Artificial Intelligence Technical Terms

Table of Contents

1. Artificial Intelligence

Term	Technical Meaning
Intelligent Agent	Entity that perceives and acts in an environment.
Knowledge Representation	Encoding information for reasoning.
Search Algorithms	Exploring state spaces to find solutions.
Heuristic Function	Approximate scoring function guiding search.
Constraint Satisfaction	Solving problems with variables + constraints.
Utility Function	Quantifies desirability of outcomes.
Decision Tree	Rule-based hierarchical decision model.
Bayesian Network	Probabilistic graph-based dependency model.
Markov Decision Process	Framework for sequential decision making.
Reinforcement Learning	Learning through rewards and actions.
Expert System	Rule-based inference system.
Inference Engine	Component performing logical reasoning.
Fuzzy Logic	Reasoning under uncertainty and vagueness.
Planning Algorithms	Multi-step strategy generation.
Cognitive Architecture	Models simulating human thought processes.
Symbolic AI	Logic-based deterministic reasoning.
Subsymbolic AI	ML-based statistical reasoning.
Autonomous Systems	Self-governing intelligent systems.
Multi-Agent System	Multiple interacting intelligent agents.
Explainable AI (XAI)	Techniques to interpret model decisions.

2. Data Science

Term	Technical Meaning
ETL Pipeline	Extract, transform, load data process.
Data Cleaning	Removing errors, duplicates, inconsistencies.

Feature Engineering	Creating features from raw data.
Statistical Inference	Deriving conclusions using statistical methods.
Hypothesis Testing	Validating claims using significance tests.
Correlation Analysis	Measuring feature relationships.
Data Normalization	Standardizing feature scales.
Outlier Detection	Identifying anomalous values.
Dimensionality Reduction	Reducing feature count (PCA/UMAP).
Sampling Techniques	Drawing representative data subsets.
Time Series Analysis	Forecasting based on temporal data.
Data Visualization	Graphical representation of insights.
Regression Modeling	Predicting continuous values.
Classification	Predicting categorical outputs.
Clustering	Unsupervised grouping of similar data points.
AB Testing	Controlled experimentation.
Data Pipeline Orchestration	Managed workflow execution.
Feature Store	Centralized repository for ML features.
Data Governance	Policies ensure quality & security.
BI Dashboards	Visual analytics for business decisions.

3. Machine Learning

Term	Technical Meaning
Supervised Learning	Train on labeled data.
Unsupervised Learning	Discover patterns without labels.
Semi-Supervised Learning	Mix of labeled and unlabeled data.
Reinforcement Learning	Reward-based optimization.
Overfitting	Memorizing data instead of generalizing.
Underfitting	Model too simple to capture relationships.
Gradient Descent	Optimization method minimizing loss.

Loss Function	The objective model tries to minimize.
Regularization	Techniques to reduce overfitting.
Hyperparameter Tuning	Configuring algorithm settings.
Cross-Validation	Robust model validation method.
Decision Trees	Rule-based learning algorithm.
Ensemble Methods	Combining multiple models.
Boosting	Ensemble with sequential learning.
Bagging	Ensemble with parallel sampling.
Feature Importance	Ranking predictive power of features.
ROC Curve	Classification performance metric.
Confusion Matrix	Breakdown of predictions vs truth.
ML Pipeline	Full workflow from data → deployment.
Model Drift	Model performance decrease over time.

---

4. Deep Learning

Term	Technical Meaning
Neural Network	Multi-layer computational graph.
Backpropagation	Gradient computation method.
Activation Function	Non-linearity enables expressiveness.
CNN	Convolutional networks for image tasks.
RNN	Sequence modeling networks.
LSTM	RNN variant with long-term memory.
GRU	Lightweight LSTM variant.
Dropout	Regularization by random neuron removal.
BatchNorm	Normalization to stabilize training.
Transformer	Attention-based neural architecture.
Attention Mechanism	Focus on important parts of input.

Embedding Layer	Dense vector representation.
Epoch	Single pass over training data.
Batch Size	Number of samples per update.
Learned Representation	Hierarchical features in deep nets.
Autoencoder	Self-supervised encoder–decoder.
Skip Connections	Bypass connections improving gradients.
Model Quantization	Lower precision for faster inference.
Weight Initialization	Setting initial parameters.
Learning Rate Scheduling	Adaptive update of LR.

5. NLP (Natural Language Processing)

Term	Meaning
Tokenization	Breaking text into tokens.
Lemmatization	Mapping to base dictionary form.
POS Tagging	Assign grammatical labels.
NER	Detect named entities.
Dependency Parsing	Sentence structure graph.
Constituency Parsing	Phrase tree representation.
Text Classification	Assign category labels.
Sentiment Analysis	Detect emotional tone.
Machine Translation	Language-to-language conversion.
Document Embedding	Vector representation of text.
Text Summarization	Shortened version of text.
Question Answering	Extract/generate answers.
Language Modeling	Predict next word probability.
TF-IDF	Keyword-weighting technique.
BM25	Search ranking algorithm.
Word2Vec	Early embedding model.

GloVe	Global co-occurrence embeddings.
BPE	Subword tokenization algorithm.
Seq2Seq	Encoder-decoder architecture.
Co-reference Resolution	Link mentions the same entity.

6. Large Language Models (LLMs)

Term	Meaning
Transformer Decoder	Autoregressive architecture.
KV Cache	Speed-up memory for tokens.
Context Window	Max token input capacity.
Token Probability Distribution	Softmax output.
Temperature	Controls randomness.
Top-K Sampling	Sample from top K tokens.
Top-P Sampling	Nucleus sampling.
Beam Search	Parallel generation paths.
Speculative Decoding	Fast draft+verify decoding.
Function Calling	Structured JSON tool calls.
Logit Bias	Modify token probabilities.
System Prompt	High-priority instruction.
Hidden States	Intermediate layer outputs.
Embedding Space	Vector semantic space.
Instruction Tuning	Training for instruction following.
RLHF	Reward-based fine-tuning.
Chain-of-Thought	Step-by-step reasoning.
Attention Heads	Parallel attention mechanisms.
Safety Guardrails	Restrictions for safe outputs.
Quantization	Reduce precision for inference.

7. Generative AI

Term	Meaning
Generative Model	Produces new samples from learned distribution.
Latent Space	Compressed representation.
Diffusion Model	Noise → denoising generation.
VAE	Probabilistic encoder–decoder.
GAN	Generator vs discriminator.
Sampling Strategy	How outputs are generated.
Classifier-Free Guidance	Prompt strength control.
Conditioning	Guiding generation with input prompt.
Attention Mask	Controls visible tokens.
Prompt Embedding	Vector representation of prompt.
Image-to-Image	Modify images with prompts.
Style Transfer	Apply artistic properties.
Reranking	Select the best generation.
Noise Schedule	Diffusion noise curve.
ControlNet	External guidance network.
Textual Inversion	Learn custom concept embeddings.
LoRA (GenAI)	Low-rank fine-tuning.
Sampling Steps	Iterations in diffusion.
CFG Scale	Degree of prompt influence.
Inference Pipeline	Ordered stages of generation.

8. AI Agents

Term	Meaning
Agent Loop	Reason → Act → Observe.
Tool Invocation	Calling external functions.

Observation	Data returned from a tool.
Agent Memory	Storage of prior knowledge.
Planner	Generates multi-step strategy.
Executor	Performs tool actions.
Critic	Evaluates agent output.
ReAct	Reason+Act alternation.
State Representation	Agent’s internal context.
Policy	Rules for selecting actions.
Skill Registry	Available tools.
Persistent Memory	Long-term knowledge.
Working Memory	Temporary per-task memory.
Agent Swarm	Multi-agent cooperation.
Routing	Selecting an appropriate agent.
Task Decomposition	Breaking tasks into subtasks.
Self-Reflection	Evaluates its own reasoning.
Event Loop	Async agent workflow.
Shutdown Condition	When the agent stops.
Autonomy Level	Degree of self-management.

---

9. Agentic AI

Term	Meaning
Autonomy	Degree of independent behavior.
Multi-Agent Collaboration	Agent-to-agent communication.
Delegation	Assigning tasks to sub-agents.
Long-Horizon Planning	Multi-step strategic reasoning.
Execution Traces	Logs of all agent steps.
Verified Reasoning	Ensuring correctness of steps.

Self-Healing Agent	Fixes errors automatically.
Behavior Cloning	Training agents on demonstration data.
Thought Isolation	Private reasoning not exposed.
Action Graph	Directed graph of actions.
Policy Gradient Tuning	RL for agent policy refinement.
Error Recovery Policy	Steps taken when failure occurs.
Reflection Loop	Critic-guided re-evaluation.
Memory Retrieval Policy	Rules for fetching memories.
Goal Specification	Defining agent objectives.
Interrupt Control	External override triggers.
Monitoring Hook	Observability into agent state.
Execution Sandbox	Safe environment for actions.
Tool Router	Selects tool sequences.
Chain-of-Command	Hierarchical agent control.

10. Computer Vision

Term	Meaning
Convolution	Feature extraction filter.
Kernel	Matrix for convolution.
Stride	Step size.
Padding	Border extension.
Feature Map	Intermediate visual representation.
Max Pooling	Downsampling.
ROI Align	Spatial pooling for detection.
Vision Transformer	Patch-based attention model.
Patch Embeddings	Tokenize images.
Segmentation	Pixel classification.

Optical Flow	Motion estimation.
Pose Estimation	Joint/keypoint prediction.
Depth Estimation	Distance from camera.
Face Embeddings	Feature vector for faces.
Augmentation	Image transformations.
SIFT	Feature detection algorithm.
HOG	Histogram of gradients.
SSD	Single-shot detector.
R-CNN	Two-stage detector family.
Heatmap Regression	Predicting pixel-level density for objects.

11. RAG

Term	Meaning
Vector Embedding	Dense semantic vector.
Vector Store	ANN database for embeddings.
Chunking	Breaking documents into pieces.
Top-K Retrieval	Best-K similar items.
Hybrid Search	Dense + sparse retrieval.
BM25	Keyword-based scorer.
Re-ranking	Secondary scoring for precision.
Context Window Budget	Token allocation.
Metadata Filtering	Filter by tags or attributes.
Embedding Model	Encoder for text vectors.
Multi-Hop Retrieval	Multi-step retrieval process.
Query Expansion	Rewrite query for better recall.
Context Stitching	Joining chunks into prompts.
Relevance Score	Similarity measure.

Attention Over Context	LLM attention distribution.
Retrieval Drift	Degraded retrieval relevance.
Context Pruning	Remove low-value chunks.
Hallucination Reduction	Use retrieval to prevent fabrication.
Score Fusion	Combine multiple retrievers.
Semantic Matching	Deep meaning-based search.

---

12. Diffusion Models

Term	Meaning
Forward Diffusion	Add noise progressively.
Reverse Diffusion	Denoising to generate images.
Noise Schedule	Controls noise amount per step.
Timestep Embedding	Encoding diffusion step.
DDPM	Base diffusion model.
DDIM	Deterministic sampling.
Scheduler	Diffusion sampling algorithm.
VAE	Latent image compression model.
CFG	Controls prompt strength.
Latent Space	Compressed representation.
Text Encoder	Converts prompt → embedding.
LoRA	Low-rank tuning for diffusion.
DreamBooth	Custom subject fine-tuning.
ControlNet	Structural guidance network.
Inpainting	Replacing masked region.
Outpainting	Extending an image.
Upscaling	Increasing resolution.

Sampler Steps	Iterations during inference.
Inference Pipeline	Sequence from prompt to image.
Attention Map	Visualizing attention.

13. Object Detection

Term	Meaning
Bounding Box	Object region.
IoU	Intersection over union.
NMS	Remove overlapping boxes.
Confidence Score	Object presence likelihood.
Anchor Boxes	Preset box priors.
Anchor-Free	No predefined anchors.
Feature Pyramid	Multi-level features.
RPN	Region proposal network.
ROI Align	Object feature extraction.
mAP	Primary detection metric.
DIoU / CIoU	Advanced bounding box losses.
Detection Head	Final prediction layers.
Focal Loss	Handle class imbalance.
Heatmap	Pixel-level center probability.
Decoder	Box refinement stage.
Hard Negative Mining	Handle false positives.
Label Assignment	Anchor ↔ GT box matching.
Multi-Scale Training	Different image sizes.
Latency	Inference speed.
ONNX	Portable model format.

14. YOLO

Term	Meaning
CSPDarknet	YOLO backbone.
PANet	Feature aggregation neck.
Decoupled Head	Separate cls/reg branches.
Mosaic	4-way augmentation.
MixUp	Image blending.
SPPF	Fast spatial pyramid pooling.
CIoU Loss	Box regression loss.
Anchor-Free Head	Direct box predictions.
Objectness Score	Probability of object.
AutoAnchor	Anchor optimization.
TTA	Test-time augmentation.
NMS	Box suppression.
EMA Weights	Smoothed weights.
DFL Loss	Distribution focal loss.
YOLOv8 Head	Unified detection head.
YOLO-NAS	NAS-optimized YOLO.
Int8/FP16	Quantization modes.
Tile Inference	Detect small objects.
Ultralytics Engine	YOLO inference framework.
ONNX Export	Model conversion.

15. Fine-Tuning

Term	Meaning
LoRA	Low-rank adaptation.
QLoRA	Memory-efficient LoRA.

SFT	Supervised fine-tuning.
RLHF	Reinforcement tuning.
DPO	Pairwise preference optimization.
Token Masking	Exclude prompt tokens from loss.
Learning Rate	Update step size.
Weight Decay	Regularization term.
Gradient Accumulation	Simulate large batch sizes.
Layer Freezing	Lock certain layers.
Adapters	Inserted trainable modules.
Bias Tuning	Train bias parameters only.
Warmup Steps	Gradual LR ramp-up.
Mixed Precision	FP16/BF16 training.
Checkpointing	Save model snapshots.
Epoch	Full dataset pass.
Overfitting Detection	Monitor validation loss.
Merging LoRA	Combine LoRA weights.
Quantized Training	Training with 8/4-bit weights.
Parameter Count	Total tunable params.

---

16. Context Engineering

Term	Meaning
Token Budget	Allocation of available tokens.
Context Window	Max input length.
Context Compression	Summaries as replacements.
Memory Retrieval	Fetching relevant past data.
System Prompt	Primary rule.
Instruction Hierarchy	Ordering of rules.
Context Stitching	Combining multiple blocks.

Context Pruning	Removing irrelevant text.
Semantic Chunking	Meaning-based segmentation.
Attention Biasing	Steering model attention.
Delimiter-Based Prompting	Structure separators.
Relevance Scoring	Rank contextual elements.
Context Drift	Accumulated noise.
Context Interference	Conflicting signals.
Metadata Tags	Structured context descriptors.
Role Conditioning	Enforcing persona.
Prompt Wrappers	Templates for tasks.
Summary Memory	Condensed long history.
Context Overflow	Tokens exceeding limit.
Positional Importance	Order-based priority.

---

17. Prompt Engineering

Term	Meaning
Zero-Shot Prompting	No examples provided.
Few-Shot Prompting	Provide some examples.
CoT Prompting	Chain-of-thought.
ReACT Prompting	Reason + Act.
Persona Prompting	Assign identity/role.
Style Transfer Prompting	Modify tone/voice.
Instruction Prompting	Clear commands.
Delimiter Prompts	Structure isolation.
Output Constraints	Format rules.
JSON Mode	Enforced structured output.
Negative Prompting	What NOT to do.

Self-Critique Prompt	The model evaluates itself.
Reflection Prompt	Iterative improvement.
Decomposition Prompt	Break tasks into steps.
Meta Prompting	Instructions about how to follow instructions.
Prompt Chaining	Multi-step prompts.
Retrieval-Aware Prompt	Uses RAG input.
Safety Prompt	Avoid harmful content.
Red Team Prompt	Stress-test behavior.
Prompt Tokens	Number of prompt characters.

---

18. Recommendation Systems

Term	Meaning
User Embedding	Vector representation of user preferences.
Item Embedding	Vector for item properties.
Matrix Factorization	Latent mapping of user-item space.
Collaborative Filtering	Behavior-based recommendations.
Content-Based Filtering	Attribute-based recommendations.
Cold Start Problem	New user/item without history.
CTR Prediction	Click-through rate modeling.
Ranking Model	Scores recommendations.
Candidate Generation	First-stage retrieval.
Two-Tower Model	Dual-encoder architecture.
BPR Loss	Pairwise ranking loss.
Session-Based Modeling	Recent session-driven patterns.
Wide & Deep Model	Mix of linear + deep layers.
ANN Retrieval	Fast item search.
Re-Ranking	Final ordering.

User Profiles	Stored preference data.
Multi-Armed Bandit	Explore/exploit algorithm.
NDCG	Ranking quality metric.
Diversity Metric	Content variety measurement.
Explore–Exploit Ratio	Balancing new vs known items.

19. Evaluation

Term	Meaning
Accuracy	Correct predictions rate.
Precision	TP / (TP+FP).
Recall	TP / (TP+FN).
F1 Score	Harmonic mean of accuracy+recall.
ROC Curve	Threshold evaluation.
AUC	Area under ROC.
MAE	Mean absolute error.
MSE	Mean squared error.
RMSE	Root mean squared error.
mAP	Detection performance metric.
Perplexity	LLM performance metric.
BLEU	Translation evaluation.
ROUGE	Summarization evaluation.
METEOR	Text similarity scoring.
Hit Rate	Recommendation metric.
NDCG	Ranking evaluation metric.
Latency	Time to respond.
Throughput	Queries per second.
Confidence Calibration	Mapping probabilities to real likelihood.
A/B Testing	Compare model variants.

20. Optimization

Term	Meaning
Learning Rate	Step size for updates.
Momentum	Smooth optimization.
Adam	Adaptive optimizer.
AdamW	Decoupled weight decay.
Warmup	Gradual LR ramp-up.
Scheduler	LR adjustment strategy.
Gradient Clipping	Prevent exploding gradients.
Dropout	Regularization technique.
BatchNorm	Activation normalization.
Weight Decay	Penalize large weights.
Early Stopping	Prevent overfitting.
Regularization	Reduce overfitting.
Quantization	FP32 → FP16/INT8.
Pruning	Remove redundant weights.
Distillation	Teacher → student training.
Mixed Precision	Use FP16/BF16.
Checkpointing	Save intermediate states.
Distributed Training	Multi-GPU scaling.
Pipeline Parallelism	Split model across devices.
AllReduce	Synchronize gradients.

21. Deployment

Term	Meaning

Model Serving	API-based model inference.
Containerization	Dockerizing ML workloads.
gRPC	High-performance communication.
API Gateway	Routing model requests.
Load Balancing	Distribute traffic.
Model Versioning	Track deploy iterations.
Canary Deployment	Test model with small traffic.
A/B Deployment	Compare two live models.
Feature Rollout	Stage-wise release.
Inference Optimization	Improve speed/latency.
GPU Serving	GPU-based model execution.
ONNX Runtime	Cross-platform inference.
TensorRT	NVIDIA accelerated inference.
Request Batching	Combine multiple queries.
Autoscaling	Scale up/down compute.
Logging	Track system behavior.
Monitoring	Live analytics of performance.
Failover	Backup system on failure.
Rate Limiting	Prevent overload.
Edge Deployment	On-device inference.

22. MLOps

Term	Meaning
CI/CD Pipeline	Automated build + deploy.
Model Registry	Store model versions.
Feature Store	Centralized feature hub.
Data Drift	Change in input distribution.
Concept Drift	Change in input–output relationship.

Monitoring Dashboard	Track live metrics.
Orchestration	Pipeline automation (Airflow).
Experiment Tracking	Save model configs + metrics.
Canary Deployment	Gradual rollout.
Shadow Mode	Test model silently.
Model Lineage	End-to-end tracking.
Retraining Pipeline	Automatic updates.
A/B Testing	Online model comparisons.
Logging System	Persisted event logs.
Model Governance	Security & compliance.
Drift Detection	Identify performance drops.
Feedback Loop	Continuous data integration.
Resource Scaling	Autoscaling compute.
SLA Compliance	Meeting latency/uptime targets.
Fail-Safe Mechanism	Auto-fallback strategy.

---

23. Hardware Acceleration

Term	Meaning
GPU	Parallel processor for ML.
TPU	Google ML accelerator.
VRAM	GPU memory used in training.
CUDA	NVIDIA GPU programming platform.
cuDNN	NVIDIA deep learning library.
Tensor Cores	Specialized matrix multiplication hardware.
BF16	Efficient numeric format.
FP16	Half-precision format.
Quantized INT8	Low precision for inference.
Mixed Precision	Blend FP16 + FP32.

Memory Bandwidth	Data transfer capacity.
Throughput	Compute per second.
PCIe	GPU-to-CPU interface.
NVLink	High-speed GPU interconnect.
Multi-GPU Parallelism	Model split across GPUs.
Distributed Worker	Node in cluster training.
Compute Kernel	Low-level GPU operation.
Tiled Matrix Multiply	Efficient GPU math pattern.
FlashAttention	IO-optimized attention.
TensorRT	Hardware-accelerated inference.