Deep Gandhi

+1 (403) 401-2563 · drgandhi@ualberta.ca · deep1401 · deep1401.github.io · deep1401

Education

University of Alberta Sep 2022 - Mar 2024
MSc. in Computing Science GPA: **3.9/4.0**

Thesis Topic: Debiasing Large Language Models (LLMs) using Distributionally Robust Optimization (DRO).
Graduate Teaching Assistant (GTA) for CMPUT 267: Basics of **Machine Learning** for Fall 2022 and Fall 2023
Graduate Teaching Assistant (GTA) for CMPUT 200: Ethics of **Data Science** for Winter 2023
Awarded the **Alberta Graduate Excellence Scholarship** 22-23.

Experience

SimPPL, Research Engineer Aug 2022 - April 2023
- Worked on tracking online misinformation spread (supported by **Wikimedia**, **Google**, **AWS**).
- Led an analysis project for **Yale Daily News** and **Vermont Digger** to track news spread on social media.
- Set up CI/CD pipelines and deployment scripts for **Kubernetes** (Kind and **AWS EKS**) using Docker and Helm.
- Architected and developed **microservices** in Python and JavaScript, configured NGINX and API Gateways.

Research Collaboration, Independent Researcher June 2022 - Apr 2023
- Curated a dataset of ∼1M tweets in Hindi & conducted emoji prediction using bi-LSTM, mBERT, IndicBERT, Hindi-Electra, XLM-R, etc. **(Accepted at EMNLP 2022)**
- Standardized 9 hate-speech datasets and experimented with AWD-LSTM, BERT, FNet, DistilBERT, RoBERTa, etc. to obtained **14.52%** improvement in F1-score. **(Accepted at EACL 2023)**

Unicode Research, Research Student Jan 2021 - Apr 2023
- Served as TA for **Google Research** funded 9-week Machine Learning Course UMLSC 2021 with 100+ students.
- Maintained and helped in setting up **CI/CD workflows, Heroku/Netlify deployment scripts**, API documentation, & logging using GitHub Actions, ApiDoc.js, Morgan, and Winston.
- Structured mongoose schemas, built backend controllers, wrote unit-tests, handled integration with React-client using **Node.js, Mongoose, Passport.js, Mocha.js, & Chai.js**.
- Utilized Huggingface and Langchain to modify the RLHF rewards policy to train debiased LLMs using bias data.

JPMorgan Chase & Co., Software Engineer Intern June 2020 - Aug 2020
- Worked with the Investment Banking team to automate validation checks for every release using pandas.
- Designed an automated system using **GCP** and **SQL** for evidence store creation for ServiceNow tickets reducing the processing time by **90%**.

Margosatree Technologies, Software Engineer Intern Jan 2020 - June 2020
- Experimented with **Hadoop** & **Spark** to detect 129 anomalies in 2TB of real-time data using iForests algorithm.
- Developed React.js dashboard by reading dynamic data from Rpy using **chart.js, Flask** and **SQLite**.

Projects

Software Engineering & ML | Vidalytics - Video Analysis for Marketing
- Developed video analysis & marketing dashboard using **Azure** for keyword identification, & sentiment analysis.
- Utilized **Flask, Azure, mySQL and Reddit Praw** data mining tool to provide targeted marketing insights.

ML | Automotive Component Failure Prediction *Guide: Dr. Kriti Srivastava*
- Collaborated with team at **Deloitte** to predict tyre life in vehicles using models such as MLP, XGB, etc..
- Designed a case study for the firm regarding tyre life uncertainty after extensive data analysis.

Skills

Languages:	Python, JavaScript, TypeScript, C, C++, HTML
Libraries/Frameworks:	CSS, Node.js, React.js, Redux, Electron.js, Socket.io, Puppeteer, Mocha.js, Chai.js, Jest, jQuery, Flask, PyTorch, Tensorflow, Keras, Huggingface, Pandas, scikit-learn, NumPy, Matplotlib
Databases:	MongoDB, Redis, MySQL, PostgreSQL
Tools:	Git, GCP, Azure, AWS EC2, S3, Docker, Jupyter, Bash, Heroku, Figma, CI/CD, LaTeX

Selected Publications [* *represents equal contribution*]

[1] **Deep Gandhi***, Jash Mehta*, Nirali Parekh, Karan Waghela, Lynette D'Mello, and Zeerak Talat, "A Federated Approach to Predict Emojis in Hindi Tweets," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* ***(EMNLP)***, Association for Computational Linguistics, 2022.

[2] Jay Gala*, **Deep Gandhi***, Jash Mehta*, and Zeerak Talat, "A Federated Approach for Hate Speech Detection," in *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* ***(EACL)***, *Short Papers*, Association for Computational Linguistics, 2023.