Shreyans Dhankhar

📞 +91-9929943636   @ shreyansdhankhar@gmail.com

Education

INDIAN INSTITUTE OF SCIENCE

M.Tech, Computer Science

📅 June 2017   📍 Bangalore, India

SKIT, JAIPUR

B.Tech in Computer Science

📅 Jun 2014   📍 Jaipur, Rajasthan

Links

GitHub *Shreyans92*

Linkedin *shreyans-dhankhar*

Patent

US PCT/US2020/017244: *Codifying Domain Knowledge of Expert Call Support Agents for Efficient Case Resolution.*

Skills

Python • PyTorch • HuggingFace • Transformers • Reinforcement Learning • Generative AI • Prompt Tuning & Engineering • Langchain • Machine Learning • Deep Learning • Cllasification • NER • Retrieve and Ranking • MLOPs • AWS • Sagemaker • FastApi • Parameter Efficient Fine Tuning(PEFT) • Gradio • Git • C/C++ • LaTeX

Co-Curricular Activities and Honors

GATE SCHOLARSHIP

Indian Institute of Science (IISc)

Awarded the Graduate Student Scholarship for M.Tech

CODE CONTRIBUTER

Practical Natural Language Processing

Helped to develop the code base for book which features a foreword from Professor Julian MCAuley and praise blurbs from Professor Zackary.

Experience

APPLIED RESEARCH SCIENTIST

Thomson Reuters Research Labs

📅 Aug 2021 – Present   📍 Bangalore, India

- **Contract Drafting Assistant:** Finetuned open-source LLM(Flan-T5-XXL) to build drafting assistant for contracts using 8 bit training with LORA, on the instructions data generated by ChatGPT. Improved the model performance using Reinforcement Learning. Implemented the *Retrieval Augmented Generation(RAG)* pipeline to make the assistant adaptable to a variety of contracts.
- **Prompt Tuning:** Implemented prompt tuning strategy on *multi-label classification* task using parameter efficient finetuning technique for Legal-Bert model.
- **Document Ranker:** Designed a retrieval and re-rank strategy to rank documents based on an incoming document. The retrieval algorithm utilized the *BM25/Biencoder*, and the re-ranker employed *Cross Encoders*. I fine-tuned the Cross Encoder to adapt to domain-specific data. Furthermore, I devised strategies to evaluate the model performance via SMEs. As a result, the top-5 metric was 82.3%, and three SMEs agreed to 63% in Cohen Kappa score.
- **Named Entity Recognition(NER):** Fine-tuned a Legal Bert model to identify important tax-related entities in lengthy documents. To overcome the restriction of 512 tokens in Bert based model, I employed a sliding window approach to cover the entire document.

DATA SCIENTIST III

Hewlett Packard Inc.

📅 Aug 2018 – Aug 2021   📍 Bangalore, India

- **Battery Swell Prediction:** Developed a machine learning model to predict battery swelling in advance within a specific Time to Fail Window using non-uniform telemetry data that had no proper ground truth labels.Increased the coverage of swelling batteries by 150% compared to a rule-based system.
- **NLP-based Assistance for Customer Support Agents:** Created an LSTM-based next troubleshooting step prediction that takes the response of the current step into account. This approach reduced the Average Call Handling Time by 17% and increased customer satisfaction by 20%.
- **Steps Extraction from Unstructured Data:** Developed a probabilistic method with minimal supervision to extract the troubleshooting steps and customer response to each step from the unstructured case notes written by the call agents. Utilized the Byte Pair Encoding technique to handle the out-of-vocabulary problem. This approach reduced manual labeling effort by 80%.

SOFTWARE ENGINEER II

Dell EMC

📅 Jun 2017 – Aug 2018   📍 Pune, Maharashtra