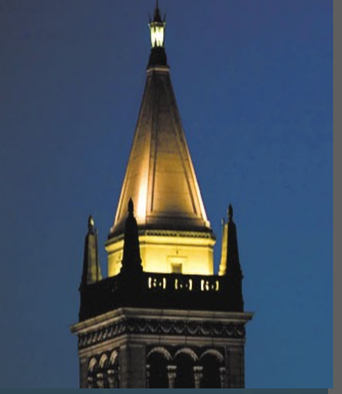
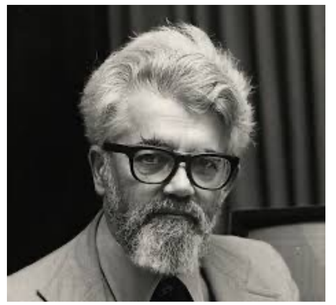


An Alternative View on AI: Collaborative Learning, Incentives, and Social Welfare

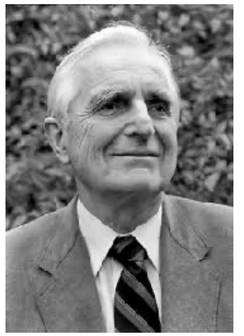
Michael I. Jordan

University of California, Berkeley

Perspectives on AI



- The classical “human-imitative” perspective
  - cf. AI in the movies



- The “intelligence augmentation” (IA) perspective
  - cf. search engines, recommendation systems, natural language translation
  - the system need not be intelligent itself, but it reveals patterns that humans can make use of
- The “intelligent infrastructure” (II) perspective
  - cf. transportation, intelligent dwellings, urban planning
  - large-scale, distributed collections of data flows and loosely-coupled decisions
  - novel market mechanisms and novel deliberative mechanisms based on data flows

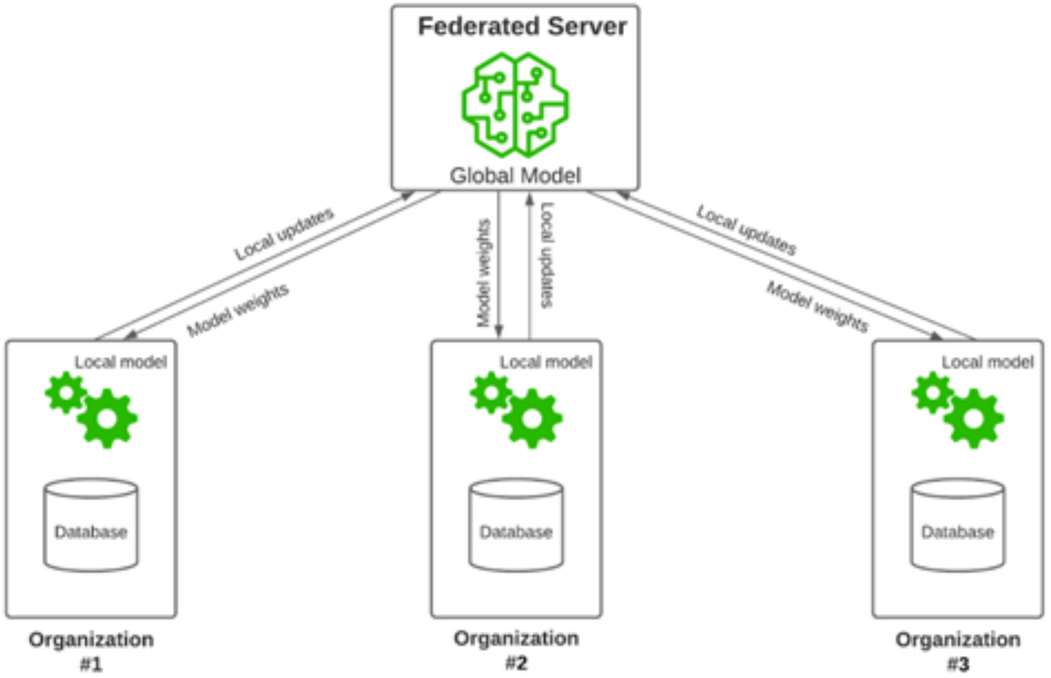
The 1950s AI Perspective

- A goal of understanding the intelligence of an individual human and building computers that mimic such intelligence
  - and possibly improve on it
- Not very clear what the overall engineering goal is
  - what kind of systems will such intelligences be embedded in
  - what kind of problems will such systems solve?
  - seems naïve to expect to solve real-world problems---in domains such as health care, climate change, commerce, etc---with such a vague premise

A Counterpoint

- Intelligence is as much about the collective as it is about the individual
- In terms of establishing goals for an emerging engineering field, thinking in terms of collectives seems at least as urgent and promising as thinking in terms of individual intelligence
- There may be new forms of collectives that can emerge if we put our minds to it

Federated Learning Paradigm



Purported to aim at collective mechanisms, but does it?

Data, Creators, Values, and Collaborations

- In real life, the “nodes” are often people, and their data is not something to simply be streamed and aggregated
- People often value their data
- They may wish to reveal aspects of their data if (and only if) they obtain commensurate benefits
- One way to start to understand this is to develop blends of microeconomics and machine learning
- Learning-aware mechanisms and mechanism-aware learning

Music in the Data Age

- Use data to structure a two-sided market; e.g., by providing a dashboard to musicians, letting them learn where their audience is
  - the musician can give shows where they have an audience
- I.e., consumers and producers become linked, and value flows: a market is created
  - the company that creates this market profits simply by taking a cut from the transactions
- Bring in brands and create a three-way market
  - the brands can partner with specific musicians based on affinities
- The company *United Masters* is doing precisely this; www.unitedmasters.com

Some Problems at the Interface of ML and Econ

- Relationships among optima, equilibria, and dynamics
- Exploration, exploitation, and incentives in multi-way markets
- Information asymmetries, contracts and statistical inference
- Strategic classification
- Uncertainty quantification for black box and adversarial settings
- Calibrating predictions for inference and decision-making
- Mechanism design with learned preferences

Statistical Contract Theory

Stephen Bates

Michael Sklar

Jake Soloff

The Theory of Incentives

- Contract theory is one branch of the theory of incentives (auction theory is another branch)
- In contract theory, agents possess private information and a principal wishes to incentivize them to take actions that depend on that private information
  - –the goal is overall social welfare, or revenue
- For example, services such as airlines have “business fares” and “economy fares”
  - –this allows them to offer different prices to agents who have different willingness to pay, without requiring agents to reveal their private values
- The design problem is to determine a menu of options, of the form (service, price), from which agents select

Clinical Trials

Average Cost of Clinical Trial

Department of Health and Human Services, 2014

Therapeutic Area	Phase 1	Phase 2	Phase 3
Anti-Infective	$4.2 (5)	$14.2 (6)	$22.8 (5)
Cardiovascular	$2.2 (9)	$7.0 (13)	$25.2 (3)
Central Nervous System	$3.9 (6)	$13.9 (7)	$19.2 (7)
Dermatology	$1.8 (10)	$8.9 (12)	$11.5 (13)
Endocrine	$1.4 (12)	$12.1 (10)	$17.0 (9)
Gastrointestinal	$2.4 (8)	$15.8 (4)	$14.5 (11)
Genitourinary System	$3.1 (7)	$14.6 (5)	$17.5 (8)
Hematology	$1.7 (11)	$19.6 (1)	$15.0 (10)
Immunomodulation	$6.6 (1)	$16.0 (3)	$11.9 (12)
Oncology	$4.5 (4)	$11.2 (11)	$22.1 (6)
Ophthalmology	$5.3 (2)	$13.8 (8)	$30.7 (2)
Pain and Anesthesia	$1.4 (13)	$17.0 (2)	$52.9 (1)
Respiratory System	$5.2 (3)	$12.2 (9)	$23.1 (4)

(in millions of dollars)

Immense social investment in clinical trials

Contract Theory

principal

- Has only partial knowledge
- Must incentivize the agents

agent

- Has private information
- Strategic and self-interested

How Should the FDA Test?

	type	P(approve)	P(non-approve)	
bad drugs	$\theta = 0$	0.05	0.95	(5% type-1 error)
good drugs	$\theta = 1$	0.80	0.20	(80% power)

Is this a good protocol?

Case 1: small profit. $20 million cost to run trial. $200 million if approved.

$\mathbb{E}[\text{profit}|\theta = 0] =$ -$10 million

All approvals are good drugs!

Case 2: large profit. $20 million cost to run trial. $2 billion if approved.

$\mathbb{E}[\text{profit}|\theta = 0] =$ $80 million

Many bad drugs are approved!

Statistical Contracts

Denote the agent's private information as $\theta \in \Theta$

Present the agent with the following opt-in protocol:

our task: **design this menu**

1. Agent pays R
2. Agent chooses payout function f from menu $\mathcal{F}$
3. Statistical trial yields random variable $Z \sim P_\theta$
4. Agent receives payoff $f(Z)$
   Principal receives utility $u(\theta, f(Z))$

Agent acts to maximize their payoff: $f^{\text{br}} = \text{argmax}_{f \in \mathcal{F}} \mathbb{E}_{Z \sim P_\theta}[f(Z)]$

Incentive Alignment

null agents: $\Theta_0 \subset \Theta$ $u(\theta_0, f(Z)) \leq 0$, decreasing in $f(Z)$ for $\theta_0 \in \Theta_0$

nonnull agents: $\Theta \backslash \Theta_0$ $u(\theta_1, f(Z)) \geq 0$, increasing in $f(Z)$ for $\theta_1 \notin \Theta_0$

The principal wants to transact as much as possible with good agents

Definition (Incentive-aligned contract)

A menu $\mathcal{F}$ is *incentive-aligned* if for all $f \in \mathcal{F}$ and $\theta_0 \in \Theta_0$

$$\mathbb{E}_{Z \sim P_{\theta_0}}[f(Z) - R] \leq 0$$ agent's expected profit

note: $p \leq .05$ protocol not incentive aligned

On average, null drugs are not profitable, so null agents are incentivized to drop out

E-values: Statistical Evidence on the Right Scale

Definition

A random variable $X \geq 0$ is an *E-value* for null hypothesis Θ_0 if for all $\theta_0 \in \Theta_0$

$$\mathbb{E}_{Z \sim P_{\theta_0}}[X] \leq 1$$

Theorem

A contract is incentive-aligned if and only if all payoff functions are E-values.

Incentivizing Data Sharing in Federated Learning

- Multiple agents cooperate with each other and with a principal to build a better statistical model than anyone could do unilaterally
  - mostly this literature has developed without considering incentives
  - free riding is a practical concern
- We adapt our statistical contract theory perspective to the problem
  - we design an incentive-compatible mechanism that incentivizes agents to contribute a maximum amount of data (rather than eliciting private types)
  - a key tool is statistical accuracy shaping
- See Karimireddy, P., Guo, W., and Jordan, M. I. (2022). Mechanisms that incentivize data sharing in federated learning. *arXiv:2207.04557*

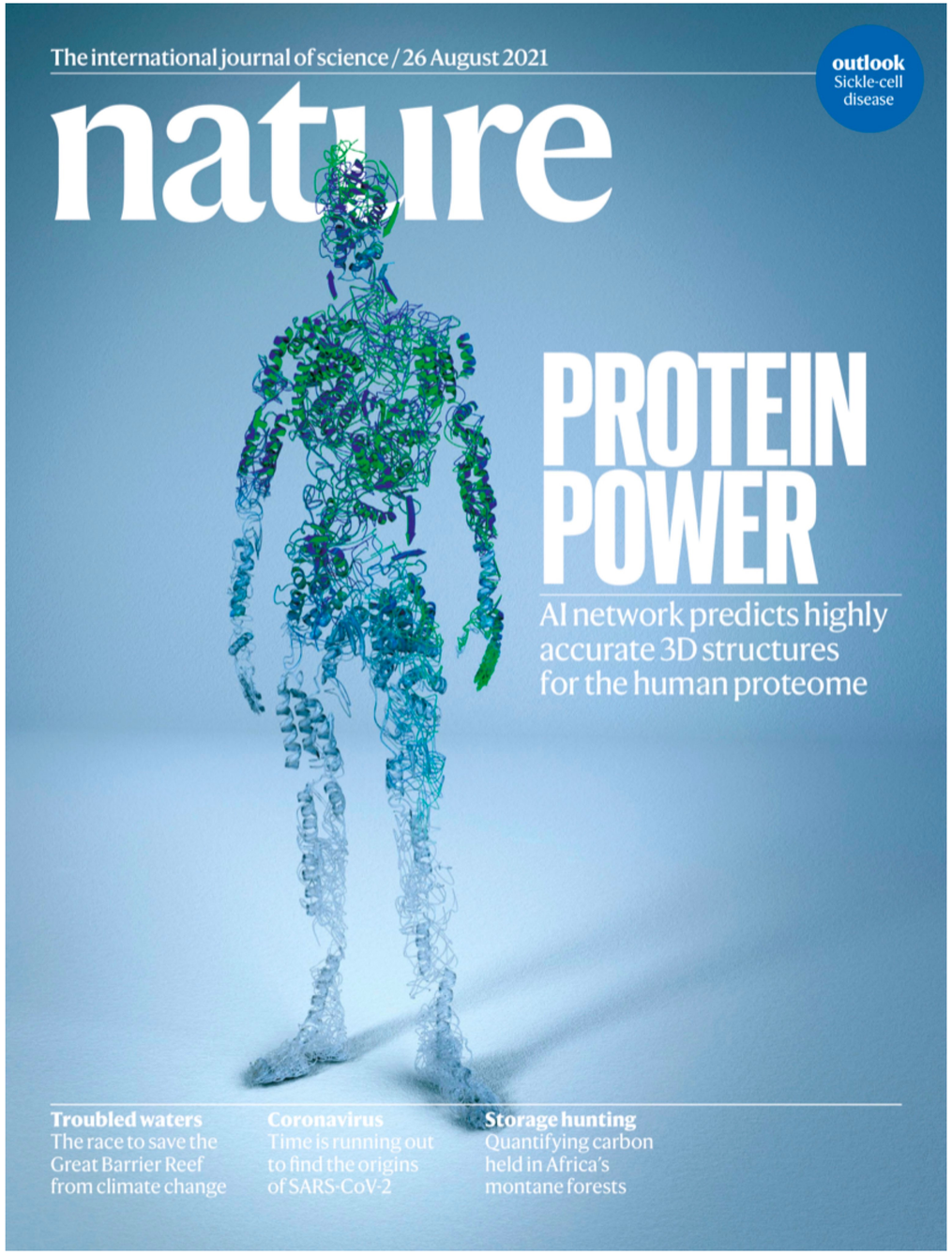
Prediction-Powered Inference

Anastasios Angelopoulos

Stephen Bates

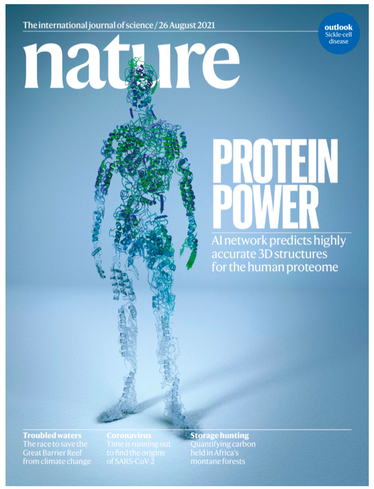
Clara Fannjiang

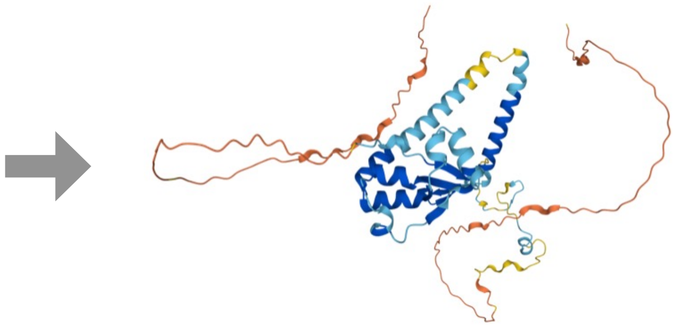
Tijana Zrnic

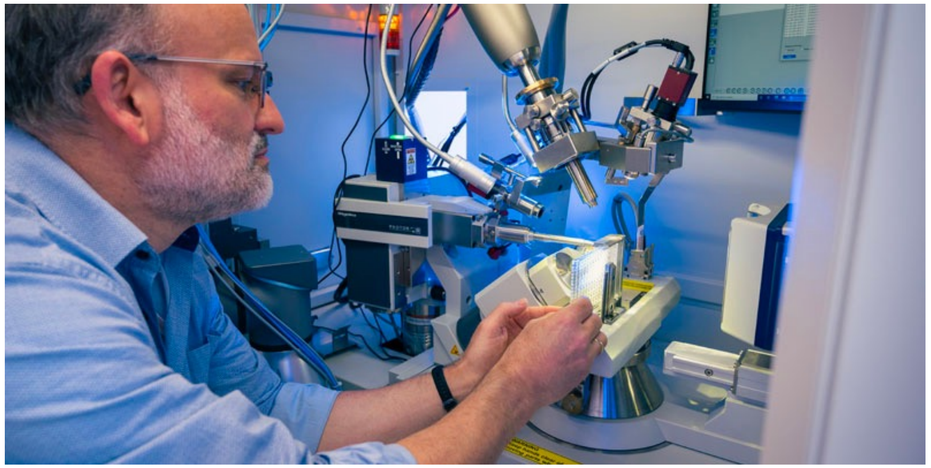

The international journal of science / 26 August 2021
outlook
Sickle-cell disease
nature
PROTEIN POWER
AI network predicts highly accurate 3D structures for the human proteome
Troubled waters
The race to save the Great Barrier Reef from climate change
Coronavirus
Time is running out to find the origins of SARS-CoV-2
Storage hunting
Quantifying carbon held in Africa's montane forests


www.nature.com/nmeth / January 2022 Vol.19 No.1
nature methods
Method of the Year 2021:
Protein structure prediction

Protein structure studies







Hundreds of millions of amino acid sequences with protein structures predicted by AlphaFold

Hundreds of thousands of amino acid sequences with protein structures from X-ray crystallography

Goal: correlate sequence information with structural information

Nucleic Acids Research, 2004, Vol. 32, No. 3 ***1037–1049***
DOI: 10.1093/nar/gkh253

The importance of intrinsic disorder for protein phosphorylation

Lilia M. Iakoucheva, Predrag Radivojac[1], Celeste J. Brown, Timothy R. O'Connor, Jason G. Sikes, Zoran Obradovic[1] and A. Keith Dunker*

Nucleic Acids Research, 2004, Vol. 32, No. 3 ***1037–1049***
DOI: 10.1093/nar/gkh253

The importance of intrinsic disorder for protein phosphorylation

Lilia M. Iakoucheva, Predrag Radivojac[1], Celeste J. Brown, Timothy R. O'Connor, Jason G. Sikes, Zoran Obradovic[1] and A. Keith Dunker*

2004

Not enough structures overlapping
with post-translational modification (PTM) data.

2004

~10k structures in PDB

Nucleic Acids Research, 2004, Vol. 32, No. 3 ***1037–1049***
DOI: 10.1093/nar/gkh253

The importance of intrinsic disorder for protein phosphorylation

Lilia M. Iakoucheva, Predrag Radivojac[1], Celeste J. Brown, Timothy R. O'Connor, Jason G. Sikes, Zoran Obradovic[1] and A. Keith Dunker*

2004

Not enough structures overlapping
with post-translational modification (PTM) data.

METHODS AND RESOURCES PLOS BIOLOGY **Published:** May 16, 2022

The structural context of posttranslational modifications at a proteome-wide scale

Isabell Bludau[1], Sander Willems[1], Wen-Feng Zeng[1], Maximilian T. Strauss[2], Fynn M. Hansen[1], Maria C. Tanzer[1], Ozge Karayel[1], Brenda A. Schulman[3], Matthias Mann[1,2]*

2022

Quantify association between PTMs and IDRs by computing:

$$\text{odds ratio } \frac{\mathbb{P}(IDR|PTM)}{\mathbb{P}(IDR|no\ PTM)}$$

2022
200+ million AlphaFold-predicted structures

2004
~10k experimental structures
in PDB

Nucleic Acids Research, 2004, Vol. 32, No. 3 ***1037–1049***
DOI: 10.1093/nar/gkh253

The importance of intrinsic disorder for protein phosphorylation

Lilia M. Iakoucheva, Predrag Radivojac[1], Celeste J. Brown, Timothy R. O'Connor, Jason G. Sikes, Zoran Obradovic[1] and A. Keith Dunker*

2004

Not enough structures overlapping
with post-translational modification (PTM) data.

METHODS AND RESOURCES PLOS BIOLOGY **Published:** May 16, 2022

The structural context of posttranslational modifications at a proteome-wide scale

Isabell Bludau[1], Sander Willems[1], Wen-Feng Zeng[1], Maximilian T. Strauss[2], Fynn M. Hansen[1], Maria C. Tanzer[1], Ozge Karayel[1], Brenda A. Schulman[3], Matthias Mann[1,2]*

2022 predicted IDRs

Quantify association between PTMs and IDRs by computing:

predicted IDRs

$$\text{odds ratio} \quad \frac{\mathbb{P}(IDR|PTM)}{\mathbb{P}(IDR|no\ PTM)}$$

2022
200+ million AlphaFold-predicted structures

2004
~10k experimental structures
in PDB

Predictions are being used for scientific inquiry.

Nucleic Acids Research, 2004, Vol. 32, No. 3 ***1037–1049***
DOI: 10.1093/nar/gkh253

The importance of intrinsic disorder for protein phosphorylation

Lilia M. Iakoucheva, Predrag Radivojac[1], Celeste J. Brown, Timothy R. O'Connor, Jason G. Sikes, Zoran Obradovic[1] and A. Keith Dunker*

2004

Not enough structures overlapping
with post-translational modification (PTM) data.

METHODS AND RESOURCES PLOS BIOLOGY **Published:** May 16, 2022

The structural context of posttranslational modifications at a proteome-wide scale

Isabell Bludau[1], Sander Willems[1], Wen-Feng Zeng[1], Maximilian T. Strauss[2], Fynn M. Hansen[1], Maria C. Tanzer[1], Ozge Karayel[1], Brenda A. Schulman[3], Matthias Mann[1,2]*

2022 predicted IDRs

Quantify association between PTMs and IDRs by computing:

predicted IDRs

$$\text{odds ratio} \quad \frac{\mathbb{P}(IDR|PTM)}{\mathbb{P}(IDR|no\ PTM)}$$

2022
200+ million AlphaFold-predicted structures

2004
~10k experimental structures
in PDB

Predictions are being used for scientific inquiry.

Article

nature

Disease variant prediction with deep generative models of evolutionary data

https://doi.org/10.1038/s41586-021-04043-8

Received: 18 December 2020

Jonathan Frazer[1,4], Pascal Notin[2,4], Mafalda Dias[1,4], Aidan Gomez[2], Joseph K. Min[1], Kelly Brock[1], Yarin Gal[2]✉ & Debora S. Marks[1,3]✉

METHODS AND RESOURCES PLOS BIOLOGY

Published: May 16, 2022

The structural context of posttranslational modifications at a proteome-wide scale

Isabell Bludau[1], Sander Willems[1], Wen-Feng Zeng[1], Maximilian T. Strauss[2], Fynn M. Hansen[1], Maria C. Tanzer[1], Ozge Karayel[1], Brenda A. Schulman[3], Matthias Mann[1,2]*

2022 ~~predicted IDRs~~

Quantify association between PTMs and ~~IDRs~~ by computing:

~~predicted IDRs~~

odds ratio $\frac{\mathbb{P}(\sout{IDR}|PTM)}{\mathbb{P}(\sout{IDR}|no\ PTM)}$

RESEARCH ARTICLES

ECONOMICS

Combining satellite imagery and machine learning to predict poverty

Neal Jean,[1,2]* Marshall Burke,[3,4,5]*† Michael Xie,[1] W. Matthew Davis,[4] David B. Lobell,[3,4] Stefano Ermon[1]

Article

Using machine learning to assess the livelihood impact of electricity access

https://doi.org/10.1038/s41586-022-05322-8

Received: 1 September 2021

Nathan Ratledge[1,2], Gabe Cadamuro[3], Brandon de la Cuesta[4], Matthieu Stigler[5] & Marshall Burke[6,7,8]✉

Article

The evolution, evolvability and engineering of gene regulatory DNA

https://doi.org/10.1038/s41586-022-04506-6

Received: 8 February 2021

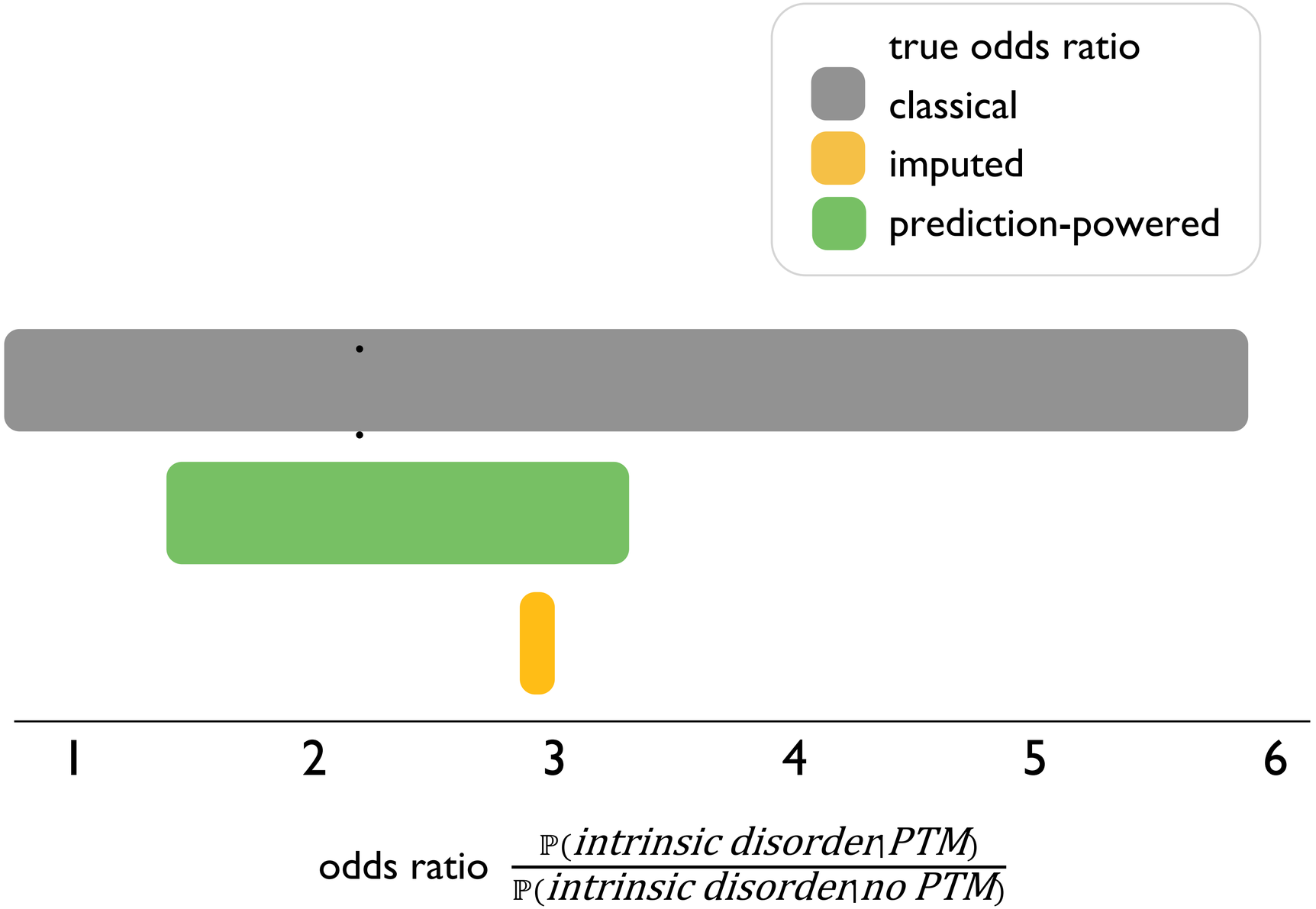
Eeshit Dhaval Vaishnav[1,2,12]✉, Carl G. de Boer[3,4,12]✉, Jennifer Molinet[5,6], Moran Yassour[4,7,8], Lin Fan[2], Xian Adiconis[4,9], Dawn A. Thompson[2], Joshua Z. Levin[4,9], Francisco A. Cubillos[5,6] & Aviv Regev[4,10,11]✉

Research and Applications *Journal of the American Medical Informatics Association*

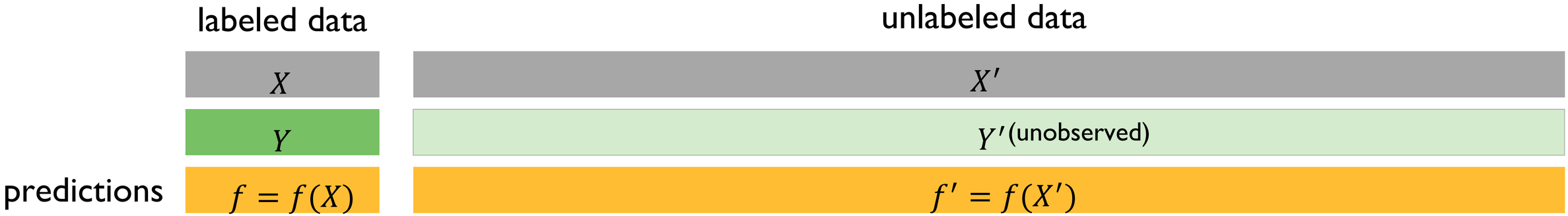
POPDx: an automated framework for patient phenotyping across 392 246 individuals in the UK Biobank study

Lu Yang[1], Sheng Wang[2], and Russ B. Altman[1,3,4]

Prediction-powered inference



Prediction-powered inference: problem setting



Estimand of interest (mean, quantile, regression coefficient, *etc.*): $\theta^\star$

Goal: construct confidence set, C_α^{PP}, that are **valid**:

$$\mathbb{P}(\theta^\star \in C_\alpha^{\mathrm{PP}}) \geq 1 - \alpha$$

classical approach

use only labeled data

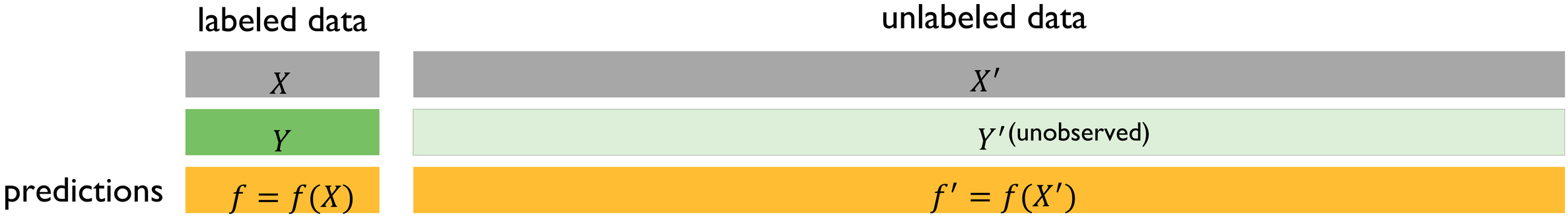
valid, but lose out on information from abundant predictions

imputed approach

treat predictions as gold-standard labels

abundant predictions, but **invalid** because predictions can contain systematic errors

Prediction-powered inference: problem setting



Estimand of interest (mean, quantile, regression coefficient, *etc.*): $\theta^\star$

Goal: construct confidence set, C_α^{PP}, that are **valid**:

$$\mathbb{P}(\theta^\star \in C_\alpha^{\mathrm{PP}}) \geq 1 - \alpha$$

classical approach

use only labeled data

valid, but lose out on information from abundant predictions

imputed approach

treat predictions as gold-standard labels

abundant predictions, but **invalid** because predictions can contain systematic errors

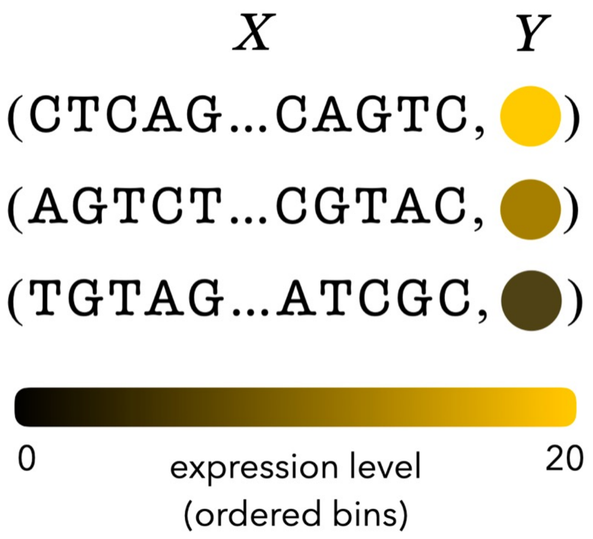
We want the best of both worlds.

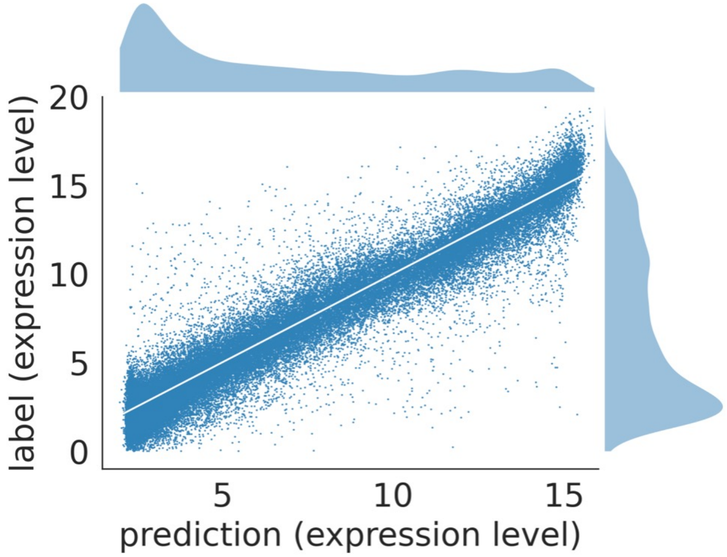
Gene expression

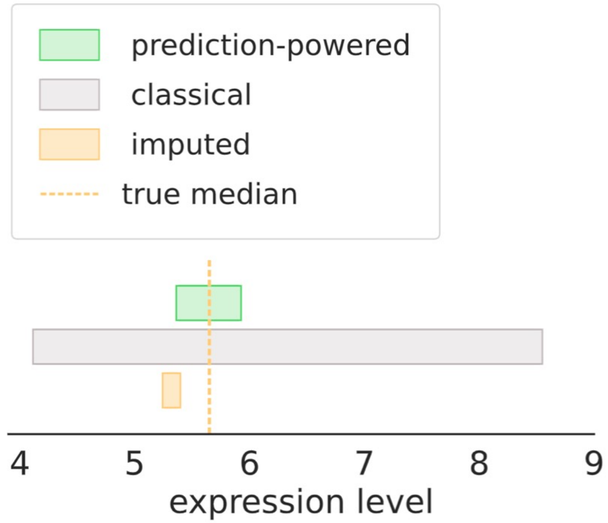
- Want to estimate median gene expression level with differing *promoters* (regulatory DNA)

(Vaishnav et. al. *Nature* '22)

- Predictive model: transformer developed in Vaishnav et. al.

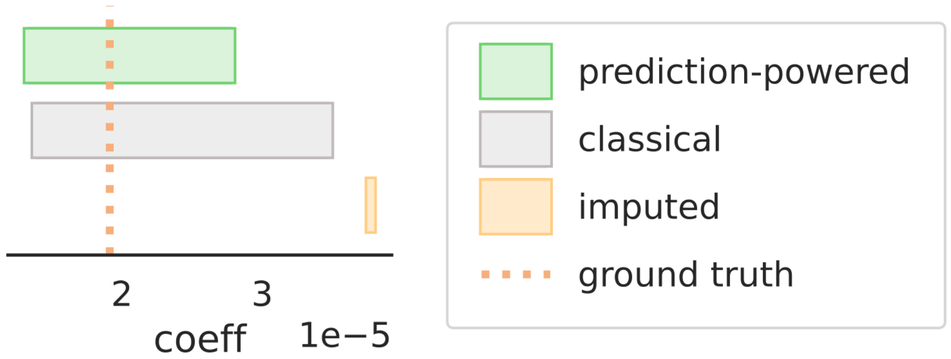






California census

- 2018 CA census data
- Estimand: logistic regression coefficient of income when predicting whether person has private health insurance
- Boosting model based on ten other covariates



Principle of prediction-powered inference

For the mean value of Y:

rectifier is the **bias**

1. Identify Rectifier

The rectifier, Δ^f, is a estimand-specific notion of error.

We give a general recipe for identifying the rectifier.

2. Confidence Set on Rectifier

Use the labeled data to construct a confidence set, R, for the rectifier.

3. Prediction-Powered Confidence Set

Construct C^{PP} by including all possible rectified values of θ^f.

Convex Estimation Problems

$\theta^\star = \arg\min_\theta \mathbb{E}[\ell_\theta(X, Y)]$ e.g. mean, median, quantiles; linear, logistic regression coefficients

gradient of loss $g_\theta(X, Y) \equiv \frac{\partial}{\partial \theta} \ell_\theta(X, Y)$

Build confidence set that contains $\theta^\star$: the value of θ such that $\mathbb{E}[g_\theta(X, Y)] = 0$.

rectifier $\boldsymbol{\Delta}_\theta^f$

estimate using only predictions

$$\mathbb{E}[g_\theta(X, f)] - \mathbb{E}[(g_\theta(X, f) - g_\theta(X, Y))] = 0$$

build confidence set R_θ for rectifier using labeled data: $g_\theta(X_i, f_i) - g_\theta(X_i, Y_i)$

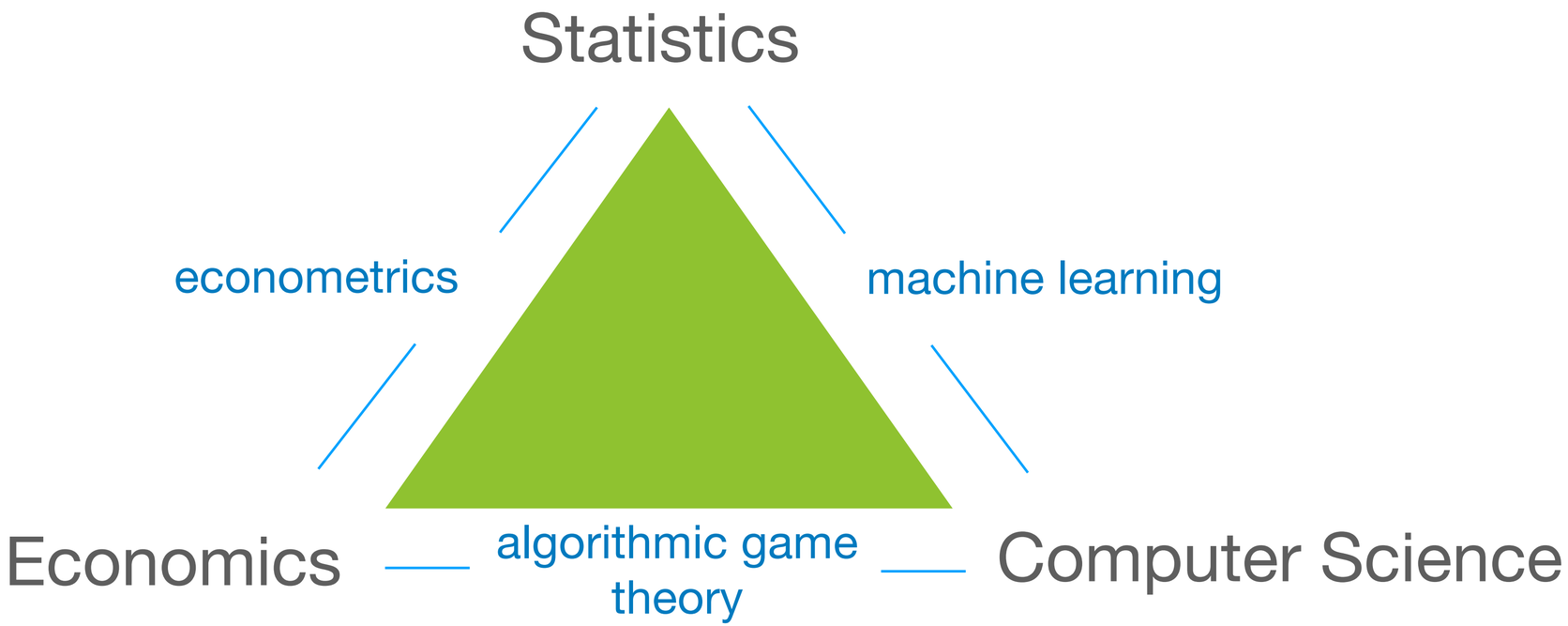
Theorem. Take $C^{\text{PP}} = \{\theta : 0 \in \mathbb{E}[g_\theta(X, f)] - R_\theta\}$, where for each θ, the confidence set R_θ contains the rectifier Δ_θ^f with probability at least $1 - \alpha$. Then, C^{PP} is valid:

$$\mathbb{P}(\theta^\star \in C^{\text{PP}}) \geq 1 - \alpha.$$

A Personal View on “AI”

- It reflects the emergence of a new engineering field, embodied in large-scale systems that link humans in new ways
- Cf. chemical engineering in the 40s and 50s
  - built on chemistry, fluid mechanics, etc
  - driven by the possibility of building chemical factories
- Cf. electrical engineering in the late 19th century
  - built on electromagnetism, optics, etc
  - clear goals in terms of human welfare
- The new field builds on inferential ideas, algorithmic ideas, and economic ideas from the past three centuries
- But its emergence is being warped by being cast in terms of poorly thought-through, naïve, old-style AI aspirations

Three Foundational Disciplines



Some Further Reading

(see www.cs.berkeley.edu/~jordan/publications.html)

- Incentive-theoretic Bayesian inference for collaborative science. S. Bates, M. I. Jordan, M. Sklar, and J. A. Soloff. *arxiv.org/abs/2307.03748*, 2023.

- Evaluating and incentivizing diverse data contributions in collaborative learning. B. Huang, S. P. Karimireddy, and M. I. Jordan. *arXiv:2306.05592*, 2023.

- A unifying perspective on multi-calibration: Unleashing game dynamics for multi-objective learning. N. Haghtalab, M. I. Jordan, and E. Zhao. *arXiv:2302.10863*, 2023.