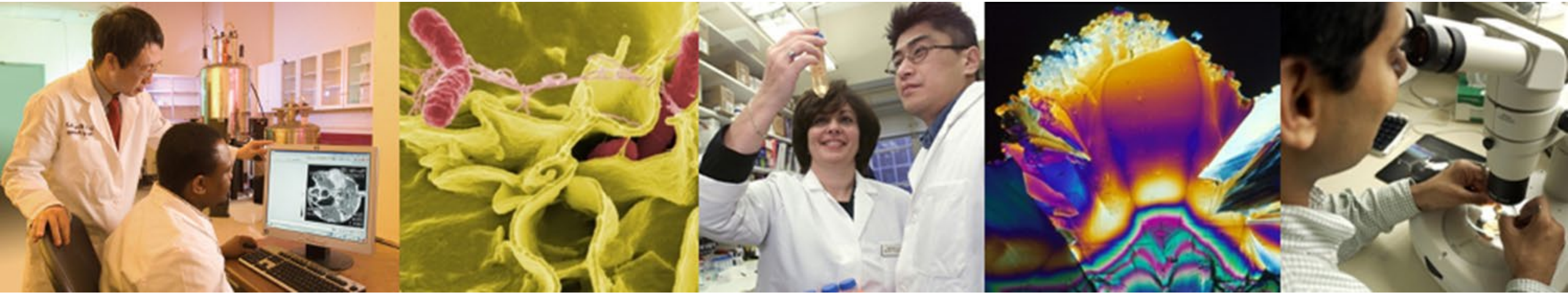
Artificial Intelligence Working Group Update

117th Meeting of the Advisory Committee to the Director (ACD)

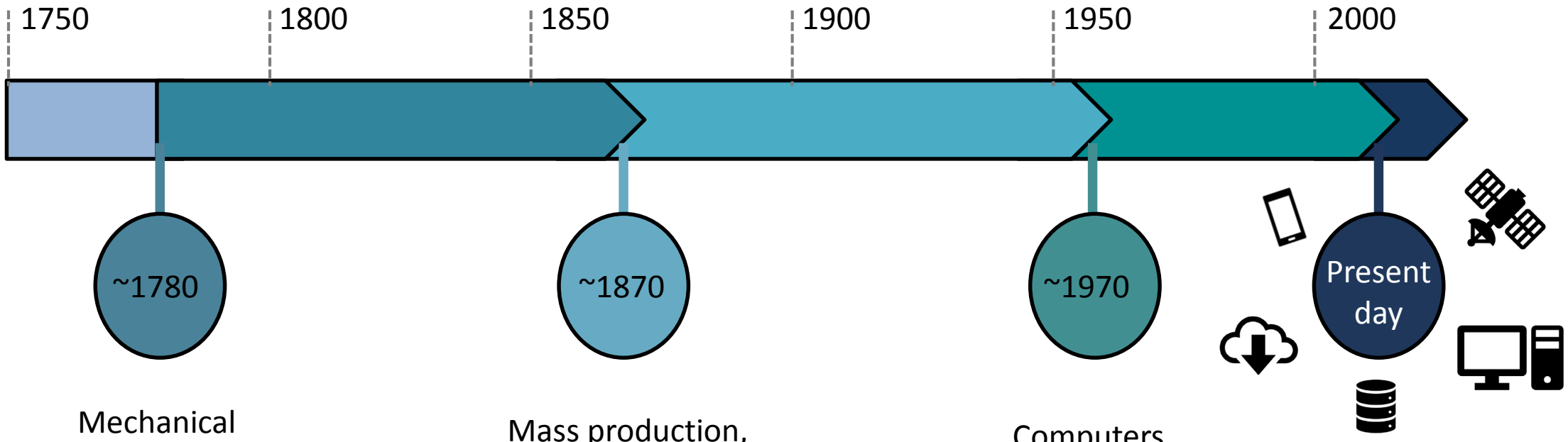
December 14, 2018



David Glazer
Engineering Director, Verily

Lawrence A. Tabak, DDS, PhD
Principal Deputy Director, NIH
Department of Health and Human Services

Industrial Revolutions



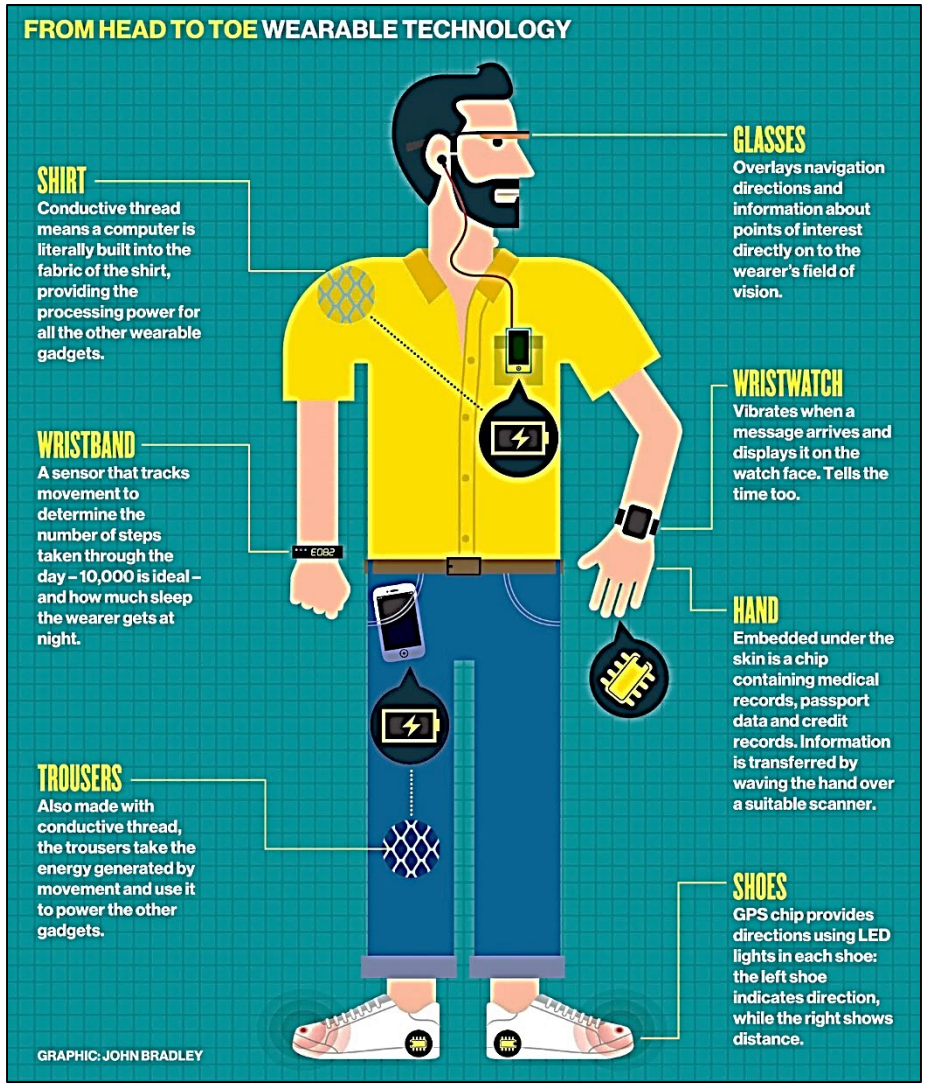
Mechanical production, railroads, steam power

Mass production, assembly line, electric

Computers, automated production, electronics

Big data, artificial intelligence, clouds, robotics…

We Generate Enormous Volumes of Data Daily



https://people.rit.edu/sml2565/iimproject/wearables/index.html

Biomedical Big Data

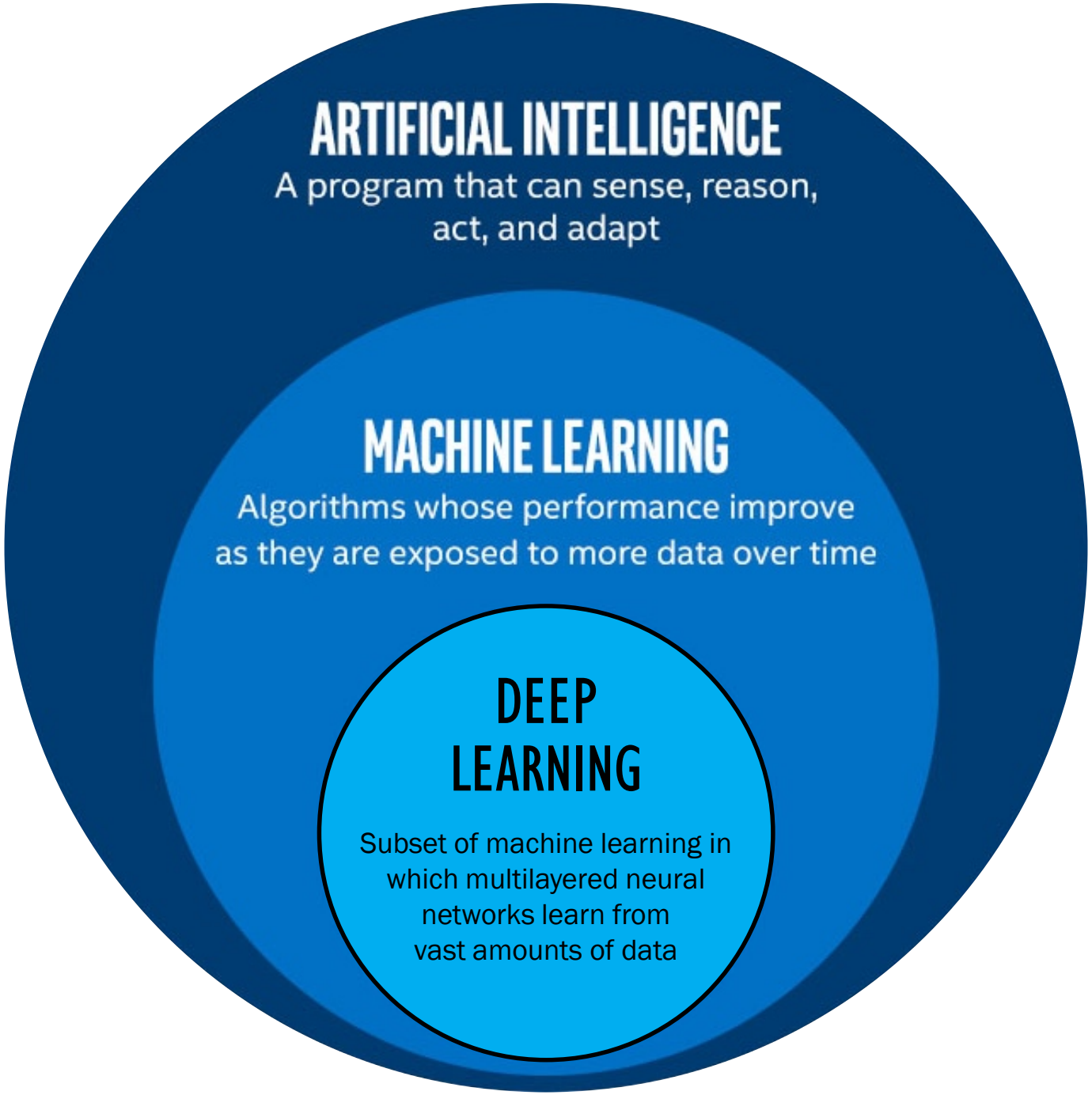
- Multimodal data being generated by researchers, hospitals, and mobile devices
- Exponential growth in biological sciences data production
  - Doubling every 10 months!
  - Imaging, phenotypic, molecular, exposure, health, behavioral, and many other types
- Challenges: amount of information, lack of organization and access to data and tools, and insufficient training in data science methods
- Opportunities: maximize the potential of existing data, enable new directions for research, increase accuracy, support precision methods for healthcare.

Data Science at NIH: A Snapshot

- CIT supports a 100GB Network moving 4PB of data per day
- Datasets and resources
  - List of extramural programs generating datasets (only a subset)
  - Datasets supported across IC and topic area
  - Range in size from several hundred terabytes to several petabytes
    - SRA and dbGaP, ~15 PB of genomic sequence data
      - Controlled access ~8 PB
      - Open access ~6 PB
    - GTEx, ~200 TB

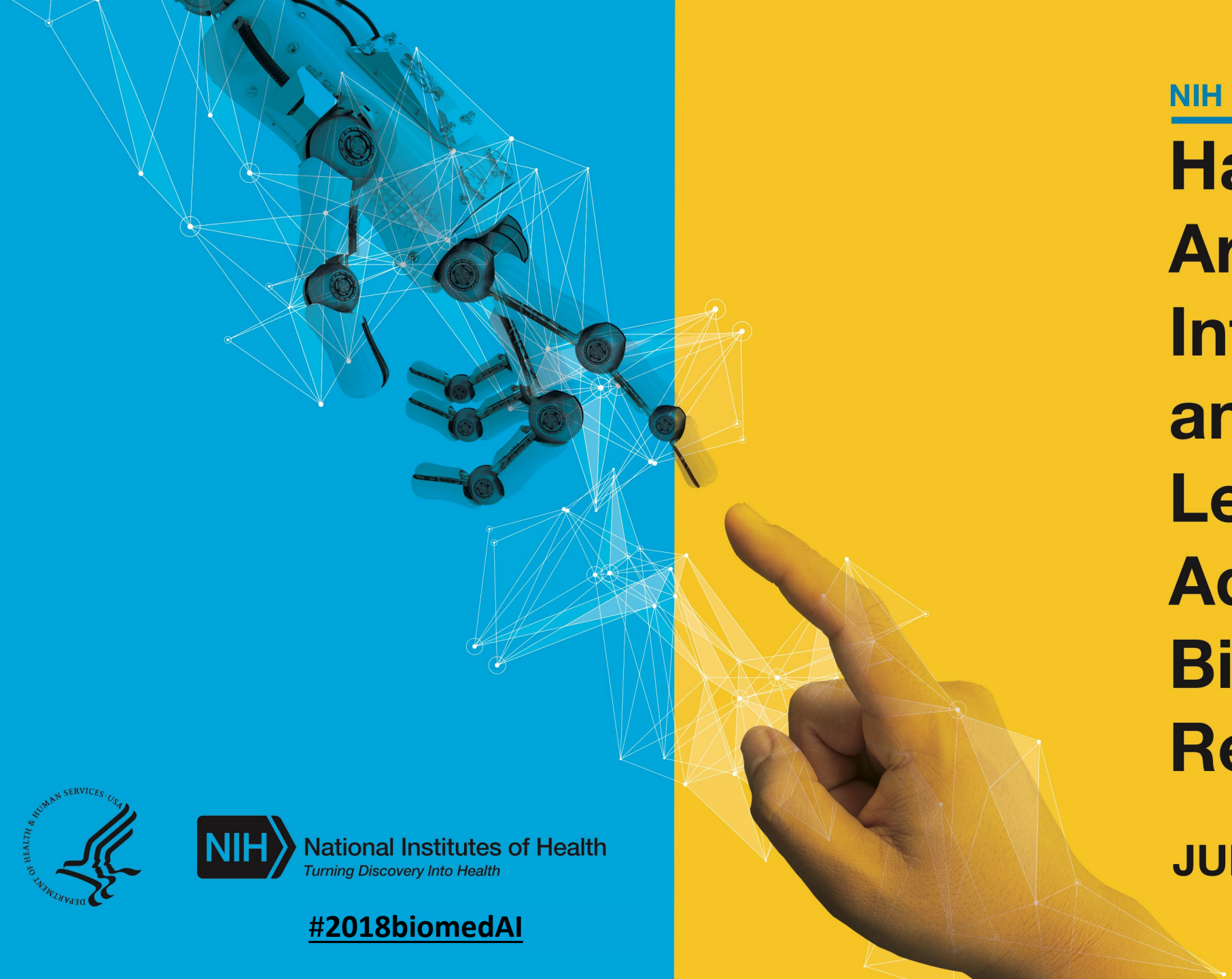
DATASET	Primary IC
ABCD (Adolescent Brain Cognitive Development)	MH
Accelerating Medicine Partnership - Parkinson's Disease (AMP PD)	NS
Age-Related Eye Disease Study (AREDS2)	EY
All of Us Research Program	OD
BRAIN Initiative	many
Biomedical Translational Research Information System (BTRIS)	CC
dbGAP	NL
Framingham Studies	HL
Gabriella Miller Kids First Pediatric Research Program	CF/HL
Genotype-Tissue Expression (GTEx)	CF/HG
Cancer Genome Characterization Initiative (CGCI)	NCI
Analysis, Visualization, and Informatics Lab-space (AnVIL)	HG
Chest and Cardiac Image Archive	HL
Genetics of Alzheimer's Disease Project (NIAGADS)	NIA
RSNA Radiology Image Share	EB
The Cancer Genome Atlas Project (TCGA)	NCI
TOPMed	HL
Alliance for Genome Resources Model Organism Databases (MODs)	HG
ClinVar	NL
dbSNP	NL
ENCODE	HG
Gene Expression Omnibus (GEO)	NL
MACS/WIHS Longitudinal AIDS Data	AI
Neuroimaging Tools & Resources (NITRC)	EB
SRA	NL
UniProt	HG/GM

Every Day Artificial Intelligence Applications


ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

AI/ML/DL in Biomedicine

- Clinical applications
  - Pathology diagnostics
  - Dermatology, ophthalmology diagnostics
  - Radiology – mammograms, CXRs,…
  - Inferring treatment options for cancer
  - Robotic surgery
  - Natural language processing of EHR data
- Basic science applications
  - Interpretation of images: cryo-EM, confocal, etc.
  - Neuroscience and the BRAIN initiative
  - Genomics: variants and risk of disease, gene structure
  - Microbiome/metagenomics
  - Epigenomics: histone marks, TF binding, enhancers, DNA methylation



NIH WORKSHOP

Harnessing Artificial Intelligence and Machine Learning to Advance Biomedical Research

JULY 23, 2018

National Institutes of Health
Turning Discovery Into Health

#2018biomedAI

Newly Formed AI Working Group Members

AI Ethics expert, membership to be finalized

Rediet Abebe
Cornell

Greg Corrado
Google

David Glazer
Verily (Co-Chair)

Daphne Koller, PhD
Stanford

Eric Lander, PhD
Broad Institute

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

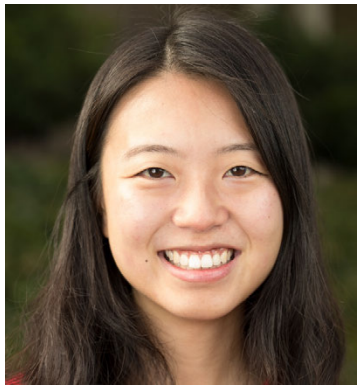
Michael McManus, PhD
Intel

Barbara Engelhardt, PhD
Princeton

Dina Katabi, PhD
MIT Computer Science & AI Lab

Anshul Kundaje
Stanford University

Jennifer Listgarten, PhD
Berkeley



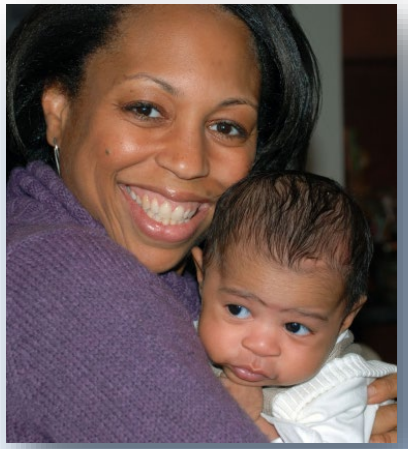
Serena Yeung, PhD
Harvard

Charge to the AI Working Group

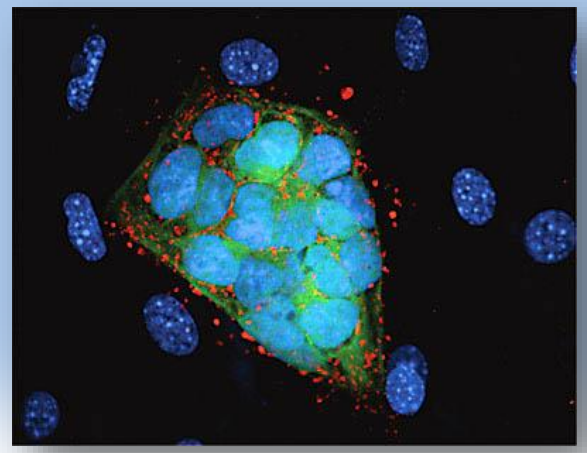
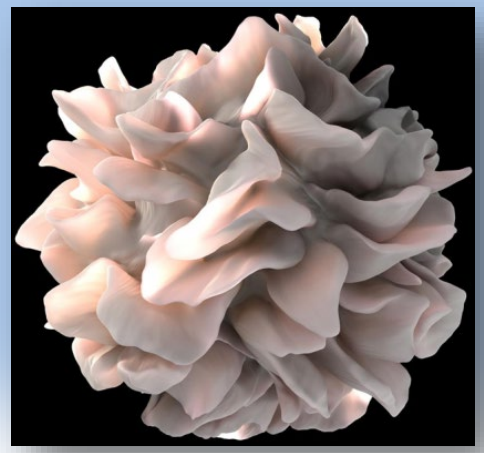
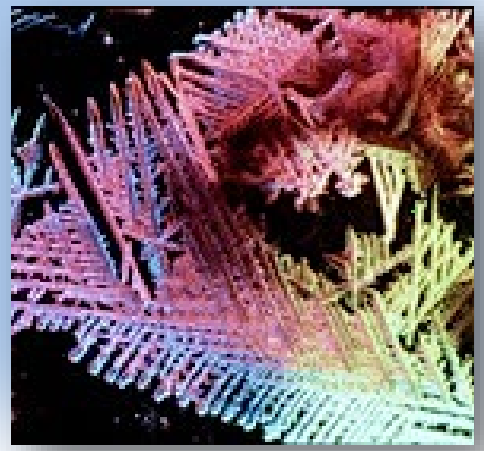
- Are there opportunities for cross-NIH effort in AI? How could these efforts reach broadly across biomedical topics and have positive effects across many diverse fields?
- How can NIH help build a bridge between the computer science community and the biomedical community?
- What can NIH do to facilitate training that marries biomedical research with computer science?
  - Computational and biomedical expertise are both necessary, but careers may not look like traditional tenure track positions that follow the path from PhD to post-doc to faculty
- Identify the major ethical considerations as they relate to biomedical research and using AI/ML/DL for health-related research and care, and suggest ways that NIH can build these considerations into its AI-related programs and activities

Timeline

- Interim recommendations June 2019 ACD meeting
- Final recommendations December 2019 ACD meeting
- Beyond 2019, the group will convene intermittently, as needed but infrequently, for updates and continued guidance



NIH…

Lawrence.Tabak@nih.gov

Turning Discovery Into Health