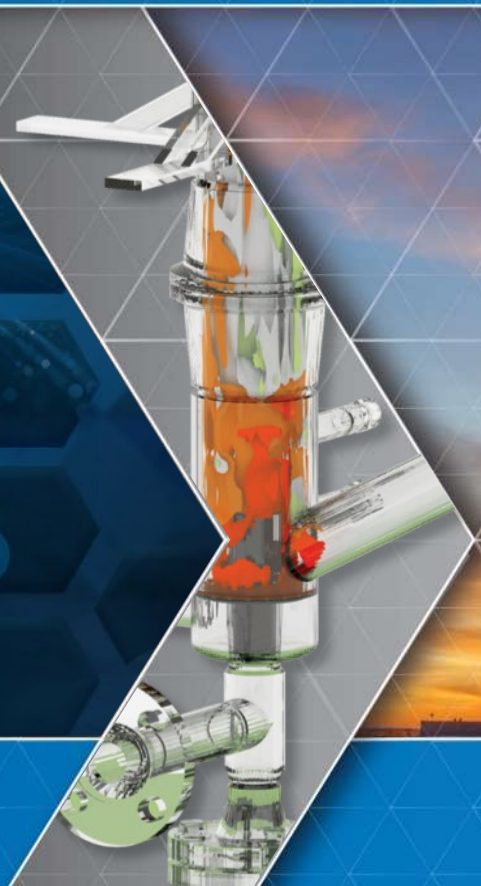
SCIENCE-BASED AI/ML INSTITUTE

An Introduction to NETL's Science-based AI/ML Institute

May 13, 2021



U.S. DEPARTMENT OF ENERGY | NATIONAL ENERGY TECHNOLOGY LABORATORY

NETL's Science-based AI/ML Institute (SAMI)

Applied energy science and physics informed AI/ML solutions

HEADQUARTERED AND LED BY THE NATIONAL ENERGY TECHNOLOGY LABORATORY

Kelly Rose
TECHNICAL DIRECTOR

Jimmy Thornton
CHAMPION

Madhava Syamlal
CHAMPION

Grant Bromhal
CHAMPION

Jonathan Pfautz
SENIOR TECHNICAL ADVISOR

Chung Shih
TECHNICAL LEAD

- SAMI was established in 2020
- Combines the strengths of NETL's energy computational scientists, data scientists, and subject matter experts with strategic partners
- SAMI is innovating solutions to today's energy challenges

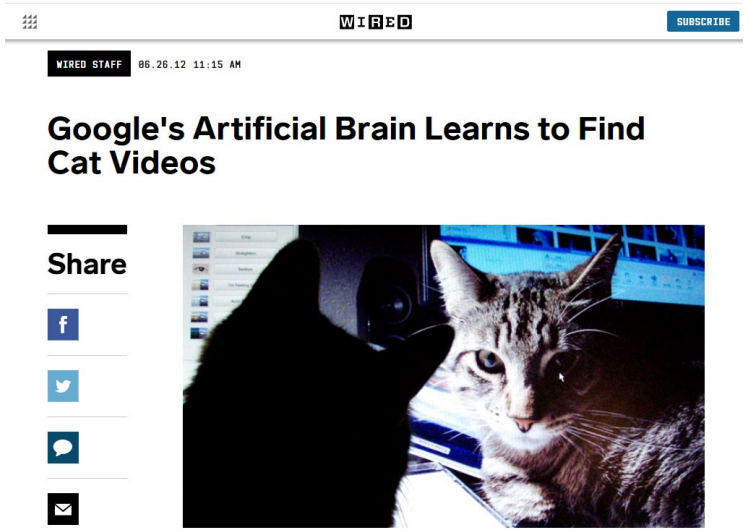
Learn more about SAMI:
https://edx.netl.doe.gov/sami/

Leverage science-based models, AI/ML methods, data analytics, and high-performance computing to accelerate applied technology development for clean, efficient, and affordable energy production and utilization

— SAMi SCIENCE-BASED AI/ML INSTITUTE —

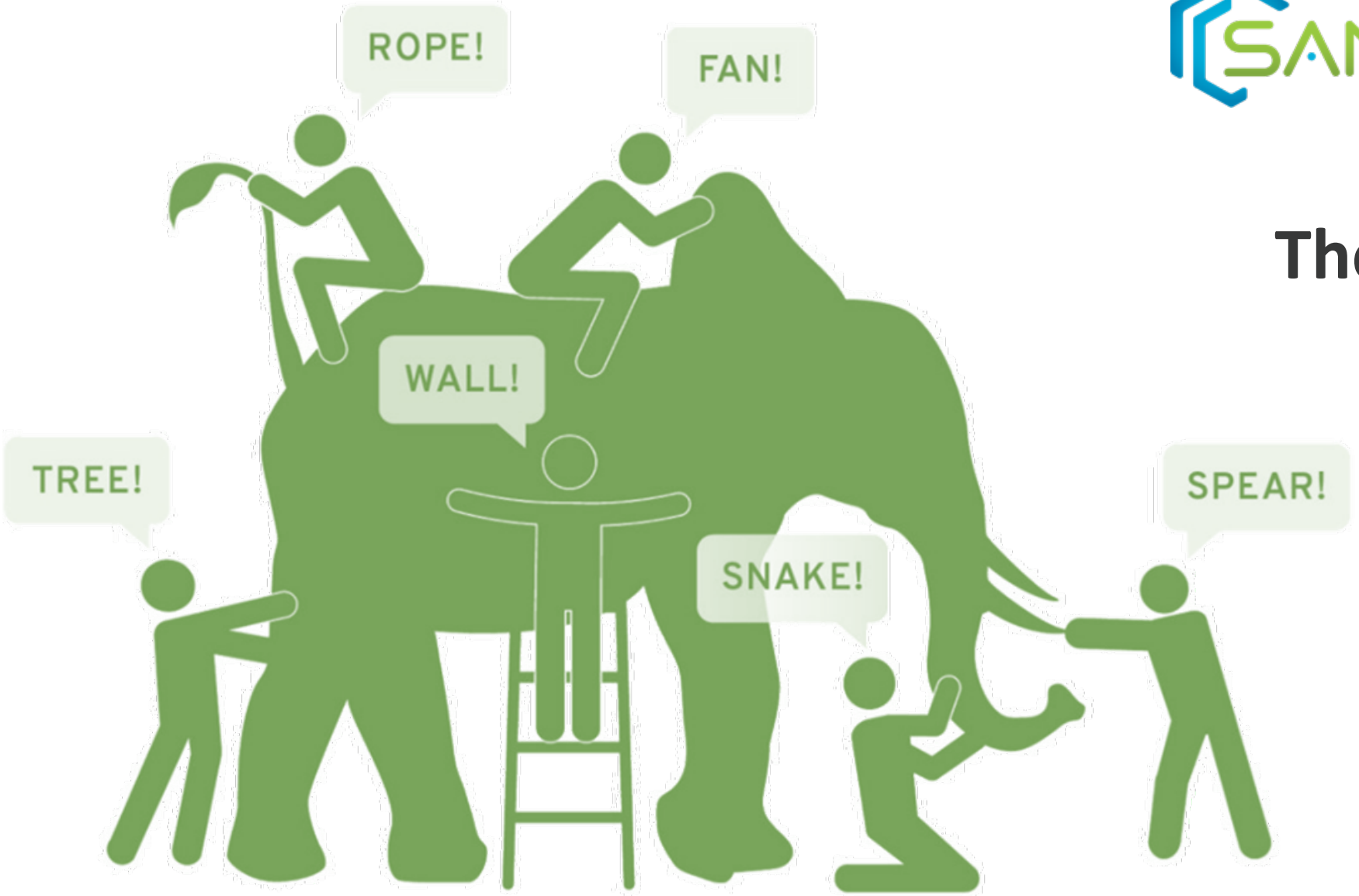
A few definitions…and why AI/ML Now?

- **Artificial Intelligence**
  - "Programmed" intelligence
- **Machine Learning**
  - **Supervised ML**, the machine is trained, taught
  - **Unsupervised ML**, the machine learns on it's own (Google's cat video experiment)
- **Big Data**
  - Large volumes, variety, variability, velocity of data
- **Big Data Computing**
  - Computing engineering & systems to handle big data



JOULE
NETL SUPERCOMPUTER

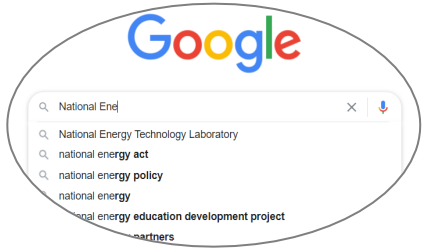
WATT
The Power of AI



The Opportunity

AI/ML offers new ways to explore old problems

AI/ML use in daily life has sky rocketed



Auto-complete

Virtual Assistant

Online Shopping

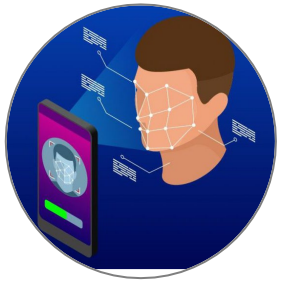


Image Recognition

Social Media Ads

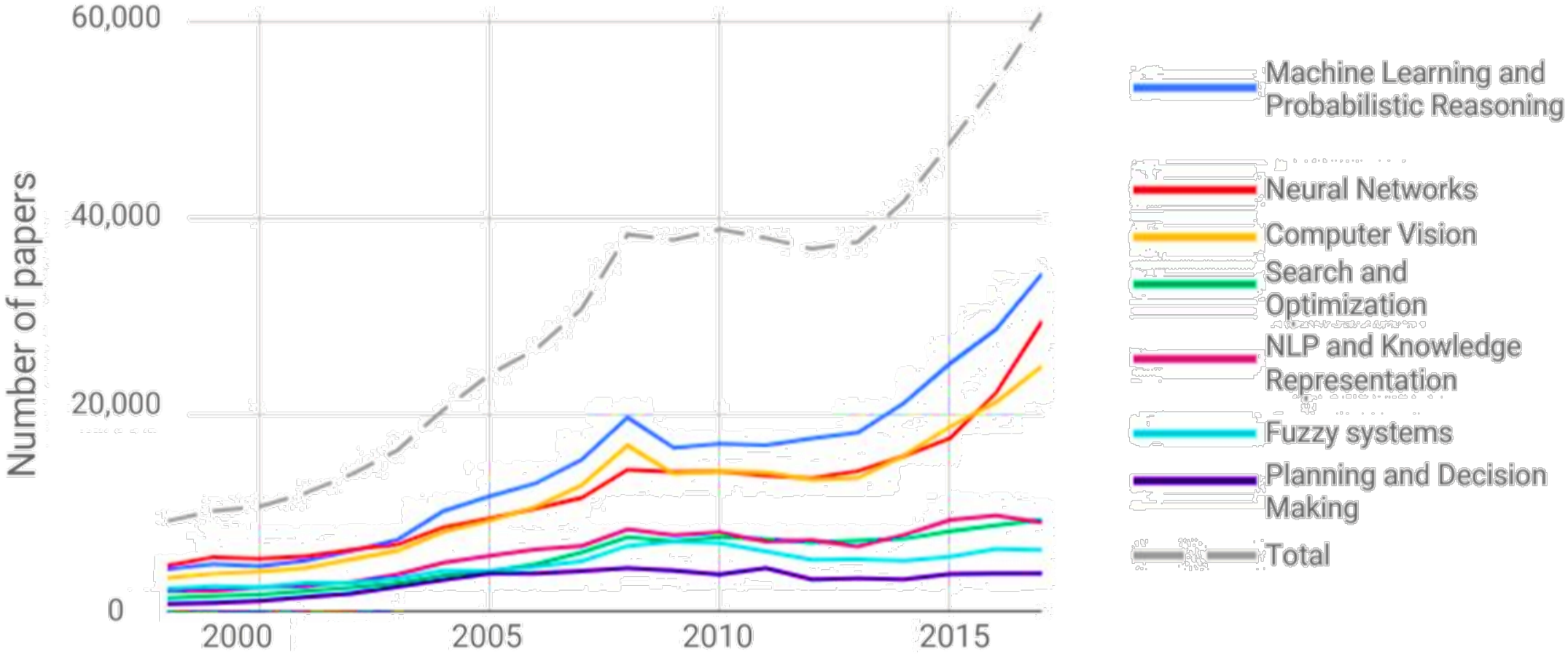
Suggester Models for TV Shows

Driving Route Guidance

Self-driving Cars

AI/ML R&D Breakthroughs are Accelerating Too

Number of AI Papers on Scopus (1998-2017), Source: Elsevier



3 MAJOR CHALLENGES in Applying AI/ML

SCIENTIFIC DISCOVERY IS DEPENDENT ON DATA

Data offers significant opportunities to drive advanced AI/ML breakthroughs, but research is often impeded by data access, quality, formats, gaps, multi-source, multi-scale, and integration challenges.

COMPLEX AI/ML MODELS CAN BE DIFFICULT TO UNDERSTAND

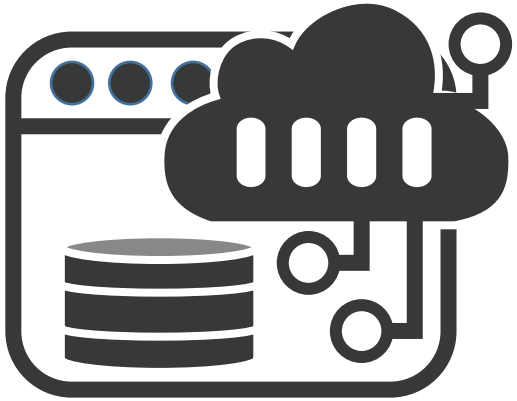
The logic of complex AI/ML, such as deep learning models, can be difficult to understand. It is critical to have explainable and trustworthy AI.

INCONSISTENT PRACTICES MAKE EVALUATION DIFFICULT

Even with training, validation, and testing processes, unnoticed incorrect workflow details can still significantly impact the results.

3-PRONGED APPROACH

Injecting science and physics to guide AI/ML



SMART DATA PLATFORM (EDX++)

R&D data & knowledge refinery with connectivity to computing resources

SCIENCE-BASED AI/ML MODELING

Address previously unanswerable problems

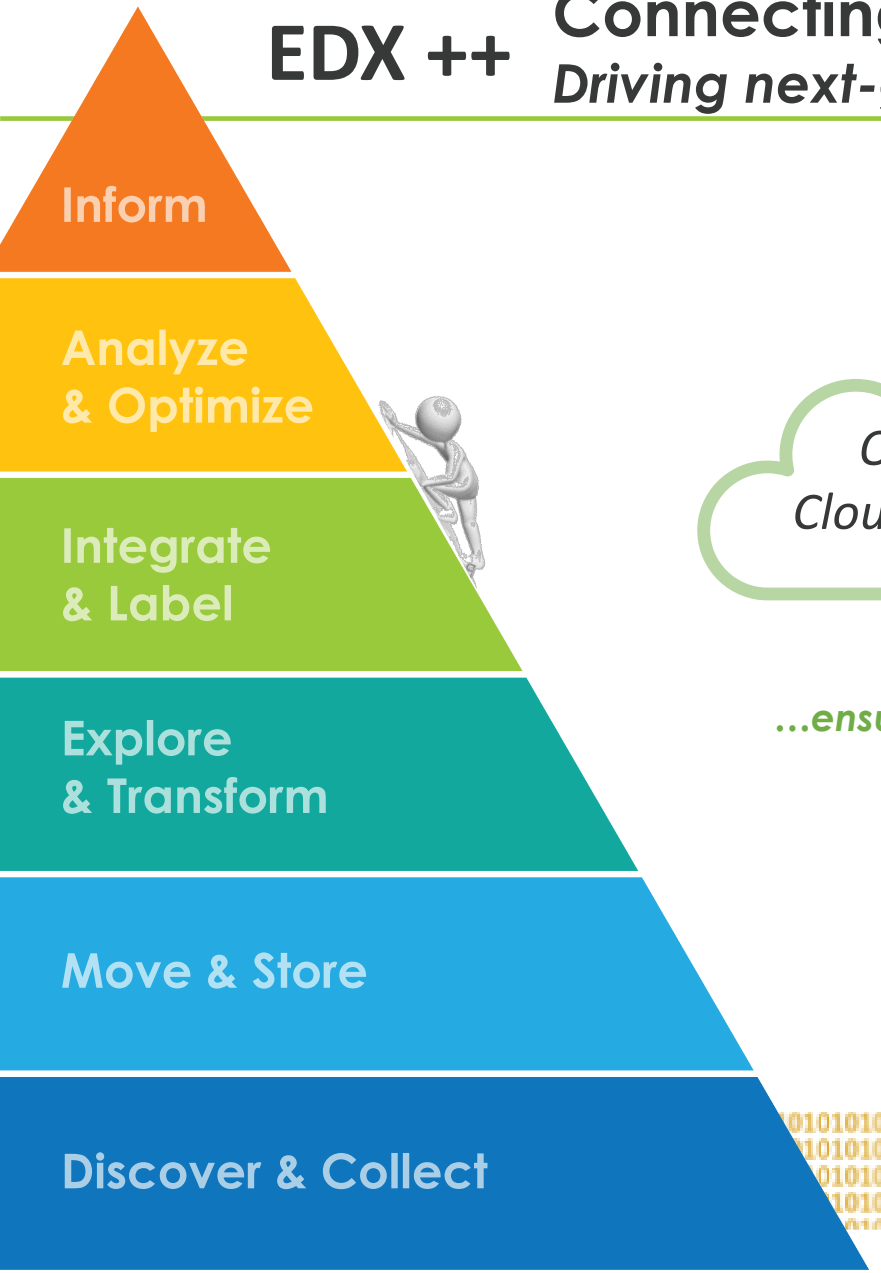
FOUNDATIONAL STANDARDS & PLATFORMS

Establish responsible use of AI/ML

EDX ++ Connecting data with scientific computing

Driving next-gen R&D…

NATIONAL ENERGY TECHNOLOGY LABORATORY



EDX++ FRAMEWORK

Off-Prem Cloud Compute

SAMi

NATIONAL ENERGY TECHNOLOGY LABORATORY

JOULE NETL SUPERCOMPUTER

WATT The Power of Ai

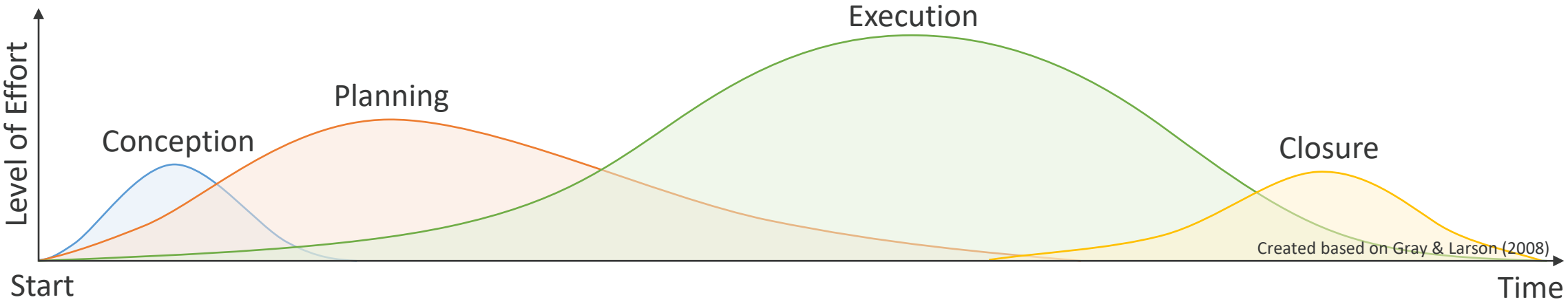
On-Prem Compute

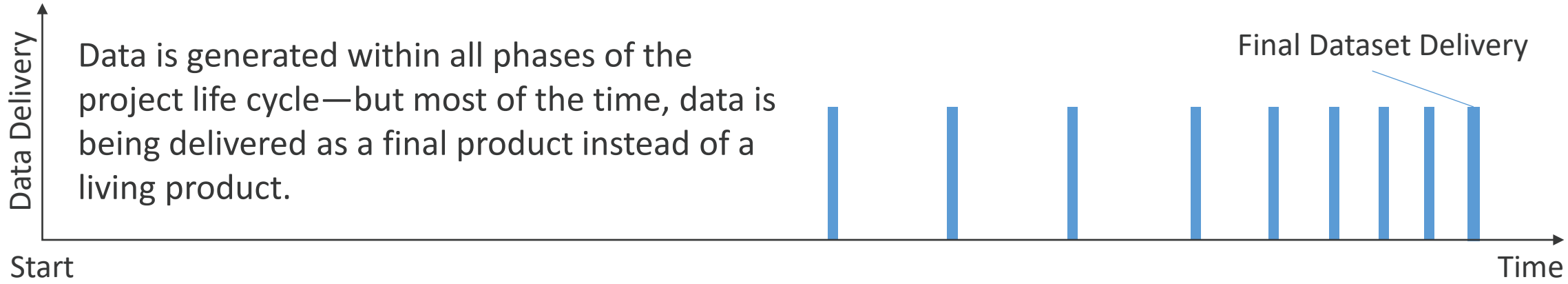
…ensuring compliance with Federal/DOE regulations

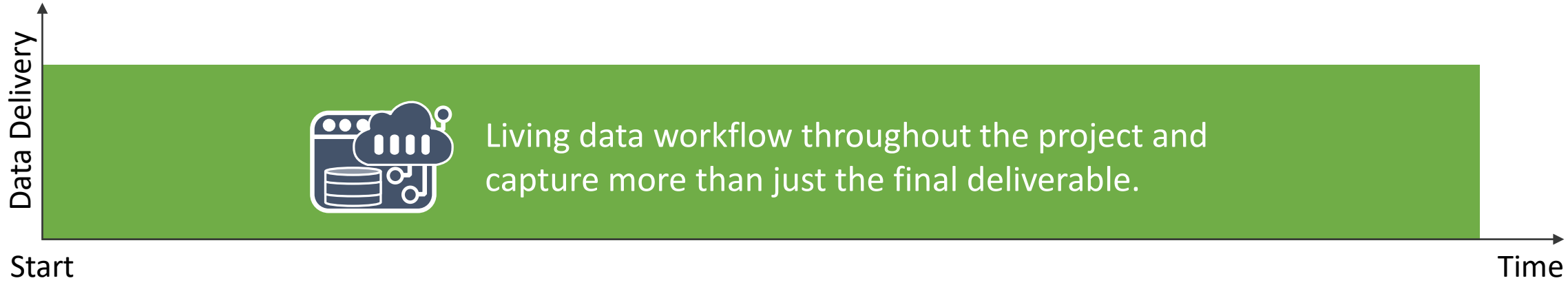
EDX® Energy Data eXchange

…ensuring preservation and access to DOE FE knowledge and data resources

Big Data


Level of Effort
Conception
Planning
Execution
Closure
Created based on Gray & Larson (2008)
Start
Time


Data Delivery
Data is generated within all phases of the project life cycle—but most of the time, data is being delivered as a final product instead of a living product.
Final Dataset Delivery
Start
Time


Data Delivery
Living data workflow throughout the project and capture more than just the final deliverable.
Start
Time

EDX is more than just a data repository

Search

Submissions within the public search on EDX provide access to many forms of information including but not limited to **presentations**, **publications**, **tools**, and **data**.

Sort

Submissions within the public search on EDX can be sorted **spatially**, by **keyword**, and **file format** connecting users to the appropriate data and information quickly and efficiently.

Groups

Submissions within the public search on EDX can be clustered into Groups of related data. Some popular EDX Groups include the **Kimberlina Data Group**, **Appalachian Basin Data Group**, and various **RCSPs**.

Tools

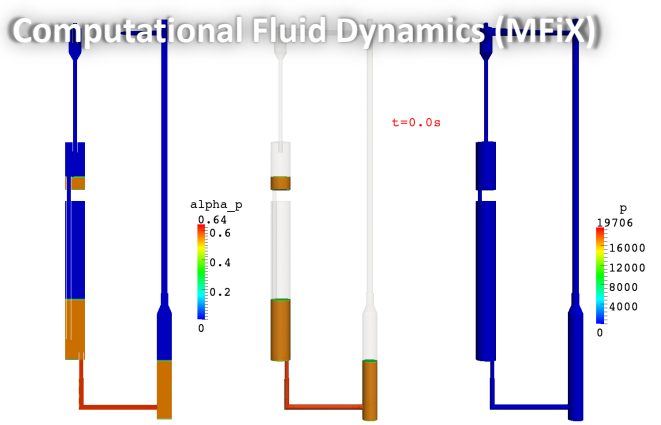
EDX Tools provide access to, management of, and interaction with data through a collection of tools including **SmartSearch**, **Natcarb Viewer 2.0** , **CSIL** and **NRAP Tools**.

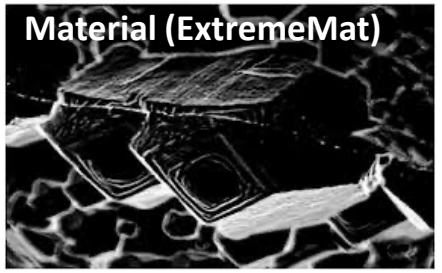
Spatial

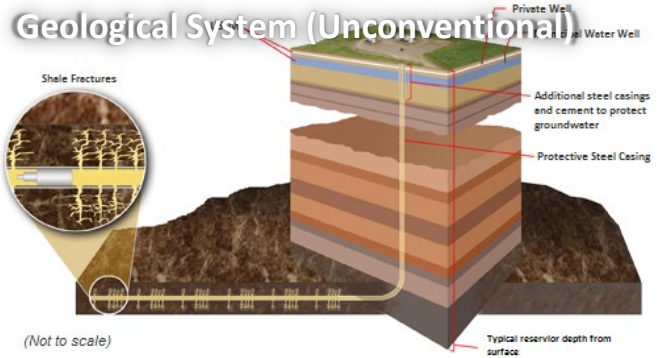
EDX Tools like **Geocube**, **Natcarb Viewer**, and **Blosom** allow users to find, sort, visualize, and interact with geospatial data.

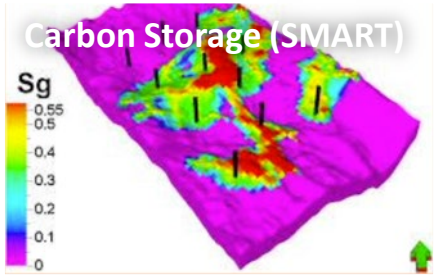
Visualize

EDX Tools provide visualization of data through various tools including **ParaView**, **Papaya**, and **RokData** (coming soon).









There are two ends on the machine learning problem spectrum

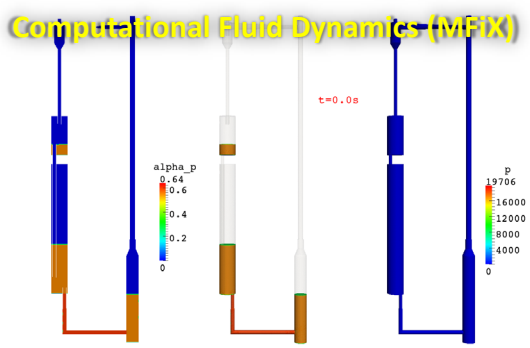
Data Driven

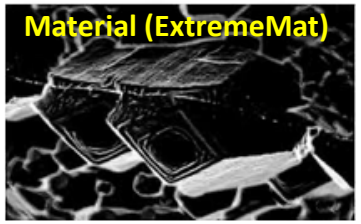
- Object Detection and Classification
- Spam Identification
- Recommendation Systems
- Autonomous Driving
- Natural Language Processing
- More…

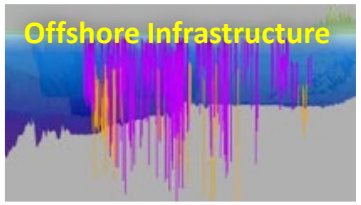
With Scientific and Physics Prior Knowledge

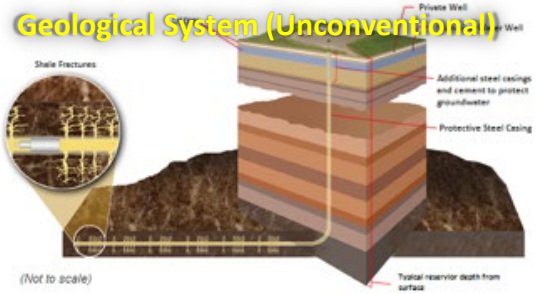
- Computational Science and Engineering
- Computational Fluid Dynamics
- Geological and Environmental Systems
- Material Engineering and Manufacturing
- Energy Conversion Engineering
- More…

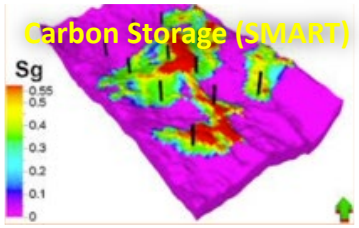
Challenges with this end of the spectrum











With Scientific and Physics Prior Knowledge

- Data acquisition can be costly and time-consuming, resulting in limited available data for AI/ML training
- Complex and high-dimensional space (with many variables)
- Multi-scale tasks (e.g., time, geospatial, atom-to-process)
- Complex AI/ML models can be difficult to understand and thus trust

Science-based AI/ML Modeling

Address previously unanswerable problems

Incorporate Science/Physics into AI/ML

Generate synthetic data to overcome the challenge of limited observation data based on scientific simulations

Enhance AI/ML loss functions with scientific knowledge

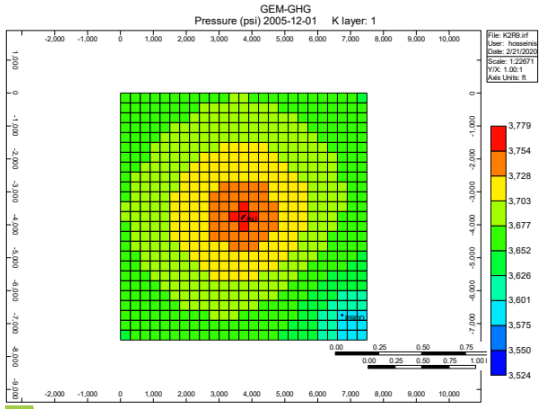
- **Loss = Prediction Data Fitness + Physics Fitness**

Benefits

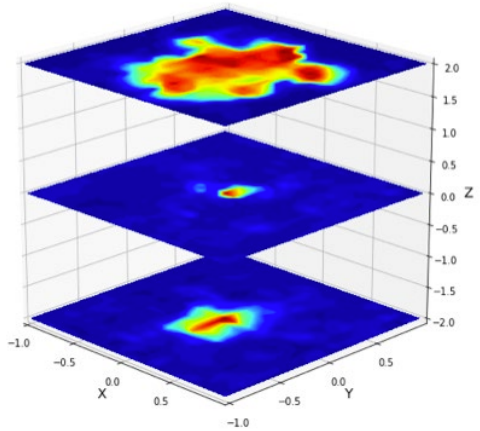
- Use less data to guide model training
- With the same amount of data, improve the overall accuracy
- Increase the confidence with the results

Part of the SMART Initiative to build a learning system to predict CO_2 storage behaviors

Dataset Dev w/ Scientific Simulations



2D Toy Problem



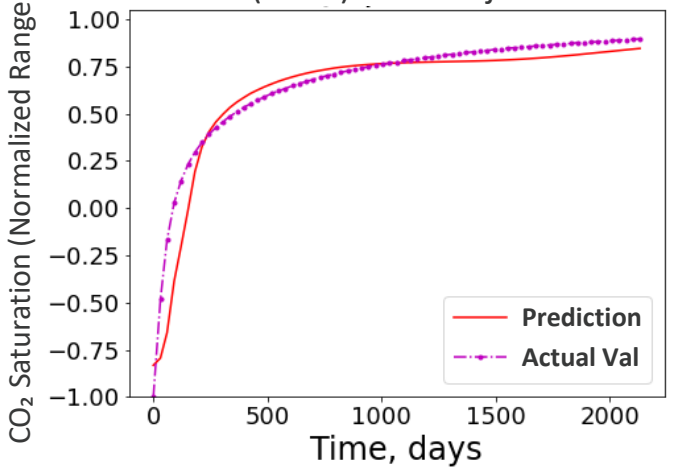
3D Toy Problem

Field Simulation
(CarbonSafe, SACROC, GOM)

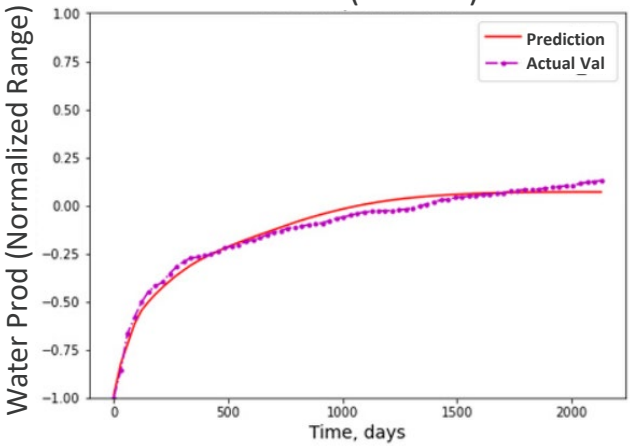
Approach & Preliminary Results

- Designed deep learning models to capture the spatiotemporal changes of CO_2 saturation, pressure, and water production
- Leveraged NETL's High Performance Computers (Joule and Watt)
- Constraints and regularizations are essential
- Started to incorporate physics into models

Test Case (K1R2) at the Injection site

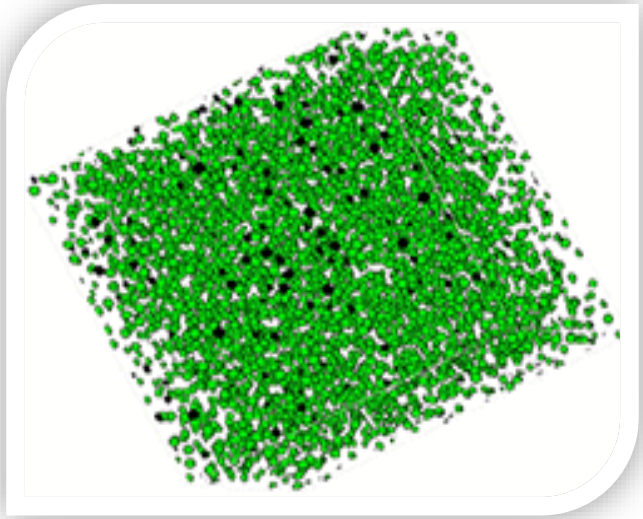


Test Case (K3R5-h)

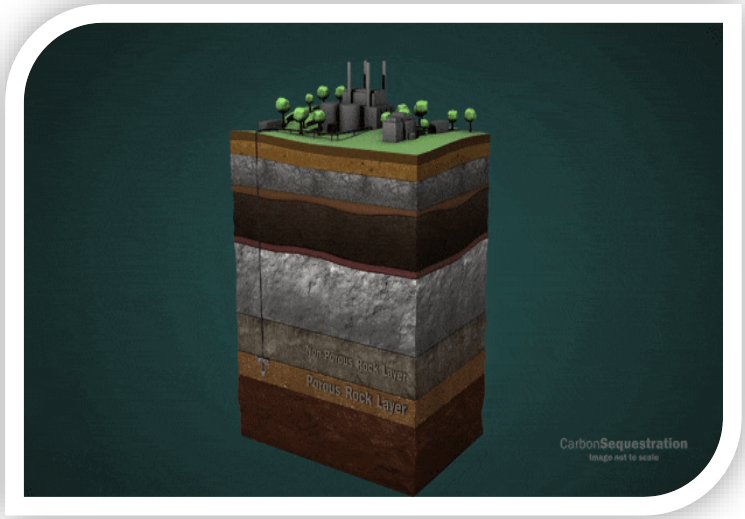


AI/ML Combined with Science-based Models is Enabling Breakthroughs in Applied Energy R&D

Power Plant Resiliency and Reliability



Advanced Materials Discovery

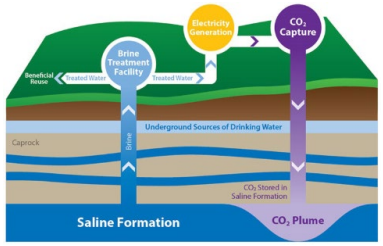


Spatial and Subsurface Analysis, Prediction, and Visualization

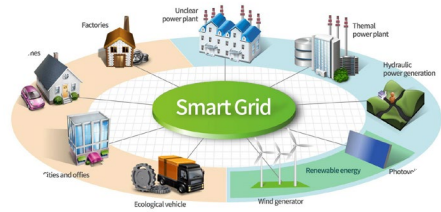
SMART · NETL Carbon Storage · Offshore R&D · eXtremeMAT · NETL SmartSearch · MFS · MFiX · Nodeworks · Computational Science & Engineering · NRAP National Risk Assessment Partnership · REE Rare Earth Elements · CCSI² · IDAES Institute for the Design of Advanced Energy Systems · NETL SmartParse · EDX Energy Data eXchange · Visualization Center · JOULE NETL Supercomputer · WATT The Power of AI

NETL supported with AI/ML capabilities

Power Plant Maintenance & Operations



Predictive Methods for Carbon Storage



Smart Grid Monitoring

Energy Efficient Decision Making

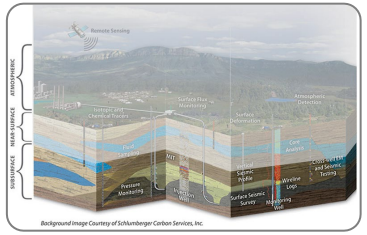
Cybersecurity

Offshore Pipeline and Infrastructure Integrity

Predictive Methods to ID Rare Earth Elements

NETL SmartSearch

Data Discovery, Labeling, Transformation



Prediction of Subsurface Properties and Geohazards

WATT The Power of Ai

AI/ML Scientific Computing

AI/ML must be applied appropriately

BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

Large er
avoidabl
of big da

David Lazer,[1,2*] Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[2,5,6]

In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. *Nature* reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (*1*, *2*). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (*3*, *4*), what lessons can we draw from this error?

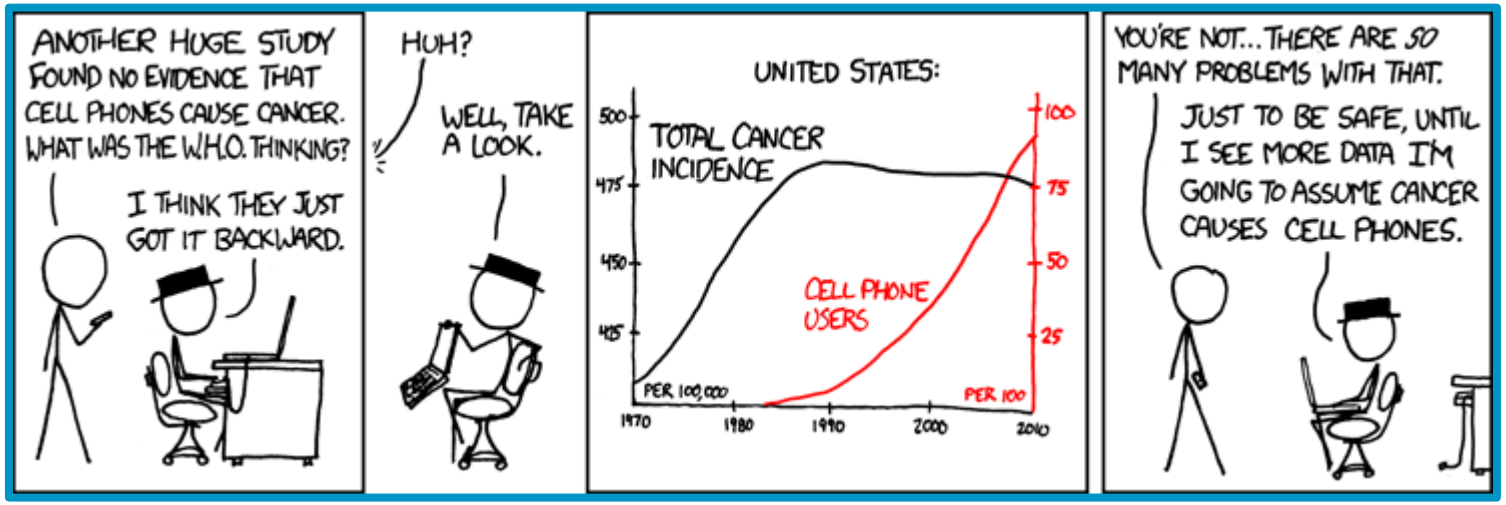
The problems we identify are not limited to GFT. Research on whether search or social media can predict *x* has become commonplace (*5*–*7*) and is often put in sharp contrast with traditional methods and hypotheses

surement and construct validity and reliability and dependencies among data (*12*)

Even
the com

Lazer et al 2014, the parable of google flu: traps in big data analysis, *Science*

- AI/ML methods & technologies must be **appropriate for the goal**
- Results need to be **explainable & validated**



- **Correlation does not equal causation**
  - Just because you have an analysis doesn't mean the results are meaningful
- **Uncertainty is critical**
  - Capture, reduce if possible, represent, utilize, quantify

Accelerating FE R&D with AI and ML

FECM formed an internal AI Guidance Team* to examine existing gaps in AI capabilities and untapped opportunities for AI to broadly accelerate FECM research.

*Team included staff and PMs from FE HQ, NETL, LLNL, and LANL.

SAMi CROSSCUTTING BENEFITS

COMMUNICATE

Share the latest AI **news** and **breakthroughs**

CREATE

Design critical foundational **models, platforms, and standards** to be applied to various fields

CATALYZE

Act as a **hub** with **AI/ML experts** to help **SME's** R&D
Enhance work force **AI/ML fluency**

CONNECT

Coordination with **partnerships**, **IT**, **communication**, and others to benefit the SAMI community

Catalyzing Opportunities & Recognition

https://edx.netl.doe.gov/sami

S A M I SUCCESSES

- **EDX selected as the first DOE geospatial data repo by OCIO**, elevating EDX and SAMI brand
- **Development of hi-fidelity 3D real-time models** for improving prognostics and diagnostics in advanced power system with Cerebras
- **Join the Watt User Group meetings** and virtual (EDX-based) Watt community https://edx.netl.doe.gov/workspace/dashboard/watt-users
- **SMART Support** – Providing HPC lesson-learned and information, software engineering, and AI/ML capabilities
- **Workshops and Trainings:** AI4CM, NVIDIA bootcamp; **Summer Seminar Series "How NETL is using AI/ML to…"**
- **Supporting proposal teams** for AITO, DOE SC (BER), EERE, ARPA-e, and others (working with intramural and extramural researchers and partners)
- **Funding AI/ML IT for** (1) Supporting Watt enhancements, (2) AI computing expert to work with R&D teams, (3) Get WATT ready for hosting services (e.g., Jupyter Notebook)

HOW CAN YOU CONNECT?

JOINING SAMI IS STRAIGHTFORWARD

Contact Us: SAMI@netl.doe.gov

Learn more about SAMI:
https://edx.netl.doe.gov/sami/