

Dhruv Kumar

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

EXPERIENCE

Grammarly, Vancouver, Canada

Senior Applied Research Scientist

Jul 2021 - Feb 2024

- Working on RAG and integrating tools into LLMs such as calendar, email APIs, internal Wiki, etc.
- Led a team of five, including linguists and ML engineers, to build the core text-generation model behind [Voice Composer](#), a voice-to-text [application](#), currently being used by millions of users. Improved the model on multiple aspects (such as robustness, fluency, coverage, naturalness, etc) by over 10% points. [\[EMNLP 2023\]](#) [\[Patent\]](#)
- Led a team to build a [paraphrasing](#) model and service that is being used by over 100k users per week.
- Developed and open-sourced [CoEdit](#) - a series of small instruction-tuned [LLMs](#) for text revision and editing ranging from 770M to 11B parameters that perform competitively to the biggest LLMs available. [\[EMNLP 2023\]](#)
 - The models have been downloaded over 100k times on HuggingFace. [\[Models\]](#) [\[Data\]](#)
 - Led to the creation of an internal team to build a similar model that now serves millions of users.
 - Extended the work to a **multilingual** setting covering 6 language families. [\[Paper under review\]](#)
- Developed the first version of meaning change classifier for sentence and paragraph level text for internal use.
- Mentored (and co-mentored) four research interns on:
 - Iterative text editing [\[ACL 2022, In2Writing@ACL 2022, EMNLP 2022, Code\]](#) — LLMs for Self-contradictions in long documents [\[ArXiv 2023, Code, Under Review\]](#) — Personalization in LLMs [\[Personalize@EACL 2024, Code\]](#)
- Published 8+ papers and three patents ([Google Scholar](#)) and took over 75 interviews for ML positions.

Borealis AI, Toronto, Canada

Machine Learning Research Intern

Sep 2020 - Dec 2020

Turing - text-to-SQL semantic parsing

- Showed how explicit schema linking and regularization techniques can improve cross-domain generalizability.
- Showed efficacy of data-dependent initialization in training deeper transformer models on small datasets. [\[Code\]](#) [\[ACL 2021\]](#)

University of Waterloo, Waterloo, Canada

Graduate Research Assistant

Sep 2018 - Aug 2020

- **GEM Benchmark**: Worked with Prof. Wei Xu and Dr. Sebastian Gehrmann on the evaluation framework. [\[Code\]](#) [\[ArXiv 2021\]](#)
- **Unsupervised Sentence Simplification**: Designed unsupervised edit-based algorithms that are controllable and interpretable and achieve SARI scores competitive with those of supervised models. [\[ACL 2020\]](#) [\[Code 1\]](#) [\[ACL 2022\]](#) [\[Code 2\]](#)
- **Music-conditioned lyrics generation**: Designed bimodal(text/audio) models using conditional variational autoencoders (VAE) and generative adversarial networks (GAN) to generate lyric lines for audio clips. [\[NLP4MusA@ISMIR 2020\]](#)[\[ICCC 2021\]](#)[\[Demo\]](#)
- **Consentio: Managing Consent to Data Access using Permissioned Blockchains**: Permissioned blockchain-based consent management system that handles up to 6000 transactions per second. [\[Code\]](#) [\[ICBC 2020\]](#)

Arcesium (DE Shaw Group), Hyderabad, India

Software Engineer, Fund and Investor Accounting

Jul 2016 - May 2018

- Developed the post-trade automation platform (written in Java) for funds operated by J.P. Morgan and D.E. Shaw.

EDUCATION

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

Sep 2018 - Aug 2020

Master of Mathematics (Thesis), Computer Science

Thesis: [Iterative Edit-based Unsupervised Sentence Simplification](#)

Indian Institute of Information Technology, Allahabad, India

Jul 2012 - Jun 2016

Bachelor of Technology (Hons.), Information Technology

Thesis: [Compressed Knowledge transfer via Factorization Models in Recommender Systems](#)

TECHNICAL SKILLS

- **Interests**: Natural Language Processing, Machine Learning, Large Language Models (LLMs), Generative AI
- **Programming Languages and Frameworks**: Python, Java, Pytorch, HuggingFace, AWS, GCP, Scikit, NLTK

ACADEMIC SERVICE

- Reviewer: 1) ARR (ACL rolling reviews), 2) GEM: Natural Language Generation, Evaluation, and Metrics workshop, 3) In2Writing: Workshop on Intelligent and Interactive Writing Assistants, 4) AISG (AI for Social Good) Workshop