

Introduction to Machine Learning

AI · ML · DEEP LEARNING · GENERATIVE AI · ML TYPES · ML PIPELINE · TOOLS · CHALLENGES

STUDY & REVIEW

01 MUST KNOW

if you remember five things, remember these

- 01 The nesting: AI $\supset$ ML $\supset$ deep learning $\supset$ generative AI.** AI is the broad branch of computer science for building machines that do tasks humans typically do; **machine learning** is the subset that learns a task by *inference* from data instead of hand-coded instructions; **deep learning** is the ML subdomain built on layered neural networks; **generative AI** is the slice of deep learning that creates new content from pretrained foundation models.
- 02 Learn from data, not rules:** instead of writing a huge set of if/else rules (a spam filter), you **train a model** on labeled examples; it learns the patterns that predict the answer, then predicts on input it has never seen. That is the essence of ML — *a task performed by inference instead of instructions*.
- 03 Three types of ML:** **supervised** learns from *labeled* data $\rightarrow$ classification and regression; **unsupervised** finds structure in *unlabeled* data, creating its own labels $\rightarrow$ clustering; **reinforcement** learns by *trial and error* from rewards and penalties in an environment. **Most business problems are supervised.**
- 04 The pipeline is iterative:** business problem $\rightarrow$ data preparation $\rightarrow$ feature engineering $\rightarrow$ model training $\rightarrow$ evaluation & tuning $\rightarrow$ deployment. The middle loops — *data processing, training, evaluation* repeat until the model meets the goal. Train on about **80%** of the data and hold back the rest (~**10–20%**) to test.
- 05 Match the tool to the effort:** the Amazon ML stack has **three layers** — low-level **frameworks** (TensorFlow, PyTorch) for full control, **Amazon SageMaker** for the managed lifecycle, and **managed services** (Rekognition, Polly, Comprehend, Lex...) that need *no ML experience*. And not every problem needs ML — reach for it with large datasets, many variables, and no clear rules.

02 KEY TERMS

TERM	DEFINITION
Artificial intelligence (AI)	The broad branch of computer science focused on building machines that can perform tasks a human would typically perform
Machine learning (ML)	A subset of AI: the study of algorithms and statistical models that perform a task using inference instead of explicit instructions
Deep learning	A subdomain of ML built on layered artificial neural networks inspired by the brain; practitioners focus on network architectures
Artificial neural network (ANN)	Layers of artificial neurons — input, hidden, and output — joined by weighted connections; each neuron takes one or more inputs and fires a single output
Generative AI	A subset of deep learning that creates new content — text, images, video, music — powered by pretrained foundation models
Foundation model (FM)	A very large ML model pretrained on vast data; one FM can handle many tasks (generation, summarization) without further training
Supervised learning	Learning from already-labeled data (known inputs and outputs); its problems are classification and regression
Unsupervised learning	Learning with no labels provided — the machine must uncover and create the labels itself; clustering is a common subcategory
Reinforcement learning	An agent learns through trial and error by interacting with an environment, guided by rewards and penalties
Feature engineering	Selecting or creating the features (columns of data) used to train a model to estimate the target value

TERM	DEFINITION
Hyperparameters	Values you set before training that alter how the learning algorithm works
Overfitting / underfitting	Overfitting: good on training data but poor on new data (it memorized). Underfitting: poor even on training data (it can't capture the X→Y relationship)

03 THE ML PIPELINE

the middle stages loop

01 Business problem — formulate it as an ML problem → **02 Data preparation** — extract, clean, reconcile sources → **03 Feature engineering** — select and encode features → **04 Model training** — ~80% of data, set hyperparameters → **05 Evaluation & tuning** — test data → metrics → adjust → **06 Deployment** — serve predictions in production

04 THE THREE TYPES OF MACHINE LEARNING

match the type to the data you have

TYPE	TRAINING DATA	PROBLEM TYPES	EXAMPLE USES
Supervised	Already labeled (known in/out)	Classification, regression	Fraud detection, image recognition, weather & sales forecasting
Unsupervised	Unlabeled — labels discovered	Clustering, structure discovery	Customer segmentation, targeted marketing, gene sequencing
Reinforcement	Rewards & penalties from an environment	Trial-and-error control	Game AI, robotics, self-driving cars, AWS DeepRacer

05 CLASSICAL PROGRAMMING OR MACHINE LEARNING?

sometimes basic programming wins

CLASSICAL PROGRAMMING FITS WHEN <i>the logic is knowable</i> — Business logic and procedures are clear — Rules can be written out as if/else statements — The number of variables is limited — Basic programming solves it just as well	REACH FOR ML WHEN YOU HAVE <i>and the support to sustain it</i> — Large datasets and a large number of variables — No clear procedure to reach the answer — Existing machine learning expertise — Infrastructure in place and management support
--	---

06 THE AMAZON ML STACK — THREE LAYERS

know a service's layer

LAYER	WHAT IT IS	EXAMPLES
ML frameworks & infrastructure	Low-level libraries and ML-tuned compute for full control	TensorFlow, PyTorch, Keras, MXNet; EC2 P3, C5/C5n
Amazon SageMaker	Managed environment for the full ML lifecycle	Ground Truth, Notebook, Training, Inference, Marketplace
Managed AI services	Pre-trained services needing no ML experience	Rekognition, Textract, Polly, Transcribe, Comprehend, Translate, Lex, Personalize

- × “Deep learning is separate from machine learning.” — Deep learning is a **subdomain of ML**, which is a subset of AI; generative AI is a subset of **deep learning**. They nest — they don't compete.
- × “A generative AI model must be retrained for each new task.” — Unlike single-task traditional models, one pretrained **foundation model** handles many tasks — text generation, summarization — with no further training.
- × “Supervised learning only does classification.” — Supervised splits into **classification** (discrete categories) *and* **regression** (a continuous value, like a predicted stock price).
- × “Binary and multiclass are different learning types.” — Both are **classification** inside supervised learning: binary chooses between two categories, multiclass among three or more.
- × “In unsupervised learning the labels are given to find.” — No labels are provided; the **machine must uncover and create the labels itself**, which is why you use it when the patterns are unknown.
- × “Overfitting means the model does badly on the training data.” — That's **underfitting**. Overfitting is the opposite: great on training data, poor on new data because the model *memorized* instead of generalizing.
- × “Reinforcement learning needs labeled training data.” — It learns by **trial and error** from rewards and penalties — best when the desired outcome is known but the path to it is not.
- × “You need data-science expertise to add AI to an app.” — AWS **managed services** (Rekognition, Polly, Transcribe, Comprehend, Lex, Personalize) require *no ML experience*; you supply data and call an API.

08 SELF-QUIZ

answers are upside-down below

- Q1** In the nesting of AI concepts, which field is the innermost subset — a type of deep learning?
- Q2** ML is described as performing a task using what, rather than following explicit instructions?
- Q3** Which type of machine learning trains on data that is already labeled?
- Q4** Supervised learning problems fall into which two broad categories?
- Q5** Which type of machine learning must uncover and create the labels itself, with none provided?
- Q6** In reinforcement learning, what term names the entity that drives the learning by interacting with the environment?
- Q7** Roughly what share of the data does the module suggest holding back to test the model (training on about 80%)?
- Q8** What are the values called that you set to alter how a training algorithm works?
- Q9** A model performs well on training data but poorly on new data — what is this problem called?
- Q10** Which AWS service hosts Jupyter notebooks and provides Ground Truth, Training, and Inference capabilities?

ANSWER KEY (rotate the page)

Q1 generative AI Q2 inference Q3 supervised learning Q4 classification and regression Q5 unsupervised learning (clustering) Q6 the agent Q7 10–20% Q8 hyperparameters Q9 overfitting Q10 Amazon SageMaker