

RUNTIME BEHAVIOR CONTROL

Steer

Adjust model behavior in real-time without retraining. Steering vectors modify activations directly, changing outputs with precision.

Behavior control shouldn't require retraining

When model outputs aren't right, your options are limited. Fine-tuning is expensive and slow. Prompt engineering is fragile and breaks with model updates. You need a way to adjust behavior in production, immediately.

<5ms

Steering overhead

50+

Pre-trained vectors

0

Retraining required

Instant changes

Modify behavior via API call, no deployment needed

Composable

Layer multiple vectors for nuanced control

Model-agnostic

Works with any LLM provider or open source

A/B testing

Built-in experimentation framework

BUILT ON

rotalabs-steer open source toolkit. Peer-reviewed research on activation steering.

INTEGRATES WITH

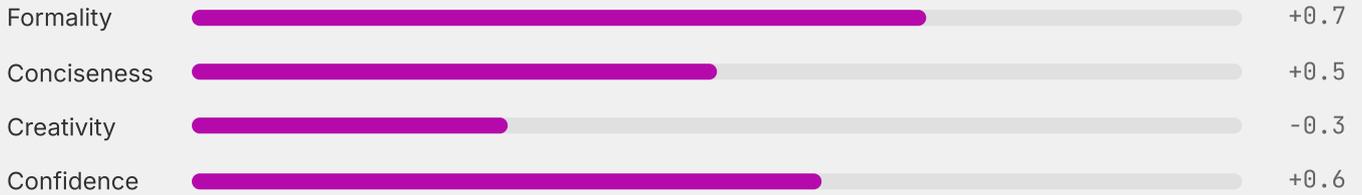
OpenAI, Anthropic, Cohere, Llama, Mistral, and custom fine-tuned models

01 - HOW STEERING WORKS

Modify activations, not weights

Steering vectors operate on model activations during inference. No retraining, no fine-tuning, no weight modification. Just precise behavioral adjustments in real-time.

Example: Adjusting Output Style



Pre-trained Vector Library

50+ ready-to-use vectors for common behaviors: formality, conciseness, helpfulness, creativity, assertiveness, and more. Start controlling behavior immediately.

Custom Vector Training

Train custom steering vectors for your specific use cases. Provide contrastive examples of desired vs. undesired behavior, and Steer learns the activation direction.

Composable Vectors

Combine multiple steering vectors for nuanced control. Layer formality with conciseness, or creativity with accuracy. Adjust strength of each independently.

Real-time Adjustment

Change steering parameters via API without redeployment. Adjust vector strength on the fly. Respond to user feedback or business requirements instantly.

Research Foundation: Activation Steering

Steering vectors work by adding direction vectors to model activations at specific layers. Research shows this can reliably modify behavior without degrading overall model quality.

Open source: [rotalabs-steer](https://github.com/rota-labs/steer) toolkit available at rotalabs.ai. Train your own vectors, verify our methods.

02 - IMPLEMENTATION

From integration to behavior change

Steer integrates with your existing inference pipeline. Add steering in minutes, not days.

01

Select Vectors

Choose from the pre-trained vector library or train custom vectors for your specific behaviors. Browse vectors by category: tone, style, safety, domain-specific.

02

Integrate API

Add Steer to your inference pipeline with a few lines of code. Python SDK, JavaScript SDK, or REST API. Works with any model provider - just wrap your existing calls.

03

Experiment

A/B test different vectors and strengths using the built-in experimentation framework. Measure impact on quality metrics, user satisfaction, and business outcomes.

04

Deploy & Iterate

Roll out behavior changes to production instantly. No deployment needed - just API configuration. Adjust steering parameters on the fly as you learn what works.

03 - USE CASES

Where Steer delivers value

CUSTOMER SERVICE

Tone Calibration

Adjust chatbot formality based on customer segment. Premium customers get more formal tone; casual brands get friendlier responses. Change instantly without retraining.

CONTENT GENERATION

Brand Voice Control

Ensure generated content matches brand guidelines. Steer toward or away from specific writing styles, vocabulary patterns, and tonal characteristics.

LEGAL/COMPLIANCE

Hedging & Certainty

Control confidence levels in AI responses. Legal contexts need hedged language; some contexts need assertive responses. Adjust on a per-request basis.

SALES/MARKETING

Persuasion Tuning

Dial persuasiveness up or down based on context. More persuasive for sales outreach, more neutral for support queries. Same model, different behaviors.

04 - TECHNICAL SPECIFICATIONS

Enterprise-ready architecture

CAPABILITY	SPECIFICATION
Steering Latency	<5ms overhead on inference calls
Supported Models	OpenAI, Anthropic, Cohere, Llama, Mistral, custom fine-tuned
Vector Library	50+ pre-trained vectors across tone, style, safety, domain
Custom Training	Train vectors from 10-100 contrastive examples
Integration	REST API, Python SDK, JavaScript SDK
Deployment	SaaS, private cloud (AWS, GCP, Azure), on-premise
A/B Testing	Built-in experimentation framework with statistical significance

05 - PRICING

Plans for every scale

<p>Dev</p> <p>\$200/mo</p> <p>50K inferences, API access, 10 vectors. For teams experimenting with behavior control.</p>	<p>Pro</p> <p>\$1,000/mo</p> <p>500K inferences, full vector library, A/B testing, custom vectors. For production apps.</p>	<p>Enterprise</p> <p>Custom</p> <p>Unlimited inferences, on-premise deployment, custom vectors, dedicated support.</p>
---	--	---

Pricing is indicative. Contact us for volume discounts and custom requirements.

See Steer in action

Interactive demo showing real-time behavior modification.

[Request Demo](#)[Learn More](#)