Pang Wei Koh

pangwei@cs.washington.edu / **www.koh.pw**

A note on my name: 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–

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
Thesis: Identifying Genetic Drivers of Cancer Morphology
Advised by Daphne Koller

Work Experience

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

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.

Retrieval-based Language Models Using a Multi-domain Datastore NeurIPS DistShift 2023
Rulin Shao, Sewon Min, Luke Zettlemoyer, and Pang Wei Koh

The Generative AI Paradox: "What It Can Create, It May Not Understand" arXiv 2023
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

OpenFlamingo: An open-source framework for training large autoregressive vision-language models arXiv 2023
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

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

DataComp: In search of the next generation of multimodal datasets NeurIPS D&B 2023
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
Oral presentation

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

Are aligned neural networks adversarially aligned? NeurIPS 2023
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

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

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

Impossibility theorems for feature attribution PNAS 2023
Blair Bilodeau, Natasha Jaques, Pang Wei Koh, and Been Kim

Leveraging domain relations for domain generalization arXiv 2023
Huaxiu Yao*, Xinyu Yang*, Xinyi Pan, Shengchao Liu, Pang Wei Koh, Chelsea Finn

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

Extending the WILDS benchmark for unsupervised adaptation ICLR 2022
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
Oral presentation

WILDS: A benchmark of in-the-wild distribution shifts ICML 2021
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
Oral presentation

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

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

Supporting COVID-19 policy response with large-scale mobility-based modeling KDD 2021
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
Best paper award (Applied Data Science track)

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

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

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

Mobility network models of COVID-19 explain inequities and inform reopening Nature 2020
Serina Y Chang*, Emma Pierson*, Pang Wei Koh*, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec
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 ICML 2020
Pang Wei Koh*, Thao Nguyen*, Yew Siang Tang*, Steve Mussmann, Emma Pierson, Been Kim, and Percy Liang
Spotlight talk at the ICML 2020 Workshop on Human Interpretability in Machine Learning

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

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

Toward trustworthy AI development: Mechanisms for supporting verifiable claims arXiv 2020
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

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

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

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

Inferring multi-dimensional rates of aging from cross-sectional data AISTATS 2019
Emma Pierson*, Pang Wei Koh*, Tatsunori B. Hashimoto*, Daphne Koller, Jure Leskovec, Nicholas Eriksson, and Percy Liang
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 NeurIPS 2017
Jacob Steinhardt*, Pang Wei Koh*, and Percy Liang

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

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

An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development Scientific Data 2016
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

Mapping the pairwise choices leading from pluripotency to human bone, heart, and other mesoderm cell types Cell 2016
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

Denoising genome-wide histone ChIP-seq with convolutional neural networks Bioinformatics 2017
Pang Wei Koh*, Emma Pierson*, and Anshul Kundaje
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

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

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

Identifying genetic drivers of cancer morphology Undergraduate honors thesis 2012
Pang Wei Koh, Andrew Beck, and Daphne Koller.
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 NeurIPS 2011
Jiquan Ngiam, Pang Wei Koh, Zhenghao Chen, Sonia Bhaskar, and Andrew Y. Ng
Spotlight talk

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

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

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

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

Advising

Students mentored at Stanford

Irena Gao (Undergraduate)
Kendrick Shen (Undergraduate, now engineer at Genesis Therapeutics)
Erik Jones (Undergraduate; now PhD student at UC Berkeley)
Thao Nguyen (Undergraduate; now PhD student at the University of Washington)
Hubert Teo (Undergraduate; now ML Engineer at Nuro)
Kai-Siang Ang (Undergraduate; now ML Engineer at Nuro)
Henrik Marklund (Master's student; now PhD student at Stanford University)
Yew-Siang Tang (Master's student; now Senior Software Engineer at You.com)
Joon Sung Park (PhD student)
Serina Chang (PhD student)
Evan Zheran Liu (PhD student)
Kaitlyn Zhou (PhD student)
Shikhar Murty (PhD student)

Teaching

CS221 (Artificial Intelligence: Principles and Techniques), Stanford Fall 2020
Head Teaching Assistant
Managed a team of 14 TAs. Adapted the class to an online format because of COVID. This involved breaking live lectures into smaller online modules; revamping problem sessions; replacing exams with weekly quizzes; adding weekly fireside talks with remote guest speakers; and facilitating individual and group remote office hours.

CS228 (Probabilistic Graphical Models), Stanford Winter 2012
Head Teaching Assistant
Managed a team of 8 TAs. Revamped the class to make it application-focused and auto-gradable. Adapted the class to the Coursera platform, where we have taught 100,000+ online learners since 2012.

Service

Conferences and workshops

Organizer, NeurIPS Workshop on Distribution Shifts	2023
Reviewer, Science Advances	2023
Reviewer, ICLR Workshop on Mathematical and Empirical Understanding of Foundation Models	2023
Area chair, ICLR	2022
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

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 research talks and guest lectures

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