

ANNUAL SYMPOSIUM
ON THE FUTURE OF COMPUTATION IN SCIENCE AND ENGINEERING

PRESENTED BY
THE INSTITUTE FOR APPLIED COMPUTATIONAL SCIENCE

THE DIGITAL DOCTOR

Health Care in an Age of AI
and Big Data

January 16, 2018

9:00am – 5:00pm

Harvard Science Center / Hall B
1 Oxford Street, Cambridge, MA

www.compute fest.seas.harvard.edu/symposium



Institute for Applied Computational Science
Harvard John A. Paulson
School of Engineering and Applied Science



Center for Research on Computation and Society (CRCS)
Harvard Data Science Initiative (HDSI)



HDSI | Harvard Data
Science Initiative



SCHEDULE

THE DIGITAL DOCTOR: HEALTH CARE IN AN AGE OF AI AND BIG DATA

TUESDAY, JANUARY 16, 2018 | 9:00AM-5:00PM
SCIENCE CENTER HALL B

"With individualized medicine, data mining of electronic health records, and AI-augmented diagnosis, computation and data science are reshaping modern medicine."

8:30 COFFEE

9:00 **WELCOME: TIM KAXIRAS**, John Hasbrouck Van Vleck
Professor of Pure and Applied Mathematics & Director,
Institute for Applied Computational Science,
Harvard University

9:05 **INTRODUCTION: FRANCESCA DOMINICI**, Professor of
Biostatistics & Co-Director of the Harvard Data Science
Initiative, Harvard University

9:15 **Finale Doshi-Velez**, Assistant Professor of Computer
Science, Harvard University

9:55 **Matt Might**, Director of the Hugh Kaul Personalized
Medicine Institute, University of Alabama at Birmingham
School of Medicine

10:35 BREAK

SCHEDULE

- 10:55** **John Brownstein**, Professor of Biomedical Informatics at Harvard Medical School & Chief Innovation Officer of Boston Children's Hospital
- 11:35** **Marzyeh Ghassemi**, Visiting Researcher at Google's Verily & MIT CSAIL Post Doctoral Fellow
- 12:15** **LUNCH BREAK**
- 2:00** **Jennifer Chayes**, Technical Fellow and Managing Director, Microsoft Research New England, Microsoft Research New York City and Microsoft Maluuba, Montreal
- 2:40** **Emery Brown**, Edward Hood Taplin Professor of Medical Engineering and Computational Neuroscience at MIT; Warren M. Zapol Professor of Anaesthesia at Harvard Medical School; and practicing anesthesiologist at Massachusetts General Hospital
- 3:20** **COFFEE BREAK**
- 3:45** **PANEL DISCUSSION: CHALLENGES AND OPPORTUNITIES**
- 4:30** **CLOSING REMARKS**



Emery N. Brown is the Edward Hood Taplin Professor of Medical Engineering and Computational Neuroscience at MIT; the Warren M. Zapol Professor of Anaesthesia at Harvard Medical School; and a practicing anesthesiologist at MGH.

Brown received his B.A. (*magna cum laude*) in Applied Mathematics from Harvard College, his M.A. and Ph.D. in statistics from Harvard University and his M.D. (*magna cum laude*) from Harvard Medical School. Professor Brown completed his internship in internal medicine at the Brigham and Women's Hospital and his residency in anesthesiology at MGH. He joined the faculty at Harvard Medical School in 1992 and at MIT in 2005.



John Brownstein, PhD is Professor of Biomedical Informatics at Harvard Medical School and is the Chief Innovation Officer of Boston Children's Hospital. He directs the Computational Epidemiology Lab and the Innovation and Digital Health Accelerator, both at Boston Children's. Dr. Brownstein was trained as an epidemiologist at Yale University. He is also Uber's healthcare advisor and co-founder of digital health companies Epidemico and Circulation.



Jennifer Tour Chayes is Technical Fellow and Managing Director of Microsoft Research New England in Cambridge, Massachusetts, which she co-founded in 2008; Microsoft Research New York City, which she co-founded in 2012; and, Maluuba, a recent acquisition which just joined Microsoft Research. Before joining Microsoft in 1997, Chayes was a Professor of Mathematics at UCLA. She is the author of over 130 academic papers and the inventor of over 30 patents. Her research areas include phase transitions in discrete mathematics and computer science, structural and dynamical properties of self-engineered networks, graph algorithms, algorithmic game theory and computational biology. Chayes is one of the inventors of the theory of graph limits, which is widely used for machine learning of massive networks.

She received her B.A. in biology and physics at Wesleyan University, where she graduated first in her class, and her Ph.D. in mathematical physics at Princeton.



Finale Doshi-Velez is an Assistant Professor at the Harvard Paulson School of Engineering and Applied Sciences. She received her PhD from MIT and her MSc from the University of Cambridge as a Marshall Scholar. Her research interests are at the intersection of machine learning and healthcare.



Marzyeh Ghassemi is a Visiting Researcher with Google's Verily and a post-doc in the Clinical Decision Making Group at MIT's Computer Science and Artificial Intelligence Lab (CSAIL) supervised by Dr. Peter Szolovits. She will join the University of Toronto as an Assistant Professor in Computer Science and Medicine in Fall 2018, and will be affiliated with the Vector Institute.

Dr. Ghassemi's research focuses on machine learning with clinical data to predict and stratify relevant human risks, encompassing unsupervised learning, supervised learning, and structured prediction. Her work has been applied to estimating the physiological state of patients during critical illnesses, modelling the need for a clinical intervention, and diagnosing phonotraumatic voice disorders from wearable sensor data.

Ghassemi received B.S. degrees in computer science and electrical engineering as a Goldwater Scholar at New Mexico State University, and a MSc degree in biomedical engineering from Oxford University as a Marshall Scholar.



Matt Might is the Director of the Hugh Kaul Personalized Medicine Institute at the University of Alabama at Birmingham, where he is the Hugh Kaul Endowed Chair of Personalized Medicine and both a Professor of Medicine and a Professor of Computer Science. Dr. Might has served as a Strategist in the Executive Office of the President at the White House for both the prior and current administration. Additionally, he has served on the faculty in the Department of Biomedical Informatics at the Harvard Medical School since 2015. Dr. Might is the Chief Science Officer of NGLY1.org and a Co-Founder and Scientific Advisor to Paimomix, LLC. He previously served as Associate

Professor of Computer Science and Adjunct Associate Professor of Pharmaceutical Chemistry at the University of Utah. Dr. Might received his Ph.D. in Computer Science from Georgia Tech in 2007. He tweets from @mattmight and blogs at blog.might.net

**CATHERINE CHUTE****SYMPOSIUM ORGANIZER**

Cathy Chute is the Executive Director of the Institute for Applied Computational Science (IACS) and Assistant Dean for Professional Programs at the Harvard Paulson School of Engineering and Applied Sciences. Chute's work at IACS focuses on growing IACS connections to industry partners and collaborators across the university. Chute was previously the Publisher of Harvard Magazine as well as the Executive Director of the Ivy League Magazine Network. Prior to coming to Harvard, she was a director of new business development at the New York Times.

**FRANCESCA DOMINICI****SYMPOSIUM ORGANIZER**

Dr. Francesca Dominici is Professor of Biostatistics at the Harvard T.H. Chan School of Public Health and Co-Director of the Data Science Initiative at Harvard University. She received her PhD in Statistics from the University of Padua, Italy, in 1997. From 1999 to 2009 she was a Professor at the Bloomberg School of Public Health at Johns Hopkins University. In 2009 she moved to Harvard T.H. Chan School of Public Health as a tenured Professor of Biostatistics and was appointed Associate Dean of Information Technology in 2011. In Fall 2013, she was appointed Senior Associate Dean for Research and in February 2017, she was appointed as co-director of the Harvard Data Science Initiative.

Dr. Dominici's research has focused on the development of statistical methods for the analysis of large and complex data. She is a passionate data scientist; her expertise is in the development of statistical methods for the analysis of large, messy data and for combining information across heterogeneous data sources. She leads several interdisciplinary groups of scientists with the ultimate goal of addressing important questions in environmental health science, climate change, comparative effectiveness research in cancer, and health policy.



BRIAN HAYES

SYMPOSIUM ORGANIZER

Brian Hayes is a writer with interests in mathematics, computing, and the sciences. He has written for American Scientist, Scientific American, The Sciences, Natural History, and other publications, as well as his own web site <http://bit-player.org>. He is the author of Infrastructure: A Guide to the Industrial Landscape (2005, 2014) and Group Theory in the Bedroom, and Other Mathematical Diversions (2008). Formerly he was an editor on the staff of Scientific American and the editor in chief of American Scientist. Since 2010 he has been an Associate of the Harvard John A. Paulson School of Engineering and Applied Sciences. His recent work is supported by a Y Combinator Research Fellowship.



ZAK KOHANE

SYMPOSIUM ORGANIZER

Isaac Kohane, MD, PhD is the inaugural Chair of the Department of Biomedical Informatics at Harvard Medical School. He develops and applies computational techniques to address disease at multiple scales: From whole healthcare systems as "living laboratories" to the functional genomics of neurodevelopment with a focus on autism. Kohane's i2b2 project is currently deployed internationally to over 120 major academic health centers to drive discovery research in disease and pharmacovigilance (including providing evidence on drugs which ultimately contributed to "black boxing" by the FDA). Dr. Kohane has published several hundred papers in the medical literature and authored a widely used book on Microarrays for an Integrative Genomics. He is a member of the Institute of Medicine and the American Society for Clinical Investigation.



MARGO SELTZER

SYMPOSIUM ORGANIZER

Margo I. Seltzer is the Herchel Smith Professor of Computer Science in the Harvard John A. Paulson School of Engineering and Applied Sciences.

She received an A.B. degree in Applied Mathematics from Harvard/Radcliffe College in 1983 and a Ph. D. in Computer Science from the University of California, Berkeley, in 1992.

She is the author of several widely-used software packages including database and transaction libraries and the 4.4BSD log-structured file system. Dr. Seltzer was a founder and CTO of Sleepycat Software, the makers of Berkeley DB and is now an Architect for Oracle Corporation.

Before pursuing an academic career, professor Seltzer spent several years working at startup companies designing and implementing file systems and transaction processing software and designing microprocessors. She is a Sloan Foundation Fellow in Computer Science, a Bunting Fellow, and was the recipient of the 1996 Radcliffe Junior Faculty Fellowship, and the University of California Microelectronics Scholarship. She is recognized as an outstanding teacher and won the Phi Beta Kappa teaching award in 1996 and the Abrahamson Teaching Award in 1999.

EMERY BROWN

Edward Hood Taplin Professor of Medical Engineering and Computational Neuroscience at MIT; Warren M. Zapol Professor of Anaesthesia at Harvard Medical School; and practicing anesthesiologist at MGH

DECIPHERING THE DYNAMICS OF THE ANESTHETIZED BRAIN: A CASE STUDY IN STATISTICAL AND MATHEMATICAL MODELING

General anesthesia is a drug-induced, reversible condition comprised of five behavioral states: unconsciousness, amnesia (loss of memory), analgesia (loss of pain sensation), akinesia (immobility), and hemodynamic stability with control of the stress response. Our clinical experience, basic science experiments, statistical modeling research and mathematical modeling research show that a primary mechanism through which anesthetic drugs create these altered arousal states is by initiating and maintaining highly structured oscillations that impair the ability of different brain regions to communicate. These insights into the mechanisms of anesthetic action suggest new paradigms for personalized anesthesia care: using the electroencephalogram to monitor in real-time the brain states of patients receiving general anesthesia; devising principled strategies to dose anesthetic drugs; and developing strategies to control anesthetic states. The study of general anesthesia is a valuable testbed for characterizing dynamic processes in neuroscience.

JOHN BROWNSTEIN

Professor of Biomedical Informatics at Harvard Medical School & Chief Innovation Officer of Boston Children's Hospital

OPTIMIZING THE PATIENT JOURNEY WITH AI

Through social media, forums and online communities, wearable technologies and mobile devices, there is a growing body of health-related data that can shape our assessment of human illness. Collectively, this data comprises an individual's 'digital phenotype' - unique, unsolicited and real-time information about a person's health. Our current research focuses on using digital phenotypes for population health surveillance, specifically to identify and analyze specific sub-populations over space and time with the goal of better understanding patient behavior and disease dynamics. Some current research topics include foodborne illness, insomnia, autism, febrile illness, and patient experience.

JENNIFER CHAYES

Technical Fellow and Managing Director, Microsoft Research New England, Microsoft Research New York City and Microsoft Maluuba, Montreal

CHALLENGES AND OPPORTUNITIES FOR MACHINE LEARNING IN CANCER IMMUNOTHERAPY

Cancer immunotherapy is one of the most exciting developments in healthcare. By enlisting our own immune systems to go after cancer, immunotherapy is much more focused than other cancer therapies – when effective, it kills the cancer cells without the damage of chemo or radiation. Cancer immunotherapy is now a first line treatment for stage four melanoma and lung cancer, leading to long-term survival in many patients. Unfortunately, existing cancer immunotherapies only work for a relatively small subset of cancer patients. How do we properly identify the patients for whom existing immunotherapies are likely to be effective, and not have serious side effects, and how do we develop new cancer immunotherapies for other patients? The relevant data is multimodal: genomic, immunological, metabolic, clinical and more. Trials involve relatively few patients with extremely high-dimensional data per person. This poses unique challenges and opportunities for machine learning and statistics.

FINALE DOSHI-VELEZ

Assistant Professor of Computer Science, Harvard University

REINFORCEMENT LEARNING IN HEALTHCARE: CHALLENGES AND PROMISE

Automated decision-making agents have recently seen great successes of Alpha-Go and playing Atari. Can these advances be applied in healthcare? As with those games, many healthcare problems require thinking not only about the immediate effect of a treatment, but possible long-term ramifications. For example, a certain drug cocktail may cause an immediate drop in viral load in HIV but also cause the resistance mutations that will reduce the number of viable treatment options in the future. Keeping a patient on a ventilator now might prevent an immediate degradation of their status but also increase the risk of muscle atrophy. However, unlike with games, we cannot simply experiment with patients! Thus, when applying reinforcement learning to healthcare problems, we must first learn and evaluate not from our own exploration, but retrospective data collected from the practice and exploration of previous doctors. In this talk, Professor Doshi-Velez will discuss some innovations in her research group to apply these paradigms to real problems in managing HIV, and also highlight challenges, pitfalls, and open problems.

MARZYEH GHASSEMI

Visiting Researcher at Verily Cambridge & MIT CSAIL Post
Doctoral Fellow

PHENOMICS IS THE NEW GENOMICS

The explosion of clinical data provides an exciting new opportunity to use machine learning to discover new and impactful clinical information. Among the questions that can be addressed are establishing the value of treatments and interventions in heterogeneous patient populations, creating risk stratification for clinical endpoints, and investigating the benefit of specific practices or behaviors. However, there are many challenges to overcome. First, clinical data are noisy, sparse, and irregularly sampled. Second, many clinical endpoints (e.g., the time of disease onset) are ambiguous, resulting in ill-defined prediction targets.

We tackle these problems by learning "phenotypes" - abstractions that generalize across applications despite missing and noisy data. Dr. Ghassemi's work spans coded records from administrative staff, vital signs recorded by monitors, lab results from ordered tests, notes taken by clinical staff, and accelerometer signals from wearable monitors. The learned representations capture higher-level structure and dependencies between multi-modal time series data and multiple time-varying targets. Dr. Ghassemi focuses on learning techniques that transform diverse data modalities into a consistent intermediate that improves prediction in clinical investigation.

In this talk, she will discuss the need for practical, evidence-based medicine, and the challenges of creating multi-modal representations for prediction targets varying both spatially and temporally.

MATT MIGHT

Director of the Hugh Kaul Personalized Medicine Institute, University
of Alabama at Birmingham School of Medicine

THE ALGORITHM FOR PRECISION MEDICINE

Powered by the ascent of genomics, precision medicine promises to bring the right treatment to the right patient at the right time. The question is: "How?" This talk explores the development of an "algorithm" for delivering precision medicine in practice -- optimizes the healthcare outcomes for patients with respect to their medical data. Rather than develop the algorithm theoretically, this talk will focus on the development of the algorithm as a consequence of learning to deliver precision medicine to patients.

ABOUT US

THE INSTITUTE FOR APPLIED COMPUTATIONAL SCIENCE

The Institute for Applied Computational Science (IACS) is an interdisciplinary education and research program. Founded in 2010 to leverage Harvard University's strengths in applied mathematics and computer science, the Institute is home to master's degree programs in applied computational science, a field that brings together mathematical models, algorithms, systems innovations, computational techniques, and statistical tools; and, data science. The IACS holds annual symposiums, biweekly seminars, public workshops, and conferences. For more information, visit iacs.seas.harvard.edu.

COMPUTEFEST

The IACS hosts its annual ComputeFest during Harvard's January intersession. The annual symposium on the future of computation in science and engineering is the final event of this two-week long program. For more information, visit computeifest.seas.harvard.edu.

SPONSOR INFORMATION

CRCS

The Center for Research on Computation and Society (CRCS), at the Harvard John A. Paulson School of Engineering and Applied Sciences, brings together computer scientists and scholars from a broad range of fields to make advances in computational research that serve public interest.

HDSI

The Harvard Data Science Initiative will unite data science efforts across the university, foster collaboration in both research and teaching, and catalyze research that will benefit our society and economy. It will be home to a research platform to accelerate the pace of discovery. It will strengthen the fabric of connections among departments to create an integrated data science community, all to empower research progress and education across the University.

NOTES



Institute for Applied Computational Science
Harvard John A. Paulson
School of Engineering and Applied Science



Center for Research on Computation and Society (CRCS)
Harvard Data Science Initiative (HDSI)



HDSI | Harvard Data
Science Initiative