Priyesh Vakharia

https://priyesh1202.github.io | priyeshvakharia09@gmail.com | github.com/Priyesh1202 | linkedin.com/in/priyeshvakharia

Education

- **University of California Santa Cruz**   **June 2024-Present**
  PhD in Computer Science

- **University of California Santa Cruz**   **Sept. 2024**
  Master of Science in Natural Language Processing   GPA : 4.00

Experience

- **Summer Intern, UC Santa Cruz**   **(July 23 – Sept. 23)**
  - Developed and deployed **CruzChat**: a full stack ChatBot using NextJS, Django and AWS with streaming capabilities. Integrated with **OpenAI** and **LLAMA**.
  - Implemented an "**Experiment Site**" that facilitates NLP research with LLMs such as **context aware interactions**, **long-term personalization**, **hallucination detection**, etc.
  - Integrated the ChatBot frontend with the "Experiment Site" to allow for continuous monitoring and automatic evaluation of deployed models.

- **Software Engineer Intern, J. P. Morgan**   **(May 2021 - July 2021)**
  - Developed a ETL Pipeline: Large data fetching from API, efficient distributed processing using Big Data tools and storage in the database.
  - Converted the existing manual workflow which included fetching records from the vendors' API, extracting and processing the relevant records and then storing it into the database to an automated workflow using Spark, improving operation speed by **23%**.

Skills

Languages:

Python, C++, SQL (MySQL, Postgres), JavaScript, HTML, CSS, NextJS, Django, Git CI/CD.

Other Skills::

Explainability, Huggingface, PyTorch, LangChain, Prompt Engineering, Machine Learning, Natural Language Processing (Parsing, Machine Translation, Generative Chatbots, QA and Summarization, LLMs, Linguistics), Transformers, Big Data, Data Science, Data Warehousing, Streamlit, AWS, CUDA.

Soft Skills:

Research, Teamwork, Project Management, Technical communication, Teaching, Adaptability.

Publications

- **ProSLM: A Prolog Synergized LM for explainable Domain Specific Knowledge Based QA**   **NeSy 2024**
  - Led the development of ProSLM, a neurosymbolic framework designed to combat LM-caused misinformation in question-answering tasks by integrating a **Prolog based** logical reasoning with existing large language models (LLMs).
  - Enabled **explainable** reasoning to generate relevant and trustworthy context for queries and validation to ensure factual accuracy against the knowledge base.
  - Introduced a novel approach in neurosymbolic generative AI, enhancing text validation and **user personalization** in LLMs.
  - Selected for a **spotlight presentation** at NeSy. 2024

- **Low-Resource Formality Controlled NMT Using Pre-trained LM**   **IWSLT 2023(ACL)**
  - Modified a multilingual translation LLM to conditionally generate formal and informal sentences in low data and computation resource settings.
  - Acheived an average **gain** of **20 BLEU** score compared to last year's submissions.

- **Preprocessy: A Customisable Data Preprocessing Framework with High-Level APIs**   **CDMA 2022**
  - Developed a novel, automated, yet customizable data preprocessing pipeline.
  - Reduced pipeline inference time by an average of **38%** by using native python instead of libraries for each data processing stage.

Projects

- **Model Interpretability for detection of hallucination in dialogue summarization, Cisco**   **Apr 23 - Dec 23**
  Advised by Prof Leilani Gilpin, Francis Kurupacheril and Prof Ian Lane.
  - Lead the research and implementation of an interpretable approach to detect hallucinations on an NVIDIA A100.
  - Created a special dataset of summaries annotated at the token level with hallucination categories.
  - Performed **distributed training** of the final model on a cluster of Nvidia GPUs, using **PyTorch**, reducing training time by **43%**.
  - Acheived a **gain** of **+4 ROUGE** score.