Introduction to Machine Learning

Qing-Yuan Jiang

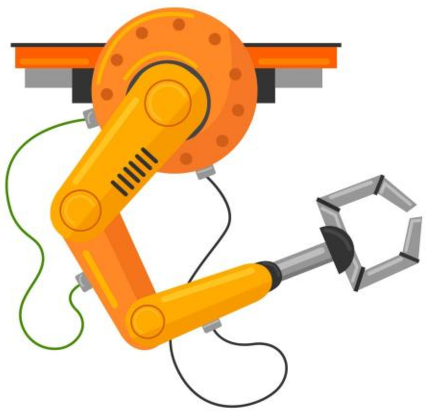
The school of Computer Science and Engineering, NJUST

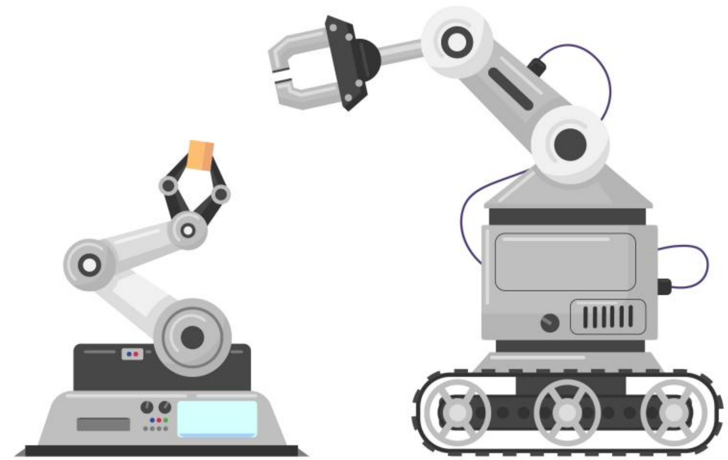
2026.09.17

What is Machine Learning?

Machine Learning

Machine

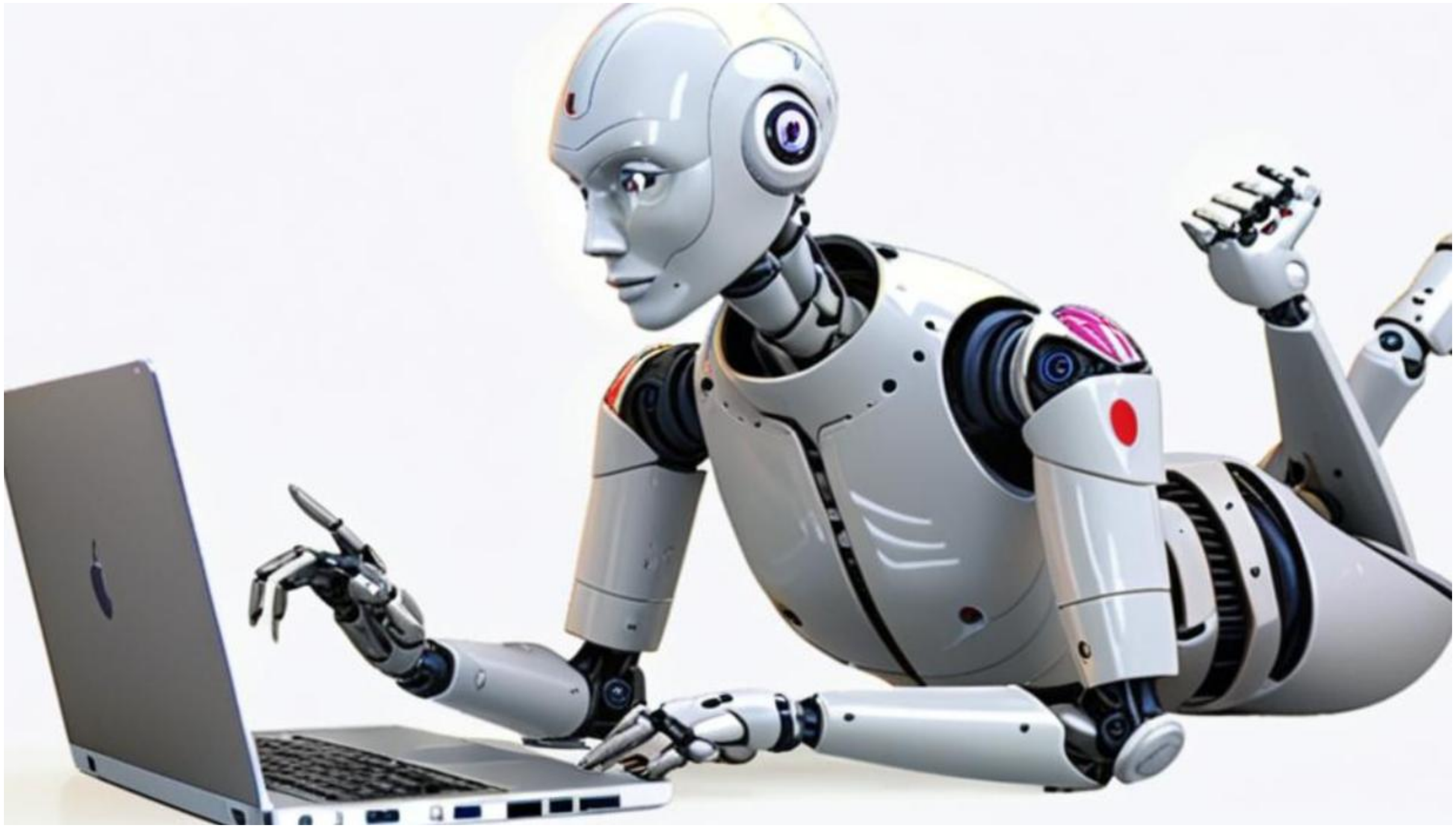




Learning

Machine Learning

- Humans continuously improve their capabilities by learning from past experiences
- Machines learn skills from the data to improve their performance

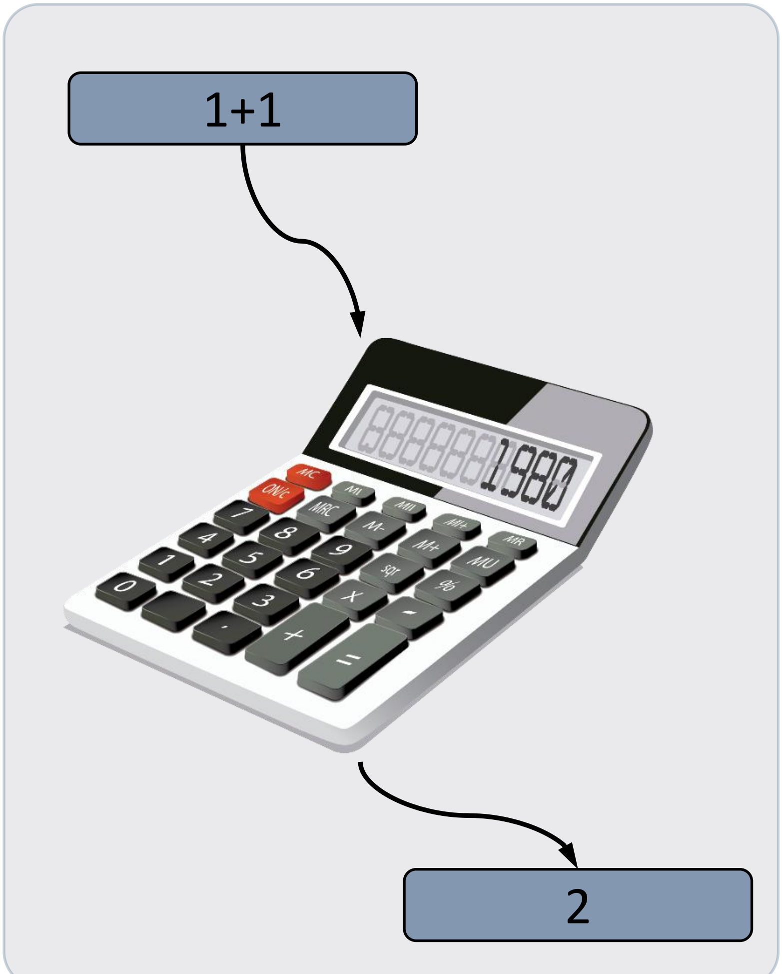


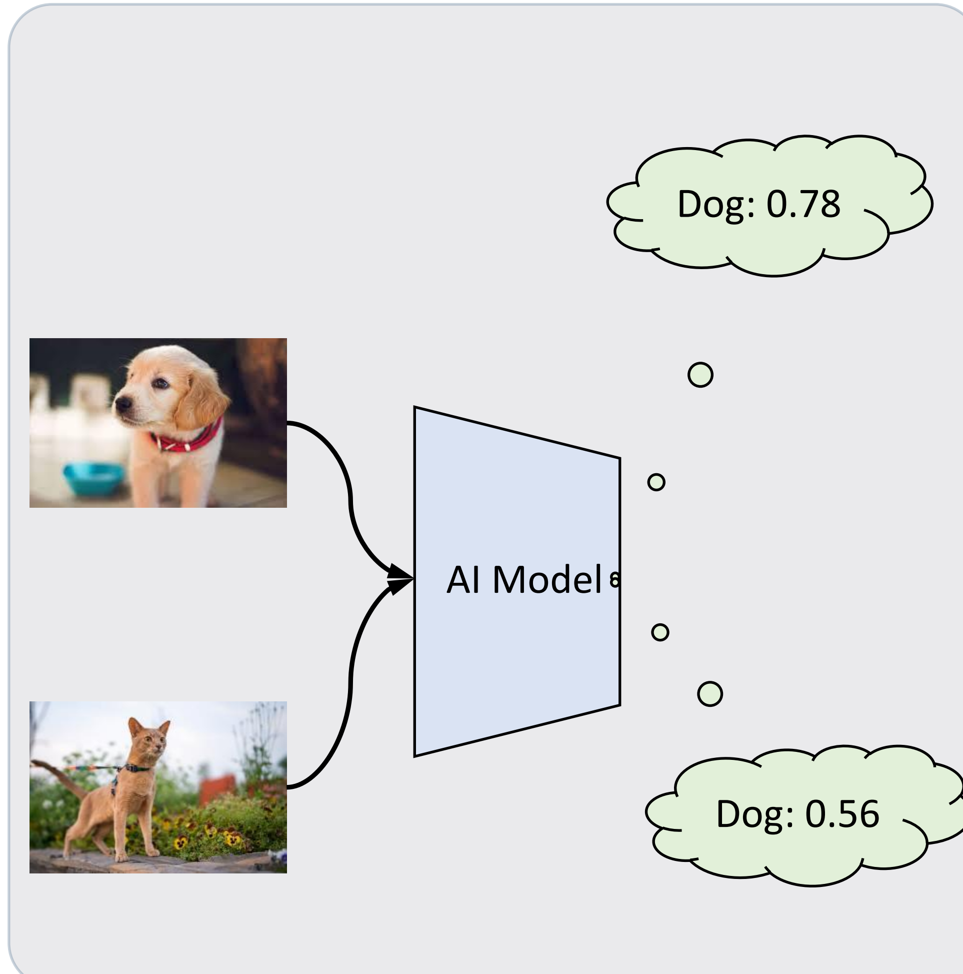
Machine Learning

Definition

Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Semuel, 1950s





Why do we need Machine Learning?

- Some tasks are easy for humans but hard to describe with **explicit rules**: recognizing a face, understanding speech, judging whether an email is spam
- Machine learning lets computers **learn the rules from data** instead of being hand-coded

	Rule-Based Programming	Machine Learning
How rules are obtained	Written by human experts	Learned from data automatically
When it struggles	Rules are incomplete, or patterns keep changing (e.g., spam)	Requires sufficient, representative data

- **Typical cases where rules fail:** handwritten digits vary in style; spam patterns change over time; speech differs across accents and environments
- **Learning from data** adapts to these variations far better than hand-written rules

Applications of Machine Learning

Computer Vision
Face recognition, medical imaging, autonomous driving

Natural Language Processing
Machine translation, chatbots, text summarization

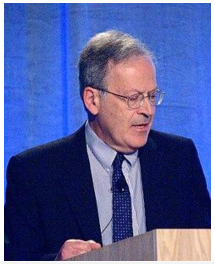
Speech
Speech recognition, voice assistants, speech synthesis

Recommendation
E-commerce, video and music platforms, advertising

Healthcare
Drug discovery, disease diagnosis, protein structure prediction

Finance
Fraud detection, credit scoring, algorithmic trading

Machine Learning



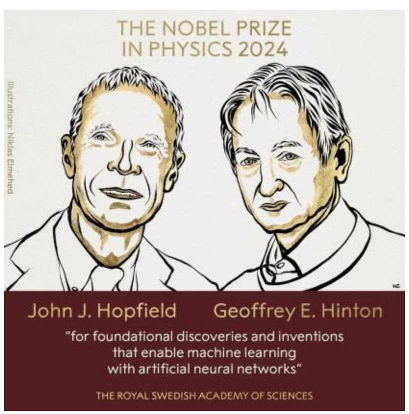
Leslie G. Valiant
(1949 -)
(Harvard Univ.)
Turing Award

2010

Judea Pearl
(1936 -)
(UCLA)
Turing Award

2011

The Pioneers of Deep Learning
2018 ACM A.M. Turing Award



John J. Hopfield
(1933 -)
(Princeton Univ.)

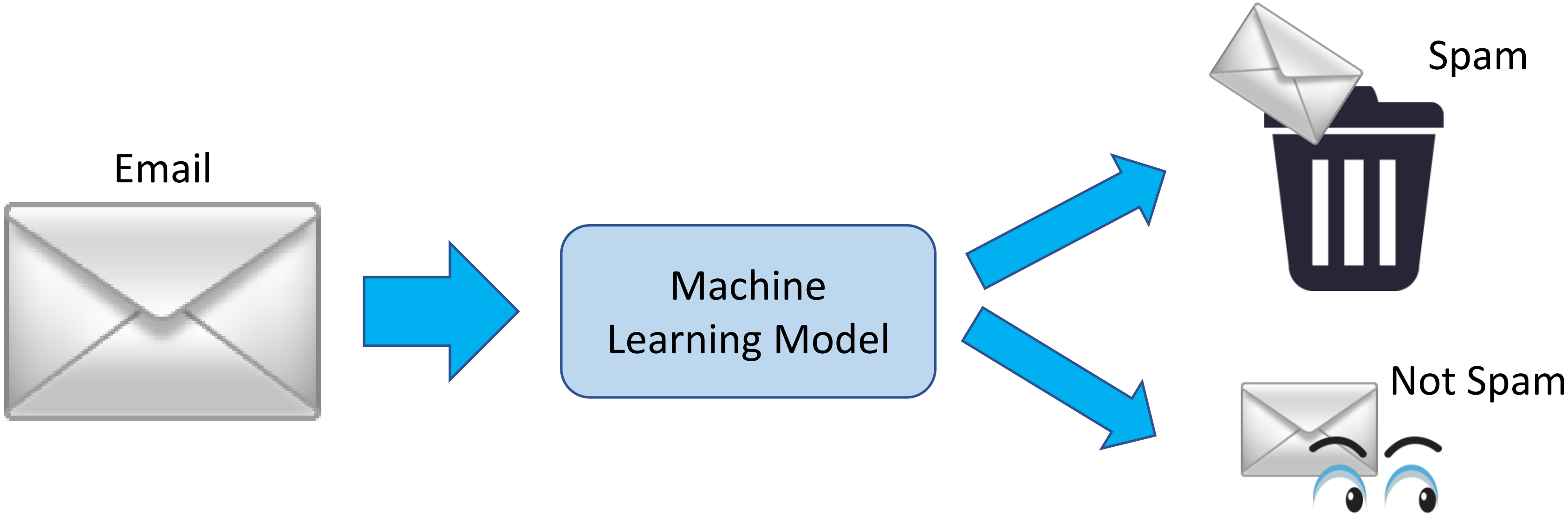
Geoffrey E. Hinton
(1947 -)
(Univ. of Toronto)

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

Applications of Machine Learning

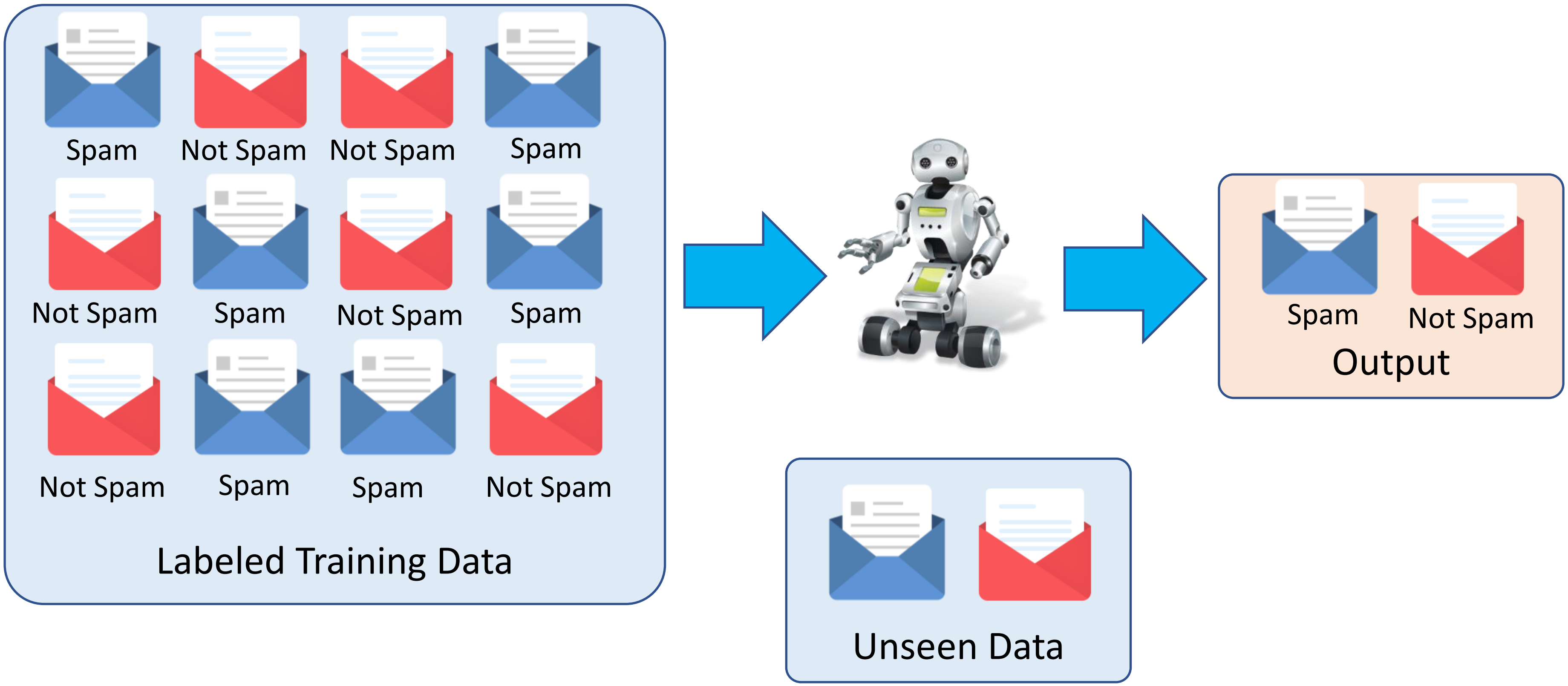
Example: Spam Email

- Decide which emails are spam and which are important.



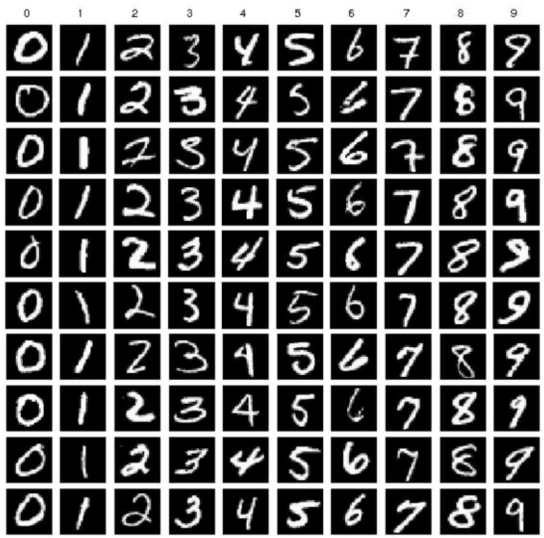
- Use the emails observed so far to learn a prediction rule for future **unseen** data.

Example: Spam Email



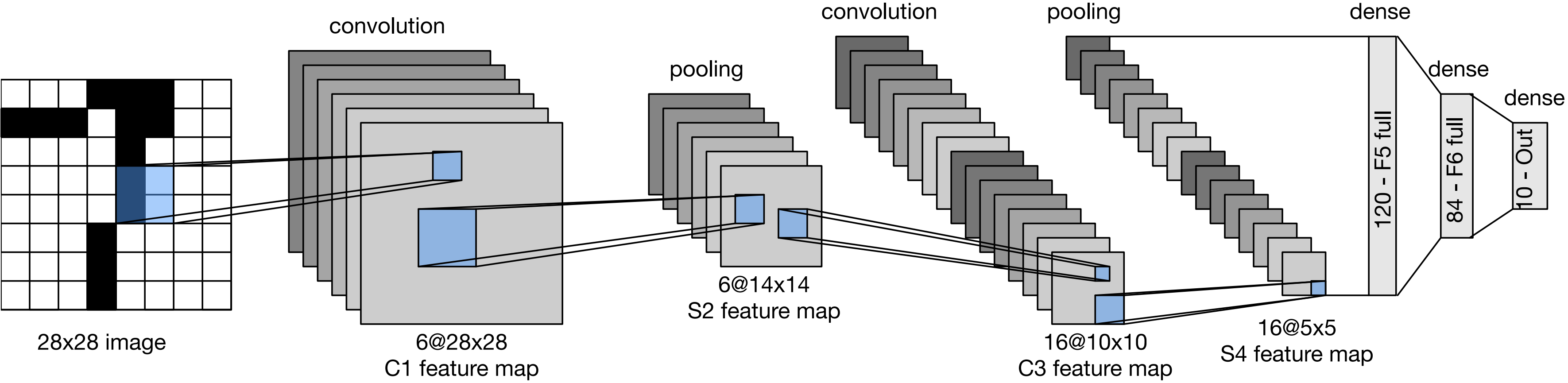
Example: Image Classification

- Handwritten digit recognition (convert hand-written digits to discrete 0,1,2,3,4,5,6,7,8,9)



- Natural Image Classification

Example: Image Classification



LeNet

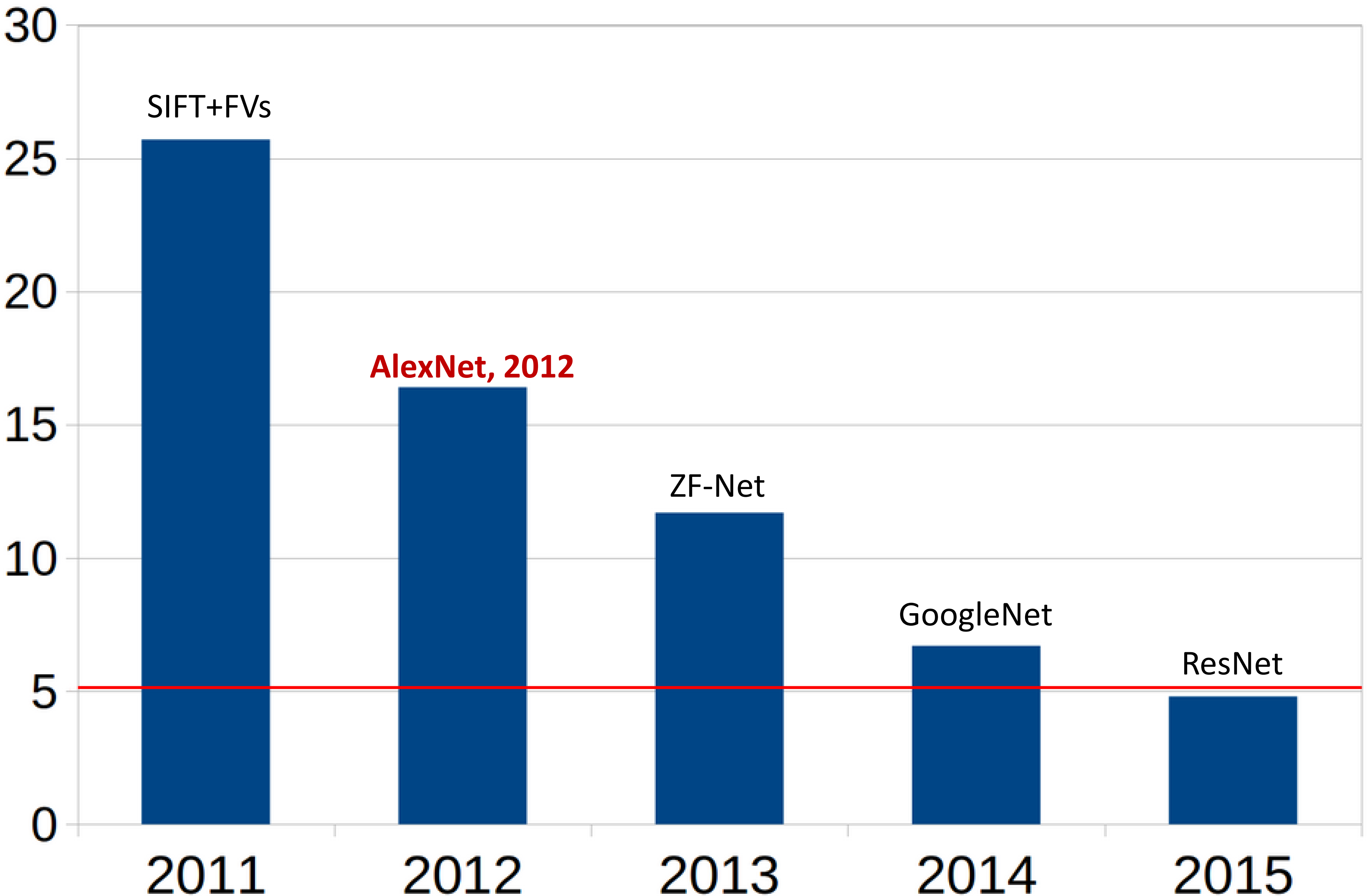
- Achieving 99% accuracy on MNIST (Yann LeCun 1998)
- Elements:
  - convolutions, maxpooling, fully connected layers
  - Activation layer: sigmoid
  - Backpropagation algorithm

Example: Image Classification

ImageNet

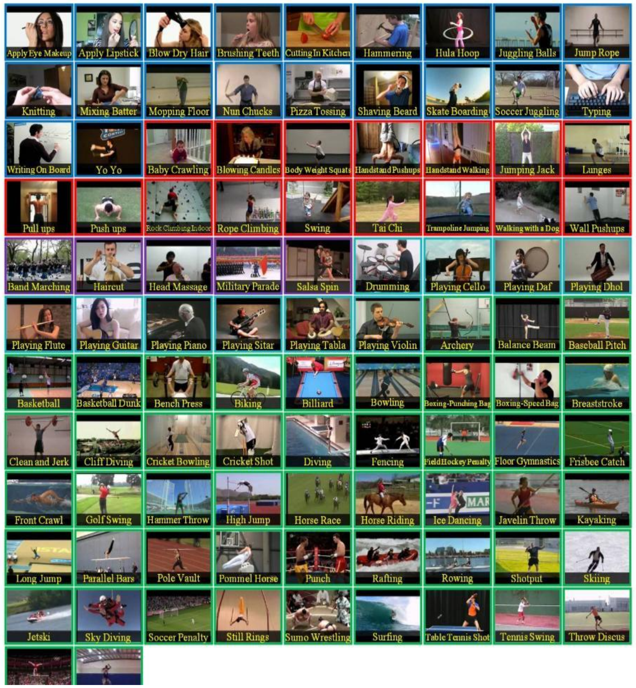
- A large scale image dataset collected by Princeton
- 1.2 million images, 1000 categories
- AlexNet: a breakthrough in large-scale image classification

Example: Image Classification



Example: Video Understanding

- **Action recognition:** assign an action label to a video clip (e.g., Playing Cello)
  - Video = image sequence + temporal information (motion across frames)
- **Benchmark datasets**
  - UCF-101: 13,320 clips, 101 action classes
  - Kinetics: 400+ classes, collected from YouTube
- **Evolution of methods**
  - Two-Stream CNN (RGB + optical flow)
  - 3D CNN (C3D, I3D)
  - Video Transformer



UCF-101

Example: Video Understanding

- Why is video understanding **harder** than image classification?

Temporal dynamics

Actions are defined by how frames change over time; a single frame is often ambiguous

Data volume

A 1-minute clip at 30 fps has about 1,800 frames, with heavy redundancy between adjacent frames

Annotation cost

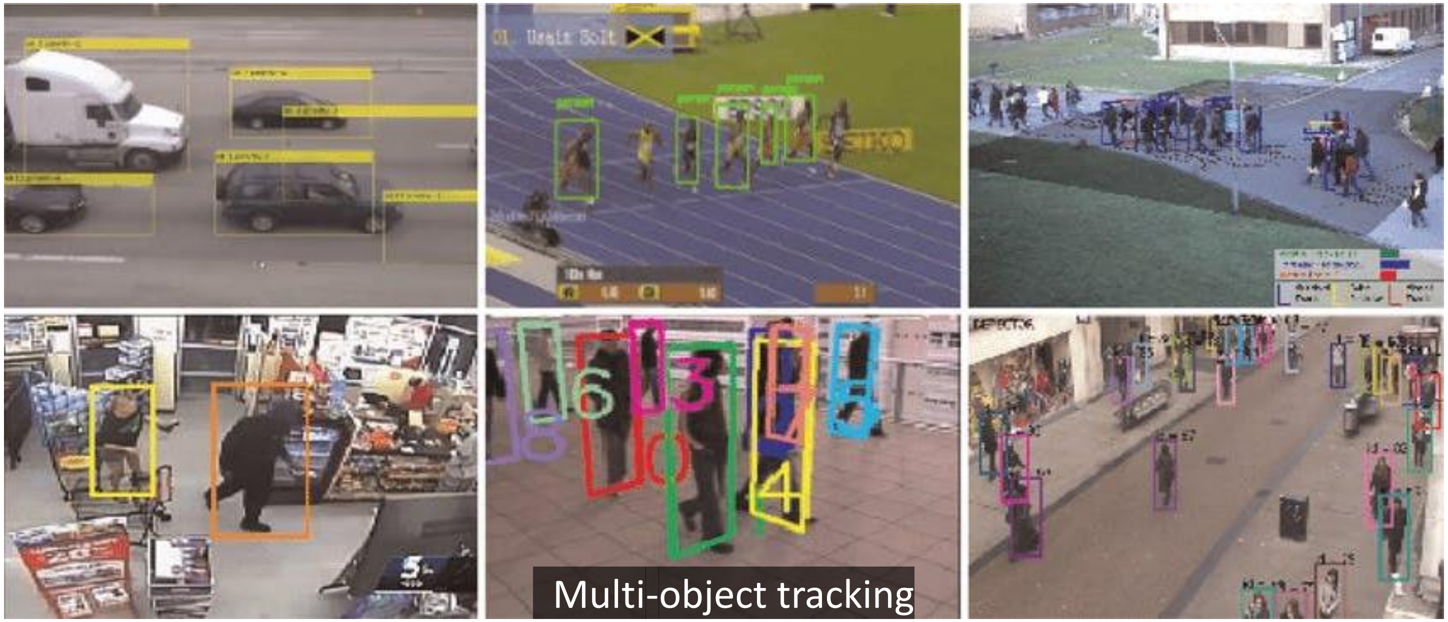
Labeling when and where actions happen in untrimmed videos is slow and expensive

Computation

Processing both space and time makes models (3D CNNs, video Transformers) costly to train

Example: Video Understanding

- **Video object detection:** locate objects with bounding boxes in every frame
- **Multi-object tracking (MOT):** link detections across frames so each object keeps a consistent identity
- **Mainstream paradigm:** tracking-by-detection — detect per frame, then associate over time using motion and appearance cues
- **Applications:** autonomous driving, video surveillance, sports analysis

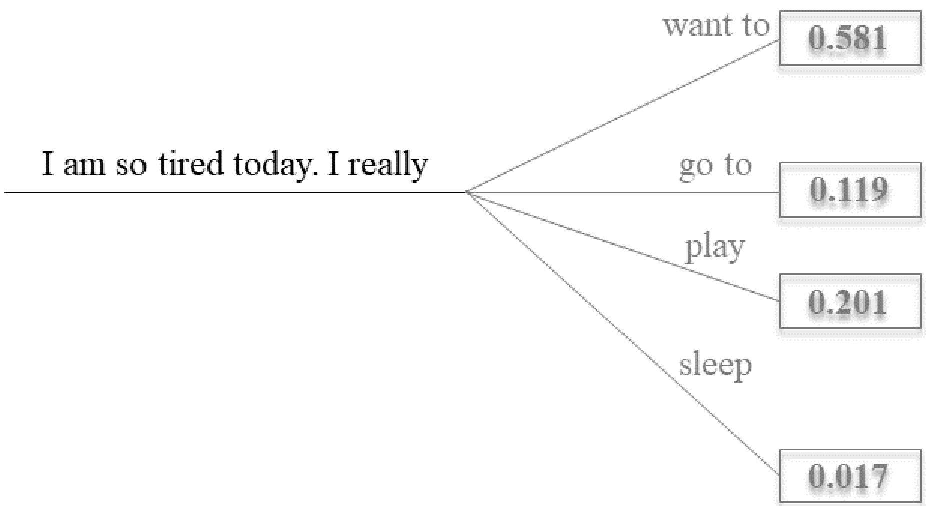


Multi-object tracking

Example: Large Language Models

- LLM: next token prediction task

$$p(\text{next\_token} \mid \text{“I am so tired today ”}; \Theta)$$



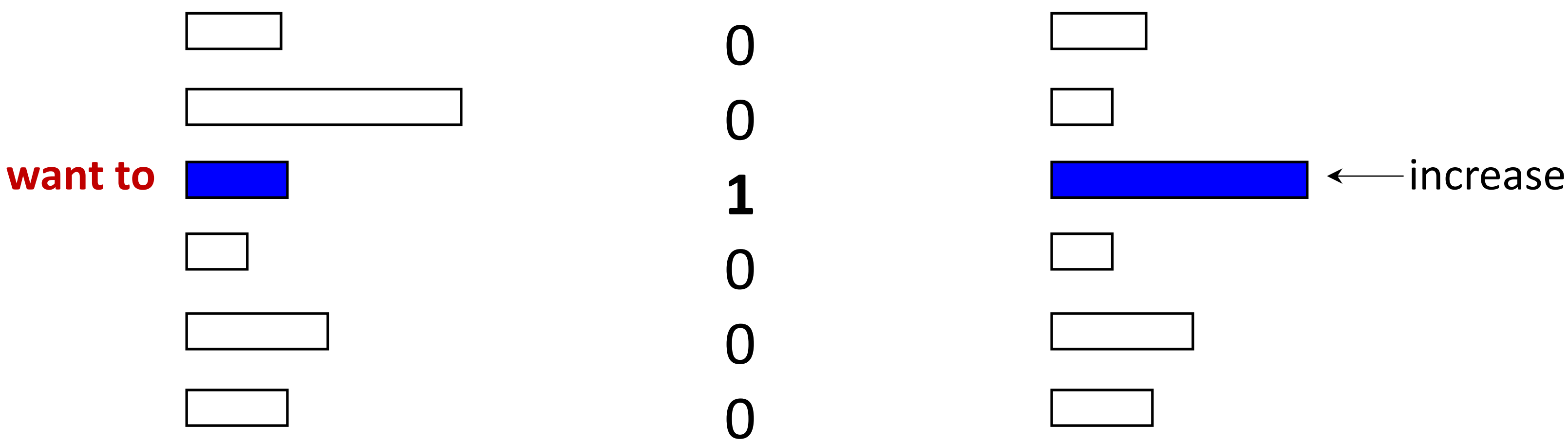
Example: Large Language Models

- How to train LLM models?

We want the model to predict this

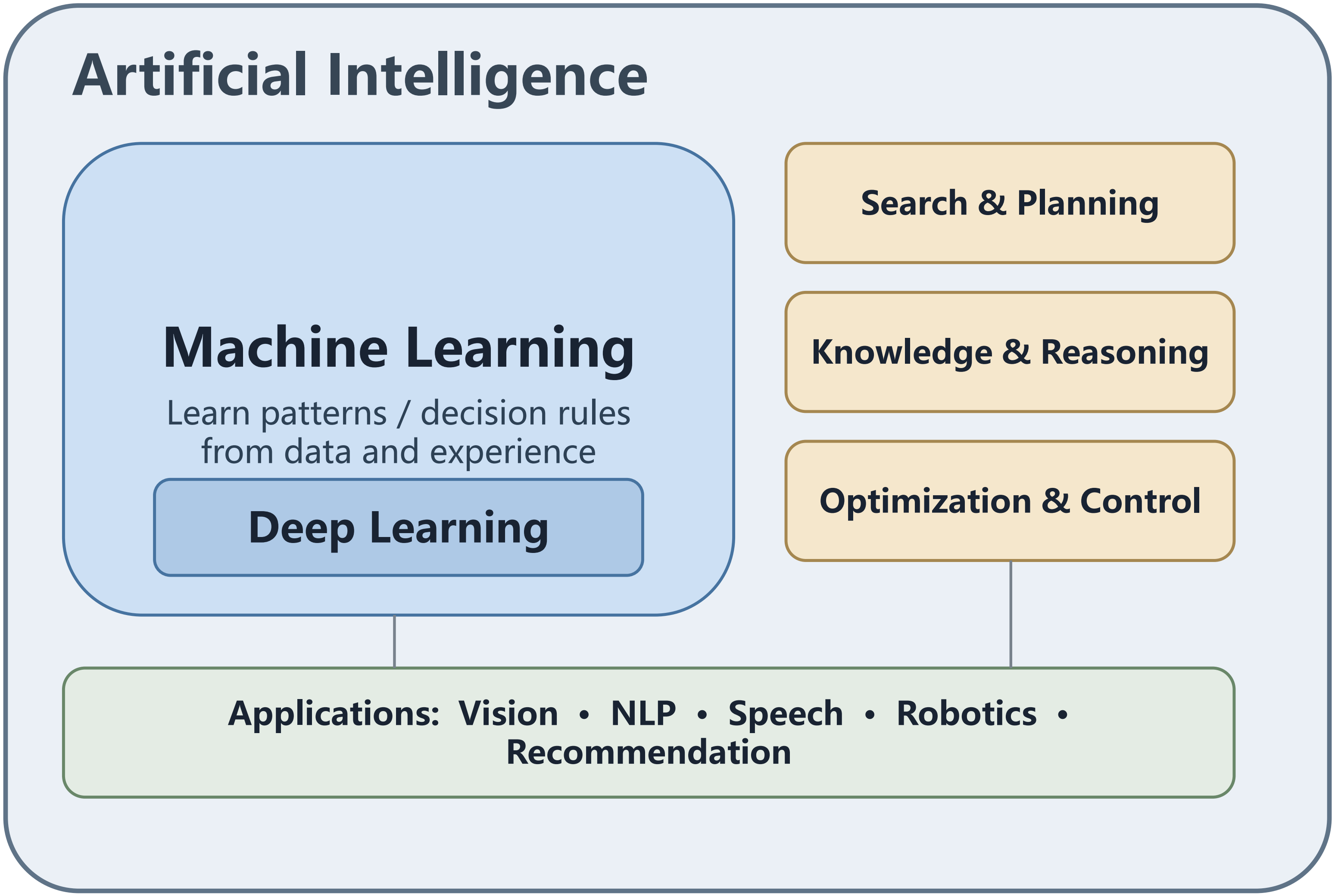
Training example: I really **want to** ……

Model prediction: $P(* \mid \text{“I really”})$ Target Loss: $-\log(P(\text{“want to”}))$



The Role of Machine Learning

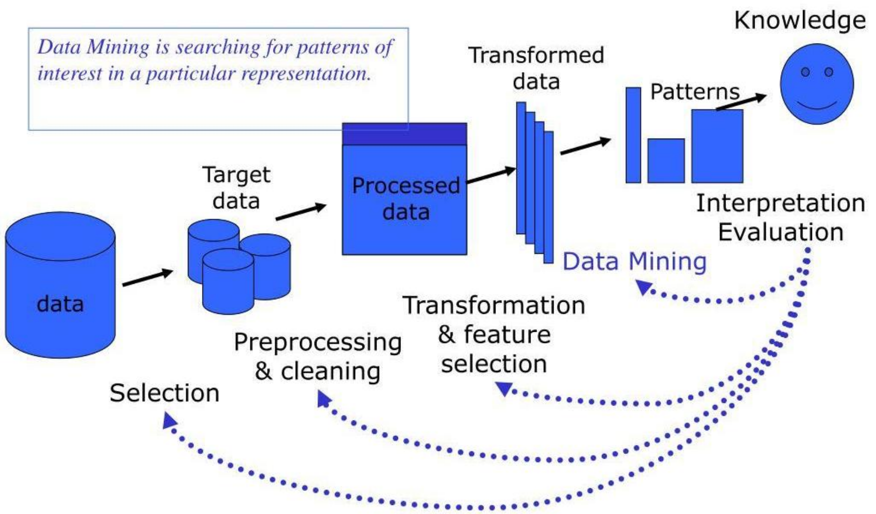
Machine Learning in Artificial Intelligence



Machine Learning & Data Mining

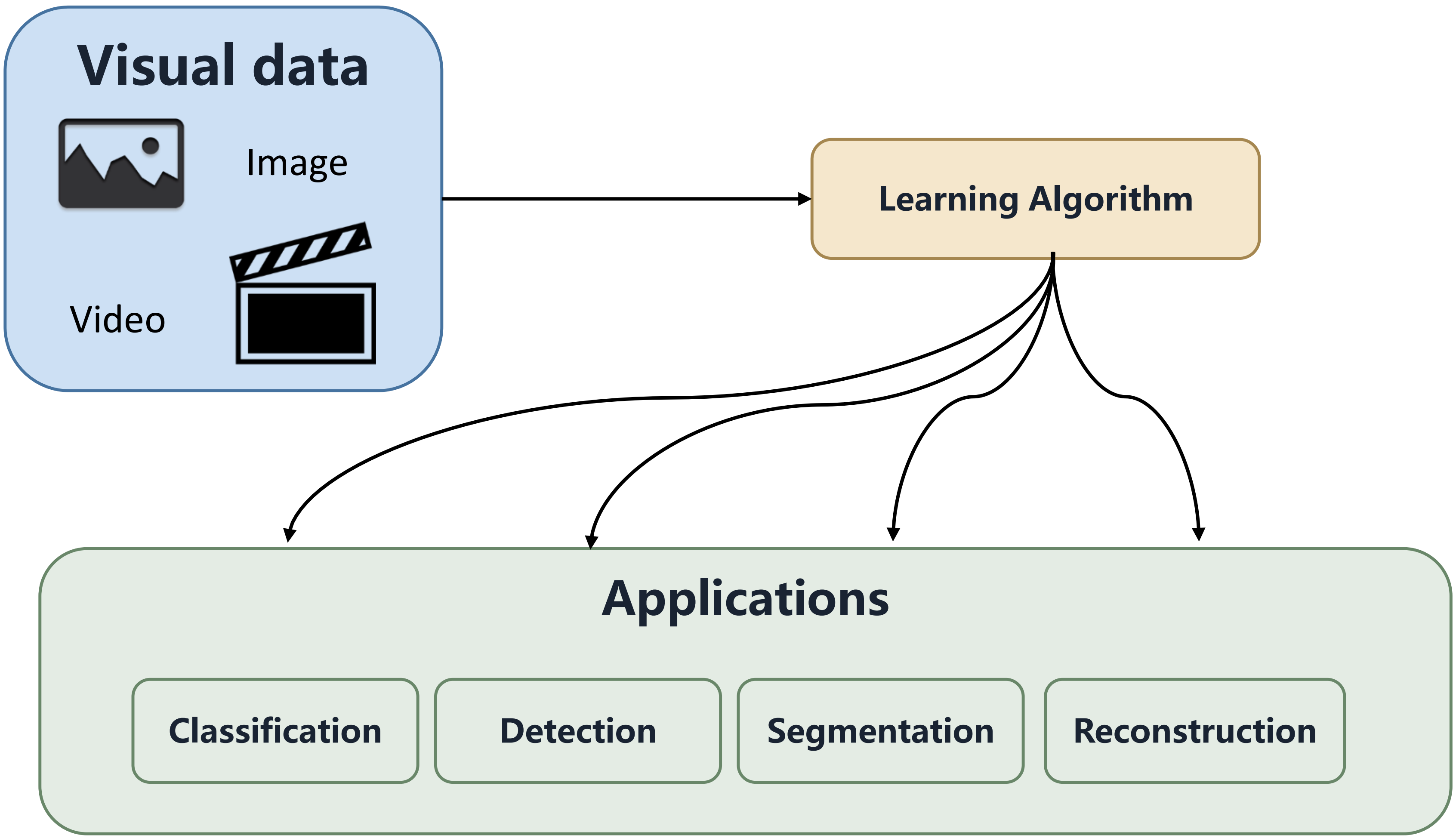
	Machine Learning	Data Mining
Focus	Learning methods	Knowledge discovery
Orientation	Method-oriented	Application-oriented
Goal	Learn patterns/prediction rules	Discover useful patterns/knowledge
Role	A key technical foundation	An important application area

- Machine learning provides powerful tools for data mining, while data mining provides important application scenarios for machine learning.



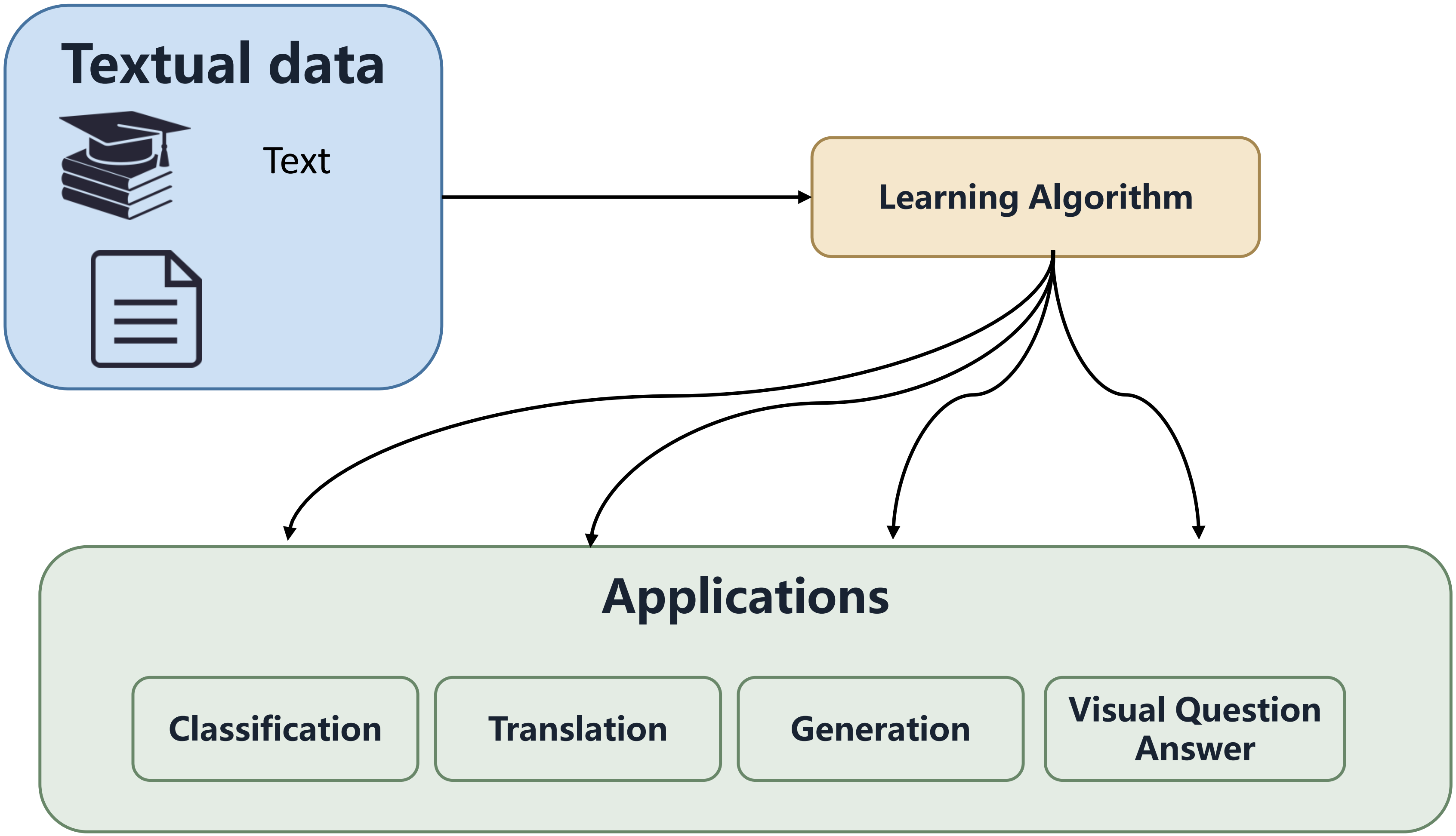
Machine Learning & Computer Vision

- Computer Vision focuses on **understanding visual information**, while Machine Learning provides methods to learn from data.



Machine Learning & Nature Language Processing

- Natural Language Processing focuses on **understanding and generating language**, while Machine Learning provides powerful methods to learn from language data.

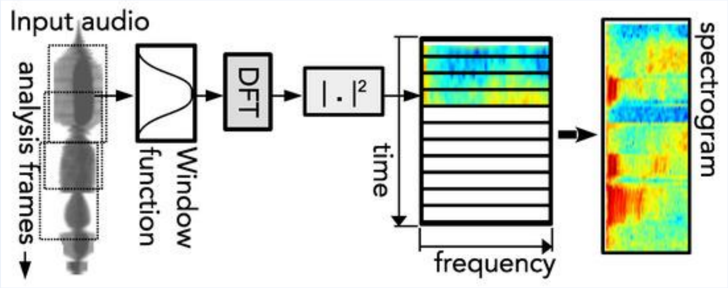


Machine Learning & Speech / Robotics

Speech

Machine learning turns acoustic signals into usable information

Tasks: speech recognition (ASR), text-to-speech (TTS), speaker identification, keyword spotting

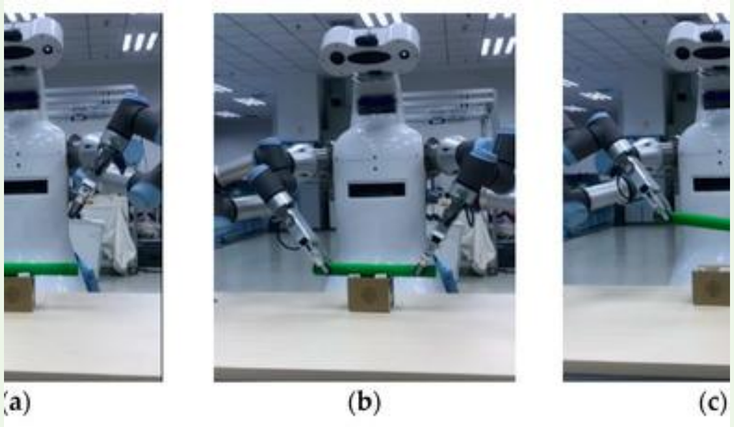


Robotics

Learning replaces hand-designed controllers for complex tasks

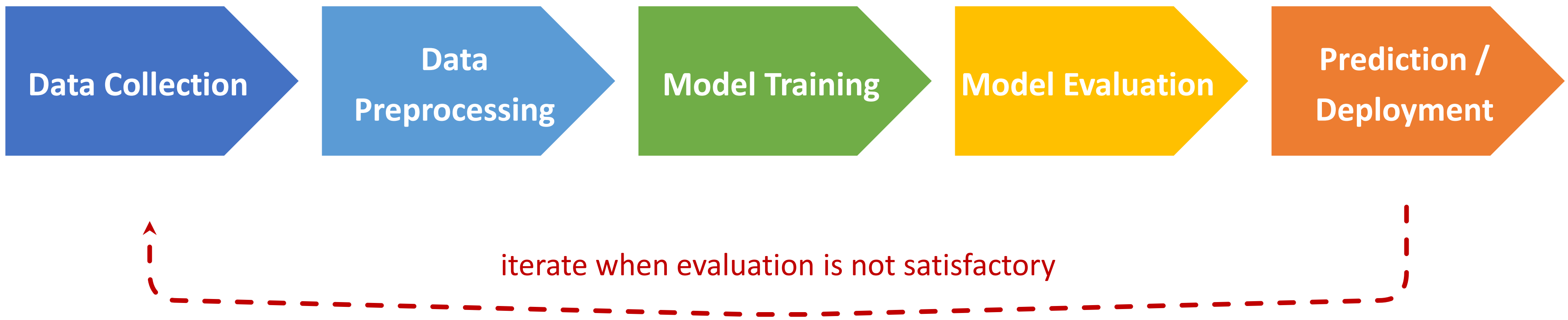
Methods: reinforcement learning, imitation learning, sim-to-real transfer

Tasks: grasping, locomotion, navigation



Machine Learning Terminology

Machine Learning Process



- A typical machine learning pipeline: collect data → preprocess → train a model → evaluate → deploy
- **Data collection & preprocessing**: gather raw data, clean it, and extract useful features
- **Training**: learn model parameters from training data by minimizing a loss function
- **Evaluation**: measure performance on held-out data
- The process is **iterative**: evaluation results guide the adjustment of data, features, and models

Data

- Data is the foundation of machine learning — **garbage in, garbage out**
- Each **sample** is described by features (attributes); the label is the prediction target
- Both quality and quantity matter: errors, noise, bias, and coverage of the data directly limit model performance

Modality	Example Datasets	Typical Tasks
Image	MNIST, ImageNet	Classification, detection
Text	WikiText	Question answering, language modeling
Video	UCF-101, Kinetics	Action recognition, tracking
Audio	LibriSpeech	Speech recognition

Model, Loss, Optimization

Model (hypothesis): a parameterized function that maps input to prediction, e.g.,

$y = f(x;\theta)$

θ : parameters learned from data; the choice of f defines the hypothesis space

Loss function $\ell(\hat{y}, y)$: measures how wrong a prediction is — e.g., cross-entropy for classification, squared error for regression

Training objective: $\theta^* = \mathrm{argmin}_\theta \frac{1}{N} \sum_{i=1}^{N} \ell(f(x_i;\theta); y_i)$

Optimization: update parameters to reduce the loss, usually by gradient descent and its variants (SGD, Adam)

$\theta_t = \theta_{t-1} - \eta \nabla_\theta \ell(\cdot)$, where η is the learning rate; gradients in deep networks are computed by **backpropagation**

Training & Testing

- Split available data into training / validation / test sets
- **Training set**: used to learn model parameters
- **Validation set**: used to tune hyperparameters and select the model
- **Test set**: kept **unseen** during training and tuning; used to estimate performance on future data
- When data is limited: use **cross-validation** — rotate which fold serves as validation
- **Never tune on the test set** — otherwise the reported performance is overly optimistic

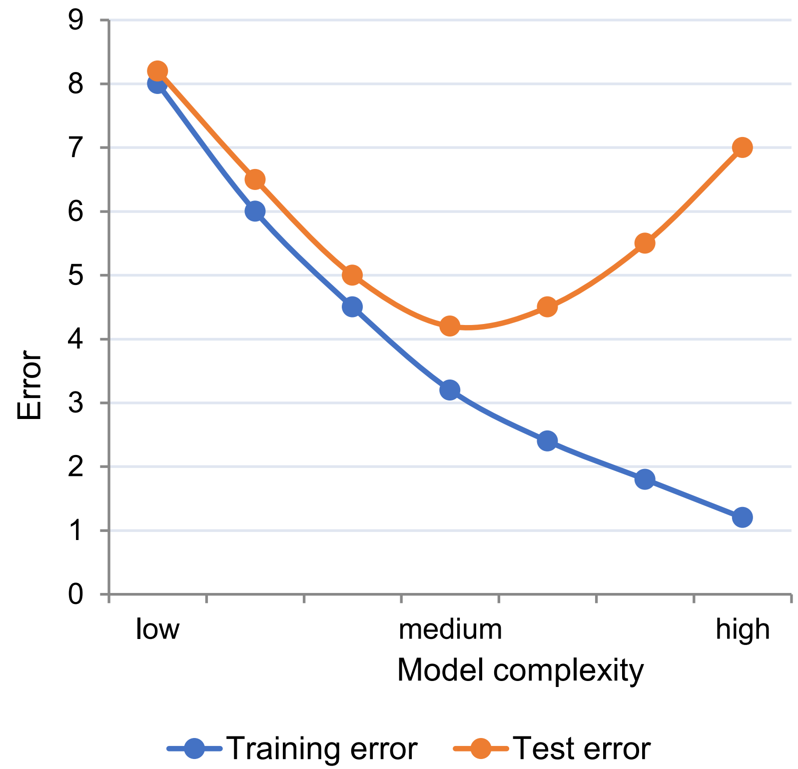
Supervised/Semi-Supervised/Unsupervised

Supervised	Unsupervised	Semi-Supervised
Data: fully labeled (x, y)	**Data**: unlabeled x only	**Data**: a few labeled samples + many unlabeled ones
Tasks: classification, regression	**Tasks**: clustering, dimensionality reduction	**Goal**: supervised-quality performance with less labeling cost
Examples: spam filtering, image classification, house price prediction	**Examples**: customer grouping, topic discovery	**Example**: medical image analysis (labeling is expensive)

- **Self-supervised learning**: supervision signals are generated from the data itself
- Example: next-token prediction for large language models — no manual labels needed

The Goal of Learning

- The goal is to perform well on **unseen data** — this is called **generalization**
- Memorizing training data is not enough (a spam filter must work for **future** emails)
- **Underfitting**: model too simple — poor on both training and test data
- **Overfitting**: model too complex — good on training data, poor on test data
- Occam's razor: prefer the **simplest** model that explains the data well



Training error vs. test error

Evaluation Metrics

Classification

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}\ , \text{Recall} = \frac{TP}{TP+FN}$$

$\text{F1} = \frac{2PR}{P+R}$ — balances precision and recall

Accuracy alone is misleading on **imbalanced** data (e.g., spam)

Confusion matrix	Pred. +	Pred. –
Actual +	TP	FN
Actual –	FP	TN

Regression

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|\ , \text{MSE} = \frac{1}{N}\sum_{i=1}^{N}||y_i - \hat{y}_i||^2,\ \text{RMSE} = \sqrt{\text{MSE}}$$

Ranking / retrieval: Precision@K, Recall@K, AUC

Choose the metric that matches the real cost of errors in the application

Tracking the Frontiers of Machine Learning

Top Journals

Journal	Full Name	Field
Nature MI	Nature Machine Intelligence	Machine Learning, AI
TPAMI	IEEE Transactions on Pattern Analysis and Machine Intelligence	Machine Learning
JMLR	Journal of Machine Learning Research	Machine Learning
TKDE	IEEE Transactions on Knowledge and Data Engineering	Data Mining
IJCV	International Journal of Computer Vision	Computer Vision
AI	Artificial Intelligence (Elsevier)	Artificial Intelligence

Top Conferences

Conference	Full Name	Field
NeurIPS	Conference on Neural Information Processing Systems	Machine Learning
ICML	International Conference on Machine Learning	Machine Learning
ICLR	International Conference on Learning Representations	Deep Learning
CVPR / ICCV / ECCV	IEEE/CVF, Int'l, and European Conferences on Computer Vision	Computer Vision
ACL / EMNLP	Annual Meeting of the ACL / Empirical Methods in NLP	Natural Language Processing
AAAI / IJCAI	AAAI Conference / Int'l Joint Conference on AI	Artificial Intelligence
KDD	ACM SIGKDD Conference on Knowledge Discovery and Data Mining	Data Mining

How to Track the Latest Research

Preprints: arXiv

cs.LG, cs.CV, cs.CL — new results appear here months before formal publication

Literature: Google Scholar

Set keyword and author alerts; follow citations of key papers forward and backward

Benchmarks: Papers with Code

State-of-the-art results on standard datasets, linked to open-source implementations

Community

Conference talks and tutorials, lab blogs, survey papers for a structured overview of a subfield

Current Trends in Machine Learning

Large Language Models & Agents

Reasoning, tool use, multi-step task solving

Multimodal Learning

Joint understanding of text, image, audio, and video

Generative AI

Image and video generation (diffusion models, video generators)

AI for Science

Protein structure, weather forecasting, materials discovery

Efficient Learning

Small models, quantization, on-device deployment

Alignment & Safety

RLHF, interpretability, responsible deployment

Summary

- Machine learning gives computers the ability to learn from data without being explicitly programmed
- The same learning paradigm drives progress across **text, image, video, speech**, and beyond
- Machine learning is a core component of AI and deeply connected to data mining, computer vision, and NLP
- A standard pipeline: **data → training → evaluation → deployment**, with generalization as the ultimate goal
- Follow top **journals and conferences** to track the frontiers of the field

Thanks