




HAEUN YU

PhD student ~ NLP

 haeunyu.github.io

 hayu@di.ku.dk

 +45 55 26 29 20

 /in/haeunyu

 Copenhagen, Denmark

ABOUT ME

I am a second year Ph.D. student at NLP Section, Department of Computer Science, University of Copenhagen. I work on explainability focusing on language models' factuality.

RESEARCH INTERESTS

Explainability of Language Models

Mechanistic Interpretability

Factuality

Knowledge Conflict, Parametric Knowledge Elucidation, Knowledge Editing/Unlearning

EDUCATION

- 9/2023 - 8/2026 **Ph.D. in Computer Science** University of Copenhagen, Denmark
Supervised by Prof. Isabelle Augenstein and Prof. Pepa Atanasova
Work on 'Explainable and Robust Automatic Fact Checking (ExplainYourself)' project
- 3/2021 - 2/2023 **M.Sc. in Computer Science and Engineering** Sungkyunkwan University, South Korea
Supervised by Prof. Youngjoong Ko
Thesis: Dialogue State Tracking System with Graph-structured Discourse Information
Honors: Korea Telecom (KT) AI Scholarship
GPA: 4.39/4.5
- 3/2017 - 2/2021 **B.A. in Global Korean Studies & B.Sc. Bigdata Science** Sogang University, South Korea
Honors: Cum Laude
GPA: 3.6/4.3

EXPERIENCE

- 2/2023 - 8/2023 **Assistant Research Engineer** Korea Telecom (KT) AI2XL Lab.
Developed Question Answering System with passage retriever.
Developed a Korean knowledge grounded conversation dataset in financial domain.
LLMs / Retrieval-Augmented Generation / Information Retrieval

PUBLICATIONS

Conferences

- [1] **Haeun Yu**, P. Atanasova, and I. Augenstein. "Revealing the Parametric Knowledge of Language Models: A Unified Framework for Attribution Methods". In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by L.-W. Ku, A. Martins, and V. Srikumar. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 8173–8186. URL: <https://aclanthology.org/2024.acl-long.444>.
- [2] C. Park, E. Ha, Y. Jeong, C. Kim, **Haeun Yu**, and J. Sung. "CopyT5: Copy Mechanism and Post-Trained T5 for Speech-Aware Dialogue State Tracking System". In: *Proceedings of The Eleventh Dialog System Technology Challenge*. Ed. by Y.-N. Chen, P. Crook, M. Galley, S. Ghazarian, C. Gunasekara, R. Gupta, B. Hedayatnia, S. Kottur, S. Moon, and C. Zhang. Prague, Czech Republic: Association for Computational Linguistics, Sept. 2023, pp. 89–94. URL: <https://aclanthology.org/2023.dstc-1.11>.
- [3] S. Park, K. Choi, **Haeun Yu**, and Y. Ko. "Never Too Late to Learn: Regularizing Gender Bias in Coreference Resolution". In: *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, WSDM 2023, Singapore, 27 February 2023 - 3 March 2023*. Ed. by T. Chua, H. W. Lauw, L. Si, E. Terzi, and P. Tsaparas. ACM, 2023, pp. 15–23. URL: <https://doi.org/10.1145/3539597.3570473>.
- [4] **Haeun Yu**, T. Hong, and Y. Ko. "Adapting Pre-trained Language Model for Dialogue State Tracking on Spoken Conversations". In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2022. URL: https://github.com/shanemoon/dstc10/raw/main/papers/dstc10_aai22_track2_25.pdf.
- [5] B. Kim, H. Choi, **Haeun Yu**, and Y. Ko. "Query Reformulation for Descriptive Queries of Jargon Words Using a Knowledge Graph based on a Dictionary". In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2021. URL: <https://doi.org/10.1145/3459637.3482382>.

Journals

- [6] **Haeun Yu** and Y. Ko. "Enriching the dialogue state tracking model with a syntactic discourse graph". In: *Pattern Recognition Letters* 169 (2023), pp. 81–86. ISSN: 0167-8655. URL: <https://www.sciencedirect.com/science/article/pii/S0167865523000958>.

- [7] B. Kim, H. Choi, **Haeun Yu**, and Y. Ko. "Graph-based query reformulation system for descriptive queries of jargon words using definitions". In: *Expert Systems with Applications* 214 (2023), p. 119149. ISSN: 0957-4174. URL: <https://www.sciencedirect.com/science/article/pii/S0957417422021674>.
- [8] T. Hong, J. Cho, **Haeun Yu**, Y. Ko, and J. Seo. "Knowledge-grounded dialogue modelling with dialogue-state tracking, domain tracking, and entity extraction". In: *Comput. Speech Lang.* 78 (2023), p. 101460. URL: <https://www.sciencedirect.com/science/article/pii/S0885230822000833>.
- [9] **Haeun Yu** and Y. Ko. "Entity Graph Based Dialogue State Tracking Model with Data Collection and Augmentation for Spoken Conversation". In: *Journal of KIISE* 49.10 (Oct. 2022), pp. 891–897. ISSN: 2383-6296. URL: <http://dx.doi.org/10.5626/JOK.2022.49.10.891>.

Preprints

- [10] S. V. Marjanović*, **Haeun Yu***, P. Atanasova, M. Maistro, C. Lioma, and I. Augenstein. *From Internal Conflict to Contextual Adaptation of Language Models*. 2024. arXiv: 2407.17023 [cs.CL]. URL: <https://arxiv.org/abs/2407.17023>.

ADDITIONAL ACADEMIC ACTIVITIES

2022	<p>Participation in Dialogue System Track Challenge 11 (DSTC11) Track 3. Augmented the dialogue using new ontology crawled from wikipedia to train the model with various named entities.</p> <p>Teaching Assistant of Data Structure & Algorithms at Sungkyunkwan University</p>
2021	<p>Participation in Dialogue System Track Challenge 10 (DSTC10) Track 2 Task 1. Achieved rank 5th out of 11 entries in Track 2 Task 1 Dialogue State Tracking (Team A09)</p> <p>Teaching Assistant of Natural Language Processing at Sungkyunkwan University</p>
2019	<p>Exchange student at Umeå University, Umeå, Sweden</p>

SKILLS

- Extensive programming experience with Python
- Experience working with NLP, machine learning algorithms and tools, including HuggingFace, PyTorch, scikit-learn, NLTK, deep graph library and spaCy
- Goal-oriented, organized, excellent at prioritization