

Ronak Pradeep

Email 1 : ronakice@hotmail.com

Email 2 : rpradeep@uwaterloo.ca

Website : ronakice.github.io

Mobile : 519-781-6942

EDUCATION

- **University of Waterloo** Waterloo, ON
PhD - Computer Science *Sep 2020 - Ongoing*
 - Neural Information Retrieval and Fact Verification with Professor Jimmy Lin
 - Coursework: Differential Privacy, Optimization for Data Science, High Stakes Information Retrieval, Affective Computing
 - Teaching Assistant: Data Structures
- **University of Waterloo** Waterloo, ON
BMath - Double Major in Computer Science and Combinatorics and Optimization *Jan. 2016 – Apr. 2020*
 - Graduate Level Coursework: Deep Reinforcement Learning, Randomized Algorithms, Formal Languages and Parsing (100%), Computational Vision, Statistical Learning, Dependent Types and Software Verification (100%)
 - Part of the Term Dean's Honours List; Graduate Level Coursework Average 96.5%

RESEARCH INTERESTS

Open-Domain Question Answering, Reading Comprehension, Paragraph and Document Ranking, Fact Verification, Biomedical Natural Language Processing, Graph Representation Learning

EXPERIENCE

- **University of Waterloo** Waterloo, ON
Undergraduate Researcher *Apr 2017 - Aug 2020*
 - Worked with Professor Jimmy Lin and Rodrigo Nogueira on Paragraph Retrieval and Ranking
 - Worked with Professor Pascal Poupard on Reading Comprehension tasks
 - Worked with Professor Jeff Orchard on a Deep Biologically Plausible Vision Model
- **Montreal Institute for Learning Algorithms (MILA)** Montreal, QC
Visiting Researcher *May 2019 - Dec 2019*
 - Worked with Professor Chris Pal and Jie Fu on Open Domain Question Answering and Graph Representation Learning
- **Wish** San Francisco, CA
AI Research Intern *Jan 2018 - Apr 2018*
 - Worked on Neural Title Generation for e-Commerce Products using various Encoder-Decoder Architectures
 - Built various Neural Models for Product and Attribute Categorization
 - Curated the iMaterialist Challenge for the FGVC Workshop at CVPR 2018
- **Royal Bank of Canada** Toronto, ON
Research Developer *Aug 2016 - Dec 2016*
 - Worked on Document Ranking and Question Answering using Dual Embedding Space and Seq2Seq models
- **University of Waterloo** Waterloo, ON
Undergraduate Teaching Assistant for Math 136 - Linear Algebra *Jan 2017 - Apr 2017*

PUBLICATIONS

- **Vera: Prediction Techniques for Reducing Harmful Misinformation in Consumer Health Search** (SIGIR 2021)
Ronak Pradeep, Xueguang Ma, Rodrigo Nogueira, and Jimmy Lin
- **Chatty Goose: A Python Framework for Conversational Search** (SIGIR 2021 Demo)
Edwin Zhang, Sheng-Chieh Lin, Jheng-Hong Yang, Ronak Pradeep, Rodrigo Nogueira, and Jimmy Lin
- **Pyserini: An Easy-to-Use Python Toolkit to Support Replicable IR Research with Sparse and Dense Representations** (SIGIR 2021 Resource)
Jimmy Lin, Xueguang Ma, Sheng-Chieh Lin, Jheng-Hong Yang, Ronak Pradeep, and Rodrigo Nogueira
- **H₂oloo at TAC 2020: Epidemic Question Answering** (TAC 2020 Proceedings)
Justin Borromeo, Ronak Pradeep*, Jimmy Lin*
- **Exploring Listwise Evidence Reasoning with T5 for Fact Verification** (ACL 2021)
Kelvin Jiang, Ronak Pradeep*, Jimmy Lin*
- **H₂oloo at TREC 2020: When all you got is a Hammer... Deep Learning, Health Misinformation, and Precision Medicine** (TREC 2020 Proceedings)
Ronak Pradeep, Xueguang Ma, Xinyu Zhang, Hang Cui, Ruizhou Xu, Rodrigo Nogueira, Jimmy Lin
- **Scientific Claim Verification with VerT5erini** (LOUHI 2021: The 12th International Workshop on Health Text Mining and Information Analysis colocated with EACL 2021)
Ronak Pradeep, Xueguang Ma, Rodrigo Nogueira, Jimmy Lin
- **A Replication Study of Dense Passage Retriever** (Will be submitted to a suitable venue)
Xueguang Ma, Ronak Pradeep, Kai Sun, Jimmy Lin
- **Covidex: Neural Ranking Models and Keyword Search Infrastructure for the COVID-19 Open Research Dataset** (Scholarly Document Processing @ EMNLP 2020)
Edwin Zhang, Nikhil Gupta, Raphael Tang, Xiao Han, Ronak Pradeep, Kuang Lu, Yue Zhang, Rodrigo Nogueira, Kyunghyun Cho, Hui Fang, Jimmy Lin
- **The Expando-Mono-Duo Design Pattern for Text Ranking with Pretrained Sequence-to-Sequence Models** (Will be submitted to a suitable venue)
Ronak Pradeep, Rodrigo Nogueira, Jimmy Lin
- **Document Ranking with a Pretrained Sequence-to-Sequence Model** (EMNLP 2020 Findings)
Rodrigo Nogueira, Zhiying Jiang, Ronak Pradeep, Jimmy Lin
- **Modular Diversity-Seeking Query Reformulation for Open-Domain Question Answering**
Ronak Pradeep, Jie Fu*, Xingdi Yuan, Zhouhan Lin, Yi Tay, Chris Pal*
- **Foveated Down-Sampling Techniques** (CVIS 2020)
Parsa Torabian, Ronak Pradeep, Jeff Orchard, Bryan Tripp

ACCOMPLISHMENTS

- **SCIVER: Verifying Scientific Claims with Evidence:** Top submission based on primary metric
- **Fact Extraction and VERification (FEVER) - 1st (As of Jan 14th 2021):** State of the Art model in a widely popular Fact Verification dataset
- **TREC Health Misinformation 2020:** A task that studies search technologies that promote credible and correct information over incorrect information - Top submission in the AdHoc Retrieval task.
- **TREC Deep Learning 2020:** A track that studies information retrieval in a large training data regime - Top submission in the Document Ranking task

- **TREC-COVID 2020:** A multi-round COVID-19 Literature Ranking Task - Best Round 4, 5 Automatic Run, Best Round 3 Feedback run
- **Fact Extraction and VERification (FEVER) - 1st (As of Jan 14th):** State of the Art model in a widely popular Fact Verification dataset
- **MS MARCO Document Ranking - 1st (As of Sep 8th 2020):** State of the Art model in a widely popular Neural Document Ranking dataset
- **MS MARCO Passage Ranking - 1st (As of May 20th 2020):** State of the Art model in a widely popular Neural Passage Ranking dataset
- **DiMarco Undergraduate Scholarship in Computational Rhetoric:** Annually awarded to a single student based on academic achievement combined with a well-demonstrated interest in the area of Computational Rhetoric
- **Terminal AI - Winner:** Developed an heuristic-based AI game bot that placed 1st among teams of top Waterloo students. Globally ranked 2nd among 15k players (at the time of submission)
- **Citadel Datathon - NYC:** Placed 2nd among teams from top universities in North America
- **HackPrinceton - Top 10:** Implemented a tool for the Sentiment Analysis of Twitter and Guardian News using Vader Lexicon and Encoder-Decoder LSTMs and visualized the trends
- **University of Waterloo President's Scholarship of Distinction and Research Award:** Awarded based on high academic average and research terms

PROJECTS

- **Covidex.ai :** A neural search engine that applies state-of-the-art neural network models to answer questions using the COVID-19 Open Research Dataset (CORD-19) corpus provided by the Allen Institute for AI
- **PyGaggle :** A framework providing a gaggle of deep neural architectures for Text Ranking and Question Answering inference
- **Pyserini :** A framework supporting sparse retrieval, dense retrieval, as well as hybrid retrieval that integrates both approaches
- **NumBERT :** An open-sourced neural Passage Ranking framework integrated closely with Castorini's Anserini and HuggingFace's Transformers
- **Deep RL for Generative Conversational Agents :** Course Project for Deep Reinforcement Learning
- **Approximate Nearest Neighbour Algorithms:** Course Project for Randomized Algorithms
- **Traffic Sign Recognition using Convolutional Neural Networks:** Course Project for Computational Vision
- **Halite - Two Sigma:** A heuristic-based agent that uses clustering as well as multi-agent path finding techniques like Windowed Hierarchical Cooperative A* to perform well in the game environment