SaitejDeepKumar Bodapati

Generative AI Engineer

+1 862-214-0815 | Bodapatisaitej@gmail.com | www.linkedin.com/in/tejchowdary

Accomplished Generative AI Engineer with strong background in Python, TensorFlow, PyTorch, and cloud-based AI platforms**,** seeking to apply advanced knowledge of text, image, and speech generation to develop scalable, ethical, and high-performing AI-driven applications.

PROFILE SUMMARY:

- Over **4.5 years** of experience in developing and deploying **Generative AI models** using Transformer architectures, **Large Language Models (LLMs), and Diffusion Models**, with a focus on creating scalable AI solutions for enterprise.
- Expertise in **Natural Language Processing (NLP)** and **Conversational AI**, implementing **BERT, GPT, and T5-based models** for tasks like text generation, summarization, semantic search, and intelligent virtual assistants.
- Designed and optimized **Generative Adversarial Networks (GANs)** and **Variational Autoencoders (VAEs)** for creating high-quality synthetic data, image generation, and deep fake detection, enhancing data augmentation.
- Skilled in **Deep Learning** frameworks including **TensorFlow, PyTorch, and Kera's**, with hands-on experience in fine-tuning pre-trained models, custom architecture design, and accelerating training through **CUDA** and **GPU/TPU** optimization.
- Strong background in **Machine Learning Operations (MLOps)**, leveraging **Kubeflow, MLflow, and Docker** for automating model training pipelines, experiment tracking, and managing production-scale generative AI deployments.
- Proficient in **Cloud Platforms** with extensive work in **GCP Vertex AI, AWS SageMaker, and Azure Machine Learning**, ensuring secure, scalable, and cost-efficient deployment of generative AI solutions.
- Implemented advanced **Reinforcement Learning with Human Feedback (RLHF)** to enhance **LLM alignment, fine-tuning, and safety**, ensuring ethical and compliant AI models in line with enterprise governance.
- Experienced in **Multimodal AI** development combining **text, image, and audio generation**, utilizing **CLIP, Stable Diffusion, and Whisper** models for cross-domain intelligent systems in healthcare, finance, and insurance.
- Developed and integrated **Vector Databases** such as **Pinecone, Weaviate, and FAISS** with **RAG (Retrieval-Augmented Generation)** workflows to enable scalable knowledge retrieval and context-aware generative applications.
- Strong expertise in **Programming Languages** like **Python, R, and Scala**, along with **API integration, RESTful services, and CI/CD pipelines**, ensuring seamless end-to-end AI application delivery and continuous improvement.
- Implemented **LoRA, PEFT, and Reinforcement Learning with Human Feedback (RLHF)** techniques to optimize both **LLMs and LHMs**, reducing compute cost while improving contextual accuracy.

TECHNICAL SKILLS:

- **Programming:** Python, R, Java, C++, JavaScript, Scala
- **Frameworks:** TensorFlow, PyTorch, JAX, Kera's
- **Generative AI:** Large Language Models (LLMs), Large Hybrid Models (LHMs), Transformers, GANs.
- **Optimization:** LoRA, PEFT, RLHF, Model Pruning, Quantization, Knowledge Distillation
- **NLP & AI Tools:** Hugging Face, Lang Chain, Llama Index, RAG, Prompt Engineering
- **MLOps & Deployment:** MLflow, Kubeflow, Vertex AI, SageMaker, Azure ML, Docker, Kubernetes
- **Data & Storage:** Spark, Hadoop, Databricks, Kafka, Airflow, Snowflake, Big Query
- **Vector Databases:** Pinecone, Weaviate, FAISS, Milvus, Chroma DB
- **Cloud Platforms:** AWS, Azure, GCP
- **Others:** Git, Jenkins, Tableau, Power BI, Gradio, Stream lit, Fast API

WORK EXPERIENCE:

Chevron Corporation | Generative AI Engineer | Houston, Texas | April 2025 – Present

Description: Chevron Corporation is one of the leading integrated energy companies. Design, build, and fine-tune generative models (e.g., LLMs, GANs, transformers) for specific energy applications and develop and implement novel AI

algorithms and techniques for energy-specific challenges and manage and optimize data flow for large-scale model training and inference and develop APIs to integrate AI models into existing energy systems and applications.

Responsibilities

- Designed and deployed advanced **Large Language Models (LLMs),** Transformer Architectures, and Retrieval-Augmented Generation **(RAG)** to automate technical documentation, optimize energy operations.
- Built scalable **Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Stable Diffusion** solutions to generate realistic geological imagery, aiding reservoir analysis and exploration decision-making.
- Designed **Conversational AI, Chatbots,** and Intelligent Assistants powered by **LLMs** to support field engineers with real-time troubleshooting and knowledge-based guidance and implemented **Explainable AI (XAI), SHAP, and LIME** frameworks to ensure transparency and trust in decision-making processes involving generative AI outputs.
- Leveraged **Data Lakes, Apache Spark, and Distributed Data Pipelines** to preprocess, cleanse, and feed large-scale datasets into generative AI systems efficiently and developed **Generative Adversarial Networks (GANs), Diffusion Models, and Variational Autoencoders (VAEs)** to simulate geological formations and forecast drilling scenarios.
- Enhanced enterprise workflows using **MLOps Pipelines, Kubeflow, and MLflow** for seamless generative model training, testing, and deployment at scale and collaborated on **Cloud-Native AI Infrastructure, AWS SageMaker, GCP Vertex AI)** to ensure scalable and cost-effective generative AI solutions across Chevron's global operations.
- Streamlined AI deployment pipelines using **MLOps, GitOps, and Continuous Integration/Continuous Deployment (CI/CD)** practices to ensure scalable, reliable, and automated management of generative AI model.
- Developed and deployed generative AI solutions using **Python** within **Microservices architecture**, orchestrated with **Kubernetes**, while implementing **CI/CD pipelines** to enable scalable, automated, and reliable delivery of AI-driven applications across Chevron's enterprise systems.
- Designed and optimized **Machine Learning pipelines** using **TensorFlow** and **PyTorch**, applying feature engineering and hyperparameter tuning to improve Chevron's predictive analytics for energy forecasting and operational efficiency.

Environment: LLMs, RAG, GANs, SHAP, LIME, Chatbots, Data Lakes, Apache Spark, **GANs, VAEs,** MLOps Pipelines, Kubeflow, MLflow, AWS SageMaker, GCP Vertex, GitOps, CI/CD, Python, Microservices, Kubernetes, TensorFlow**,** PyTorch.

Comerica Incorporated | Machine Learning Engineer| Dallas, Texas, USA | April 2024 – March 2025

Description: Comerica Incorporated is a leading financial services company. Design, build, and train machine learning models using various algorithms and frameworks like TensorFlow or PyTorch and Create and optimize features from raw data to improve model accuracy and predictive power and deploy trained models into production environments.

Responsibilities:

- Designed and implemented advanced **Machine Learning Models** using **Python** and **Scikit-learn** to predict customer credit risk, enhance fraud detection, and improve financial decision-making processes.
- Built scalable **Data Pipelines** with **Apache Spark**, **SQL**, and **ETL frameworks** to process high-volume financial transactions and deliver real-time insights across core banking operations and applied **Natural Language Processing (NLP)**, Transformers, and **Generative AI** for intelligent document processing, customer sentiment analysis.
- Developed and deployed Deep Learning Architectures using **TensorFlow** and **PyTorch** to optimize loan approval workflows, automate risk modelling, and support predictive analytics initiatives.
- Leveraged **Cloud Platforms** along with **MLOps** frameworks to streamline model deployment, monitor model drift, and ensure reproducibility of financial AI solutions and integrated **Kubernetes**, **Docker**, and **Microservices** with robust **CI/CD pipelines** for scalable model deployment in production environments with minimal downtime.
- Conducted advanced **Statistical Modelling**, **Feature Engineering**, and **Big Data Analytics (Hadoop, Hive, Spark SQL)** to extract actionable insights from structured and unstructured banking data.
- Implemented and optimized **Azure Machine Learning**, **Azure Data Factory**, and **Azure Synapse Analytics** to orchestrate end-to-end ML workflows, enabling secure data ingestion, automated model training.

Environment: Python**,** Scikit-learn, Apache Spark**,** SQL**,** ETL frameworks, NLP, TensorFlow**,** PyTorch, MLOps, Kubernetes**,** Docker**,** Microservices, CI/CD, Hadoop, Hive, Azure Machine Learning, Azure Data Factory, Azure Synapse Analytics.

Citibank India (Wipro) | Data Scientist | Bangalore, India | April 2022 – July 2023

Description: Citigroup Inc. is a leading global financial services corporation. Design, build, and optimize machine learning algorithms and statistical models to forecast trends, predict market movements, and assess risk and Implement anomaly detection and behavioural analytics to identify and prevent fraudulent activities in real-time.

Responsibilities:

- Designed and implemented advanced **Machine Learning, Deep Learning, and Natural Language Processing (NLP)** models to analyse large-scale financial datasets, enabling fraud detection, risk scoring with high Accuracy.
- Utilized **Python, R, and SQL** for data preprocessing, statistical modelling, and algorithm development, ensuring efficient feature engineering, model interpretability, and deployment of predictive analytics solutions in critical banking operations.
- Applied **Big Data technologies such as Hadoop, Spark, and Hive** to process and analyse structured and unstructured financial data, improving decision-making processes for investment strategies and credit risk management.
- Developed interactive **Power BI, Tableau, and Data Visualization** dashboards to communicate insights effectively to business leaders, helping Citi improve portfolio management, client segmentation, and revenue forecasting.
- Implemented **Cloud-based AI solutions on Azure, GCP** to build scalable machine learning pipelines, enabling real-time data processing, model deployment, and integration with Citi's digital banking platforms.
- Leveraged **MLOps, CI/CD, and Kubernetes** for automating model training, versioning, and deployment in production, ensuring continuous monitoring, governance, and compliance with financial regulations.
- Built **Generative AI and Large Language Models (LLMs) such as GPT and BERT** to automate compliance document analysis, regulatory reporting, and enhance customer support with AI-driven chatbots.

Environment: Machine Learning, NLP, Python, R, SQL, Hadoop, Spark, Hive, Power BI, Tableau, Data Visualization, Azure, GCP, MLOps, CI/CD, Kubernetes, LLMs, GPT, BERT.

Aurobindo Pharma Limited | Data Scientist | Bangalore, India | June 2020 – March 2022

Description: Aurobindo Pharma Limited is a leading Indian multinational pharmaceutical company. Designing, building, and optimizing machine learning and statistical models to predict trends, identify potential drug candidates, and forecast trial outcomes and creating clear, insightful visualizations, dashboards, and reports.

Responsibilities:

- Implemented **Time Series Forecasting, ARIMA, and LSTM models** to predict drug demand, optimize supply chain management, and reduce inventory costs and Applied Image Processing, Computer Vision, and **CNNs** on medical imaging datasets to assist in disease diagnosis and accelerate drug research.
- Integrated **Big Data frameworks (Hive, Pig, and Spark SQL)** for querying and processing large volumes of patient and drug trial data and collaborated with cross-functional teams to deploy **MLOps pipelines, CI/CD for ML models, and Docker/Kubernetes** for scalable pharma analytics solutions.
- Designed and deployed **Natural Language Processing (NLP, BERT, and Transformers)** models to analyse clinical trial reports, research papers, and patient feedback for drug safety insights.
- Built **Knowledge Graphs, Graph Neural Networks (GNNs), and Neo4j** frameworks to map drug–gene interactions and identify new therapeutic targets and automated **ETL Pipelines with Airflow, Pyspark, and SQL** for seamless integration of pharmaceutical data from research labs, clinical trials, and manufacturing units.
- Enhanced predictive accuracy using **Ensemble Methods (XGBoost, LightGBM, Cat Boost)** to model adverse drug reactions and patient outcomes and collaborated with R&D teams to implement **Generative AI models (GANs, Variational Autoencoders)** for molecular structure simulation and accelerated drug discovery.

Environment: Time Series Forecasting, ARIMA, LSTM models, CNNs, Hive, Pig, Spark SQL, MLOps pipelines, CI/CD, NLP, BERT, GNNs, Neo4j, Airflow, Pyspark, SQL, XGBoost, LightGBM, Cat Boost, GANs.

EDUCATION: **New Jersey Institute of Technology,** Masters in Data Science, Newark – NJ, USA from 2023 – 2024