

Summit Singh Thakur

Austin, TX | hello@summitsingh.com | summitsingh.com
linkedin.com/in/summitsingh | github.com/summitsingh | X: @SummitSThakur

PROFESSIONAL SUMMARY

Senior Software Engineer with 8+ years of experience building AI-powered applications, backend systems, and cloud infrastructure. Former startup CTO with experience scaling products from prototype to production. Focused on LLM applications, agent workflows, and developer platforms. Strong track record in shipping production systems across AI, APIs, and distributed cloud environments.

CORE SKILLS

AI Systems: LLM applications, MCP, agent workflows, RAG, embeddings, multimodal AI, GPT, LangChain, LlamaIndex, Pinecone

Backend Systems: Python, Java, TypeScript, Go, C++, REST APIs, OpenAPI, microservices, distributed systems

Infrastructure: OCI, AWS, GCP, Azure, Kubernetes, Docker, Terraform, Helm, CI/CD (Jenkins, GitLab), AWS Lambda

Developer Platforms: API design, SDKs, CLIs, documentation, developer workflows, internal tooling

Data & Storage: PostgreSQL, MongoDB, Oracle Database, SQL, Firebase Realtime DB, Amazon RDS, S3, Supabase, Pinecone

AI PROJECTS & WRITING

AI Instagram Organizer

Open-source AI application | Python, Llama, Gemini, Ollama | 75+ GitHub stars

- Built a multimodal AI pipeline that scores, organizes, and transforms photo collections into Instagram-ready posts with captions, hashtags, scheduling recommendations, and multi-platform content.
- Implemented multi-provider model support across Llama, Gemini, and Ollama for both cloud and local inference workflows with config-driven execution and test utilities.
- Added duplicate detection, semantic context analysis, engagement prediction, image quality scoring, format conversion, analytics, and smart scheduling recommendations.
- Designed modular analyzers and workflow scripts across photo analysis, AI provider selection, content generation, and output packaging.

The Role of Generative AI in Enhancing Cloud Operations

RTInsights technical article | Oct 2024

- Published technical writing on applying GenAI to infrastructure automation, IaC generation, resource optimization, incident response, cloud security, and DevOps workflows.

PROFESSIONAL EXPERIENCE

Oracle America, Inc.

Senior Software Engineer | Jul 2021 - Present

- Built reusable Terraform modules and Helm deployment frameworks that standardized OCI environments across 8+ engineering teams and reduced environment provisioning time by 50%.
- Developed Python/OCI SDK cloud optimization tooling that analyzed utilization, automated resource cleanup workflows, reduced compute costs by 22%, and saved \$10K+ per month.
- Designed and maintained Java/Spring Boot backend services and REST APIs for Oracle Primavera/P6 platforms, improving reliability and reducing production support incidents by 25%.
- Documented APIs using OpenAPI/Swagger, creating clear service contracts that improved integration quality and enabled parallel development across teams.

- Modernized monolithic service components into independently deployable microservices, improving fault isolation and accelerating feature delivery.
- Automated CI/CD workflows for containerized services using Jenkins, Docker, and Kubernetes, reducing deployment time by 40%.
- Created developer-facing docs, demos, architecture diagrams, and internal tooling that improved onboarding by 20% and accelerated platform adoption.

Koru Shield

Founder | 2026 - Present

- Building Koru Shield from concept to production beta: a consumer DNS-security product that blocks phishing, malware, scams, ads, and trackers for the whole household before they load.

TruckBux, Inc. (now ZIKI Kitchens)

Co-Founder, CTO | Jan 2018 - Jul 2021

- Co-founded and led engineering for a \$6M-funded food-tech startup, scaling team from 2 to 12 developers and launching customer, vendor, and admin products.
- Architected Java/Spring microservices on AWS (EC2, Lambda, S3, RDS, EKS) with Firebase auth, supporting 10x monthly order growth while reducing infrastructure costs by 40%.
- Unified native Android/iOS apps into a React Native codebase, cutting release cycles by 50% and accelerating user-facing product iteration.
- Integrated DoorDash and Postmates APIs, increasing delivery revenue by 60% across 50+ vendors.
- Built API versioning, usage analytics, and Metabase KPI dashboards for orders, users, and vendors, supporting investor updates and a \$100K pre-seed raise.
- Mentored engineers and drove product execution across engineering, operations, sales, and business stakeholders.

EDUCATION, CERTIFICATIONS & AWARDS

The University of Texas at Austin - Master of Science, Artificial Intelligence - Aug 2025 - Present

Drexel University - Bachelor of Science, Computer Science - Sep 2016 - Jun 2021

Certifications: OCI 2025 Professional (6x): GenAI, Data Science, Architect, Developer, DevOps, Multicloud; Architecting Solutions on AWS; AWS Cloud Technical Essentials; Kubernetes Administration - The Linux Foundation

Awards: Hackathon awards: AWS-sponsored Best Voice User Experience using Amazon Alexa; GIPHY API award; Drexel CS Senior Design Second Place; STAR Scholar.