

Aadit Deshpande

linkedin.com/in/aadit-deshpande/

Email : aaditd@andrew.cmu.edu

Mobile : (717)-917-9082

EDUCATION

Carnegie Mellon University

Master's Degree, LTI, School of Computer Science — **GPA: 4.0**

- **Courses** : Introduction to ML, Advanced NLP, Speech Processing, Search Engines, Subword Modeling, Linguistics

Pittsburgh, PA

Expected Dec 2024

Birla Institute of Technology and Science, Pilani

Bachelor of Engineering Computer Science — **GPA: 9.64/10**

Pilani, India

May 2023

EXPERIENCE

Siemens Digital Industries Software

Software Development Intern, NX LLM Integration Team

Cincinnati, OH

May 2024 - Present

- Scraped a database of video transcripts for the Retrieval Augmented Generation (RAG) chatbot for Siemens NX.
- Identified the best ChromaDB embedding models and distance functions for the RAG pipeline using TREC Eval.
- Prompted Claude-3-Sonnet to generate video descriptions and synthetic QA pairs to augment the transcript data.
- Performed prompt engineering with the Claude-3-Sonnet/Haiku models and GPT-4 for the chatbot component.
- Evaluated the end-to-end RAG pipeline using RAGAS framework and GPT-4 (using Azure OpenAI Service API).

Teaching Assistant, Machine Learning Department

10601 Introduction to Machine Learning

Carnegie Mellon University, Pittsburgh, PA

Jan 2024 - May 2024

- Designed and graded homework assignments and exams for a class of 400 students under Prof. Matt Gormley
- Improved weekly office hour attendance by 50% and mentored undergraduate and graduate students.
- Led a skill team of TAs to create exam questions on ML topics like logistic regression, significance testing, and reinforcement learning algorithms for three exams.

American Express, AI Labs

Analyst Intern, Advanced NLP Team

Bangalore, India (Remote)

Jul 2022 - Dec 2022

- Created novel scrapers using Reddit API to build a consumer complaints database for the 'External Perspectives' division of project 'VoCAL' (Voice of Customer by AI Labs).
- Improved efficiency by 30% by customizing unsupervised intent detection codebase to summarize Reddit Threads.
- Developed an unsupervised aspect-based sentiment analysis module (sentenceBERT, nltk.vader) and conceptualized a scoring mechanism to retrieve high-engagement posts.
- Collaborated with a cross-functional team and integrated the end-to-end Reddit pipeline with project 'VoCAL'.

PROJECTS

Evaluating the rhyming capabilities of Large language models (Prof. D Mortensen) Mar 2024 – May 2024

- Mined English, Dutch datasets with 5000 rhyme word pairs using pronouncing dictionaries (CMUDict, Celex2).
- Inferred open-source LLMs (Llama2-7b, Llama3-8b, and CrystalChat-7b) (experiments with prompt variation).
- Achieved 0.687 and 0.649 F1 scores on English and Dutch using Llama3-8b, the best performing model.
- Proposed an error typology based on LLM mistakes due to implicit and explicit reasoning mismatch.

Cascaded vs. End-to-End Speech-to-Speech Translation systems

Feb 2024 – May 2024

- Built real-time S2ST systems: Cascaded using the pre-trained OWSM v3.1 (Speech-to-text translation) and VIT TTS (text-to-speech) models, and a discrete unit E2E ESPNet model trained on CVSS-C.
- Achieved ASR scored of 14.90 (E2E) and 17.69 (Cascaded) by forward feeding the models on the CVSS-C Spanish-to-English dataset.
- Fine-tuned the Cascaded model on the CVSS-C dataset (with LoRA), for a fairer comparison with the E2E model.
- Used new modality and ASR-agnostic translation metrics like BLASER2.0, COMET, and METEOR, more aligned with human judgement.

Maintaining Consistency in extended LLM text generation (Prof. C Rose)

Feb 2024 – May 2024

- Proposed a novel architecture for turn-wise text generation using a cascade of Llama2-7b-chat-hf generator and Maximal Marginal Relevance (MMR) summarizer models.
- Performed prompt engineering with parameteric medical persona blog posts to evaluate turn-wise text consistency.
- Analyzed the generations (blogs) qualitatively, and quantitatively by proposing 5 new factual plausibility metrics.

TECHNICAL SKILLS

Languages: Python, Java, C/C++, SQL, R

Deep Learning: PyTorch, TensorFlow, Keras

Libraries: ESPnet, Sci-kit learn, NLTK, Spacy, Huggingface transformers

Tools: Slurm scheduler, Git, Unity, L^AT_EX, PowerBI, Data Studio