Applied AI Portfolio: Independent R&D

Tyler Stahl - tylerstahl@gmail.com - 206.402.1346

Overview

This document serves as a living repository of technical briefs for primary projects completed during my "AI Sabbatical." These projects focus on bridging advanced LLM reasoning with production-ready infrastructure, emphasizing hybrid inference routing, durable memory, and human-in-the-loop (HITL) system design.

Project Name	Core Technology	Key Focus
Market Intelligence Analysts	OpenClaw, Python, SQLite	Multi-agent synthesis and reporting on real world and World of Warcraft commodity markets
Hybrid Inference Pilot	Gemini 2.5 Pro, Gemma4:26b/e2b (Local), Ollama, Antigravity	Cost/Latency optimization
iBlueprint Embeddings	Vector DBs, iBlueprint.ai	Data embedding stress-testing
AgentOS: Chief of Staff Agent	Cursor Agent SDK, Python	Autonomous Workflow Orchestration
Cognitive Bridge	Firebase, Google AI Studio, Gemini	AI Personality & User Alignment (OCEAN)

1. OpenClaw Multi-Agent Ecosystem

An autonomous team designed to crawl, parse, and structure market intelligence. Addresses the primary challenge of handling long-running agentic tasks without losing state or context.

- **Durable Memory:** Implemented a SQLite persistence layer allowing agents to resume interrupted tasks and maintain a multi-turn history over days of research.
- **Autonomous ETL:** Python-based workers that transform unstructured web data into structured relational formats with high schema fidelity.

- **Governance:** Integrated HITL (Human-in-the-Loop) terminal gates to ensure high-stakes data is validated before database commits.

2. Tri-Tier Hybrid Inference Strategy

A routing system designed to solve the "hallucination vs. latency" tradeoff. It intelligently distributes tasks between heavy cloud models and high-speed local models.

- **Tier 1 (Reasoning):** Routes complex planning and cross-referencing to **Gemini 2.5 Pro**.
- **Tier 2 (Verification):** Utilizes local **Gemma4:26b** for consistency checks, keeping validation logic on-premise for privacy.
- **Tier 3 (Formatting):** Offloads high-velocity JSON schema parsing to **Gemma4:e2b** to minimize operational costs and latency.

3. iBlueprint Collaboration (In-Progress)

Active collaboration with Stephen Rockwell (Humanservices.ai) focused on the technical implementation of vector data infrastructure.

- **Embedding Optimization:** Stress-testing different embedding model implementations to determine accuracy thresholds for domain-specific data.
- **Vector Persistence:** Evaluating the integration of vector stores within existing blueprinting workflows to improve RAG (Retrieval-Augmented Generation) performance.

4. AgentOS: Chief of Staff Agent

A multi-agent autonomous workflow orchestrator designed to act as a "Chief of Staff" by managing complex, long-running tasks, and developing anticipatory intuition.

- **Autonomous Orchestration:** Built with the Cursor Agent SDK, managing subagents for different functions (e.g., feature evaluation, testing).
- **HITL Workflow:** Implemented mandatory "try it out handoffs" to ensure human validation and approval at critical stages.
- **Modular Design:** Architecture allows for the splitting of evaluation and testing subagents to optimize for continuous integration and deployment workflows.

5. Cognitive Bridge: AI Personality Alignment

An experimental tool to align AI agent personality with a user's personality based on the OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) model. The goal is to optimize human-AI collaboration by matching personality traits.

- **Personality Modeling:** Uses a brief user interview to generate a personality map.
- **Alignment Engine:** Built in Google AI Studio to generate aligned and unaligned chat responses for demonstration and evaluation.
- **Research Focus:** Visual demonstration of the impact of sub/optimal personality trait matching, drawing on academic research.

- ## Architectural Pivot: Prototyping to Production
-
- ### Core Technologies
- * **Prototyping Framework:** Google AI Studio (Prompt engineering, system instruction validation, parameter tuning)
- * **Production Backend Runtime:** Python 3.11+ via Firebase Cloud Functions (v2)
- * **Durable Orchestration Layer:** Vertex AI for Firebase Python SDK
- * **State & Memory Management:** Cloud Firestore (NoSQL Document Store)
- * **Security & Gateway:** Firebase Authentication with secure Environment Variables
-
- ### Enterprise Portfolio Talking Points
- * **Productionized AI Prototypes:** Successfully migrated an ephemeral behavioral analysis prototype out of Google AI Studio into a scalable, production-ready serverless Python backend architecture.
- * **Secured LLM Data Pipelines:** Designed and deployed serverless cloud endpoints wrapped in explicit user authentication frameworks, ensuring user-scoped data privacy and protecting upstream Gemini API tokens from client-side exposure.
- * **Enforced Deterministic Responses:** Utilized `Pydantic` runtime data validation alongside Gemini's structured output configurations (`response_mime_type="application/json"`) to guarantee type-safe JSON returns, seamlessly driving live UI states and database mutations.
- * **Engineered Durable Agent Memory:** Developed a persistent, real-time contextual memory layer within Cloud Firestore, tracking user behavioral trait variations across independent interaction sessions to dynamically update the agent's personalized response parameters.
-