Krishnapriya Vishnubhotla

vkpriya@cs.toronto.edu • https://priya22.github.io/

I am a PhD candidate in the Computational Linguistics group at the University of Toronto, supervised by **Prof. Graeme Hirst**, and co-supervised by **Prof. Frank Rudzicz**. My main research area is **Natural Language Processing**. My expertise is in developing datasets and models for the computational analysis of literary texts, and quantifying variations in style, meaning, and emotion. I am particularly interested in the patterns of similarity in the writing styles of people and communities.

Education

- **University of Toronto** — **Toronto**
  PhD in Computer Science, Computational Linguistics Group, GPA: 3.97/4.0 — *2019–Present*
- **University of Toronto** — **Toronto**
  Master of Science in Computer Science, Thesis option, GPA: 4.0/4.0 — *2017–2019*
- **National Institute of Technology Karnataka-Surathkal** — **Mangalore, India**
  B.Tech Computer Science and Engineering , CGPA 8.93/10 — *2013–2017*
- **K.V.I.I.Sc** — **Bangalore, India**
  Primary, Secondary and High School (Grades 1 to 12), 97% AISSCE — *2001–2013*

Work Experience

- **National Research Council (NRC) Canada** — **Toronto**
  Research Intern — *June 2023 – Present*
  As a part-time research intern at NRC-Canada, I am working on formulating metrics to characterize emotional expression and variation among different demographic groups using social media utterances.
- **AI4Good Lab** — **Toronto**
  Teaching Assistant — *April 2022 – June 2022*
  The AI4Good Lab is a 7-week introductory machine learning program for under-represented groups in the field. As a TA, I designed and ran daily tutorial sessions on ML theory, programming, and mentored teams on an applied project.
- **Georgian Partners** — **Toronto**
  Research Intern — *May 2020 – Dec 2020*
  As a research intern at Georgian, I worked on developing unsupervised clustering models of text embeddings for internal company applications.
- **Samsung AI Research Center** — **Toronto**
  Research Intern — *May 2019 – September 2019*
  As a research intern at Samsung AI, I worked on multi-modal representation learning and semi-supervised methods of text and video alignment. I was a part of the winning submission for the Samsung Retail Robot Challenge.

Research Themes

- **Quotation Attribution and Character Voice in Literary Texts (Sept 2019-Current)**
  In this thesis project, I am exploring models that can capture the stylistic variation across characters and authors in literary novels. This partly involved designing neural models that can accurately attribute quotations within texts to their speakers. Outcomes include a web-platform for annotating quotations and related information, and a large dataset of English-language novels annotated for various aspects of quotation within them.
- **Characterizing Emotional Dynamics in Twitter data (Oct 2021-Current)**
  Developing computational metrics to measure characteristics of emotional expression and variation in social media data, particularly tweets. These metrics will be used to evaluate correlations between emotional profiles of demographic groups and the mental, physical and emotional health of populations. This work builds on prior studies in psychology and affective science on emotion dynamics and emotional granularity.
- **A Dataset of Semantic Textual Relatedness (March 2020-Jan 2021)**
  Created a dataset of sentence pairs annotated for semantic relatedness using Best–Worst Scaling. Explored the contribution of various linguistic features to semantic relatedness, and evaluated state-of-the-art sentence representation models on the dataset.

- **Disentangling Content and Style in Texts (Sept 2019-Jan 2020)**

  Evaluated autoencoder variants that learn disentangled representations of content and style on a highly-structured Natural Language Generation dataset. Our findings highlight the data requirements, effectiveness, and limitations of current learning methods.

Publications

- Vishnubhotla, K., Rudzicz, F., Hirst, G., and Hammond, A., 2023. Improving Quotation Attribution in Literary Novels. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp 737–746)
- Abdalla M., Vishnubhotla, K. and Mohammad, S.M., 2023. What Makes Sentences Semantically Related: A Textual Relatedness Dataset and Empirical Study. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (pp 782–796)
- Vishnubhotla, K. and Mohammad, S.M., 2022. Tweet Emotion Dynamics: Emotion Word Usage in Tweets from US and Canada. In *Proceedings of the 13th Language Resources and Evaluation Conference: LREC 2022* (pp 4162-4176)
- Vishnubhotla, K., Hammond, A. and Hirst, G., 2022. The Project Dialogism Novel Corpus: A Dataset for Quotation Attribution in Literary Texts. In *Proceedings of the 13th Language Resources and Evaluation Conference: LREC 2022* (pp 5838-5848)
- Vishnubhotla, K., Hirst, G. and Rudzicz, F., 2021, August. An Evaluation of Disentangled Representation Learning for Texts. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (pp. 1939-1951).
- Vishnubhotla, K., Hammond, A. and Hirst, G., 2019, June. Are Fictional Voices Distinguishable? Classifying Character Voices in Modern Drama. In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature 2019* (pp. 29-34).
- Budhkar, A., Vishnubhotla, K., Hossain, S. and Rudzicz, F., 2019, August. Generative Adversarial Networks for Text Using Word2vec Intermediaries. In *Proceedings of the 4th Workshop on Representation Learning for NLP 2019* (pp. 15-26).

Academic Service

- I have served as a reviewer for:
  - ACL Rolling Review: 2023, 2022, 2021
  - *ACL Conferences (ACL, NAACL): 2022, 2021, 2020, 2019, 2018
  - EMNLP: 2020, 2019
- Volunteer mentor for the Graduate Application Assistance Program, 2021.
- Served as a Triager for the DCS Admissions Program in 2021 and 2020.
- Maintained the official webpage of the Computational Linguistics group from 2018-2020.

Teaching Experience

I have served as a Teaching Assistant for various courses at the University of Toronto. My duties include preparing and delivering tutorials, quizzes and taking lab hours.
○ Introductory Computation and Data Science for the Social Sciences, Life and Physical Sciences, and Literature ○ Natural Language Computing ○ Introduction To Computer Programming ○ Introduction to Computer Science ○ Programming on the Web.

Relevant Courses

○ (Advanced) Computational Linguistics ○ Natural Language Computing ○ Introduction to Machine Learning ○ Learning Discrete Latent Structure ○ Algorithms for Private Data Analysis ○ Topics in Computational Social Science ○ Computational Models of Semantic Change

Technical skills

- **Machine Learning Frameworks:** PyTorch, TensorFlow, FastAI, HuggingFace
- **Programming Languages:** Proficient in: C, C++, Python, Javascript, Ruby
  Basic ability with: Java.