Mehak Preet Dhaliwal

+1 (510) 927-1246 | mdhaliwal@uscd.edu | mehak-preet | mehak126 | Mehak Preet

EDUCATION

University of California, San Diego — Fall 2021-Spring 2023
Master's in Computer Science — *GPA: 4/4*

Indian Institute of Technology, Delhi — July 2015-May 2019
Bachelor of Technology, Computer Science & Engineering — *GPA: 8.33/10*

PUBLICATIONS

- Noveen Sachdeva, **Mehak Preet Dhaliwal**, Carole-Jean Wu, Julian McAuley *"Infinite Recommendation Networks: A Data-Centric Approach"* NeurIPS '22
- Muskaan*, **Mehak P. Dhaliwal***, A. Seth. *"Fairness and Diversity in the Recommendation and Ranking of Participatory Media Content"* Intelligent Information Feed, KDD, 2019 & ASONAM 2021
- **Mehak Preet Dhaliwal**, Hemant Tiwari, Vanraj Vala. *"Automatic Creation of a Domain Specific Thesaurus Using Siamese Networks"* IEEE 15th International Conference on Semantic Computing (ICSC), 2021
- **Mehak Preet Dhaliwal***, Rishabh Kumar*, Mukund Rungta, Hemant Tiwari, Vanraj Vala *"On-Device Extractive Text Summarization"* IEEE 15th International Conference on Semantic Computing, 2021
- Mukund Rungta*, Rishabh Kumar*, **Mehak Dhaliwal***, Hemant Tiwari, Vanraj Vala *"TransKP: Transformer based Key-Phrase Extraction"* International Joint Conference on Neural Networks, 2020
- Mukund Rungta*, Praneet Prabhakar Sherki*, **Mehak P. Dhaliwal***, Hemant Tiwari, Vanraj Vala *"Two-Phase Multimodal Neural Network for App Categorization using APK Resources"* IEEE 14th ICSC, 2020

* indicates equal contribution

WORK EXPERIENCE

University of California, San Diego — San Diego, CA
Teaching Assistant — *Jan 2022 - Present*

- **Spring 2022:** CSE 256: Statistical Natural Language Processing with Prof. Ndapa Nakashole
- **Fall 2022:** CSE 258: Web Mining and Recommender Systems with Prof. Julian McAuley

University of California, San Diego — San Diego, CA
Graduate Student Researcher, Advisor- Prof. Julian McAuley — *Jan 2022 - Present*

- **Infinite Recommender Networks**
  - Explored the use of **Neural Tangent Kernels** to train an infinite-width auto-encoder (∞-AE) with a closed-form solution and a single hyper-parameter for recommendation, achieving state-of-the-art performance on multiple datasets.
  - Utilised ∞-AE to develop **data distillation for collaborative filtering datasets**, resulting in 96-105% of ∞-AE's performance on the full dataset with as little as 0.1% of the original dataset size. See publication.
- **De-biasing Large Language Models**
  - Analysing **social biases in Large Language Models** (GPT-3, GPT-2, CTRL etc.) under unified fairness frameworks and metrics.
  - Working on **prefix-tuning and prompting** techniques for reducing social biases while maintaining performance metrics on downstream tasks.

Tonita — New York City, New York
Software Engineering Intern — *June 2022 - Sept 2022*

- Developed a novel **question-answering** model for handling multiple questions for a context and heterogeneous answer types (extractive, boolean, numerical, no-answer etc.) for reduced inference latency and increased use-case flexibility.
- Demonstrated an end to end search pipeline utilising developed models for **NLP-augmented semantic search**.
- Explored text **data augmentation** techniques including GPT-3 for data generation, paraphrasing and table-to-text.

Samsung R&D Institute — Bangalore, India
Senior Engineer — *June 2019 - Aug 2021*

- Developed a character-level model for **extractive text summarization**, outperforming several baselines and state-of-the-art models with a 97%+ reduction in parameters. Model ported on-device for enhancing search & assistance features
- Designed a novel method for automatic creation of a **domain-specific thesaurus** using Siamese Networks. Constructed lexical database is under commercialisation for use in improving keyword based object retrieval.
- Conceptualised and developed solutions for **fine-grained entity recognition**, **Key-phrase extraction** using transformers and **classification of applications**. Solutions outperformed several state-of-the-art methods and were integrated on-device.

Samsung R&D Institute — Bangalore, India
Software Research Intern — *May 2018 – July 2018*

- Implemented a Bi-LTSM model with attention using Keras framework for **emoji prediction from text**
- Achieved accuracy comparable to the state of art model with model size reduced to 1/6 (pre-quantization)

Indian Institute of Technology Delhi, India

Research Intern, Advisor- Prof. Subodh Sharma *May 2017 – July 2017*

- Set up a client-server model using Raspberry Pis where client received data from servers via TCP connection
- Detected anomalous network data sent by servers through outlier detection using k-means clustering on client
- Analysed k-means initialisation algorithms, distance metrics and cluster boundaries for outlier detection

COURSE PROJECTS

Constituency Parsing

Prof. Taylor Berg-Kirkpatrick (UCSD) *Structured Prediction for NLP, Spring 2022*

- Implemented a neural Conditional Random Field (CRF) over a Bi-LSTM in Pytorch for Constitutency Parsing on the English WSJ Penn TreeBank (PTB) data set
- Performed a qualitative and quantitative analysis of the model's performance (See report)

Machine Translation

Prof. Taylor Berg-Kirkpatrick (UCSD) *Structured Prediction for NLP, Spring 2022*

- Implemented a GRU-based encoder-decoder with attention in Pytorch for German to English translation on the Multi30K dataset
- Performed a quantitative and qualitative comparison of greedy decoding, beam search and nucleus sampling decoding methods. (See report)

Prototype Selection for Nearest Neighbors

Prof. Taylor Berg-Kirkpatrick (UCSD) *ML: Learning Algorithms, Winter 2022*

- Proposed and implemented an algorithm for prototype selection for nearest neighbor classification following a two-step process: eliminating noisy data points followed by K-means clustering based prototype set construction.
- Evaluated the performance of the approach on the MNIST dataset for different sizes of the prototype set and showed consistent improvement over the baselines. (See report)

Learning General Purpose Sentence Representations via Multi-task Learning

Vineet Kumar, IBM Research (IIT, Delhi) *Deep Learning, Fall 2018*

- Trained RNN based Seq2Seq models on Natural Language Inference, Constituency Parsing & Machine Translation via multi-task learning with a shared sentence encoder
- Evaluated encoder performance on text classification, paraphrase identification and semantic similarity tasks (See project)

Attention Models in CNNs

Dr. Raghavendra Singh, Director (Oyla)(IIT, Delhi) *Deep Learning, Fall 2018*

- Added attention layers to the AlexNet CNN model to improve classification of CIFAR-10 images
- Visualized attention maps demonstrating the model's focus on the main object while ignoring background (See project)

AIDS Detection in Molecules

Prof. Sayan Ranu (IIT, Delhi) *Data Mining, Fall 2018*

- Identified discriminative subgraphs from a data set containing molecules active and inactive against HIV virus
- Modelled the presence/absence of each extracted subgraph as a binary feature vector for every molecule
- Classified molecules as active/inactive for the HIV virus by training a linear LIBSVM kernel on the data set

TECHNICAL SKILLS

Python, C/C++, Java, PyTorch, TensorFlow, TF Lite, Keras, Hugging Face, Git, Android, LaTeX, SQL

AWARDS AND HONORS

- **Excellence Award- Innovator | Samsung R&D Institute, Bangalore:** For work on research to market, Patents and Publications, building creative culture & enhancing software development ecosystem *(June 2021)*
- **Citizen Award | Samsung R&D Institute, Bangalore:** For excellence in work, extraordinary commitment, extension of support and achievement beyond functional scope & process improvements *(Jan 2021)*
- **Spot Award | Samsung R&D Institute, Bangalore:** For excellent contribution to research output *(Apr 2020)*
- **Suresh Chandra Memorial Trust Award | Indian Institute of Technology, Delhi:** Awarded for the best software project in Computer Science and Engineering discipline during the session 2018-2019 *(Nov 2019)*

RELEVANT COURSEWORK

- **University of California, San Diego:** Probabilistic Reasoning and Decision Making, Search and Optimization, Web Mining and Recommender Systems, ML: Learning Algorithms, Convex Optimization, Structured

Prediction for Natural Language Processing, Principles in Computer Architecture

- **Indian Institute of Technology, Delhi:** Principles of Artificial Intelligence, Machine Learning, Data Mining, Optimization Methods and Applications, Analysis and Design of Algorithms, Operating Systems, Introduction to Database Management Systems, Data Structures & Algorithms, Discrete Mathematical Structures

OTHER INITIATIVES

Allyship Programs Chair, Grad Women in Computing (GradWIC) San Diego, CA
University of California, San Diego *September 2022 - Present*

- Planned events for the academic year focussing on allyship, diversity and fostering an inclusive community at UCSD
- Collaborated with the community centers at UCSD (Women's Center, Cross Cultural Center, Black Resource Center, LGBT Resource Center, Raza Resource Centro etc.) for a tour of the centers and resources on campus
- Acting as a mentor as part of the GradWIC mentorship program

Samsung CSR Bangalore, India
Samsung R&D Institute *August 2019*

- Volunteered in Samsung's CSR initiative- distributed books and stationary to school children in rural Karnataka

National Association for the Blind (NAB) Delhi & Chandigarh, India
National Association for the Blind *Summer 2017 & Winter 2018*

- Volunteered at NAB, South Delhi, assisting visually impaired students with college applications, digitizing attendance records and preparing posters for stalls
- Volunteered at NAB, Chandigarh in the Inclusive Education and Computer Education program for familiarizing students with the voice assistive software

Activity Head, Social Campaign, Rendezvous IITD Delhi, India
Indian Institute of Technology *April 2017 - Oct 2017*

- Worked in the social campaign, 'The Red War' with the aim of raising awareness and support for thalassemia patients
- Co-organized a pan-India blood donation drive with BloodConnect and held awareness sessions in Delhi
- Collaborated with the National Thalassemia Welfare Society (NTWS) and organised a performance event by children with thalassemia

Project Executive Delhi, India
Enactus - IIT Delhi Chapter *Feb 2016 - Aug 2016*

- **Project Titli:** Project to improve menstrual health and hygiene conditions of women in rural areas of India
  - Visited and conducted surveys in different regions of Delhi to identify problems related to awareness, accessibility and affordability of menstrual products
  - Held meetings with doctors and gynecologists to understand medical issues with poor menstrual hygiene
  - Helped in establishing the first low cost sanitary napkin manufacturing hub in Tilak Nagar, Delhi run by women from the community
  - Collaborated with local NGOs for product distribution and to hold awareness sessions in their communities to help eradicate social stigma and myths surrounding menstruation
- **Project Nirmalya:** Project to improve waste disposal system of Delhi through decentralized waste management
  - Conducted campaigns and door to door awareness drives in residential societies of Delhi to educate and encourage residents for segregating waste
  - Collaborated with the South Delhi Municipal Corporation to convert local waste storage depots into composting plants for organic waste and connected with E-Kabadi for recycling dry waste
- Represented the team in the Enactus National Competition, 2016 and were declared National runners up from among 72 teams