Ji-Ung Lee

ji-ung.lee@tu-darmstadt.de Darmstadt, Germany

Education

PhD Student Darmstadt, Germany
Ubiquitous Knowledge Processing (UKP) Lab, Technische Universität Darmstadt *Sep. 2017 – present*
Thesis: Constrained Generation and Adaptive Selection of C-Tests (defense planned for Jul. 2024)

M.Sc. Computer Science Darmstadt, Germany
Technische Universität Darmstadt, Cumulative grade 1.69 *Oct. 2013 – Apr. 2017*
Thesis: Automatic Annotation of Argument Components (Grade 1.3)

B.Sc. Computer Science Darmstadt, Germany
Technische Universität Darmstadt, Cumulative grade 2.07 *Oct. 2010 – Oct. 2013*
Thesis: Transductive Pairwise Classification (Grade 1.0)

Experience

Intel Collaboration, *UKP Lab, TUDa* *Jun. 2022 – present*
ELM Center for Efficiency in Very Large AI Models: Increasing fine-tuning Efficiency

Athene (AVSV), *UKP Lab, TUDa* *Feb. 2023 – Dec. 2023*
Code Transformers and Knowledge Graphs for Vulnerability Detection

Athene (SenPAI), *UKP Lab, TUDa* *Jan. 2022 – Dec. 2022*
Adversarial Attacks on NLP systems and robust training [3]

TexPrax, *UKP Lab, TUDa* *Feb. 2020 – Dec. 2022*
- Ethical and responsible data collection in a production environment [4, 7]
- Demonstrator on GitHub
- Evaluated with the highest grade

Evidence, *UKP Lab, TUDa* *Jan. 2020 – present*
- Investigating the interactive extraction of good and diverse dictionary examples from large corpora
- Sampling strategies for jointly optimizing system and user demands [9]

a! – Automated Language Instruction, *UKP Lab, TUDa* *Sep. 2017 – Apr. 2019*
- Adaptive selection of language learning exercises [9]
- Automated difficulty manipulation of language learning exercises [10]

Freelancing Software Developer *Jul. 2016 – Jun. 2017*
- Developed various Python web crawlers for a startup company

Student Research Assistant, *Department of Psychology, TUDa* *Nov. 2014 – Jul. 2015*
- Developed a robot arm providing haptic feedback in an interdisciplinary project
- Assisted in a follow-up user study for investigating the perception shift with haptic feedback

Student Research Assistant, *CASED, TUDa* *Apr. 2014 – Jul. 2014*
- Developed a sand-boxed web application for teaching basic hacking skills

Teaching

Project Coordinator, *UKP Lab, TUDa* *Oct. 2020 – Mar. 2021*
Organized and coordinated 54 software projects for undergraduate students

Project Coordinator, *UKP Lab, TUDa* *Oct. 2019 – Mar. 2020*
Organized and coordinated 51 software projects for undergraduate students

Lecturer, *UKP Lab, TUDa* *Apr. 2018 – Oct. 2018*
Text analytics course on active learning for undergraduate and graduate students

Student Thesis Supervision, *UKP Lab, TUDa* *Recurring Activity*
Igor Cherepanov (M.Sc.), Jonathan Gruhle (B.Sc.), Tom Halecker (M.Sc.),
Johanna Heinz (M.Sc.), Thorsten Hollstein (M.Sc.), Erik Schwan (B.Sc.),
Darjush Siadohoni (M.Sc.), Hanna Sterz (B.Sc.), Marian Thull (M.Sc.)

Independent Projects

Darmstadt kocht! *Co-organizer & Software developer* *Jan. 2015 – present*

- Organization of a (bi)annual and free of charge city-wide dinner event with up to 350 participants
- Development and maintenance of software for optimizing paths and matching cooking partners

Research Areas

Active learning [9], computer-assisted language learning [8, 10], efficient data annotation [4, 5, 6, 7, 2], efficient NLP [1, 14], transfer learning [11, 12, 3]

Skills

- Programming languages: Proficient in Python and Java, Familiar with JavaScript and SQL, Basic experience with PhP, Matlab, and C++
- AI-related: Proficient in PyTorch, huggingface, and numpy (and other commonly-used frameworks). Proficient in high-performance computing (SLURM) and multi-GPU experiments (PyTorch FSDP, Ray)
- Other technologies & IDEs: Git, Linux, Docker, PyCharm, VisualStudio, IntelliJ, Eclipse
- Proficient languages: English, German, Korean

Community Engagement

7.1 Reviewing Activities

2024 AAAI, EACL$_{\texttt{ARR}}$ (great reviewer), LREC, NAACL$_{\texttt{ARR}}$, ACL$_{\texttt{ARR}}$, COLM, NeurIPS Workshops
2023 AAAI, ACL, EMNLP, EACL, BEA, SustaiNLP (outstanding reviewer), Eval4NLP, NeurIPS Workshops
2022 ACL$_{\texttt{ARR}}$, EMNLP, COLING, BEA, NLP4PI
2021 ACL, EMNLP, NAACL, EACL (outstanding reviewer), SustaiNLP, MRL
2020 ACL (outstanding reviewer), EMNLP, BEA, SustaiNLP
2019 ACL, BEA

Since 10/2021: Registered as a reviewer at `ARR`

7.2 Workshop Organization

- InterNLP 2022 co-located with NeurIPS 2022: Yoav Artzi, Kianté Brantley, Soham Dan, Khanh Nguyen, **Ji-Ung Lee**, Edwin Simpson, Alane Suhr
- InterNLP 2021 co-located with ACL 2021: Kianté Brantley, Soham Dan, Iryna Gurevych, **Ji-Ung Lee**, Filip Radlinski, Hinrich Schütze, Edwin Simpson, Lili Yu

7.3 Talks

- Invited talk: Facets of efficiency in NLP., KUIS AI, Koç University. September, 2022.
- Spotlight talk: Investigating rational activation functions to train Transformer models. Dagstuhl Seminar on *Efficient and Equitable Natural Language Processing in the Age of Deep Learning*. June, 2022.

Publications

Accepted publications (latest first, * marks equal contribution works.)

[1] Marcos Treviso*, **Ji-Ung Lee***, Tianchu Ji*, Betty van Aken, Qingqing Cao, Manuel R. Ciosici, Michael Hassid, Kenneth Heafield, Sara Hooker, Colin Raffel, Pedro H. Martins, André F. T. Martins, Jessica Zosa Forde, Peter Milder, Edwin Simpson, Noam Slonim, Jesse Dodge, Emma Strubell, Niranjan Balasubramanian, Leon Derczynski, Iryna Gurevych, Roy Schwartz, Efficient Methods for Natural Language Processing: A Survey, *Transactions of the Association for Computational Linguistics*, 11: pages 826–860. July, 2023. MIT Press

[2] Jan-Christoph Klie, **Ji-Ung Lee**, Kevin Stowe, Gözde Gül Şahin, Nafise Sadat Moosavi, Luke Bates, Dominic Petrak, Richard Eckart De Castilho, Iryna Gurevych, Lessons Learned from a Citizen Science Project for Natural Language Processing, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3594–3608, May, 2023. Dubrovnik, Croatia

[3] Haishuo Fang, **Ji-Ung Lee**, Nafise Sadat Moosavi, and Iryna Gurevych, Transformers with Learnable Activation Functions, *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2382–2398, May, 2023. Dubrovnik, Croatia

[4] Lorenz Stangier*, **Ji-Ung Lee***, Yuxi Wang, Marvin Müller, Nicholas Frick, Joachim Metternich, and Iryna Gurevych, TexPrax: A Messaging Application for Ethical, Real-time Data Collection and Annotation, *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing: System Demonstrations* (AACL), pages 9–16, November, 2022. Taipei, Taiwan.

[5] **Ji-Ung Lee***, Jan-Christoph Klie*, and Iryna Gurevych, Annotation Curricula to Implicitly Train Non-Expert Annotators, *Computational Linguistics*, Volume 48 (2), pages 343–373, June, 2022. MIT Press

[6] Tilman Beck, **Ji-Ung Lee**, Christina Viehmann, Marcus Maurer, Oliver Quiring, and Iryna Gurevych, Investigating label suggestions for opinion mining in German Covid-19 social media, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (ACL), pages 1–13, July, 2021. Online.

[7] Marvin Müller, **Ji-Ung Lee**, Nicholas Frick, Lorenz Stangier, Iryna Gurevych, and Joachim Metternich, Extracting problem related entities from production chats to enhance the data base for assistance functions on the shop floor, *9th CIRP Global Web Conference – Sustainable, resilient, and agile manufacturing and service operations : Lessons from COVID-19* (Procedia CIRP), Volume 103, pages 231–236, October, 2021. Online.

[8] Marianne Grace Araneta, Gülşen Eryiğit, Alexander König, **Ji-Ung Lee**, Ana Luís, Verena Lyding, Lionel Nicolas, Christos Rodosthenous, and Federico Sangati, Substituto–A Synchronous Educational Language Game for Simultaneous Teaching and Crowdsourcing, *In Proceedings of the 9th Workshop on NLP for Computer Assisted Language Learning* (NLP4CALL) pages 1–9, November, 2020. Online.

[9] **Ji-Ung Lee**, Christian M. Meyer, and Iryna Gurevych Empowering Active Learning to Jointly Optimize System and User Demands, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (ACL), pages 4233–4247, July, 2020. Online

[10] **Ji-Ung Lee**, Erik Schwan, and Christian M. Meyer, Manipulating the Difficulty of C-Tests, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*(ACL),pages 360–370, July, 2019. Florence, Italy

[11] Steffen Eger, Gözde Gül Şahin, Andreas Rücklé, **Ji-Ung Lee**, Claudia Schulz, Mohsen Mesgar, Krishnkant Swarnkar, Edwin Simpson, and Iryna Gurevych, Text Processing Like Humans Do: Visually Attacking and Shielding NLP Systems, *Annual Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL-HLT), pages 1634–1647, June, 2019. Minneapolis, USA

[12] **Ji-Ung Lee**, Steffen Eger, Johannes Daxenberger, and Iryna Gurevych, UKP TU-DA at GermEval 2017: Deep Learning for Aspect Based Sentiment Detection, *Proceedings of the GermEval 2017 - Shared Task on Aspect-based Sentiment in Social Media Customer Feedback*, September, 2017. Berlin, Germany

Preprints (latest first, * marks equal contribution works.)

[13] **Ji-Ung Lee**, Marc Pfetsch, Iryna Gurevych. Constrainted C-Test Generation using Mixed-Integer Programming. *arXiv:2404.08821*. 2024. (under review)

[14] **Ji-Ung Lee**, Haritz Puerto, Betty van Aken, Yuki Arase, Jessica Zosa Forde, Leon Derczynski, Andreas Rücklé, Iryna Gurevych, Roy Schwartz, Emma Strubell, Jesse Dodge Surveying (Dis)Parities and Concerns of Compute Hungry NLP Research *arXiv preprint arXiv:2306.16900*, 2023.

[15] Ulf A Hamster, **Ji-Ung Lee**, Alexander Geyken, Iryna Gurevych Rediscovering Hashed Random Projections for Efficient Quantization of Contextualized Sentence Embeddings *arXiv preprint arXiv:2304.02481*, 2023.