**[Sai Chaitanya Pachipulusu](https://www.datascienceportfol.io/saichaitanya)**

[linkedin.com/in/psaichaitanya](https://www.linkedin.com/in/psaichaitanya/) | [github.com/chaitanyasaip](https://github.com/chaitanyasaip) | [pachipulusu.vercel.app](https://pachipulusu.vercel.app/)

Email : [siai.chaitanyap@gmail.com](mailto:siai.chaitanyap@gmail.com)

Mobile : [+1 (551) 344-5967](tel:+15513445967)

Machine Learning Engineer with 4+ years of experience delivering scalable AI solutions, including LLMs, RAG systems, and data pipelines, across diverse industries. Proven track record of reducing costs, improving efficiency, and driving innovation through cutting-edge ML and data engineering techniques

**SKILLS & ACCOMPLISHMENTS**

* **Programming Languages:** Python, Java, C++, R, SQL
* **Technical**: Statistical Analysis, Probability and Inference, Analytical Methods, Machine Learning Algorithms, MLOps, Neural Networks (CNN, RNN, LSTM), GANs, VAEs, Autoencoders, Transformers, LLM, NER, knowledge graphs, Hive, Scala, GCP, Clustering, Regression, Statistics, Mathematics, Selenium, BeautifulSoup, Linux/UNIX, Advanced Analytics, Predictive Models, Data Science, Natural Language Processing, Data Manipulation, Data Visualization, Critical thinking, GenAI solutions, Prompt engineering, Optimization, Agile, Data analysis, Data pipelines, Data integration, data classification, data governance, Data lakes
* **Frameworks/Libraries**: PyTorch, TensorFlow, Snowflake, Streamlit, Keras, Scikit-learn, XGBoost, PySpark, NumPy, Pandas, spaCy, Natural Language Tool Kit, Apache Spark, Matplotlib, Hadoop, Seaborn, Apache Kafka, LightGBM, Plotly, JAX
* **Generative AI**: LLMs, RAG Pipelines, LangChain, Llamaindex, Weaviate, Pinecone, ChromaDB, Fine-tuning (LoRA/PEFT), FAISS
* **Database Management Systems**: Relational-Databases (MySQL, PostgreSQL), MariaDB, NoSQL, SparkSQL, SQL
* **Tools/Platforms**: AWS (S3, EC2, Lambda, SageMaker, Bedrock), Databricks, Tableau, PowerBI, Visual Studio, WEKA, phpMyAdmin, Red Hat Network Satellite, Excel, Version Control (Git), CI/CD (Docker, Kubernetes), A/B testing, Terraform, BigQuery
* **Certifications**: Getting Started with PowerBI (Coursera), Machine Learning Specialization (Coursera), Python for Everybody (Coursera), SPSS certified professional in Data mining and Warehousing

**EXPERIENCE**

• **Community Dreams Foundation** Remote, USA Machine Learning Engineer Jul 2024 - Present

* Architected an HR tool using BERT, RoBERTa, and Sentence-BERT embeddings to match resumes with job descriptions, cutting manual screening by 90% from 10 hours to 1 hour per job and speeding up the hiring by 40% through context-aware ranking
* Built an end-to-end cloud-native pipeline with Python, FastAPI, and Kubeflow on Kubernetes for automated interview scheduling, achieving 92% candidate selection precision measured in pilot with 50 companies
* Automated rejection emails with sentiment-aware templates with VADER score above 0.6, handling 200+ weekly communications and reducing admin work by 90% from 15 hours/week to 1.5 hours/week while ensuring empathetic, bias-free communication
* Developed a fine-tuned BERT model and Flask microservice to automate personalized rejection communications, processing 200+ weekly applicants with 94% positive feedback rate while reducing HR workload by 90% through contextual response generation
* Designed a RAG chatbot with Mistral, PyTorch on AWS SageMaker, achieving 90% response relevance and reducing ticket escalations by 25%, saving an estimated 200+ hours annually for the support team

• **CGI Information Systems and Management Consultants Pvt. Ltd** Bengaluru, India Data Engineer/ Associate Software Engineer (Client: Shell Corporation) Sep 2020 - Jun 2022

* Developed a real-time ingest layer using Kafka Connect to capture 8 sensor data streams (400 events/second) from factory equipment, reducing data availability lag from overnight batch to less than 5 minutes for maintenance teams
* Wrote and maintained 15+ Python ETL scripts to process daily Shell refinery data files (CSV, JSON) into a centralized SQL data warehouse, enabling dashboard KPIs previously unavailable to operating teams
* Led migration of 52 legacy servers to AWS EC2 (t3.xlarge instances) using Terraform, reducing monthly costs by $18k and ensuring 99.9% uptime over 6-month period
* Optimized Databricks medallion architecture with schema validation and incremental processing, reducing data pipeline failures from 12 to 3 (75% decrease) per week and cutting recovery time from 4 hours to 45 minutes for Shell’s refinery operations
* Awarded Best Employee for Q4 2021 for exceptional contributions to project efficiency and collaboration with interdisciplinary teams, leading 30% of the overall data migration effort

• **Imbuedesk Pvt. Ltd** Hyderabad, India Machine Learning Engineer May 2018 - Aug 2020

* Achieved 97% accuracy on FER2013 dataset by developing facial expression recognition system using OpenCV and TensorFlow
* Processed 50K vehicle plates daily by designing an image processing pipeline with Tesseract OCR orchestrated with Kubernetes
* Reduced equipment downtime by 28% through implementation of predictive maintenance dashboards with AWS Beanstalk
* Decreased image processing latency by 66% by building a Kafka-based pipeline handling 35MB/hour with a 3-node consumer topology
* Implemented back-pressure handling for peak traffic periods when processing volume increased by 300%

**TECHNICAL PROJECTS**

* **Career Roadmap Generator** | Generative AI, RAG System, Streamlit, ChromaDB | [Huggingface](https://huggingface.co/SaiChaitanya/Career_Roadmap)

Integrated GPT 3.5, GPT-4o models with ChromaDB for context-aware recommendations to the job description achieving 0.87 hit@5 score on an evaluation set of 200 anonymized job transitions from tech professionals. Async implementation reduced API latency from 12 to 4 seconds for 10+ concurrent users

* **SaaS Chatbot** | Large Language Model, Llama-2-7B, Langchain, Streamlit, Python

Developed support chatbot by fine-tuning Llama-2-7B using parameter-efficient PEFT/LoRA techniques on 5k support conversations. Deployed on a single A10 GPU achieved 84% accuracy while reducing inference costs by 65% compared to hosted API solutions

**EDUCATION**

|  |  |  |
| --- | --- | --- |
| • | **Stevens Institute of Technology** | Hoboken, NJ |
|  | Master of Science in Machine Learning; CGPA: 3.9 | Sep 2022 - May 2024 |
| • | **Sreenidhi Institute of Science and Technology** | Hyderabad, India |
|  | Bachelor of Technology in Information Technology | Aug 2016 - May 2020 |