Introduction to Deep Learning

by Dr. Rishikesh Yadav

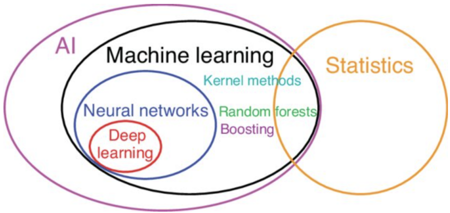
September 10, 2025

Assistant Professor, School of Mathematical and Statistical Sciences, IIT Mandi, India

Table of Contents – Introduction to Deep Learning

What are AI, ML, and DL?

Relationship: AI, ML, and DL?



Relation between AI, ML, DL[1].

- AI = Broad goal: build intelligent systems.
- ML = A way to achieve AI using data-driven learning.
- DL = A modern ML technique using deep neural networks.
- Hierarchy: AI $\supset$ ML $\supset$ DL.

Mathematics is everywhere and plays a key role in building the theoretical foundation.

Mathematicians provide the backbone: optimization, probability, high-dimensional analysis.

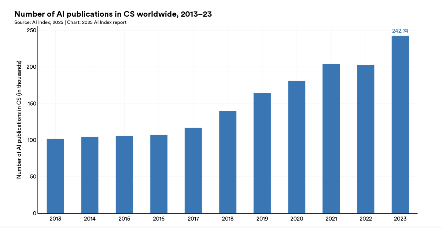
[1] WW Hseih (2022). Evolution of machine learning in environmental science—A perspective. Environmental Data Science

AI, ML, and DL Overview

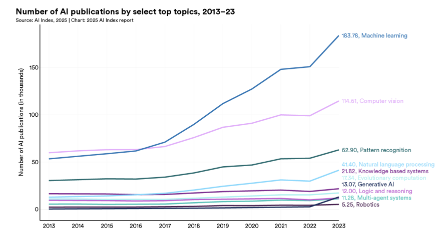
- **What is Artificial Intelligence (AI)?**
  - Artificial Intelligence (AI) is the broad field of making machines "think and act like humans".
  - Includes: reasoning, problem-solving, learning, planning.
  - AI is the umbrella term that covers both Machine Learning (ML) and Deep Learning (DL).
- **What is Machine Learning (ML)?**
  - Machine Learning (ML) is a subset of AI.
  - Focuses on algorithms that learn patterns from data without being explicitly programmed.
  - Examples:
    - Supervised learning: classification, regression.
    - Unsupervised learning: clustering, dimensionality reduction.
    - Reinforcement learning: decision-making through rewards.
- **What is Deep Learning (DL)?**
  - Deep Learning (DL) is a subset of ML.
  - Uses artificial neural networks with many layers (depth) inspired by the structure of the human brain.
  - Good at handling high-dimensional data such as images, text, speech.

Why Deep Learning is Popular Today

- Availability of big data.
- Increase in computational power (GPUs, TPUs).
- Advances in algorithms (better training methods, optimizers).
- Success in real-world applications: computer vision, natural language processing, speech recognition and many more.



Number of publications in CS, worldwise 2013-23.

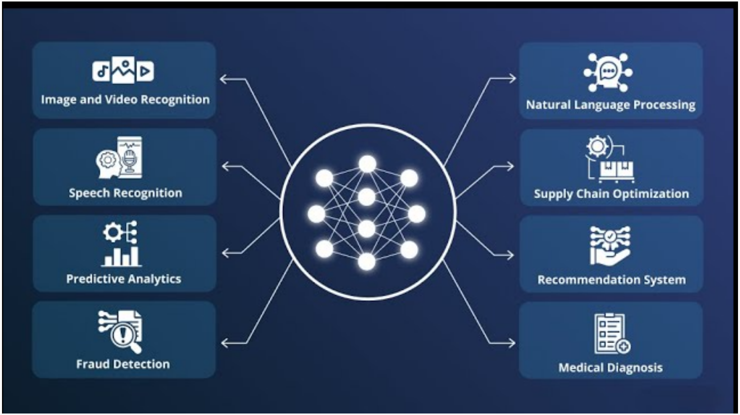


Number of publications by topics, 2013-23.

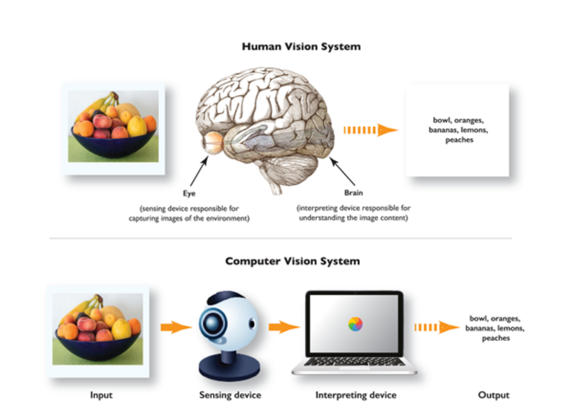
Applications of Deep Learning

Applications of Deep Learning

- Deep Learning is everywhere in our daily lives.
- Different neural network architectures are suited to different tasks.
- Applications span across science, industry, medicine, entertainment, and daily utilities.

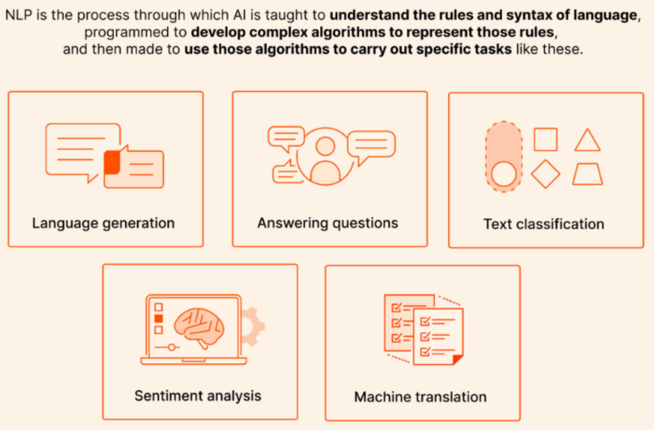


Computer Vision Applications



- Image Classification & Recognition – Face unlock on phones, airport screening. Algorithm: Convolutional Neural Networks (CNNs).
- Object Detection – Self-driving cars recognizing pedestrians and vehicles. Algorithms: YOLO, Faster R-CNN.
- Medical Imaging – Detecting tumors, X-ray analysis. CNNs, U-Net (segmentation).

Natural Language Processing (NLP)

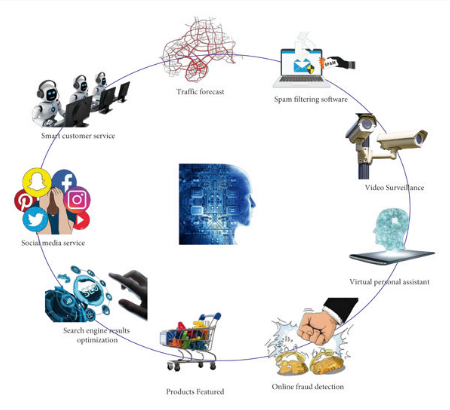


- Machine Translation – Google Translate. Algorithms: Recurrent Neural Networks (RNNs), Transformers.
- Chatbots & Assistants – Siri, Alexa, ChatGPT. Transformers, Large Language Models (LLMs).
- Text Summarization & Sentiment Analysis. BERT, GPT-family models.

Other Applications of Deep Learning

- **Speech and Audio Applications**
  - Speech Recognition – Voice typing, dictation apps. RNNs, LSTMs, Transformers (Wav2Vec).
  - Voice Assistants – Alexa, Google Assistant. End-to-End Speech Models.
  - Music Generation – AI composing songs. GANs, Sequence Models.
- **Healthcare and Life Sciences**
  - Disease Diagnosis – AI detecting diabetic retinopathy, cancers. CNNs, Ensemble Models.
  - Drug Discovery – Predicting molecule properties. Graph Neural Networks (GNNs).
  - Personalized Medicine. Deep Reinforcement Learning, Bayesian DL.
- **Science and Engineering Applications**
  - Climate Modeling & Weather Forecasting. Spatiotemporal DL Models.
  - Physics Simulations – AI accelerating fluid dynamics, particle simulations. Physics-Informed Neural Nets (PINNs).
  - Astronomy – Galaxy classification, exoplanet detection. CNNs, Autoencoders.

Deep Learning in Daily Life



- Recommendation Systems – Netflix, YouTube, Amazon. DL + Collaborative Filtering, Embedding Models.
- Finance & Banking – Fraud detection, credit scoring. RNNs, Autoencoders, Graph DL.
- Transportation – Google Maps predicting traffic. Spatiotemporal Graph Networks.
- Social Media – Instagram filters, auto-captioning. GANs, CNNs.

Historical Development of Deep Learning

Early Origins of Neural Networks (1940s–1960s)

- McCulloch & Pitts (1943): First neuron model[2].
  - Inspired by biology: proposed neurons as simple binary threshold logic units.
  - Key idea: networks of such units can compute logical functions.
- Hebbian Learning (1949): "Neurons that fire together, wire together."[3]
  - Core principle: connections between neurons strengthen when activated together.
  - Limitation: unsupervised and biologically inspired, but mathematically vague at the time.
- Rosenblatt's Perceptron (1958): First trainable neural network[4].
  - Innovation: introduced a supervised learning rule for adjusting weights.

[2] W. McCulloch, W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, 1943.

[3] D. Hebb, *The Organization of Behavior*, 1949.

[4] F. Rosenblatt, "The Perceptron, A probabilistic model for information storage and organization in the brain," *Psychological Review*, 1958.

First AI Winter (1970s)

- Minsky & Papert (1969): Limitations of Perceptrons[5].
  - Contribution: provided mathematical analysis of perceptrons.
  - Key finding: single-layer perceptrons **cannot solve non-linearly separable problems** (e.g., XOR).
  - Long-term value: highlighted the need for multi-layer architectures.
- Consequence: Loss of funding and interest in neural networks.
  - Funding agencies (DARPA, NSF): reduced support for AI research due to unmet expectations.
  - Public perception: AI seen as overhyped and underdelivering.
  - Shift in focus: research moved toward symbolic AI (logic-based systems) during this period.
- Mathematics' Role: Formal proofs of limitations.
  - Provided clarity: precise theorems about linear separability and perceptron capabilities.
  - Influenced theory: laid groundwork for computational learning theory.
  - But: reinforced skepticism about neural networks' practical usefulness at the time.

[5] M. Minsky, S. Papert, *Perceptrons: An Introduction to Computational Geometry*, 1969.

Backpropagation Revolution (1980s)

- Werbos (1974): Introduction of Backpropagation[6].
  - Core idea: used the **chain rule of calculus** to efficiently compute gradients in multi-layer networks.
  - Importance: allowed training of deeper models compared to single-layer perceptrons.
  - Limitation: remained largely unnoticed in the 1970s, as computing power and datasets were limited.
- Rumelhart, Hinton & Williams (1986): Popularization of Backpropagation[7].
  - Breakthrough: demonstrated successful training of multi-layer NNs.
  - Challenge: still slow to train on large, real-world datasets of the time.
- Enabled Multilayer Neural Networks
  - Overcame: limitations of single-layer perceptrons (e.g., XOR).
  - Led to: revival of neural networks research after the First AI Winter.
  - Issue: still vulnerable to vanishing gradients in very deep networks.

[6] P. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," PhD thesis, Harvard, 1974.
[7] D. Rumelhart, G. Hinton, R. Williams, "Learning Representations by Back-propagating Errors," *Nature*, 1986.

Second AI Winter (1987–1995)

- Neural Nets vs. Symbolic AI Debate
  - Symbolic AI: focused on logic, rules, and expert systems (knowledge-based reasoning).
  - Neural networks: emphasized learning from data via connectionist models.
  - Conflict: symbolic AI dominated research funding, while neural nets were still seen as limited.
- Hardware and Data Insufficient for Neural Nets
  - Computing power: CPUs of the era too slow for large-scale training.
  - Data availability: datasets were small and not representative of real-world complexity.
  - Result: neural nets struggled to scale beyond small toy problems.
- In this time, **Neural networks were largely sidelined in favor of statistical approaches**.

Resurgence via Statistical Learning (1990s–2000s)

- Vapnik: Support Vector Machines (SVMs)[8].
  - Core idea: maximize the margin between classes for better generalization.
  - Kernel trick: enabled solving non-linear problems by mapping inputs into higher-dimensional feature spaces.
  - Limitation: computationally expensive for very large datasets; less effective with raw high-dimensional data (e.g., images).
- Pearl: Probabilistic Graphical Models[9].
  - represented complex dependencies between variables using graphs (Bayesian networks, Markov networks).
- Jordan (1990s): Variational Methods and Bayesian Networks.
- Mathematics Foundations: Convex optimization, functional analysis, probability.
  - Convex optimization: ensured global minima for SVMs and related models.
  - Functional analysis: supported kernel methods and reproducing kernel Hilbert spaces (RKHS).

[8] V. Vapnik, *The Nature of Statistical Learning Theory*, 1995; Recent edition in 2013.

[9] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, 1988; Recent edition in 2014.

Deep Learning Boom (2006–2015)

- Hinton et al. (2006): Deep Belief Networks (DBNs)[10].
  - Core idea: stacked Restricted Boltzmann Machines trained layer-by-layer in an unsupervised manner. This reintroduced **deep learning** to the research community after the second AI winter.
- **ImageNet Breakthrough** (Krizhevsky, Sutskever, Hinton, 2012)[11].
  - AlexNet: deep convolutional neural network with ReLU activations, dropout, and GPU acceleration. Key Achievement: reduced ImageNet classification error by nearly half, a massive leap forward.
- Mikolov et al. (2013): Word2Vec Embeddings[12].
  - Innovation: represented words as dense vectors capturing semantic meaning. This revolutionized NLP, enabling analogies.
- Goodfellow et al. (2014): Generative Adversarial Networks (GANs)[13].
  - Idea: generator vs. discriminator game to learn realistic data generation.

[10] G. Hinton, S. Osindero, Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, 2006.

[11] A. Krizhevsky, I. Sutskever, G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, 2012.

[12] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv:1301.3781.

[13] I. Goodfellow et al., "Generative Adversarial Nets," NIPS, 2014.

Modern Era (2017–Present)

- Vaswani et al. (2017): Transformers[14].
  - Key idea: replaced recurrence and convolutions with **self-attention**. Around 200k citations
  - Impact: enabled efficient parallel training and captured long-range dependencies.
- Large Language Models (LLMs): GPT, BERT.
  - Advances: pretraining on massive corpora with fine-tuning or in-context learning.
  - Impact: revolutionized NLP, coding, reasoning, and multimodal AI (e.g., GPT-4, Gemini, Claude).
- Multimodal & Reinforcement Learning Advances.
  - AlphaGo/AlphaZero (2016-2018): deep RL + search, breakthrough in strategic games.
  - Diffusion models (2020+): new generative paradigm for images (e.g., Stable Diffusion, DALL·E).
  - Multimodal AI: unifying vision, text, speech (e.g., CLIP, Flamingo, GPT-4V).

[14] A. Vaswani et al., "Attention is All You Need," NeurIPS, 2017.

The Big 3 Heroes of Deep Learning

(a) Yann LeCun – Chief AI Scientist at Meta, Prof. at NYU
Pioneer of CNNs (Convolutional Neural Networks)[a]

[a] Y. LeCun et al., "Gradient-based learning applied to document recognition," Proc. IEEE, 1998.

(b) Geoffrey Hinton – Prof. Emeritus at Toronto, Adviser at Vector Inst.
Backpropagation, deep belief networks[a]
Received 2024 Nobel Prize in Physics.

[a] Rumelhart, Hinton, Williams, "Learning Representations by Back-Propagating Errors," Nature, 1986.

(c) Yoshua Bengio – Prof. at Université de Montréal, Mila Institute
Foundations of deep learning, generative models

All three received the 2018 Turing Award.

Other Ground-Breaking Contributors

- Jürgen Schmidhuber – IDSIA, Switzerland
  Long Short-Term Memory (LSTM)[15]
- Sepp Hochreiter – JKU Linz, Austria
  Co-inventor of LSTM, now driving AI in bioinformatics.
- Ian Goodfellow – Apple, ex-OpenAI/Google
  Invented GANs (Generative Adversarial Networks)[16]
- Andrew Ng – Stanford Univ., DeepLearning.AI
  Massive impact via online education and applied deep learning.

[15] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997.
[16] Goodfellow et al., "Generative Adversarial Nets," NeurIPS, 2014.

Historical Foundations

- Frank Rosenblatt – Inventor of the *Perceptron* (1958).
- Marvin Minsky & Seymour Papert – Showed limitations of perceptrons (1969).
- Paul Werbos – First described *backpropagation* (1974, PhD thesis).

Required Tools to Learn DL

Mathematical Tools

- Linear Algebra – vectors, matrices, eigenvalues.
  - Neural nets = compositions of linear maps + nonlinearities.
  - Key ops: dot products, SVD, eigendecomposition.
- Probability & Statistics – distributions, Bayesian methods.
  - Uncertainty modeling, Bayesian priors, likelihood.
  - Applications: dropout, variational inference.
- Calculus – differentiation, gradients, optimization.
  - Backpropagation = repeated chain rule.
  - Gradient descent = discrete dynamical system.
- Optimization Theory – convex analysis, gradient descent.
  - Convex optimization ensures global minima (e.g., SVMs).
  - Non-convex DL optimization studied via dynamical systems.

These are the mathematical backbone of Deep Learning. No doubt you all know much more than that.

Programming Foundations

- Programming Language: Python – most widely used in AI/ML/DL. (we will use throughput)
  - Reason: clean syntax, huge ecosystem, numerical + scientific packages.
  - Alternative: R (good for statistics, but less popular for ML/DL).
- Mathematical Libraries: NumPy, SciPy, SymPy.
  - NumPy: linear algebra, tensors (good enough for basic use).
  - SciPy: optimization, special functions.
  - SymPy: symbolic algebra (useful for derivations).
- Data Handling: Pandas, SQL, basic wrangling.
  - Pandas: tabular data (like matrices for applied mathematicians).
  - SQL: structured data queries.

For documentation: Jupyter Notebooks (we will use throughput)

Deep Learning Frameworks

- TensorFlow (Google) – scalable for research & production. (This is the one we will use throughout the reading group)
- PyTorch (Meta) – flexible, Pythonic, favored by researchers.
- JAX (Google) – high-performance, automatic differentiation.
- Keras: simple API for beginners.
- Why? Provide ready-made building blocks (autograd, layers, optimizers).

Computational Tools

- GPUs & TPUs – parallel linear algebra operations.
  - NVIDIA CUDA GPUs dominate current DL.
  - TPUs (Google): specialized for tensor ops.
- Cloud Platforms:
  - Google Colab (This is the one we will use throughout the reading group)
  - AWS
  - Azure
  - Kaggle.
- Local Workstations: CUDA-enabled NVIDIA GPUs.
- Why? **Training DL models without GPUs is prohibitively slow**.

Dont worry! You will get all of this in your Google Colab, mostly for free, but for limited uses

Our Plans for the Reading Group

Resources and Code

- All slides, code, and reading material will be available on GitHub:

https://github.com/yadavrishikesh/Deep-Learning-Slides-Code

- This will include organized topics every week with structured notes, code, and references.