

SUZAL REGMI

470-265-8523 | suzalregmii@protonmail.com | linkedin.com/in/sreg | git.motaphe.dev/Motaphe

Experience

Paramount Staffing

Atlanta, GA

Lead Software Engineer

Feb 2026 – Present

- Own engineering and operations across 48 repositories and 29 Azure resource groups for products used by candidates, clients, and internal teams.
- Took a Flask and PostgreSQL hiring hub for TTEC from handoff to production in one day, adding 140+ automated tests and GitHub Actions deployment through Azure OIDC.
- Built an Azure Functions pipeline mirroring 9 Bullhorn entities nightly; exactly matched live row counts for 1.48 million candidates and 1.05 million job submissions.
- Reduced Azure costs 43% from peak while adding three client hubs and a dedicated 6.8 GB analytics database.
- Administer Microsoft 365, Entra ID, endpoints, and production DNS alongside product engineering; migrated four domains, including one with 83 records, with DNSSEC.
- Built scoped agent workflows with Claude for Azure, Bullhorn, and Microsoft 365 operations, enforcing credential boundaries and verification gates.

Software Engineer Intern

Aug 2023 – Aug 2025

- Built a Flask video interview platform on Azure Container Apps, Blob Storage, and Redis for candidates, recruiters, and client contacts; it served 1,445 submissions over 30 days, up 35% month over month.
- Automated candidate assessments in Python and TypeScript that write results directly to the Bullhorn ATS, plus REST APIs, SQL reporting and React dashboards for recruiters.
- Developed Python ETL that filtered 264,666 candidate records and loaded 72,715 deduplicated profiles into Bullhorn.

Georgia State University

Atlanta, GA

Software Developer, ARCTIC Lab

Jan 2025 – Dec 2025

- Built a RAG chatbot over the lab wiki using vector search and a local Mistral endpoint.
- Supported Slurm and Kubernetes research platforms, and documented Proflex, an Angular and Django orchestration portal, with diagrams and local build guides.

Projects

Local AI Lab | *llama.cpp, CUDA, Open WebUI, SearXNG, ComfyUI*

- Serve a Q4_K_XL build of Qwen3.8-27B with a 114K context on an RTX 5090 through llama.cpp and llama-swap; tuned VRAM allocation and speculative decoding to raise throughput from 62 to 111 tokens/sec.
- Built a private agent with SearXNG, Playwright retrieval and sandboxed code execution; exposed search over MCP and curated 6 validated ComfyUI image and video workflows.
- Connected Home Assistant to a local LLM with device tools and SearXNG search, Parakeet STT and Kokoro TTS; the warm voice round trip measured 1.32 seconds, with GPU wake on demand.

Homelab | *Proxmox, ZFS, Docker, Caddy, Headscale, AdGuard Home*

git.motaphe.dev/Motaphe/homelab

- Operate roughly 100 Docker containers across 10 Proxmox guests on a 27.3 TB raw ZFS RAIDZ2 pool, behind Caddy ingress and a private Headscale mesh, with 62 service checks and 9 heartbeat monitors.
- Built a configuration backup and restore pipeline with explicit allowlists, generated environment templates, secret redaction, and a pre-commit gate that fails closed.

dotFiles | *Arch Linux, Hyprland, Bash, fish, Lua, GNU Stow*

git.motaphe.dev/Motaphe/dotFiles

- Maintain a reproducible Arch Linux desktop with Hyprland, GNU Stow and Bash/Lua automation; a matugen palette pipeline retints the shell, editor, launcher and RGB hardware without restarting the session.

Skills

Languages: Python, TypeScript, SQL, Bash, Lua

Application and Data: Flask, React, PostgreSQL, Redis, Bullhorn ATS

Cloud and Systems: Azure, Docker, Kubernetes, GitHub Actions, Linux, Proxmox, ZFS

AI and Agent Systems: Claude Code, Codex, MCP, llama.cpp, CUDA, ComfyUI, Playwright

Education

Georgia State University, Honors College

Atlanta, GA

B.S. Computer Science, GPA 4.0 | Generation Google Scholar, NCR Scholar, Basta G-SWEP Fellow

Dec 2025