

# Joon Sik Kim

Last Updated: May 14, 2023

I recently graduated from CMU with a PhD in machine learning. My research focuses on methods that can assist our understanding of complex machine learning models, and their application to useful downstream tasks involving human users. Methods include efficient decomposition of neural network predictions, algorithmic fairness assessments; applications include dataset debugging, spurious correlation detection, and assisting human users in document matching. Primarily focused on deep learning models applied to text and image data, the approaches studied include model-agnostic perspectives that can be extended to different data modalities.

## CONTACT

Email: [davidkim21c@gmail.com](mailto:davidkim21c@gmail.com)  
Website: <https://wnstlr.github.io>

## EDUCATION

**Carnegie Mellon University** Pittsburgh, PA  
Doctor of Philosophy, Machine Learning May 2023  
*Thesis: Methods and Applications of Explainable Machine Learning*  
Advisor: Ameet Talwalkar

**Carnegie Mellon University** Pittsburgh, PA  
Master of Science, Machine Learning (GPA 3.86/4.00) May 2020

**California Institute of Technology** Pasadena, CA  
Bachelor of Science, Computer Science (GPA 4.0/4.3) June 2017

## EXPERIENCE

**Microsoft Research** Redmond, WA  
*Summer Research Intern* June 2021 - Aug 2021

**J.P.Morgan** New York, NY  
*Summer Intern at AI Research* May 2019 - Aug 2019

**SK Inc.** Seoul, South Korea  
*Machine Learning Research Intern* June 2017 - Aug 2017

**Computer Vision Lab, California Institute of Technology** Pasadena, CA  
*Undergraduate Research Student for Prof. Pietro Perona* July 2016 - Dec 2016

**California Institute of Technology** Pasadena, CA  
*Undergraduate Research Student for Prof. Yisong Yue* June 2015 - June 2016

**California Institute of Technology** Pasadena, CA  
*Undergraduate Research Student for Dr. Vasumathi Raman* July 2014 - Nov 2014

**Criminal Investigation Command, Ministry of National Defense** Seoul, South Korea  
*Soldier, Software Developer, Interpreter* Oct 2012 - July 2014

**Laser Interferometer Gravitational-wave Observatory (LIGO)** Pasadena, CA  
*Undergraduate Research Student for Dr. Parameswaran Ajith* Jan 2012 - Aug 2012

## PUBLICATIONS<sup>1</sup>

8. Keegan Harris, Valerie Chen, **Joon Sik Kim**, Ameet Talwalkar, Hoda Heidari, Steven Wu. *Bayesian Persuasion for Algorithmic Recourse*. Neural Information Processing Systems (NeurIPS), 2022.
7. **Joon Sik Kim**, Gregory Plumb, Ameet Talwalkar. *Sanity Simulations for Saliency Methods*. International Conference on Machine Learning (ICML), 2022. (Spotlight)
6. Valerie Chen\*, Jeffrey Li\*, **Joon Sik Kim\*\***, Gregory Plumb\*\*, Ameet Talwalkar. *Interpretable Machine Learning: Moving From Mythos to Diagnostics*. ACM Queue 19.6 (2022): 28-56.
5. Kwangho Kim, Jisu Kim, Manzil Zaheer, **Joon Sik Kim**, Frederic Chazal, Larry Wasserman. *PLLayer: Efficient Topological Layer based on Persistence Landscapes*. Neural Information Processing Systems (NeurIPS), 2020.
4. **Joon Sik Kim**, Jiahao Chen, Ameet Talwalkar. *FACT: A Diagnostic for Group Fairness Trade-offs*. International Conference on Machine Learning (ICML), 2020.
3. David I. Inouye, Leqi Liu, **Joon Sik Kim**, Bryon Aragam, Pradeep Ravikumar. *Automated Dependence Plots*. Uncertainty in Artificial Intelligence (UAI), 2020.
2. Chih-Kuan Yeh\*, **Joon Sik Kim\***, Ian E.H. Yen, Pradeep Ravikumar. *Representer Point Selection for Explaining Deep Neural Networks*. Neural Information Processing Systems (NeurIPS), 2018.
1. Matteo Ruggero Ronchi, **Joon Sik Kim**, Yisong Yue. *A Rotation Invariant Latent Factor Model for Movement Discovery from Static Poses*. IEEE International Conference on Data Mining (ICDM), 2016.

## PRE-PRINTS

2. **Joon Sik Kim**, Valerie Chen, Danish Pruthi, Nihar Shah, Ameet Talwalkar. *Assisting Human Decisions in Document Matching*. arXiv 2023.
1. Valerie Chen, Muyu Yang, Wenbo Cui, **Joon Sik Kim**, Ameet Talwalkar, Jian Ma. *Best Practices for Interpretable Machine Learning in Computational Biology*. bioRxiv 2022.10.28.513978

## PROFESSIONAL ACTIVITIES

Reviewer (FAccT 2023, ICML 2022, FAccT 2022, ICML 2021); Program Committee (Workshop on Fair AI in Finance, NeurIPS 2020).

## FELLOWSHIPS, AWARDS

Kwanjeong Foundation Graduate Research Fellowship	2017 - 2021
Caltech Bhansali Family Prize in Computer Science	2017
Caltech CMS Celebration of Undergraduate Research Best Demo/Poster Award	2015
Caltech Summer Undergraduate Research Fellowship	2014, 2015, 2016
Caltech Physics 11 Summer Undergraduate Research Fellowship	2012

## TEACHING

10-725 Convex Optimization (Carnegie Mellon University)	Spring 2023
10-737 Creative AI (Carnegie Mellon University)	Fall 2019

---

<sup>1</sup>\* indicates equal contribution; includes conference workshop papers