KAVYA SREE CHANDHI

Generative AI Engineer

Mail id: chandhikavyasree@gmail.com | **Contact Number:** +1 9406298898 | **Location:** Herriman, Utah, 84096

LinkedIn: https://www.linkedin.com/in/chandhi-kavya-sree/ |

GitHub: https://github.com/kavya-sree-chandhi |

Portfolio: https://kavya-sree-chandhi.github.io/ |

PROFESSIONAL SUMMARY

Generative AI & ML Engineer with 3+ years of experience delivering enterprise AI solutions across **education, finance, healthcare, and automotive** domains. Specialized in **Generative AI, Conversational AI, Multi-Agent Systems, RAG pipelines, and prompt engineering**, with hands-on expertise in **LangChain, LangGraph, Hugging Face, OpenAI, and Ollama**. Led the development of **UWorld's AI-Enhanced Test Prep Platform**, boosting student engagement by **40%** through AI-driven essay grading, adaptive simulations, and gamified learning. Skilled in **Python, SQL, and deep learning frameworks (PyTorch, TensorFlow, Keras)**, with proven impact in **fraud detection, risk analytics, and medical imaging**. Experienced in **end-to-end AI/ML delivery** — from data engineering and model training to **cloud deployment (Azure, AWS, Oracle, GCP), MLOps (MLflow, Kubernetes, CI/CD), and production monitoring** — delivering scalable, compliant, and high-performance AI systems.

TECHNICAL SKILLS:

- **Programming Languages:** Python, C, JavaScript, SQL
- **Generative AI:** LLMs, RAG, Prompt Engineering, Fine-Tuning, AI Agents, LangChain, LangGraph, Hugging Face, OpenAI, Ollama, DALL·E, Vector Databases (Milvus, ChromaDB, Pinecone)
- **Machine Learning & Data Science:** Supervised/Unsupervised Learning, Predictive Modeling, Feature Engineering, Cross-Validation, Hyperparameter Tuning, Statistical Analysis, Scikit-learn
- **Deep Learning & Computer Vision:** ANN, CNN, RNN, LSTM, ResNet, VGG, MobileNet, Transfer Learning, GANs, Autoencoders, OpenCV, PyTorch, TensorFlow, Keras
- **NLP:** Sentiment Analysis, Summarization, NER, Emotion Detection, Text Classification, spaCy, NLTK
- **Recommender Systems & Optimization:** Collaborative Filtering, Matrix Factorization, Bayesian Optimization, Optuna, Hyperopt
- **Big Data & Streaming:** Apache Spark, PySpark, Hadoop, Kafka, Spark Streaming, Snowflake, Redshift, Azure Data Lake, BigQuery
- **MLOps & DevOps:** Docker, Kubernetes, Jenkins, GitHub Actions, MLflow, W&B, DVC, Terraform, Airflow, Helm, CI/CD Pipelines
- **Responsible AI & Compliance:** SHAP, LIME, Fairlearn, HIPAA, GDPR, PCI DSS, RBI Guidelines
- **Visualization & Analytics:** Power BI, Tableau, Grafana, Matplotlib, Seaborn, Plotly, Dash, Excel, EDA, Predictive Analytics
- **Cloud Platforms:** Microsoft Azure (Prompt Flow, Bot Framework, Cosmos DB, Data Factory), AWS (SageMaker, Lambda, EC2, S3, RDS, Glue), Oracle Cloud (OCI Generative AI, Autonomous DB), GCP (Vertex AI, BigQuery, Firestore, Cloud Functions, Dataflow)
- **Databases:** SQL, MySQL, PostgreSQL, MongoDB, DynamoDB, NoSQL, Neo4j
- **API & Web Development:** RESTful APIs, FastAPI, Flask, Streamlit, Django
- **Collaboration & Tools:** Git, Jira, Confluence, Jupyter Notebook, VS Code, PyCharm, Anaconda, Orange

WORK EXPERIENCE:

Client - UWorld, USA — Aug 2024 – Present

Role - Generative AI Engineer

Project Description: Leading the development of an **AI-Enhanced Test Prep Platform** for SAT, ACT, and AP exams, integrating **Generative AI–driven tutoring, automated essay grading, adaptive test simulations, quiz generation, and gamified learning** to improve student engagement and learning outcomes.

Key Responsibilities:

- Developed an **AI-driven essay grading system** aligned with College Board rubrics, ensuring consistent and efficient evaluation of student submissions.
- Delivered **real-time feedback** on structure, grammar, and content, reducing revision cycles and accelerating student improvement.
- Integrated **SAT/AP print guides into a hybrid digital platform**, enabling blended learning and expanding study material usage.
- Engineered an **adaptive test simulator** replicating official exam timing and stress conditions to improve readiness.
- Designed **gamified challenges and leaderboards**, increasing daily active student engagement by **40%**.
- Built an **AI-powered quiz generator**, reducing manual content creation by **60%** and improving question diversity.
- Leveraged **Azure OpenAI and LangGraph pipelines** to implement scalable AI workflows with 99.9% uptime.
- Applied **Retrieval-Augmented Generation (RAG) with ChromaDB** to deliver precise subject-matter answers and improve accuracy.
- Optimized latency and response quality in AI tutoring systems by implementing **prompt optimization and caching strategies**, improving response speed by 35%.
- Mentored junior engineers and collaborated with product managers to align **AI capabilities with educational goals**, accelerating roadmap execution.
- Established **end-to-end monitoring and evaluation pipelines** with Azure Application Insights and custom metrics, ensuring model performance, accuracy, and reliability in production at scale.

Client – DXC Technology, India **Aug 2022 – Aug 2023**

Role – Associate Professional Software Engineer

Project Description: Contributed to the enhancement of **Product Lifecycle Management (PLM) systems** by implementing **Python automation, predictive analytics, and machine learning models** to streamline workflows, optimize reporting, and improve system reliability for global automotive engineering teams.

Key Responsibilities:

- Automated **PLM workflows using Python**, reducing repetitive manual tasks by **30%** and streamlining operations across global teams.
- Integrated **ML models for predictive maintenance**, lowering system downtime by **15%** through early fault detection.
- Designed and implemented **user management modules** supporting 5,000+ engineering accounts with secure, role-based access.
- Optimized **SQL + Python reporting pipelines**, accelerating data extraction and decision-making.
- Built **interactive dashboards in Tableau and Grafana** for license tracking, usage, and performance insights.
- Explored **GenAI-based prototypes** to enable natural language queries on PLM data, improving accessibility for non-technical users.
- Authored **technical documentation and SOPs**, streamlining onboarding and knowledge transfer.
- Partnered with automotive engineering teams to deliver **data-driven insights** that improved PLM system adoption and reduced operational bottlenecks globally.
- Implemented **CI/CD pipelines with Docker and Jenkins**, reducing deployment cycles for PLM automation scripts from weeks to hours.
- Collaborated with cross-functional teams across IT and engineering to ensure **compliance with ISO/IEC standards**, strengthening data security and system reliability.

Client – Pelican IT Solutions Pvt. Ltd., India **Aug 2021 – July 2022**

Role – Data Analyst

Project Description: Supported a leading banking client in developing a **Fraud Detection & Risk Analytics Platform**, applying **statistical modeling, anomaly detection, ETL automation, and real-time dashboards** to identify high-risk transactions, reduce fraud losses, and ensure compliance with BFSI regulations (KYC/AML).

Key Responsibilities:

- Conducted **fraud detection analysis** using advanced statistical models (Random Forest, Autoencoders, Isolation Forest), improving detection accuracy by **20%**.
- Automated **ETL pipelines in Python (Pandas, NumPy, Scikit-learn)** to prepare and standardize high-volume financial transaction data, reducing manual effort by **30%**.
- Designed **fraud monitoring dashboards** with Flask and Grafana to provide real-time visibility into high-risk activities for stakeholders.
- Developed **fraud risk indicators** such as transaction velocity, seasonal trends, and user profiling to strengthen anomaly detection and reporting.
- Applied **predictive analytics techniques** including cross-validation and parameter tuning to ensure data-driven fraud risk assessments were robust.
- Collaborated with **compliance and risk teams** to ensure analytics outputs aligned with BFSI regulations (KYC/AML).
- Worked with **data engineering teams** to streamline pipelines and improve real-time scoring performance across fraud detection workflows.

ACADEMIC PROJECTS:

AI Research Agent for Healthcare Diagnostics

- Built an AI-powered research assistant using **Python, Streamlit, LangChain, and LLaMA3** with a **RAG pipeline (FAISS + PyMuPDF)**, incorporating citation validation to reduce hallucinations and generate structured healthcare diagnostics reports.

Career Navigation AI

- Developed an AI career copilot with **LangChain + GPT-4** that generated personalized learning paths, recommended certifications, and improved career readiness through skill gap analysis and progress tracking.

Automatic Attendance Monitoring System (Face Recognition)

- Created a **Python + OpenCV + Django** system with ML classifiers for automated attendance, role-based access, and dashboards, eliminating proxy attendance and reducing manual workload.

Image Classification for Disease Detection

- Designed deep learning models (**CNNs, VGG16, ResNet50**) for MRI and X-ray scans, achieving **98% accuracy** in brain tumor detection and improving robustness with transfer learning.

Youth Smoking and Drug Use Analysis

- Conducted **EDA and ML modeling on 8,993 records** using Python, revealing peer influence and school-based interventions as key factors in reducing substance use.

Securing Data Using Blockchain and Artificial Intelligence

- Built **SecNet architecture** with Blockchain + AI anomaly detection on a **Django prototype**, enabling patient-controlled healthcare data access and tamper-proof record management.

CERTIFICATIONS:

- **Oracle Cloud Infrastructure Generative AI Professional** – **Oracle**
- **Microsoft Azure Fundamentals (AZ - 900) - Microsoft**
- **Artificial Intelligence Fundamentals** - **IBM**
- **Google Technical Support Fundamentals** – **Coursera/ Google**

ACADEMICS:

Master of Science in Computer Science | GPA: 3.916/4 **Aug 2023 – May 2025**

University of North Texas, Denton, Texas