

Sang Yun Kwon

PH.D. STUDENT · SCHOOL OF INFORMATION · UNIVERSITY OF BRITISH COLUMBIA (UBC)

2075, West Mall, Vancouver, BC, Canada V6T 1Z2

☎ (+1) 604-773-6305 | ✉ skwon01@mail.ubc.ca | 🌐 skwon01-ubc.github.io/website/ | 🔗 LinkedIn | 📄 Google Scholar

Research Interests

My research interest lie in **AI, NLP, Human AI interaction**, aiming to build AI systems that can interpret nuance, understand human behaviour, and align with diverse human requirements.

- **Human AI Collaboration:** Developing NLP systems that enable interactions between AI systems and human users. Focusing on handling implicature and ambiguity, understanding human behavior, and aligning with the diverse requirements humans have from technology. Actively designing collaboration frameworks that allow users from diverse backgrounds to co-create and iteratively refine AI-driven processes particularly in the context of education and writing assistance.
- **Inclusive NLP Models for Low-Resource Languages:** Designing NLP models that promote greater inclusivity, particularly for underrepresented communities that cater to a wide range of cultures and linguistic needs.

Education

University of British Columbia (UBC)

PH.D. IN INFORMATION

- Advisor: **Prof. Muhammad Abdul-Mageed**
- GPA: 4.00/4.00

Vancouver, BC, Canada

Sep. 2024 - Nov. 2028 (Expected)

University of British Columbia (UBC)

M.A. LIBRARY AND INFORMATION STUDIES

- Advisor: **Prof. Muhammad Abdul-Mageed**
- GPA: 4.00/4.00

Vancouver, BC, Canada

Sep. 2021 - Aug. 2024

Yonsei University (YU)

B.A. INTERNATIONAL COMMERCE

- GPA: 3.30/4.00

Seoul, South Korea

Match. 2015 - Aug. 2020

Publications

CONFERENCE PUBLICATIONS

- [C.4] AbdelRahim Elmadany, **Sang Yun Kwon**^{*}, FaHawau Olamide Toyin, Alcides Alcoba Inciarte, Hanan Aldarmaki, and Muhammad Abdul-Mageed. “**Voice of a Continent: Mapping Africa’s Speech Technology Frontier**”. *The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2025.
- [C.3] Samar Magdy^{*}, **Sang Yun Kwon**^{*}, Fakhraddin Alwajih, Safaa Taher Abdelfadil, Shady Shehata, and Muhammad Abdul-Mageed. “**JAWA-HER: A Multidialectal Dataset of Arabic Proverbs for LLM Benchmarking**”. *Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*. Nov. 2025.
- [C.2] Samar Magdy, **Sang Yun Kwon**, Fakhraddin Alwajih, Reem Abdel-Salam, and Muhammad Abdul-Mageed. “**Gazelle: An Instruction Dataset for Arabic Writing Assistance**”. *The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Aug. 2024. [\[Arxiv\]](#)
- [C.1] **Sang Yun Kwon**, Gagan Bhatia, El Moatez Billah Nagoudi, and Muhammad Abdul-Mageed. “**Beyond English: Evaluating LLMs for Arabic Grammatical Error Correction**”. *Arabic Natural Language Processing Conference (ArabicNLP)*. Aug. 2023. [\[Arxiv\]](#)

WORKSHOP PUBLICATIONS AND POSTERS

- [W.1] **Sang Yun Kwon**, Gagan Bhatia, El Moatez Billah Nagoudi, Alcides Alcoba Inciarte, and Muhammad Abdul-Mageed. “**SIDLR: Slot and Intent Detection Models for Low-Resource Language Varieties**”. *Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial’23)*. May. 2023. [\[Arxiv\]](#)

Honors and Awards

2021-2023 Graduate Scholarship of School of Information
2023 Masters Travel Award of School of Information, University of British Columbia (UBC)
2018-2019 Yonsei University Academic Honors (Top 25% in Class)

Canada
Canada
South Korea

Experience

Master of Data Science in Computational Linguistics

Vancouver, BC, Canada

TEACHING ASSISTANT

Sept. 2025 - Present

- DSCI-511 Python Programming

UBC Deep Learning and Natural Language Processing Group, University of British Columbia (UBC)

Vancouver, BC, Canada

GRADUATE RESEARCH ASSISTANT

June. 2022 - Present

- Advisor: Prof. Muhammad Abdul-Mageed
- Research in NLP, Grammatical Error Correction, Text generation, and Language modeling.
- Creative text generation and collaborative writing.
- Developed a state-of-the-art (SoTA) Grammatical Error Correction model specifically tailored for the Arabic language.
- Engaged in language modeling and dataset creation focused on Slot and Intent Detection for Low-resource Languages.
- Publishing papers in top-tier NLP venues and delivering presentations.

Talks

Apr. 2025 **Guest Lecture: LLM Agents**, UBC, Master of Data Science in Computational Linguistics

Vancouver

Dec. 2023 **Beyond English: Evaluating LLMs for Arabic Grammatical Error Correction**, at ArabicNLP'23

Singapore

Service

Aug. 2024 **Reviewer**, at ArabicNLP'25

Suzhou, China

Aug. 2024 **Reviewer**, at ArabicNLP'24

Bangkok, Thailand

Dec. 2023 **Reviewer**, at ArabicNLP'23

Singapore

Skills

Programming Languages | Tools R, Python, Pytorch, DSPy, LangChain, CrewAI

Relevant Coursework

- Applied Machine Learning, UBC
- Designing Interactive Computational Technology for People, UBC
- Database Design, UBC
- Machine Learning and Data Mining, UBC
- Topics in HCI, UBC
- Python Programming, UBC