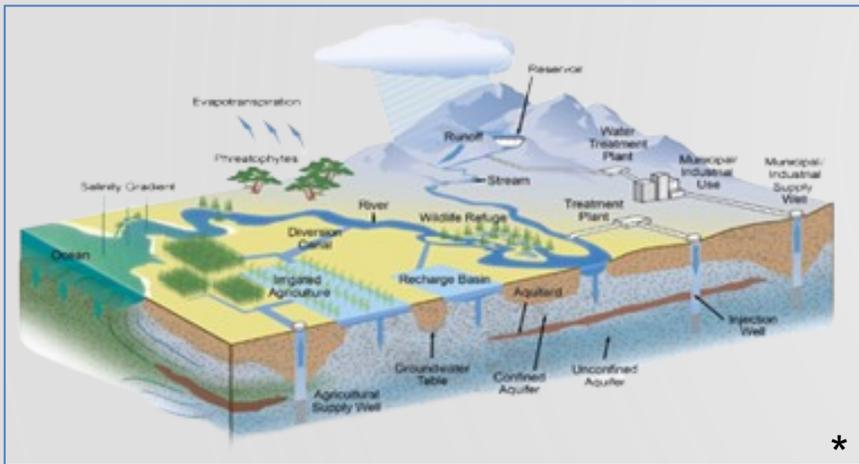


## Session 16: Machine Learning for Water & Environmental Modeling

	Title	Presenter(s)
1	Machine Learning Protocols	Kevin He (DWR)
2	Machine Learning-Based Harmful Algal Blooms Modeling	Gourab Saha (DWR)
3	Machine Learning for Long-Range Streamflow Forecasting	Peyman Namadi (DWR)
4	Exploring Deep Learning Techniques in Daily Streamflow Prediction	Yu-Chieh (Jay) Chao (UCD) & Wyatt Arnold (DWR)
5	Towards Generalization in CalSim Surrogates for Drought, Sea Level Rise, and Landscape Change	Eli Ateljevich (DWR)

# Machine Learning Protocols for Water and Environmental Modeling

CWEMF Annual Meeting, May 13, 2025



Water and Environmental System \*



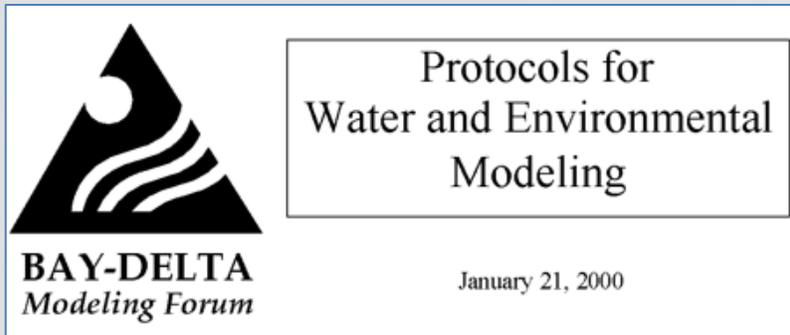
Machine Learning \*\*





# Why do we need ML modeling protocols?

## ➤ Existing Protocols



Model  
Credibility

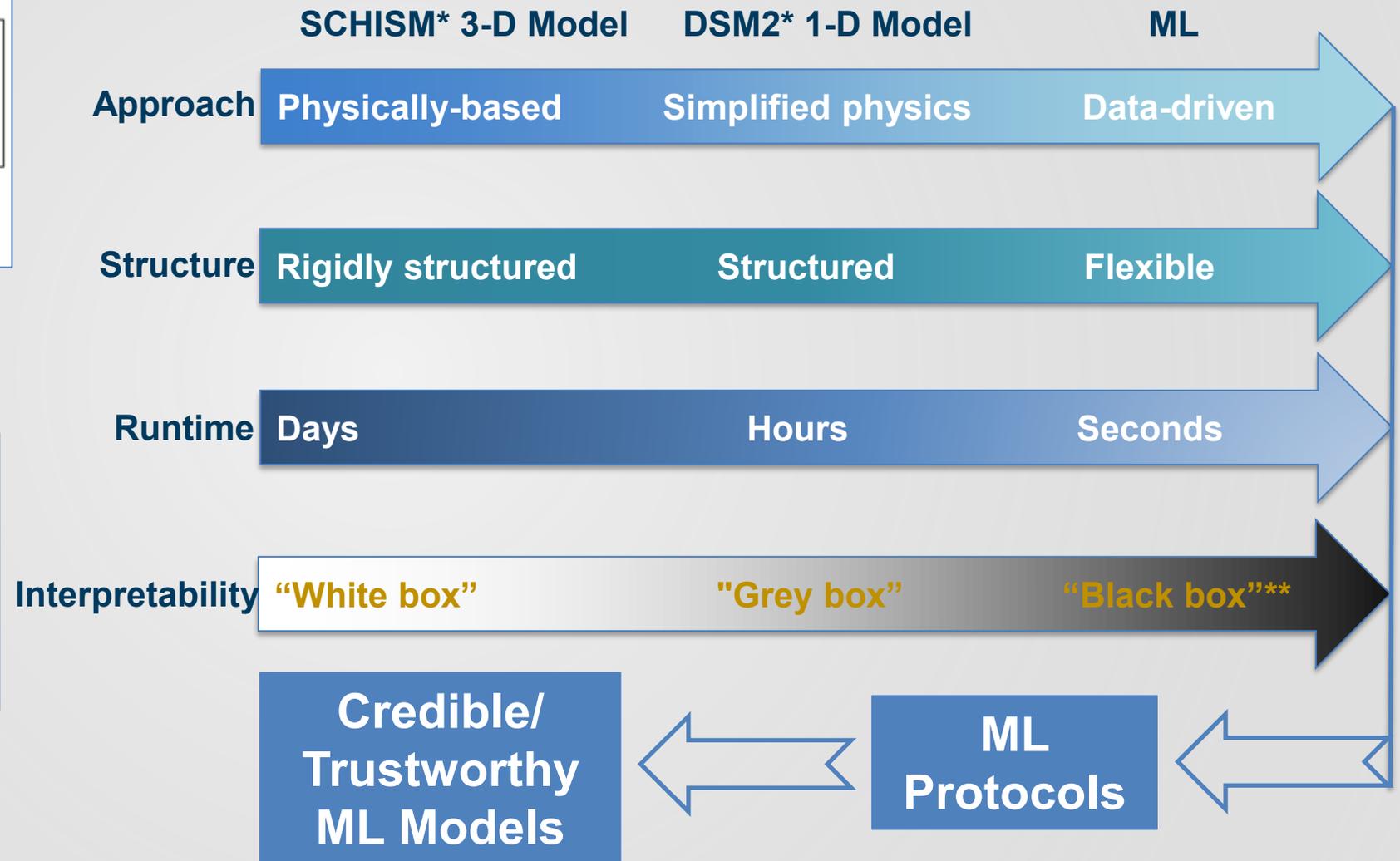


## ➤ AI Terms

**AI:** Artificial Intelligence  
**ML:** Machine Learning  
**DL:** Deep Learning

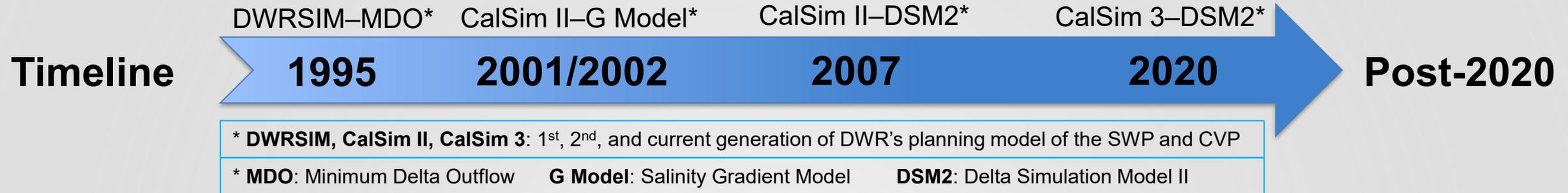


## ➤ Conventional Bay-Delta models vs. ML models

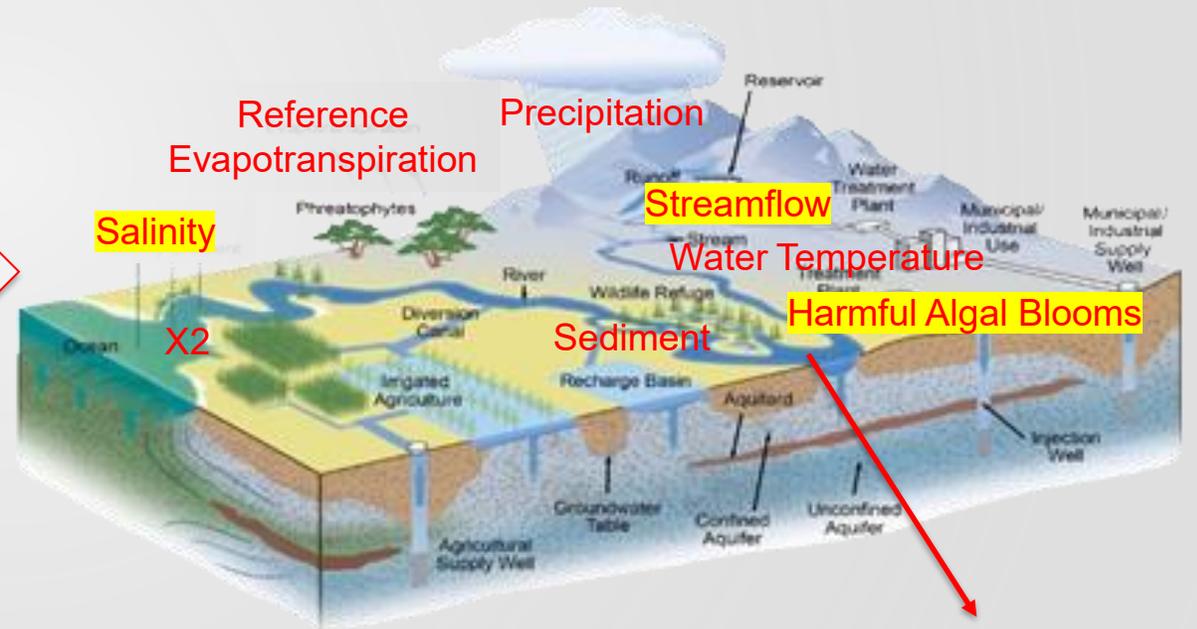


\*SCHISM: Semi-implicit Cross-scale Hydroscience Integrated System Model DSM2: Delta Simulation Model II  
\*\*\* Active research area (e.g., physics-informed ML)

# ML experience of the Modeling Support Office

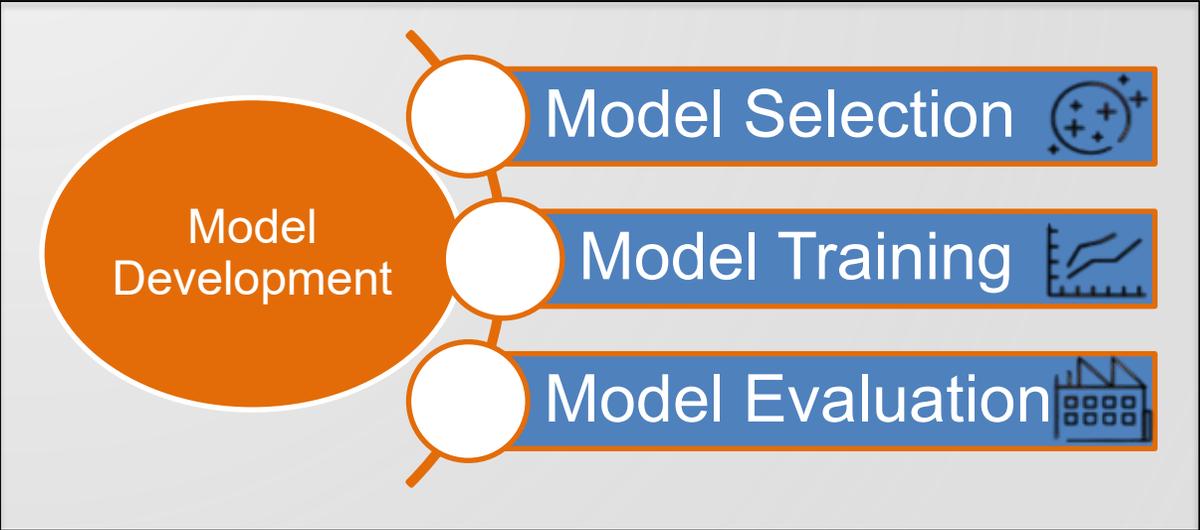
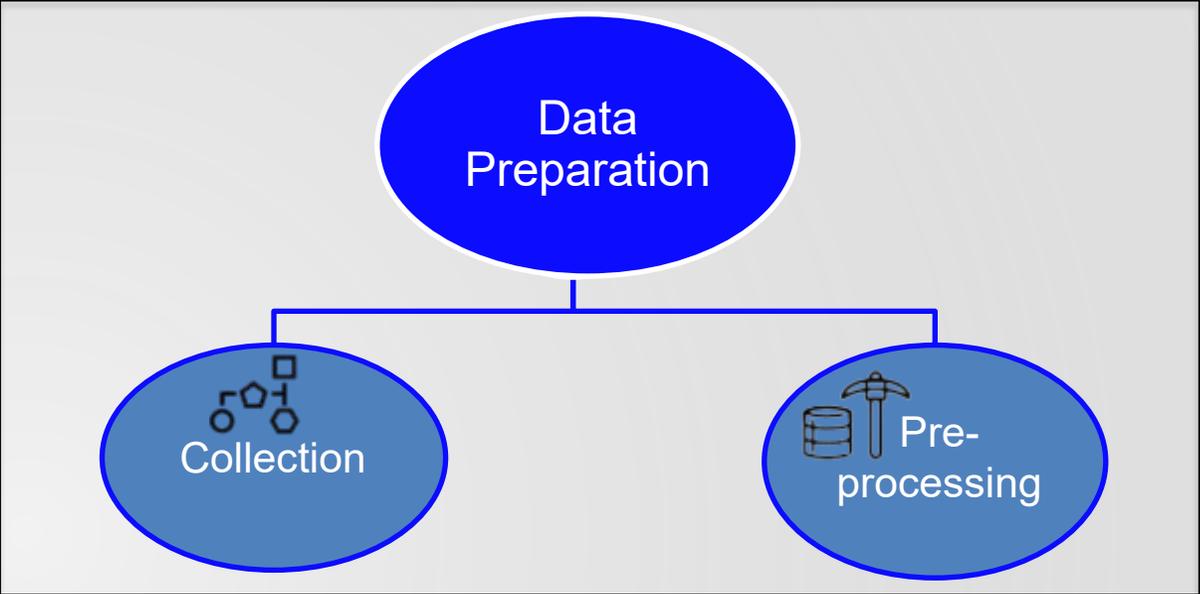
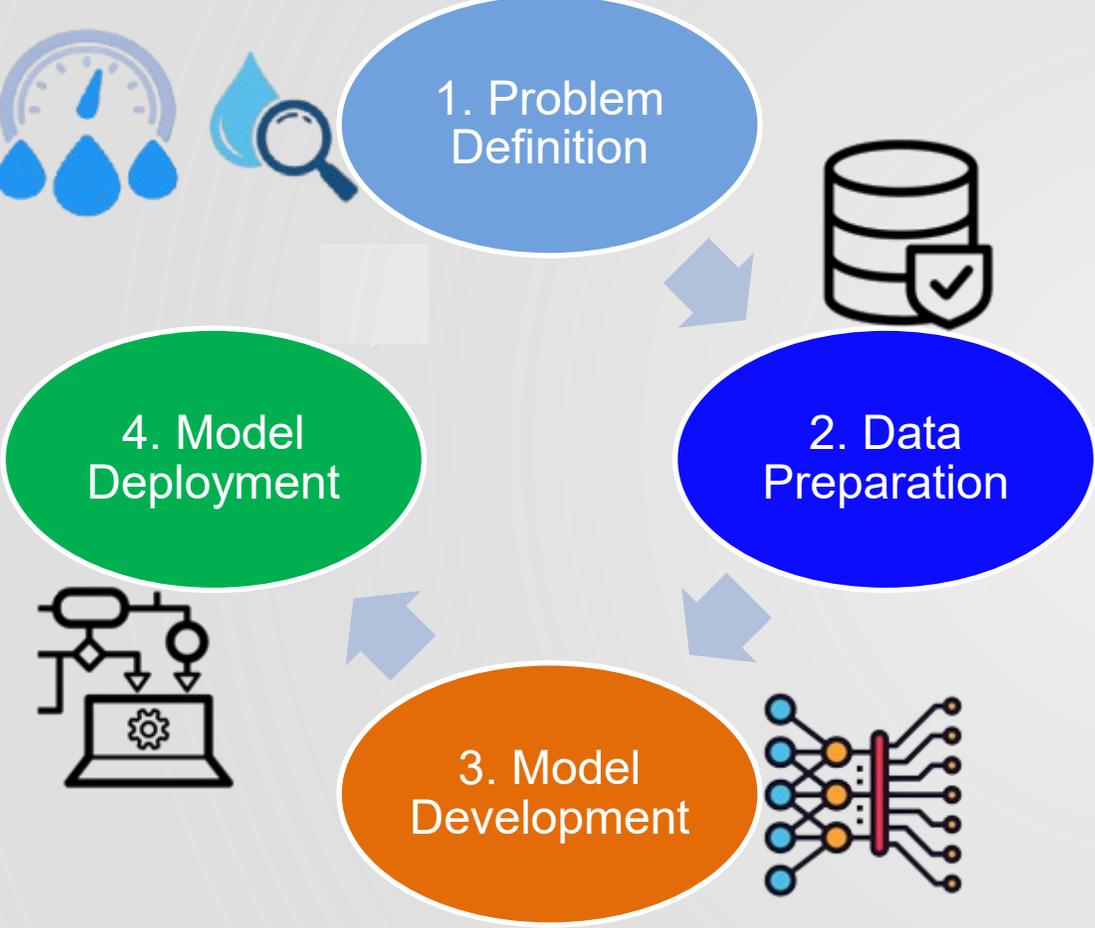


## 2020 - Present

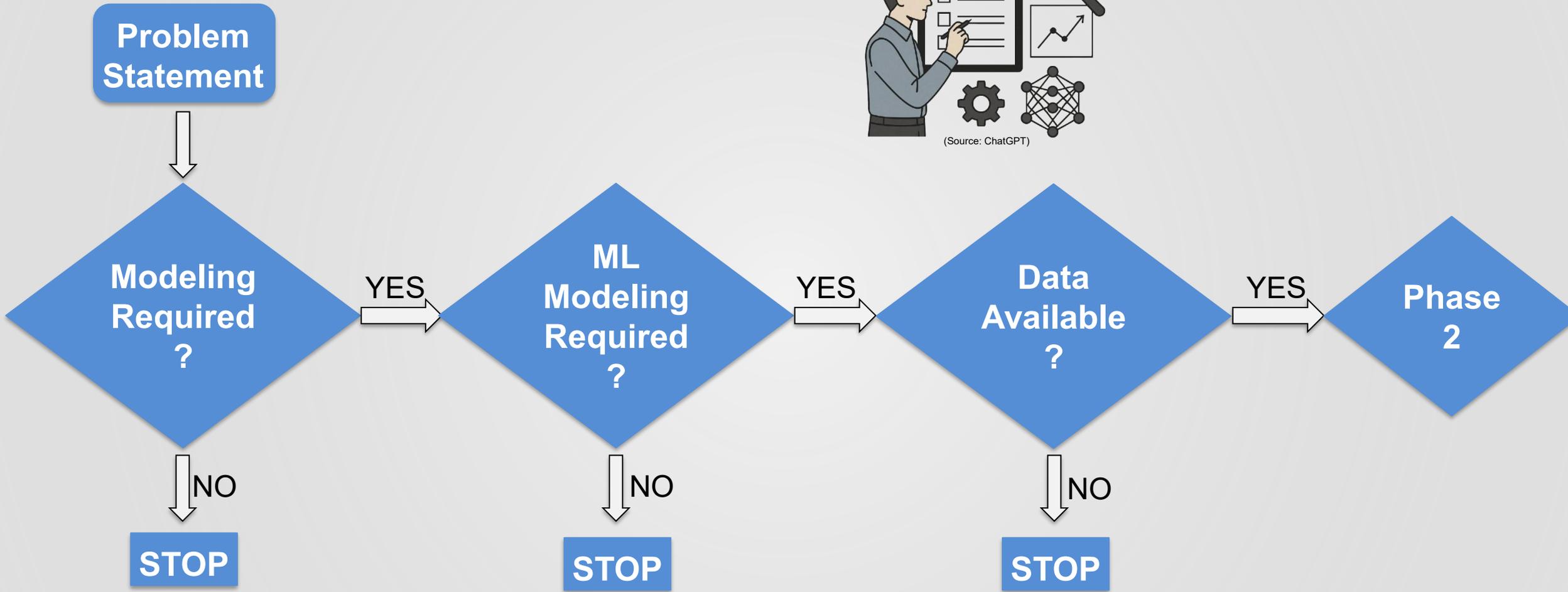


- Alkalinity
- Br<sup>-</sup>
- K<sup>+</sup>
- Mg<sup>2+</sup>
- Ca<sup>2+</sup>
- Hardness
- Na<sup>+</sup>
- SO<sub>4</sub><sup>2-</sup>
- Cl<sup>-</sup>
- Dissolved Boron
- Dissolved Nitrate
- Total Dissolved Solids

# Life-cycle of a typical ML model



# Phase 1: Problem Definition



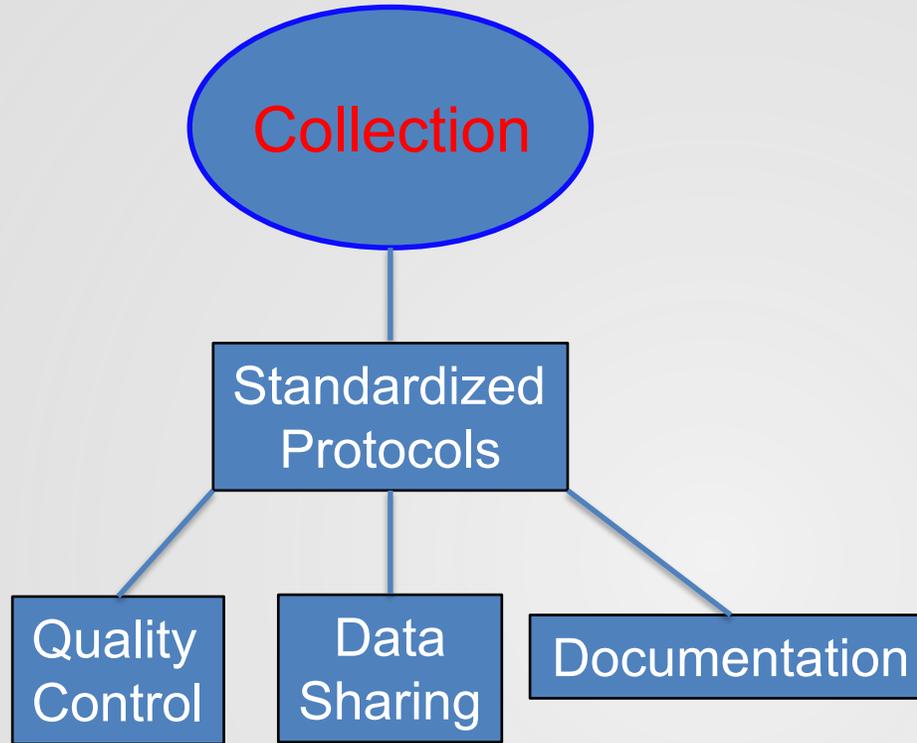
# Phase 2: Data Preparation

## Focus

- Numerical
  - Categorical
- Water quantity;  
water quality;  
fish abundance  
and distribution,  
.....

## Not Included

- Text
- Image
- Audio
- Video



**Example: Assembly Bill 1755 (AB1755)**  
**Open and Transparent Water Data Act**

**California Natural Resources Agency Open Data**

Our mission is to restore, protect and manage the state's natural, historical and cultural resources for current and future generations.

<https://data.cnra.ca.gov/>



**Cleaning**  
-Missing values  
-Outlier  
-Noise



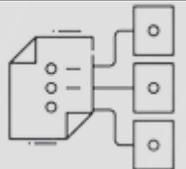
**Transformation**  
-Normalization  
-Standardization

## Pre-processing

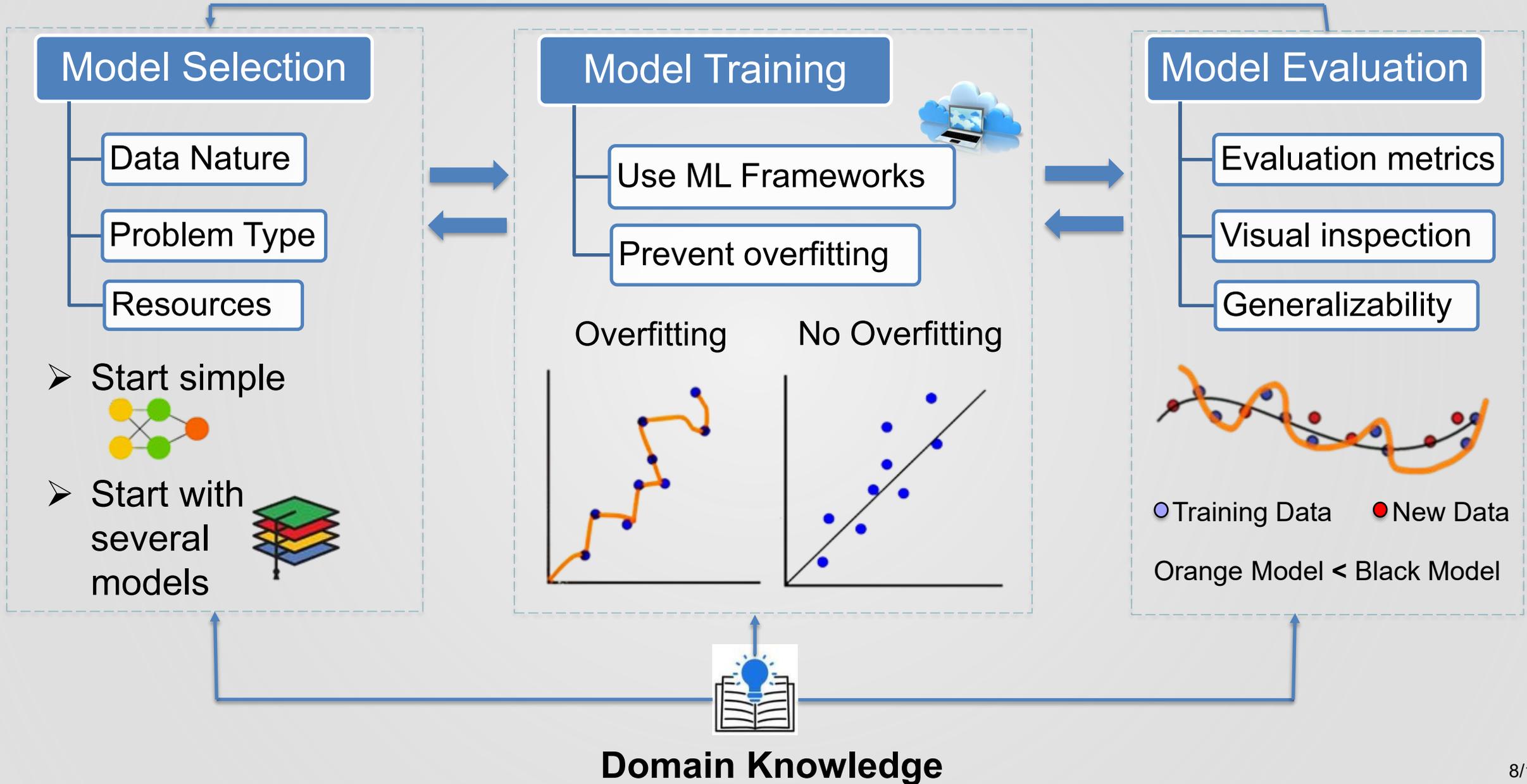
**Feature Engineering**  
-Select inputs  
-Create new inputs



**Splitting**  
-Training  
-Validation  
-Testing



# Phase 3: Model Development

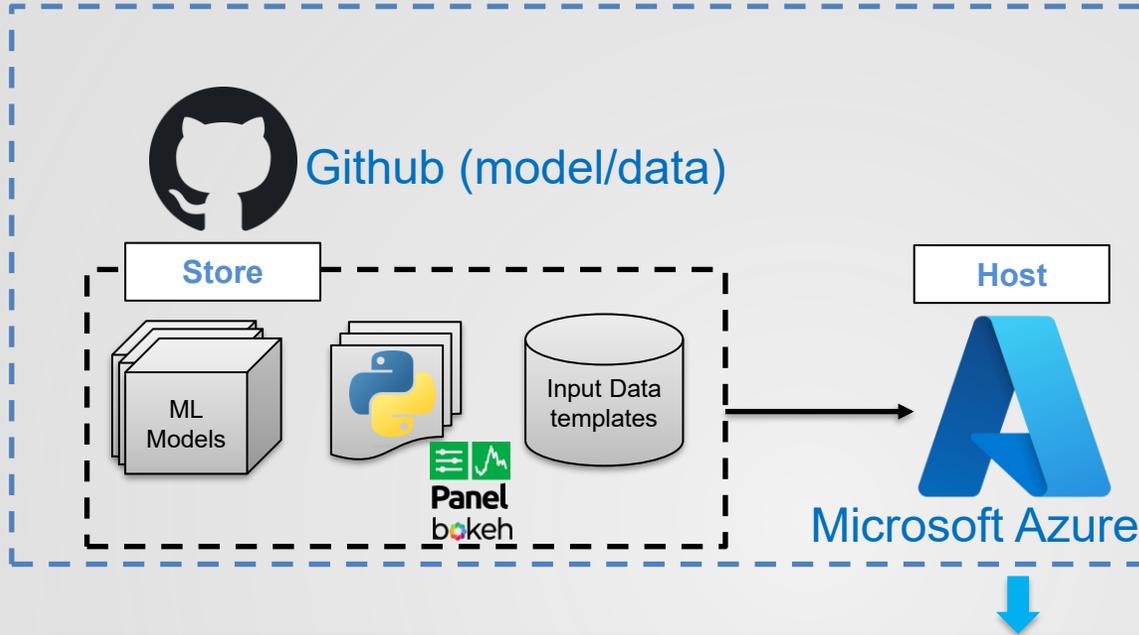


# Phase 4: Model Deployment

## 1. Product

Users

Accessibility



## 2. Documentation

Transparency

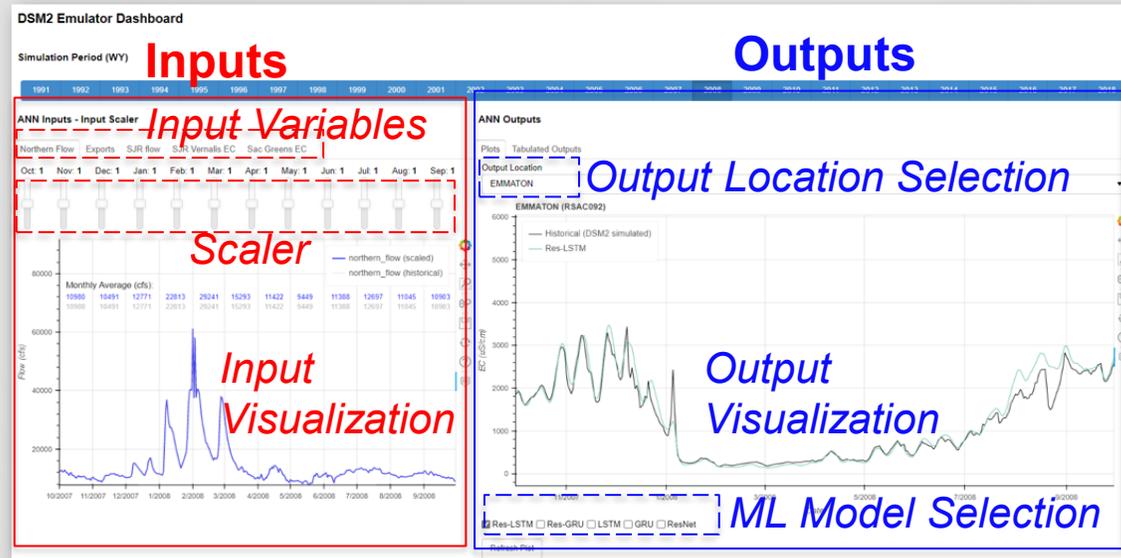
Peer-review

Docs

Code & Data



Example: Salinity Dashboard



## 3. Training

Workshop

Hands-on

Recording



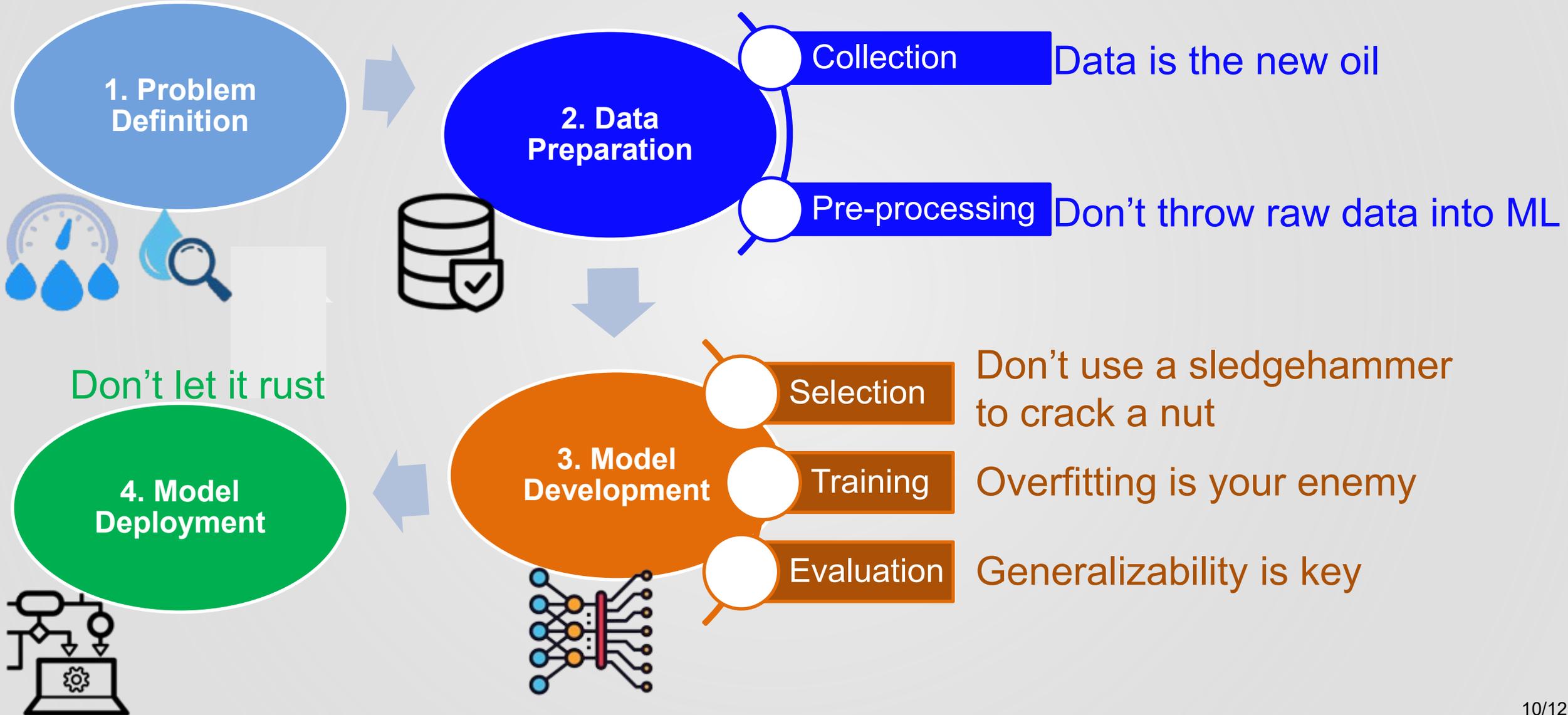
## 4. Maintain

Monitor

Improve

# Take-Home Messages

ML → powerful, fast, but not a replacement



# Technical Report



## Machine Learning Protocols for Water and Environmental Modeling

October 2024



# Peer-Reviewed Article



Open Access Feature Paper Review

## Protocols for Water and Environmental Modeling Using Machine Learning in California



Minxue He, Prabhjot Sandhu, Peyman Namadi, Erik Reyes, Kamyar Guivetchi and Francis Chung

## Acknowledgements

**DWR Modeling Support Office**

**Bay-Delta Modeling Forum (BDMF)**

–Modeling Protocols Committee (2000)

**California Water and Environmental Modeling Forum (CWEMF)**

–Modeling Protocols Committee (2021)

**Machine Learning Community**

# Questions?

