Crash Lecture into AI's Research Opportunities and Challenges

Hany Abdel-Khalik,
CYNICS Research Group, Purdue University

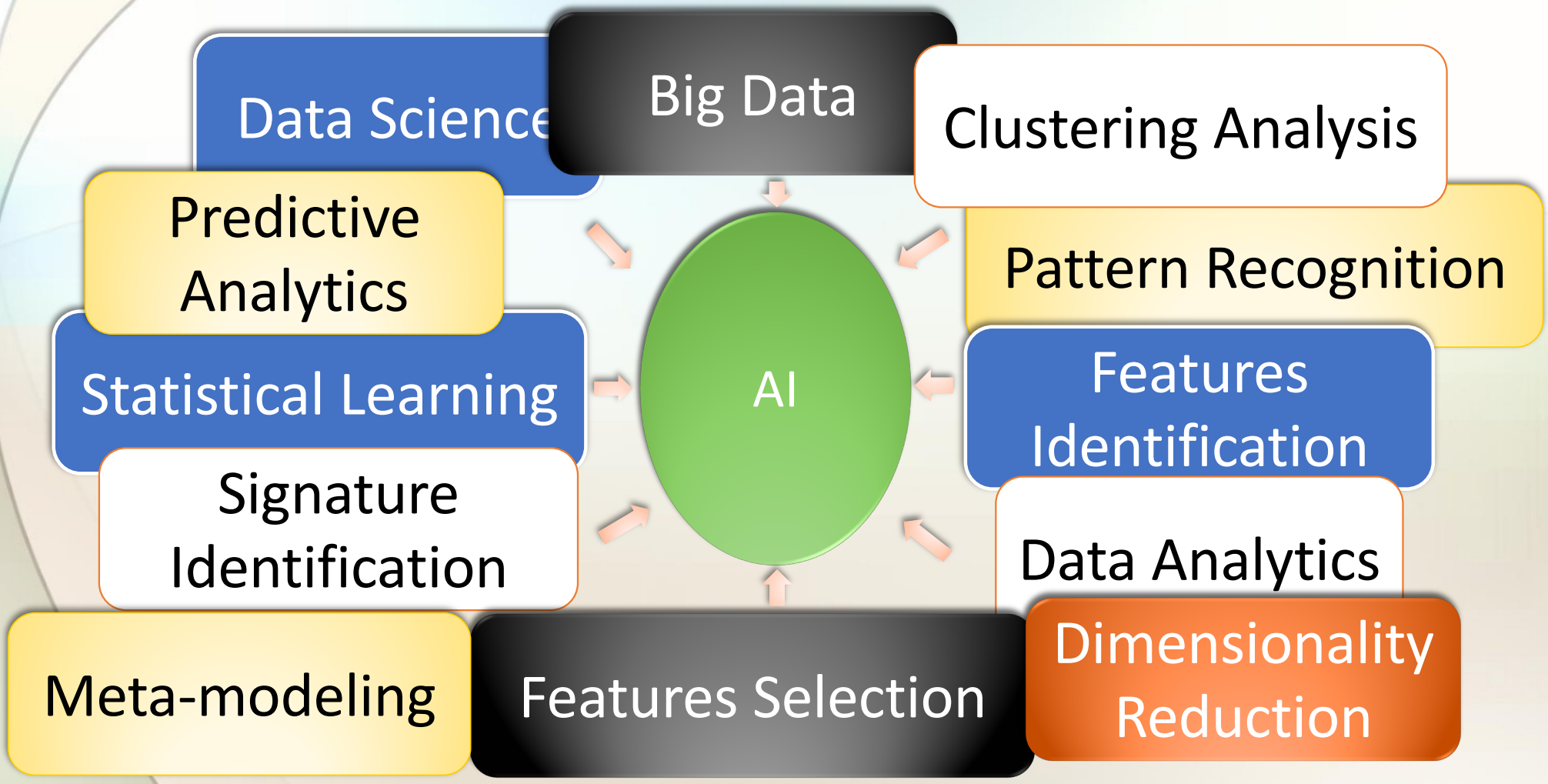
IAEA-ICTP School, Trieste Italy, May 25th, 2022

What's Artificial Intelligence (AI)?

- Computer science (CS) techniques enabling machines to complete tasks that usually requires human intelligence
- **Automating Human-based Intelligence** or
- **Artificial Emulation of Human Intelligence**

VS

Other commonly-related terms



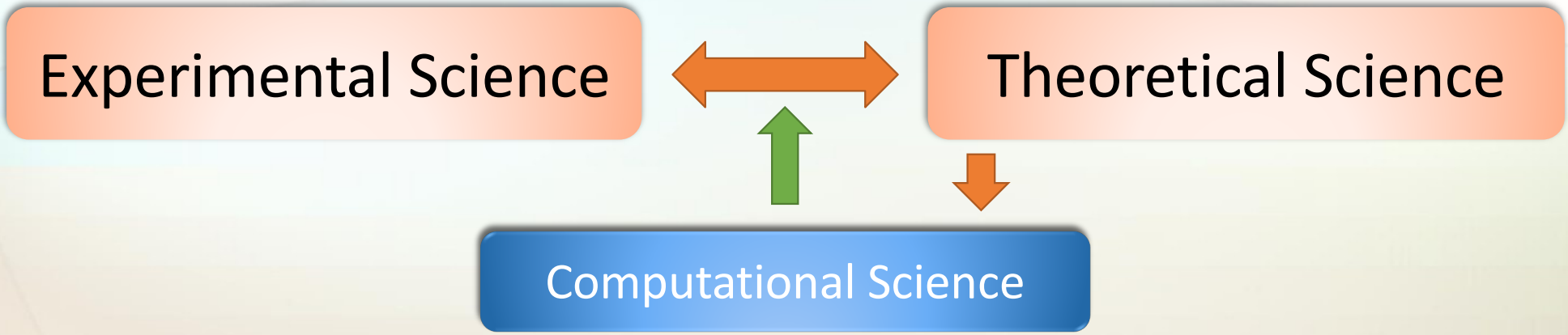
Motivation

- Use of AI techniques has increased significantly in past decade, with several industries reaping rewards for **complex** engineering systems in terms of improved economy, reliability, and safety.
- Material is extremely mathematical, and fairly complicated, forcing practitioners to treat AI techniques as **magic tool-boxes**, sometimes proposed as **panacea** for any problem containing lots of data.
- AI techniques performance can be hard to understand/predict; they sometimes fail abruptly and unexpectedly
- Interestingly, new line of research emerged to predict failure, and even how to cause it.
- For complex systems, off-the-shelve AI techniques can be challenged, and may require tuning by domain experts to get them to work.

AI– Teaching Challenges

- A comprehensive introduction requires sophisticated background in information theory, functional analysis, linear algebra, random matrix theory, advanced statistics, regression techniques, calculus of variations, ..
- Topics of close proximity to AI techniques, e.g., sensitivity analysis, inference techniques, etc., require advanced knowledge in numerical analysis, calculus of variations, probability theory.
- Course work can easily justify a four-years undergraduate degree, which some universities are attempting to implement.
- Given its importance to many engineering fields, this massive knowledge must be data-mined somehow to compress it in a form suitable for introductory course that maximizes student's inference on field of AI state-of-the-art, recent advances, and challenges

Pillars of Scientific Learning



Data Science?

Pillars of Scientific Learning

- ES and TS represent conventional pillars which are responsible for most of discoveries until late 20s century.
- CS emerged in late 20th century with rapid increase in computer power acting as acts as bridge between TS and ES to optimally utilize knowledge to propose updates to theory, better design of experiments, and improved design of associated engineering systems.
- Computational scientists rely on a number of tools, including best-available theories, codes, calculus, statistics, linear algebra, probability theory, geometry, etc., to achieve their goals.
- Struggled in beginning for recognition by traditional scientists, but now considered third pillar of science.

Data Science

- Depending on who you ask, data science is viewed anywhere between basic statistical learning, and a revolutionary leap towards scientific learning.
- It employs statistics, mathematics, computer science to improve all three pillars and their communications, e.g.,
  - ES - Improved experiments – what's optimum design of experiment?
  - TS - Improved theory – can computer guess what are missing physics?
  - CS - Improved models in terms of better accuracy, efficiency, and with credible measures of confidence.
  - ES/TS – Is experiment really relevant to physics?
  - CS/TS – Are models capturing dominant physics?
  - CS/ES – Are experiments providing new value to improve predictions?

AI-Empowered CS

- Ultimate goal is to enable predictions
  - with demonstrated accuracy:
    - models can predict past observed performance to small tolerance
    - "requires intelligent recall of past knowledge and development of new theories"
  - with defendable confidence
    - Confidence: ability to estimate errors/uncertainties
    - Defendable confidence: ability to develop rigorous mathematical arguments to support the noted confidence
    - *"requires ability to describe the unknown"*
  - with efficiency:
    - Cost (in terms of computing needs, e.g., runtime, computing power, storage requirements) must be affordable for routine engineering evaluations
    - *"requires intelligent ways to calculate things in simpler manners"*

Is it a Science?

- “Data science has become a fourth approach to scientific discovery, in addition to experimentation, modeling, and computation,” said Provost Martha Pollack (University of Michigan, 2015)
- DSI website states: “This coupling of scientific discovery and practice involves the collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and interdisciplinary applications.”

- Reference: “50 years of Data Science,” David Donoho, MIT, 2015

Is it a Science?

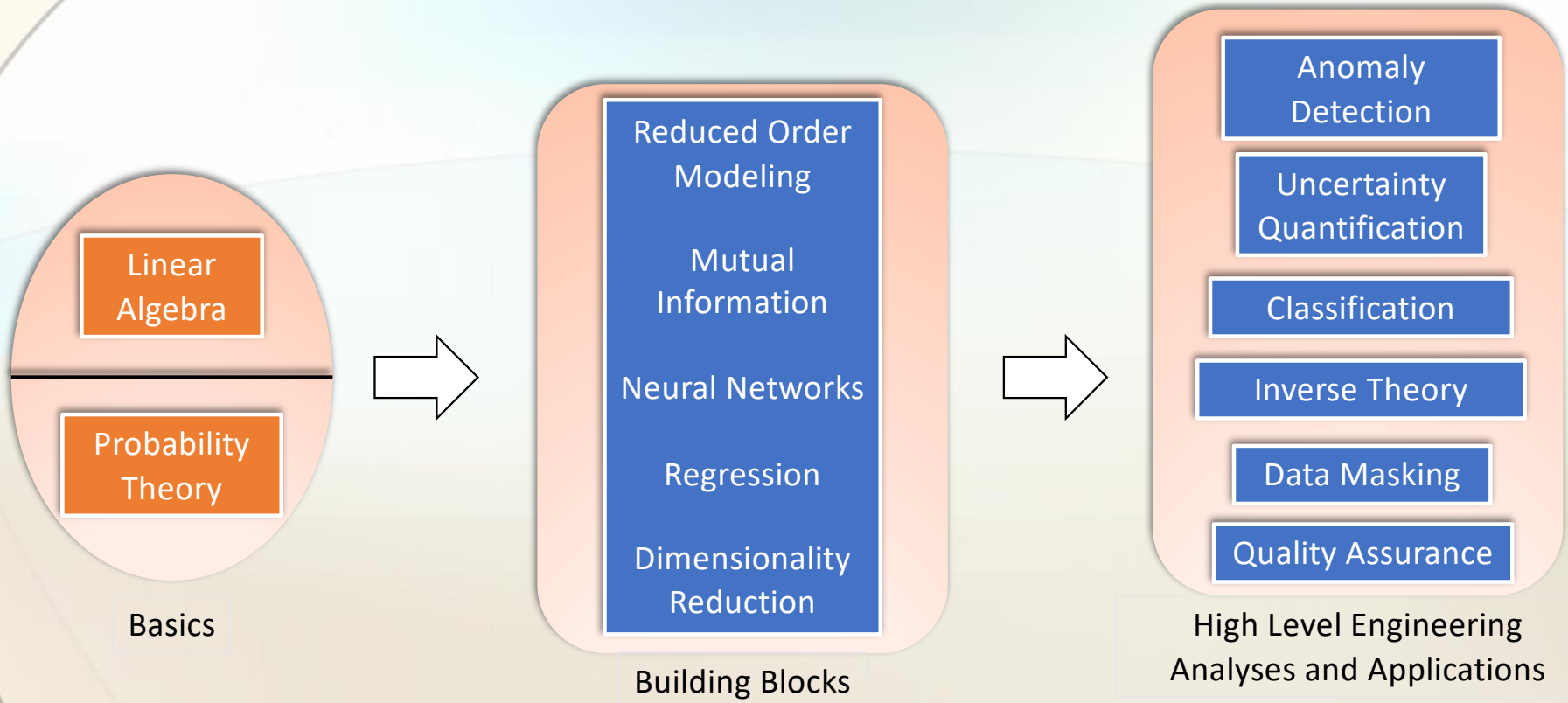
- Data Science is statistics. When physicists do mathematics, they don't say they're doing number science. They're doing math. If you're analyzing data, you're doing statistics. You can call it data science or informatics or analytics or whatever, but it's still statistics. ... You may not like what some statisticians do. You may feel they don't share your values. They may embarrass you. But that shouldn't lead us to abandon the term ''statistics''. Karl Broman, Univ. Wisconsin

- Reference: "50 years of Data Science," David Donoho, MIT, 2015

My own opinion on AI & DS

- As we will learn in this course, concept of “independence” is very critical to scientific thinking.
- Science always tries to break/explain any difficult phenomena into basic “irreducible” components, such that any new observation can be described using these components
  - Think of atoms, elements, basic blocks of matter; humans, family, basic blocks in society; etc.
- DS hybridizes all knowledge acquired by TS, CS, and ES, mathematics, statistics, etc. It is perfectly described using components from other sciences, i.e., it contains no new components, it only optimizes how we use everything we know.
- For DS to become its own science, it must introduce NEW/INDEPENDENT component such as human-like reasoning combining both emotions and logic, intelligence, ability to abstract, etc.

Background Material

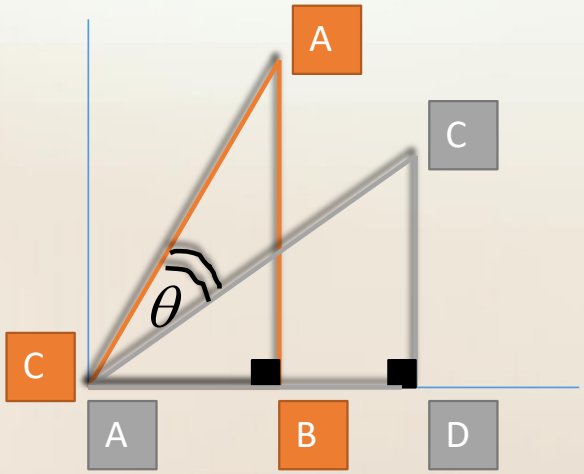


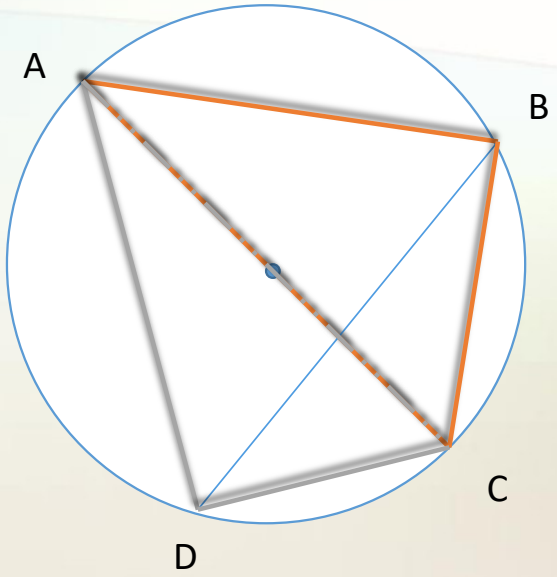
Basics: Linear Algebra

- Ptolemy's theorem ~ 100 A.D.

$$AB \,.\, DC + BC \,.\, AD = AC \,.\, DB$$

$$\cos\theta = AB \,.\, DC + BC \,.\, AD$$

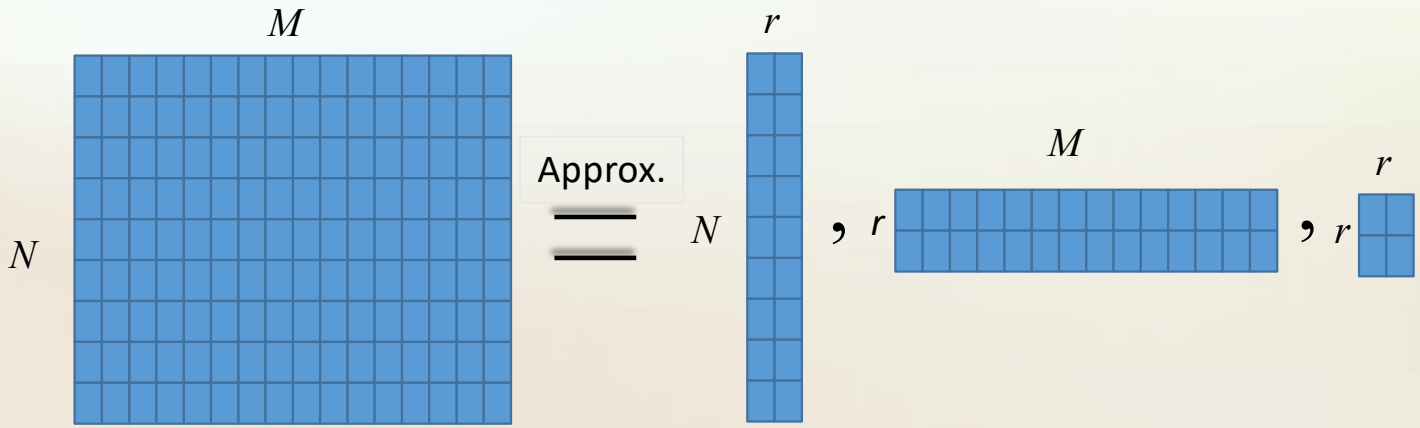




AC is oriented in two different ways, using two DOFs.
How similar are these DOFs?
Concept of inner product provides measure of similarity

Value of Vector Representation

- Finding patterns in large amounts of data
- Data can be represented by M vectors, each vector of length N, then reduced/compressed to r DOFs as follows:



Value of Vector Representation

- Matrix-operators can help identify cause-effect relationships by identifying four fundamental subspaces for given operator $\mathbf{A}$
  - **Range of $\mathbf{A}$**, describes effects
  - **Range of $\mathbf{A}^T$**, descries causes (associated with effects)
  - **Null space of $\mathbf{A}$**, describes non-causes (cannot affect changes)
  - **Null space of $\mathbf{A}^T$**, describes non-effects (effects unexplained by causes)
  - Fundamental Theorem of Linear Algebra

$$R(\mathbf{A}) \oplus N(\mathbf{A}^T) = {}^{\sim M} \text{ and } R(\mathbf{A}^T) \oplus N(\mathbf{A}) = {}^{\sim N}$$

$\sim M$ Space of possible causes

$\sim N$ Space of possible effects

Basics: Probability Theory

- About 300 years ago, scientists started to seriously think about how to reason about uncertain events.

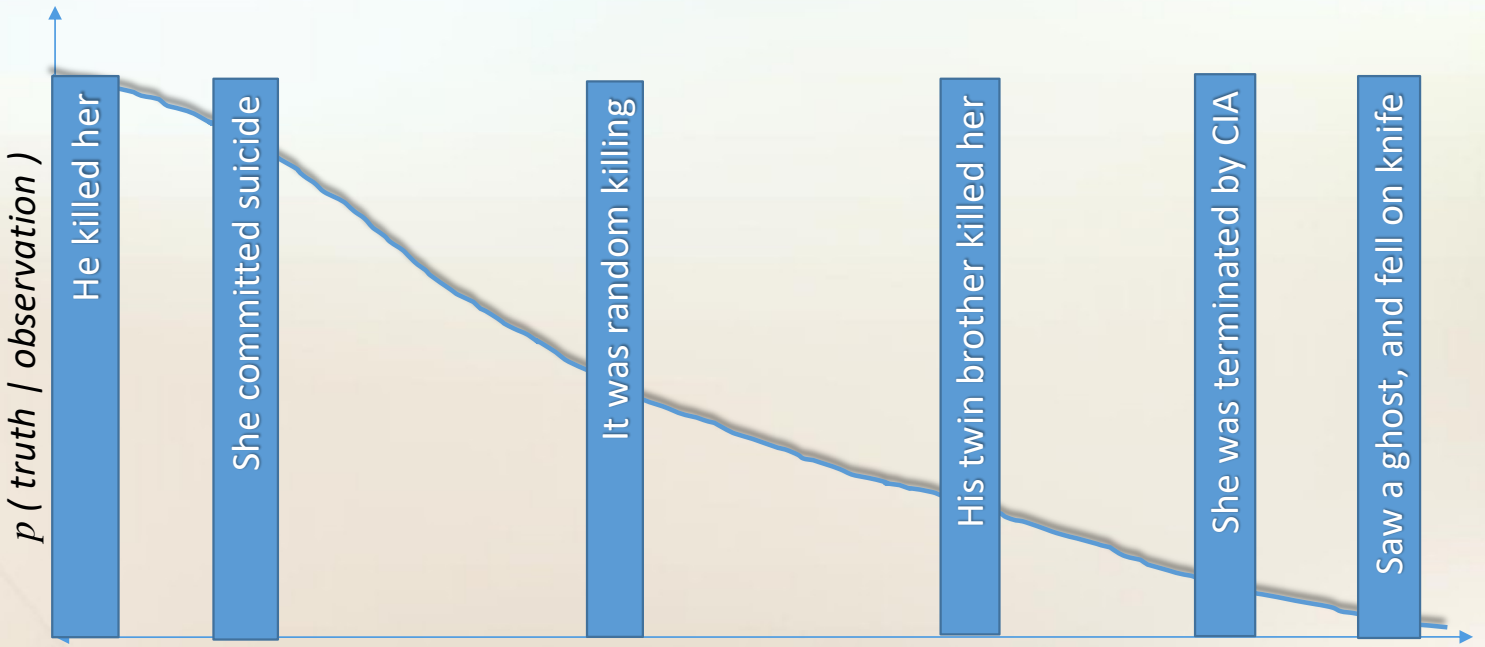
- Deterministic statements:
  - The gravitational force on given object is equal to 3.0 N (unit for force)
  - Particle is travelling at 3 m/s
- Probabilistic statements:
  - There is 90% chance it will rain tomorrow
  - We are 95% confident that DNA found on victim matches that of defendant.

Need for Probabilities

- You don't really understand something unless you can describe it with numbers
- Probabilities have been used to describe two entirely different concepts, yet both called 'probability':
  - Frequency/chance of occurrence:
    - If you toss a fair coin, there is 50% chance you will get heads and 50% tails.
    - If you test negative, there is 5% chance you have the illness
  - Degree of belief:
    - Given a coin of unknown origin; you tossed it twice and got tails, what is the probability you get head/tail on the third toss?
    - I measured the thermal conductivity k 100 times, I believe the probability of finding the true value for k between k_1 and k_2 is 90%.

What is Likelihood? - Social Example

- You walk into your neighbor's house, and find her dead with husband's leaning over her and holding the knife and crying
- What happened? What is the truth given what was observed?
- Each new observation adds a myriad of possibilities (with probabilities)



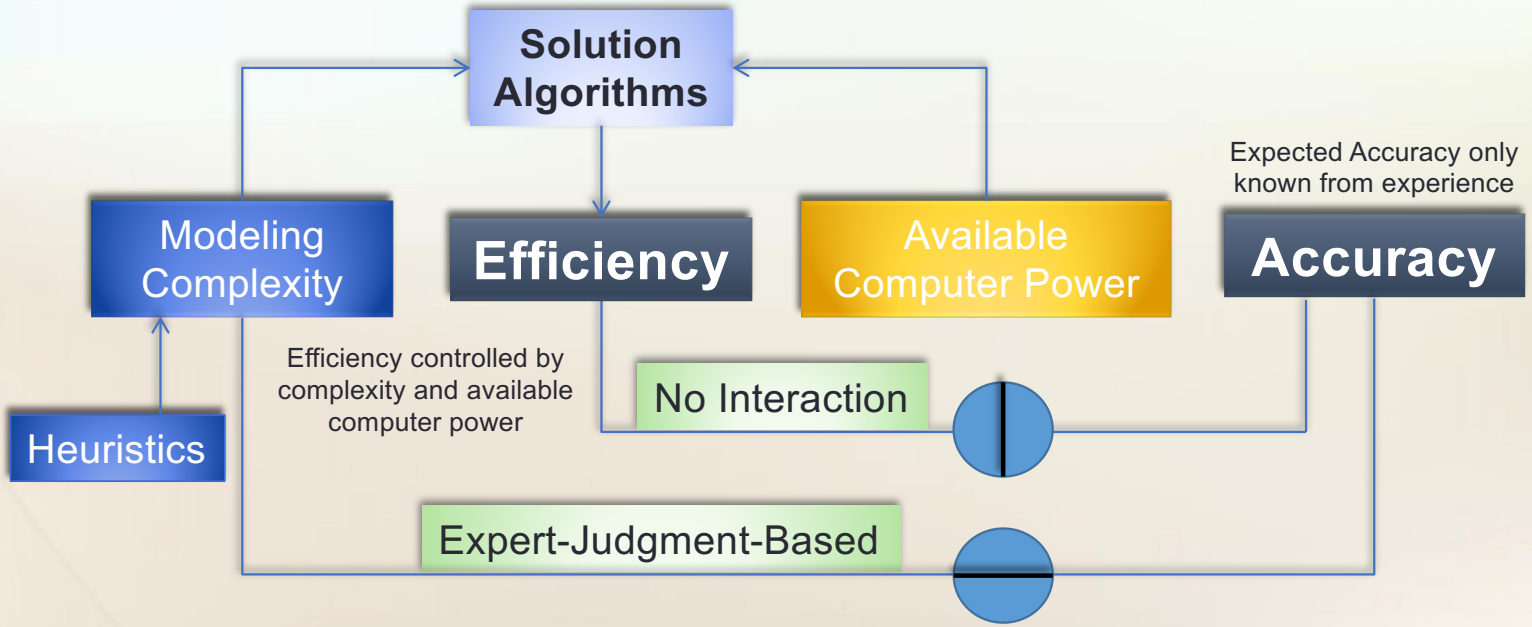
What is ROM?

- Any algorithm/approach/process that reduces cost of simulation is a form of ROM
  - Using coarse mesh instead of fine mesh
  - Using simplified physics instead of first principles physics
  - Using loose physics coupling instead of tight physics coupling.
- Reduction is essential to render engineering calculations practical; but introduces errors, that are difficult to quantify
- Research focuses on rendering practical construction of ROM models and quantification of reduction errors
- Why cannot models be born reduced? “open research question”

Reduction Philosophy

In modeling any natural or man-made system, one is often thinking about two issues: accuracy and efficiency.

How much one willing to pay for accuracy?



$[accuracy, efficiency] = f(complexity)$

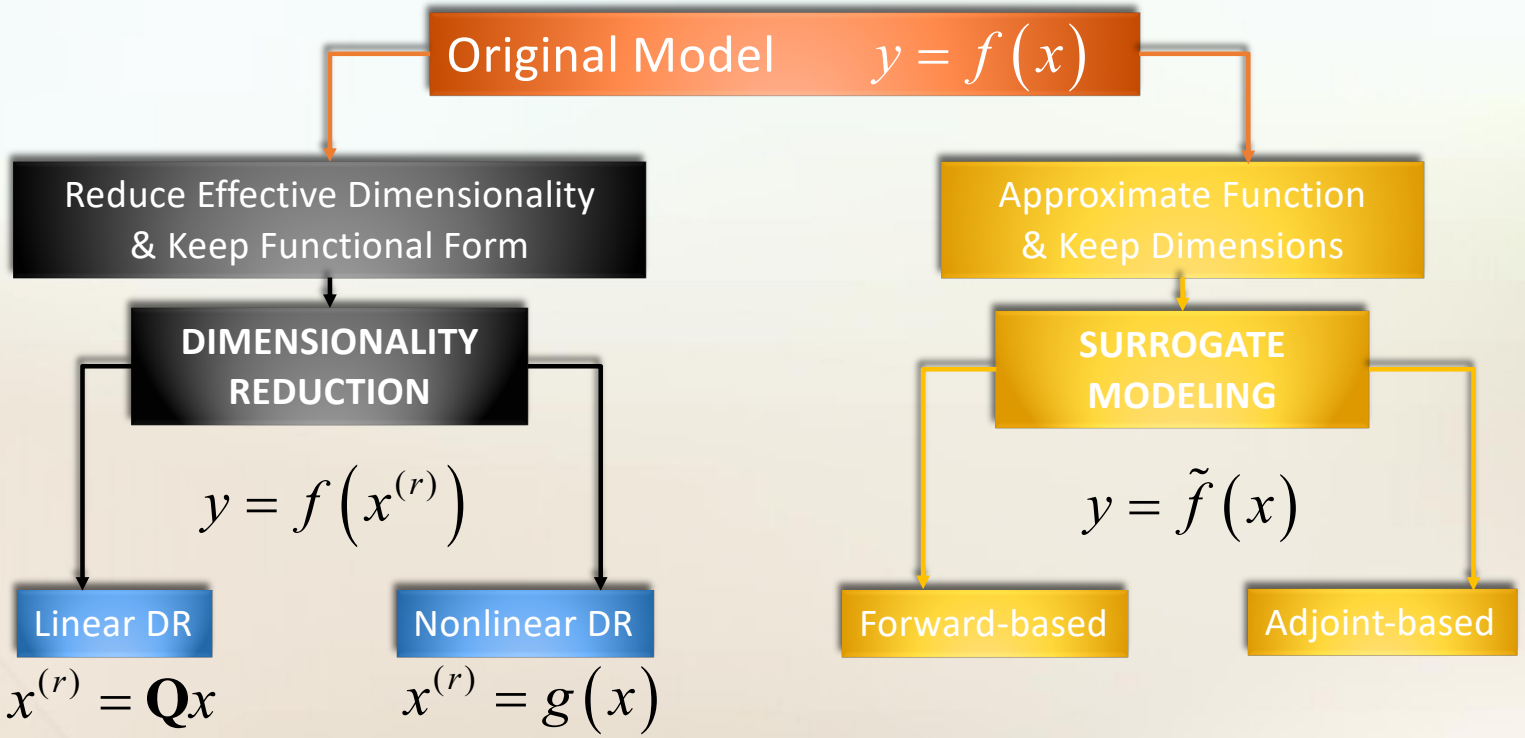
- Complexity is increased heuristically based on modeler's intuition until acceptable accuracy is reached
- Heuristics: increasing mesh size, order of numerical error, order of expansion, etc.
- Given complexity determined by modeler, efficiency is improved employing better iterative techniques, and powerful computers

Modeling heuristics grossly overestimate required degree of complexity

Measured by DOFs used to describe the model

Reduced Order Modeling

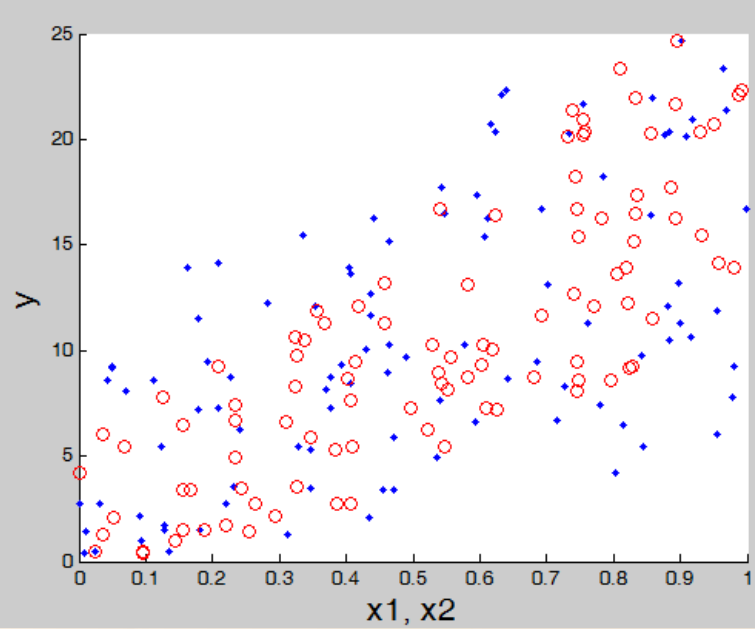
- Any attempt to reduce complexity (i.e., order) of model to minimize execution time for UC application

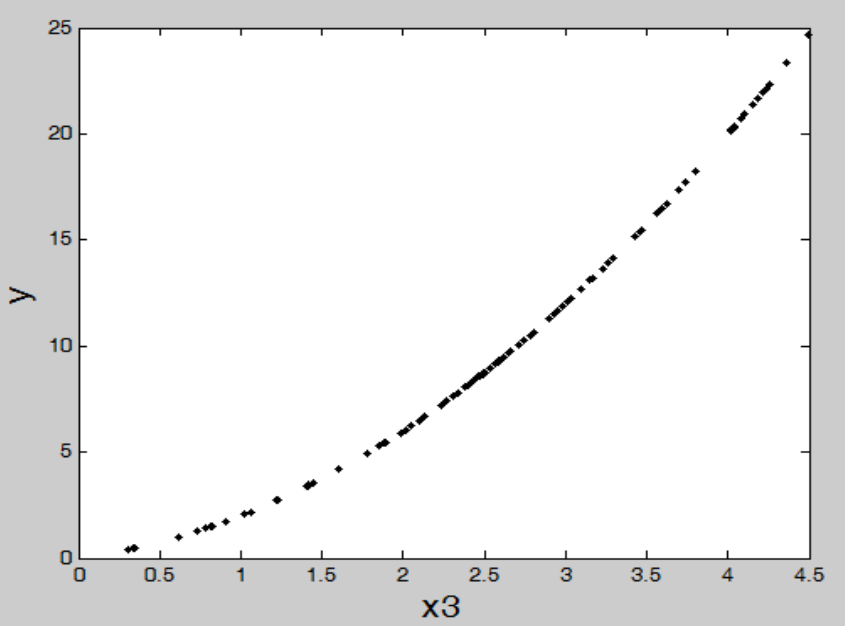


Simple DR MATLAB Example

$$y = 2x_1 + 3x_2 + 4x_1^2 + 12x_1x_2 + 9x_2^2$$

```
>> x1=rand(100,1);x2=rand(100,1);x3=2*x1+3*x2; %x3-DR_Variable
>> y=2*x1+3*x2+4*x1.^2+12*x1.*x2+9*x2.^2;
>> plot(x1,y,'b.');hold on; plot(x2,y,'ro'); plot(x3,y);
```





AI Explainability

Right to explanation

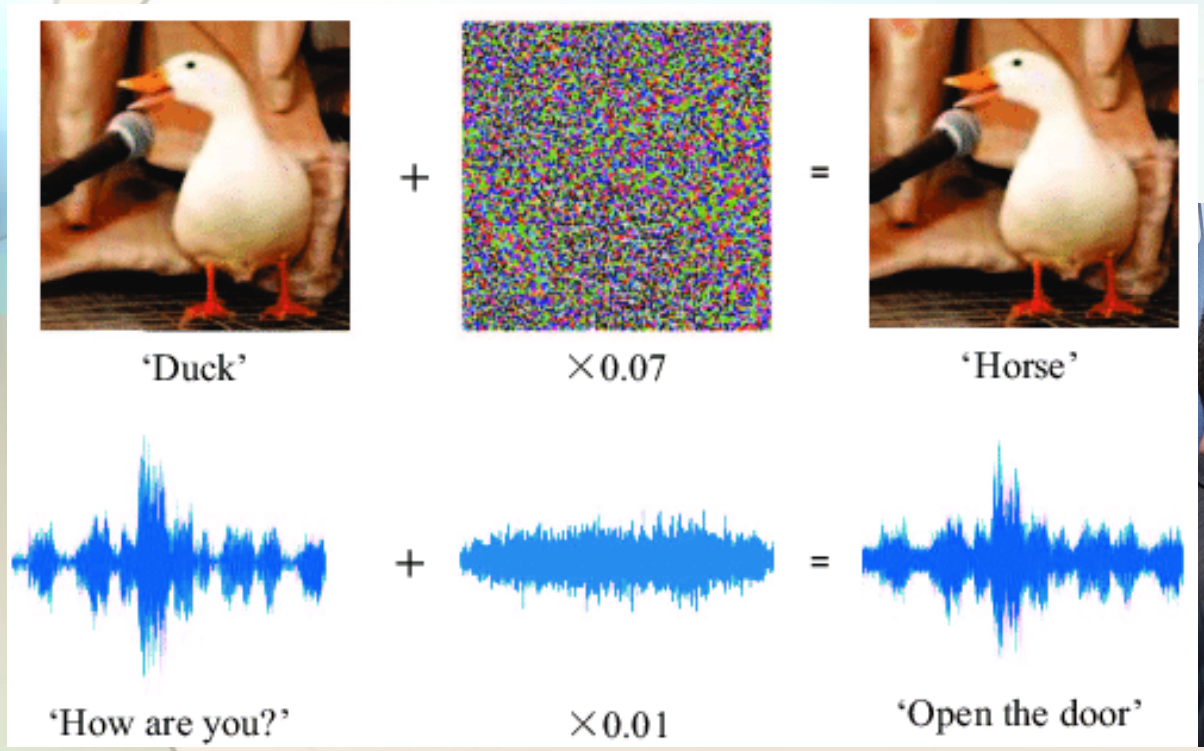
From Wikipedia, the free encyclopedia

In the regulation of algorithms, particularly artificial intelligence and its subfield of machine learning, a **right to explanation** (or **right to *an* explanation**) is a right to be given an explanation for an output of the algorithm. Such rights primarily refer to individual rights to be given an explanation for decisions that significantly affect an individual, particularly legally or financially. For example, a person who applies for a loan and is denied may ask for an explanation, which could be "Credit bureau X reports that you declared bankruptcy last year; this is the main factor in considering you too likely to default, and thus we will not give you the loan you applied for."

Some such legal rights already exist, while the scope of a general "right to explanation" is a matter of ongoing debate. There have been arguments made that a "social right to explanation" is a crucial foundation for an information society, particularly as the institutions of that society will need to use digital technologies, artificial intelligence, machine learning.[1] In other words, that the related automated decision making systems that use explainability would be more trustworthy and transparent. Without this right, which could be constituted both legally and through professional standards, the public will be left without much recourse to challenge the decisions of automated systems. One of the emerging problems is how to communicate an explanation to a user, should it be through text, a high-level visual diagram, video or some other medium, and how can an explainable system scope the explanation in a reasonable way?

If we do not know how it works, when it fails, when it succeeds, if it is biased, how do we trust it?

AI Brittleness



AI succeeds and fails spectacularly

AI Brittleness

Why does this happen?

- Black box approach
- Neural nets ≠ human brain
- Overparameterized
- Overfitting / Memorization

AI Brittleness

How does this affect the real world?

The Adversarial Turtle

- Use a 3D printer to print a turtle
- Place different “textures” on the shell
- Most image recognitions fail
- CNNs: “On a very fundamental level, our work highlights how far current CNNs are from learning the 'true' **structure** of the world”

rifle

https://www.theverge.com/2017/11/2/16597276/google-ai-image-attacks-adversarial-turtle-rifle-3d-printed

99% of modern image recognition is just simple (but precise) texture matching

Malicious AI

Purposeful misdirection of AI with the intent to cause harm

AI decides to “go” instead of “stop”

Malicious AI

How can we deploy AI models if we cannot trust them to work as intended?

Malicious AI

Prevent adversaries from accessing the model

- Not resilient to insiders
  - Know everything about the AI model used
- Membership inference attacks
  - Reconstruct the training data
- Black-box attacks
  - Learn the true AI model from scratch using input-output examples (public)
  - Construct adversarial examples (FGSA, Carlini & Wagner)

Deepfakes

AI is a double-edged sword

- AI used to learn can be used to mimic and create fakes
- It gives the same model for the same data
- Cannot tell intent of user

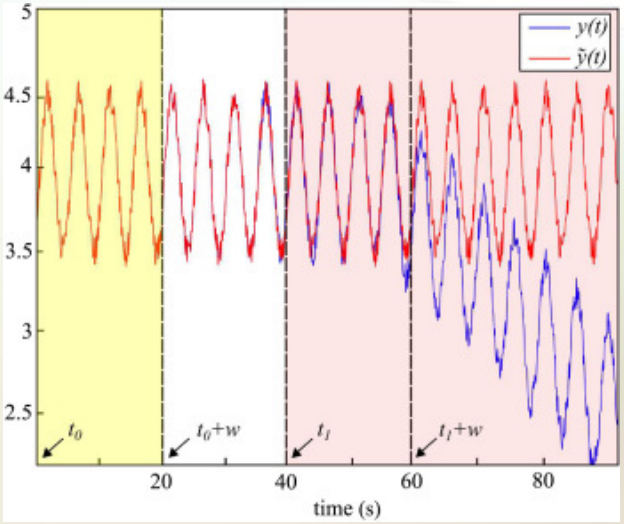
Deepfake videos

Deepfake videos

The Engineering Perspective

Deepfakes can be applied to time-series data to fool operators and passive defense mechanisms

+



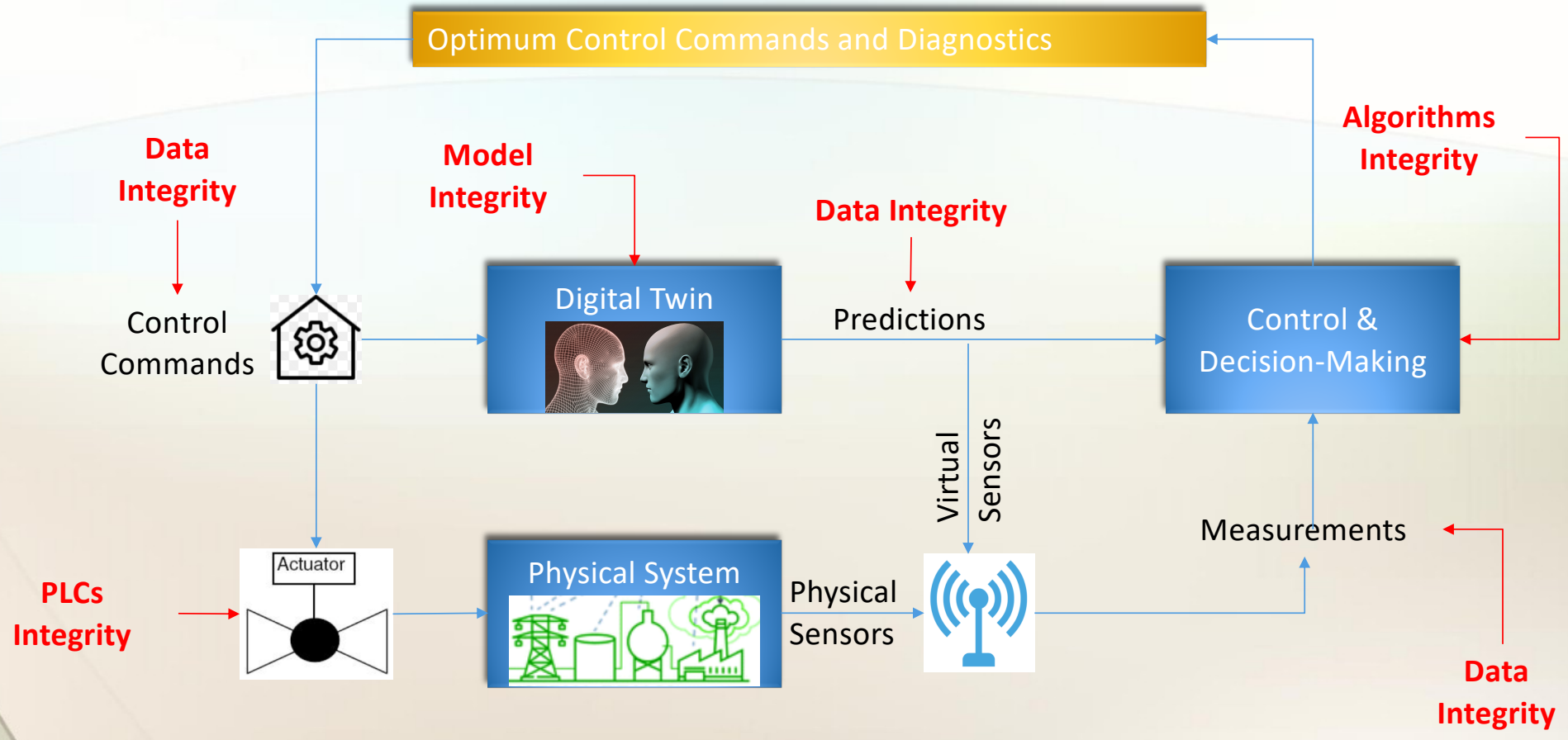
=

An Indian nuclear power plant suffered a cyberattack. Here's what you need to know. – *The Washington Post*

Hackers breached a dozen US nuclear plants, reports say – *BBC*

Stuxnet worm 'targeted high-value Iranian assets' – *BBC*

Data Integrity Challenges



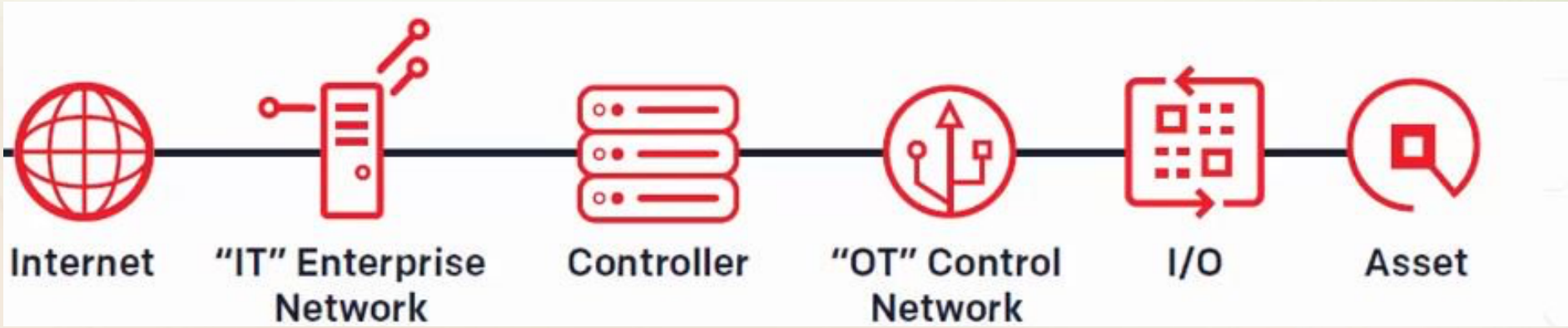
Current R&D Efforts

Information Technology (IT) Defense

- Build ‘walls’ to stop unauthorized access
- Use generic methods
- Can be bypassed: Stuxnet, 2010; Ukrainian Electric Grid attack , 2015

Operational Technology (OT) Defense

- Complementary to IT defense
- Use customized methods
- Protect system at the physical process level

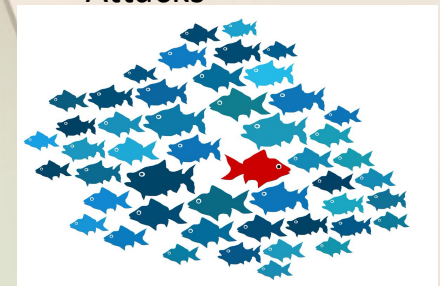


https://forums.xilinx.com/t5/Adaptable-Advantage-Blog/Protect-Your-Assets-in-Healthcare-and-Industrial-Domain-Watch/ba-p/1014289

Current R&D Efforts

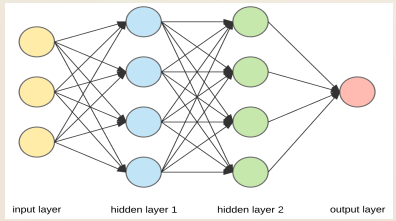
Outlier/Anomaly Detection Techniques

- Inconsistent with majority
- Straightforward FDI Attacks



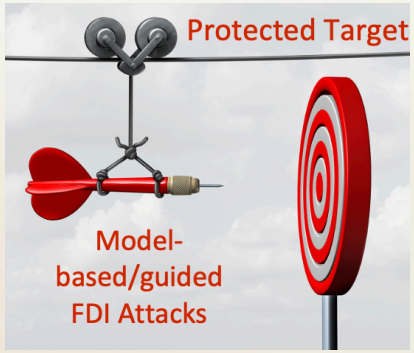
Data-driven Techniques

- Build predictive models
- Auto-correlation-type regression
- Neural Network



Model-based Defense

- Rely on physics model
- Establish basis for normal behavior



High-order model-based defense

- Irreproducible signature
- Signature sensitive to system behavior

Figure source: https://towardsdatascience.com/a-brief-overview-of-outlier-detection-techniques-1e0b2c19e561; https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6; https://www.growteam.com/blog/topic/digitalmarketing; https://www.123rf.com/photo_66137198_man-with-magnifier-on-white-background-isolated-3d-image.html

Challenges

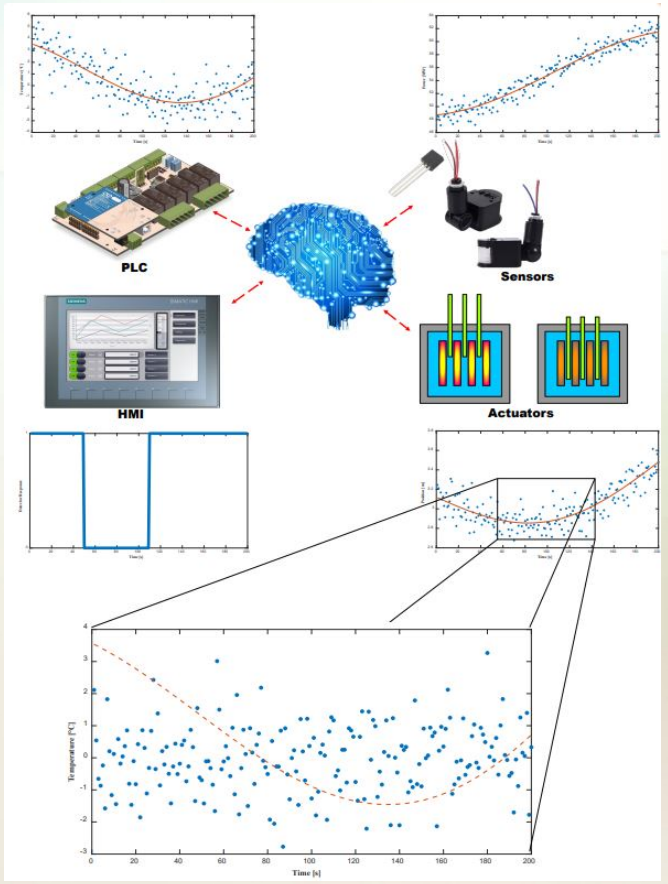
Model = Keys to the castle

- If the model is known or can be learned, defense is bypassed
- AI and physics are used to find the model – both known to insiders

How to differentiate between data from system and data faked “perfectly”?

Covert Cognizance (C^2)

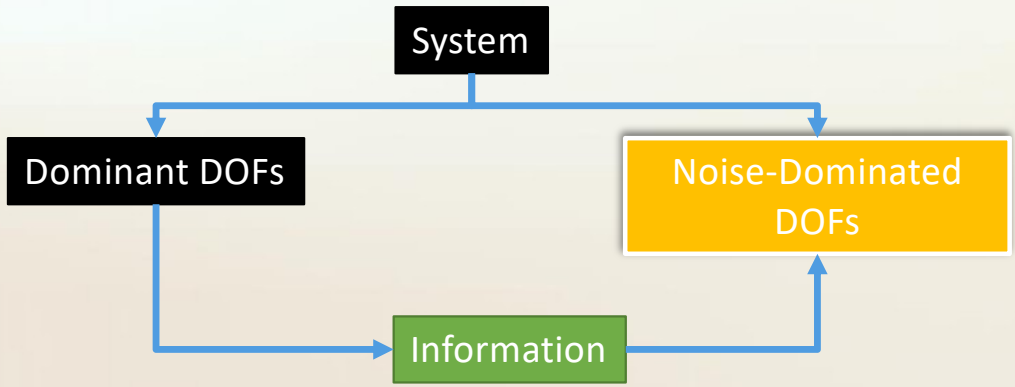
- Physical Process Defense
  - Focuses on the noise, not just the patterns
  - Noise is inherent redundancy in the system
  - Store information about other systems along the noise
  - Randomness for security (OTP)
  - No footprint on system operation, and no additional variables



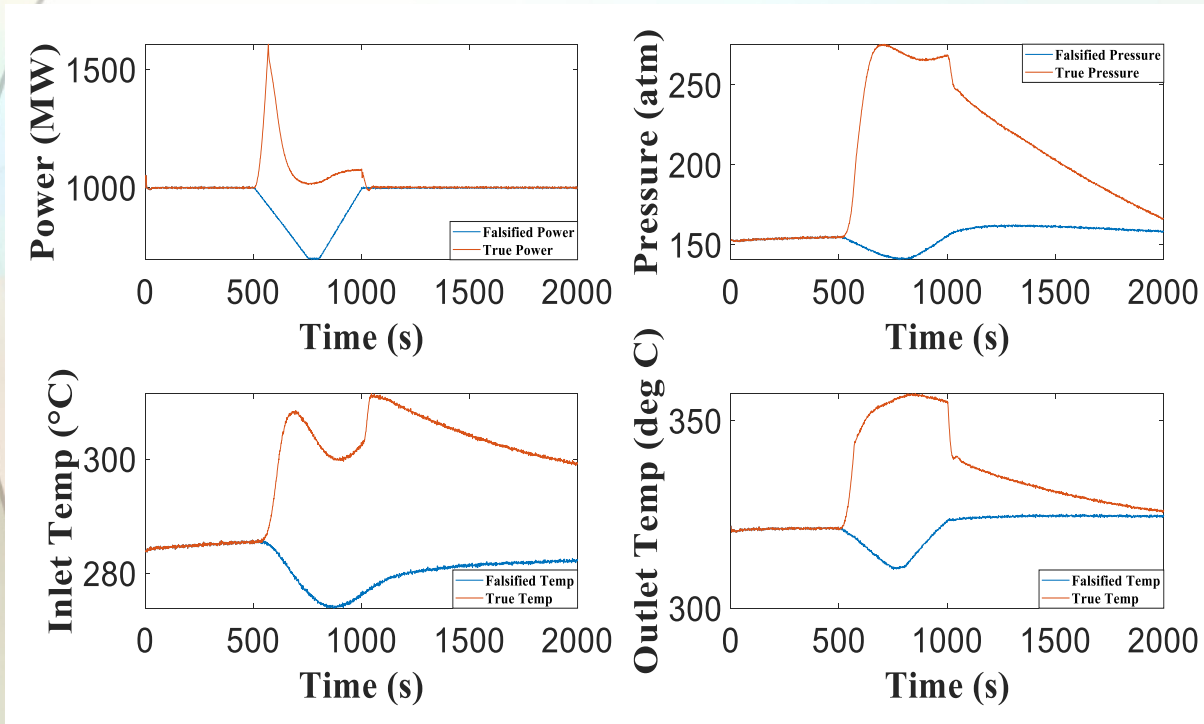
C^2 Motivation

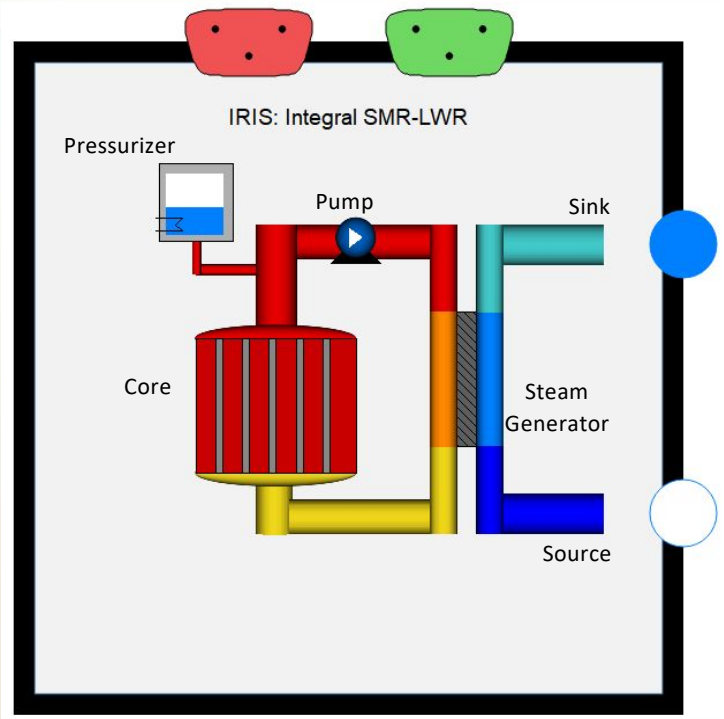
Industrial systems are reducible:

- Dominant behavior can be described using small no. DOFs,
- Leaving huge number of “unused” noise-dominated DOFs, that can serve as courier variables



Replay Attack





AI Collaboration/Testing

How to improve AI collaboration when dealing with data from critical infrastructure?

- Mask sensitive characteristics of data
- Leverage full benefits of open-source AI/ML

Figure Source: https://www.tolpagorni.com/blog/professional-networking-and-collaboration

Proprietary Data – Nuclear Power

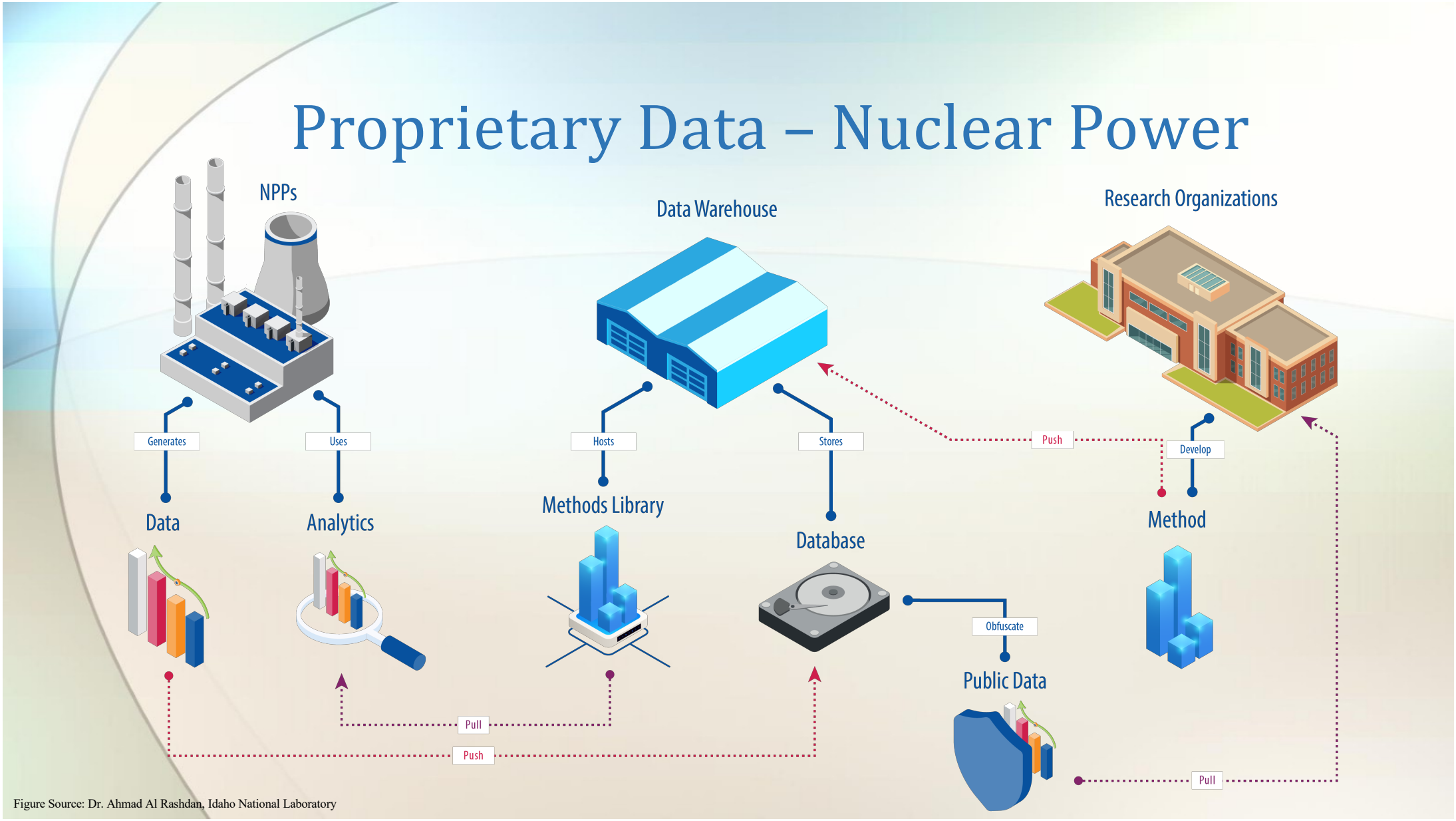


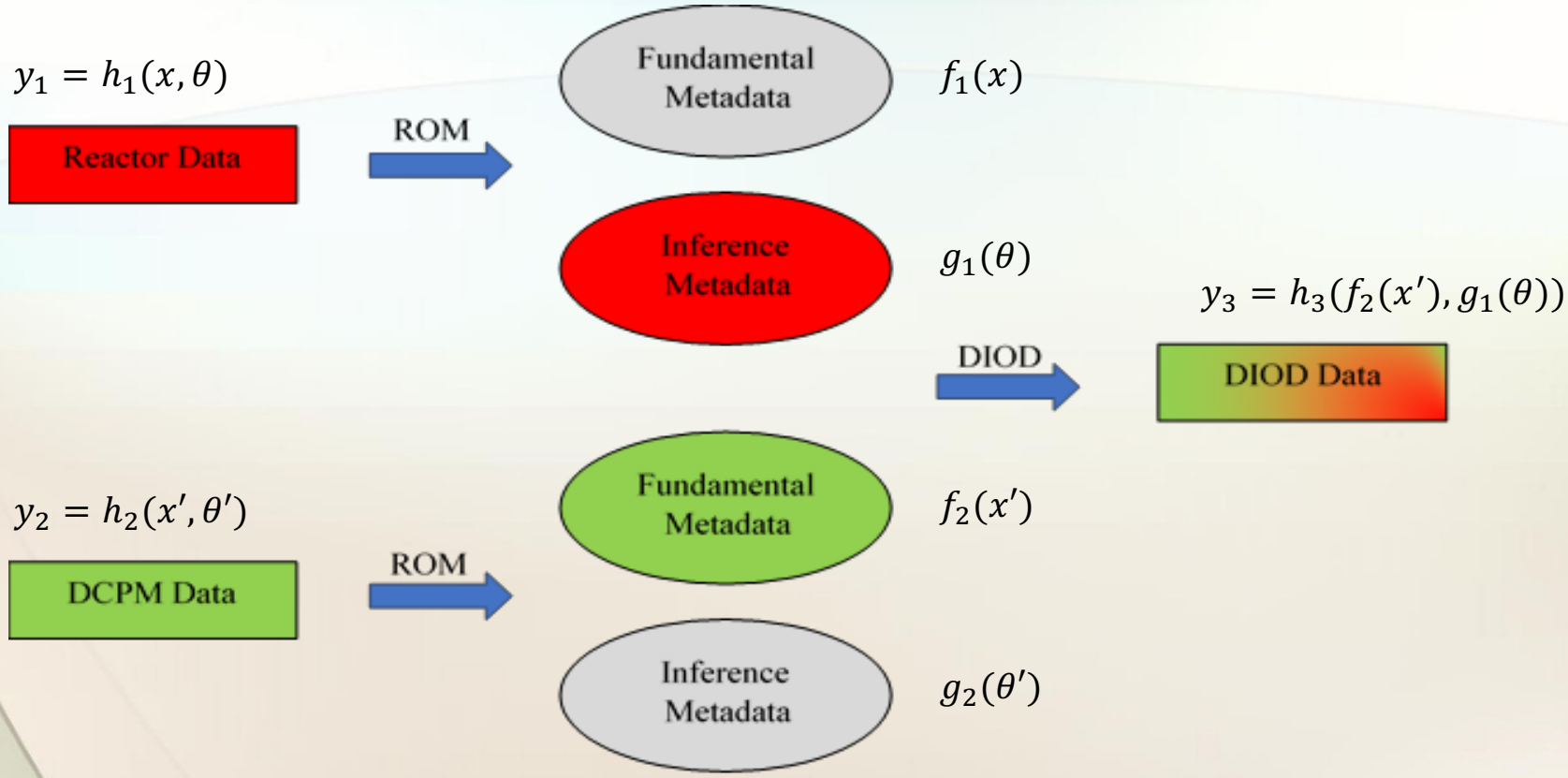
Figure Source: Dr. Ahmad Al Rashdan, Idaho National Laboratory

DIOD Data Masking Paradigm

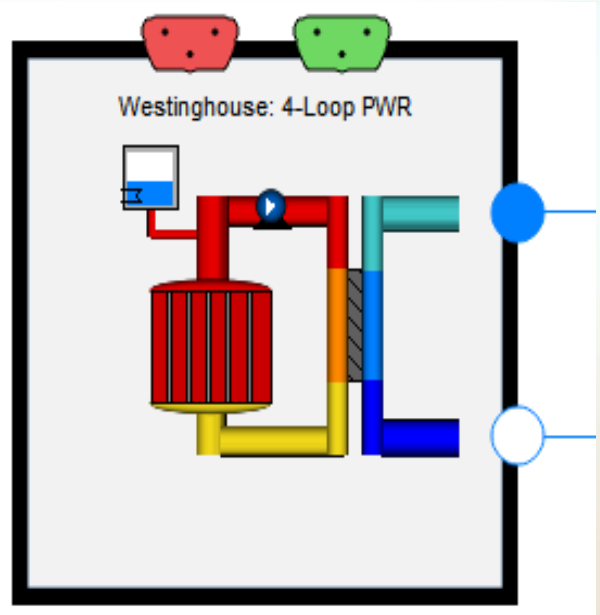
Deceptive Infusion of Data (DIOD)

- Splits dataset into fundamental metadata and inference metadata
  - Fundamental metadata denotes information pertaining to system identity
  - Inference metadata denotes information relevant for target AI/ML task
- Obfuscates system identity by mounting inference metadata onto fundamental metadata of a different generic system; generate DIOD version of data
- Ensures same theoretical inference on original and DIOD version through mutual information guarantees
- Cannot reverse-engineer DIOD data to decipher system identity

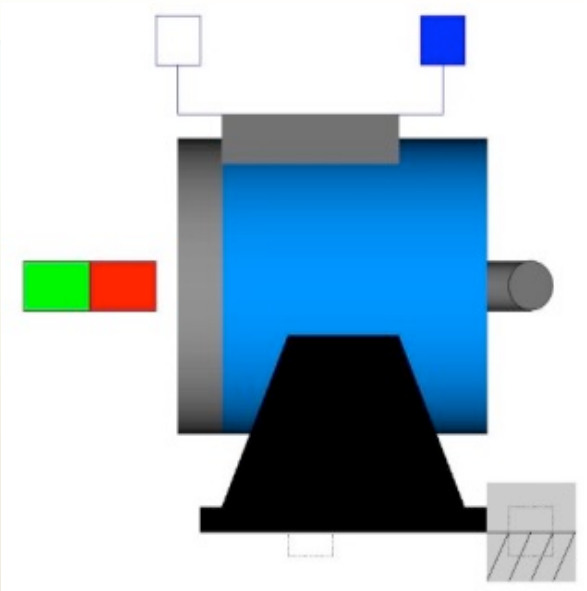
DIOD Algorithm



DIOD: Nuclear Application

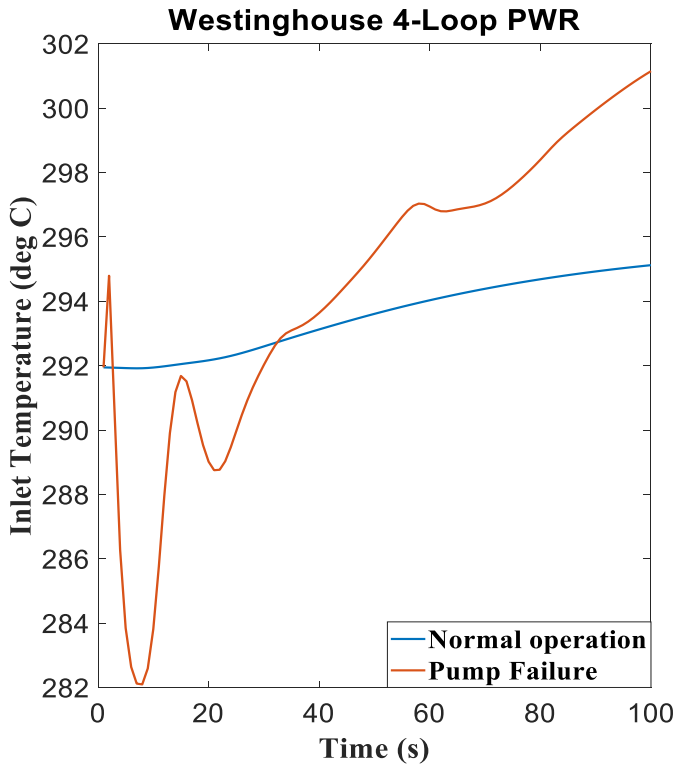


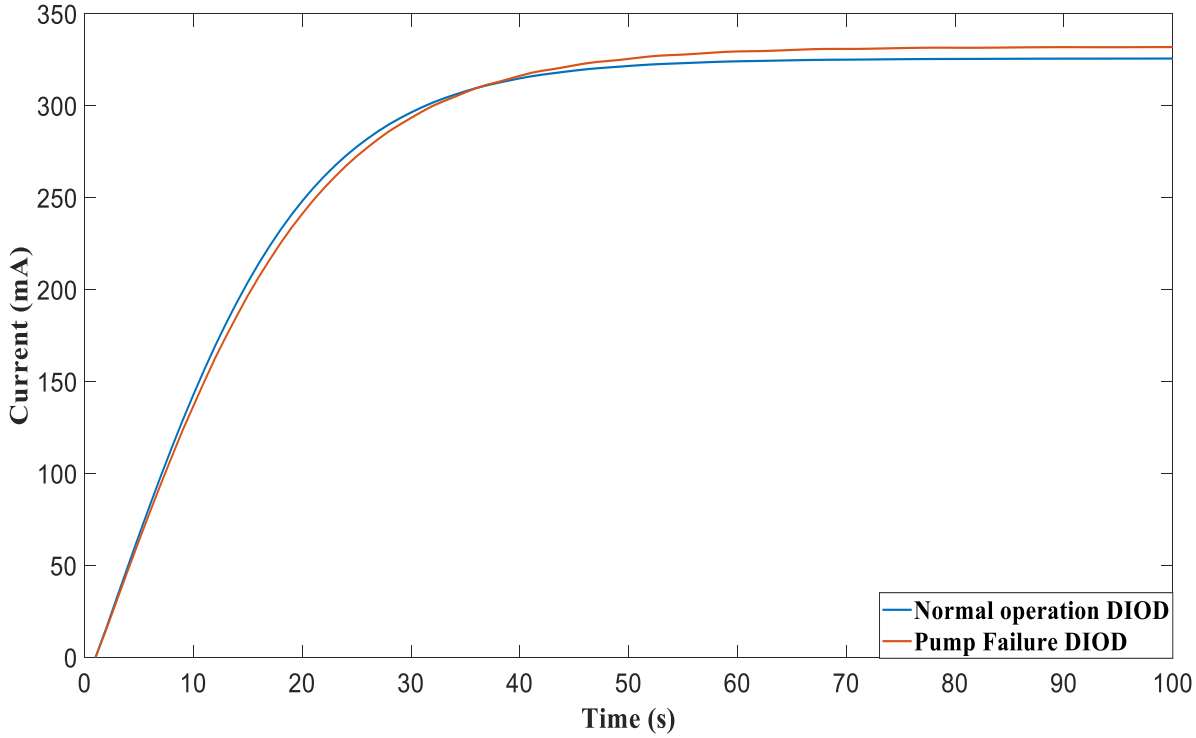
Inference Metadata



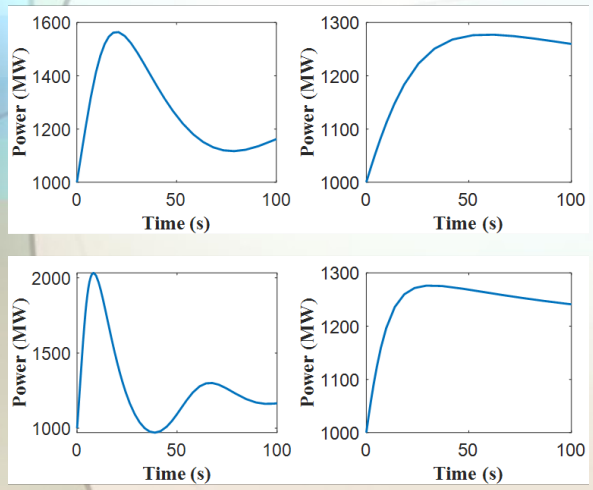
Current Controlled DCPM

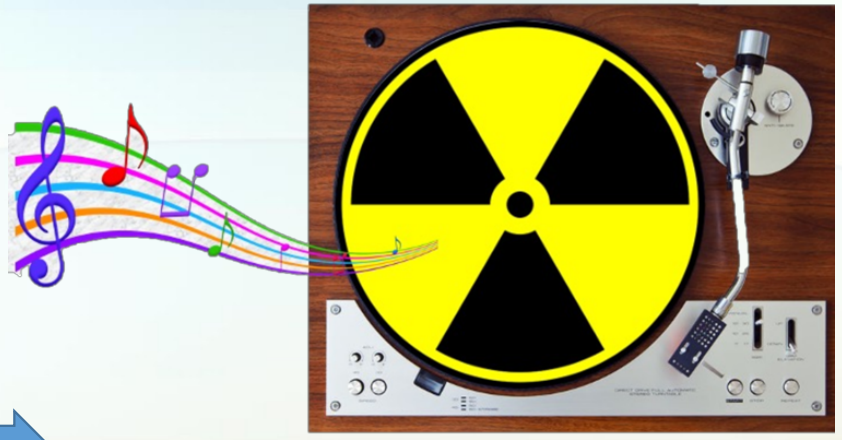
Nuclear Data appears like it's from a motor

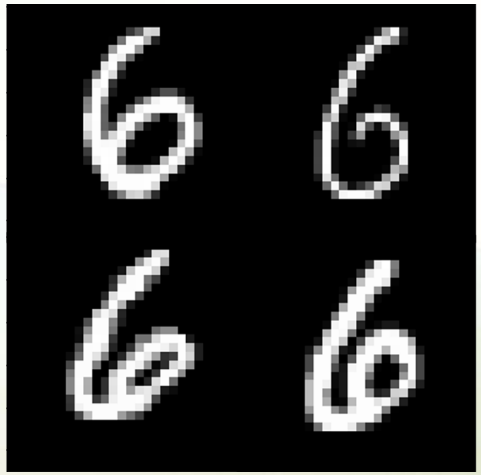


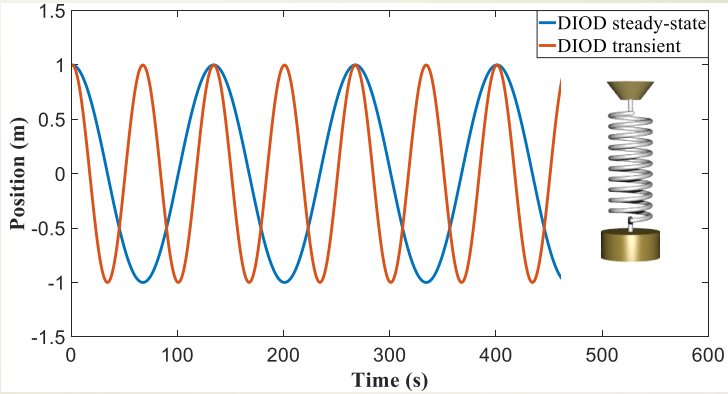


DIOD Applications









- **Automation is going to replace many jobs,**
- **unlikely to replace the humans designing the automation algorithms.**

- Do not hesitate to contact me for more information

- abdelkhalik@purdue.edu