# The Role of Mathematics in Artificial Intelligence and Machine Learning

**Dr. Mrinal Sarma**
Assistant Professor, Department of Mathematics
Narangi Anchalik Mahavidyalaya, Guwahati, Assam, India
E-Mail: mrinal1973@narangianchalikmahavidyalaya.ac.in

**How to Cite this Article**:

Sarma, M. (2026). The Role of Mathematics in Artificial Intelligence and Machine Learning. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(03). https://doi.org/10.55041/ijcope.v2i3.145



https://doi.org/10.55041/ijcope.v2i3.145

## Abstract

Artificial Intelligence (AI) and Machine Learning (ML) have transitioned from theoretical constructs to pervasive technologies shaping our modern world. Behind their sophisticated capabilities lies a foundational bedrock of mathematical principles. This paper explores the critical and multifaceted role of mathematics in the development, implementation, and advancement of AI and ML. We delve into core mathematical disciplines, including linear algebra, calculus, probability and statistics, and discrete mathematics, examining how each contributes to fundamental AI/ML concepts such as data representation, model optimization, learning algorithms, and decision-making processes. Furthermore, we discuss the evolving landscape of mathematical research integral to pushing the boundaries of AI, emphasizing the symbiotic relationship between mathematical innovation and AI progress. This paper aims to provide a comprehensive overview for researchers and practitioners, highlighting the indispensable nature of mathematics in understanding and developing intelligent systems.

**Keywords:** *Artificial Intelligence, Machine Learning, Mathematics, Linear Algebra, Calculus, Probability, Optimization, Algorithms.*

## 1. Introduction

The rapid proliferation of Artificial Intelligence (AI) and Machine Learning (ML) has ignited a revolution across diverse sectors, from healthcare and finance to autonomous systems and natural language processing. These technologies empower machines to perform tasks that typically require human intelligence, such as learning, problem-solving, perception, and decision-making. However, the seemingly magical capabilities of AI/ML artifacts are not born from an ethereal source; rather, they are meticulously engineered through the application of sophisticated mathematical frameworks.

Mathematics serves as the fundamental language and toolkit that underpins every aspect of AI and ML, from the representation of data to the intricate algorithms that enable learning and prediction. This paper asserts that mathematics is not merely a supporting discipline for AI/ML but its indispensable cornerstone. Without a robust understanding of mathematical principles, the effective design, implementation, interpretation, and advancement of AI/ML systems would be impossible.

We will systematically explore the key mathematical domains that contribute to AI/ML, illustrating their direct impact on core concepts and algorithms. Our journey will begin by establishing the fundamental relationship between mathematics and AI/ML, followed by a review of related literature, a detailed examination of specific mathematical disciplines and their applications, a discussion of emerging mathematical frontiers, and a reflection on the future interplay between mathematics and intelligent systems.

**2. Literature Review**

The rapid development of Artificial Intelligence (AI) and Machine Learning (ML) has been closely associated with advances in mathematical theory and computational methods. Scholars across disciplines have emphasized that modern AI systems are fundamentally built upon mathematical structures that enable learning, reasoning, and decision-making.

One of the most influential contributions to the mathematical understanding of AI is the work of **Goodfellow, Bengio, and Courville (2016)**, who provided a comprehensive framework explaining how deep learning architectures rely heavily on linear algebra, optimization, and probability theory. Their work demonstrated that neural networks can be interpreted as compositions of mathematical functions operating in high-dimensional vector spaces. Similarly, **Hastie, Tibshirani, and Friedman (2009)** explored statistical learning theory, presenting a rigorous mathematical foundation for supervised and unsupervised learning algorithms. Their research emphasized the role of statistical inference, regularization, and model complexity in determining predictive performance.

The probabilistic perspective of machine learning has been extensively developed by **Murphy (2012)**, who argued that probabilistic modelling offers a unified framework for representing uncertainty and learning from data. Probabilistic graphical models, including Bayesian networks and Markov random fields, rely on probability theory and graph theory to represent complex dependencies among variables. Earlier work by **Koller and Friedman (2009)** further established the mathematical basis of graphical models, demonstrating how probabilistic inference can be formulated as an optimization and combinatorial problem.

Another major mathematical pillar of AI research is **optimization theory**. Training machine learning models requires minimizing loss functions defined over large parameter spaces. **Boyd and Vandenberghe (2018)** provided a rigorous treatment of convex optimization techniques that are widely used in machine learning, including gradient descent, interior-point methods, and dual optimization. These methods enable efficient parameter estimation in large-scale learning systems. Optimization has become particularly important in deep learning, where non-convex objective functions require sophisticated numerical algorithms for stable training.

Research in **deep learning** has further emphasized the importance of mathematical modelling in understanding representation learning. **LeCun, Bengio, and Hinton (2015)** highlighted how deep neural networks use hierarchical representations to capture increasingly abstract features from raw data. Their work showed that convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks rely on mathematical principles such as matrix operations, gradient-based optimization, and probabilistic inference.

Dimensionality reduction techniques have also played a crucial role in the development of machine learning. **Hinton and Salakhutdinov (2006)** demonstrated that neural networks can perform nonlinear dimensionality reduction, extending traditional methods such as Principal Component Analysis (PCA). Their research illustrated how deep autoencoders can compress high-dimensional data into lower-dimensional representations while preserving meaningful structures.

More recently, the integration of **geometric and topological methods** into machine learning has opened new research directions. **Bronstein et al. (2021)** introduced the concept of geometric deep learning, which generalizes neural network architectures to non-Euclidean domains such as graphs and manifolds. This framework uses tools from differential geometry, spectral graph theory, and topology to analyze structured data such as social networks, molecular structures, and 3D shapes.

Another important theoretical perspective was proposed by **Jordan (2015)**, who emphasized the role of mathematics in bridging the gap between statistical inference, optimization, and large-scale data analysis. Jordan argued that modern AI systems should be viewed as mathematical systems integrating statistical modelling with algorithmic efficiency.

In addition to statistical and geometric approaches, **causal inference** has emerged as an important mathematical framework for AI research. **Pearl (2009)** introduced a formal theory of causality using directed acyclic graphs and structural equation models. His work highlighted the limitations of purely correlational learning systems and emphasized the importance of causal reasoning in building intelligent systems capable of understanding real-world relationships.

Reinforcement learning provides another area where mathematics plays a central role. **Sutton and Barto (2018)** demonstrated that reinforcement learning algorithms are grounded in Markov Decision Processes (MDPs), dynamic programming, and stochastic optimization. Concepts such as the Bellman equation, value functions, and policy gradients rely on probability theory and numerical optimization techniques.

Recent studies have also focused on **information-theoretic perspectives** in machine learning. Information theory provides quantitative measures such as entropy, mutual information, and Kullback–Leibler divergence, which are widely used to evaluate model uncertainty and information flow in neural networks. These measures help explain learning dynamics and guide the design of generative models such as variational autoencoders (VAEs) and generative adversarial networks (GANs).

Another emerging direction in the literature is the use of **optimal transport theory** in machine learning. Originally developed in pure mathematics, optimal transport has been applied to problems such as domain adaptation, generative modelling, and distribution alignment. These approaches provide mathematically principled methods for comparing probability distributions in high-dimensional spaces.

Furthermore, the increasing scale of modern AI systems has highlighted the importance of **numerical linear algebra and high-dimensional statistics**. Efficient matrix decomposition, sparse representations, and randomized algorithms are now essential for training large neural networks and processing massive datasets.

Overall, the literature clearly demonstrates that AI and machine learning are deeply rooted in mathematical theory. From statistical learning and optimization to geometry and information theory, mathematics provides the conceptual tools required to design, analyze, and improve intelligent systems. As AI technologies continue to evolve, ongoing mathematical research will remain essential for addressing challenges related to interpretability, robustness, scalability, and ethical deployment.

## 3. The Fundamental Interplay: Mathematics as the Language of Intelligence

Artificial Intelligence (AI) seeks to develop computational systems capable of performing tasks that normally require human cognitive abilities such as reasoning, learning, perception, and decision-making. Machine Learning (ML), a core subfield of AI, focuses on enabling systems to improve their performance automatically through experience and data. At the heart of these capabilities lies mathematics, which provides the formal language, structure, and analytical tools required to design intelligent algorithms.

Mathematics enables AI systems to transform raw data into meaningful representations, construct predictive models, learn patterns, and make decisions under uncertainty. Without mathematical abstraction and formal reasoning, the design and analysis of AI algorithms would lack precision and reliability. In essence, mathematics provides both the theoretical framework and the computational mechanisms that allow AI systems to function effectively.

The interplay between mathematics and AI can be understood through several fundamental processes: data representation, model formulation, learning and optimization, and decision-making.

### 3.1 Mathematical Representation of Data

One of the first steps in any AI or machine learning task is converting real-world information into mathematical structures that can be processed by algorithms. Data originating from various domains—such as images, text, speech signals, or sensor measurements—must be encoded into numerical formats. These representations typically take the form of vectors, matrices, tensors, or graphs.

For example, in computer vision, an image is represented as a matrix of pixel values, where each entry corresponds to the intensity of a pixel. In natural language processing, words or sentences are represented using numerical embeddings, which map linguistic units into high-dimensional vector spaces. These embeddings allow algorithms to measure semantic similarity using geometric operations such as vector distance or cosine similarity.

Mathematical structures enable AI systems to manipulate and analyze data efficiently. Linear algebra plays a particularly significant role in this context, as operations such as matrix multiplication, vector transformations, and eigenvalue decomposition form the backbone of many machine learning algorithms.

### 3.2 Mathematical Formulation of Learning Models

AI models are essentially mathematical functions designed to approximate relationships between input data and desired outputs. A machine learning model can be expressed as a function

$$y = f(x, \theta)$$

where $x$represents the input data, $y$represents the predicted output, and $\theta$denotes the set of parameters that define the model.

The objective of learning is to determine the optimal values of the parameters $\theta$such that the model accurately captures patterns in the data. Different AI models correspond to different mathematical formulations. For instance:

- **Linear regression** models relationships using linear equations.
- **Neural networks** approximate complex nonlinear functions using layers of interconnected nodes.
- **Decision trees** use recursive partitioning of the feature space.
- **Probabilistic models** represent relationships using probability distributions.

These mathematical formulations enable models to generalize from observed data and make predictions about unseen data. The ability to formalize learning problems mathematically allows researchers to analyze model performance, identify limitations, and design improved algorithms.

### 3.3 Learning Through Optimization

Learning in machine learning is essentially an optimization problem. The goal is to find parameter values that minimize a predefined loss or cost function measuring the discrepancy between predicted outputs and actual observations.

If a dataset contains input-output pairs $(x_i, y_i)$, a typical objective is to minimize a loss function $L(\theta)$, defined as

$$L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i, \theta), y_i)$$

where $n$is the number of training samples and $\ell(\cdot)$represents the error between predicted and true outputs.

Optimization techniques such as **gradient descent** are widely used to solve this problem. Gradient descent iteratively updates model parameters according to

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t)$$

where $\eta$is the learning rate and $\nabla L(\theta_t)$represents the gradient of the loss function.

The process of backpropagation in neural networks is fundamentally based on the chain rule of calculus, which allows efficient computation of gradients across multiple layers of a network. Through repeated optimization steps, the model gradually improves its ability to capture patterns within the data.

### 3.4 Mathematical Foundations of Decision-Making

After training, AI models use mathematical reasoning to generate predictions or decisions. These decisions may involve classification, regression, clustering, or control actions. Mathematical functions determine how input features are transformed into output predictions.

For instance, in classification problems, models often use probability distributions to estimate the likelihood of different classes. A classifier might output probabilities such as

$$P(y = k \mid x)$$

for each possible class $k$. The final decision is typically made by selecting the class with the highest probability.

Decision-making under uncertainty relies heavily on probability theory and statistical inference. Bayesian approaches allow models to update beliefs about unknown variables as new information becomes available. This probabilistic reasoning is particularly important in domains such as medical diagnosis, autonomous vehicles, and financial forecasting.

### 3.5 Mathematical Abstraction and Generalization

Another critical contribution of mathematics to AI is the ability to generalize beyond observed data. Generalization refers to a model's capability to perform well on new, unseen data rather than simply memorizing training examples.

Statistical learning theory provides mathematical tools for analyzing this phenomenon. Concepts such as bias–variance trade-off, regularization, and model complexity help researchers understand why certain models generalize better than others. Mathematical frameworks also provide guarantees about the convergence and stability of learning algorithms.

By analyzing models through theoretical principles, researchers can develop algorithms that are not only accurate but also reliable and interpretable.

### 3.6 Mathematics as a Unifying Framework for AI

Mathematics ultimately serves as the unifying language that integrates diverse components of AI systems. Linear algebra enables efficient representation of high-dimensional data, calculus drives optimization and learning, probability theory models uncertainty, and discrete mathematics supports algorithmic reasoning.

This unified mathematical framework allows AI researchers to systematically analyze algorithms, prove theoretical properties, and design new computational methods. As AI systems grow increasingly complex, mathematical theory continues to play a crucial role in ensuring their reliability, efficiency, and scalability.

Thus, the relationship between mathematics and AI is not merely supportive but fundamentally structural. Mathematics provides the conceptual architecture that allows intelligent systems to learn from data, reason about uncertainty, and make informed decisions in complex environments.

## 4. Core Mathematical Disciplines and Their Contributions to AI/ML

### 4.1 Linear Algebra

Linear algebra provides the framework for representing and manipulating high-dimensional data. Vectors, matrices, and tensors define relationships between features, allowing complex operations such as matrix multiplication, eigenvalue decomposition, and singular value decomposition (SVD). Techniques like Principal Component Analysis (PCA) rely on eigenvectors to identify dominant data directions.
*Example:* In convolutional neural networks (CNNs), images are treated as matrices and processed via convolution operations to extract features like edges and patterns.

### 4.2 Calculus

Calculus, especially differential calculus, enables learning through optimization. Derivatives and gradients guide parameter updates in algorithms like Gradient Descent or Adam. The chain rule underpins backpropagation in deep learning.
*Example:* In logistic regression, the derivative of a loss function (cross-entropy) directs how model weights are updated to minimize prediction error.

### 4.3 Probability and Statistics

These disciplines handle uncertainty, data variability, and inference. Bayesian reasoning, hypothesis testing, and maximum likelihood estimation are fundamental. Information theory concepts such as entropy and KL divergence measure uncertainty and model quality.
*Example:* Naive Bayes classifiers use Bayes' theorem to estimate the likelihood of an email being spam based on word probabilities.

### 4.4 Discrete Mathematics

Crucial for algorithmic reasoning, discrete mathematics includes graph theory, set theory, logic, and combinatorics. Graphs model relationships between entities, enabling social network analysis, recommendation systems, and pathfinding algorithms.
*Example:* Recommender systems use graph-based collaborative filtering to suggest content based on user–item connections.

## 5. Advanced Mathematical Concepts in Modern AI/ML

While classical mathematical disciplines such as linear algebra, calculus, and probability form the foundation of Artificial Intelligence and Machine Learning, recent advances in AI research increasingly rely on more sophisticated mathematical frameworks. These advanced mathematical concepts enable the development of powerful algorithms capable of handling complex data structures, large-scale optimization problems, and dynamic learning environments. Modern AI systems incorporate ideas from optimization theory, geometry, functional analysis, dynamical systems, and stochastic processes to achieve improved performance, interpretability, and scalability.

### 5.1 Optimization Theory

Optimization theory plays a central role in training machine learning models. Most learning algorithms can be formulated as optimization problems where the objective is to minimize a loss function that quantifies the difference between predicted outputs and actual observations.

In mathematical terms, the training process can be expressed as

$$\theta^* = \arg\min_{\theta} L(\theta)$$

where $\theta$represents model parameters and $L(\theta)$denotes the loss function.

Traditional optimization methods include gradient descent, stochastic gradient descent (SGD), and Newton-type methods. However, deep learning models often involve highly non-convex optimization landscapes with numerous local minima and saddle points. Advanced optimization techniques such as Adam, RMSProp, and momentum-based methods have been developed to address these challenges.

Convex optimization has also played a significant role in machine learning. Convex problems possess unique global minima, allowing efficient and reliable solutions. Many classical learning algorithms—including support vector machines (SVMs), logistic regression, and LASSO regression—are based on convex optimization frameworks.

In large-scale machine learning, distributed optimization and parallel algorithms have become increasingly important, enabling training of models on massive datasets across multiple computing nodes.

**5.2 Information Geometry**

Information geometry is an emerging mathematical field that combines differential geometry with probability theory. It studies statistical models as geometric objects defined on manifolds, allowing researchers to analyze learning algorithms from a geometric perspective.

In this framework, probability distributions can be viewed as points on a manifold, and distances between distributions are measured using divergence metrics such as the Kullback–Leibler (KL) divergence. The Fisher information matrix plays a key role in defining the geometric structure of statistical models.

Information geometry has been applied in several areas of machine learning, including:

- Natural gradient descent algorithms
- Variational inference
- Deep generative models
- Statistical manifold learning

By interpreting learning processes geometrically, researchers gain insights into how models evolve during training and how parameter spaces are structured.

**5.3 Functional Analysis and Kernel Methods**

Functional analysis provides the mathematical framework for studying infinite-dimensional vector spaces and operators acting on them. This area of mathematics underlies kernel-based learning algorithms and many modern statistical methods.

Kernel methods allow complex nonlinear relationships to be modeled by implicitly mapping data into high-dimensional feature spaces. Instead of computing coordinates in this space directly, kernel functions measure similarity between data points.

A typical kernel function can be expressed as

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

where $\phi(x)$represents a transformation of the input data into a higher-dimensional feature space.

Kernel-based algorithms include:

- Support Vector Machines (SVMs)
- Kernel Principal Component Analysis (KPCA)
- Gaussian Process Regression

These techniques enable powerful nonlinear modelling while maintaining strong theoretical guarantees derived from functional analysis.

**5.4 Dynamical Systems Theory**

Dynamical systems theory provides mathematical tools for analyzing systems that evolve over time. Many AI models, particularly recurrent neural networks (RNNs) and continuous-time neural networks, can be interpreted as dynamical systems.

A dynamical system can be described by a differential equation

$$\frac{dx}{dt} = f(x,t)$$

where $x$represents the system state and $f$defines its evolution.

In machine learning, dynamical systems theory helps researchers understand the stability and convergence of learning algorithms. For example:

- Recurrent neural networks exhibit temporal dynamics that can be studied using stability analysis.
- Neural ordinary differential equations (Neural ODEs) treat neural networks as continuous dynamical systems.
- Training dynamics of deep networks can be analyzed using gradient flow equations.

This perspective provides deeper insights into how learning processes unfold over time and how stable solutions emerge.

**5.5 Reinforcement Learning and Control Theory**

Reinforcement learning (RL) is an important branch of AI that focuses on learning optimal decision-making strategies through interaction with an environment. The mathematical foundation of RL is based on **Markov Decision Processes (MDPs)**, which model sequential decision problems under uncertainty.

An MDP consists of:

- A set of states $S$
- A set of actions $A$
- Transition probabilities $P(s' \mid s, a)$
- A reward function $R(s, a)$

The goal of an RL agent is to learn a policy $\pi(a \mid s)$that maximizes the expected cumulative reward.

 

The Bellman equation, which plays a central role in reinforcement learning, is given by

$$V(s) = \max_{a} \left[ R(s, a) + \gamma \sum_{s'} P(s' \mid s, a) V(s') \right]$$

where $V(s)$represents the value of state $s$and $\gamma$is a discount factor.

Reinforcement learning connects closely with control theory, which studies how systems can be regulated through feedback mechanisms. Modern AI applications such as robotics, autonomous vehicles, and game-playing agents rely heavily on these mathematical principles.

**5.6 Topology and Geometric Deep Learning**

Another emerging mathematical direction in AI involves the use of topology and geometric methods to analyze complex data structures. Traditional machine learning models often assume data lies in Euclidean spaces, but many real-world datasets possess non-Euclidean structures.

Geometric deep learning extends neural networks to data defined on graphs, manifolds, and other geometric domains. Tools from spectral graph theory, differential geometry, and topology are used to design algorithms that respect these underlying structures.

For instance:

- Graph neural networks (GNNs) operate on relational data such as social networks and molecular graphs.
- Manifold learning techniques attempt to uncover low-dimensional geometric structures embedded in high-dimensional datasets.
- Topological data analysis (TDA) studies persistent features of data using concepts such as homology and simplicial complexes.

These mathematical approaches enable AI systems to process structured and relational data more effectively.

**5.7 Optimal Transport Theory**

Optimal transport theory has recently gained attention in machine learning due to its ability to compare and transform probability distributions. Originally developed in the context of mass transportation problems, it provides a rigorous framework for measuring distances between distributions.

The Wasserstein distance, a key concept in optimal transport, is widely used in generative modelling. In particular, Wasserstein Generative Adversarial Networks (WGANs) employ this distance metric to improve training stability compared to traditional GANs.

Optimal transport methods are also used in:

- Domain adaptation
- Distribution alignment
- Generative modelling
- Robust machine learning

These applications highlight how theoretical mathematics can directly influence the development of modern AI algorithms.

## 6. The Symbiotic Relationship: Mathematics Enabling AI and AI Driving Mathematical Research

The interplay between mathematics and AI is bidirectional. While mathematics provides the foundation, AI's computational demands inspire new mathematical frameworks.

- **New Mathematical Problems:** The need for interpretable and robust models fosters new theoretical challenges—such as explainability and fairness in algorithms.
- **Rediscovery of Classical Fields:** The rise of geometric deep learning has revitalized interest in graph theory, topology, and differential geometry.
- **Optimal Transport Theory:** Once a pure mathematical topic, it now underpins modern generative models (e.g., diffusion models, GANs) by defining how probability distributions can be optimally transformed.
- **Attention Mechanisms:** The mathematical formalization of attention in transformers has led to advances in sparse matrix theory and combinatorial optimization.
- **Computational Mathematics:** Handling massive datasets requires efficient numerical methods, encouraging progress in matrix factorization, approximation, and high-dimensional calculus.

## 7. Challenges and Future Directions

Despite progress, key challenges are

- **Interpretability:** Deep learning models often act as "black boxes." Mathematical frameworks for explainability remain underdeveloped.
- **Robustness and Generalization:** Understanding model stability and adversarial resistance demands advanced mathematical analysis.
- **Causality:** Shifting from correlation to causal inference is a frontier challenge requiring sophisticated probabilistic modelling.
- **Quantum Mathematics:** Quantum computing introduces new paradigms for optimization and probabilistic amplitude estimation, potentially revolutionizing AI learning.
- **Ethical and Fair AI:** Mathematical bias detection and statistical fairness metrics are vital for trustworthy AI systems.
- **Theoretical Guarantees:** Mathematical proofs of safety, stability, and performance are increasingly essential in high-stakes AI applications.

**Summary Table**

| Mathematical Field | AI/ML Role | Example Application |
|---|---|---|
| Linear Algebra | Data representation, feature extraction | CNNs, PCA |
| Calculus | Optimization and learning dynamics | Gradient Descent, Backpropagation |
| Probability & Statistics | Uncertainty modelling, inference | Bayesian Networks, Naive Bayes |
| Discrete Mathematics | Logical reasoning, structured data | Knowledge Graphs, Pathfinding |
| Optimization Theory | Efficient convergence, robustness | Adam, Convex Optimization |
| Information Theory | Decision metrics, generative models | Decision Trees, VAEs |
| Geometry & Topology | Manifold learning, structure discovery | Geometric Deep Learning |

## 8. Conclusion

Mathematics is not merely an auxiliary tool for Artificial Intelligence and Machine Learning—it is their language, architecture, and engine. From the algebraic representation of data to the probabilistic reasoning of decision-making, every layer of AI depends on mathematical precision.

As AI evolves, the need for deeper mathematical literacy will only intensify. The partnership between mathematics and AI is inherently symbiotic: mathematical innovation propels AI forward, while AI's complexity fuels new mathematical discoveries. Bridging these disciplines is not only essential for technological progress but also for building interpretable, ethical, and human-aligned intelligence.

Mathematics remains the invisible architect of artificial intelligence—its logic, its structure, and its conscience.

## References

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2018.

[2] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," *Nature*, vol. 594, no. 7863, pp. 345–350, 2021.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[5] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[6] M. I. Jordan, "Taming the world-identity beast," *Communications of the ACM*, vol. 58, no. 6, pp. 78–87, 2015.

[7] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[8] Y. LeCun, L. Bottou, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[10] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2009.

[11] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.