DONGHEE CHOI, PH.D. @ donghee.choi@imperial.ac.uk ⌗ github.com/dongheechoi

SUMMARY

Research Interest: Food Science, Natural Language Processing, Data Mining, Knowledge Base Construction

I am passionate about developing practical AI applications in the culinary and biomedical domains. My academic projects are closely aligned with industry needs, aiming to implement applications for real-world usage. I have completed an internship and collaborated on projects with Sony, in addition to founding startups.

EDUCATION

Korea University   Seoul, Korea
Ph.D. in Computer Science, advised by Prof. Jaewoo Kang. (On leave: Startups, Oct 2014–Mar 2019)   Mar 2014–Aug 2023
M.S. in Bioinformatics, advised by Prof. Jaewoo Kang   Mar 2012–Feb 2014
B.S. in Computer Science and Engineering (On leave: Military service, Aug 2007 - Jul 2009)   Mar 2006–Feb 2012

WORK EXPERIENCE

Imperial College London   London, United Kingdom
Research Associate   Nov 2023 – Present
- Currently working on the CoDiet project (COMBATTING DIET RELATED NON-COMMUNICABLE DISEASE THROUGH ENHANCED SURVEILLANCE) as a UKRI project under the supervision of Dr. Joram Posma.
- Relation extraction with large language models.

Korea University   Seoul, Korea
Postdoctoral Researcher   Sep 2023 – Nov 2023
- Artificial intelligence applications for food computing, automatic portfolio selection using artificial intelligence, and clinical chatbot adaptation to create a new biomedical large language model.

Sony Research   Tokyo, Japan
Research Intern   Dec 2022 – June 2023
- Focusing on gathering and processing large volumes of text data to extract food and cooking-related information, including biomedical, cultural, and other aspects, utilizing pipelines with large language models.

LYZE, Inc (Dispatched from Kono Labs, Returned to complete Ph.D.)   Seoul, Korea
Co-founder & Research Lead   Jun 2018 – Mar 2019
- Created a backend system with a large-scale cryptocurrency transaction data repository, incorporating real-time data crawling and processing.

Kono Labs   Seoul, Korea
Data Scientist & Backend Developer   Apr 2016 – Mar 2019
- Worked as a Lead Backend engineer based on Java & Spring server developments.
- Created a near real-time scheduling agent based on email requests, managing accounts with Google and Microsoft APIs, and overseeing the scheduling data repository composed of numeric and text data.

Opinion8 (Dispatched from Korea University, merged into Kono Labs)   Seoul, Korea
Co-founder & CIO   Oct 2014 – Apr 2016
- Implemented an opinion, location, and time-based entity (restaurants, attractions, and hotels) search engine based on large-scale web data. Managing a large-scale text repository and its search engine.

Korea Army Training Center, KATC   Nonsan, Korea
Instructor   Aug 2007 – Jul 2009
- Developed leadership skills by instructing and managing a diverse group of recruits.

⋆: First or Co-first author paper; †: Equal contributions, ‡: Corresponding author.

Journal Articles

- Hyunjae Kim, Hyeon Hwang, Jiwoo Lee, Sihyeon Park, Dain Kim, Taewhoo Lee, Chanwoong Yoon, Jiwoong Sohn, Jungwoo Park, Olga Reykhart, Thomas Fetherston, **Donghee Choi**, Soo Heon Kwak, Qingyu Chen, Jaewoo Kang *"Small Language Models Learn Enhanced Reasoning Skills from Medical Textbooks."* npj Digital Medicine (2025, Accepted). Preprint: https://arxiv.org/abs/2404.00376

- ⋆ **Donghee Choi**, Mogan Gim, Samy Badreddine, Hajung Kim, Donghyeon Park, and Jaewoo Kang. *"KitchenScale: Learning to predict ingredient quantities from recipe contexts."* Expert Systems with Applications 224 (2023): 120041. https://doi.org/10.1016/j.eswa.2023.120041

- ⋆ Jinkyu Kim†, **Donghee Choi**†, Mogan Gim, and Jaewoo Kang. *"HADAPS: Hierarchical Adaptive Multi-Asset Portfolio Selection."* IEEE Access 11 (2023): 73394-73402. https://doi.org/10.1109/ACCESS.2023.3285613

- Sunwon Lee, Donghyeon Kim, Kyubum Lee, Jaehoon Choi, Seongsoon Kim, Minji Jeon, Sangrak Lim, **Donghee Choi**, Sunkyu Kim, Aik-Choon Tan, and Jaewoo Kang. *"BEST: next-generation biomedical entity search tool for knowledge discovery from biomedical literature."* PloS one 11, no. 10 (2016): e0164680. https://doi.org/10.1371/journal.pone.0164680

Full Conference Papers

- ⋆ **Donghee Choi**†, Jinkyu Kim†, Mogan Gim, Jinho Lee, and Jaewoo Kang. *"DeepClair: Utilizing Market Forecasts for Effective Portfolio Selection"* Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024. https://doi.org/10.1145/3627673.3680008

- Heedou Kim, Dain Kim, Jiwoo Lee, Chanwoong Yoon, **Donghee Choi**, Mogan Gim, and Jaewoo Kang. *"LAPIS: Language Model-Augmented Police Investigation System"* Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024. https://doi.org/10.1145/3627673.3680044

- ⋆ **Donghee Choi**, Mogan Gim, Donghyeon Park, Mujeen Sung, Hyunjae Kim, Jaewoo Kang, and Jihun Choi. *"CookingSense: A Culinary Knowledgebase with Multidisciplinary Assertions."* In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pp. 3983-3996. 2024. https://aclanthology.org/2024.lrec-main.354/

- ⋆ Mogan Gim†, **Donghee Choi**†, Kana Maruyama, Jihun Choi, Hajung Kim, Donghyeon Park, and Jaewoo Kang. *"Recipemind: Guiding ingredient choices from food pairing to recipe completion using cascaded set transformer."* In Proceedings of the 31st ACM international conference on information & knowledge management, pp. 3092-3102. 2022. https://doi.org/10.1145/3511808.3557092

- Donghyeon Kim, Jinhyuk Lee, **Donghee Choi**, Jaehoon Choi, and Jaewoo Kang. *"Learning user preferences and understanding calendar contexts for event scheduling."* In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 337-346. 2018. https://doi.org/10.1145/3269206.3271712

- Jinhyuk Lee, Hyunjae Kim, Miyoung Ko, **Donghee Choi**, Jaehoon Choi, and Jaewoo Kang. *"Name Nationality Classification with Recurrent Neural Networks."* In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Main Track, vol. 17, pp. 2081-2087. 2017. https://www.ijcai.org/proceedings/2017/289

Demo Conference Papers

- Jaehoon Choi, Donghyeon Kim, **Donghee Choi**, Sangrak Lim, Seongsoon Kim, Jaewoo Kang, and Youngjae Choi. *"Smith search: Opinion-based restaurant search engine."* In Proceedings of the 24th International Conference on World Wide Web, pp. 187-190. 2015. https://doi.org/10.1145/2740908.2742829

Conference Workshop Papers

- ⋆**Donghee Choi**, Antoine D Lain, Joram M. Posma, Mark Kozdoba, Binyamin Perets, Shie Mannor. *"From Medical Literature to Predictive Features: An Evidence-based Knowledge Graph Approach"* (Accepted at LMRL Workshop at ICLR 2025).

Preprints

- Abdurrahim Yilmaz, Furkan Yuceyalcin, Ece Gokyayla, **Donghee Choi**, Ozan Erdem Ali Anil Demircali, Rahmetullah Varol, Ufuk Gorkem Kirabali, Gulsum Gencoglan, Joram M. Posma, Burak Temelkuran *"DermaSynth: Rich Synthetic Image-Text Pairs Using Open Access Dermatology Datasets"* arXiv preprint arXiv:2502.00196 (2025). https://www.arxiv.org/abs/2502.00196

- ★ Hoonick Lee, Mogan Gim, Donghyeon Park, **Donghee Choi**‡, and Jaewoo Kang‡ *"Culinary Class Wars: Evaluating LLMs using ASH in Cuisine Transfer Task"* arXiv preprint arXiv:2411.01996 (2024).
  https://arxiv.org/abs/2411.01996

Patents

- Jaewoo Kang, Jaehoon Choi, Jinhyuk Lee, and **Donghee Choi**. *"DEVICE AND METHOD FOR PROVIDING NATIONALITY INFORMATION OF USER NAME USING NEURAL NETWORKS"* (2017, Korea)
  https://patents.google.com/patent/KR20190051574A/ko
- Jaewoo Kang, Jaehoon Choi, Sangrak Lim, and **Donghee Choi**. *"SYSTEM AND METHOD FOR MANAGING CONTENTS"* (2014, Korea) https://doi.org/10.8080/1020120078866

OTHERS

Reviewer for Conferences: KDD (2025), ACL Rolling Review (2025 Feb, April), ACL (Secondary: 2021), EMNLP (2021, Secondary: 2023), NAACL (2024), ICDM (Secondary: 2019, 2020)

Reviewer for Journals: Information Processing & Management (Secondary: 2024), JMIR (2024), Discover Computing (2024), IEEE Transactions on Industrial Informatics (2024)

Invited Talks: Domain-specific AI applications: Food and Finance to Assistant Professor Minji Jeon's lab (Mar, 2024)

Public Science Outreach & Service:
CoDiet at the Great Exhibition Road Festival 2024 (CoDiet Homepage), Conference Presentations (Oral: CIKM 2024 / Poster: LREC-COLING 2024, CIKM 2022, WWW 2015), Session Chair(CIKM 2024: Biomedical & Health), Ph.D. Symposium Mentor (CIKM 2024)

Teaching: Supervised 3 Research Projects of MSc Students in 2024 at Imperial College London: 1) Combining Machine- And Human-Annotations In A Multi-Layer Network To Predict Probabilities Of Biomedical Entities In Scientific Texts, 2) Investigating Diet And Food Relationships With Language Model Perspectives, 3) Using level of expertise as prior in optimising and fine-tuning biomedical language models

Marking of oral presentations of BSc module on Biomedical Data Science at Imperial College London (2024).
Presentation on the current biomedical language models at the BDSC seminar at Imperial College London (17 July 2024).