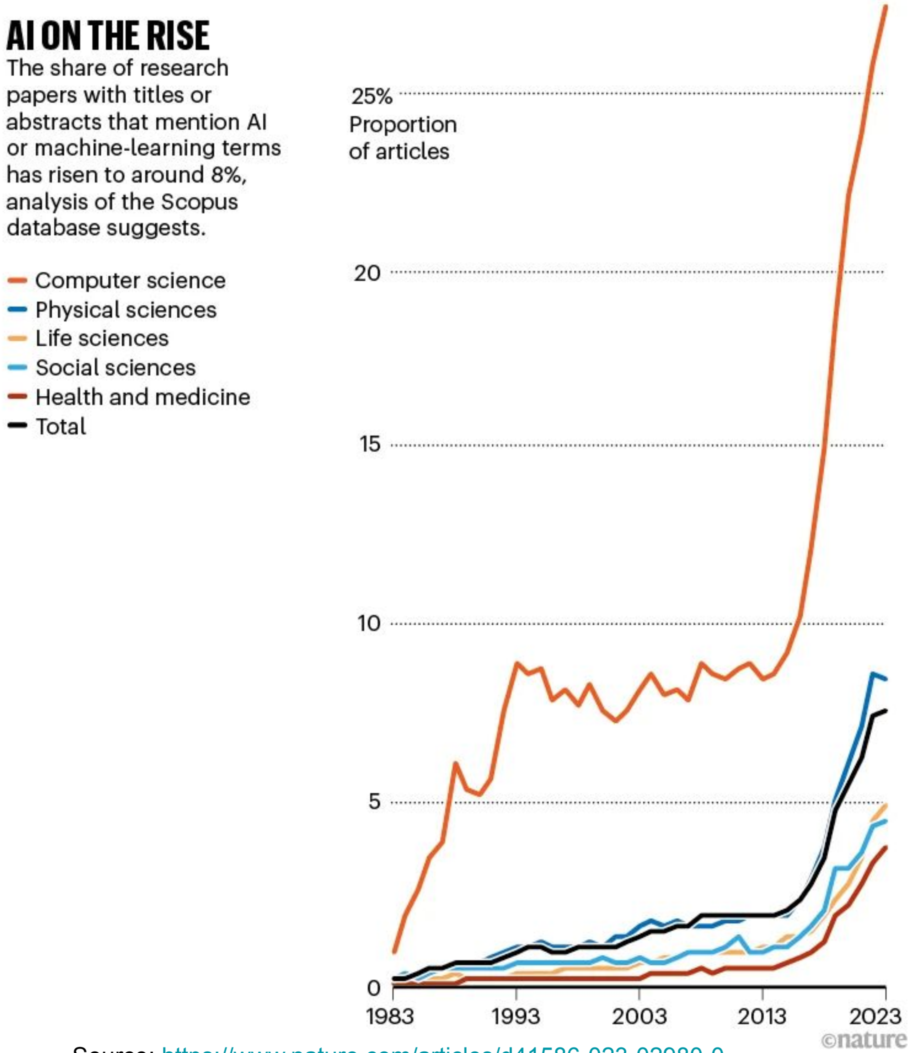
Supervised Machine Learning in Science

Justification and puzzle pieces for machine learning in science

Christoph Molnar & Timo Freiesleben

UQSay #64, Nov 2nd 2023

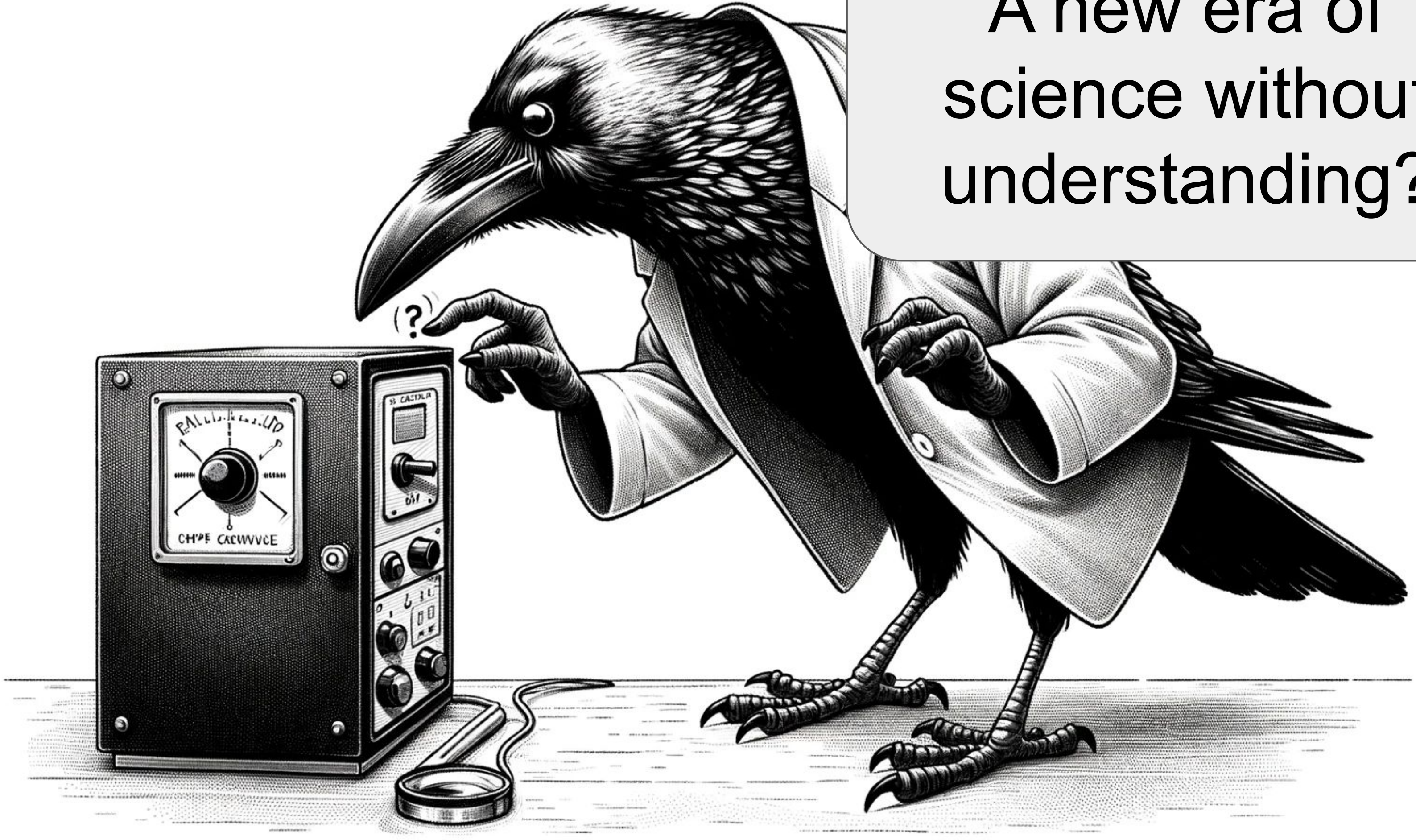
Machine Learning has arrived in Science



Source: https://www.nature.com/articles/d41586-023-02980-0

Biggest Benefit:
Saving time

Second Benefit:
Enabling new
research

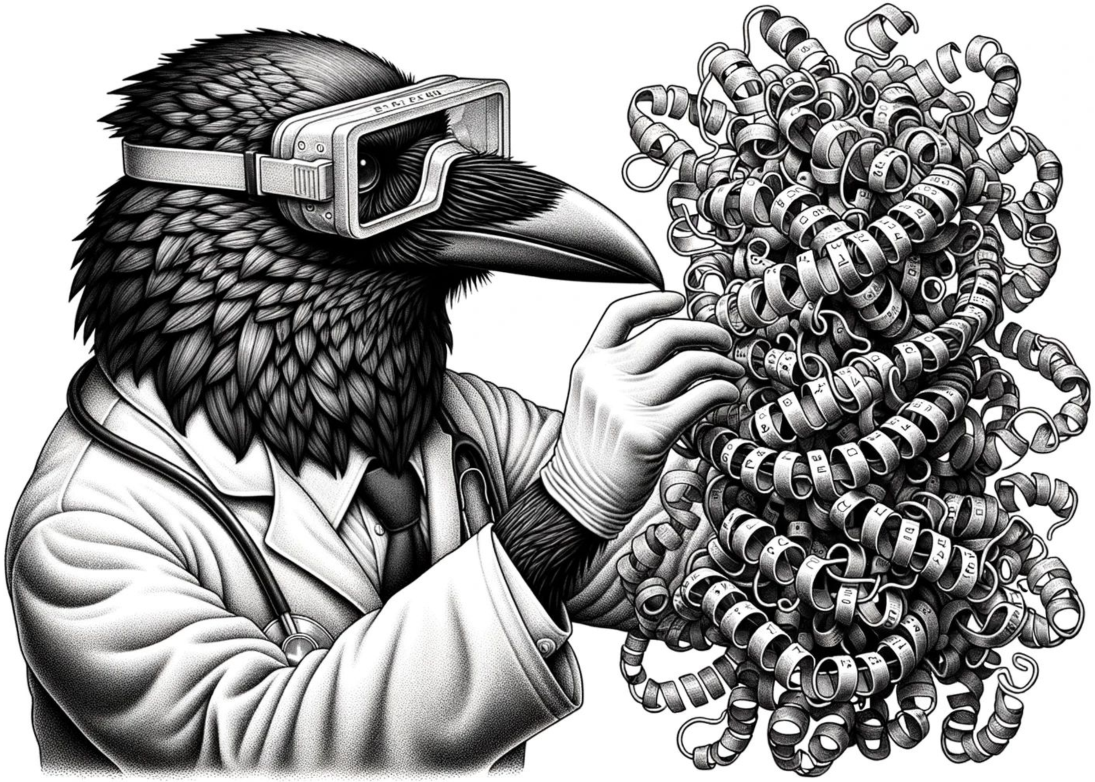


A new era of science without understanding?

Predict Tornadoes

Lagerquist, et al.
“Deep Learning on Three-Dimensional Multiscale Data for Next-Hour Tornado Prediction.”
Monthly Weather Review (2020)

Alpha Fold



Jumper et al.
“Highly Accurate Protein Structure Prediction with AlphaFold.”
Nature
(2020)

Predict Almond Yield

Zhang et al. “California Almond Yield Prediction at the Orchard Level with a Machine Learning Approach.” *Frontiers in Plant Science* (2019)

The role of prediction in science

Clash with other goals

- Machine learning….
  - is just curve fitting but does not guide actions
    ~Pearl
  - predicts well but does not explain
    ~Shmueli
  - does not help us reason about language and semantics
    ~Chomsky

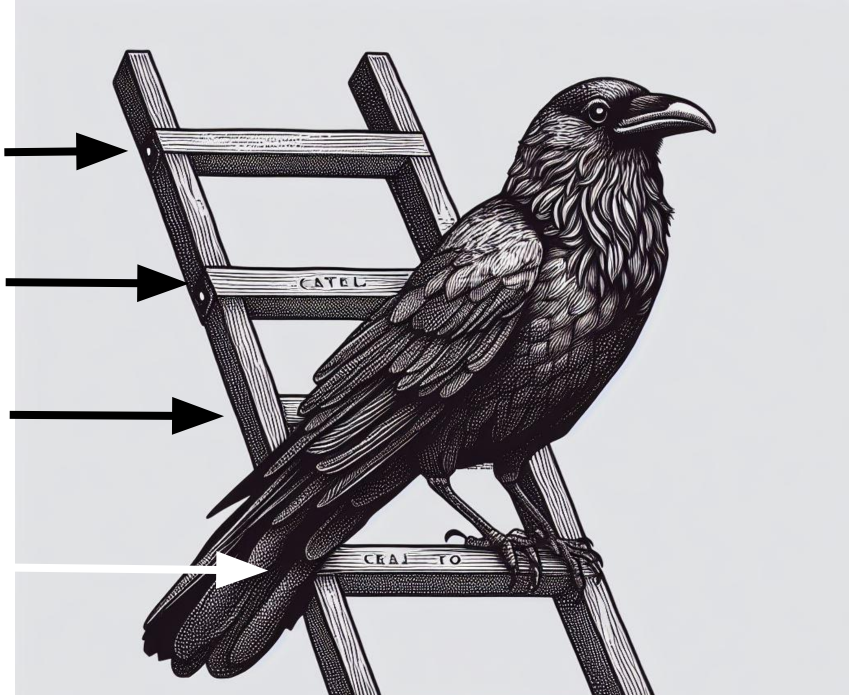
Climbing up the ladder

reasoning

explanation

control

observation



All in all, we think it’s justified to use ML in science…

- ML adapts your model to the world and not the world to your model
- ML can handle all data structures
- ML allows us to work on new questions

- Other justifications:
  - time-efficiency
  - computationally cheaper than solving differential equations
  - basis for theory building

…but we have to fix the following!

- Domain knowledge is overlooked
- Lack of interpretability and explanations
- No causal understanding
- Limited robustness

- Other limitations:
    - No uncertainty quantification
    - Garbage in, garbage out
    - Lack of standards for reporting model results
    - Lack of standards for reproducibility

Infusing Domain Knowledge

Infuse Domain Knowledge

Obvious

- Choice of target
- Choice of features
- Choice of task
- Choice of data
- Feature engineering

Less Obvious:

- Model constraints
  - Linearity
  - Monotonicity
  - Cyclicity
- Multi-objective optimization
- Think about inductive biases
- Design the loss function

Domain Knowledge ⇔ Model is a two-way street. We don’t only infuse domain knowledge but also get an evaluation of it (for the prediction task).

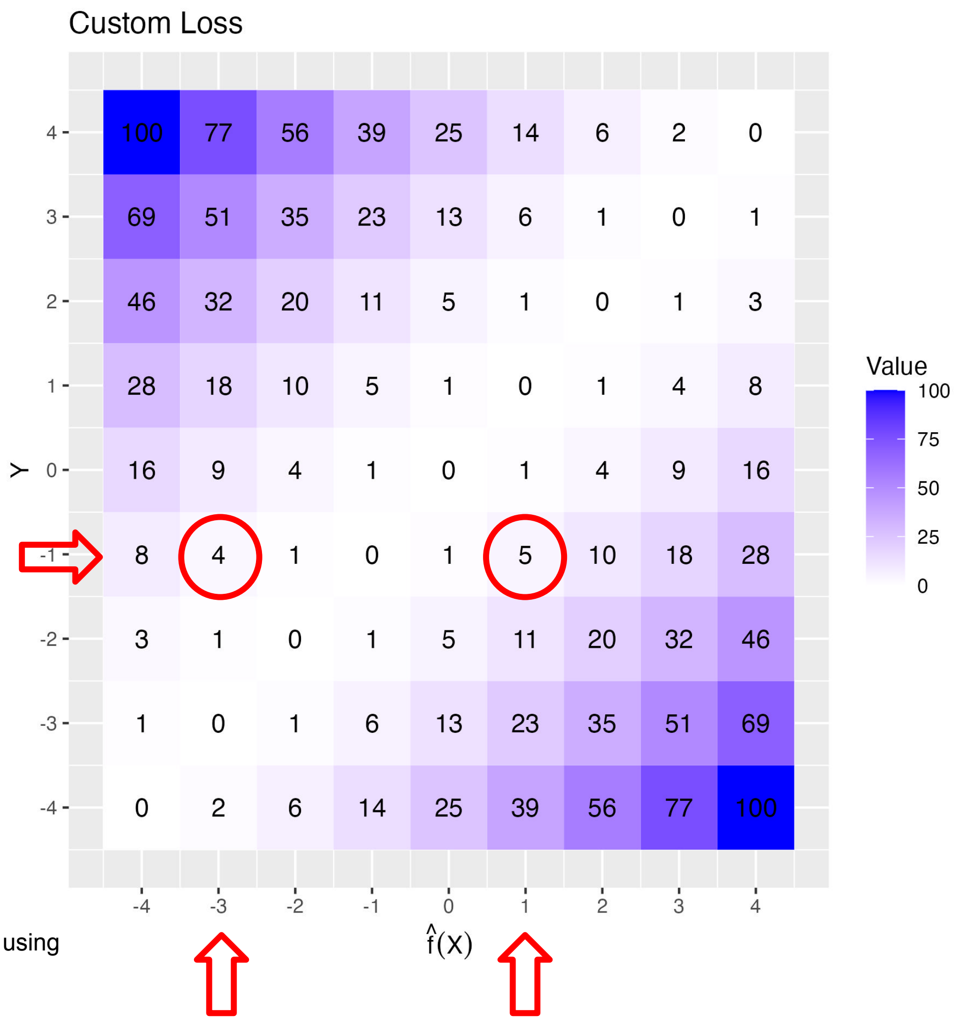
Example: Custom Loss

Predict Parkinson disease severity from

- -4 (severe slowing of movements) to
- 0 (okay) to
- +4 (severe excessive movements)

Custom loss:

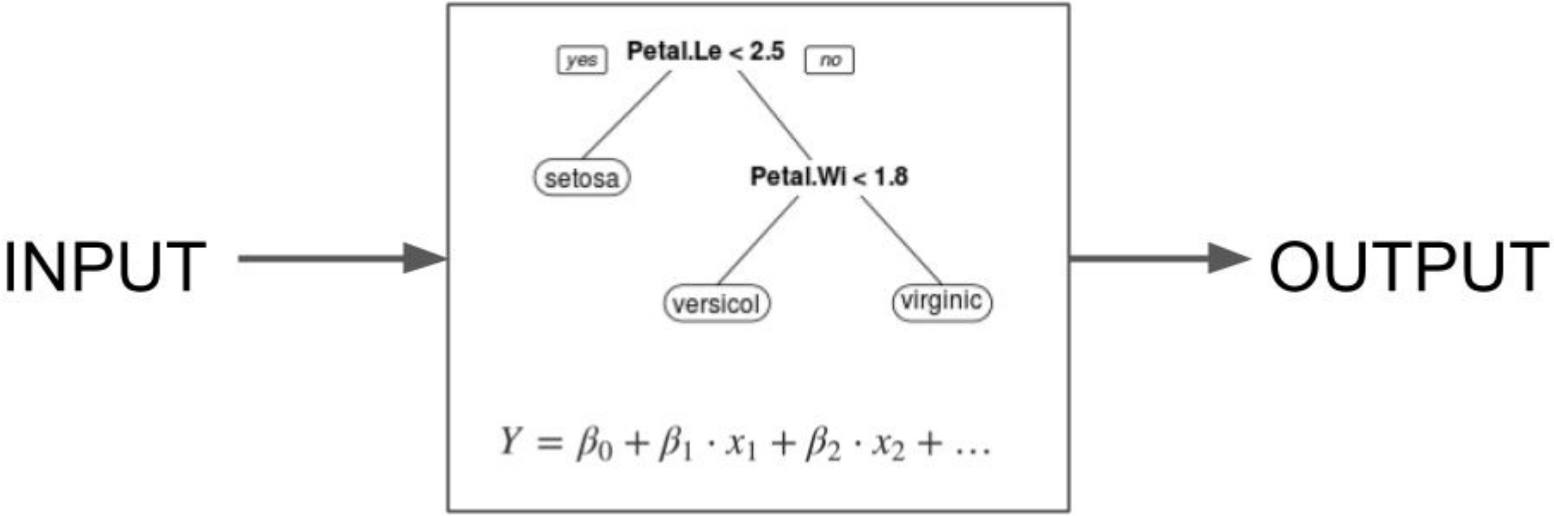
- L2 as starting point
- Asymmetry added: Underestimation of severity expensive
- Penalty for wrong direction
- Scaled between 0 and 100



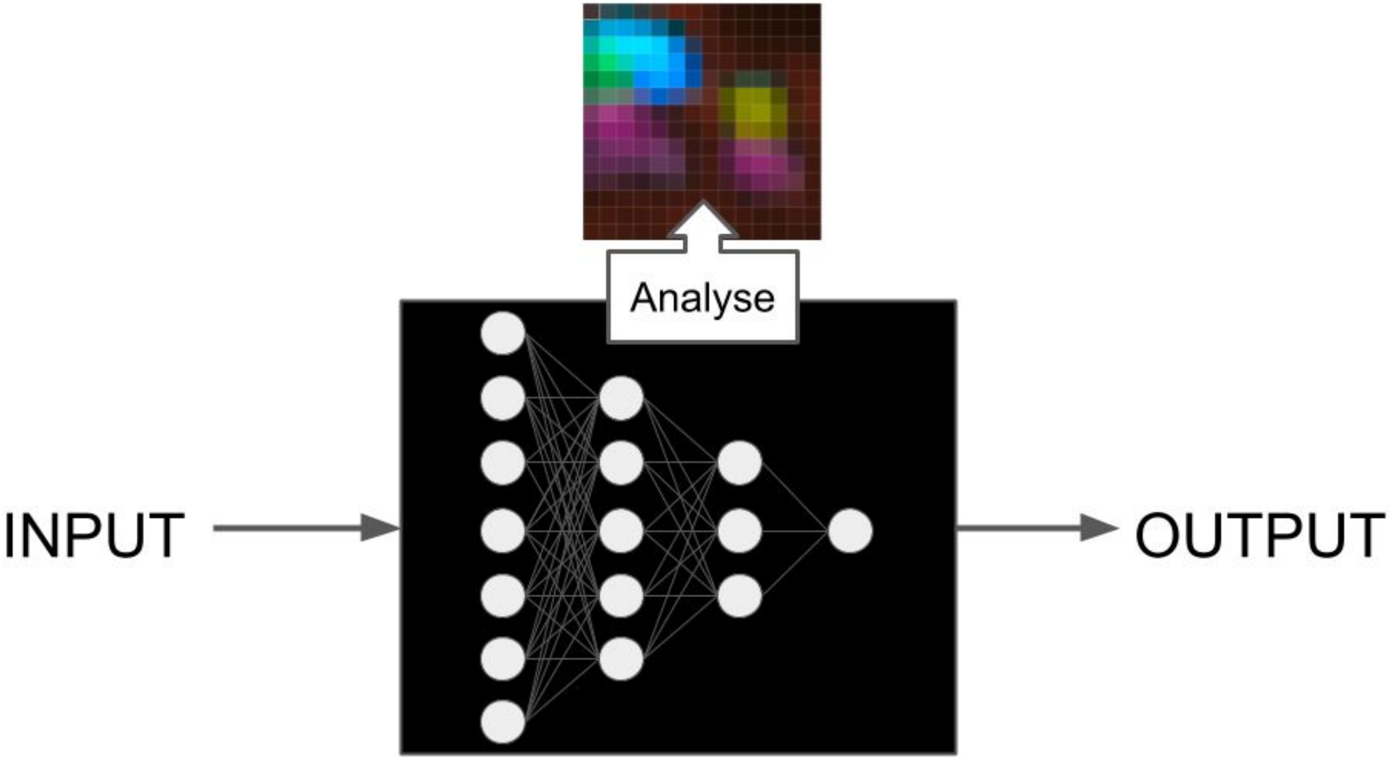
Goschenhofer, Jann, et al. "Wearable-based parkinson's disease severity monitoring using deep learning." *ECML* (2019)

Interpretability

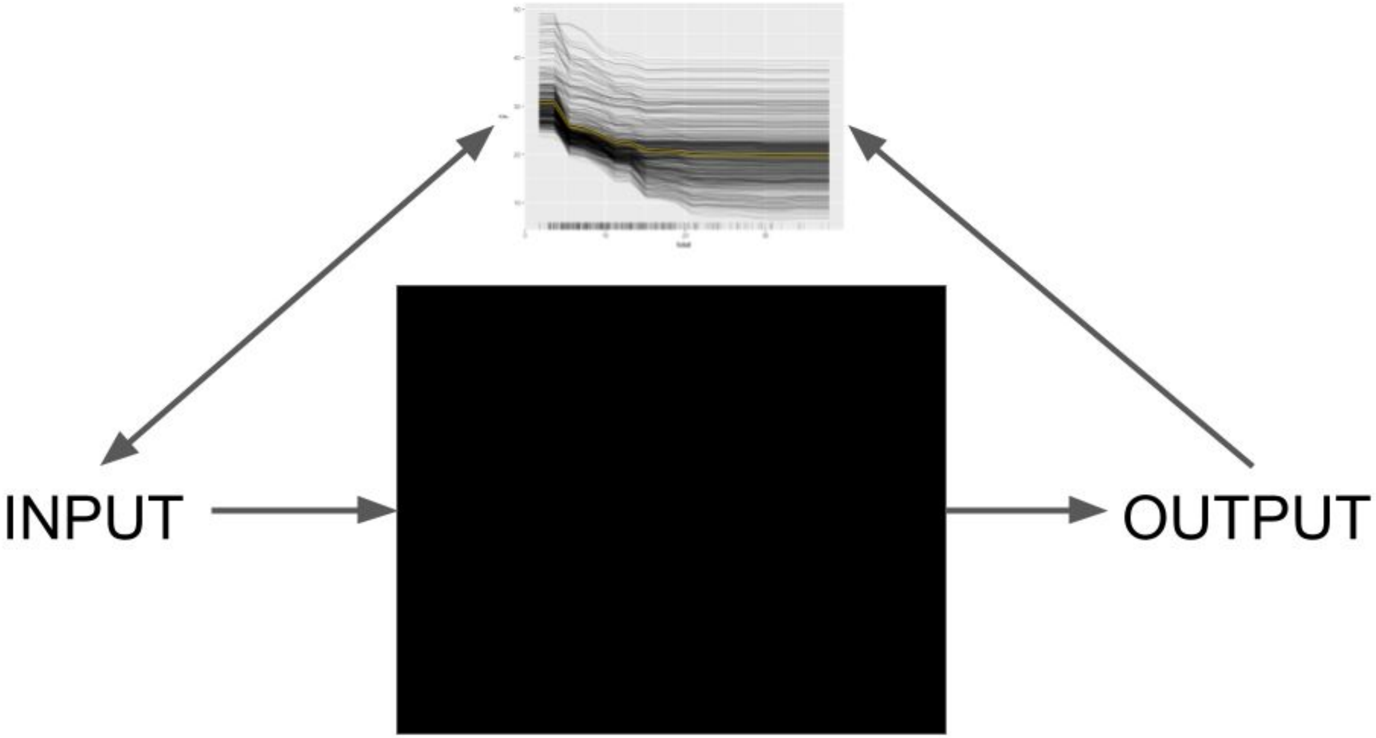
Inherently Interpretable Models



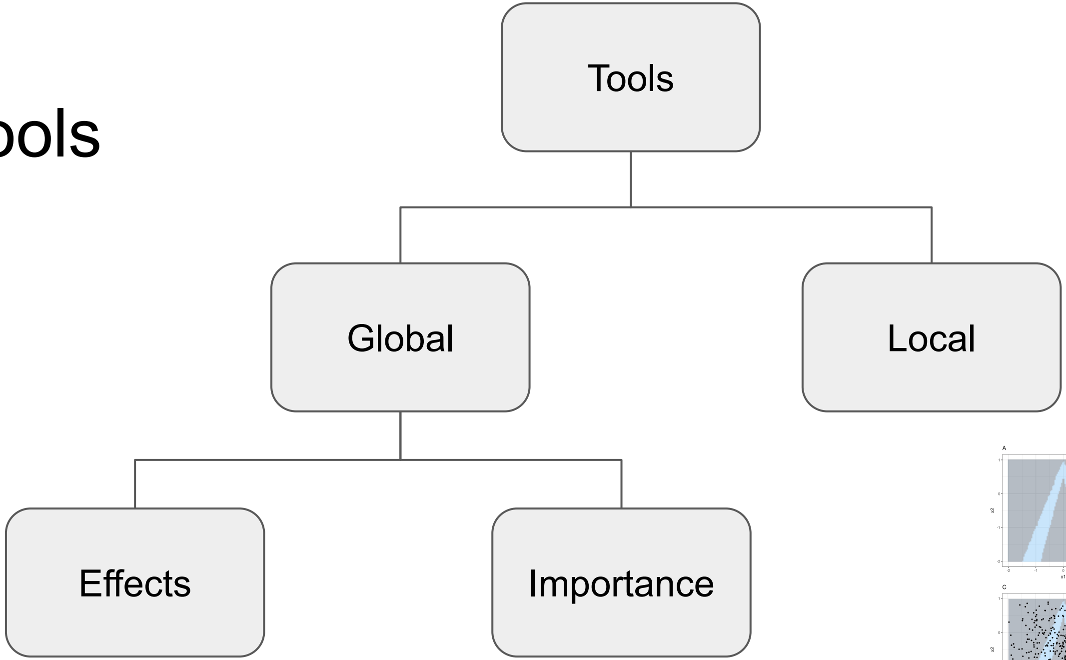
Model-Specific Interpretability

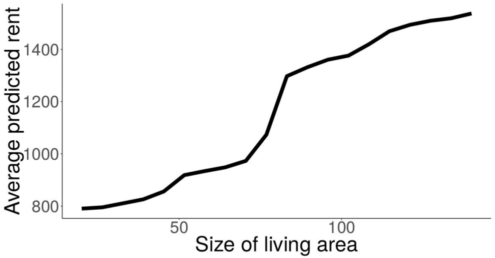


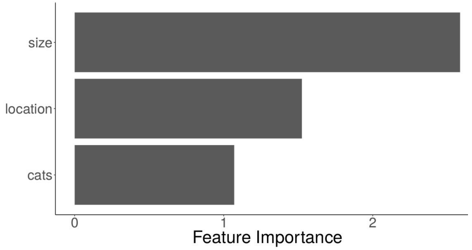
Model-Agnostic Interpretability

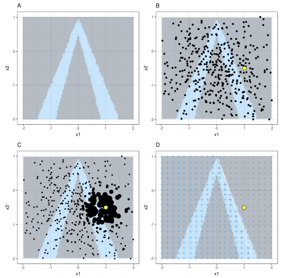


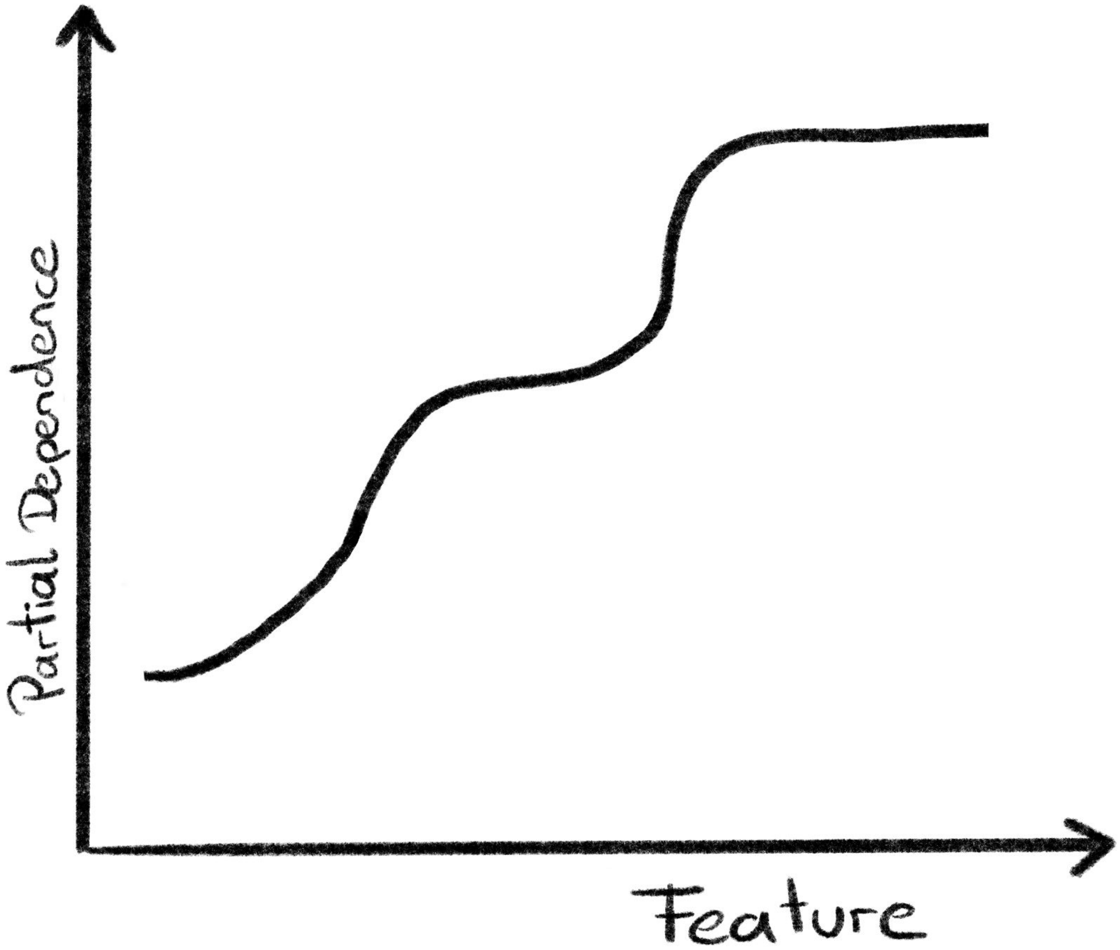
Model-agnostic tools

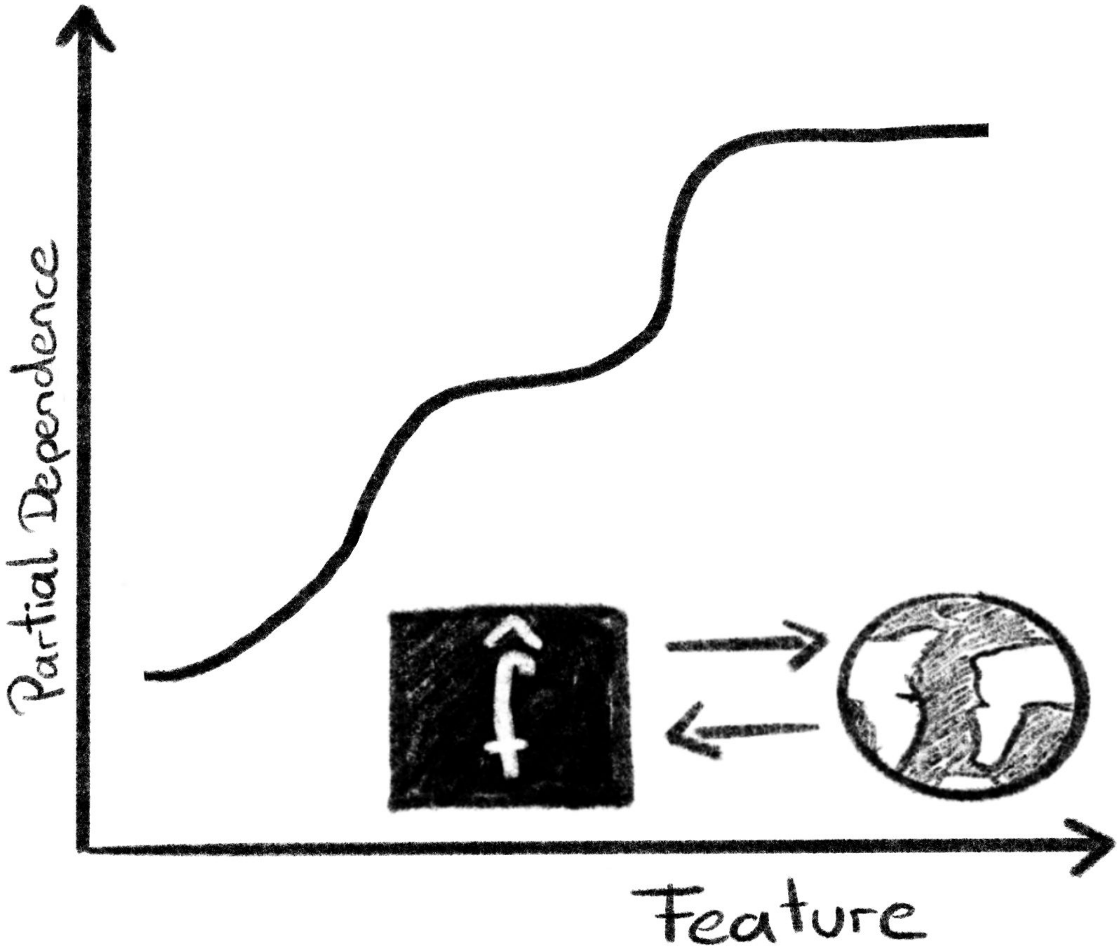


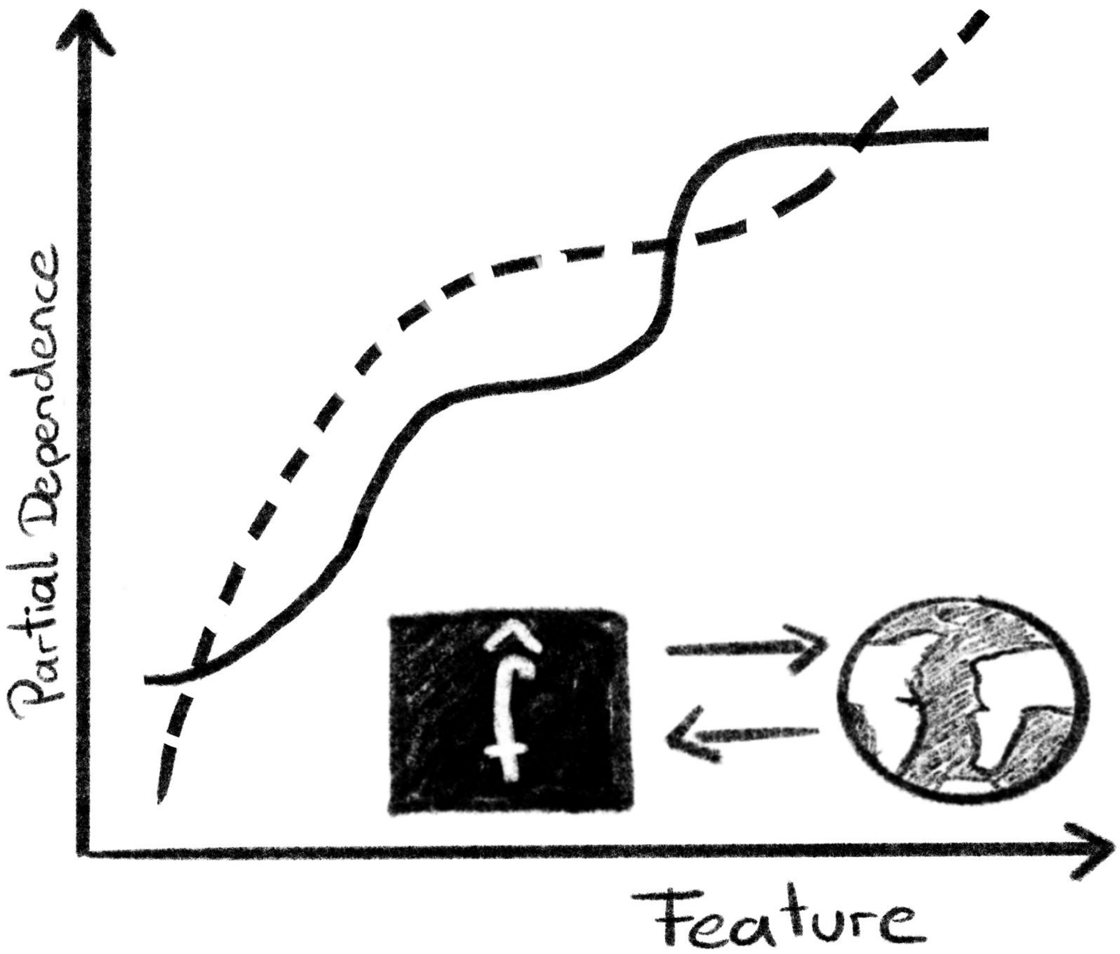


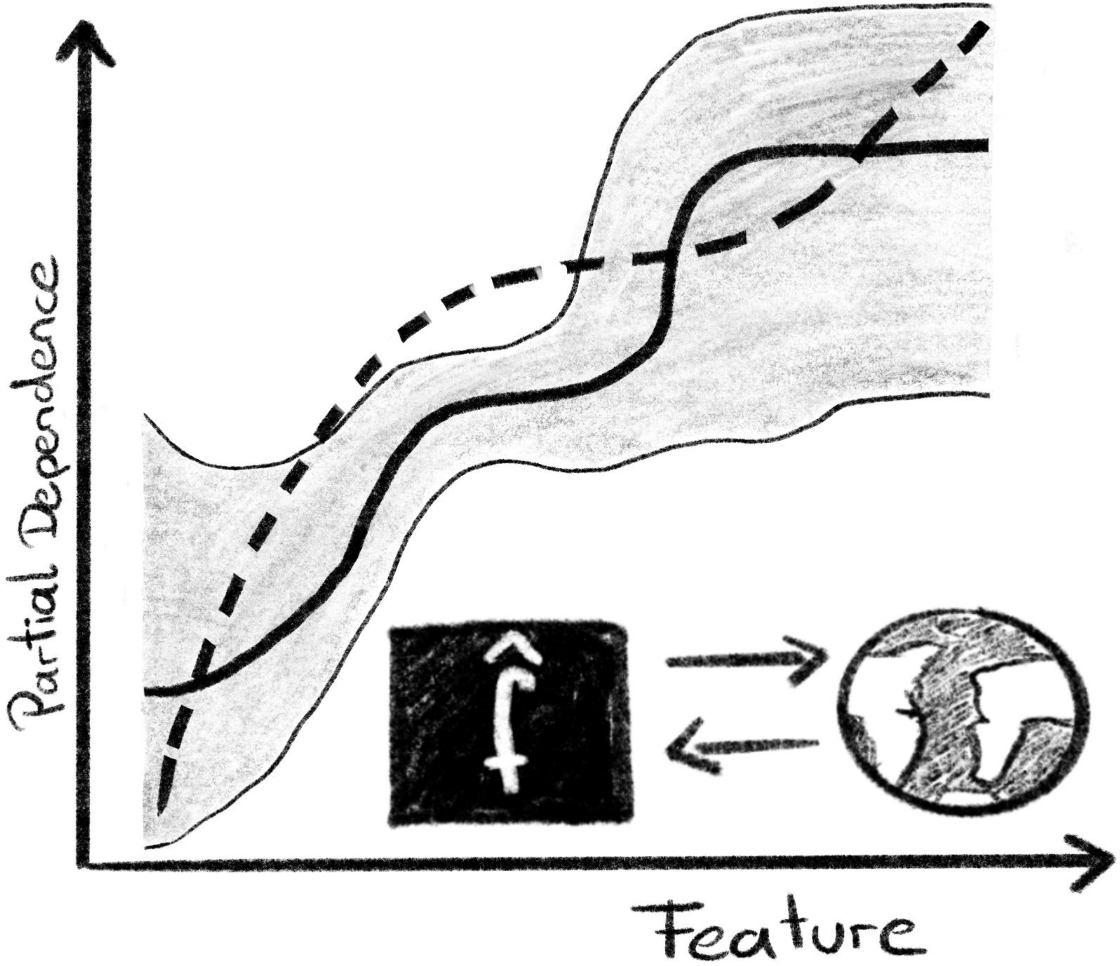





Partial Dependence
Feature

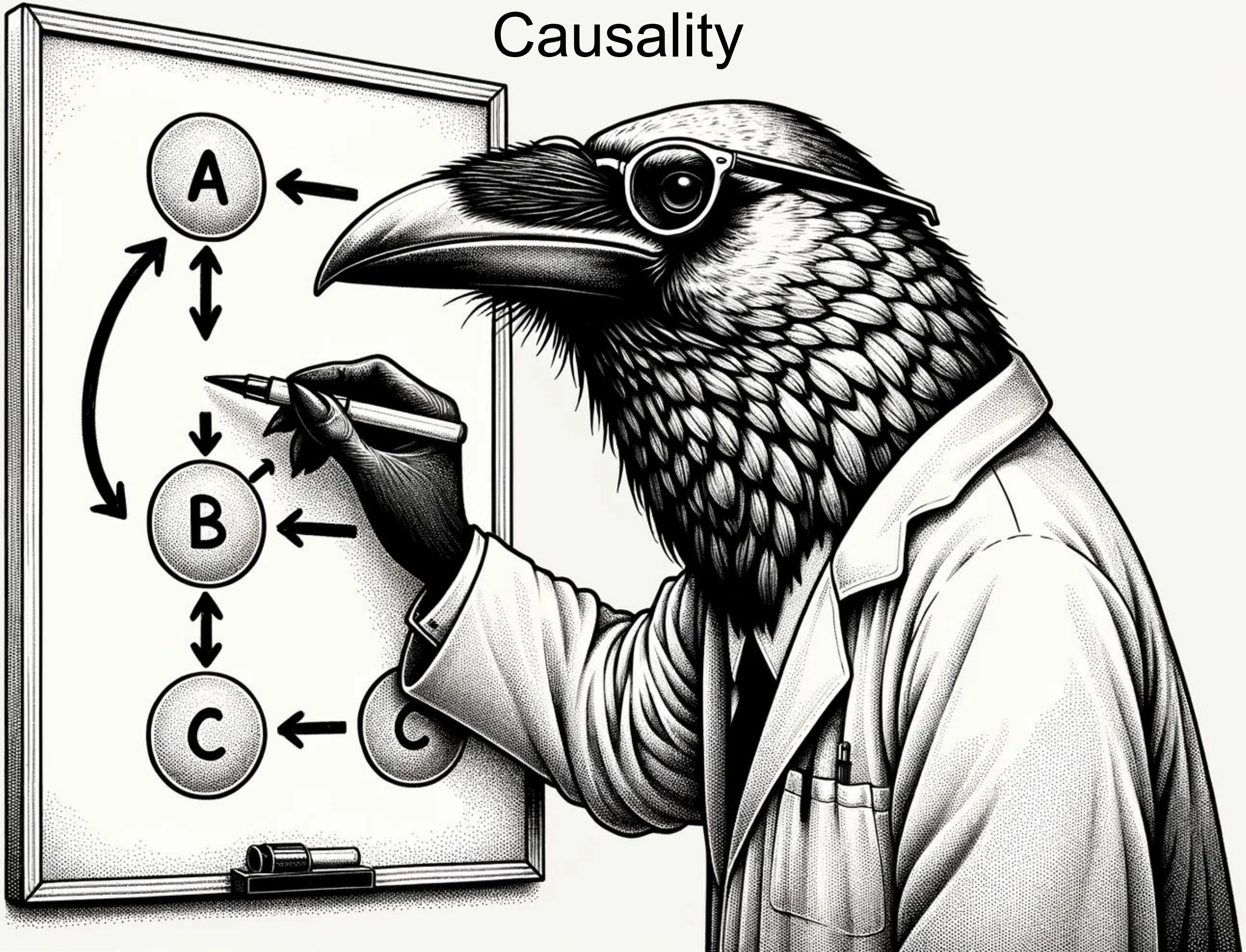

Partial Dependence
Feature

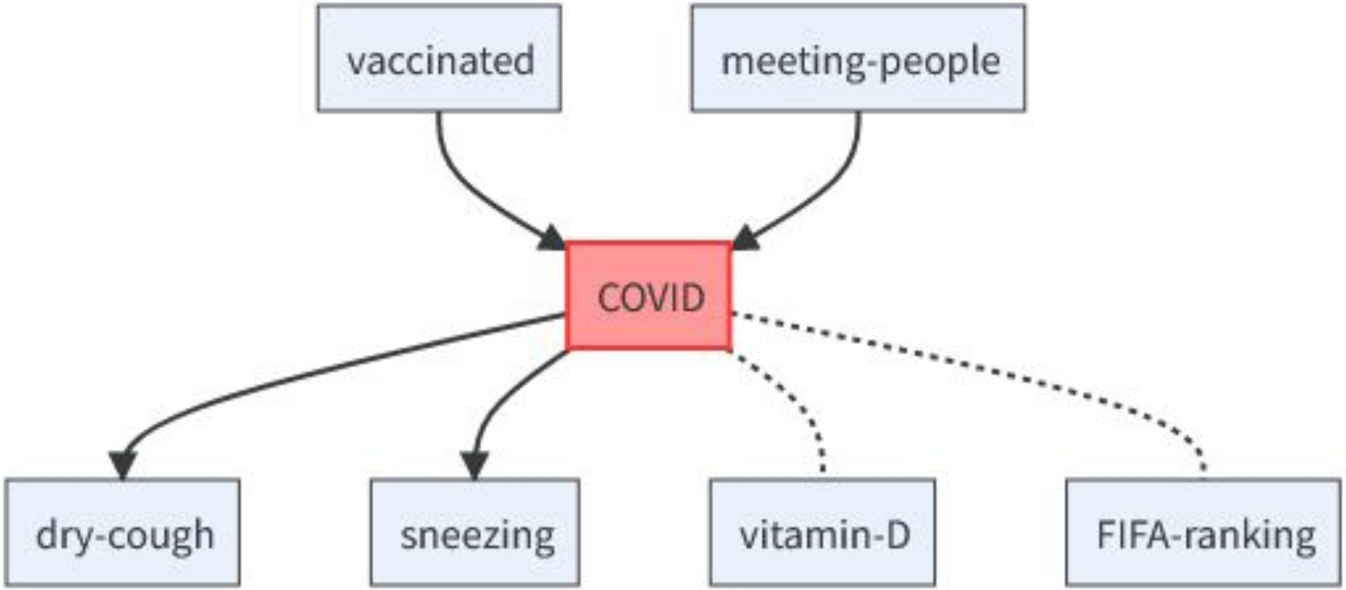

Partial Dependence
Feature



Molnar, Freiesleben, König et al. "Relating the partial dependence plot and permutation feature importance to the data generating process.", *World XAI Conference* (2023).

Causality

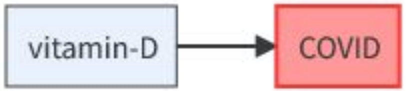


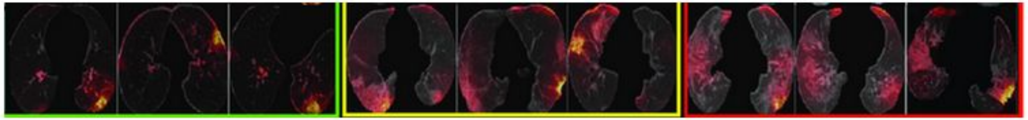


König, Freiesleben &Grosse-Wentrup
"Improvement-Focused Causal Recourse",
AAAI
(2023).

Can ML help to learn causality?

1. *Forming causal hypotheses from associations*



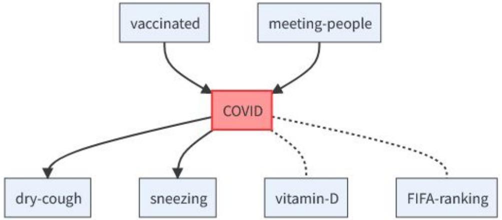


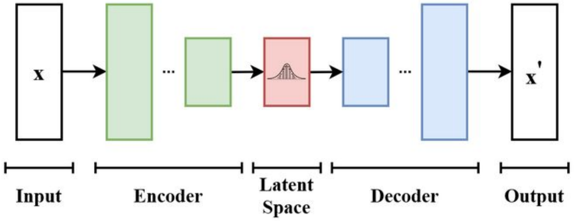
2. *Estimating causal effects with causal graphs*

$$\text{ATE} = \mathbb{E}[Y \mid do(T) = 1] - \mathbb{E}[Y \mid do(T) = 0]$$

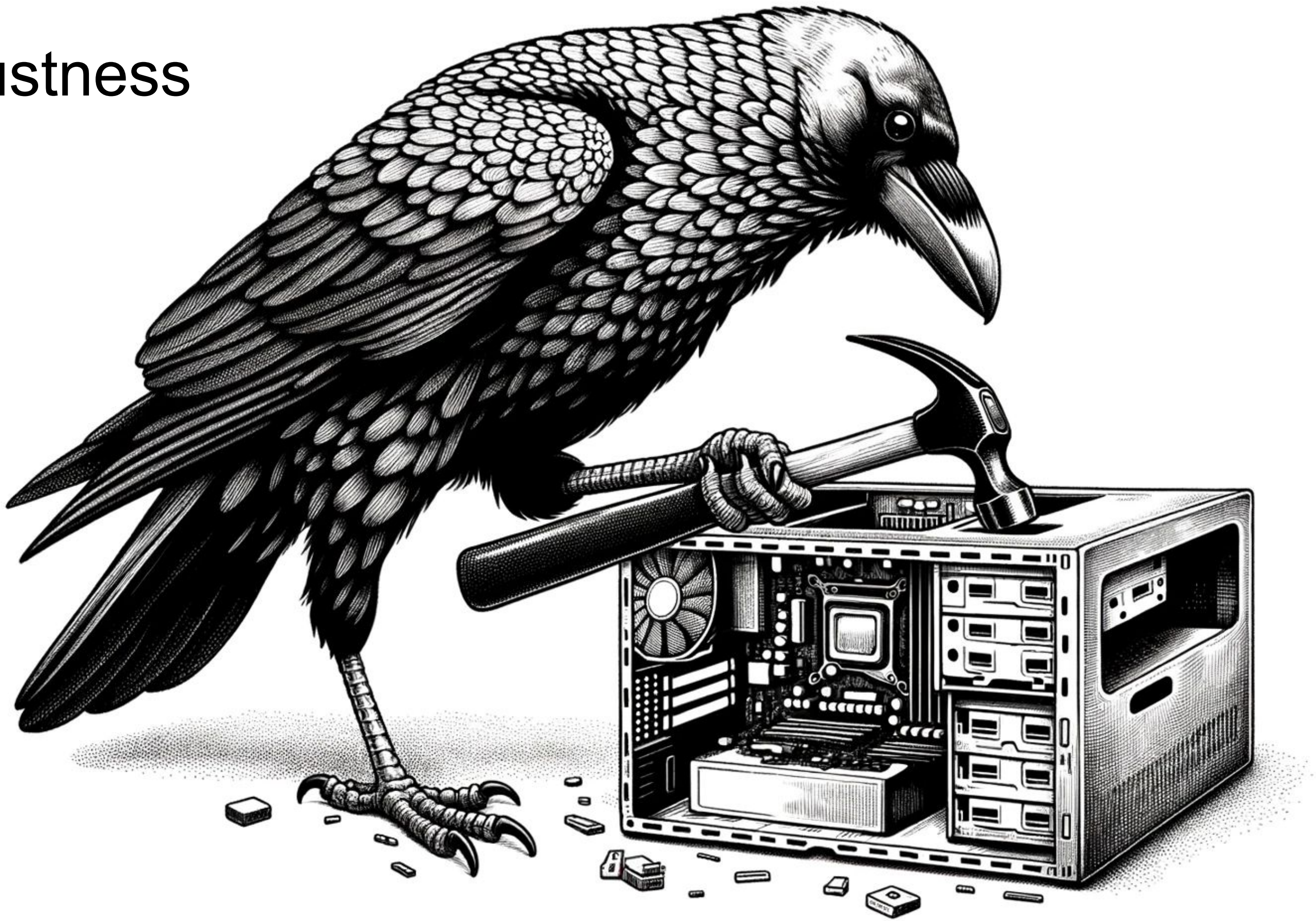
S-Learner, T-Learner, Double ML, etc.

3. *Learning causal relations and variables using assumptions*





Robustness



What means robustness?

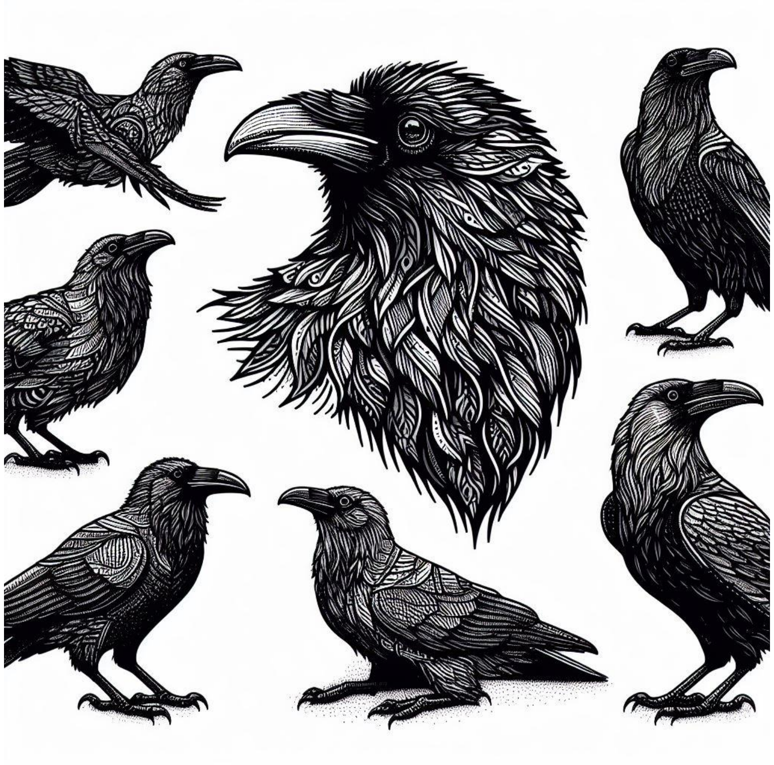
- *Robustness target* = what should be robust? (e.g. deployment perf.)
- *Modifier* = to what is it robust? (e.g. deployment dist.)
- *Modifier domain* = what changes do we expect? (e.g. natural shift)
- *Target tolerance* = how stable must the target be? (e.g. ε=1%)

Freiesleben & Grote
"Beyond Generalization:
A Theory of Robustness
in Machine Learning"
Synthese (2023).

How to deal with robustness?

- While (0!=1)
    i) (Reactive) Analyze the source of the distribution shift, or
    (Proactive) Anticipate new shifts
    ii) Audit your model
    iii) Robustify your model

Data augmentation, invariances, and generators



Other Topics

- Representativeness
- Time and Space
- Ablation Studies
- Reproducibility
- Reporting Results
- ???

Open Book Coming Soon!