Governed Fine-Tuning With Verifiable Provenance

by Nick Clark | Published March 27, 2026 | PDF

Fine-tuning adapts a base model to specific tasks using specialized data. Governed fine-tuning extends the full training governance framework to fine-tuning operations, producing cryptographically verifiable lineage that links every parameter change in the fine-tuned model back to the specific training examples that caused it, through which governance policies they were admitted.

What It Is

Governed fine-tuning applies the complete training governance framework, including admissibility evaluation, depth profiling, gradient routing, and memorization detection, to the fine-tuning process. Additionally, it produces a verifiable provenance record that cryptographically links fine-tuned parameter changes to their source training examples.

The provenance record enables any party to verify what data contributed to a fine-tuned model's behavior and under what governance policies.

Why It Matters

Fine-tuning is often performed on sensitive, proprietary, or rights-governed data. Without provenance, there is no way to verify that a fine-tuned model was trained only on authorized data or that training governance was properly applied. Verifiable provenance provides this assurance, enabling trust in fine-tuned models by making their training process auditable.

How It Works

During fine-tuning, the governance framework records each training example's admission, depth profile assignment, gradient routing configuration, and resulting parameter changes. These records are cryptographically linked to form a verifiable chain from training corpus to model parameters.

A verification protocol allows any authorized party to check that the chain is complete and consistent: every parameter change is attributable to an admitted training example, and every admission complies with the applicable governance policy.

What It Enables

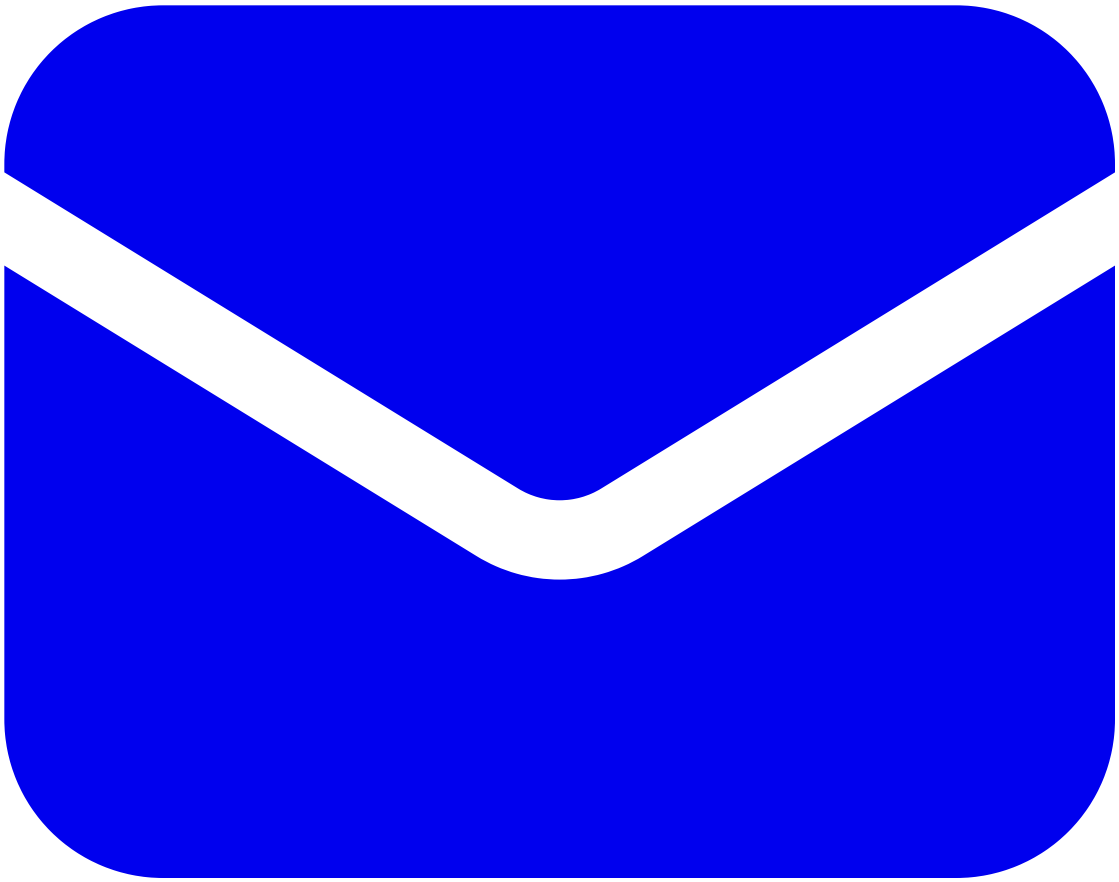
Verifiable fine-tuning provenance enables a market for trusted fine-tuned models. Model consumers can verify that fine-tuning was performed responsibly, on authorized data, under appropriate governance. This is essential for regulated industries where model provenance is a compliance requirement and for rights-governed domains where training corpus authorization must be demonstrable.

Training Governance All 21 steps →

Govern what the model learns, at what depth, with what provenance.

Primary Technical Disclosure
○ Depth-Selective Training Governance for Machine Learning Systems
Secondary Technical
○ Training Examples as Proposed Semantic Mutations○ Entropy-Band-Indexed Training Depth Profiles○ Depth-Selective Gradient Routing for Governed Training○ Training-Level Memorization Detection○ Differential Privacy Through Depth-Selective Routing● Governed Fine-Tuning With Verifiable Provenance○ The Training Loop as a Governed Execution Environment○ Policy-Governed Knowledge Retention and Suppression○ Provenance-Traceable Training Dynamics○ Curriculum-Integrated Depth Scheduling○ Affect-Modulated Training Depth○ Training-Inference Governance Integration○ Training Governance for Human-Relatable Agents
Applications (General)
○ Rights-Compliant Model Training Through Depth-Selective Routing○ Regulated Industry Model Governance With Provenance○ Training Governance for Medical AI○ Training Governance for Legal AI○ Training Governance for Financial Model Training○ Training Governance for Defense AI○ Training Governance for Educational AI Models○ Training Governance for Creative AI
Applications (Specific)
○ OpenAI's Training Pipeline Has No Depth-Selective Governance○ Constitutional AI Training Lacks Depth-Selective Control○ Stable Diffusion's Training Has No Provenance Layer○ Midjourney Trains Aesthetics Without Governed Depth○ Scale AI Labels Data Without Governing What Models Learn○ Labelbox Manages Annotation Workflows, Not Learning Dynamics○ Snorkel AI Programs Labels but Does Not Govern Gradient Depth○ Weights & Biases Tracks Experiments, Not Learning Governance○ Determined AI Orchestrates Compute, Not Learning Depth○ MosaicML Optimizes Training Efficiency, Not Learning Governance
Training Governance overview →
AQ
deterministic
autonomy

Last updated: 2026-03-03



-
- nick@qu3ry.net
- 72 28 14 36 01

Invented by Nick Clark | Founding Investors: Devin Wilkie