# Chethan H N

+91-7259578979 | chethanhn29@gmail.com | LinkedIn | GitHub | Portfolio

Experienced Data Scientist/AI Developer specializing in Machine Learning, Deep Learning, Generative AI and Large Language Models (LLMs). Proficient in Python, PyTorch, TensorFlow, NLP, CNNs, Docker, and cloud deployments. Seeking a challenging role to drive innovation and deliver impactful results in AI projects.

## PROFESSIONAL EXPERIENCE

**Data Scientist** Nov 2024– Present
*Fractal Analytics* *Bengaluru, Karnataka*

- **Optimized hardware operations using NLP models** to analyze and extract actionable insights from unstructured operational data, enhancing decision-making efficiency by 10%
- Integrated **Transformer-based models (BERT, GPT)** to automate report generation and provide actionable insights for hardware performance monitoring.
- Collaborated with cross-functional teams to align AI-driven insights with business objectives, driving data-informed decision-making across hardware operations.

**Artificial Intelligence Developer** Mar 2023 – Sep 2024
*DRDO Funded Project, VIT* *Chennai, Tamilnadu*

- Developed user behavior modeling models for real-time anomaly detection and insider threat prevention, achieving a 25% reduction in false positives compared to baseline systems.
- Utilized PyTorch, TensorFlow, and Transformers to optimize ML models for log analysis, achieving a 40% improvement in processing efficiency.
- Implemented federated learning concepts for decentralized model training, ensuring privacy and scalability across multiple endpoints, leading to a 50% increase in model training scalability.
- Led deployment of federated learning models, improving threat detection by 20% with Docker for seamless integration.

**Machine Learning Intern** Sep 2022 – Nov 2022
*Feynn Lab Services* *Remote*

- Implemented Named Entity Recognition (NER) techniques to analyze and extract key entities from customer support tickets, resulting in a 20% increase in critical issue identification and resolution.
- Developed and fine-tuned NER models using spaCy and custom-built algorithms, improving entity recognition accuracy by 15%.
- Conducted in-depth analysis on NER-extracted entities to identify patterns, trends, and customer sentiments, providing actionable insights for service enhancement strategies.

## PROJECTS

**User Behavior Models for Real-time Endpoint Security** *Python, Pytorch, SQL, Tensorflow, Docker*

- Project Objective: Develop advanced behavioral AI models for real-time anomaly detection and threat identification using host and network logs data.
- Utilized ML algorithms (RNN, LSTM), NLP, Transformers for behavior analysis, implementing Federated Learning for decentralized model training.
- Impact: Implemented individualized behavior models for each user, reducing false positives by 25% and proactively preventing insider threats and external attacks, resulting in a 30% decrease in security incidents.

**Retail Query Assistant Chatbot** *Python, Langchain, Hugging face, Chroma DB, Streamlit* *Link*

- Project Objective: Develop an end-to-end language model chatbot for handling complex natural language queries on a store's management MySQL database.
- Text Processing and Embedding: Utilize ChromaDB to extract and embed both queries and answers, forming a vector database.
- LLM Integration and Database Chain Creation: Integrate Langchain's Google Palm LLM with few shot prompts to create an SQL database chain.
- User Interface Development: Design a streamlined UI using Streamlit, allowing users to input queries and receive LLM-generated responses.
- Project Impact: Efficient Query Processing: Developed a tool to assist store management by accurately processing diverse and complex queries.

**Knowledge Assist: A Smart Information Retrieval System** *Python, Hugging Face,NLP,MongoDB,Streamlit,LLM* Link

- Developed and implemented an intelligent knowledge assistance system integrating retrieval-based and generative-based techniques, enhancing user search experience.
- Leveraged MongoDB for storage and retrieval, achieving a 30% improvement in search efficiency compared to traditional relational databases.
- Utilized open-source models like Sentence Transformer and MongoDB's vector search capabilities, resulting in a 50% reduction in resource consumption and a 40% increase in system scalability.
- Integrated Gemma for natural language generation (NLG) tasks, leading to a 40% reduction in response time for user queries.

## EDUCATION

**Bachelor of Engineering (BE)** Mysuru, Karnataka
*ATME College of Engineering, VTU University* *June. 2014 – June 2018*
*CGPA – **8.10***

## TECHNICAL SKILLS

**Programming Languages:** Python

**Database:** SQL, Vector DB, Pinecone, Chroma DB.

**Frameworks:** Pytorch, Tensorflow, Keras, Lang chain, Llama index

**Cloud Computing**: AWS Sage maker, Azure, Google Cloud Platform.

**DevOps Tools:** Docker, CI/CD pipeline (Github Actions)

**Libraries:**

- Data Science: Pandas, NumPy, Matplotlib, Seaborn, scikit-learn, SciPy
- Deep Learning & NLP : OpenCV, Hugging Face Transformers, Gensim, NLTK, SpaCy

**Data Science & Miscellaneous Technologies**: Data cleaning, wrangling, visualization, modeling, interpretation, ETL, Hypothesis testing, Probability.

**Area of Expertise:** Machine Learning, Deep learning, Natural language Processing, Neural Network, Computer Vision, CNN,Transfer Learning, Large Language Models, Statistics, Generative AI, Retrieval augmented generation (RAG) , LLM Fine-tuning.

**Other Skills:** Git, GitHub, Tableau, Linux, Advanced Excel, Streamlit

## COURSEWORK AND CERTIFICATIONS

- Generative AI with Large Language models
- Python
- Tableau
- Deep learning with Keras and Tensorflow
- SQL
- Pytorch
- Statistics
- Machine learning