Sai Rohith Tanuku

United States | +1-412-589-8024 | tanukusai.sai52@gmail.com

SUMMARY

Software engineer with an MS in Computer and Information Sciences and 3+ years building scalable backend systems and data pipelines. Developed FastAPI microservices handling 10K+ events/day, cut processing time by 40%, and engineered an AI-assisted autograding pipeline reducing manual workload by 30%. Proficient in Java, C/C++, Python, Linux, TCP/IP, AWS, and GCP. Familiar with microservices architecture, CI/CD, automation testing, and AI/ML infrastructure.

EDUCATION

University of Pittsburgh   Aug 2023 - Apr 2025
Master of Science, Computer and Information Sciences
- **GPA:** 3.957/4.00
- **Coursework:** Algorithms, Data Structures, Machine Learning, Neural Networks, Cloud Computing

Manipal Academy of Higher Education   Aug 2019 - May 2023
Bachelor of Science, Computer Science & Communication Engineering
- **Coursework:** Database, Operating Systems, SDLC, Computer Network, Data Mining, Web & App Dev

WORK EXPERIENCE

ChiEAC   Mar 2025 - Present
Software Engineer   *Remote*
- Engineered backend infrastructure by implementing CI/CD pipelines for automated builds, security checks, and zero-downtime deployments, and by building REST APIs that accelerated cross-team integration and cut manual workflows by 50%.
- Developed comprehensive observability stacks with Prometheus and Grafana to monitor latency, system health, and model-inference performance, partnering with various stakeholders to translate analytical needs into scalable backend architectures for classroom and outreach programs

University of Pittsburgh   Jun 2024 - May 2025
Master of Data Science Course Facilitator and TA   *Pittsburgh, PA*
- Engineered an AI-assisted autograding pipeline with automated testing workflows and redesigned test cases for 15 Coursera-based assignments, reducing TA workload by 30% while supporting 140+ students with code reviews and academic guidance over two semesters.

EAIESB (Enterprise Solutions)   Jan 2023 - May 2023
Software Engineer   *Hyderabad, India*
- Engineered microservices-based system for ingesting data, processing 10,000+ events per day. Established a FastAPI system for handling and uploading datasets exceeding 10GB and executing machine learning tasks such as clustering, regression, and classification with automated preprocessing. Designed a fault-tolerant error-handling framework with APEX, Sentry, CloudWatch, and TCP/IP networking, reducing processing time by 40% and error rates by 30%

Larsen & Toubro   Jun 2022 - Aug 2022
Artificial Intelligence Research Intern   *Chennai, India*
- Built AI/ML pipelines, refined YOLO/RetinaNet/EfficientDet models, automated data ingestion (REST APIs, TensorFlow, OpenCV, GIS), processed LiDAR data to enhance detection accuracy by 10%, reduced manual analysis time by 25%, and enabled scalable model serving for edge and cloud

PROJECTS

Fine-Tuning LLaMA and DeepSeek for SQL Generation
- Fine-tuned LLaMA-7B and DeepSeek-Coder models on 120K+ NL-to-SQL pairs through PyTorch, Hugging Face, and DeepSpeed on multi-GPU A100 clusters, improving SQL query correctness by 30% over baseline models.
- Implemented scalable, domain-adaptable training workflows to support custom SQL dialects across finance, e-commerce, and healthcare applications with custom frontend.

VillaVista:Villa Rental
- Architected a full-stack villa rental platform using NestJS, TypeScript, and PostgreSQL, serving 10,000+ users with 99.9% uptime, implementing JWT authentication and a scalable Prisma ORM database schema.
- Integrated real-time payment processing with Stripe and Razorpay, handling 5,000+ transactions in the backend, and deployed on AWS using Docker and CI/CD pipelines, achieving 90%+ test coverage with Jest.

Blockchain-Based Data Integrity System
- Built a fault-tolerant data integrity system using Ethereum, Cassandra, and Merkle Trees, detecting 97% tampering on 1.2B+ records and out-performing baseline methods by 4x with distributed MapReduce processing.

TECHNICAL SKILLS

- **Programming Languages & Backend**: Python, Java, Kotlin, C/C++, SQL, TypeScript, JavaScript, Shell Scripting, REST
- **Frameworks, Libraries & AI**: PyTorch, TensorFlow, Node.js, AI/ML Understanding, MLOps, APIs, Rest APIs
- **Backend & API Development**: Microservices, Distributed Computing, Load Balancing, Multi-threading, Caching, Backend Development, Platform Engineering, System Design
- **Cloud, Big Data & Databases**: AWS, GCP, Docker, CI/CD, PostgreSQL, MongoDB, Spark, Cloud Networking
- **Networking & Automation**: Linux, Networking Concepts, Networking Protocols, Automation Testing