PRATIK MEHTA

PROFILE

A combination of NLP research and Generative AI backend engineering

EXPERIENCE

SCALE AI

Freelance Sr. Software Engineer   JUNE 2024 - OCT 2024

- Underwent training to provide qualitative feedback on code-generation and general-purpose LLM abilities as a domain expert

DELOITTE INDIA (OFFICES OF THE US)

Technology Consultant (ML/AI)   NOV 2020 - JUNE 2024

- Served as a back-end engineer, successfully deploying APIs to production.
- Designed and implemented a micro-service from scratch, enabling structured data retrieval over enterprise-scale data.
- Developed a taxonomy and evaluation system for RAG systems, utilizing statistical testing on a synthetically generated test-set of question-answer pairs.
- Helped develop a document classification and information extraction engine, served and hosted on AWS.

CENTER FOR INDIAN LANGUAGE TECHNOLOGY - IIT BOMBAY

Research Engineer   AUG 2015 - JULY 2016

- Trained and deployed machine translation systems for English-Marathi and English-Indonesian medical domain text translation, enhancing language accessibility in healthcare.
- Designed an error analysis engine to identify and classify syntactic errors in translations, and developed a greedy algorithm for automatic post-editing, improving translation quality and efficiency.
- Co-compiled the IIT Bombay English-Hindi parallel corpus, a significant resource in the field of natural language processing.

EDUCATION

UNIVERSITY OF MASSACHUSETTS AMHERST – MS COMPUTER SCIENCE
2016-2019

UNIVERSITY OF MUMBAI – BE COMPUTER ENGINEERING
2011-2015

SKILLS

Languages: python (experienced), rust (new)
Databases: PostGreSQL, Qdrant (vector-db)
Libraries: numpy, FastAPI, SQLalchemy, pandas

PTO..

E-mail: mehtapratik@outlook.com   Phone: +91-9082150282
Github & Twitter: pratikmehta64   Linkedin: pratikmehta14

PROJECTS

LegalEase
- Developed LegalEase - an early RAG system for legal document QnA, featuring a cross-lingual search capability over a multi-lingual document collection.
- LegalEase was awarded 2nd place in the Cohere multilingual semantic search hackathon organized by lablab.ai and sponsored by Qdrant cloud and Cohere.

Semantic Segmentation of Cardiac MRI images [Class project]
- Trained a CNN model to perform semantic segmentation to identify the boundary of the right ventricle in cardiac MRI images for cardiomyopathy detection.

PUBLICATIONS

The challenges of optimizing machine translation for low resource cross-language information retrieval
[NAACL '19]
- Studied correlation between different machine translation evaluation metrics and their effectiveness in predicting downstream cross-lingual   information retrieval performance

The IIT Bombay English-Hindi Parallel Corpus
[LREC '18]
- Helped cleaning what was at the time the largest collection of English-Hindi parallel corpus dataset

Investigating the potential of post-ordering SMT output to improve translation quality
[ICON '15]
- Established a 'skyline' BLEU score for machine-generated translations in Indian language-English SMT systems by finding out the best possible word-order for them (as compared to a human-generated reference translation)
- Developed a greedy-based decoder to construct a better sequence or word-order without changing the SMT system's lexical choices (actual words in the sentences).

E-mail: mehtapratik@outlook.com   Phone: +91-9082150282
Github & Twitter: pratikmehta64   Linkedin: pratikmehta14