To Eun Kim

Information Retrieval | LLM | Retrieval-Enhanced ML | Language Agents | Evaluation

✉ toeunkim@cmu.edu   🎓 Google Scholar   🌐 https://kimdanny.github.io/

Education

Carnegie Mellon University (CMU)   Aug 2023 - (2028)
Ph.D. Language and Information Technology at Language Technologies Institute (LTI).
Advisor: Prof. Fernando Diaz
Focus: Information Retrieval, Retrieval-Enhanced Machine Learning, Language Agents

University College London (UCL)   Sep 2018 - Sep 2022
M.Eng. Computer Science (4-year master's integrated undergrad. degree)   **First-Class Honours**
Minor: Engineering Mathematics
Thesis: Multi-Task Neural User Simulator for Task Oriented Dialogue System

Publications

Tip of the Tongue Query Elicitation for Simulated Evaluation
Y. He*, T.E. Kim*, F. Diaz, J. Arguello, B. Mitra [paper] [code1] [code2]
(SIGIR 2025) 48th ACM SIGIR Conference on Research and Development in Information Retrieval

Overview of the TREC 2024 Tip-of-the-Tongue Track
J. Arguello, S. Bhargav, F. Diaz, T.E. Kim, Y. He, E. Kanoulas, B. Mitra [paper]
(TREC 2025) The 33rd Text Retrieval Conference

Towards Fair RAG: On the Impact of Fair Ranking in Retrieval-Augmented Generation
T.E. Kim, F. Diaz [paper] [code]
(NeurIPS 2024 Workshop) Algorithmic Fairness through the lens of Metrics and Evaluation Workshop
Top-3 Spotlight Presentation

Retrieval-Enhanced Machine Learning: Synthesis and Opportunities
T.E. Kim, A. Salemi, A. Drozdov, F. Diaz, H. Zamani [paper]
(SIGIR-AP 2024 Tutorial) 2nd ACM SIGIR Conference on Information Retrieval in the Asia Pacific
Manuscript Under Review for Journal Publication

A Comprehensive Survey on Generative Diffusion Models for Structured Data
H. Koo, T.E. Kim [paper]
arXiv, 2023

A Multi-Task Based Neural Model to Simulate Users in Goal-Oriented Dialogue Systems
T.E. Kim, A. Lipani [paper] [poster] [code]
(SIGIR 2022) 45th ACM SIGIR Conference on Research and Development in Information Retrieval

Condita: A State Machine Like Architecture for Multi-Modal Task Bots
J. Ramos*, T.E. Kim*, Z. Shi, X. Fu, F. Ye, Y. Feng, A. Lipani [paper]
Alexa Prize TaskBot Challenge Proceedings, 2022

When and What to Ask Through World States and Text Instructions: IGLU NLP Challenge Solution
Z. Shi*, J. Ramos*, T.E. Kim, X. Wang, H. Rahmani, A. Lipani [paper]
(NeurIPS 2022 Competition) Interactive Grounded Language Understanding in a Collaborative Environment
Research Prize Winner

Attention-based Ingredient Phrase Parser
Z. Shi, P. Ni, M. Wang, T.E. Kim, A. Lipani [paper] [code]
(ESANN 2022) European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning

Research & Industry Experience

CMU LTI - Graduate Research and Teaching Assistant   Aug 2023 - Present

- Courses taken: 11642 Search Engines (Prof. Jamie Callan); 10701 Intro to ML (Dr. Henry Chai); 11711 Advanced NLP (Prof. Graham Neubig); 11801 Quantitative Evaluation of Language Technologies (Prof. Fernando Diaz); 11797 Question Answering (Prof. Eric Nyberg, Prof. Teruko Mitamura); 11830 Ethics, Safety, and Social Impact in NLP and LLMs (Prof. Maarten Sap)

UCL Web Intelligence Group　　　　　　　　　　　　　　　　　　May 2021 - Jun 2023
Research & Teaching Assistant　　　　　　　　　　　　　　　　　　　　UCL Centre for AI
- Supervised by Prof. Emine Yilmaz and Dr. Aldo Lipani
- Alexa Prize TaskBot Challenge - Quarter-finals.
- Four research papers.
- Four TAships and lecturer for Python programming course.

Raft AI　　　　　　　　　　　　　　　　　　　　　　　　　　　　Jun 2022 - Jul 2022
Machine Learning Engineer Intern　　　　　　　　　　　　　　　　　　　London, UK
- Tabular information extraction from OCR for freight forwarding automation.
- Use of modern dev tools, e.g., Gitlab, Jira, Sentry, Docker, Kubernetes, GCP.

Amazon Alexa Prize TaskBot Challenge　　　　　　　　　　　　　　May 2021 - Mar 2022
Team Condita
- Engineered a Task-oriented Dialogue System by using Alexa Skills Kit, Evi, DynamoDB, Lambda, and CodePipeline. Constructed and maintained CI/CD pipelines, deployed multiple microservices with Flask and Docker, dealing with thousands of user conversations with feedback loop (more details in Master's thesis).

UCL (Human-Computer) Interaction Centre - bachelor's group thesis　　Oct 2020 - Apr 2021
Project *Olfactory-NLP* with group memebers and Prof. Marianna Obrist .
- Extracted smell-related phrases and clustered multiple passages into different olfactory groups.
- Literature Review: *An NLP Approach to Classify Smell Experiences.* [paper]
- *Exploring the potential of automating the process of clustering smell stories.* [paper] [code]

KB Securities　　　　　　　　　　　　　　　　　　　　　　　　Jun 2020 - Aug 2020
Data Analyst Intern　　　　　　　　　　　　　　　　　　　　　　Seoul, South Korea
- Built a mutual funds recommender system based on Network Analysis, Hierarchical Clustering and Association Rule Mining.
- Built a stock price prediction system using LSTM and news articles. [code]
- Social media sentiment monitoring system using KB's NLP Engine and Automation software.
- Use of Hadoop ecosystem (Map-reduce, Hive, and Hue).

Royal College of General Practitioners (RCGP)　　　　　　　　　　Oct 2019 - Mar 2020
Software Engineer - UCL Industrial Exchange Network　　　　　　　　　　London, UK
- Implemented a tree-based task-oriented chatbot that assists in checking software compliance regimes of British healthcare organizations based on Dialogflow.

Presentations & Invited Talks	Algorithmic Fairness through the Lens of Metrics and Evaluation @ NeurIPS 2024, ***Towards Fair RAG: On the Impact of Fair Ranking in Retrieval-Augmented Generation***, 14th Dec., 2024.
	NLP Interest Group at The Alan Turing Institute, ***Can Dialogue Systems Have Theory-of-Mind Ability?***, 24th Nov., 2022. [recording][slides]
	Knowledge Graphs Interest Group at The Alan Turing Institute (5th meet-up), ***Personal Knowledge Graphs: Towards Better Conversational Assistants***, 27th Oct., 2022. [recording] [slides]
	Search-Oriented Conversational AI Workshop (SCAI '22), ***A Multi-Task Based Neural Model to Simulate Users in Goal-Oriented Dialogue Systems***, 15th Jul., 2022. [recording] [slides]
Awards	2022 NeurIPS Interactive Grounded Language Understanding (IGLU) – Research Prize – $650
	2022 Automated Knowledge Graph Construction (AKBC) Conference – Registration fee waiver
	2022 SIGIR Student Travel Grant – Registration fee waiver

2021 UCL Student Union best societies award – *Centenary Award (UCL AI Society)*
2021 UCL CS Undergraduate Summer Research Internship Fund – £3,200

Teaching Experience	UCL Intense Python Programming course, **Course Co-Organiser and Lecturer**	March 2023

- Designing and will lead a new intensive programming course for MSc dissertations.

UCL COMP0189 Applied Artificial Intelligence, *Teaching Assistant* — Spring 2023
- Making weekly practical contents and coursework [GitHub]

UCL COMP0071 Software Engineering, *Teaching Assistant* — Spring 2023
UCL CEGE0004 Machine Learning for Data Science, *Teaching Assistant* — Spring 2023
UCL CEGE0096 Geospatial Programming, *Teaching Assistant* — Fall 2022
UCL Artificial Intelligence Society, *Lecturer* — 2020 - 2022
- **Founder, maintainer, and lecturer** of Machine Learning Tutorial Series
- Season 20/21: [GitHub] [Lecture Recordings] || Season 21/22: [GitHub]

Mentoring
CMU Paths to AI, *Mentor* — Spring 2025
CMU Paths to AI, *Mentor* — Spring 2024
CMU Graduate Application Support Program — Fall 2023

Community Involvement
TREC 2025 Tip-of-the-Tongue (ToT) Track, *Organizer*; TREC track website
The Alan Turing Institute NLP & KG Interest Group, *member* — Nov 2022
The Alan Turing Institute Knowledge Graphs Interest Group, *member* — Sep 2022
Automated Knowledge Graph Construction (AKBC) Conference, *volunteer* — Nov 2022
ICLR 2023 *(Reviewer)*, CIKM 2022 *(Sub-reviewer)* — July, Oct 2022
UCL Artificial Intelligence Society, *Head of Tutorials Team* — Sep 2019 - Jun 2022
- Co-organised an international competition, *Climate Hack.AI* [GitHub] [media coverage]

NLP paper reading group (Jiphyeonjeon), *Season 2,3 Organizer* — May 2021 - Nov 2022
- Season 2: Survey on Task-Oriented Dialog System and Knowledge Graph construction. [slides] [recording] (spoken in Korean)
- Season 3: Recent Advances and Challenges in Task-oriented Dialogue Systems. [slides] [recording] (spoken in Korean)

UCL CS Transition Mentor, *Mentor - helping 1st year CS students* — Sep 2020 - Jun 2022
Google Developer Student Club, *UCL core team - event with Marc Cohen* — Sep 2021 - Jun 2022
UCL Korean Society, *IT Director - webmail automation* — Jun 2020 - May 2021

Software
Unified API of Apache Kafka & Google Pub/Sub for model serving — Nov 2021
- Serving ML models in an event-driven way. [code]

Propositional Semantic Tableau Validator — Dec 2019
- Propositional logic satisfiability checker based on semantic tableau (implemented in C) [code]

Professional Training
Google Developers Korea Machine Learning Bootcamp — Oct 2020 - Jan 2021
Deep Learning Specialization (Prof. Andrew Ng on Coursera) — Dec 2020
UCLQ Summer School in Quantum Technology — Sep 2020
Introduction to AI in the Data Center (NVIDIA) — Mar 2020
KAIST Artificial Intelligence Bootcamp [Project 1] [Project 2] — Jun 2019 - Aug 2019

Major Skills

Computer Language:	Python, Java, C, Matlab, R, LaTeX
Research:	PyTorch, TensorFlow2, HuggingFace, Scikit-Learn
SW Dev:	Docker, SQL, Kafka, Node, Express, React, Flask
Platform:	Unix, Git, AWS, Alexa Skills Kit, GCP
Natural Language:	English (professional), Korean (native)

References — Available on request.