



# evalops

Press Kit — February 2026

*The future of  
AI engineering.*

evalops.dev  
press@evalops.dev  
github.com/evalops

# Infrastructure for AI engineering.

---

EvalOps builds the infrastructure layer for teams shipping production AI systems. We provide replayable execution graphs, transparent traces, and evaluation systems that make AI engineering rigorous.

FOUNDED  
**2025**

FOCUS  
**AI Infra**

As AI moves into production, teams face critical challenges around reproducibility and debugging—especially when customers and auditors demand evidence that systems work as intended.

PRODUCTS  
**3**

Composer generates deterministically. Conductor proves what happened. Score decides if it ships. Together, they form a pipeline that makes AI engineering possible at scale.

*“The model is the easy part. The hard part is knowing whether it works.”*

# Composer

**Specs to PRs. Auto-evaluated.**

Converts specifications into working code via replayable execution graphs. Same inputs, same outputs. Every run can be recorded and replayed for debugging or audit.

# Conductor

**Browser workflows with traces.**

Browser automation with complete execution traces. Every action logged with DOM snapshots, element selection reasoning, and screenshots.

# Score

**Baselines. Diffs. Quality gates.**

Evaluation as infrastructure. Produces versioned baselines, diff reports, gating policies, and audit trails. Ships with evidence, not hope.

# Logo



evalops

The logomark is three horizontal bars forming an abstract "E"—representing evaluation layers and structured execution. Always pair with the wordmark.

## Usage

**Do:** Use on light or dark backgrounds. Maintain 0.5in clear space. Minimum size 0.75in wide.

**Don't:** Recolor the bars. Separate mark from wordmark. Rotate or distort. Add effects.

# Colors



Foreground  
#0C0A09



Accent  
#C2410C



Muted  
#57534E



Background  
#FAFAF9

## Short — 50 words

EvalOps builds infrastructure for AI engineering. Composer generates code via replayable execution graphs. Conductor proves what happened with full traces. Score decides if it ships. Together, they form a pipeline for shipping production AI with evidence.

## Long — 100 words

EvalOps is building the infrastructure layer for AI engineering. As organizations move AI into production, they face critical challenges around reproducibility and debugging—especially when customers and auditors demand evidence that systems work as intended.

EvalOps addresses this with three products that form a pipeline: Composer generates code via replayable execution graphs; Conductor captures browser automation with DOM snapshots, screenshots, and reasoning traces; Score produces versioned baselines, diff reports, and gating policies. Ships with evidence, not hope.



# evalops

Compose. Conduct. Score.

[evalops.dev](https://evalops.dev)

© 2026 EvalOps. All rights reserved.