Report of the ACD Working Group Ad Hoc Virtual Meeting on AI/ML Electronic Medical Records for Research Purposes

Special Meeting of the Advisory Committee to the Director (ACD)

May 6, 2021

Lawrence A. Tabak, DDS, PhD
Principal Deputy Director, NIH
Department of Health and Human Services
December 2019 – ACD Artificial Intelligence WG

Rediet Abebe, PhD
Cornell

Greg Corrado, PhD
Google

Kate Crawford, PhD
AI Now Institute

Barbara Engelhardt, PhD
Princeton

David Glazer
Verily (Co-Chair)

David Haussler, PhD
USCS

Dina Katabi, PhD
MIT Computer Science & AI Lab

Daphne Koller, PhD
insitro

Anshul Kundaje, PhD
Stanford University

Eric Lander, PhD
Broad Institute

Jennifer Listgarten, PhD
Berkeley

Michael McManus, PhD
Intel

Lawrence Tabak,
DDS, PhD
NIH (Co-Chair)

Serena Yeung, PhD
Harvard
The opportunity is huge
- including to drive discontinuous change

We need new data generation projects
- NOT business-as-usual

The single best way to attract the right people is with the right data
- “Show me the data”

The time to invest in ethics is now
- Before we dig a deeper hole
<table>
<thead>
<tr>
<th>Rediet Abebe, Ph.D.</th>
<th>Barbara Engelhardt, Ph.D.</th>
<th>Ryan Luginbuhl, M.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistant Professor of Computer Science</td>
<td>Associate Professor, Department of Computer Science</td>
<td>Principal, Life Sciences Division</td>
</tr>
<tr>
<td>University of California, Berkeley</td>
<td>Princeton University</td>
<td>MITRE Corporation</td>
</tr>
<tr>
<td>Atul Butte, M.D., Ph.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor, Pediatrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priscilla Chan and Mark Zuckerberg Distinguished</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of California, San Francisco</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kate Crawford, Ph.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior Principal Researcher at Microsoft Research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labs – New York City</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visiting Chair of AI and Justice at the École</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normale Supérieure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miegunyah Distinguished Visiting Fellow at the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Melbourne</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joel Dudley, Ph.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chief Scientific Officer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tempus Labs, Inc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>David Glazer</td>
<td>Jianying Hu, Ph.D.</td>
<td>Brad Malin, Ph.D.</td>
</tr>
<tr>
<td>Engineering Verily</td>
<td>IBM Fellow; Global Science Leader</td>
<td>Accenture Professor Biomedical Informatics, Biostatistics,</td>
</tr>
<tr>
<td></td>
<td>AI for Healthcare</td>
<td>and Computer Science</td>
</tr>
<tr>
<td></td>
<td>Director, Center for Computational Health</td>
<td>Vanderbilt University</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dina Katabi, Ph.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andrew and Erna Viterbi Professor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MacArthur Fellow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts Institute of Technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abel Kho, M.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Director, Institute for Public Health and Medicine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northwestern University</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serena Yeung, Ph.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistant Professor, Biomedical Data Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistant Professor, by courtesy, of Computer Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and of Electrical Engineering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanford University</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ACD AI/ML Electronic Medical Records for Research Purposes

NIH MEMBERS

Francis Collins, M.D., Ph.D.
Director

Susan Gregurick, Ph.D.
Associate Director for Data Science

Helene Langevin, M.D.
Director, National Center for Complementary and Integrative Health

Jon Lorsch, Ph.D.
Director, National Institute of General Medical Sciences

Andrea Norris, M.B.A
Director, Center for Information Technology
Chief Information Officer

James Anderson, M.D., Ph.D.
Deputy Director for Program Coordination, Planning, and Strategic Initiatives

Patricia Brennan R.N., Ph.D.
Director, National Library of Medicine

Michael Chiang, M.D.
Director, National Eye Institute

Joshua Denny, M.D., M.S.
Chief Executive Officer, All of Us Research Program

Eric Dishman
Chief Innovation Officer, All of Us Research Program

Susan Gregurick, Ph.D.
Associate Director for Data Science

Helene Langevin, M.D.
Director, National Center for Complementary and Integrative Health

Jon Lorsch, Ph.D.
Director, National Institute of General Medical Sciences

Andrea Norris, M.B.A
Director, Center for Information Technology
Chief Information Officer

Joni Rutter, Ph.D.
deputy director, national center for advancing translational sciences

Tara Schwetz, Ph.D.
Associate Deputy Director

Norman Sharpless, M.D.
Director, National Cancer Institute

Lawrence Tabak, D.D.S., Ph.D.
Principal Deputy Director

Carrie D. Wolinetz, Ph.D.
Acting Chief of Staff
Associate Director for Science Policy
Charge to the Working Group

- Identify unique research opportunities for NIH to apply resources in a practical way to make electronic health records (EHRs).
- Identify EHR research challenges that AI/ML could have the greatest impact.
- Determine barriers to the widespread use/deployment of AI/ML capabilities can NIH support help overcome.
- Identify the partners that would be needed and incented to help scale these capabilities.
Impactful AI/ML Research Challenges

- Start by addressing the needs of marginalized communities. If these communities remain an afterthought, the same inequities will be repeated.
- Address the significant errors, gaps, and racial and gender inequities in EHR data.
- Using problematic data for models will amplify the gaps. The foundation that the models are being built on must be addressed
  - Example: 50,000 COVID-19 patients, more than half of whom are Black, however, rigorous methods to analyze these data do not exist, because of confounding factors such as disparate treatment by physicians.
Impactful AI/ML Research Challenges

- Community-enabled real-world effort to encourage solicitation of networks of clinics/systems
- Create consortia among all groups that participate to enrich data collection and sharing
- Create clinical trials networks for the rapid validation of algorithms
- Analyses of dense data (e.g. medical images) and to study diseases that do not translate well from models
- Design methods to impute or fill in missing data, social determinants of health and actual outcomes
Barriers to the Widespread Use/Deployment

- Information infrastructure that is needed includes not only EHRs but also the network of large-scale data repositories and imaging studies
- Clinical data capture is human intensive, biased, incomplete, and sparse
- EHRs provide only a partial picture of health and health care experience
- Inconsistent quality of the content, integrity of data, research usefulness of EHRs
- EHR accessibility is a serious problem
Addressing Barriers

- Create a body that could be the translator and take data and de-identify data, and then provide the data to the end users
- Create an infrastructure to facilitate designing and testing algorithms and comparing them across institutions
- Invest in earlier-stage research that private industry is reluctant to fund
- Clarity is needed on ownership and sharing of the ML produced after public shared data are used
Strategic Partnerships

• Who are the partners that would be needed and incented to help scale these capabilities
  • AI Industries, Computing Partners, Academic, Research, and Nontraditional Partners

• NIH investment needs to add meaningful complements to what is already being done by the community

• Focus on sustainable infrastructure through strategic public/private partnerships
Compelling Ideas Emerge

- Support for AI/ML work to redress the challenges of health disparities, health inequality, and minority health
- Develop the infrastructure and training of diverse researchers
- Catalyzing access to high quality diverse data sets
- Address the errors and biases in EHR data and linking EHR with Social Determinants of Health
### Quality of the content in EHRs
- Innovative ways to use AI strategies (robotics, computer vision) when capturing the patient experience
- Computational approaches to debiasing data to avoid perpetuating biased understanding of health and the patient experience

### Research usefulness of EHRs
- Identify high priority problems worthy of solutions
- Model validity and verification
- Security, equity, and privacy

### Tools for clinical care
- Integrative data models that enrich understanding of health beyond the data available in the EHR
- Methodology that supports translation and transparency
- Ensure that models are fit for purpose

### Integrity of data in EHRs
- Advance the use of standards in the capture of clinical data
- Deploy AI methods to support computational coding at scale
NIH...

Turning Discovery Into Health

Lawrence.Tabak@nih.gov