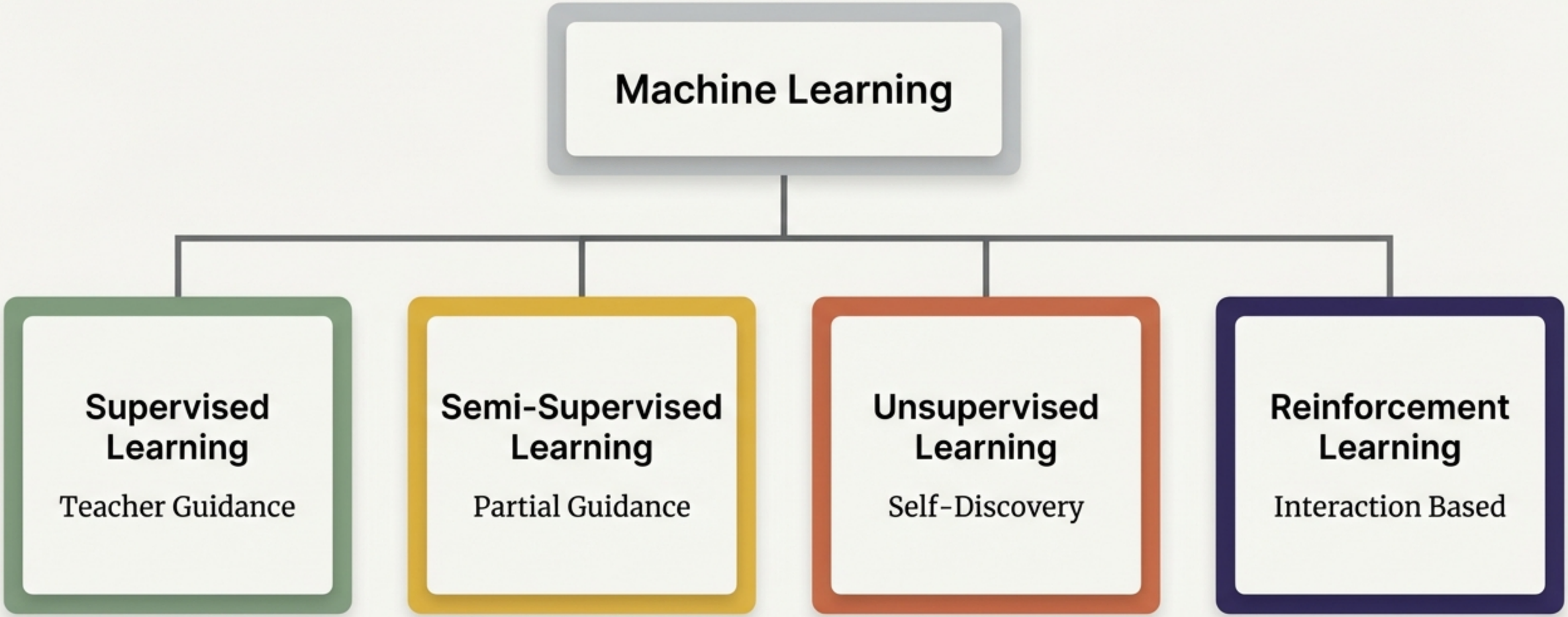
The Landscape of Machine Learning

A taxonomic guide to the four dominant algorithms based on supervision requirements.



Classification variable: The amount of external supervision required for training.

Full Supervision

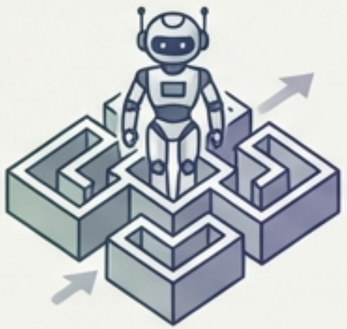
Zero Data / Interaction

The Core Differentiator: **Supervision**

All Machine Learning depends on data, but the nature of that data determines the learning style. To choose the right algorithm, we must ask one fundamental question:

“Does the training data include the answers (labels), or must the machine find the answers itself?”

Supervised



Reinforcement

Supervised Learning: The Teacher and the Student

Definition:
Learning where the system is provided with both **Input (Questions)** and **Output** (Answers).

The Process:

1. Algorithm analyses labelled student data (5,000+ rows).
2. It maps the mathematical relationship between **IQ/CGPA** and **Placement.**

Goal: When given a new student's input (e.g., IQ 7.3), it predicts the output based on the learned rules.

	Input (X)		Output (Y)
	IQ	CGPA	Placement?
1	87	7.1	Yes
2	111	8.9	Yes
3	75	6.3	No
4	**73**	**7.3**	**???**

Two Paths of Supervision

The type of output data determines the sub-category.

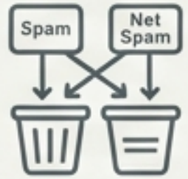
Regression

Output is Numerical / Continuous.

Rooms	Area	Price (Output)
3	1200 sqft	£450,000

Used for predicting quantity (e.g., Salary, House Price).

Classification



Output is Categorical / Class.

Subject	Keywords	Is Spam? (Output)
Win Lottery	Click here	Spam

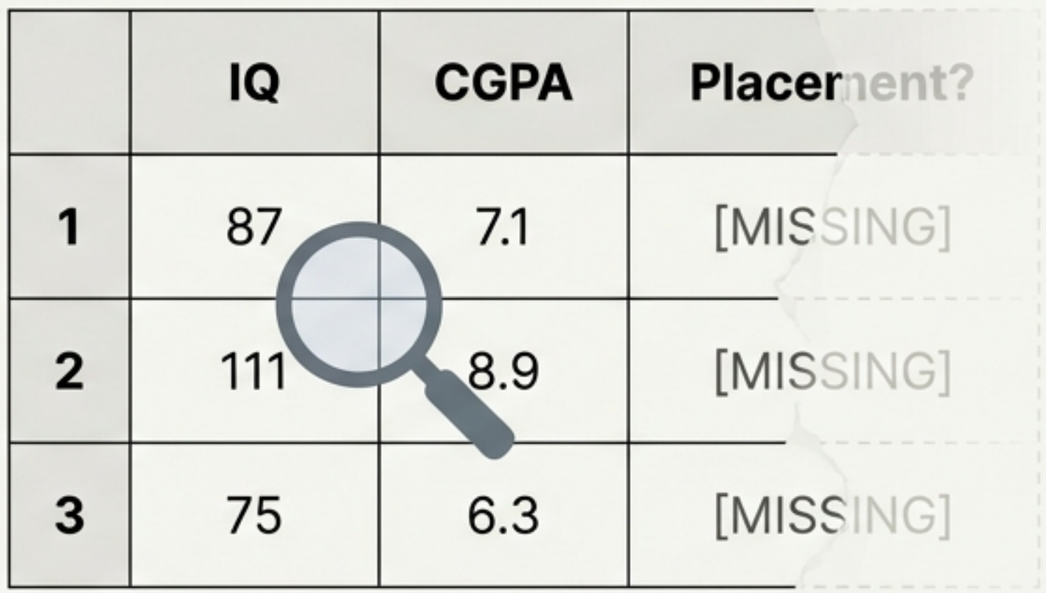
Used for sorting into groups (e.g., Yes/No, Dog/Cat).

Unsupervised Learning: The Explorer

Definition: Learning with Input only. No ‘Teacher’ and no labelled outcome exists.

The Shift: Since prediction is impossible (no answers provided), the goal shifts to Discovery.

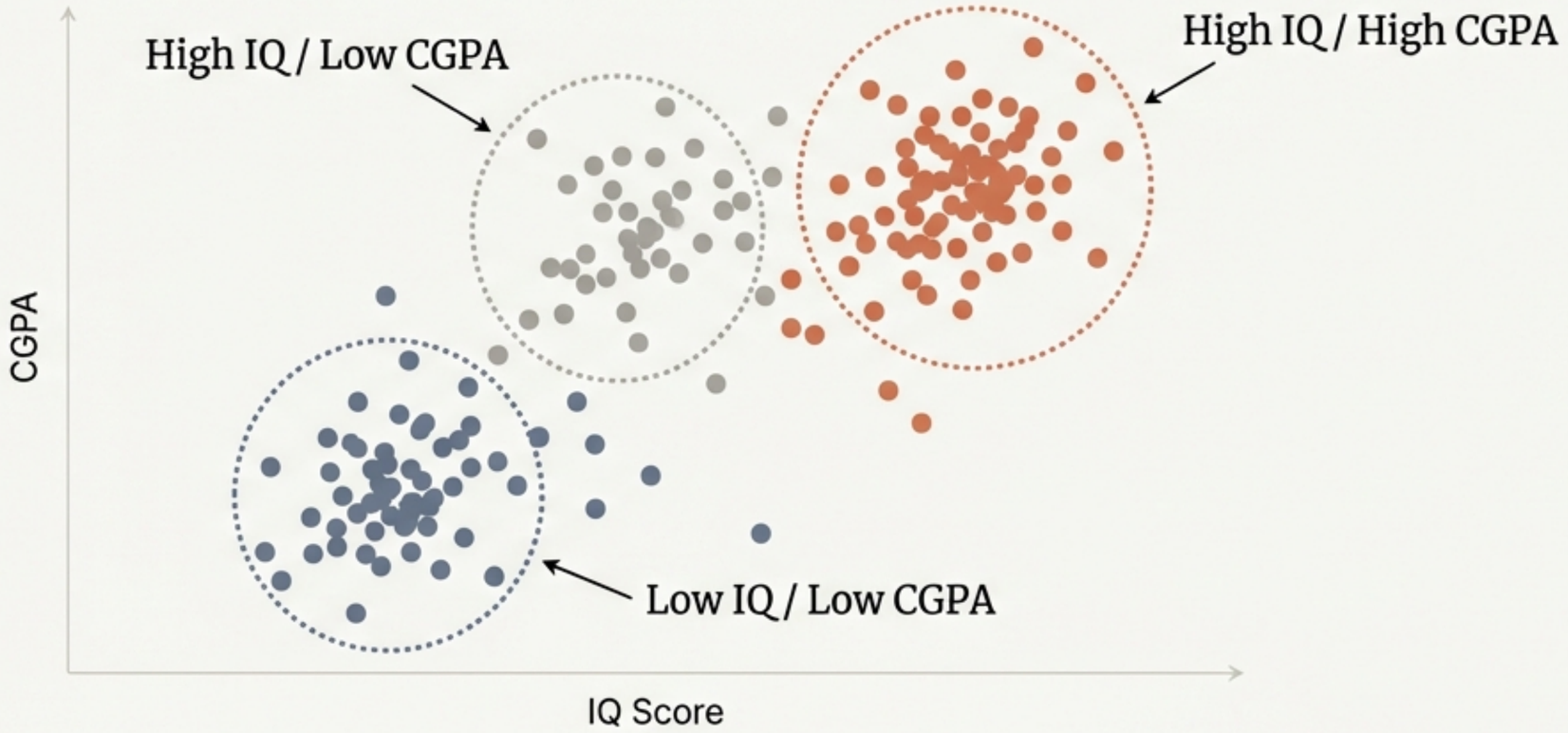
Objective: To find hidden patterns, underlying structures, and natural groupings within the raw data.



	IQ	CGPA	Placement?
1	87	7.1	[MISSING]
2	111	8.9	[MISSING]
3	75	6.3	[MISSING]

Task A: Clustering

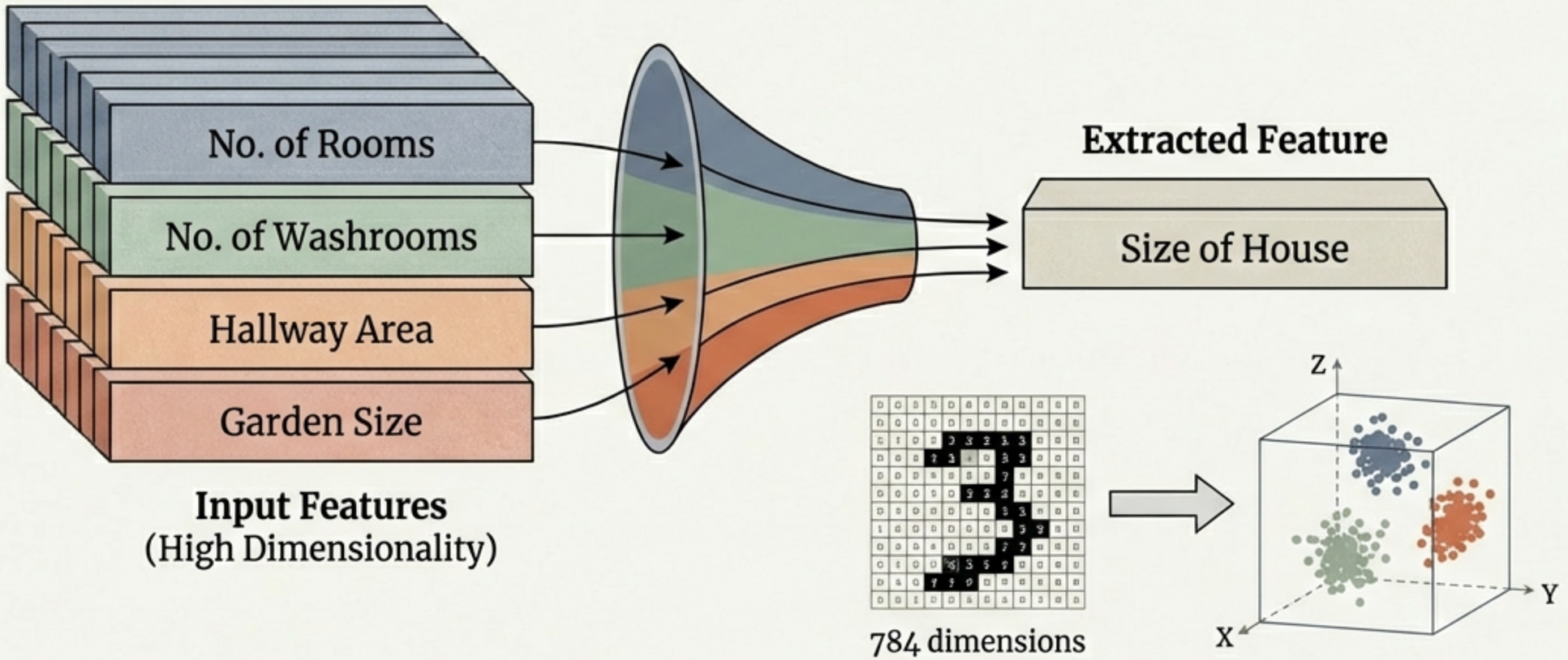
Grouping similar data points based on inherent characteristics.



Real-World Use Case: Customer Segmentation. E-commerce platforms group users by purchasing behaviour (e.g., ‘Budget Shoppers’ vs. ‘High Spenders”) to target marketing without explicit demographic labels.

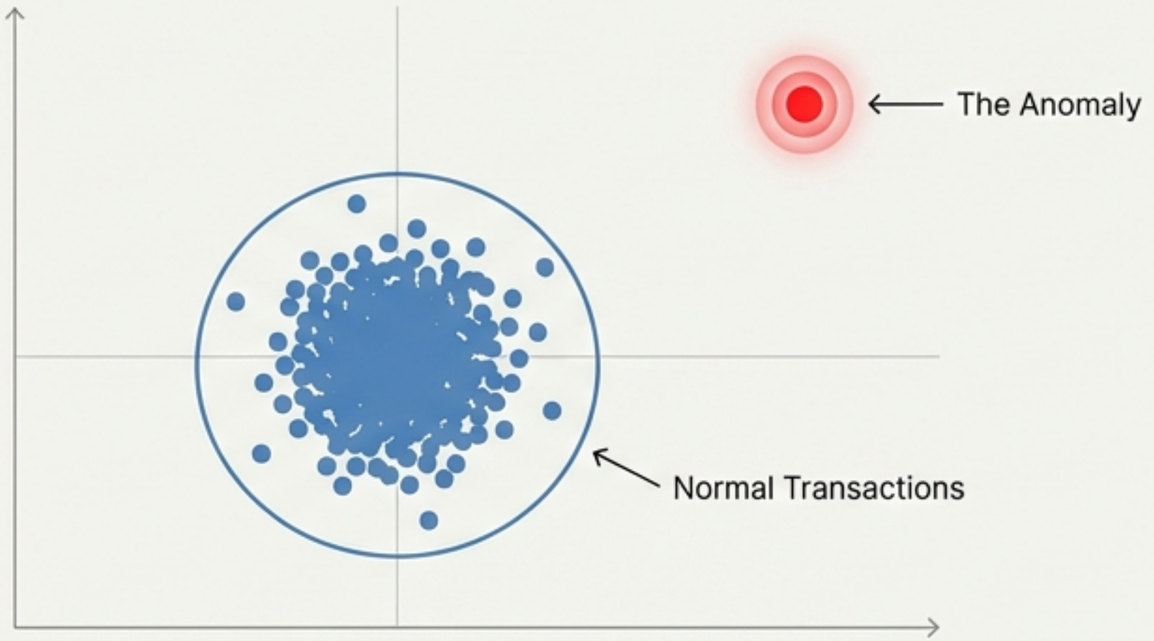
Task B: Dimensionality Reduction

Simplifying complexity by combining correlated features.



Example: Compressing 784 pixels of a handwritten digit into a readable 3D coordinate.

Task C: Anomaly Detection



The Concept

The algorithm builds a model of what “normal” looks like. Anything deviating significantly from this density is flagged.

Applications

- **Fraud Detection:** Credit card spending that differs from user history.
- **Manufacturing:** Defective products on an assembly line.
- **IT Operations:** Sudden spikes in server error rates.

Task D: Association Rule Learning

Discovering the ‘People who bought X also bought Y’ rule.

The Wal-Mart Insight

The Case Study

Analysis of millions of transactions revealed a hidden correlation: Men buying nappies on Friday evenings were highly likely to also buy beer.

The Action

Placing these items adjacent to each other increased sales for both.

The Logic

Finding hidden co-occurrences in transactional data to optimise layouts and recommendation engines.

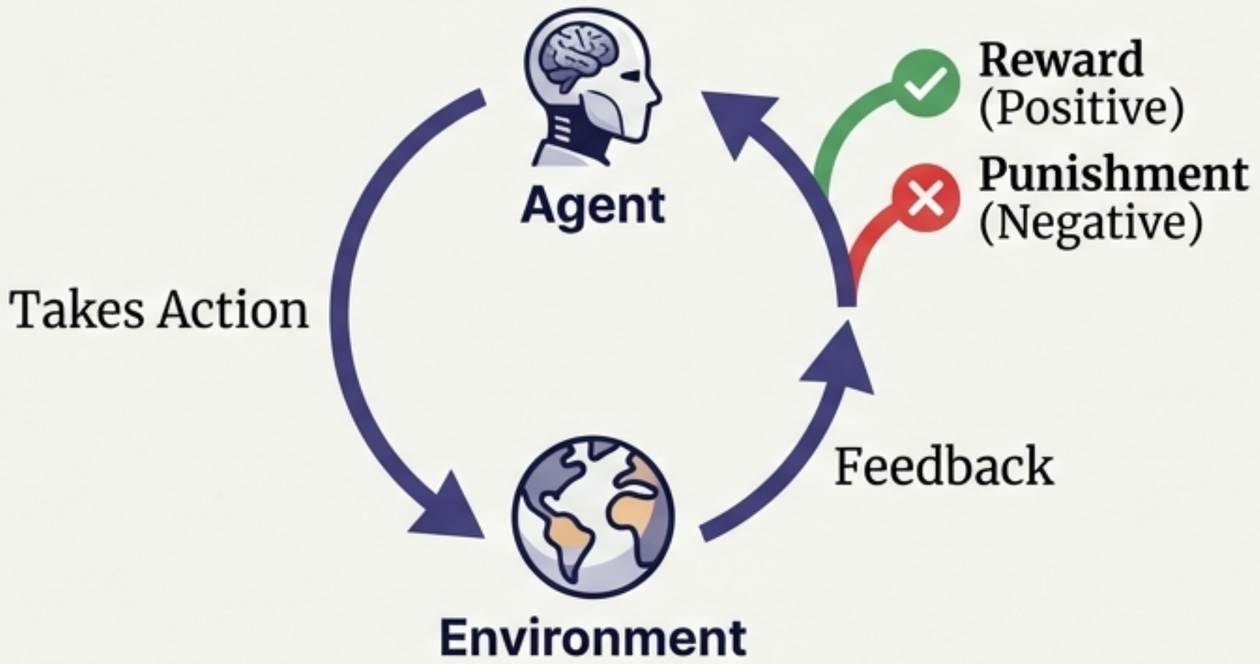
Semi-Supervised Learning: The Hybrid

Labelling data is expensive. Raw data is cheap. This approach bridges the gap.

Example: Google Photos. You tag a face once (Supervised), and the AI infers the identity across your entire library (Unsupervised).

Reinforcement Learning: The Agent

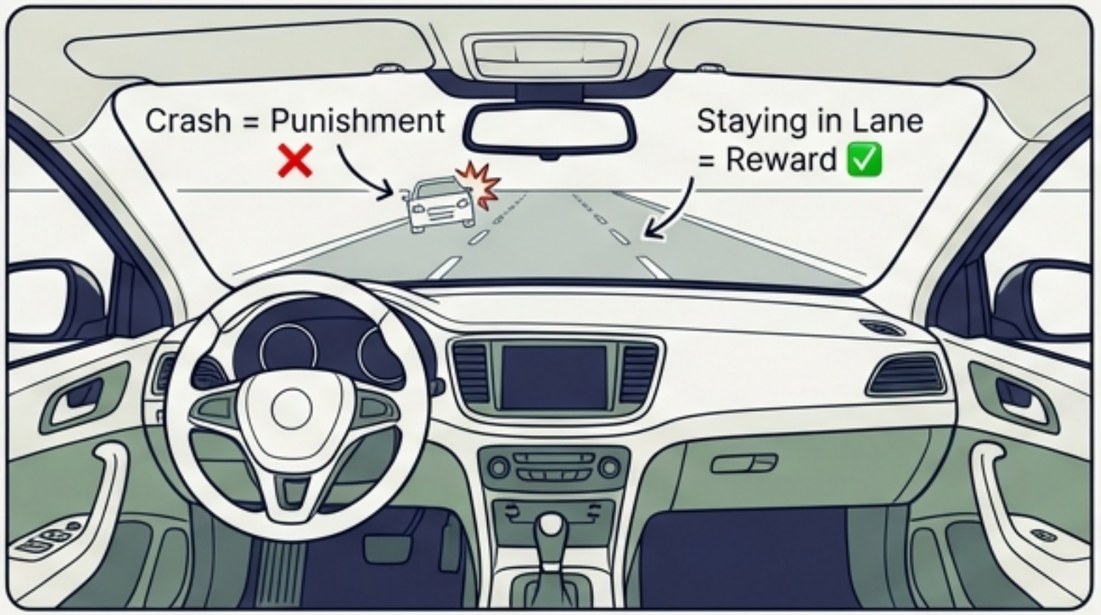
Learning through interaction, not static data.



Think of training a dog. There is no manual. You provide a treat (**Reward**) when it sits, and a scolding (**Punishment**) when it misbehaves. The agent updates its policy to maximise rewards.

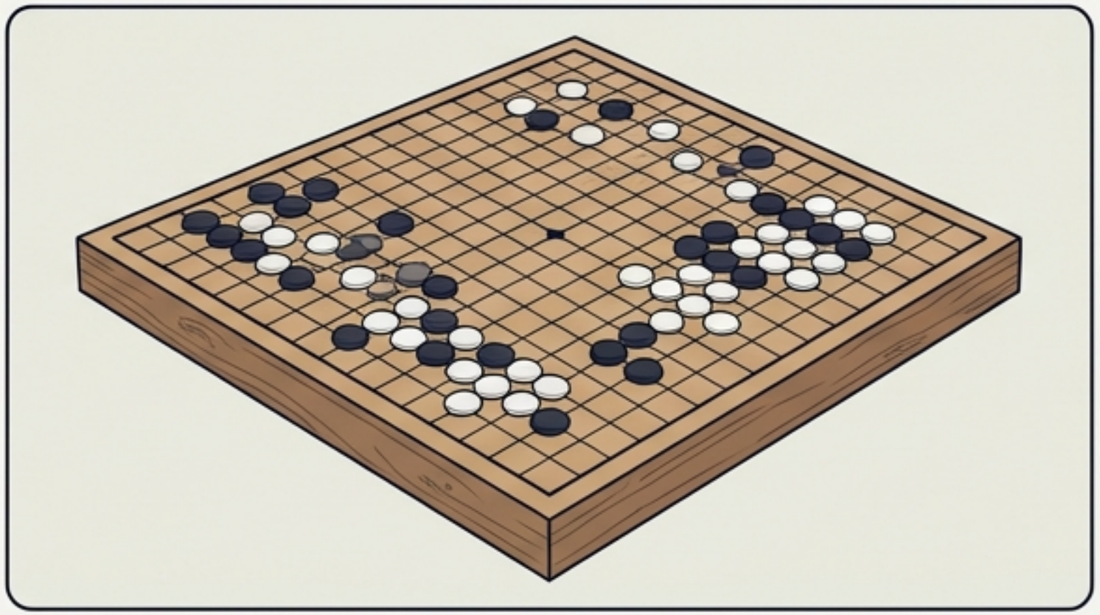
Mastering Complexity through Reinforcement

Self-Driving Cars



Learning the rules of the road from scratch through millions of simulated miles.

AlphaGo



Defeated the world champion by playing millions of games against itself, discovering strategies human masters had never conceived.

Summary: Choosing the Right Approach

The choice is dictated by the data you have and the problem you solve.

Type	Data Available	Primary Goal	Key Example
Supervised	Input + Output (Labelled)	Prediction	House Prices / Spam Filter
Unsupervised	Input Only (Unlabelled)	Discovery & Structure	Customer Segments / Market Basket
Semi-Supervised	Small Labelled + Large Unlabelled	Efficiency	Google Photos Face Tagging
Reinforcement	No Data (Interaction Based)	Optimal Strategy	Self-Driving Cars / Chess