

Lokesh Pulivarthi

(913) 742-9950 | lokesh.pulivarthi@hotmail.com | [linkedin.com/in/lokeshpulivarthi](https://www.linkedin.com/in/lokeshpulivarthi) | github.com/LXP86050 | lokeshpulivarthi.com

SUMMARY

Software Engineer (SDE II) with around 7 years building scalable distributed systems, agentic AI applications, and cloud-native platforms. Hands-on experience designing Retrieval-Augmented Generation (RAG) pipelines with Large Language Models (LLMs), Kubernetes-based microservices on Azure, and high-throughput REST APIs. Track record of reducing latency, improving system reliability and high availability, and scaling production systems to handle thousands of concurrent requests with secure, observable, production-grade deployments. AWS and Google Cloud certified.

CORE SKILLS

Languages: Python, JavaScript, TypeScript, C#, Java, SQL

Backend: Django, FastAPI, .NET, .NET Core, ASP.NET Core, RESTful APIs, Microservices Architecture, Entity Framework, Dapper

Frontend: React, Next.js, Redux, Tailwind CSS, Fluent UI, Material-UI, HTML5, CSS3, Responsive Design

Databases: PostgreSQL, SQL Server, MongoDB, Cosmos DB, Redis, Azure AI Search (vector database), Vector Embeddings

Cloud & DevOps: Microsoft Azure (AKS, App Services, Functions, Azure SQL, Azure DevOps, Bicep, Blob Storage, Azure AI Search, Azure OpenAI), Docker, Kubernetes (K8s), Terraform, GitHub Actions, CI/CD Pipelines, Infrastructure as Code (IaC)

AI & ML: Large Language Models (LLM, LLMs), Retrieval-Augmented Generation (RAG), Agentic AI, AI Agents, Prompt Engineering, Fine-tuning, Vector Search, LangChain, Model Context Protocol (MCP), Azure OpenAI (GPT-4), scikit-learn (sklearn), NLTK, TensorFlow, Model Deployment, Data Preprocessing

Security: Azure AD, Microsoft Entra, OAuth 2.0, JWT, SAML, RBAC, Authentication & Authorization, SSO, Okta

Concepts: System Design, Distributed Systems, High Availability, Fault Tolerance, Scalability, Observability, Performance Optimization, API Integration, Full Stack, Production Engineering, Agile/Scrum

Tools & Testing: Git, Postman, VS Code, Visual Studio (2017/2019/2022), WSL, Xunit, Jest, PyTest

EXPERIENCE

Software Engineer II | Infosys

Mar 2024 – Present

- Architected agentic AI assistants powered by Large Language Models (LLMs) using Azure OpenAI (GPT-4), LangChain, and Azure AI Search vector embeddings (Retrieval-Augmented Generation); reduced analyst lookup time by 55% and served 8K+ weekly queries with prompt-engineering guardrails, hybrid retrieval, and token-cost tracking.
- Designed Azure AD / Microsoft Entra Authentication & Authorization with RBAC and SSO for sensitive enterprise data, achieving 100% compliance with internal security standards.
- Automated CI/CD across distributed microservices on AKS (Azure Kubernetes Service) using Azure DevOps pipelines, Bicep templates, and Infrastructure-as-Code; decreased deployment time by 30% while improving system reliability and observability.
- Developed full-stack web applications using Python, Django, and React; refactored database queries and code paths to achieve 40% improvement in API response times.
- Owned end-to-end system design across frontend, backend APIs, databases, and Kubernetes-based microservices on Azure, ensuring high availability and fault tolerance.

Full Stack Developer | UCLA

Aug 2022 – Mar 2024

- Migrated legacy ASP.NET applications to .NET Core, re-architecting the codebase with Entity Framework Core, Dapper, and modular dependency injection to improve application performance by 35% and increase

maintainability for future development.

- Designed and developed RESTful APIs in C# / .NET Core for secure client-server integration; Dapper-based query optimization improved API response times by 30% and raised user satisfaction.
- Optimized database interactions to handle large datasets, reducing redundant calls with Entity Framework Core and Dapper to cut database load by 25% and improve overall application throughput.

Dotnet Developer | Cognizant

Dec 2018 – Aug 2021

- Built secure API integrations in .NET for payment gateways, messaging systems, and analytics services, expanding application functionality and user engagement.
- Modularized React frontend using component-based architecture, improving code readability and maintainability while reducing development time for future updates.
- Delivered full-stack applications with .NET Core backends and React frontends, implementing RESTful APIs and modern state-management libraries to streamline UI rendering and performance.

PROJECTS

Enterprise RAG-Based Knowledge Assistant | *Python, LangChain, Azure OpenAI (GPT-4), Azure AI Search, REST API*

- Architected and deployed a Retrieval-Augmented Generation (RAG) AI assistant using Large Language Models, processing 8K+ weekly queries with grounded, low-hallucination responses, prompt-engineering safeguards, and cost monitoring.
- Deployed as secure REST APIs with Azure AD RBAC for cross-team access, achieving <1.5s average latency on 8K+ monthly queries while ensuring data privacy.
- Implemented hybrid retrieval (vector + keyword search), caching layers, and evaluation pipelines that reduced token costs by 25% and improved overall system efficiency.

AI-Powered Sentiment Analysis Platform | *Python, Django, scikit-learn, PostgreSQL, JavaScript, Microsoft Azure*

- Built a web-based sentiment classifier using machine learning models trained on a large product-review dataset, achieving over 90% classification accuracy.
- Integrated the trained ML model into a Django backend with a real-time web interface for uploading or typing text for instant analysis.
- Engineered a PostgreSQL backend to store feedback and sentiment results, powering analytics dashboards with sentiment distribution charts.
- Deployed on Microsoft Azure for scalable multi-user access with actionable customer-insight reports.

EDUCATION

Master of Science in Computer Science — University of Central Missouri, 2022

Bachelor of Technology in Electronics & Communication Engineering — SR University, 2018

CERTIFICATIONS

- AWS Training & Certification — **Fundamentals of Machine Learning and Artificial Intelligence** (May 2026)
- Google Cloud Skills Boost — **Introduction to Large Language Models** (May 2026)
- Google Cloud Skills Boost — **Introduction to Generative AI** (May 2026)
- Google Cloud Skills Boost — **Introduction to Responsible AI** (May 2026)