

KISHOR KUMAR PAROI

AI Systems Architect | (Dev/ML/LLM/AI)Ops | Founder@KoomAI | FDE(LangGraph, Kubernetes, Kubeflow, GCP, Hybrid RAG<35ms, Agent Mesh, AI Security) | vLLM, Finetune | Data Engineer(Azure, AWS, Databricks, Snowflake) | Rust SIMD

Founder @ KoomAI (Live: koom.kishor.tech | GitHub) • B.Sc. in CSE @ RUET • Dhaka, Bangladesh / Global Remote

Core Architect Pillars: Multi-Cloud Agentic AI (MCP) | Distributed MLOps/LLMOps | High-Throughput SIMD Scoring | Zero-Trust DevSecOps IaC

1703053@student.ruet.ac.bd | kishor.ruet.cse@gmail.com | +880 1724-714156 | koom.kishor.tech | Kishor.tech | macos-portfolio | 3d | hacker | 15m Zoom | linkedin | github | x.com/Mr_paroi

EDUCATION

Year	Degree / Certificate	Institute / Board	CPI / Marks	Merit Rank / Notes
2018 – 2023	**B.Sc. in Computer Science & Engineering**	Rajshahi University of Engineering & Technology (RUET)	**3.25 / 4.00 (D.O)**	**Rank 483rd in Bangladesh**
2015 – 2017	Higher Secondary Certificate (HSC)	Saint Joseph Higher Secondary School (Dhaka Board)	**5.00 / 5.00**	**Rank 85th** in Dhaka Board
2013 – 2015	Secondary School Certificate (SSC)	S.J Govt Model High School (Gopalganj, Dhaka Board)	**5.00 / 5.00**	**Rank 90th** in Dhaka Board

WORK EXPERIENCE

phi0.ai — July 2026 – Oct 2026
Forward Deployed AI Engineer (Production RAG, GenAI & MLOps) — *Europe | Remote*

- Orchestrated forward deployment of **5 concurrent autonomous AI agents** executing multi-step tools, cyclic state machines (**LangGraph**), and dynamic routing; attained **99.4% task completion** under complex enterprise workloads, accelerating client onboarding by **4x**.
- Devised high-throughput **Hybrid RAG engine** combining dense HNSW vector search, sparse BM25, and multi-stage re-ranking (**FlashRank ONNX + Cohere v3.5**) with semantic caching (**<35ms latency**); curtailed hallucinations by **46%** and diminished LLM token expenditure by **$18K/month**.
- Codified continuous **MLOps & GenAI flywheel** featuring NeMo Guardrails, autonomous PII scrubbing, and canary model rollouts on **Kubernetes**; eradicated model regressions and spearheaded phi0.ai's enterprise AI transformation, securing **$1.4M in enterprise contracts**.

Samsung AI Core / Care Health — March 2026 – June 2026
Data Engineering & AI Systems Intern — *Gurgaon / Remote*

- Formulated batch & streaming pipelines via **AWS S3, EMR, Athena, PySpark**; slashed operational compute overhead by **$4.2K/run (34% savings)** via event-driven AWS Lambda/SQS asynchronous worker dispatch.
- Accelerated 8B/70B LLM inference on NVIDIA H100 GPUs via **FlashAttention-2** and custom fused CUDA kernels; realized **3.8x throughput** on >32K context windows, mitigating p99 latency by **54%**.
- Spearheaded multi-node fine-tuning with **PyTorch FSDP and DeepSpeed ZeRO-3** alongside HIPAA-compliant de-identification; trimmed training wall-clock time by **42% ($14K saved/run)** and preparation overhead by **65%**.
- Deployed real-time health insurance claim fraud detection pipeline over **AWS Kinesis and OpenTelemetry**; preempted false-positive claim investigations by **52%**, safeguarding **$380K annually across high-risk audits**.

Independent AI Specialization & Engineering Sabbatical — Oct 2025 – Feb 2026
Hands-on Immersion (Full-time career break dedicated to mastering mathematical AI foundations, Big Data & MLOps) — *Self-Directed / Remote*

- **Mathematical Foundations:** Rigorous deep dive into **Linear Algebra** (SVD, Eigenvalues, PCA, Matrix Decompositions) and **Multivariate Calculus** (Gradient Descent dynamics, Backpropagation, loss manifolds, convex optimization).
- **Hands-on ML, DL & NLP:** Crafted architectures from first principles in **PyTorch** (Transformers, Multi-Head Self-Attention, CNNs, tokenizers, custom loss functions), mastering modern representation learning.
- **GenAI, RAG & Agentic AI:** Built production pipelines using **Hybrid RAG (HNSW + BM25 + Cross-Encoder Reranking)**, **LangGraph cyclic state machines**, FastMCP tool execution, and NVIDIA NeMo guardrails.
- **Big Data Engineering & Lakehouses:** Mastered distributed data systems & Medallion Lakehouses (Zach Wilson, Mayank Agarwal) across **Apache Spark, Delta Lake, Snowflake, dbt Core SCD-2, Apache Kafka** streaming, and Debezium CDC.
- **Production MLOps:** Instituted continuous experimentation and deployment pipelines using **MLflow, DVC, Feast Feature Store**, Docker, and Kubernetes for low-latency serving.

kepi.ai — Jan 2022 – Sep 2025
Senior Full Stack Engineer (3 yrs 9 mos) — *Rajshahi, Bangladesh (Remote)*

- Architected zero-downtime microservices migration to **Kubernetes on AWS EKS**; scaled platform throughput by **300%** while curtailing p95 latency by **48%** at 99.99% multi-region availability.
- Spearheaded enterprise AI initiatives: devised autonomous multi-agent pipelines and scalable LLM microservices, driving a **22% sustained increase in customer LTV**.
- Fortified zero-trust mesh infrastructure with **Istio mTLS and fine-grained RBAC**; precluded unencrypted inter-service traffic and diminished cross-zone egress costs by **38%**.
- Designed event-driven asynchronous processing core using **Redis Streams and Celery**; sustained **15M+ daily async background tasks** with zero message loss under 4x traffic surges.

foodpanda — 2021 – 2022
Full Stack Engineer — *Dhaka / Remote*

- Crafted high-traffic checkout microservices with **Next.js/MERN**; trimmed latency by **34%** and sustained **100K+ daily peak orders** at 99.98% uptime, averting **$120K in checkout cart abandonment**.

FOUNDER EXPERIENCE & STARTUP PLATFORM: KOOMAI

KoomAI — Enterprise Asynchronous Video Intelligence Platform — Jan 2025 – Dec 2025
Founder & Principal Architect | Orchestrating Desktop, Chrome Extension, Backend & Full-Stack Next.js App with AWS, CDN

🚀 **Live: koom.kishor.tech** / **Full-Stack Next.js** / **Desktop Client** / **Chrome Extension** / **Backend API**

- **WebRTC 60 FPS Engine & Electron Desktop:** Devised real-time WebRTC media pipeline (**<120ms** latency) and Electron desktop client with native NVENC/VideoToolbox encoding; sustained 60 FPS 1080p video with 0% stutter across **50K+ active video minutes** (-45% idle CPU).
- **Autonomous 20-Agent Mesh & Video Copilot:** Deployed 20 concurrent **LangGraph** specialist agents (**<35ms p99**) and interactive video chatbot copilot generating 1-click GitHub PRs, Jira BDD stories, and Playwright E2E tests over Whisper transcripts.
- **Personal NotebookLLM Studio & AI Job Recon:** Built conversational two-host audio podcast generator, Principal Architect system design drills (Kafka, CDC, Feast, Saga), and automated **ai-job-search** Trojan Horse proposals with tailored ATS CVs.
- **Pomodoro Chrome MV3 & Client Invoicing:** Engineered screen-recording extension with integrated **Pomodoro focus sprints**, tab usage monitor, and 432Hz meditation audio; coupled video deliverables with instant **Stripe client invoicing** ($24K MRR pilot).

TECHNICAL PROJECTS • 🚀 DEPLOY: CLICK BUTTON TO COPY BASH COMMAND & RUN IN TERMINAL (BOOTSTRAP.SH | DEPLOY.SH | $0.00 CLEAR.SH)

1. Zomato AI Lakehouse & Multimodal Dispute Platform — Sept 2026 – Present
Lakehouse (35M+ Rows), Feast Feature Store, dbt SCD-2 & CV Engine

🚀 Deploy / GitHub: zomato-analysis

- Architected enterprise Lakehouse on **Snowflake & PySpark Delta Lake** processing **35M+ orders (2.3 GB)** with **dbt Core SCD-2**; curtailed warehouse compute by **42% ($18K/mo savings)**.
- Devised real-time feature streaming via **Airflow, Debezium CDC WAL**, and **Feast Store** over Redis; lowered serving latency to **<5ms**, unlocking **+$340K quarterly revenue lift**.
- Operationalized multimodal CV dispute arbitration (**dHash** anti-fraud, receipt OCR) with **G-Eval CoT**; neutralized fraudulent claims by **68%**, safeguarding **$1.2M in annual partner margins**.
- Deployed decentralized **CQRS and Transactional Outbox pattern** with Apache Kafka event streams, **Great Expectations** validation, and **AWS S3 Glacier** archiving ($7.8K/mo saved).
- Mechanized real-time **Presidio PII redaction and NeMo Guardrails** across review streams; guaranteed 100% GDPR/DPDP data privacy compliance with zero PII leaks into the gold layer.
- Instituted continuous self-healing data observability & reconciliation via Monte Carlo principles and Prometheus alerts; eradicated silent schema corruption across 14 ingestion channels, protecting **$450K in merchant payouts**.
- Formulated resilient dead-letter queue (DLQ) automated replay semantics with **AWS SQS & Kafka consumer groups**; recovered **$210K in transient webhook timeouts** with zero order transaction loss.

2. Hybrid Anime Intelligence & Continuous LLMOps Platform — Aug 2026 – Present
Polyglot Mesh (Go/Rust), RSLoRA Flywheel & Dynamic Multi-LoRA vLLM

🚀 Deploy / GitHub: Hybrid-Anime-Recommender

- Polyglot mesh: **Go API Gateway**, **Rust SIMD** vector scoring, and **FastAPI** AI core over **Kafka**; curtailed scoring latency from 25ms to **<0.5ms (50x speedup)**, saving **64% CPU cost**.
- Pioneered hybrid retrieval synthesizing **Dual-Tower NCF, ChromaDB HNSW**, and **BM25** via Reciprocal Rank Fusion (**NDCG@10 = 0.94**); catalyzed a **28% lift in user session retention**.
- Served adapters via **vLLM Multi-LoRA dynamic multiplexing** on 1x GPU with **RSLoRA flywheel**; curtailed GPU spend by **68%** while elevating token throughput to **3.4x**.
- Codified headless **Kubeflow v2 & MLflow continuous training pipeline** with DVC data versioning, NeFTune regularization, and stringent model quality gating.
- Orchestrated multi-cloud GitOps continuous delivery via **ArgoCD and Kustomize** across **AWS EKS, GCP GKE, and Azure AKS**; empowered KEDA autoscaling from 1 to 24 pods in <45s under 15K req/s spikes.
- Implemented real-time embedding concept drift observability with **Evidently AI & Prometheus**; detected semantic divergence within **90 seconds**, triggering automated LoRA adapter hot-swapping.

3. Enterprise AI Security & Production Agentic RAG Platform — Oct 2026 – Present
Defense-in-Depth AI Gateway, NVIDIA NeMo Guardrails & LangGraph

🚀 Deploy / GitHub: AI_Security_Production_App

- Built zero-trust defense with **BotShield WAF and NVIDIA NeMo Guardrails**; blocked **95%+ of prompt injections & jailbreaks** (benchmarked via **Microsoft PyRIT 0.14.0**).
- Formulated stateful agentic core in **LangGraph** with MemorySaver checkpoints and **Portkey Gateway**; saved **$8.4K/month (42%)** via **<50ms semantic similarity caching**.
- Standardized tool execution via **FastMCP** with **Human-in-the-Loop (HITL)** gates and Logfire/LangSmith tracing; guaranteed **100% audit compliance** with zero unauthorized tool runs.
- Architected hierarchical multi-agent supervisor coordinating 5 specialized sub-agents (Ingestion, Safety, Retrieval, Synthesis, Verification) with **cyclic retry loops and state rollbacks**; boosted multi-hop reasoning accuracy by **37%**.
- Instituted high-precision two-stage retrieval combining **Jina v3 1024-dim dense vectors**, BM25 sparse search, and **local ONNX FlashRank cross-encoders** with **Cohere Rerank v3.5** fallback; garnered **96.4% Context Precision** at sub-35ms p95 latency.
- Constructed automated red-teaming CI/CD evaluation harness executing **500+ adversarial prompt payloads** per release via **PyRIT**; intercepted 100% of OWASP LLM Top 10 vulnerabilities prior to production.

4. Smart Manufacturing Industrial AI & Multi-Cloud MLOps Platform **Oct 2026 – Present**
Polyglot (Rust/Go/Next.js), PyTorch Autoencoders, GKE/K3s & ArgoCD GitOps Deploy / GitHub: smart-manufacturing

- Deployed IoT anomaly detection (**PyTorch Autoencoders**) and defect inspection (**CNN, <20ms**); mitigated part escape defect rate by **31%**, saving **$420K annually in scrap waste**.
- Built polyglot edge architecture: **Rust SIMD** edge vibration pre-filter (:8082), **Go cloud-native gateway** with rate-limiting (:8085), **Next.js 15** industrial cockpit (:3000), and **PyTorch AI core** (:8000).
- Formulated resilient Kubernetes: **GKE multi-zone cluster with seamless fallback to host-native K3s CNCF engine** for zero-dependency standalone execution on EC2.
- Orchestrated 12-stage declarative CI/CD via **Jenkins, SonarQube, Nexus 3, and ArgoCD GitOps** across AWS & GCP; curtailed deployment cycles by **96% (4h to 8m)** with zero-downtime failover.
- Pioneered single-command zero-touch cloud bootstrap (curl | bash) with deterministic **$0.00 teardown (clear.sh --force)**; coupled with **EFK Stack and Prometheus/Grafana (:3001)** real-time telemetry.
- Sustained edge offline resilience via **local SQLite write-ahead caching and asynchronous MQTT broker sync**; averted telemetry loss during plant network disconnects across **120+ simulated IoT sensors**.

5. Nexus-AI: Enterprise Multi-Cloud Lakehouse & Autonomous Agent Mesh **Sept 2026 – Present**
Multi-Cloud Lakehouse (Bronze-Silver-Gold), 5-Agent Mesh, Go Gateway & Next.js 16 Deploy / GitHub: Nexus_AI

- Architected security-first 5-agent mesh (Data NL2SQL, Business Analyst, MLOps Supervisor, OWASP Firewall, Voice Alerts) over **LangGraph**; accelerated ad-hoc enterprise analytics turnaround by **82%**.
- Constructed 3-tier Medallion Lakehouse (Bronze Parquet → Silver 32-point DQ validation → Gold Star Schema marts) with **DuckDB**; compressed downstream dashboard query latencies from 4.5s to **<120ms**.
- Crafted **Go 1.22 API Gateway** with JWT edge authentication and token-bucket rate limiting paired with Express BFF and **Next.js 16** glassmorphic cockpit; sustained 12K req/s with zero memory leaks.
- Codified multi-cloud infrastructure via **Terraform** across **AWS (EKS, S3, Glue), GCP (GKE, BigQuery), and Azure (AKS, Synapse)**; compressed environment provisioning time by 90% via master launcher.
- Governed attribute-based access control (ABAC) and row-level security (RLS) across **Medallion Delta tables**; enforced strict multi-tenant data isolation and audit trails across **40+ analytical business units**.

6. Advanced AI Chatbot: Enterprise Autonomous Agent & HITL Platform **Jul 2026 – Present**
LangGraph Cyclic State Machine, FastMCP, HITL Financial Gating & BotShield WAF Deploy / GitHub: Advanced_AI_Chatbot_Project

- Built enterprise conversational agent using **LangGraph cyclic state machines** with PostgreSQL checkpointer; empowered fault-tolerant multi-turn thread persistence, dynamic rollbacks, and **sub-50ms session recovery**.
- Standardized isolated tool execution using **Model Context Protocol (FastMCP)** with strict **Human-in-the-Loop (HITL)** approval gates on financial actions (>$50 refunds/cancellations); eradicated **100% of unauthorized transaction risk**.
- Pioneered **BotShield WAF** middleware featuring honeypot traps (/.env, /wp-admin), scanner fingerprint dropping, and sliding-window rate limiting; intercepted **99.4% of adversarial probes**.
- Operationalized two-stage hybrid retrieval with **Qdrant (BGE-Large dense vectors) and BM25 sparse keyword matching** with real-time Server-Sent Events (SSE) token streaming; elevated response relevance to **94.8%**.
- Architected semantic intent caching layer with **Redis & Cosine similarity gating (<0.15 threshold)**; served 38% of repetitive customer intents in **<12ms**, curtailing LLM token billing by **$3.2K/month**.
- Devised adaptive conversational memory with **sliding-window token compaction & vector summary checkpointing**; diminished long-context token footprints by **44% across 35+ turn dialogs**.

7. Data Agent: Enterprise Multi-Agent Business Intelligence & SQL System **Sept 2026 – Present**
Hierarchical LangGraph Mesh, Self-Reflective SQL Security Judge & ReAct ETL Deploy / GitHub: Data-Agent

- Architected a **LangGraph supervisor router** classifying business queries into specialized SQL and ETL analyst subgraphs; garnered **96.2% routing accuracy** across multi-domain operational requests.
- Instituted dynamic PostgreSQL schema introspection with an **independent LLM security judge**; sanitized generated SQL and **intercepted 100% of destructive DDL/DML injection attacks**.
- Built autonomous **ReAct tool-calling agent** for file extraction, pandas data transformation, and sandboxed python execution; compressed ad-hoc data pipeline development time by **75%**.
- Integrated **self-healing SQL execution** with automatic syntax retry & schema reflection; programmatically corrected **89% of initial query execution errors** without human intervention.
- Orchestrated multi-dialect SQL transpilations supporting **PostgreSQL, Snowflake, and BigQuery** with proactive **EXPLAIN cost estimation**; prevented runaway analytical queries exceeding cost ceilings.

8. Advanced Network Security Medallion Lakehouse & Distributed ML Platform **Aug 2026 – Present**
Azure Data Factory, ADLS Gen2, Databricks Spark, MongoDB Atlas, Synapse CETAS & Sub-4ms ML Deploy / GitHub: Advanced_ETL_Project

- Architected enterprise Lakehouse: orchestrated batch extraction via **Azure Data Factory (ADF v2)** dynamic lookup and parallel ForEach loops, landing raw logs into **ADLS Gen2 Bronze layer**.
- Conducted large-scale ETL in **Azure Databricks (Apache Spark)** with Azure AD OAuth 2.0 Client Credentials, cross-joining real-time operational metadata from **MongoDB Atlas**.
- Materialized analytical serving layer in **Azure Synapse Analytics Serverless SQL** using Managed Identity scoped credentials and **CETAS (*Create External Table As Select*)** for instant Power BI/Fabric reporting.
- Built data validation engine with schema drift traps, KNN imputation + RobustScaler, and multi-model benchmark (**XGBoost, CatBoost, Random Forest** delivering **98.4% detection accuracy**).
- Containerized sub-4ms **FastAPI ONNX inference serving** with single-command zero-touch bootstrapper (curl | bash) and deterministic **$0.00 teardown (clear.sh)**; mitigated p99 latency by **11.8x**.
- Implemented dynamic partition pruning and **Z-Order clustering on ADLS Gen2 Delta Parquet files**; expedited complex analytical join queries by **6.2x across 1.4B security log records**.

9. AWS SageMaker End-to-End MLOps Pipeline & Governance Platform **Oct 2026 – Present**
SageMaker Pipelines (DAG), Managed Spot Training, Model Registry & Serverless Serving Deploy / GitHub: Aws-Sagemaker

- Designed **SageMaker DAG Pipeline** (ProcessingStep → TrainingStep → ConditionStep → ModelStep) with **AWS Managed Spot Training**, curtailing model training compute costs by **70%**.
- Instituted programmatic quality evaluation gating (≥ 85% holdout accuracy) for promotion into **SageMaker Model Registry**, deploying to real-time and serverless endpoints with standard inference contracts (model_fn, input_fn, predict_fn, output_fn) and FastAPI fallback.
- Configured **DataCaptureConfig** streaming 100% of live inference payloads to Amazon S3 for baseline statistical drift detection, anomaly alerting, and programmatic retraining triggers.
- Devised multi-model endpoint deployments with **autoscaling policies (target tracking at 70% CPU/GPU)**; sustained **2,500 req/s** with zero cold-start timeouts under spiky production loads.
- Hardened inference endpoints with **AWS KMS customer-managed keys (CMK) envelope encryption and TLS 1.3 mutual auth**; achieved complete compliance with **NIST AI RMF governance mandates**.

10. Cloud-Native Real-Time MLOps: KServe, AWS ECS Fargate & Kubernetes **Aug 2025 – Present**
KServe v1beta1, AWS ECS Fargate, Kubernetes, Terraform IaC & S3 Deploy / GitHub: realtime-mlops / Terraform-EKS

- Deployed serverless model inference using **KServe v1beta1 InferenceService** on Kubernetes backed by Amazon S3 model storage; facilitated dynamic scale-to-zero autoscaling, curtailing idle GPU/CPU compute spend by **65%**.
- Architected containerized microservices orchestration on **AWS ECS on Fargate** with Application Load Balancers and modular **Terraform IaC**; accelerated infrastructure provisioning by **80%** with zero server maintenance overhead.
- Hardened end-to-end container security with **Trivy CVE scanning and AWS KMS envelope encryption**; ensured **100% programmatic CI/CD gating** with zero critical vulnerability deployments.
- Configured automated canary traffic routing (**90/10 split**) via Envoy sidecars with Prometheus-driven automatic rollback on p99 latency regressions (**>25ms**).

11. Core Systems Engineering: Rust Git Engine & POSIX Terminal Shell **Aug 2024**
Bare-Metal Systems Programming in Rust, Concurrency & Process Pipelines GitHub: codecrafters-git-rust

- Crafted Git internals in **Rust** (SHA-1 object graph, streaming zlib decompression with SIMD, delta packfiles); matched C Git performance within 3% while guaranteeing **100% compile-time memory safety**.
- Built POSIX shell with async signal multiplexing, fork/exec pipelines, and I/O redirection; clocked **sub-microsecond execution latency** under heavy concurrent process loads with **73% lower peak heap allocation**.

ACHIEVEMENTS & HACKATHONS

- **Competitive Programming & DSA:** Ranked **232nd in Codeforces Div-2** (**Blue Coder**), **5★ on CodeChef**, **17th in IUPC**; Qualified in **Meta Hacker Cup Round 3**; 2000+ DSA, participated system design hackathon.
- **Champion | Technobeast AI Agents Hackathon:** Pioneered Qdrant vector optimization with negative embeddings, trajectory matching with transformers, and multimodal API divergence detection.
- **7th Rank | AI Security Hackathon:** Ranked in top tier for real-time prompt injection firewall (**97.2% accuracy**, **<30ms** latency).
- **National Merit Scholarships:** Conferred merit scholarships for 85th Rank in Dhaka Board (HSC) and 90th Rank in Dhaka Board (SSC).

LEADERSHIP & MENTORING

- **In-Person AI Agents Building Masterclass:** Conducted immersive hands-on in-person workshops on architecting autonomous AI Agents, cyclic state machines (**LangGraph**), tool calling, and production deployment.
- **Technical Mentorship & Community Impact:** Mentored **100+ students & engineers** across **Machine Learning, AI Agents, DevOps, and MLOps**, leading code reviews, production architecture design sessions, and career roadmapping.

TECHNICAL SKILLS

- **Languages & Frameworks:** Python, C/C++, Rust (Actix-web), Go (Gin/Fiber), SQL, Bash, TypeScript, JavaScript, Java (Spring Boot), FastAPI, Express.js, NestJS, Next.js 16, Streamlit
- **AI, ML & LLMOps:** PyTorch, vLLM, LangGraph, LangChain, HuggingFace, LangSmith, LangFuse, DeepEval, TRL, PEFT (RSLoRA, QLoRA), NeFTune, MLflow, Feast, Kubeflow, AWS SageMaker, KServe, DVC
- **Big Data & Lakehouse:** Snowflake, Databricks, Azure Data Factory (ADF), Azure Synapse Analytics (CETAS), ADLS Gen2, dbt Core, PySpark, Delta Lake, Apache Kafka, Apache Airflow, Debezium CDC, DuckDB, ChromaDB, Qdrant
- **Cloud Platforms & Infra:** AWS (EKS, S3, KMS, EMR, Athena, ECS, Fargate, Bedrock, SageMaker), GCP (GKE, BigQuery, GCE, VertexAI, Gemini AI), Azure (Databricks, Synapse, Fabric, vLLM, AKS), Terraform, Ansible, Helm, ArgoCD, Sonatype Nexus, Docker, Kubernetes (GKE, EKS, K3s)
- **DevSecOps & Observability:** NVIDIA NeMo Guardrails, PyRIT RedTeaming, BotShield WAF, SonarQube, Trivy, Jenkins, GitLab CI/CD, GitHub Actions, Prometheus, Grafana, EFK (Elasticsearch/Kibana), Redis
- **WebRTC, 3D Graphics & Full-Stack Systems:** WebVR, WebRTC, Three.js, React Three Fiber, ElectronJS, Next.js 16, Prisma, Drizzle ORM, Stripe Connect, PostgreSQL, MongoDB Atlas

CERTIFICATIONS & SPECIALIZATIONS

- **Deep Learning Specialization (DeepLearning.AI & Stanford – Andrew Ng):** Neural Networks, Hyperparameter Tuning, CNNs, Sequence Models, Attention Mechanisms & Transformers.
- **Machine Learning Specialization (Stanford Online & DeepLearning.AI – Andrew Ng):** Supervised Learning, Advanced Learning Algorithms, Unsupervised Learning, Recommenders & Reinforcement Learning.
- **Enterprise Production MLOps & AI Systems Architecture:** End-to-End MLOps Pipelines (Krish Naik/CampusX), Enterprise AI Systems Architecture (Vivian Aranha), Agentic-AI, n8n (Mayank Agarwal), Computer Vision (Sourangshu Paul), AI Data Analyst (Ansh Lamba), Automation & MCP (Ed Donner)
- **Distributed Systems, Big Data & AI System Design:** High-Throughput Distributed Systems & AI Architecture (Arpit Bhayani, Gaurav Sen, ByteByteGo, Hussein Nasser), Big Data Engineering & Lakehouses (Mayank Agarwal, Zach Wilson), LLM Foundations (Karpathy, Aurimas), LeetCode algorithmic mastery.
- **Digital Portfolios & Systems Architecture Showcases:** vantage.kishor.tech | GitHub: vantage-portfolio — Clean academic AI architecture showcase presenting complete 38-module AI curriculum, multi-agent lakehouses, sub-35ms semantic search & FastMCP governance. • 3d.kishor.tech | GitHub: re-invent — React 19.2, Rolldown Vite, GSAP 3D choreography & Supabase real-time visitor telemetry.
- **Core CS & Engineering Foundations (RUET):** Advanced Data Structures & Algorithms, Object-Oriented Programming (C++/Java), Operating Systems & System Programming, Database Systems (DBMS), Distributed Systems & Cloud Computing, Computer Networks & Internet Protocols, Theory of Computation & Automata, Compiler Design, Software Engineering & System Analysis, Computer Architecture, Artificial Intelligence, Microprocessors & Interfacing, Discrete Mathematics, and Network Security & Cryptography.