

# Wei-Rui Chen

wrchen012@gmail.com | 

## HIGHLIGHTS

---

- 10+ years of ML, 5+ years of NLP. Published in top-tier venues. Experienced in training and evaluating LLMs/VLMs. Currently focusing on multimodal and multilingual AI, RAG, and Agentic systems.
- **Programming:** Python, Bash, R, C++, C
- **NLP:** Huggingface, PyTorch, Fairseq, LangGraph
- **Tools:** Linux, Clouds, HPC, SQL, Docker, Spark

## EDUCATION

---

- PhD, Computational Linguistics**, University of British Columbia (UBC), Canada Sep 2020 - Feb 2026  
Multimodal and Multilingual AI, Large Language Models, Cross-lingual Transfer Learning
- MSc, Computer Science**, National Chengchi University (NCCU), Taiwan Sep 2016 - Jul 2018  
Distinguished Academic Award; GPA 4.21/4.30

## WORK EXPERIENCE

---

- Graduate Research Assistant** Sep 2020 - Present  
Deep Learning and NLP Lab, UBC NLP Group, UBC Advisor: Dr. Muhammad Abdul-Mageed

- Enhancing the understanding and reasoning capabilities of vision-language models (VLMs) across video, image, and text modalities, as well as their text generation abilities by leveraging agentic workflows and Retrieval-Augmented Generation (RAG). Hands-on Experience with LLaVA, Llama3.2, MiniCPM, and more.
- Specialized in multilingual NLP with a focus on cross-lingual transfer, machine translation (MT), and large language models. Developed robust, efficient automated pipelines for training 5000+ models in multi-GPU HPC environments. Possess solid knowledge of Unicode and tokenization, with experience processing data from 600+ languages. Achieved state-of-the-art results in several MT tasks.

- Machine Learning Engineer** Aug 2018 - Sep 2020  
Trend Micro
- Led an international team of 12 engineers in the research and design of graph-based anomaly detection systems.
  - Designed and implemented robust time-series anomaly detection systems deployed across the globe.

## PUBLICATIONS

---

- Interplay of Machine Translation, Diacritics, and Diacritization, NAACL 2024** [Link](#)  
[WR Chen](#), I Adebara, M Abdul-Mageed

- Fumbling in Babel:**  
**An Investigation into ChatGPT's Language Identification Ability, NAACL 2024** [Link](#)  
[WR Chen](#), Ife Adebara, Khai Duy Doan, Qisheng Liao, M Abdul-Mageed

- Findings of the Second AmericasNLP Competition**  
**on Speech-to-Text Translation (the best model), NeurIPS 2022 Competition Track** [Link](#)  
[WR Chen](#), P Sullivan, I Adebara, B Talafha, AA Inciarte, M Abdul-Mageed  
(skipping organizers and other teams)

- Improving Neural Machine Translation of Indigenous Languages**  
**with Multilingual Transfer Learning, LoResMT at EACL 2023** [Link](#)  
[WR Chen](#), M Abdul-Mageed

- Machine Translation of Low-Resource Indo-European Languages, WMT 2021** [Link](#)  
[WR Chen](#), M Abdul-Mageed

- IndT5: A Text-to-Text Transformer for 10 Indigenous Languages, AmericasNLP 2021** [Link](#)  
EMB Nagoudi, [WR Chen](#), M Abdul-Mageed, H Cavusoglu

- NLP Methods for Information Extraction from Biographies: An Exploration**  
**with the Elite Biographies in the Extended Taipei Gazetteers, DADH 2019** [Link](#)  
YS Wang, [WR Chen](#), CL Liu