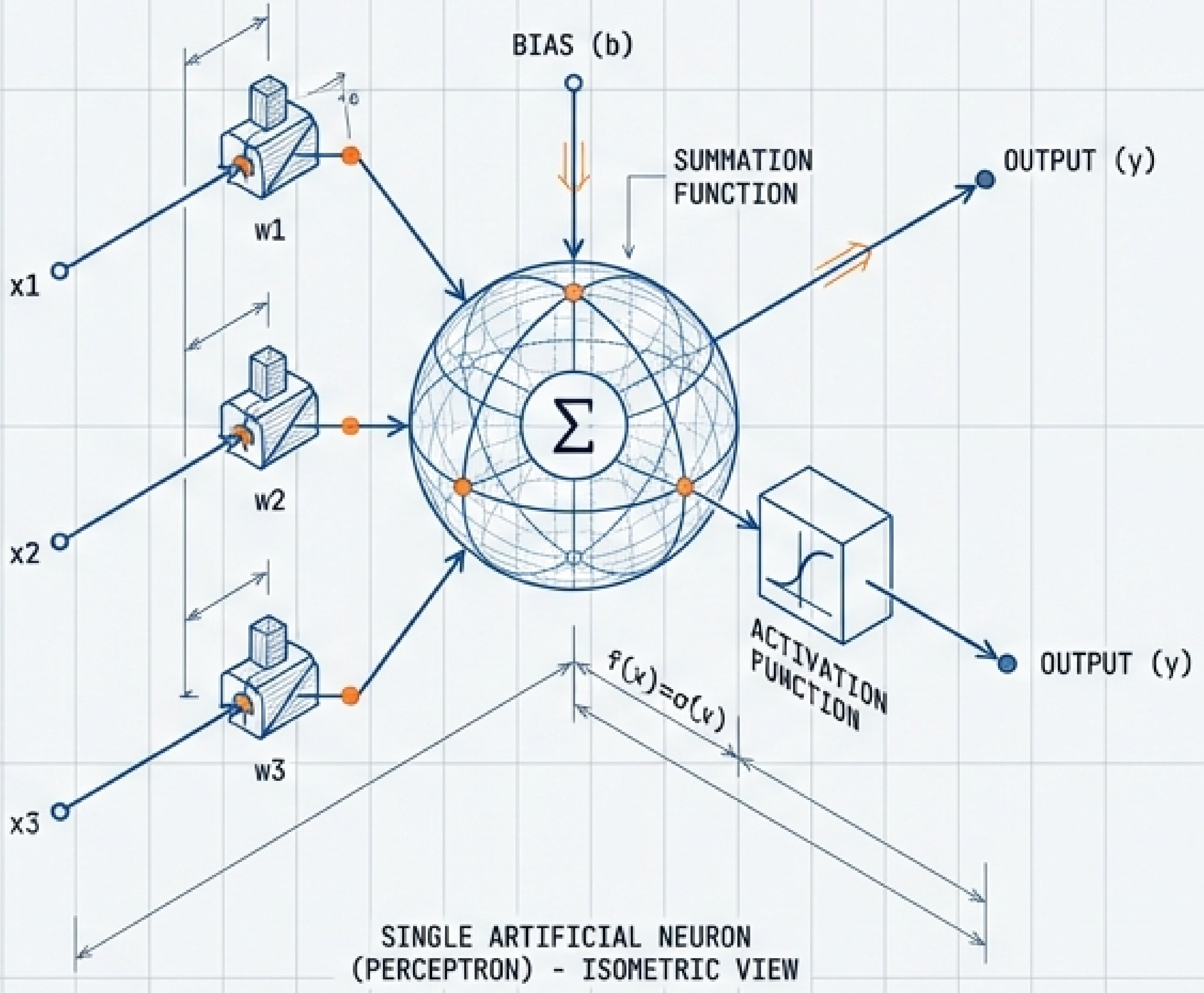
DAY 03

MACHINE LEARNING PRINCIPLES AND PRACTICE

// From Explicit Programming to Learned Patterns

DOC. VER: 1.0 // REV: A



INSTRUCTOR: Anshul | AIML Trainer, Skilloceans

TARGET AUDIENCE: Engineering Year 3

THE ENGINEERING PATH TO MACHINE LEARNING

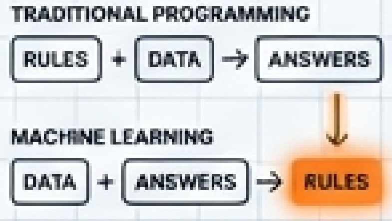
SYSTEM ARCHITECTURE & LEARNING ARC

01.

JetBrains Mono

THE PARADIGM SHIFT

INPUT/OUTPUT INVERSION

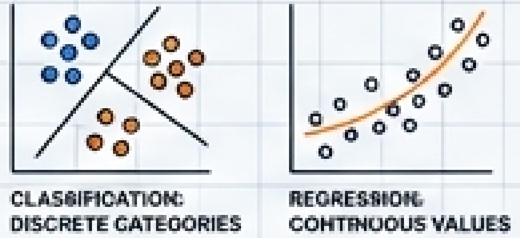


JetBrains Mono

THE LOGIC

CLASSIFICATION VS. REGRESSION

02.

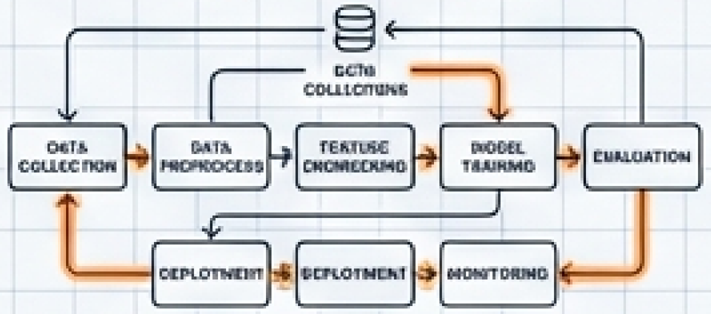


03.

JetBrains Mono

THE ARCHITECTURE

LIFECYCLE & PIPELINES

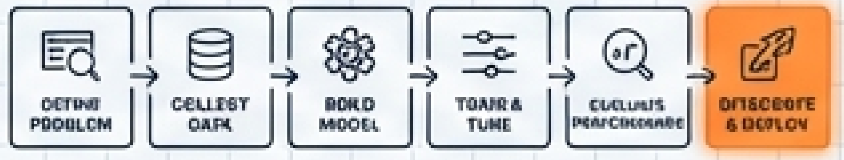


JetBrains Mono

THE APPLICATION

BUILDING A PREDICTOR (WALKTHROUGH)

04.

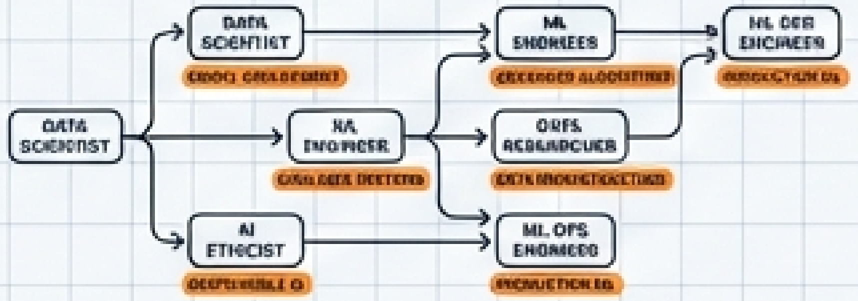


05.

JetBrains Mono

THE CAREER

ROLE ROADMAP

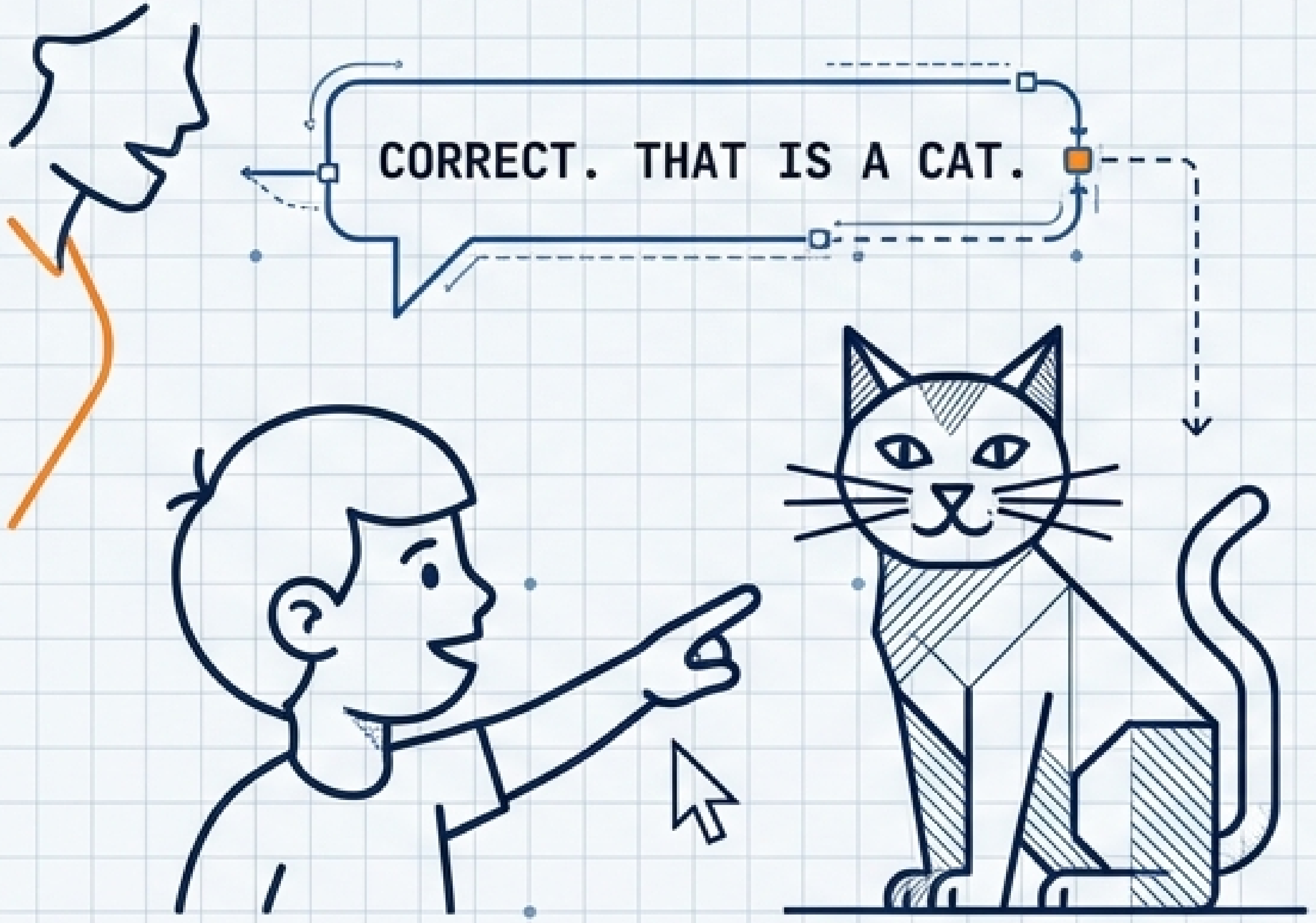


BLUEPRINT BLUE

INTUITION: LEARNING WITHOUT INSTRUCTION

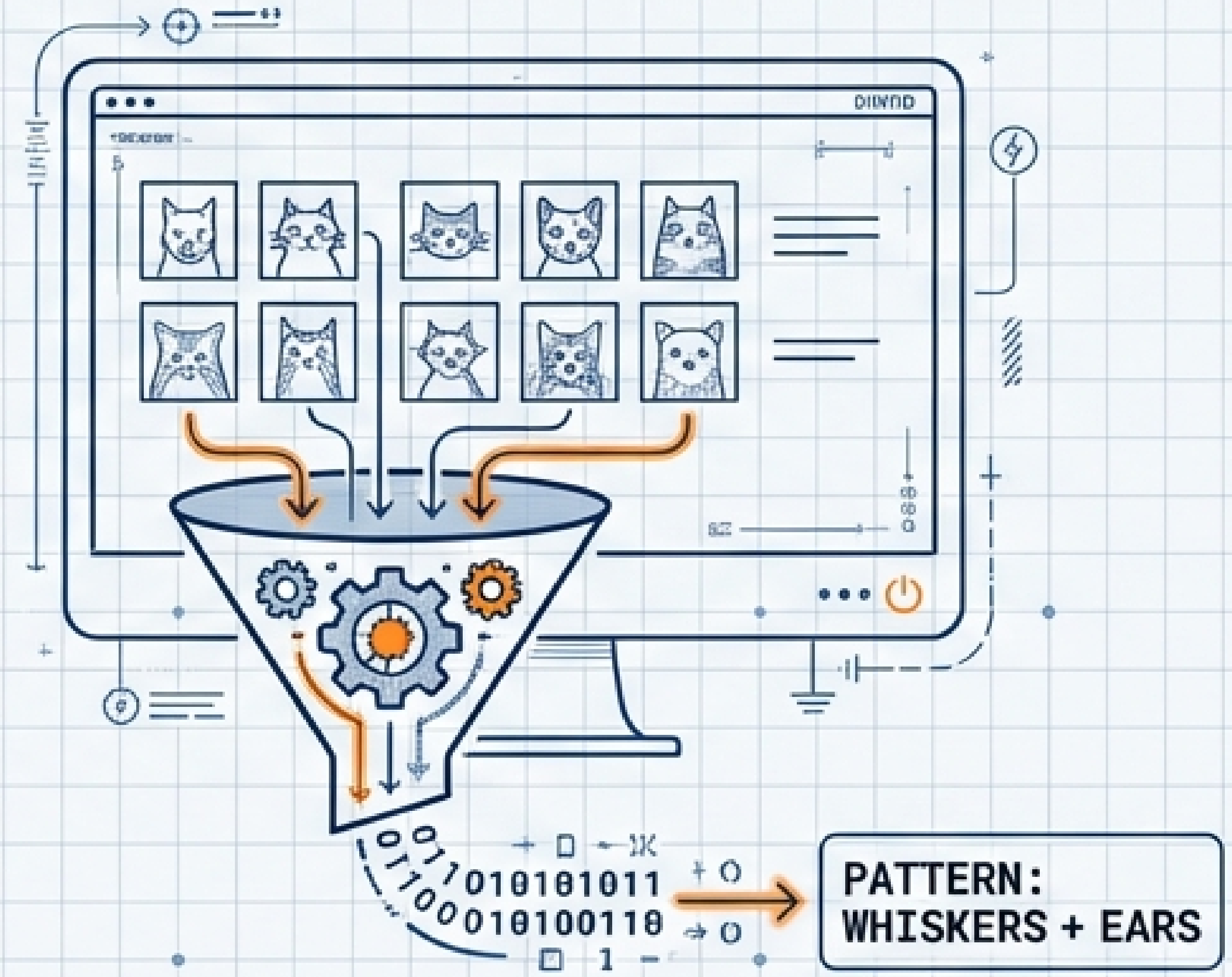
Implicit Pattern Recognition vs. Explicit Rule Following

HUMAN LEARNING (TRIAL & ERROR)



Feedback Loop: Positive reinforcement strengthens the neural connection.

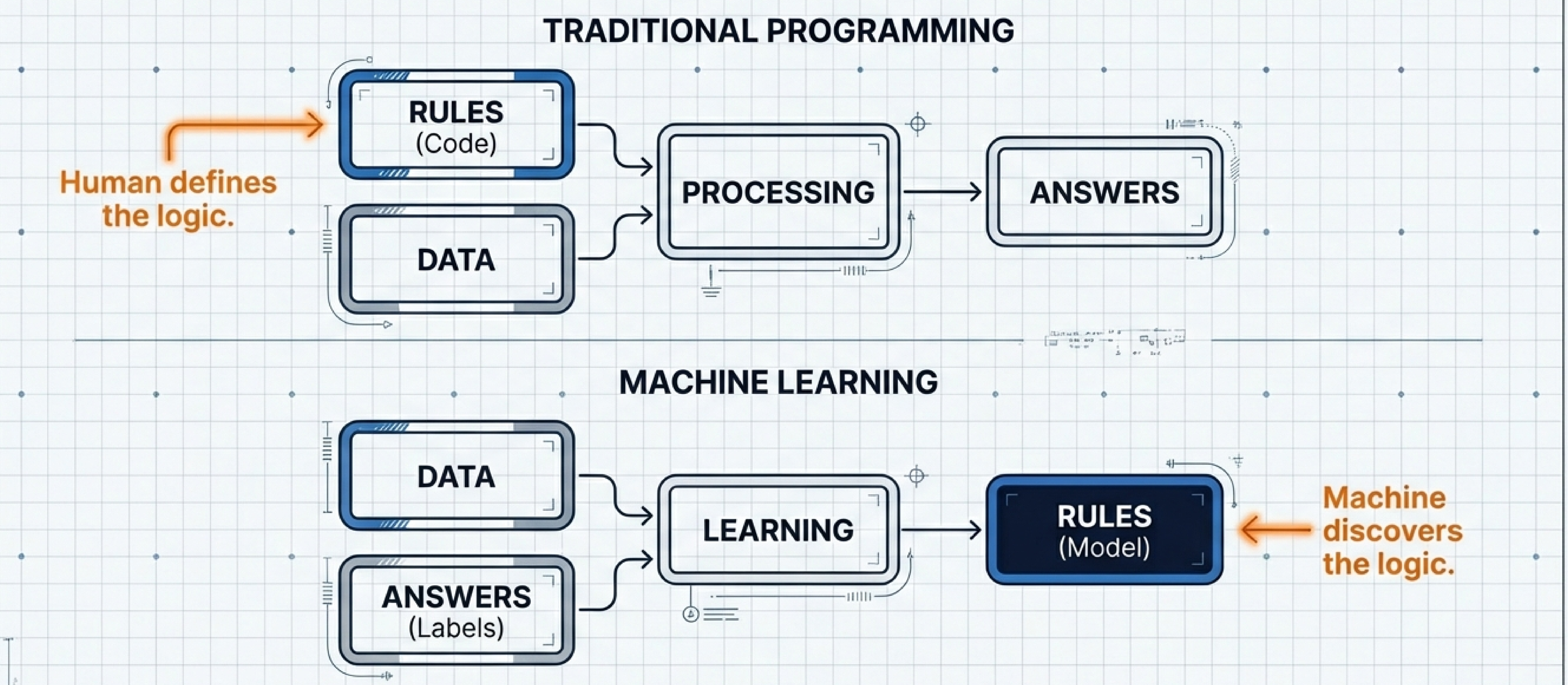
MACHINE LEARNING (DATA PROCESSING)



Pattern Loop: Statistical correlation strengthens the mathematical weight.

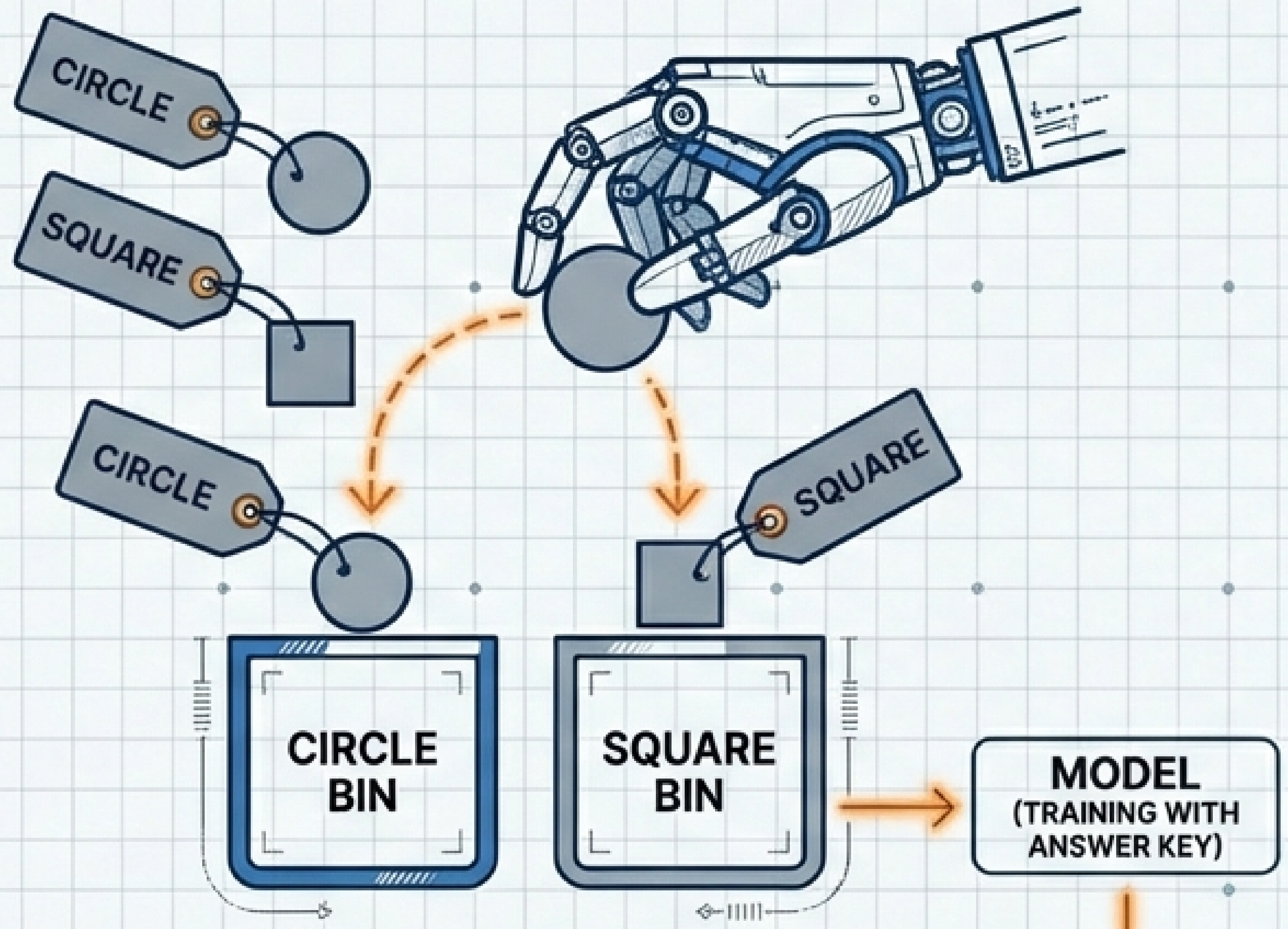
DEFINITION: Machine Learning is a subset of AI where computers have the ability to learn without being explicitly programmed. — Arthur Samuel

THE PARADIGM SHIFT: INVERTING CONTROL



TAXONOMY: SUPERVISED VS. UNSUPERVISED

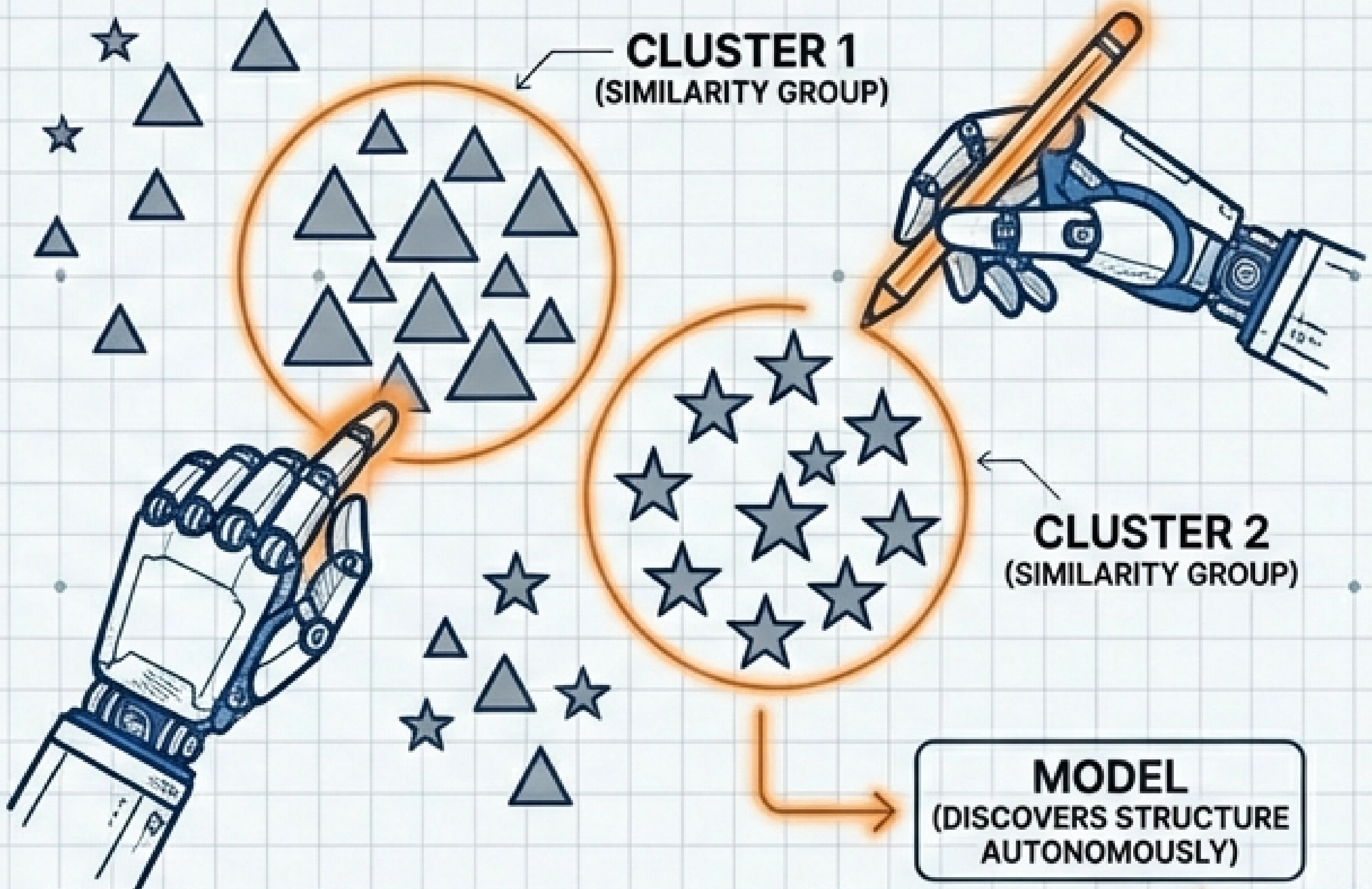
SUPERVISED LEARNING



Input data is **LABELED.** The model trains with an **answer key**.

Example: Image Classification, Spam Detection

UNSUPERVISED LEARNING



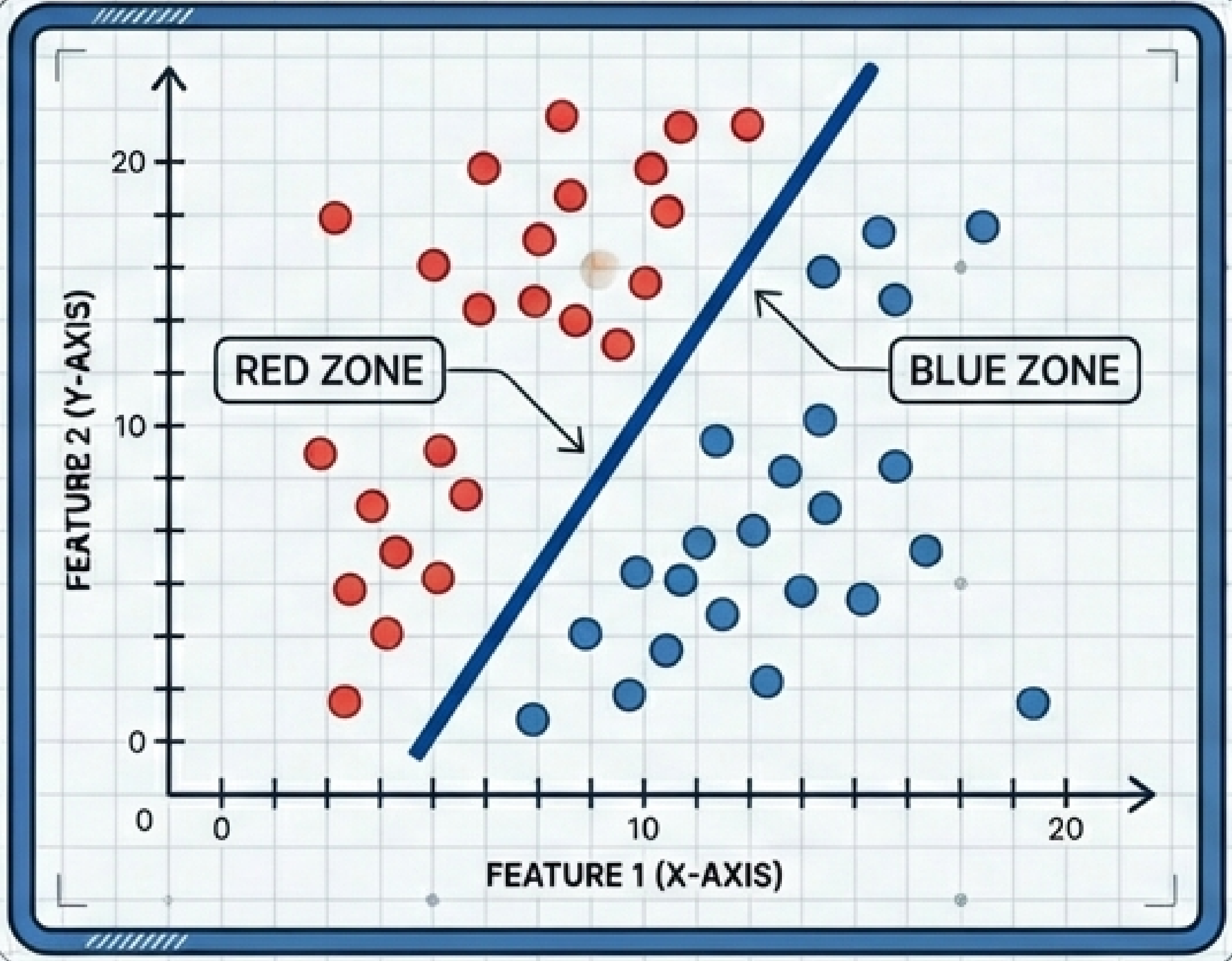
Input data is **UNLABELED.** The model **finds structure** autonomously.

Example: Customer Segmentation, Anomaly Detection

DECISION DRIVERS: CLASSIFICATION VS. REGRESSION

Choosing the math based on the output type.

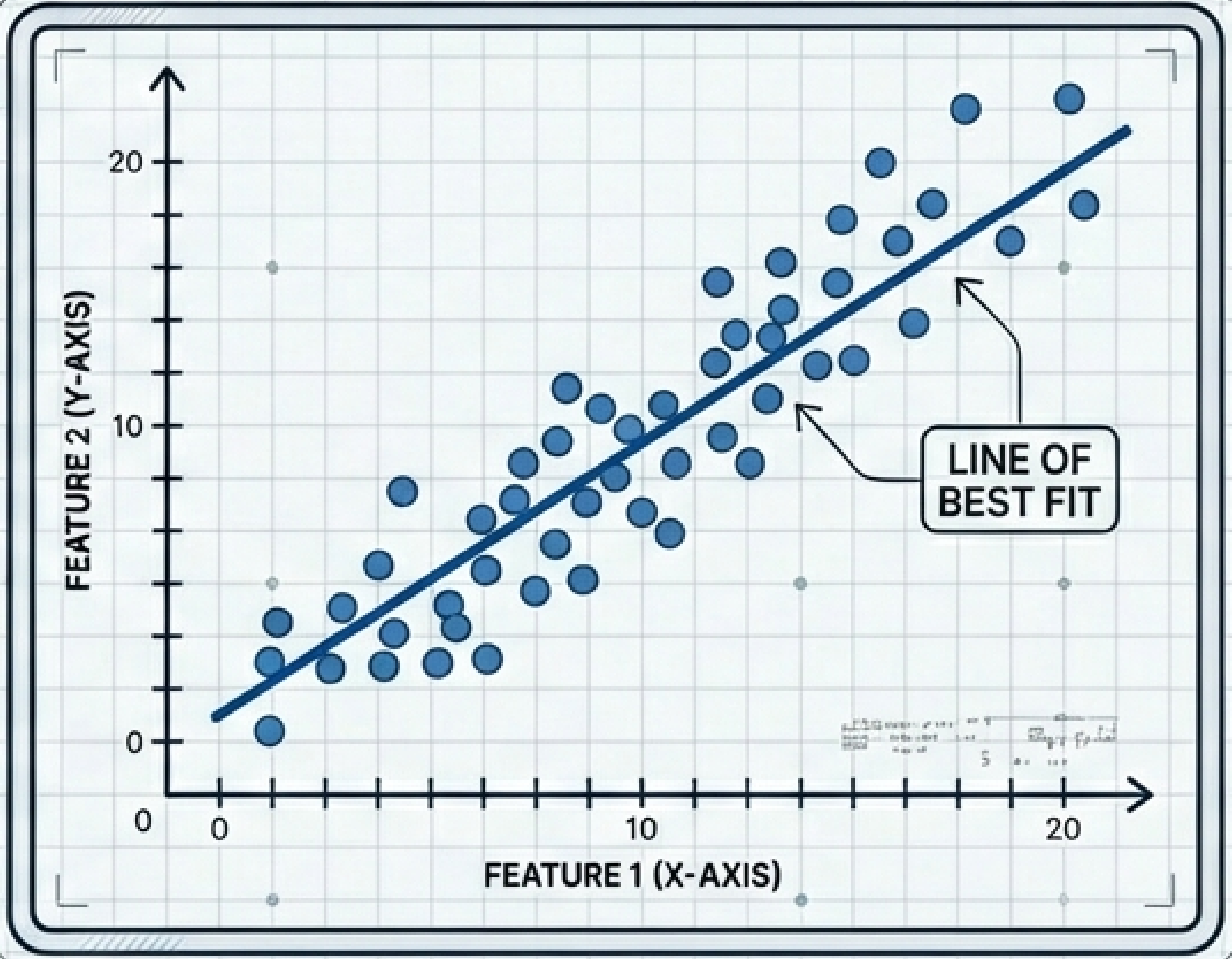
CLASSIFICATION



Output = Discrete Category (Class A or Class B).

Is it spam? (Yes/No)

REGRESSION



Output = Continuous Value (Quantity).

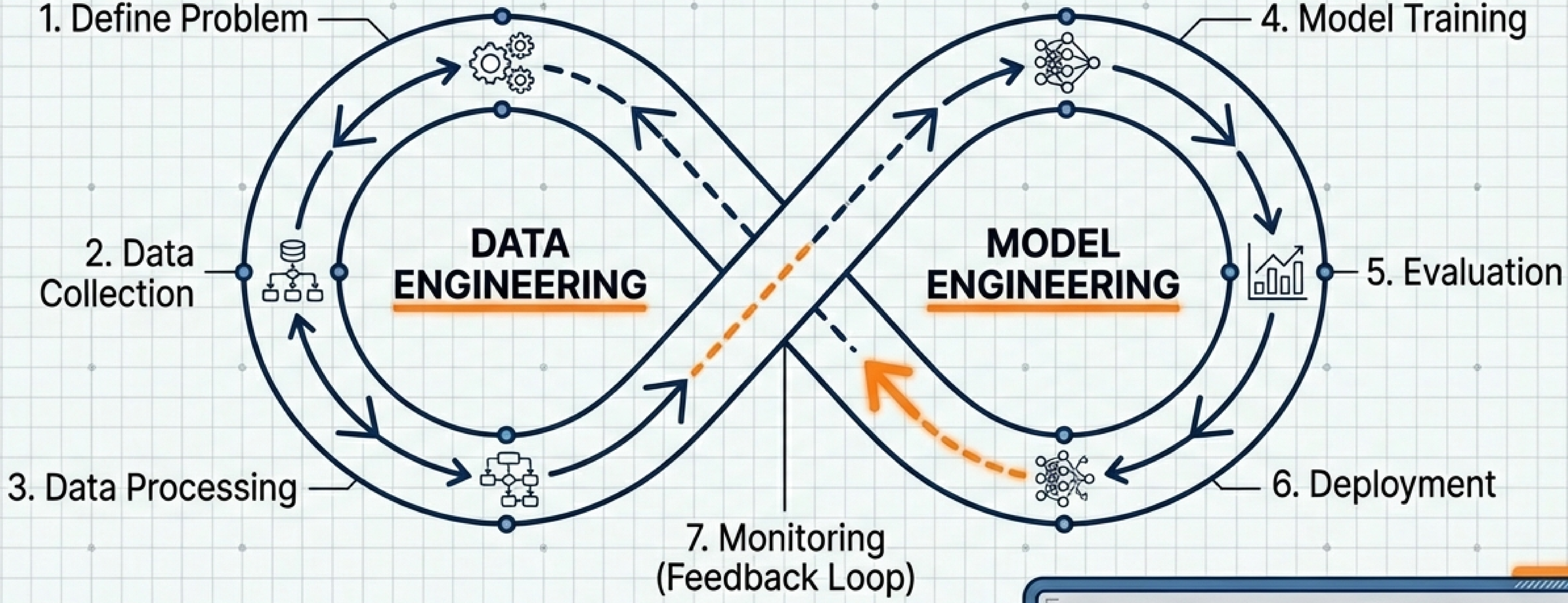
How much will it cost? ($)

PROBLEM MAPPING MATRIX

Matching industry problems to ML techniques.

PROBLEM	QUESTION ASKED	ML TECHNIQUE	TYPE
Email Filter	Is this spam?	Classification	Supervised
Housing Prices	What is the value?	Regression	Supervised
Credit Card Fraud	Is this transaction weird?	Anomaly Detection	Unsupervised
Market Marketing	Who buys similar items?	Clustering	Unsupervised

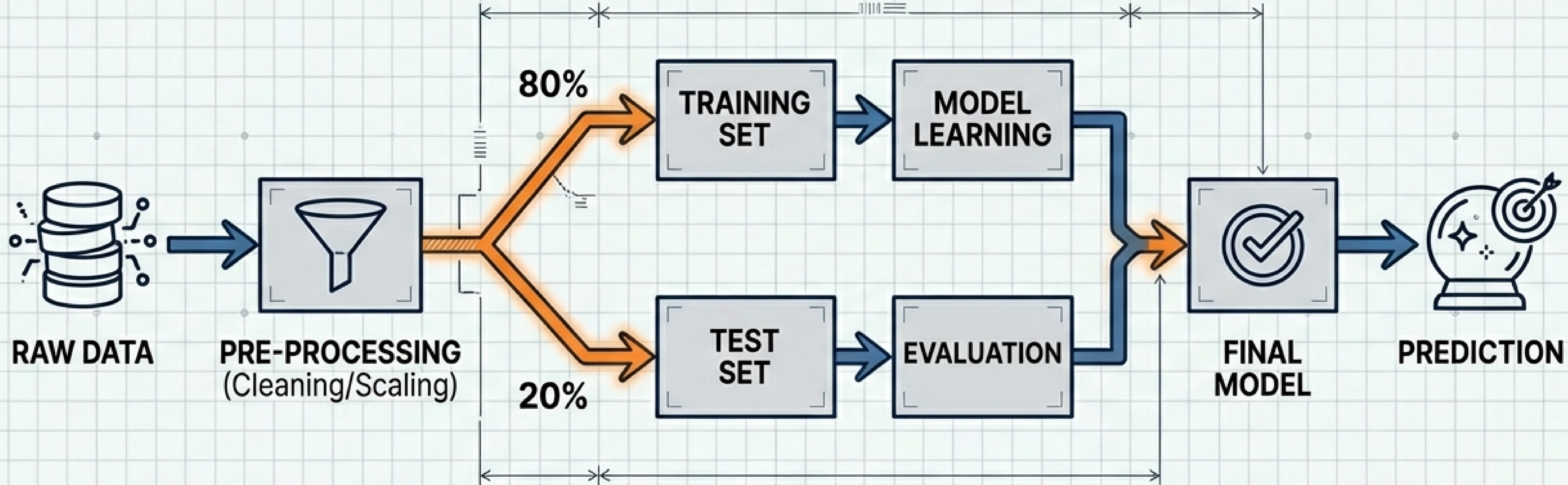
THE AI PROJECT LIFECYCLE (MACRO VIEW)



Note: Modeling is only ~20% of the effort. Data prep and monitoring comprise the majority of the engineering lifecycle.

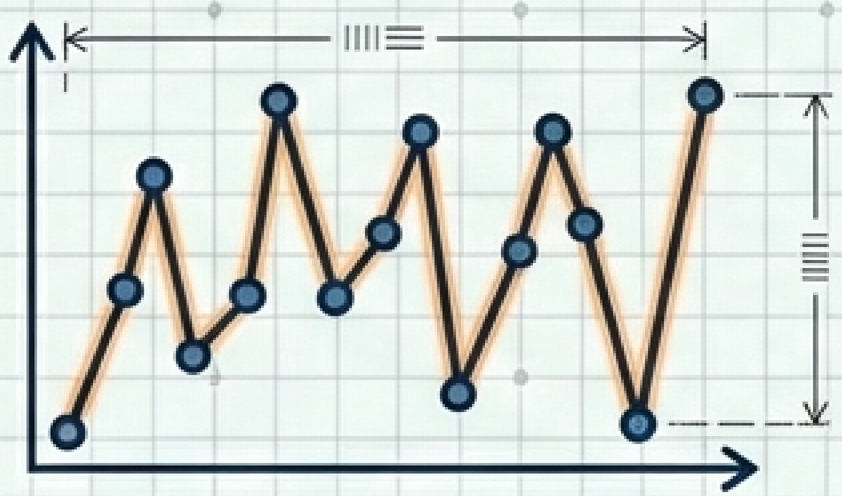
THE TECHNICAL PIPELINE (MICRO VIEW)

Data Transformation Workflow



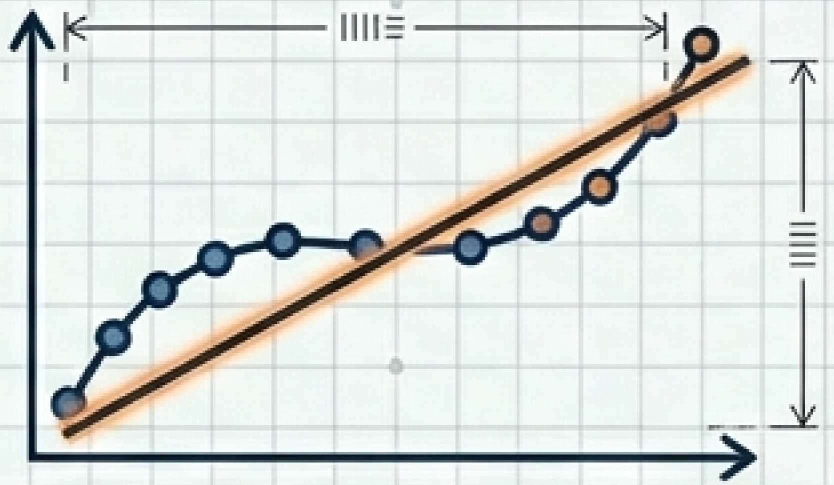
ENGINEERING CONSTRAINTS: FAILURE MODES

1. OVERFITTING



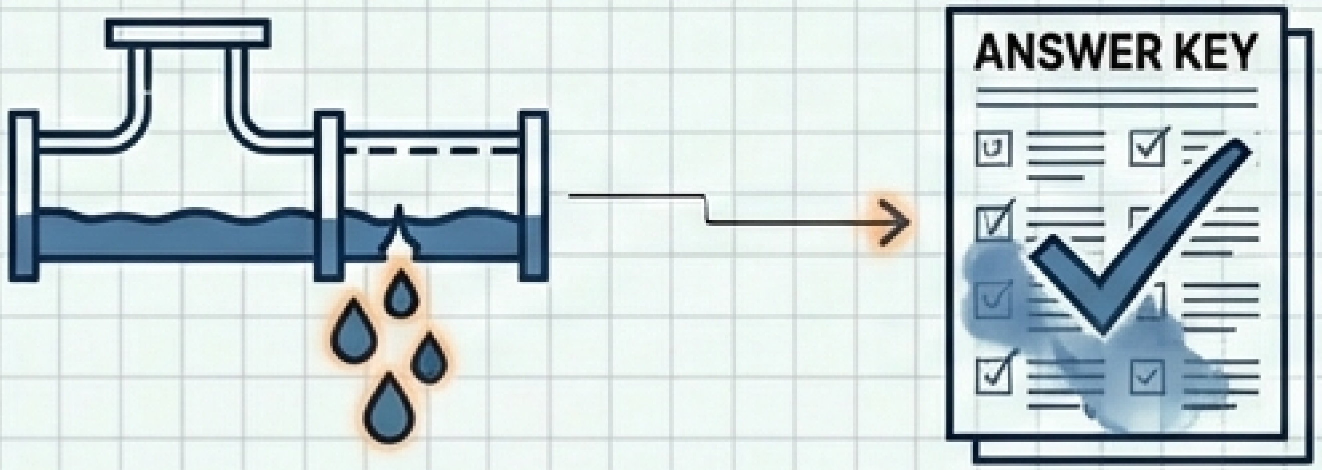
Memorizing noise instead of learning patterns.
High accuracy on training, fails in production.

2. UNDERFITTING



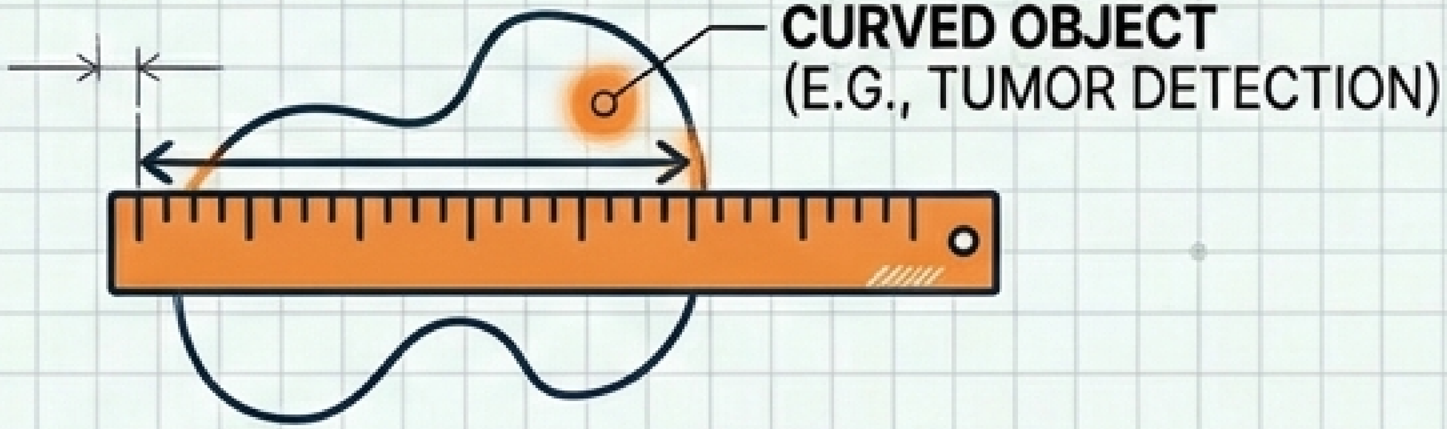
Model is too simple to capture the complexity.

3. DATA LEAKAGE



Training with information that won't be available at prediction time (Cheating).

4. METRIC MISMATCH



Optimizing for Accuracy when Recall is needed (e.g., Cancer detection).

VISUAL WALKTHROUGH: BUILDING A PREDICTOR

1. THE DATA

CSV

SqFt	Bedrooms	Age	PRICE
1500	3	10	350000
2200	4	5	500000
1200	2	15	280000
1800	3	8	420000
1800	3	8	420000
...	...	...	...

FEATURE MATRIX (X)

TARGET VECTOR (y)

2. THE CODE

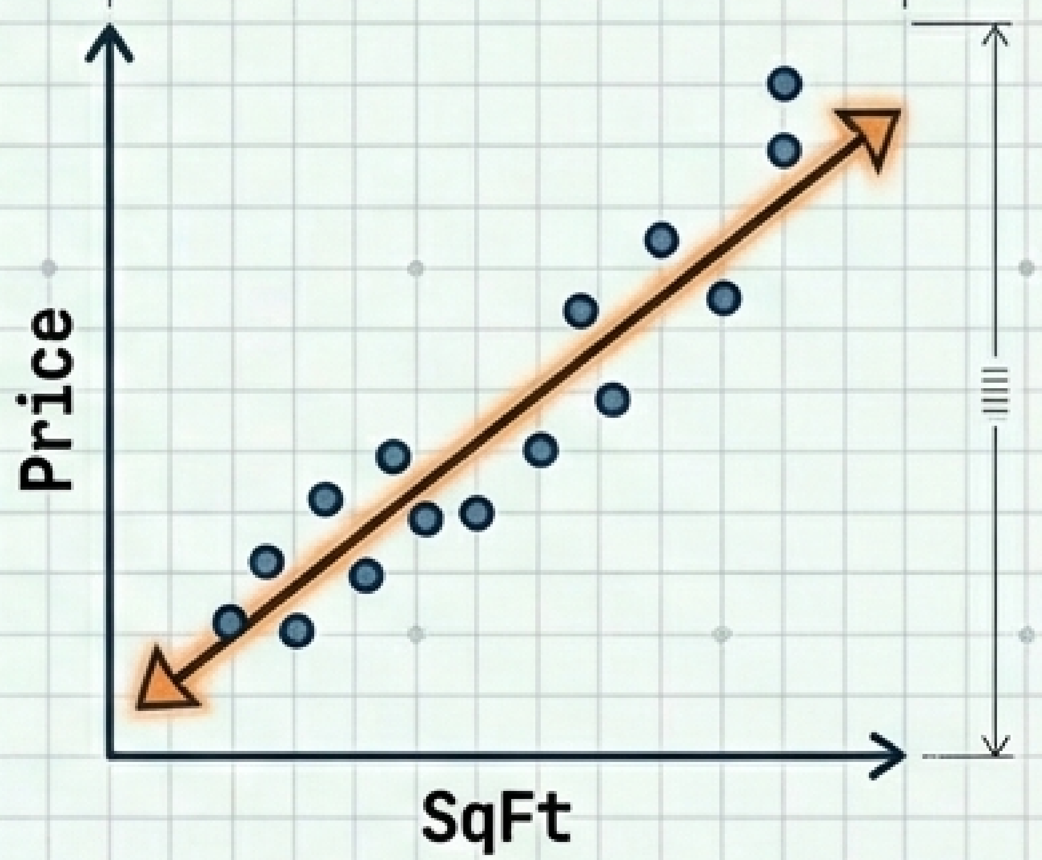
>_ PYTHON CONSOLE

```python
from sklearn.linear_model import
LinearRegression

model = LinearRegression()

model.fit(X_train, y_train)

prediction = model.predict(X_test)

_
```

TRAINING PHASE

PREDICTION PHASE

3. THE RESULT



Model learns the coefficient weights to predict Price based on SqFt.

ARCHITECTURE CHALLENGE: LOAN APPROVAL SYSTEM

1. CLASSIFICATION OR REGRESSION?

(Prompt: Are we predicting a number or a decision?)

PROBLEM STATEMENT:

Bank requires automated approval system to reduce defaults.

2. FEATURE SELECTION

(Prompt: Income, Credit Score, Debt-to-Income Ratio...)

3. SUCCESS METRIC

(Prompt: Minimize False Positives or False Negatives?)

THE CAREER ROADMAP

1. DATA ANALYST

Technical: JetBrains, SQL, Visualiau/PowerBI), Statistics

Descriptive Analytics, SQL, Visualization (Tableau/PowerBI), Statistics

2. ML ENGINEER

Technical: JetBrains, Scikit-Learn, MLOps, Model Deployment, Python

Predictive Modeling, Scikit-Learn, MLOps, Model Deployment, Python

3. AI SCIENTIST

Technical: Learning, Neural Networks, TensorFlow/PyTorch, Algorithm R&D

Deep Learning, Neural Networks, TensorFlow/PyTorch, Algorithm R&D

FOUNDATION: MATHEMATICS (LINEAR ALGEBRA, CALCULUS) & PROGRAMMING

ESSENTIAL TECHNICAL BASELINE

PROJECT: CAREER PROGRESSION MAP		**DRAWING NO:** CAR-003
REV: A	**DATE:** 2024-05-20	**ENGINEER:** SWISS SCHEMATIC DESIGN

CRITICAL ENGINEERING TAKEAWAYS

Data > Algorithms

Clean, relevant data is more valuable than a complex model. Garbage in, garbage out.

Inductive Logic

Shift your mindset from writing rules (if/else) to curating training sets.

Iterative Cycle

Deploying is not the end. Models drift and require constant monitoring.

Right Tool, Right Job

Don’t use Deep Learning when a simple Regression will solve the problem.

PROJECT: CRITICAL ENGINEERING TAKEAWAYS		DRAWING NO: ENG-004
REV: A	DATE: 2024-05-20	ENGINEER: SWISS SCHEMATIC DESIGN

CONTINUE THE JOURNEY

RECOMMENDED RESOURCES:

- **Datasets:** Kaggle, UCI Repository
- **Tools:** Scikit-Learn, TensorFlow, Jupyter
- **Practice:** "The best way to learn Machine Learning is to build."

DOCUMENTATION

DATASETS

COMMUNITY