

Yo Joong Choe

Postdoctoral Scholar

Data Science Institute, University of Chicago

Updated March 2024

Email: yjchoe@uchicago.edu

Website: <https://yjchoe.github.io>

RESEARCH INTERESTS

Game-theoretic statistics: sequential inference; anytime-validity; e-values and e-processes; confidence sequences; testing by betting; nonparametric inference; evaluation of forecasters and black-box predictors

Science of large language models: causal representations; understanding text embeddings; Transformers; (mechanistic) interpretability; alignment

EMPLOYMENT

Postdoctoral Scholar

Sep 2023 – current

Data Science Institute, University of Chicago

Mentor: Victor Veitch

Research Scientist

Aug 2019 – Nov 2020

Kakao Brain

Research Scientist

Jul 2017 – Jul 2019

Kakao

EDUCATION

Ph.D. in Statistics and Machine Learning

2015 – 2023

Carnegie Mellon University

(On leave: 2017 – 2020)

Advisor: Aaditya Ramdas

Thesis: Comparing Forecasters and Abstaining Classifiers ([thesis](#), [slides](#))

M.S. in Machine Learning

2015 – 2017

Carnegie Mellon University

Advisors: Aarti Singh, Sivaraman Balakrishnan, and Timothy Verstynen

Thesis: Finding Relationships Between Structural and Functional Brain Networks via Connectome Fingerprinting

B.S. in Mathematics and Computer Science, Honors

2011 – 2015

University of Chicago

AWARDS

Runner-up Poster Award, Citadel Securities PhD Summit

2022

Franklin V. Taylor Memorial (Best Paper) Award, IEEE SMC

2018

Phi Beta Kappa, University of Chicago

2014

PREPRINTS

Combining Evidence Across Filtrations ([arXiv](#))

Yo Joong Choe, Aaditya Ramdas

arXiv, 2024

The Linear Representation Hypothesis and the Geometry of Large Language Models (arXiv, code)

Kiho Park, Yo Joong Choe, Victor Veitch

arXiv, 2023; Presented at *NeurIPS Workshop on Causal Representation Learning*

JOURNAL PAPERS

Comparing Sequential Forecasters (paper, arXiv, code, slides, poster)

Yo Joong Choe, Aaditya Ramdas

Operations Research, 2023; Presented at *SAVI 2022* and *JSM 2021*

Runner-up Poster Award (\$10K Research Grant), Citadel Securities PhD Summit

Predicting Drug-Target Interaction Using a Novel Graph Neural Network with 3D Structure-Embedded Graph Representation (paper, arXiv)

Jaechang Lim, Seongok Ryu, Kyubyong Park, Yo Joong Choe, Jiyeon Ham, Woo Youn Kim

Journal of Chemical Information and Modeling, 2019

CONFERENCE PAPERS

Counterfactually Comparing Abstaining Classifiers (arXiv, code, slides, poster)

Yo Joong Choe, Aditya Gangrade, Aaditya Ramdas

Advances in Neural Information Processing Systems (NeurIPS), 2023

Presented (oral and poster) at the *ICML 2023 Workshop on Counterfactuals*

KorNLI and KorSTS: New Benchmark Datasets for Korean Natural Language Understanding (paper, data)

Jiyeon Ham^{*}, Yo Joong Choe^{*}, Kyubyong Park^{*}, Ilji Choi, Hyungjoon Soh

Findings of the Association for Computational Linguistics: EMNLP, 2020

word2word: A Collection of Bilingual Lexicons for 3,564 Language Pairs (paper, code)

Yo Joong Choe^{*}, Kyubyong Park^{*}, Dongwoo Kim^{*}

Language Resources and Evaluation Conference (LREC), 2020

Jejueo Datasets for Machine Translation and Speech Synthesis (paper, code)

Kyubyong Park, Yo Joong Choe, Jiyeon Ham

Language Resources and Evaluation Conference (LREC), 2020

Discovery of Natural Language Concepts in Individual Units of CNNs (paper, poster, code)

Seil Na, Yo Joong Choe, Dong-Hyun Lee, Gunhee Kim

International Conference on Learning Representations (ICLR), 2019

Local White Matter Architecture Defines Functional Brain Dynamics
([paper](#), [arXiv](#), [slides](#))

Yo Joong Choe, Sivaraman Balakrishnan, Aarti Singh, Jean Vettel, Timothy Verstynen

IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018
Franklin V. Taylor Memorial Award (for Best Paper and Oral Presentation)

WORKSHOP PAPERS
(^[P]: PUBLISHED)

An Empirical Study of Invariant Risk Minimization ([arXiv](#), [code](#), [slides](#))

Yo Joong Choe, Jiyeon Ham, Kyubyong Park

ICML Workshop on Uncertainty and Robustness in Deep Learning, 2020

^[P]**A Neural Grammatical Error Correction System Built On Better Pre-training and Sequential Transfer Learning** ([paper](#), [code](#))

Yo Joong Choe^{*}, Jiyeon Ham^{*}, Kyubyong Park^{*}, Yeoil Yoon^{*}

Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2019

Runner-up, Workshop Shared Task (Restricted & Low Resource Tracks)

Learning Diverse Overcomplete Dictionaries via Determinantal Priors
([abstract](#))

Maruan Al-Shedivat, Yo Joong Choe, Neil Spencer, Eric P. Xing

ICML Workshop on Geometry in Machine Learning, 2016

MISCELLANEOUS

Probabilistic Interpretations of Recurrent Neural Networks ([report](#))

Collaborators: Jaehyeok Shin, Neil Spencer

Technical Report, 2017

A Statistical Analysis of Neural Networks ([report](#))

Technical Report, 2016

Sparse Additive Models with Shape Constraints ([report](#), [slides](#), [code](#))

Mentors: John Lafferty, Sabyasachi Chatterjee, Min Xu

University of Chicago Computer Science REU, 2014

TALKS

Comparing Sequential Forecasters

Eurandom, the Netherlands

Workshop on Safe, Anytime-Valid Inference (SAVI)

2022

A Casual Intro to Machine Learning & Deep Learning

Seoul, S. Korea

Seoul Tech Society Meetup

2019

REVIEWING

International Conference on Machine Learning (ICML)

2024

ACM T. on Asian and Low-Resource Language Information Processing

2023

Journal of the Royal Statistical Society: Series C (Applied Statistics)

2022

ICML Workshop on Adaptive & Multitask Learning

2019

TEACHING

Postdoctoral Mentor for Student Projects

University of Chicago, Data Science Institute Clinic

(In the clinic, student teams work with external clients' data science projects.

Below is the list of clients whose projects I mentored.)

Argonne National Laboratory, AI for Operations	Winter 2024
Argonne National Laboratory, AI for Operations	Fall 2023

Teaching Assistant

Carnegie Mellon University

UnitedHealth Group Bridges to Healthcare Technology (Summer Undergraduate Research Program)	Summer 2023
Simulation Methods for Option Pricing	Spring 2023
Financial Time Series Analysis	Spring 2022; 2023
Data Science in Sports & Healthcare (Summer Undergraduate Research Program)	Summer 2022
Special Topics: Methods of Statistical Learning	Spring 2021
Statistical Machine Learning II	Fall 2016
Introduction to Probability Modeling	Spring 2016
Statistical Computing	Fall 2015

University of Chicago

Machine Learning and Large-Scale Data Analysis	Spring 2015
Foundations of Computational Data Analysis	Winter 2015
Calculus I-II-III	Fall 2012 – Spring 2013