PROJECTS

# Agentry

## A Native macOS Control Plane for Local AI

Lead Architect · 2026

*Agentry is a native macOS application that runs as a persistent local inference server, hosting multiple MLX models and agents at once. Each agent is an independently addressable endpoint with its own API key, system prompt and MCP tool access — a server-management dashboard for local AI, not a chat app.*

---

## 1. Overview

Local inference tools force a tradeoff: LM Studio is polished but single-model and single-chat; Ollama runs many models but has no real GUI, no per-model auth and no native feel; MLX directly means writing Python. Agentry is the missing control-plane paradigm — run multiple MLX models concurrently on one Mac, expose each as an authenticated OpenAI-compatible endpoint, and manage it all from a dashboard built in the spirit of Server.app, not ChatGPT.

## 2. Key Features

- **Concurrent MLX Models** Run multiple MLX models on a single Apple-Silicon Mac at once, with per-model lifecycle controls — load, unload, hot-swap and memory budgets.
- **Agents as Endpoints** An agent = base model + system prompt + MCP servers + key + metadata, independently addressable from any tool that speaks OpenAI-compatible HTTP.
- **Per-Agent Authentication** Each agent gets its own API key with one-click creation and copy-to-clipboard — no shared, unauthenticated localhost endpoint.
- **Native Server Dashboard** A macOS-native dashboard surfaces server health, active agents, request volume and per-agent logs at a glance.

- **Always-On Background Service** Runs as a LaunchAgent that survives login, with a menu bar item for status and quick controls.

## 3. Architecture

Agentry is a native Swift macOS app that hosts a persistent local inference server over Apple's MLX framework, running multiple models concurrently and exposing each agent as an authenticated, OpenAI-compatible HTTP endpoint. It is designed to be consumed by external clients — Claude Code wrappers, agentic coding tools, n8n, chat apps and custom shell aliases — rather than chatted with directly. It ships as a background LaunchAgent with a menu bar item, Apple-Silicon-only, MLX-only in v1.

---

TECH STACK

Swift · SwiftUI · MLX · Apple Silicon · OpenAI-compatible HTTP · MCP · LaunchAgent

LINKS

https://www.jonathandumitru.com/projects/agentry