Zero-Shot Learning (ZSL)

How machines recognise things they've never seen

In this presentation, you'll discover what zero-shot learning is, why it's revolutionising machine learning, and how it works at a conceptual level.

The Big Idea Behind ZSL

> "Recognising something new without seeing any examples during training."

🧠 **Analogy:** Imagine you've studied cats, dogs, and birds throughout your life. Then one day, you encounter a zebra for the first time. Even though you've never seen one before, you can infer it's like a horse with distinctive black and white stripes.

➡️ That's precisely **zero-shot reasoning** in action — connecting prior knowledge to understand something entirely new.

Why Zero-Shot Learning Matters

Traditional ML Approach

- Requires labelled examples for every single class
- Expensive and time-consuming data collection
- Fails completely on new categories
- Rigid and inflexible system

Zero-Shot Learning

- Learns to reason about unseen classes intelligently
- Highly cost-efficient approach
- Adapts dynamically to new scenarios
- Scales without constant retraining

💡 **Key insight:** ZSL bridges the crucial gap between "what the model knows" and "what it hasn't yet encountered" — making AI systems more flexible and practical.

Breaking Down the Terminology

Term	Meaning
Shot	The number of training examples provided per class
Zero-Shot	Absolutely 0 training examples for the new class
One-Shot	Just 1 single example to learn from
Few-Shot	A small handful of examples (typically 5–10)

🧩 **The ZSL challenge:** The model receives *zero* training examples but must still accurately recognise and classify new items based on semantic understanding.

The Core Question

How can a model classify something it has never seen?

The answer lies in providing the model with **extra knowledge** about the relationships and characteristics between different concepts:

Shared attributes

Visual and semantic properties that connect related objects

Language semantics

Understanding how words and concepts relate in natural language

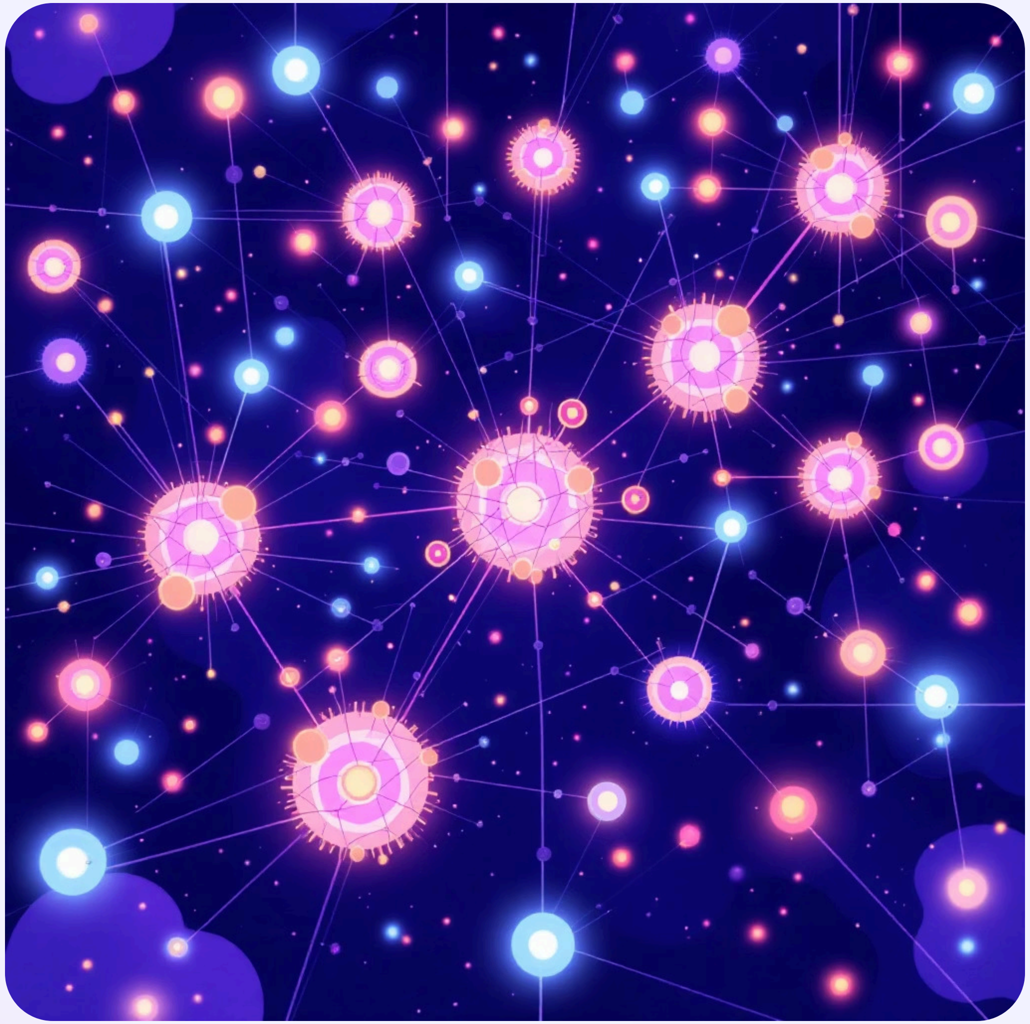
Descriptive text

Rich textual descriptions that capture the essence of concepts

Word embeddings

Dense vector representations that encode semantic meaning

The Two Phases of Zero-Shot Learning



Phase 1: Training (Seen Classes)

The model learns from normal, familiar categories like cats, dogs, and birds. During this phase, it also learns their attributes, characteristics, and detailed text descriptions to build a knowledge foundation.

Phase 2: Inference (Unseen Classes)

When presented with a completely new image or text (for example, a zebra), the model matches its extracted **visual features** with the **semantic information** of unseen categories to make an intelligent prediction.

How ZSL Works: Concept Flow

Input Image

Raw visual data enters the system

Feature Extractor

Deep neural network extracts meaningful patterns and creates a feature vector

Semantic Space Mapping

Features are mapped to a shared semantic representation space

Compare with Class Descriptions

System compares against word embeddings and descriptions of all possible classes

Predicted Class

The closest semantic match becomes the final prediction

Auxiliary Information Used in ZSL

Attributes

Example: "Has fur", "Has stripes", "Four legs"

Purpose: Describes specific visual traits and characteristics that distinguish objects from one another

Semantic Embeddings

Example: Word2Vec, GloVe, or BERT vectors

Purpose: Captures rich semantic meaning in high-dimensional numerical representations

Text Descriptions

Example: "A zebra is a striped horse-like animal"

Purpose: Provides natural language clues and context for unseen items

Understanding Through Analogy

📘 Traditional Machine Learning

Like memorising every single flashcard individually without understanding the underlying patterns or relationships. Time-consuming and inflexible.

📗 Zero-Shot Learning

Like learning the roots, prefixes, suffixes, and logical rules — so you can intelligently guess new words you've never encountered before.

Concrete example: Once you understand that *tele* means "far" and *phone* means "sound", you can logically deduce that "telephone" refers to sound transmitted from far away — even if you've never seen the word before.

Summary: Key Takeaways

Learns from Seen Classes

Builds foundational knowledge using semantic information from familiar categories

Applies Logic to Unseen Classes

Uses reasoning and semantic understanding to handle novel scenarios

Uses Multiple Information Sources

Leverages embeddings, attributes, and textual descriptions effectively

Reduces Data Requirements

Dramatically cuts down the need for expensive labelled training data

---

➡️ **Coming up next:** We'll dive deeper into how ZSL works technically – exploring the architecture, mathematical foundations, and real-world applications that are transforming industries today.