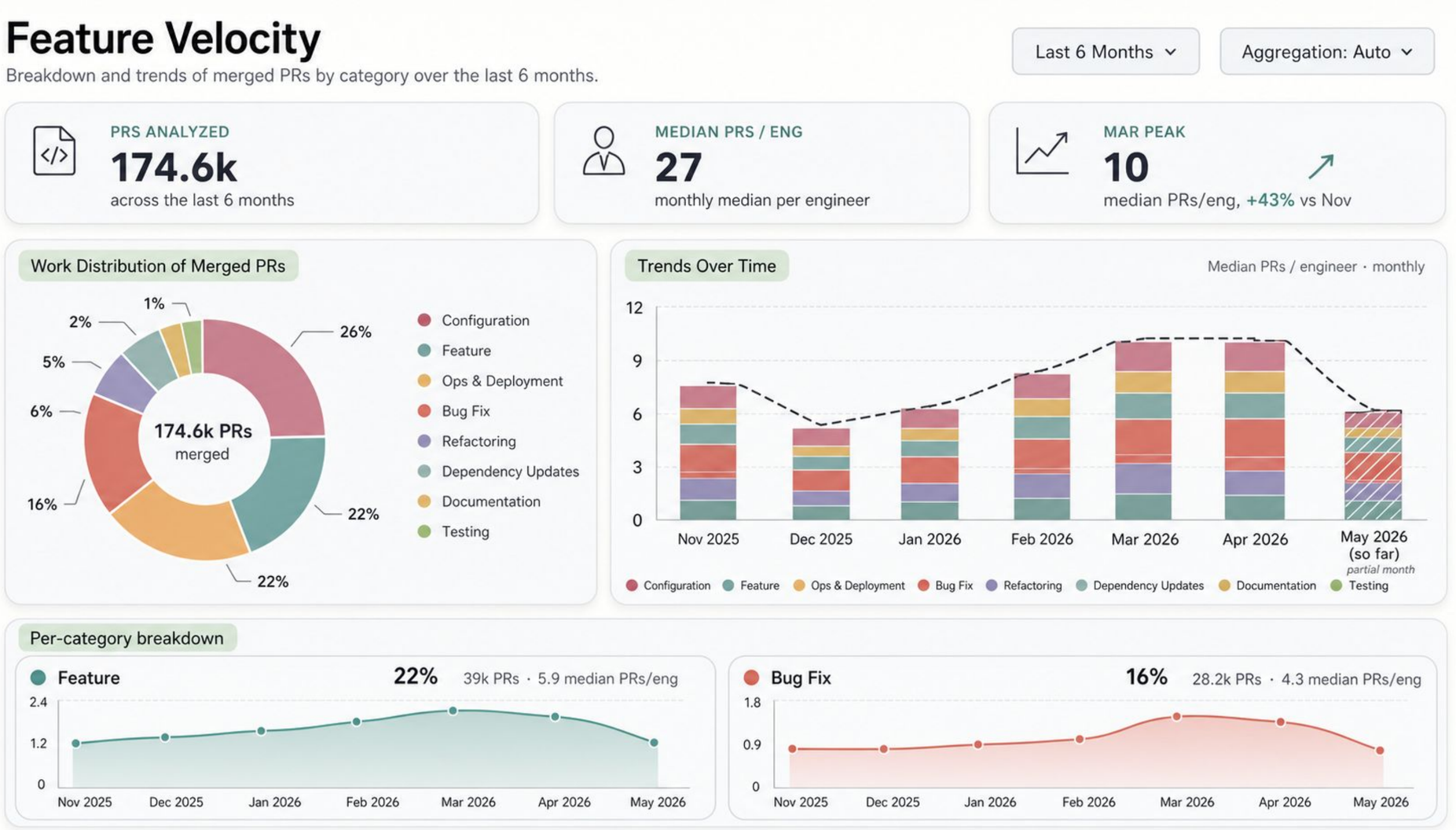
VANITY · OUTPUT	SIGNAL · OUTCOME
~~Tokens · completions · suggestions~~	**Agent quality** Are the agents producing high quality results
~~Lines of code generated~~	**Rollback & incident rate** the cost of being wrong
~~PRs opened~~	**Feature velocity** delivered behavior, not authored text
~~Engineers "using AI"~~	**Autonomous-completion %** work done without a human in the diff

Measure what ships — not what is *generated*.

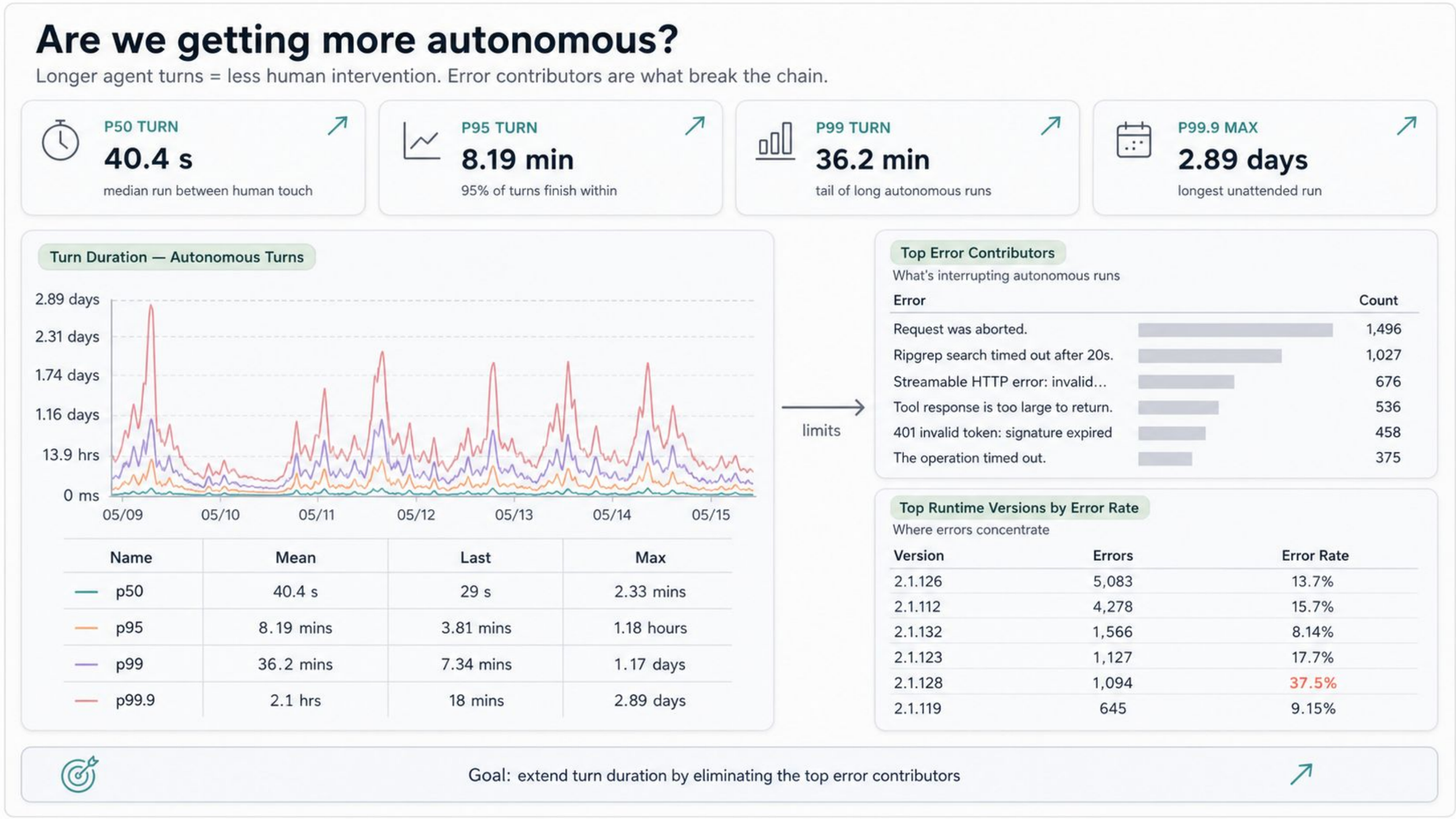
Agent quality matters.



Feature Velocity



Agent Autonomy



WHAT I HOPE YOU'LL WALK AWAY WITH

Three takeaways on the way to 24x7 AI

TAKEAWAY · 01

Alignment & Guardrails

Encoding expert judgment as operational systems.

TAKEAWAY · 02

Security & Access

Architecting an agentic SDLC that securely runs on its own

TAKEAWAY · 03

Measuring What Matters

Productivity metrics for an Agentic native organization.

INSTRUMENT · CURATE · GOVERN · TRUST

Thank you.

SPEAKER

Andrew Swerdlow

Sr. Director · Roblox

FIND ME

linkedin.com/in/andrewswerdlow

andrewswerdlow.net

BOOK

Tech Leadership:

The Blueprint for Evolving from Individual Contributor to Tech Leader