



# Bharatbhushan Dwarkeyasi

DATA SCIENTIST/ GENERATIVEAI ENGINEER



[bharat.genai@gmail.com](mailto:bharat.genai@gmail.com)



7058628834



[LinkedIn](#)



[Github](#)



[Certifications](#)

## PROFESSIONAL SUMMARY

- With **3.6 years** of experience as a Data Scientist, I possess expertise in data modeling, business intelligence, data pipeline and transfer learning techniques.
- Proficient in Python and data processing, I have a strong foundation in NLP (BERT, GPT), Generative AI, Large Language Models, LangChain, and Machine Learning techniques, including AI Chatbots, Conversational AI, and Data Retrieval.
- Created and deployed a chatbot product for a company using the DistilBERT Transformer-based model.
- This chatbot improved user experience with fast and accurate responses, handling diverse input variations due to its training on extensive data.
- To enhance operations, automated customer interactions, minimizing human intervention.
- Developed a standardized machine learning pipeline incorporating five essential components: data ingestion, transformation, validation, evaluation, and model training and deployment.
- Ensured rapid and efficient data access and boosting retrieval performance by leveraging MongoDB for data storage.
- To safeguard pipeline artifacts, implemented robust security measures by securely storing them in an S3 bucket, preventing data loss and unauthorized access.
- Area of Expertise: Data Science, Generative AI, and Data Analytics. I possess proficiency in Python Scripting and Data Science Analytics.
- Experienced in regular interaction with clients/stakeholders for analyzing business requirements and translating such understanding into technical solutions.
- I have a proven track record of achieving results and possess exceptional analytical research and communication skills with a talent for translating complex technical ideas into actionable business strategies.

## Skills

**Programming Languages** (Python,SQL) | **Feature Engineering** (Pandas | Numpy) | **Frameworks** (Tensorflow | PyTorch) | **Databases** (MySQL | MongoDB | Pinecone | ChromaDB | Faiss | Qdrant | Weaviate) | **MLOps** (Git | DVC | MLFLOW | Docker | CircleCI | Github Action | Langsmith | Langserve | VEXT | Amazon BedRock | Amazon Q) | **Rest API** (Flask | FastAPI) | **Frontend** (Streamlit | Chainlit | Gradio) | **Cloud** (AWS | AZURE | GCP) | **LLM Platforms** (Langchain | LlamaIndex | Google AI Studio | Vertex AI | Groq | Hugging Face | OpenAI) | **Generative AI** (Prompt Engineering | RAG | Quantization | Agents | Chains | Vector DB | LLM | Haystack) | **Natural Language Processing** (BERT | GPT | ROBERTA | ALBERT | T5 | Encoder-Decoder | NER | Spacy | NLTK) | **Deep Learning** (ANN | RNN | LSTM | GRU | Transformers) | **Machine Learning Algorithms** (Linear Regression | Logistic Regression | Decision Tree | Random Forest | Xg Boost)(supervised and unsupervised learning) | **Statistics** (Descriptive | Inferential) | **Soft Skills** (Leadership | Event Management | Writing | Public Speaking | Time Management)

## PROJECTS

### Automated Document Summarization using NLP

*BrainVision Technology*

- Developed a document summarization parser using **TensorFlow** and **T5-small HuggingFace** pre-trained model.
- Text data from PDF is extracted using the PyPdf library.
- Evaluated the performance of the summarization algorithm using metrics such as **ROUGE** and **BLEU** scores.
- Conducted **A/B testing** to evaluate the performance of the algorithm against other state-of-the-art summarization techniques.
- Utilized **AWS** cloud for project implementation and **EC2** instance for model deployment.
- **CI/CD** was implemented using **GitHub Actions**.

Video-Document based Conversational AI system using LLM and LangChain

BrainVision Technology

- Built a conversational AI system utilizing **LangChain** and **LLM** to extract key information from video transcripts or documents.
- **Whisper Base 7b** is used for audio transcription.
- Used **OpenAI embeddings** class for generating the embeddings.
- Used **Chroma DB** to save the embeddings and **cosine similarity** is used to find out the similar chunks of documents.
- A group of tasks including **video summarisation**, and **chat-with-video/documents** are obtained using **text-davinci-003 LLM model**.

ML Pipeline Standardization

- Introduced a consistent machine learning pipeline comprising five key elements to enhance the efficiency of upcoming projects at Brain Vision.
- The pipeline encompasses phases such as Data Ingestion, Data Transformation, Data Validation, Model Training, Model Evaluation and Deployment.
- Utilized MLFlow Experiment Tracker to facilitate better comparison of experiments and enhance the tracking of parameters efficiently.
- Implemented the storage of processed data in MongoDB, optimizing access speed and overall data retrieval efficiency.
- Ensured secure storage of pipeline artifacts in an S3 bucket, providing dependable retrieval and safeguarding against potential data loss.

PROFESSIONAL EXPERIENCE

Data Scientist, BrainVision Technology

01/2022 – Present  
Pune, India

- Collaborated with the backend AI team to develop doubt clustering, question answering and generative AI modules.
- Utilized Hugging Face pipelines for rapid prototyping and experimentation: Leveraged pre-trained models and tools to quickly test and evaluate new ideas.
- Implemented NLP techniques, leveraging transformers and LLMs for enhanced performance.
- Implemented advanced NLP methodologies for efficient data processing (**30% improvement** in categorization).
- Pioneered NLP-driven chatbot development, leveraging Generative AI and transformer architectures to achieve a **40% increase** in user engagement and **25% faster response times**. This innovative
- Approach to dialogue interaction empowered users with more natural and efficient communication.
- Prepared and analyzed data for model training Ensured high-quality data for training and optimized model performance. Monitored and evaluated model performance, driving continuous improvement.

GenerativeAI Freelancer, Remotask(scale.ai)

09/2023 – Present  
Remote, US

- Leveraged data science expertise to train and improve large language models (LLMs) applied domain knowledge to provide high-quality training data and feedback for LLM development.
- Achieved **Diamond Tasker** status with in Bulba Experts Chat Rewrite Computer Science demonstrated exceptional performance and expertise in this project.
- Earned **Gold Tasker** status with **80% accuracy** in Amendment Eight consistently delivered high-quality work met project requirements.

EDUCATION

G.H.Raisoni College of Engineering,Nagpur

06/2016 - 05/2019

Bachelor of Engineering : CGPA - 7.86

Hackthons & Honored

- Google Generative AI Hackathon Idea Submitted (AI-powered Debt Repayment Strategies) March - 2024
- Honored to be Recognized as a Achiever of the year at GHRCE May - 2019