1. Intro to Deep Learning

Que. What is Deep Learning?

Deep learning is a subset of machine learning, under the broader umbrella of artificial intelligence that is inspired by the structure of human brain.

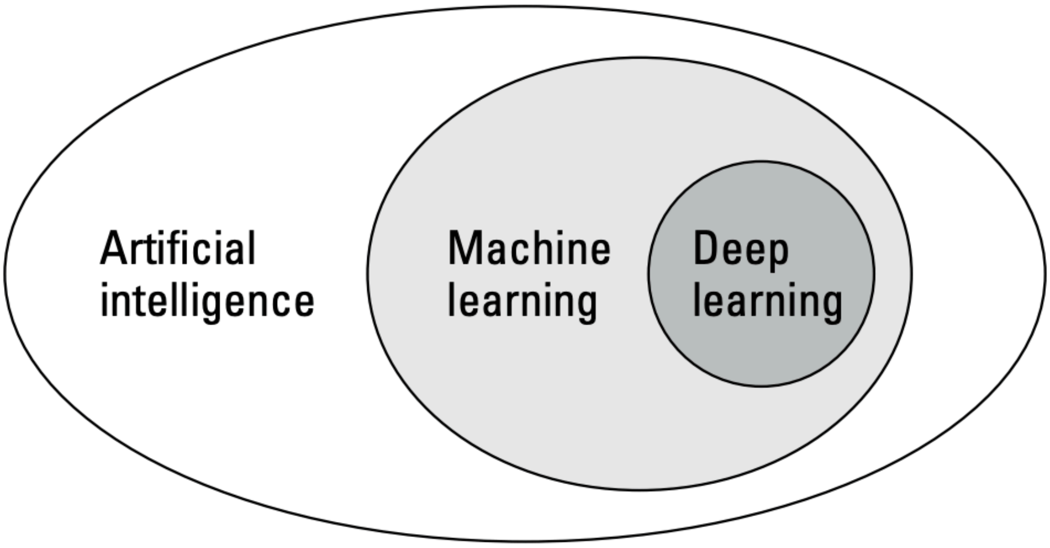


FIGURE 1-1: Deep learning, a subset of a subset of AI.

Deep Learning algorithms attempt to draw similar conclusions as human would by continually analyzing the data with a given logical structure called neural network.

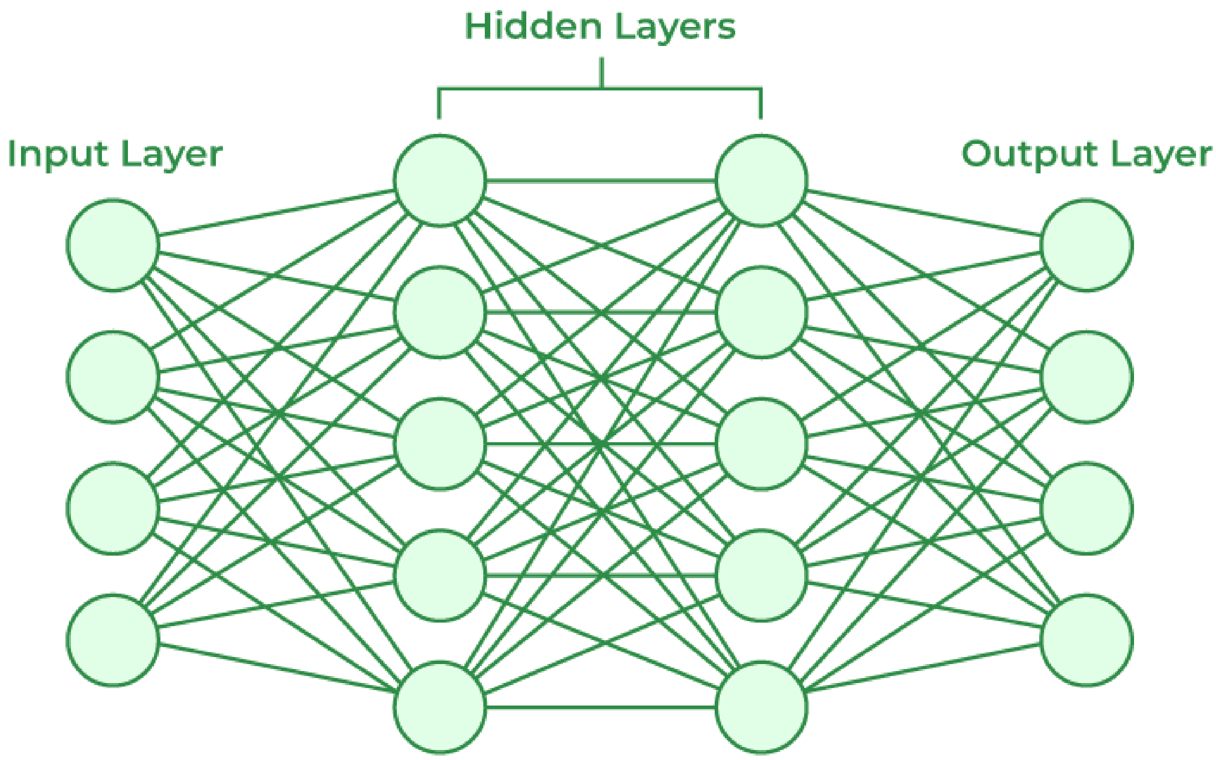
What is Neural Network ?

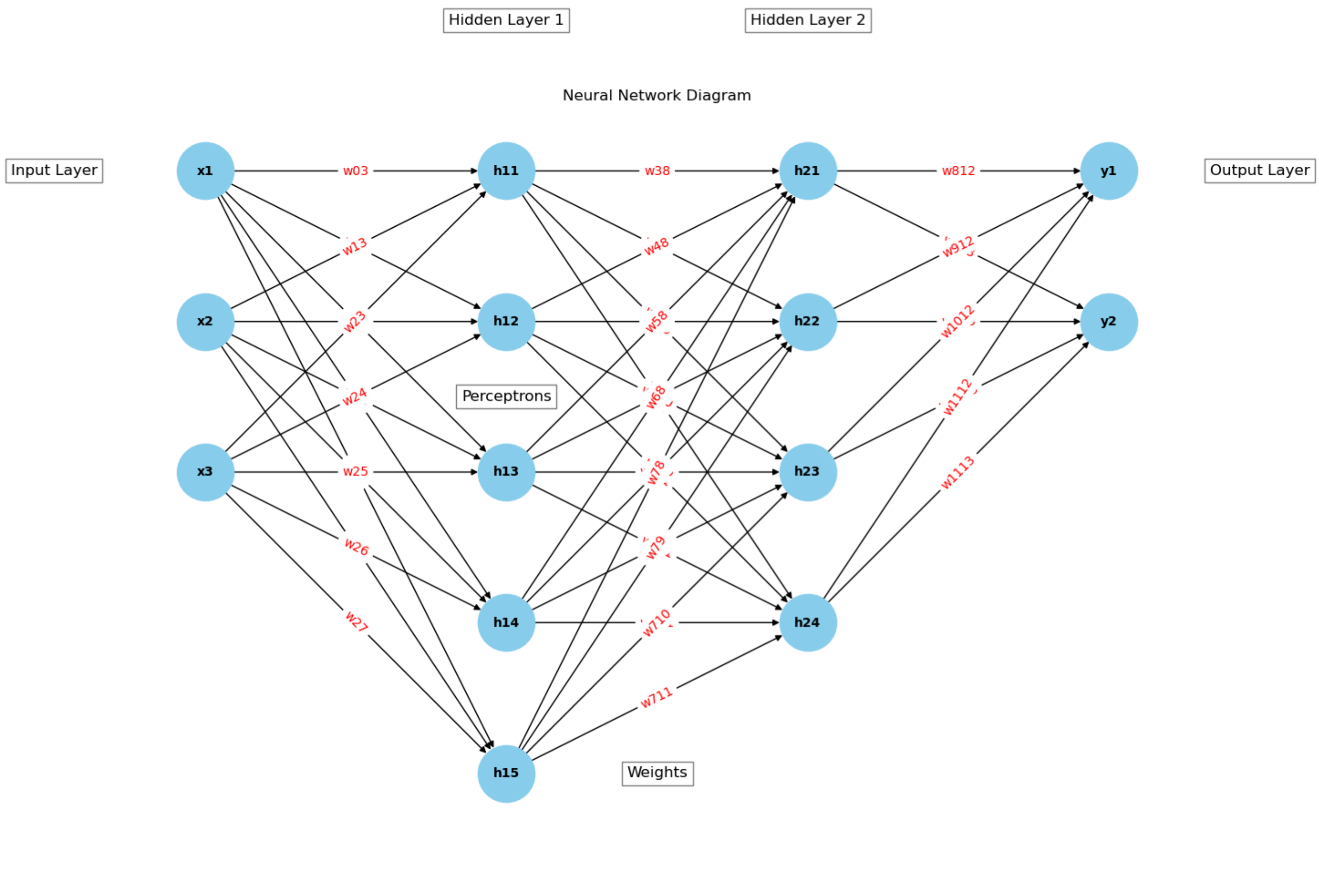
A neural network is a type of computer system designed to learn from data and make decisions, similar to how the human brain works. It's made up of layers of interconnected units called neurons. These neurons are inspired by the way our brain cells (neurons) work.

Key Components:

- **Neurons**: The basic units of a neural network. Each neuron takes in information, processes it, and passes it on to the next layer.
- **Layers**: Neural networks are organized in layers. The input layer takes in data, hidden layers process the data, and the output layer produces the final result.
- **Connections and Weights**: Neurons are connected to each other, and each connection has a weight that adjusts as the network learns. These weights determine the strength of the signal between neurons.

- **Learning**: Neural networks learn by adjusting the weights based on the difference between the predicted result and the actual result, a process called training. Over time, the network gets better at making accurate predictions or decisions.

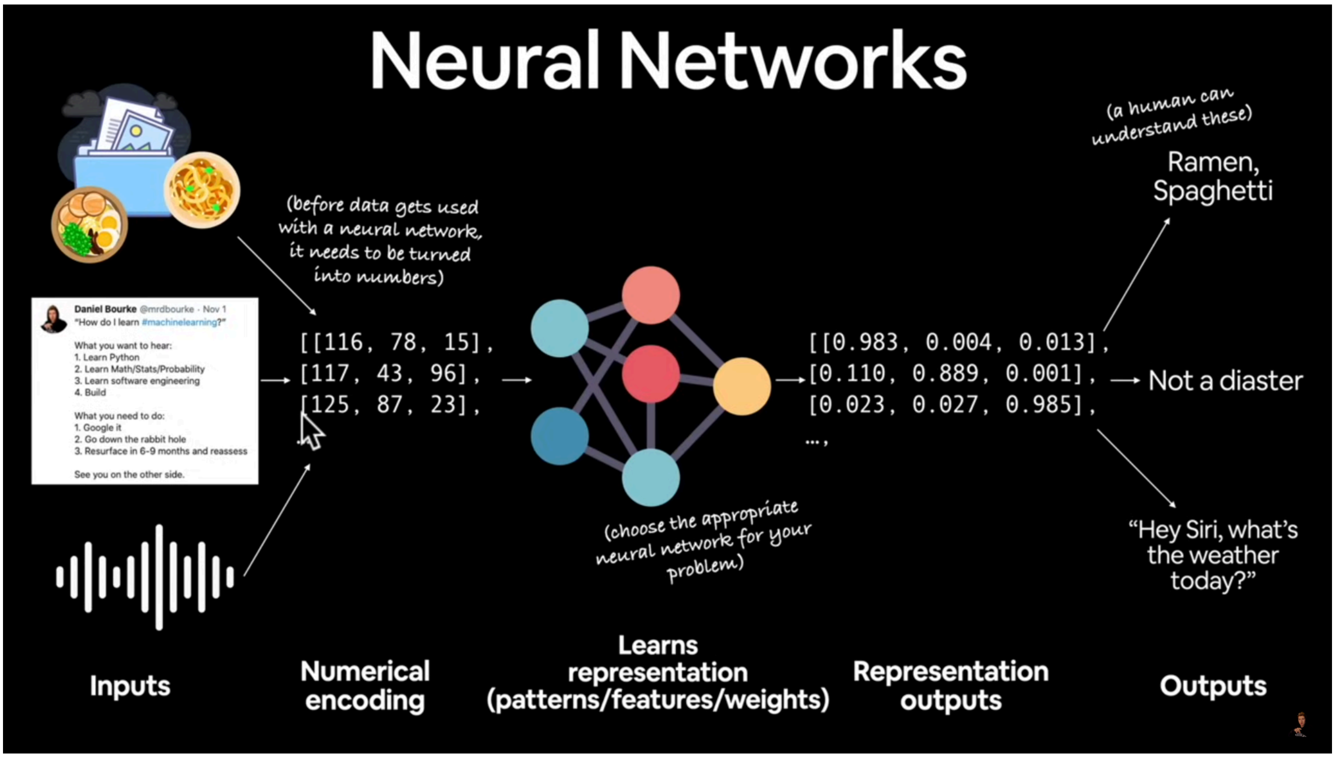




These many-layered networks prompted the move to name this field of AI research “deep learning” since algorithms now process data several layers deep in order to reach an answer.

1. ANN - Artificial NN (simplest NN)
2. CNN - Convolutional NN (Used in image classification)
3. RNN - Recurrent NN (for Speech / text)
4. GAN - Generative adversarial Network (to develop new conclusions from the input) ’ and many more…

   A basic idea of how a Nural Network work is



Why Deep Learning ?

- Applicability - Compared to traditional machine learning methods, however, deep learning has the potential to solve much more complicated problems and intuit new solutions. It requires less upfront work with feature engineering and dataset labelling. e.g., Bioinformatics & Drug Discovery, Natural Language & Speech, Face & Object Recognition etc.
- Performance - Deep learning shines in very complex domains where solutions aren’t simple. And performance is state-of-the-art even more than human intelligence.

When Deep Learning ?

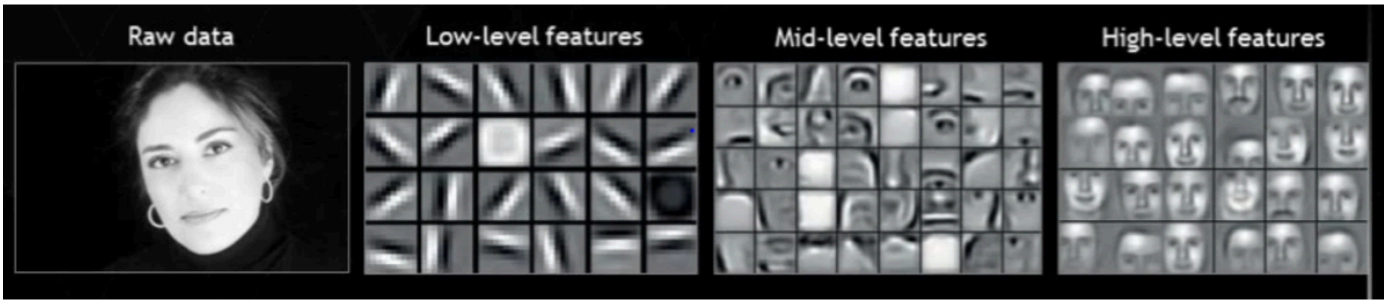
- Problem with long list of rules.
- Continually changing enviornments.
- Discovering insights within large collection of data.

When not Deep Learning ?

- When you need explainability.
- When the traditional approach is better option.
- When you don't have much data.

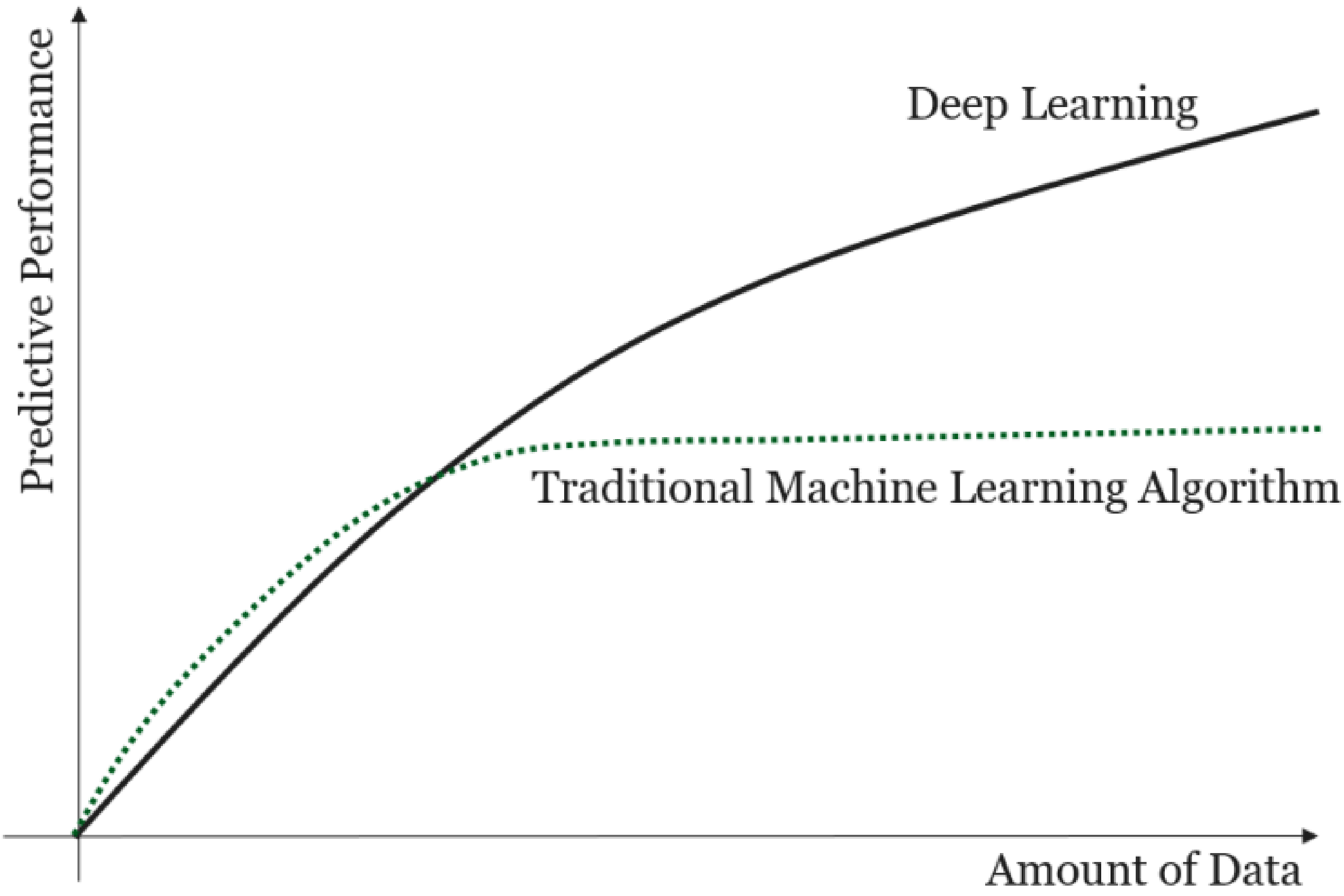
How Deep Learning works ?

- Deep learning is a part of a broader family of machine learning based on Artificial Neural Network with representation learning.
- Deep learning algorithms uses multiple layers to progressively extract higher-level features from the raw input. e.g., in image processing, lower layer may identify edges, while higher layer may identify the concepts relevant to humans such as digits, faces, words or letters. We can also say that, starting layers detect primitive features like edges, but the deep layer detect complex features too like face.



ML vs DL

- Process method: Machine learning uses statistical and mathematical techniques to learn the relationship between Input and Output, whether Deep learning uses Neural Network to achieve the same.
- At its core, machine learning generally is about pattern recognition. Once a computer learns to recognize patterns in a given domain, it can classify new data that it receives based on what it learned from past trials. In contrast, a deep learning model is an end-to-end problem solver. Using layers of networked nodes, the deep learning model does the feature extraction and identification of multiple patterns all at once. This makes deep learning useful for much more complicated problems.
- Data Dependency : Deep learning requires huge amount of data in comparison of Machine Learning. As per below graph, for less data, ML has more accuracy than Deep learning. But with more data, ML models accuracy gets constant but Deep Learning models increases.



- Hardware Dependency : Deep learning models needs GPU to run faster, but ML models can run on CPU easily. Modern deep-learning applications would not be possible if there were no GPUs! A GPU parallelizes operations on data because of its large number of cores. So, costly hardware needed for Deep learning.
- Training time : DL models training time is very high as compared to ML models. Sometimes DL model takes a whole week and months to train. ML models hardly need an hour to get trained. In contrary, DL models prediction time is less than some ML models prediction time like KNN.
- Feature Selection : Deep learning uses representation learning, no manual feature extraction is needed. But, with ML, feature extraction is done manually.
- Interpretability : Deep learning models are not interpretable. Whatever happen in the hidden layers is like a black box. Whereas ML models interpretability is high.

Why Deep Learning is so famous only now after 2012?

Research on Deep learning started from 1960s during the time of alan turing, but it is only got famous now because:

(i) Datasets : Deep learning is data hungary. After smartphone revolution of 2010 and high speed internet in cheap rate, huge amount of data getting generated by humans. But this was unlabelled data. To convert it into labelled one, companies like Microsoft and Google spent money and labelled these datasets for research and made them open source. These open source datasets made possible this so much research in the field of Deep learning because to make such datasets isn't everyone's cup of tea. e.g., MS-COCO : ~25 GB (Compressed) : 330K images, 80 object categories, 5 captions per image, 250,000 people with key points

ImageNet: ~150GB : 1,500,000 images
Free Music Archive (FMA) : ~1000 GB : 100,000 HQ audio music tracks
google AudioSet - A sound vocabulary and dataset : 2,084,320 human-labeled 10-second sound clips

(ii) Hardware : Moore's Law refers to Gordon Moore's observation that the number of transistors on a single chip would double every two years at minimal costs.

Deep Learning -> Lot of data -> matrix operations -> i.e., parallel computing

- GPUs can perform multiple, simultaneous computations. This enables the distribution of training processes and can significantly speed deep learning operations. With GPUs, you can accumulate many cores that use fewer resources without sacrificing efficiency or power.
  Once NVIDIA introduced CUDA framework in 2007, several deep learning frameworks were developed, such as Theano and TensorFlow. These frameworks have made GPU processing accessible to modern deep learning implementations.
  GPUs boost computations by an impressive 15-20x compared to CPUs,

There are also custom made hardwares for deep learning, like

- There are also FPGAs which are theoretically better than GPUs to deploy Deep Learning models, but they are complex and costly. Microsoft is already using FPGAs for Bing search ranking,
- ASICs (Application-Specific Integrated Circuits) are computer chips that combine several different circuits all on one chip allowing it to be custom programmed. And because all of the circuits are on the same chip, instead of spread out over different chips, an ASIC can execute that task much faster than a less focused circuit. e.g.,
  Google TPU (Tensors processing Unit) for for neural network machine learning,
  Edge TPU (to run deep learning model on smart devices like smartwatches, drones),
  NPU (Neural Processing Unit) to run a deep learning model on smartphones etc.

These custom made hardwares made deep learning research very fast.

(iii) Frameworks : To train deep learning models is really difficult. Also, writing a code from scratch to run a DL model is extremely hard. So, we need frameworks for DL just like scikit-learn for ML. There are two main frameworks used in Deep learning:

- Tensorflow + keras by google (History)
- PyTorch + caffe2 by Meta : PyTorch has become a preferred machine learning framework for many AI researchers due to its research flexibility. Over half of Facebook AI projects run on PyTorch. (History)

(iv) Deep learning architecture : Deep learning architectures are at the leading edge of transforming artificial intelligence (AI) by introducing innovative capabilities. These advanced structures, inspired by the human brain's neural networks, empower machines to comprehend, learn, and make independent decisions.

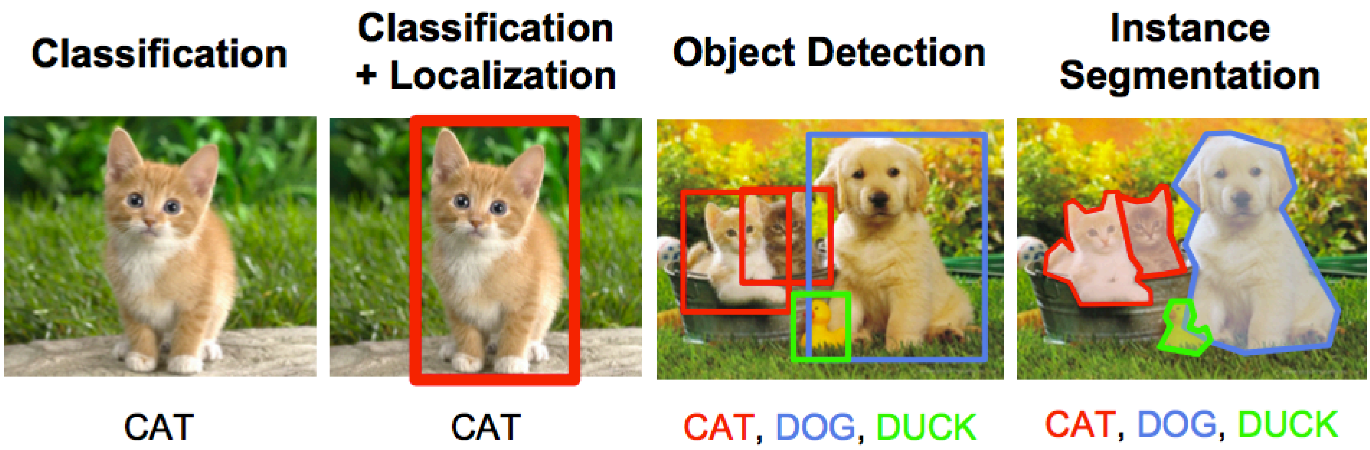
Different deep learning architectures exist because different types of problems require different approaches for the most effective solution. Each architecture is designed to handle specific types of data, tasks, or computational challenges, making it better suited for certain applications than others.

Here's why we have a variety of deep learning architectures:

> **1. Different Data Types:**

- **Images** : For tasks involving image data, such as object recognition or image classification, Convolutional Neural Networks (CNNs) are the go-to architecture.
- **Sequences (Text, Speech, Time-Series)** : When working with sequential data, such as text, speech, or time-series data, architectures like Recurrent Neural Networks (RNNs) or Long Short-Term Memory networks (LSTMs) are used.
- **Graphs** : For tasks that involve data structured as graphs (like social networks, molecules, etc.), Graph Neural Networks (GNNs) are more appropriate.

> **2. Different Tasks:**



- **Classification** : When the task is to categorize data into predefined classes (e.g., determining if an image contains a cat or a dog), architectures like CNNs and feedforward neural networks are commonly used.
- **Object Detection** : For detecting and locating objects within an image, architectures like YOLO (You Only Look Once) or R-CNN (Region-based Convolutional Neural Networks) are specialized.
- **Language Modeling and Translation** : For understanding and generating human language, Transformer architectures (like GPT or BERT) are used.
- **Generative Tasks** : For tasks that involve generating new data, such as creating images or text, Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs) are used.

> **3. Scalability and Efficiency:**

- **Efficiency** : Some architectures are designed to be more computationally efficient, making them suitable for real-time applications or deployment on devices with limited resources (like mobile

phones). For example, MobileNets are lightweight CNN architectures optimized for mobile and embedded vision applications.
- **Scalability** : Some tasks require processing massive datasets or very large models. Architectures like Transformers have been scaled up to create large models like GPT-3, which can handle a wide range of language tasks with a single model.

> **Examples of Popular Architectures** :

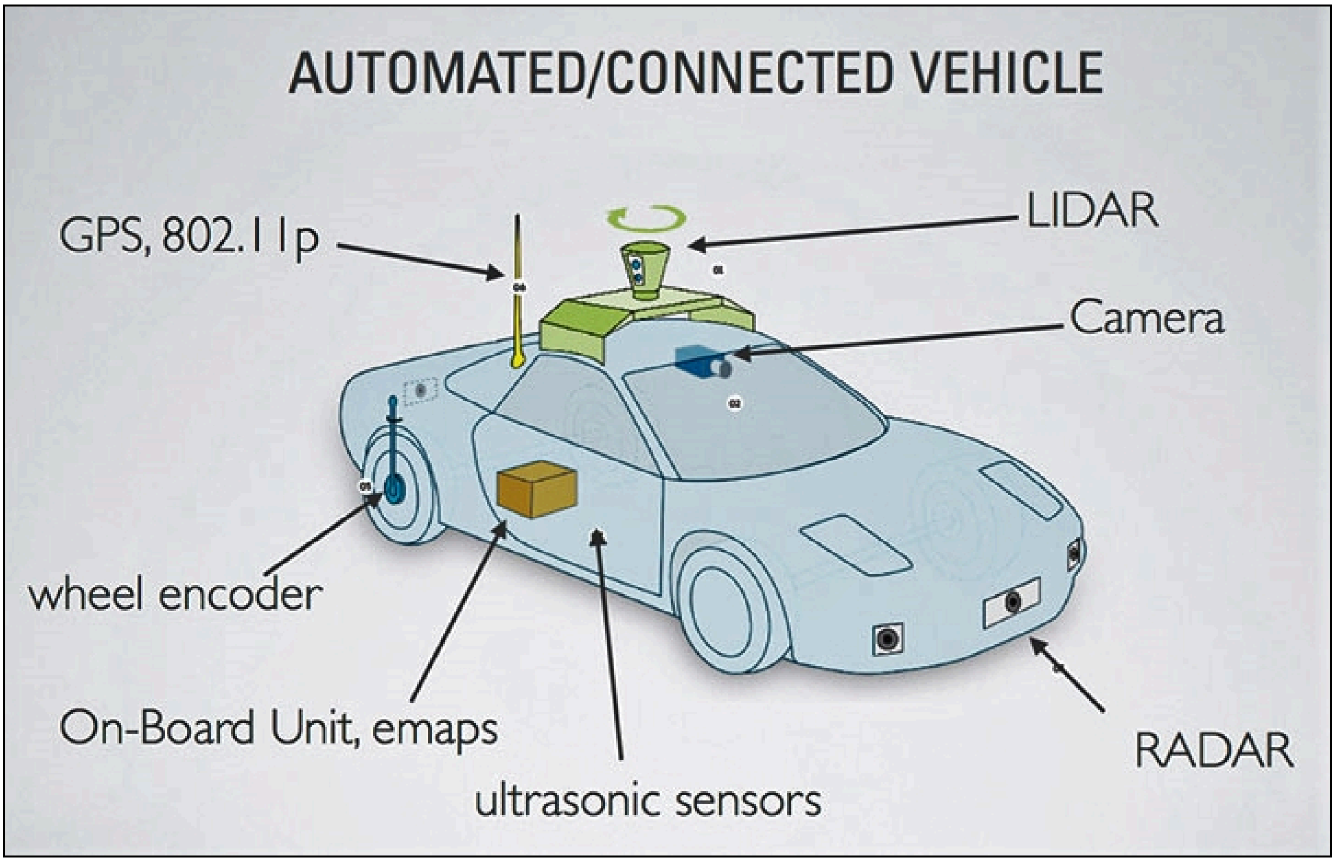
- CNNs: Best for image processing tasks.
- RNNs and LSTMs: Ideal for sequence-based tasks like text and speech recognition.
- Transformers: State-of-the-art for natural language processing and tasks requiring long-range dependencies.
- GANs: Excellent for generating new data, like realistic images or videos.
- GNNs: Best for data represented as graphs, such as social networks or molecular structures.

(v) Open Source community : The open-source community has played a crucial role in the development and advancement of deep learning. e.g.,

- development of Deep Learning Frameworks like TensorFlow, PyTorch
- Sharing of Pre-Trained Models like BERT for natural language processing and ResNet for image recognition. These models save time and resources, as others can build upon them instead of starting from scratch.
- Collaboration and Knowledge Sharing : Platforms like GitHub allow developers and researchers to share code, collaborate on projects, and contribute to existing repositories. This has led to rapid iteration and improvement of deep learning tools and models.
  The community regularly shares tutorials, notebooks, and educational content, making it easier for newcomers to learn deep learning.
- Publication of Research Papers: Researchers often publish their findings in open-access formats, such as arXiv. Along with the papers, they frequently share the corresponding code and models, allowing others to replicate, validate, and build upon their work.
- Competitions and Benchmarks: Open competitions like those hosted on Kaggle and challenges like ImageNet's Large Scale Visual Recognition Challenge (ILSVRC) have driven innovation. The community collaboratively works on these problems, and the best solutions are shared publicly.
- Forums and Discussion Platforms: Online forums like Stack Overflow, Reddit, and specialized communities like the TensorFlow or PyTorch forums allow people to ask questions, share solutions, and collaborate on problems. This support network helps both beginners and experts overcome challenges more efficiently.

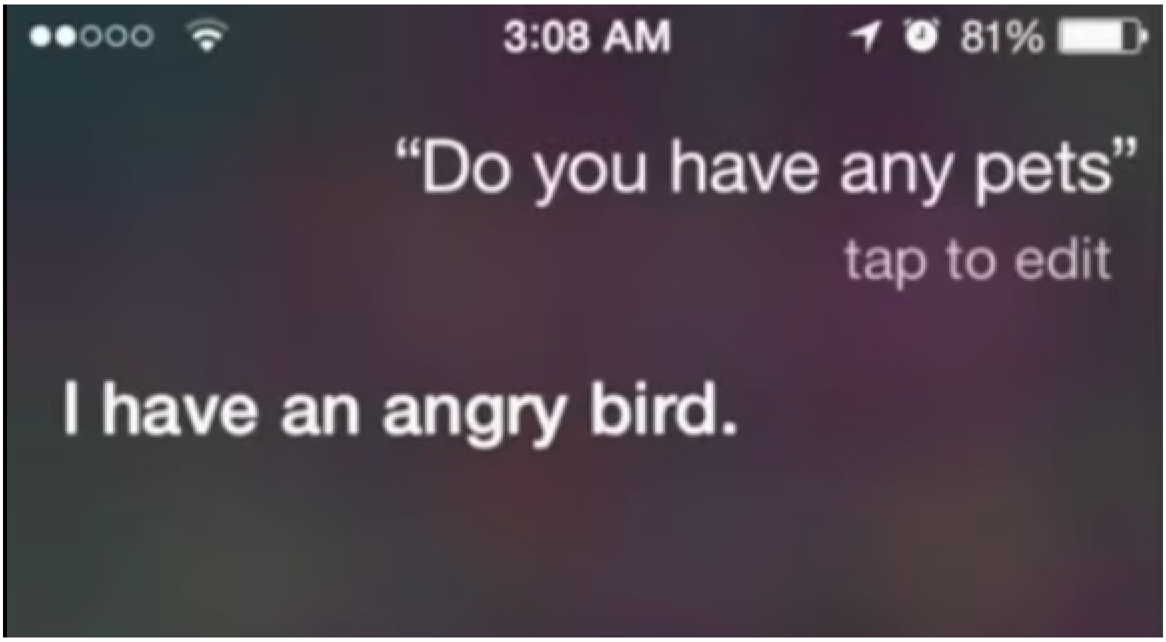
Use cases of Deep Learning

- self driving cars.



imagine you book a cab and you get is a car without driver.

- alpha go using deep reinforcement learning which won 4-5 against lee se dol. He quitted playing in 2019 saying “Even if I become the number one, there is an entity that cannot be defeated.”
- Virtual assistant :

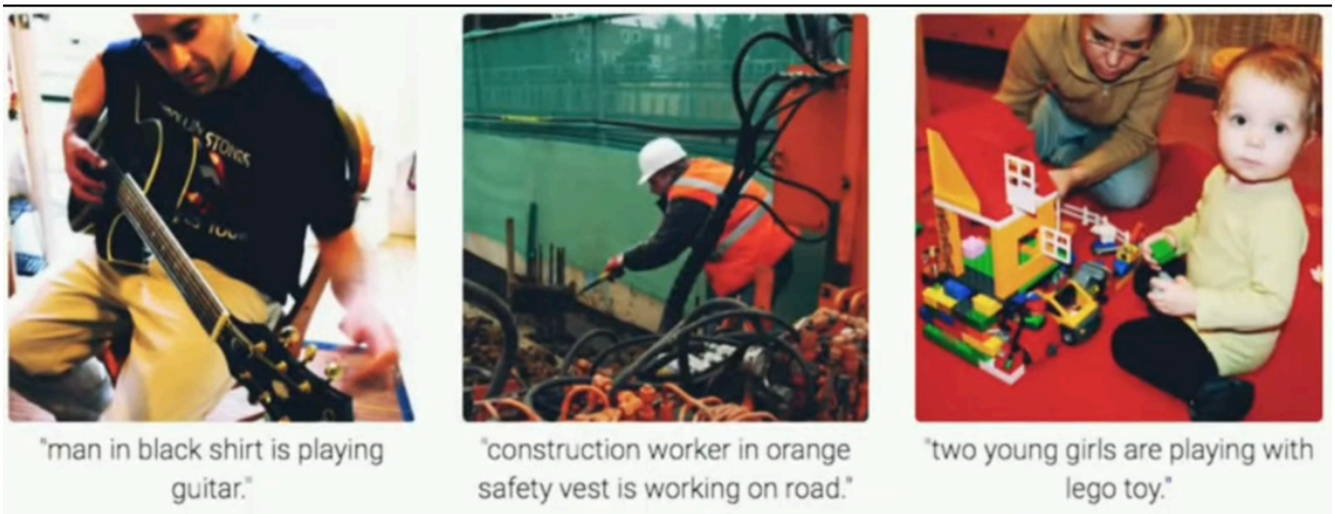


The sarcastic replies and context reading is only possible because of RNN and LSTM etc. After 2015, performance of chatbots is increased significantly.

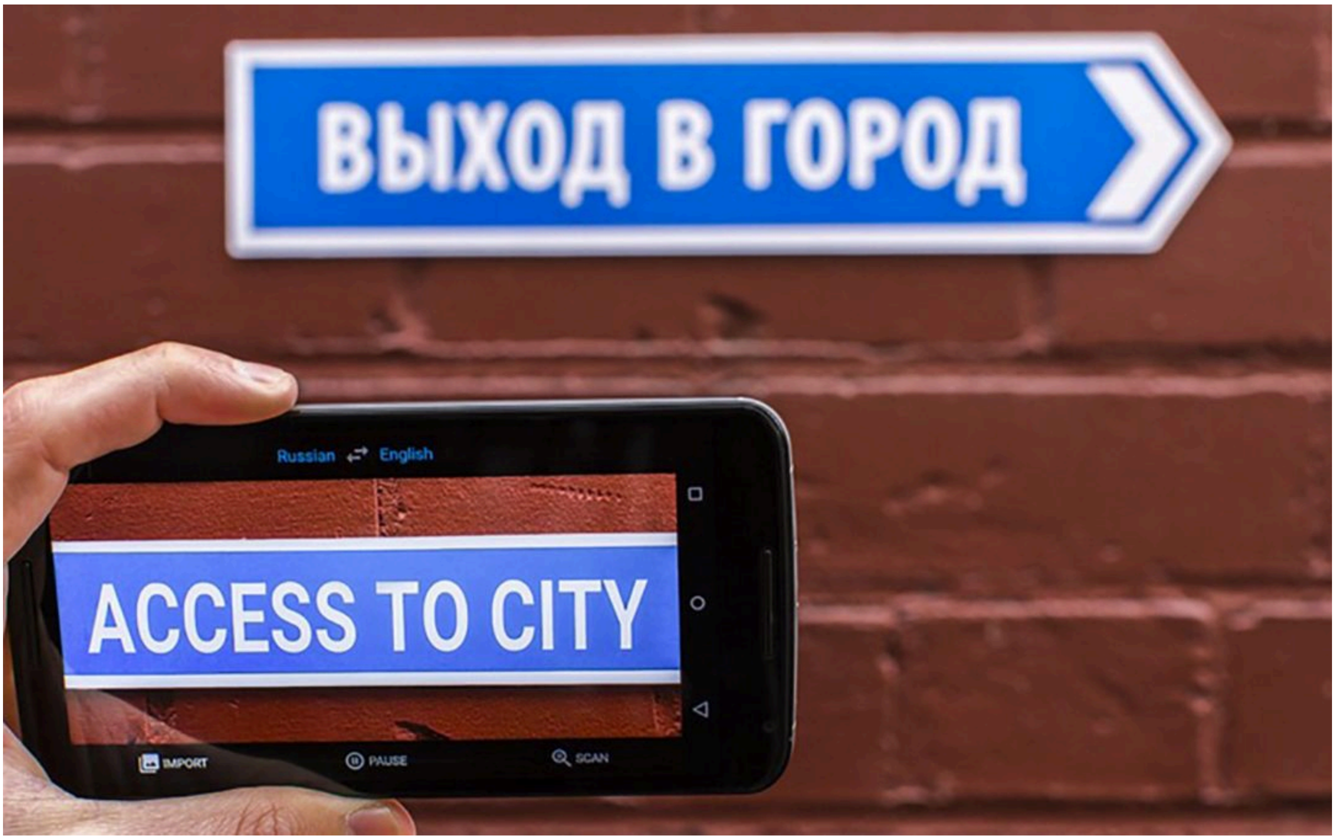
- Image colorisation of images and movies:

  using CNN with Opencv

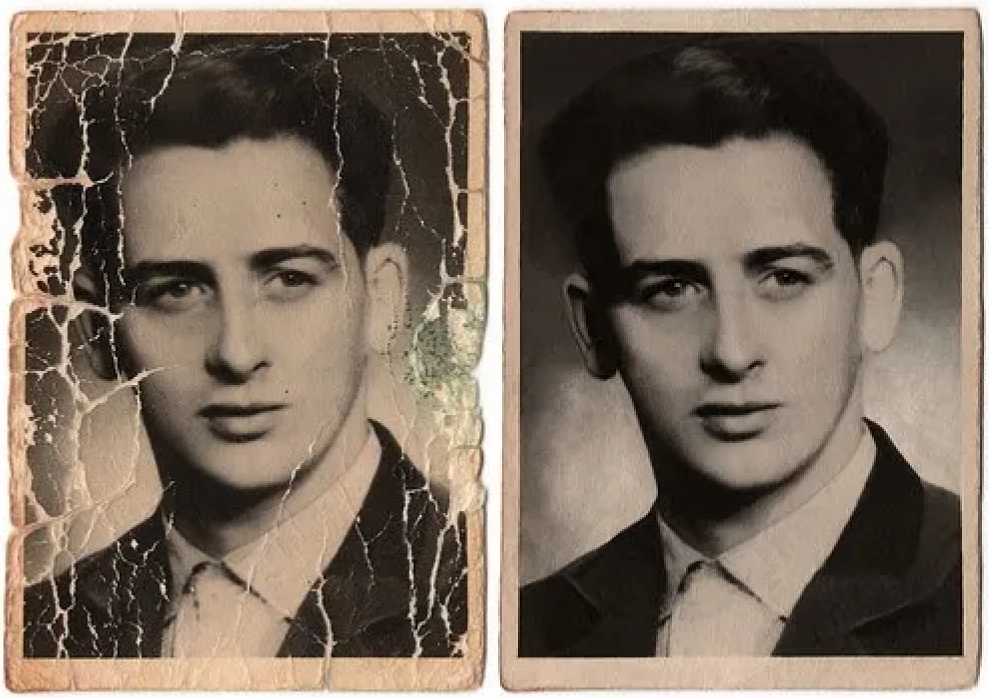
- Generating and Embedding audio in a mute video as per the scene.
- Image Caption generation :



- Real time text translation from images.

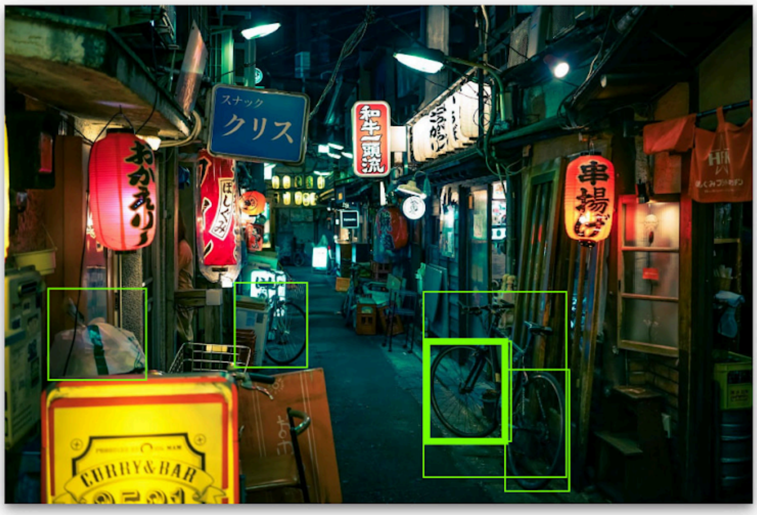


- Pixel restoration : The art of image restoration lies in finding missing information inside the image and returning it back to its original state. Deep learning-based methods have achieved remarkable success in image restoration.



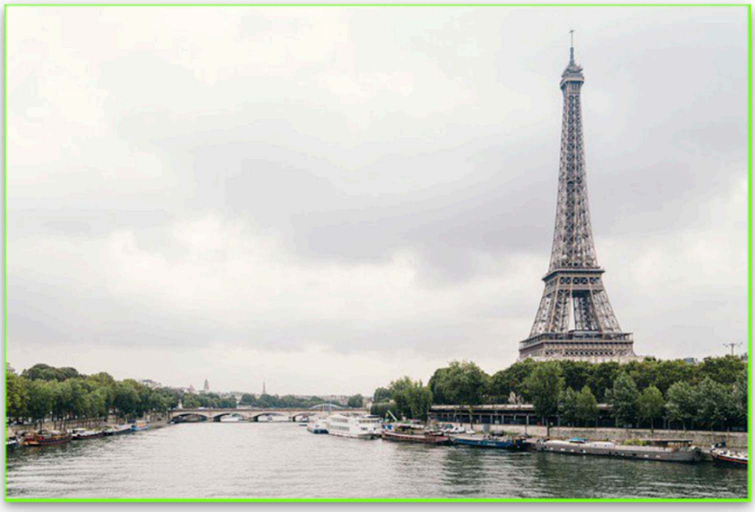
- Object detection identification: Google Photos automatically categorize all photos that have the faces of other people in it or say, **Face Grouping**.
  It also Label your photos automagically with Vision API.

Objects | Labels | Text | Properties | Safe Search

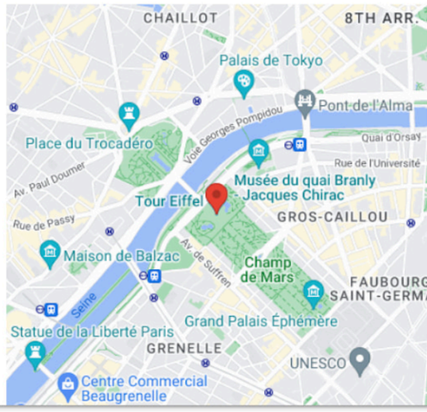


Tire	93%
Bicycle	93%
Tire	92%
Bicycle	75%
Wheel	73%

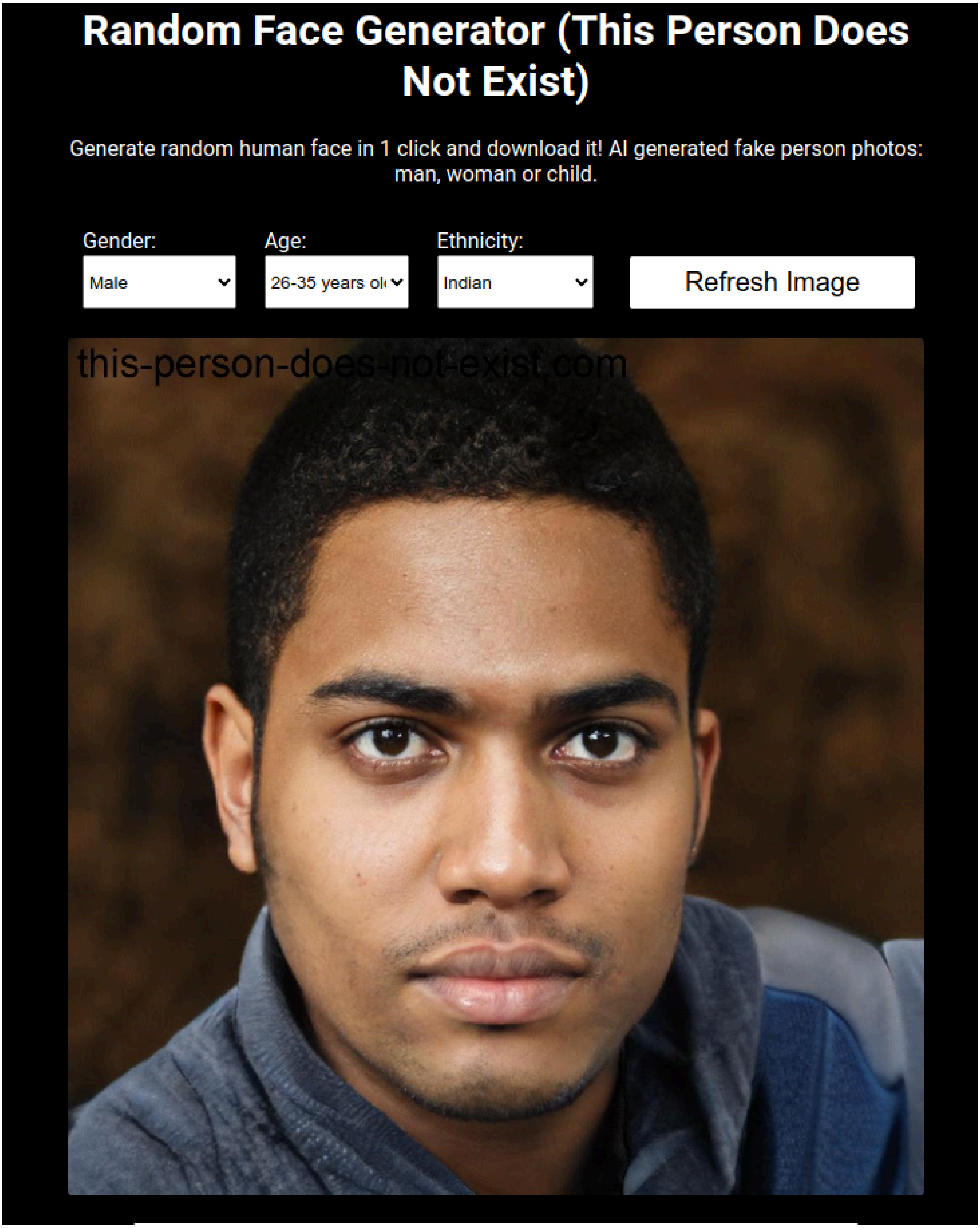
Landmarks | Objects | Labels | Text | Properties | Safe Search



Eiffel Tower 74%



- This person doesn't exist : Random face generator



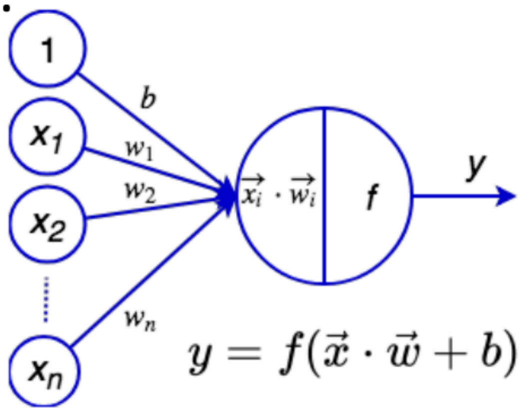
It is done using a GAN network.

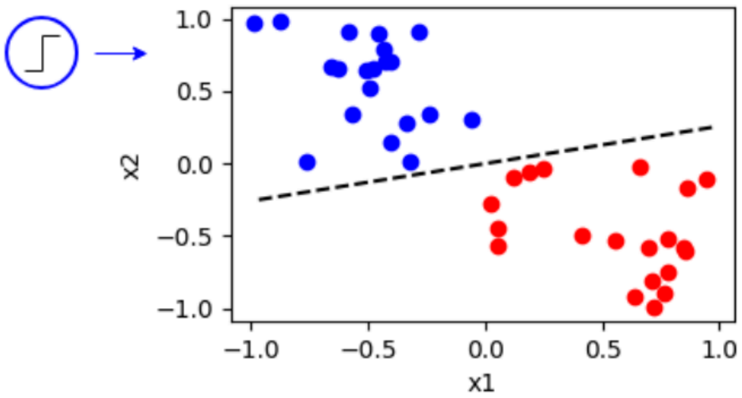
- Handwriting generator, Deep dreaming and many more…

History of Deep Learning

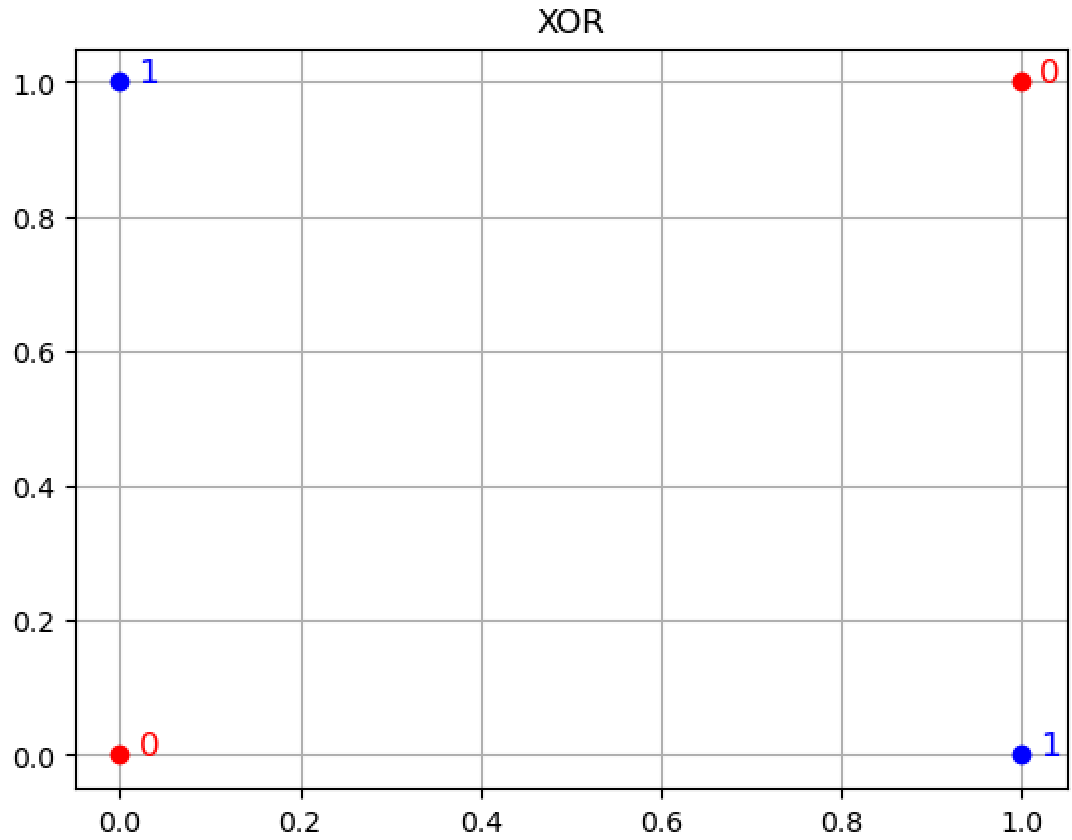
The history of deep learning is a fascinating journey that spans decades, filled with brilliant ideas, setbacks, and eventual breakthroughs.

- **1950s-1960s: The Birth of Neural Networks**
  In the early 1950s, during the dawn of artificial intelligence, scientists were inspired by the human brain's ability to learn and adapt. This led to the development of the first artificial neuron, called the Perceptron, by Frank Rosenblatt in 1958. The Perceptron was a simple model that could learn to classify inputs into two categories.





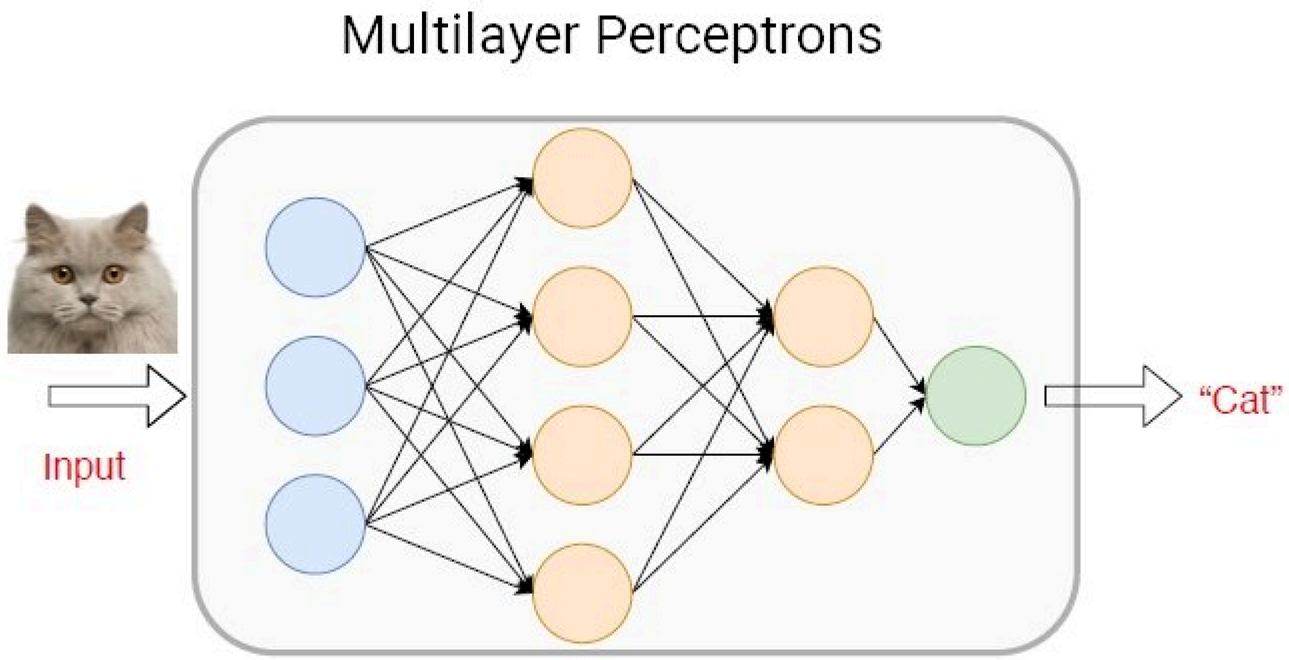
  The excitement was palpable, with hopes that machines could soon replicate human intelligence. However, the Perceptron was limited; it could only solve linearly separable problems. In 1969, Marvin Minsky and Seymour Papert published a book, "Perceptrons," showing that the Perceptron couldn't solve problems like the XOR problem (where outputs are not linearly separable). This realization led to a decline in interest and funding in neural networks, known as the "AI Winter."



- **1970s-1980s: The Quiet Years and the Rise of Backpropagation**
  Despite the setbacks, a small group of researchers kept the flame alive. In 1975, Paul Werbos developed the Backpropagation algorithm, which allowed neural networks to adjust their weights by calculating the gradient of the loss function. However, it didn't gain much attention until the mid-1980s.

In 1986, Geoffrey Hinton, David Rumelhart, and Ronald Williams popularized backpropagation, showing it could train multi-layer neural networks, which are now known as Multilayer Perceptrons (MLPs). This breakthrough renewed interest in neural networks, but the enthusiasm was still limited by the computational power and data availability of the time.

> - Single-Layer Perceptrons (without hidden layers) can only handle linearly separable data.
> - Multilayer Perceptrons (with hidden layers and non-linear activation functions) can handle non-linear data and model more complex patterns.



- **1990s: The Struggles and the "Second AI Winter"**
  The 1990s were a period of mixed feelings for neural networks. While researchers had developed better techniques, including the Convolutional Neural Network (CNN) by Yann LeCun for image recognition, the broader AI community shifted focus towards other methods like Support Vector Machines (SVMs) and decision trees. These methods often outperformed neural networks on the available datasets and were easier to train.
  The limitations of computational power and the difficulty in training deep networks led to another decline in neural network research. This period is sometimes referred to as the "Second AI Winter."
- **2000s: The Revival and the Birth of Deep Learning**
  At the turn of the millennium, the field of neural networks saw a revival, largely driven by three factors:

1. Increased computational power, particularly with the advent of GPUs (Graphics Processing Units).
2. Availability of large datasets, especially with the rise of the internet.
3. Algorithmic improvements, like better initialization techniques, activation functions (such as ReLU), and optimizers (like Adam).
   In 2006, Geoffrey Hinton, along with his students, introduced the concept of Deep Belief Networks (DBNs), a type of neural network that could be pre-trained layer by layer, making it easier to train

deep networks. Deep Belief Networks (DBNs) addresses issues with classic neural networks in deep layered networks.
For example – slow learning, becoming stuck in local minima owing to poor parameter selection, and requiring a large number of training datasets of these given input layer. This was a pivotal moment that marked the beginning of Deep Learning as we know it today.

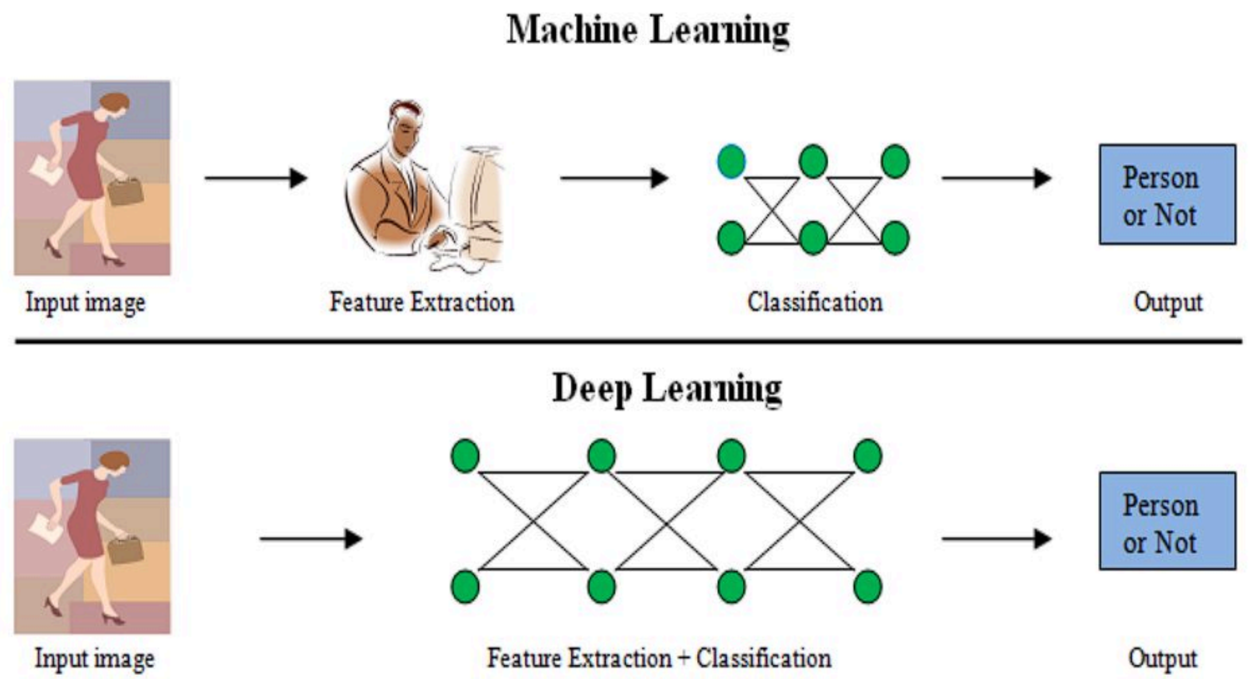
- **2010s: The Golden Age of Deep Learning**
  The 2010s were truly the golden age of deep learning. In 2012, a deep CNN called AlexNet, developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, won the ImageNet competition by a significant margin, revolutionizing the field of computer vision. The error rate achieved by AlexNet was 15.3%, while the second-best entry had a error rate of 26.2%. This success demonstrated the power of deep learning on large-scale problems.
  Soon after, deep learning architectures like RNNs (Recurrent Neural Networks) and LSTMs (Long Short-Term Memory networks) began to dominate sequential data tasks like language modeling and speech recognition. New architectures like GANs (Generative Adversarial Networks), introduced by Ian Goodfellow in 2014, opened up possibilities in creative tasks, such as generating realistic images and videos.
  Transformers, introduced in the 2017 paper "Attention is All You Need," by Vaswani et al., transformed the field of natural language processing (NLP). Transformers, and their derivatives like BERT and GPT, became the backbone of modern NLP systems.
  Deep learning's success stories spanned across industries: from healthcare to finance, from self-driving cars to game-playing AI like AlphaGo, developed by DeepMind, which defeated the world champion in the complex board game Go in 2016.
- **2020s: Scaling Up and Ethical Concerns**
  As the 2020s began, deep learning models became larger and more powerful. OpenAI's GPT-3, released in 2020, with its 175 billion parameters, stunned the world with its ability to generate human-like text. However, this era also brought challenges, such as the ethical implications of AI, including bias, privacy concerns, and the environmental impact of training massive models.
  The open-source community played a critical role during this time, with frameworks like TensorFlow and PyTorch making deep learning accessible to a broader audience. These tools, along with the democratization of knowledge through online courses and research papers, accelerated innovation in the field.
  Despite these advancements, researchers began to recognize that simply scaling up models wasn't enough. Issues like explainability, robustness, and fairness became central to ongoing research. Efforts are now focused on making AI not only powerful but also trustworthy and aligned with human values.

With tools like Colab, Keras and Tensorflow, virtually anyone can solve a problem in a day, with no initial investment, that would required an engineering team working for a quarter and $20K in hardware in 2014.

What is representation learning? Also its role in ML and Deep Learning.

- Representation Learning is a process that simplifies raw data into understandable patterns for machine learning. It enhances interpretability, uncovers hidden features, and aids in transfer learning.
- Data in its raw form (words and letters in text, pixels in images) is too complex for machines to process directly. Representation learning transforms the data into a representation that machines can use for classification or predictions.
- This success of Deep Learning heavily relies on the advancements made in representation learning.
- Previously, manual feature engineering constrained model capabilities, as it required extensive expertise and effort to identify relevant features. Whereas Deep learning automated this feature extraction.

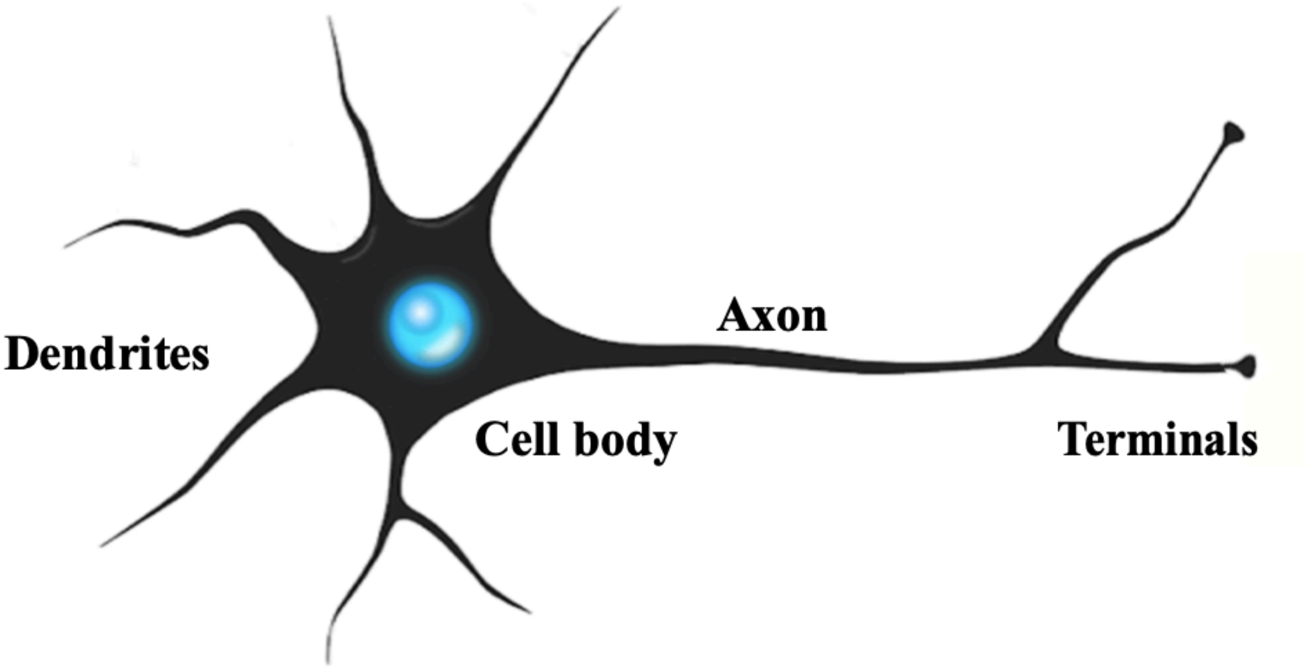
Hence, Deep learning uses representation learning, no manual feature extraction needed.

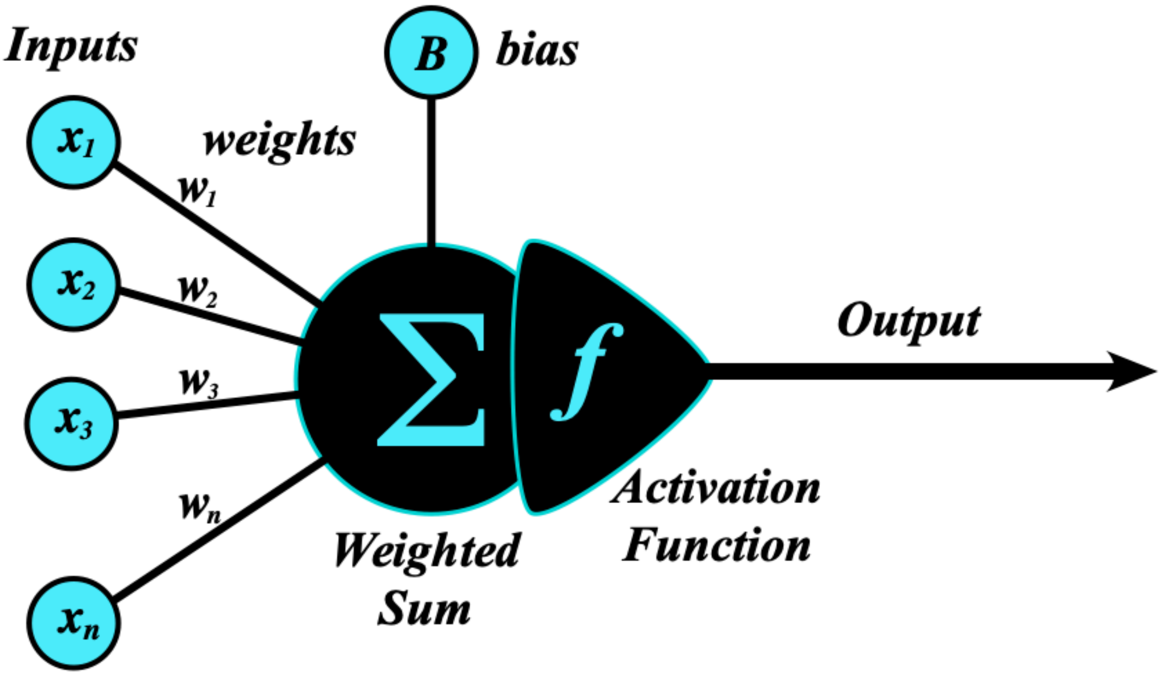


Neural Network Definition and Components

An artificial neural network (AAN) is a type of machine learning model inspired by the structure of the human brain. It consists of a set of interconnected processing nodes (artificial neurons), organized into layers that work together. The initial layer receives input data and the final layer produces the output. In between the input and output layers are one or more hidden layers. The hidden layers perform a series of nonlinear transformations on the input data, allowing the network to learn complex patterns in the data.

AANs with just a small number (1-3) hidden layers are known as shallow networks; those with many more layers are called deep networks.





The fundamental component of artificial neural networks, loosely modeled after biological nerve cells, is called the perceptron, alternatively known as an (artificial) neuron, unit, or node. In network diagrams like the one above, individual neurons are represented by colored circles.

Biologic and artificial neurons

A biological nerve cell receives input stimuli from neighboring nerves through its dendrites. If the sum of these stimuli is sufficient to create membrane depolarization in the neuron's cell body, an electrical output signal will be transmitted down the axon to its terminals (which in turn may stimulate dendrites of other nerves).

The artificial neuron receives a set of weighted inputs $(w_1x_1 + w_2x_2 + ... + w_nx_n)$ plus a constant bias (B). This weighted sum is then fed into an activation function that produces an output for the node.

The purpose of the activation function is to introduce nonlinearities between units. Without this, the output of all neuron groups would simply be linear combinations of the others and the more interesting and powerful behaviors of neural networks would not be possible.