Desai Brave

+1 347 614 7903 | brave.d302001@gmail.com | linkedin.com/in/bd4

Education

Stony Brook University 08/2025 – 05/2027
Master of Science in Data Science *Stony Brook, NY*
Data Structures and Algorithms, Data Analysis, Introduction to Computer Vision

Pandit Deendayal Energy University – *GPA: 3.8* 08/2019 – 05/2023
Bachelor of Technology in Information and Communication Technology *Gujarat, India*
OS, DBMS, Cloud Computing, Machine Learning, Probability and Statistics, Big Data Analytics, AI Systems

Skills

Programming: Python (*Pandas, NumPy, SciPy, Scikit-Learn, PyTorch, TensorFlow, OpenCV*), R, SQL, Java
Deep Learning & Language Systems: NLP, Generative AI, Agentic Systems & User Interaction Modeling, LLMs, Retrieval-Augmented Generation, Fine-tuning (LoRA, QLoRA), Prompt Engineering, LangGraph, Pydantic AI, DeepEval, Comet Opik, LlamaIndex, LlamaParse
ML & Modeling Frameworks: PyTorch, PyTorch Lightning, TensorFlow, Scikit-Learn, PySpark, Model Deployment & Monitoring, Regression, Classification
Deployment, Tools & Visualization: Docker, Git, Flask, FastAPI, Talend, Hugging Face, Power BI, Tableau
Databases & Cloud Platforms: PostgreSQL, MongoDB, Snowflake, Vector DBs, Azure (Data Factory, SQL, Synapse), AWS (S3, EC2, SageMaker, Bedrock, Lambda)

Professional Experience

Growexx 08/2023 – 05/2025
Data Scientist | Python, PyTorch, SQL, RAG, Pinecone DB, AWS Bedrock, AWS S3 *India*

- Built AI platforms (Ojavix & Hirin.ai) using **advanced NLP & agentic reasoning systems** (Claude 3.7, Llama 3.3); deployed **production-ready ML systems** on AWS Bedrock, achieving **98%+ retrieval accuracy.**
- Led multi-layered RAG pipeline initiatives with **ReAct-style agent planning, user-interaction loops, and agent performance evaluation**; achieved **¡3s latency** in production.
- Improved reliability by integrating **continuous evaluation (Comet Opik, DeepEval)**; **documented model design and monitoring practices**, achieving **96% response accuracy** in HIPAA-secure deployments.
- Developed a **Text-to-SQL system** enabling non-technical users to query large proprietary datasets; achieved **90%+ accuracy** through schema-aware retrieval and iterative feedback.

Growexx 01/2023 – 08/2023
Data Engineer Intern | Python, Selenium, LLM, SQL, Talend, Snowflake, Azure, Power BI *India*

- Maintained and executed **Python & SQL scripts** for large-scale data ingestion pipelines, improving reliability and reducing failures.
- Designed and deployed **scalable NLP/ML applications** with GPT and Azure Cognitive Search; automated evaluation workflows & improved user engagement by **50%**; **documented data and model pipelines** for cross-functional review.
- **Documented and improved operational workflows** for scalable ML/LLM pipelines; collaborated cross-functionally to test software releases and resolve data pipeline issues.

Projects and Publications

HubermanGPT | *Python, LLM-Fine tuning, QLoRA, Microsoft Phi-2* 07/2025 – 08/2025

- Generated domain-specific Q&A datasets by chunking Andrew Huberman's podcasts and fine-tuned the **Microsoft Phi-2** model using **QLoRA** for efficient domain adaptation.
- Constructed an interactive chatbot pipeline with **agentic reasoning and user feedback loops** to improve answer quality and engagement, delivering context-aware responses that replicate Huberman's explanatory style while ensuring scientific accuracy.

Transformer Architecture Implementation | *Python, PyTorch* 06/2024

- Implemented the **Transformer architecture from scratch** in **PyTorch**, including **multi-head self-attention**, positional encoding, and encoder–decoder blocks, based on Vaswani et al. (2017).

Facial Recognition Using Siamese Neural Network and Data Augmentation Techniques 07/2024

- Built a **Siamese** CNN–based facial recognition model with **data augmentation**, presented at **IEEE WCONF 2024**, achieving accuracy, robustness, and generalisation in computer vision and identity verification tasks.

Flight Price Predictor | *Python, Machine Learning* 04/2024

- Processed and standardized **heterogeneous datasets (structured + unstructured)** to build predictive models, reinforcing ability to work with diverse data sources.