# 6.7960 Agentic Coding Tutorial

September 22, 2026 and September 24, 2026

# Agenda

1. Getting Set Up: ORCD Open On-Demand
2. Getting Set Up: Parley
3. Codex Walkthrough
4. Agent Harnesses

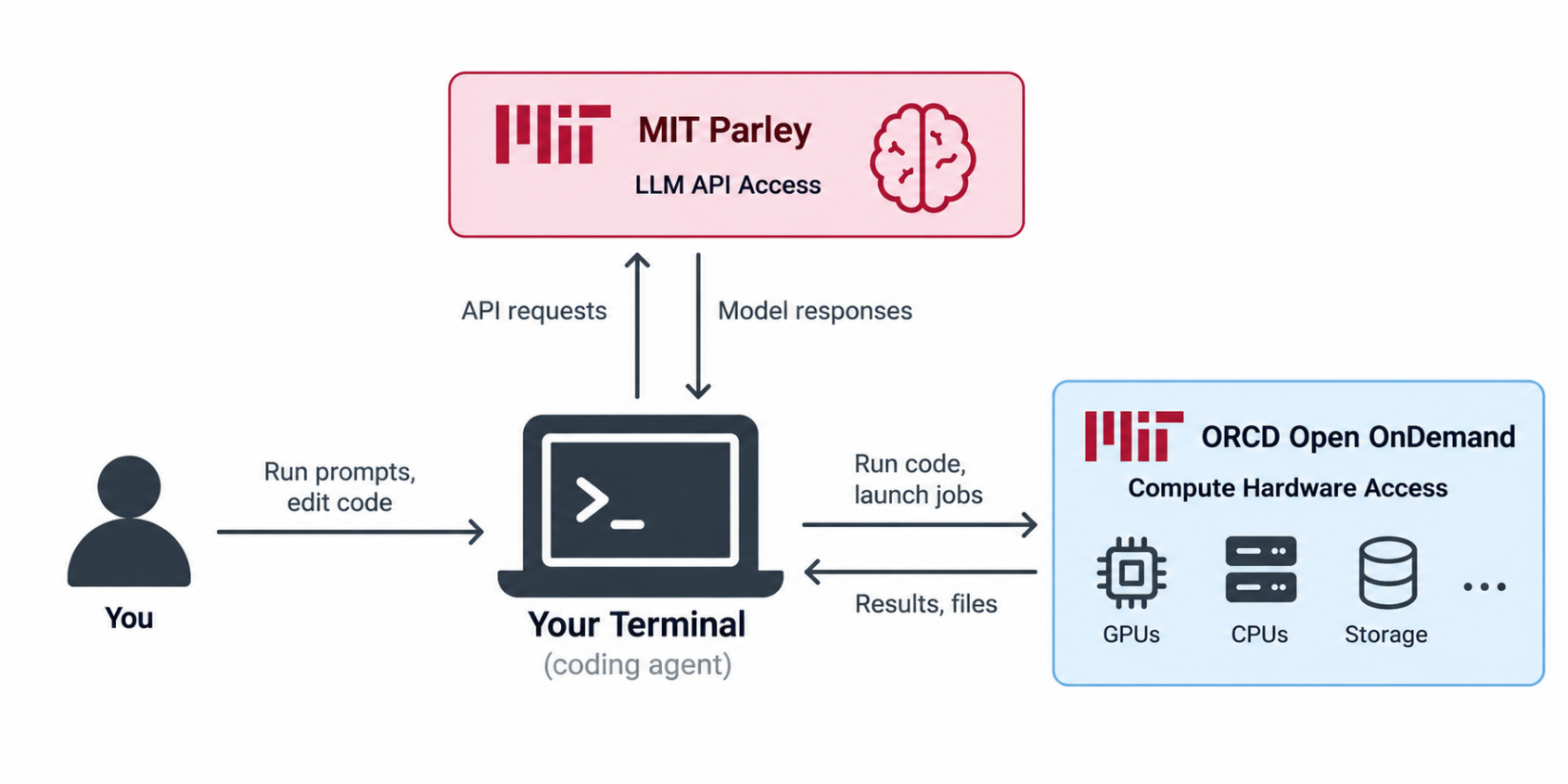
# Getting Set Up: ORCD Open On-Demand

- ORCD Setup Guide
- What is it?
  - ORCD is the Office of Research Computing and Data
  - Engaging is ORCD's shared HPC cluster.
  - The cluster includes CPU nodes, GPUs, and shared storage.
  - Users can access Engaging through SSH or Open OnDemand.
- Usage: GPU/Compute Hardware Access - ML training, model inference, large-scale experiments

# Getting Set Up: Parley

- Parley Setup Guide
- What is it?
  - Provides access to LLMs through an MIT-managed interface and API.
  - It can be used as the model backend for agentic coding tools such as Claude Code and Codex CLI.
- Usage: LLM API Access - agentic coding, code generation, reasoning, model inference

# Workflow


MIT Parley
LLM API Access
API requests
Model responses
Run prompts,
edit code
You
Your Terminal
(coding agent)
Run code,
launch jobs
Results, files
ORCD Open OnDemand
Compute Hardware Access
GPUs
CPUs
Storage
...

# Codex Walkthrough

- Inspect an existing codebase (Github)
- Set up context
- Implement with coding agent
- Run experiment with coding agent
- Review changes with `git diff`, `\code-review`

# Can a neural network learn a boundary a line cannot?

**Task:** Classify points on an inner circle versus an outer ring.
**Data:** 384 training points; 128 held-out test points.

**Linear classifier**

$$f_{\text{linear}}(x) = w^\top x + b$$

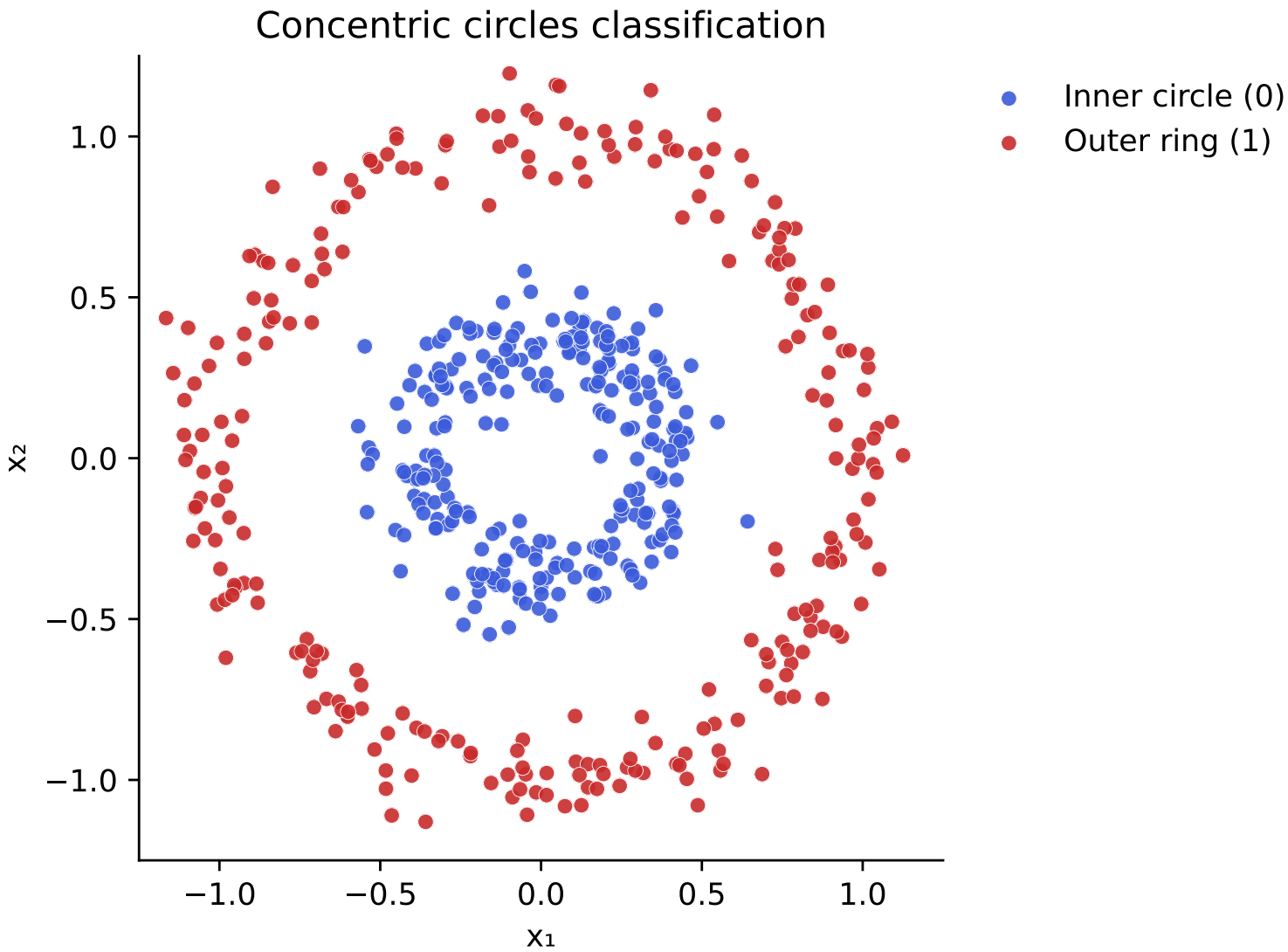
**Two-layer MLP**

$$h = \mathrm{ReLU}(W_1 x + b_1)$$
$$f_{\text{MLP}}(x) = W_2 h + b_2$$

$$W_1 \in \mathbb{R}^{16\times 2}, \quad W_2 \in \mathbb{R}^{1\times 16}.$$

**Experiment:** Compare test accuracy and learned decision boundaries.

# Experiment Setup



# Context Engineering

| | |
|---|---|
| `AGENTS.md` | How to work in this repository |
| Skills (`SKILL.md`) | How to perform recurring tasks |
| `EXPERIMENT.md` | What we are currently investigating |
| Prompt | What to do right now |

# Demo

- Question / hypothesis - complete
- Context Setup - complete
- Code Implementation
- Experiment Launch
- Conclusion / Documentation

**Demo: agent runs and documents an experiment**

# Extensions

- Agent harnesses: sandboxes, tool execution, context management, retries, multi-agent coordination
- Experiment loops: modify code, launch runs, inspect metrics, iterate, record results
- Evaluation loops: tests, benchmarks, regression checks, success criteria
- Parallelization: multiple agents, hyperparameter sweeps, candidate implementations, ablations
- Tool extensions: Git/GitHub, Slurm, documentation search, experiment tracking, external APIs