

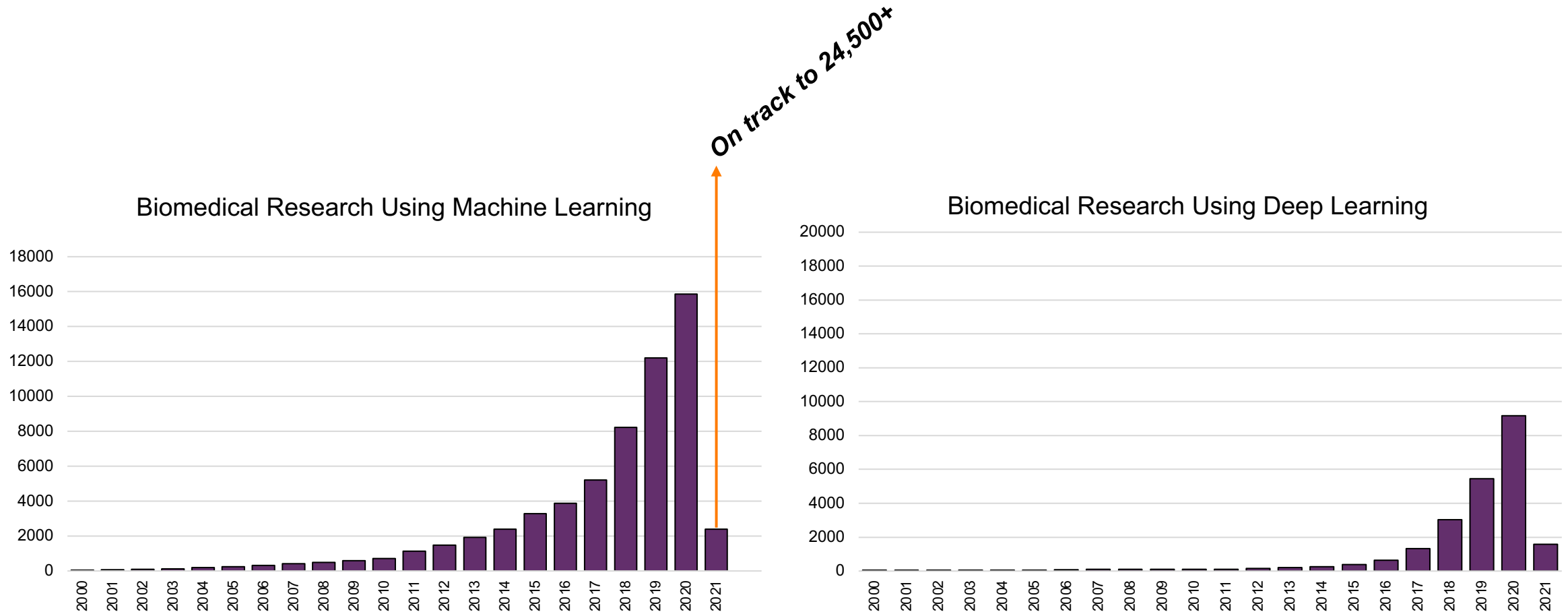
National Institutes of Health efforts in Advancing AI/ML in Biomedical Research & Healthcare

Chris M Hartshorn, Ph.D.
Acting Chief, CTSA Program Branch

*The role of advanced computation, predictive technologies, and big data analytics related to food and nutrition research – **NASEM Meeting**, 11 October 2023*



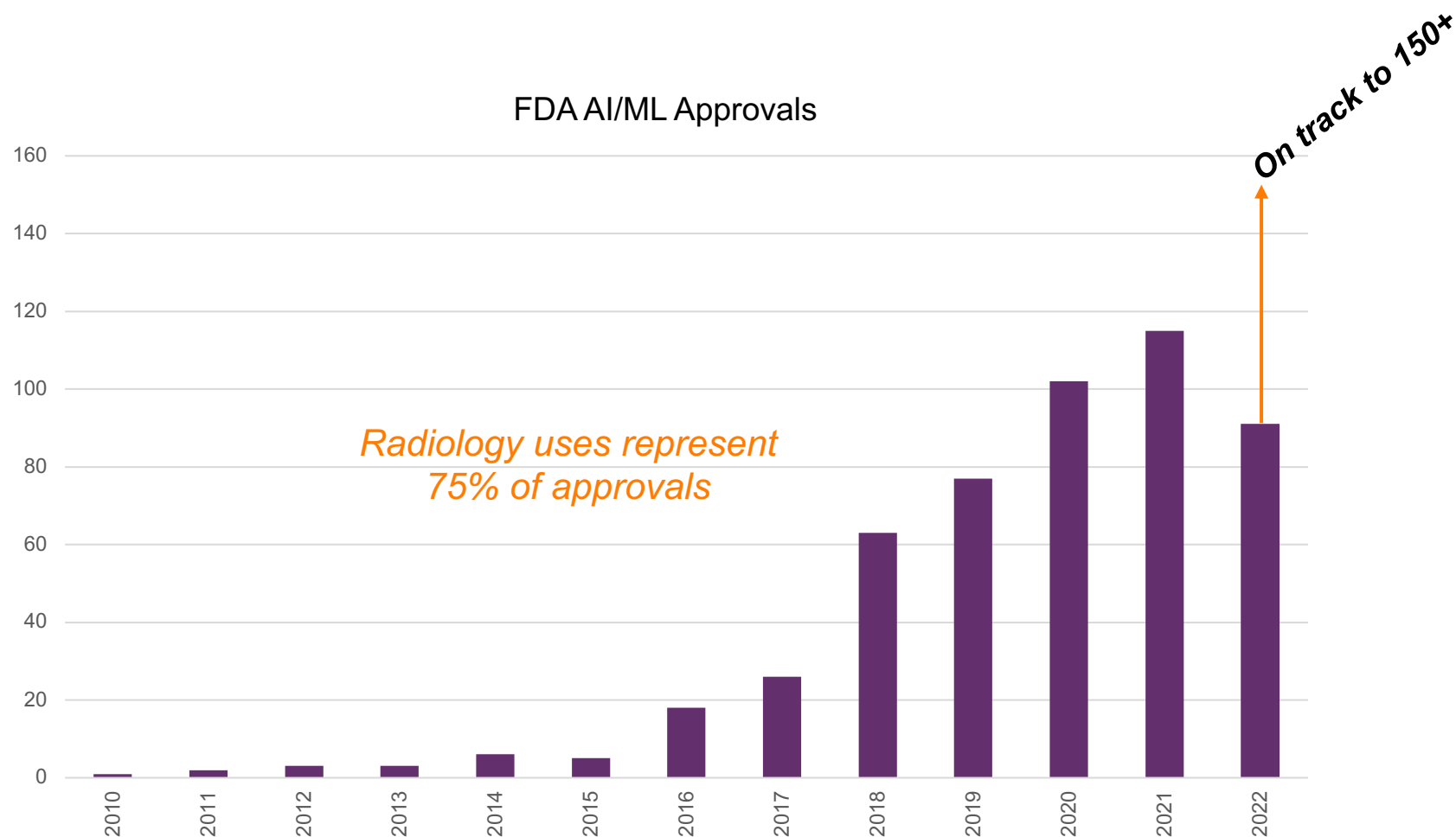
'AI' use in Clinical Research: Publication Acceleration - Rough Indicator



Pubmed publications by search terms as of Feb 1, 2021



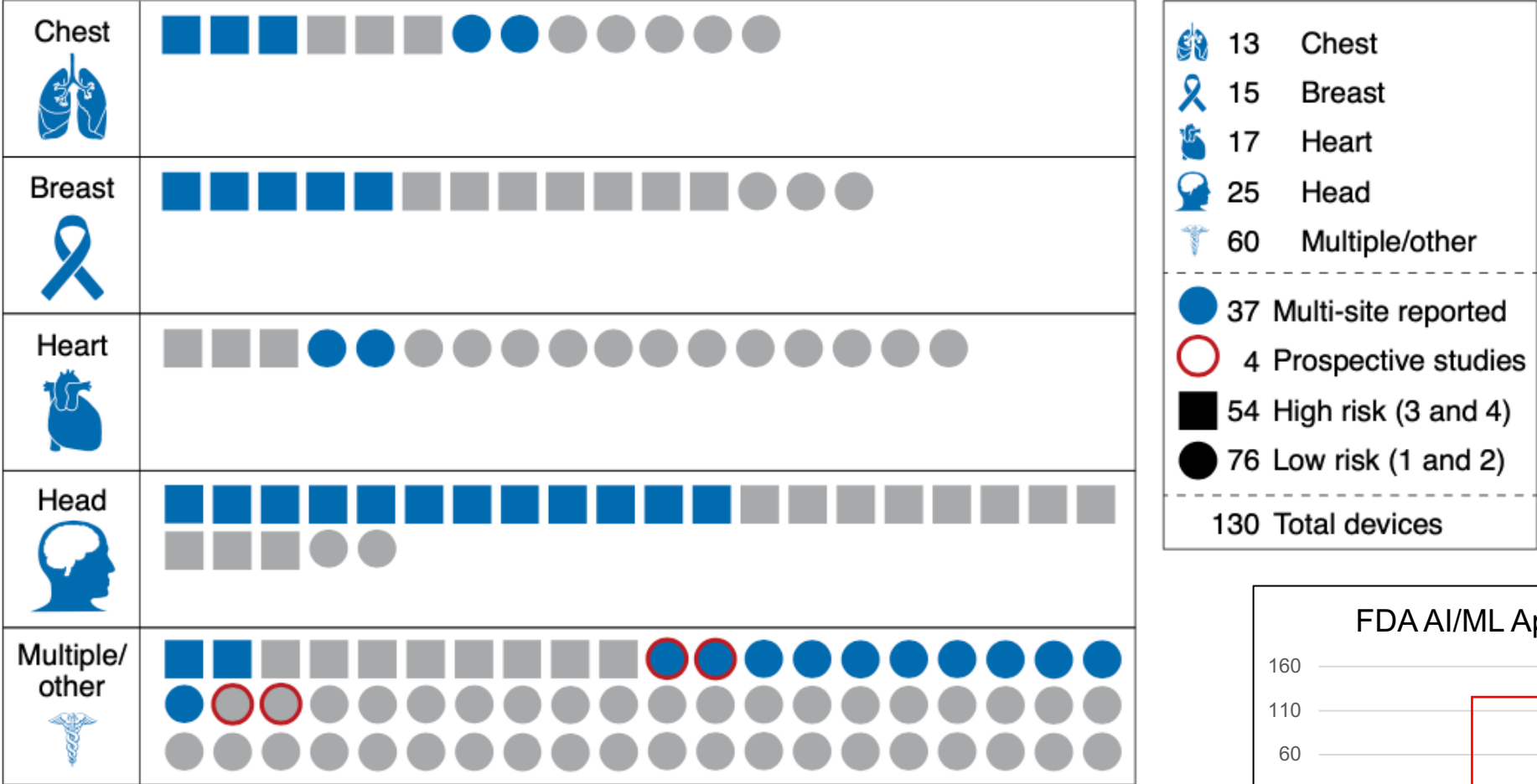
'AI' use in Clinical Research: Publication Acceleration + Regulatory Approvals = Substantial Indicator



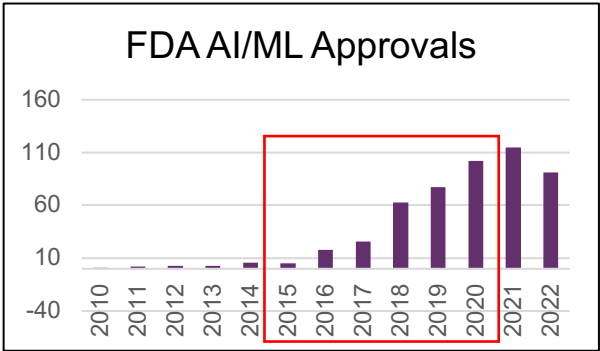
FDA identified AI/ML approvals as of July 29, 2022



Considering what we know of AI/ ML challenges – the rigor of data supporting current approved devices is



Wu et al, Nature Medicine 2021



Contemporary-to-Precision-to-Predictive Medicine: 'Weaving the Fabric' of all Medical Evidence Available

Major Challenge *Inter-modal/scale* **'Dot-Connecting'** *For AI*

Data Integration Methods

*Machine Learning
Algorithms and
modeling*

*Informed Dot-connecting /
Predictive Analytics ("AI")*

+ Patient Reporting , Comorbidity, and Environmental Factors

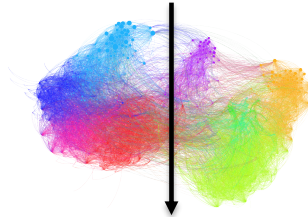
*Integration of imaging, EHR, lab assays, -omics, and other Real-world data, with respect
to time, will greatly enhance disease understanding and its earlier, proactive mitigation*

Ingest

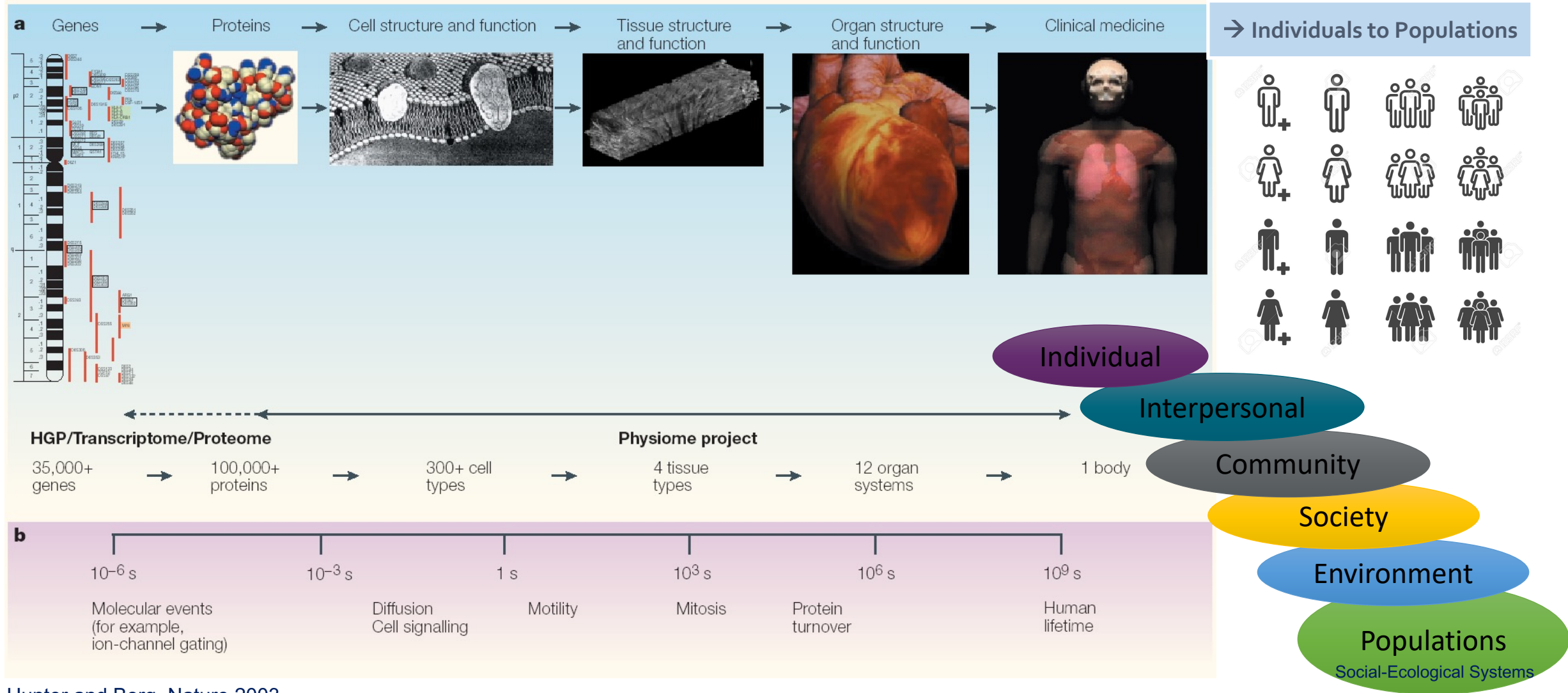
Analyze

Act

Reiterate / Learn



Biomedical Data is, Biologically, Multiscale

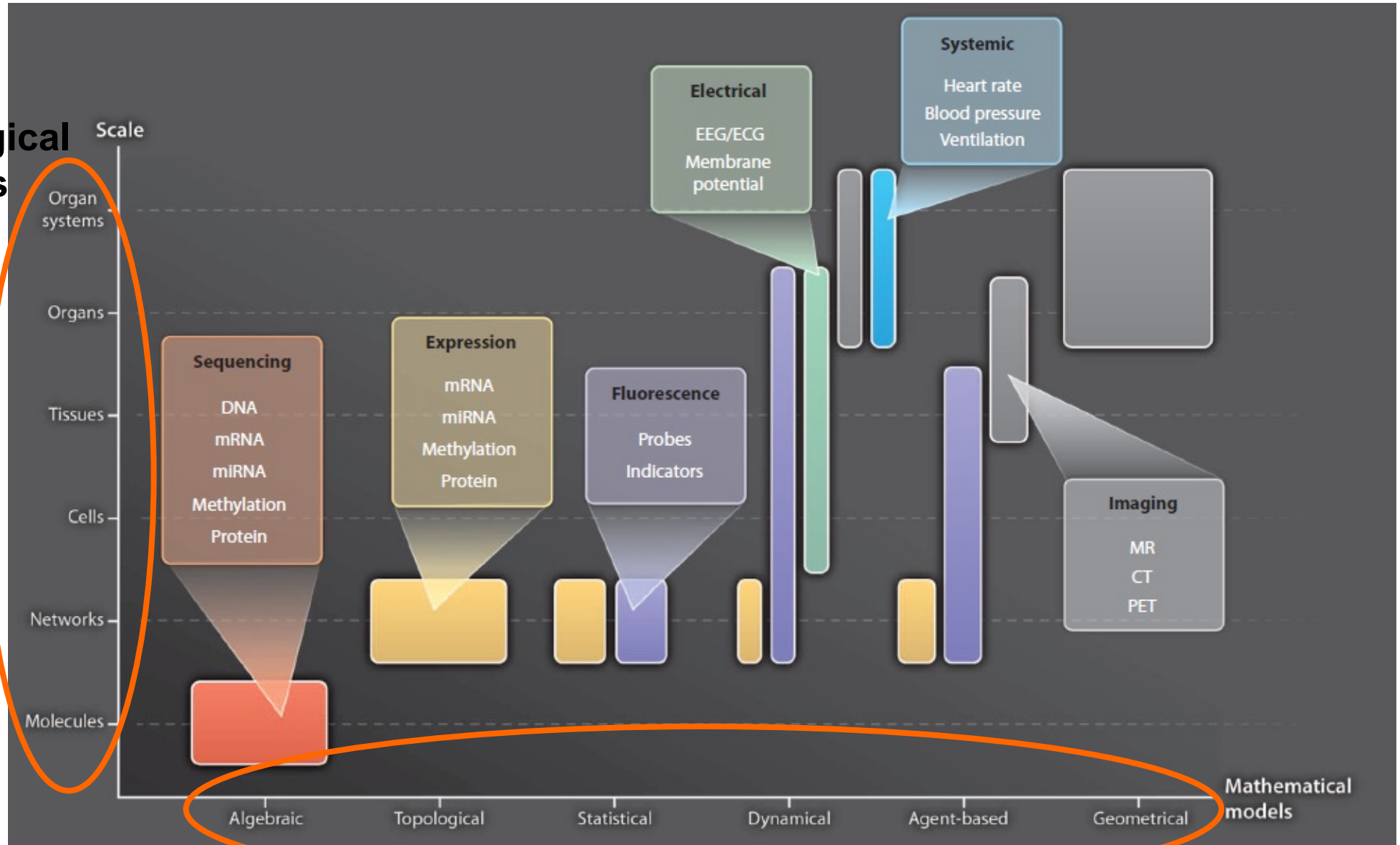


Hunter and Borg, Nature 2003



Biomedical Data is Multimodal, Measurement Tools Unique

**Biological
Scales**



**Analysis
Methods**

Winslow et al., 2012



National Center
for Advancing
Translational Sciences

Further Challenging the Utility of ‘AI’ in Biomedical / Clinical ‘Big Data’ Research

- Applying AI to biological questions requires models that can handle:
 - longitudinal data,
 - noisy data,
 - patient data and privacy issues,
 - incomplete data,
 - Categorical / nominal / ordinal,
 - quantitative: discrete / continuous,
 - structured / unstructured / semi
- ML algorithms are rarely interpretable so the logic and underlying mechanisms behind these models cannot be used to discern novel biological insight – only connected to existing knowledge
- Prospective research to generate ethically (*aka* unbiased) and statistically relevant, AI-friendly data, abiding by FAIR (Findability, Accessibility, Interoperability, and Reusability) principles with standardized data provenance, has not been adequately addressed leaving much of prior data less usable to drive current / future clinical research
- EHR data in terms of participants (demographics, comorbidity, etc.) is severely lacking for contemporary clinical trials/studies and clinical trials are not designed with appropriate sample size in mind
- Industry leaders (Google, Amazon, Microsoft, etc) develop AI services and platforms that biomedical researchers use – these tools were not developed specifically for biomedical data
- In general, ‘Real-world’ and user-generated data is not yet validated to the degree needed to add – only confound



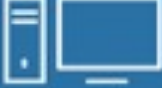
NIH Efforts to Drive Future AI and Predictive Analytics for Biomedical Research

Large Recent Initiatives for Biomedical AI: Trans-NIH and interagency efforts

National Cancer Institute &
Department of Energy Collaborations

Joint Design of Advanced Computing
Solutions for Cancer Pilots

CANcer Distributed Learning
Environment (CANDLE)





Frederick National Lab
Lawrence Livermore National Lab
UC – San Francisco
GSK



Bridge to Artificial Intelligence – Bridge2AI

National Institutes of Health
Office of Strategic Coordination–The Common Fund

BRIDGE2AI

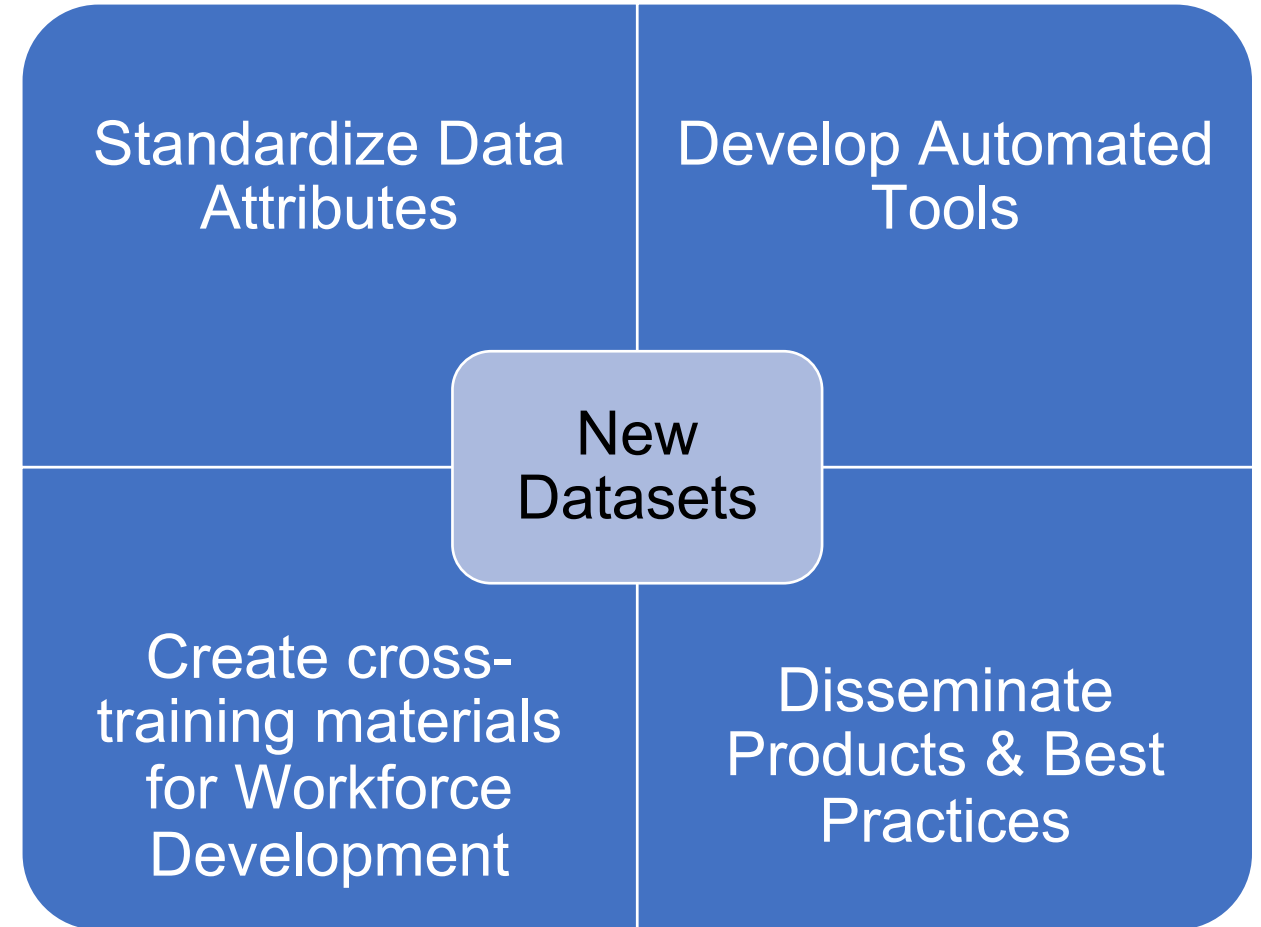
Nutrition for Precision Health
powered by the *All of Us* Research Program

NIH Common Fund: Bridge2AI Program Goals

BRIDGE2AI

Determining how to:

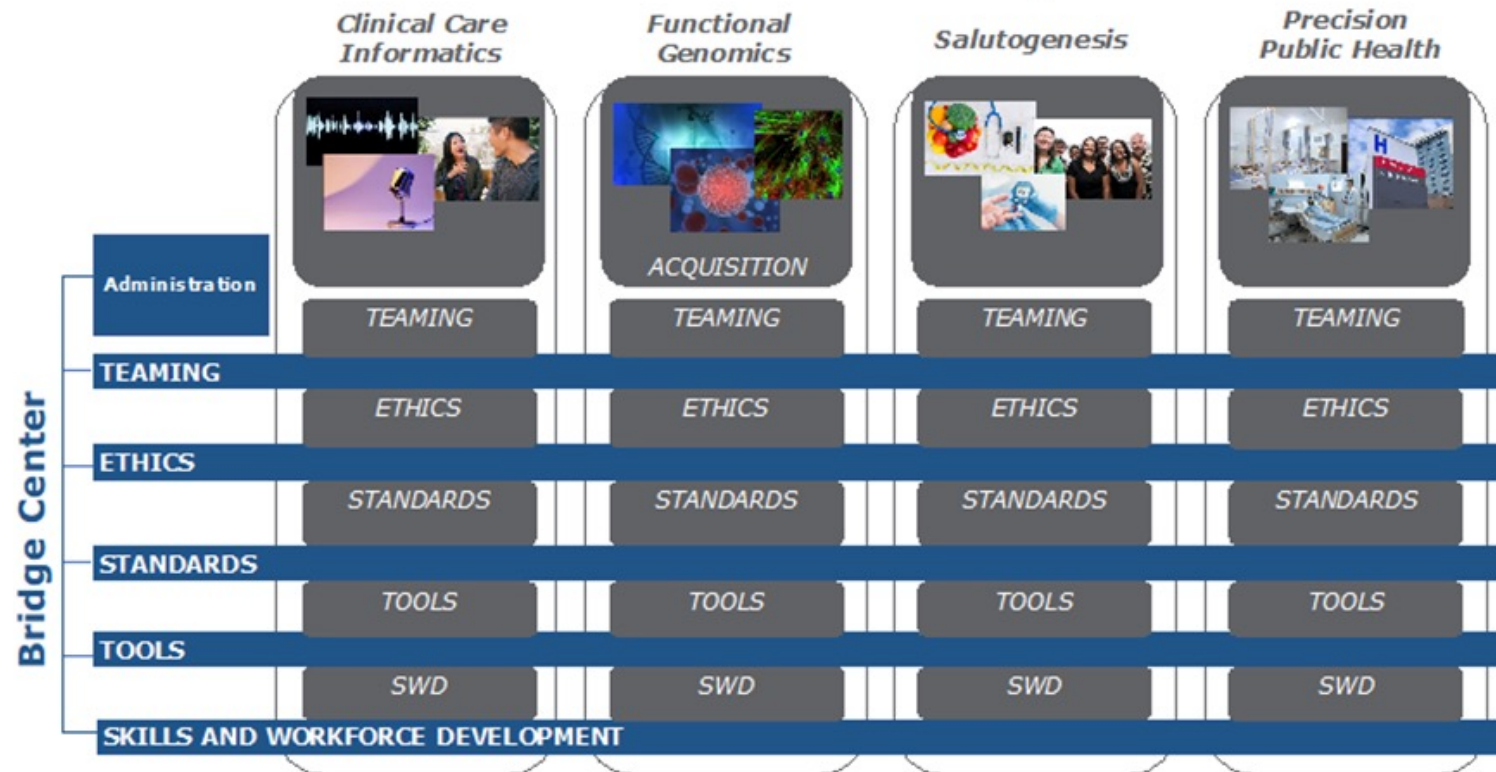
- Use biomedical and behavioral research grand challenges to generate **flagship data sets**
- **Prepare** AI/ML-friendly data
- Emphasize **ethical** best practices
- Promote **diverse teams**



NIH Common Fund: Bridge2AI Program Structure



Data Generation Projects



Clinical Care Informatics

Using imaging, clinical, and other data collected in an ICU setting for diagnosis and risk prediction



Functional Genomics

Mapping spatiotemporal architecture of human cells to interpret cell structure/function in health and disease



Salutogenesis

Uncovering the details of how human health is restored after disease, using type 2 diabetes as a model



Precision Public Health

Using voice as a biomarker for human health, revealing how genomic variation, behavioral, and environmental factors affect individual and population health



NIH Common Fund: Bridge2AI Program Funded Awards



Bridge Center

Consists of three awards:

- **Admin and Ethics** cores (contact PI: *Lucila Ohno-Machado**, UC San Diego)
- **Standards and Teaming** cores (contact PI: *Monica Munoz-Torres*, U. Colorado, Denver)
- **Tools and S&WD** cores (contact PI: *Alex Bui*, UC Los Angeles)

* Lead PI: Lucila Ohno-Machado

Data Generation Projects

- **Voice as a Biomarker of Health** (contact PI: *Yael Bensoussan*, U. South Florida)
- **Cell Maps for AI** (contact PI: *Trey Ideker*, UC San Diego)
- **Salutogenesis** (contact PI: *Aaron Lee*, U. Washington, Seattle)
- **Critical Care** (contact PI: *Eric Rosenthal*, Massachusetts General Hospital)

<https://www.commonfund.nih.gov/bridge2ai>



NIH National Center
for Advancing
Translational Sciences

NIH Common Fund: Nutrition for Precision Health (NPH) program – powered by the All of Us Research Program

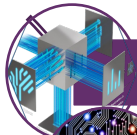
The challenge

- Diet-related diseases remain a major cause of death and disease in the US
- No single diet to optimize health in all individuals exist
- Current guidance is one-size-fits-most
- Interactions between dietary intake, microbiome ecology, metabolism, nutritional status, genetics, and the environment are still poorly understood

The need

- Research that investigates more precise nutrition recommendations
- Additional longitudinal studies that more precisely link health, evidence-based outcomes to diet input

Why
now?



Improvements in multi-omic data generation, throughputs, costs, and analysis methods



Advances in artificial intelligence and deep learning



Advances in analyzing and understanding microbiome ecology



Development and refinement of digital health technologies for dietary assessment



Readiness of the *All of Us* Research Program for ancillary studies



National Center
for Advancing
Translational Sciences

NIH Common Fund: NPH Program Structure

First major initiative to advance the goals of the 2020-2030 Strategic Plan for NIH Office for Nutrition Research

First ancillary study of All of Us research program

≈ \$170M NIH-wide effort over 5 years

15 awardees around the US



Overall Goal

To develop algorithms that predict individual responses to food and dietary patterns



NIH Common Fund: NPH Program Study Overview

1



Examine baseline diet and physiological responses to meal challenges

10,000 *All of Us* participants

2



Examine responses to 3 short-term intervention diets in free-living controlled feeding studies

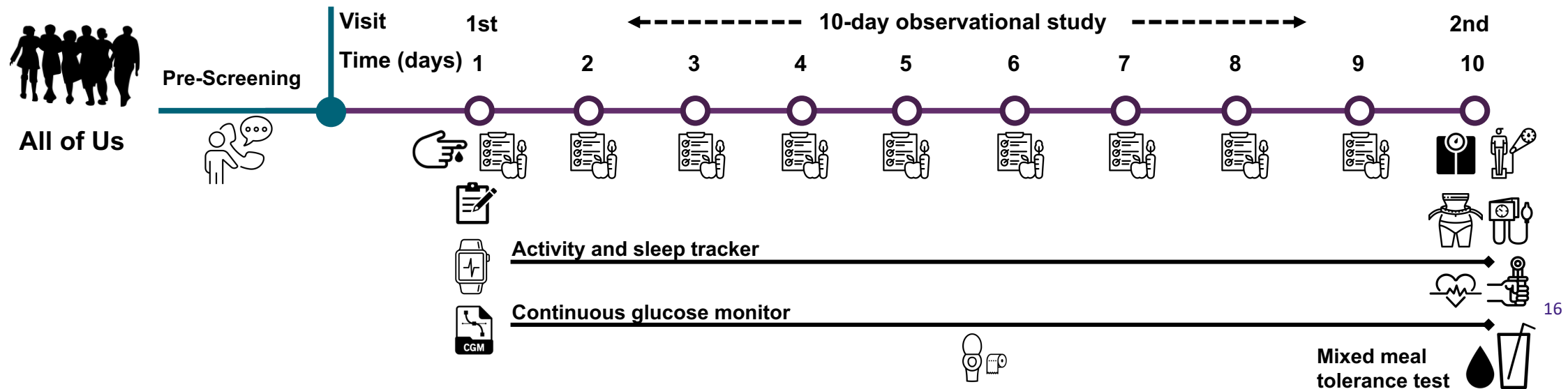
1,500-2,000 Module 1 participants

3



Examine responses to 3 short-term intervention diets in domiciled controlled feeding studies

500-1,000 Module 1 participants



NIH AI/ML Funding and Resources

Artificial Intelligence at NIH



Office of Data Science Strategy » Home » Artificial Intelligence at NIH

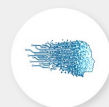


Artificial Intelligence at the NIH

The National Institutes of Health (NIH) makes a wealth of biomedical data available to research communities and aims to make these data findable, accessible, interoperable, and reusable—or FAIR. Additionally, the NIH seeks to make these data usable with artificial intelligence and machine learning (AI/ML) applications.

NIH has unique needs that can drive the development of novel approaches and application of existing tools in AI/ML. From electronic health record data, omics data, imaging data, disease-specific data, and beyond, NIH is poised to create and implement large and far-reaching applications using AI and its components.

Learn more about artificial intelligence activities at the NIH below.



ODSS-Led Initiatives

Catalyzing new opportunities in AI and data science



Bridge2AI

Propelling biomedical research by setting the stage for widespread adoption of AI



AI/ML Consortium to Advance Health Equity and Researcher Diversity

Increasing the participation and representation of researchers and communities currently underrepresented in the development of AI/ML models



Institute- and Center-Funded Initiatives

Developing and implementing AI/ML technologies across biomedical research domains

<https://datascience.nih.gov/artificial-intelligence>



National Center
for Advancing
Translational Sciences

Thank you and
Questions?

christopher.hartshorn@nih.gov



NCATS

COLLABORATE. INNOVATE. ACCELERATE.



NIH National Center
for Advancing
Translational Sciences

Challenges and Opportunities for AI in Agricultural Production & Food: Considerations for Ethics & Diverse Populations

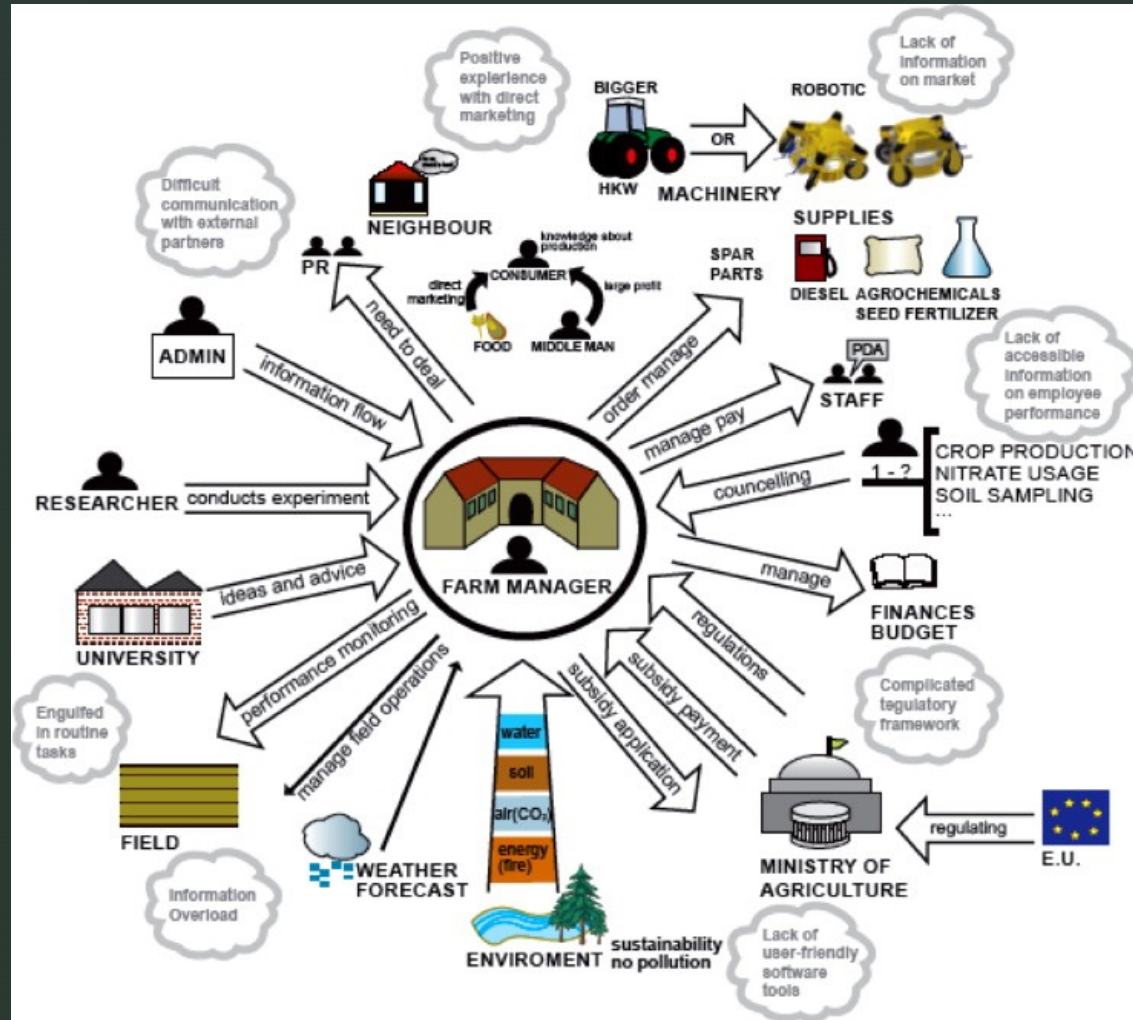
Janie Simms Hipp, J.D., LL.M.
CEO, Native Agriculture Financial Services

with assistance from

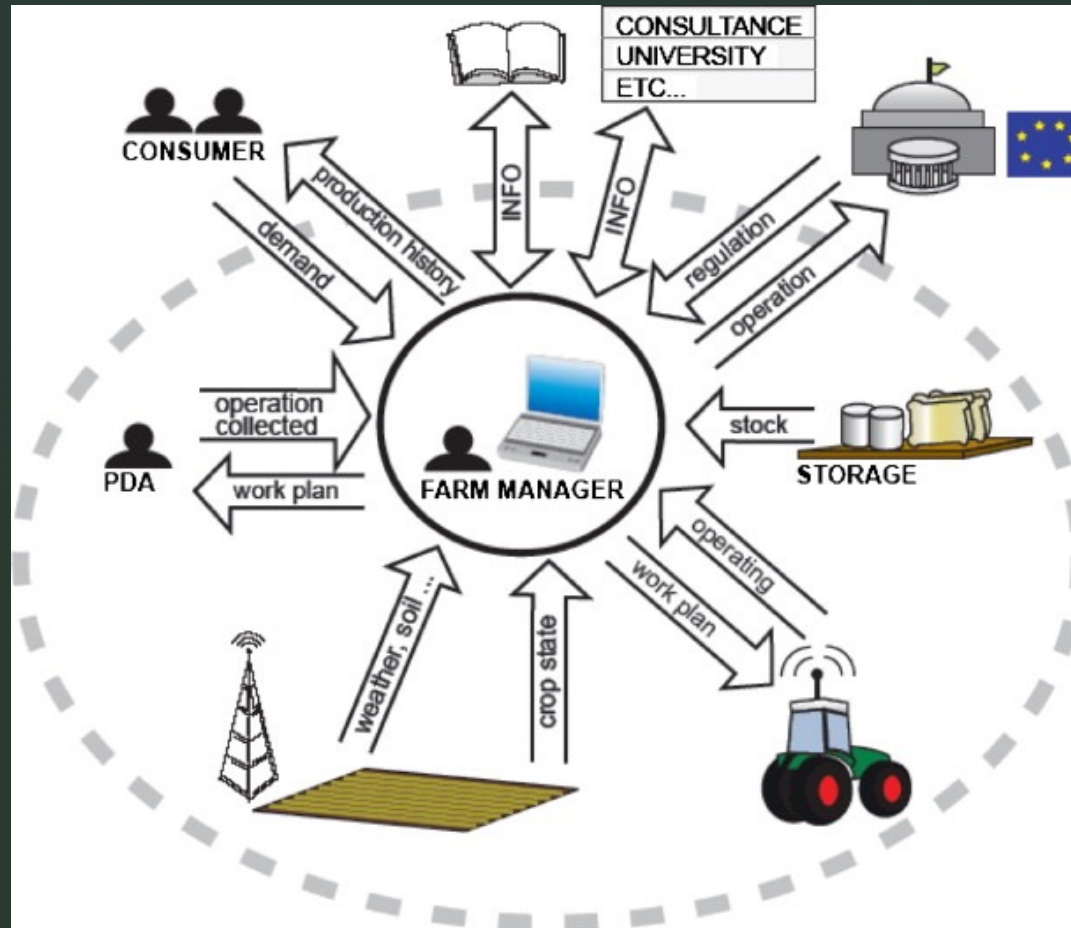
Dr. Marty Matlock
University of Arkansas

Janie Simms Hipp, J.D., LL.M.
CEO, Native Agriculture Financial Services
With Deep Support & Assistance From
Dr. Marty Matlock
University of Arkansas

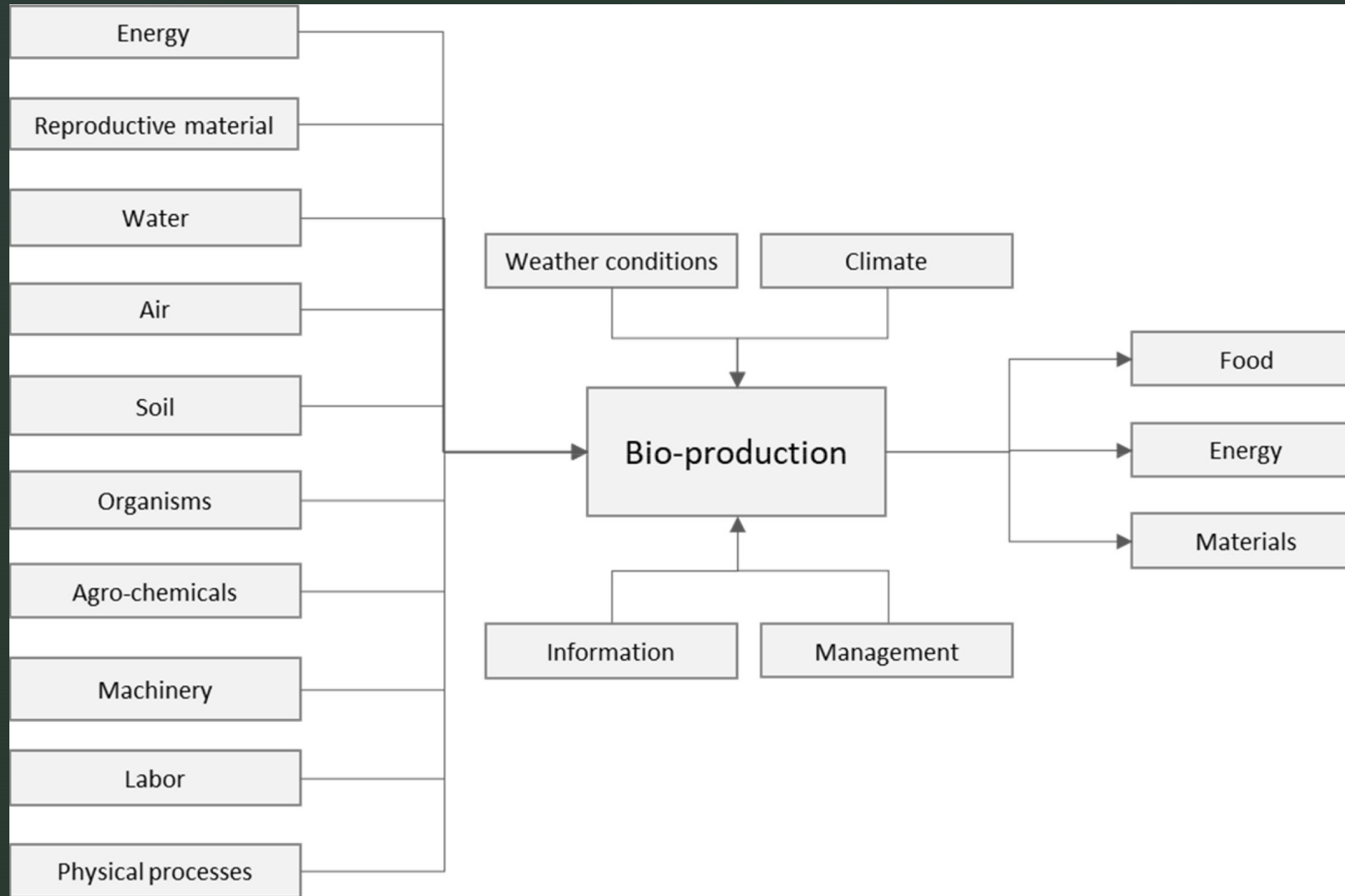
Ag producers are facing unprecedented complexity in decision-making



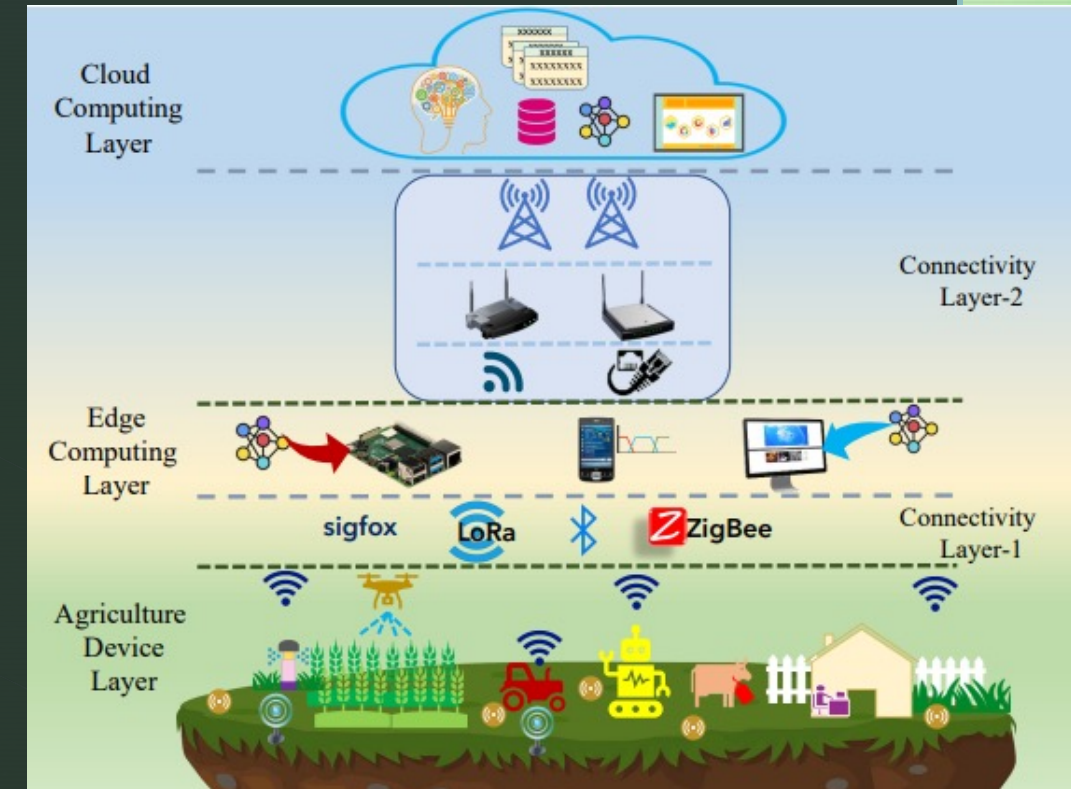
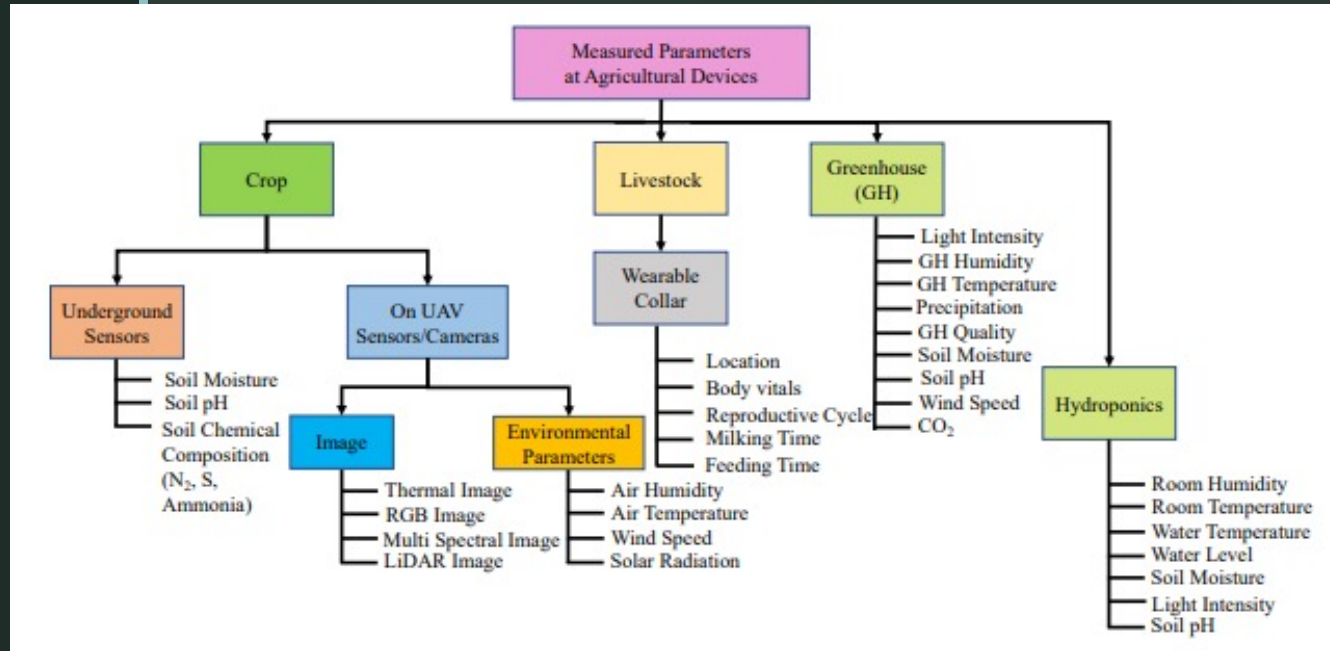
Agricultural AI has the potential to support (but not simplify) many of these decisions



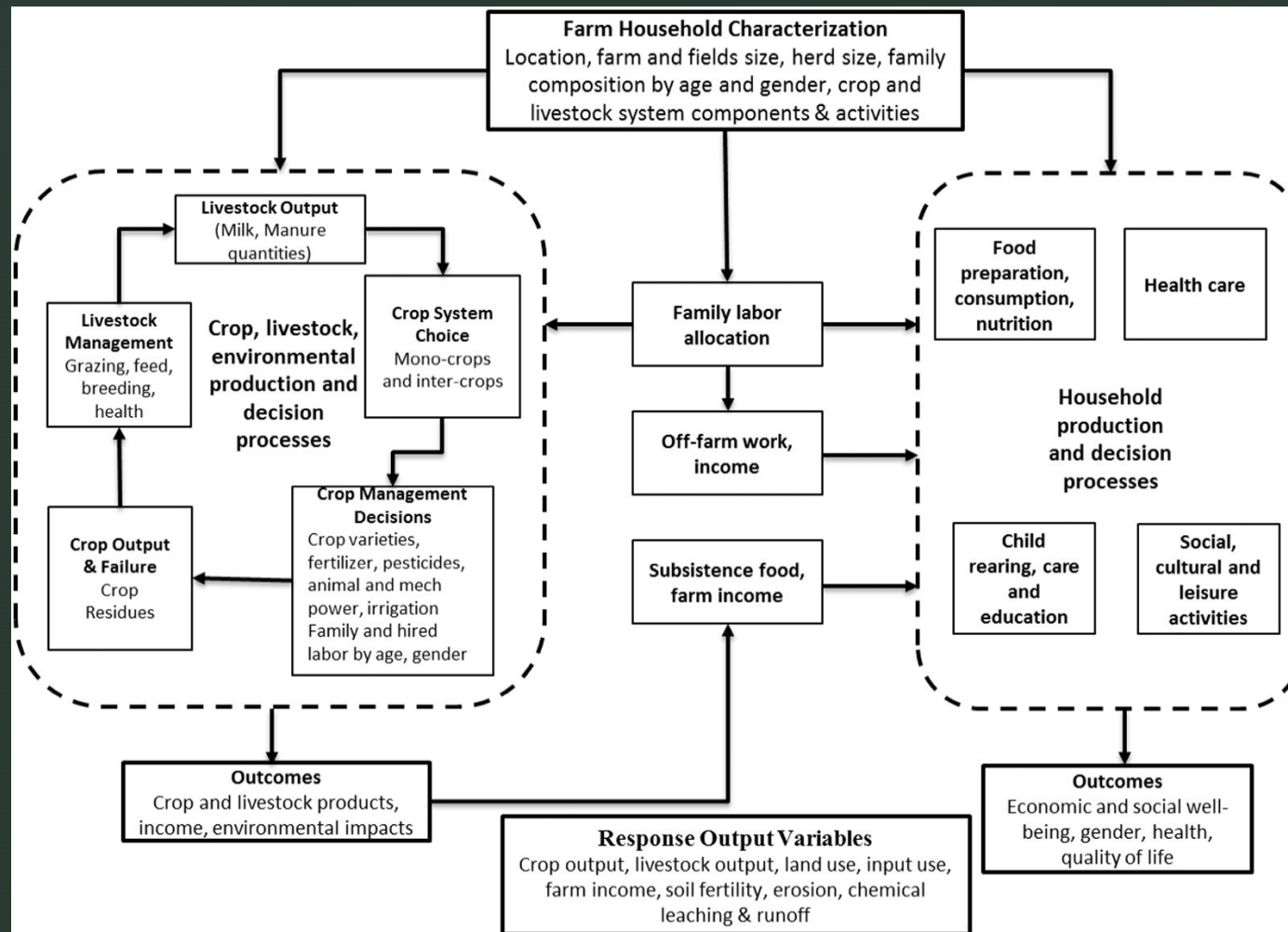
Agricultural production is a data and algorithm intensive process



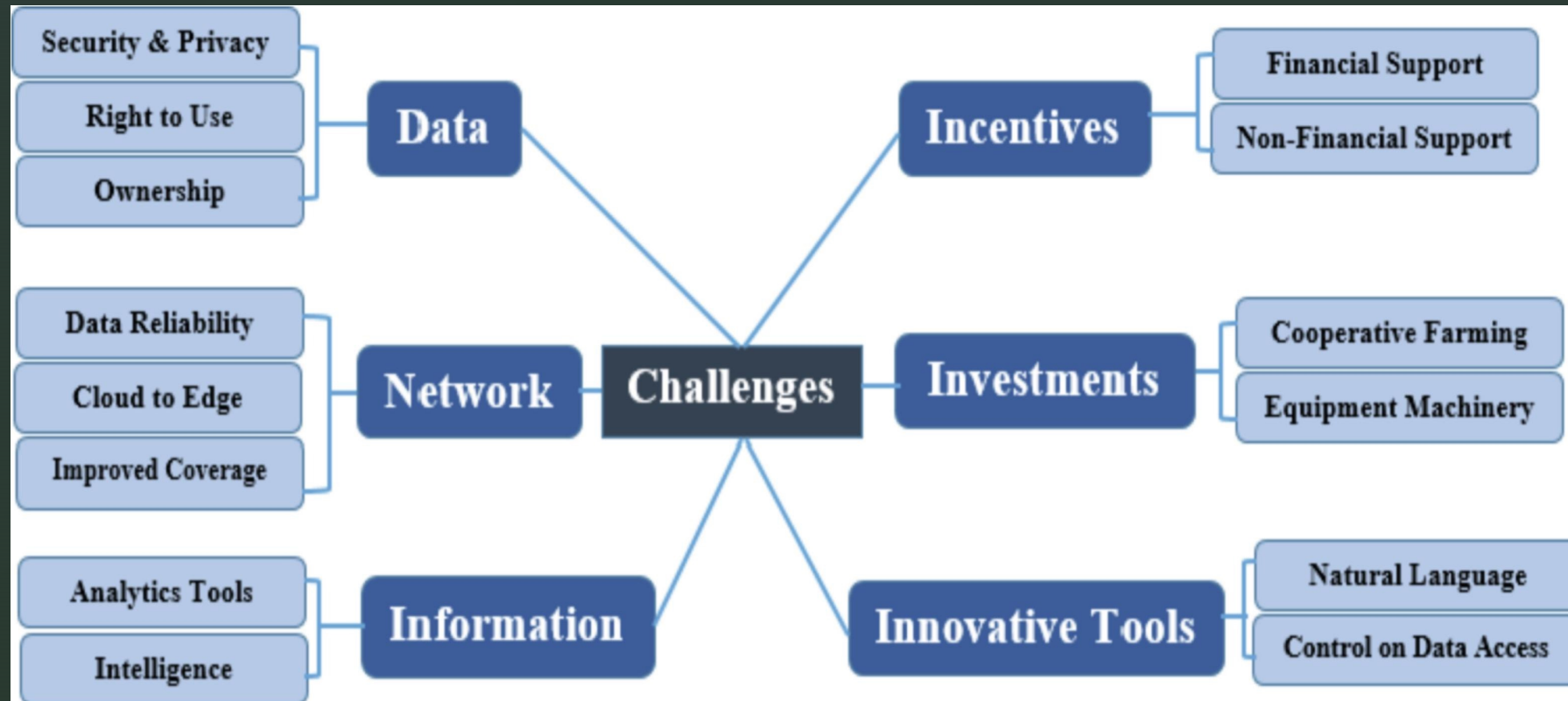
Architecture of smart agriculture



Farmer decisions are driven by environmental, market, and cultural conditions



Artificial Intelligence in agriculture encounters every challenge with big data systems



Pitfalls, Avoidance of Pitfalls, Role of Government Regulation, Safeguards

- Who owns the data that will drive these innovations?
- Who owns the innovations derived from these data?
- Who is responsible for insuring the veracity of the AI systems?
- Who is responsible for security of the AI systems?
- What is the legal framework that surrounds AI?
- What is the ethical framework that surrounds AI?

Data Rich Systems & Pitfalls

- AI works because of data-rich systems...clearly...and agriculture is inherently data-rich...
 - Farmers create data all day every day...from the land, equipment, and inputs (including labor) they own or manage
 - Q: If you are building something with stuff that I own, what are my ownership rights in what you build?
 - Who owns “AI” and at which stage in the AI creation does ownership or “shared ownership” attach?
 - Where does the government come in?
 - What happens when data is bad? Garbage in...garbage out. How does bad data impact the farmer whose raw data goes into the AI application?
- What are the national security implications of AI in agriculture, since food security is national security and AI will impact food security, access, availability, and success of the sector?
 - What aspects of data/AI in agriculture has national security implications, and what does not?

Currently, ownership of agricultural data generated on the farm is most likely governed by contract

- Farmers often face unbalanced contract negotiations because many Ag Tech Providers are large multinational corporations with powerful legal teams.
- Farmers and ranchers want to control the information collected from their fields and livestock....and rightfully so.
- Most (81%) think they own their data and 77% of them are worried about data security.
- Ownership and control of farming data is a significant concern for farmers if others could use their information for commodity market speculation without their consent – or attempt to hold them liable for something outside their control – or restrict their access to farm programs based on bad data.

In the US, farm generated data are “facts”

- US law classifies data, including farming data, as “facts.”
- US law provides for a regime if seeking patent, trademark, copyright, or trade secret protection.
- As the basic fact underlying certain agriculture commodities, farming data lacks a creative element which can be defined as an intellectual property (IP) whose ownership could be protected by copyright laws (17 U.S.C. § 102(a)), such as published books or commercial programs.
- Therefore, legally speaking, farmers may not own the “raw” data generated from their land.
- There is a challenge in protecting this data in all arenas except “trade secret”...even “copyright” concerns require publishing activities to solidify protection.
- How many farmers will actually “publish” the data.
- Trade Secrets are a creature of state law, primarily.
- What is the role of government; and which government is involved and through which arm of government?

American Farm Bureau Federation. 2016. "Privacy and Security Principles for Farm Data." Retrieved July 9, 2021. <https://www.agdatatransparent.com/principles>

<https://edis.ifas.ufl.edu/publication/AE564>

Ethical and fundamental questions

- The driving force behind centuries of agriculture production improvement, has been, and still is...stemming the tide of hunger and malnutrition in the human species....however, agriculture is not just an economic act.
 - It is at the same time an act of community, sacredness, protection, and can be involved in exploitation.
- When considering the most underserved actors in any equation...what is their interest and impact in the AI equation?
 - The vast majority of agriculture production in many countries is concentrated in the hands of just a few very large players...but the vast number of agriculture producers are not the large players.
 - How do you target AI both at the larger players and simultaneously at the vast numbers of smaller players?
 - How do you stabilize access to data and information?
 - How do you ensure AI is size-specific?
- If it is too “hard” to consider ethics in AI, then are we prepared to live with the outcomes?

Legal frameworks

- Legal frameworks must be reformed at the same time we embrace our AI future in agriculture
 - What is the regulatory regime and who holds it? USDA is the only informed choice.
 - How do we ensure privacy of data generated on-farm?
 - How do we ensure permissions/authorizations for use of data occur, and if compensation for data occurs, how do we ensure the amount of compensation for data is fair?
 - How do we focus on incentives in AI (and not just regulation of AI)...are incentives the only approach in agriculture that allows for the continued presence of smaller to mid-sized players?
 - What are the unintended consequences to various types/sizes of producers as AI proceeds?
 - How do we ensure informed consent for the use of data?
 - A solely “regulatory” approach to the output of AI will further ensure uneven distribution of AI.
 - Environmental management, risk management, economic analysis, resiliency, adaptation to climate change can all be impacted by AI – which of these have the greatest potential for harm and need the greater level of regulation and scrutiny?

Pitfalls and Challenges

- Downsides if we do not consider ethics, access, legal frameworks, and fairness in AI
 - Failure to ensure equal access & failure to guard against exploitation.
 - Failure to ensure and comply with IRB frameworks (ex: Indigenous IRB).
 - **Inherent Knowing and Sacredness.** How are these issues dealt with in a data or AI context (Ex: Data Sovereignty movement among Indigenous communities)?
 - Permissions, consent, inclusion of data creators will lag or be non-existent.
 - Recognition and quantifying of harm caused by use of AI will be difficult.
 - Rise of state-level push-back on AI (much like “ag-gag” laws of many years back) will lead to patchwork of policy; are we prepared for an “uneven” landscape.
 - Rights to “clean up” data – how do we deal with “bad” data and bad actors.
 - Who is at the table throughout – how do we embed fairness?
 - Current data sets – particularly in nutrition, food access – are already significantly flawed.
 - Current data sets in agriculture production are almost unusable by the farmers who generate the data and who can experience the most improvement in the resiliency and viability of their operations.
 - Who is creating the decision support systems that allow producers to stop drowning and start using data?

Final Thoughts

- We need to focus on the diversity, equity, inclusion, accessibility, and fairness of AI systems – as they are being built, modified, enlarged.
- We need competent lawyers at the table immediately.
- We also need to impress upon lawmakers at every level that this train is moving fast and unless we focus on achieving the proper balance today, we are headed for rapid exacerbation of exclusion, exploitation, lack of privacy, and failures of control.
- Finally, if we have any hope of a fair and just food system that takes into account national and food security and that achieves the highest good for the most people, we must get this right.