# AI Fundamentals

April 2026

## 1. What Is Artificial Intelligence?

**Artificial intelligence (AI)** is a branch of computer science focused on building systems that can perform tasks typically requiring human cognition — pattern recognition, decision-making, language understanding, and prediction. Rather than a single technology, AI is an umbrella term covering a broad set of methods and techniques aimed at replicating or augmenting specific cognitive capabilities.

It is important to distinguish between two categories of AI:

- **Narrow AI (Weak AI)** — Systems designed for a specific task or a constrained set of tasks. Every AI system deployed in the real world today — from search engines to voice assistants to medical imaging tools — is narrow AI. These systems can be remarkably capable within their domain but have no understanding or ability outside it.
- **General AI (AGI)** — A theoretical form of AI with human-level reasoning, learning, and adaptability across all cognitive domains. AGI does not yet exist. It remains an active area of research and debate, but no current system meets this threshold.

Historically, the field has been shaped by two major approaches. **Symbolic AI** (dominant from the 1950s through the 1980s) relied on handcrafted rules and logic — expert systems that encoded human knowledge as if/then statements. The **statistical/machine learning approach**, which began gaining traction in the 1990s, learns patterns directly from data rather than following pre-programmed rules. Modern AI is overwhelmingly dominated by this statistical paradigm, particularly through deep learning.

## 2. How AI Learns — The Core Process

**Machine learning (ML)** is the dominant method by which modern AI systems are built. Rather than being explicitly programmed with rules for every scenario, ML systems learn patterns and relationships from data, then apply those learned patterns to new, unseen inputs. The process follows a structured pipeline.

### The Training Pipeline

1. **Data Collection and Preparation** — The process begins with gathering a dataset relevant to the task. Data may be **labeled** (each example tagged with its correct answer) or **unlabeled** (raw data without annotations). Data quality is critical: noisy, incomplete, or biased data will produce unreliable models. This stage typically involves cleaning (removing errors and duplicates), formatting, and splitting data into training, validation, and test sets.
2. **Model Training** — During training, the model processes input data, generates predictions, and compares those predictions against the expected output. The difference between prediction and reality is quantified as a **loss** (also called error). The model then adjusts its internal **parameters** (numerical weights) to reduce that loss. This optimization process is driven by an algorithm called **gradient descent**, which iteratively nudges parameters in the direction that minimizes error. Training typically requires many passes (epochs) through the dataset.
3. **Evaluation and Testing** — Once trained, the model is tested against data it has never seen before. This step measures how well the model **generalizes** — that is, whether it has learned true underlying patterns rather than simply memorizing the training examples. A model that performs

well on training data but poorly on new data is said to be **overfitting**. Evaluation metrics vary by task (accuracy, precision, recall, F1 score, etc.).

4. **Deployment and Inference** — A validated model is deployed into a production environment where it processes real-world inputs and returns predictions. This operational phase is called **inference**. Deployed models require ongoing monitoring for performance degradation, data drift, and edge cases not represented in training data.

### Three Learning Paradigms

Machine learning methods are broadly organized into three paradigms, suited to different types of problems:

- **Supervised Learning** — The model learns from labeled examples, where each input is paired with its correct output. The goal is to learn a mapping from inputs to outputs that generalizes to new data. Use cases include email spam detection, image classification, and credit risk scoring.
- **Unsupervised Learning** — The model works with unlabeled data, seeking to discover hidden structure, groupings, or patterns without being told what to look for. Use cases include customer segmentation, anomaly detection, and dimensionality reduction.
- **Reinforcement Learning (RL)** — An agent learns by interacting with an environment, taking actions, and receiving rewards or penalties based on outcomes. Over time, the agent develops a strategy (policy) that maximizes cumulative reward. Use cases include game-playing AI, robotics control, and resource optimization.

## 3. Neural Networks — The Engine Behind Modern AI

**Neural networks** are computational models loosely inspired by biological neurons, and they form the backbone of modern AI. A neural network is organized into layers: an **input layer** (which receives raw data), one or more **hidden layers** (which process and transform the data), and an **output layer** (which produces the final prediction or classification).

Data flows forward through the network in a process called **forward propagation**. Each layer applies mathematical transformations — weighted sums followed by nonlinear activation functions — that extract increasingly abstract features from the input. Early layers might detect simple patterns (edges in an image, common word pairs in text), while deeper layers identify complex, high-level representations (faces, sentence meaning).

The term **"deep learning"** simply refers to neural networks with many hidden layers. Depth is what gives these models their power — more layers allow the network to learn more nuanced and hierarchical representations of data.

### Key Architectures

- **CNNs (Convolutional Neural Networks)** Specialized for image and spatial data. CNNs use convolutional filters that slide across input data to detect local patterns (edges, textures, shapes) and build up to full-object recognition. Widely used in facial recognition, medical imaging, autonomous vehicles, and video analysis.
- **RNNs / LSTMs (Recurrent Neural Networks / Long Short-Term Memory)** Designed for sequential data where order matters. These architectures maintain an internal memory state that allows information to persist across time steps. Used in speech recognition, time-series forecasting, and machine translation (though increasingly replaced by Transformers).
- **Transformers** — The architecture behind modern large language models. Unlike RNNs, Transformers process entire sequences in parallel rather than step-by-step. Their core innovation is the **attention**

**mechanism**, which allows the model to dynamically weigh which parts of the input are most relevant to each part of the output. This parallelism and selective focus make Transformers highly efficient and scalable.

# 4. Large Language Models and Generative AI

**Large language models (LLMs)** are transformer-based neural networks trained on massive text corpora — often spanning significant portions of the publicly available internet. Their fundamental mechanism is **next-token prediction**: given a sequence of tokens (word pieces), the model predicts the most probable next token. Despite this seemingly simple objective, the scale of training data and model parameters produces systems capable of sophisticated language understanding, reasoning, and generation.

LLMs are built through a two-phase process:

- **Pre-training** — The model learns general language patterns, grammar, facts, and reasoning capabilities from large, diverse datasets. This phase is computationally expensive and produces a broad, general-purpose language model.
- **Fine-tuning** — The pre-trained model is further trained on smaller, task-specific or curated datasets to specialize its behavior. A key technique is **Reinforcement Learning from Human Feedback (RLHF)**, in which human evaluators rank model outputs and the model is trained to align with those preferences — improving helpfulness, accuracy, and safety.

**Generative AI** refers broadly to models that produce new content — text, images, code, audio, or video — rather than simply classifying or analyzing existing data. While LLMs are the most prominent example for text generation, other architectures serve different modalities. **Diffusion models**, for instance, generate images by starting with random noise and iteratively refining it into coherent visual output, and are the technology behind image generation tools.

Several practical considerations are essential when working with these systems:

- **Hallucination** — LLMs can generate outputs that are fluent and confident but factually incorrect. The model has no inherent mechanism for verifying truth; it produces statistically plausible text.
- **Context windows** — Every LLM has a maximum input size (measured in tokens). Information beyond this limit is not processed. Context window sizes vary by model and directly affect what the model can consider when generating a response.
- **Training knowledge vs. real-time information** — An LLM's knowledge is frozen at the time of its last training update. It does not have access to real-time information unless connected to external tools or retrieval systems.

# 5. Key Terms Reference

| Term | Definition |
|---|---|
| **Algorithm** | A defined sequence of steps or rules that a computer follows to solve a problem or complete a task. In AI, algorithms govern how models learn from data and make decisions. |

| Term | Definition |
|---|---|
| **Machine Learning** | A subset of AI in which systems learn patterns from data rather than being explicitly programmed with rules. Performance improves with exposure to more and better data. |
| **Neural Network** | A computational model organized in layers of interconnected nodes (neurons) that process data by applying weighted transformations. The foundation of deep learning. |
| **Deep Learning** | A subset of machine learning that uses neural networks with many hidden layers to learn hierarchical representations of data. Responsible for most recent breakthroughs in AI. |
| **Training Data** | The dataset used to train a machine learning model. The quality, size, and representativeness of training data directly impact model performance and bias. |
| **Model** | The learned mathematical representation produced by a machine learning algorithm after training. A model encodes patterns from training data and applies them to new inputs. |
| **Supervised Learning** | A learning paradigm where the model is trained on labeled data — input-output pairs where the correct answer is provided during training. |
| **Unsupervised Learning** | A learning paradigm where the model discovers patterns or structure in unlabeled data without being given correct answers. |
| **Reinforcement Learning** | A learning paradigm where an agent learns optimal behavior through trial and error, receiving rewards or penalties from an environment based on its actions. |
| **Natural Language Processing (NLP)** | The branch of AI focused on enabling machines to understand, interpret, and generate human language. Encompasses tasks like translation, summarization, and sentiment analysis. |
| **Transformer** | A neural network architecture that processes sequences in parallel using attention mechanisms. The dominant architecture behind modern LLMs and many state-of-the-art AI systems. |
| **Token / Tokenization** | A token is a unit of text (a word, subword, or character) that a language model processes. Tokenization is the process of splitting raw text into these units before feeding it to a model. |
| **Embedding** | A dense numerical representation (vector) of data — such as a word, sentence, or image — in a continuous space where similar items are positioned closer together. |
| **Overfitting** | A condition where a model performs well on training data but poorly on new, unseen data — indicating it has memorized specific examples rather than learning generalizable patterns. |
| **Inference** | The process of using a trained model to generate predictions or outputs on new input data. This is the operational phase after training is complete. |

| Term | Definition |
|---|---|
| **Fine-Tuning** | The process of further training a pre-trained model on a smaller, task-specific dataset to adapt its behavior for a particular use case or to align it with human preferences. |

## 6. Practices for Best Results

**It's a Tool**
AI is a tool, and like any tool, the ability to use it correctly to get the best results, is understanding the tool use or application of it, whether it's a physical object, wrench, knife etc. or computational, getting the best results is understanding how to apply the tool.

**Garbage in Garbage out (GiGo)**
Clearly state objective(s), too much generalization results in generated responses that are broad and/or drift from the goal/objective.

**Know the Topic**
If you don't know what you are asking for (GiGo), you won't be able to understand the output (response). Having fundamental knowledge is essential in understanding the result of the query and being able to apply the result to your objective.

**Analysis**
You need to keep the AI honest. It will give erroneous information and direction that on the surface looks good, don't trust it, your experience and knowledge is essential to making this tool work for your benefit.

**Socialization**
These tools are very "socialized", they are "engineered" to give you a comfort level of interaction that mimic's human responses. For lay people this can be and is very misleading, believing they are interacting with a friend or expert in any given field e.g. Medical, Diet, Advice or any number of topics that we experience in everyday life, even friendly social chatting. It is very subtle and seductive. It is a machine, a tool.

Eugene Beckett
IT – Network Consultant
EPB Enterprises