Veeky Kumar

Jalandhar, Punjab, India 141001 — +91 75976 05761 — vk133162@gmail.com

linkedin.com/in/veeky-kumar — github.com/Veeky-kumar

PROFESSIONAL SUMMARY

AI Engineer with **4+ years** of experience in **Python, Machine Learning, and NLP**, including hands-on experience with **LLMs, generative AI models, and agentic frameworks**. M.Tech in AI with expertise in designing scalable AI solutions, fine-tuning transformers, and deploying cloud-based models. Strong foundation in **data science, feature engineering, and MLOps**.

TECHNICAL SKILLS

AI/ML & Generative AI: PyTorch, TensorFlow, Hugging Face Transformers, LLMs (GPT, BERT), GANs, VAEs, Diffusion Models, Fine-tuning, RLHF, Transfer Learning, RAG, NLP, Computer Vision, Agentic AI Frameworks

Programming: Python, C++, JavaScript, TypeScript, SQL, FastAPI, REST APIs

Data Science: Statistical Analysis, Feature Engineering, Data Preprocessing, ETL Pipelines, OCR, Document Intelligence

Databases: MongoDB, PostgreSQL, FAISS, Pinecone, Redis, Vector Databases, NoSQL, Similarity Search

Cloud & DevOps: Azure, AWS, Docker, Kubernetes, Linux/Unix, CI/CD, MLOps, Model Deployment

Core Competencies: Distributed Systems, System Design, Optimization Techniques, Agile Development

EDUCATION

Master of Technology in Artificial Intelligence, NIT Jalandhar — CGPA: 8.04/10 — Expected 2027

Relevant Coursework: Machine Learning, Deep Learning, NLP, Computer Vision, Distributed Computing

PROFESSIONAL EXPERIENCE

Full Stack Software Development Engineer — May 2024 – Feb 2025

Protolabz eServices, Mohali, India

- Designed and deployed **scalable AI-powered distributed systems** serving **10,000+ concurrent users** with optimized database indexing, caching strategies, and advanced algorithms for improved performance
- Applied **data science methodologies** for system optimization including statistical analysis, feature engineering, and performance monitoring in fast-paced agile environment
- Collaborated with cross-functional teams to deliver full-stack solutions using modern frameworks, implementing JWT authentication, RBAC security, and conducting comprehensive code reviews
- Solved complex problems across distributed architectures in 2-week sprint cycles, ensuring seamless integration and production-ready deployment

AI/ML PROJECTS & ACHIEVEMENTS

Pathway-Enhanced Narrative Consistency Evaluator — Generative AI & LLMs — Jan 2026

GitHub — **Finalist, IIT Kharagpur Data Science Hackathon**

- **Led team presentation** at IIT Kharagpur, designing **scalable generative AI pipeline** using **LLMs for long-context reasoning** and logical consistency validation across 100,000+ word narratives, achieving 63.75% accuracy
- Implemented advanced **NLP techniques with Sentence Transformers** for semantic embeddings and **Top-K similarity search with vector databases**, optimizing retrieval accuracy for generative AI outputs
- Developed **distributed ETL pipelines** using Pathway framework for real-time data processing with fault-tolerant architecture; containerized with **Docker** and deployed on **Linux** with MLOps best practices

AI-Powered Resume Matching System — NLP & Vector Search — Mar 2025

GitHub — Live Demo

- Architected end-to-end AI system utilizing **FAISS vector database** for semantic matching, processing 1,000+ resumes with **advanced ranking algorithms** and optimized similarity search achieving **sub-200ms response times**
- Built scalable **FastAPI backend with React.js frontend**, implementing RESTful APIs and comprehensive **data engineering pipeline for document intelligence** including OCR integration and feature extraction
- Integrated **MongoDB for NoSQL storage** ensuring cost-effective, fault-tolerant architecture suitable for **cloud deployment on Azure/AWS**

Healthcare Glucose Prediction System — Time-Series ML — Ongoing

NIT Jalandhar Research Project

- Developing ML system using **LSTM neural networks** for time-series forecasting with focus on **feature engineering, statistical analysis**, and data preprocessing for healthcare applications
- Implemented evaluation metrics and applied **transfer learning, domain adaptation** techniques; following **agile methodologies with CI/CD pipelines** ensuring compliance with data privacy and ethical AI considerations

KEY QUALIFICATIONS

- Strong understanding of **Azure OpenAI GPT models** and Hugging Face ecosystem for leveraging pre-trained models and enhancing generative AI capabilities with domain-specific fine-tuning
- Experience with **optimization techniques** including model compression and **trusted AI practices** ensuring fairness, transparency, and accountability in AI systems
- Excellent problem-solving and analytical abilities with proven track record in translating **business requirements into technical AI solutions** and collaborating with stakeholders at various levels