

Research Interests

My research interests are in machine learning and natural language processing. In particular, I work in Bayesian models and artificial neural networks to simulate human behaviors and discover latent patterns in large datasets such as conversational corpora¹.

Employment

Assistant Professor

Suwon, South Korea

Sungkyunkwan University (SKKU)

Sep 2020 - Present

I am working as a professor in the College of Computing and Informatics at Sungkyunkwan University. I am the leader of the Human Language Intelligence Lab².

Junior Data Scientist

Jakarta, Indonesia

United Nations Global Pulse Lab

June 2016 - Aug 2016

I interned with Dr. Jonggun Lee in the United Nations Global Pulse Lab Jakarta. I developed a topic model based methodology for expanding relevant keywords for specific subjects to identify related tweets in Twitter. This work is published at ICML 2017 workshop [23].

Research Intern

Beijing, China

Microsoft Research Asia

Sep 2013 - Feb 2014

I interned with Dr. Chin-Yew Lin in the Knowledge Mining group at Microsoft Research Asia. I developed a topic model to identify self-disclosure in Twitter conversations. This work is published at EMNLP 2014 as a full paper [25].

Education

KAIST, Ph.D., School of Computing, Mar 2013 - Aug 2020

Dissertation: Speaker oriented Conversation Model and its Evaluation [16]

Dissertation committee: Prof. Alice Oh, Dr. Chin-Yew Lin, Prof. Kee-Eung Kim,
Prof. Meeyoung Cha, and Prof. Sung-Hyon Myaeng

Outstanding PhD Thesis Award

KAIST, M.S., Computer Science, Feb 2011 - Feb 2013

GPA: 3.98/4.30

Thesis: Distributed Online Learning for Topic Models [26]

Thesis committee: Prof. Alice Oh, Prof. Sung-Eui Yoon, and Prof. Sung-Hyon Myaeng

Sungkyunkwan University, B.S., Computer Engineering, Mar 2004 - Feb 2011

GPA: 4.45/4.50

Graduation project: On-Line LECTure Allocation³, Worked with Byung Il Woo,
Best Project Award

Publications

Google Scholar: https://scholar.google.com/citations?user=oYK9Z_IAAAAJ

DBPia (Korean): <https://www.dbpia.co.kr/author/authorDetail?ancId=5086204>

Last updated: November, 2022

¹Full research statement: https://nosyu.kr/assets/JinYeong_Bak_ResearchStatement.pdf

²<https://hli.skku.edu>

³Online system for lecture allocation and mileage management

- [1] Minsoo Park, Dai Quoc Tran, **JinYeong Bak**, and Seunghee Park. Advanced wildfire detection using generative adversarial network-based augmented datasets and weakly supervised object localization. volume 114, page 103052, 2022.
- [2] Haneul Yoo, Jiho Jin, Juhee Son, **JinYeong Bak**, Kyunghyun Cho, and Alice Oh. HUE: Pre-trained model and dataset for understanding hanja documents of Ancient Korea. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1832–1844, Seattle, United States, July 2022. Association for Computational Linguistics.
- [3] HyeJoon Jang and **JinYeong Bak**. Detoxifying toxic comments using text style transfer. In *Proceedings of the Korea Computer Congress*, pages 2081–2083. The Korean Institute of Information Scientists and Engineers, June 2022.
- [4] ChaeYun Jang and **JinYeong Bak**. Multi-turn question generation using past and future information. In *Proceedings of the Korea Computer Congress*, pages 1976–1978. The Korean Institute of Information Scientists and Engineers, June 2022.
- [5] Dai Quoc Tran, Minsoo Park, Yuntae Jeon, **JinYeong Bak**, and Seunghee Park. Forest-fire response system using deep-learning-based approaches with cctv images and weather data. In *IEEE Access*, volume 10, pages 66061–66071, 2022.
- [6] Ida Ayu Putu Ari Crisdayanti, **JinYeong Bak**, YunSeok Choi, and Jee-Hyong Lee. la-bert: Context-aware sarcasm detection by incorporating incongruity attention layer for feature extraction. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, SAC '22*, page 1084–1091, New York, NY, USA, 2022. Association for Computing Machinery.
- [7] JinUk Cho, MinSu Jeong, **JinYeong Bak**, and Yun-Gyung Cheong. Genre-controllable story generation via supervised contrastive learning. In *Proceedings of the ACM Web Conference 2022, WWW '22*, page 2839–2849, New York, NY, USA, 2022. Association for Computing Machinery.
- [8] HyunJin Kim and **JinYeong Bak**. Function name prediction using binary code with transformer. In *Proceedings of the Korea Software Congress*, pages 449–451. The Korean Institute of Information Scientists and Engineers, December 2021.
- [9] DongJin Jeong and **JinYeong Bak**. Extracting emotion-cause information from conversation data using the graph structure. In *Proceedings of the Korea Software Congress*, pages 515–517. The Korean Institute of Information Scientists and Engineers, December 2021.
- [10] YeongJun Hwang and **JinYeong Bak**. User attribute inference using knowledge graph. In *Proceedings of the Korea Software Congress*, pages 551–553. The Korean Institute of Information Scientists and Engineers, December 2021.
- [11] Ji Woo Kim and **JinYeong Bak**. Finding the hidden relation through text-mining the annals of the joseon dynasty. In *Proceedings of the Korea Software Congress*, pages 1473–1439. The Korean Institute of Information Scientists and Engineers, December 2021.
- [12] Yohan Jo, Haneul Yoo, **JinYeong Bak**, Alice Oh, Chris Reed, and Eduard Hovy. Knowledge-enhanced evidence retrieval for counterargument generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3074–3094, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [13] YunSeok Choi, **JinYeong Bak**, CheolWon Na, and Jee-Hyong Lee. Learning sequential and structural information for source code summarization. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, Online, August 2021. Association for Computational Linguistics.
- [14] **JinYeong Bak** and Alice Oh. A Leader's Final Decision Classification Model Tested on Meeting Records with BERT. In *Journal of KI/SE*, volume 48, 2021.
- [15] Bonggun Shin, Sungsoo Park, **JinYeong Bak**, and Joyce C. Ho. Controlled molecule generator for optimizing multiple chemical properties. In *Proceedings of the Conference on Health, Inference, and Learning*, 2021.
- [16] **JinYeong Bak**. Speaker oriented conversation model and its evaluation. PhD dissertation, KAIST, Daejeon, South Korea, 2020.
- [17] **JinYeong Bak** and Alice Oh. Speaker sensitive response evaluation model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.

- [18] **JinYeong Bak** and Alice Oh. Variational hierarchical user-based conversation model. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 2019.
- [19] **JinYeong Bak** and Alice Oh. Conversational decision-making model for predicting the king's decision in the annals of the Joseon dynasty. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.
- [20] Gabriel Lima and **JinYeong Bak**. Speech emotion classification using raw audio input and transcriptions. In *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, 2018.
- [21] Sungjoon Park, **JinYeong Bak**, and Alice Oh. Rotated word vector representations and their interpretability. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017.
- [22] Jongin Lee, Daeki Cho, Junhong Kim, Eunji Im, **JinYeong Bak**, Kwan Hong Lee, John Kim, and others. Itchtector: A wearable-based mobile system for managing itching conditions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017.
- [23] **JinYeong Bak**, Imaduddin Amin, Jong Gun Lee, and Alice Oh. Keyword expansion for understanding crisis events in Indonesian tweets. In *ICML Workshop on Interactive Machine Learning and Semantic Information Retrieval*, 2017.
- [24] **JinYeong Bak** and Alice Oh. Five centuries of monarchy in Korea: Mining the text of the annals of the Joseon dynasty. In *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, 2015.
- [25] **JinYeong Bak**, Chin-Yew Lin, and Alice Oh. Self-disclosure topic model for classifying and analyzing Twitter conversations. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014.
- [26] **JinYeong Bak**. Distributed online learning for topic models. Master's thesis, KAIST, Daejeon, South Korea, 2012.
- [27] **JinYeong Bak**, Dongwoo Kim, and Alice Oh. Distributed online learning for latent Dirichlet allocation. In *Proceedings of Workshop on Big Learning : Algorithms, Systems, and Tools at the Neural Information Processing Systems*, 2012.
- [28] **JinYeong Bak**, Suin Kim, and Alice Oh. Self-disclosure and relationship strength in Twitter conversations. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, 2012.
- [29] Suin Kim, **JinYeong Bak**, and Alice Oh. Do you feel what I feel? social aspects of emotions in Twitter conversations. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*, 2012.
- [30] Suin Kim, **JinYeong Bak**, Yohan Jo, and Alice Oh. Do you feel what I feel? social aspects of emotions in twitter conversations. In *Proceedings of Workshop on Computational Social Science and the Wisdom of Crowds*, 2011.
- [31] Rae Young Ko, Duk Sun Kim, **JinYeong Bak**, and Sang Gu Lee. Development of mobile sage-math and its use in linear algebra. In *J. Korea Soc. Math. Ed. Ser. E: Communications of Mathematical Education*, 2009.
- [32] Duk Sun Kim, **JinYeong Bak**, and Sang Gu Lee. The educational models using enhanced mathematics ict in the Korean it environments. In *J. Korea Soc. Math. Ed. Ser. E: Communications of Mathematical Education*, 2008.

Academic Services - Reviewer and Program Committee

[AAAI] 2020, 2021, 2022, 2023

[AACL] 2020, 2022

[ACL] 2015, 2016, 2017, 2018, 2019, 2020, 2021

[ACML] 2016

[ARR] {Jul, Sep, Oct, Nov} 2021, {Jan, Jul, Oct} 2022

[COLING] 2020, 2022

[EACL] 2016, 2023

[EMNLP] 2015, 2016, 2019, 2020, 2021, 2022

[ICLR] 2021, 2022, 2023

[ICML] 2020

[ICWSM] 2015, 2016

[JOK] 2020, 2021, 2022

[LaTeCH-CLfL] 2016, 2017, 2018, 2019, 2020, 2021, 2022

[LREC] 2022

[NAACL] 2019, 2021

[NeurIPS] 2021, 2022

[SAC] 2021

[TACL] 2019

[TASLP] 2022

[The Web Conference] 2019, 2020, 2022

[W-NUT] 2020, 2021, 2022

Academic Services - Others

Virtual co-chair, ACM FAccT 2022

Mentor of group mentoring, ACL 2020

Volunteer of birds of a feather meetup, ACL 2020

Volunteer of micro-blogging, ACL 2020

Session chair, {POLTEXT 2019, The Web Conference 2022}

Student volunteer, {ACL 2012, EMNLP 2018, EMNLP 2019, ICLR 2020}

Projects

“A study on explainable automatic evaluation model for generated natural language texts in terms of factual consistency”, NRF, June 2022 - Present

“Abductive inference framework using omni-data for understanding complex causal relations”, IITP, Apr 2022 - Present

“National Program for Excellence in SW”, IITP, Apr 2021 - Present

“A study on automatic evaluation model for generated natural language texts in terms of topic consistency”, NRF, June 2021 - May 2022

“Digital SOC D.N.A. based Health Care Laboratory for Carbon Neutralization”, NRF, June 2021 - Present

“Development of Digital Therapeutics for Depression from COVID19”, KEIT, June 2021 - Present

“Standard Development of Blockchain based Network Management Automation Technology”, IITP, Sep 2020 - May 2021

“Artificial Intelligence Graduate School Program”, IITP, Sep 2020 - Present

“Explainable Human-level Deep Machine Learning”, IITP, Sep 2017 - Aug 2020

“Machine Learning Center”, IITP, Mar 2013 - Mar 2017

“Distributed online topic modeling for big data analysis”, Samsung Electronics, Dec 2012 - Nov 2013

Talks

“Natural Language Inference with Knowledge and Emotion-Cause Pair Extraction”, RIKEN, Tokyo, Japan, 2022.11.01.

“Mining the Text of the Annals of the Joseon Dynasty”, Seoul National University, Seoul, South Korea, 2022.10.28.

“NLP for Conversations”, Yonsei University, Seoul, South Korea, 2022.09.15.

“Introduction to NLP for Conversations”, University of Seoul, Seoul, South Korea, 2022.09.06.

“AI and Ethics”, The AI Korea 2022, Seoul, South Korea, 2022.08.17.

“Mining the Text of the Annals of the Joseon Dynasty”, KAIST, Daejeon, South Korea, 2021.12.14.

“Evaluation and Scalability of Conversation models”, Korea Institute of Science and Technology Information, Daejeon, South Korea, 2021.08.20.

“Automatic evaluation methods for conversation models”, SOCAR, Seoul, South Korea, 2021.07.28.

“Automatic evaluation methods for conversation models”, 2021 Summer Conference, Korean Artificial Intelligence Association, Online, 2021.07.09.

“2021 NLP Research Trends”, CJ Olive Networks, Seoul, South Korea, 2021.06.08.

“Automatic evaluation methods for conversation models”, AI Frontiers Summit 2021, The Korean Institute of Communications and Information Sciences, Seoul, South Korea, 2021.05.21.

“Scaling laws in Natural Language Processing”, 2021 Spring Colloquium, SKKU Department of Physics, Suwon, South Korea, 2021.03.03.

“Conversation Model and its Evaluation”, 2020 Fall Conference, Korean Artificial Intelligence Association, Online, 2020.11.20.

“Speaker-oriented Conversation Model and its Evaluation”, Online Seminar, Dongguk University, Seoul, South Korea, 2020.11.18.

“Speaker-oriented Conversation Model and its Evaluation”, BK21 Artificial Intelligence Colloquium, Ajou University, Online, 2020.11.10.

“Speaker-oriented Conversation Model and its Evaluation”, Fall 2020 CSE GSAI Seminar Series, Postech, Online, 2020.09.09.

“Speaker-oriented Conversation Model and its Evaluation”, Machine Learning for Language, NYU, Online, 2020.06.16.

“Variational Hierarchical User-based Conversation Model”, KAIST-NAVER Clova AI Workshop, KAIST, Daejeon, South Korea, 2019.06.19.

“Conversational decision-making model for predicting the kings decision in the annals of the Joseon dynasty”, NAVER, Seongnam, South Korea, 2018.12.14.

“Conversational decision-making model for predicting the kings decision in the annals of the Joseon dynasty and Word vector Interpretability”, Samsung Electronics, Suwon, South Korea, 2018.11.30.

“Self-disclosure in Twitter conversations”, Qatar Computing Research Institute, Doha, Qatar, 2014.10.23.

“Bayesian Nonparametric Topic Modeling”, Korean machine learning summer school, Seoul, South Korea, 2013.08.22.

Teaching Experience

Lecturer, “Machine Learning Algorithms and Applications”, SKKU, Spring 2022

Lecturer, “Smart Factory Application Programming”, SKKU, Fall 2021

Lecturer, “AI and Ethics”, SKKU, Fall 2021

Lecturer, “Introduction to Artificial Intelligence”, SKKU, Fall 2021

Lecturer, “Bayesian Learning”, SKKU, Spring 2021

Lecturer, “Natural Language Processing”, SKKU, {Spring 2021, Fall 2021, Spring 2022}

Lecturer, “Natural Language Processing”, DSME, {Oct 2020, June 2021, Dec 2021}

Lecturer, “Probability and Random Process”, SKKU, {Fall 2020, Spring 2021, Spring 2022}

Lecturer, “Artificial Intelligence for Data Scientists”, elice, {Apr 2017 - Aug 2017, Jun 2018 - Sep 2018}

Teaching assistant, “Artificial Intelligence and Machine Learning”, KAIST, {Fall 2011, Spring 2013, Spring 2015, Fall 2018}

Teaching assistant, “Introduction to Programming”, KAIST, {Spring/Fall 2014, Fall 2015, Spring/Fall 2016, Spring/Fall 2017}

Teaching assistant, “IT Convergence User-centered Service Design”, KAIST - Microsoft Design Expo, Spring 2015

Teaching assistant, “Operating Systems”, SKKU, Fall 2010

Teaching assistant, "Discrete Mathematics", SKKU, Spring 2010

Graduate Student Research Supervision

Academic Advisor

- DongJun Kang, ehdwns2356@g.skku.edu, Master, 2025 (Expected)
- Jihee Kim, mii2026@g.skku.edu, Master, 2024 (Expected)
- Taekhyun Kim, treecko.kth@g.skku.edu, Master, 2024 (Expected)
- HoKeun Yoon, hkyoon95@g.skku.edu, Master, 2024 (Expected)
- Youngjin Jo, yjspecial.jo@g.skku.edu, Master, 2023 (Expected)
- DongJin Jeong, jdjin3000@g.skku.edu, Master, 2023 (Expected)
- YeongJun Hwang, hmtyj2@g.skku.edu, Master, 2023 (Expected)
- HyunJin Kim, khyunjin1993@g.skku.edu, Master, 2023 (Expected)

Graduate Committee Member

- HyunJu Kim, julia1028@skku.edu, Master, 2022
- Dr. Hwanhee Lee, wanted1007@snu.ac.kr, Ph.D., 2022
- Jong Hyeok Park, realkaya@g.skku.edu, Master, 2022
- Jiwung Hyun, kabbi159@gmail.com, Master, 2022
- Sanghee Park, parksangheeeee@naver.com, Master, 2022
- Hyeonjong Ha, wnajrqkq94@g.skku.edu, Master, 2022
- Dr. Byeongchang Kim, byeongchang.kim@gmail.com, Ph.D., 2022
- Dr. DongHyun Choi, cdh4696@gmail.com, Ph.D., 2022
- Donghyun Kim, donghyun.kim@g.skku.edu, Master, 2022
- Sangwoo Han, uoo723@g.skku.edu, Master, 2022
- Jeon Hyun-Kyu, hkjeon13@g.skku.edu, Master, 2022
- Min Su Jeong, omicro03@gmail.com, Master, 2022
- Hoseung Kim, tree901024@skku.edu, Master, 2022
- Samuel Kim, lemonl7@naver.com, Master, 2022
- Youngrok Song, id2thomas@gmail.com, Master, 2022
- Dr. Suin Kim, suin.kim@kaist.ac.kr, Ph.D., 2021
- Ida Ayu Putu Ari Crisdayanti, dayu.crish@gmail.com, Master, 2021
- Jaewoo Choi, yhjgoldhair@naver.com, Master, 2021
- Jinho Lee, most323@naver.com, Master, 2021
- Lee Jung Hoon, vhrehfdl@gmail.com, Master, 2021
- Lee Won Kyu, stbaker517@g.skku.edu, Master, 2021

Programming Skills

Language: Python, PHP, JavaScript

Library: TensorFlow, PyTorch, NumPy, scikit-learn, KoNLPy

GitHub repositories: <https://github.com/NoSyu>

Extracurricular Activities

Organizer, State, Limitations, and Future of Large Language Models, KOFST, 2022

Advisory committee member, Korea Tourism Organization, 2021

Students representative of department of computer science in student council, KAIST, 2014

Social service personnel, Banyeo library, Busan, South Korea, May 2005 - July 2007

Member and representative, Linux & Open-source learned club (SKKULUG), SKKU, 2004 - 2010

References

Prof. Alice Oh, KAIST, alice.oh@kaist.edu

Mr. Byung Il Woo, SK C&C, outerlight@gmail.com

Dr. Chin-Yew Lin, Microsoft Research, cyl@microsoft.com

Prof. Eunseok Lee, SKKU, leees@skku.edu

Dr. Jonggun Lee, Enterprise Blockchain, jonggunlee@gmail.com

Prof. Sang Gu Lee, SKKU, sglee@skku.edu