



Stanford | Stanford Data Science Initiative

Stanford Data Science Initiative Retreat

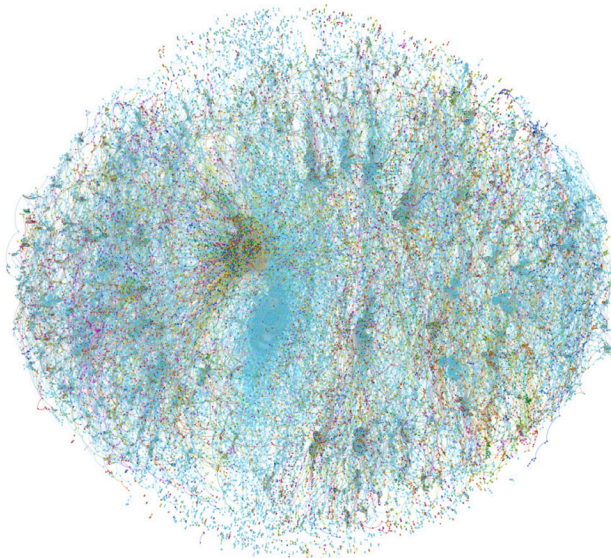
Li Ka Shing Conference Center, Stanford University

NOVEMBER 2, 2017

ABOUT THE EVENT

Welcome to the Stanford Data Science Initiative Fall 2017 Retreat! This event presents some of the most exciting, timely, and socially impactful data science research. Presentations include research on machine learning, the Internet of Things, and using the cloud for multi-institution clinical trials. Panel discussions will address data science for social benefit and industry best practices for dealing with the revolution in data science and artificial intelligence. Student and postdoc posters provide an opportunity for discussion.

SDSI brings together industry and academia around a shared interest in data science, its application, and domains where it is used. Thank you for your participation.



Cover illustration. The cover art is a diffusion cascade from an image meme that resurfaced in popularity multiple times in a social network. Studies of information sharing are improving society's understanding of how content spreads in a network. From "Do Cascades Recur?" J. Cheng, L. Adamic, J. Kleinberg, and J. Leskovec, ACM International Conference on World Wide Web (WWW), 2016.

2017 RETREAT AGENDA

8:00 am	Breakfast and registration	
9:00 am	Welcome and introductions	Steve Eglash , <i>Executive Director of Strategic Research Initiatives, Computer Science</i>
9:10 am	Vision for data science	Jure Leskovec , <i>Associate Professor, Computer Science; Director, Stanford Data Science Initiative</i>
9:30 am	Safe distributed computation for reluctant data sharers: At home in the Microsoft Azure cloud	Philip Lavori , <i>Professor, Biomedical Sciences</i> Balasubramanian Narasimhan , <i>Senior Research Scientist, Statistics and Biomedical Sciences</i> John Mark Agosta , <i>Principal Data Scientist, Microsoft</i>
10:10 am	Break	
10:40 am	Democratizing machine learning in the Stanford DAWN project	Matei Zaharia , <i>Assistant Professor, Computer Science</i>
11:00 am	Panel Discussion: Industry approaches for dealing with disruptive technology <i>Moderated by: Steve Eglash</i>	Peter Gunder , <i>Chief Business Development Officer, American Family Insurance</i> Peter Norvig , <i>Research Director, Google</i> Luca Marighetti , <i>Managing Director, Swiss Re</i>
12:00 pm	Lunch	
1:15 pm	Welcome back	
1:20 pm	After DAWN: The potential of usable machine learning today and tomorrow	Peter Bailis , <i>Assistant Professor, Computer Science</i>
1:40 pm	Finding earthquakes in a spaghetti of seismic streams	Philip Levis , <i>Associate Professor, Computer Science and Electrical Engineering</i>
2:00 pm	Panel Discussion: Socially impactful and ethical data science <i>Moderated by: Euan Ashley, Professor, Medicine (Cardiovascular); Director, Stanford Data Science Initiative</i>	Emma Brunskill , <i>Assistant Professor, Computer Science</i> Rumman Chowdhury , <i>Senior Principal, AI, Accenture</i> Thomas Kalil , <i>Senior Advisor to the Eric and Wendy Schmidt Group</i> Rob Reich , <i>Professor, Political Science</i>
3:15 pm	Break	
3:45 pm	Analytics on time-series data	Jure Leskovec , <i>Associate Professor, Computer Science; Director, SDSI</i>
4:05 pm	Security in the era of big data	Dan Boneh , <i>Professor, Computer Science and Electrical Engineering</i>
4:25 pm	Poster preview presentations	
4:45 pm	Closing comments	
4:50 pm	Poster viewing & cocktail reception	

POSTER TITLES

Presenter	Poster Title	Adviser
Imon Banerjee	Intelligent embedding of free-text clinical narratives: Key to unlock rich source of information	Daniel Rubin
Cody Coleman	DAWNBench: An End-to-End Deep Learning Benchmark and Competition	Peter Bailis Matei Zaharia
Animesh Garg	Neural task programming: Learning to generalize across hierarchical tasks	Fei-Fei Li
Leying Guan	A direct approach for predicting platelet use at Stanford Hospital	Robert Tibshirani
David Hallac	Toeplitz inverse covariance-based clustering of multivariate time series data	Jure Leskovec Stephen Boyd
Will Hamilton	Inductive representation learning on large graphs	Jure Leskovec Dan Jurafsky
Neal Jean	Enabling rapid screening of bacterial blood infections with machine learning	Stefano Ermon
Daniel Kang	NoScope: Querying videos 1,000x faster with deep learning	Peter Bailis Matei Zaharia
Volodymyr Kuleshov	Time series super-resolution using deep neural networks	Stefano Ermon
Himabindu Lakkaraju	Human-centric machine learning	Jure Leskovec
Deepak Narayanan	Kostos: A cost-based optimizer for data science workloads	Matei Zaharia
Shoumik Palkar	Accelerating data analytics end-to-end: From parsing to compute	Matei Zaharia
Alex Ratner	Snorkel: Rapid training set creation with weak supervision	Christopher Ré
Anna Shcherbina	Population-level analysis of physical activity patterns and impact on health: determining the roles of genes and environment	Euan Ashley Anshul Kundaje
Paroma Varma	Enabling efficient image and video training set generation	Christopher Ré
Somalee Datta Amir Bahmani	Genomic Data Commons: AI Assisted Genomic Data Interpretation Engine	n/a

SPEAKER BIOS



John Mark Agosta

Principal Data Scientist, Microsoft

John Mark Agosta is a principal data scientist in IMML at Microsoft. Over his career, he has worked with startups and labs in the Bay Area, including the original Knowledge Industries, and was a researcher at Intel Labs, where he was awarded a Santa Fe Institute Business Fellowship in 2007, and at SRI International after receiving his PhD from Stanford. He has participated in the annual Uncertainty in AI conference since its inception in 1985, proving his dedication to probability and its applications.

Euan Ashley

Professor of Medicine (Cardiovascular); Director, Stanford Data Science Initiative

Born in Scotland, Dr. Ashley graduated with 1st class Honors in Physiology and Medicine from the University of Glasgow. He completed medical residency and a PhD in molecular physiology at the University of Oxford before moving to Stanford University where he trained in cardiology and advanced heart failure, joining the faculty in 2006. His group is focused on the science of precision medicine. In 2010, he led the team that carried out the first clinical interpretation of a human genome. The paper published in the Lancet was the focus of over 300 news stories, became one of the most cited articles in clinical medicine that year, and was featured in the Genome Exhibition at the Smithsonian in DC. The team extended the approach in 2011 to a family of four and now routinely applies genome sequencing to the diagnosis of patients at Stanford hospital where Dr. Ashley directs the Clinical Genome Service and the Center for Inherited Cardiovascular Disease. He was a recipient of the National Innovation Award from the American Heart Association as well as the NIH Director's New Innovator Award. Dr. Ashley is Principal Investigator of the MyHeart Counts study, developed in collaboration with Apple Inc. in 2015, and one of the fastest recruiting studies ever launched. He is co-founder of Personalis Inc., a genome scale genetic diagnostics company and advises many other biotechnology and pharmaceutical companies. In 2013, Dr. Ashley was recognized by the Obama White House for his contributions to Personalized Medicine.



Peter Bailis

Assistant Professor of Computer Science

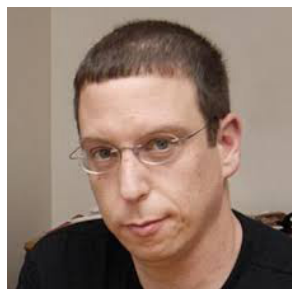
Peter Bailis is an Assistant Professor of Computer Science at Stanford University. Peter's research in the Future Data Systems group focuses on the design and implementation of next-generation, post-database data-intensive systems. His work spans large-scale data management, distributed protocol design, and architectures for high-volume complex decision support. He is the recipient of an NSF Graduate Research Fellowship, a Berkeley Fellowship for Graduate Study, best-of-conference citations for research appearing in both SIGMOD and VLDB, and the CRA Outstanding Undergraduate Researcher Award. He received a PhD from University of California at Berkeley in 2015 and an A.B. from Harvard College in 2011, both in Computer Science.



Dan Boneh

Professor of Computer Science and Electrical Engineering

Professor Dan Boneh heads the applied cryptography group at the Computer Science department at Stanford University. Professor Boneh's research focuses on applications of cryptography to computer security. His work includes cryptosystems with novel properties, web security, security for mobile devices, digital copyright protection, and cryptanalysis. He is the author of over a hundred publications in the field and a recipient of the Packard Award, the Alfred P. Sloan Award, and the RSA award in mathematics. Last year Dr. Boneh received the Ishii award for industry education innovation. Professor Boneh received his PhD from Princeton University and joined Stanford in 1997.





Emma Brunskill

Assistant Professor of Computer Science

Emma Brunskill is an Assistant Professor of Computer Science at Stanford University. She is a Rhodes Scholar, a Microsoft Faculty Fellow, a NSF CAREER awardee and an ONR Young Investigator Program recipient. Her work focuses on interactive machine learning and reinforcement learning, with a particular focus on algorithms for high stakes domains like healthcare, education and consumer interactions.



Rumman Chowdhury

Senior Principal, Artificial Intelligence, Accenture

Her passion lies at the intersection of artificial intelligence and humanity. She comes to data science from a quantitative social science background. Currently, she is a Senior Principal at Accenture, working on cutting-edge applications of Artificial Intelligence and leading their Strategic Global Initiative on Responsible Artificial Intelligence. In her job, she advises companies on ethical AI practices and works with organizations such as the World Economic Forum, and the IEE on standards and best practices for Ethical Artificial Intelligence and Autonomous Systems. In 2017, Chowdhury was selected as one of the BBC 100 women as a part of #Teamlead tasked with tackling the glass ceiling by creating an app that can teach women to 'lean in' during meetings. In her spare time, she serves on the Board of Directors for multiple AI startups and an AI mentor for Katapult Accelerator, an impact tech accelerator in Oslo, Norway. Chowdhury is also a teacher and an advisor at a data camp startup that helps refugees learn skills for the digital economy. She's a Forbes Tech contributing author and has been named by InformationWeek as one of 10 influential AI and machine learning experts to follow on Twitter.



Steve Eglash

Executive Director of Strategic Research Initiatives, Computer Science

Steve Eglash is Executive Director of Strategic Research Initiatives in the Computer Science Department at Stanford University, where he develops and manages research programs in big data, artificial intelligence, Internet of Things, security, autonomous vehicles, robotics, and machine learning. Prior to joining Stanford, Steve was CEO of Cyrium Technologies, venture capitalist at Worldview Technology Partners, vice president at SDL (JDSU), and member of the technical staff at MIT Lincoln Laboratory. As vice president at SDL, Steve was part of the management team that grew the company to \$1 billion in annual revenue and engineered one of the largest high-tech acquisitions in history. He has published more than forty papers in peer-reviewed journals and has four patents. Steve received a PhD and MS from Stanford University, and a BS from the University of California at Berkeley, all in Electrical Engineering. Steve is chair of the Santa Clara University College of Arts & Sciences Leadership Board.



Peter Gunder

Chief Business Development Officer, American Family Insurance

Peter Gunder joined the American Family Insurance Group in 2008. "AmFam" is a top 10 personal lines auto and homeowners insurer. Its brands include The General and American Family. In January 2013, Peter was named Executive Vice President responsible for product management for home and auto insurance, enterprise strategy, mergers and acquisition, investments and innovation. In January 2014 he was named Chief Business Development Officer. He is the executive responsible for American Family Ventures. He is a board observer for startup Clearcover and CoverHound and a board member of Gen.Life. Before joining American Family, Peter was co-founder, chief investment officer and managing director of Cardinal Investment Advisors in Chicago. Peter earned a bachelor's degree in mechanical engineering from Stanford University, later receiving an MBA in Finance from the University of Chicago Graduate School of Business. He is a chartered financial analyst.



Thomas Kalil

Senior Advisor, Eric and Wendy Schmidt Group

Previously, Thomas Kalil served as the Deputy Director for the White House Office of Science and Technology Policy and Senior Advisor for Science, Technology and Innovation for the National Economic Council. Working with agencies across the federal government, OSTP's Technology and Innovation Division developed dozens of White House initiatives that are designed to foster American leadership in innovation.



Kevin Koy

Managing Director, Data Science and AI Affiliates Programs

Kevin is responsible for operating and administering Stanford's data science and artificial intelligence industrial affiliates programs that generate critical funds to help support the university's research mission in data science, artificial intelligence, and computation. Prior to joining Stanford, Kevin served as the founding Executive Director for the Berkeley Institute for Data Science (BIDS) where he helped to establish an active interdisciplinary environment at the University of California at Berkeley that connects people with the data, methods, and tools to advance their research. Kevin also has previous experience in geospatial research with a focus on analysis and visualization of spatial data for natural systems at Berkeley, the American Museum of Natural History, and the Smithsonian Institution.



Philip Lavori

Professor of Biomedical Sciences

Philip W. Lavori grew up on a small farm in New York City. He received a PhD in Mathematics from Cornell University in 1974, taught in the Mathematics and Electrical Engineering Departments at MIT, and established a biostatistics unit at the Massachusetts General Hospital in 1982, when he joined the faculty of Harvard Medical School, and the Harvard School of Public Health (Biostatistics). He became Professor and Vice Chair for Research in the Brown University Department of Psychiatry and Human Behavior in 1989. In 1992 he moved to Stanford University where he is Professor of Biomedical Data Science emeritus. He served as Director of the Palo Alto Coordinating Center for the VA Cooperative Studies Program, Acting Director of the National VA Cooperative Studies Program where he helped launch the VACSP's national DNA banking effort. He founded the Biostatistics Core of the Stanford Cancer Center, served as co-director of Spectrum, Stanford's Clinical and Translational Sciences Award, as Chair of the Department of Health Research and Policy, and as Vice Chair of the Department of Biomedical Data Science. His current interests center on innovative clinical trials design, especially for seamless phase II-III trials, biomarker-guided treatments, point-of-care randomization for comparative effectiveness research, sequential multiple assignment randomized trials for adaptive treatment strategies and distributed computation as a substitute for aggregated data.



Jure Leskovec

Associate Professor of Computer Science; Director, Stanford Data Science Initiative

Jure Leskovec is an Associate Professor of Computer Science at Stanford University where he is a member of the InfoLab and the AI lab. He joined the department in September 2009. He is also working as Chief Scientist at Pinterest, where he focuses on machine learning problems. He co-founded Kosei, a machine learning startup, acquired by Pinterest. In 2008/09 he was a postdoctoral researcher at Cornell University working with Jon Kleinberg and Dan Huttenlocher. He completed his PhD in Machine Learning Department, School of Computer Science at Carnegie Mellon University under the supervision of Christos Faloutsos in 2008. He completed his undergraduate degree in computer science at University of Ljubljana, Slovenia in 2004. He also works with the Artificial Intelligence Laboratory, Jozef Stefan Institute, Ljubljana, Slovenia.



Philip Levis

Associate Professor of Computer Science and Electrical Engineering

Philip Levis is an Associate Professor in the Computer Science and Electrical Engineering Departments of Stanford University. He heads the Stanford Information Networking Group (SING), co-directs the Secure Internet of Things Project, and holds the Fletcher Jones Faculty Development Chair. His research is in operating systems, networks, and software design, especially wireless networks, sensor networks, and embedded systems. He likes building stuff other people use and writing code. He's drawn to excellent engineering and has a self-destructive aversion to low hanging fruit.

Luca Marighetti

Managing Director, Swiss Re

Luca has nearly 30 years of professional experience in a large variety of industries, in geopolitical matters, and in tech. He held a number of roles first in Academia, then at Procter & Gamble (European Brand Manager), and SC Johnson (Commercial Director Central Europe), McKinsey & Company (Assembly, R&D, Chemicals, Banking), Deutsche Bank (on-line trading, investment products), Allianz Group (market management across banking), Zurich Insurance (strategy and tech), and now Swiss Re where he developed the tech strategy for the Group and he now leading its implementation. Luca studied, Medieval Literature (with Hans Robert Jauss), History (with Arno Borst), Sociology (with Thomas Luckmann), Philosophy (with Jürgen Mittelstrass), with masters in the first two) at the University of Constance. There he also earned a PhD in Philosophy of science, with a thesis on the epistemic contribution delivered by the extension of the thinkable ("possible worlds" and "language") as performed in the literary space, and mentored by the same academic teachers. Luca attended scientific High School in Italy and Germany.



Balasubramanian Narasimhan

Senior Research Scientist, Statistics and Biomedical Sciences

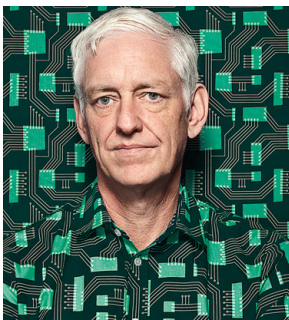
Balasubramanian Narasimhan is a Senior Research Scientist in the Department of Biomedical Data Science and the Department of Statistics. He is also the Director of the Data Coordinating Center in the School of Medicine, Stanford University. He obtained his doctorate in Statistics in the Department of Statistics at Florida State University under George Marsaglia. He taught at the University of Minnesota, Morris and Penn State University, Erie, before moving to Stanford University. His research interests are in computational statistics, clinical trial design and methodology and design, and optimization.



Peter Norvig

Research Director, Google

Peter Norvig is a Director of Research at Google Inc. Previously he was head of Google's core search algorithms group, and of NASA Ames's Computational Sciences Division, making him NASA's senior computer scientist. He received the NASA Exceptional Achievement Award in 2001. He has taught at the University of Southern California and the University of California at Berkeley, from which he received a PhD in 1986 and the distinguished alumni award in 2006. He was co-teacher of an Artificial Intelligence class that signed up 160,000 students, helping to kick off the current round of massive open online classes. His publications include the books *Artificial Intelligence: A Modern Approach* (the leading textbook in the field), *Paradigms of AI Programming: Case Studies in Common Lisp*, *Verbmobil: A Translation System for Face-to-Face Dialog*, and *Intelligent Help Systems for UNIX*. He is also the author of the Gettysburg Powerpoint Presentation and the world's longest palindromic sentence. He is a fellow of the AAI, ACM, California Academy of Science and American Academy of Arts & Sciences.





Rob Reich

Professor of Political Science

Rob Reich is Professor of political science and, by courtesy, professor of philosophy and at the Graduate School of Education, at Stanford University. He is the director of the Center for Ethics in Society and faculty co-director of the Center on Philanthropy and Civil Society (publisher of the Stanford Social Innovation Review), both at Stanford University. He is the author or editor of several books on education and two books on the relationship between philanthropy, democracy, and justice: *Philanthropy in Democratic Societies* (edited with Chiara Cordelli and Lucy Bernholz) the forthcoming *Just Giving*. His current work focuses on ethics and technology, and he is editing a new volume called Digital Technology and Democratic Theory. He is the recipient of multiple teaching awards and is a board member of GiveWell.org and the magazine Boston Review.

More details at his personal webpage: <http://robreich.stanford.edu>

Twitter: @robreich



Erika Strandberg

Program Director, Data Science and AI Initiatives

Erika Strandberg is Program Director for Data Science and AI Initiatives at Stanford where she is developing new programs and initiatives to support the relationship between industrial affiliates and Stanford faculty. Prior to becoming Program Director, Erika completed a PhD in Biomedical Informatics at Stanford, focusing on machine learning and rare event clinical predictions using electronic medical record data with high amounts of missingness. She also has a Masters degree in Statistics from the University of California at San Diego where she studied deviations from assumptions in survival analysis and extensive experience in project management for medical device research and development.



Matei Zaharia

Assistant Professor of Computer Science

Matei Zaharia is an Assistant Professor at Stanford Computer Science, where he works on computer systems and big data. He is also co-founder and Chief Technologist of Databricks, the big data company commercializing Apache Spark. Prior to joining Stanford, he was an assistant professor of computer science at MIT.

CURRENT SDSI CORPORATE MEMBERS

We are pleased to acknowledge the generous support of our corporate members

FOUNDING MEMBERS



REGULAR MEMBERS



The Stanford Data Science Initiative is Stanford's university-wide program for big data. SDSI conducts research on massive data, new algorithms, and advanced analytics with strong ties to application areas across Stanford, with industry, and worldwide.

sdsi.stanford.edu