Pavan Kumar Gannoju

+916300422984 | pavan9542644804@gmail.com | LinkedIn | Portfolio | Hyderabad, India

Professional Summary

Computer Science undergraduate (B.Tech, AI/ML) with a strong foundation in data structures, algorithms, and object-oriented programming. Proven experience in building scalable distributed systems, optimizing backend performance, and deploying secure microservices. Adept in Python, Java, and C++. Delivered real-world impact during internships by reducing infra cost, enhancing fault tolerance, and ensuring 99%+ uptime. Passionate about solving ambiguous problems, working in agile teams, and contributing to scalable, secure, and inclusive digital systems at Mastercard.

Experience

1. **Software Engineering Intern (Gen AI) -** Digital Nexus AI, Bangalore May 2025 – Oct 2025
   - Developed scalable backend systems using FastAPI + AWS Lambda that handled >100 RPS, reducing infra costs by 31%.
   - Secured endpoints with JWT, RBAC, and Redis throttling; supported 1,000+ daily users.
   - Improved response time by 23% through async I/O, caching, and optimized database queries.
   - Maintained 99.97% uptime with observability stack (Prometheus, CloudWatch, structured logging).
   - Automated CI/CD with GitHub Actions for daily deploys and <5-min rollback recovery.
2. **Software Engineering Intern (Agentic AI)-** Prodigal AI Technologies Pvt. Ltd., Delhi March 2025
   - Delivered 6 Dockerized microservices with 99.5% uptime, serving 10K+ weekly API requests.
   - Implemented a FAISS + SQL hybrid retrieval system, reducing query latency by 34% across 40K+ documents.
   - Reduced debugging time by 50% via retry logic, circuit breakers, and Slack-based SQL alerts.
   - Collaborated with QA using PyTest and Postman to reduce production bugs by 40%.
3. **Research – Autonomous Vehicle Data Pipeline (JNTUH) -** Hyd Feb 2024 – July 2024
   - Processed 1.5M+ vehicle logs using Python + SQL; reduced manual review time by 45%.
   - Enhanced anomaly detection accuracy by 10% via time-window tuning and signal aggregation.
   - Delivered dashboards (Seaborn, Pandas) and modular code adopted in internal research tools.

Projects

1. **AI-Powered Document Retrieval System** January 2025 – March 2025
   - Engineered a scalable retrieval engine combining FAISS (Vector DB) and SQL, enabling 5,000+ daily document queries with 98.6% semantic accuracy.
   - Integrated embedding-based semantic search and custom query routing to reduce average latency by 34%.
   - Released 4 microservices using retry, timeout, and circuit breaker patterns to ensure 99.5% production reliability.
   - Handled 100+ concurrent sessions efficiently using async I/O and memory-optimized indexing.
2. **Cloud Auto-Scaler with Real-Time Monitoring** May 2025 – July 2025
   - Designed a lightweight cloud auto-scaling solution using Prometheus + AWS Lambda, reducing idle compute cost by 43%.
   - Monitored 100+ microservices and applied CPU/memory thresholds to auto-scale container workloads.
   - Built Grafana dashboards and configured 12+ CloudWatch alerts to track system health (e.g., memory leaks, latency spikes).
   - Achieved <60s response time for failures with Slack-triggered alerts and GitHub Actions-based rollbacks.

Skills

- **Programming Languages:** Python, Java, C/C++
- **Computer Science:** Data Structures, Algorithms, OOP, Complexity Analysis, Operating Systems, Networking (TCP/IP)
- **Backend & Distributed Systems:** RESTful APIs, Microservices, Async I/O, Circuit Breakers, Load Testing
- **Cloud & DevOps:** AWS Lambda, GitHub Actions, CI/CD, CloudWatch, Prometheus, Grafana, Auto-Scaling
- **Databases & Search:** PostgreSQL, SQLite, Redis, FAISS (Vector DB), SQL Query Optimization
- **Monitoring & Observability:** Structured Logging, Alerting (Slack), Dashboarding (Grafana), Real-Time Metrics
- **Security Practices:** JWT Authentication, RBAC, Rate Limiting, Retry/Fallback Logic
- **Tools & Environments:** Docker, Unix/Linux, Bash, Git, Agile Methodologies
- **AI/ML & Data Processing:** Embedding Retrieval, Anomaly Detection, Pandas, Seaborn

Education

- **Visvesvaraya College of Engineering & Technology -** Bachelor of Technology - Computer Science (AI/ML)
  June 2023 – May 2026 | Hyderabad
- **Indian Institute of Technology, Ropar – Minor in Artificial Intelligence**
  Sept 2024 - Jul 2025 | Punjab

Extra Curricular Activities

1. J.P. Morgan Virtual Software Engineering Program – Forage (June 2024)
2. Organizer, *Code for Fun* Hackathon – Led event for 50+ participants, handling logistics and mentoring.