ISTA ANNUAL MEETING · CALGARY 2026

A.I. models in seed testing

How it works and the statistical pitfalls

Thomas Michelon, PhD
ISTA Statistics Committee

How it works

Human	**Eye** information capture	→	**Brain** processing	→	**Decision**
Machine	**Machine vision** information capture	→	**Machine learning** processing	→	**Decision**

Supervised learning: the model copies labelled expert decisions from a training sample.

Every pitfall today traces back to one word: sample.

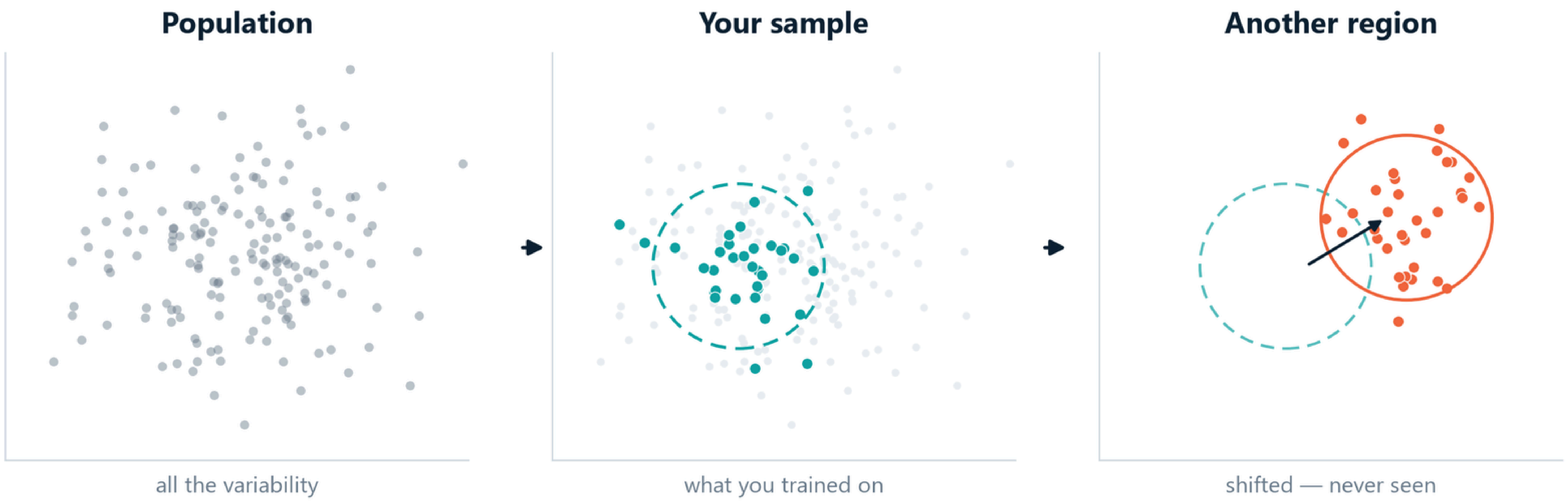
Source: ESL, Ch. 2 (Hastie, Tibshirani & Friedman, 2009)

Four pitfalls

1. The model only knows your sample
2. A model can memorize instead of learning
3. Accuracy depends on which classes you compare
4. A high accuracy can still miss the contaminant

1 Pitfall #1: The model only knows your sample

You train on a sample, so the model only ever learns the variability it actually saw.



CONSEQUENCE the model stops working on other lots, seasons, or regions it was never sampled from.

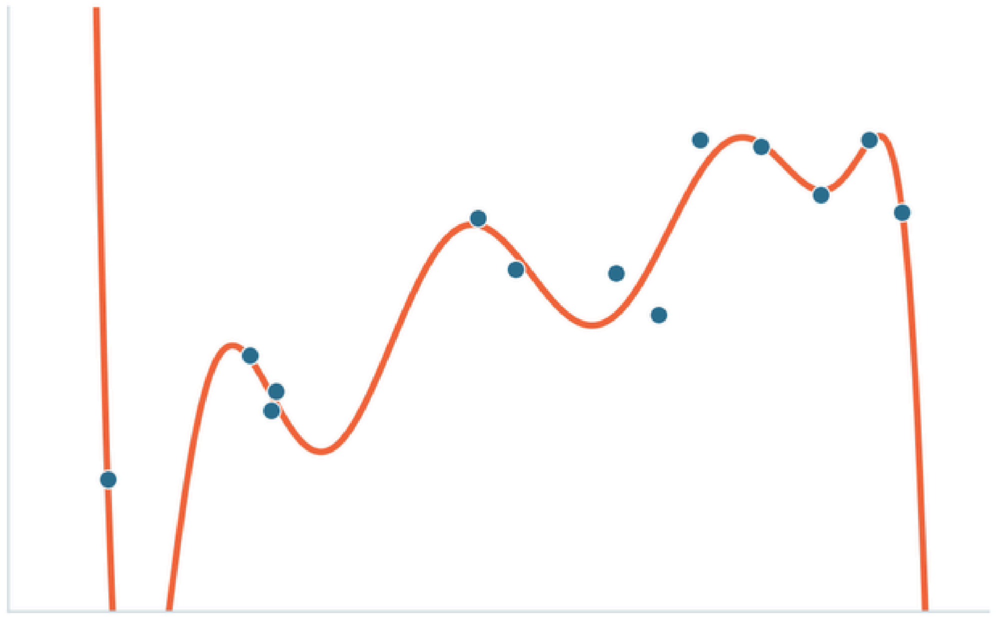
Source: ESL Ch. 2 & 7

2

Pitfall #2: A model can memorize instead of learning

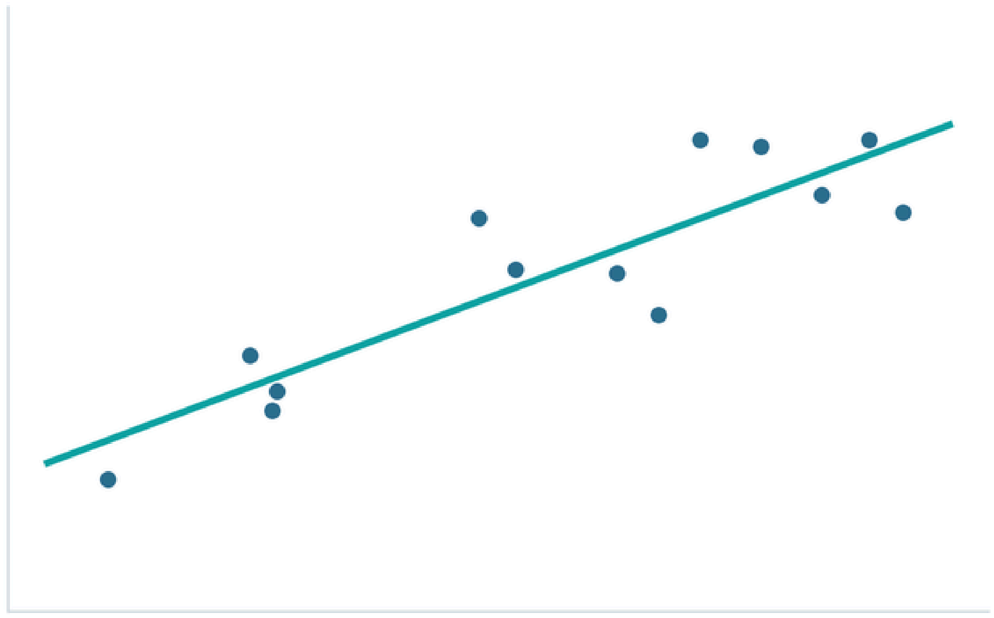
A model can memorize your data instead of learning the pattern behind it.

Memorizes the sample



fits every point — even the noise

Learns the pattern



captures the trend, ignores noise

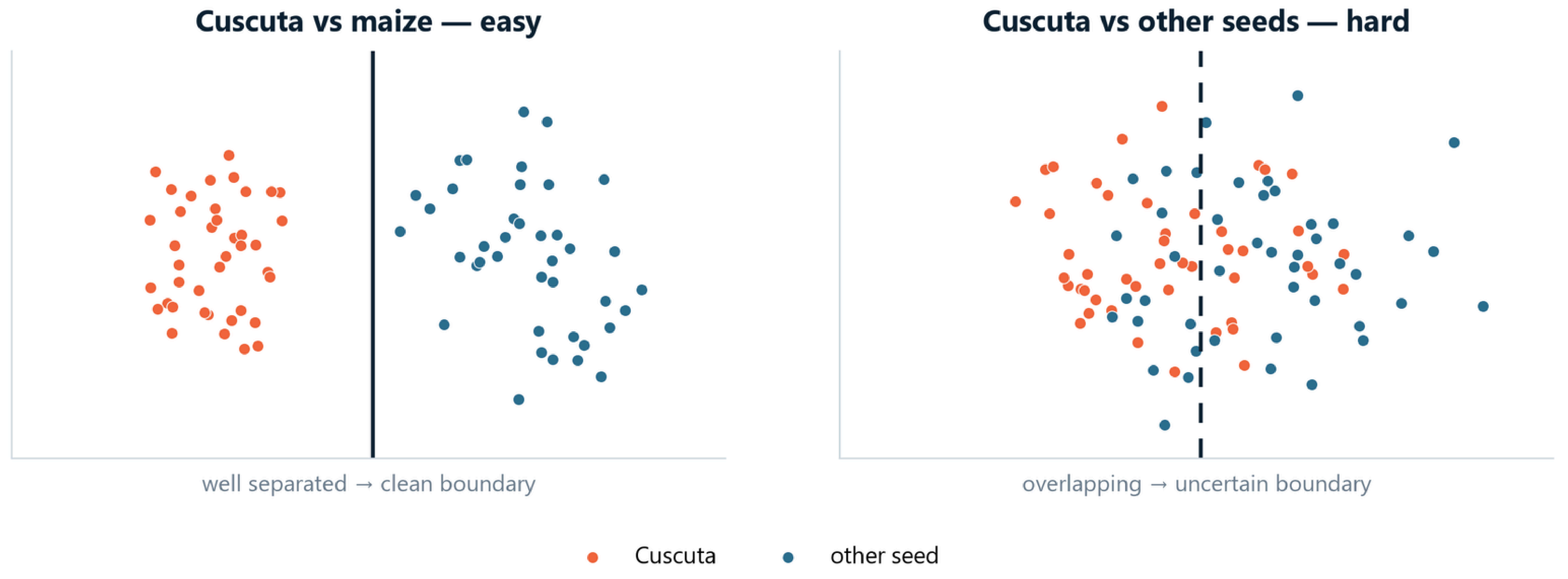
Catch it: independent hold-out set · k-fold cross-validation · bootstrap CI

CONSEQUENCE it can fail even on new samples from the same region, it memorised, it did not learn.

Source: ESL §7.2–7.4 · ISL Ch. 5

3 Pitfall #3: Accuracy depends on which classes you compare

How accurate a model can be is set by how different the classes are — not by the model itself.



CONSEQUENCE a tool proven for one pair of classes gives false results once other species appear.

Source: ESL §2.4

4 Pitfall #4: A high accuracy can still miss the contaminant

When the target is rare, overall accuracy can look excellent while the tool catches nothing.

	Pred: Cuscuta	Pred: not
Actual: Cuscuta	0	5 missed!
Actual: not	0	995

Accuracy
99.5%

Sensitivity
0%

CONSEQUENCE a “highly accurate” tool can pass contaminated lots — measure sensitivity, not accuracy.

Source: ISL Ch. 4

Takeaways

	🛠️ If you build the model	👤 If you use the model
1 Scope	**Document the region & variety it covers.** Eval → representative, diverse sample	**Check it was trained for your seeds.** Eval → local blind test, your region first
2 Learning	**Prove it learned — no leakage.** Eval → cross-validation + held-out; report a CI	**Don’t trust the reported accuracy.** Eval → ask for independent validation; test fresh seeds
3 Classes	**Have a well document scope and per-class metrics.** Eval → documentation and confusion matrix	**Confirm it covers your species.** Eval → match class list; test local look-alikes
4 Rare targets	**Report sensitivity & FNR, not accuracy.** Eval → spiked test; test vs threshold	**Ask the miss rate, not the accuracy.** Eval → spiked blind test, known counts

Same four pitfalls — what to get right (builder) and what to check before trusting it (user).

Thank you!

Questions & discussion welcome

Thomas Michelon, PhD
ISTA Statistics Committee
thomasbrunomichelon@gmail.com

ISTA ANNUAL MEETING
22-25 June 2026
Calgary, Canada

LinkedIn