Advances in Multimodal Artificial Intelligence to Enhance Environmental and Biomedical Data Integration

June 14-15, 2023

THE NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE
500 5TH ST NW, WASHINGTON, DC 20001

FOR REGISTRATION AND MORE INFORMATION, PLEASE VISIT THE WORKSHOP WEBSITE
ALL TIMES LISTED ARE IN EASTERN TIME
This workshop will be webcast

How can artificial intelligence be used to accelerate data integration to further understanding of complex health challenges? Driven by increasing growth in data collection and availability as well as computing power, artificial intelligence and machine learning applications are rapidly growing in all areas, including research and health applications. In this workshop, experts will discuss opportunities for leveraging new and cutting-edge developments in artificial intelligence and machine learning to help integrate heterogenous data (eg., genomics, epigenomics, exposomics, geospatial information, data from wearables, electronic health records). Workshop speakers will also highlight emerging technologies and explore the ethical, social, and policy implications of large-scale environmental health and biomedical data collection and integration.

WEDNESDAY JUNE 14, 2023 10:00 AM – 4:30 PM ET Room 106

9:15 Registration

10:00 Welcome and Opening Remarks—**Andrea Baccarelli***, Columbia University, Co-Chair of the Standing Committee on the Use of Emerging Science for Environmental Health Decisions

10:10 **SESSION 1: ON THE HORIZON: THE CONVERGENCE OF ARTIFICIAL INTELLIGENCE, ENVIRONMENTAL HEALTH, AND BIOMEDICINE**

Patrick Breysse, John Hopkins University
Lucila Ohno-Machado, Yale University
Marzyeh Ghassemi, Massachusetts Institute of Technology
Moderator: **Megan W. Latshaw*†**, John Hopkins University

10:45 Session 1 Panel Discussion

11:20 Break

11:25 **SESSION 2: LEVERAGING AI/ML FOR ENVIRONMENTAL HEALTH AND BIOMEDICAL DATA INTEGRATION**

Chirag Patel, Harvard University
Aidong Zhang, University of Virginia
Heidi Hanson, Oak Ridge National Laboratory
Moderator: **David Reif**, National Institute of Environmental Health Sciences

12:00 Session 2 Panel Discussion

12:35 Break

2:00 **SESSION 3: EMERGING METHODS FOR DATA INTEGRATION**

Karim Lekadir, University of Barcelona
Joyce Ho, Emory University
Thomas Hartung, John Hopkins University
Moderator: **Yao-Yi Chiang*†**, University of Minnesota

2:35 Session 3 Panel Discussion

3:10 Break

3:20 **REMARKS FROM RICHARD WOYCHIK, NATIONAL INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES**

3:30 **KEYNOTE ADDRESS: ERIC TOPOL, THE SCRIPPS RESEARCH INSTITUTE**

3:45 **FIRESIDE CHAT: ERIC TOPOL, THE SCRIPPS RESEARCH INSTITUTE, AND RICHARD WOYCHIK, NATIONAL INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES**

Moderator: **Lucila Ohno-Machado**, Yale University

4:15 Closing Remarks—**Carmen Marsit***, Emory University

4:30 Adjourn Day 1

THURSDAY JUNE 15, 2023 10:00 AM – 1:30 PM ET Room 100

10:00 Welcome—**Andrea Baccarelli***, Columbia University, Co-Chair of the Standing Committee on the Use of Emerging Science for Environmental Health Decisions

10:05 **SESSION 4: PERSPECTIVES ON AI AND DATA GOVERNANCE AND INFRASTRUCTURE**

Susan Gregurick, National Institutes of Health
Jorge Calzada, Centers for Disease Control and Prevention
Janet Haven, Data and Society; Member of the National AI Advisory Committee

Suzanne Dorsey, Maryland Department of Environment
Moderator: **Gwen Ottinger†**, Drexel University

10:45 Panel Discussion

11:15 **SESSION 5: TECHNOLOGIES AND TOOLS TO ADVANCE ENVIRONMENTAL HEALTH AND BIOMEDICAL RESEARCH**

Lorenzo Hankla, Department of Defense
Akane Sano, Rice University
Gengchen Mai, University of Georgia
Nicholas Skaff, Centers for Disease Control and Prevention
Moderator: **Alison Motsinger-Reif**, National Institute of Environmental Health Science

11:55 Panel Discussion

12:20 Closing Remarks—**Carmen Marsit*†**, Emory University

12:40—1:20 Breakout Groups and Informal Networking
JOIN

1:30 Adjourn Workshop[#]

---

* Member of the Standing Committee on the Use of Emerging Science for Environmental Health Decisions.
† Member of the workshop Organizing Committee for Advances in Multimodal Artificial Intelligence to Enhance Environmental and Biomedical Data Integration
[#] The Standing Committee on the Use Emerging Science for Environmental Health Decisions Business Meeting is from 2:50–4:00pm. This meeting is open to committee members and government liaisons.

Workshop Organizing Committee

This workshop was organized by the following experts: **Carmen Marsit (Workshop Chair)**, Emory University; **Yao-Yi Chiang,** University of Minnesota; **Christopher Duncan,** National Institute of Environmental Health Sciences; **Anindita Dutta**, University of Chicago; **Megan Latshaw,** John Hopkins University; **Gwen Ottinger**, Drexel University

About Us

The National Academies' Standing Committee on the Use of Emerging Science for Environmental Health Decisions (ESEHD) examines and discusses issues on the use of new science, tools, and research methodologies for environmental health decisions. The ESEHD committee is organized under the auspices of Board on Life Sciences and the Board on Environmental Studies and Toxicology of the National Academies of Sciences, Engineering, and Medicine, and sponsored by the National Institute of Environmental Health Sciences.

SPEAKER BIOS

Patrick Breysse

Dr. Breysse has over 40 years of extensive academic and governmental leadership experience conducting health studies and leading nationwide efforts to address the most pressing environmental health issues of our times. A highly successful and prolific researcher, mentor, director, and leader, Dr. Breysse has over 270 publications investigating the impact of chemical, biologic, and physical hazards on human health, along with a robust record of conducting health studies, attracting external funding, and advising students. In 2014, Dr. Breysse joined the Centers for Disease Control and Prevention as the Director of both the National Center for Environmental Health and the Agency for Toxic Substances and Disease Registry (NCEH/ATSDR). While serving as Director, Dr. Breysse led a staff of over 1,000 professionals, serving a national leadership role in the CDC's efforts to address critically important national environmental health issues, including responding to the Flint water crisis; addressing widespread per- and polyfluorinated alkyl substances (PFAS) contamination; responding to hurricanes, wildfires, and other natural disasters; conducting assessments of drinking water contamination and health at Camp Lejeune, NC; revising cancer cluster investigation guidelines; and responding to an outbreak of lung injury associated with the use of E-cigarette (vaping) products. Through NCEH's Climate and Health Program, Dr. Breysse provided leadership support to state and local health departments as they responded to many climate-related health concerns such as increasing occurrences of harmful algal blooms, extreme heat and cold events, and wildfire-associated air pollution. Working with the EPA, Dr. Breysse and staff developed guidance on wildfire smoke community protections. Dr. Breysse led the development of the first CDC-wide Climate and Health Framework. After a productive and exciting eight-year tenure, Dr. Breysse retired from governmental service in 2022.

Jorge Calzada

Jorge Calzada is the Acting Associate Director for Platforms within the Office of Public Health Data, Surveillance, and Technology at the Centers for Disease Control and Prevention (CDC). The new office within CDC seeks to advance the Public Health Data Strategy and Data Modernization Initiatives by

bringing together technology leaders with public health domain experts. Prior to joining CDC, Jorge has spent the last 20 years building data science and machine learning expertise for organizations across several different sectors including Energy, Market Research, and Artificial Intelligence Software Start-ups. Jorge received his undergraduate degree in Operations Technology from Northeastern University, a Master of Science in Information Systems also from Northeastern University, and a Master of Business Administration from the Massachusetts Institute of Technology.

Suzanne Dorsey, PhD

Dr. Suzanne Dorsey is Deputy Secretary of the Maryland Department of the Environment. She manages a portfolio that encompasses regulation and enforcement of state and federal environmental laws, multi-jurisdictional restoration projects, climate policy and environmental justice. Before this role, Dr. Dorsey worked with the agency's Water and Science Administration on Chesapeake Bay restoration and on major issues that require cross-agency collaboration on climate resiliency. She previously was executive director of the Harry R. Hughes Center for Agro-Ecology at the University of Maryland and also was executive director of the Bald Head Island Conservancy and Smith Island Land Trust for 11 years. Dr. Dorsey has been a former commissioner of the North Carolina Division of Coastal Management and a professor at University of North Carolina at Wilmington and Salem College. She has her bachelor's degree in biology from Drew University, her master's degree in marine-estuarine environmental science from University of Maryland and her Ph.D. in oceanography from State University of New York at Stony Brook.

Marzyeh Ghassemi, PhD

Dr. Marzyeh Ghassemi is an Assistant Professor at MIT in Electrical Engineering and Computer Science (EECS) and the Institute for Medical Engineering & Science (IMES), and a Vector Institute faculty member holding a Canadian CIFAR AI Chair and Canada Research Chair. Professor Ghassemi focuses on creating and applying machine learning to understand and improve health in ways that are robust, private and fair. Professor Ghassemi has previously served as a NeurIPS Workshop Co-Chair and General Chair for the ACM Conference on Health, Inference and Learning (CHIL). She also founded the non-profit

Association for Health Learning and Inference. Professor Ghassemi has published across computer science and clinical venues, including NeurIPS, KDD, AAAI, MLHC, JAMIA, JMIR, JMLR, AMIA-CRI, Nature Medicine, Nature Translational Psychiatry, and Critical Care. Her work has been featured in popular press such as Fortune, MIT News, NVIDIA, and The Huffington Post. She holds MIT affiliations with the Jameel Clinic and CSAIL. Professor Ghassemi holds a Herman L. F. von Helmholtz Career Development Professorship, and was named a CIFAR Azrieli Global Scholar and one of MIT Tech Review's 35 Innovators Under 35. Previously, she was a Visiting Researcher with Alphabet's Verily and an Assistant Professor at University of Toronto. Prior to her PhD in Computer Science at MIT, she received an MSc. degree in biomedical engineering from Oxford University as a Marshall Scholar, and B.S. degrees in computer science and electrical engineering as a Goldwater Scholar at New Mexico State University.

Susan Gregurick, PhD

Dr. Susan K. Gregurick is Associate Director for Data Science and Director of the Office of Data Science Strategy at the National Institutes of Health (NIH). Under Dr. Gregurick's leadership, the ODSS leads the implementation of the NIH Strategic Plan for Data Science through scientific, technical, and operational collaboration with the institutes, centers, and offices that comprise NIH. Dr. Gregurick advance research in computational biology, biophysics and data sciences, mathematical and biostatistical methods, and biomedical technologies in support of the NIGMS mission to increase our understanding of life processes. Dr. Gregurick received the 2020 Leadership in Biological Sciences Award from the Washington Academy of Sciences for her work in this role. She was instrumental in the creation of the ODSS in 2018 and served as a senior advisor to the office until being named to her current position. Dr. Gregurick received her undergraduate degree in chemistry and mathematics from the University of Michigan and her Ph.D. in physical chemistry from the University of Maryland. She completed a Lady Davis postdoctoral fellowship at Hebrew University in Israel and a Sloan postdoctoral fellowship at the University of Maryland's Center for Advanced Research in Biotechnology, now the Institute for Bioscience & Biotechnology Research, in Shady Grove, Md.

Lorenzo Hankla, MS

Lorenzo Hankla --who goes by "Loren"-- is the Wearables Program Lead at the Army's Joint Program Executive Office for Chemical, Biological, Radiological, and Nuclear Defense, headquartered in Aberdeen Proving Ground, MD. As Program Lead he oversees the Department of Defense's largest single investment in wearable technology, which has a specific focus on using physiological monitoring capabilities to transform how the Department monitors and predicts warfighter readiness, health, and performance. He holds a bachelor of science in mechanical engineering from the Georgia Institute of Technology and a master of science in mechanical engineering from the University of South Florida. When he's not busy working, he has his hands full raising two small children with his loving wife near the national capital region.

Heidi Hanson, PhD

Dr. Heidi Hanson is Group Lead of the Biostatistics and Multiscale Systems Modeling Group in the Computing and Computational Sciences Directorate at Oak Ridge National Laboratory (ORNL). Her research is focused on disentangling the interactions of genetic and environmental influences on disease risk throughout the life course. Dr. Hanson’s training and experience are in the fields of demography, statistics, familial analyses, biomedical informatics, -omics, and life course epidemiology and she aims to link her findings to clinical measures that can be used to improve precision strategies for screening and treatment. She is currently the technical lead on DOE-National Cancer Institute MOSSAIC, focused on advancing computing, predictive machine learning/deep learning (ML/DL) models, and large-scale computational simulations for NCI-supported cancer research. Previous projects have taken advantage of large health databases such as the UPDB, CMS Medicare, Demographic Health Survey, National Health and Nutrition Examination Survey (NHANES), and the Surveillance, Epidemiology, and End-Results (SEER) Program. Dr. Hanson holds adjunct appointments in the Department of Population Health Sciences, Sociology, and Surgery at the University of Utah. She completed her bachelor’s degree in Behavioral Science and Health and MS and PhD in Sociology at the University of Utah.

Thomas Hartung, MD, PhD

Dr. Thomas Hartung is the Doerenkamp-Zbinden-Chair for Evidence-based Toxicology in the Department of Environmental Health and Engineering at Johns Hopkins Bloomberg School of Public Health, Baltimore, with a joint appointment at the Whiting School of Engineering. He also holds a joint appointment for Molecular Microbiology and Immunology at the Bloomberg School. He is adjunct affiliate professor at Georgetown University, Washington D.C. In addition, he holds a joint appointment as Professor for Pharmacology and Toxicology at University of Konstanz, Germany; he also is Director of Centers for Alternatives to Animal Testing (CAAT, http://caat.jhsph.edu) of both universities. As PI, he headed the Human Toxome project funded as an NIH Transformative Research Grant. He is Chief Editor of Frontiers in Artificial Intelligence. He is Consulting Vice-President of AxoSim Inc., New Orleans. He is the former Head of the European Commission's Center for the Validation of Alternative Methods (ECVAM), Ispra, Italy, and has authored more than 640 scientific publications.

Janet Haven

Janet Haven is Executive Director of Data & Society and member of the National Artificial Intelligence Advisory Committee, which advises President Biden and the National AI Initiative Office on a range of issues related to artificial intelligence. She has worked at the intersection of technology policy, governance, and accountability for twenty years, both domestically and internationally. Before joining Data & Society, where she previously served as Director of Programs and Strategy, Janet spent more than a decade at the Open Society Foundations. There, she oversaw funding strategies and grant-making related to technology's role in strengthening civil society, and played a substantial role in shaping the field of data and technology governance. Janet started her career in technology start-ups in Central Europe and lived in the region for more than ten years, deepening her understanding of the ways the internet and algorithmic technologies impact communities outside the US. She sits on the board of the Public Lab for Open Technology and Science and advises a range of non-profit organizations. She holds an MA from the University of Virginia, and a BA from Amherst College.

Joyce Ho, PhD

Dr. Joyce Ho is associate professor in the Computer Science Department at Emory University. Her research focuses on the development of novel data mining and machine learning algorithms to address problems in healthcare. Recent projects include identifying patient subgroups or phenotypes, integration of new streams of data, fusing different modalities of data (e.g., structured medical codes and unstructured text), and dealing with conflicting expert annotations. Joyce previously co-founded a successful healthcare analytics company (Accordion Health) and worked at Lawrence Livermore National Laboratory. She received a Ph.D. in Electrical and Computer Engineering from the University of Texas at Austin, and an M.A. and B.S. in Electrical Engineering and Computer Science from Massachusetts Institute of Technology.

Karim Lekadir, PhD

Dr. Karim Lekadir is a Ramon y Cajal Researcher and Director of the Artificial Intelligence in Medicine Lab at the University of Barcelona. His current research focuses on the development of data science and machine learning approaches for the analysis of large-scale biomedical data, including imaging, clinical, lifestyle, and mobile data. The software he developed during his PhD for cardiac functional quantification has been CE marked and commercialized by CMRtools, and is now used in more than 250 clinical centers worldwide. He is the Coordinator of the following Horizon 2020 projects: euCanSHare, developing a big data platform for cardiovascular research; EarlyCause, which investigates multi-morbidity using experimental and data science approaches; and EuCanImage, which is building a federated artificial intelligence environment for cancer imaging. He is also work package leader in the longITools H2020 project, developing a mobile app for cardio-metabolic risk prediction based on exposome data. In addition, Karim is General Chair for the MICCAI 2024 Conference (Medical Image Computing and Computer-Assisted Intervention) which for the first time will take place in Marrakesh, Morocco. He is an Associate Editor of IEEE Transactions on Medical Imaging. He holds a PhD from Imperial College London (UK) and was previously a visiting scholar at Stanford University (USA).

Gengchen Mai, PhD

Dr. Gengchen Mai is Assistant Professor for the Department of Geography at the University of Georgia. He is also an Affiliated Professor of UGA School of Computing, UGA AI Institute, and UGA Institute for Integrative Precision Agriculture. Dr. Mai is specialized in Spatially Explicit Machine Learning, Geospatial Artificial Intelligence, Geographic Knowledge Graphs, Geographic Question Answering, Spatial Data Science, and Geographic Information Science. He received a Ph.D. in Cartography and Geographic Information Science from the University of California, Santa Barbara. Before coming to UGA, he was a postdoc at Stanford Artificial Intelligence Lab, Department of Computer Science, Stanford University. He is the recipient of many prestigious awards including AAG 2021 Dissertation Research Grants, AAG 2022 William L. Garrison Award for Best Dissertation in Computational Geography, AAG 2023 J. Warren Nystrom Dissertation Award, Top 10 WGDC 2022 Global Young Scientist Award, and the Jack and Laura Dangermond Graduate Fellowship.

Lucila Ohno-Machado, MD, PhD

Lucila Ohno-Machado is the Deputy Dean for Biomedical Informatics and the Chair of Biomedical Informatics and Data Science at the Yale School of Medicine. Her research focuses on privacy - preserving distributed analytics for healthcare and biomedical sciences. Prior to this role, she was Associate Dean for Informatics and Technology, and the founding chair of the UCSD Health Department of Biomedical Informatics at UCSD, where she led a group of faculty with diverse backgrounds in medicine, nursing, informatics, and computer science. Also, she is the PI for the California Precision Medicine Consortium for the NIH All of Us Research Program. Dr. Ohno-Machado is an elected member of the American College of Medical Informatics, the American Institute for Medical and Biological Engineering, the American Society for Clinical Investigation and the National Academy of Medicine. She directed the patient-centered Scalable National Network for Effectiveness Research, a large clinical data research network covering more than 40 million patients and 14 healthcare systems, and she was one of the founders of UC-Research exchange, a clinical data research network that connected the data warehouses of the five University of California medical centers. She received her medical degree from the University of São Paulo and her doctoral degree in medical information sciences and computer science from Stanford.

Chirag J Patel, Ph.D.

Dr. Chirag Patel is Associate Professor in the Department of Biomedical Informatics at Harvard Medical School (HMS). His primary research interests include developing multi-scale computational and data science methods to dissect the role of environmental exposures and genetic factors in complex traits and disease, with an emphasis on the trajectory from obesity to diabetes and its complications. His portfolio is supported by the National Science Foundation and the National Institutes of Health (National Institutes of Environmental Health Sciences and National Institutes of Allergy and Infectious Disease). Dr. Patel is a leader in "exposome" science, developing methods to map systems of dietary and environmental exposure factors with disease. He is an active researcher in "meta-science", studying the science of science, with an affiliation and active collaboration in the Meta-Research Innovation Center at Stanford University (METRICS). His work in meta-science has led to participation in an international committee developing recommendations for dietary intake. Dr. Patel was a previous member (2014-2017) of the National Academies of Sciences, Engineering, and Medicine's Standing Committee on Emerging Sciences for Environmental Health Decisions. He serves as an active member of HMS's faculty council. He received his Ph.D. in biomedical informatics from Stanford University.

Akane Sano, PhD

Dr. Akane Sano is an Assistant Professor at Rice University in the Department of Electrical Computer Engineering, Computer Science, and Bioengineering. She directs the Computational Wellbeing Group and is a member of Rice Digital Health Initiative. Her research includes data science, machine learning, and human-centered intelligent systems for health and wellbeing and spans the field of affective computing, ubiquitous and wearable computing, and biobehavioral sensing and analysis/modeling. She has been developing tools, algorithms, and systems to measure, forecast, understand and improve health and wellbeing using multimodal data from mobile and wearable devices in daily life settings, and clinical assessment. Recent awards include the NSF Career Award, the Best of IEEE Transactions on Affective Computing 2021, and the Best Paper Award at IEEE BHI 2019 conference. Sano received a B.Eng and M.Eng from Keio University, Japan and a Ph.D. from the Massachusetts Institute of Technology.

Nicholas Skaff, PhD

Dr. Nicholas Skaff is a Health Scientist at the Centers for Disease Control and Prevention. Working with the Environmental Public Health Tracking Program, he applies principles of ecology, epidemiology, and data science to develop methods, tools, and data products that support the surveillance of environmental hazards and exposures. Previously, as a PhD student at Michigan State University and postdoctoral scholar at the University of California, Berkeley, he applied machine learning methods to understand the environmental drivers of West Nile virus transmission. Dr. Skaff is deeply interested in topics at the intersection of environmental health, data science, and machine learning.

Eric Topol, MD

Dr. Eric Topol is Professor, Molecular Medicine, Executive Vice-President of Scripps Research, and Founder and Director of the Scripps Research Translational Institute. He has published over 1300 peer-reviewed articles, with more than 300,000 citations, elected to the National Academy of Medicine, and is one of the top 10 most cited researchers in medicine. His principal scientific focus has been on the use of genomic and digital data, along with artificial intelligence, to individualize medicine. He is also a practicing cardiologist. In 2016, Topol was awarded a $207M grant from the NIH to lead a significant part of the Precision Medicine (All of Us) Initiative, a prospective research program enrolling 1 million diverse participants in the US. Prior to coming to Scripps in 2007, he led the Cleveland Clinic to become the #1 center for heart care and was the founder of a new medical school there. Topol was commissioned by the UK 2018-2019 to lead planning for the National Health Service's integration of AI and new technologies. He has published 3 bestseller books on the future of medicine: The Creative Destruction of Medicine, The Patient Will See You Now and latest Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again.

Rick Woychik, PhD

Dr. Rick Woychik became the Director of the National Institute of Environmental Health Sciences (NIEHS), one of the National Institutes of Health (NIH), and the National Toxicology Program (NTP) on

June 7, 2020. In these roles, Woychik oversees federal funding for biomedical research to discover how the environment influences human health and disease. Woychik and NIEHS/NTP staff receives input from several advisory boards and councils to accomplish this significant task. Prior to becoming Director and since 2011, Woychik served as Deputy Director of NIEHS. In this role, he assisted the former NIEHS Director, Linda Birnbaum, Ph.D., in the formulation and implementation of plans and policies necessary to carry out the NIEHS missions and the administrative management of NIEHS.

Aidong Zhang, PhD

Dr. Aidong Zhang is Thomas M. Linville endowed professor of Computer Science in the School of Engineering and Applied Sciences at University of Virginia (UVA). She also holds joint appointments with the Department of Biomedical Engineering and School of Data Science at University of Virginia. Dr. Zhang's research interests include machine learning, data science, bioinformatics, and health informatics. Dr. Zhang served as a program director at the National Science Foundation from 2015 to 2018. Dr. Zhang has authored more than 380 refereed publications. She was the Editor in-Chief of the IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) from 2017 to 2021 and served as the founding Chair of ACM Special Interest Group on Bioinformatics and Computational Biology (SIGBio) from 2011 to 2015 and also served as the Chair of its advisory board from 2015 to 2018. She was also the founding and steering chair of ACM international conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB) from 2010 to 2019. She was General Co-Chair for ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD2022). Dr. Zhang is a fellow of ACM (Association for Computing Machinery), AIMBE (the American Institute for Medical and Biological Engineering), and IEEE (Institute of Electrical and Electronics Engineers). Dr. Zhang received her Ph.D. in computer science from Purdue University.