Rishi Shreshtham

Machine Learning / Generative AI Engineer

📞 *9381876084*   ✉ *rishi.sh.dev@gmail.com*   in *Rishi*   ⚓ *github.com/fluire*

EXPERIENCE

Tredence   **September 2024 – Present**
Machine Learning Engineer   *Hyderabad*

LangGraph Agents for Inventory Analysis

* Worked on implementation of inventory root cause analysis using LangGraph agents and integrated with **Neo4j Graph DB**, enabling stakeholders to visually explore causal relationships.
* Reduced manual RCA effort by **40%** by automating agent workflows for dynamic inventory issue tracing.
* Optimized graph queries and schema design, improving traversal performance by **60%** and enhancing interactive dashboard responsiveness.

RAG System Development

* Designed and developed a Retrieval-Augmented Generation (RAG) system from scratch using Python, LangChain, and vector databases, supporting real-time question answering over unstructured product documentation.
* Reduced retrieval latency by **35%** through efficient document chunking and metadata filtering.
* Delivered a robust and scalable API architecture enabling integration with internal analytics platforms, accelerating product insight delivery across teams.

Brane Enterprises   **May 2022 – September 2024**
Data Scientist   *Hyderabad*

Continuous Document Extraction

* Delivered project achieving **90% on document extraction** accuracy, enabling clients to process thousands of documents in minutes.
* **Fine tuning LLM** (llama 2 13b) using **LORA** for Natural Language Understanding and **summary generation**.
* Fine-tuned **image recognition** model (YOLO) for image extraction from documents.
* Developed an advanced chatbot for document question answering utilizing Retrieval-Augmented Generation (**RAG**) and a vector database such as Quadrant, enabling efficient management and retrieval of information from over 1,000 pages of PDF content..

Redaction

* **Automated Document Redaction with 95%** accuracy, reducing manual effort and processing by 80%.
* Fine-tuned BERT model for token classification NLP task, monitoring and maintaining the model registry.
* Cross team integration for end-to-end delivery of feature currently serving Hospitals like Medaxioms and Call Health.

Automatic speech recognition (ASR)

* Deployed **audio to text model** (whisper) for generating transcriptions with being able to process more than 1000 hours of videos in minutes.
* Scripted a feature to overlay subtitles on videos, significantly enhancing viewer comprehension.

Assisted Rehab

* Developed scalable **API** for image recognition service specializing in pose estimation, utilizing Python, **Fastapi** improved response time by 30% and facilitated integration with multiple cross projects applications.
* Authored highly efficient Python scripts for intermediate data processing.

Wind Forecast and Power Prediction

* Developed and evaluated models for predicting power generation at wind farms.
* Achieved **92%** accuracy, securing 2nd position among competing vendors.

Brane Enterprises   **November 2021 – May 2022**
Data Scientist Intern   *Hyderabad*

* Database Management, Migration, and Integration
* Developed a resume parser to extract and import key information.

SKILLS

Development

* **Language:** Python, C.
* **Tools and framework:** Fastapi, Flask, Torch, Docker, Linux, MongoDB, SQL.
* ML Algorithms, Neural Networks, NLP, Generative AI, system design and Maintenance.

Soft Skills

* Fluent in teamwork, professional overthinker (critical thinker), part-time peace negotiator (conflict resolver), full-time decision-maker (usually under caffeine influence), champion of knowledge sharing (even the memes), and skilled in organizing chaos like it's an Olympic sport.

EDUCATION

CVR College Of Engineering   **Aug-2018 – May-2022**
B.Tech   *GPA – 8.6*
* Participated in 8 Kaggle image recognition challenges and being at top 10% in 3 and top 4% in one.

CERTIFICATIONS

* Natural Language Processing using Tensorflow ⌟   * Sequences, Time Series and Prediction ⌟

* Machine Learning in Production ⌟   * Deep Learning ⌟