Last update - June 28, 2024

MyeongJun Erik Jang

Oxfordshire, United Kingdom

(+82)1040690669/(+44)7535129502 • myeongjun.jang@cs.ox.ac.uk
mj-jang.github.io

Expertise: Natural Language Processing, Computational Linguistics, Trustworthy AI

Research Interest

- Developing trustworthy and consistent language models
- Application of NLP techniques to industry

Education

University of Oxford — **Oxford, United Kingdom**
D.Phil. in Computer Science — *Oct. 2020 – Present (Thesis submitted)*

- Specialisation: natural language processing, machine learning, and deep learning
- Supervisor: Thomas Lukasiewicz

Korea University — **Seoul, Republic of Korea**
M.Eng in Industrial Management Engineering — *Mar. 2017 – Feb.2019*

- Specialisation: natural language processing, machine learning, and deep learning
- Supervisor: Pilsung Kang
- GPA: 4.31 / 4.5

Korea University — **Seoul, Republic of Korea**
B.S in Industrial Management Engineering — *Mar. 2013 – Feb.2017*

- GPA: 3.95 / 4.5

Work Experience

London NLP team, Machine Learning CoE — **J.P. Morgan Chase, United Kingdom**
AI & Data Science Associate, Ph.D. Intern — Sep. 2023 – Mar. 2024

1. Automatic Data Drift Detection
- Responsible for implementing automatic data drift detection tool for text data.
2. FAQ based on Document Retrieval
- Responsible for implementing FAQ model using query-to-query retrieval.

Data Science Lab — **Clifford Chance, United Kingdom**
PhD NLP Scientist, Alan Turing Internship Network — Feb. 2023 – Aug. 2023

1. AI-Assisted Due Diligence
- Responsible for applying neural NLP techniques and LLMs for a due diligence task for M&A.

AI Language Tech Labs — **SK Telecom, Republic of Korea**
NLP Scientist & Engineer — Nov. 2019 – Sep. 2020

1. Developing T-World Communication bot for supporting customer service
- Responsible for developing and deploying NLU and small-talk engines.
- Constructed a dialogue policy maker.
- Connected user-info API to an NLG system.
- The public beta service is launched on 27th Aug 2020

Data Machine Intelligence Group — **SK Telecom, Republic of Korea**

NLP & Machine Learning Engineer Apr. 2019 – Oct. 2019

1. Data Annotation Framework System for Human-involved Active Learning
   - Responsible for building a system for training and deploying semi-labeling ML models and retraining after human annotation.
   - Apr. 2019 - Aug. 2019
2. T-direct Communication bot
   - Responsible for developing and deploying an NLU engine.
   - Apr. 2019 - Oct. 2019

Research Papers

1. **Myeongjun Erik Jang**, Antonios Georgiadis, Yiyun Zhao, Fran Silavong. (2024). DriftWatch: A Tool that Automatically Detects Data Drift and Extracts Representative Drifted Examples, *In Proceedings of The 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Industry Track (NAACL 2024), Mexico City, Mexico,*, Association for Computational Linguistics [DOI].
2. **Myeongjun Erik Jang**, Gábor Stikkel. (2024). Leveraging Natural Language Processing and Large Language Models for Assisting Due Diligence in the Legal Domain, *In Proceedings of The 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Industry Track (NAACL 2024), Mexico City, Mexico,*, Association for Computational Linguistics [DOI].
3. Vid Kocijan, **Myeongjun Jang**, Thomas Lukasiewicz*. (2024). Pre-training and Diagnosing Knowledge Base Completion Models, *Artificial Intelligence*, *VOLUME 329*, 104081 [DOI].
4. **Myeongjun Erik Jang**, Thomas Lukasiewicz*. (2023). Consistency Analysis of ChatGPT, *In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023), Singapore*, Association for Computational Linguistics [DOI].
5. **Myeongjun Erik Jang**, Thomas Lukasiewicz*. (2023). Improving Language Models' Meaning Understanding and Consistency by Learning Conceptual Roles from Dictionary, *In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023), Singapore*, Association for Computational Linguistics [DOI].
6. Chloe H. Lee, Jaesung Huh, Paul R. Buckley, **Myeongjun Jang**, Mariana Pereira Pinho, Ricardo A. Fernandes, Agne Antanaviciute, Alison Simmons, Hashem Koohy. (2023). A robust deep learning platform to predict CD8+ T-cell epitopes, *Genome Medicine 15, 70*, [DOI]
7. **Myeongjun Jang**, Bodhisattwa Prasad Majumder, Julian McAuley, Thomas Lukasiewicz, Oana-Maria Camburu. (2023). Know How to Make Up Your Mind! Adversarially Detecting and Alleviating Inconsistencies in Natural Language Explanations, *In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023), Toronto, Canada* pp. 540-553, Association for Computational Linguistics. [DOI].
8. **Myeongjun Jang**, Deuk Sin Kwon, Thomas Lukasiewicz*. (2022). BECEL: Benchmark for Consistency Evaluation of Language Models, *In Proceedings of the 29th International Conference on Computational Linguistics (COLING 2023), Gyeongju, Republic of Korea*, pp. 3680-3696, International Committee on Computational Linguistics. [DOI].
9. **Myeongjun Jang**, Dohyung Kim, Deuk Sin Kwon, Eric Davis. (2022). KoBEST: Korean Balanced Evaluation of Significant Tasks, *In Proceedings of the 29th International Conference on Computational Linguistics (COLING 2023), Gyeongju, Republic of Korea*, pp. 3697-3708, International Committee on Computational Linguistics. [DOI]. (*co-first author.*)
10. **Myeongjun Jang**, Thomas Lukasiewicz*. (2022). NoiER: An Approach for Training More Reliable Fine-Tuned Downstream Task Models, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *VOLUME 30*, 2514-2525. [DOI].
11. **Myeongjun Jang**, Frank Mtumbuka, Thomas Lukasiewicz*. (2022). Beyond Distributional Hypothesis: Let Language Models Learn Meaning-Text Correspondence, *In Findings of the Association for Computational Linguistics: NAACL 2022, Seattle, United States*, pp. 2030-2042, Association for Computational

Linguistics. [DOI]

12. **Myeongjun Jang**, Pilsung Kang*. (2022). Sentence Transition Matrix: An efficient approach that preserves sentence semantics, *Computer Speech & Language*, *VOLUME 71*, 101266. [DOI]
13. Czangyeob Kim, **Myeongjun Jang**, Seungwan Seo, Kyeongchan Park, Pilsung Kang*. (2021). Intrusion Detection based on Sequential Information Preserving Log Embedding Methods and Anomaly Detection Algorithms, *IEEE Access*, *VOLUME 9*, 58088-58101. [DOI]
14. **Myeongjun Jang**, Pilsung Kang*. (2021). Learning-free Unsupervised Extractive Summarization Model, *IEEE Access*, *VOLUME 9*, 14358-14368. [DOI]
15. **Myeongjun Jang**, Pilsung Kang*. (2020). Paraphrase Thought: Sentence Embedding Module Imitating Human Language Recognition, *Information Sciences*, *Volumn 541*, 123-145. (SCI) [DOI]
16. Seungwan Seo, Deokseong Seo, **Myeongjun Jang**, Jaeyun Jung, Pilsung Kang*. (2020). Unusual customer response identification and visualization based on text mining and anomaly detection, *Expert Systems With Applications*, *Volume 144*, 113111. (SCIE) [DOI]
17. **Myeongjun Jang**, Seungwan Seo, Pilsung Kang*. (2019). Recurrent Neural Network-Based Semantic Variational Autoencoder for Sequence to Sequence Learning, *Information Sciences*, *Volume 490*, 59-73. (SCI) [DOI]

Patents

1. Method and Apparatus for Detection of Anomaly on Computer System
   - Y.J Yoon, H.S Yoon, H.S Lee, S.D Heo, K.S Kim, Y.J Jeong, P.S Kang, C.Y Kim, **M.J Jang**, S.W Seo, K.C Park
   - Korea - Registration No.10-2088509

Technical Strength

Advanced	Python, R
Intermediate	Java
Novice	Julia, C++

Language Proficiency

Working proficiency in **English**, native in **Korean**

- IELTS: overall 8.0 (Listening 8.5, Reading 9.0, Writing 6.5, Speaking 7.5)
- iBT TOEFL: 99 (Reading 29, Listening 24, Speaking 22, Writing 24)