Sung Min (Sam) Park

sungminp@stanford.edu | www.sungminpark.com

Current Position	**Postdoctoral Scholar, Department of Computer Science, Stanford University** Advisors: Tatsunori Hashimoto, Percy Liang, James Zou Sep 2024 – Present
Education	**Massachusetts Institute of Technology** *Ph.D. in EECS* Advisor: Aleksander Mądry Dissertation: Machine Learning through the Lens of Data Aug 2018 - Sep 2024 (on leave 2016-18 for military service) **Massachusetts Institute of Technology** *S.M. in EECS* Advisor: Guy Bresler Thesis: On the Equivalence of Sparse Statistical Problems Aug 2014 - Aug 2016 **Cornell University** *B.S. in Computer Science, GPA: 4.22/4.3* Aug 2011 - May 2014
Research Focus	My broad goal is build a principled, scalable, and robust toolkit to understand and improve large-scale machine learning models, often with a focus on data. I have developed state-of-the-art data attribution methods that scale reliably to modern generative models.
Honors & Awards	**Samsung Scholarship for Graduate Studies** (five year fellowship), 2014 **MIT Akamai Presidental Fellowship**, 2014 **Cornell CS Prize for Academic Excellence** (one of two seniors), 2014 **Cornell Merrill Presidential Scholars** (top 1% of graduating class), 2014
Publications	(* denotes equal contribution; α-β indicates alphabetical order)

Conference Publications

Kristian Georgiev*, Roy Rinberg*, **Sung Min Park***, Shivam Garg*, Andrew Ilyas, Aleksander Mądry, Seth Neel. "Attribute-to-Delete: Machine Unlearning vis Datamodel Matching." International Conference on Learning Representations, 2025.

S.M. Park*, K. Georgiev*, A. Ilyas*, G. Leclerc, A. Mądry. "TRAK: Attributing model behavior at scale." International Conference on Machine Learning, 2023. **Oral.**

H. Shah*, **S.M. Park***, A. Ilyas*, A. Mądry. "Modeldiff: A framework for comparing learning algorithms." International Conference on Machine Learning, 2023.

S. Jain*, H. Salman*, A. Khaddaj*, E. Wong, **S.M. Park**, A. Mądry. "A Data-Based Perspective on Transfer Learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern, 2023.

G. Leclerc, A. Ilyas, L. Engstrom, **S.M. Park**, H. Salman, A. Mądry. "FFCV: Accelerating training by removing data bottlenecks." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern, 2023.

A. Ilyas*, **S.M. Park***, L. Engstrom*, G. Leclerc, A. Mądry. "Datamodels: Predicting predictions from training data." International Conference on Machine Learning, 2022.

(α-β) G. Bresler, **S.M. Park**, M. Persu. "Sparse PCA from sparse linear regression." Advances in Neural Information Processing Systems (NeurIPS), 2018.

A. Fix, T. Joachims, **S.M. Park**, R. Zabih. "Structured learning of sum-of-submodular higher order energy functions" Proceedings of the IEEE International Conference on Computer Vision, 2013.

Preprints

K. Georgiev*, J. Vendrow*, H. Salman, **S.M. Park**, A. Mądry. "The Journey, Not the Destination: How Data Guides Diffusion Models." 2023.

S.M. Park, K.A. Wei, K. Xiao, J. Li, A. Mądry. "On Distinctive Properties of Universal Perturbations" `https://https://arxiv.org/abs/2112.15329`, 2021.

Work Experience	**Waymo, LLC.**

Perception R&D Intern, May 2020 - August 2020
 - Designed and implemented new end-to-end DNN models for structured object detection from LiDAR data. Led to the patent "Region Detection and Geometry Prediction."

Republic of Korea Army
Signals Intelligence Researcher, September 2016 - June 2018
 - Conducted research as part of a select R&D unit in signals intelligence for military service. Discharged as sergeant.

Dropbox, Inc.
Software Engineering Intern, May 2013 - August 2013
 - Built a new internal API for storing and managing contacts.

Google, Inc.
Software Engineering Intern, June 2012 - August 2012
 - Built photos backend pipeline for face tag suggestion in Google+ Events.

Other Research Experience

Cornell University, CS Department
Research Assistant, January 2013 - May 2013
 - Advisor: Ramin Zabih
 - Implemented an efficient variant of max flow in C++ for minimizing sum-of-submodular functions, and applied to binary de-noising and interactive segmentation.

Cornell University, CS Department
Research Assistant, August 2012 - December 2012
 - Advisor: Robert Kleinberg
 - Analyzed LP-based approximation algorithms for the k-max coverage problem in set systems of half spaces and limited VC-dimension.

Academic Service	**Refereeing**

Refereeing

Reviewer, *International Conference on Machine Learning (ICML)*, 2022-24
Reviewer, *Neural Information Processing Systems (NeurIPS)*, 2022-24
Reviewer, *International Conference on Learning Representations (ICLR)*, 2024-25

Organizing

Presenter and organizer, *Tutorial on Data Attribution at Scale (ICML) 2024*
 - 2000+ audience at top ML conference
Organizer, *Workshop on Attributing Model Behavior at Scale (ATTRIB @ NeurIPS)*, 2023-24