

Allfin Ahsan

Plantation, FL | (571) 842-2187 | allfinahsan@outlook.com | linkedin.com/in/allahsan | github.com/allahsan | cars4sure.com

SUMMARY

AI engineer with production experience across founder, freelance, and enterprise contexts. Architected Cars4Sure end-to-end: a multi-tenant marketplace on PostgreSQL with row-level security, JWT-embedded RBAC across 13 roles, and real-time AI content moderation. Built production agentic systems with RAG, persistent memory, and dual-layer guardrails for client engagements. Day job in Windows systems administration grounds AI work in production discipline: Active Directory, Azure, observability, and zero-trust network segmentation. Security+ certified, BS Computer Science (University of Miami, 2023).

EXPERIENCE

Founder & Engineer

Late 2024 – Present

Cars4Sure

Remote

- Architected and built Cars4Sure (cars4sure.com), a verified-dealer automotive marketplace currently live with onboarded dealer partners and active vehicle inventory
- Designed system architecture: 5-portal product (buyer, dealer, inspector, admin, auth-router microservice) over a shared PostgreSQL data layer with row-level security for multi-tenant isolation, Redis caching tier, and a multi-agent development workflow using specialized Claude subagents coordinated through MCP integrations
- Implemented production engineering practices end-to-end: CI/CD on GitHub Actions, deploys on Vercel, observability and error tracking, iterative security review covering input validation, rate limiting, and OWASP Top 10 categories
- Operate the platform as sole engineer including customer onboarding, dealer verification flows, billing, and cron-driven lifecycle automation

Independent AI Engineer

Oct. 2023 – Present

Self-employed

Remote

- Built and deployed production customer-support agent for advertising agency: RAG retrieval over client knowledge base, Mem0 for persistent conversational memory, dual-layer guardrails screening prompts on input and responses on output for prompt injection and PII exposure, with Anthropic and OpenAI providers routed through LiteLLM for cost optimization and provider failover, currently serving live customer traffic
- Built and deployed preference-aware route optimization platform in daily use by a field sales team: multi-stop itinerary ingestion via Google Places API, Gemini layer for route verification and natural-language reasoning, with user-override signals captured as embeddings and injected into context on subsequent requests for personalization across user, team, and similar-cohort dimensions
- Advised three client engagements on AI adoption strategy: identifying high-leverage automation targets, framing build-vs-buy decisions, and scoping pilots to measurable outcomes

Windows Systems Administrator

Oct. 2023 – Present

Titan America

Deerfield Beach, FL

- Deployed 7 production AI agents across HelpDesk, Sales, Operations, Customer Service, Procurement, Accounting, and Credit using Microsoft Copilot Studio, integrated with SharePoint knowledge bases and Microsoft Teams, with calibrated personality per department, topic and scope guardrails, and human-handoff protocols for out-of-scope queries
- Reduced IT support ticket volume approximately 40% by automating high-frequency flows: hardware-alert ticket creation, certificate renewal via PowerShell (cutting a 30-minute manual process to under 5 minutes), and tier-1 user questions handled by context-aware Copilot agents before reaching the queue
- Built real-time observability stack with Prometheus for metrics collection, InfluxDB for time-series storage, and Grafana dashboards for a 350+ user environment
- Architected Azure Active Directory migration and implemented 5-VLAN network segmentation with zero observed inter-VLAN leakage, enforcing zero-trust principles across user, server, IoT, and management traffic
- Wrote Python and PowerShell automation across user provisioning, account lifecycle, and routine administration, replacing manual click-throughs with idempotent scripts and Group Policy templates

Junior Systems Administrator

Mar. 2021 – Oct. 2023

Innovative IT Solutions, Inc.

Miami, FL

- Built automated deployment pipelines for endpoint provisioning across a 200+ device environment, replacing manual imaging with scripted workflows over Active Directory and Group Policy
- Managed Active Directory at scale with Group Policy automation, standardizing configuration across users and devices and reducing repeat support escalations

PROJECTS

Cars4Sure Marketplace | *Next.js 16, React 19, TypeScript, Supabase, PostgreSQL, Redis, Vercel* | cars4sure.com Late 2024 – Present

- Architected a 5-portal platform (buyer marketplace, dealer portal, inspector portal, admin console, auth-router microservice) sharing a single PostgreSQL database with row-level security enforcing multi-tenant isolation across all portals
- Designed RBAC system with 223 permissions across 13 roles, embedded into JWTs via a custom Supabase Auth hook for zero-latency permission checks at the edge without per-request database round-trips
- Built dual-engine AI content moderation pipeline (OpenAI Moderation API + GPT-4o-mini) running in real-time against listings to detect marketplace fraud patterns, scam schemes, and policy violations before publication
- Implemented 7-phase sale attribution system with cross-portal notifications, buyer confirmation flows, automated billing, and cron-driven lifecycle management
- Engineered Redis caching layer with tiered TTLs, graceful fallback to Postgres on cache miss, and atomic counter operations to eliminate race conditions on shared state
- Ran iterative security-review and remediation cycles using Claude Code as tooling, hardening the platform across input validation, rate limiting, RLS policy correctness, auth-flow edge cases, and OWASP Top 10 categories

Cars4Sure Internal AI Operations Layer | *Python, Node.js, Ollama, Qwen, Microsoft Graph API, Mem0/Qdrant, Cognee/Neo4j*

- Designed a 7-agent autonomous operations system to run Cars4Sure as a solo founder: each agent owning a distinct business function across operations, growth, marketing, strategy, customer support, and DevOps, with cross-agent task delegation
- Built Microsoft 365 integration bridge (Node.js) enabling agents to monitor and respond across Teams, Email, OneDrive, SharePoint, and Calendar via Graph API with delegated ROPC auth flow, plus an email classification engine resolving customer identity across M365 and Supabase
- Architected multi-model inference pipeline with automatic fallback chains across local models (Qwen 14B on Ollama), cloud APIs (Anthropic Claude, OpenAI GPT), and Gemini-grounded web search, optimizing for cost on routine traffic and latency on critical paths
- Implemented multi-tier security model: input/output sanitization, canary tokens for prompt-injection detection, exec allowlists, per-agent clearance levels, and identity-resolved authorization checks before privileged actions
- Deployed memory and knowledge layer with Mem0 (Qdrant vector DB) for semantic facts and Cognee (Neo4j graph DB) for relational knowledge; fine-tuned Qwen 2.5 14B with QLoRA (rank 16) on a curated training set, exported to GGUF for local inference

Production Homelab | *Proxmox, WireGuard, Prometheus, Grafana, InfluxDB* Ongoing

- Multi-VLAN segmented network on Proxmox virtualization with WireGuard VPN for remote access and full observability stack mirroring production patterns; sandbox for threat-hunting experiments and infrastructure trials before production rollout

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, SQL, PowerShell, Bash

Full-Stack & Databases: Next.js, React, Tailwind CSS, Supabase, PostgreSQL (row-level security, real-time subscriptions, edge functions), Redis, Neo4j, Qdrant, REST APIs, Vercel

AI & LLM Engineering: Multi-agent systems, MCP (Model Context Protocol), Claude API, OpenAI API, Gemini API, LiteLLM, Ollama, Microsoft Copilot Studio, RAG, Mem0, Cognee, QLoRA fine-tuning, prompt and context engineering, AI guardrails (prompt injection, PII)

Infrastructure & Cloud: Azure (Active Directory, Intune), Microsoft Graph API, Docker, GitHub Actions, Proxmox, Linux, Windows Server, Prometheus, Grafana, InfluxDB

Security & Networking: RBAC & JWT-based auth, OWASP Top 10, Active Directory, Group Policy, VLAN segmentation, WireGuard VPN, zero-trust architecture, CompTIA Security+

CERTIFICATIONS

CompTIA Security+ (valid through 2027) | Microsoft Azure Fundamentals (AZ-900) | Microsoft Azure Administrator (AZ-104) – in progress

EDUCATION

University of Miami

Bachelor of Science in Computer Science

Coral Gables, FL

2023