

Clove: Failure-Aware Drug Research ML Pipeline Demo

This document describes a realistic, infrastructure-focused demonstration of Clove applied to drug discovery and pharmaceutical research workflows. The demo intentionally avoids clinical decision-making and instead focuses on data processing, machine learning experimentation, and reproducibility — core requirements in modern drug research.

Context: Drug Research Pipelines

Drug discovery workflows involve long-running, failure-prone computational pipelines such as molecule featurization, model training, hyperparameter sweeps, and validation. These pipelines frequently fail due to resource exhaustion, unstable code, or invalid inputs, yet must remain fully auditable and reproducible for regulatory and research integrity reasons.

Demo Objective

Demonstrate how Clove can be used as a research-grade runtime to execute drug discovery machine learning pipelines with strong guarantees around isolation, failure handling, artifact preservation, and experiment traceability.

Pipeline Workflow

- Load Molecular Dataset (public / synthetic)
- Molecule Featurization (e.g., fingerprints, graphs)
- Train Predictive Model (activity / toxicity proxy)
- Model Evaluation & Validation
- Generate Research Report
- Archive Models and Artifacts

Execution Model

Each pipeline stage is executed as an isolated OS process supervised by the Clove runtime. Processes are constrained by CPU, memory, and runtime limits. The runtime monitors execution, captures exit signals, and enforces retry or fallback policies when failures occur.

Failure Scenarios

The demo intentionally introduces realistic failures such as memory exhaustion during featurization or training instability due to malformed molecular inputs. Clove detects these failures, safely terminates the affected process, and retries execution with adjusted parameters while preserving intermediate artifacts.

Auditability and Reproducibility

Each experiment produces a complete audit trail including dataset identifiers, feature extraction parameters, model configurations, random seeds, evaluation metrics, and execution logs. This ensures that every result can be independently reproduced and inspected.

Why This Matters for Drug Research

Drug discovery research demands computational rigor, traceability, and robustness. By treating ML pipelines as supervised OS-level workflows, Clove provides a foundation for scalable and trustworthy research infrastructure without making clinical or medical claims.