

Pratik Mehta

+91 908-215-0282 | mehtapratik@outlook.com | linkedin.com/in/pratikmehta14 | github.com/pratikmehta64

EDUCATION

University of Massachusetts Amherst

Master of Science in Computer Science

Amherst, MA

Aug 2016 – May 2019

D. J. Sanghvi College of Engineering, University of Mumbai

Bachelor of Engineering

Mumbai, India

Aug 2011 – Jun 2015

EXPERIENCE

Consultant, Machine Learning & GenAI

Deloitte Digital USI

Nov 2020 – Jun 2024

Mumbai, India

- Successfully designed & developed REST APIs using FastAPI, PostgreSQL, and Llama served over AWS to answer user questions about live quantitative enterprise KPI data for Beta testing by a few thousand users
- Personally initiated driving deep changes to the in-house black-box testing practice for LLM-in-the-loop applications for use across clients
- Developed use-cases and AI-adoption strategies for AI-aspirant enterprise clients

Research Engineer

Indian Institute of Technology Bombay

Aug. 2015 – Jul. 2016

Mumbai, India

- Trained and deployed phrase-based statistical machine translation systems for English-Marathi and English-Indonesian using medical-domain textbooks
- Designed an error analysis engine to identify and classify syntactic errors in translations, and developed a greedy algorithm for automatic post-editing, evaluated via BLEU scores and human evaluators

Research Assistant

Information Sciences Institute, University of Southern California

June 2018 – Sept 2018

Waltham, MA

- Studied correlation between different evaluation metrics for machine-translation systems and their downstream impact on cross-lingual search applications

PUBLICATIONS

The IIT Bombay English-Hindi parallel corpus | *LREC '18*

The challenges of optimizing machine translation for low-resource CLIR | *EMNLP '19*

Investigating the potential of post-ordering SMT output to improve translation quality | *ICON '15*

PROJECTS

Sony Persona-Grounded Commonsense Reasoning challenge - API track

May 2025

- Tweaked prompts, API, and parameter choice to improve tool-calling accuracy for a videogame NPC agent

Alfred: LLM powered dispatch of hyper-personalized notifications

Jan 2024

- Conceptualized and developed a proof-of-concept system to track users' plain-language notification preferences and notify them based on live-data events in the application

LegalEase: Cross-lingual semantic search engine

Mar 2023

- Developed an application that answers questions factually based on lengthy legal documents (now called Retrieval Augmented Generation)
- Also developed a multi-lingual semantic search engine for any-language-query to any-language-document retrieval

TECHNICAL SKILLS

Languages: Python, Rust, SQL (Postgres), HTML

Frameworks: FastAPI, gradio, streamlit, qdrant, pytorch

Developer Tools: Git, Docker, VS Code, RustRover

Libraries: pandas, NumPy, scikit-learn