PAVAN AKKINENI

📞 6027567565 | ✉ pavanchowdary.akkineni@gmail.com | in LinkedIn | 💼 Portfolio | 📕 Databricks Cert | 🔷 AWS Cert

EDUCATION

University of North Carolina at Chapel Hill — North Carolina, USA
Masters in Computer Science, CGPA: 4.0/4.0 — *Expected May 2026*
- **Courses:** Deep Learning, LLMs for Software, Visual Computing, Image Processing, Multimodal AI, Reinforcement Learning and Markov Decision Processes, Machine Learning, Software Engineering.

WORK EXPERIENCE

UNC School Of Data Science and Society — Chapel Hill, NC, USA
Graduate Teaching Assistant — *January 2025 – Present*
- Mentored students in **statistics, machine learning, and neural networks**; designed and led hands-on workshops on clinical datasets; guided teams to build ML pipelines (ingestion, feature engineering, model selection, and evaluation).

UNC School of Medicine — Chapel Hill, NC, USA
Summer AI Engineering Intern — *June 2025 - August 2025*
- Engineered a predictive model for RNA MFE leveraging **Rinalmo(RNA LM)** context and K-mer pos-embeddings.
- Skilfully created antisense RNA-based therapeutic candidates for HIV TAR1 through **Evolutionary Algorithm**.

Bajaj Finance Limited — Pune, IN
Senior Software Engineer — *September 2023 – August 2024*
- Strategically steered **data-privacy** controls for **11+** databases, leveraging Securiti.ai, Databricks, and EDW(ASA) to achieve compliance with DPDP Act with RBAC/least-privilege IAM, PII masking, KMS-backed encryption.
- Engineered a high-availability Django platform (Securiti.ai–style) with async I/O, DB read-replicas, Celery queues, and Prometheus/Grafana observability—zero-downtime releases and 3.6× faster scan cycles.
- Pioneered a data catalog/lineage service with REST/GraphQL contracts, RESTful microservices, schema versioning, and audit logging, enabling real-time metadata flow and idempotent writes securing $80K/yr vendor savings.
- Automated an **OCR**-driven KYC ingestion pipeline using OpenCV, Tesseract with queue-backed workers, idempotent retries/backoff, PII redaction and structured logging; Graphana SLIs with alerts, enabling **60%** hands-free onboarding.
- Revamped legacy ingestion scripts to **Databricks Lakehouse** with **modular ETL**, schema enforcement **Delta**, **Unity Catalog ACLs** and lineage-aware audit logs, plus **AQE/shuffle tuning** across **24** production sources.

Data Engineer — *July 2022 – August 2023*
- Replatformed the entire production environment and scripts onto **Databricks**, reducing batch-upload time by 30%.
- Instituted expert-grade data-quality controls across OfferMart pipelines, eliminating **35%** of unnecessary processing.
- Overhauled scalable **Spark ETL** pipelines using **spark and Airflow** with lazy-loading, checkpointing and dynamic compute provision, cutting batch p95 by **25%** while orchestrating concurrent tasks for peak resource utilization.
- Led campaign offer generation with **DAG**-based policy workflows, queue-backed optimization and monitoring campaign dissonance through cron jobs with **SLA**-backed anomaly **alerts** and apply canary patches/rollback playbooks.

Data Engineer Intern — *January 2022 – June 2022*
- Contributed to Delta Lake data purging through VACUUM, retention policies, daily reports to save storage by **30%**.
- Facilitated partitions, z-order indexing strategies for **Hive** & Delta Lake tables, boosting p90 query latency by **25%**.

PROJECTS

- **MatchPlayer** - Masterminded a **YOLO**-based face-detection + **FaceNet**-embedding pipeline within an AI soccer commentary, driving player-ID accuracy to **52% Top-2 / 60% Top-3—about 15 × above the 4%** random baseline.
- **Prism LLM** - Expertly architected an **agentic** planner web app with **rate-limiting, retry/backoff** that classifies user query and dynamically routes it to the optimal LLM (DeepSeek, Mistral, Qwen, Gemma), boosting answer relevance.

TECHNICAL SKILLS

- **Data:** Spark SQL, Transact-SQL, PySpark, CDC (MERGE/Auto Loader), ADF, HDFS, MapReduce, DynamoDB, Azure Synapse Analytics, Azure Databricks, Azure Cost Management, ADF, Azure Devops, kafka, DynamoDB, MS Excel.
- **AI:** NLP, Computer Vision, Gen AI, LangChain, PyTorch, TensorFlow, Keras, scikit-learn, SciPy, NLTK, Hugging Face Transformers, MLflow, Data Acquisition and Handling
- **Other:** Python, C/C++, Java, Data Structures, OOP, Docker, Kubernetes, JavaScript, FastAPI, React, HTML, CSS, JQuery, Linux, Git, CRM, Cloud Computing, SDLC, CI/CD, SLIs/SLOs, Prometheus, Distributed tracing (OpenTelemetry), Agile/Scrum, Mentorship, Problem solving and Effective integration of related AI capabilities.

CERTIFICATIONS & AWARDS

- **Databricks certified Data Engineer Professional** -Structured streaming, Lakehouse, Testing, deployment, security.
- **AWS certified Machine Learning Associate** - Glue, S3, Kinesis, SageMaker, ECS, CloudWatch, Athena, Redshift.
- **Oracle AI Vector Search Certified Professional** - RAG, vector indexing, similarity search, Select AI, GoldenGate.
- **Bajaj Finserv Kudos Award** - Owning & leading the enterprise-wide high-stakes production replatforming to Databricks.