AI/ML for BioEnergy Research (AMBER) Workshop

August 23-25, 2022

Organizing Committee

Huimin Zhao (Chair) University of Illinois, Urbana-Champaign

Nathan Hillson (co-Chair) Lawrence Berkeley National Laboratory

Kerstin Kleese van Dam (co-Chair) Brookhaven National Laboratory

Deepti Tanjore (co-Chair) Lawrence Berkeley National Laboratory

Resham Kulkarni DOE Office of Science

R. Todd Anderson DOE Office of Science

Jay Fitzgerald DOE Office of Energy Efficiency and Renewable Energy

Gayle Bentley DOE Office of Energy Efficiency and Renewable Energy

Wayne Kontur DOE Office of Science

Ramana Madupu DOE Office of Science

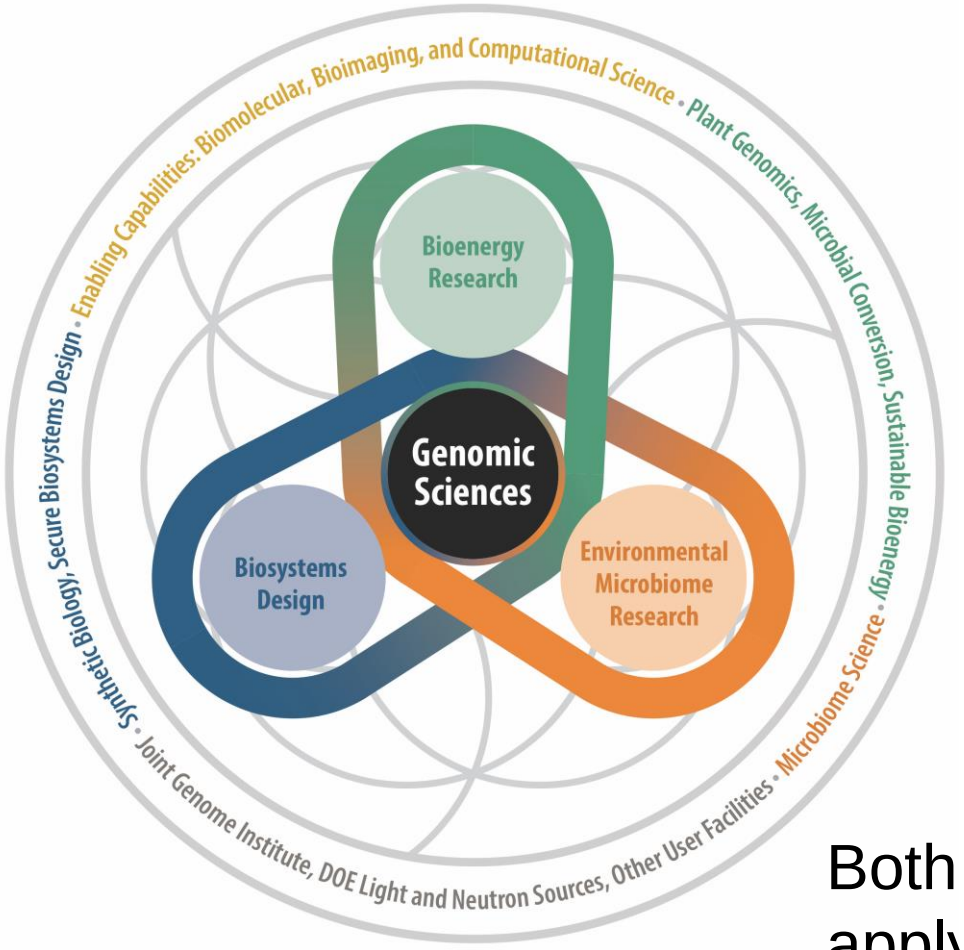
Pablo Rabinowicz DOE Office of Science

U.S. Department of Energy Office of Science and Office of Energy Efficiency and Renewable Energy. https://doi.org/10.2172/1968870.

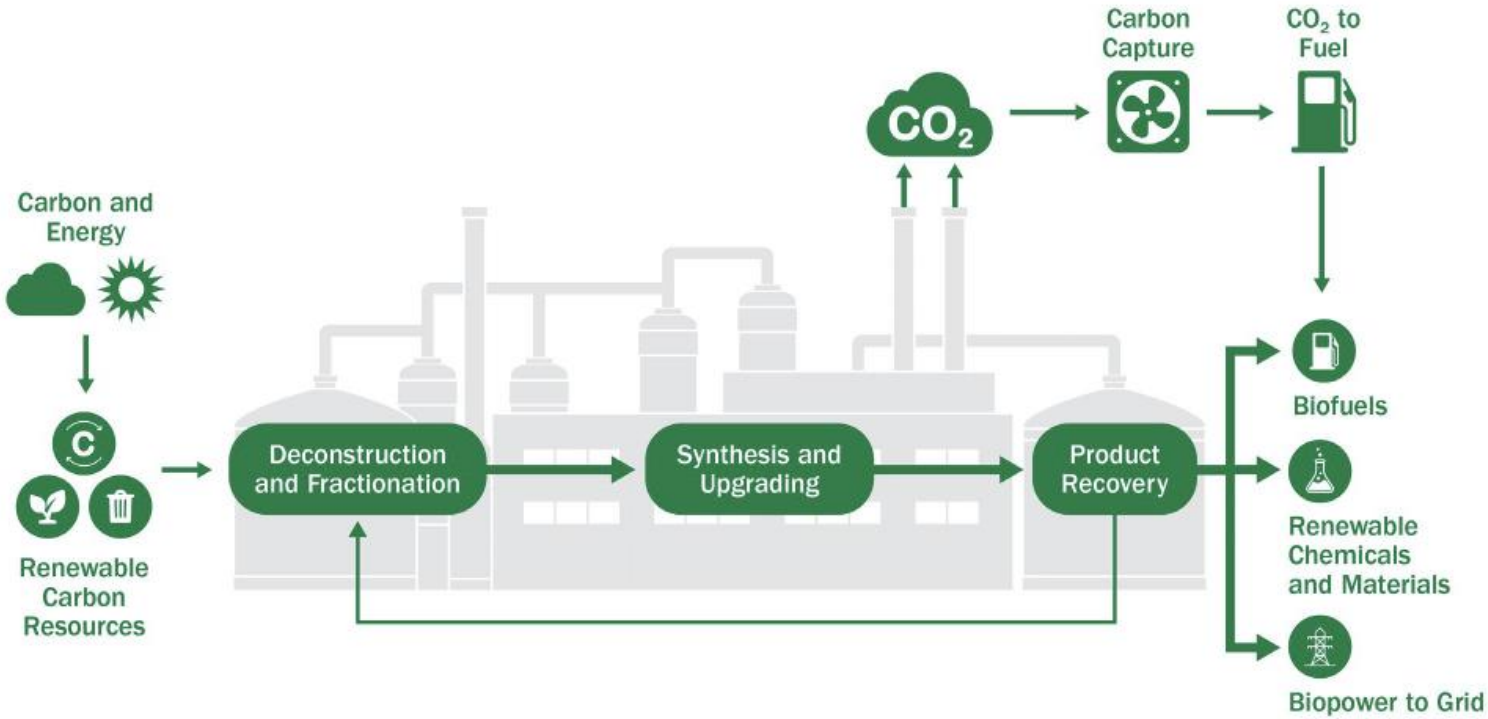
U.S. DEPARTMENT OF ENERGY

BER-BETO Joint Virtual Workshop

BER/BSSD

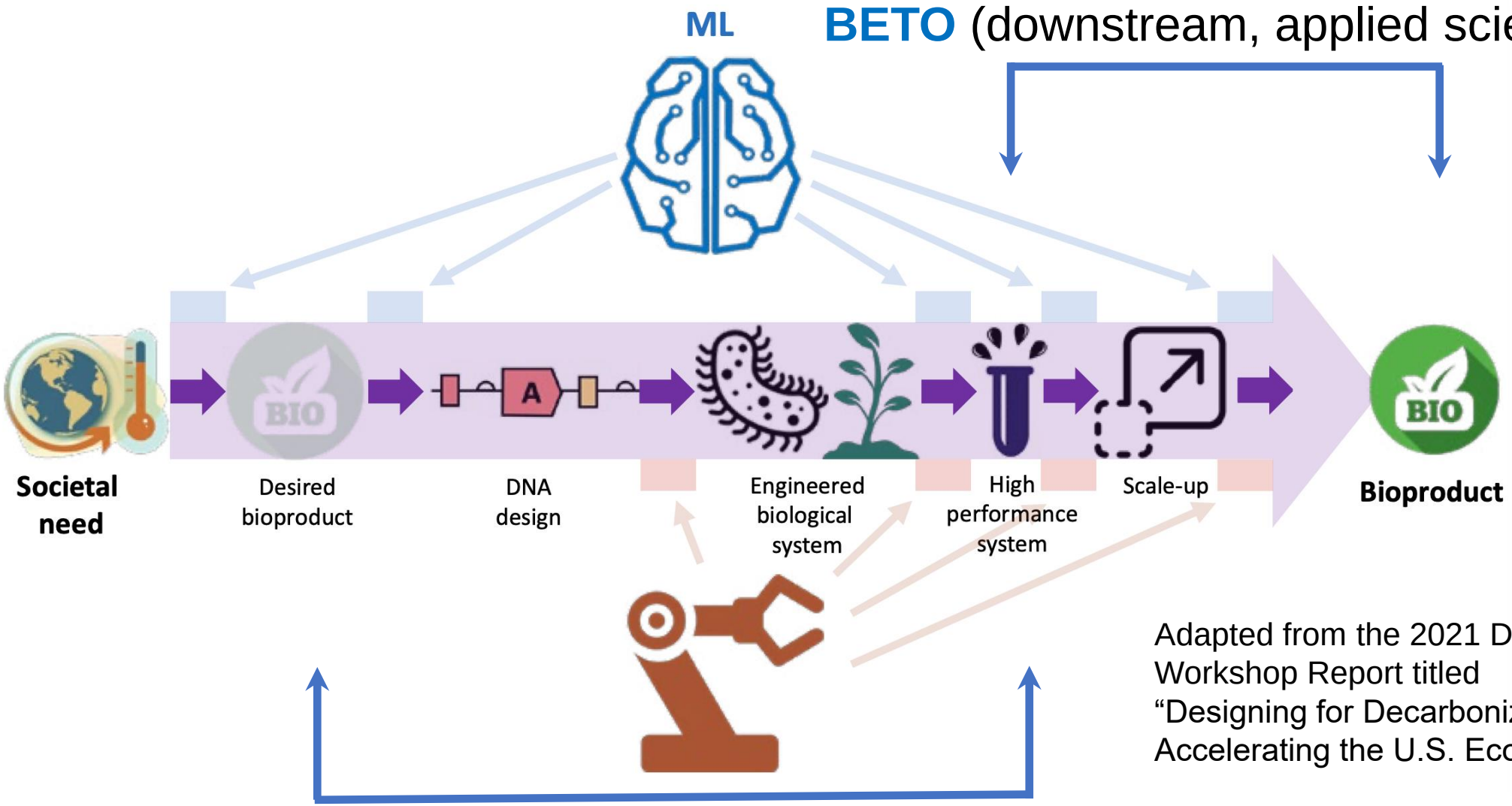


BETO



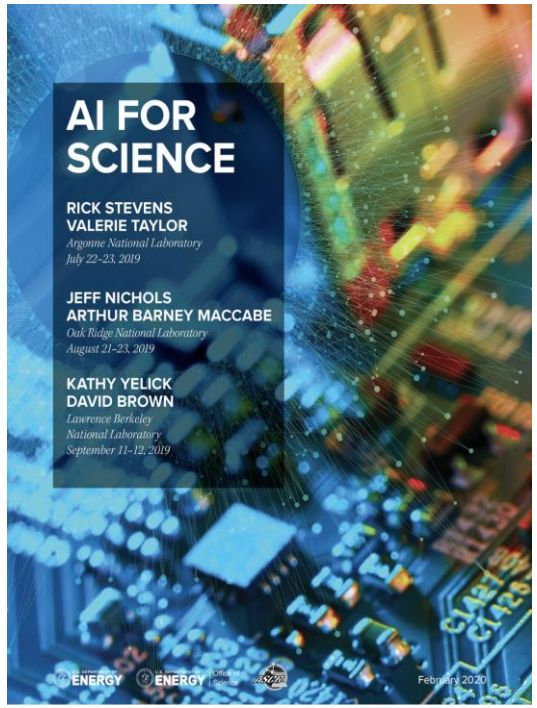
Both BER and BETO are interested in developing and applying AI/ML tools to address grand challenges in bioenergy research.

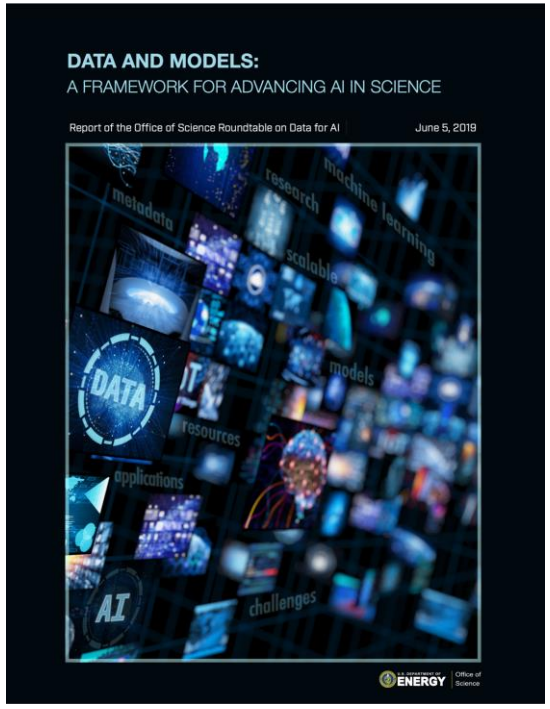
AI/ML & Automation for Bioenergy Research

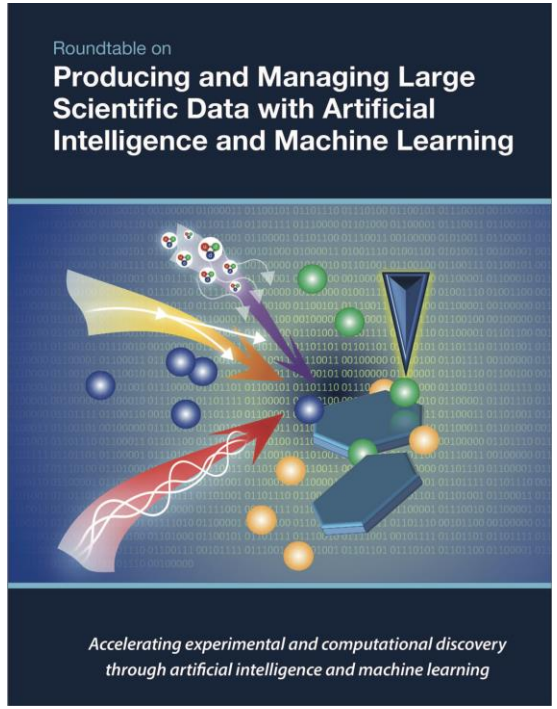


Adapted from the 2021 DOE Workshop Report titled "Designing for Decarbonization: Accelerating the U.S. Economy."

DOE's "AI for Science" Initiative







AMBER Workshop

Goal:

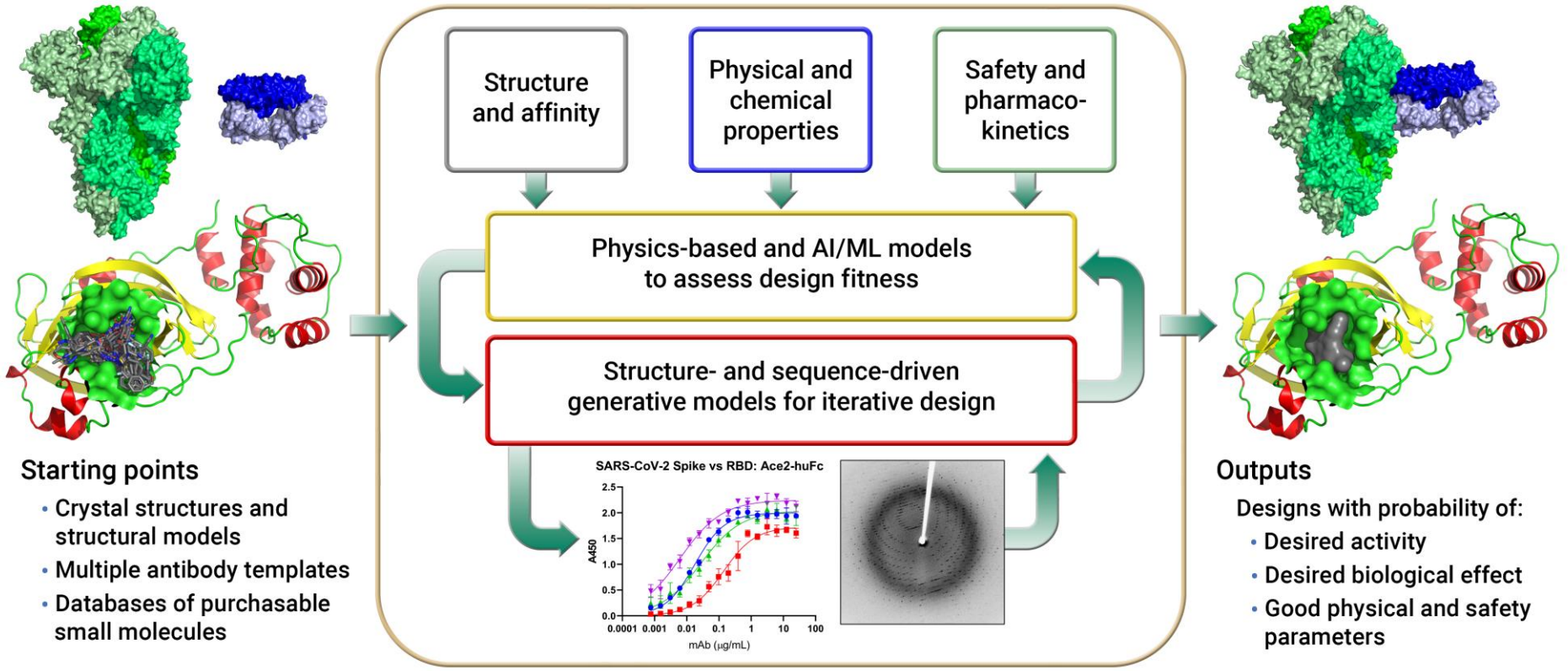
“Big picture” for integration and implementation of AI/ML approaches into experimentation to speed basic sciences discoveries in genome biology and design of new biological systems.

Sessions

1. Workshop Goals and Introduction to Artificial Intelligence/Machine Learning
2. Defining Focus on Applications of AI/ML for Bioenergy Research
3. AI/ML Approaches to Meet Bioenergy Research Needs
4. Data and Compute Infrastructure Needed
5. Community Development Including Outreach, Engagement, and Training

- ~50 participants from academia, industry, and DOE national labs
- 9 position papers
- 14 elevator pitch presentations

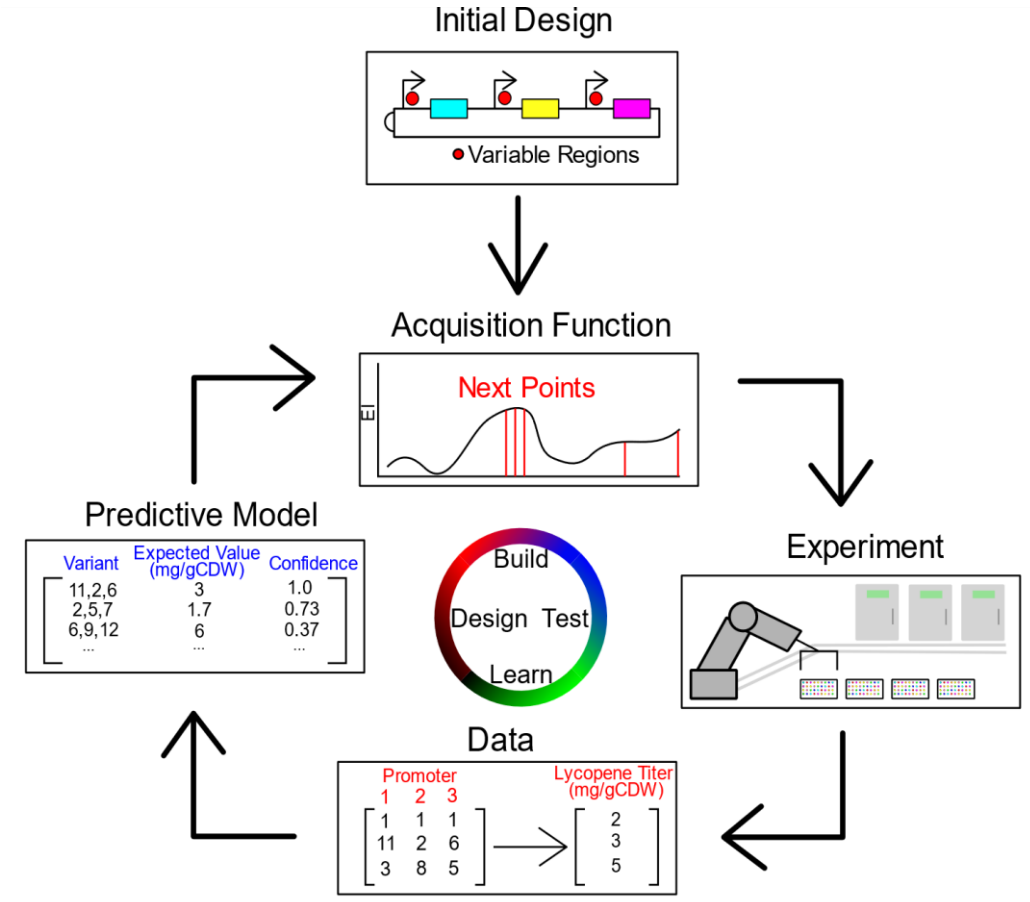
Early Successful Examples



AI/ML Functions can Speed High-Performance Computing. The National Virtual Biotechnology Laboratory project on molecular therapeutics created an integrated computational and experimental platform for designing COVID-19 therapeutics. [Courtesy Oak Ridge National Laboratory]

Early Successful Examples

BioAutomata: a Self-driving biofoundry



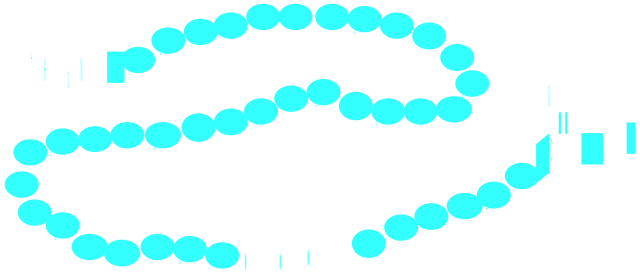
AI/ML can Enable Closed DBTL Loop. BioAutomata was used to rapidly optimize the lycopene biosynthesis pathway in *E. coli.* [Courtesy CABBI]

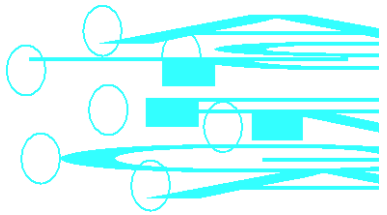
HamediRad et al. *Nature Communications* 2019

Early Successful Examples

Two Biggest Fundamental Problems in Protein Science

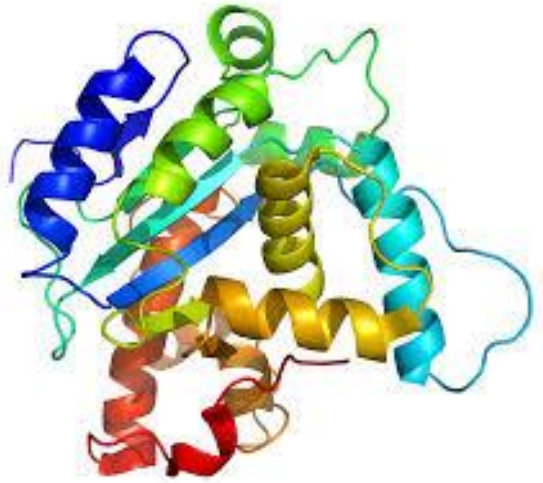
20 natural amino acids, assembled in the cell according to DNA instructions





Nature 2021

?



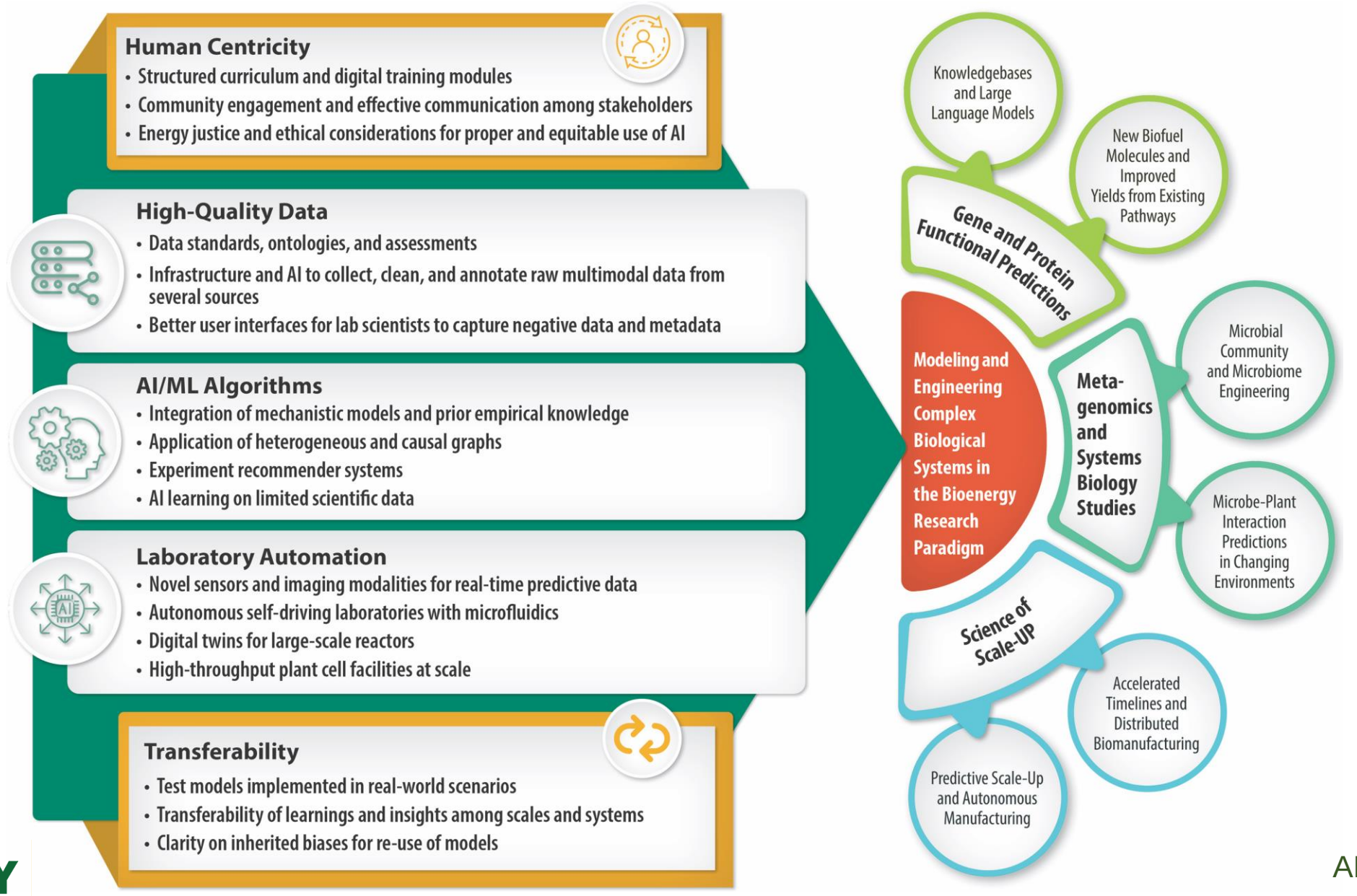
Functions

?

CLEAN

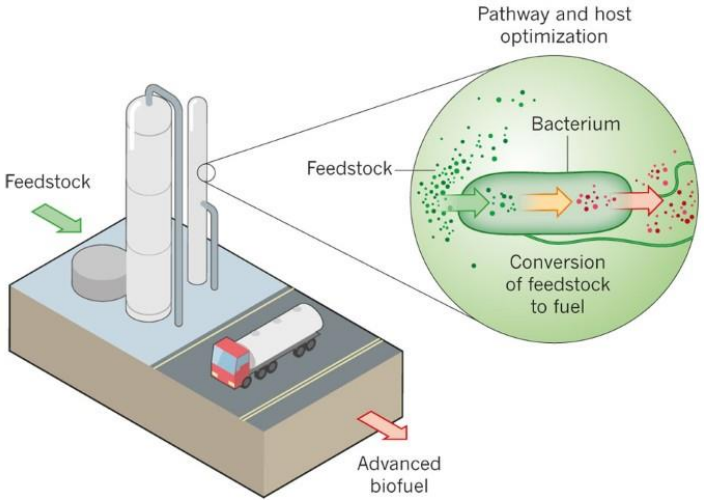
Science 2023

AI/ML Needs for Bioenergy Research

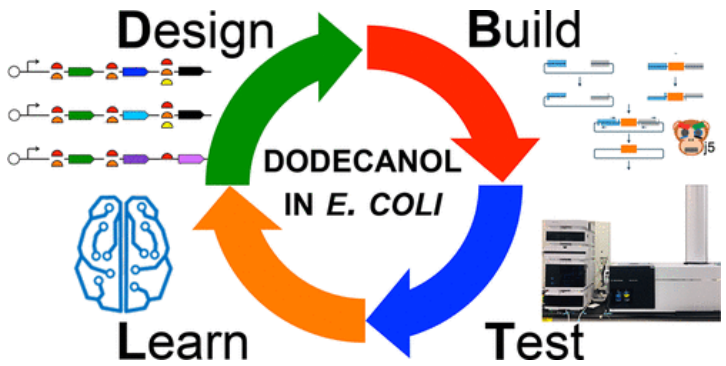


AMBER Workshop Takeaways

1. Numerous AI/ML and automated experimentation applications exist for a variety of DOE mission needs in energy and the environment. Exemplary research grand challenges for which AI/ML could provide solutions include:



building microbes and microbial communities to specifications



developing closed-loop autonomous design and control for biosystems design

advancing scale-up and automation for biomanufacturing

AMBER Workshop Takeaways

2. Lack of sufficient high-quality, annotated data hinders the development of AI/ML applications.

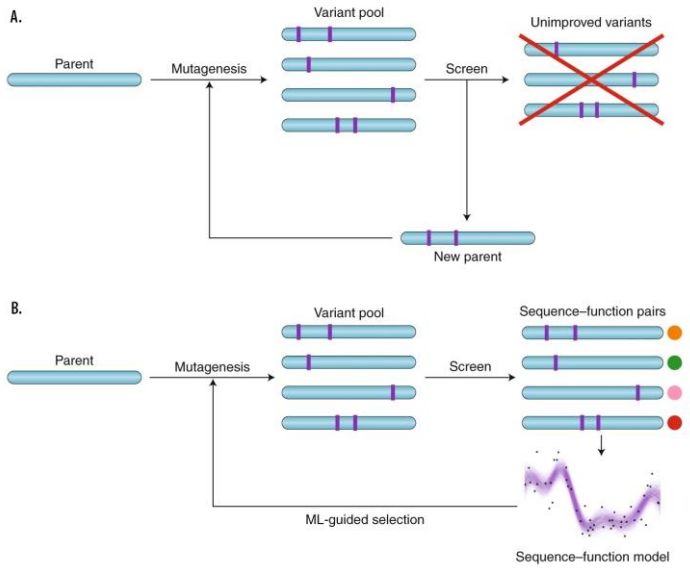
data exchange standardization: Integrating heterogeneous software, data, and automation across vendors and developers is difficult, partly due to a lack of standardized metadata formats, vocabularies, and syntaxes.

data quality: A key challenge is the need for very large, high-quality datasets suitable for the research questions at hand.

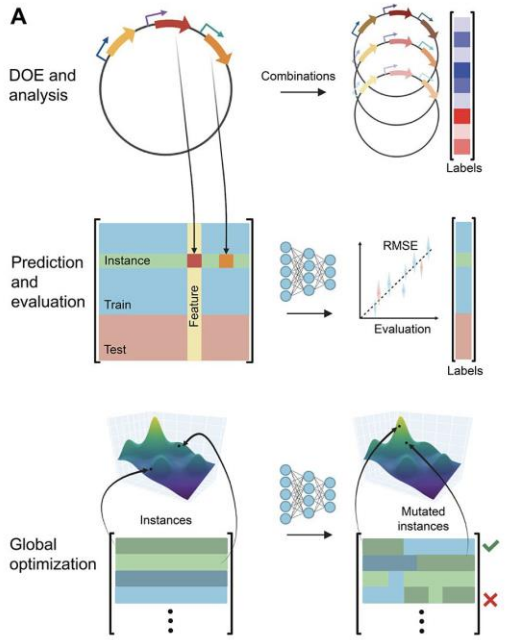
data privacy: How to build models on top of a foundation of private (e.g., company-owned) primary data and make the trained models available to the public without revealing the primary data and creating issues with intellectual property or copyrights.

AMBER Workshop Takeaways

3. New and improved AI/ML tools are needed, particularly those meeting the specific needs of the BER/BSSD and BETO research communities. For example,



enzyme engineering

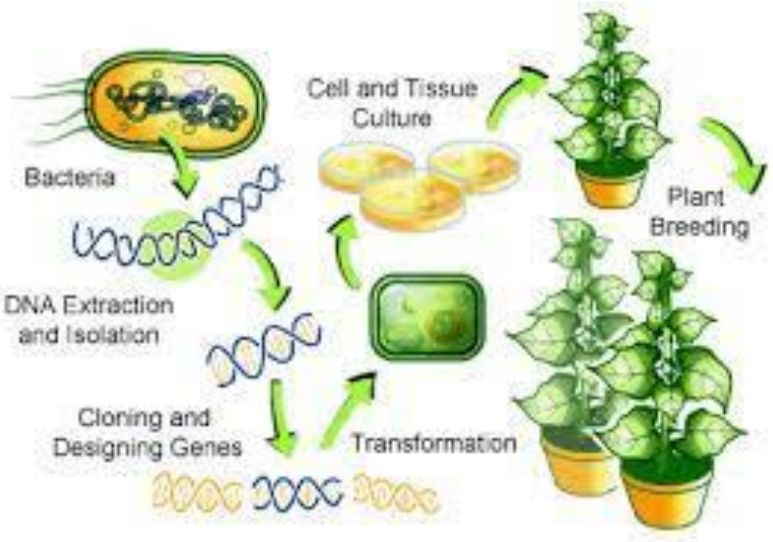


metabolic engineering

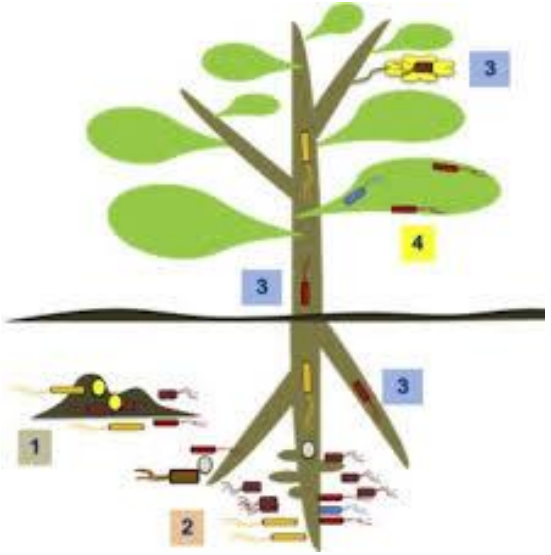
bioprocess development

AMBER Workshop Takeaways

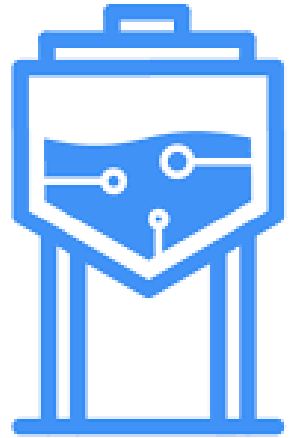
3. New and improved AI/ML tools are needed, particularly those meeting the specific needs of the BER/BSSD and BETO research communities. For example,



plant engineering



microbiome engineering



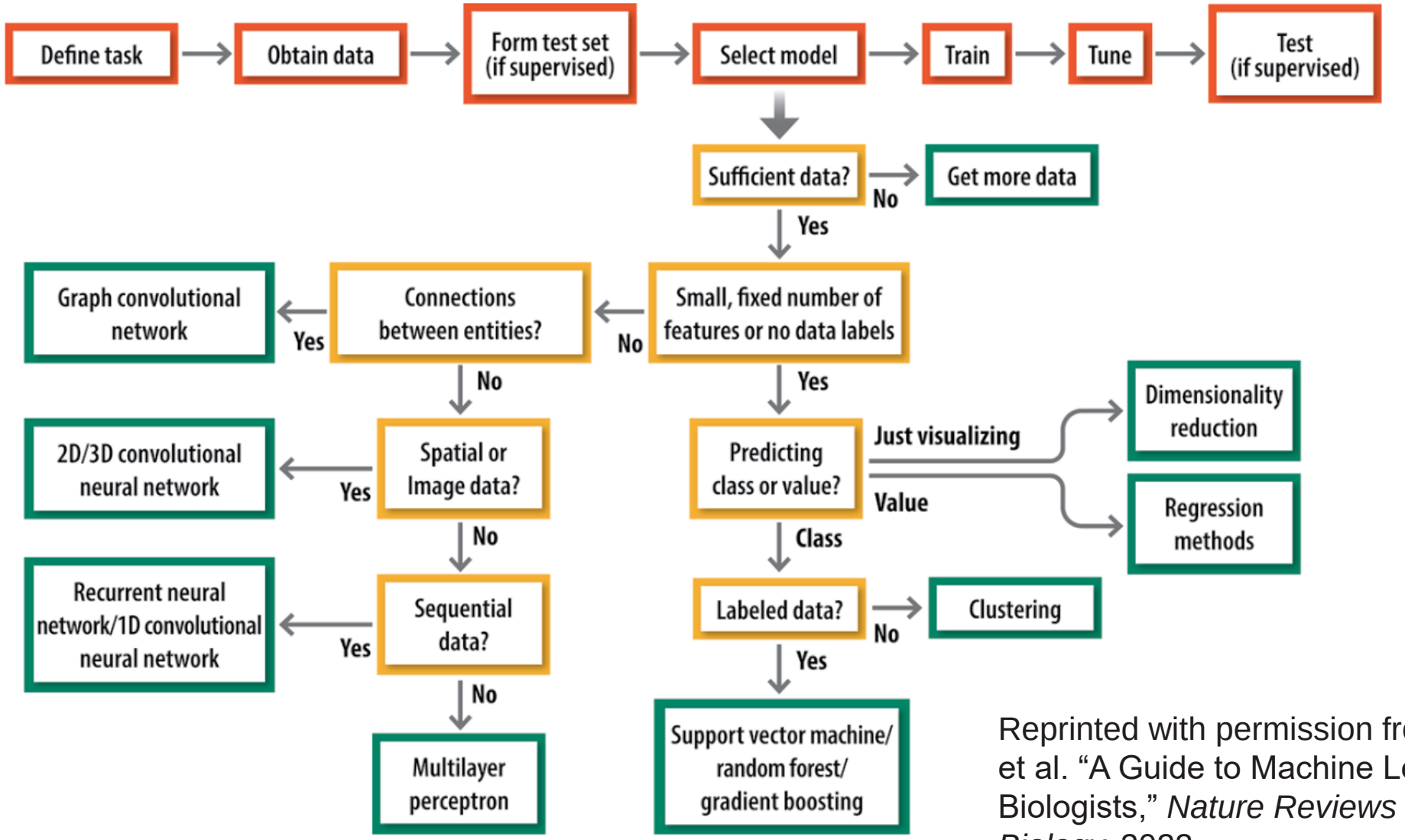
bioreactor digital twins

AMBER Workshop Takeaways

3. New and improved AI/ML tools are needed, particularly those meeting the specific needs of the BER/BSSD and BETO research communities.

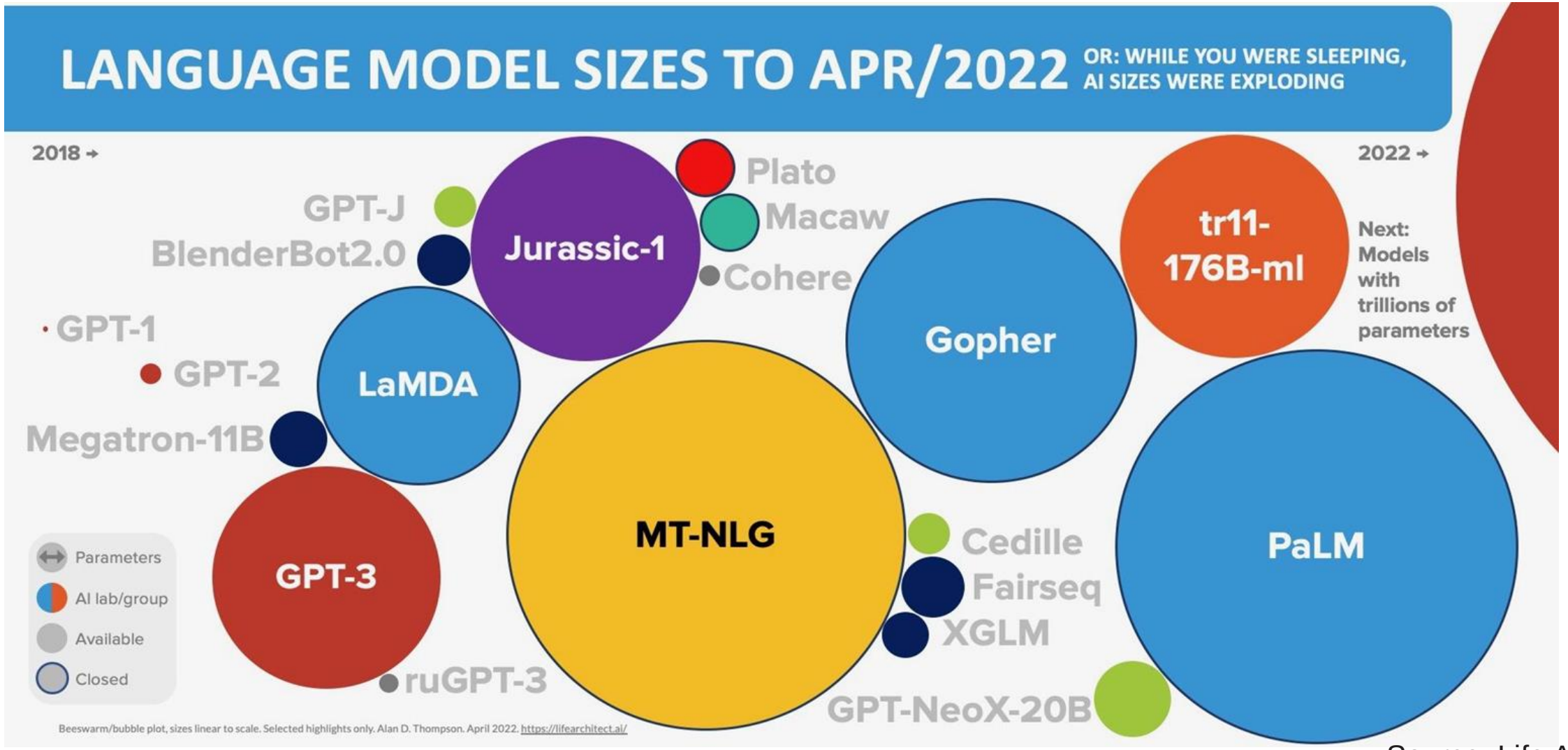
- matching AI/ML models to problems of interest,
- merging AI/ML predictive capabilities with mechanistic insight,
- overcoming the limited data problem,
- integrating data from various resources,
- quantifying the predictive capacity of AI/ML models,
- developing generally applicable large language models/foundation models.

Flowchart Summarizing How to Select a ML Model



Reprinted with permission from Greener, G., et al. "A Guide to Machine Learning for Biologists," *Nature Reviews Molecular Cell Biology,* 2022

Large Language Models/Foundation Models



Source: Life Architect

Do we need a ChatGPT like language model for bioenergy research?

AMBER Workshop Takeaways

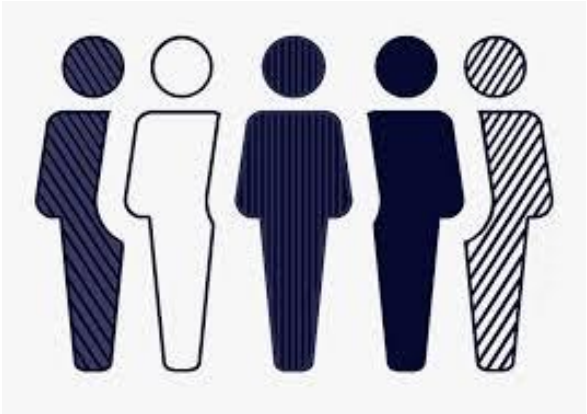
4. Trade-offs in performance, cost, and reliability exist between deploying commercially available versus building custom-developed instrumentation and software for automated or autonomous experimentation; translation of manual to automated or autonomous methods is often a nontrivial endeavor.

- Capture expert knowledge to drive autonomous experiments and laboratories.
- Establish environmentally hardy technology for field-scale autonomous experiments and laboratories.
- Address increased complexity due to scale for autonomous experiments and production
- Facilitate training of AI/ML models for bioenergy scenarios:
  - Data archives
  - Computing resources
  - New AI/ML training infrastructure
  - Integrative technology test labs

AMBER Workshop Takeaways

5. Training a new generation of young scientists who can develop and apply AI/ML tools is needed to solve long-standing scientific challenges in bioenergy research.

Workforce development



Diversity & inclusion

Outreach

Social responsibility and ethics

A New Bioenergy Research Paradigm

AI Enabled
Design Workflows
(what to make)

→ Biofuels, bioproducts, enzymes, organisms, microbiomes, plants,

AI Enabled
Experimental Workflows
(how to make it)

→ Self-driving labs

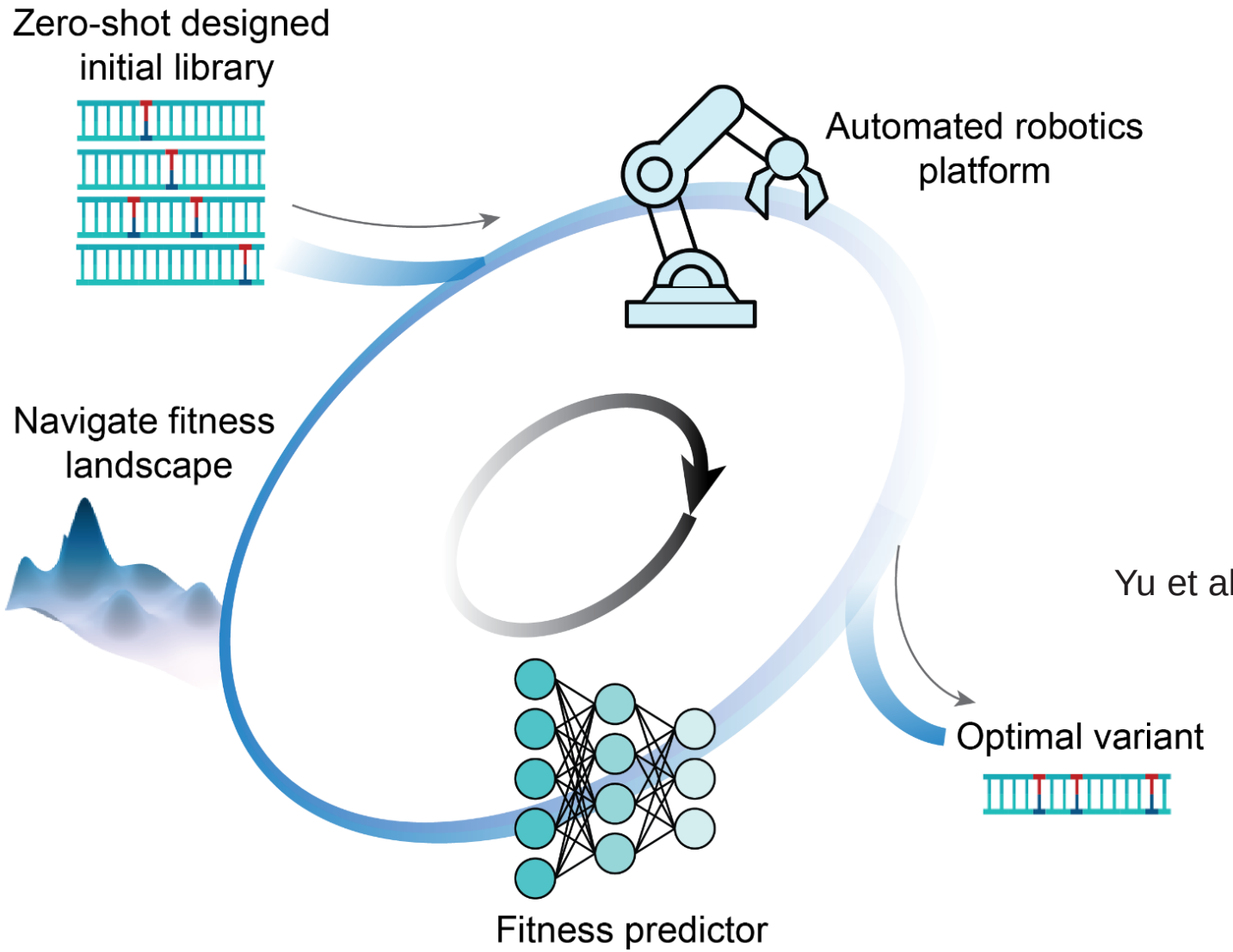
AI Enabled
Scientific Comprehension
(what it means)

→
- datasets
- literature
- science “news”
- strategy

→ *Cleaned*
Updated
Annotated
Aggregated
Interpreted

→ **Insight**?

A New Bioenergy Research Paradigm: Autonomous Experimentation



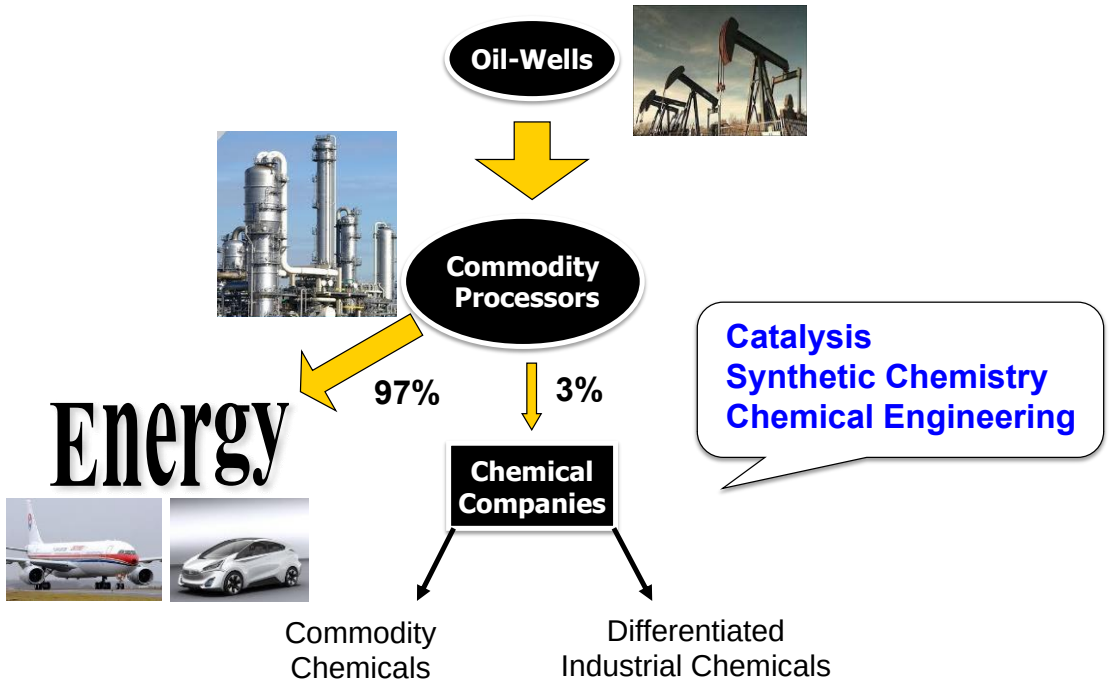
Yu et al. *Cell Systems* in press.

Future of Bioenergy Research

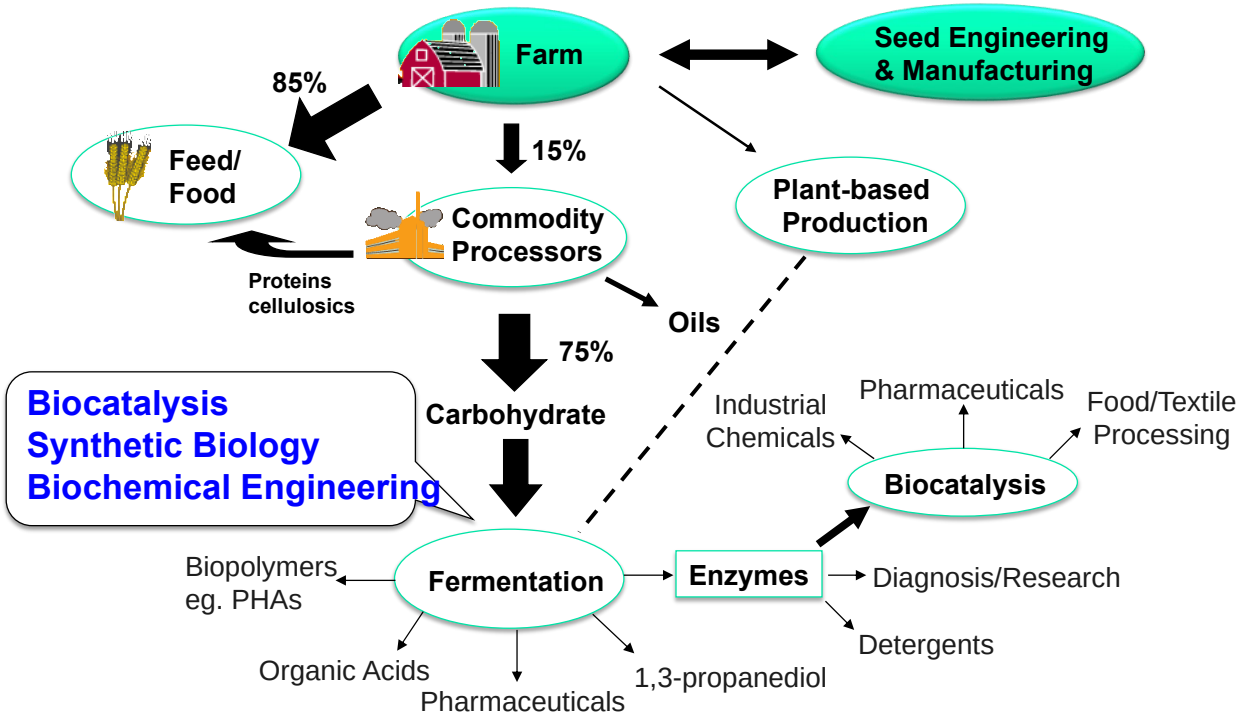
Courtesy Rick Stevens

Towards a Carbon-negative Sustainable Bioeconomy

Petroleum-based



Biomass-based



Refinery → **10-50 years?** → **Biorefinery**