Goutham Raju Kosuru Srinivasa

gkosurus@asu.edu | +1 (602) 696-9216 | linkedin.com/in/gouthamraju11 | github.com/Gouthamraju11

Education

Arizona State University, Tempe, AZ — Dec 2025
Master of Science in Information Technology — GPA: 4.0/4.0

Technical Skills

Languages: Python, C#, Java, JavaScript, TypeScript, PowerShell, HTML/CSS
Frameworks/Tools: React.js, Angular, Node.js, .NET Core, Django, TensorFlow, Docker, Git, Azure DevOps
Databases/Cloud: Microsoft Azure, AWS (Lambda, EC2, S3, DynamoDB), MySQL, MongoDB, SQLite
Focus Areas: Distributed Systems, Cloud Computing, Data Engineering, Machine Learning, Performance Optimization

Experience

Neudesic (An IBM Company) — Associate Software Engineer — Apr 2022 – Nov 2023
Tech: .NET Core, Angular, React.js, Azure, Node.js, REST APIs, xUnit, Jest

- Designed, developed and deployed a large-scale internal field-service management system for a Fortune 500 client using Angular and .NET Core.
- Optimized backend logic and caching layers to improve API throughput by 25% and reduce page latency by 30%.
- Re-architected microservice-based modules for high availability using Azure App Services and Azure SQL with sharded data models, enabling concurrent operations for over 2,000 daily active users without downtime.
- Built and integrated CI/CD pipelines with Azure DevOps using YAML automation and parallel job execution, reducing manual deployment effort by 60% and increasing release frequency by 40%.
- Enhanced code reliability by refactoring critical backend services, implementing dependency injection and expanding unit test coverage from 25% to 75% using xUnit and Jest, reducing regression bugs post-deployment.
- Implemented Azure Application Insights and Log Analytics for distributed tracing, improving incident response and root cause analysis, which cut production issue resolution time by 30%.

The Sparks Foundation — Software Engineer Intern — Aug 2021 – Oct 2021
Tech: AWS CDK, Lambda, API Gateway, DynamoDB, Python, Docker, CloudWatch

- Developed and deployed a serverless transaction monitoring API using AWS Lambda, API Gateway and DynamoDB, improving request processing reliability and maintaining near-constant uptime during load testing.
- Automated infrastructure provisioning with AWS CDK templates, building reproducible environments integrated with CodePipeline and CodeBuild, reducing setup time from 2 hours to under 20 minutes.
- Containerized the service with Docker for isolated environment builds and CI integration, improving build consistency.
- Configured CloudWatch alarms and logging for event-driven triggers and performance metrics, reduce debugging time and improving fault tolerance under concurrent invocations.

Projects

Multi-Agent AI Workflow Automation Platform — AWS Bedrock, SageMaker, FastAPI, React.js — Aug 2025

- Designed and implemented an intelligent workflow automation system using LangChain and CrewAI for orchestrating collaborative AI agents capable of autonomous data extraction, summarization and decision-making across enterprise datasets.
- Integrated AWS Bedrock for fine-tuned foundation models and Amazon SageMaker for real-time model inference, enabling scalable LLM serving and low-latency parallel processing across distributed inference endpoints.
- Built a modular backend with FastAPI and vector databases (FAISS) for semantic retrieval and contextual memory, improving multi-agent task coherence by 40% and reducing API latency to under 150ms per query.
- Developed a responsive React.js dashboard for live task monitoring, model analytics and reinforcement feedback loops, facilitating adaptive prompt tuning and continuous agent improvement through user interaction data.

Book Recommendation System — Python, Pandas, NumPy, SciPy — Dec 2024

- Built a collaborative filtering recommendation system using cosine similarity on user-item matrices for over 60K records, leveraging sparse matrix computation for memory-efficient similarity lookup.
- Designed modular data pipelines with pandas for normalization, filtering and transformation, improving similarity computation speed by 25%. Used NumPy broadcasting for optimized linear algebra operations.
- Implemented performance profiling with SciPy vectorization and caching mechanisms, enabling real-time recommendation generation within 100ms per query on commodity hardware.

Reader and Object Detector for the Visually Impaired — Python, TensorFlow, OpenCV, OCR, TTS — Jul 2022

- Developed an end-to-end assistive AI system that detects objects and reads printed text aloud using SSD MobileNet, Tesseract OCR and text-to-speech conversion, enhancing navigation assistance for visually impaired users.
- Integrated asynchronous frame capture, GPU acceleration and model quantization to achieve 45% faster inference and real-time performance (15+ FPS) on standard hardware.
- Implemented object prioritization algorithms to identify critical elements like vehicles and obstacles, boosting classification precision by 18% and significantly improving situational awareness.