

Maxwell Sutton

Anaheim, CA • maxwellcsutton@gmail.com • [LinkedIn](#) • [GitHub](#)

Senior SDET (8 yrs) building AI agent systems at work and in open source — LLM test triage, multi-model evaluation pipelines, and local-LLM inference infrastructure.

EXPERIENCE

NBA — New Jersey (Remote) April 2025 – Present

Sr. SDET

- Built the AQA department's AI harness powering automated test creation, code review, documentation generation, and Jira workflow automation
- Designed and built an AI test-triage agent (Claude Opus on AWS Bedrock) that reviews failed CI pipeline runs, diagnoses root cause with a reported confidence level, recommends fixes or drafts bug tickets, and self-updates its own review skills as new failure patterns emerge
- Led TypeScript/Playwright migrations via a custom AI migration workflow that converts legacy tests, auto-selects optimal locator strategies, redesigns tests to new department patterns, and self-validates each migration
- Designed an AI agent that self-maintains AQA documentation; authored the AQA code style guide and upskilled engineers on maintainable TypeScript
- Build and maintain Playwright web/API automation for large-scale services, plus internal tools and CI/CD pipelines

Chipotle — Newport Beach, CA June 2023 – April 2025

Sr. SDET, Performance and Quality Engineering

- Performance tested front-end, back-end, and mobile services; determined resource allocation for Azure Kubernetes containers based on results
- Led TypeScript migrations of automation repositories; redesigned API automation classes for parallel runs, reducing flakiness; initiated the front-end performance framework for Chipotle web

Groundswell — Sr. QA Engineer (React Native automation with Detox) Sept 2022 – Feb 2023

AppOnboard — QA Engineer II (Appium/WebdriverIO + TS API frameworks) April 2021 – Sept 2022

Tapcart — QA Engineer (manual + automated mobile/web testing) June 2020 – April 2021

AppOnboard — QA Engineer (WebdriverIO + Jest web/backend automation) July 2018 – May 2020

OPEN SOURCE & INDEPENDENT PROJECTS

Homunculus — local LLM agent runtime — TypeScript, llama.cpp, Postgres

- Built a continuously running local agent runtime: heartbeat scheduler, multi-lane inference (chat/work/embed/vision as isolated llama-server processes), self-curated tiered memory, and a hardened OpenAI-compatible client
- Designed a warm-base KV caching scheme (byte-stable prompt base + volatile tail with content-hash diffs) saving ~12–14K tokens of prefill per turn; deployed across a 3-machine Tailscale mesh with launchd/NSSM service persistence

agent-validation-harness — agent validation pipeline — Claude, Kimi, Qwen

- Designed an LLM-as-judge pipeline gating empirical model changes: premise-verification agents, blind pre-op/post-op judges with packet-contamination controls, pre-registered predictions with falsification clauses to prevent fitting to data, and a 2-of-2 dual-judge landing rule gating implement/log/reject
- Built cross-family review gates (plans and diffs are always reviewed by a different model family than their author, with provenance-stamped verdicts) plus Claude Code hooks enforcing hard limits on autonomous runs: uncommitted-churn caps, image-read budgets, destructive-command guards
- Built a self-hosted agent console: a private-mesh web app (Tailscale, zero public ingress) for launching and piloting headless Claude/Kimi/Qwen CLI sessions against any local repo from any device

Idle-Tower — open-source game driven by an autonomous AI player — Next.js, tRPC, Prisma

- Built an idle game where a local LLM autonomously manages build optimization via a brain/body repo split over an HTTP tool-calling contract, with byte-stable tool catalogs for cache warmth and deterministic testing

devenv-boilerplate-for-claude-code — agentic dev-environment tooling — Claude Code

- Built an OSS starter kit for safe agentic development: fail-closed PreToolUse permission governor, doc-drift hooks, a memory→skills self-maintenance loop, and unattended-run guardrails (kill-switch, draft-PR-only, scope ceilings)

SKILLS

AI / LLM: Claude API, AWS Bedrock, Claude Code, LLM agent design, LLM-as-judge evaluation, multi-agent orchestration, llama.cpp local inference, KV-cache optimization, prompt engineering

Languages & Automation: TypeScript, JavaScript, SQL • Playwright, WebdriverIO, Cypress, Appium, Detox, Jest, Vitest, Postman

Cloud, Perf & Observability: AWS, Azure, GCP, Kubernetes, Postgres • LoadRunner • Splunk, Datadog, Grafana, JIRA, TestRail