Dhruv Kumar

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

OBJECTIVE

NLP researcher with strong academic and industrial research experience in Text Generation, Evaluation, and Large Language Models, particularly specialized in generating high-quality text for robust and reliable applications. I am now looking for Research Engineer/Applied Scientist positions where I can leverage my expertise in NLP/ML.

EXPERIENCE

Grammarly, Vancouver, Canada

Senior Applied Research Scientist **Jul 2021 - Feb 2024**

- Used RAG and integrated external tools such as calendars and KBs into LLMs to make them more functional and reliable.
- Led a cross-functional team of five (SW/ML engineers and linguists) to build a text-generation model for a new Grammarly feature Voice Composer, now in production with 5M+ users. Improved the model to the production quality on multiple aspects such as **robustness, sensitivity,** and **coverage** by 10%+ points. [EMNLP '23, Patent]
- Created and deployed a paraphrasing model using **constrained decoding** techniques, now serving 100k+ users/month.
- Developed and open-sourced CoEdIT—a series of instruction-tuned LLMs for **text revision** (from 770M to 11B parameters) that perform competitively with the biggest LLMs available. [EMNLP '23]. **100k+** downloads on HuggingFace. [Models/Data]
    - Extended it to a **multilingual** setting covering six language families, using **PEFT** techniques like **LoRA**. [NAACL '24]
    - Helped launch a new internal team to develop a production-ready small model, which serves 10M+ users/month..
- Developed the first version of meaning change classifier for sentence and paragraph level text which is used in the product and internally to ensure the safety and accuracy of generated text.
- Mentored four research interns onIterative text editing [ACL '22, In2Writing@ACL '22, EMNLP '22, Code]; Detecting self-contradictions in long documents using LLMs [NAACL '24, Code]; Personalization in LLMs [Personalize@EACL '24, 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

- Analyzed how explicit schema linking and regularization techniques can improve cross-domain generalizability.
- Analyzed efficacy of data-dependent initialization in training deeper Transformer models on small datasets. [ACL '21, Code]

University of Waterloo, Waterloo, Canada

Graduate Research Assistant **Sep 2018 - Aug 2020**

- **GEM Benchmark**: Built the evaluation infrastructure for text simplification for the workshop. [ArXiv '21, Code]
- **Unsupervised Sentence Simplification:** Designed unsupervised edit-based algorithms that are controllable and interpretable and achieve SARI scores competitive with those of supervised models.[ACL '20, Code] [ACL '22, Code]
- **Music-conditioned lyrics generation:** Designed bimodal(text/audio) models using variational autoencoders (VAE) and generative adversarial networks (GAN) to generate lyric lines for audio clips. [NLP4MusA@ISMIR '20, ICCC '21, Demo]
- **Consentio:** Permissioned blockchain-based consent management system handling 6000 transactions per second. [ICBC '20]

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-learn, NLTK, Deepspeed

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