Korivi Chetan Kumar

korivichetan5@gmail.com | +91-7981685655 | GitHub | Hugging Face | LinkedIn

PROFESSIONAL SUMMARY

AI/ML Engineer and Data Developer with expertise in LLM fine-tuning, Retrieval-Augmented Generation, and Big Data pipelines. Deployed 6+ production-ready applications, improving accuracy by 12% over baselines, cutting incident response times by 40%, and delivering predictive systems with $R^2 = 0.73$. Skilled in full-stack AI, cloud deployment (AWS, GCP, Azure), Explainable AI, and Data Engineering.

EDUCATION

Degree	Specialization	Institute	Year	CPI
B.Tech	*Computer Science & Engineering*	VIT-AP University	2021-2025	8.02
BIEAP	*Physics, Chemistry, & Mathematics*	Narayana College	2021	7.44
SSC	-	Indus School	2019	8.8

WORK EXPERIENCE

- **AI-Native Data Engineer – iCustomer** *Full-Time | Mar 2026 – Ongoing*
  - Architected scalable Big Data infrastructure and synthetic databases for proof-of-concept environments, driving enterprise-grade product demonstrations and ETL/ELT pipeline optimization.
  - Engineered robust data engineering pipelines to process and validate high-volume datasets (95K+ records), leveraging external APIs (Bright Data) for precise data segmentation and enrichment.
  - Developed enterprise reporting dashboards by writing customized **LookML** code for **Looker**, delivering actionable business intelligence and tailored analytics for client AEO.
  - Delivered 59+ successful technical tasks, focusing on precise fit score evaluations, complex SQL query optimization, and generating comprehensive workflow documentation.
  - Resolved critical backend bugs and optimized user onboarding flows by implementing domain-matcher guards and efficiently managing pull request merge conflicts.
- **Machine Learning Engineer – GENIE AI, INC. Canada** *Contract | Sep 2025 – Jan 2026*
  - Built a multimodal, multilingual AI assistant with LLM-based reasoning, voice integration (TTS), and a custom backend flow engine for intent detection and stateful form workflows.
  - Deployed a private HuggingFace inference pipeline (merged Qwen2-0.5B) behind production REST APIs, implementing a fallback chain (HF → OpenAI → Groq) to ensure high reliability and deterministic JSON outputs.
  - Engineered external RAG data pipelines for real-time semantic search, including normalization, summarization, and dynamic entity lookups.
  - Spearheaded product enhancements by building functional prototypes for senior stakeholders and authoring comprehensive architectural documentation, accelerating leadership approval for new feature rollouts.
- **Motion Graphic Designer – Information Data Systems (IDS)** *Internship | Sep 2023 – Mar 2024*
  - Delivered motion graphics and interactive visual assets aligned with product and UX requirements.
  - Collaborated with cross-functional teams to translate design concepts into production-ready visuals.
  - Applied Transcription workflows to asset creation Team and reduce turnaround time.
  - Recorded and documented **Minutes of Meetings (MoMs)** and communicated deliverables directly with senior stakeholders and leadership.
  - Implemented iterative design changes based on executive feedback to ensure alignment with business and marketing objectives.
  - Acted as a **content strategist**, designing high-impact marketing visuals and campaigns in alignment with current industry trends.

PROJECT EXPERIENCE

- **WatchTowerAI – Enterprise Log & Incident Orchestrator** *2025*
  Live Demo on Hugging Face
  - Built an AI-native observability system using **LangChain** to classify logs and automate workflows; reduced incident response time by **40%**.
  - Integrated backend services with **AWS S3** and Slack, demonstrating ability to debug distributed systems and ship resilient tooling.
- **MedIntel – RAG-Enabled Biomedical QA System** *2025*
  - Architected a domain-specific QA solution using **SciBERT** and **FAISS Vector DB**, improving answer retrieval accuracy by **12%**.
  - Delivered a production-grade Dockerized application, solving complex information retrieval challenges for medical professionals.
- **CodeMentor-AI – Interactive Coding Assistant** *2025*
  Live Demo on Hugging Face
  - Fine-tuned **Flan-T5** on 20k+ samples to create a developer-focused assistant; solved deployment blockers (Git LFS) to release a stable MVP.
  - Focused on developer experience (DevEx) by optimizing response relevance for technical interview scenarios.
- **LLMGuard – AI Security & Governance Toolkit** *2025*
  Live Demo on Hugging Face
  - Developed a security layer for Large Language Models, achieving **92% precision** in detecting prompt injections and jailbreaks.
  - Aligns with enterprise requirements for deploying secure, regulated AI systems in production environments.
- **StackGenie – Cloud Architecture Analytics** *2025*
  - Connected **Google BigQuery** data pipelines to a frontend dashboard, visualizing 20k+ cloud release notes to inform technical strategy.
  - Demonstrated product sensibility by building intuitive filters and search capabilities for complex technical data.
- **PriceSense – Predictive Modeling with Explainability** *2025*
  - Implemented a Random Forest pipeline with **SHAP explainability** ($R^2 = 0.73$), focusing on model transparency and drift monitoring.

RESEARCH

- **BirdZ – Computer Vision Dataset & Model** *Ongoing*
  - Curated a dataset of 6,900+ images and trained YOLOv12 (**mAP@50 = 81%**), demonstrating rigor in data handling and model evaluation.

TECHNICAL SKILLS

- **Forward Deployed Engineering:** Full Stack Development (React.js, Node.js), REST/GraphQL APIs, UI/UX Integration, System Architecture, Customer-Facing Solutions.
- **Generative AI & LLMs:** LangChain, RAG pipelines, Vector Databases (FAISS), Semantic Search, Fine-tuning, Prompt Engineering, Agentic Workflows, Hugging Face.
- **Data Engineering & BI:** ETL/ELT pipelines, Big Data processing, Looker, LookML, SQL Optimization.
- **Cloud & DevOps:** AWS, Google Cloud Platform (GCP), Docker, Kubernetes, CI/CD pipelines, Infrastructure-as-Code (Terraform), Model Observability.
- **Languages & Tools:** Python, JavaScript/TypeScript, SQL, NoSQL (MongoDB), Git, Agile/Scrum Methodologies.