

Zhi (Bruce) Wen

✉ zhiwen120@gmail.com

🌐 zhi-wen.net

Last update: 01/02/2022

Profile Summary

- Applied research scientist at Mila with demonstrated expertise in Machine learning, Natural Language Processing (NLP), and their applications in healthcare.
- Theoretical and hands-on knowledge in machine learning, NLP, Graph Neural Networks, and Reinforcement Learning in research and industry settings.
- Great communicator and team player with peer-reviewed publications at high-impact venues such as Nature Communications and cross-discipline collaboration experience.

Skills

Technical Python, PyTorch, Hugging Face libraries, PyTorch Geometric, Pandas, Matplotlib, Bash
General Scientific and technical writing, research, project management
Languages English, Mandarin

Education

- 2019 – 2021 **McGill University - Computer Science, Montreal, Canada.**
MSc in Computer Science (Thesis), supervised by Prof. Yue Li. GPA 3.95/4.0.
- 2015 – 2019 **Wuhan University - Electronic Information Engineering, Wuhan, China.**
BEng in Electrical Engineering (Excellent Engineer Program)

Experience

Industry

- Aug. 2021 – **Applied research scientist, Mila – Quebec Artificial Intelligence Institute.**
- Participate in challenging applied machine learning research projects with companies, academic or research institutes.
 - Support companies through consulting / coaching to introduce AI in their products or improve their AI-based solutions.
- Oct. 2018 – **R&D Intern, Horizon Robotics, Algorithm Research and Development Team.**
- April 2019
- Designed and implemented a pipeline for large-scale evaluation of audio recordings' quality and data selection basing on audio and textual features.
 - Implemented a large-scale audio data processing platform with Hadoop and web-interface.

Academic and Research

- Sept. 2019 – **Graduate Researcher, McGill University.**
- Aug. 2021
- Partly sponsored by Mitacs Globalink Graduate Fellowship and conducted research in machine learning and Natural Language Processing in healthcare.
 - Participated as core members in several projects including predicting prolonged mechanical ventilation, large-scale public medical text dataset construction, and assessing COVID-19 public health responses from media news.
 - Published in Nature Communications, EMNLP 2020 Clinical NLP, ACM-BCB, PLOS ONE, etc.
- Sept. 2020 – **Teaching Assistant, McGill University, COMP 550 Natural Language Processing.**
- Dec. 2020
- Helped Prof. Jackie Cheung deliver the course to about 150 students through assignments grading, office hours, and final project grading.

- Nov. 2019 – **Participant**, *Neurips 2019 Reproducibility Challenge*.
- Dec. 2019
- Conducted extensive ablation study of *Controllable Unsupervised Text Attribute Transfer via Editing Entangled Latent Representation* to examine its claims' soundness and its model's robustness.
 - Project report available here: <https://openreview.net/forum?id=XibbOqrT4q>.
- July 2018 – **Research Intern**, *University of Toronto*.
- Oct. 2018
- Sponsored by Mitacs Globalink Research Internship program as research student for Dr. Andrea Kassner of University of Toronto and The Hospital for Sick Children.
 - Supported multiple projects by analyzing and visualizing fMRI data, including BOLD signal and DKI.
 - Constructed a pipeline in MatLab for experimental data processing and analysis which continued to be used at the end of the internship.
- April 2017 – **Undergraduate Researcher**, *Wuhan University*.
- Nov. 2018
- Sponsored by the National Innovation and Entrepreneurship Training Program of Ministry of Education of China.
 - Designed and implemented the core algorithm for elevator's movements anomaly detection based on Kalman filter.
 - Developed a comprehensive elevator monitoring system for both anomaly detection and usage pattern analysis.

Selected Publications

- 2022 **Inferring global-scale temporal latent topics from news reports to predict public health interventions for COVID-19**, *Patterns*, **Zhi Wen**, Guido Powell, Imane Chafi, David L Buckeridge, Yue Li.
- 2021 **Mining heterogeneous clinical notes by multi-modal latent topic model**, *Plos ONE*, **Zhi Wen**, Pratheeksha Nair, Chih-Ying Deng, Xing Han Lu, Edward Moseley, Naomi George, Charlotta Lindvall, Yue Li.
- 2020 **MeDAL: Medical Abbreviation Disambiguation Dataset for Natural Language Understanding Pretraining**, *EMNLP 2020 Clinical NLP Workshop*, **Zhi Wen**, Xing Han Lu, Siva Reddy.
Code and data at <https://github.com/BruceWen120/medal>.
- 2020 **Global Surveillance of COVID-19 by mining news media using a multi-source dynamic embedded topic model**, *11th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, Yue Li, Pratheeksha Nair, **Zhi Wen**, Imane Chafi, Anya Okhmatovskaia, Guido Powell, David Buckeridge.
- 2020 **Inferring Multimodal Latent Topics from Electronic Health Records**, *Nature Communications*, Yue Li, Pratheeksha Nair, Xing Han Lu, **Zhi Wen**, Yuening Wang, Amir Ardalan Kalantari Dehaghi, Yan Miao, Weiqi Liu, Tamas Ordog, Joanna M. Biernacka, Euijung Ryu, Janet E. Olson, Mark A. Frye, Aihua Liu, Liming Guo, Ariane Marelli, Yuri Ahuja, Jose Davila-Velderrain, Manolis Kellis.

Awards

- 2019 – 2020 **Mitacs Globalink Graduate Fellowship**
- 2018 **Wuhan University Outstanding Student**
- 2017 **Interdisciplinary Contest in Modeling – Honorable Mention**
- 2016 **China College Students' Entrepreneurship Competition – Golden Award** *highest award*