

# Aakash Regmi

410-227-5677 · [aakashregmi123@gmail.com](mailto:aakashregmi123@gmail.com) · [linkedin.com/in/aakashregmi](https://linkedin.com/in/aakashregmi) · [aregmi.net](https://aregmi.net) (AI & engineering docs) · Azure Certified

---

## Professional Summary

Lead software engineer and systems generalist who builds AI applications end-to-end — from business discovery and solution design to prompt engineering, secure implementation, CI/CD, and production operations.

Deep experience delivering cloud-native and agentic AI services on Azure/AKS (Java/Spring Boot, Python/FastAPI), including MCP-backed tool integrations, multi-agent orchestration, RAG, evals, and enterprise security architecture with OAuth and Azure Entra.

For the last 3 years, served as a lead developer and architect for GenAI initiatives — guiding engineers on new technologies, architecture decisions, and production best practices — while leading AI governance and cross-functional delivery (Business, Security, Data) to production and growing adoption from 0 usage in 2023 to ~150K API calls/hour while creating practices for observability, cost awareness, and risk management in production AI systems.

---

## Education

**Towson University** — M.S. Computer Science (December 2020)

**Towson University** — B.S. Computer Information Systems / Business (May 2019)

---

## Professional Development

Certification
Microsoft Certified Azure AI Engineer
Microsoft Certified Azure AI-900
Microsoft Certified Azure Administrator
Microsoft Certified AZ-900
AWS Certified Cloud Practitioner
Student Employee Technology Corp, Towson University
Hardware and Network Technician, CRTT

---

## Experience

### Lead Software Engineer — AT&T

June 2024 – Present · Atlanta, GA

**Tech Stack:** Azure, AKS, Spring Boot, FastAPI (Python), Spring AI, LangChain4j, MCP, Azure Communication Services, Twilio, Arize, LiteLLM Trace, Spring AI OTEL, Azure AD (App Registrations), Akamai, Azure DevOps, GitHub Actions

- Spearheaded the development and deployment of GenAI products and platform capabilities on Azure/AKS, including MCP server and client integrations, tool routing, and multi-agent workflows aligned with emerging agent-to-agent interoperability patterns to enhance customer experience and streamline business operations
- Designed operational insight workflows using orchestration agents, delegated subagents, and blackboard-style collaboration to analyze incidents, change requests, and logs, reducing troubleshooting time by 40%
- Built voice agents leveraging Azure Communication Services and Twilio, integrating with multiple IVR partners to deliver agentic call flows that reduced call-center agent handle time and lowered operational costs

- Designed and implemented Retrieval-Augmented Generation (RAG) patterns, prompt optimization strategies, and evaluation guardrails across multiple flows to ground responses in knowledge-base content, reduce hallucination risk, and improve answer quality and information accessibility
  - Built and launched GenAI-powered digital flow services (e.g., bill explanations, service recommendations, intent analyzers) and internal support tools, helping reduce customer service call drivers and support sales enablement
  - Developed and deployed automation for AI Use case tracking and governance workflow using Power Automate, Power Apps, improving cross-functional visibility and efficiency of the approval process while ensuring compliance with risk and security requirements
  - Leveraged Power Automate and Power Apps to build connectors for Microsoft Dynamics CRM for customer support and collecting data for making customer service efficient and automated.
  - Instrumented GenAI services with Arize, LiteLLM Trace, and Spring AI OTEL to trace agent execution, monitor latency and token usage, and improve cost-performance visibility for production AI systems
  - Built proof-of-concept fine-tuning workflows for customer-facing tone and response quality, then translated those learnings into safer prompt and response-pattern design for production releases
  - Architected service-to-service security for GenAI and platform services using OAuth and Azure AD app registrations, standardizing authentication/authorization patterns for production deployments
  - Troubleshoot Akamai-related network and security issues impacting application traffic, partnering with security and infrastructure teams to restore service and prevent recurrences
  - Built and maintained CI/CD pipelines using Azure DevOps and GitHub Actions to automate build/test and deployments to AKS, improving release consistency and reducing manual steps
  - GenAI governance lead: partnered with business stakeholders and cross-functional teams (Security, Data, etc.) to define controls, align risk/compliance requirements, and secure approvals to move AI flows into production
  - Drove enterprise adoption of GenAI from no production usage in 2023 to ~150K API calls/hour, establishing scalable patterns for agentic flows, tool routing, and production delivery
  - Led Cricket's GenAI engineering efforts across both the Sr. Systems Administrator and Lead Software Engineer roles — starting with early prototypes in 2023, then scaling to production as lead developer and architect, mentoring engineers on system design and delivery best practices while driving multiple use cases aligned to business goals
  - Created and facilitated a GenAI Community of Practice (COP) and internal hackathons to teach developers across teams agentic patterns, guardrails, and implementation best practices, accelerating adoption beyond the core team
- 

## Sr. Systems Administrator (Developer) — AT&T

March 2022 – June 2024 · Atlanta, GA

**Tech Stack:** Azure, Azure Functions, Intune, Jamf, Power Automate, Azure Automation, PowerBI, SentinelOne, SAML, Conditional Access/MFA

- Initiated and led the first GenAI proof-of-concept projects at Cricket — prototyping LLM-based customer support flows and internal automation tools using Python and OpenAI, which demonstrated enough business value to justify a dedicated GenAI engineering role.
- Built early RAG and prompt engineering prototypes using FAISS that became the foundation for production GenAI services after transitioning to the Lead Software Engineer role
- Built text to sql prototypes using OpenAI SDK , langchain to demonstrate how GenAI could help non-technical users query databases, which informed the design of later production releases with Spring AI
- Developed and deployed a Linux web app for API management on an Azure VM using an open-source platform
- Led a cost-efficient replacement of McAfee with SentinelOne, delivering ~\$400K annual savings while improving endpoint security across 10,000+ endpoints
- Led R&D and proof-of-concept efforts for new technologies and process improvements, accelerating delivery of production-ready solutions
- Served as lead developer/architect for a cross-functional internal project, coordinating a 13-person team to deliver milestones and production outcomes
- Automated role-based app access requests, improving turnaround time while strengthening access controls
- Contributed to successful CricketOne Team campaigns, reaching 2.6 million users, increasing customer engagement and satisfaction
- Enabled cross-tenant reporting for the Compass app, facilitating data collaboration between AT&T and Cricket
- Enhanced device management and reporting by resolving Aktivite reporting issues for iPad names with Intune
- Delivered SAML integrations end-to-end (design through deployment) to improve secure access and SSO enablement

- Assisted the infrastructure team in upgrading and patching Linux RHEL 8 VMs and deploying the Docker application for Hopscotch
- Played a key role in launching the Cricket One customer data platform and personalization experience application
- Implemented Autopilot deployment, reducing PC reset time and increasing the success rate
- Strengthened enterprise security by deploying SentinelOne threat protection and implementing Conditional Access/MFA controls for internal applications
- Automated the wiping of unauthorized devices using Azure Functions and redesigned the security architecture for Azure and Intune
- Tested different Microsoft licenses to optimize user experience for app access with AAD
- Automated repetitive tasks using Power Automate and Azure Automation, improving turnaround time and reducing manual effort
- Developed an automated process for PC enrollment with a one-time password, improving previous enrollment methods
- Enhanced system security by deploying secure local Azure admin accounts across endpoints using endpoint protection
- Delivered a business reporting solution with device name for transactions within a two-week timeframe
- Facilitated cross-tenant reporting between AT&T and Cricket for the PowerBI app and ensured successful Checkpoint (firewall) installations in 18 stores

---

## Engineer III — Omniforce Solutions (Client: CenturyLink/Lumen)

May 2021 – May 2022 · Remote

**Tech Stack:** Windows Server, Linux (RHEL), VMware ESXi/vSphere, Azure, O365, Active Directory, IIS, PowerShell, SQL Server

- Managed approximately 15,000 Windows and Linux servers for Fortune 500 companies
- Managed customers' Azure and O365 subscriptions including management of users, groups, and applications
- Performed troubleshooting and diagnosis of application, Active Directory, IIS, OS performance, and networking issues via GUI and PowerShell
- Coordinated VMware ESXi patching and cluster migrations from 5.1/5.5 to 6.7 and newer for multiple customers
- Installed VMware tools and troubleshooted increased CPU/RAM and disconnected hosts in VMware
- Migrated servers and storage to different hosts, clusters, and data centers and set up VLANs
- Led customer change requests for adding servers to domain, drive expansion, patch installation, third-party application installation, and unresponsive websites
- Involved in the execution of client change requests in a timely manner as planned by Client Service Partner, Engineering, and Hosting Compute Manager
- Mentored lower tiers on resolution of OS and VMware issues
- Coordinated and scheduled maintenance activities with clients for 100% success rate
- Coordinated with storage team on customer disk LUN growth, including SQL failover clusters
- Configured DRS affinity rules for VMs
- Installed and configured antivirus and VPN for customers' domains
- Coordinated with third-party vendors to replace faulty hardware
- Installed and renewed SSL certificates on IIS
- Investigated and diagnosed priority-one incidents involving server and application down issues and engaged crisis management
- Resolved third-level support cases; 95% resolved without escalation
- Ensured clients and tickets were always up to date on incident progress
- Provided detailed feedback to other groups on all incident resolutions, ensuring full details were entered into ticket case notes
- Reviewed open cases daily and updated Operations Management on current cases and status
- Provided comprehensive detail to global peers on shift handovers
- Attended daily operations and change review meetings as required
- Logged cases to third-party vendors requesting assistance on unresolved issues
- Compliance: worked with the CTL client base to organize and schedule compliance items such as hardware and software updates including firmware revisions and OS and application patching

---

**Systems Administrator Jr. — Towson University**

**Tech Stack:** SCCM, Jamf Pro, Active Directory, Azure AD, VMware ESXi/vSphere, Hyper-V, Azure, PowerShell, Python, Jenkins, Git

## Systems

- Configured software, OS, and security patches for 12,000 Windows desktops, servers, and Mac devices on a monthly basis with 95% success rate using SCCM and Jamf Pro
- Designed and implemented automated steps for imaging new PCs and laptops using SCCM and MDT to reduce imaging time and personnel used
- Implemented Windows migration and BitLocker encryption with 97% accuracy
- Successfully enrolled, managed, and integrated new Windows and Mac devices to domain
- Created and deployed monthly updates including endpoint protection to all Macs using Composer and Jamf Admin
- Managed user accounts, groups, and access control using MS Active Directory and Azure AD
- Engineered and modified GPOs for creating new policies as well as consistency across all machines
- Installed, configured, and maintained MS SharePoint and SQL servers
- Installed, configured, and administered VMware ESXi servers with minimal downtime
- Configured, managed, and created Windows Hyper-V servers
- Created, managed, and edited WordPress and SharePoint sites and modified web redirects with IIS
- Configured patches, snapshots, and templates in VMware vSphere
- Developed wireframes, mockups, and technical flowcharts using Visio to describe technical details of WebApp with ETL processes
- Interacted with help desk and other teams for troubleshooting, identifying root cause, and providing technical support when needed
- Performed deep log analysis for troubleshooting of emergency desktop and application issues
- Maintained hardware inventory and scheduled preventative maintenance schedules, resulting in greater uptime than the contract's SLA
- Analyzed data using Splunk and SQL reports with Python and RStudio for data analysis
- Experience with data reporting, BI, and visualization tools like Power BI, Tableau, and SSRS
- Experienced in ETL for developing and managing websites with database backend using Python and JavaScript
- Monitored infrastructure through Azure Monitor and System Center Operations Manager (SCOM)
- Developed project plans and schedules using MS Project; contributed to weekly agile scrum meetings with a 6-person team and sprint planning/review meetings

## Cloud

- Developed PowerShell scripts and ARM templates to automate the provisioning of Azure resources, increasing onboarding speed and removing human error
- Built and deployed VM labs based on client requests on Azure Labs and vSphere
- Monitored Azure cloud-based systems for availability, performance, reliability, and security using Azure Monitor, Azure Log Analytics, and Azure Trusted Advisor
- Designed and configured Azure Virtual Networks (VNets), subnets, Azure network settings, DHCP address blocks, DNS settings, security policies, and routing
- Built and deployed ML web apps by leveraging Flask and SQL database on Microsoft Azure
- Automated creation and enforcement of policies using Azure Policies
- Automated reporting and scripting using PowerShell, PowerCLI, Python, and Azure CLI
- Deployed Azure virtual machines and cloud services into secure VNets and subnets
- Configured CI/CD using Jenkins and Git in Linux/Windows for application processes and production support

---

## Technical Skills

Category	Skills
Cloud & Containers	Azure, AKS, Azure OpenAI, Azure AI Services, Intune, Jamf, AWS, VMware, Akamai, MS Dynamics
Languages	Java, Python, SQL, PowerShell, Bash
Backend	Spring Framework, FastAPI, Flask

<b>Category</b>	<b>Skills</b>
<b>AI/Automation</b>	Spring AI, LangChain4j, prompt engineering, Evals, OTEL frameworks, agentic workflows, multi-agent orchestration, RAG, MCP, A2A-aligned patterns, fine-tuning , Power Automate, Power App
<b>Observability</b>	Arize, LiteLLM Trace, Spring AI OTEL, Azure Monitor, Log Analytics, Splunk, MLflow
<b>Databases</b>	SQL Server, MongoDB, SQLite, Azure CosmosDB, Databricks, Snowflake
<b>DevOps</b>	Azure DevOps, GitHub Actions, Git, Jenkins, IIS, Power Automate, Visio, MS Project
<b>MDMs</b>	Intune, Jamf, SCCM, Autopilot, SentinelOne, Entra
<b>OS</b>	Windows, Windows Server, macOS, RHEL, Ubuntu
<b>ITSM/CRM</b>	ServiceNow, Remedy, SolarWinds, Zoho