

## Dhruv Kumar

ddhruvkr@gmail.com | +1 226-978-3751 | Website | GitHub | LinkedIn | Google Scholar

### EDUCATION

**University of Waterloo, School of Computer Science, Waterloo, Canada**

Sep 2018 - Aug 2020

Master of Mathematics (Thesis), Computer Science

GPA: 86.75/100 (3.88/4)

**Thesis:** Iterative Edit-based Unsupervised Sentence Simplification

**Indian Institute of Information Technology, Allahabad, India**

Jul 2012 - Jun 2016

Bachelor of Technology (Hons.), Information Technology

GPA – 8.68/10 (9.07 - Last 2 years)

**Thesis:** Compressed Knowledge transfer via Factorization Models in Recommender Systems

### SKILLS

- **Relevant Coursework (Graduate and Undergraduate):** Deep Learning, Trust Modeling and Online Social Networks, Privacy and Fairness in Data Science, Information Retrieval, Artificial Intelligence, Natural Language Processing, Cognitive Process Modelling, Optimization Techniques, Data Mining, Probability & Statistics, Mathematics(I, II, III).
- **Languages and Frameworks:** Python, Pytorch, Keras, Java, SQL, C/C++
- **Interests:** Natural Language Processing, Machine Learning

### SELECTED PUBLICATIONS

- S. Gehrmann et al., The GEM Benchmark: Natural Language Generation, its Evaluation and Metrics, ArXiv 2021.
- O. Vechtomova, G. Sahu, **D. Kumar**, Generation of lyrics lines conditioned on music audio clips, In Workshop on NLP for Music and Audio (NLP4MusA) at ISMIR 2020.
- **D. Kumar**, L. Mou, L. Golab, O. Vechtomova, Iterative Edit-based Unsupervised Sentence Simplification, In Proceedings of the 58th annual meeting of the Association for Computational Linguistics (ACL 2020) - Long Paper
- R. Cohen, R. Agarwal, **D. Kumar**, A. Parmentier, T. H. Leung, Sensitivity to risk profiles of users when developing AI systems, In 33rd Canadian Conference on Artificial Intelligence (Canadian AI 2020)
- R. Agarwal\*, **D. Kumar\***, L. Golab, S. Keshav, Consentio: Managing Consent to Data Access using Permissioned Blockchains, IEEE International Conference on Blockchain and Cryptocurrency (ICBC) 2020 - Full Paper
- **D. Kumar**, R. Cohen, L. Golab, Online abuse detection: the value of preprocessing and neural attention models, In 10th workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA) at NAACL-HLT 2019

### PROFESSIONAL EXPERIENCE

**Borealis AI, Toronto, Canada**

*Machine Learning Research Intern, Alan Research*

Sep 2020 - Jan 2021

- Worked on explicit schema linking and regularization techniques to improve cross-domain generalizability and performance of a state-of-the-art semantic parser on the Spider dataset.
- Showed the efficacy of data-dependent initialization for transformers in improving generalization and for training deeper models on small datasets.
- The paper is under review.

**Arcesium(DE Shaw Group), Hyderabad, India**

*Software Engineer, Fund and Investor Accounting*

Jul 2016 - May 2018

- Enhanced the post-trade automation platform for funds operated by J.P. Morgan and D.E. Shaw.

**Citigroup, Pune, India**

*Software Engineering Intern, Equities*

May 2015 - Jul 2015

- Designed and implemented the first prototype of the Trading Controls application. Declined the full-time offer.

### ACADEMIC PROJECTS

**GEM Benchmark**

Sep 2020 - Present

- Working with Prof. Wei Xu, Sebastian Gehrmann, Mounica Maddela, Prof. Ondrej Dusek and many others on the GEM benchmark for natural language generation, evaluation, and metrics, to be held as a workshop at ACL 2021.

## **University of Waterloo, Waterloo, Canada**

### **Music audio conditioned lyrics generation**

**Sep 2019 - Apr 2020**

- Designed bimodal neural network models based on variational autoencoders(VAEs) to generate lines of lyrics for instrumental pieces of music.

### **Iterative Edit-based Unsupervised Sentence Simplification**

**Jul 2019 - June 2020**

- Designed an unsupervised algorithm building on edit-based text generation techniques for sentence simplification. The model is more controllable and interpretable and achieves SARI scores of 30.44 and 37.85 on the Newsela and Wikilarge datasets respectively.

### **Unsupervised neural models to anonymize personal attributes in text**

**Oct 2019 - Dec 2019**

- Human written text contains implicit linguistic information that can be used to identify our attributes such as gender, age and political leanings. Built an unsupervised neural model for multi-attribute style transfer that uses a combination of denoising, cycle consistency and classification losses.

### **Consentio: Managing Consent to Data Access using Permissioned Blockchains**

**Jan 2019 - Aug 2019**

- Designed a consent management system based on permissioned blockchains that can handle up to 6000 transactions per second.

### **Attention-based Text classification**

**Dec 2018 - Jan 2019**

- Implemented various deep learning models (e.g. Co-attention, Self-attention, Hierarchical attention) for text classification. The models were inspired by research conducted for different tasks in NLP.

### **Detecting Incivility in online social networks**

**Oct 2018 - Feb 2019**

- Proposed a co-attention based neural model for online abuse detection. The model achieves F1 scores of 82.41, 77.75 and 76.07 for the minority abuse class on the Wikipedia toxicity/attack/aggression datasets respectively.

## **Universität Paderborn, Paderborn, Germany**

### **Compressed Knowledge transfer via Factorization Models in Recommender Systems**

**Jan 2016 - Jun 2016**

- Developed an algorithm to incorporate metadata in Factorization Machines, lowering RMSE value to 0.836 as compared to 0.853 when using a Joint Matrix Factorization method on the MovieLens 1M dataset.

## **Indian Institute of Information Technology, Allahabad, India**

### **Analysis of Time-Aware and Semantic Feature-Based Music Recommender System**

**Jul 2015 - Dec 2015**

- Proposed a Joint Matrix Factorization algorithm for Music Recommendation that utilizes geographical and time-based tagging information of artists, in addition to implicit user feedback (user clicks).

## **SERVICE**

---

- Program Committee / Reviewer: NeurIPS 2019 AISG (AI for Social Good) Workshop
- Program Committee / Organizer: GEM: Natural Language Generation, Evaluation, and Metrics- ACL 2021

## **TEACHING ASSISTANT EXPERIENCE**

---

- CS 115: Introduction to Computer Science, Fall 18, University of Waterloo.
- CS 230: Introduction to Computers and Computer Systems, Winter 19, Spring 20, University of Waterloo.
- CS 231: Algorithmic Problem Solving, Spring 19, University of Waterloo.
- CS 241: Foundations of Sequential Programs, Fall 19, University of Waterloo.
- CS 489/698: Topics in Computer Science, Neural Networks, Winter 20, University of Waterloo.

## **ACHIEVEMENTS**

---

- Received International Masters Student Award and University of Waterloo Entrance Scholarship for graduate studies.
- Got Graduate Studies Research Travel Assistantship to attend the NAACL-HLT Conference, 2019.
- Accepted in the Deep Learning and Reinforcement Learning Summer School, 2019 (less than 25% acceptance rate).
- Served as the Events Head of the annual cultural cum technical festival - Effervescence 2014.
- Represented the college band as a drummer.
- Stood in the top 0.5% in the All India Engineering Entrance Examination 2012.