

Grace (Sungjoo) Byun

122 West Trinity Place no.1327, Decatur, Georgia, United States, 30030

+1 (404) 406-4823 / gbyun@emory.edu / [LinkedIn](#)

Education

- Emory University**, Georgia, United States Sep 2024 - Present
- ❖ PhD in Computer Science and Informatics
 - ❖ Advisor: Prof. Jinho D. Choi | [EmoryNLP Lab](#)
- Seoul National University**, Seoul, South Korea Sep 2022 – Aug 2024
- ❖ Master's in computational Linguistics *GPA 4.2 / 4.3*
 - ❖ *Dissertation: Simplifying Large Language Model Alignment and Detoxification: Comprehensive Instruction and Preference Data Solutions*
- Korea University**, Seoul, South Korea Mar 2018 – Aug 2022
- ❖ Bachelor of Arts in Linguistics *GPA 4.35 / 4.5 (Linguistics 4.5 / 4.5)*

Work Experience

- Seoul National University**, Seoul, South Korea Mar 2023 – Present
- Humanities Data Science (Interdisciplinary major) Administrative Assistant*
- ❖ Led department events, managing logistics for alumni gatherings, and workshops.
- Kakao Enterprise**, Gyeonggi-do, South Korea Oct 2021 – Jan 2022
- Knowledge Graph Base TF Team*
- ❖ Determined how knowledge graph should be modeled by meticulously analyzing classes and properties within wiki data, which are integral to semantic web technologies and data modeling.
 - ❖ Involved in decision-making process regarding ontology or schema used in knowledge graph.
- Knowledge Graph AI Team* Apr 2021 – Oct 2021
- ❖ Constructed RDF triples for a knowledge graph and analyzed using SQL to organize the collected information.

Teaching Experience

- Emory University**, Georgia, United States Jan 2025 – Present
- TA of Computational Linguistics course (CS329)*
- Seoul National University**, Seoul, South Korea Nov 2023 – Aug 2024
- Special Lecture for Data Science Humanities major*
- ❖ Delivered special lectures on Large Language Model (LLM) training methods and practical usage of the ChatGPT API to 20 undergraduate students majoring in Data Science Humanities.
- Application of Computing Teaching Assistant* Sep 2023 – Feb 2024
- ❖ Earned a distinguished academic scholarship for assisting “Basic of Natural Language Processing” course.
- Humanities Data Science Teaching Assistant* Sep 2022 – Feb 2023
- ❖ TA of “Humanities Data Science Practice” course, grading assignments and developing textbooks.

Extracurricular Activities

- Hyundai Crash Engineering Generative AI Project** Sep 2024 - Present
- Emory University*
- ❖ Developed a Generative AI Assistant to analyze crash test data and support engineering decision-making for vehicle safety improvements. (Project in collaboration with Hyundai Motor Company.)
- Korean Large Language Model (LLM) Project** Dec 2023 – Jul 2024
- Seoul National University*
- ❖ Developed multilingual LLM specialized in Korean, leveraging methods like Mixture of Experts ([MoE](#)), Tokenizer Expansion, Instruction Tuning, DPO, and ORPO.
 - ❖ Implemented [Korean instruction datasets](#) to develop multilingual LLM accessible to public (<https://dag.snu.ac.kr/>), which involves rigorous testing and parameter adjustments for optimal performance.
- Bank of Korea Financial Large Language Model (LLM) Project** Jul 2023 – Jan 2024

Seoul National University

- ❖ Project commissioned by the Bank of Korea, the central bank of South Korea.
- ❖ Developed domain-specific LLM using provided financial raw data.

Medical and Legal Big Data Project

Mar 2022 – Jun 2023

Team Lead at Korea University and Seoul National University

- ❖ Orchestrated large-scale project involving compilation of medical and legal resources spanning a decade.
- ❖ Participated in meetings with professors from KU Medical and Law School.

Korea University Computer Club (KUCC)

Mar 2021 – Aug 2021

Mentor

- ❖ Provided Python training to a group of 10 students, meeting once a week throughout the semester.

Korea University Cyber Law and Policies Society (CLPS)

Jul 2020 – Jan 2022

- ❖ Engaged in educational activities related to cyber law and policies with a 40-member student society at KU.

Volunteer Experience

Korea University International Student Assistant (KUISA)

Jul 2019 – Jun 2020

- ❖ Conducted weekly meetings to brainstorm volunteer events, improving international student experiences.

Presentations and Publications

❖ Published Works

Seo, J., Park, M., Byun, S., et al. (2024). *The Manchu milestone: Wav2vec 2.0 and the future of minority language speech recognition*. Proceedings of the EMNLP 2024 Field Matters Workshop.

<https://aclanthology.org/2024.fieldmatters-1.2/>

Byun, S., et al. (2024). *Korean Bio-Medical Corpus (KBMC) for named entity recognition with span representation analysis*. Proceedings of LREC-COLING 2024. <https://aclanthology.org/2024.lrec-main.868/>

Jang, D., Byun, S., Jo, H., & Shin, H. (2024). *KIT-19: A comprehensive Korean instruction toolkit on 19 tasks for fine-tuning Korean large language models*. Proceedings of LREC-COLING 2024. <https://aclanthology.org/2024.lrec-main.853/>

Jang, D., Byun, S., & Shin, H. (2024). *A study on how attention scores in the BERT model are aware of lexical categories in syntactic and semantic tasks on the GLUE benchmark*. Proceedings of LREC-COLING 2024. <https://aclanthology.org/2024.lrec-main.148/>

Lee, S., Byun, S., et al. (2024). *ManNER & ManPOS: Pioneering NLP for endangered Manchu language*. Proceedings of LREC-COLING 2024. <https://aclanthology.org/2024.lrec-main.961/>

Byun, S., & Shin, H. (2023). *Automatic construction of a Korean toxic instruction dataset for ethical tuning of large language models*. NeurIPS 2023 Instruction Workshop (Awarded: DEI Scholarship). <https://arxiv.org/abs/2311.18215>

Seo, J., Byun, S., Kang, M., & Lee, S. (2023). *Mergen: The first Manchu-Korean machine translation model trained on augmented data*. Proceedings of the EMNLP 2023 MRL Workshop. <https://aclanthology.org/2023.mrl-1.10/>

❖ Oral Presentation

Byun, S. (2023). *Phonetic Comparison Analysis of Spanish Speech Production by Korean Spanish Learners and Native Speakers & L2 Sound Pedagogy*. ICKL.

❖ Under Review / Preprint

Byun, G., Lee, S., Choi, N., & Choi, J. (2025). *Secure multifaceted-RAG for enterprise: Hybrid knowledge retrieval with security filtering*. Manuscript under review.

Byun, G., & Choi, J. (2025). *D-GEN: Efficient distractor generation and evaluation for model benchmarking*. Manuscript under review.

Choi, N., Byun, G., Chung, A., et al. (2025). *Trustworthy answers, messier data: Bridging the gap in low-resource retrieval-augmented generation for domain expert systems*. [arXiv:2502.19596](https://arxiv.org/abs/2502.19596)

Jang, D., Seo, J., Byun, S., et al. (2023). *CARBD-Ko: A contextually annotated review benchmark dataset for aspect-level sentiment classification in Korean*. [arXiv:2402.15046](https://arxiv.org/abs/2402.15046)

Jang, D., Lee, S., Byun, S., et al. (2023). *DaG LLM ver 1.0: Pioneering instruction-tuned language modeling for Korean NLP*. [arXiv:2311.13784](https://arxiv.org/abs/2311.13784)

Awards and Certifications

Academic Excellence Award

Mar 2018 – Dec 2021

- ❖ Obtaining a GPA of 4.0+/4.5 in six out of eight semesters at Korea University.

Dean's List

Mar 2020

- ❖ Received academic recognition for having a perfect 4.5/4.5 GPA at Korea University.

Scholarship

- ❖ ICKL 2023 Travel Award: \$ 700
- ❖ NeurIPS DEI Scholarship (Sponsored by Ubiquant, SambaNovaSystems, Apple): \$ 475

Skills

- ❖ Languages: Korean (native), English (native), and Spanish (intermediate)
- ❖ Technical: Python, R, SQL, SPARQL, Praat, LaTeX