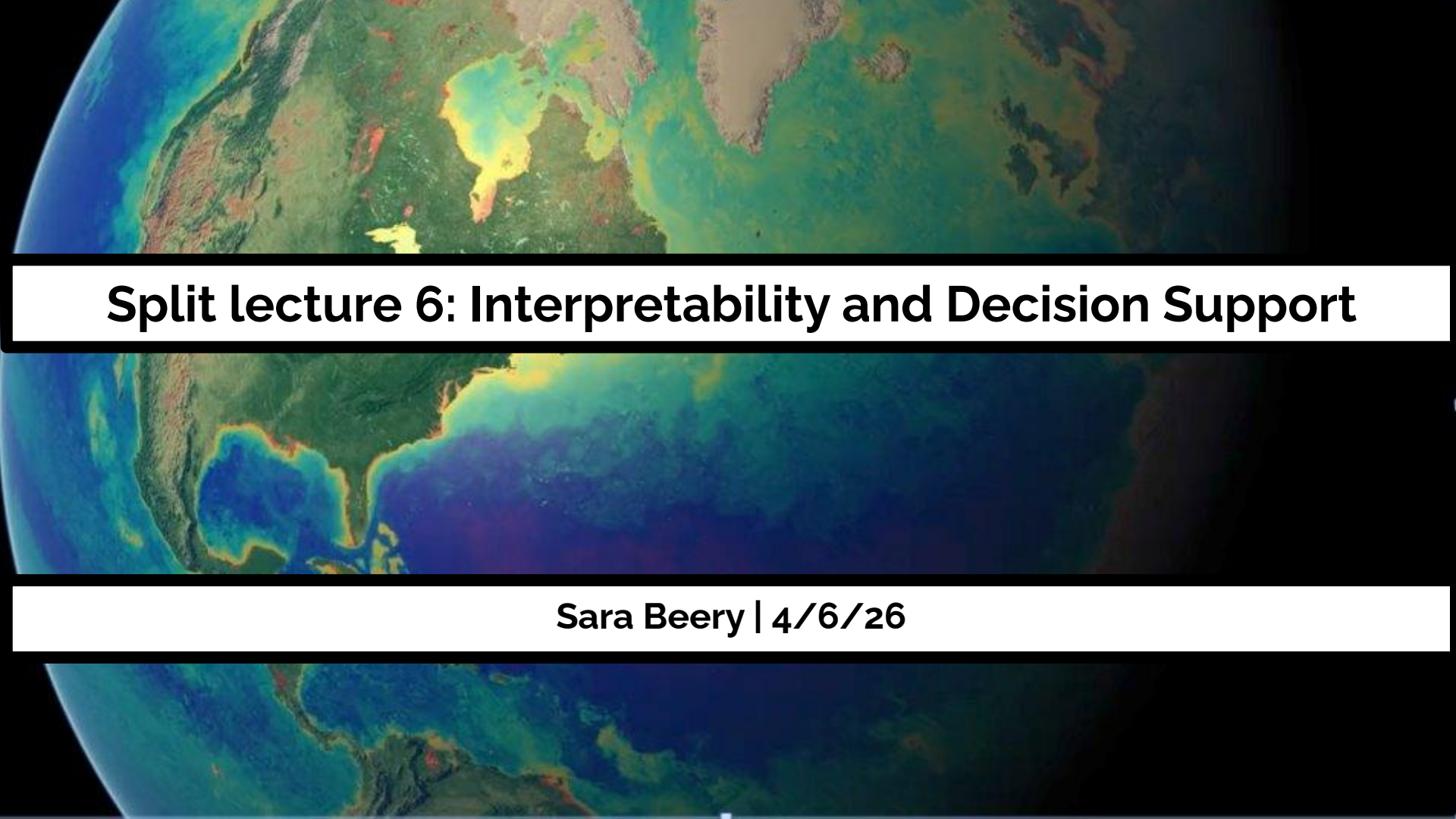
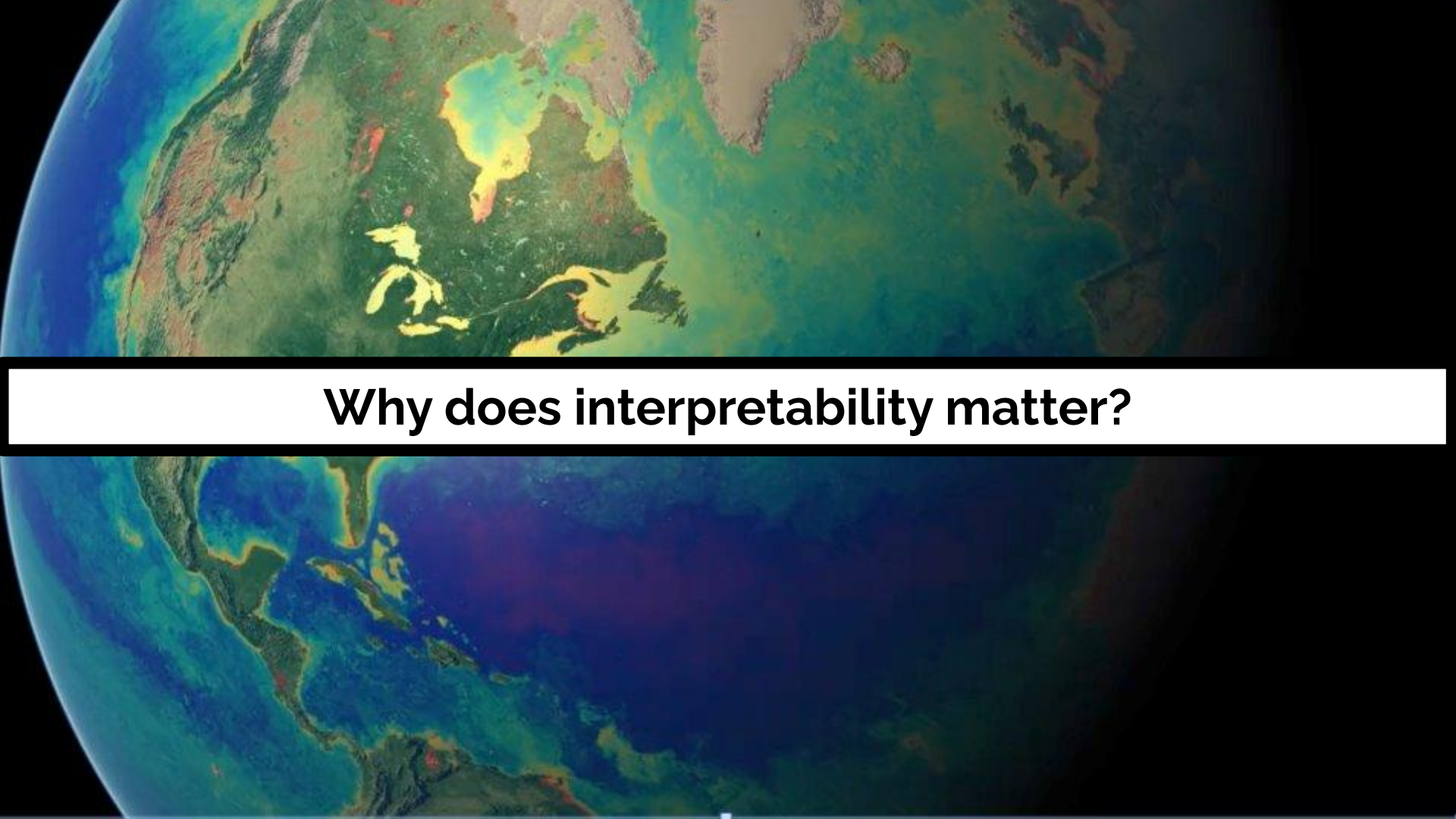


Split lecture 6: Interpretability and Decision Support

Sara Beery | 4/6/26



Why does interpretability matter?

5 Reasons interpretability matters in AI + Ecology

- **Accountability and trust**
  - Understanding *how and why* an AI system made a decision enables stakeholders (policymakers, conservation agencies, the public) to know when to trust AI-based systems.
- **Error detection and bias mitigation**
  - Detect/identify when models exploit spurious correlations, encode biases, or fail in edge cases that affect vulnerable populations or ecosystems.
- **Scientific validity**
  - Scientific analysis based on interpretable models may be more easy to validate (for example calibration can be viewed as interpretability).
- **Regulatory and legal compliance**
  - High-stakes applications (environmental policy, resource allocation) need to meet governance standards and legal liability requirements, which may require interpretability.
- **Knowledge extraction**
  - Uncover novel insights about underlying phenomena (e.g., identifying previously unknown relationships between hyperspectral signatures and ecosystem health).

Types of interpretable AI

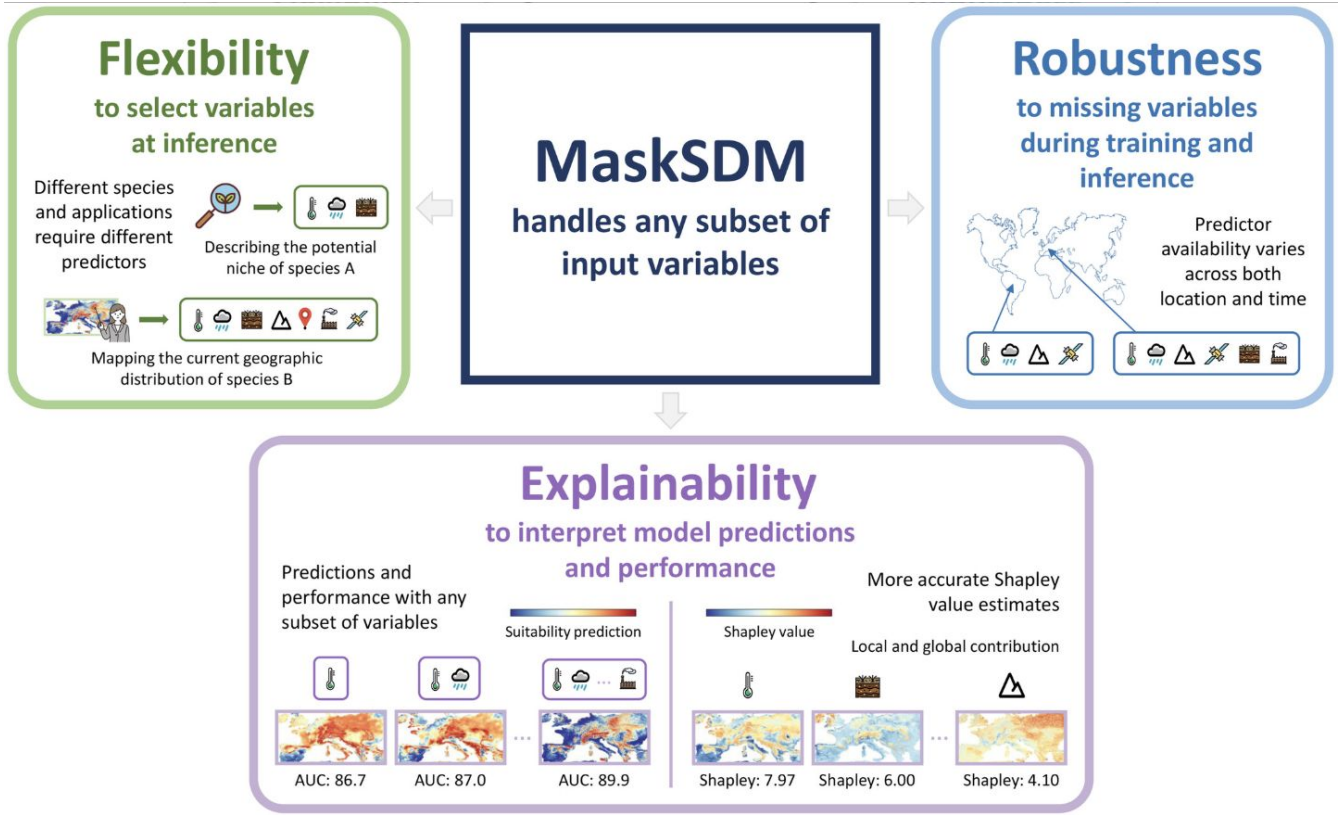
- **Post-hoc explanation techniques/attribution methods**
  - Methods applied after training that approximate “decision logic” (e.g., LIME, SHAP, attention visualizations, saliency maps, SAEs).
  - Flexible but can introduce approximation artifacts, or be misinterpreted.
- **Inherently interpretable models**
  - Architectures/methods that are transparent by construction (decision trees, rule-based systems, sparse linear models, additive models (GAMs), CBMs)
  - Often these have lower predictive accuracy than deep learning, or have been shown to enable shortcuts/routing (eg CBMs).
- **Prototype-based explanations**
  - Explains “decisions” via representative examples, prototypes, or identifying similar cases in training data (influence functions, case-based reasoning, retrieval-augmented explanations).
  - These are more intuitive in spatial/imagery applications.
- **Causal models**
  - Identify causal relationships and mechanisms (causal graphs, process-based models, structural equation modeling).

Diagnosing model bias with counterfactual generation

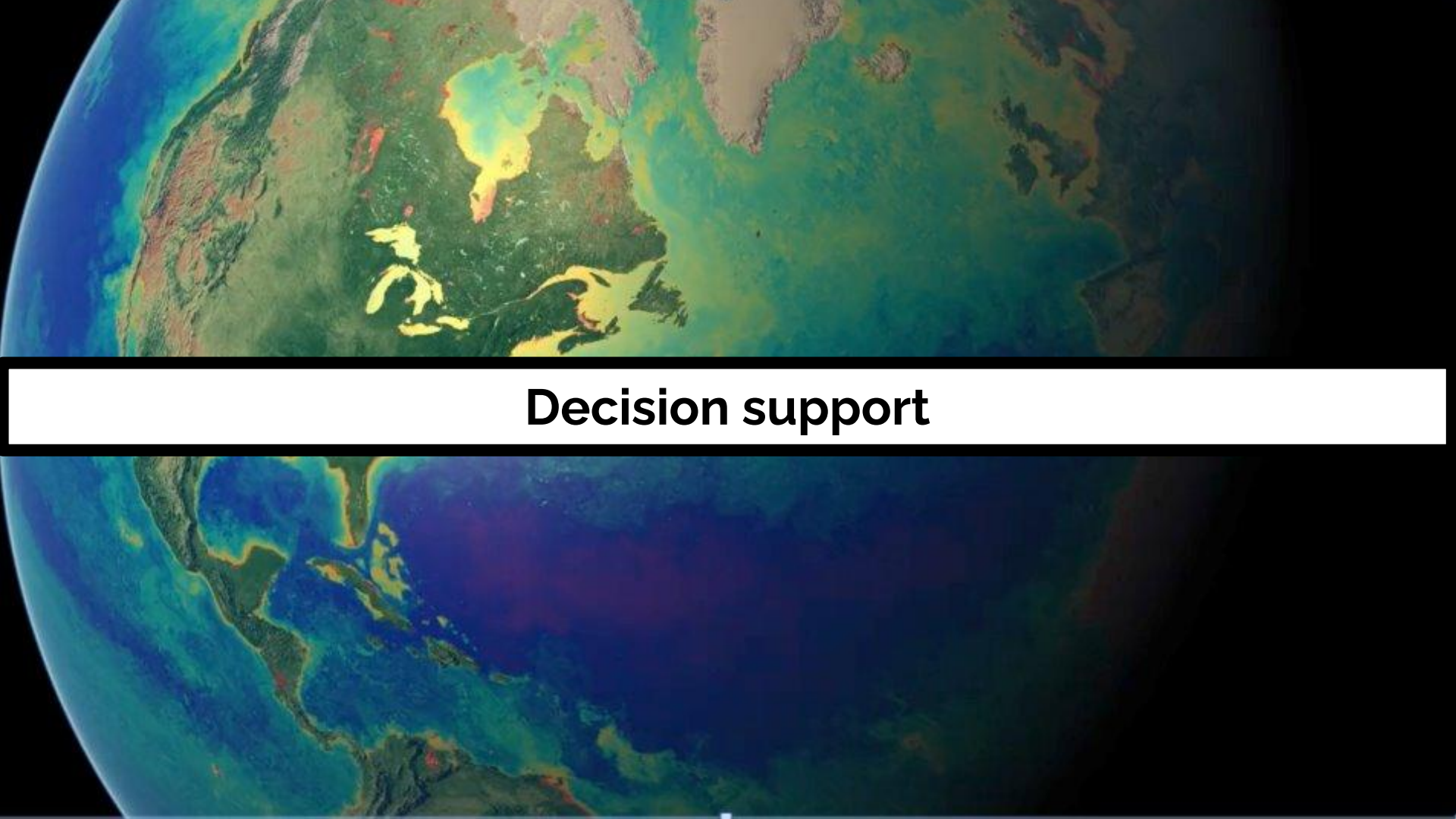
Figure 8: Real images of plates, with and without food and either on a table or in the grass. Below each image is the predicted class by an ImageNet-trained ResNet50.

https://arxiv.org/abs/2302.07865

Masking + Shapley values for Species Distribution Model interpretability (both by construction and as explanation)



hhttps://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210x.70200



Decision support

Adaptive management

A systematic approach to natural resource management that treats each action as a monitored experiment, using observed outcomes to iteratively refine management strategies and reduce uncertainty about ecosystem responses.

- First introduced back in 1978 (Walters and Hilborn)
- Clear complimentarity with AI-based monitoring

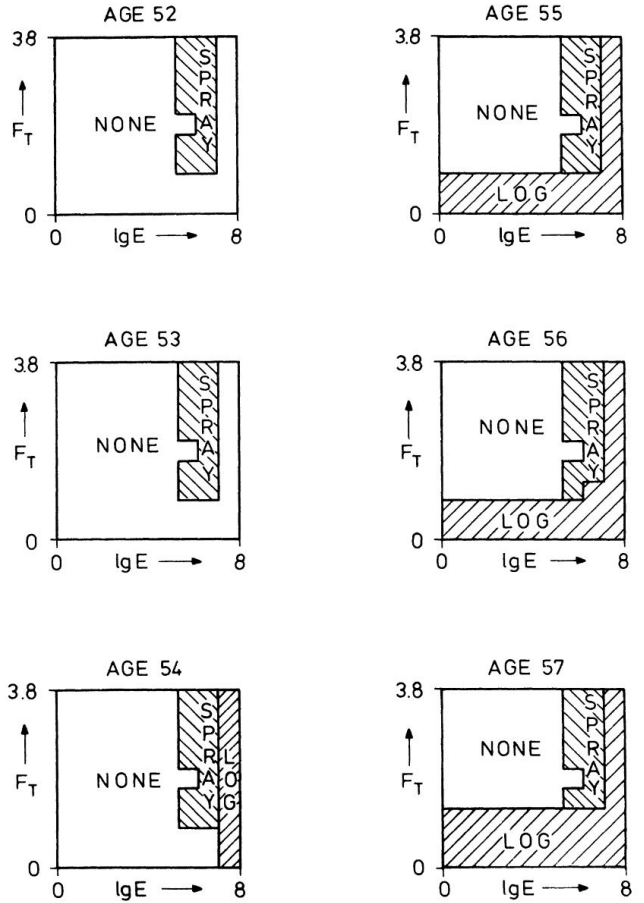


Figure 1 Optimal management policies for different-aged tree stands from Winkler's dynamic programming model. The X axis is the natural log of budworm egg density, the Y axis is the foliage level. Reprinted from (125).

Multi-agent systems to plan ranger routes

Figure 1: Rangers searching for snares (right) near a waterhole (left) in Srepok Wildlife Sanctuary in Cambodia. The waterhole is frequented by deer, pig, and bison, which are targeted by poachers.

Algorithm 1: LIZARD

Inputs: Number of targets N, time horizon T, budget B, discretization levels Ψ, target features $\vec{y}_i$
$n(i, \psi_j) = 0$, $\texttt{reward}(i, \psi_j) = 0 \quad \forall i \in [N], j \in [J]$
for $t = 1, 2, \ldots, T$ **do**
  Compute UCB_t using Eq. 4
  Solve $\mathcal{P}(\text{UCB}_t, B, N, T)$ to select super arm $\vec{\beta}$
  Observe rewards $X_1^{(t)}, X_2^{(t)}, \ldots, X_n^{(t)}$
  for $i = 1, 2, \ldots, N$ **do**
    $\texttt{reward}(i, \beta_i) = \texttt{reward}(i, \beta_i) + X_i^{(t)}$
    $n(i, \beta_i) = n(i, \beta_i) + 1$

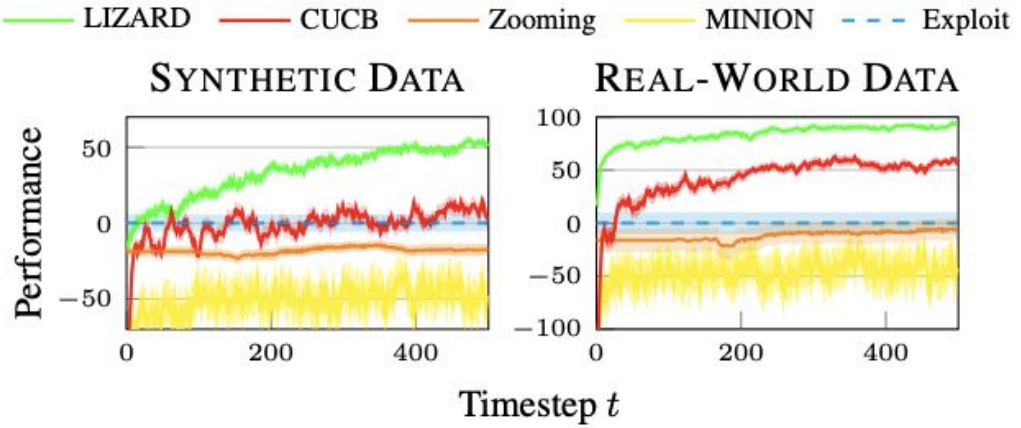
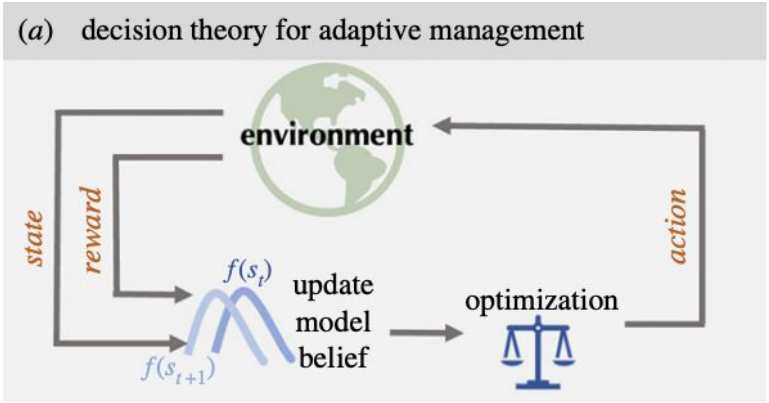
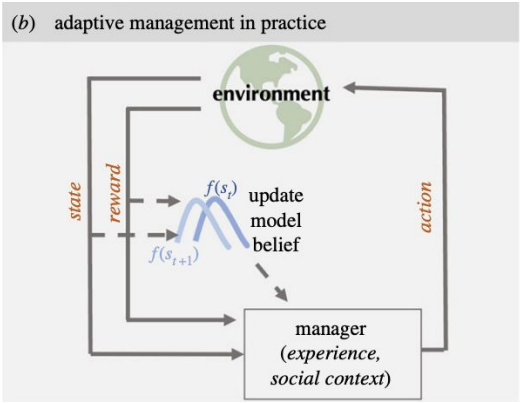


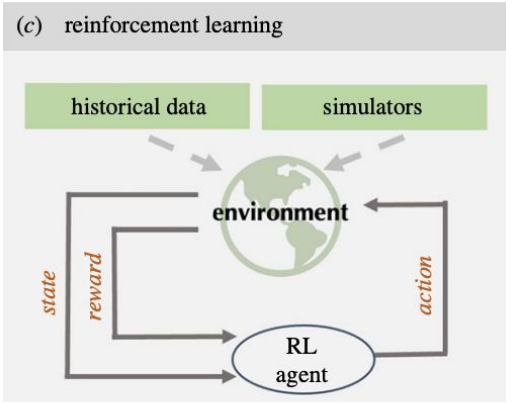
Figure 5: Performance, measured in terms of percentage of reward achieved between OPTIMAL − EXPLOIT, over time. Shaded region shows standard error. Setting shown is $N = 25$, $B = 1$. LIZARD (green) performs best.

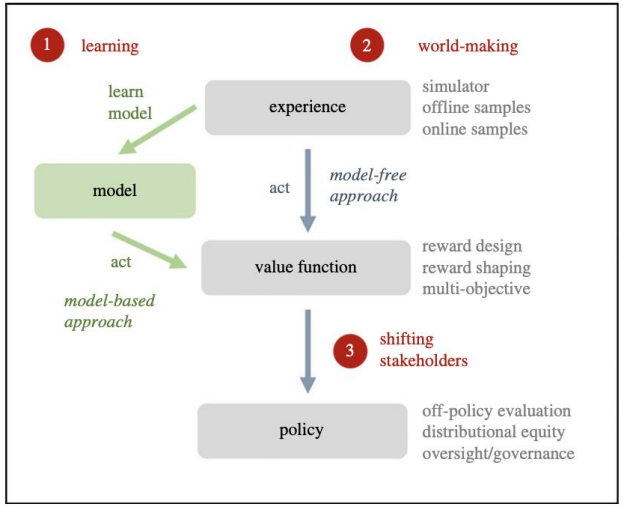
https://ojs.aaai.org/index.php/AAAI/article/view/17757

Bridging RL and Adaptive management





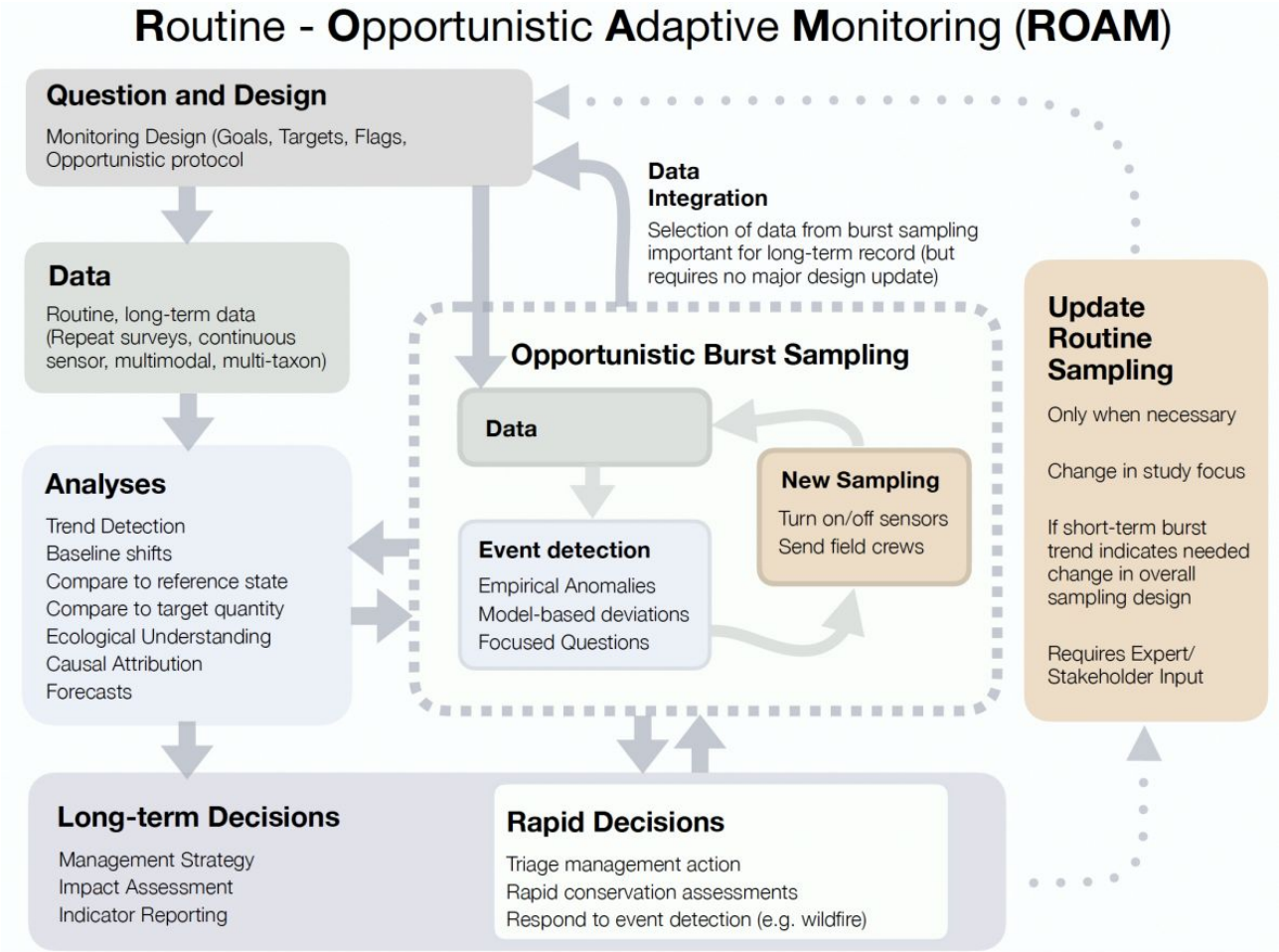




https://arxiv.org/abs/2303.08731

Adaptive monitoring

- Focuses on resource allocation when monitoring
- Requires understanding of what data modalities cost and what they capture



https://ecoevorxiv.org/repository/view/11640/

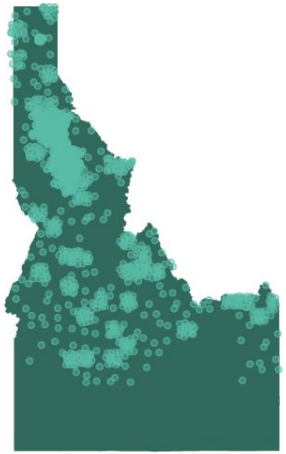
Challenge: AI models require expensive hardware

These are inaccessible to many potential AI users in biodiversity

Increased efficiency reduces cost

MegaDetector is used to process data for NGOs and conservation organizations globally

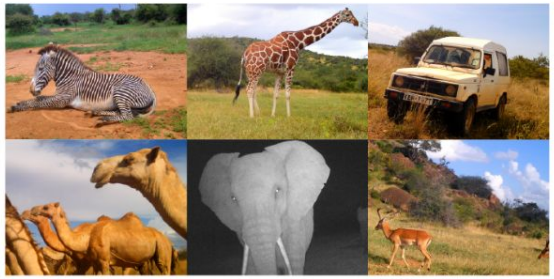
Idaho Dept. of Fish and Game



WOLF
pop. mgmt

2,000
cameras

11M
images



The MegaDetector

animal | empty | human | vehicle

Less than 15% of images require human review

human review

alert

Wildlife Protection Solutions

WILDLIFE CRIME PREVENTION

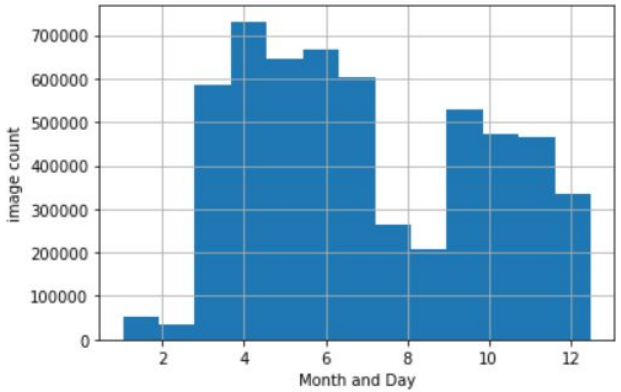
18 nations | 800 cameras | 900K images

Real-time alerts

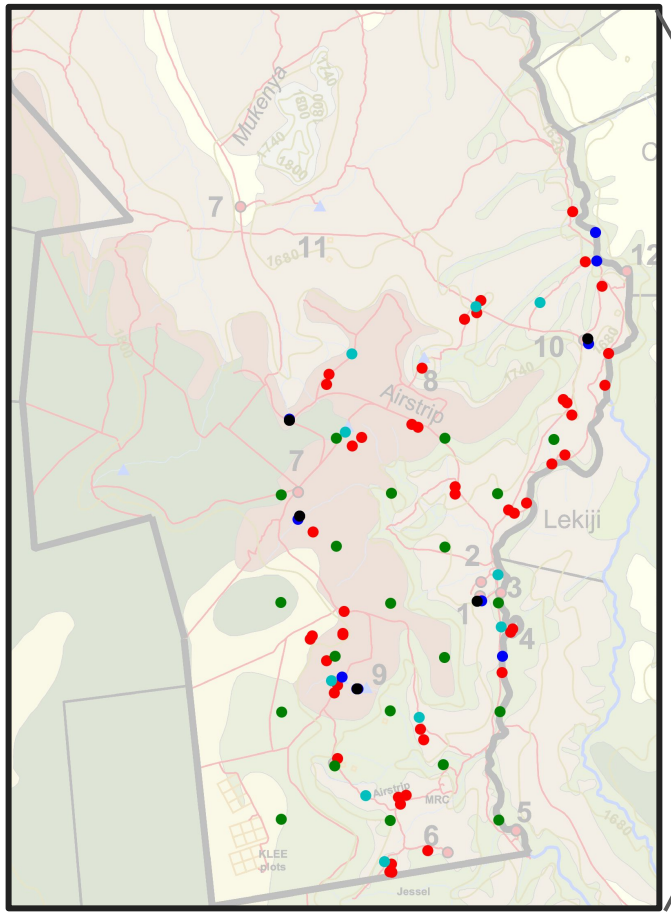
Detects one real wildlife threat per week on average

Seeing biodiversity: perspectives in machine learning for wildlife conservation, Tuia*, Kellenberger*, Beery*, Costelloe*, et al., Nature Communications (to appear)

Bandwidth is limited in the field



Up to 700K hi-resolution images per month



Commercial edge-based AI camera traps in development

ZSL
LET'S WORK FOR WILDLIFE

Instant Detect

Focused on anti-poaching

RESOLVE

TRAILGUARD AI

Focused on anti-poaching & human-wildlife conflict

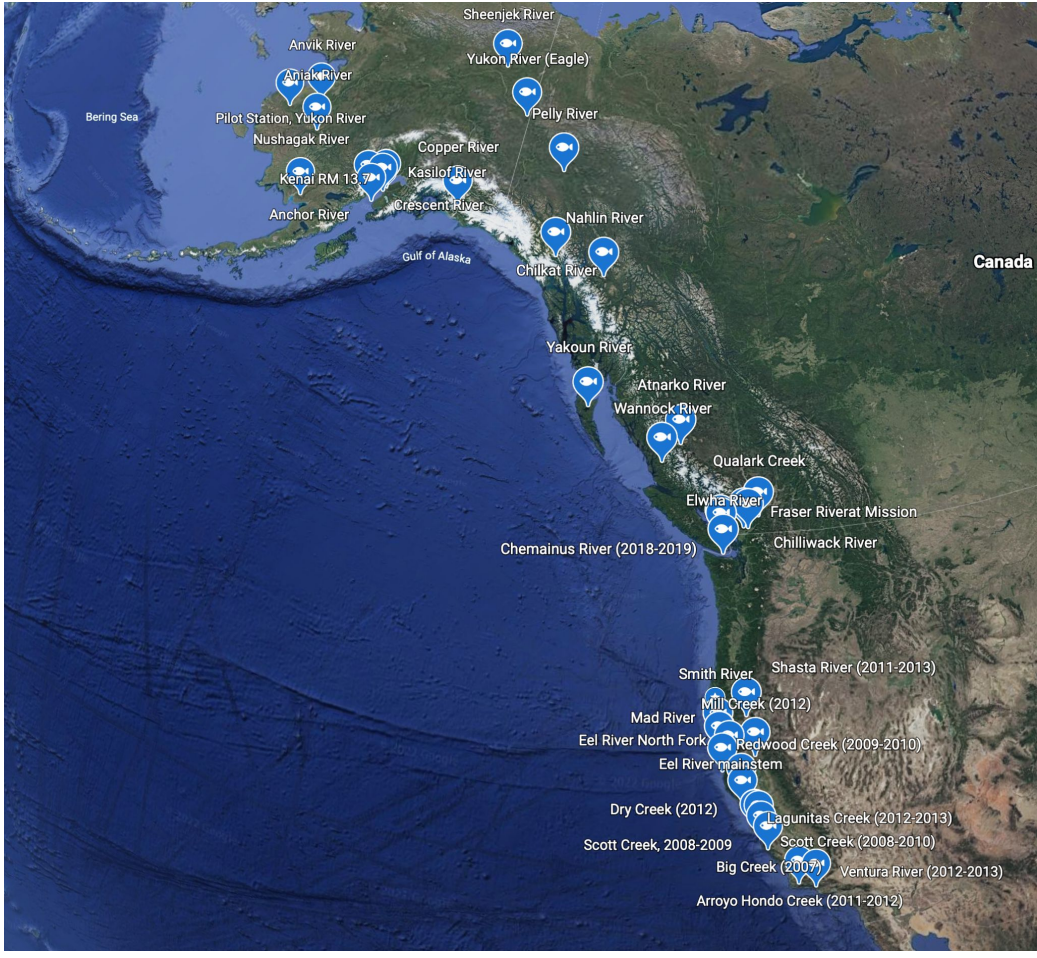
CONSERVATION X LABS

The SENTINEL

Focused on real-time animal behavior monitoring

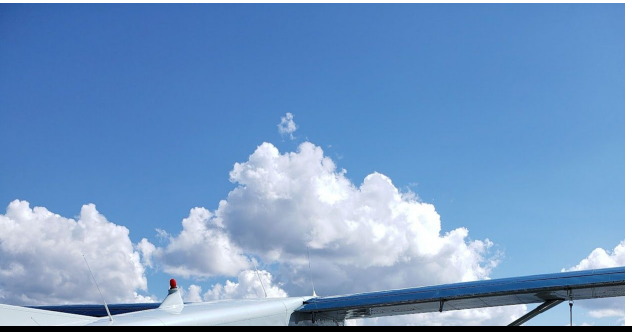
Sonar deployment to monitor salmon returns

Photos courtesy of ADFG



We need near-real-time counts from remote field sites

We need near-real-time counts from remote field sites



This requires edge-based models that are robust and reliable even as environmental conditions change



www.inaturalist.org

iNaturalist

CALIFORNIA ACADEMY OF SCIENCES

NATIONAL GEOGRAPHIC

iNaturalist is a joint initiative of the California Academy of Sciences and the National Geographic Society.

How It Works

1 Record your observations

2 Share with fellow naturalists

3 Discuss your findings

iNaturalist Explore Community More Log In or Sign Up

Observations

Species Location Go Filters

The World | 90,060,114 OBSERVATIONS | 344,629 SPECIES | 234,007 IDENTIFIERS | 2,015,371 OBSERVERS

Map Grid List Places of Interest

South African Ground Squirrel
(*Geosciurus inauris*)
University Rd, Par... • Jun 20, 2007
3m

Ring-necked Duck
(*Aythya collaris*)
Texas, US • Today
3m

South African Ground Squirrel
(*Geosciurus inauris*)
University Rd, Par... • Jun 15, 2007
3m

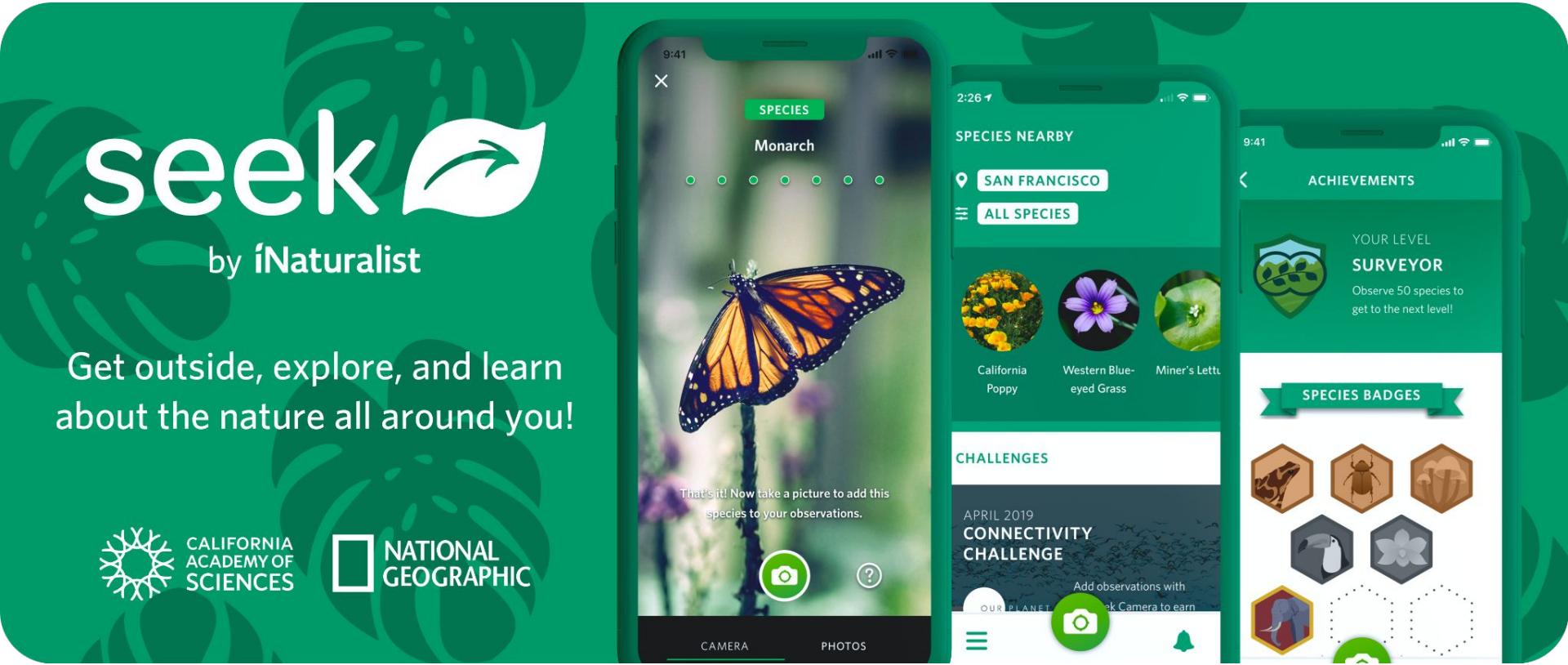
South African Ground Squirrel
(*Geosciurus inauris*)
University Rd, Par... • Jun 20, 2007
3m

South African Ground Squirrel
(*Geosciurus inauris*)
University Rd, Par... • Jun 20, 2007
3m

South African Ground Squirrel

Map Legend

Real-time, on-device fine grained categorization



Training efficiency

- Fine tuning often decreases training time, and thus costs
- Reducing sample overlap or removing noisy training data can reduce training costs without impacting performance, or sometimes improving performance (Coresets, DataComp)
- Smaller models train faster

Training on the edge

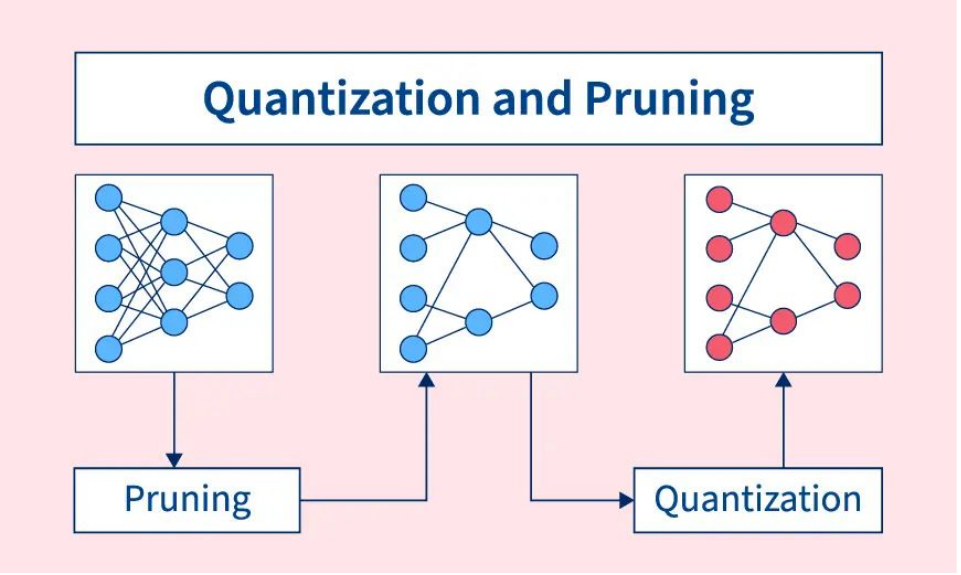
- Power
- Hardware
- Bandwidth
- Verification

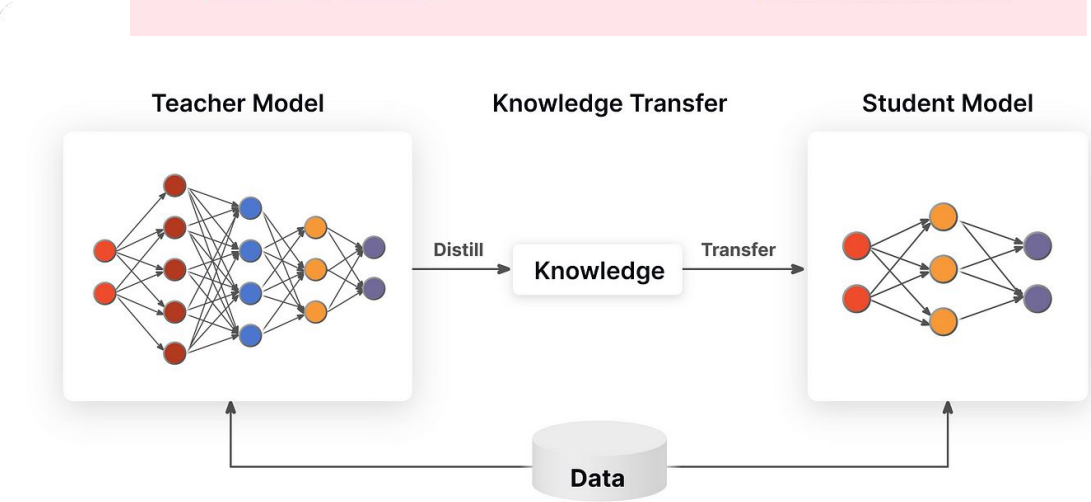
Evaluation efficiency

- Inexpensive proxy tests (ie brightness)
- Small (but representative) test sets
- Striation and multiple metrics
  - Make the most of your inference calls
- Active testing (soon!)

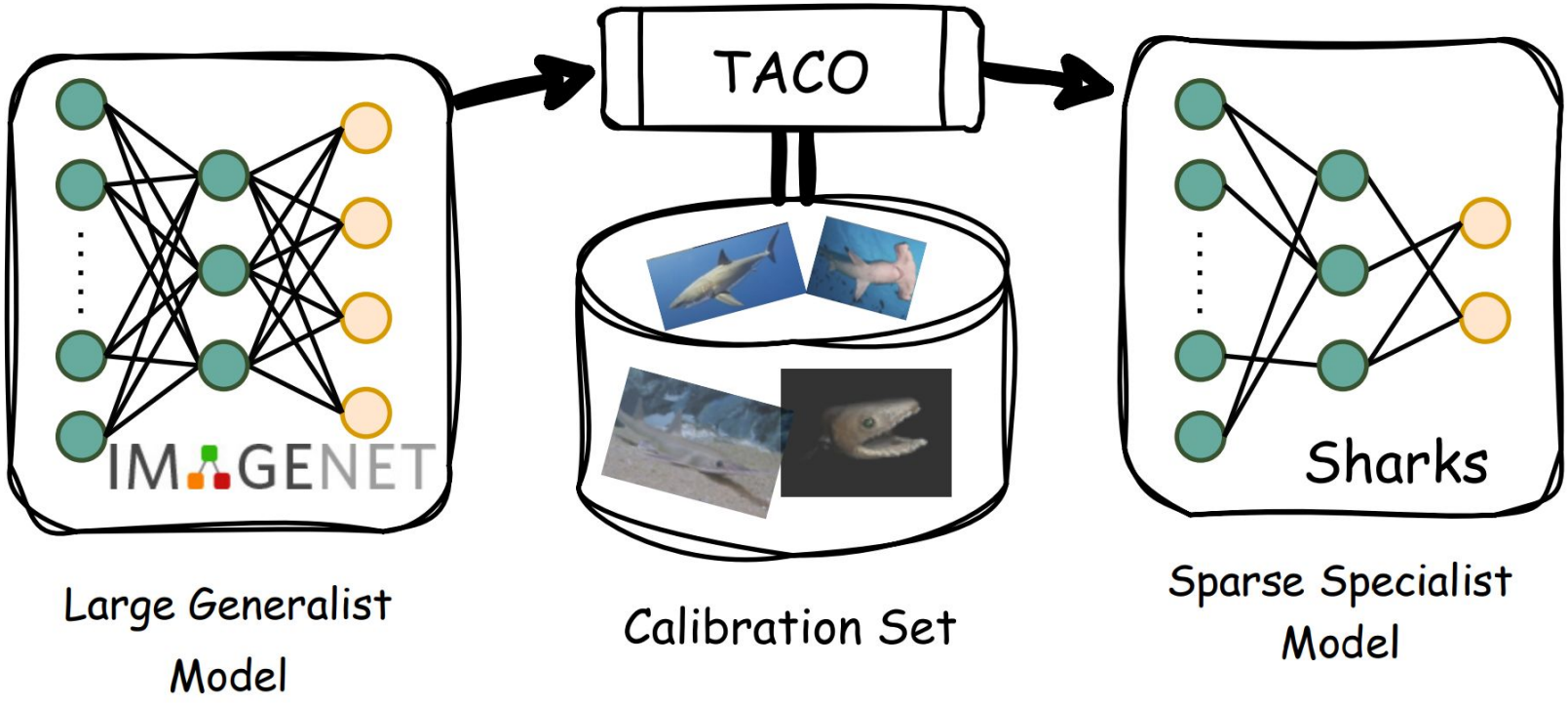
Inference efficiency

- Quantization
- Pruning
- Distillation
- Routing





We can quickly compress large generalist models into accurate and efficient specialists



Institute of Science and Technology Austria

Skoltech
Skolkovo Institute of Science and Technology

Vision Models Can Be Efficiently Specialized via Few-Shot Task-Aware Compression, Kuznedelev,..., Beery, Kurtic, Alistarh et al. 2023

Federated learning

- Maintains data privacy
- Can be efficient at the edge
- Requires bandwidth and synchronization