BROTOTYPE TAMIL

Machine Learning

📞 +91 9995270355 — ✉ admissions@brototype.com — 🔗 Linkedin — ⚫ GitHub link here

Profile Summary

Machine Learning Engineer focusing on Computer Vision and Natural Language Processing, with hands-on experience developing deep learning models using PyTorch and TensorFlow. Skilled in implementing state-of-the-art architectures including Transformers, CNNs, and LSTMs for real-world applications. Passionate about combining theoretical ML knowledge with practical engineering solutions to build models that scales.

Technical Skills

Programming Languages: Python, Java, C, JavaScript, SQL
Machine Learning: TensorFlow, PyTorch, Keras, Scikit-learn, Transformers, CNNs, LSTMs, GANs, AutoEncoders, Reinforcement Learning
NLP & Computer Vision: HuggingFace Transformers, BERT, GPT, NLTK, SpaCy, OpenCV, YOLO, ResNet, MobileNet, VGG
Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Plotly
Cloud & Development Tools: AWS, Google Cloud, Docker, Git, CI/CD, PostgreSQL, MongoDB
Web Development: Django, FastAPI, REST APIs
Other: Data Structures, Algorithms, Jupyter Notebooks, Streamlit

Projects

AIDA | *'"AI-Driven Assessment" - A Complete Technical Interview Simulation Platform.*   Add Live Link here

An AI-powered interview simulation platform leveraging speech-to-speech technology and multiple AI personas to conduct, evaluate, and provide real-time feedback for those preparing for technical interviews.

- Engineered an ultra-low latency (250-500ms) speech-to-speech technical interview simulation platform for real-time interaction and evaluation helping candidates to practice for technical interviews.
- Implemented LLaMA-70b for question generation, evaluation, and summarization, achieving 90% relevance compared to original interview questions.
- Developed Advanced Chain of Thought (CoT) prompts to maintain interviewer persona with 98% accuracy.
- Trained a custom ResNet-based CNN for emotion tracking in interviews, achieving 95 percent accuracy.
- Developed a custom evaluation system combining syntactic (tree-based + Edit Distance) and semantic (text embeddings) similarity metrics, reducing score inflation in low-quality responses by 60% and enhancing overall evaluation accuracy.
- Scaled platform to handle 99% of concurrent users without performance degradation.
- Orchestrated containerized deployment, applied CI/CD pipelines, and utilized AWS Lambda & EC2, ensuring 100%uptime.
- **Technologies Used:** Python, LLaMA-70b, TensorFlow, PyTorch, AWS Lambda, Docker, Nginx, FastAPI, React.

JobSum | *AI platform for analyzing thousands of job descriptions seamlessly.*   Add Live Link here

A comprehensive job market analysis tool that leverages AI to extract and analyze job posting data, providing actionable insights for job seekers.

- Built an end-to-end job market analysis platform using Python, BeautifulSoup, Plotly, and Gemini API for automated data extraction and insights delivering relevant insights to candidates.
- Constructed a web scraper to extract 1000+ job descriptions from major platforms.
- Designed an interactive data analytics dashboard with Plotly, improving data interpretation efficiency by 40%.
- Integrated conversational AI using Gemini API, reducing manual analysis time by 99.93% and making the platform 1500x more efficient.
- **Technologies Used:** Python, Gemini API, Web Scraping, Data Visualization, NLP, Streamlit Cloud.

Mini Projects

Scriptr | *A Fined-tuned model for YouTube Script Generation.*

Fine-tuned LLM designed to generate coherent and contextually relevant YouTube scripts from comma-separated keywords, seamlessly integrating even unrelated keywords into compelling narratives.

- Fine-tuned Gemma-2 model on 924K tokens using QLoRA and 4-bit quantization, reducing model size by 73%.
- Optimized for single T4 GPU deployment while maintaining 96% coherence score in script generation.
- Implemented efficient prompt engineering techniques for coherent script generation from unrelated keywords.
- **Technologies Used:** Huggingface Transformers, PyTorch, PEFT, QLoRA, Weights and Biases, Gemma2 Model, SFTTrainer, YouTube API.

Face Recognition Model | *A model for face identification and detection over multiple classes.*

A deep learning-based face recognition system using Siamese Neural Networks for multi-class identification and annotation.

- Achieved 90% classification accuracy across multiple identity classes using FaceNet architecture.
- Implemented transfer learning and model optimization techniques to reduce training computational requirements.
- **Technologies Used:** TensorFlow, Siamese Neural Network Architecture, FaceNet, Scikit-learn, OpenCV, MTCNN, Keras, FastAPI.

LSTM NMT Model | *A Machine Translation model implemented from scratch.*

An Neural Machine Translation system for English-to-Hindi language translation.

- Developed LSTM-Attention architecture achieving 80% translation accuracy on T4 GPU.
- Implemented an LSTM encoder-decoder with an attention mechanism for improved context handling in translations.
- **Technologies Used:** TensorFlow, Pandas, Keras.

Publications and Articles

- "SVM Kernels: Polynomial Kernel Implementation in Python" - Referenced in California State University, Northridge thesis (2023).
- "Multi-Layer Perceptron Explained: A Beginner's Guide" - Cited in ResearchGate publication (2024).
- "Top-5 Most Advanced AI Systems" - Cited by National Institute of Health (NIH) publication.

Education

Machine Learning Student Intern   **2025 – Present**
Brototype, Chennai

Bachelor of Computer Applications   **2020 – 2023**
University of Chennai

CGPA: 8.4/10.0