RSAConference2020

San Francisco | February 24 – 28 | Moscone Center

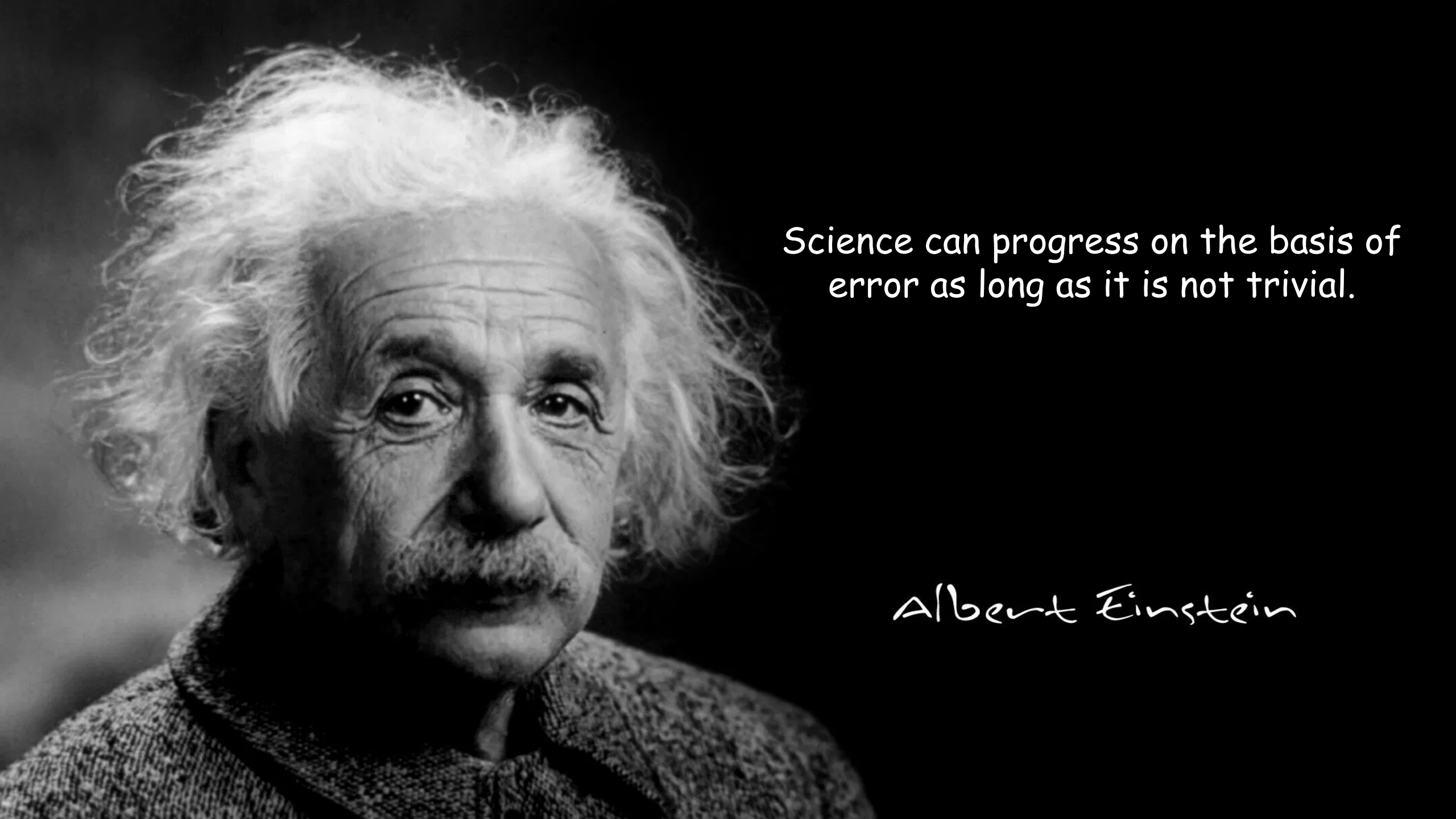
HUMAN ELEMENT

SESSION ID: **MLAI1-T11**

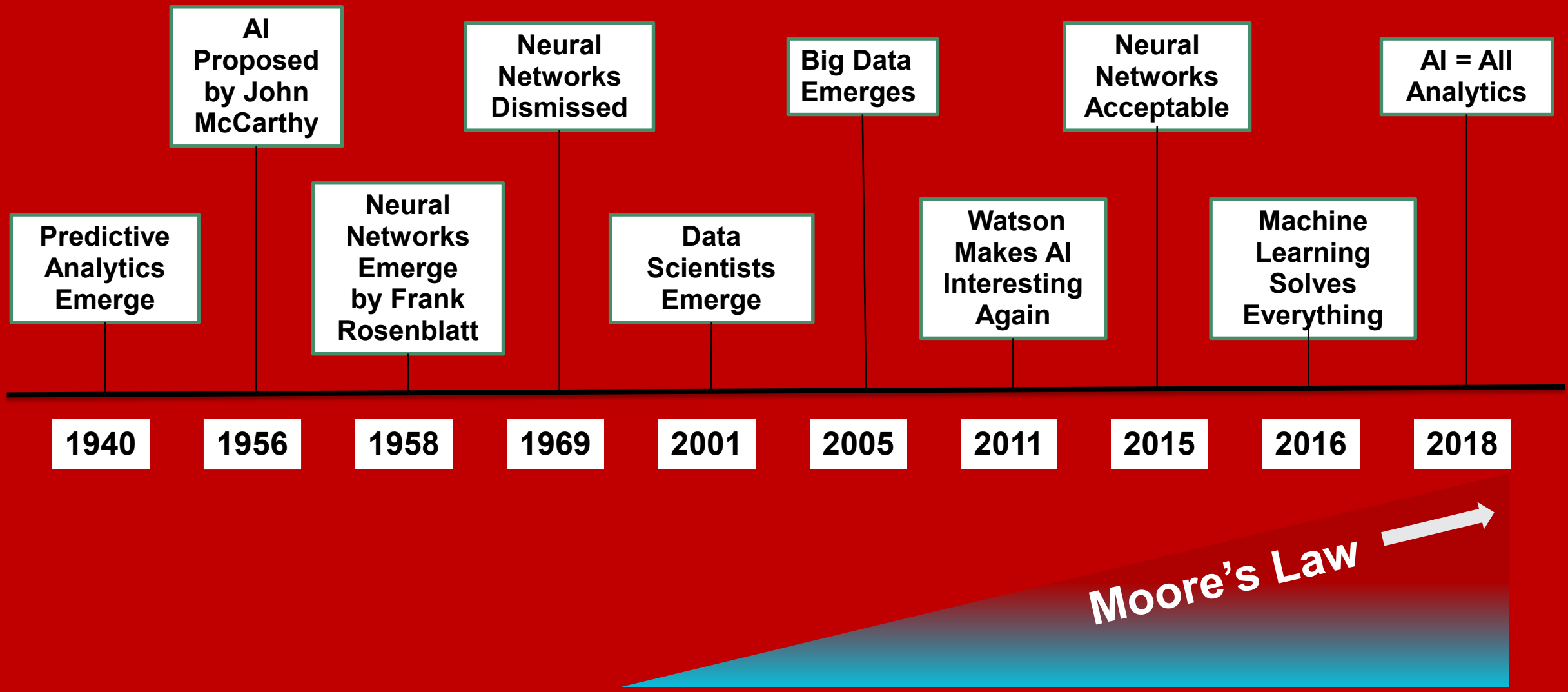
Reproducibility: The Life and Risks of a Model

Celeste Fralick, Ph.D.

Chief Data Scientist / Senior Principal Engineer
McAfee LLC, Office of the CTO | www.mcafee.com
@purkinje16 | celeste_fralick@mcafee.com
v2


Science can progress on the basis of error as long as it is not trivial.
Albert Einstein

The Analytics Hype-line

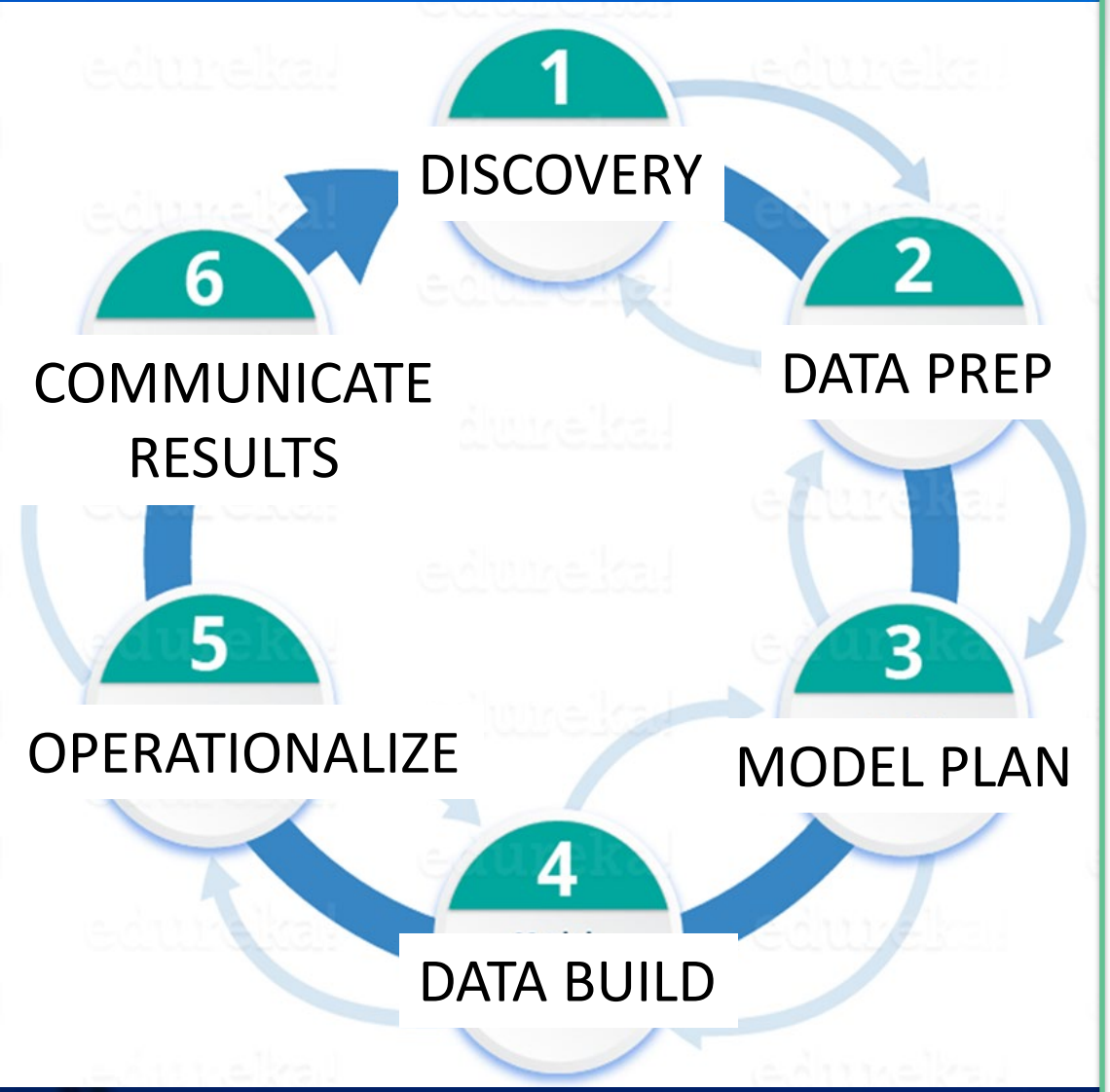


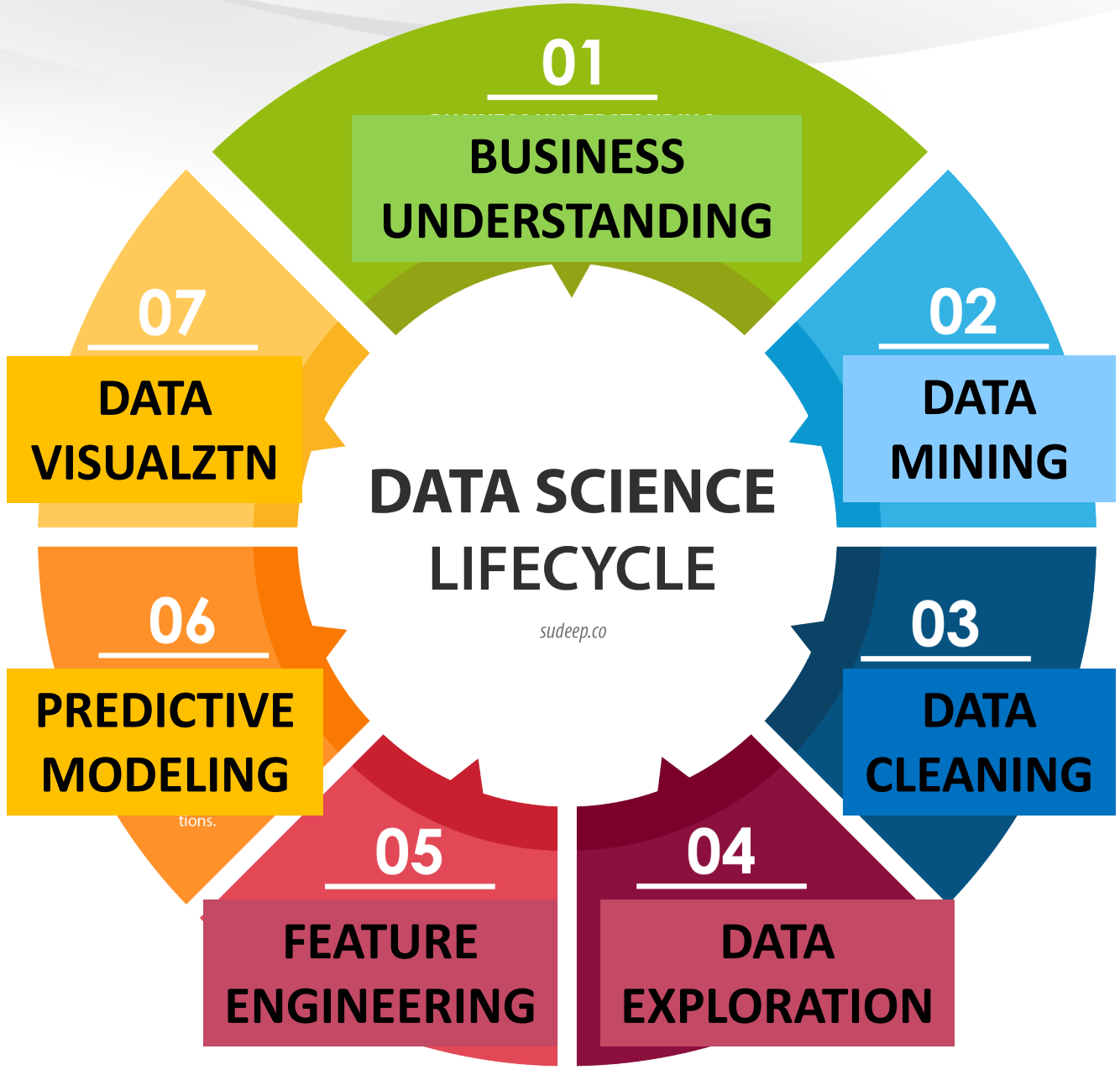
Not to scale

Loosely based on https://en.wikipedia.org/wiki/Timeline_of_machine_learning

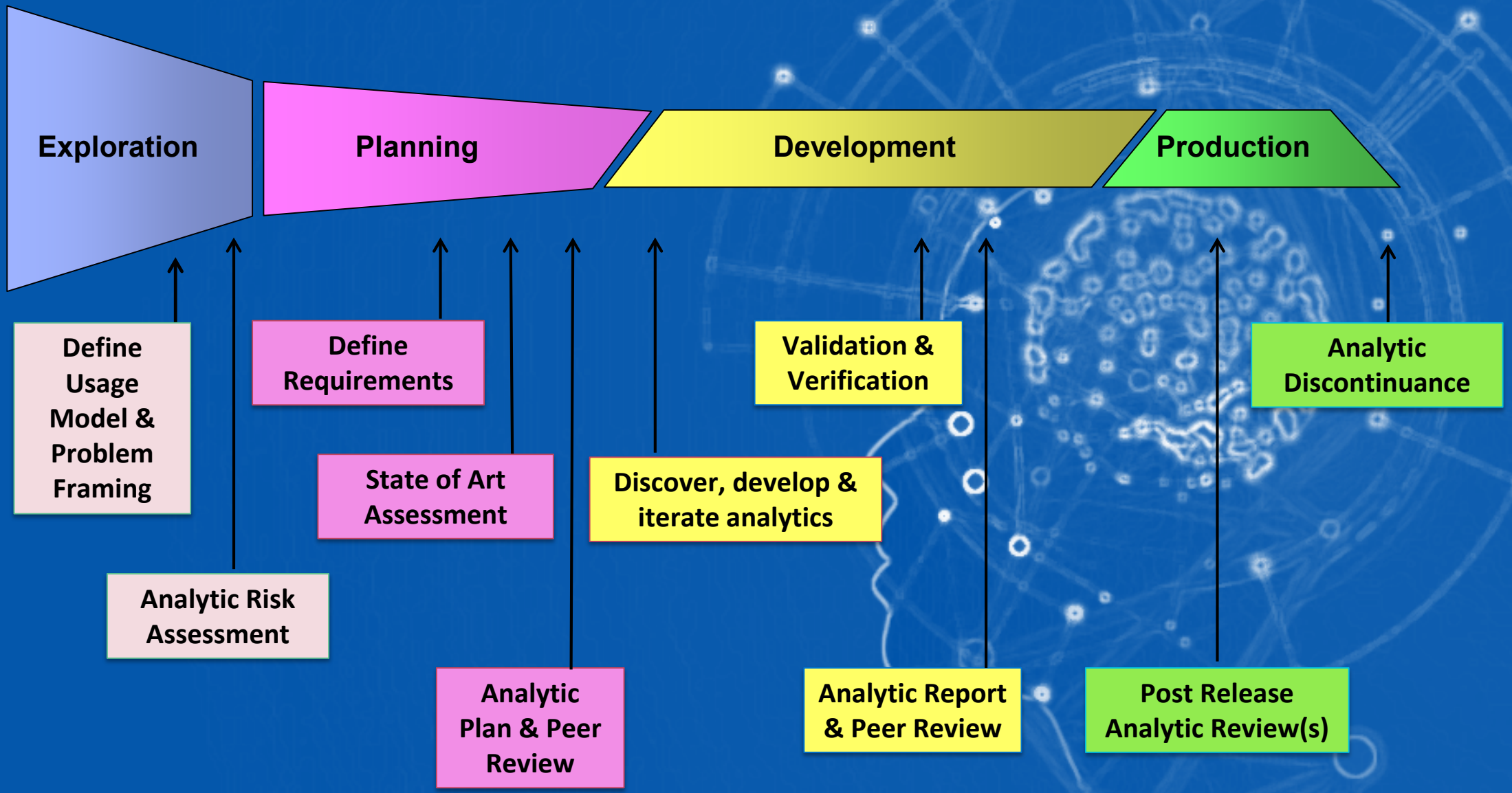
https://twitter.com/drjuliashaw/status/874293864814845952

(chicken) Muffin photo courtesy of @teenybiscuit.com

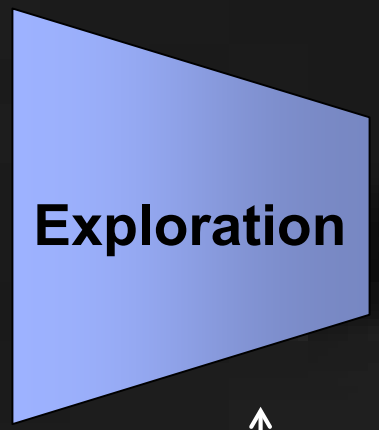

1
DISCOVERY
2
DATA PREP
3
MODEL PLAN
4
DATA BUILD
5
OPERATIONALIZE
6
COMMUNICATE RESULTS


01
BUSINESS UNDERSTANDING
02
DATA MINING
03
DATA CLEANING
04
DATA EXPLORATION
05
FEATURE ENGINEERING
06
PREDICTIVE MODELING
07
DATA VISUALZTN
DATA SCIENCE LIFECYCLE
sudeep.co

Analytic Development Process (Waterfall)



Identify, Quantify, Mitigate, and Learn Analytic Risks



Analytic Risk Assessment

- 1 Model ≠ Data Scientist
- Multiple error rates
- Compute footprint
- Explainability (XAI)
- **Bias**
- **Adversarial ML attacks (AML)**
- **Model Reliability**

Measurement Bias

Algorithmic Bias

Bias is the difference between the average prediction of our model and the correct value which we are trying to predict

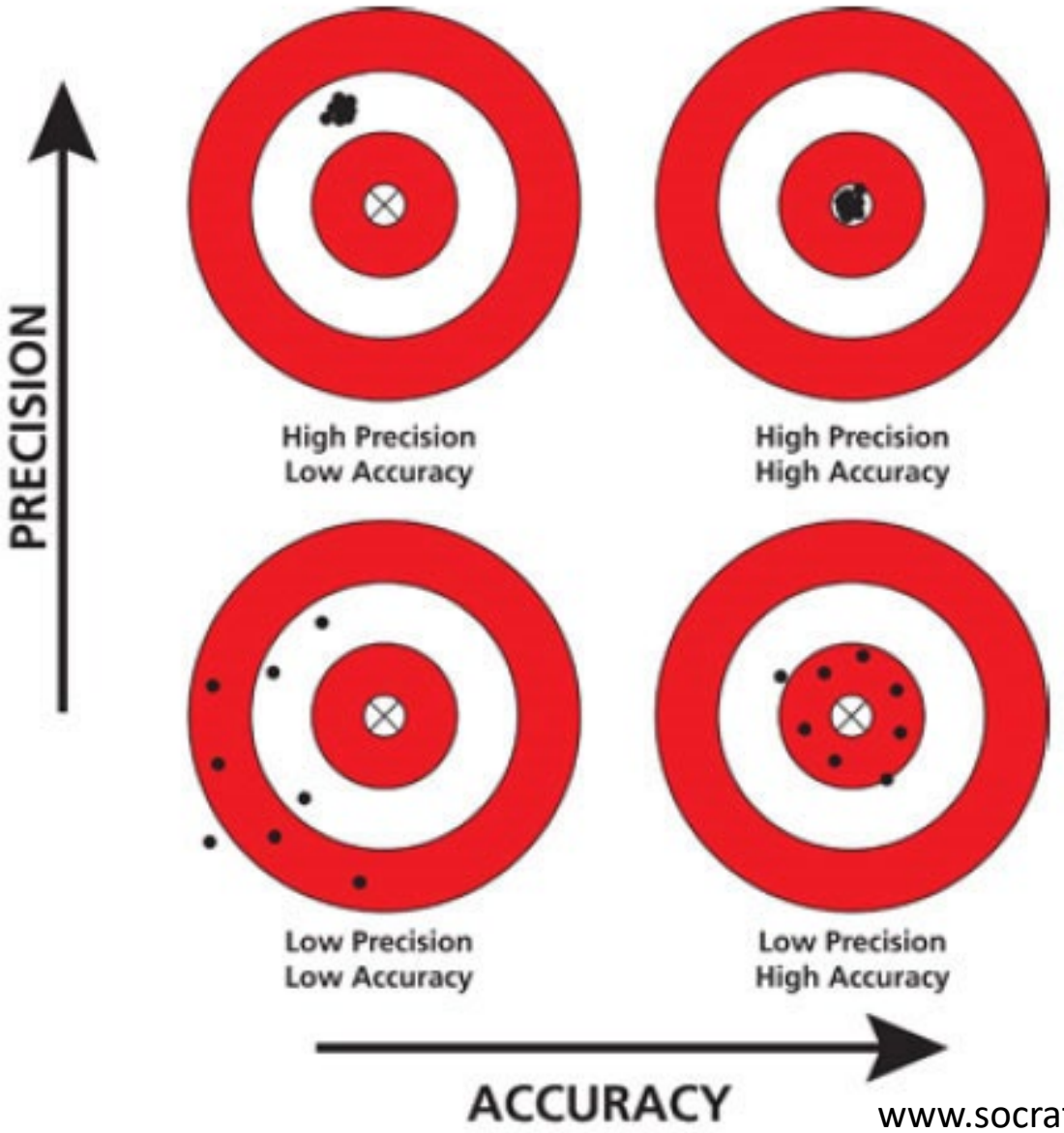
(Seema Singh)

Prejudicial Bias

Sample Bias

Measurement Bias

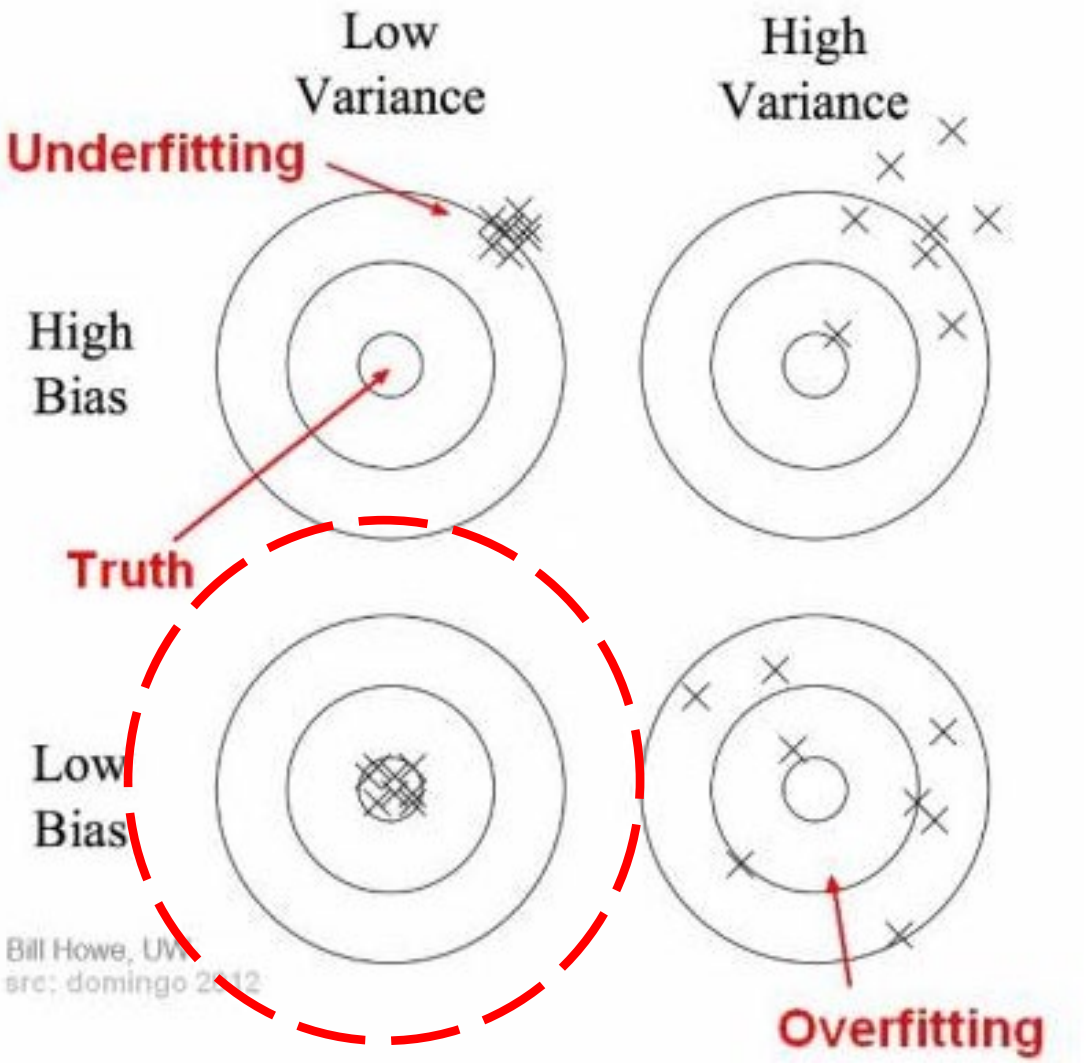
Different people can measure the same thing and get different results



PREVENTION

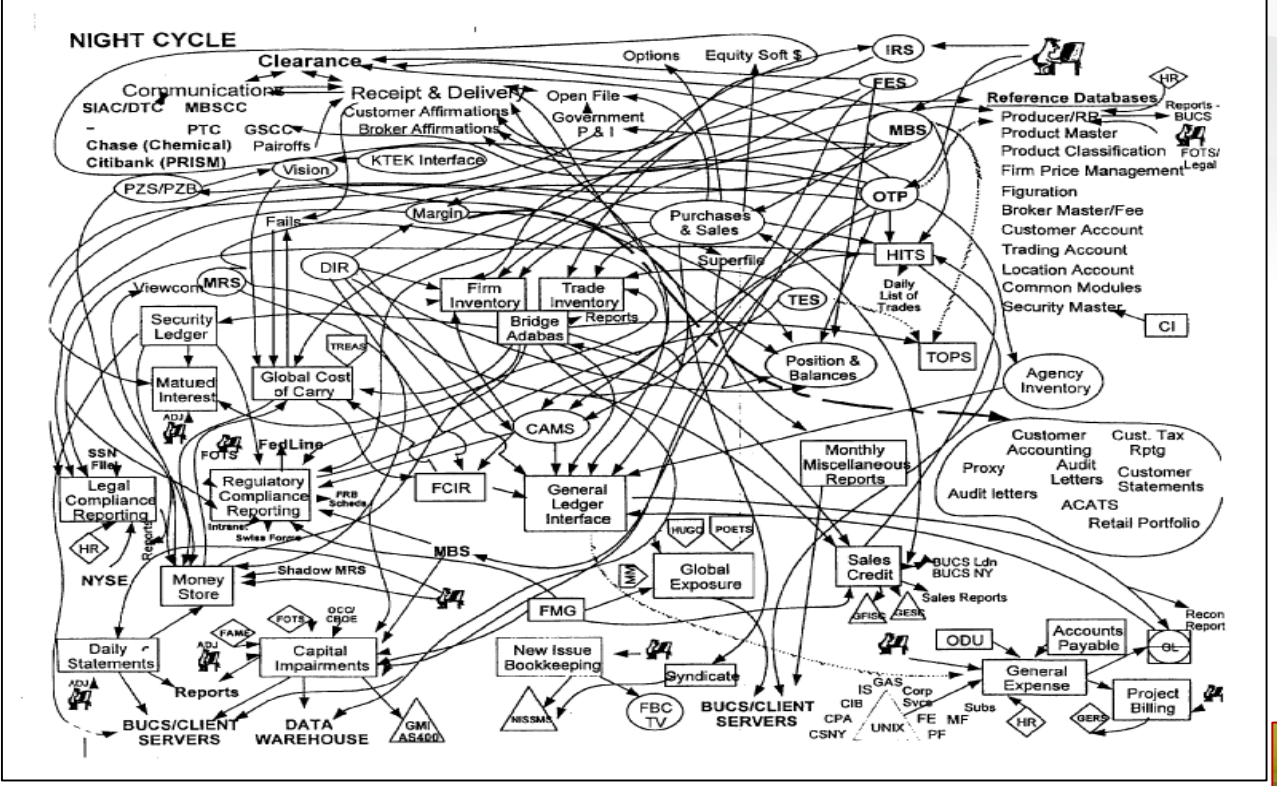
- Utilize tool with highest precision & accuracy
- Ensure data collection is balanced and appropriate
- Perform a “Gauge R&R”
- Have a ground truth and/or calibration

Prediction errors (bias & variance) require optimization



PREVENTION

- Use multiple error rates
  - E.g., RMSE, R^2, GenR^2
- Ensure Training accuracy </= Validation accuracy
- Reduce noise in preprocessing
- Beware automated & open source tools



Sample inference means that the model's sample reflects its larger population where it will be applied

PREVENTION

- Understand math at each point in data flow
- Use automated sampling
- Select samples randomly
- Train models on all cases

McAfee™
Together is power.

The inherent conscious or unconscious bias that impacts resulting model

Prejudicial Bias

PREVENTION

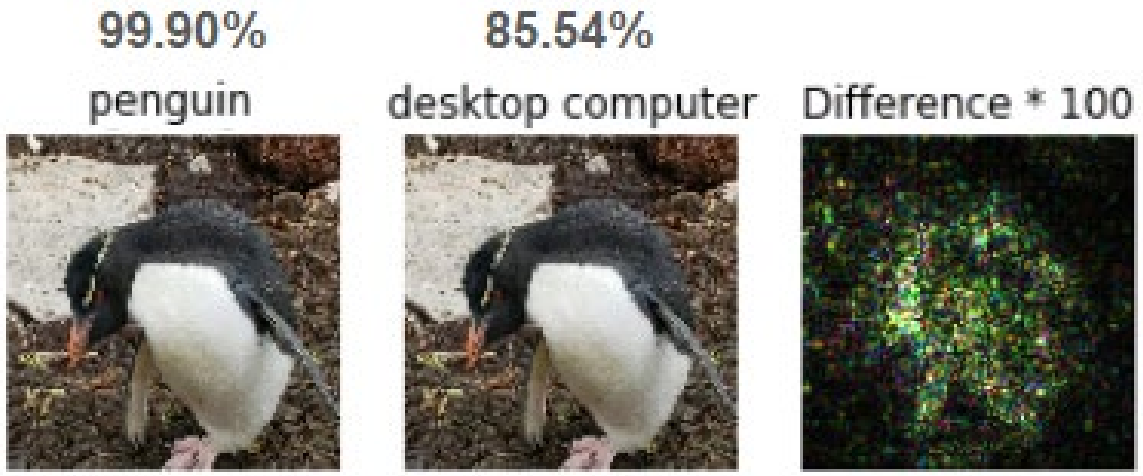
- Examine domain of feature
- Consider interactions, combinations, polymorphisms
- Re-consider cleaning/outliers
- Accountability
- Use FairML, LIME, other XAIs

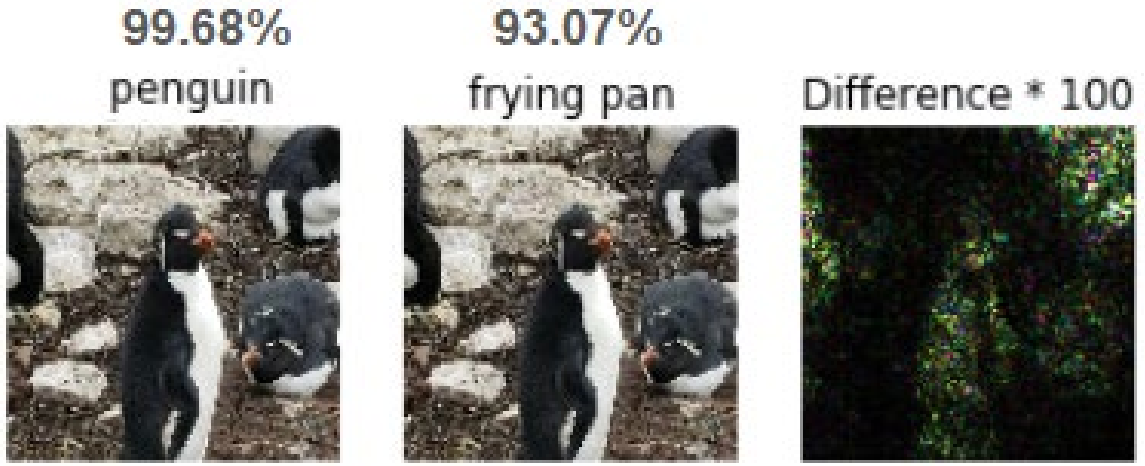
Risk: Adversarial Machine Learning (AML)

The study and design of machine learning algorithms that can resist attacks

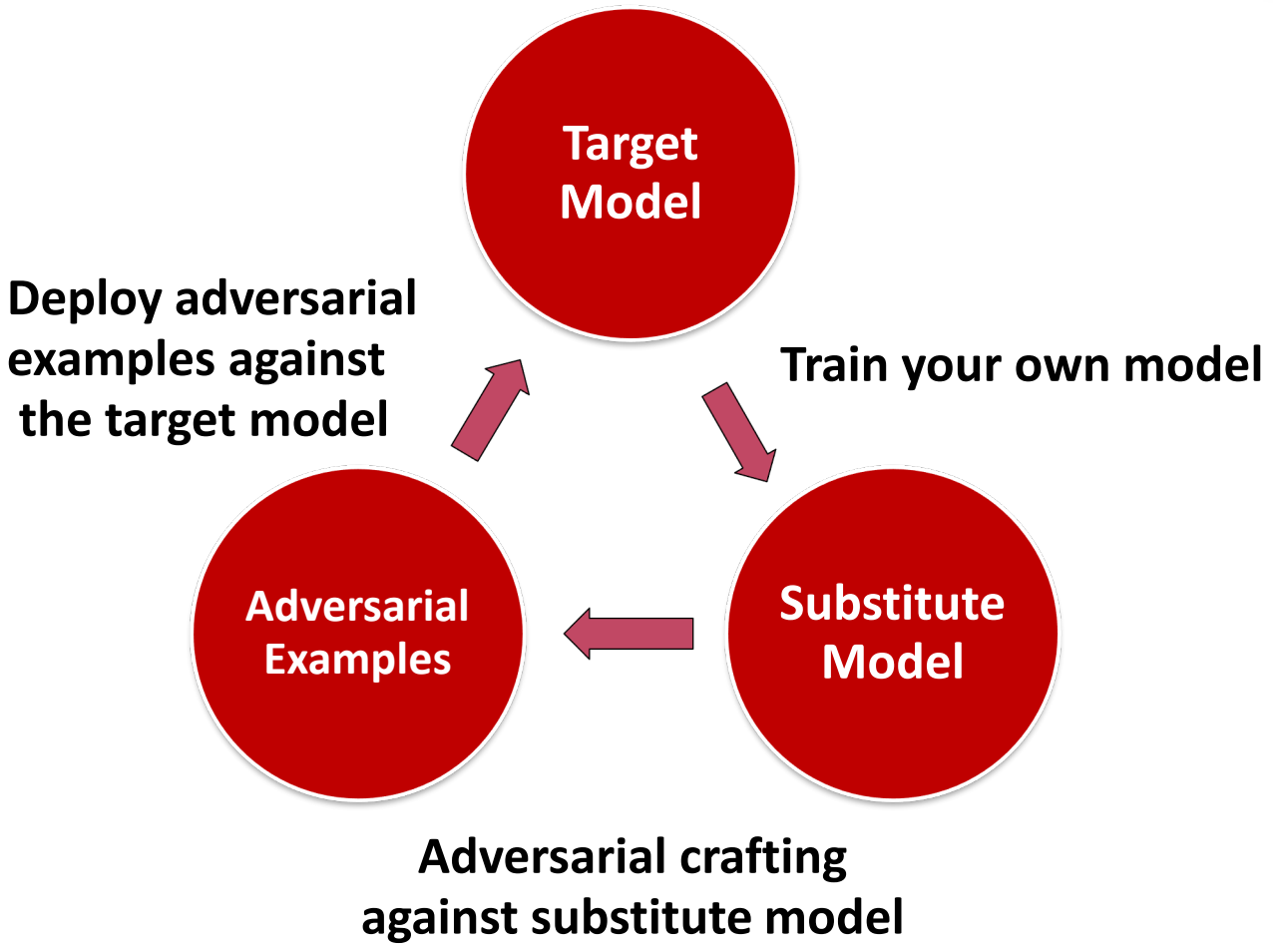
"Model Hacking"

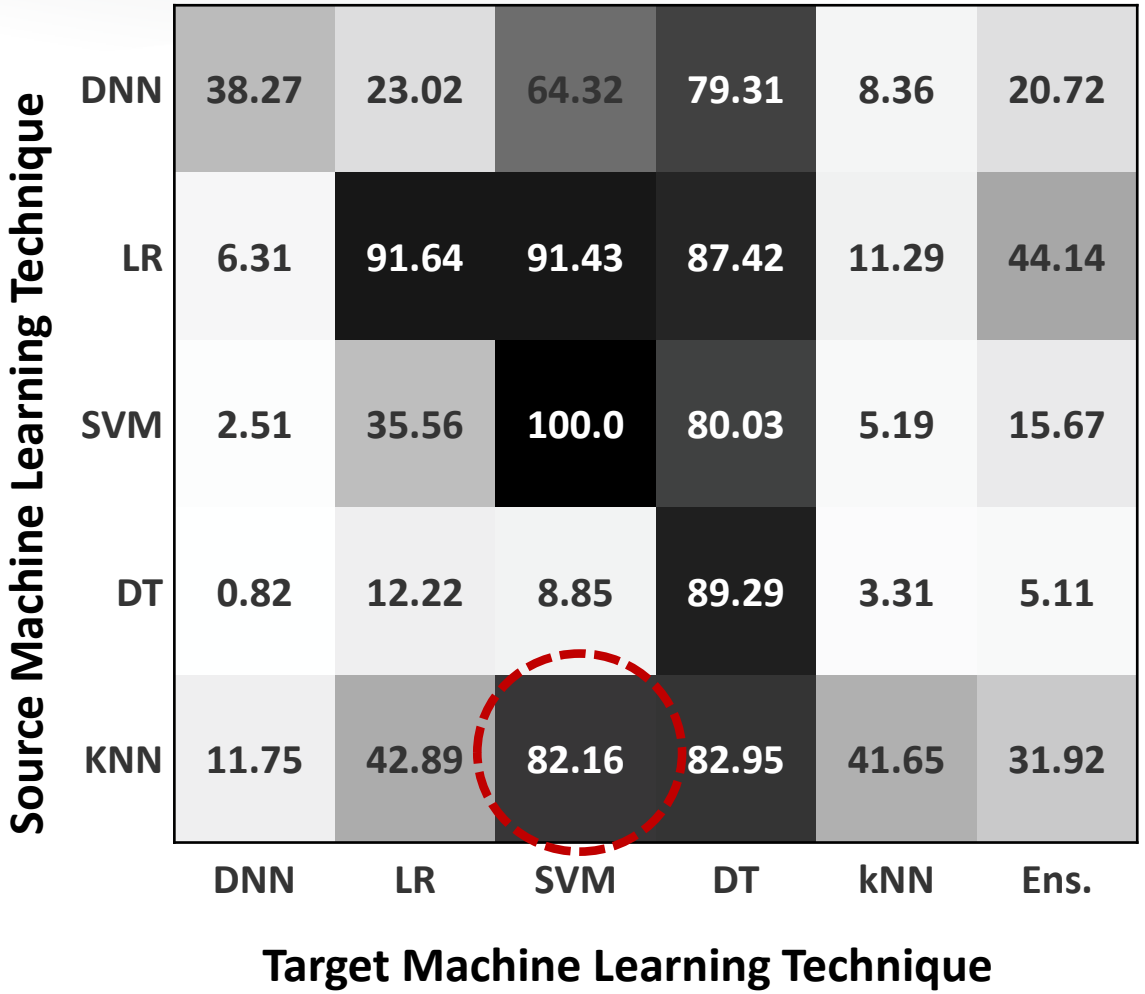
- **Poisoning Attacks** at Training can change model parameters
- **Evasion Attacks** at Test can misclassify a model decision, ↑ False Negatives





Transferability of Adversarial Examples





[Papernot 2016]

Defending Against AML Attacks

- Apply various analytic techniques:
  - Distillation
  - Feature Squeezing
  - Noise Addition
  - Adversarial Samples
  - Reject on Negative Impact
  - Fast Gradient Sign Method
- Frequent re-Training
- Monitor drift of key analytic metrics
- Human-Machine Teaming
- Explainability monitoring
- Monitor Data Decay

Risk: Model Reliability

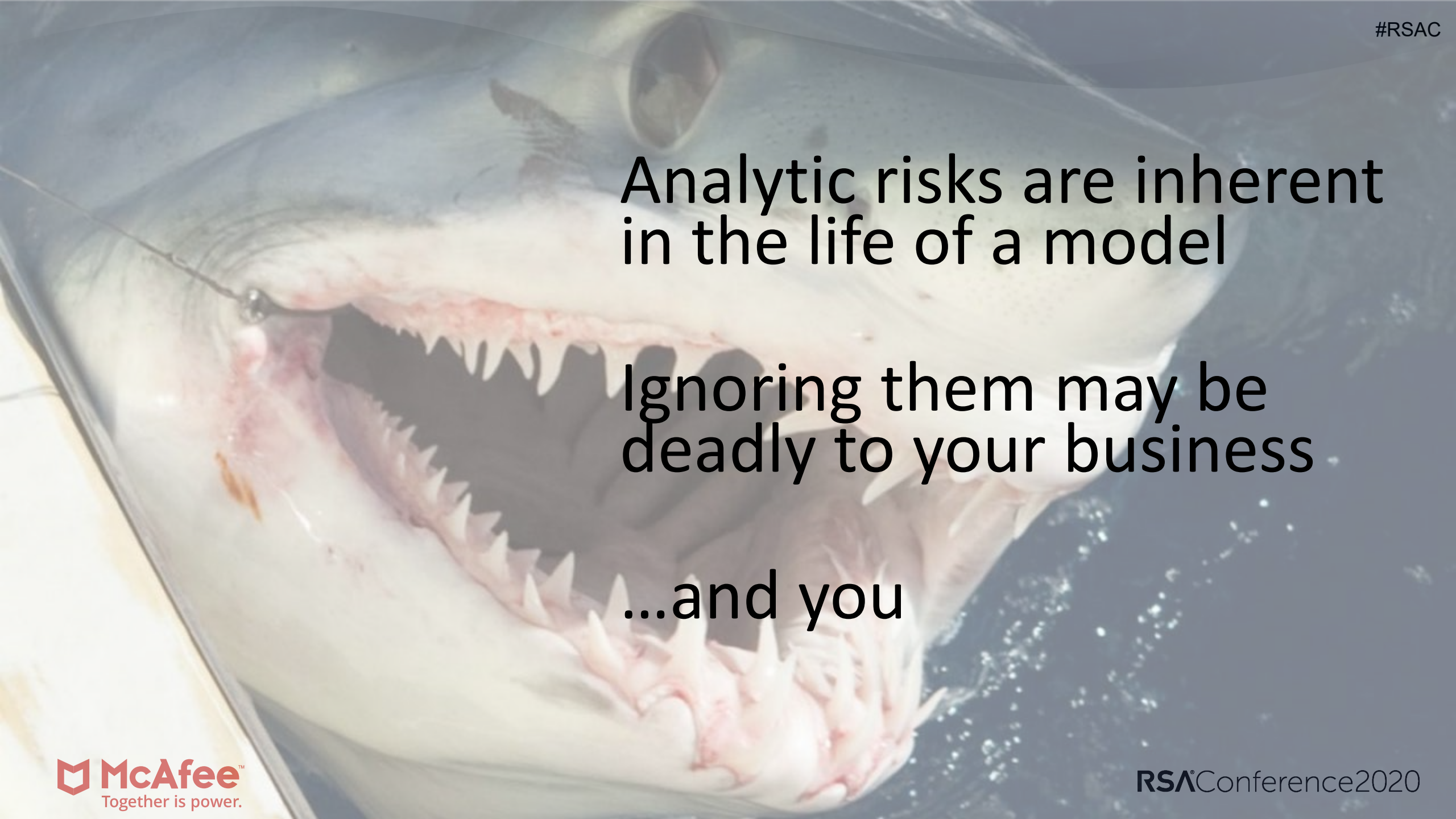
Poor reliability of the model in the field results in poor performance over time

- How often does the model “Learn”?
- Is the drift of key metrics monitored?
- Are actions & tolerances statistically derived?
- Has the dataflow (lineage) changed?
- Has the customer’s process changed?
- Has the ground truth evolved?
- Has there been a recent Post-release Analytic Review?
- Has the contributions of the Features changed?

If the answer to any of these questions is “I don’t know”, it is time for a Post-release Analytic Review!

Applied Learnings to Take Away

1. Do you have an analytic life cycle? Does it include risk ID?
2. Are multiple models and error rates examined?
3. How has the customer's compute footprint integrated into model development?
4. Can you explain how your model reached its decision?
5. How do you minimize bias in your model?
6. How do you detect AML / model hacking?
7. How do your models perform over time in the field ("model reliability")



Analytic risks are inherent in the life of a model

Ignoring them may be deadly to your business

…and you

Analytic Life Cycle (Agile)