

# GAVIN BELSON

San Francisco, CA | gavin.belson@email.com | (555) 123-4567 | linkedin.com/in/gavinbelson | github.com/gavinbelson

## SUMMARY

---

Agentic AI specialist with 4 years of experience designing and deploying multi-agent systems for enterprise process automation. Proficient in LangGraph orchestration, tool-use architecture, and human-in-the-loop workflow design across GPT-5.5 and Claude 4.6. Reduced manual processing time by 70% across a five-agent document review pipeline handling 10,000+ monthly tasks at a Series B HR technology company.

## SKILLS

---

**Agent Frameworks and Orchestration:** LangGraph, LangChain, AutoGen, CrewAI, custom orchestration with FastAPI

**Models and APIs:** GPT-5.5, Claude 4.6 Sonnet, Gemini 3.1 Pro, function calling, structured outputs, model routing

**Memory and Retrieval:** Pinecone, Weaviate, Chroma, episodic memory design, RAG within agentic pipelines, hybrid search

**Tool Use and Integrations:** Function calling, REST/GraphQL API orchestration, browser use (Playwright), code execution sandboxing, Slack/Email/Calendar integrations

**Observability and Reliability:** LangSmith, Weights & Biases Weave, Arize, agent tracing, human-in-the-loop checkpoints, retry and fallback routing, cost management

**Supporting Technical Skills:** Python, TypeScript, Docker, Kubernetes, AWS (Lambda, ECS, S3), Redis, PostgreSQL, JSON schema design

## EXPERIENCE

---

### Agentic AI Specialist — Aviato (Series B HR Tech) Mar 2023 – Present

*New York, NY*

- Designed and maintained a five-agent recruitment automation system using LangGraph and GPT-5.5, owning full pipeline design from candidate sourcing through interview scheduling for 10,000+ monthly hiring workflows.
- Architected an orchestrator-subagent pattern where a planning agent decomposed each workflow into discrete tasks — resume screening, outreach drafting, calendar coordination, status tracking, and offer prep — delegated to specialized subagents with isolated tool access.
- Implemented a human-in-the-loop checkpoint system using LangSmith for trace monitoring, requiring recruiter approval before any candidate-facing communication, reducing unauthorized outreach incidents to zero across 18 months of production.
- Built short-term memory in Redis and long-term candidate history retrieval in Pinecone, enabling agents to maintain context across multi-session hiring workflows without exceeding GPT-5.5 context window limits.
- Reduced average time-to-interview from 12 days to 4 days, achieving 94% end-to-end task completion rate with a fallback rate under 6%.
- Cut token cost per workflow by 38% through prompt compression, model routing (GPT-5.5 for planning, GPT-5.5-mini for routine subtasks), and aggressive caching of repeated retrievals.
- Mentored two junior engineers onto the team and codified agent design review checklists now used across the org.

### Prompt Engineer (LLM Agent Infrastructure and Tool Use) — Hooli Aug 2021 – Feb 2023

*Palo Alto, CA*

- Designed multi-step prompt chains with output validation and re-prompting loops for an internal contract analysis tool — work that retrospectively constituted single-agent task decomposition with tool grounding.
- Built and maintained the function-calling layer connecting LLM outputs to internal services (CRM, ticketing, document store), establishing a tool schema design pattern adopted by the broader applied AI team.
- Instrumented evaluation pipelines for trajectory accuracy and tool-use correctness, surfacing systematic failures in date-extraction tools that had been masked by happy-path testing.
- Reduced average prompt cost per contract by 45% through prompt compression and selective context retrieval.

## **PROJECTS**

---

**Agent Trajectory Eval Harness — Open Source** 2024 – Present

*[github.com/gavinbelson/traj-eval](https://github.com/gavinbelson/traj-eval)*

- Open-source library for measuring trajectory accuracy and tool-use correctness across LangGraph and AutoGen agents.
- Currently 480+ GitHub stars and used by three external companies; maintainer of issues, PR review, and a monthly release cycle.

## **EDUCATION**

---

**B.S. Computer Science — University of California, Berkeley** May 2021