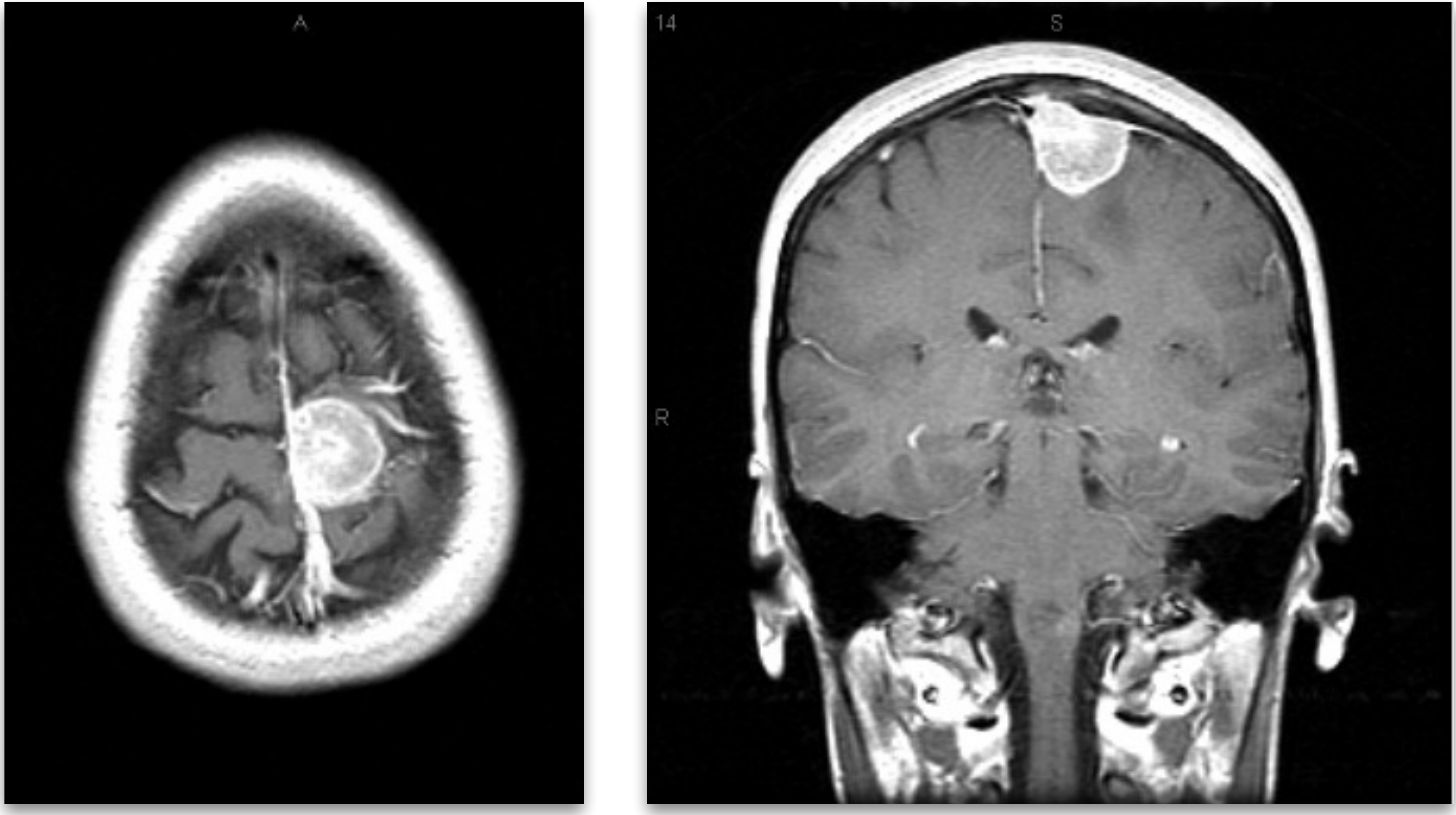
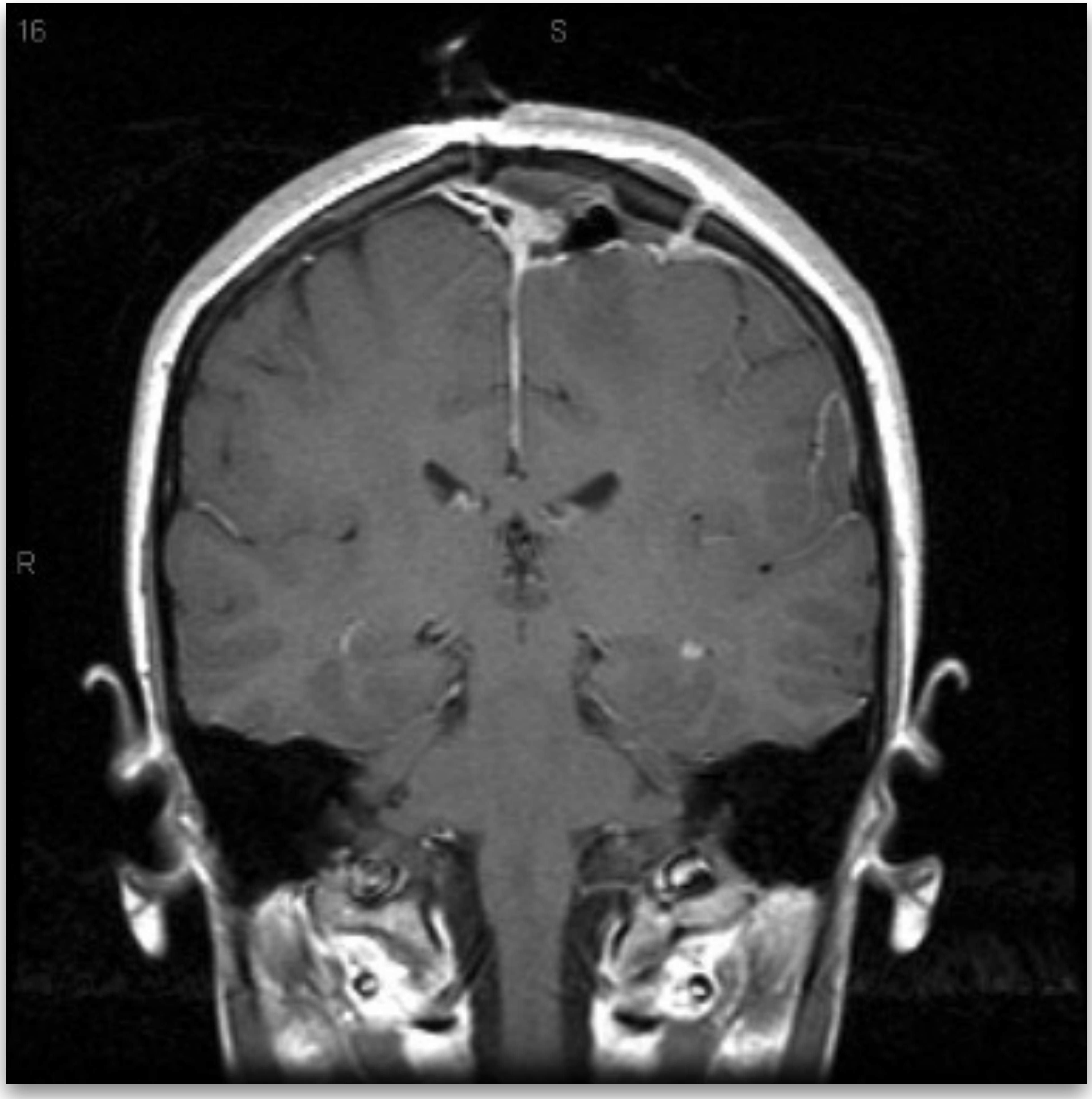
Machine Learning Foundations Accelerate Innovation and Promote Trustworthiness

Rebecca Willett
University of Chicago


MAGNETOM Spectra
SIEMENS


A
14
S
R


A


16
S
R

AI will affect every step of this process

Groundbreaking Discoveries and Translation

- Develop a new understanding of the laws of nature and rules of life
- Accelerate affordable drug development
- Engineer green materials
- Build quantum computers
- Develop sustainable climate policies

image credit: https://www.greenbiz.com/article/whats-your-sustainability-moonshot

Developing **applied machine learning** without understanding **math, stats, & CS foundations** is like developing **biotech** without understanding **biology**.

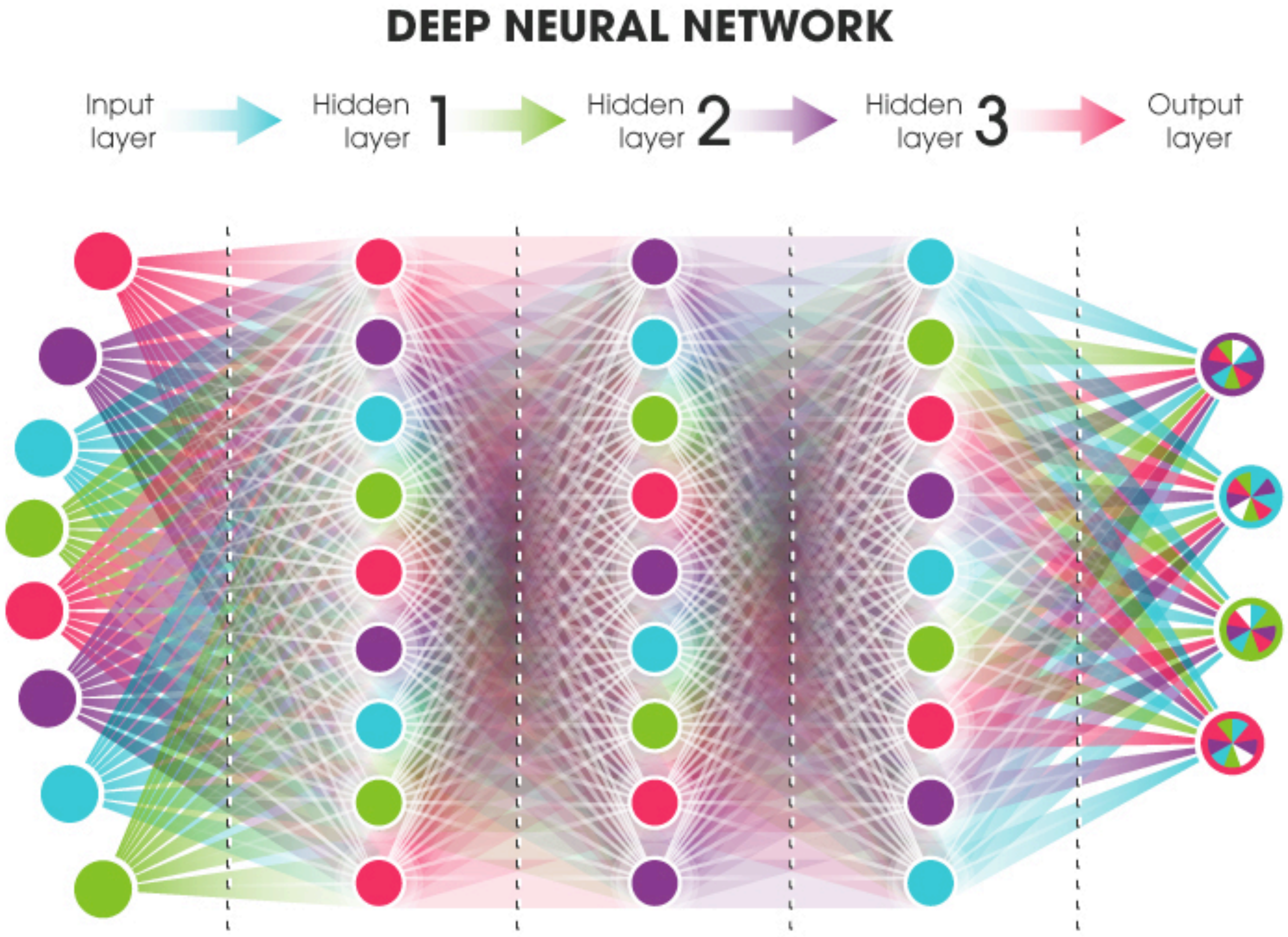
Machine learning foundations' impact

Emerging and future directions

Machine learning foundations' impact

Emerging and future directions

Faster optimization methods for ML



neuralnetworksanddeeplearning.com - Michael Nielsen, Yoshua Bengio, Ian Goodfellow, and Aaron Courville, 2016.

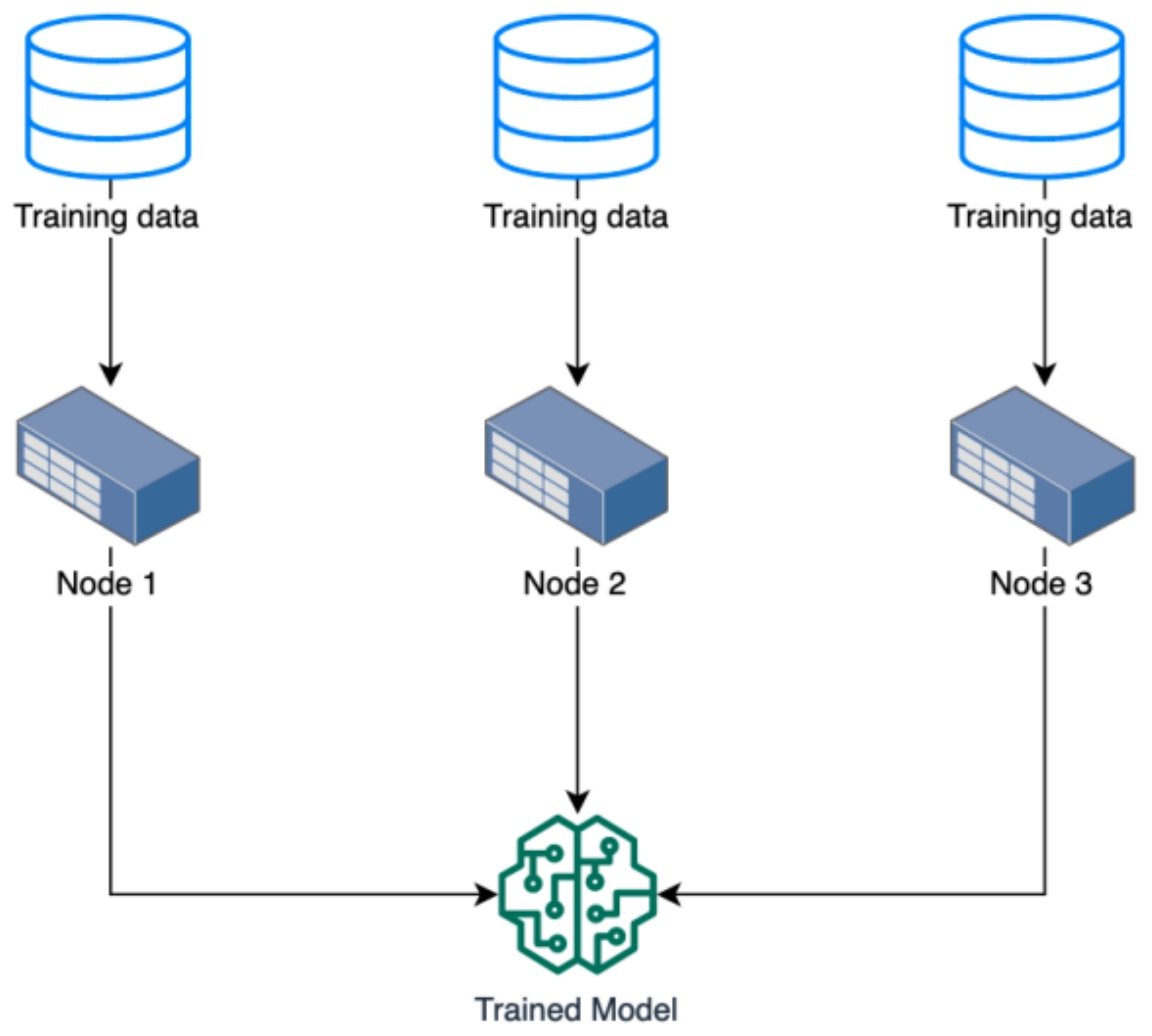
- Machine learning uses training data to set parameters of a model (e.g. neural network weights)
- We train models using optimization methods which update parameters based on gradients of the loss

Adagrad adapts gradients to past estimates, accelerating training; foundation of popular methods like **Adam**

Duchi, Hazan, Singer, 2011
Kingma and Ba, 2014

Faster optimization across multiple computers

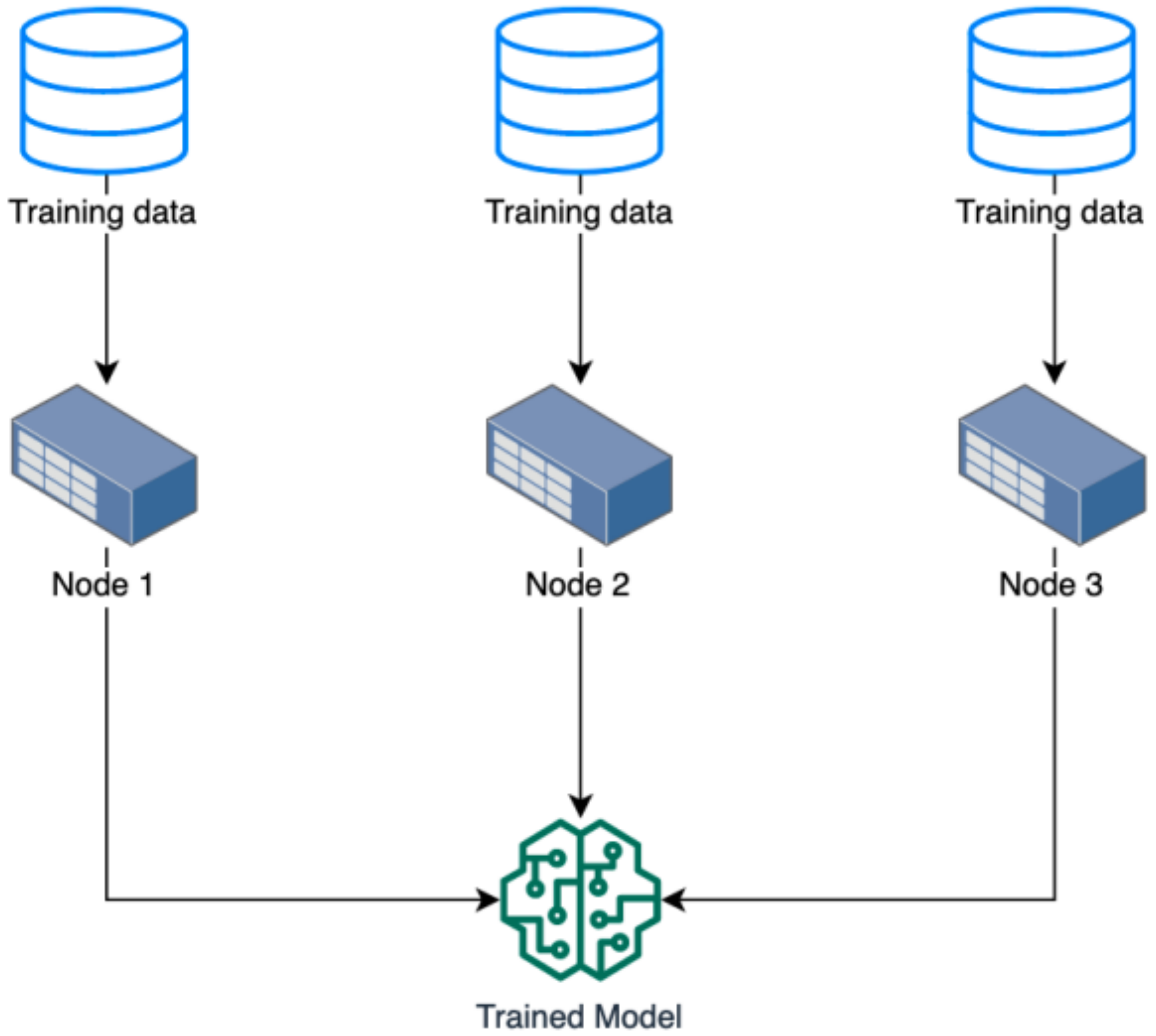
- Large-scale machine learning is typically distributed across multiple machines
- **Expectation**: more machines = **faster** computation
- **Reality** with naïve distributed optimization: more machines = **diminishing returns**



Recht, Re, Wright, Niu, 2011

Faster optimization across multiple computers

- Large-scale machine learning is typically distributed across multiple machines
- **Expectation**: more machines = **faster** computation
- **Reality** with naïve distributed optimization: more machines = **diminishing returns**

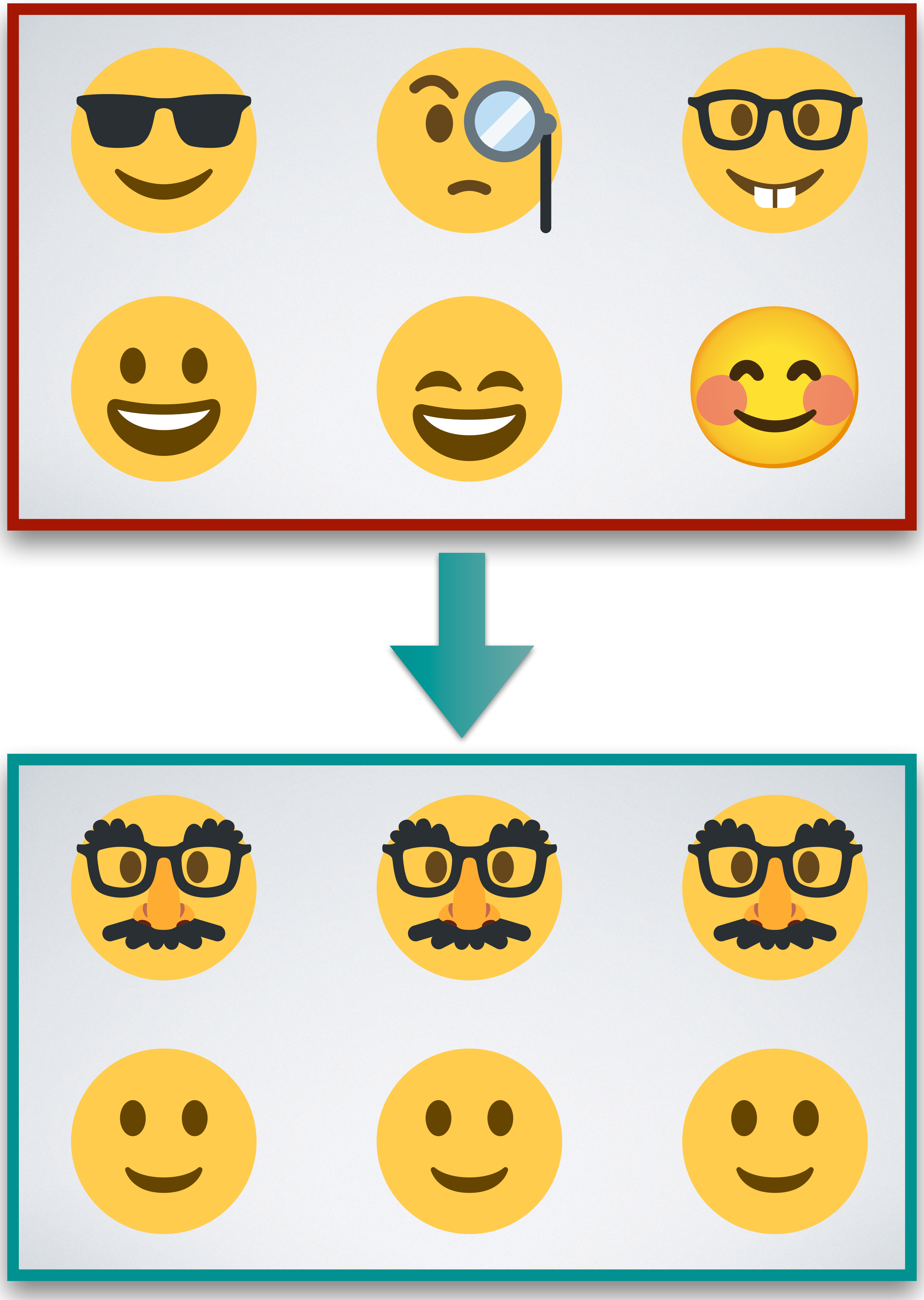


Hogwild: theoretically-grounded **asynchronous** distributed optimization ⇒ faster computation with more machines

Recht, Re, Wright, Niu, 2011

Privacy guarantees

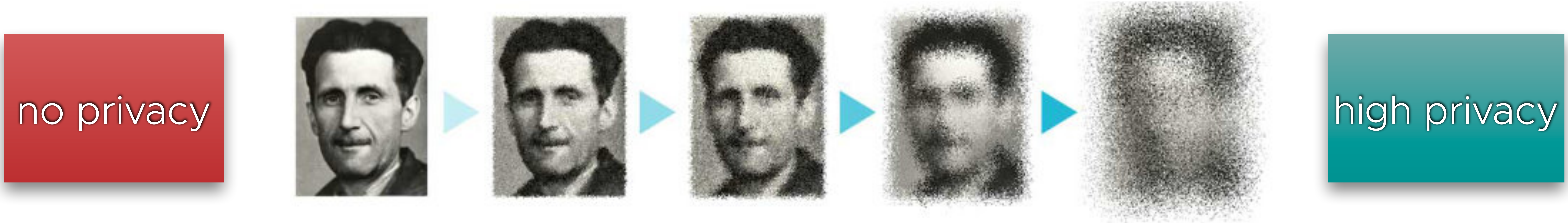
- Common standard: k-anonymity, which transforms data just enough to make each individual indistinguishable from k others in the data set.
- Legally sufficient for fulfilling privacy-protection regulations such as HIPAA and GDPR
- Foundational insight: users redact the *minimum possible* to satisfy k-anonymity. Knowing they redacted the minimum provides additional information about what was redacted.



Cohen, 2022

Privacy guarantees

- How can we preserve the privacy of people represented by our data?
- Classical approach: **aggregate data**
  - E.g., only release summary statistics for 10 or more people.
  - Without more conditions, very easy to break
- More recent: **differential privacy**
  - E.g., randomly perturb data
  - Guarantees that someone seeing algorithm output cannot tell if a particular person's data was used



Dwork, McSherry, Nissim and Smith, 2006
Image: https://aircloak.com/explaining-differential-privacy/

Quantifying uncertainty in predictions

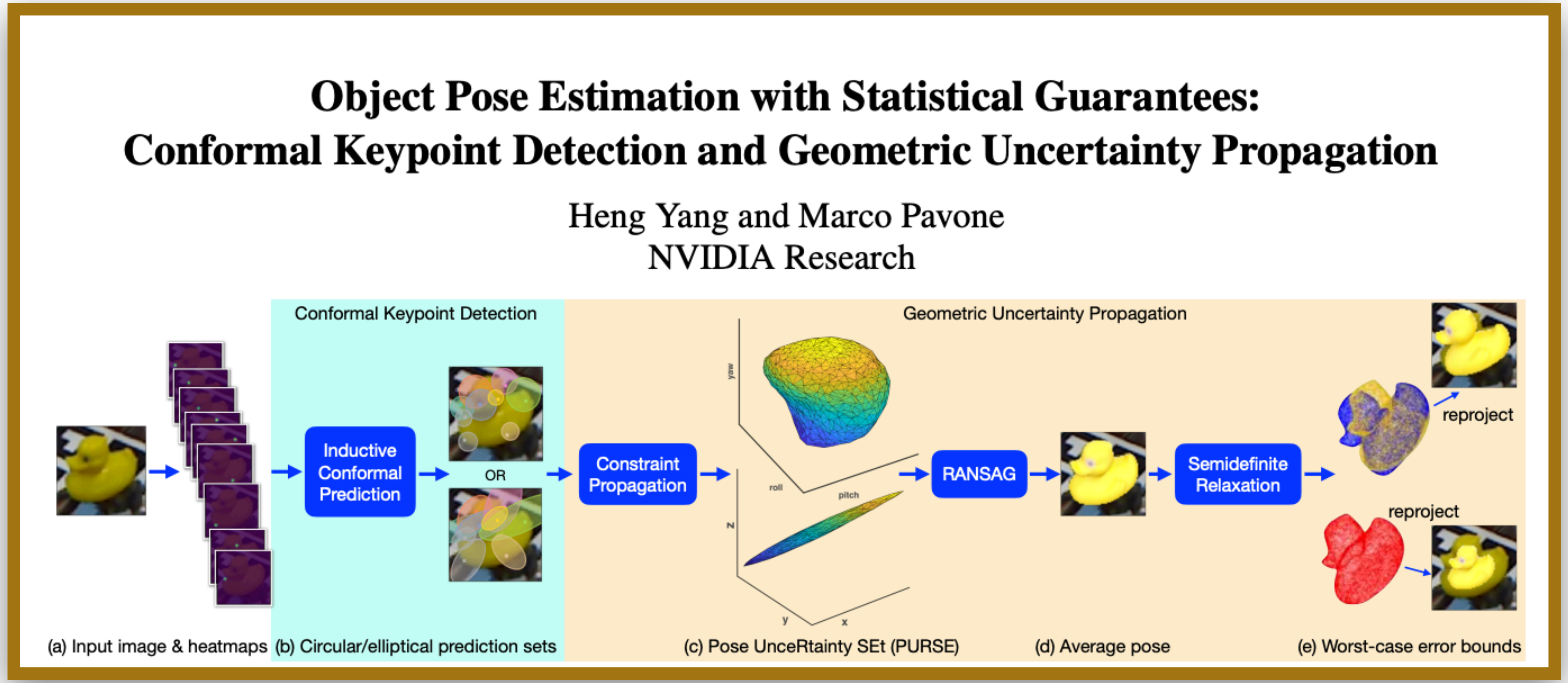
https://www.allianz.com/en/press/news/commitment/environment/190912_Allianz-drones-and-hurricanes-flying-masters-of-disasters.html

- We need not only **raw ML predictions**; we also want to know **how certain the ML model is about its prediction**
- Essential in climate analysis, model predictive control, automatic translation…
- Classical methods required either simple models (i.e., no neural networks) or strong prior knowledge

Conformal prediction allows us to assess uncertainties of ML predictions with theoretical guarantees and minimal assumptions

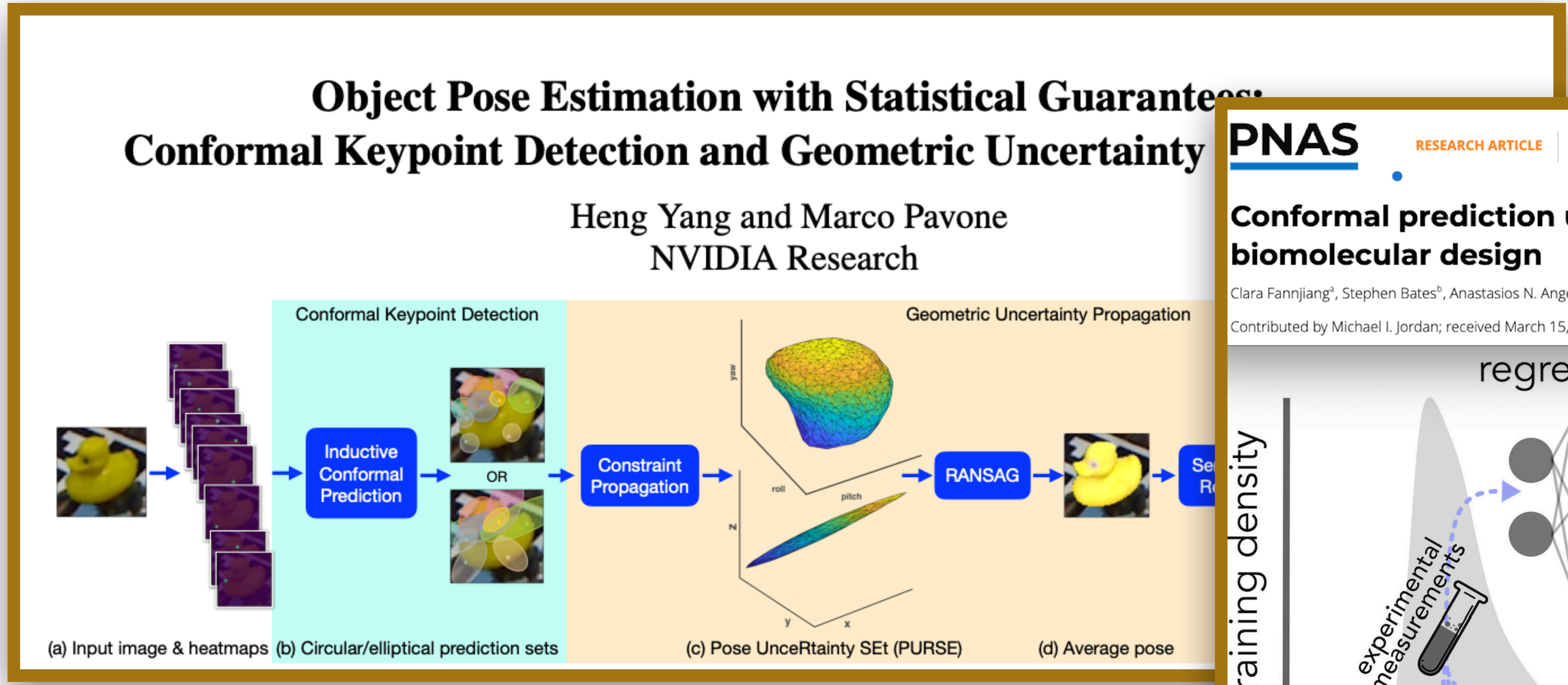
Lei, G'Sell, Rinaldo, Tibshirani, Wasserman, 2016
Tibshirani, Barber, Candès, and Ramdas, 2019

Conformal prediction allows us to assess uncertainties of ML predictions with theoretical guarantees and minimal assumptions

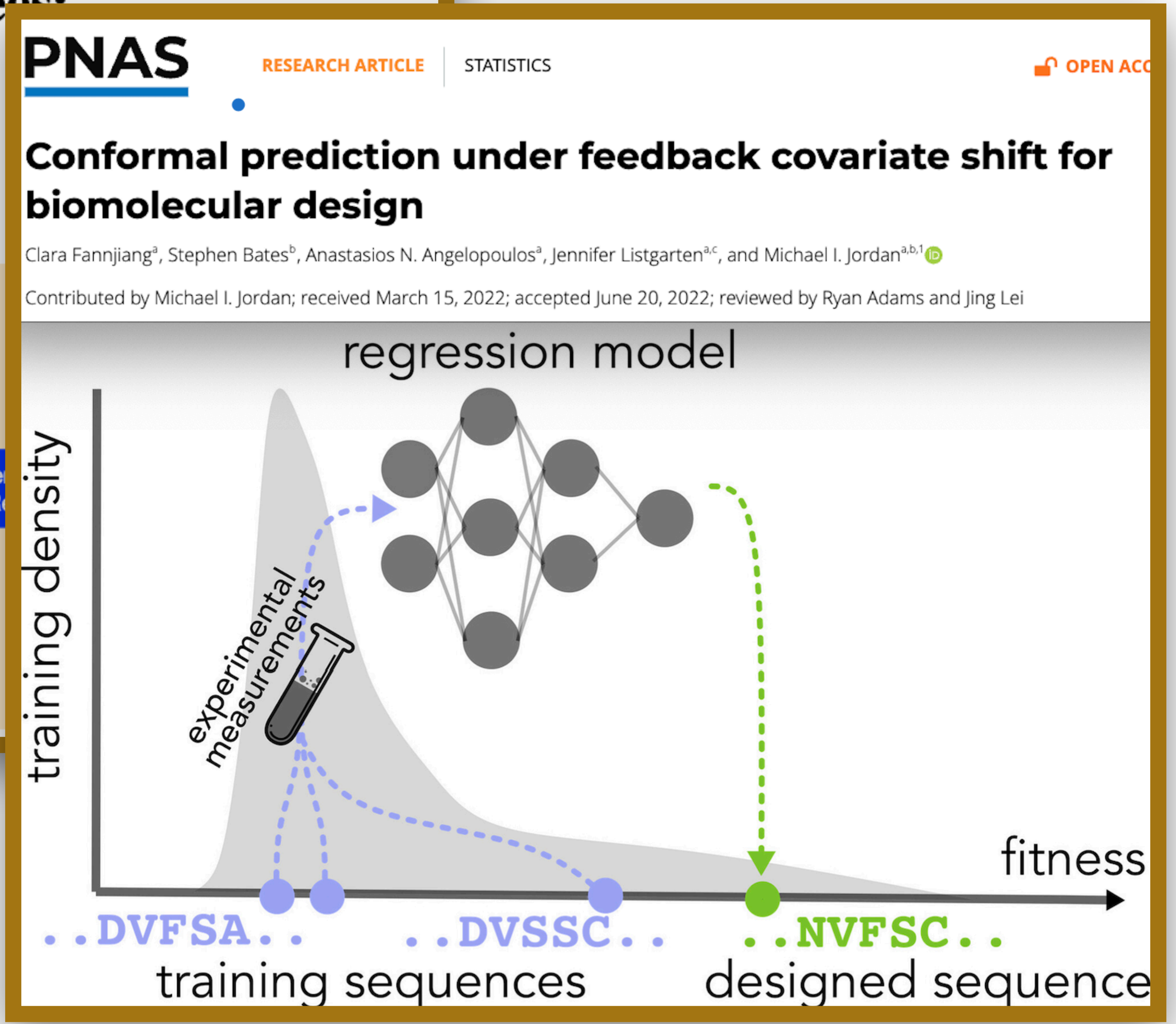


Lei, G'Sell, Rinaldo, Tibshirani, Wasserman, 2016
Tibshirani, Barber, Candès, and Ramdas, 2019

Conformal prediction allows us to assess uncertainties of ML predictions with theoretical guarantees and minimal assumptions


Object Pose Estimation with Statistical Guarantees: Conformal Keypoint Detection and Geometric Uncertainty

Heng Yang and Marco Pavone
NVIDIA Research


PNAS RESEARCH ARTICLE | STATISTICS

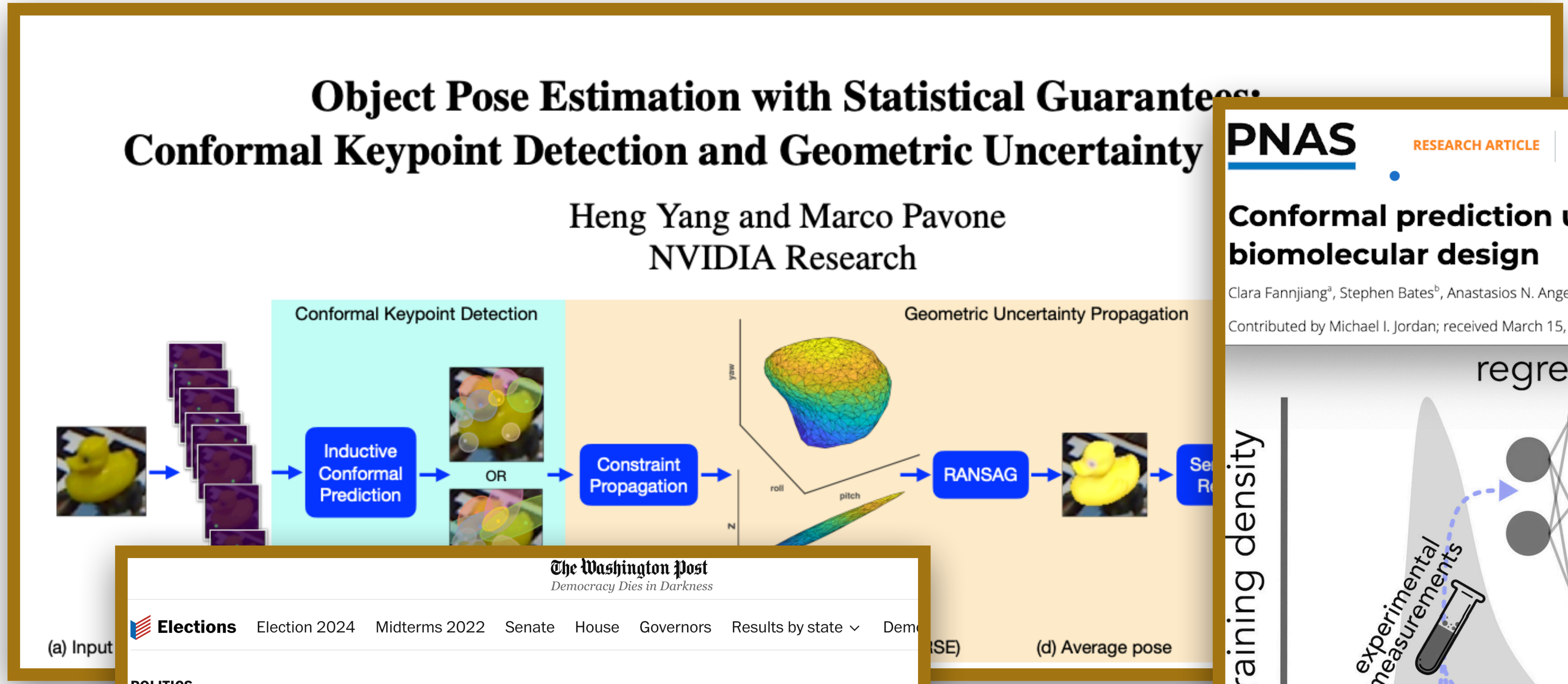
Conformal prediction under feedback covariate shift for biomolecular design

Clara Fannjiang[a], Stephen Bates[b], Anastasios N. Angelopoulos[a], Jennifer Listgarten[a,c], and Michael I. Jordan[a,b,1]

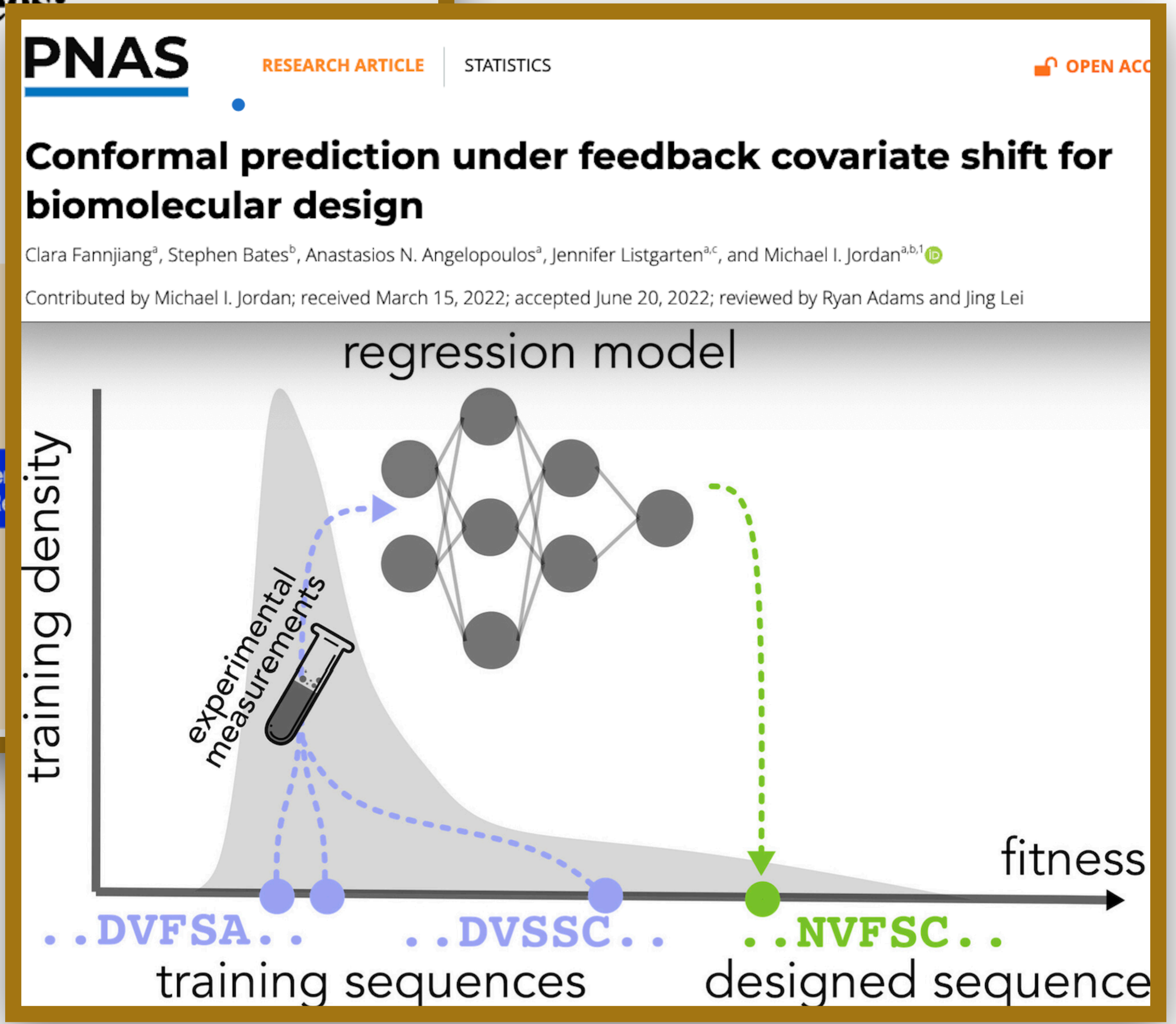
Contributed by Michael I. Jordan; received March 15, 2022; accepted June 20, 2022; reviewed by Ryan Adams and Jing Lei

Lei, G'Sell, Rinaldo, Tibshirani, Wasserman, 2016
Tibshirani, Barber, Candès, and Ramdas, 2019

Conformal prediction allows us to assess uncertainties of ML predictions with theoretical guarantees and minimal assumptions


Object Pose Estimation with Statistical Guarantees: Conformal Keypoint Detection and Geometric Uncertainty

Heng Yang and Marco Pavone
NVIDIA Research


PNAS RESEARCH ARTICLE STATISTICS OPEN ACC

Conformal prediction under feedback covariate shift for biomolecular design

Clara Fannjiang[a], Stephen Bates[b], Anastasios N. Angelopoulos[a], Jennifer Listgarten[a,c], and Michael I. Jordan[a,b,1]

Contributed by Michael I. Jordan; received March 15, 2022; accepted June 20, 2022; reviewed by Ryan Adams and Jing Lei


The Washington Post
Democracy Dies in Darkness

Elections Election 2024 Midterms 2022 Senate House Governors Results by state Dem

POLITICS

How The Washington Post will model possible outcomes in the Virginia governor's race

By Lenny Bronner

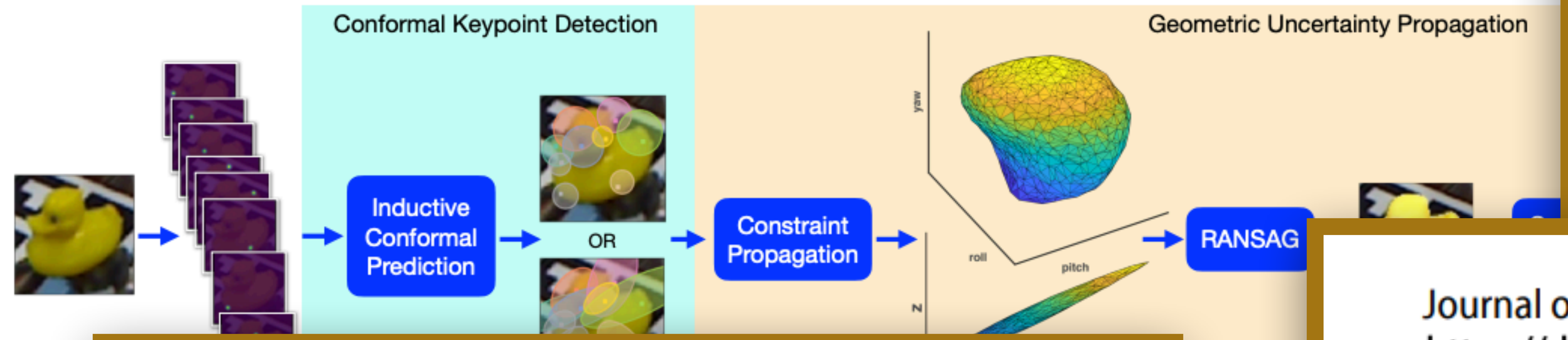
November 2, 2021 at 1:57 p.m. EDT

Lei, G'Sell, Rinaldo, Tibshirani, Wasserman, 2016
Tibshirani, Barber, Candès, and Ramdas, 2019

Conformal prediction allows us to assess uncertainties of ML predictions with theoretical guarantees and minimal assumptions

Object Pose Estimation with Statistical Guarantees: Conformal Keypoint Detection and Geometric Uncertainty

Heng Yang and Marco Pavone
NVIDIA Research



(a) Input ... SE)

PNAS RESEARCH ARTICLE | STATISTICS OPEN ACC

Conformal prediction under feedback covariate shift for biomolecular design

Clara Fannjiang[a], Stephen Bates[b], Anastasios N. Angelopoulos[a], Jennifer Listgarten[a,c], and Michael I. Jordan[a,b,1]

Contributed by Michael I. Jordan; received March 15, 2022; accepted June 20, 2022; reviewed by Ryan Adams and Jing Lei

regression model

The Washington Post
Democracy Dies in Darkness

Elections | Election 2024 | Midterms 2022 | Senate | House | Governors | Results by state | Dem

POLITICS

How The Washington Post will model possible outcomes in the Virginia governor's race

By Lenny Bronner
November 2, 2021 at 1:57 p.m. EDT

Journal of Healthcare Informatics Research (2022) 6:241–252
https://doi.org/10.1007/s41666-021-00113-8

REVIEW ARTICLE

Conformal Prediction in Clinical Medical Sciences

Janette Vazquez[1] · Julio C. Facelli[1]

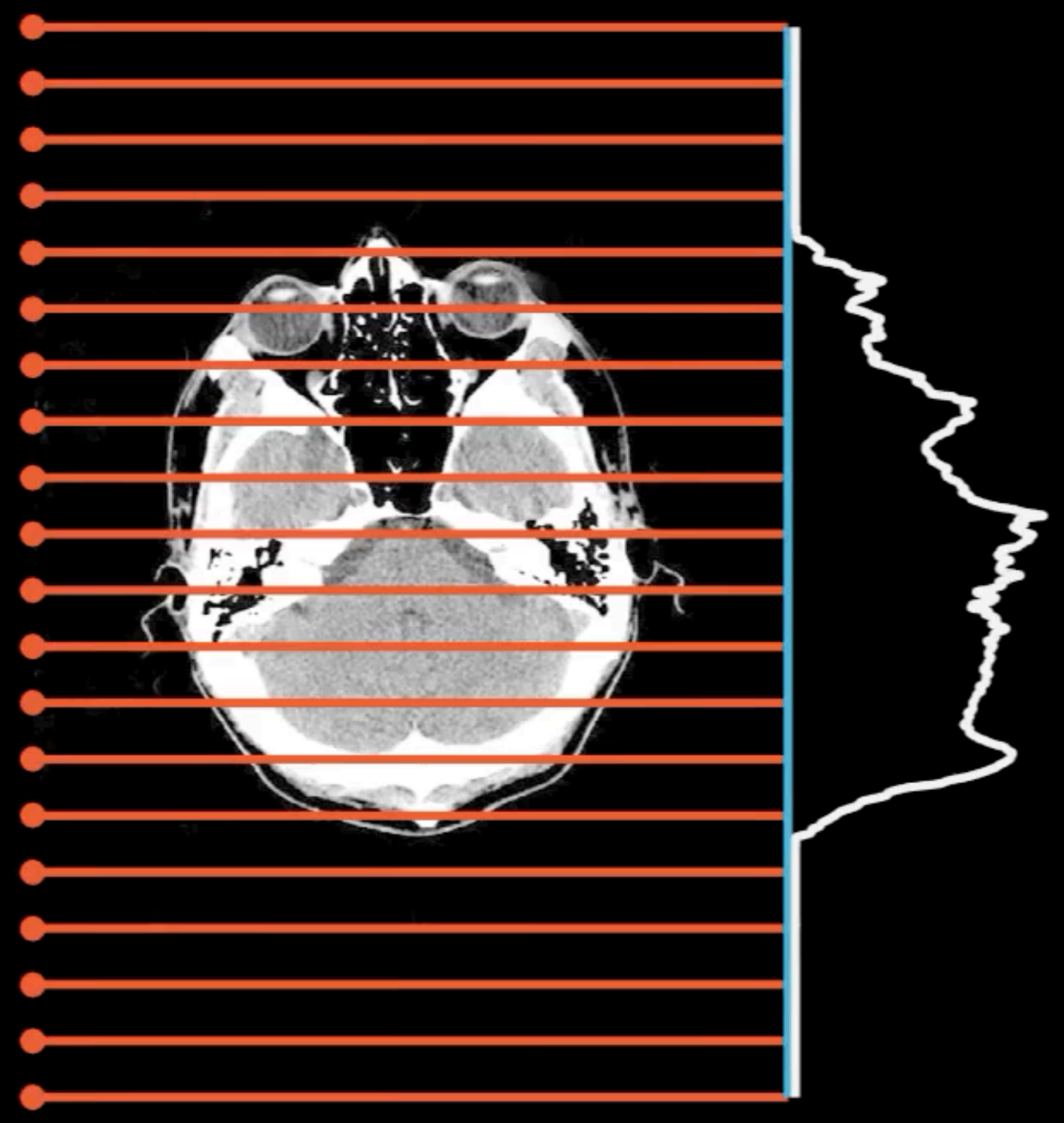
Lei, G'Sell, Rinaldo, Tibshirani, Wasserman, 2016
Tibshirani, Barber, Candès, and Ramdas, 2019

Allocating data collection resources

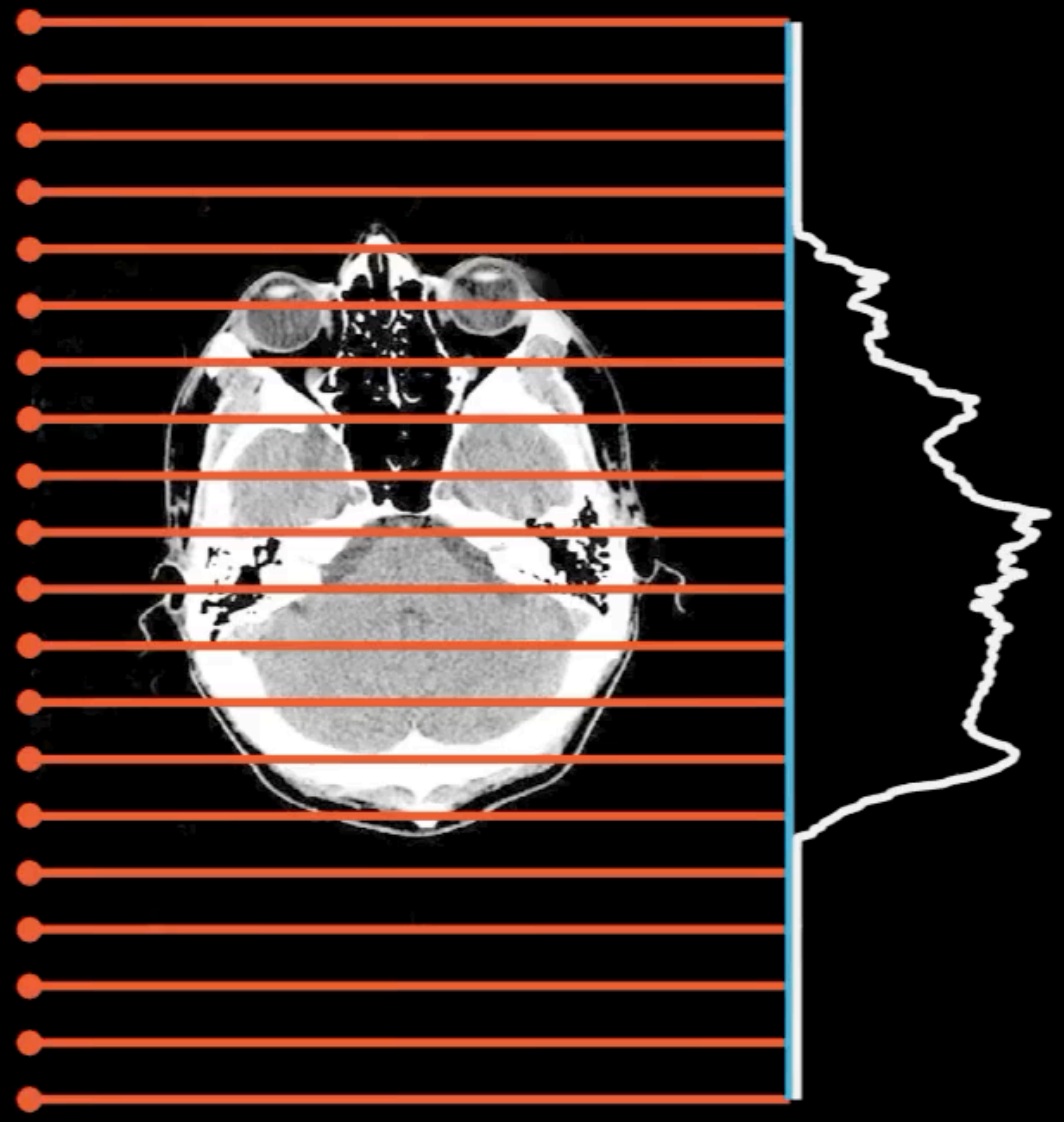
- Collecting data and assigning labels for training data is laborious and expensive
- Bandit algorithms, active learning, and Bayesian optimization guide data collection and labeling
- Widely used throughout industry (e.g., for ad placement)

https://en.wikipedia.org/wiki/File:Las_Vegas_slot_machines.jpg

Thompson, 1933
Russo, Van Roy, Kazerouni, Osband & Wen, 2018



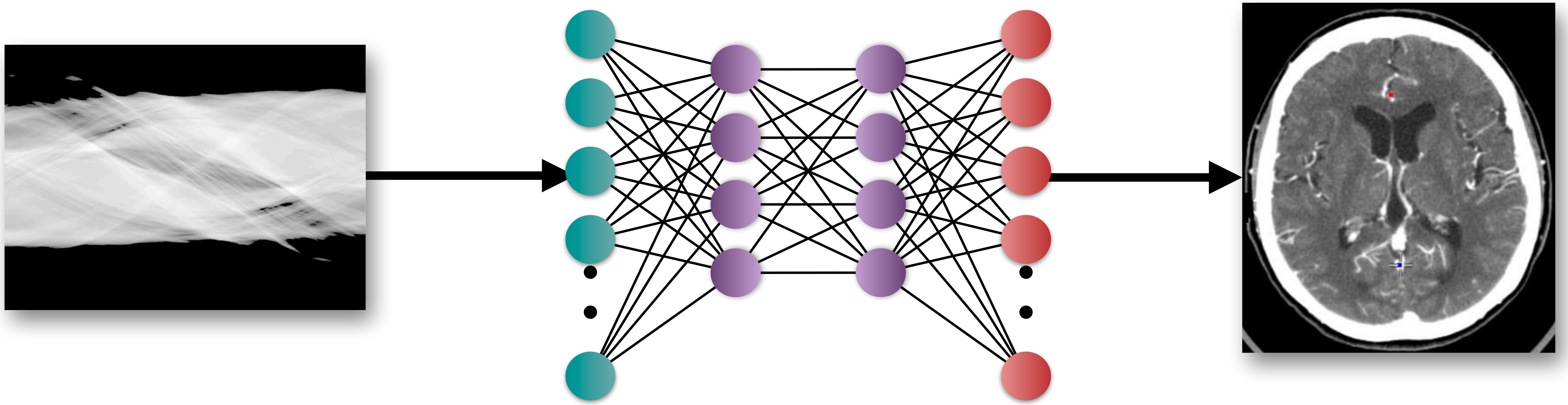
Video credit: Samuli Siltanen https://www.youtube.com/watch?v=q7Rt_OY_7tU



Video credit: Samuli Siltanen https://www.youtube.com/watch?v=q7Rt_OY_7tU

Can machine learning help reconstruct images?

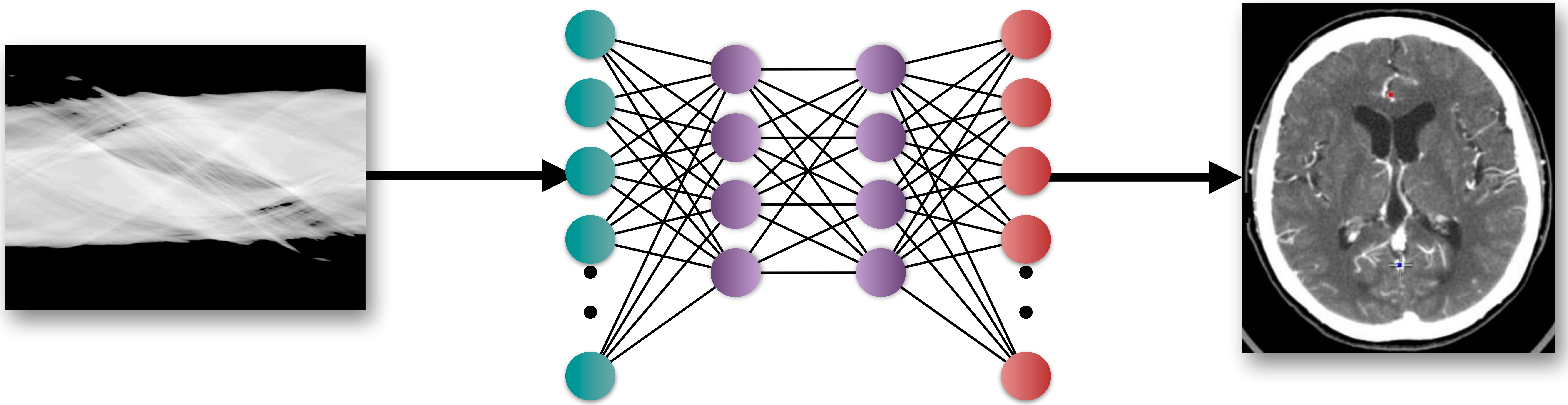
Train deep neural network to reconstruct CT images from sinogram measurements



Zhu, Liu, Rosen, Rosen, 2017; Arridge, Maass, Öktem, Schönlieb, 2019; Ongie, Jalal, Metzler, Baraniuk, Dimakis, Willett, 2020; Akçakaya, Yaman, Chung, Ye, 2022; Sahel, Bryan, Cleary, Farhi, Eldar, 2022; Kamilov, Bouman, Buzzard, Wohlberg, 2022

Can machine learning help reconstruct images?

Train deep neural network to reconstruct CT images from sinogram measurements



This approach can require ***many*** training samples.

It also ignores everything we know about the data collection process.

Zhu, Liu, Rosen, Rosen, 2017; Arridge, Maass, Öktem, Schönlieb, 2019; Ongie, Jalal, Metzler, Baraniuk, Dimakis, Willett, 2020; Akçakaya, Yaman, Chung, Ye, 2022; Sahel, Bryan, Cleary, Farhi, Eldar, 2022; Kamilov, Bouman, Buzzard, Wohlberg, 2022

Can we design neural networks to reflect our knowledge of the underlying physics?

Can we design neural networks to reflect our knowledge of the underlying physics?

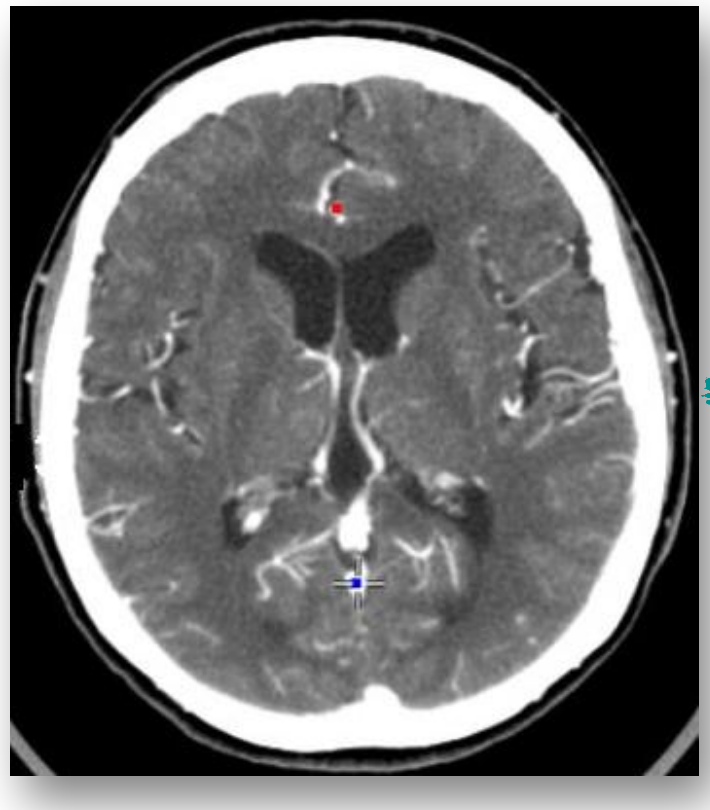
Yes! To do so, we leverage decades of accumulated knowledge of inverse problems, data assimilation, and optimization

Example: linear inverse problems in imaging

Observe: $y = Hx + \varepsilon$

Goal: Recover x from y

x



y

"forward model" H reflects the physics of the imaging system

Classical approach to solving inverse problems

$$\underset{x}{\text{minimize}} \; \|Hx - y\|^2 + R(x)$$

Data fit term
measures how well image x fits observation y, taking physical model H into account

Regularization function
measures to what extent an image x has expected geometry (e.g. smoothness or sharp edges)

Optimization framework

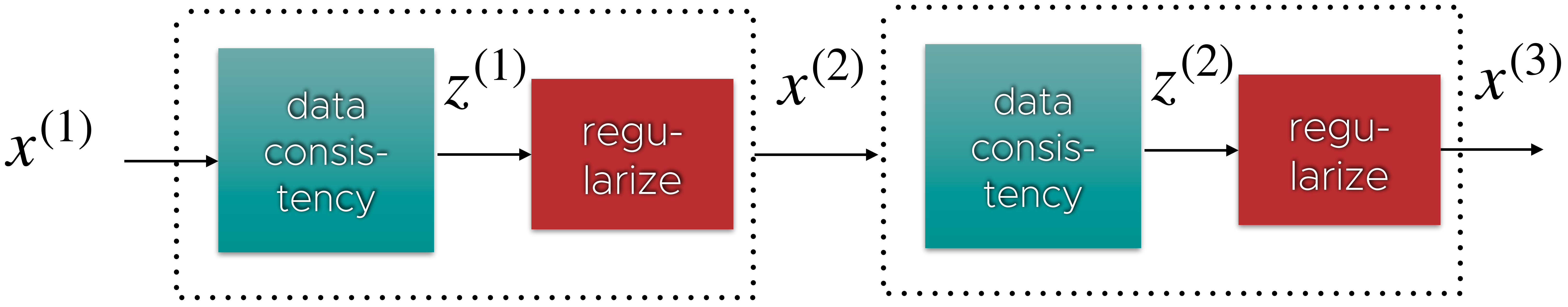
$$y \rightarrow \underset{x}{\text{minimize}} \; \|Hx - y\|^2 + R(x) \rightarrow x$$

data mismatch — irregularity

for $k = 1, 2, \ldots$

$$z^{(k)} = x^{(k)} - \eta H^\top (Hx^{(k)} - y)$$ **data consistency step**

$$x^{(k+1)} = \text{regularize}(z^{(k)}, R)$$ **regularization step (e.g. proximal operator)**



repeat until convergence

Deep Unrolling

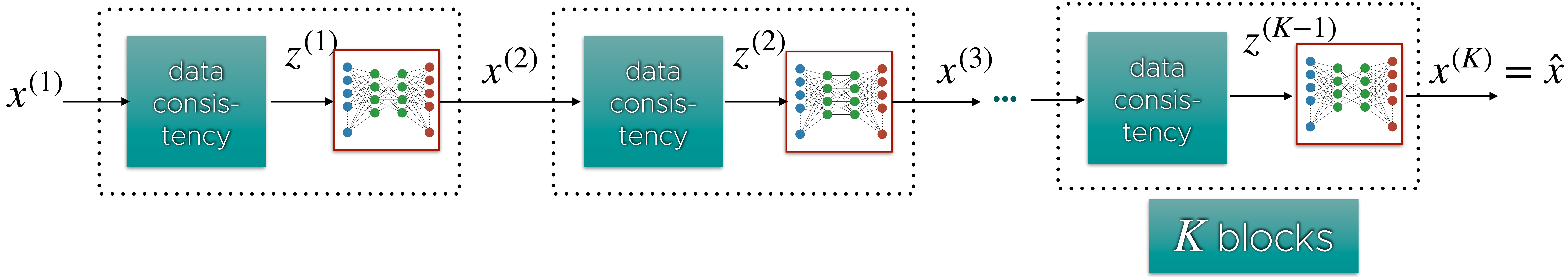
$y \longrightarrow$ $$\underset{x}{\text{minimize}} \; \|Hx - y\|^2 + R(x)$$ $\longrightarrow x$

data mismatch — irregularity

for $k = 1, 2, \ldots$

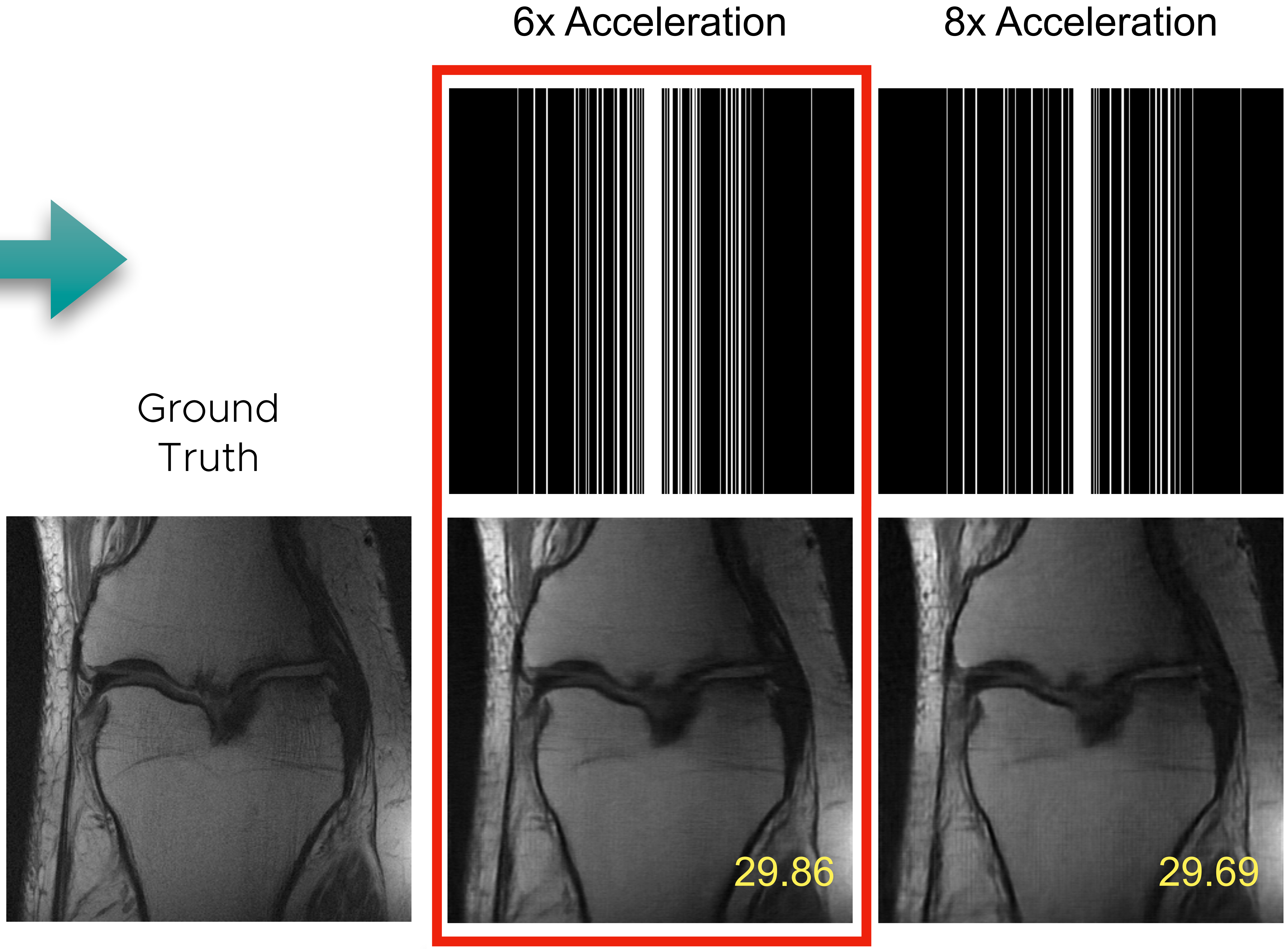
$$z^{(k)} = x^{(k)} - \eta H^\top (Hx^{(k)} - y)$$ **data consistency step**

$$x^{(k+1)} = \textbf{CNN}(z^{(k)})$$ **regularization step**



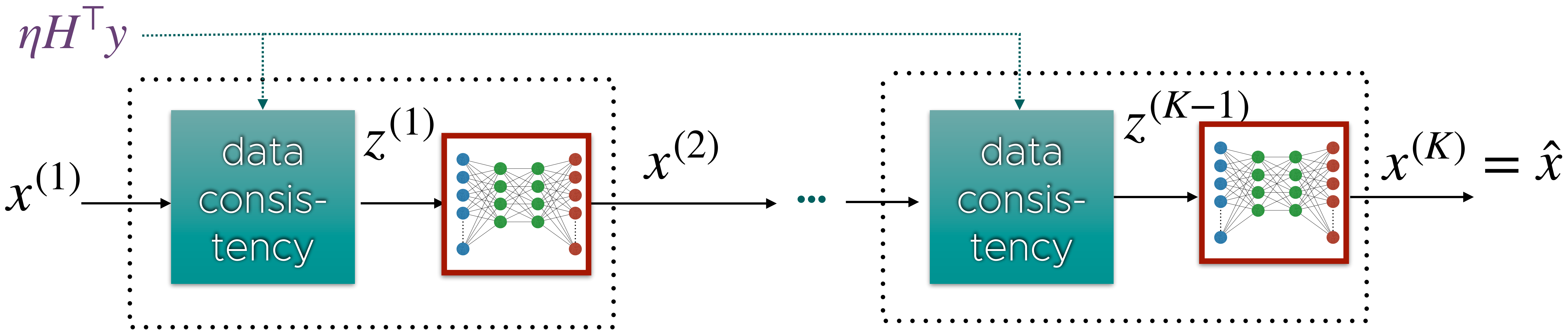
Enabling faster data acquisition and faster reconstruction

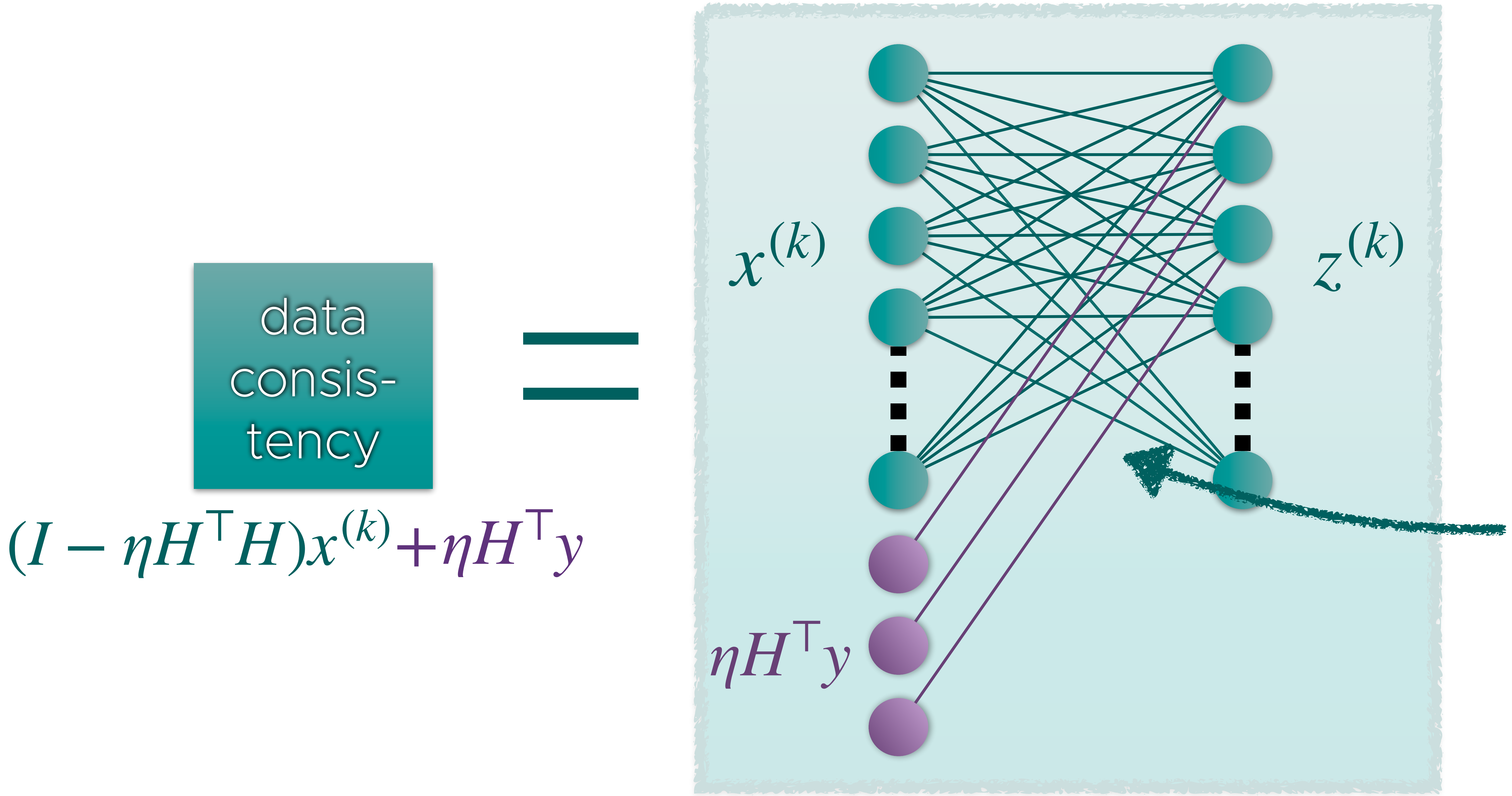
Deployment setting



Trained for 6x

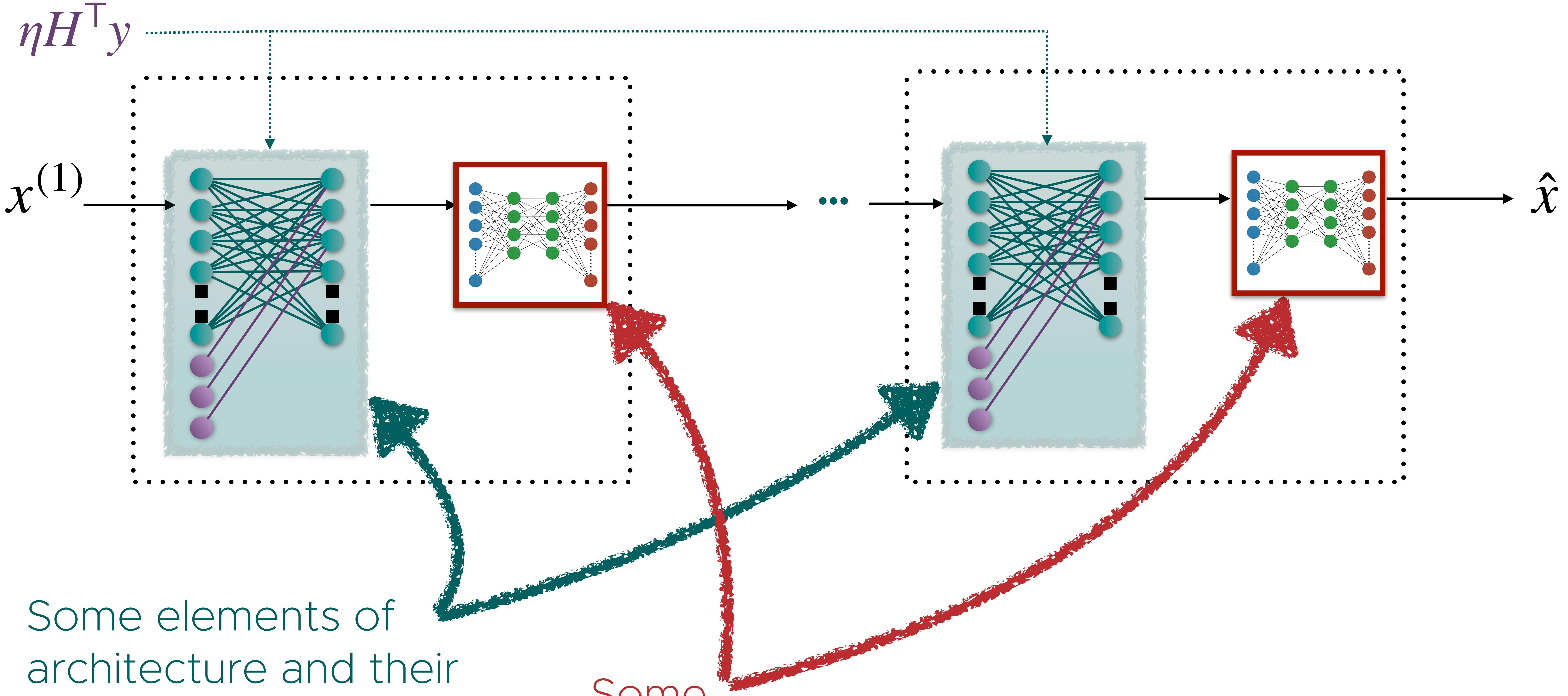
Physics-guided neural network architecture





weight matrix = $I - \eta H^\top H$ is determined by physical model instead of learned from data

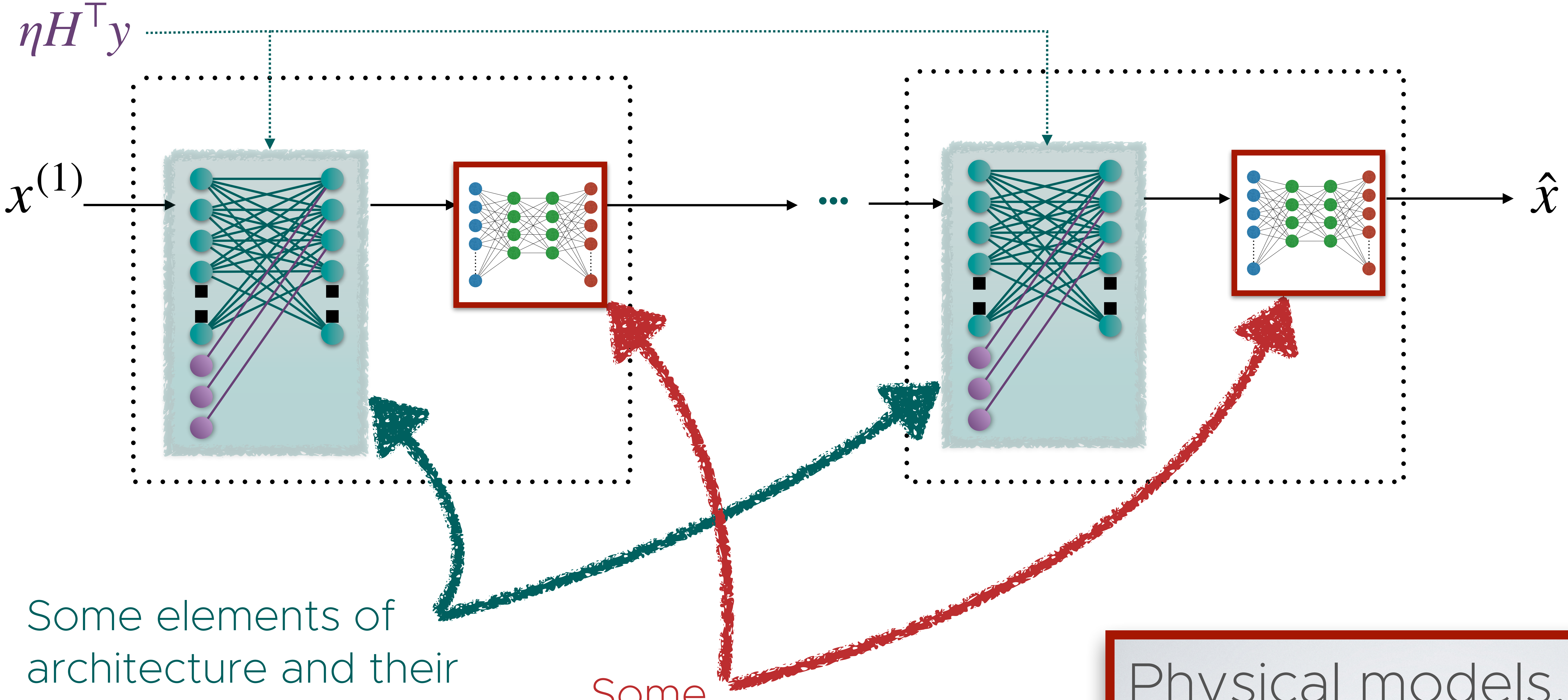
Physics-guided neural network architecture



Some elements of architecture and their weights are fixed, determined by choice of optimization method, forward model $\boldsymbol{H}$, and observed image data $\boldsymbol{y}$.

Some weights to be learned from training data

Physics-guided neural network architecture



Some elements of architecture and their weights are fixed, determined by choice of optimization method, forward model $\boldsymbol{H}$, and observed image data $\boldsymbol{y}$.

Some weights to be learned from training data

Physical models, inverse problem methods, and optimization theory lead to novel architectures

These advances depend on decades of NSF investment in foundational research

Machine learning foundations' impact

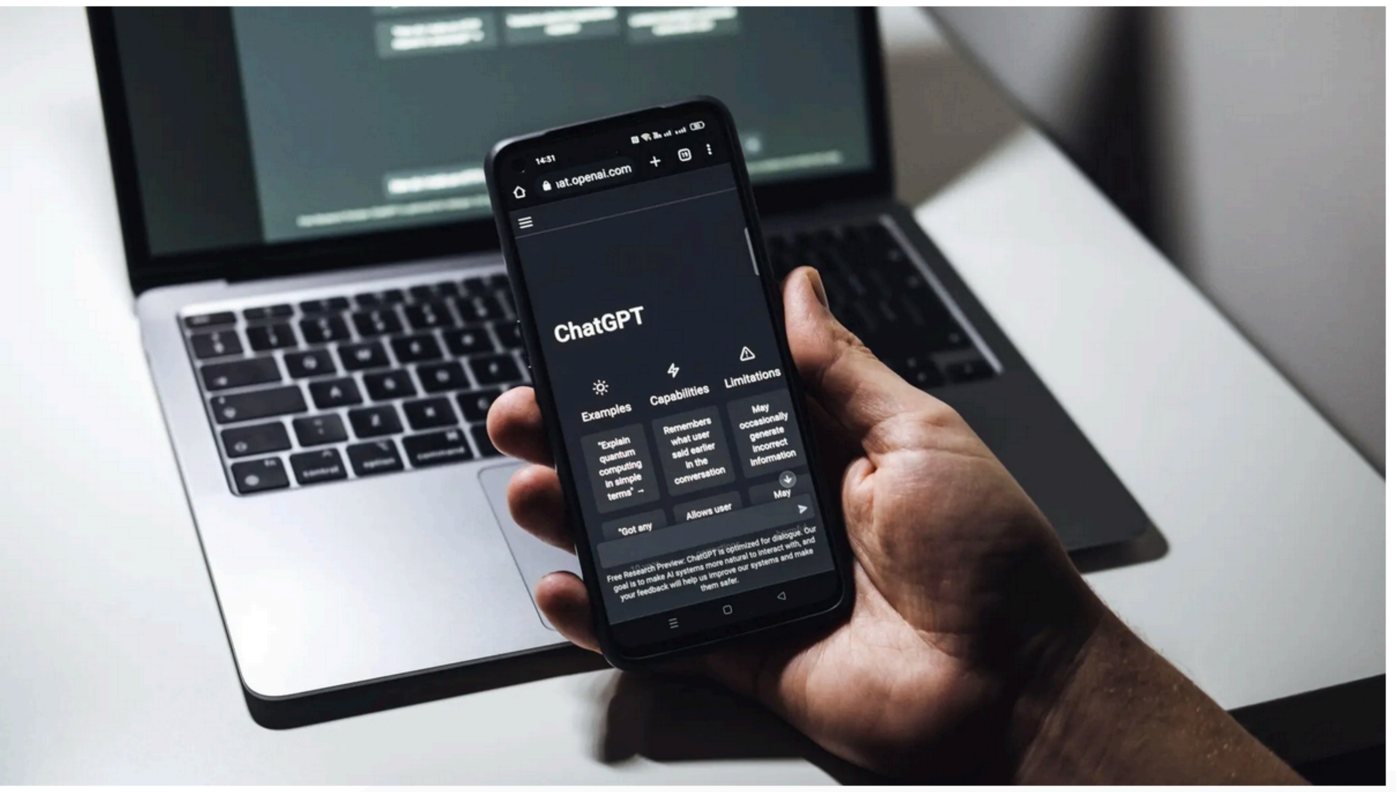
Emerging and future directions

The New York Times

Disinformation Researchers Raise Alarms About A.I. Chatbots

Researchers used ChatGPT to produce clean, convincing text that repeated conspiracy theories and misleading narratives.

Give this article 183



The Guardian

Women less likely to be shown ads for high-paid jobs on Google, study shows

Racial bias in a medical algorithm favors white patients over sicker black patients

By Carolyn Y. Johnson

October 24, 2019 at 2:00 p.m. EDT

Forbes

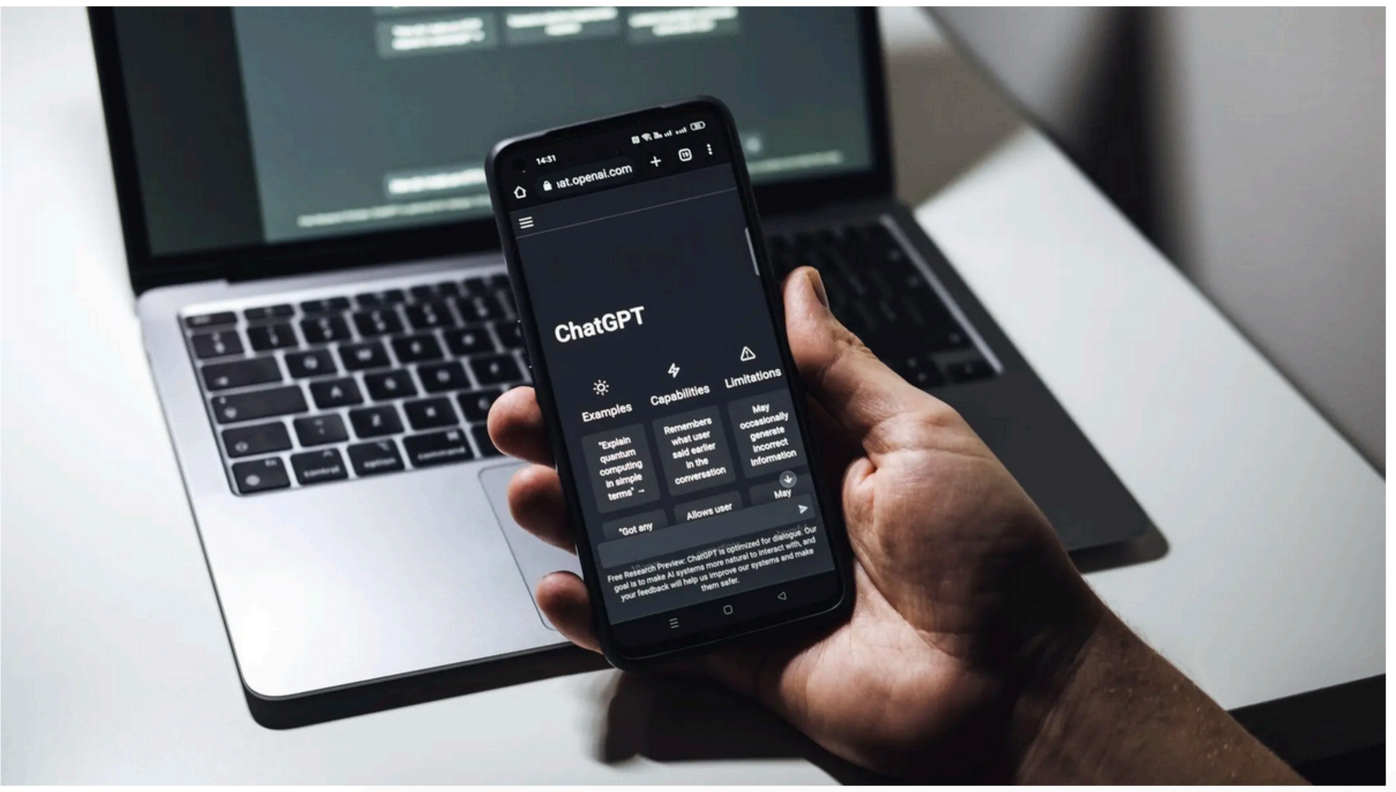
A.I. Bias Caused 80% Of Black Mortgage Applicants To Be Denied

The New York Times

Disinformation Researchers Raise Alarms About A.I. Chatbots

Researchers used ChatGPT to produce clean, convincing text that repeated conspiracy theories and misleading narratives.

Give this article 183



Forbes

A.I. Bias Caused 80% Of Black Mortgage Applicants To Be Denied

How do we design regulations and certification of ML systems?

The Guardian

Women less likely to be shown ads for high-paid jobs on Google, study shows

Racial bias in a medical algorithm favors white patients over sicker black patients

By Carolyn Y. Johnson

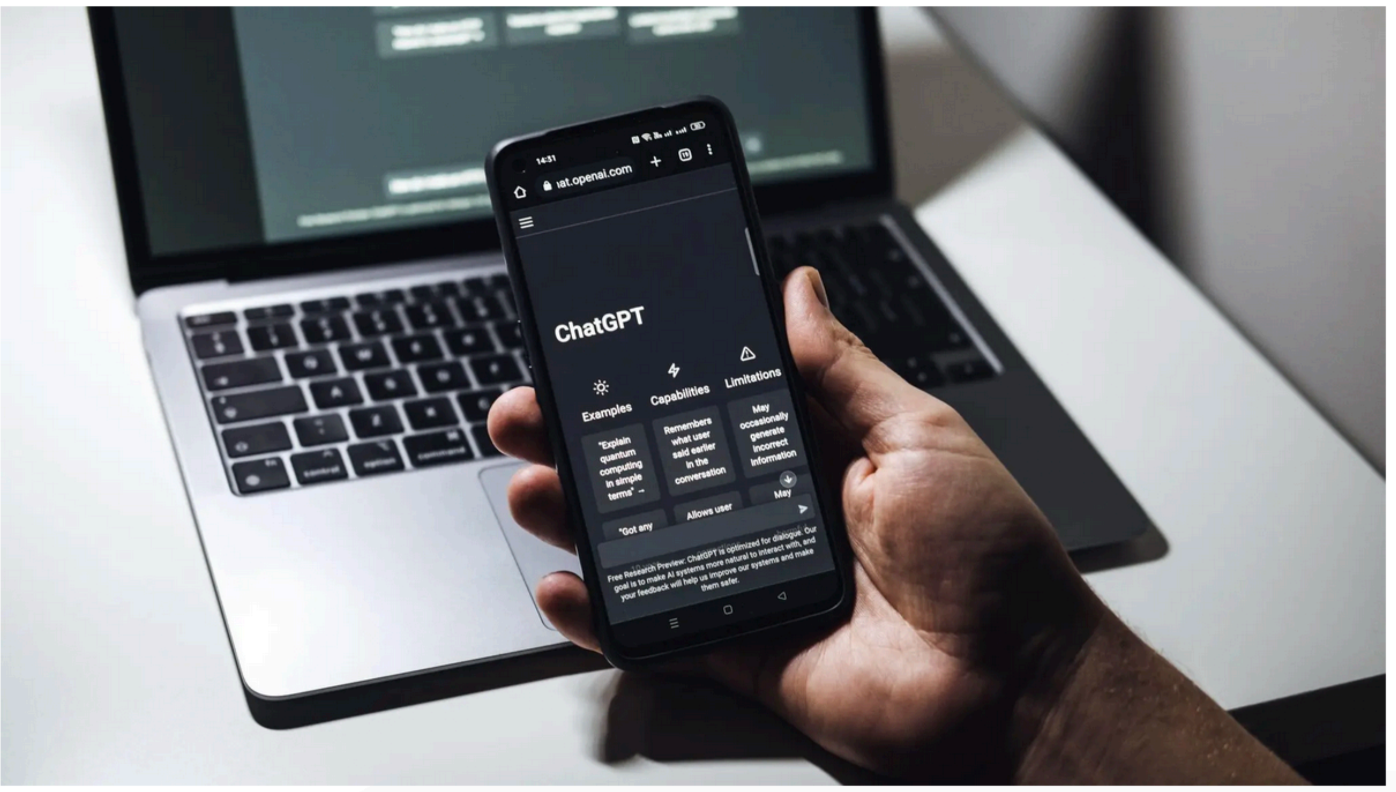
October 24, 2019 at 2:00 p.m. EDT

The New York Times

Disinformation Researchers Raise Alarms About A.I. Chatbots

Researchers used ChatGPT to produce clean, convincing text that repeated conspiracy theories and misleading narratives.

Give this article 183



The Guardian

Women less likely to be shown ads for high-paid jobs on Google, study shows

How do we design regulations and certification of ML systems?

Racial bias in a medical algorithm favors white patients over sicker black patients

By Carolyn Y. Johnson

October 24, 2019 at 2:00 p.m. EDT

Forbes

A.I. Bias Caused 80% Of Black Mortgage Applicants To Be Denied

Efficiency

Forbes

Deep Learning's Carbon Emissions Problem

Rob Toews Contributor

The bottom line: AI has a meaningful carbon footprint today, and if industry trends continue it will soon become much worse. Unless we are willing to reassess and reform today's AI research agenda, the field of artificial intelligence could become an antagonist in the fight against climate change in the years ahead.

The Washington Post

A new front in the water wars: Your internet use

In the American West, data centers are clashing with local communities that want to preserve water amid drought

By Shannon Osaka

April 25, 2023 at 6:30 a.m. EDT

A Google data center in The Dalles, Ore., seen in October 2021. (Andrew Selsky/AP)

Efficiency

Forbes

Deep Learning's Carbon Emissions Problem

Rob Toews Contributor

The bottom line: AI has a meaningful carbon footprint today, and if industry trends continue it will soon become much worse. Unless we are willing to reassess and reform today's AI research agenda, the field of artificial intelligence could become an antagonist in the fight against climate change in the years ahead.

Math, stats, & CS foundations help us optimize architectures and training efficiency

The Washington Post

A new front in the water wars: Your internet use

In the American West, data centers are clashing with local communities that want to preserve water amid drought

By Shannon Osaka

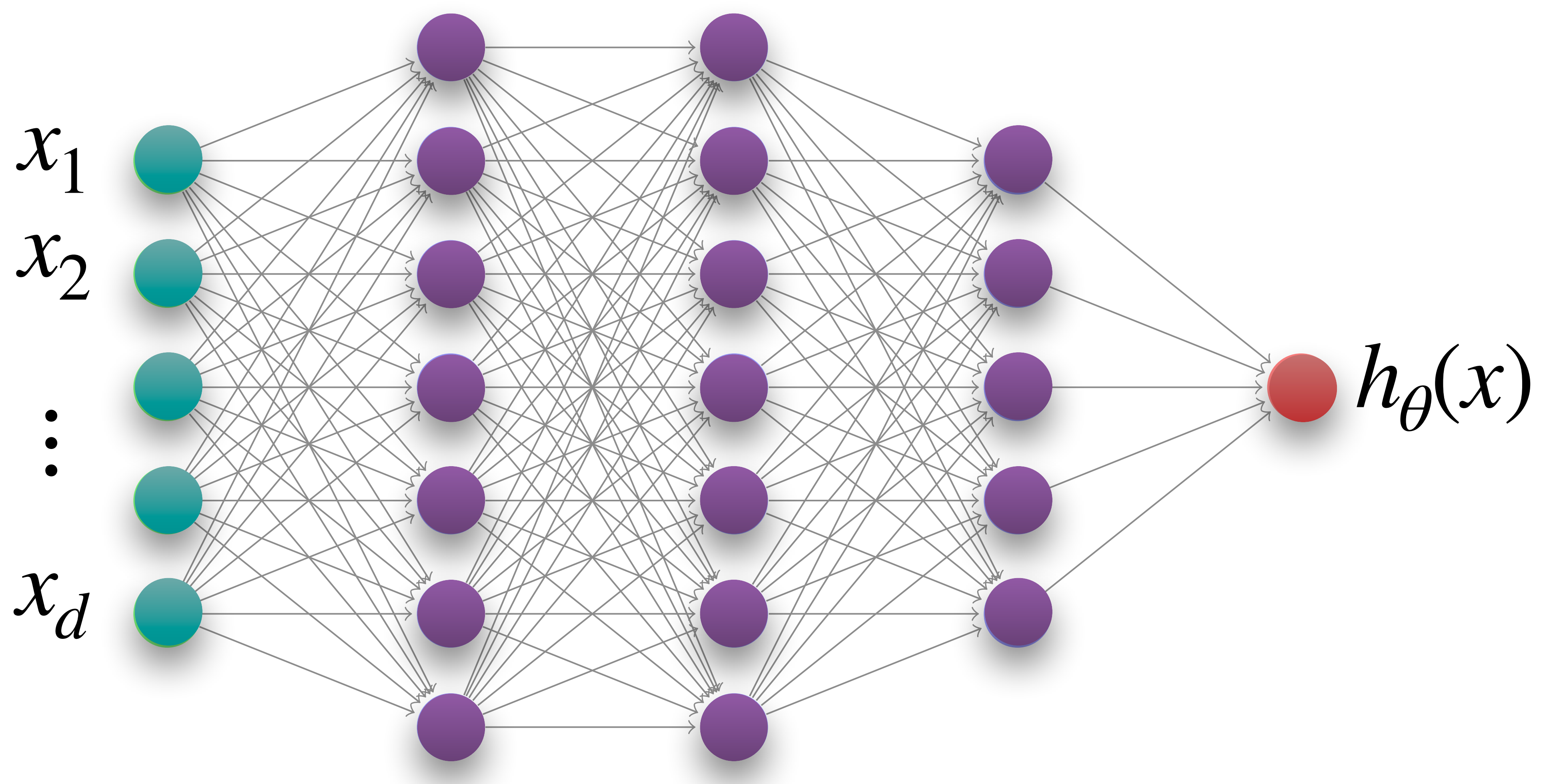
April 25, 2023 at 6:30 a.m. EDT

A Google data center in The Dalles, Ore., seen in October 2021. (Andrew Selsky/AP)

- How much data do we need?
- How can we promote robustness?
- Will models work in new settings?
- Can we make machine learning more sustainable?
- Do transformers offer special advantages?
- How do we design next-gen architectures?

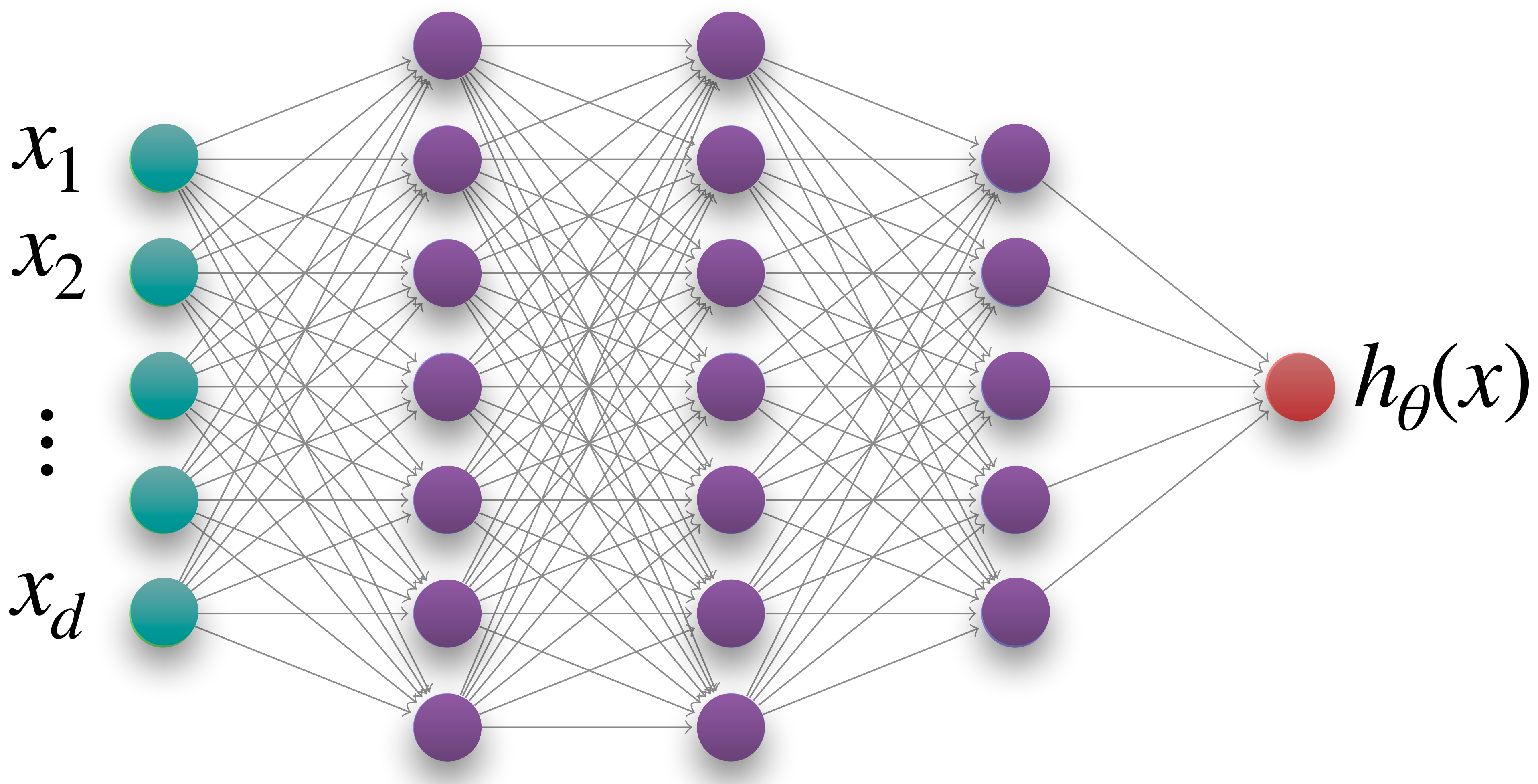
Neural networks are functions

- How much data do we need?
- How can we promote robustness?
- Will models work in new settings?
- Can we make machine learning more sustainable?
- Do transformers offer special advantages?
- How do we design next-gen architectures?



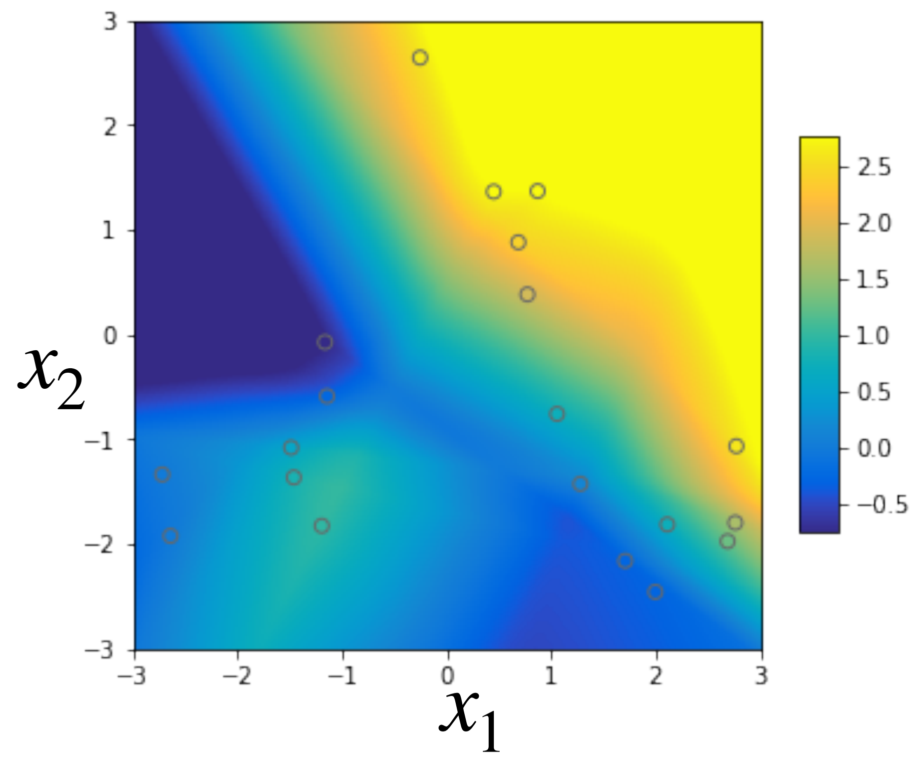
Network inputs vector x and outputs a prediction $\hat{y} = h_\theta(x)$ that depends on learned weights θ

Neural networks are functions

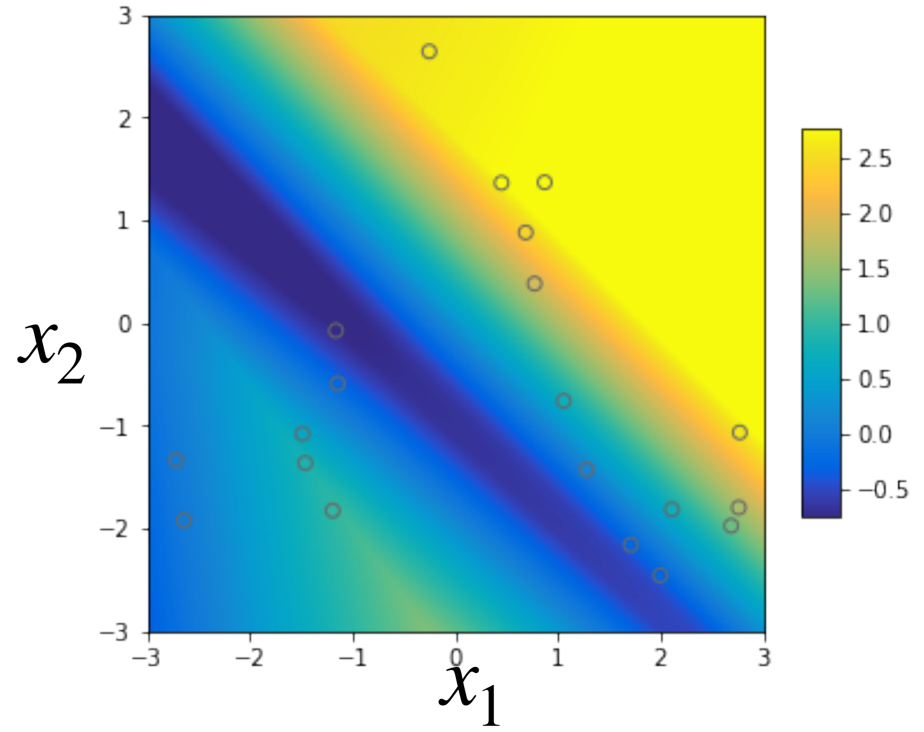


Network inputs vector x and outputs a prediction $\hat{y} = h_\theta(x)$ that depends on learned weights θ

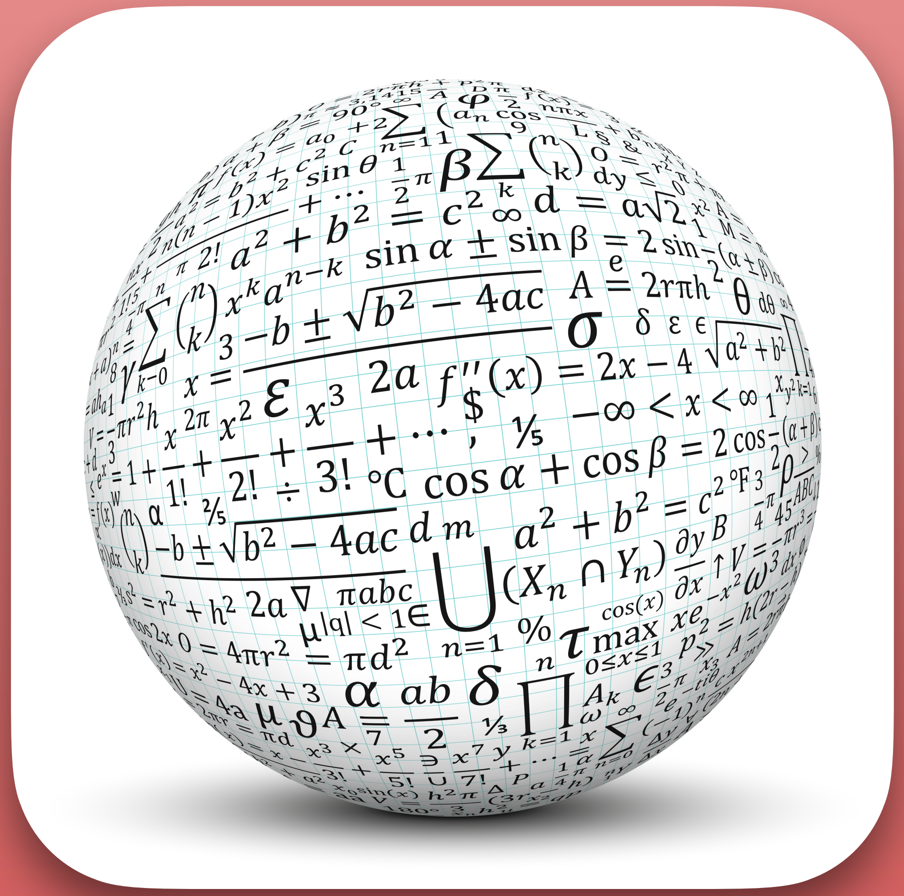
These are both functions; for every 2-d input $x = [x_1, x_2]$, the color shows what the output value $\hat{y} = h_\theta(x)$ would be.



Both functions exactly fit the same training data but with different weights θ. What determines which function will be selected when we train the neural network?



ML will fundamentally change the nature and pace of scientific discovery, influencing data analysis, hypothesis generation, simulation, and experimental design

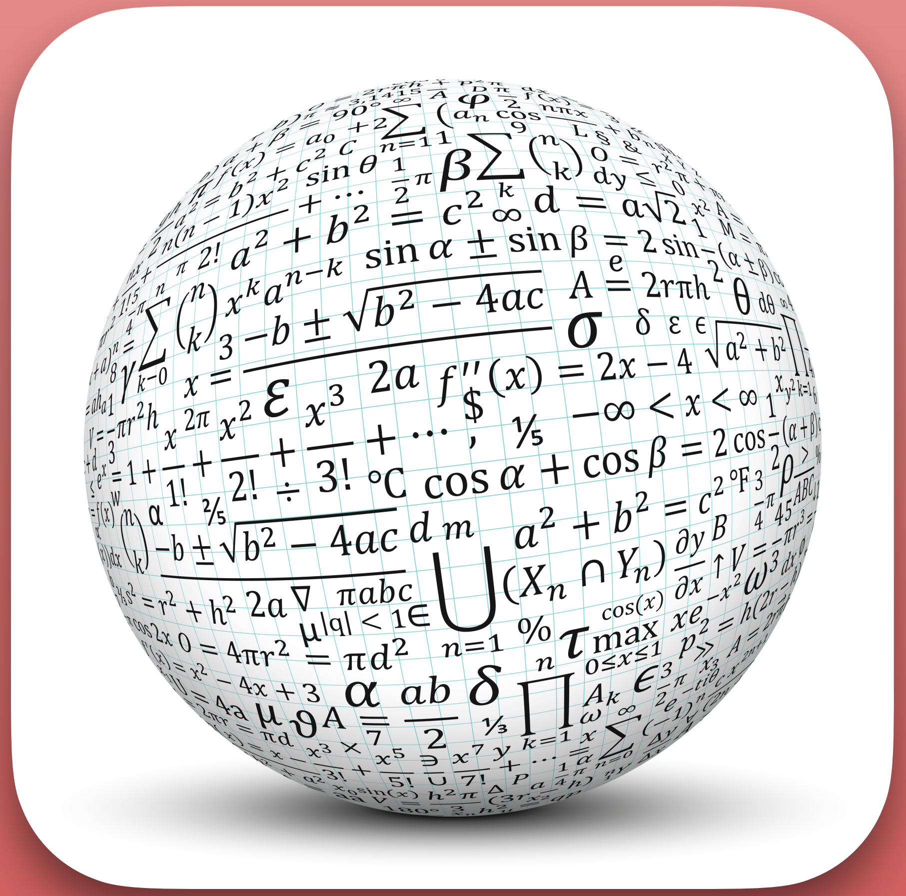


Uncovering new laws of nature

AI-guided scientific measurement

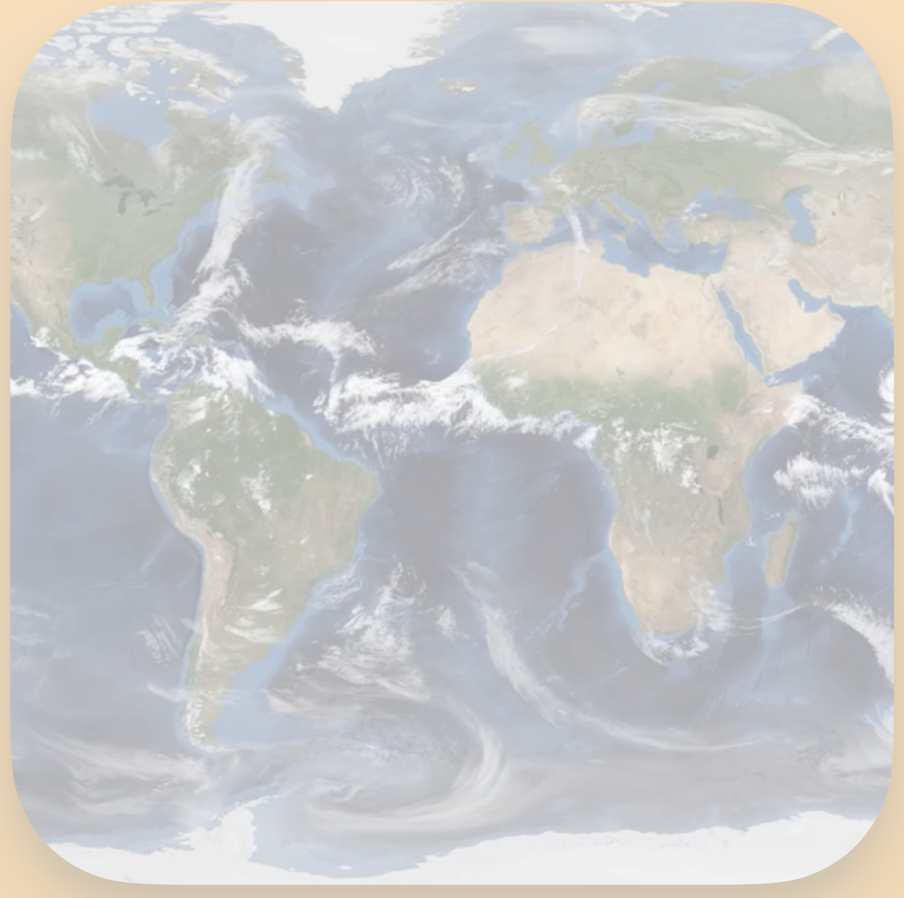
Physics-informed machine learning

Advancing ML frontiers

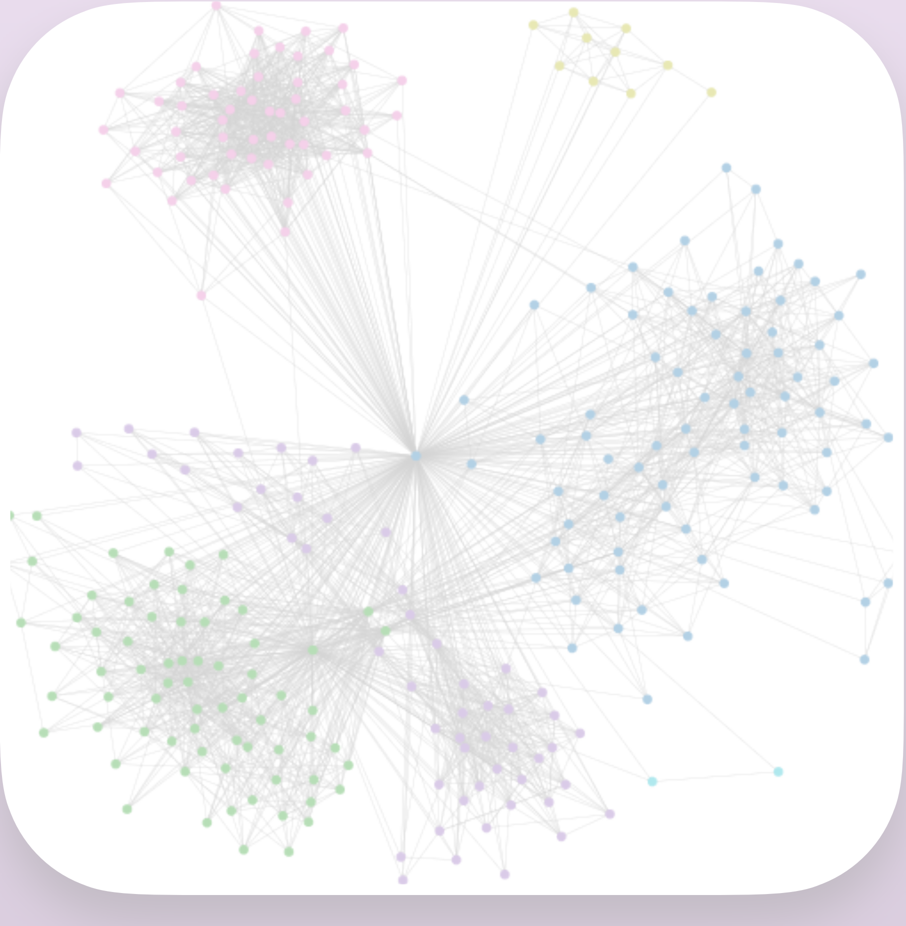


Uncovering new laws of nature

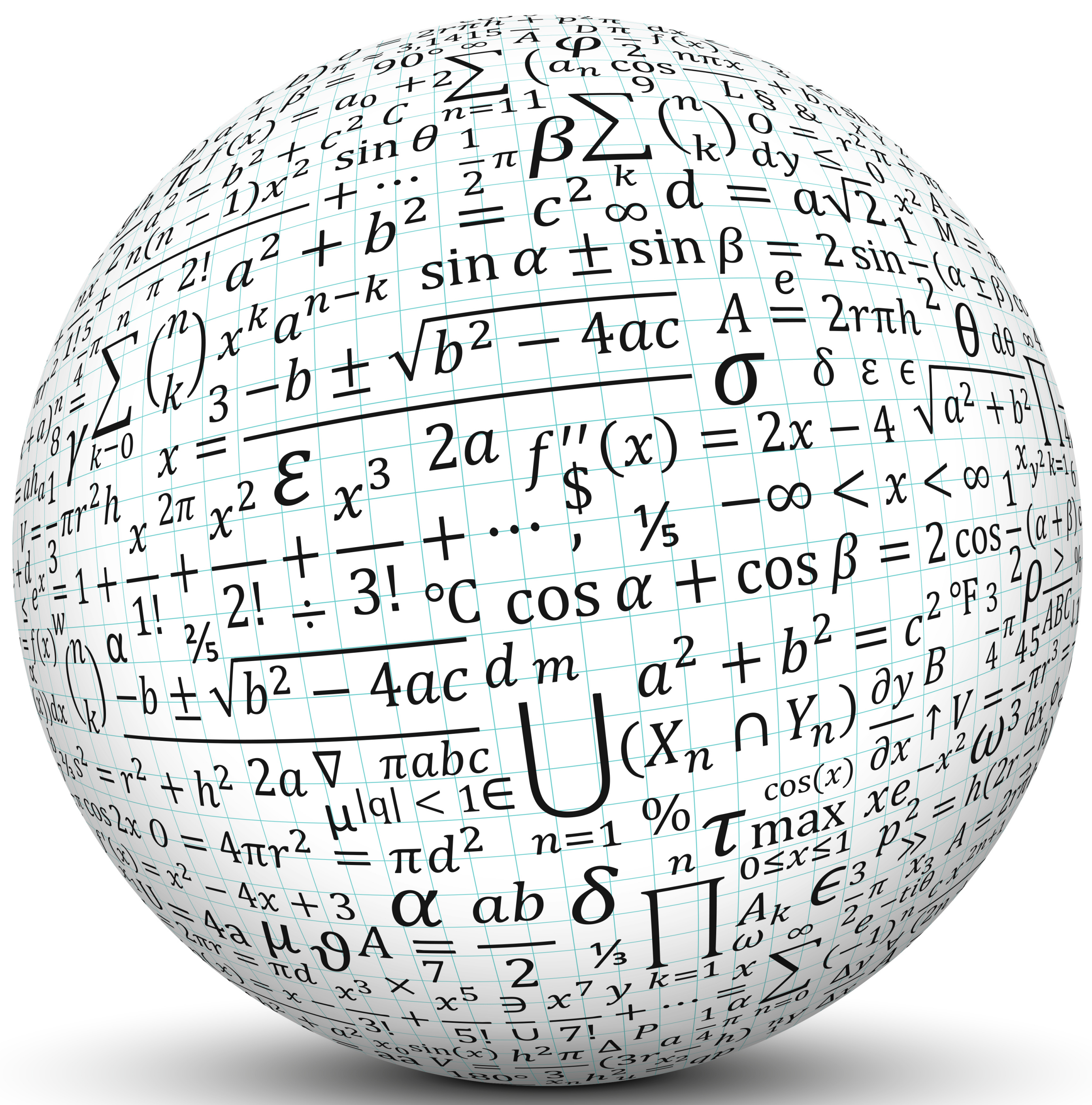
AI-guided scientific measurement



Physics-informed machine learning



Advancing ML frontiers



Given observations of a system, use AI to uncover the governing physical laws

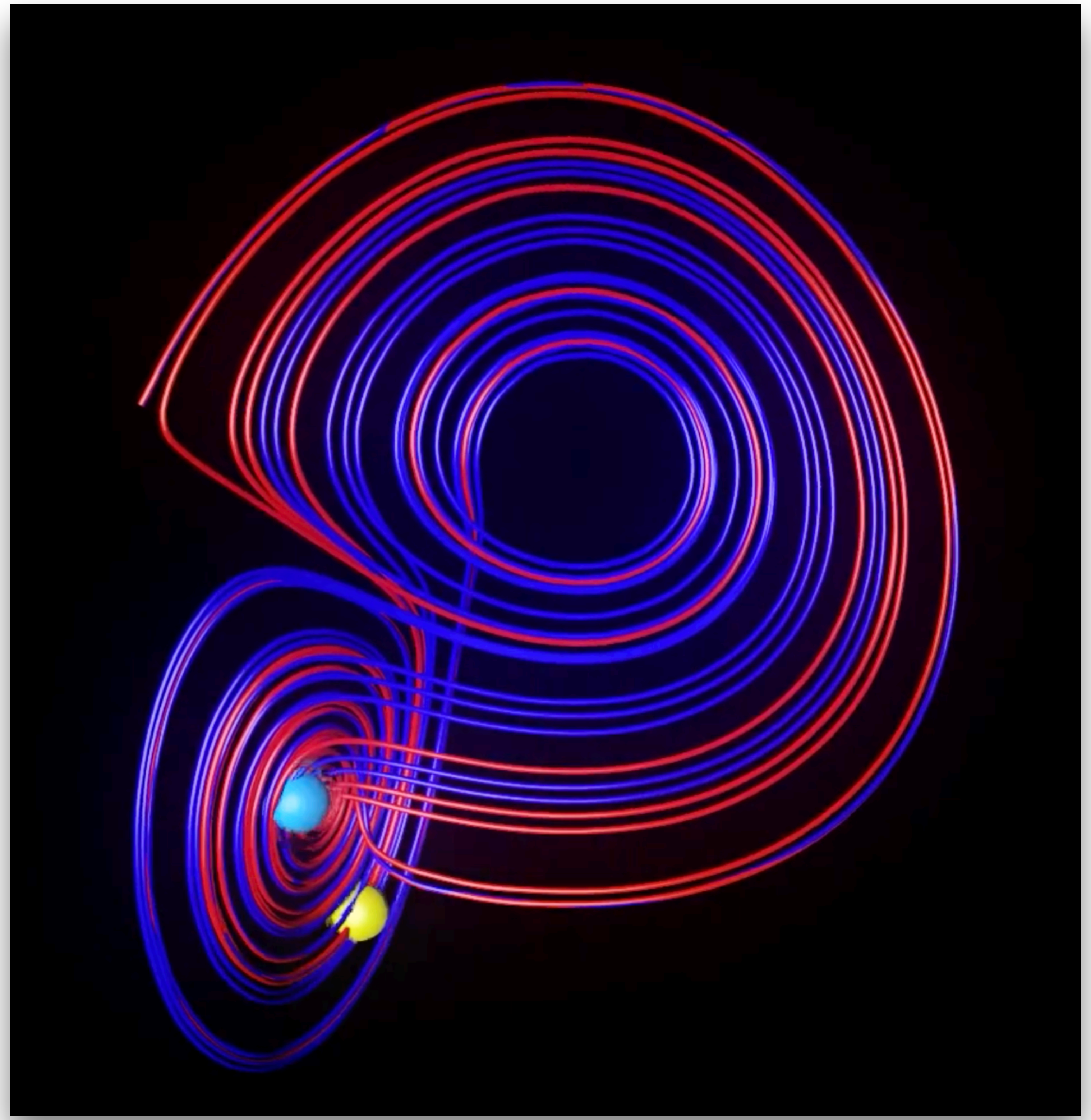
$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\rho - z) - y$$

$$\frac{dz}{dt} = xy - \beta z$$

https://www.youtube.com/watch?v=q3kNHomvU0k

Can we learn governing equations from data?

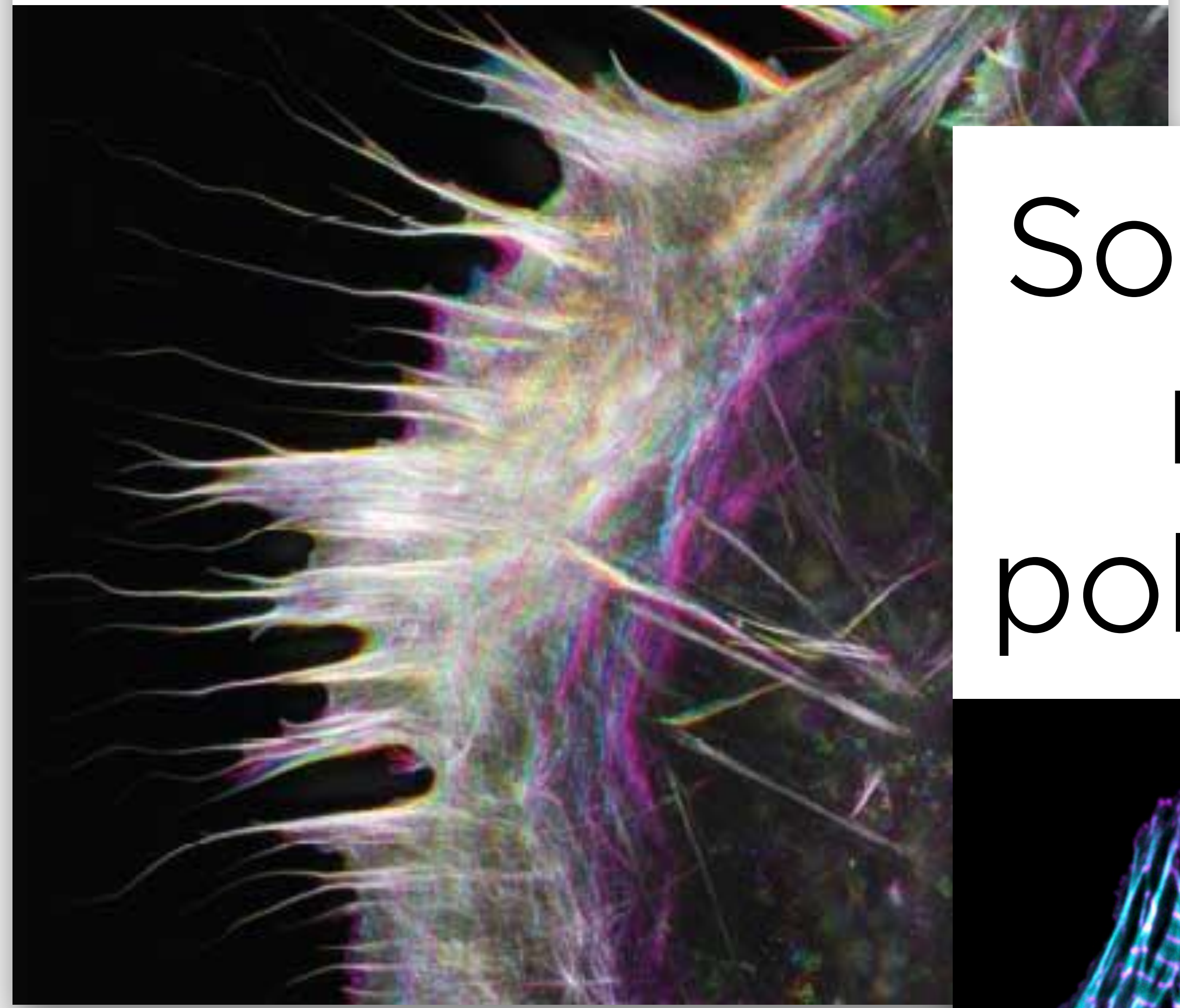


$$\frac{dx}{dt} = \sigma(y - x)$$

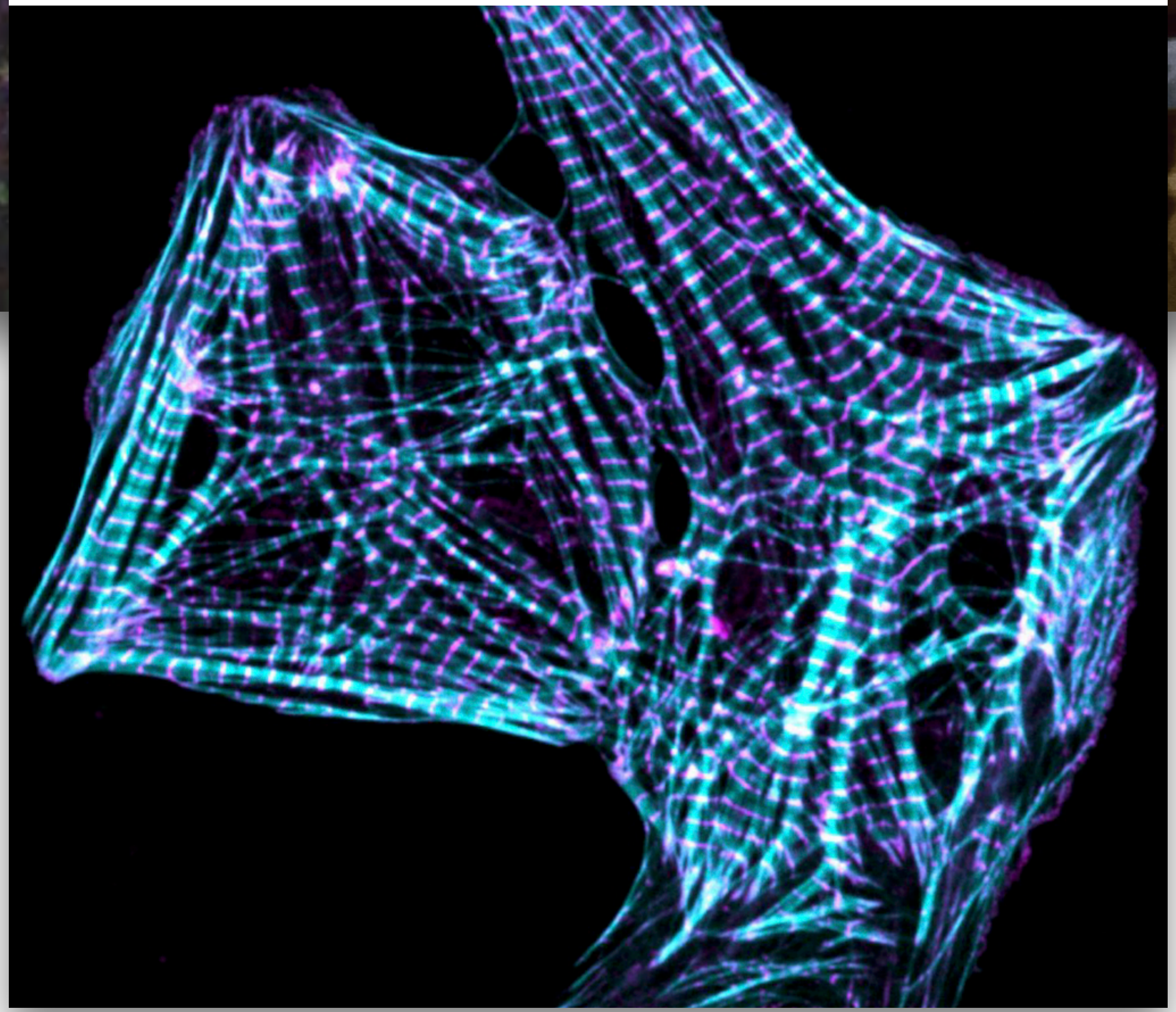
$$\frac{dy}{dt} = x(\rho - z) - y$$

$$\frac{dz}{dt} = xy - \beta z$$

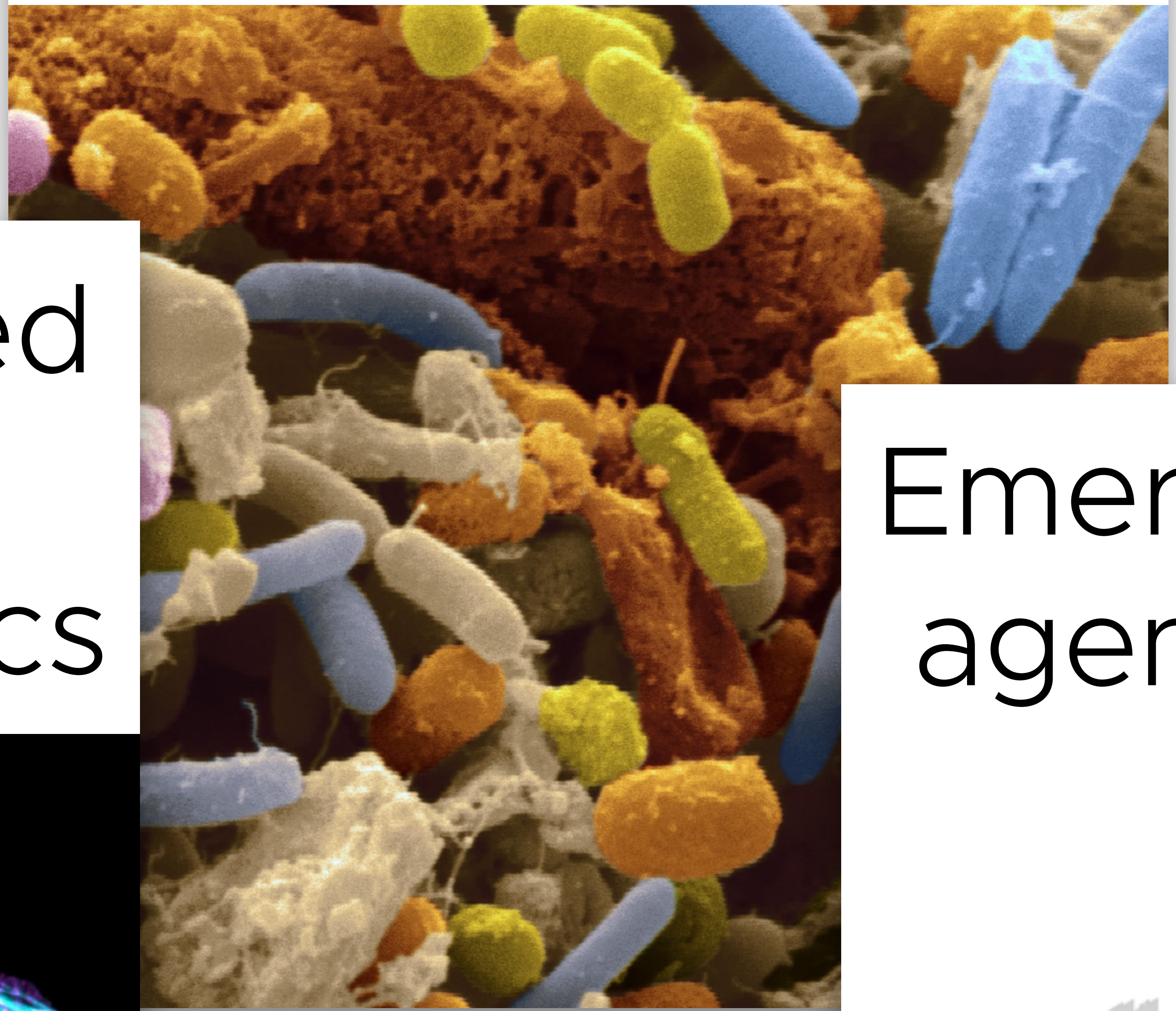
Biophysical forces of cell development and function



Soft condensed matter and polymer physics



Dynamics of microbial communities



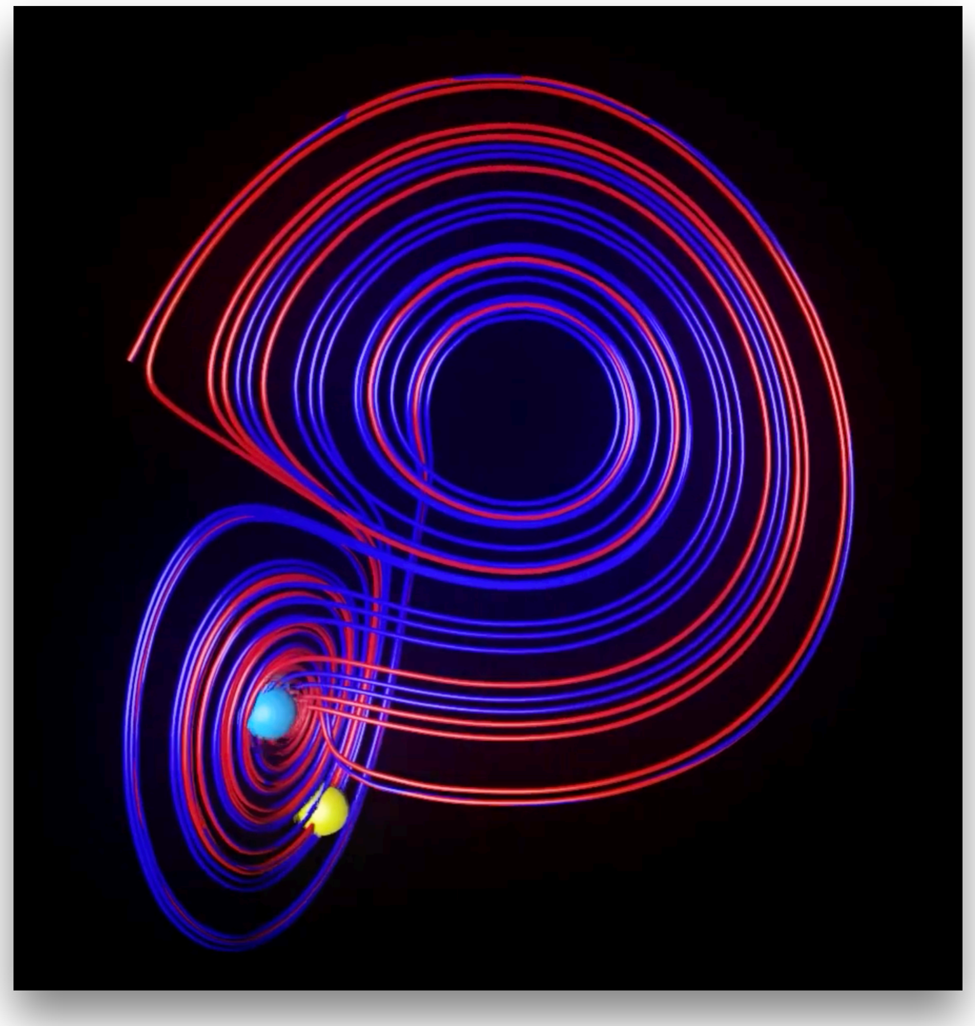
Emergent behavior of agent-based models

Sparse Identification of Nonlinear Dynamics (SINDy)

$$\frac{dx}{dt} = w_0 + w_1 x + w_2 y + w_3 z + w_4 x^2 + w_5 xy + w_6 xz + w_7 y^2 + \cdots + w_* z^5$$

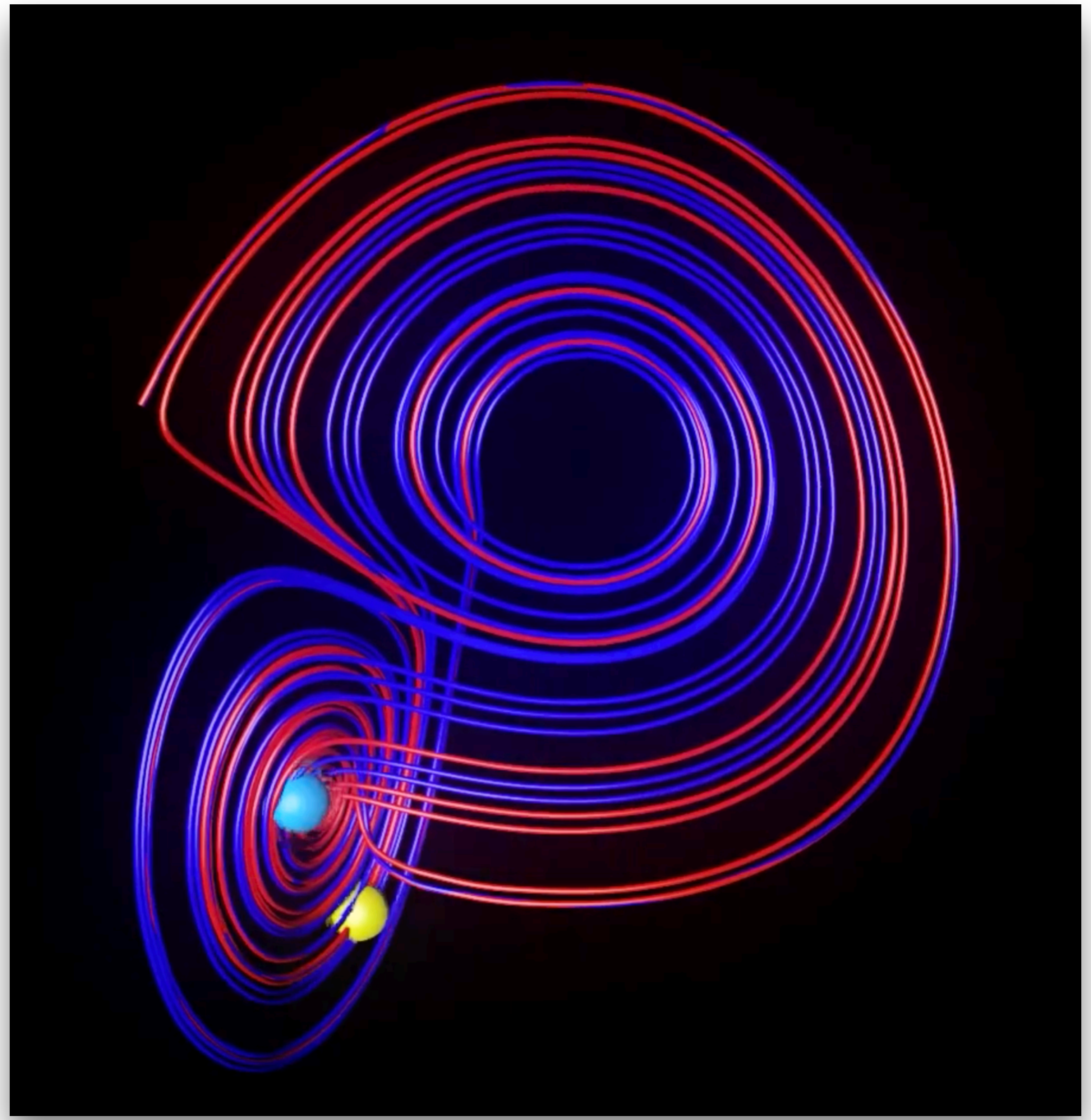
Learn weights from data. Only two are non-zero: $w_2 = -w_1 = \sigma$

Repeat for $\frac{dy}{dt}$ and $\frac{dz}{dt}$ to recover full dynamics



Brunton, Proctor, & Kutz 2016

Can we learn governing equations from data?

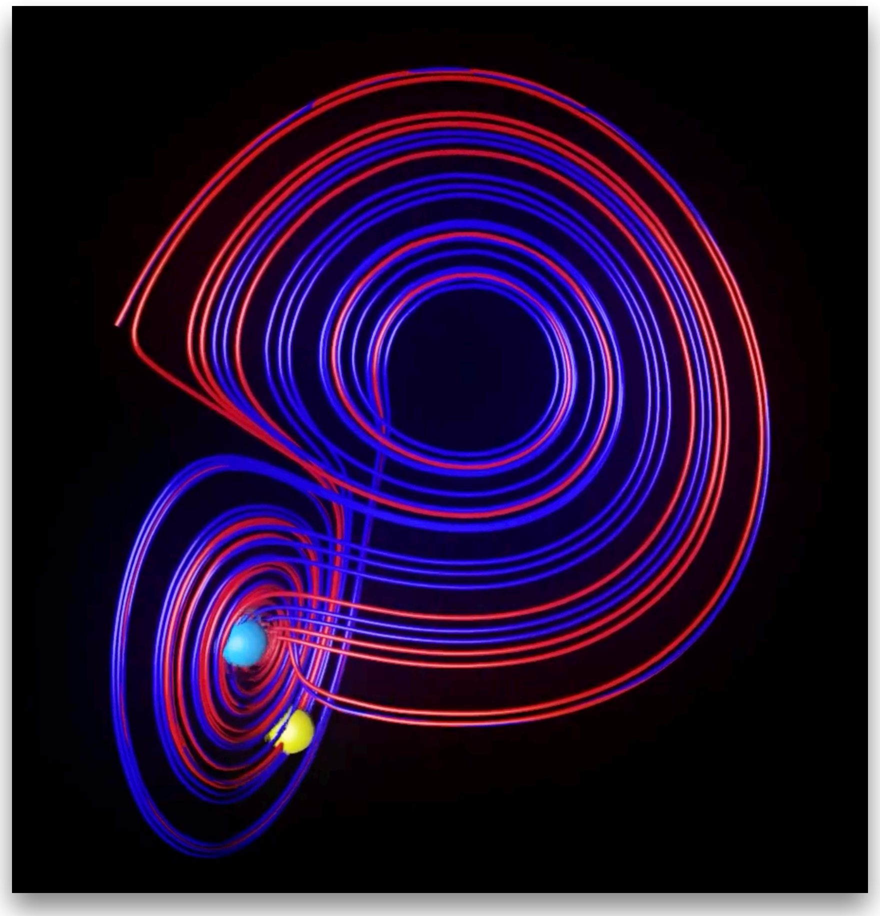


$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\rho - z) - y$$

$$\frac{dz}{dt} = xy - \beta z$$

Can we learn governing equations from data?



$$\frac{dx}{dt} = \sigma(y - x)$$
$$\frac{dy}{dt} = x(\rho - z) - y$$
$$\frac{dz}{dt} = xy - \beta z$$

arXiv > astro-ph > arXiv:2202.02306

Search...

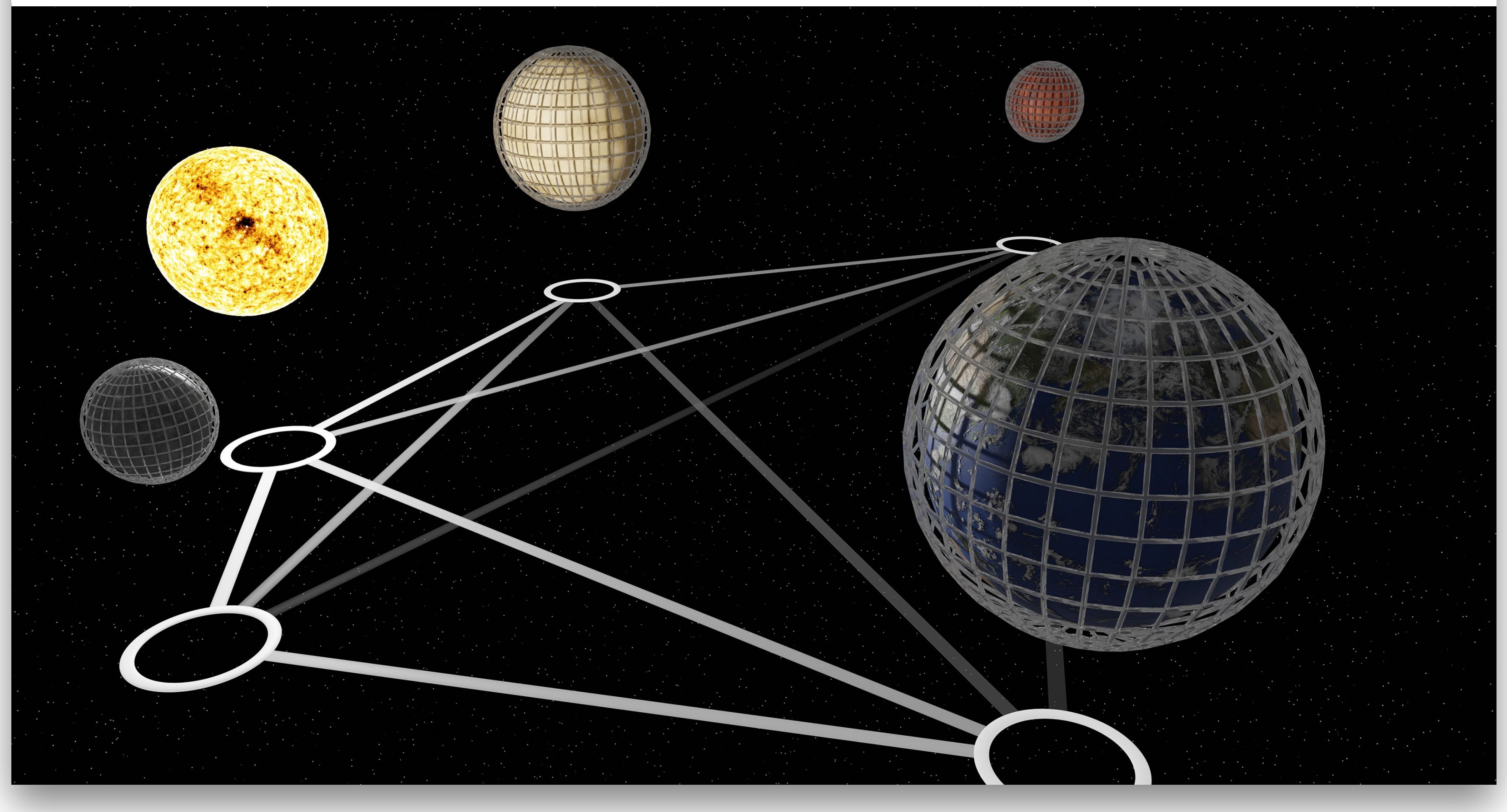
Help | Advance

Astrophysics > Earth and Planetary Astrophysics

[Submitted on 4 Feb 2022]

Rediscovering orbital mechanics with machine learning

Pablo Lemos, Niall Jeffrey, Miles Cranmer, Shirley Ho, Peter Battaglia



PNAS

Discovering governing equations from data by sparse identification of nonlinear dynamical systems

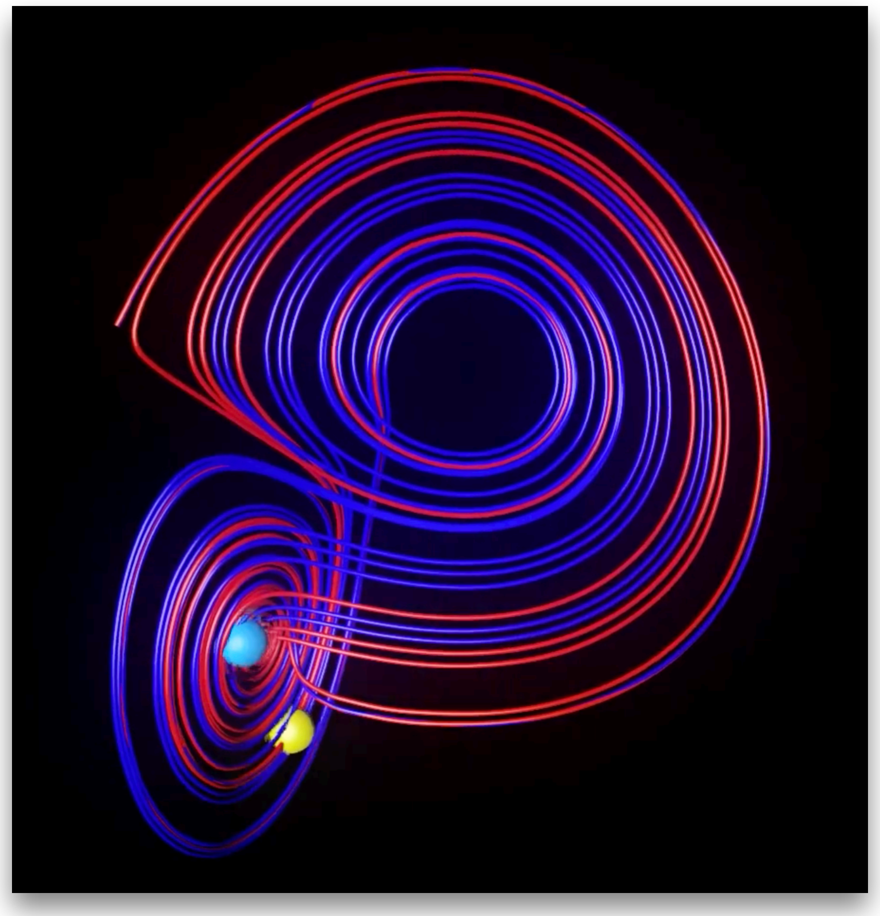
Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz Authors Info & Affiliations

ScienceAdvances

AI Feynman: A physics-inspired method for symbolic regression

SILVIU-MARIAN UDRESCU AND MAX TEGMARK Authors Info & Affiliations

Can we learn governing equations from data?



$$\frac{dx}{dt} = \sigma(y - x)$$
$$\frac{dy}{dt} = x(\rho - z) - y$$
$$\frac{dz}{dt} = xy - \beta z$$

arXiv > astro-ph > arXiv:2202.02306

Search...

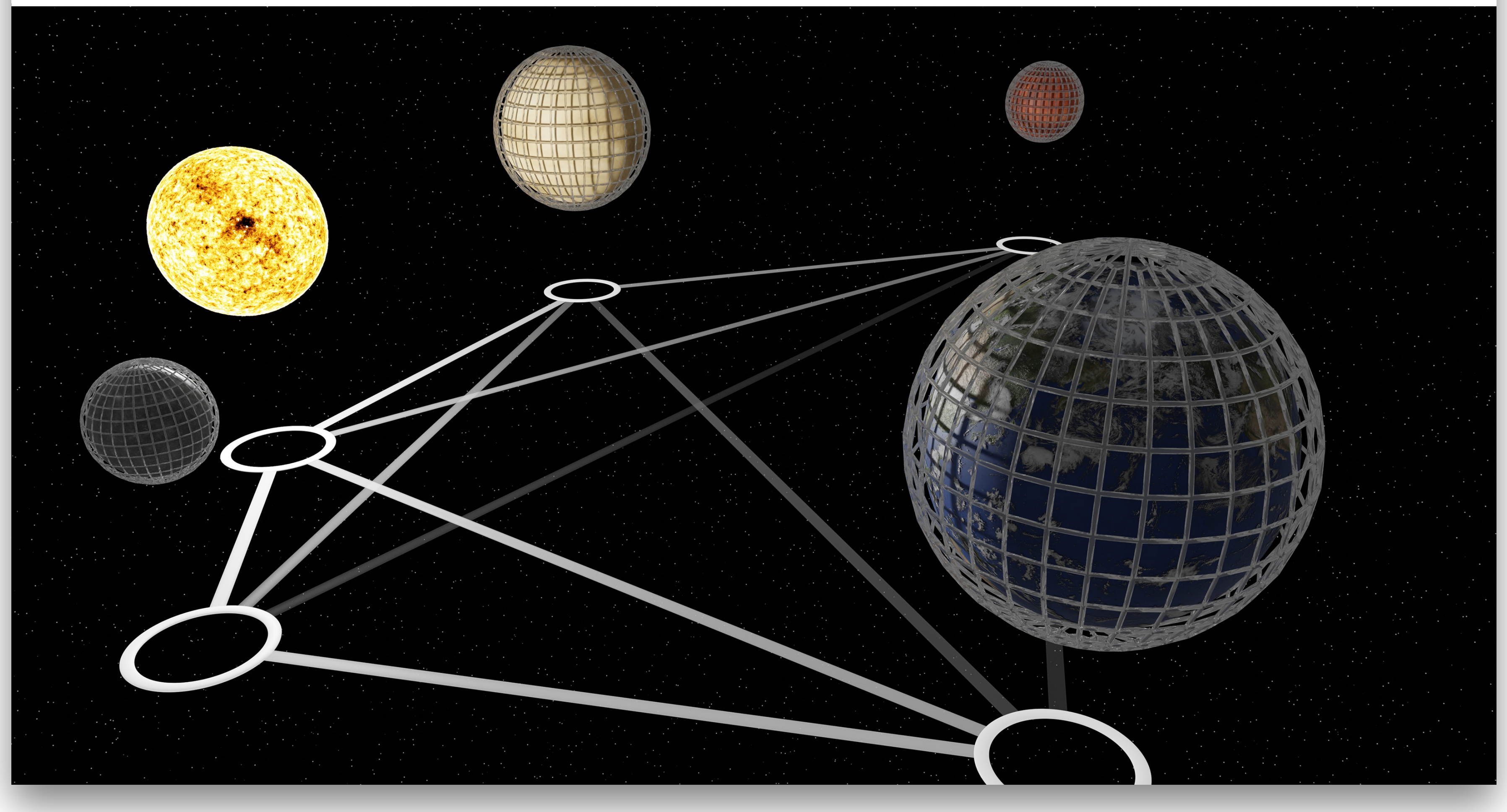
Help | Advanc

Astrophysics > Earth and Planetary Astrophysics

[Submitted on 4 Feb 2022]

Rediscovering orbital mechanics with machine learning

Pablo Lemos, Niall Jeffrey, Miles Cranmer, Shirley Ho, Peter Battaglia



PNAS

Discovering governing equations from data by sparse identification of nonlinear dynamical systems

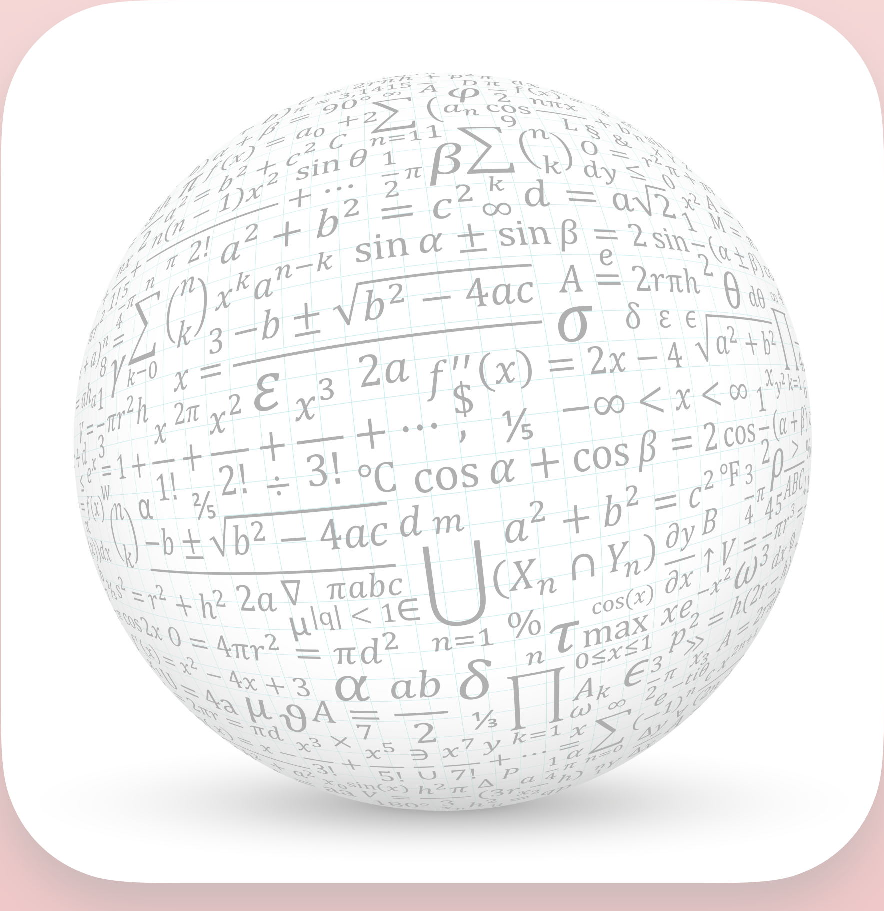
Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz Authors Info & Affiliations

Science Advances

AI Feynman: A physics-inspired method for symbolic regression

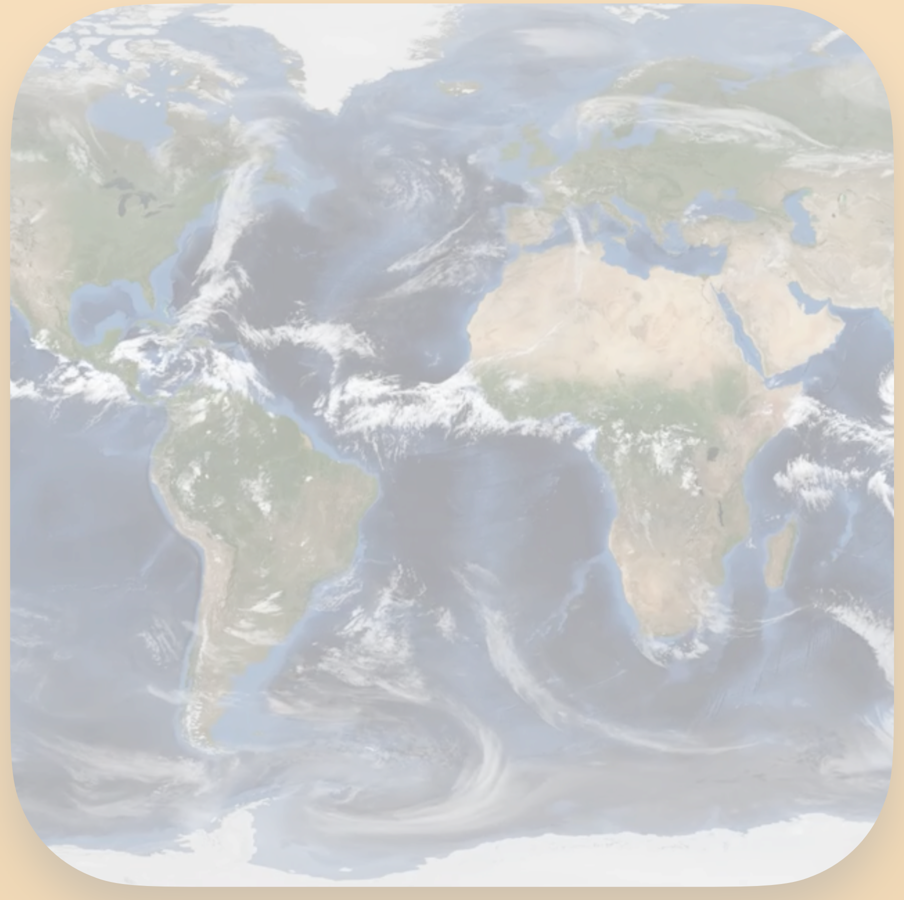
SILVIU-MARIAN UDRESCU AND MAX TEGMARK Authors Info & Affiliations

Equation discovery with high-dimensions, sparse and noisy data, etc., poses significant foundational challenges

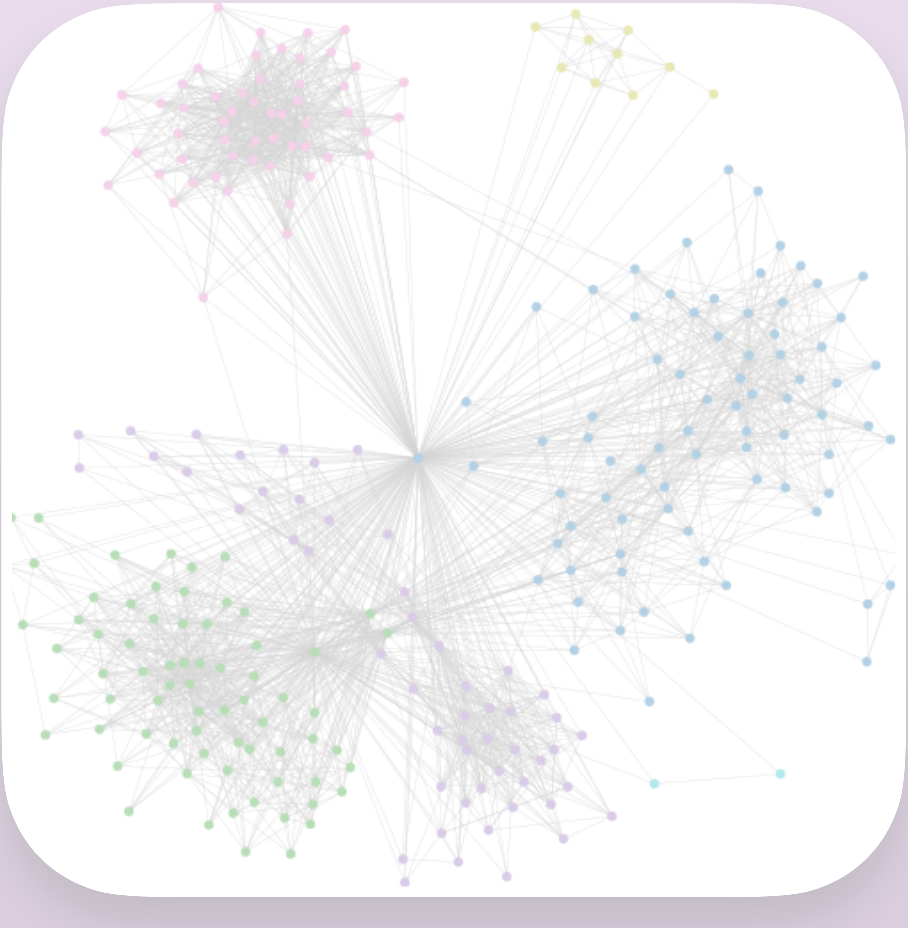


Uncovering new laws of nature

AI-guided scientific measurement



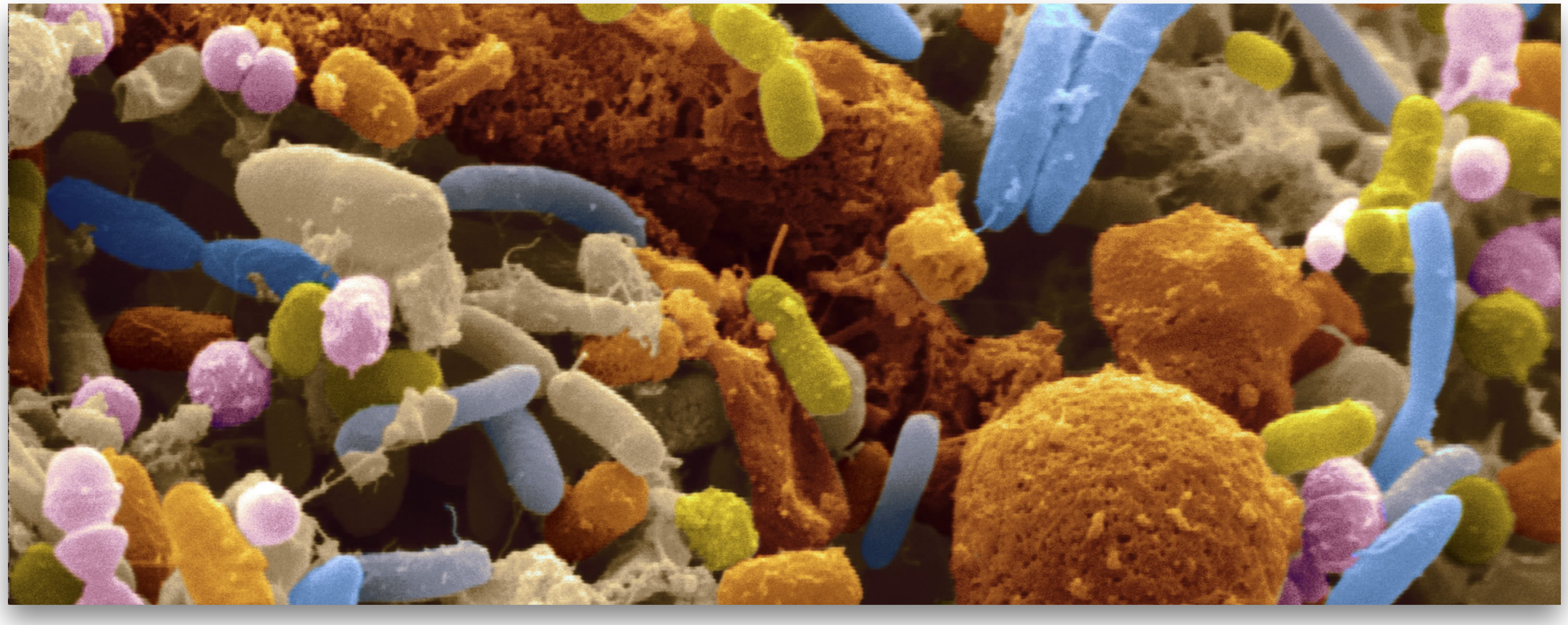
Physics-informed machine learning



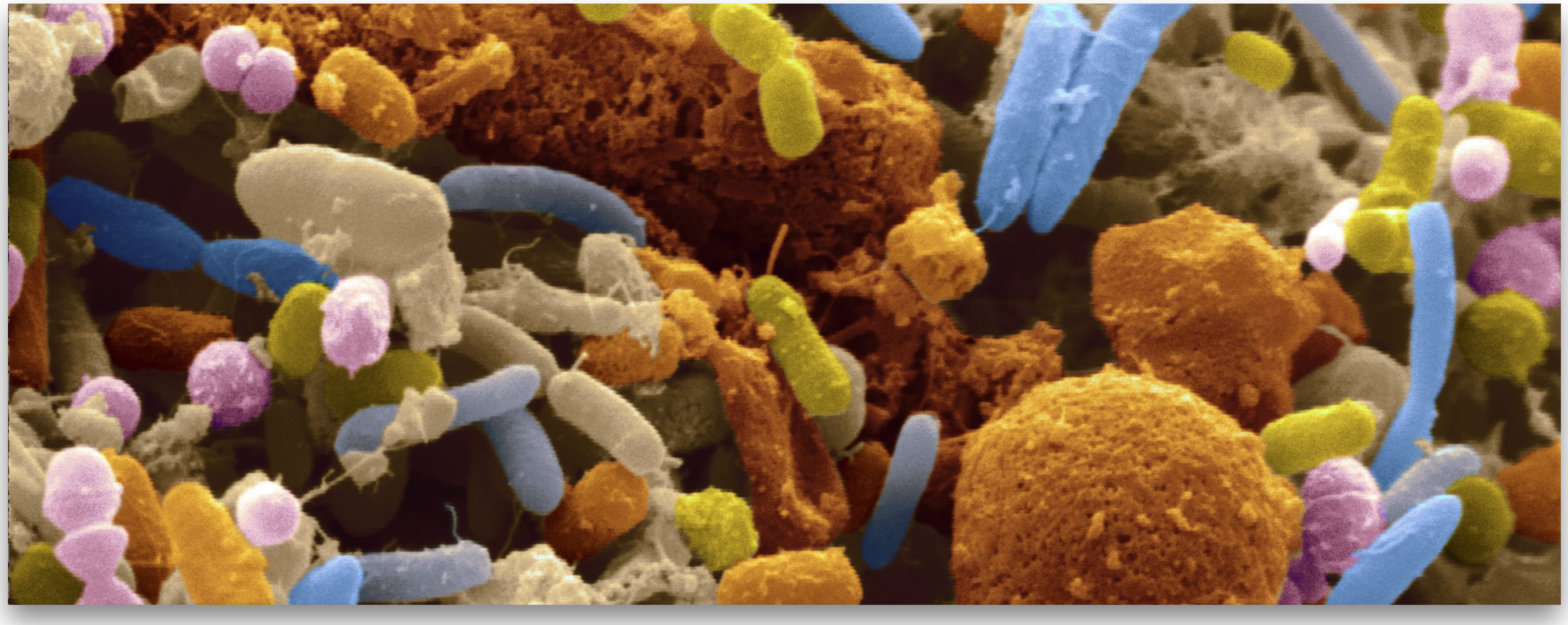
Advancing ML frontiers

Use AI to design better experiments, simulations, and sensors

How do we design a microbial community with maximum fitness?



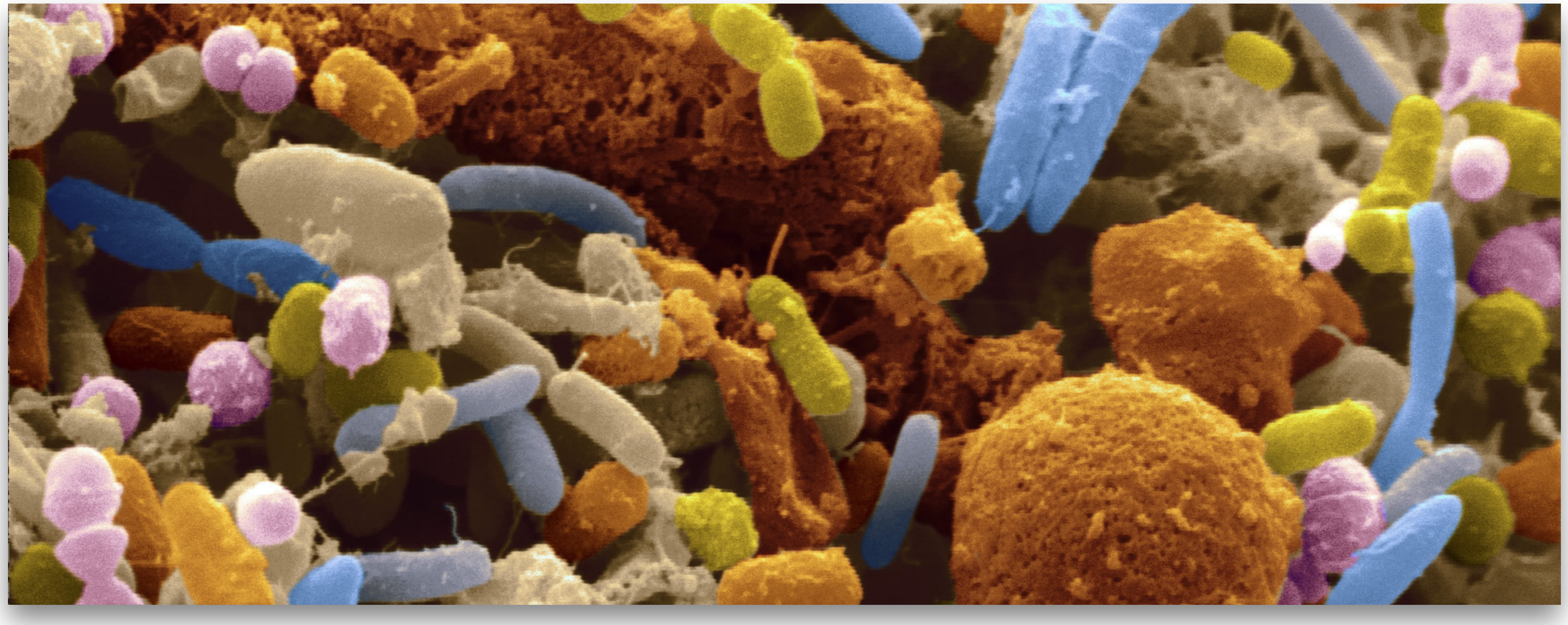
How do we design a microbial community with maximum fitness?



Community fitness = $f(S, N, A, E)$ where

- $S =$ Strain population densities
- $N =$ Nutrient sources concentration
- $A =$ Anti-microbial peptides concentration
- $E =$ Environmental conditions
- f is an unknown function we want to maximize

How do we design a microbial community with maximum fitness?



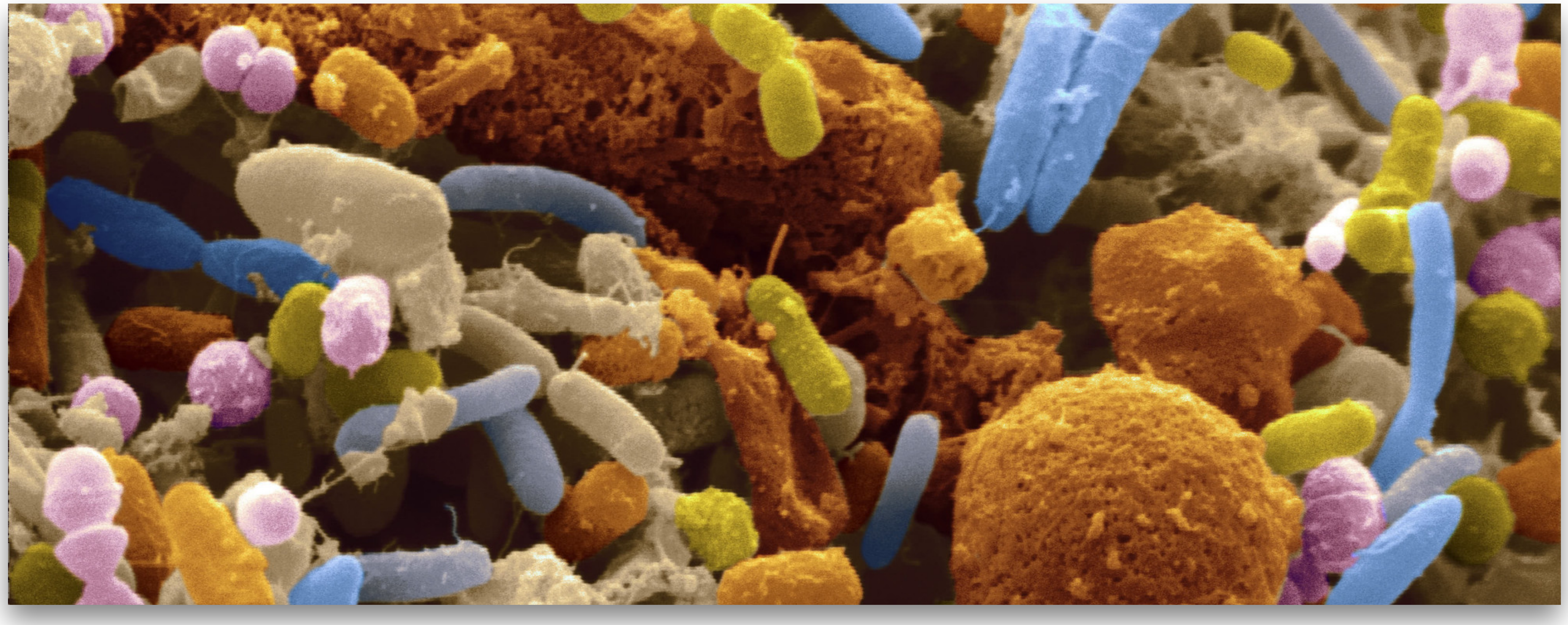
There are too many possible combinations of (S, N, A, E) to test them all

Sampling at random may mean conducting many experiments far from the maximum we seek

Community fitness = $f(S, N, A, E)$ where

- $S = S$train population densities
- $N = N$utrient sources concentration
- $A = A$nti-microbial peptides concentration
- $E = E$nvironmental conditions
- f is an unknown function we want to maximize

How do we design a microbial community with maximum fitness?



There are too many possible combinations of (S, N, A, E) to test them all

Sampling at random may mean conducting many experiments far from the maximum we seek

Community fitness = $f(S, N, A, E)$ where

- $S = S$train population densities
- $N = N$utrient sources concentration
- $A = A$nti-microbial peptides concentration
- $E = E$nvironmental conditions
- f is an unknown function we want to maximize

Foundational work in uncertainty quantification, active learning, and bandit methods help guide sequences of experiments to find maximally fit communities.

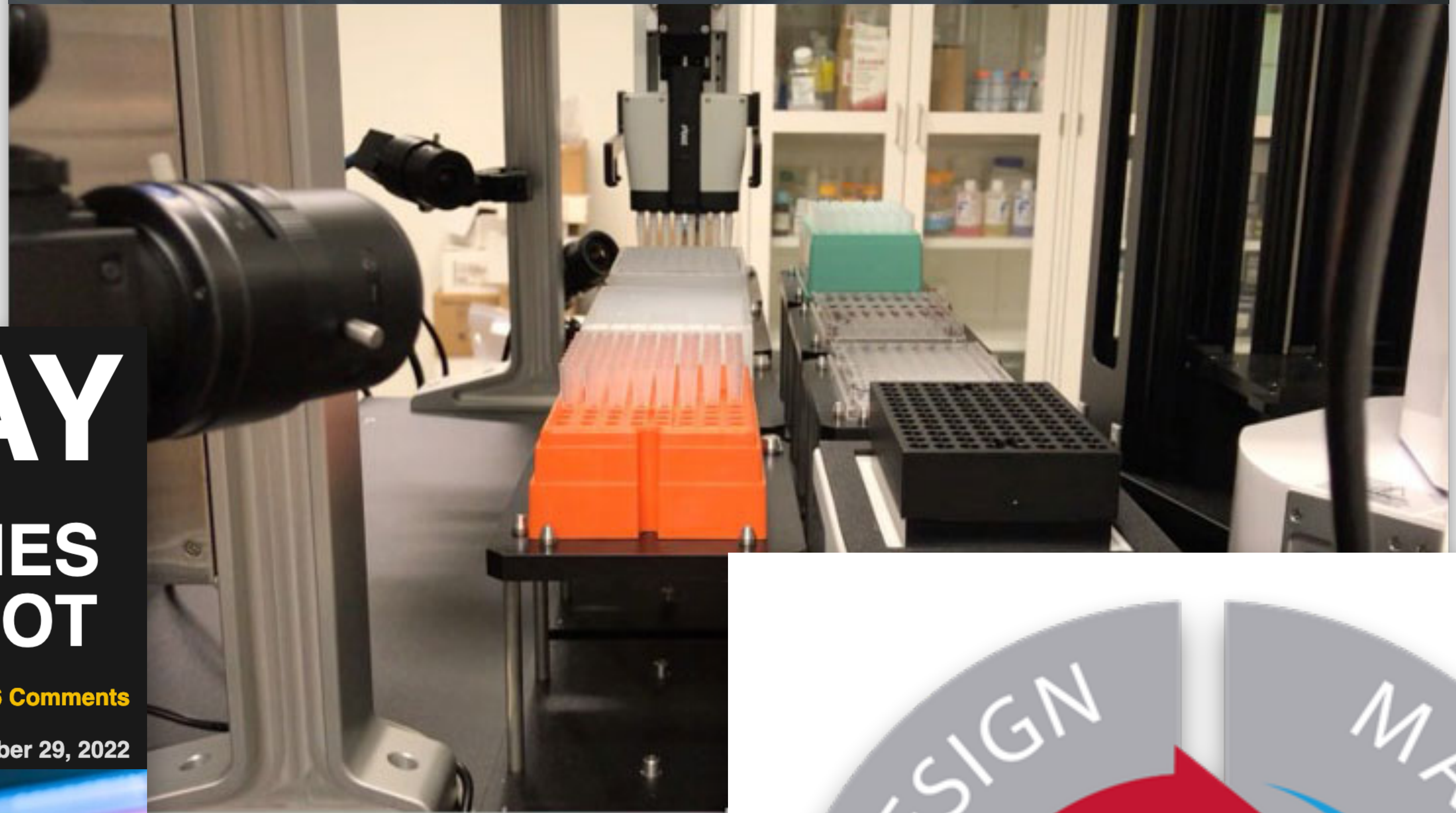
ScienceAdvances

Self-driving laboratory for accelerated discovery of thin-film materials

B. P. MACLEOD, F. G. L. PARLANE, T. D. MORRISSEY, F. HÄSE, L. M. ROCH, K. E. DETTELBACH, R. MOREIRA, L. P. E. YUNKER, M. B. ROONEY, J. R. DEETH, V. LAI, G. J. NG, H. SITU, R. H. ZHANG, M. S. ELLIOTT, T. H. HALEY, D. J. DVORAK, A. ASPURU-GUZIK, J. E. HEIN, AND C. P. BERLINGUETTE
fewer Authors Info & Affiliations

Argonne NATIONAL LABORATORY

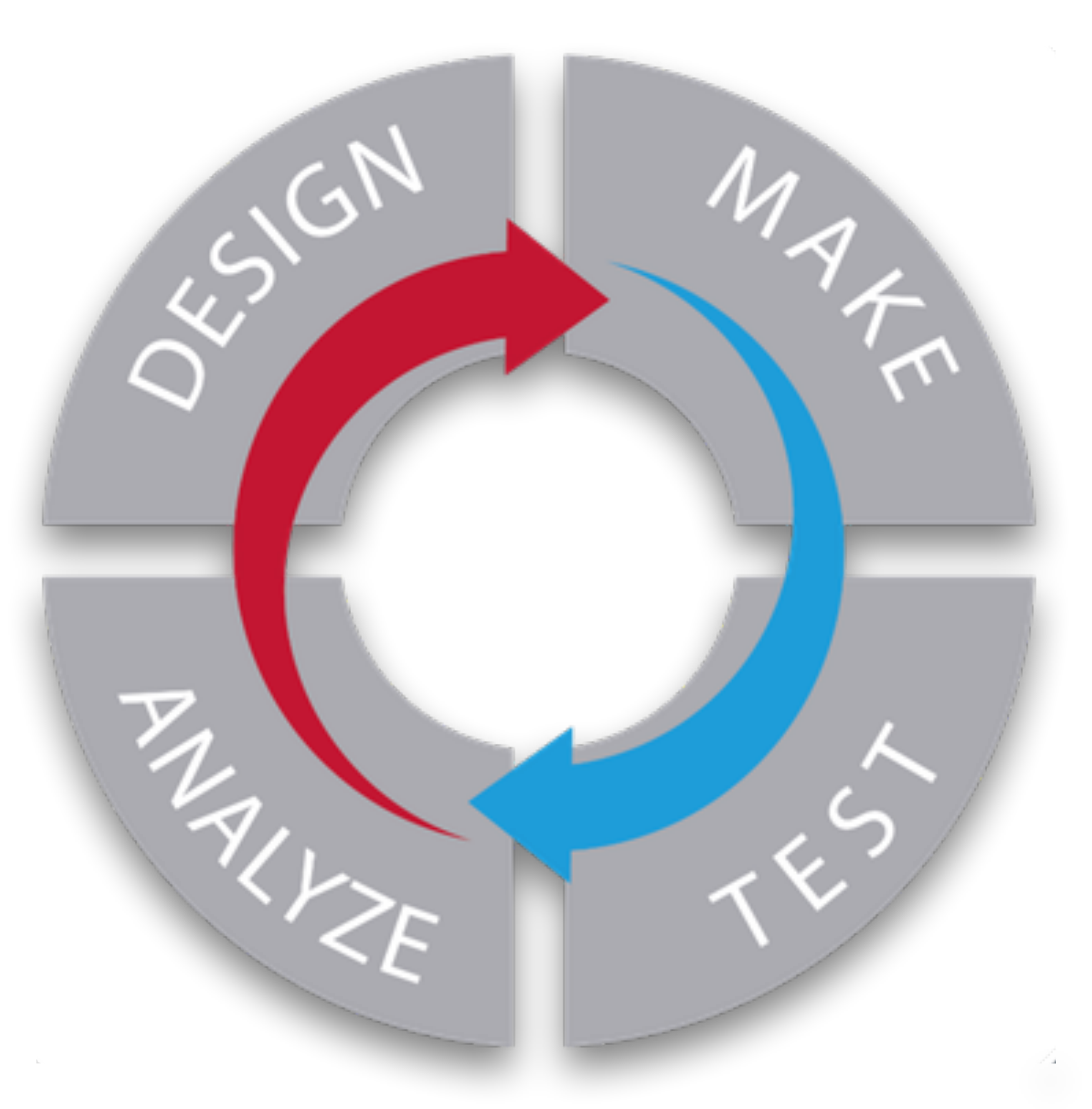
Developing a self-driving laboratory prototype

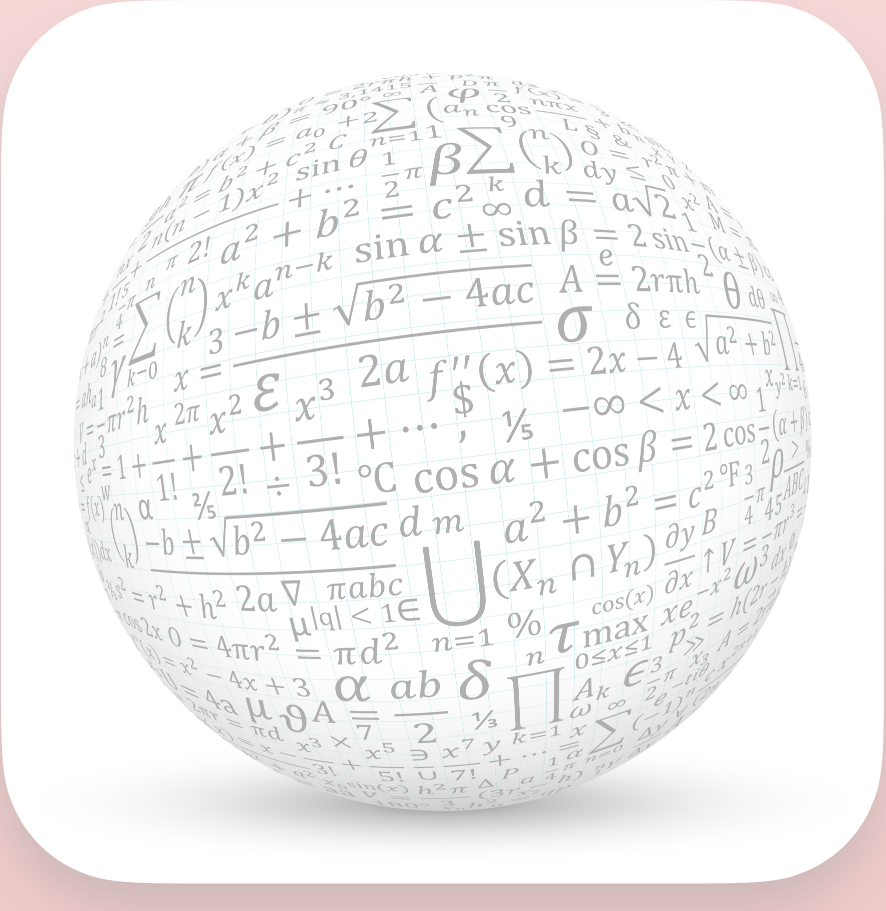


HACKADAY

SELF-DRIVING LABORATORIES DO RESEARCH ON AUTOPILOT

by: Lewin Day
16 Comments
September 29, 2022

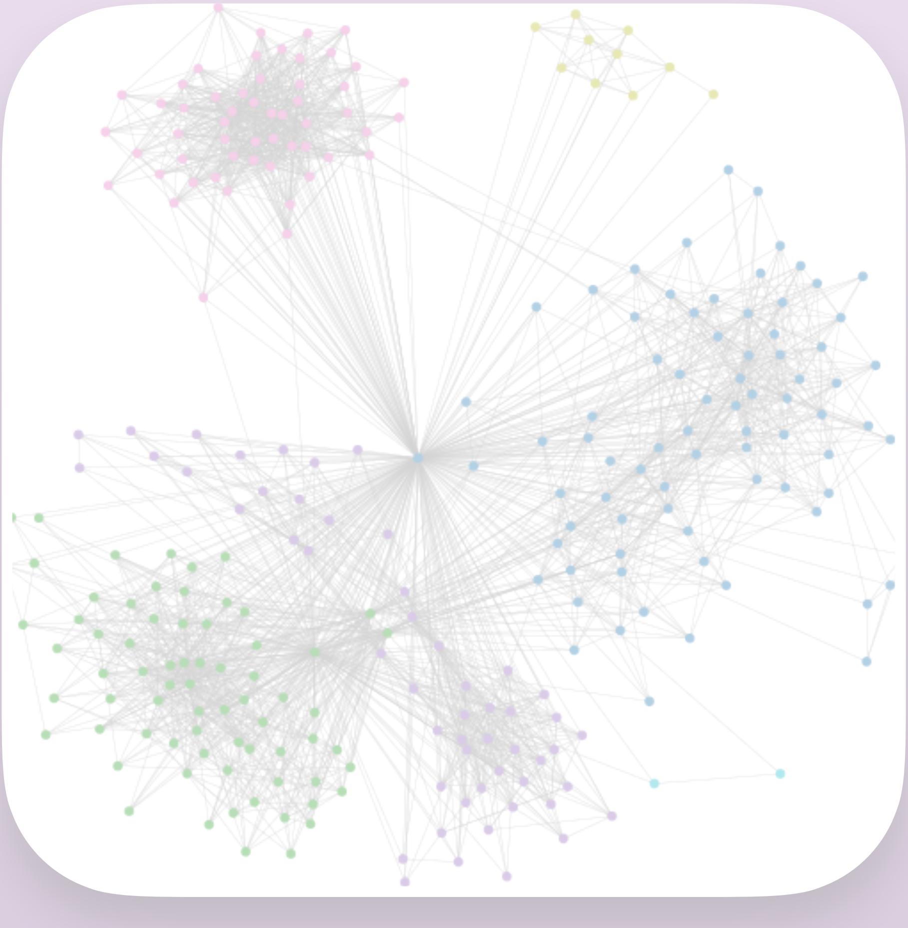




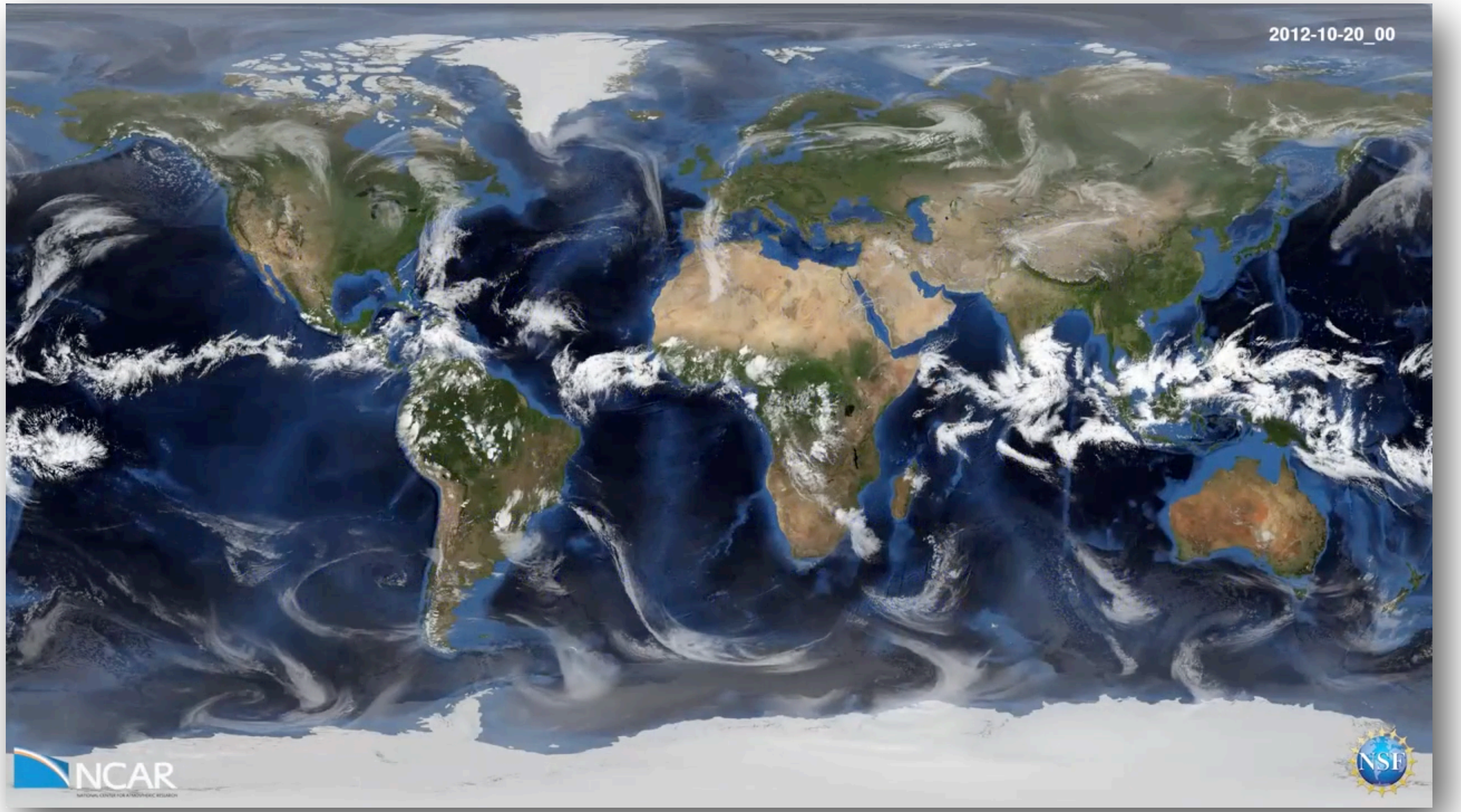
Uncovering new laws of nature

AI-guided scientific measurement

Physics-informed machine learning



Advancing ML frontiers

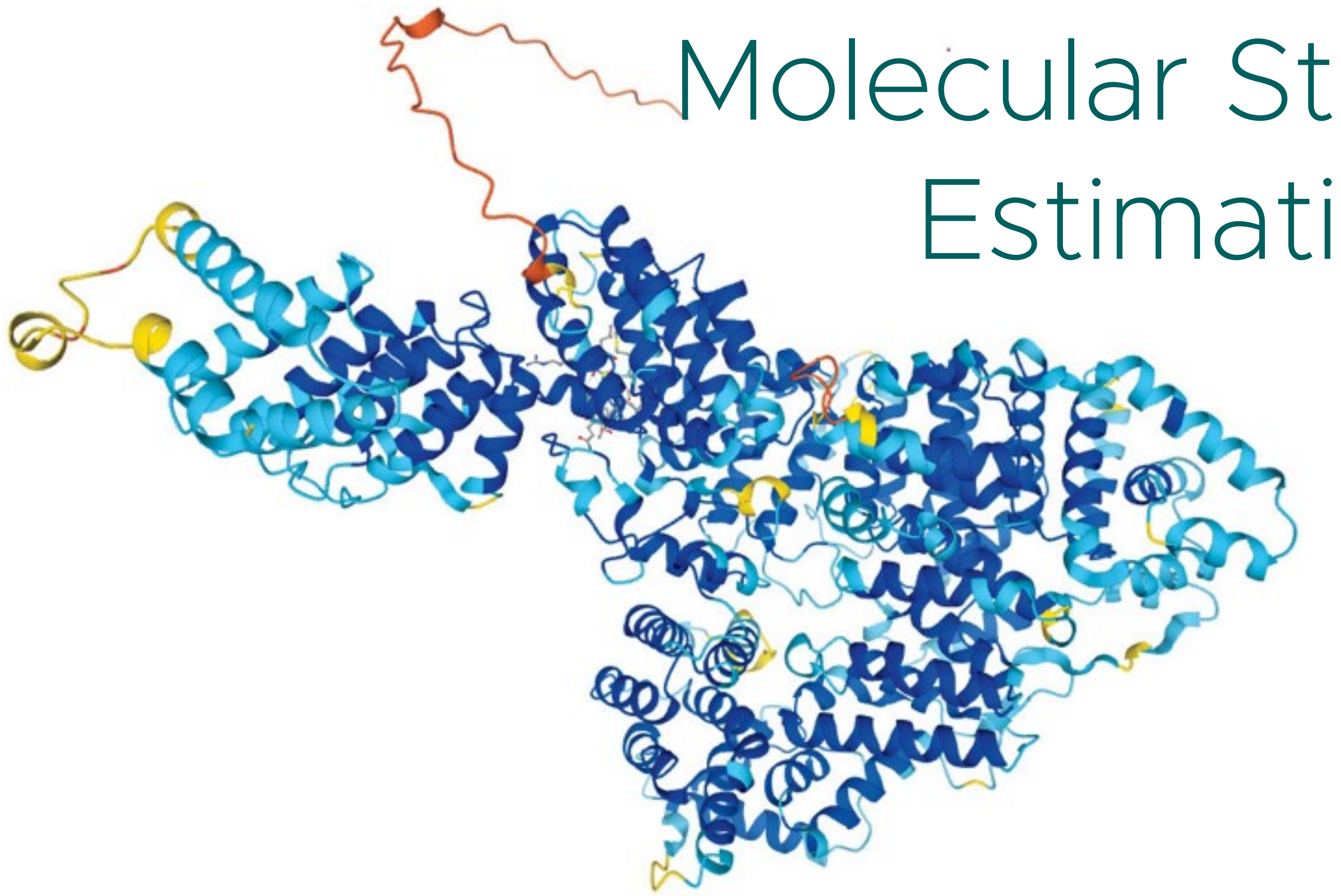


Optimally leverage physical models and experimental or observational data

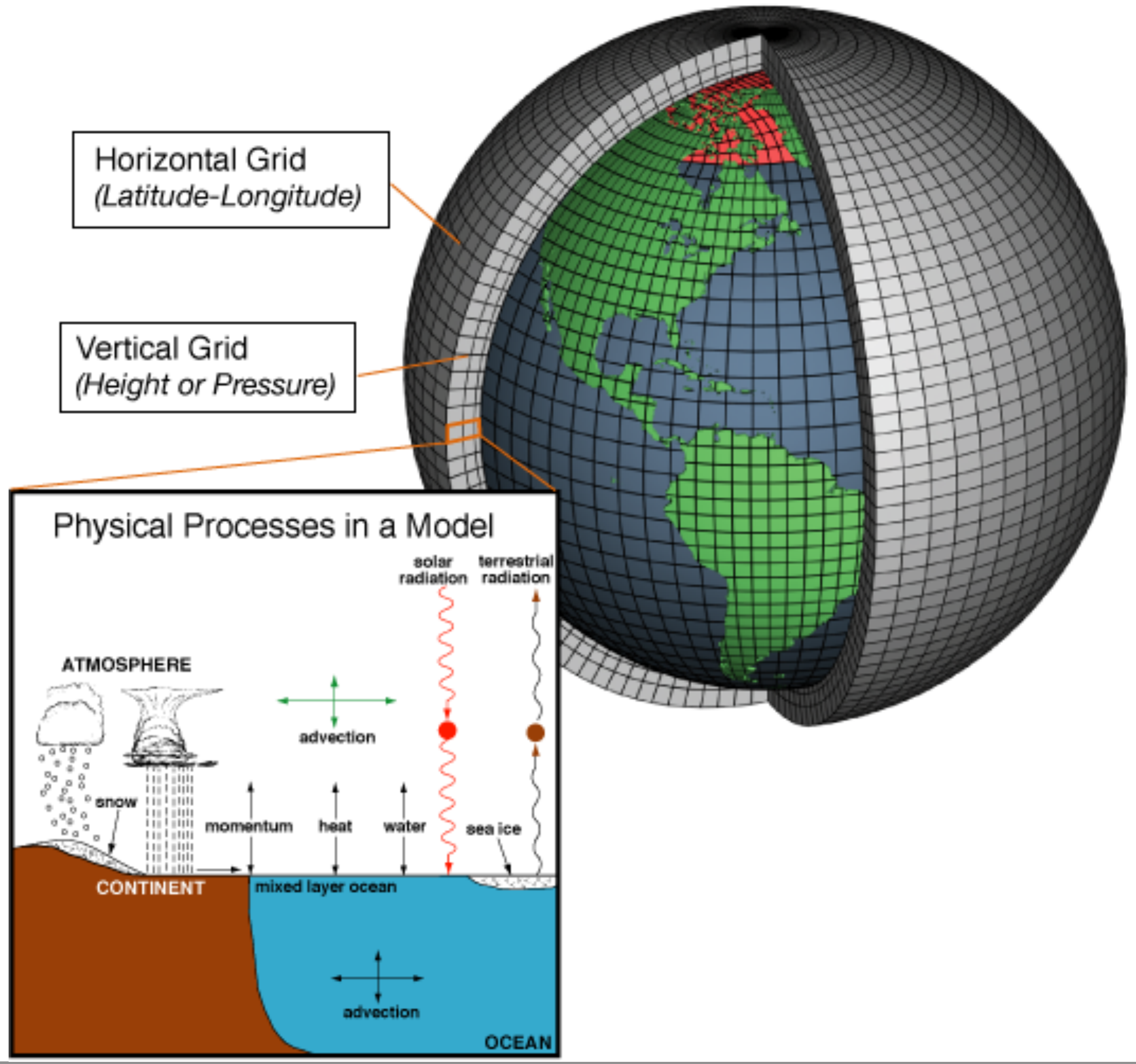
There are *many* settings in which we have both training data and physical models.

Physics Experi-ments

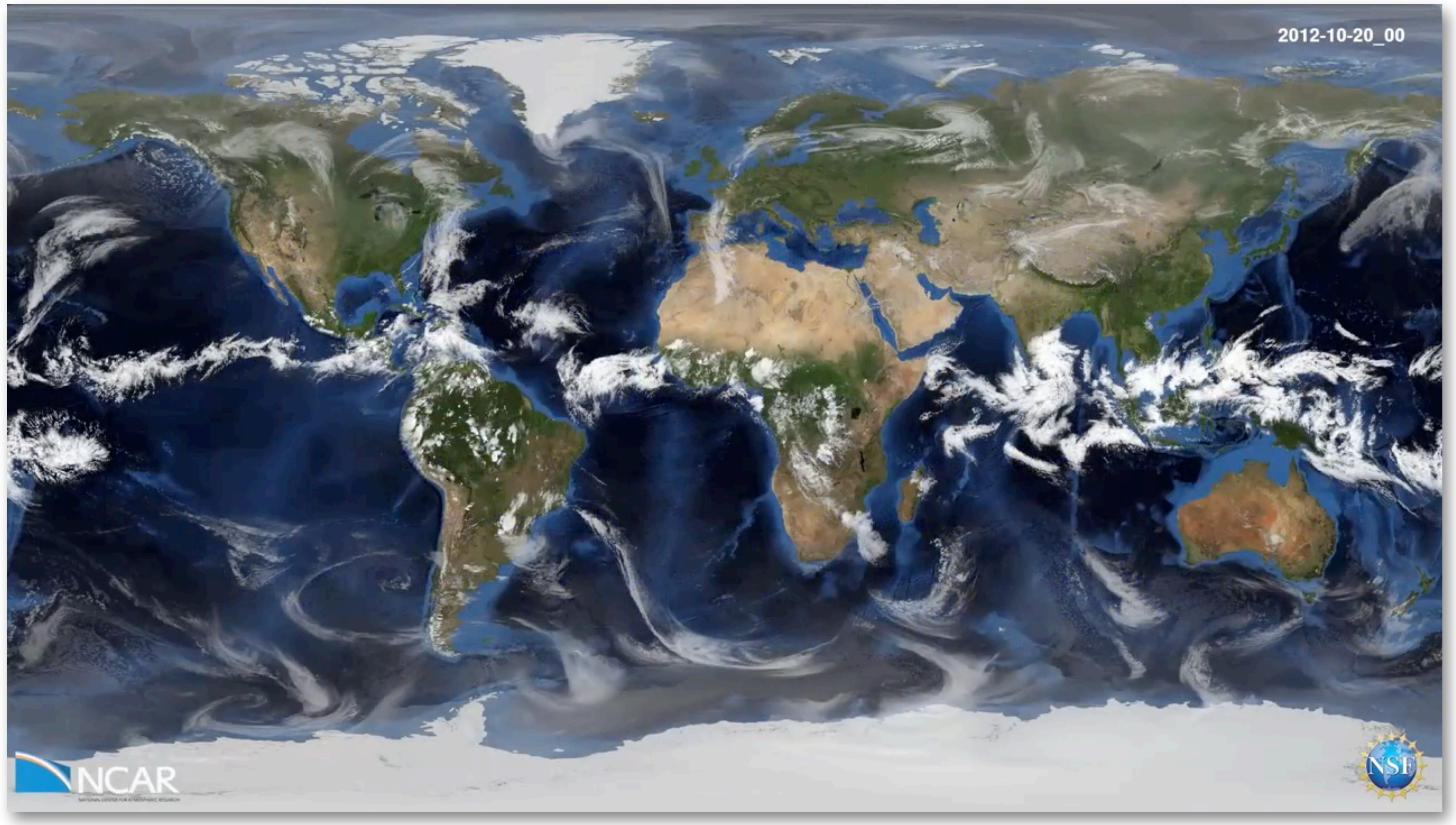
Molecular Structure Estimation



Climate Forecasting



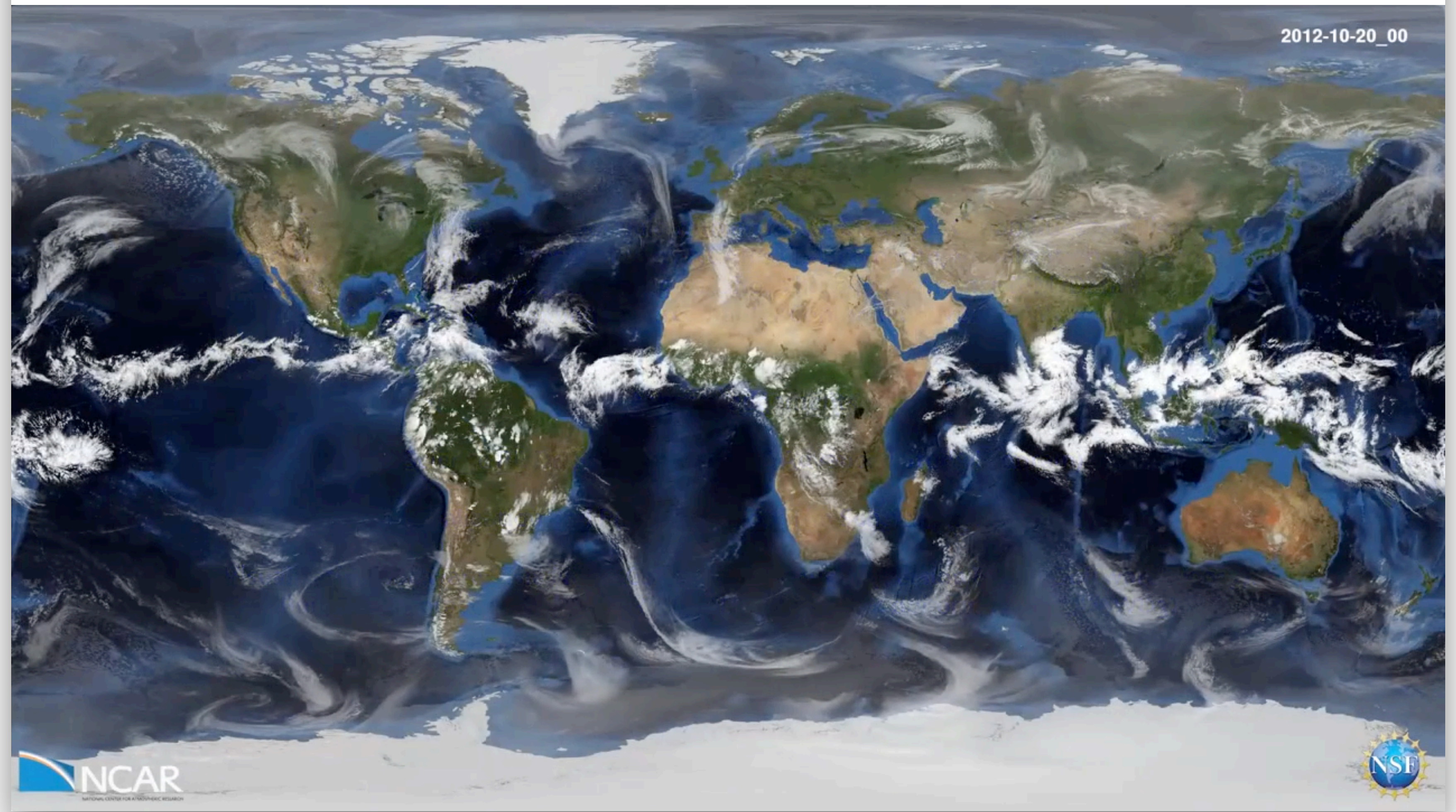
Also fluid dynamics, turbulence, particle accelerators, scattering, automatic control…



Learning from simulations

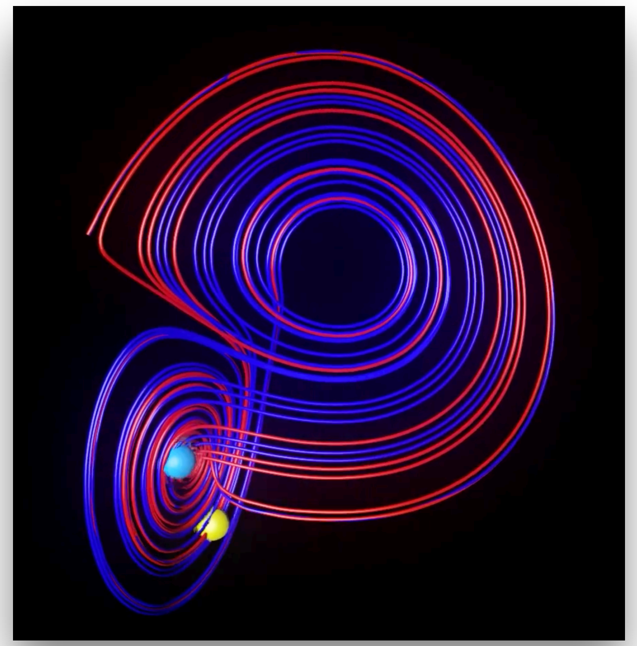
requires understanding distribution drift, transfer learning, data assimilation, reduced-order modeling, and active learning

How to jointly leverage simulations and data?

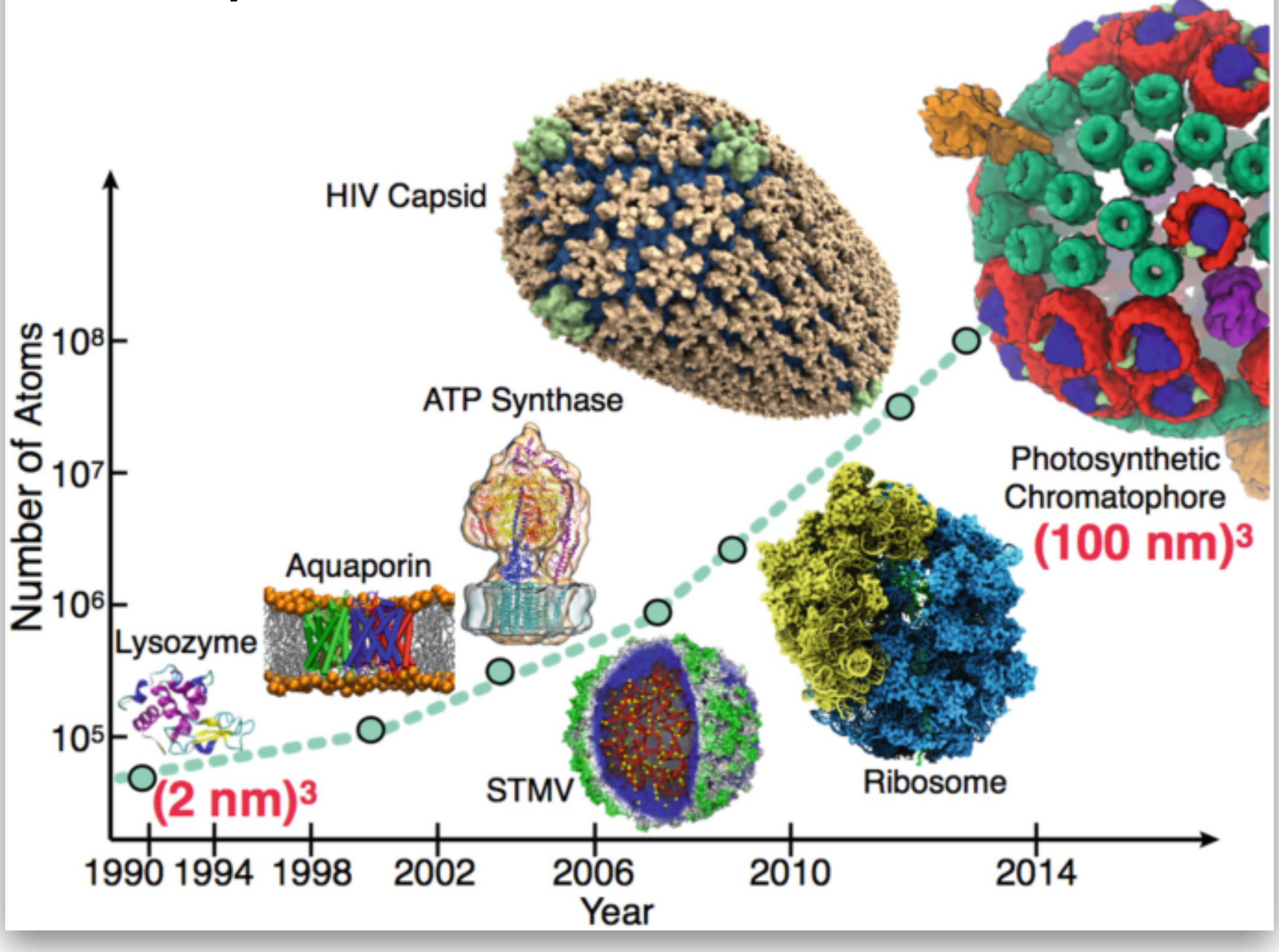


Physics-informed neural networks

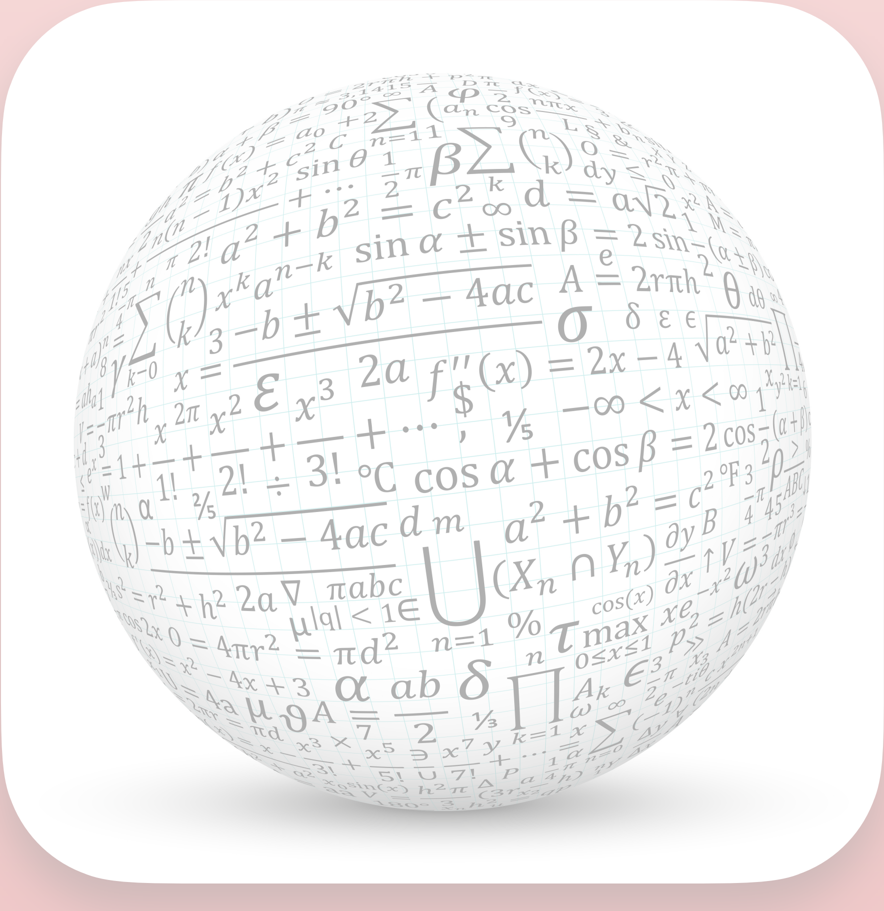
$$\frac{dx}{dt} = \sigma(y - x)$$
$$\frac{dy}{dt} = x(\rho - z) - y$$
$$\frac{dz}{dt} = xy - \beta z$$



Upscaling molecular dynamics simulations

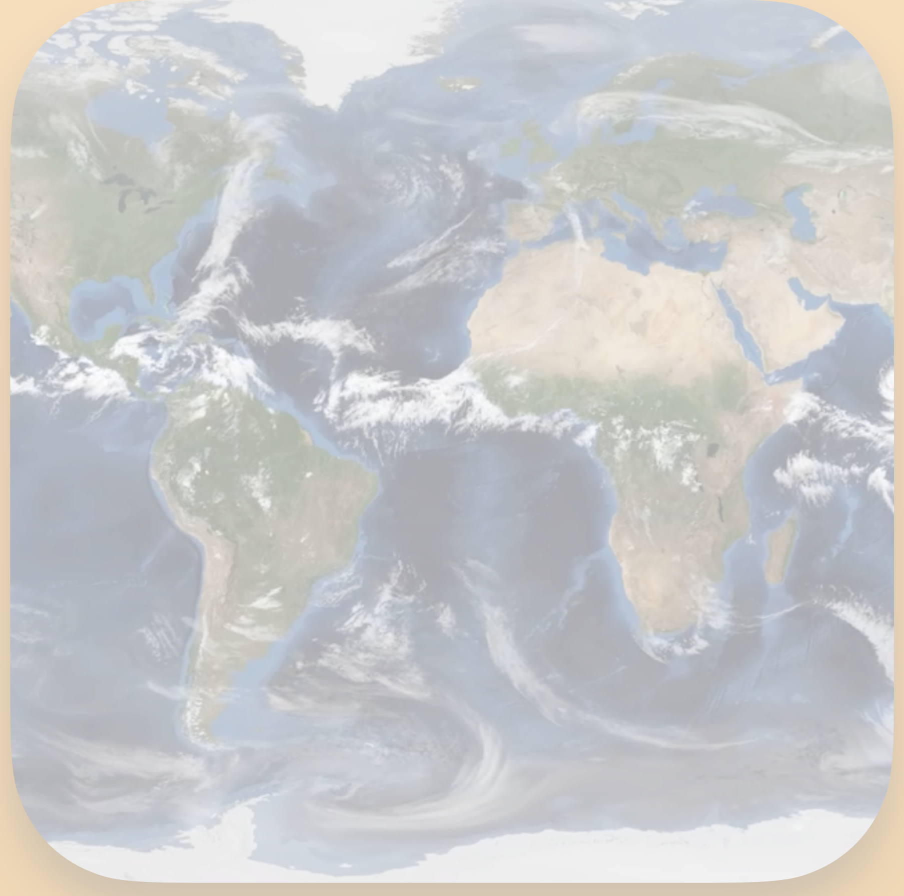


Physics-informed machine learning promotes **robustness** and **efficiency**, and is essential to **extrapolating beyond domain of training data**



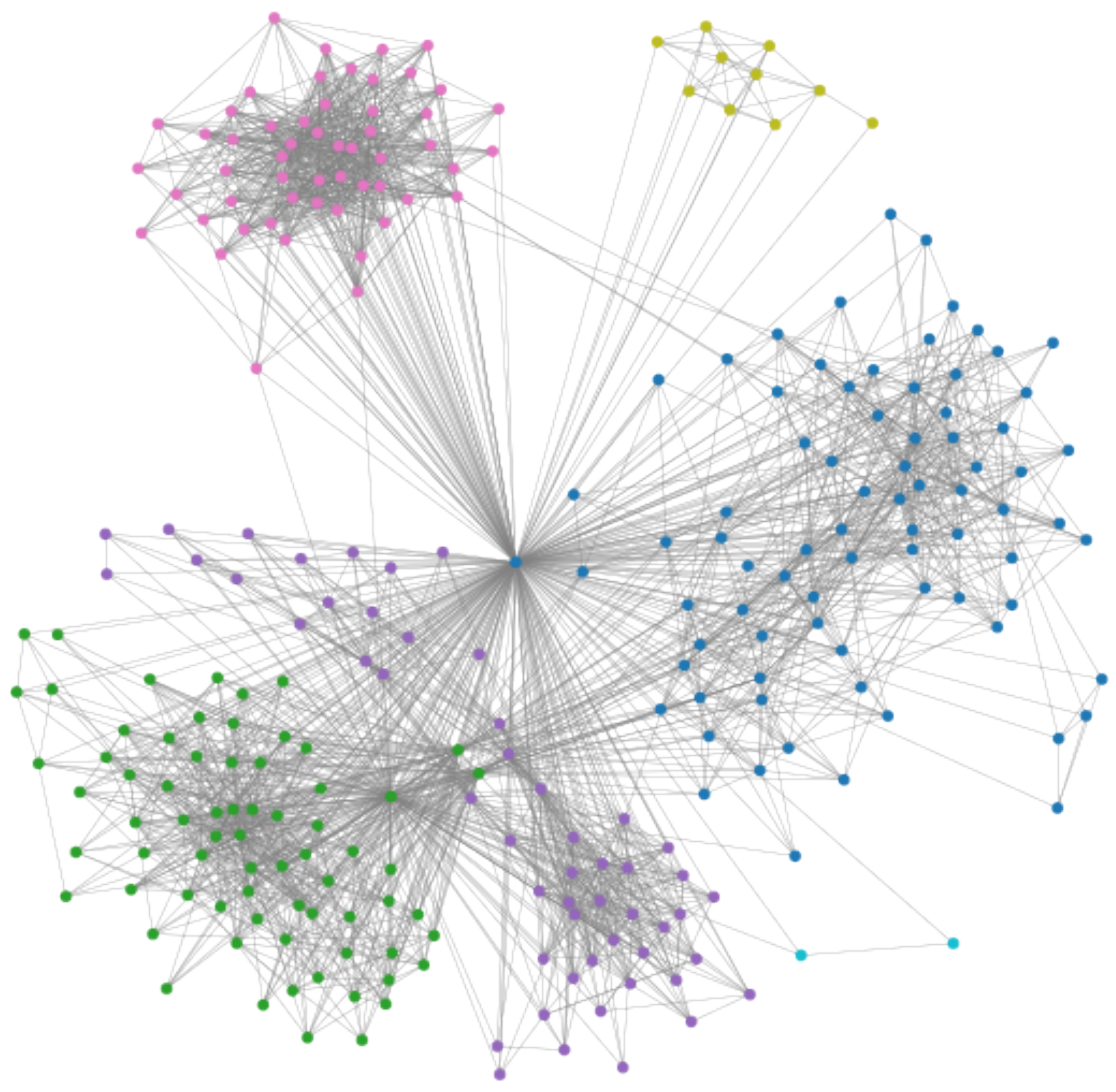
Uncovering new laws of nature

AI-guided scientific measurement



Physics-informed machine learning

Advancing ML frontiers



Develop new ML theory and methods inspired by scientific settings with broad impacts

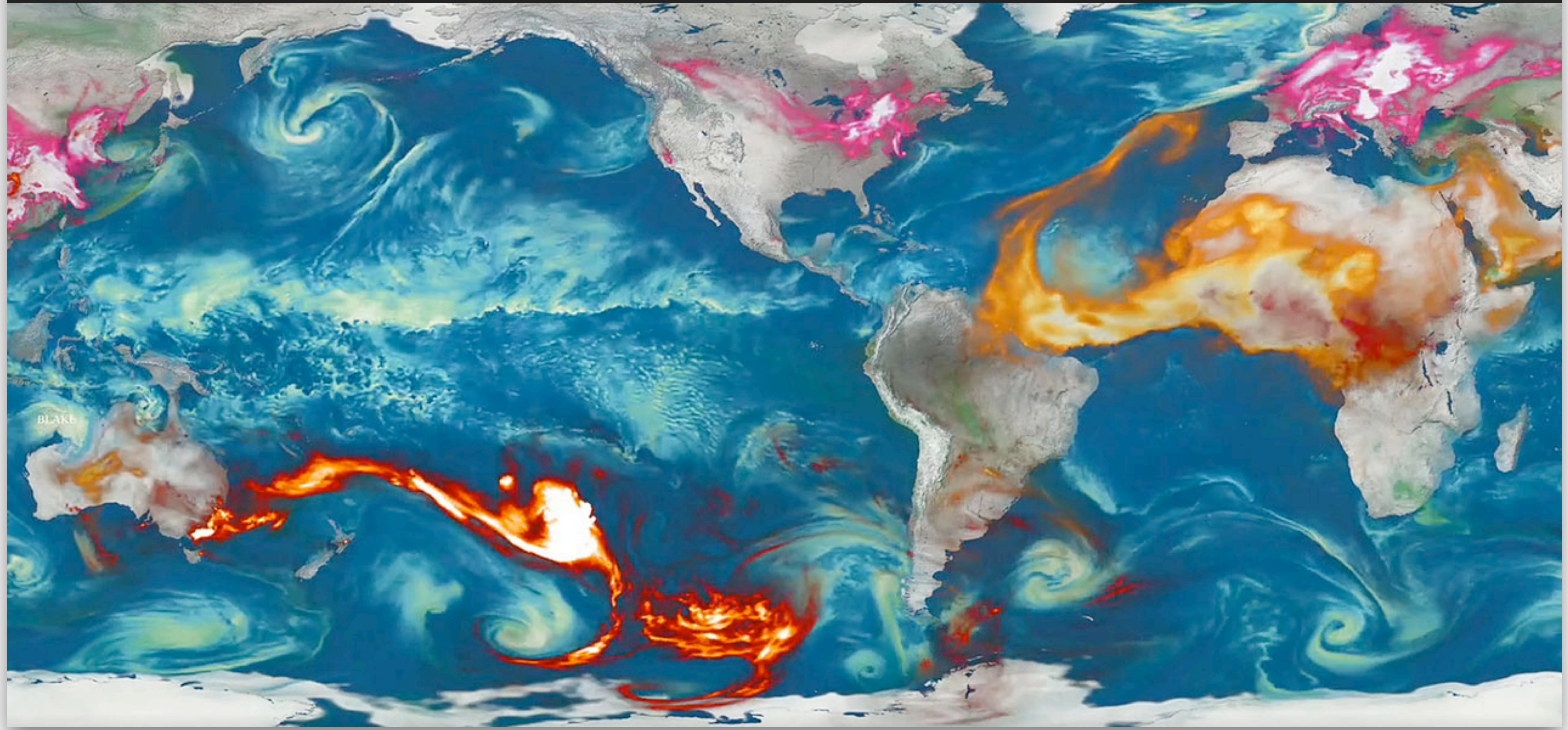
Learned emulators are models trained to mimic numerical simulations at a much lower computational cost, particularly for parameters or inputs that have not been simulated.

Science

From models of galaxies to atoms, simple AI shortcuts speed up simulations by billions of times

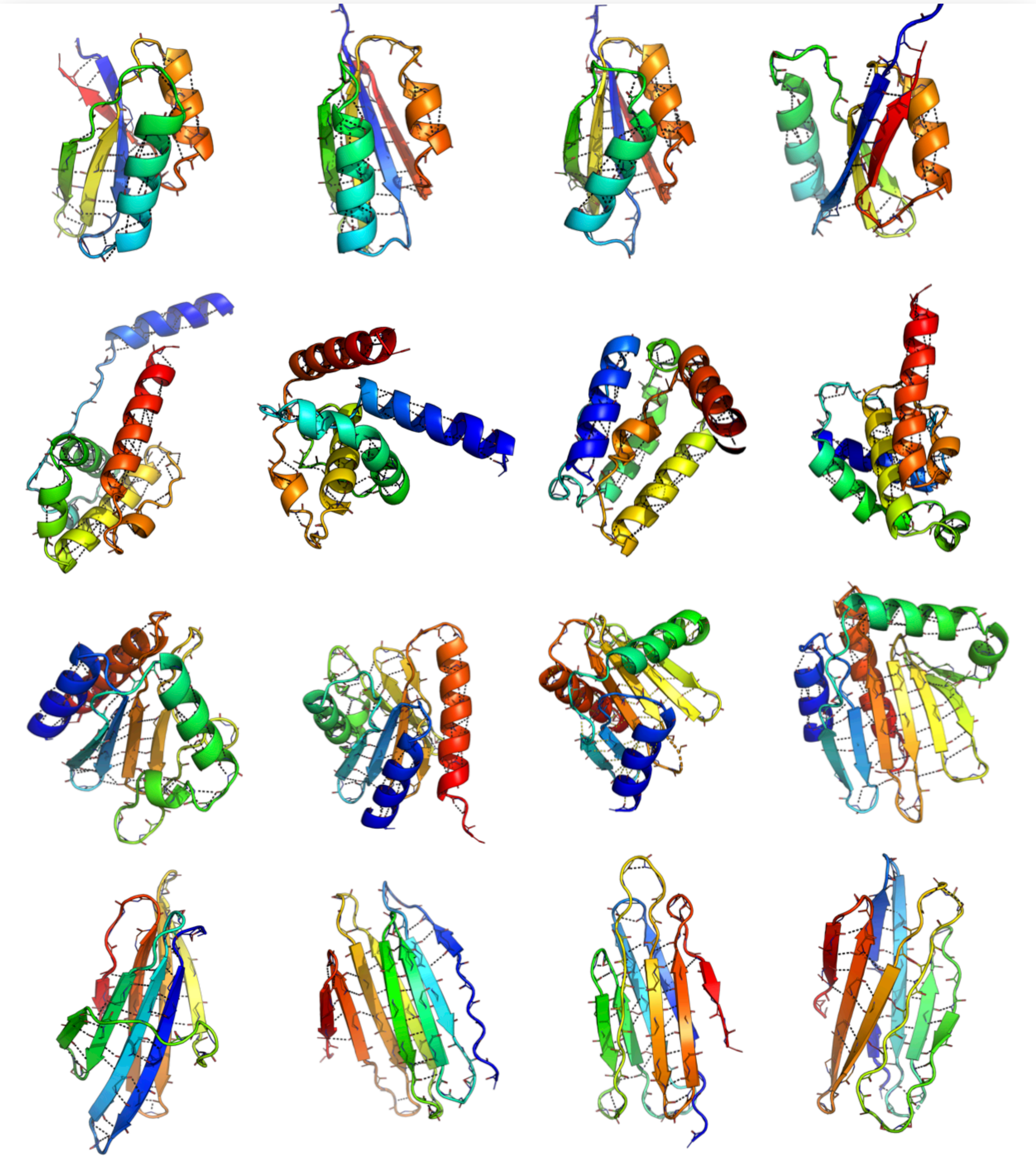
With little training, neural networks create accurate emulators for physics, astronomy, and earth science

12 FEB 2020 · BY MATTHEW HUTSON



https://www.science.org/content/article/models-galaxies-atoms-simple-ai-shortcuts-speed-simulations-billions-times

Generative Models for Science



Generative models in science must account for

- much less data than ChatGPT, DALL·E
- rare events, chaotic dynamics
- wide variations in scales and resolutions
- physical models, constraints, symmetries, & invariances

You

List 3 reasons football is safer than badminton

ChatGPT

[Some caveats]

1. **Protective Gear:** Football players often wear more extensive protective gear, including helmets, shoulder pads, and shin guards. This gear can provide additional protection against certain types of injuries, whereas badminton players typically wear minimal protective equipment.
2. **Rules and Refereeing:** Football has strict rules and regulations enforced by referees, which can contribute to a more controlled and organized game. The presence of referees helps ensure that players adhere to the rules, minimizing the potential for dangerous situations. Badminton, being a non-contact sport, relies on players' sportsmanship and adherence to rules.
3. **Physical Conditioning:** Football players undergo rigorous physical conditioning to prepare for the demands of the sport. This conditioning can help reduce the risk of injuries by enhancing strength, flexibility, and overall fitness. Badminton players also need to be fit, but the physical demands of the sport are generally less intense.

You

List 3 reasons football is safer than badminton

ChatGPT

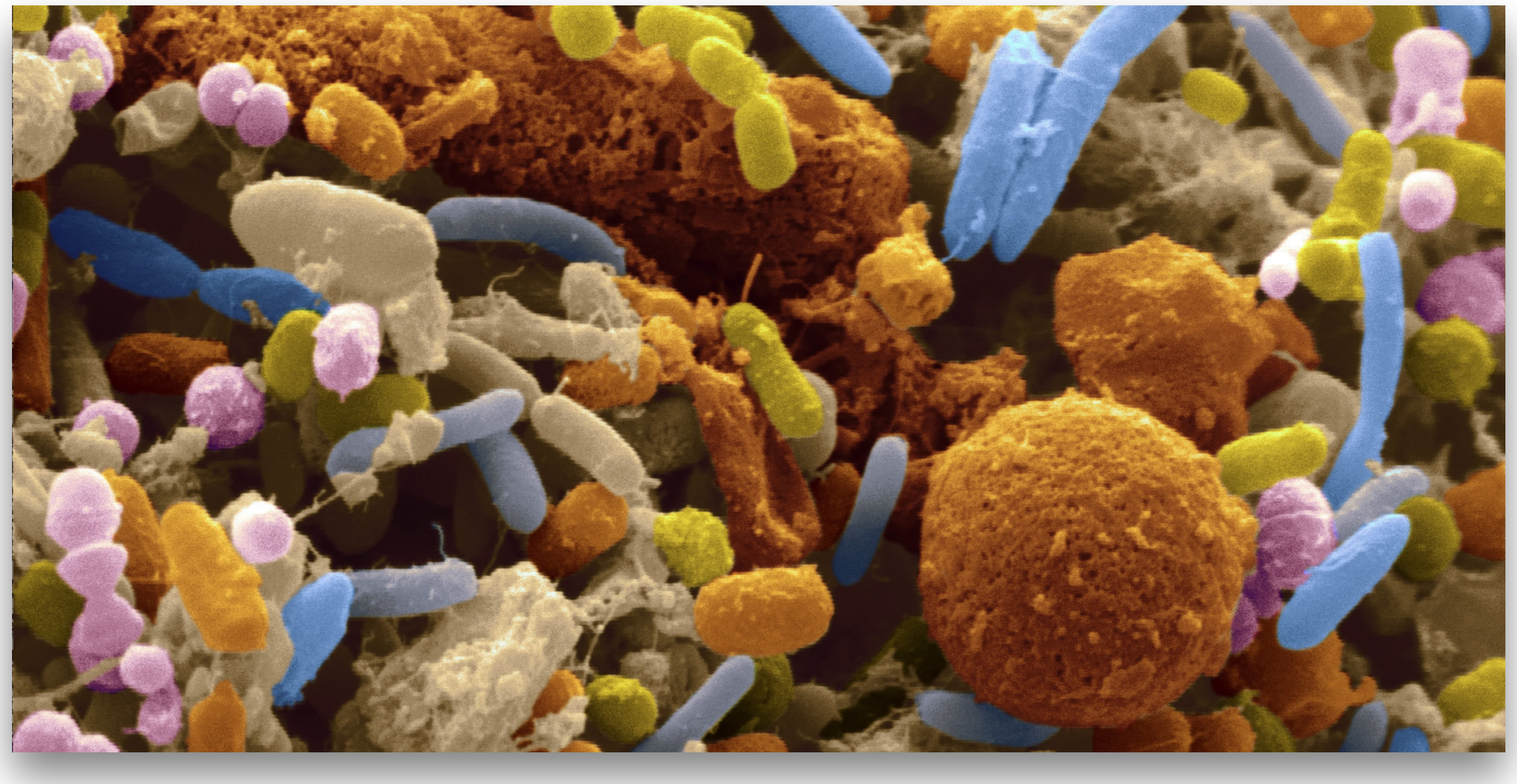
[Some caveats]

1. **Protective Gear:** Football players often wear more extensive protective gear, including helmets, shoulder pads, and shin guards. This gear can provide additional protection against certain types of injuries, whereas badminton players typically wear minimal protective equipment.
2. **Rules and Refereeing:** Football has strict rules and regulations enforced by referees, which can contribute to a more controlled and organized game. The presence of referees helps ensure that players adhere to the rules, minimizing the potential for dangerous situations. Badminton, being a non-contact sport, relies on players' sportsmanship and adherence to rules.
3. **Physical Conditioning:** Football players undergo rigorous physical conditioning to prepare for the demands of the sport. This conditioning can help reduce the risk of injuries by enhancing strength, flexibility, and overall fitness. Badminton players also need to be fit, but the physical demands of the sport are generally less intense.

Generative models can produce *plausible* results, but foundational work is needed to make them trustworthy

Uncertainty quantification

aids in validating models, guiding scientific measurement, and analyzing rare events



aws

Contact Us Sup

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Customer Enablement

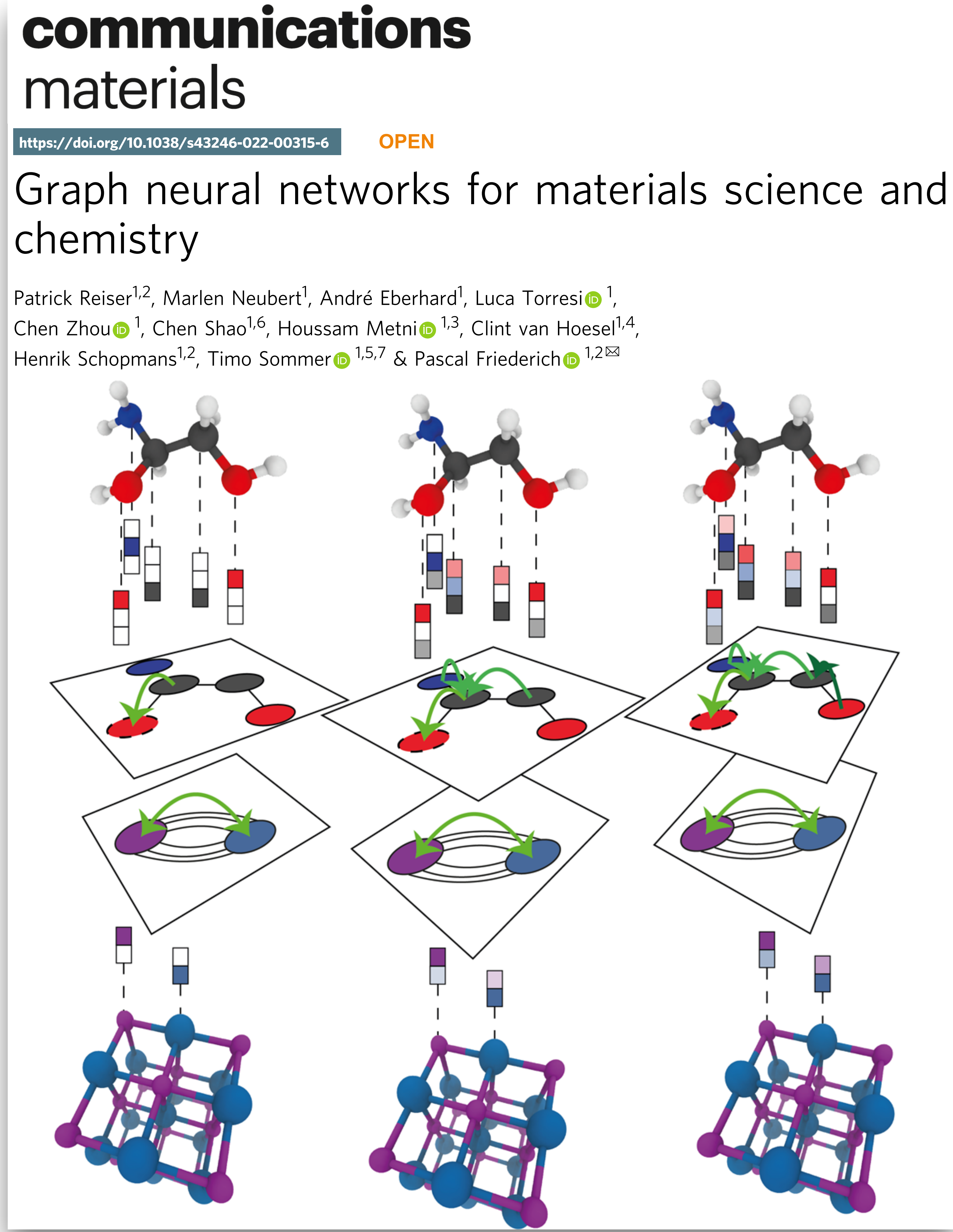
AWS Blog Home Topics ▾ Edition ▾

AWS Machine Learning Blog

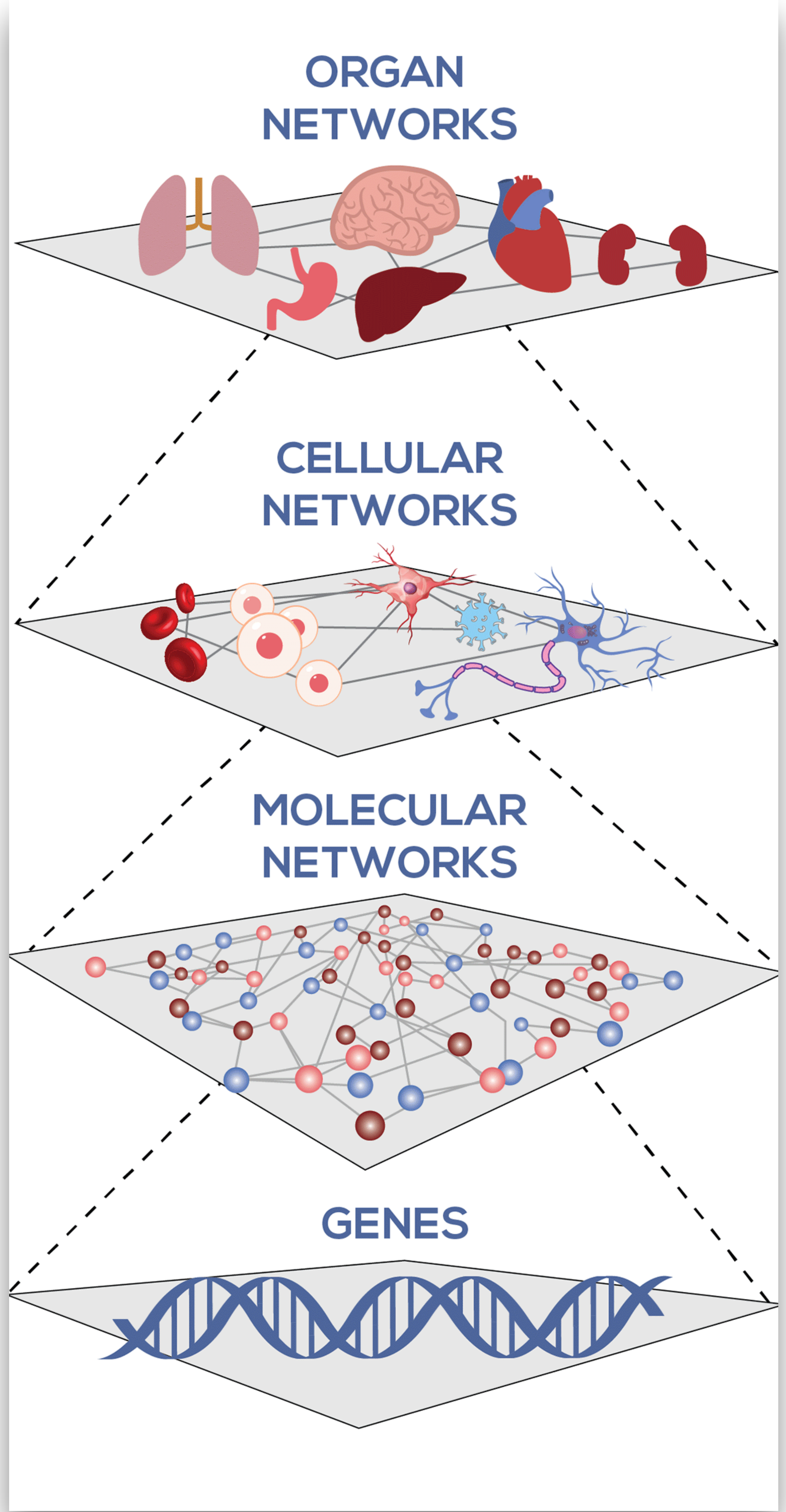
Introducing Fortuna: A library for uncertainty quantification

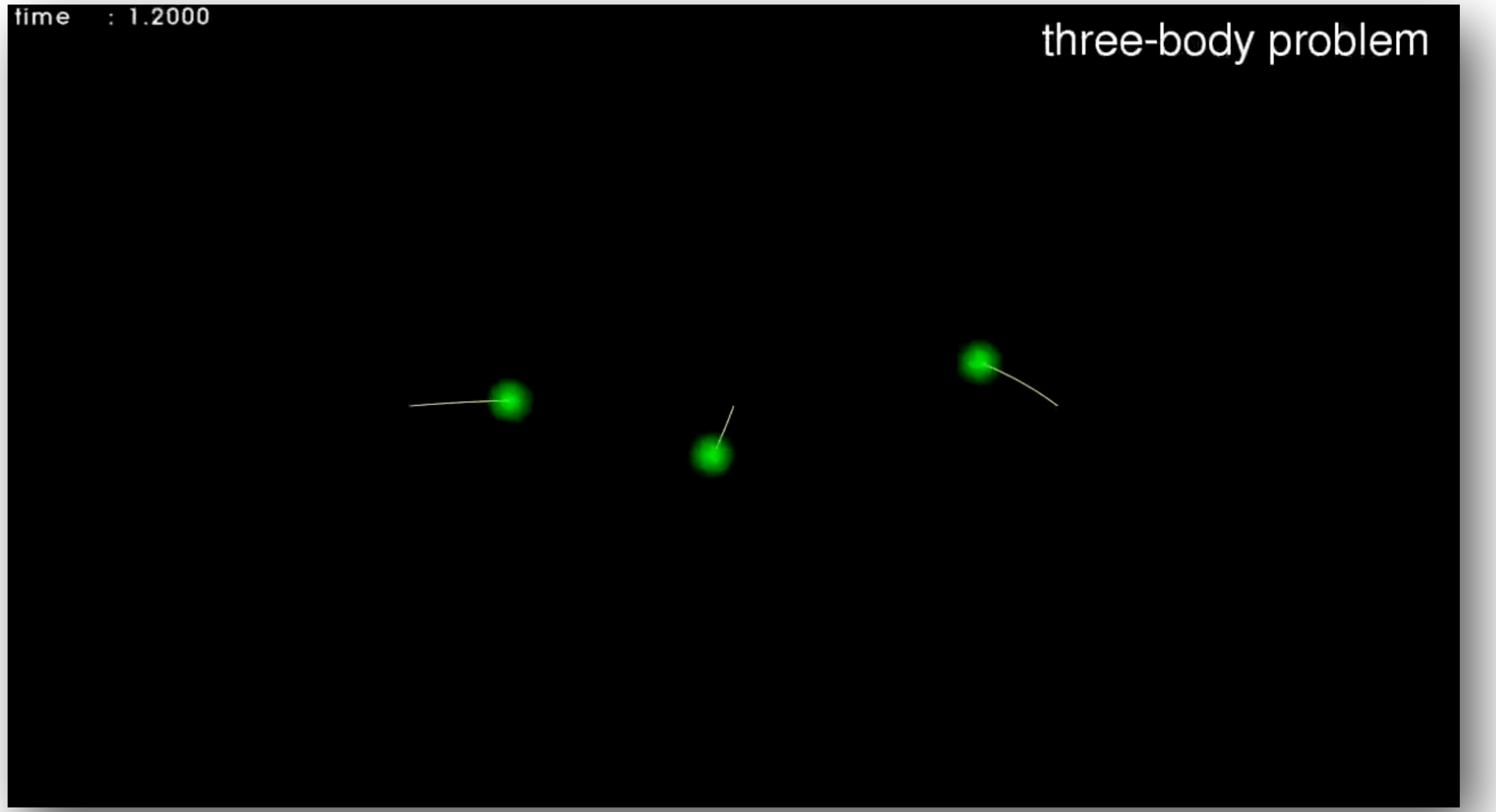
by Gianluca Detommaso, Alberto Gasparin, Cedric Archambeau, Michele Donini, Matthias Seeger, and Andrew Gordon Wilson | on 16 DEC 2022 | in Amazon Machine Learning, Artificial Intelligence, Foundational (100) | Permalink | Comments | Share

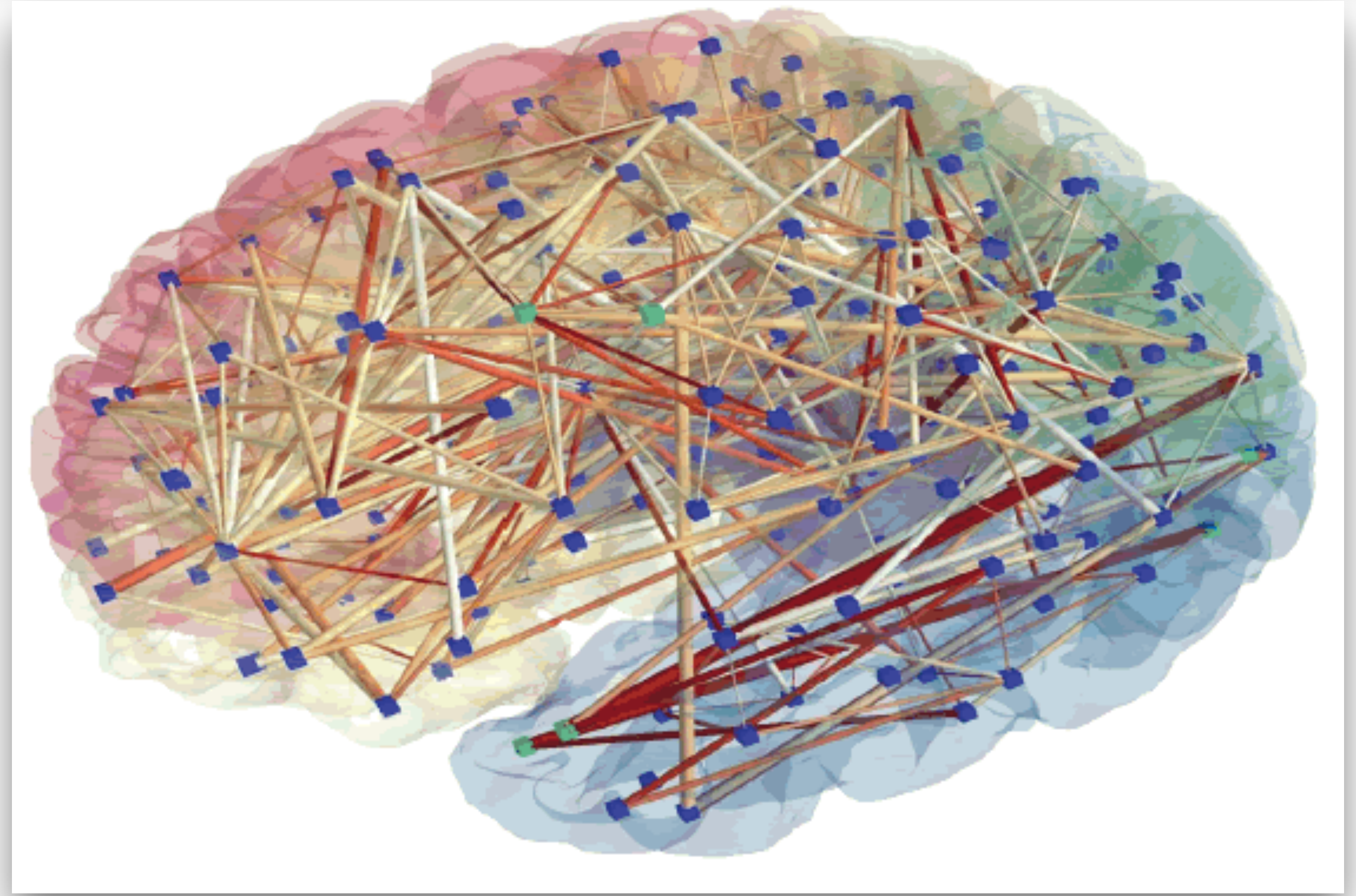
Proper estimation of predictive uncertainty is fundamental in applications that involve critical decisions. Uncertainty can be used to assess the reliability of model predictions, trigger human intervention, or decide whether a model can be safely deployed in the wild.


communications
materials
https://doi.org/10.1038/s43246-022-00315-6 OPEN
Graph neural networks for materials science and chemistry

Patrick Reiser[1,2], Marlen Neubert[1], André Eberhard[1], Luca Torresi[1], Chen Zhou[1], Chen Shao[1,6], Houssam Metni[1,3], Clint van Hoesel[1,4], Henrik Schopmans[1,2], Timo Sommer[1,5,7] & Pascal Friederich[1,2]✉



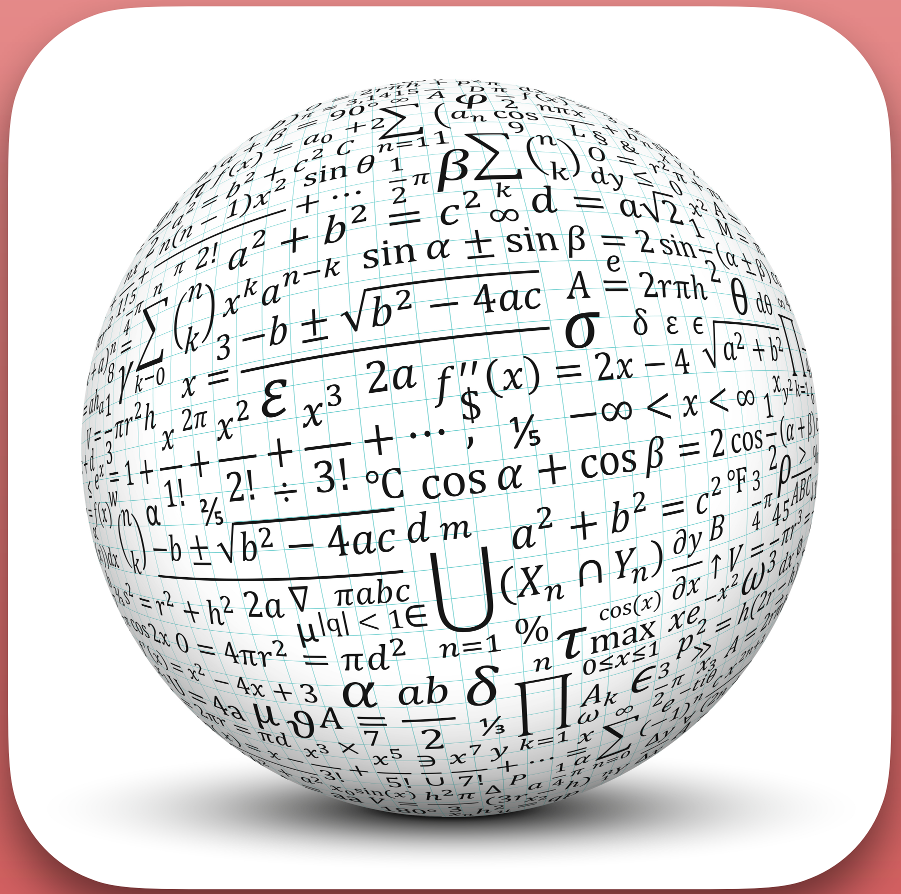




Graphs: a fundamental representation of scientific data

https://www.pulsus.com/scholarly-articles/essentials-in-brain-connectivity-3758.html, https://isbscience.org/about/what-is-systems-biology/network-biology/, https://www.youtube.com/watch?v=_3uQqrrBcrQ

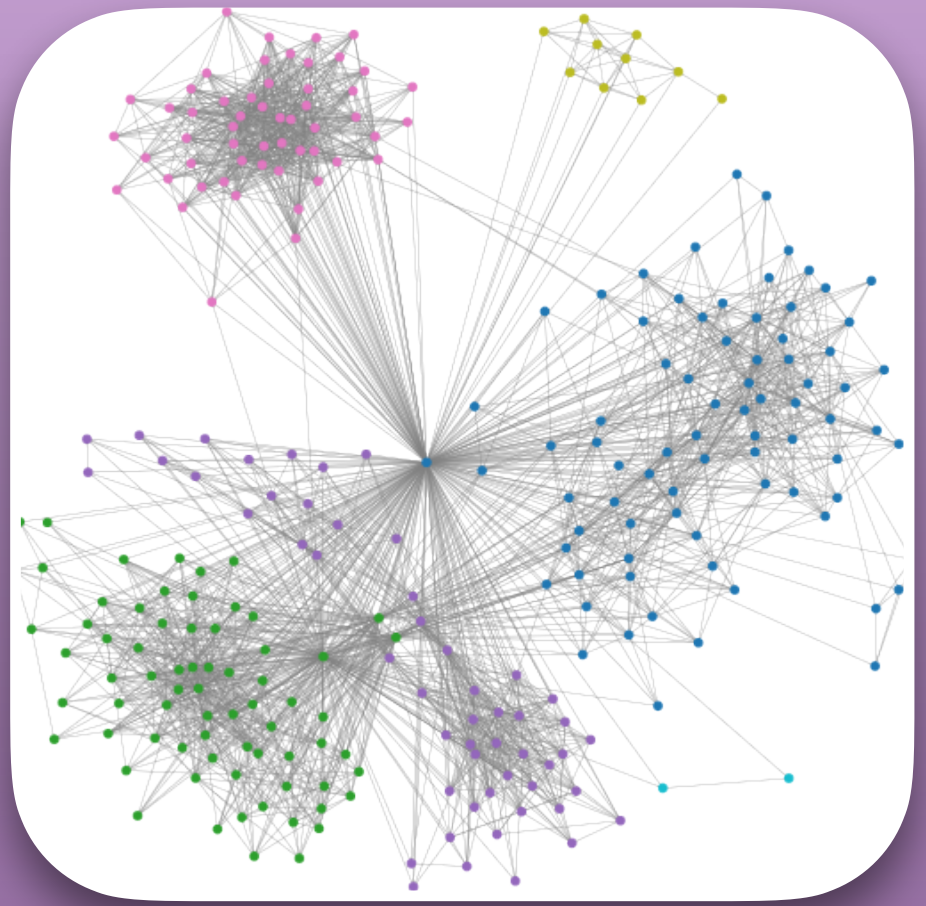
Privacy, transparency, fairness, and accountability
pose additional foundational challenges with human-centric data



Uncovering new laws of nature

ML-guided scientific measurement

Physics-informed machine learning



Advancing ML frontiers

Machine Learning Foundations

AI & ML are transforming science

Investments in AI & ML foundations are essential for high-quality, reproducible, AI-enabled scientific research

WIRED

WILL KNIGHT BUSINESS AUG 10, 2022 7:00 AM

Sloppy Use of Machine Learning Is Causing a ‘Reproducibility Crisis’ in Science

AI hype has researchers in fields from medicine to sociology rushing to use techniques that they don’t always understand—causing a wave of spurious results.

PHOTOGRAPH: PM IMAGES/GETTY IMAGES

Science

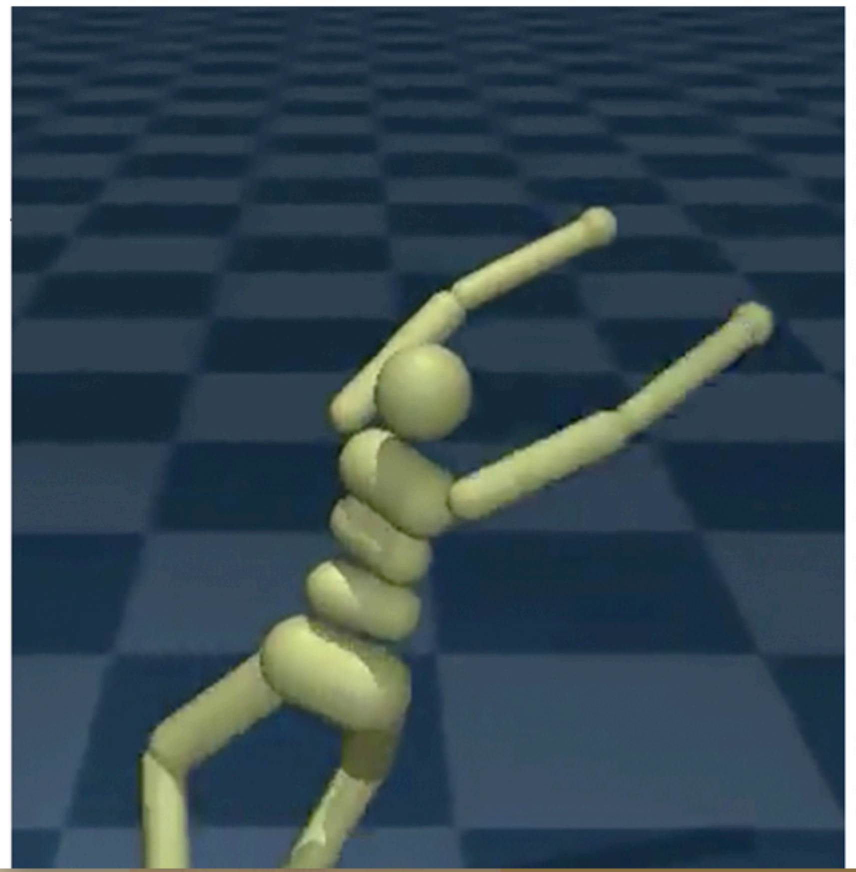
Artificial intelligence faces reproducibility crisis

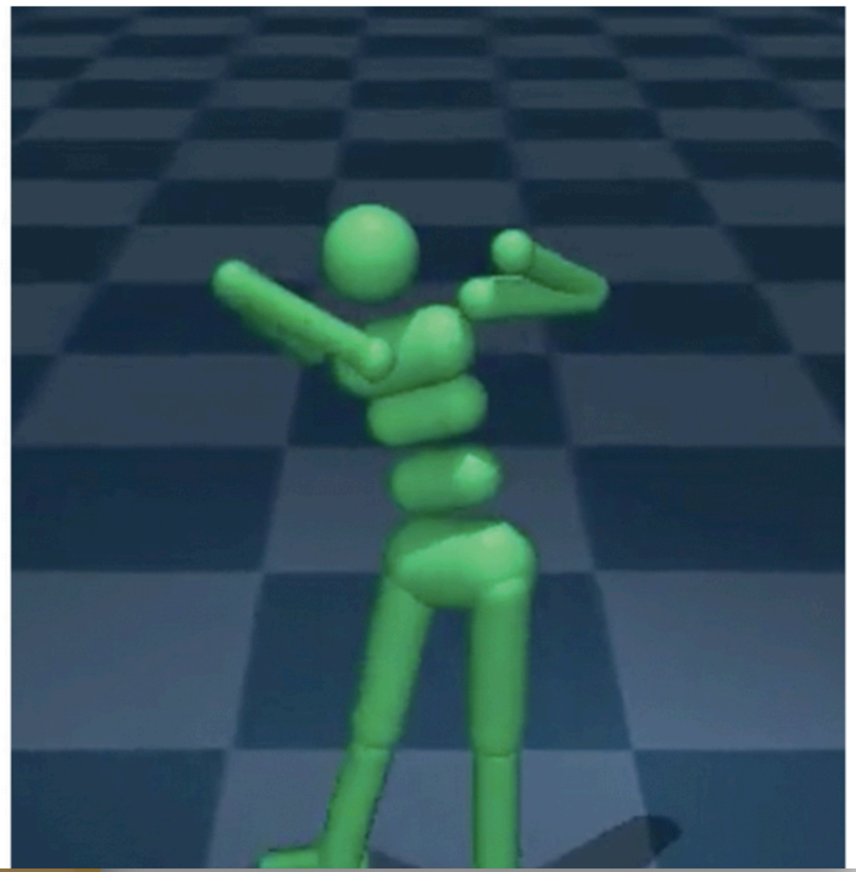
Unpublished code and sensitivity to training conditions make many claims hard to verify

MATTHEW HUTSON Authors Info & Affiliations

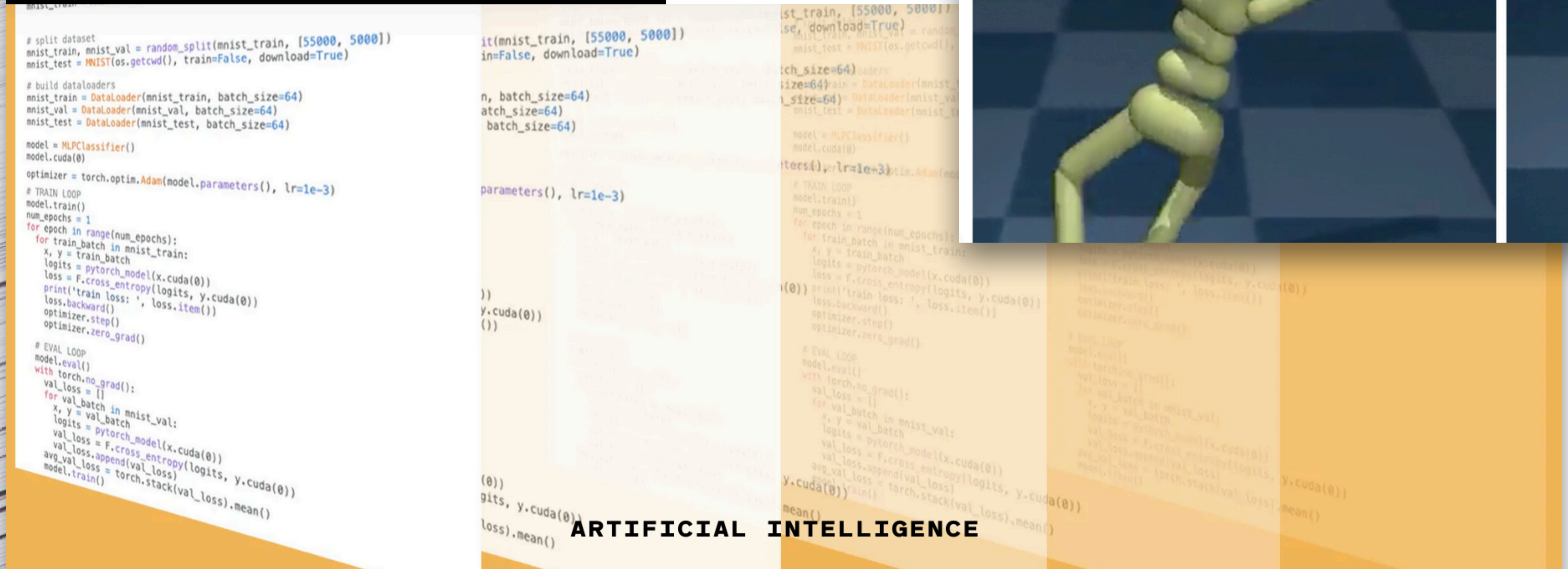
SCIENCE • 16 Feb 2018 • Vol 359, Issue 6377 • pp. 725-726 • DOI: 10.1126/science.359.6377.725

4,648 1





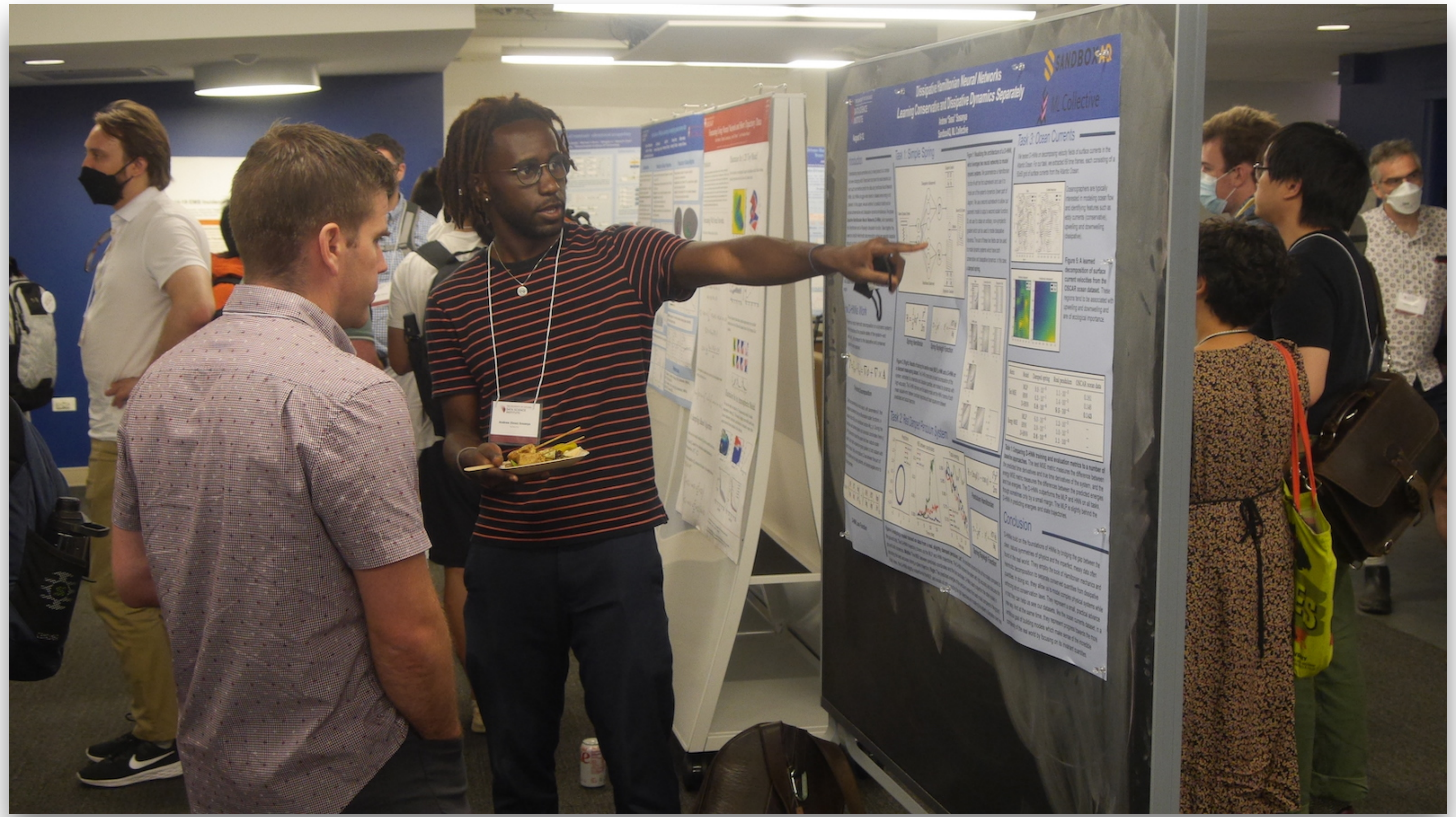
MIT Technology Review



ARTIFICIAL INTELLIGENCE

AI is wrestling with a replication crisis

Tech giants dominate research but the line between real breakthrough and product showcase can be fuzzy. Some scientists have had enough.




EXIT

Training activities (summer schools, workshops, cross-disciplinary collaborations) catalyze ground-breaking research and accelerate workforce development

Long Programs

Programs > Long Programs > Multiscale Geometry and Analysis in High Dimensions

Multiscale Geometry and Analysis in High Dimensions

SEPTEMBER 7 - DECEMBER 17, 2004

ipam

Long Programs

Programs > Long Programs > Multiscale Geometry and Analysis in High Dimensions

Multiscale Geometry and Analysis in High Dimensions

SEPTEMBER 7 - DECEMBER 17, 2004

IMSI

Institute for Mathematical and Statistical Innovation

IFDS

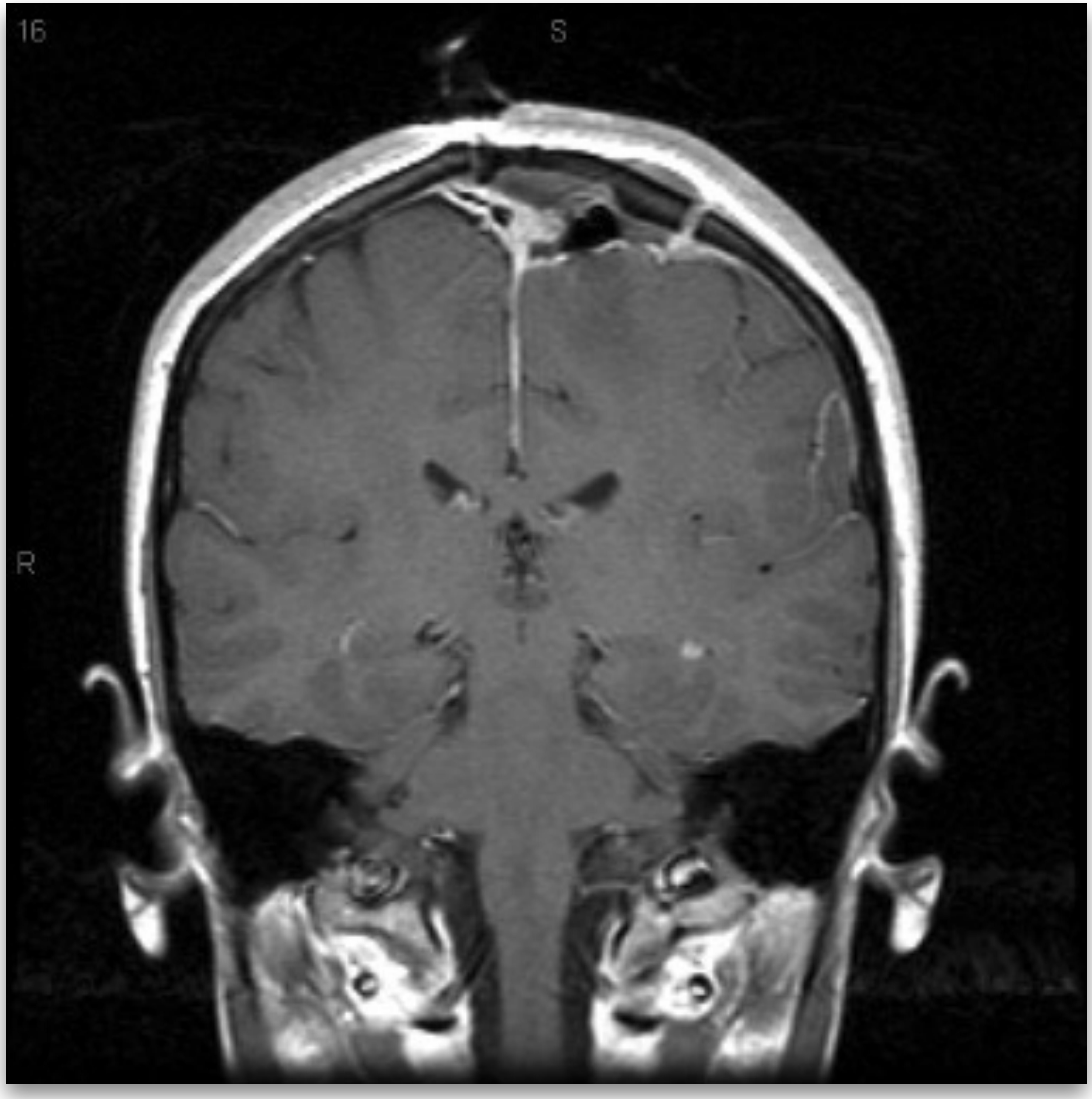
Institute for Foundations of Data Science

IFML

Institute for Foundations of MACHINE LEARNING

National Institute for Theory and Mathematics in Biology


A


16
S
R

Developing **applied machine learning** without understanding **math, stats, & CS foundations** is like developing **biotech** without understanding **biology**.

Thank you!

willett@uchicago.edu