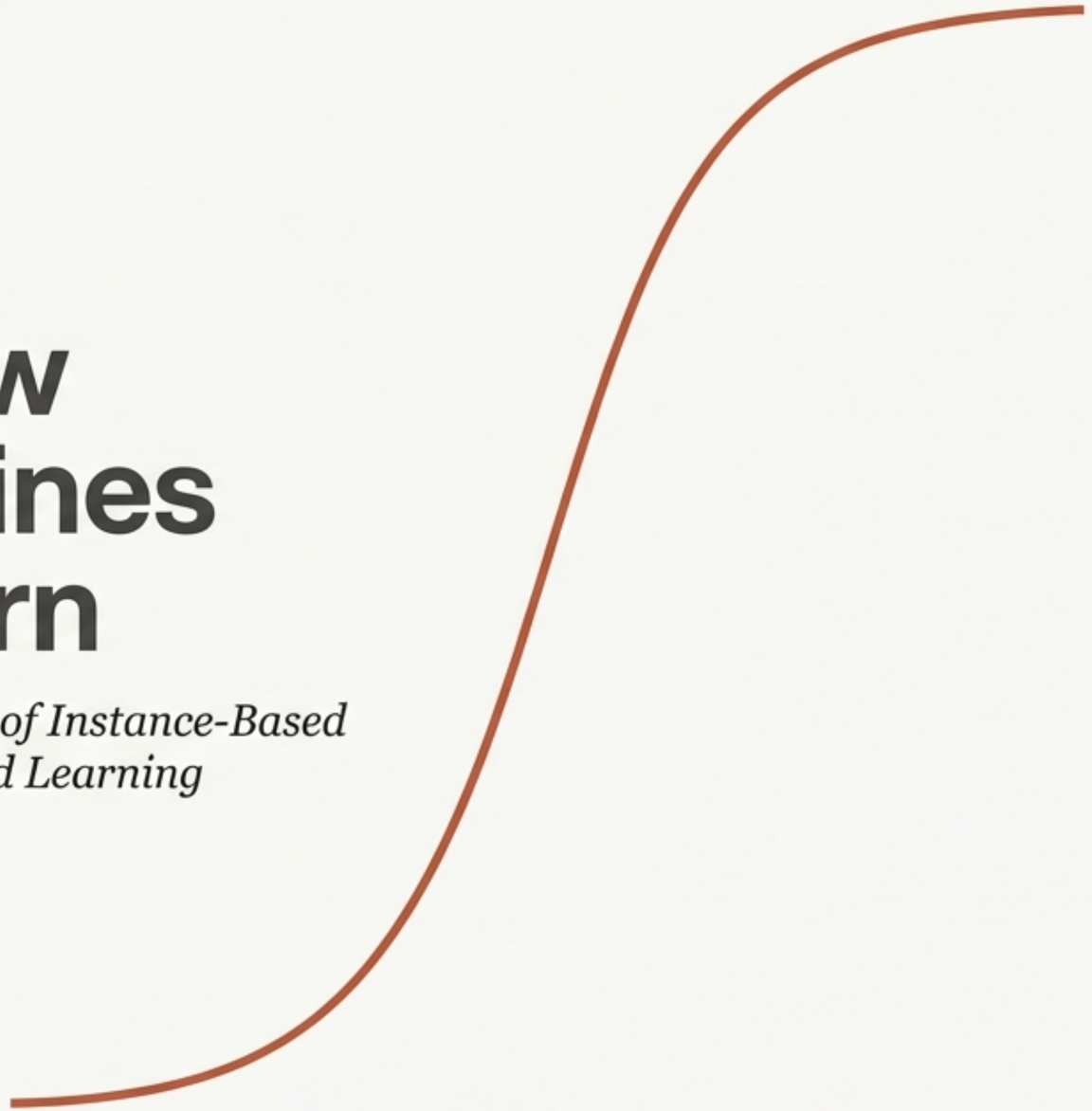
THE MEMORISER

THE GENERALISER



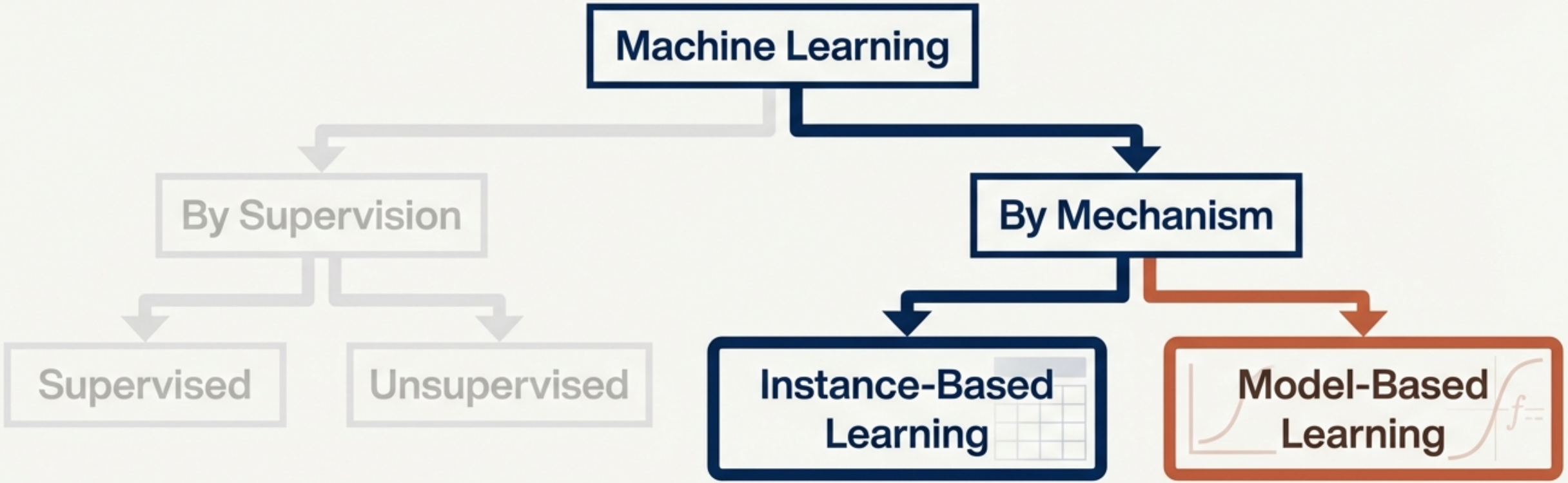
How Machines Learn

A comparative analysis of Instance-Based and Model-Based Learning

Based on insights from CampusX

Beyond Supervision: Classifying by Mechanism

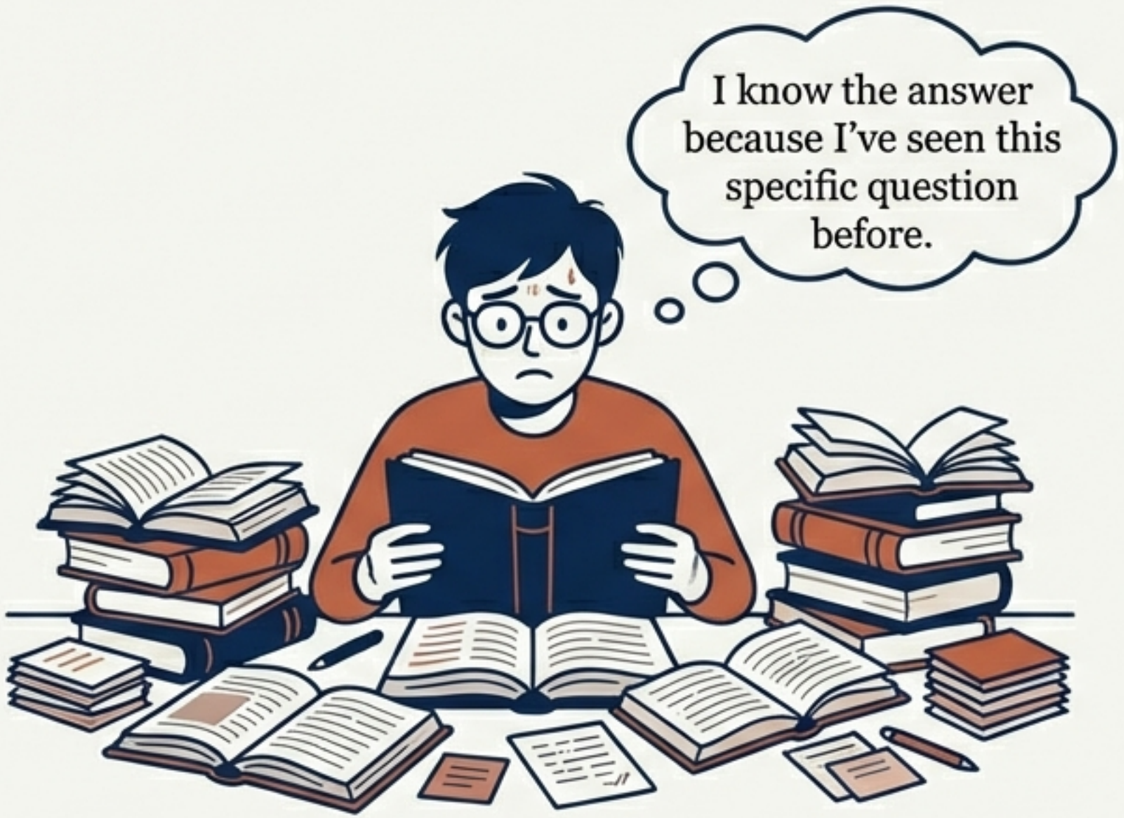
We often categorize Machine Learning by the amount of supervision required (Supervised vs. Unsupervised) or the output type. Today, we look at a more fundamental distinction: *how* the system actually learns.



When a machine encounters data, does it simply memorise the past, or does it attempt to understand the underlying principles?

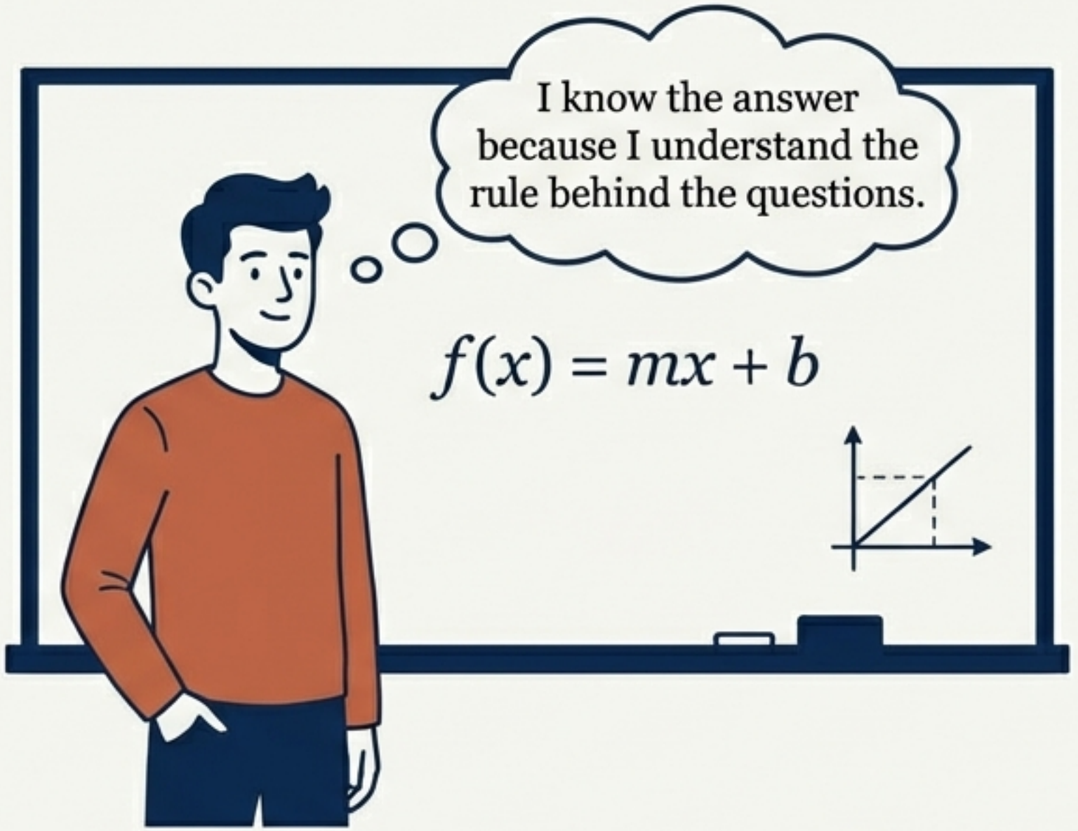
The Human Parallel: Rote vs. Conceptual Learning

The Rote Learner



Analogous to Instance-Based Algorithms

The Conceptual Learner



Analogous to Model-Based Algorithms

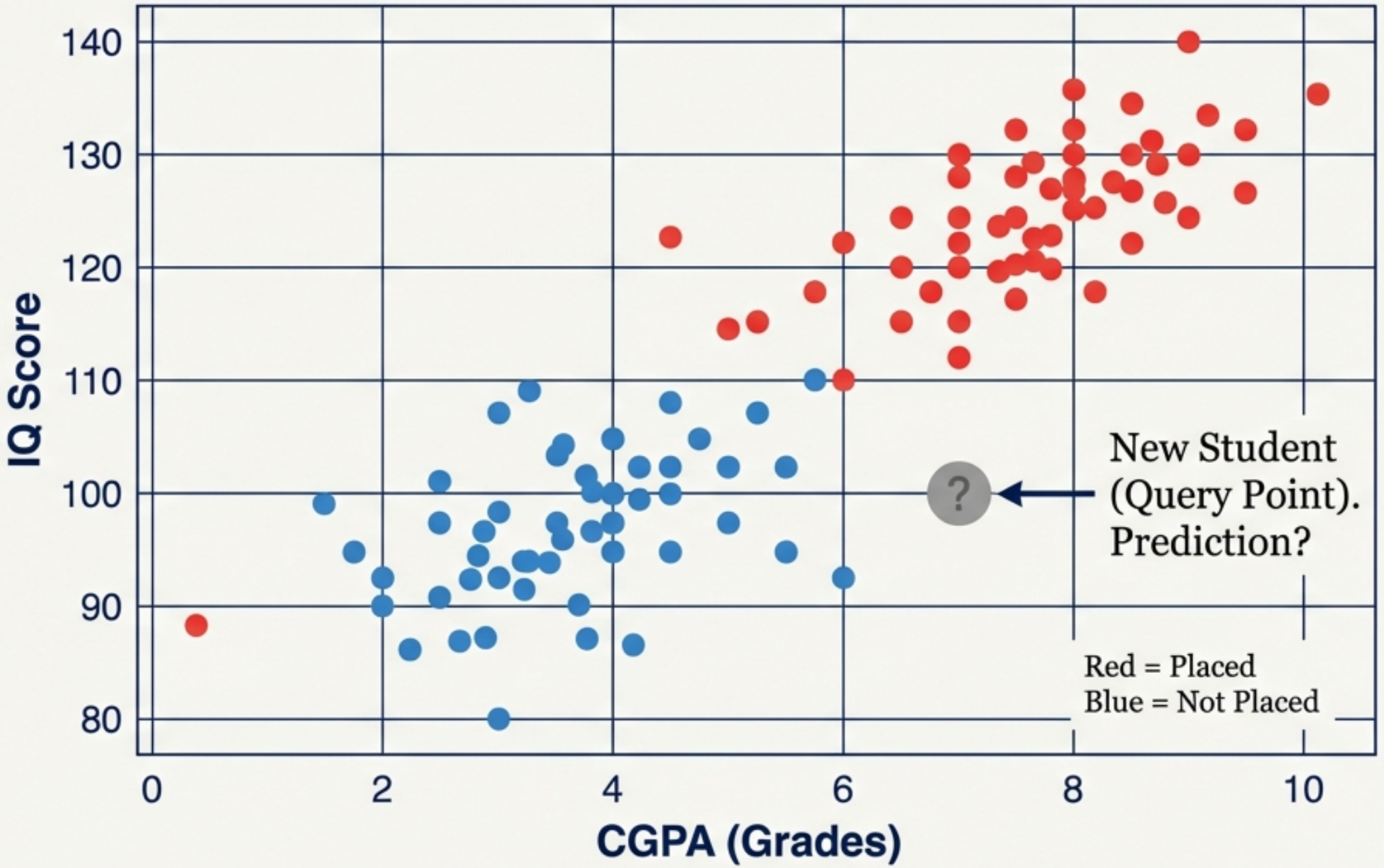
The Case Study: Predicting Job Placement

To compare these methods, we use a classic classification problem.

Input Features:
IQ & CGPA

Output Target:
Placement Status (Yes/No)

Challenge:
A new student arrives. Will they get placed?



Approach A: Instance-Based Learning

The Memoriser

Mechanism:

- 1. Ingestion: The algorithm accepts the training data.
- 2. Inaction: It performs zero calculations initially. It implies storage, not processing.
- 3. Reaction: It waits until a specific question is asked to do any work.

Key Term: **Lazy Learning**

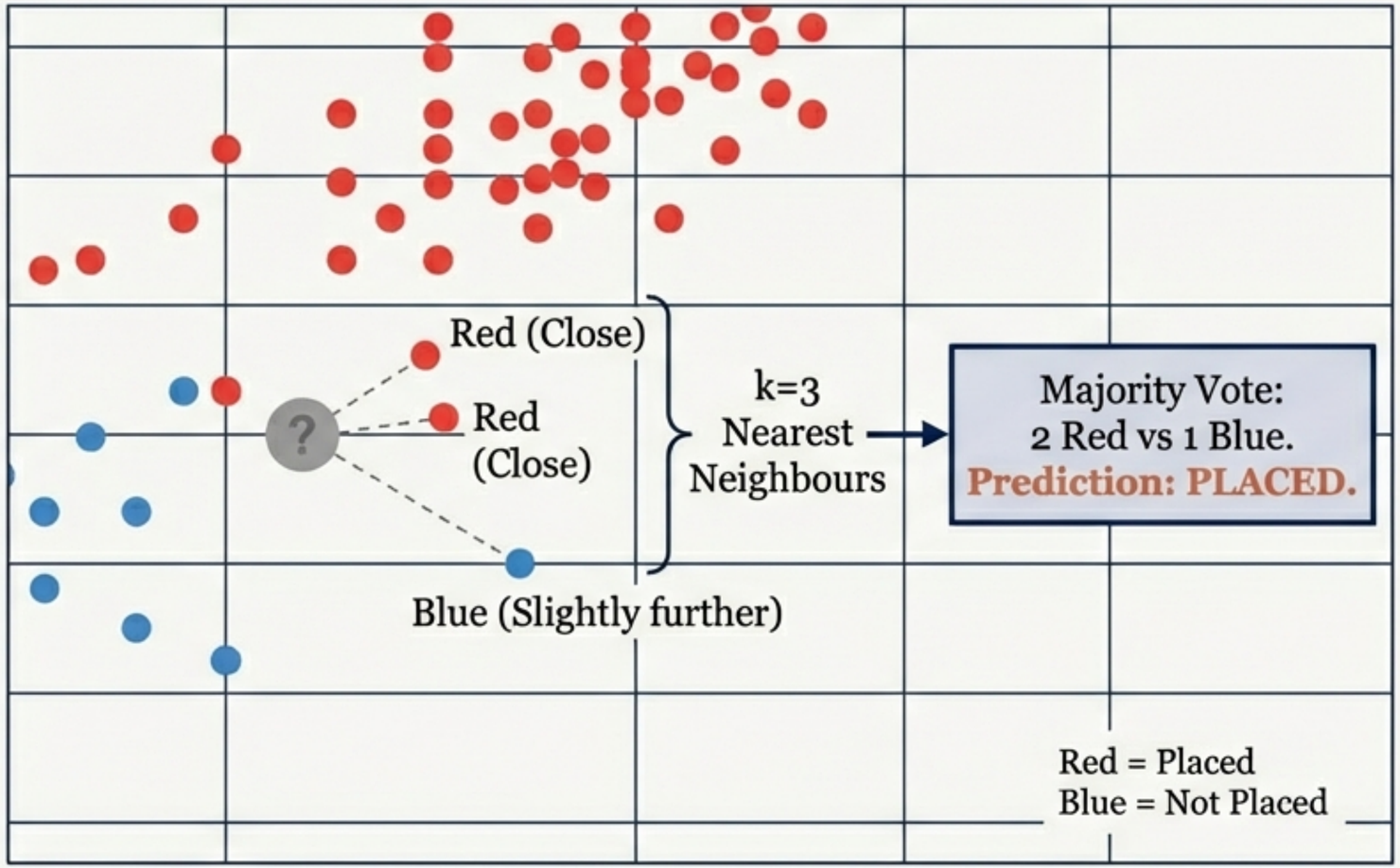
Instance-Based in Action: The Neighbourhood Logic

Mechanism:

- **Step 1:** Calculate distance to all historical points.
- **Step 2:** Identify nearest neighbours.
- **Step 3:** Adopt the majority label.

Insight:

"The algorithm assumes similarity defines reality. You are defined by the company you keep."



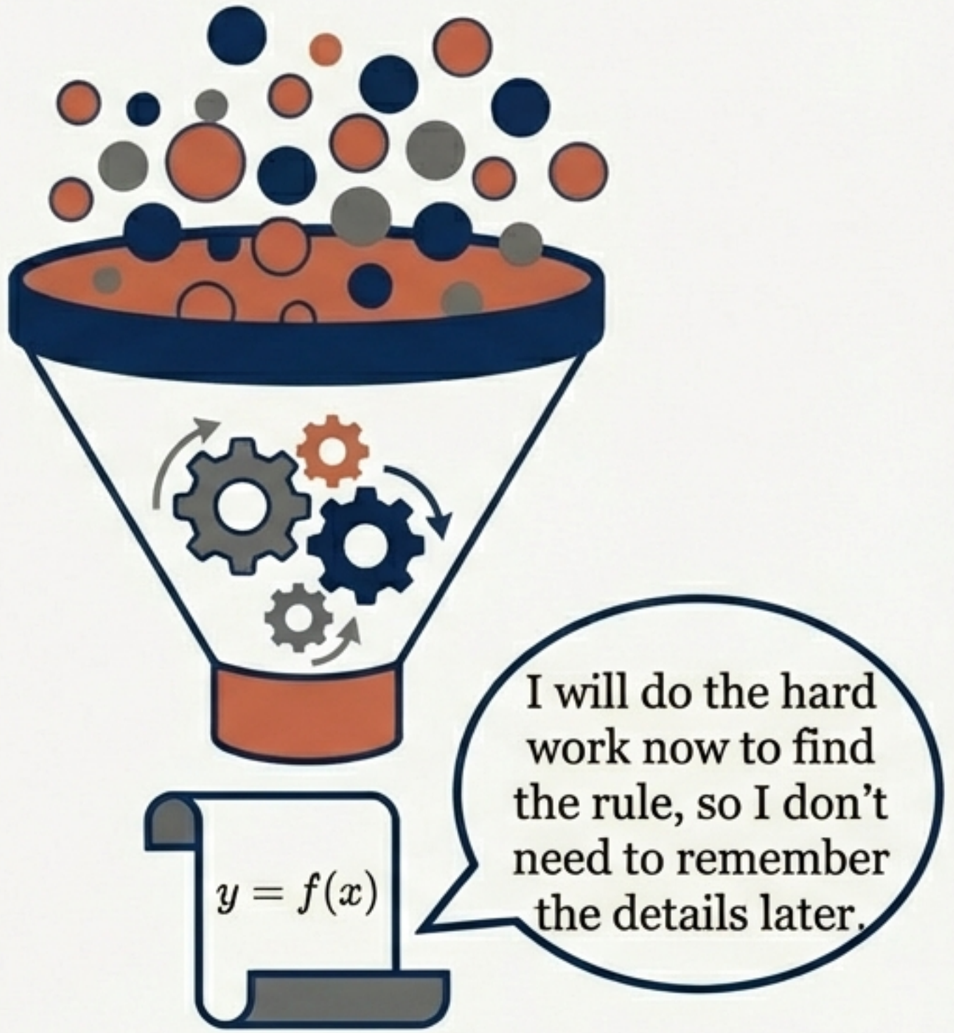
Approach B: Model-Based Learning

The Generaliser

Mechanism:

- 1. Ingestion: The algorithm accepts the training data.
- 2. Extraction: It runs an algorithm to find a relationship (a mathematical function) that explains the data.
- 3. Abstraction: It creates a Model and discards the specific data points.

Key Term: **Eager Learning**



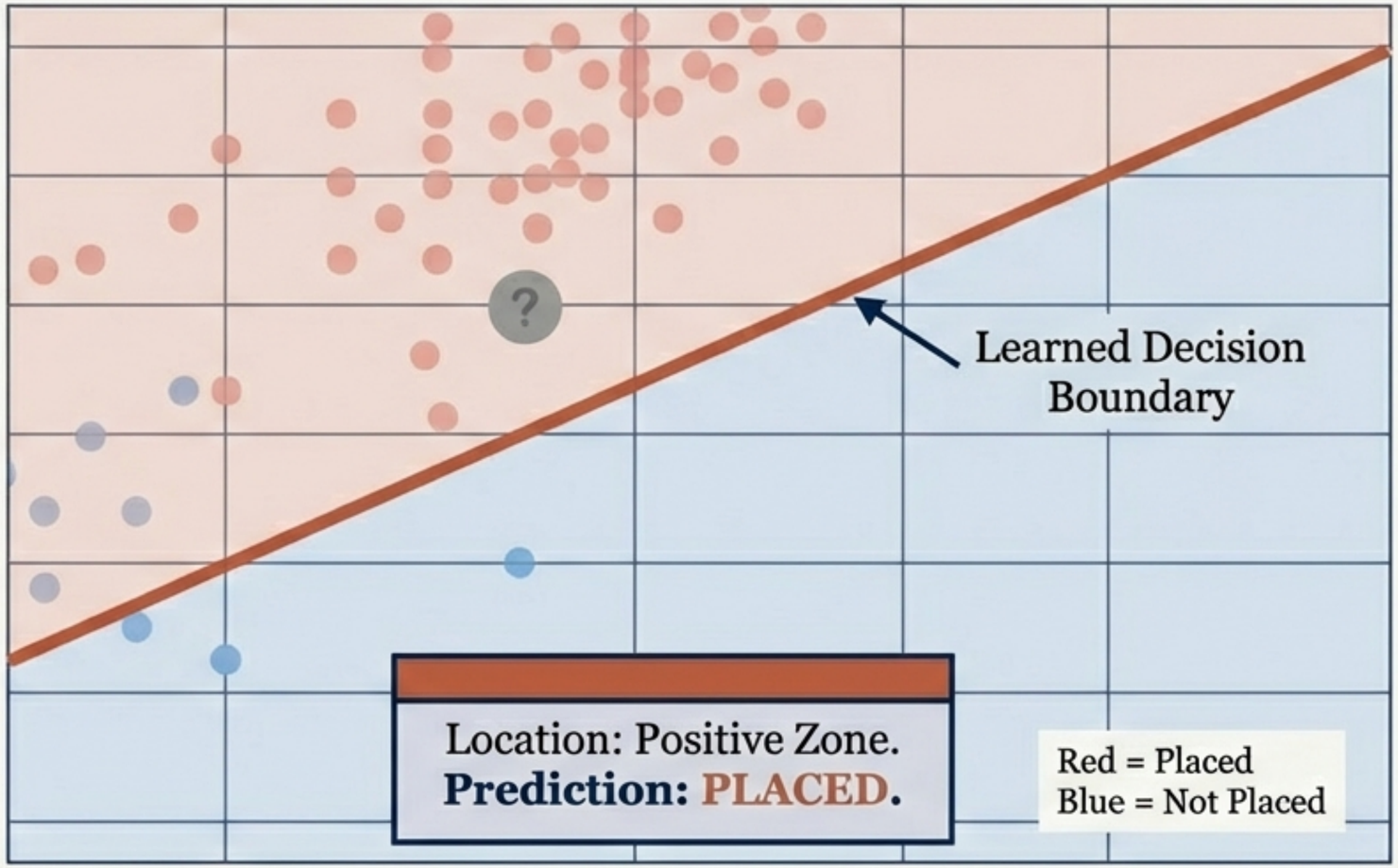
Model-Based in Action: The Decision Boundary

Mechanism:

- **Step 1:** Training determines the optimal line (slope/intercept).
- **Step 2:** Check which side of the line the new student falls on.

Insight:

The specific nearby dots no longer matter; only the mathematical rule matters.



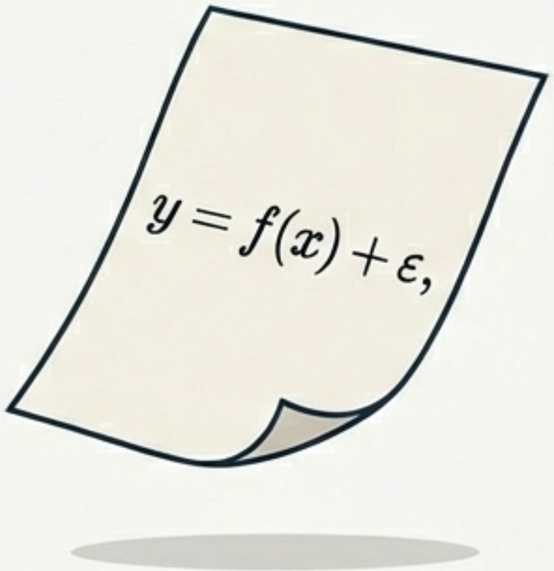
The Fundamental Shift: Data vs. Function

Instance-Based

Intelligence resides in the DATA.

Requires history to be present forever.
Delete the data, destroy the model.

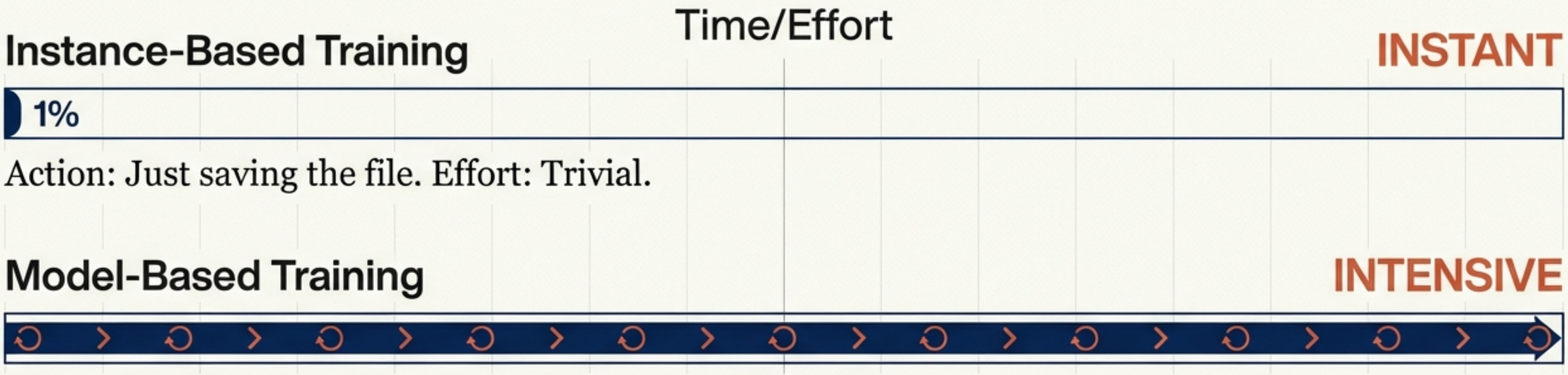
Model-Based



Intelligence resides in the FUNCTION.

Data is disposable. Once the rule is learned,
the history can be deleted.

Head-to-Head: The Training Phase



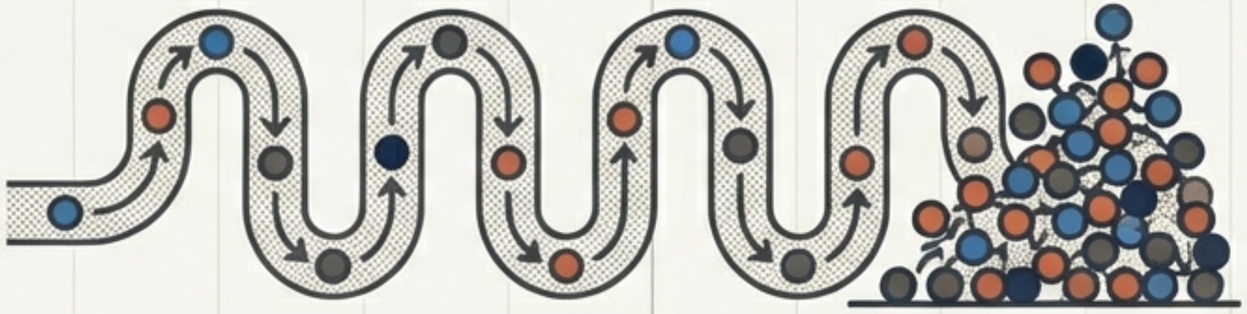
Model-based learning does the homework before the test; Instance-based learning procrastinates.

Head-to-Head: The Prediction Phase

Instance-Based

SLOW & EXPENSIVE

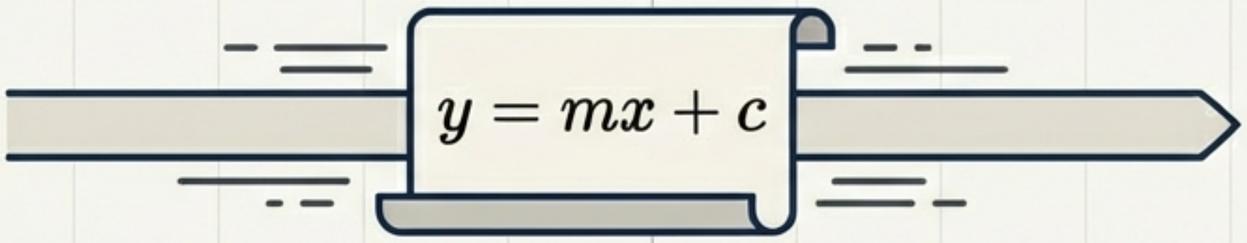
Processing Time



Must compare query against EVERY historical data point. As data grows, speed drops significantly.

Model-Based

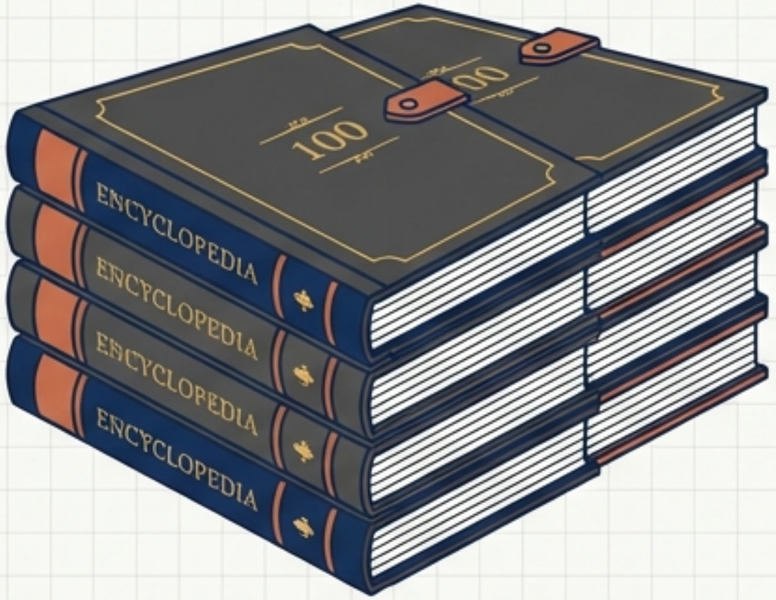
INSTANT



Simply plugs numbers into an equation. Dataset size is irrelevant.

Head-to-Head: Storage Implications

Instance-Based



Heavy Footprint

Model size = Data size.
Linear growth.
1GB Data = 1GB Model.

Model-Based

Light Footprint

Model size = Few Parameters.
Constant size.
1GB Data = ~4KB Model.

Summary Matrix

Feature	Instance-Based (The Memoriser)	Model-Based (The Generaliser)
Core Logic	Memorisation	Generalisation
Training Speed	Fast (Zero effort)	Slow (Computationally heavy)
Prediction Speed	Slow (Calculates on fly)	Fast (Instant formula)
Storage	High (Keeps all data)	Low (Keeps only rules)
Sensitivity	Sensitive to Outliers	Robust (Ignores noise)
State	Lazy	Eager

Identifying Them in the Wild

Classifying common algorithms by their nature.

Instance-Based

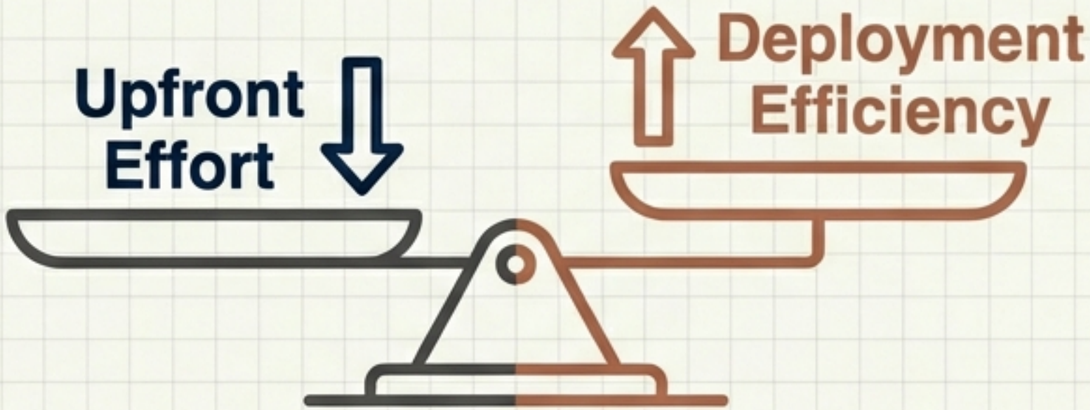
- K-Nearest Neighbours (KNN)
- Kernel Machines / RBF Networks
- LazyAssociative Classifiers

Model-Based

- Linear & Logistic Regression
- Neural Networks (Deep Learning)
- Decision Trees & Random Forests
- Naive Bayes

The vast majority of modern, complex ML systems (like Large Language Models) are Model-Based.

The Core Decision



Choose **Instance-Based** if:

- Your data is small
- Training time must be instant
- You need the system to adapt immediately to new data without retraining

Choose **Model-Based** if:

- You need fast predictions
- Low storage footprint
- The ability to find general rules that ignore the noise

In a world of Big Data, we increasingly lean toward the Generaliser—but there is always a place for the Memoriser.