A National Academies of Sciences, Engineering, & Medicine Workshop: Policy Issues for Integrating Artificial Intelligence in Cancer Research and Care

Session 3:

Examples of AI Applications to Address the Distinctive Needs for Cancer Research

A moderated panel discussion

March 9-110, 2026
Washington, D.C.

Moderator: **Usama Fayyad**
Chairman, Open Insights
SVP AI & Data Strategy, Inaugural Director, Institute for Experiential AI @ Northeastern University & Professor of Computer Sciences

Open Insights™

EAI The Institute for Experiential AI
Northeastern University

Disclosures

- I represent the Northeastern University – a private not-for-profit institution
- The faculty members of Institute for EAI receive research funding from NIH, NSF, FDA, DARPA and many other public and private research funding agencies
- The Institute of EAI works with and seeks projects with companies (private and public) in the AI+Life Sciences area – we leverage such applied projects to drive new AI research as well as provide experiential learning opportunities to students from Northestern University and learners from partner organizations
- I am also affiliated as chairman of a company I founded in 2008: Open Insights. Historically Open Insights has worked with pharma and other Healthcare tech and manufacturing companies on data, data science and AI projects. The company has and seeks many clients in related spaces

Our Panelists - *Examples of AI Applications to Address the Distinctive Needs for Cancer Research*

Emily Boja, PhD
Office of Data Sharing
National Cancer Institute

Aaron Cohen, MD, MSCE
Flatiron Health; NYU Grossman School of Medicine

Eric Horvitz, MD, PhD
Microsoft Research

Etta D. Pisano, MD, FACR, FSBI
American College of Radiology

Usama Fayyad, Chairman & Founder, ***Open Insights***
at Northeastern University: SVP, AI & Data Strategy, Sr. Advisor to President, & Professor of CS
Inaugural Director, Institute for Experiential AI (EAI) at Northeastern University

Goal: Make AI and Data usable, useful, manageable - democratize the responsible use of AI across fields

Education

Large Orgs

Startups

Education

- **Ph.D. Computer Science & Engineering (CSE) in AI/Machine Learning**
- **MSE (CSE), M.Sc. (Mathematics)**
- **BSE (EE), BSE (Computer Engin)**

Academic achievements

- Fellow: Association for the Advancement of Artificial Intelligence (AAAI) and Association for Computing Machinery (ACM)
- Over 100 technical articles on data mining, data science, AI/ML, and databases.
- Over 20 patents, 2 technical books.

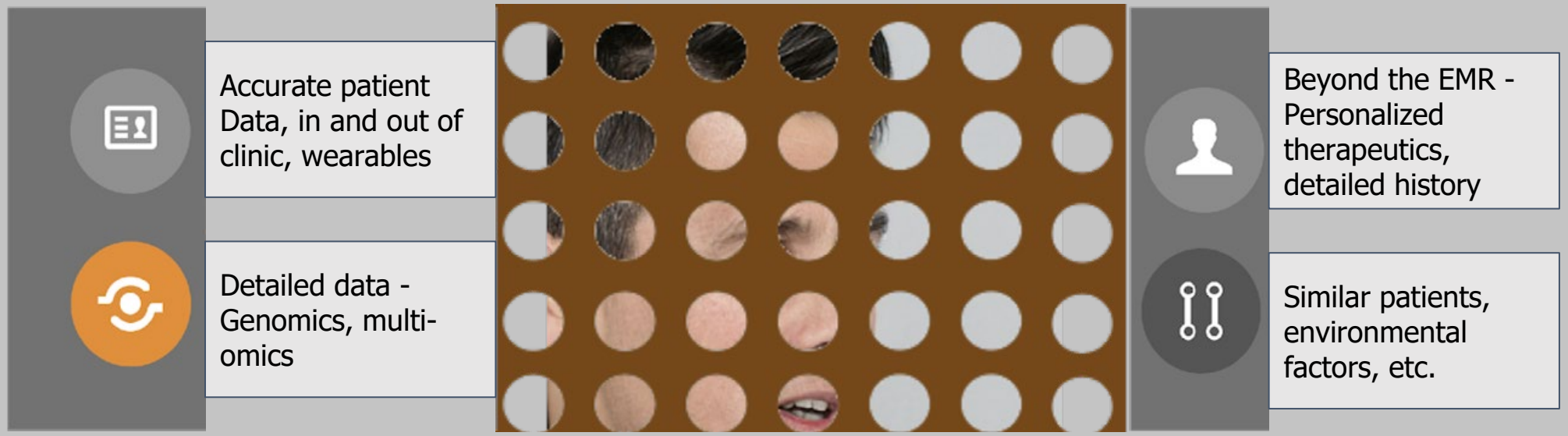
- First in industry: Chief Data Officer at Yahoo!
- First Global Chief Data Officer & Group Managing Director at Barclays Bank, London
- Chaired/started major conferences in Data Science, Data Mining, AI
- Founding Editor-in-Chief on two key journals in Data Science

Main take-away themes for this panel

1. Where are the realistic opportunities for AI impacting cancer research [acceleration, efficiency, discovery and evaluation of effectiveness] - especially discovery over an overwhelming flood of information and literature…
2. The critical role of Data as the perennial missing component and data sharing and availability always being an “afterthought”
3. While much of the research is focused on treatment and therapies, we are in an environment where Prevention is not getting the attention is should – can AI help change this dynamic?
4. Bridging the gap between the phenotype and the genotype (in fact multi-omics-type: “molecular phenotype” or “endophenotype”
5. How can we accelerate the research, especially with rapid leverage of AI to some pressing problems in cancer care?

How our Brain identifies Missing Values

Missing Data In health and cancer care



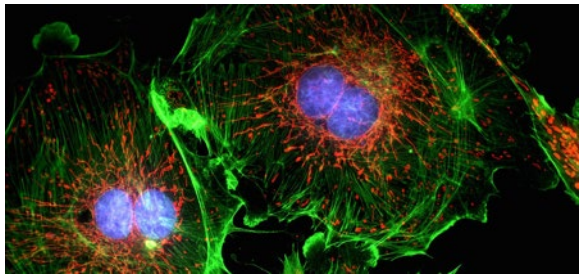
Can we use AI - to "complete" our view of the patient?
Can we track outside the clinical setting?
Can we close the loop and also think about preventative medicine?

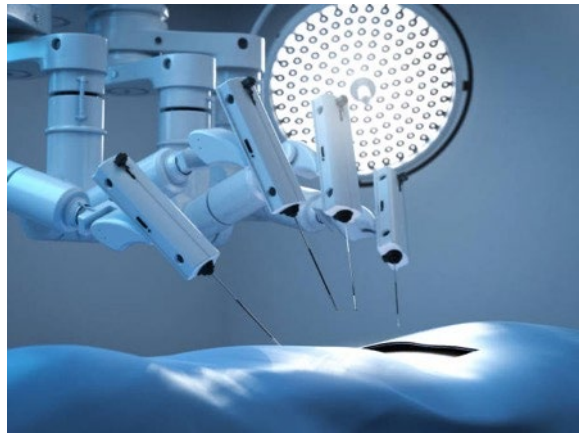
What Has Digital Transformation Looked Like in Health?

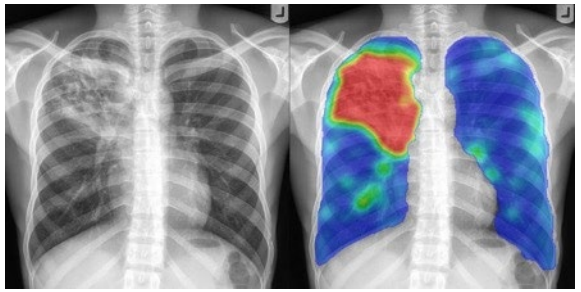
- **Slower** *digitization and digitalization* than other fields
- Great advances in device tech – from surgical to imaging
- Interesting examples of inconsistent adoption of digital & AI

Some sample examples of uses of AI:

	Routine Medical Diagnostics	Cell Imaging	The Omics
✓	Radiology: digitized and a lot of automation	Great advances in technology down to sub-single-cell	Great advances in Genomics, Proteomics, and Transcriptomics
✗	Pathology: still in the land of analog and little AI processing	Very few large-scale uses	Very little work on combining omics and advancing to metabolomics





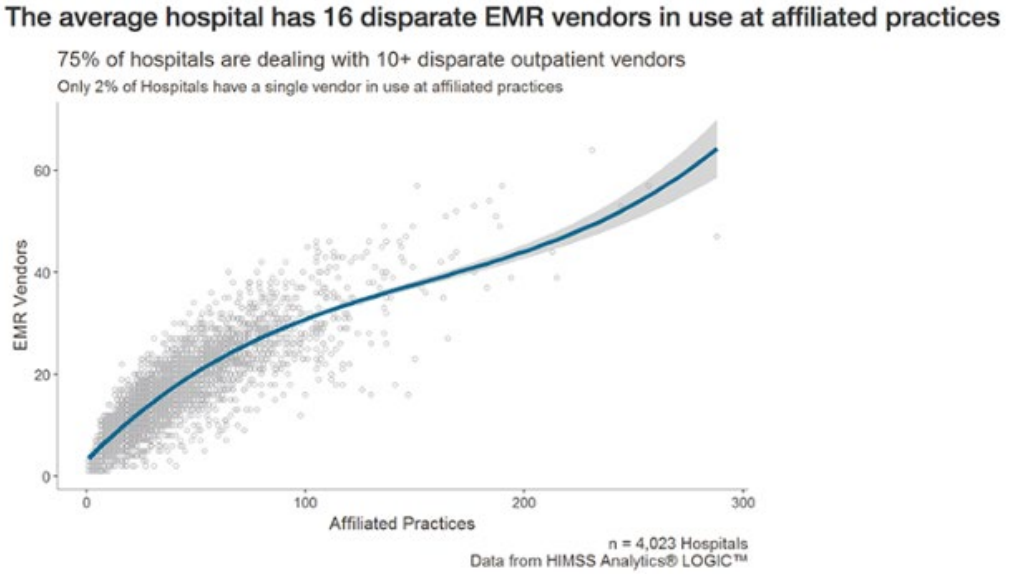


What About Electronic Medical Records?

- Mandated digital coding and medical notes
- Failure to standardize the Data
- No two EHR's look the same…
- Very fragmented space, and even the largest EHR systems do not inter-operate
- No real incentives to share data

This results in great difficulty to leverage AI/ML and automated analytics to help leverage the wealth of data

The average hospital has 16 disparate EMR vendors in use at affiliated practices Most hospitals have 10 EHR systems, and only 2% are down to only two EHR systems!



Meanwhile the platform is burning in the U.S.

- **Cost of healthcare is growing out of control as share of GDP**
- **> $5T total medical spend in U.S. (CMS 2021 – 10% annual growth),**
- **It is estimated that > $400B of administrative waste**
- **We need serious help in figuring out how to scale healthcare in a more economically sustainable model**
- **AI approaches hold a lot of promise to help**
  a. **In Healthcare in general**
  b. **In Health Sciences**
  c. **In Life Sciences and Drug Discovery**
  d. **In understanding diseases better, faster**
  e. **In evaluating therapy effectiveness better, faster**

AI could help, but there is a serious catch…

- **Working AI needs ML and Data**
- **Data requires digitization**
- **Data capture, representation, sharing, and management remains a largely unaddressed challenge in healthcare…**

Digitization Produces Much More Data

But most organizations are not equipped to effectively manage data as an asset

How do we make this Data a working asset?

New economy of Interactions is rich with unstructured data

90% of Data in any organization is UNSTRUCTURED

Without proper Data, AI cannot work: ML needs granular and high quality training data

BigData challenges are not just about size but structure & entity extraction

OVER TO OUR PANEL

- **6 minutes of introductory remarks by each panelists**
- **Initial discussion by panelists**
- **Audience Q&A**

Areas where AI Can Help

- NLP to leverage unstructured text data – LLM (large language models) and other open source methods for image and TS analysis
- Image analysis tools to leverage and retrieve related image data (*query by example, pattern recognition, etc.*)
- Graph-based and network representations
- Network Science models for understanding multi-factor interactions
- Multi-omics approaches to counter the single-omics traditions

Why Multi-omics?

- **Detect and understand interactions between different omics**
  - Genome or Exome
  - Proteomics
  - Transcriptomics
  - Metabolomics
- **Most work appears to focus on working within one "silo"**
  - Genomics and transcriptomics are "mature"
  - Metabolomics evolving and difficult but is critical in immuno-response and cell-response analysis
- **Example:** Exome or Genome "similar" patients have different responses to the same cancer therapeutic:
  - Respond well vs. no response?
  - Respond well but then stop responding after some time?

The resistance is also “cultural” and “social”

- **Clinical protocols are hard to change** (for good reason): from a Data Science/AI technologist perspective:
  - Typically simple
  - Typically ”outdated”
  - Typically do not leverage latest technology, science, or math (probabilistic modeling)
- **Consider an Example in Single-Cell Analytics**
  - Single cell metabolomics or in-vivo imaging/video
  - Can actually observe direct effect of therapy at cell-level – e.g. are tumor cells being attacked by immune cells?
  - Can evaluate effectiveness/impact in days (not weeks or months)
  - **How do you get uptake of this new evidence by a clinician?**
  - Typical outcome: follow the “weeks to months” observation cycle instead of adjusting therapy
    - ***What do we need to trust the micro and cell-level responses as much as we trust the “phenotype” observables?***