Pang Wei Koh

pangwei@cs.washington.edu / www.koh.pw
(My first name is "Pang Wei" and my last name is "Koh".)

Current Position

Assistant Professor, Allen School of Computer Science & Engineering
University of Washington 2023–

Member of Technical Staff
Microsoft AI 2026–

Education

Stanford University 2016–2022
PhD in Computer Science
Thesis: Reliable Machine Learning in the Wild
Advised by Percy Liang

Stanford University 2009–2013
MSc in Computer Science, specializations in AI and Biocomputation
BSc in Computer Science with Honors and Distinction, minor in Mathematics
Thesis: Identifying Genetic Drivers of Cancer Morphology
Advised by Daphne Koller

Work Experience

Research Manager, Allen Institute for AI 2025–2026

Research Scientist, Allen Institute for AI 2024–2025

Senior Research Scientist, Google DeepMind 2022–2023

Research Intern, Calico Life Sciences 2017–2018

Research Analyst, Kundaje Lab, Department of Genetics, Stanford University 2015–2016

Director of Partnerships & Product Manager, Coursera Inc. 2012–2015
Employee #3. Established and led the Partnerships and Course Operations functions at Coursera, building a team of 25 people working with 100+ university partners. Subsequently led product management for all university- and instructor-facing products.

Honors and Awards

Delta Institute's Rising Professors 2026
Best Paper Award, Demo Track, ACL 2025
Google ML and Systems Junior Faculty Award 2025
AI2050 Early Career Fellowship 2025
Singapore AI Visiting Professorship 2024
MIT Technology Review Innovators Under 35, Asia Pacific 2022
Best Paper Award, Applied Data Science Track, KDD 2021

Young NUS Fellow (NUS Development Grant)	2021
Meta PhD Fellowship	2018
Best Paper Award, International Conference in Machine Learning (ICML)	2017
Top 10 Papers of 2016–17 in Regulatory & System Genomics (RECOMB/ISCB)	2017
Best Poster Award, ICML Workshop on Computational Biology	2016
Frederick E. Terman Award *(for overall undergraduate GPA)*	2013
Kennedy Thesis Prize *(for best honors thesis in Stanford Engineering & Applied Sciences)*	2012
Ben Wegbreit Prize *(for best honors thesis in Stanford Computer Science)*	2012
Firestone Medal for Excellence in Research	2012
Overall Winner in CS, The Global Undergraduate Awards *(an international research award)*	2012
Craig and Susan McCaw Scholar *(full scholarship for international students)*	2009

Publications

* = equal contribution / co-first authorship. For more information, see **Google Scholar**.

Designing protein structures from natural language
Joseph Lee, Yanjing Li*, Bo Qiang*, Gregory Lau, Allyson Ettinger, Noah A. Smith, Rory Donovan-Maiye, David Baker, Hanna Hajishirzi, and Pang Wei Koh
arXiv 2026

The COUNT: Counting n-grams unlocks scaling on host memory and training-free adaptation
Ian Magnusson, Noah A. Smith, and Pang Wei Koh
arXiv 2026

Context Language Models
Rulin Shao, Shannon Zejiang Shen, Junjie Oscar Yin, Yuetai Li, Minheng Wang, Hamish Ivison, Radha Poovendran, Nathan Lambert, Teng Xiao, Mike Lewis, Wen-tau Yih, Luke Zettlemoyer, and Pang Wei Koh
arXiv 2026

Gacha Decoding: Eliciting diverse generations through instruction following
Scott Geng, Yufei Zhang, Joseph Lee, Jerry Li, Marjan Ghazvininejad, and Pang Wei Koh
arXiv 2026

Knowledge distillation during mid-training favors reasoning over factual recall
Jacqueline He, Howard Yen, Shuyue Stella Li, Margaret Li, Hanqing Zeng, Yinglong Xia, Benyu Zhang, Zhuokai Zhao, Qiang Zhang, Pang Wei Koh, Luke Zettlemoyer, and Wen-tau Yih
arXiv 2026

FrugalEvo: Towards cost-aware LLM-guided program evolution
Hui Chen, Xuan Qi, James Xu Zhao, Zhaopeng Feng, Shilong Liu, Kuang Xu, Pang Wei Koh, and Bryan Hooi
arXiv 2026

Rubric-based optimization for text-to-music generation
Ping Wang*, Guang Yang*, Shao-Rong Su*, Junkai Wu, Pang Wei Koh, and Noah A. Smith
arXiv 2026

ScholarCatalyst: A benchmark for retrieving papers that inspire new research
Sohyeon Kim*, Yoonho Lee*, Bo Liu, Dayoon Ko, Rulin Shao, Seungone Kim, Graham Neubig, Pang Wei Koh, Aakanksha Chowdhery, Akari Asai, Omar Khattab, Yejin Choi, Gunhee Kim, and Chelsea Finn
arXiv 2026

Efficient test-time adaptation through human-AI interaction
Zora Zhiruo Wang, Apurva Gandhi, Rulin Shao, Aspen Chen, Jonas Mueller, Zhiqi Liang, Jett Chen, Michael Ryan, Qianou Ma, Luxi He, Zhoujun Cheng, Andre He, Seungone Kim, Jiayi Geng, Mingqian Zheng, Weiwei Sun, Zheyuan Zhang, Xinran Zhao, Yike Wang, Abe Hou, Liwei Jiang, Pang Wei Koh, Diyi Yang, Graham Neubig, and Daniel Fried
arXiv 2026

Spokes: Optimizing for diverse pretraining data selection
Clarence Lee, Yejin Choi, Luke Zettlemoyer, Pang Wei Koh, and Hai Leong Chieu
arXiv 2026

JobBench: Aligning agent work with human will
Yuetai Li*, Yichen Feng*, Zhangchen Xu, Zixian Ma, Kaiyuan Zheng, Fengqing Jiang, Xinghua Sun, Rulin Shao, Zichen Chen, Yue Huang, Xinyang Han, Brian Lee, Kayla Xu, Shenglai Zeng, Hang Hua, Xiangliang Zhang, Basel Alomair, Ranjay Krishna, Luke Zettlemoyer, Pang Wei Koh, Bhaskar Ramasubramanian, Luyao Niu, Xiang Yue, and Radha Poovendran
NeurIPS (Evaluations and Datasets Track) 2026

Deep reasoning in general purpose agents via structured meta-cognition
Dean Light*, Michael Theologitis*, Kshitish Ghate*, Shuyue Stella Li, Benjamin Newman, Chirag Shah, Aylin Caliskan, Pang Wei Koh, Dan Suciu, and Yulia Tsvetkov
NeurIPS 2026

DR Tulu: Reinforcement learning with evolving rubrics for deep research
Rulin Shao*, Akari Asai*, Shannon Zejiang Shen*, Hamish Ivison*, Varsha Kishore, Jingming Zhuo, Xinran Zhao, Molly Park, Samuel Finlayson, David Sontag, Tyler Murray, Sewon Min, Pradeep Dasigi, Luca Soldaini, Faeze Brahman, Wen-tau Yih, Tongshuang Wu, Luke Zettlemoyer, Yoon Kim, Hannaneh Hajishirzi, and Pang Wei Koh
ICML 2026
Oral presentation

Spurious rewards: Rethinking training signals in RLVR
Rulin Shao*, Shuyue Stella Li*, Rui Xin*, Scott Geng*, Yiping Wang, Sewoong Oh, Simon Shaolei Du, Nathan Lambert, Sewon Min, Ranjay Krishna, Yulia Tsvetkov, Hannaneh Hajishirzi, Pang Wei Koh, and Luke Zettlemoyer
ICML 2026

RLVE: Scaling up reinforcement learning for language models with adaptive verifiable environments
Zhiyuan Zeng*, Hamish Ivison*, Yiping Wang*, Lifan Yuan*, Shuyue Stella Li, Zhuorui Ye, Siting Li, Jacqueline He, Runlong Zhou, Tong Chen, Chenyang Zhao, Yulia Tsvetkov, Simon Shaolei Du, Natasha Jaques, Hao Peng, Pang Wei Koh, and Hannaneh Hajishirzi
ICML 2026

Anchored decoding: Provably reducing copyright risk for any language model
Jacqueline He, Jonathan Hayase, Wen-tau Yih, Sewoong Oh, Luke Zettlemoyer, and Pang Wei Koh
ICML 2026
Workshop best paper honorable mention at DATA-FM, ICLR 2026

Privasis: Synthesizing the largest "public" private dataset from scratch
Hyunwoo Kim, Niloofar Mireshghallah, Michael Duan, Rui Xin, Shuyue Stella Li, Jaehun Jung, David Acuna, Qi Pang, Hanshen Xiao, G. Edward Suh, Sewoong Oh, Yulia Tsvetkov, Pang Wei Koh, and Yejin Choi
ICML 2026

Cold-start personalization via training-free priors from structured world models
Avinandan Bose*, Shuyue Stella Li*, Faeze Brahman, Pang Wei Koh, Simon Shaolei Du, Yulia Tsvetkov, Maryam Fazel, Lin Xiao, and Asli Celikyilmaz
ICML 2026

FabScore: Fine-Grained Evaluation of Fabrications in Automated AI Research
Hui Chen, James Xu Zhao, Dongfu Jiang, Qianyun Guo, Jiefeng Chen, Yiwei Wang, Muhao Chen, See-Kiong Ng, Pang Wei Koh, and Bryan Hooi
ICML Workshop on AI for Science 2026
Workshop best paper at AI4Science, ICML 2026

MAI-Thinking-1: Building a Hill-Climbing Machine
The Microsoft AI Team
Microsoft AI 2026

OpenScholar: Synthesizing scientific literature with retrieval-augmented LMs
Akari Asai, Jacqueline He*, Rulin Shao*, Weijia Shi, Amanpreet Singh, Joseph Chee Chang, Kyle Lo, Luca Soldaini, Sergey Feldman, Mike D'arcy, David Wadden, Matt Latzke, Minyang Tian, Pan Ji, Shengyan Liu, Hao Tong, Bohao Wu, Yanyu Xiong, Luke Zettlemoyer, Graham Neubig, Dan Weld, Doug Downey, Wen-tau Yih, Pang Wei Koh, and Hannaneh Hajishirzi
Nature 2026

Enabling federated discovery leveraging agentic AI systems: The Cancer AI Alliance and DataVoyager.
Simone E. Dekker, Stephen Salerno, Bodhisattwa P. Majumder, Reece Adamson, Rory M. Donovan-Maiye, Pang Wei Koh, Lei Deng, Keith D. Eaton, Srinivasan Yegnasubramanian, Kenneth L. Kehl, Justin Jee, Brian M. Bot, Jeff T. Leek, and Cancer AI Alliance (CAIA)
Journal of Clinical Oncology 2026

Transforming the cytokine literature into a resource for experimental analysis and discovery
Lukas Oesinghaus*, Molly Park*, Rulin Shao, Pang Wei Koh, and Georg Seelig
bioRxiv 2026

EvoLM: Self-evolving language models through co-evolved discriminative rubrics
Shuyue Stella Li, Rui Xin, Teng Xiao, Yike Wang, Rulin Shao, Zoey Hao, Melanie Sclar, Sewoong Oh, Faeze Brahman, Pang Wei Koh, and Yulia Tsvetkov
arXiv 2026

PanCanBench: A comprehensive benchmark for evaluating large language models in pancreatic oncology
Yimin Zhao, Sheela R. Damle, Simone E. Dekker, Scott Geng, Karly Williams Silva, Jesse J Hubbard, Manuel F Fernandez, Fatima Zelada-Arenas, Alejandra Alvarez, Brianne Flores, Alexis Rodriguez, Stephen Salerno, Carrie Wright, Zihao Wang, Pang Wei Koh, and Jeffrey T. Leek
arXiv 2026

ALMo: Interactive Aim-Limit-defined, Multi-Objective system for personalized high-dose-rate brachytherapy treatment planning and visualization for cervical cancer
Edward Chen, Natalie Dullerud, Pang Wei Koh, Thomas Niedermayr, Elizabeth Kidd, Sanmi Koyejo, and Carlos Guestrin
CHIL 2026

Buy versus build an LLM: A decision framework for governments
Jiahao Lu, Ziwei Xu, William Tjhi, Junnan Li, Antoine Bosselut, Pang Wei Koh, and Mohan Kankanhalli
ACM Technology Policy Council TechBrief 2026

A false sense of privacy: Evaluating textual data sanitization beyond surface-level privacy leakage
Rui Xin*, Niloofar Mireshghallah*, Shuyue Stella Li, Michael Duan, Hyunwoo Kim, Yejin Choi, Yulia Tsvetkov, Sewoong Oh, and Pang Wei Koh
SatML 2026

PrefDisco: Benchmarking Proactive Personalized Reasoning
Shuyue Stella Li, Avinandan Bose, Faeze Brahman, Simon Shaolei Du, Pang Wei Koh, Maryam Fazel, and Yulia Tsvetkov
ICLR 2026

Frustratingly simple retrieval improves challenging, reasoning-intensive benchmarks
Xinxi Lyu, Michael Duan, Rulin Shao, Pang Wei Koh, and Sewon Min
ICLR 2026

Modeling multi-objective tradeoffs with monotonic utility functions
Edward Chen, Natalie Dullerud, Thomas Niedermayr, Elizabeth Kidd, Ransalu Senanayake, Pang Wei Koh, Sanmi Koyejo, and Carlos Guestrin
AISTATS 2026

Reliable and responsible foundation models
Xinyu Yang, Junlin Han, Rishi Bommasani, Jinqi Luo, Wenjie Qu, Wangchunshu Zhou, Adel Bibi, Xiyao Wang, Jaehong Yoon, Elias Stengel-Eskin, Shengbang Tong, Lingfeng Shen, Rafael Rafailov, Runjia Li, Zhaoyang Wang, Yiyang Zhou, Chenhang Cui, Yu Wang, Wenhao Zheng, Huichi Zhou, Jindong Gu, Zhaorun Chen, Peng Xia, Tony Lee, Thomas Zollo, Vikash Sehwag, Jixuan Leng, Jiuhai Chen, Yuxin Wen, Huan Zhang, Zhun Deng, Linjun Zhang, Pavel Izmailov, Pang Wei Koh, Yulia Tsvetkov, Andrew Wilson, Jiaheng Zhang, James Zou, Cihang Xie, Hao Wang, Philip Torr, Julian McAuley, David Alvarez-Melis, Florian Tramèr, Kaidi Xu, Suman Jana, Chris Callison-Burch, Rene Vidal, Filippos Kokkinos, Mohit Bansal, Beidi Chen, and Huaxiu Yao
TMLR 2025

Olmo 3
Team Olmo, Allyson Ettinger, Amanda Bertsch, Bailey Kuehl, David Graham, David Heineman, Dirk Groeneveld, Faeze Brahman, Finbarr Timbers, Hamish Ivison, Jacob Morrison, Jake Poznanski, Kyle Lo, Luca Soldaini, Matt Jordan, Mayee Chen, Michael Noukhovitch, Nathan Lambert, Pete Walsh, Pradeep Dasigi, Robert Berry, Saumya Malik, Saurabh Shah, Scott Geng, Shane Arora, Shashank Gupta, Taira Anderson, Teng Xiao, Tyler Murray, Tyler Romero, Victoria Graf, Akari Asai, Akshita Bhagia, Alexander Wettig, Alisa Liu, Aman Rangapur, Chloe Anastasiades, Costa Huang, Dustin Schwenk, Harsh Trivedi, Ian Magnusson, Jaron Lochner, Jiacheng Liu, Lester James V. Miranda, Maarten Sap, Malia Morgan, Michael Schmitz, Michal Guerquin, Michael Wilson, Regan Huff, Ronan Le Bras, Rui Xin, Rulin Shao, Sam Skjonsberg, Shannon Zejiang Shen, Shuyue Stella Li, Tucker Wilde, Valentina Pyatkin, Will Merrill, Yapei Chang, Yuling Gu, Zhiyuan Zeng, Ashish Sabharwal, Luke Zettlemoyer, Pang Wei Koh, Ali Farhadi, Noah A. Smith, and Hannaneh Hajishirzi
arXiv 2025

Self-improving VLM Judges without human annotations
Inna Wanyin Lin, Yushi Hu, Shuyue Stella Li, Scott Geng, Pang Wei Koh, Luke Zettlemoyer, Tim Althoff, and Marjan Ghazvininejad
arXiv 2025

FlexOlmo: Open language models for flexible data use
Weijia Shi, Akshita Bhagia, Kevin Farhat, Niklas Muennighoff, Pete Walsh, Jacob Morrison, Dustin Schwenk, Shayne Longpre, Jake Poznanski, Allyson Ettinger, Daogao Liu, Margaret Li, Dirk Groeneveld, Mike Lewis, Wen-tau Yih, Luca Soldaini, Kyle Lo, Noah A Smith, Luke Zettlemoyer, Pang Wei Koh, Hannaneh Hajishirzi, Ali Farhadi, and Sewon Min
NeurIPS 2025
Spotlight paper

Precise information control in long-form text generation
Jacqueline He, Howard Yen, Margaret Li, Shuyue Stella Li, Zhiyuan Zeng, Weijia Shi, Yulia Tsvetkov, Danqi Chen, Pang Wei Koh, and Luke Zettlemoyer
NeurIPS 2025

HybridCoT: Interleaving latent and text chain-of-thought for efficient reasoning
Shannon Zejiang Shen, Rulin Shao, Chenyu Wang, Songlin Yang, Vincent-Pierre Berges, Gargi Ghosh, Pang Wei Koh, Luke Zettlemoyer, Yoon Kim, Jason E Weston, David Sontag, and Wen-tau Yih
NeurIPS Workshop on Efficient Reasoning 2025
Spotlight paper

The Delta Learning hypothesis: Preference tuning on weak data can yield strong gains
Scott Geng, Hamish Ivison, Chun-Liang Li, Maarten Sap, Jerry Li, Ranjay Krishna, and Pang Wei Koh
COLM 2025

EvalTree: Profiling language model weaknesses via hierarchical capability trees
Zhiyuan Zeng, Yizhong Wang, Hannaneh Hajishirzi, and Pang Wei Koh
COLM 2025

ReasonIR: Training Retrievers for Reasoning Tasks
Rulin Shao*, Rui Qiao*, Varsha Kishore, Niklas Muennighoff, Xi Victoria Lin, Daniela Rus, Bryan Kian Hsiang Low, Sewon Min, Wen-tau Yih, Pang Wei Koh, and Luke Zettlemoyer
COLM 2025

Establishing task scaling laws via compute-efficient model ladders
Akshita Bhagia, Jiacheng Liu, Alexander Wettig, David Heineman, Oyvind Tafjord, Ananya Harsh Jha, Luca Soldaini, Noah A Smith, Dirk Groeneveld, Pang Wei Koh, Jesse Dodge, and Hannaneh Hajishirzi
COLM 2025

ParaPO: Aligning language models to reduce verbatim reproduction of pre-training data
Tong Chen, Faeze Brahman, Jiacheng Liu, Niloofar Mireshghallah, Weijia Shi, Pang Wei Koh, Luke Zettlemoyer, and Hannaneh Hajishirzi
COLM 2025

2 OLMo 2 Furious
Team OLMo, Pete Walsh, Luca Soldaini, Dirk Groeneveld, Kyle Lo, Shane Arora, Akshita Bhagia, Yuling Gu, Shengyi Huang, Matt Jordan, Nathan Lambert, Dustin Schwenk, Oyvind Tafjord, Taira Anderson, David Atkinson, Faeze Brahman, Christopher Clark, Pradeep Dasigi, Nouha Dziri, Michal Guerquin, Hamish Ivison, Pang Wei Koh, Jiacheng Liu, Saumya Malik, William Merrill, Lester James V. Miranda, Jacob Morrison, Tyler Murray, Crystal Nam, Valentina Pyatkin, Aman Rangapur, Michael Schmitz, Sam Skjonsberg, David Wadden, Christopher Wilhelm, Michael Wilson, Luke Zettlemoyer, Ali Farhadi, Noah A. Smith, and Hannaneh Hajishirzi
COLM 2025

Fluid language model benchmarking
Valentin Hofmann, David Heineman, Ian Magnusson, Kyle Lo, Jesse Dodge, Maarten Sap, Pang Wei Koh, Chun Wang, Hannaneh Hajishirzi, and Noah A. Smith
COLM 2025
Oral presentation

The curious case of factuality finetuning: Models' internal beliefs can improve factuality
Benjamin Newman, Abhilasha Ravichander, Jaehun Jung, Rui Xin, Hamish Ivison, Yegor Kuznetsov, Pang Wei Koh, and Yejin Choi
arXiv 2025

DataDecide: How to predict best pretraining data with small experiments
Ian Magnusson*, Nguyen Tai*, Ben Bogin*, David Heineman, Jena D Hwang, Luca Soldaini, Akshita Bhagia, Jiacheng Liu, Dirk Groeneveld, Oyvind Tafjord, Noah A Smith, Pang Wei Koh, and Jesse Dodge
ICML 2025

NICE: Non-differentiable evaluation metric-based data selection for instruction tuning
Jingtan Wang, Xiaoqiang Lin, Rui Qiao, Pang Wei Koh, Chuan-Sheng Foo, and Bryan Kian Hsiang Low
ICML 2025

OLMoTrace: Tracing language model outputs back to trillions of training tokens
Jiacheng Liu, Taylor Blanton, Yanai Elazar, Sewon Min, YenSung Chen, Arnavi Chheda-Kothary, Huy Tran, Byron Bischoff, Eric Marsh, Michael Schmitz, Cassidy Trier, Aaron Sarnat, Jenna James, Jon Borchardt, Bailey Kuehl, Evie Cheng, Karen Farley, Sruthi Sreeram, Taira Anderson, David Albright, Carissa Schoenick, Luca Soldaini, Dirk Groeneveld, Rock Yuren Pang, Pang Wei Koh, Noah A Smith, Sophie Lebrecht, Yejin Choi, Hannaneh Hajishirzi, Ali Farhadi, and Jesse Dodge
ACL 2025
Best demo award

Large-scale data selection for instruction tuning
Hamish Ivison, Muru Zhang, Faeze Brahman, Pang Wei Koh, and Pradeep Dasigi
arXiv 2025

S4S: Solving for a diffusion model solver
Eric Frankel, Sitan Chen, Jerry Li, Pang Wei Koh, Lillian J Ratliff, and Sewoong Oh
ICML 2025

Metabolically purified human stem cell-derived hepatocytes reveal distinct effects of Ebola and Lassa viruses
Joseph B Prescott, Kevin J Liu, Angelika Lander, Nicole Min Qian Pek, Sawan Kumar Jha, Marcel Bokelmann, Manali Begur, Pang Wei Koh, Henry Yang, Bing Lim, Kristy Red-Horse, Irving L Weissman, Kyle M Loh, Lay Teng Ang
bioRxiv 2025

Exploring how generative MLLMs perceive more than CLIP with the same vision encoder
Siting Li, Pang Wei Koh, and Simon Shaolei Du
ACL 2025

OLMoE: Open Mixture-of-Experts language models
Niklas Muennighoff, Luca Soldaini, Dirk Groeneveld, Kyle Lo, Jacob Morrison, Sewon Min, Weijia Shi, Pete Walsh, Oyvind Tafjord, Nathan Lambert, Yuling Gu, Shane Arora, Akshita Bhagia, Dustin Schwenk, David Wadden, Alexander Wettig, Binyuan Hui, Tim Dettmers, Douwe Kiela, Ali Farhadi, Noah A Smith, Pang Wei Koh, Amanpreet Singh, and Hannaneh Hajishirzi
ICLR 2025

Language models scale reliably with over-training and on downstream tasks
Samir Yitzhak Gadre, Georgios Smyrnis, Vaishaal Shankar, Suchin Gururangan, Mitchell Wortsman, Rulin Shao, Jean Mercat, Alex Fang, Jeffrey Li, Sedrick Keh, Rui Xin, Marianna Nezhurina, Igor Vasiljevic, Jenia Jitsev, Luca Soldaini, Alexandros G. Dimakis, Gabriel Ilharco, Pang Wei Koh, Shuran Song, Thomas Kollar, Yair Carmon, Achal Dave, Reinhard Heckel, Niklas Muennighoff, and Ludwig Schmidt
ICLR 2025

Group-robust sample reweighting for subpopulation shifts via influence functions
Rui Qiao, Zhaoxuan Wu, Jingtan Wang, Pang Wei Koh, and Bryan Kian Hsiang Low
ICLR 2025

Use large language models to promote equity
Emma Pierson*, Divya Shanmugam*, Rajiv Movva*, Jon Kleinberg*, Monica Agrawal, Mark Dredze, Kadija Ferryman, Judy Wawira Gichoya, Dan Jurafsky, Pang Wei Koh, Karen Levy, Sendhil Mullainathan, Ziad Obermeyer, Harini Suresh, and Keyon Vafa
NEJM AI 2025

ICONS: Influence Consensus for Vision-Language Data Selection
Xindi Wu, Mengzhou Xia, Rulin Shao, Zhiwei Deng, Pang Wei Koh, and Olga Russakovsky
arXiv 2025

PLeaS–Merging models with permutations and least squares
Anshul Nasery, Jonathan Hayase, Pang Wei Koh, and Sewoong Oh
CVPR 2025

Negative Token Merging: Image-based adversarial feature guidance
Jaskirat Singh, Lindsey Li, Weijia Shi, Ranjay Krishna, Yejin Choi, Pang Wei Koh, Michael F Cohen, Stephen Gould, Liang Zheng, and Luke Zettlemoyer
arXiv 2024

Scaling retrieval-based language models with a trillion-token datastore
Rulin Shao, Jacqueline He, Akari Asai, Weijia Shi, Tim Dettmers, Sewon Min, Luke Zettlemoyer, and Pang Wei Koh
NeurIPS 2024

The unmet promise of synthetic training images: Using retrieved real images performs better
Scott Geng, Cheng-Yu Hsieh, Vivek Ramanujan, Matthew Wallingford, Chun-Liang Li, Pang Wei Koh, and Ranjay Krishna
NeurIPS 2024

MEDIQ: Question-asking LLMs for adaptive and reliable clinical reasoning
Shuyue Stella Li, Vidhisha Balachandran, Shangbin Feng, Jonathan Ilgen, Emma Pierson, Pang Wei Koh, and Yulia Tsvetkov
NeurIPS 2024

Uncertainty of Thoughts: Uncertainty-aware planning enhances information seeking in large language models
Zhiyuan Hu, Chumin Liu, Xidong Feng, Yilun Zhao, See-Kiong Ng, Anh Tuan Luu, Junxian He, Pang Wei Koh, and Bryan Hooi
NeurIPS 2024

Multilingual diversity improves vision-language representations
Thao Nguyen, Matthew Wallingford, Sebastin Santy, Wei-Chiu Ma, Sewoong Oh, Ludwig Schmidt, Pang Wei Koh, and Ranjay Krishna
NeurIPS 2024

DataComp-LM: In search of the next generation of training sets for language models
Jeffrey Li, Alex Fang, Georgios Smyrnis, Maor Ivgi, Matt Jordan, Samir Gadre, Hritik Bansal, Etash Guha, Sedrick Keh, Kushal Arora, Saurabh Garg, Rui Xin, Niklas Muennighoff, Reinhard Heckel, Jean Mercat, Mayee Chen, Suchin Gururangan, Mitchell Wortsman, Alon Albalak, Yonatan Bitton, Marianna Nezhurina, Amro Abbas, Cheng-Yu Hsieh, Dhruba Ghosh, Josh Gardner, Maciej Kilian, Hanlin Zhang, Rulin Shao, Sarah Pratt, Sunny Sanyal, Gabriel Ilharco, Giannis Daras, Kalyani Marathe, Aaron Gokaslan, Jieyu Zhang, Khyathi Chandu, Thao Nguyen, Igor Vasiljevic, Sham Kakade, Shuran Song, Sujay Sanghavi, Fartash Faghri, Sewoong Oh, Luke Zettlemoyer, Kyle Lo, Alaaeldin El-Nouby, Hadi Pouransari, Alexander Toshev, Stephanie Wang, Dirk Groeneveld, Luca Soldaini, Pang Wei Koh, Jenia Jitsev, Thomas Kollar, Alexandros G. Dimakis, Yair Carmon, Achal Dave, Ludwig Schmidt, and Vaishaal Shankar
NeurIPS D&B 2024

JPEG-LM: LLMs as image generators with canonical codec representations
Xiaochuang Han, Marjan Ghazvininejad, Pang Wei Koh, and Yulia Tsvetkov
arXiv 2024

CopyBench: Measuring literal and non-literal reproduction of copyright-protected text in language model generation
Tong Chen, Akari Asai, Niloofar Mireshghallah, Sewon Min, James Grimmelmann, Yejin Choi, Hannaneh Hajishirzi, Luke Zettlemoyer, and Pang Wei Koh
EMNLP 2024

Data-centric AI in the age of large language models
Xinyi Xu, Zhaoxuan Wu, Rui Qiao, Arun Verma, Yao Shu, Jingtan Wang, Xinyuan Niu, Zhenfeng He, Jiangwei Chen, Zijian Zhou, Gregory Kang Ruey Lau, Hieu Dao, Lucas Agussurja, Rachael Hwee Ling Sim, Xiaoqiang Lin, Wenyang Hu, Zhongxiang Dai, Pang Wei Koh, and Bryan Kian Hsiang Low
EMNLP Findings 2024

Merge to Learn: Efficiently Adding Skills to Language Models with Model Merging
Jacob Morrison, Noah A. Smith, Hannaneh Hajishirzi, Pang Wei Koh, Jesse Dodge, Pradeep Dasigi
EMNLP Findings 2024

Annotation alignment: Comparing LLM and human annotations of conversational safety
Rajiv Movva, Pang Wei Koh, and Emma Pierson
EMNLP 2024

Information-theoretic distillation for reference-less summarization
Jaehun Jung, Ximing Lu, Liwei Jiang, Faeze Brahman, Peter West, Pang Wei Koh, and Yejin Choi
COLM 2024

Using unlabeled data to enhance fairness of medical AI
Rajiv Movva, Pang Wei Koh, and Emma Pierson
Nature Medicine 2024

Reliable, adaptable, and attributable language models with retrieval
Akari Asai, Zexuan Zhong, Danqi Chen, Pang Wei Koh, Luke Zettlemoyer, Hannaneh Hajishirzi, and Wen-tau Yih
arXiv 2024

Instructional fingerprinting of large language models
Jiashu Xu, Fei Wang, Mingyu Derek Ma, Pang Wei Koh, Chaowei Xiao, and Muhao Chen
NAACL 2024

The generative AI paradox: "What it can create, it may not understand"
Peter West*, Ximing Lu*, Nouha Dziri*, Faeze Brahman*, Linjie Li*, Jena D. Hwang, Liwei Jiang, Jillian Fisher, Abhilasha Ravichander, Khyathi Chandu, Benjamin Newman, Pang Wei Koh, Allyson Ettinger, and Yejin Choi
ICLR 2024

Leveraging domain relations for domain generalization
Huaxiu Yao*, Xinyu Yang*, Xinyi Pan, Shengchao Liu, Pang Wei Koh, and Chelsea Finn
ICLR 2024
Spotlight paper

Impossibility theorems for feature attribution
Blair Bilodeau, Natasha Jaques, Pang Wei Koh, and Been Kim
PNAS 2024

Retrieval-based language models using a multi-domain datastore
Rulin Shao, Sewon Min, Luke Zettlemoyer, and Pang Wei Koh
NeurIPS DistShift 2023

OpenFlamingo: An open-source framework for training large autoregressive vision-language models
Anas Awadalla*, Irena Gao*, Josh Gardner, Jack Hessel, Yusuf Hanafy, Wanrong Zhu, Kalyani Marathe, Yonatan Bitton, Samir Gadre, Shiori Sagawa, Jenia Jitsev, Simon Kornblith, Pang Wei Koh, Gabriel Ilharco, Mitchell Wortsman, and Ludwig Schmidt
arXiv 2023

FActScore: Fine-grained atomic evaluation of factual precision in long form text generation
Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi
EMNLP 2023

DataComp: In search of the next generation of multimodal datasets
Samir Yitzhak Gadre*, Gabriel Ilharco*, Alex Fang*, Jonathan Hayase, Georgios Smyrnis, Thao Nguyen, Ryan Marten, Mitchell Wortsman, Dhruba Ghosh, Jieyu Zhang, Eyal Orgad, Rahim Entezari, Giannis Daras, Sarah Pratt, Vivek Ramanujan, Yonatan Bitton, Kalyani Marathe, Stephen Mussmann, Richard Vencu, Mehdi Cherti, Ranjay Krishna, Pang Wei Koh, Olga Saukh, Alexander Ratner, Shuran Song, Hannaneh Hajishirzi, Ali Farhadi, Romain Beaumont, Sewoong Oh, Alex Dimakis, Jenia Jitsev, Yair Carmon, Vaishaal Shankar, and Ludwig Schmidt
NeurIPS D&B 2023
Oral presentation

Proximity-informed calibration for deep neural networks
Miao Xiong, Ailin Deng, Pang Wei Koh, Jiaying Wu, Shen Li, Jianqing Xu, and Bryan Hooi
NeurIPS 2023
Spotlight paper

Are aligned neural networks adversarially aligned?
Nicholas Carlini, Milad Nasr, Christopher A Choquette-Choo, Matthew Jagielski, Irena Gao, Anas Awadalla, Pang Wei Koh, Daphne Ippolito, Katherine Lee, Florian Tramer, and Ludwig Schmidt
NeurIPS 2023

On the trade-off of intra-/inter-class diversity for supervised pre-training
Jieyu Zhang*, Bohan Wang*, Zhengyu Hu, Pang Wei Koh, and Alexander Ratner
NeurIPS 2023

Out-of-distribution robustness via targeted augmentations
Irena Gao*, Shiori Sagawa*, Pang Wei Koh, Tatsunori Hashimoto, and Percy Liang
ICML 2023

Wild-Time: A benchmark of in-the-wild distribution shift over time
Huaxiu Yao*, Caroline Choi*, Yoonho Lee, Pang Wei Koh, and Chelsea Finn
NeurIPS D&B 2022

Extending the WILDS benchmark for unsupervised adaptation
Shiori Sagawa*, Pang Wei Koh*, Tony Lee*, Irena Gao*, Sang Michael Xie, Kendrick Shen, Ananya Kumar, Weihua Hu, Michihiro Yasunaga, Henrik Marklund, Sara Beery, Etienne David, Ian Stavness, Wei Guo, Jure Leskovec, Kate Saenko, Tatsunori Hashimoto, Sergey Levine, Chelsea Finn, and Percy Liang
ICLR 2022
Oral presentation

WILDS: A benchmark of in-the-wild distribution shifts
Pang Wei Koh*, Shiori Sagawa*, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton A. Earnshaw, Imran S. Haque, Sara Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang
ICML 2021
Oral presentation

Just Train Twice: Improving group robustness without training group information
Evan Zheran Liu*, Behzad Haghgoo*, Annie S. Chen*, Aditi Raghunathan, Pang Wei Koh, Shiori Sagawa, Percy Liang, and Chelsea Finn
ICML 2021
Oral presentation

Accuracy on the line: On the strong correlation between out-of-distribution and in-distribution generalization
John Miller, Rohan Taori, Aditi Raghunathan, Shiori Sagawa, Pang Wei Koh, Vaishaal Shankar, Percy Liang, Yair Carmon, and Ludwig Schmidt
ICML 2021

Supporting COVID-19 policy response with large-scale mobility-based modeling
Serina Chang, Mandy L. Wilson, Bryan Lewis, Zakaria Mehrab, Komal K. Dudakiya, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, Madhav Marathe, Jure Leskovec
KDD 2021
Best paper award (Applied Data Science track)

On the opportunities and risks of foundation models
Rishi Bommasani, Drew A. Hudson, ..., Pang Wei Koh, ..., and Percy Liang (116 authors, alphabetical within ellipses)
arXiv 2021

Selective classification can magnify disparities across groups
Erik Jones*, Shiori Sagawa*, Pang Wei Koh*, Ananya Kumar, and Percy Liang
ICLR 2021
Spotlight talk at the NeurIPS 2020 ICBINB Workshop

Stronger data poisoning attacks break data sanitization defenses
Pang Wei Koh*, Jacob Steinhardt*, and Percy Liang
Machine Learning 2021

Mobility network models of COVID-19 explain inequities and inform reopening
Serina Y Chang*, Emma Pierson*, Pang Wei Koh*, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec
Nature 2021
Accompanying Nature News and Views; interactive article in The New York Times; other coverage by The New York Times; The Washington Post; The Telegraph; Bloomberg; CNN; MIT Technology Review; Wired; STAT; and Stanford News. Also presented at NetSci 2021 (**oral presentation**) and the NeurIPS 2020 COVID-19 Symposium (**invited talk**). See project webpage for data and more press coverage.

Concept bottleneck models
Pang Wei Koh*, Thao Nguyen*, Yew Siang Tang*, Steve Mussmann, Emma Pierson, Been Kim, and Percy Liang
ICML 2020
Spotlight talk at the ICML 2020 Workshop on Human Interpretability in Machine Learning

An investigation of why overparameterization exacerbates spurious correlations
Shiori Sagawa*, Aditi Raghunathan*, Pang Wei Koh*, and Percy Liang
ICML 2020

ExpBERT: Representation engineering with natural language explanations
Shikhar Murty, Pang Wei Koh, and Percy Liang
ACL 2020

Toward trustworthy AI development: Mechanisms for supporting verifiable claims
Miles Brundage*, Shahar Avin*, Jasmine Wang*, Haydn Belfield*, Gretchen Krueger*, Gillian Hadfield, Heidy Khlaaf, Jingying Yang, Helen Toner, Ruth Fong, Tegan Maharaj, Pang Wei Koh, Sara Hooker, ..., Thomas Krendl Gilbert, Lisa Dyer, Saif Khan, Yoshua Bengio, and Markus Anderljung
arXiv 2020

Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization
Shiori Sagawa*, Pang Wei Koh*, Tatsunori B. Hashimoto, and Percy Liang
ICLR 2020

On the accuracy of influence functions for measuring group effects
Pang Wei Koh*, Kai-Siang Ang*, Hubert H. K. Teo*, and Percy Liang
NeurIPS 2019

Temporal FiLM: Capturing long-range sequence dependencies with feature-wise modulations
Sawyer Birnbaum*, Volodymyr Kuleshov*, Zayd Enam, Pang Wei Koh, Stefano Ermon
NeurIPS 2019

Inferring multi-dimensional rates of aging from cross-sectional data
Emma Pierson*, Pang Wei Koh*, Tatsunori B. Hashimoto*, Daphne Koller, Jure Leskovec, Nicholas Eriksson, and Percy Liang
AISTATS 2019
Contributed talk at the ICML/IJCAI 2018 Workshop on Computational Biology
Spotlight talk at the NeurIPS 2018 Workshop on Machine Learning for Health

Certified defenses for data poisoning attacks
Jacob Steinhardt*, Pang Wei Koh*, and Percy Liang
NeurIPS 2017

Understanding black-box predictions via influence functions
Pang Wei Koh and Percy Liang
ICML 2017
Best paper award

Localized hepatic lobular regeneration by central-vein-associated lineage-restricted progenitors
Jonathan M. Tsai, Pang Wei Koh, Ania Stefanska, Liujing Xing, Graham G. Walmsley, Nicolas Poux, Irving L. Weissman, and Yuval Rinkevich
PNAS 2017

Denoising genome-wide histone ChIP-seq with convolutional neural networks
Pang Wei Koh*, Emma Pierson*, and Anshul Kundaje
Bioinformatics 2017
Spotlight talk and **best poster award** at the ICML 2016 Workshop on Computational Biology
Top 10 papers of 2016-2017 in regulatory and systems genomics at RECOMB/ISMB

An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development
Pang Wei Koh*, Rahul Sinha*, Amira A. Barkal, Rachel M. Morganti, Angela Chen, Irving L. Weissman, Lay Teng Ang, Anshul Kundaje, and Kyle M. Loh
Scientific Data 2016

Mapping the pairwise choices leading from pluripotency to human bone, heart, and other mesoderm cell types
Kyle M. Loh*, Angela Chen*, Pang Wei Koh, Tianda Z. Deng, Rahul Sinha, Jonathan M. Tsai, Amira A. Barkal, Kimberle Y. Shen, Rajan Jain, Rachel M. Morganti, Ng Shyh-Chang, Nathaniel B. Fernhoff, Benson M. George, Gerlinde Wernig, Rachel E.A. Salomon, Zhenghao Chen, Hannes Vogel, Jonathan A. Epstein, Anshul Kundaje, William S. Talbot, Philip A. Beachy, Lay Teng Ang, and Irving L. Weissman
Cell 2016

Dissecting an online intervention for cancer survivors
Zhenghao Chen, Pang Wei Koh, Philip L. Ritter, Kate Lorig, Erin O'Carroll Bantum, and Suchi Saria
Health Ed. & Behavior 2014

Peer and self assessment in massive online classes
Chinmay Kulkarni, Pang Wei Koh, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott Klemmer
TOCHI 2013

Identifying genetic drivers of cancer morphology
Pang Wei Koh, Andrew Beck, and Daphne Koller.
Undergraduate honors thesis 2012
Firestone Medal for Excellence in Research
Ben Wegbreit Prize for best undergraduate honors thesis in Stanford Computer Science
Kennedy Thesis Prize for best undergraduate honors thesis in Stanford Engineering & Applied Sciences
Overall Winner, Computer Science, The Global Undergraduate Awards.

Sparse filtering
Jiquan Ngiam, Pang Wei Koh, Zhenghao Chen, Sonia Bhaskar, and Andrew Y. Ng
NeurIPS 2011
Spotlight paper

Learning deep energy models
Jiquan Ngiam, Zhenghao Chen, Pang Wei Koh, and Andrew Y. Ng
ICML 2011

On random weights and unsupervised feature learning
Andrew Saxe, Pang Wei Koh, Zhenghao Chen, Maneesh Bhand, Bipin Suresh, and Andrew Y. Ng
ICML 2011

Tiled convolutional neural networks
Quoc V. Le, Jiquan Ngiam, Zhenghao Chen, Daniel Chia, Pang Wei Koh, and Andrew Y. Ng
NeurIPS 2010

Lower bound on the time complexity of local adiabatic evolution
Zhenghao Chen, Pang Wei Koh, and Zhao Yan
Physical Review A 2006

Advising

PhD students

Scott Geng (co-advised with Ranjay Krishna)	2023–
Jacqueline He (co-advised with Luke Zettlemoyer)	2023–
Rulin Shao (co-advised with Luke Zettlemoyer)	2023–
Rui Xin (co-advised with Sewoong Oh)	2023–
Ian Magnusson (co-advised with Noah Smith)	2024–
Zhiyuan Zeng (co-advised with Hannaneh Hajishirzi)	2024–
Joseph Lee (co-advised with Hannaneh Hajishirzi)	2025–

Undergraduates and Master's students

Yufei Zhang (undergraduate)	2025–
Molly Park (undergraduate)	2025–
Irena Gao (undergraduate; now PhD student at Stanford University)	2020–2022
Kendrick Shen (undergraduate; now engineer at Genesis Therapeutics)	2021–2022
Erik Jones (undergraduate; now PhD student at UC Berkeley)	2020–2021
Henrik Marklund (Master's student; now PhD student at Stanford University)	2020–2021
Thao Nguyen (undergraduate; now PhD student at the University of Washington)	2019–2020
Yew-Siang Tang (Master's student; now Senior Software Engineer at You.com)	2019–2020
Hubert Teo (undergraduate; now ML Engineer at Nuro)	2018–2019
Kai-Siang Ang (undergraduate; now ML Engineer at Nuro)	2018–2019

Visiting students

Qiao Rui (PhD student from NUS; now research scientist at Meta)	2024
Gregory Lau (PhD student from NUS)	2025
Jina Kim (undergraduate from KAIST)	2025

Teaching

CSE481M (Machine Learning Capstone), UW	Spring 2026
CSE446/546 (Machine Learning), UW	Autumn 2025
CSE446/546 (Machine Learning), UW	Autumn 2024
CSE599J (Data-centric Machine Learning), UW	Winter 2024

Service

Conferences and workshops

Area chair, COLM	2026
Award committee, ICLR	2026
Reviewer, ICML	2026
Reviewer, ECCV	2026
Reviewer, CVPR	2026
Area chair, ICLR	2026
Area chair, COLM	2025
Organizer, COLM Workshop on Large Language Modeling for Scientific Discovery	2025
Reviewer, CVPR	2025
Organizer, ICLR Workshop on Foundation Models in the Wild	2025
Organizer, ICLR Workshop on Navigating and Addressing Data Problems for Foundation Models	2025

Organizer, ICLR Workshop on Reasoning and Planning for Large Language Models 2025
Area chair, ICLR 2025
Area chair, COLM 2024
Reviewer, ICML Workshops 2024
Organizer, ICLR Workshop on Reliable and Responsible Foundation Models 2024
Organizer, ICLR Workshop on Data Problems for Foundation Models 2024
Organizer, NeurIPS Workshop on Distribution Shifts 2023
Reviewer, Science Advances 2023
Reviewer, ICLR Workshop on Mathematical and Empirical Understanding of Foundation Models 2023
Organizer, NeurIPS Workshop on Distribution Shifts 2022
Organizer, NeurIPS Workshop on Distribution Shifts 2021
Reviewer, AAAI 2021
Reviewer, ICLR 2021
Reviewer, ICLR Workshop on AI for Public Health 2021
Reviewer, ICLR Workshop on Robust and Reliable ML in the Real World 2021
Reviewer, ICML 2021
Reviewer, ICML Workshop on Uncertainty in Deep Learning 2021
Reviewer, NeurIPS 2021
Reviewer, ICLR 2020
Reviewer, ICML 2020
Reviewer, ICML Workshop on Human Interpretability in Machine Learning 2020
Reviewer, ICML Workshop on ML Interpretability for Scientific Discovery 2020
Reviewer, ICLR 2019
Reviewer, ICLR Workshop on Debugging ML Models 2019
Reviewer, ICML 2019
Reviewer, NeurIPS 2019
Reviewer, NeurIPS Workshop on Information Theory and Machine Learning 2019
Reviewer, ICML 2018
Reviewer, NeurIPS 2018
Reviewer, UAI 2018
Reviewer, ICML Workshop on Reliable Machine Learning in the Wild 2017

Journals

Reviewer, Transactions on Pattern Analysis and Machine Intelligence 2023
Reviewer, Transactions on Pattern Analysis and Machine Intelligence 2020
Reviewer, The American Statistician 2019
Reviewer, Transactions on Pattern Analysis and Machine Intelligence 2019
Reviewer, Distill 2018
Reviewer, Journal of Machine Learning Research 2018
Reviewer, ACM Transactions on Computational Biology and Bioinformatics 2017
Reviewer, Journal of Machine Learning Research 2017

Community

Panelist, Joint Mathematics Meetings 2025
Mentor, CURIS Undergraduate Summer Research Program 2021
Mentor, Stanford CS Undergraduate Mentoring Program 2021
Mentor, NeurIPS Workshop on Distribution Shifts Author Mentorship Program 2021
Volunteer, Stanford CS PhD Student Applicant Support Program 2021
Mentor, CURIS Undergraduate Summer Research Program 2020
Mentor, Stanford CS Undergraduate Mentoring Program 2020
Volunteer, Tapia Conference Virtual Booth 2020
Volunteer, Singapore GovTech COVID-19 Response 2020

Mentor, CURIS Undergraduate Summer Research Program	2019
Mentor, Stanford AI Lab Undergraduate Mentoring Program	2019
Mentor, CURIS Undergraduate Summer Research Program	2018
Mentor, Stanford AI Lab Undergraduate Mentoring Program	2018

Invited Talks

Academic talks

CHIL	2026
Fred Hutch Cancer Center, Faculty Retreat	2026
Lorong AI, Singapore	2026
AAAI Workshop on Deployable AI	2026
AAAI Workshop on Reliability of Foundation Models in the Real World	2026
Panelist, LLM Localization: Challenges and Opportunities, IAAI	2026
AI Institute Symposium on Agentic AI and Multiagent Systems, National University of Singapore	2026
AI Seminar, Oregon State University	2026
Microsoft Research Redmond	2025
NLP Seminar, Stanford University	2025
NLP Seminar, Massachusetts Institute of Technology	2025
Department of Computer Science, Harvard University	2025
BAIR NLP Workshop, UC Berkeley	2025
Institute for Foundations of Data Science Seminar, University of Washington	2025
ICML DataWorld Workshop: Unifying Data Curation Frameworks Across Domains	2025
ICML Workshop on Long Context Foundation Models	2025
Sloan Precision Oncology Institute AI Symposium, Fred Hutchinson Cancer Center	2025
UW-RIKEN AIP Joint Workshop	2025
RAISE.SG	2024
Fred Hutchinson Cancer Center	2024
University of Tsukuba	2024
RIKEN Center for Advanced Intelligence Project (AIP)	2024
AI Horizons Talk, National University of Singapore	2024
DSO National Labs, Singapore	2024
NLP Workshop, Korea Advanced Institute of Science & Technology (KAIST)	2024
ICML Workshop on Foundation Models in the Wild	2024
Large Model Safety Workshop, Singapore	2024
UK AI Safety Institute	2024
Department of Statistics, University of Oxford	2024
Panelist, Trustworthy ML Symposium	2022
Institute for Mathematical Sciences, National University of Singapore	2022
Brain Team, Google Research	2022
Adaptive Systems and Interaction Group, Microsoft Research	2022
Sea AI Lab	2022
Department of Computer Science, Cornell Tech	2022
Department of Computer Science, Cornell University	2022
School of Computer and Communication Sciences, EPFL	2022
Department of Computer Science, ETH Zurich	2022
School of Computer Science, McGill University	2022
School of Computer Science and Engineering, Nanyang Technological University	2022
Department of Computer Science, National University of Singapore	2022
Department of Computer Science, Princeton University	2022
Departments of Computer Science and Electrical and Computer Engineering, University of Toronto	2022
Allen School of Computer Science and Engineering, University of Washington	2022

Department of Statistics and Data Science, Yale University	2022
Machine Learning Symposium, University of Southern California	2021
Workshops in Biostatistics (BIODS/STATS 260) guest lecture, Stanford University	2021
Interpretability and Explainability in ML (COMPSCI 282BR) guest lecture, Harvard University	2021
Panelist, ICML Workshop on Human Interpretability	2020
Microsoft Research AI Breakthroughs	2020
Faculty Lunch, Computer Science Department, Stanford University	2020
Department of Computer Science and Engineering, Ohio State University	2020
Department of Statistics and Data Science, Yale University	2020
AAAI Spring Symposium: Interpretable AI for Well-Being	2019
School of Computer Science and Engineering, Nanyang Technological University	2019
AAAI Spring Symposium: Beyond Machine Intelligence	2018
Security and Fairness of Deep Learning (18-739) guest lecture, CMU Silicon Valley	2018
School of Computing, National University of Singapore	2018
Institute for Infocomm Research, Singapore	2018
Machine Learning Group, Massachusetts Institute of Technology	2017
Discrete Algorithms Group, Google	2017
ICML Workshop on Human Interpretability	2017
Machine Learning Group, University of Cambridge	2017
Microsoft Research Cambridge	2017
Alan Turing Institute, London	2017
Institute of Molecular and Cell Biology, Singapore	2016

Talks and workshops on online education

Infocomm Development Authority of Singapore	2014
Johns Hopkins University	2014
University of Illinois at Urbana-Champaign	2014
University of Maryland at College Park	2014
University of Pennsylvania	2014
Princeton University	2014
Annual American Dental Education Association Deans' Conference, Savannah	2013
Association of Academic Health Centers Annual Meeting, Boston	2013
Association of Schools of Allied Health Professions Spring Meeting, San Diego	2013
Emory University	2013
European MOOC Summit, EPFL	2013
European MOOCs in a Global Context Workshop, University of Wisconsin-Madison	2013
Georgia Tech	2013
Ministry of Education, Singapore	2013
National University of Singapore	2013
Ohio Higher Education Trustees Conference, Columbus	2013
The Pennsylvania State University	2013
University of Pittsburgh	2013
Vanderbilt University	2013