Deep Learning

Network Application Diagnostics
B2M32DSAA

Radek Mařík

Czech Technical University
Faculty of Electrical Engineering
Department of Telecommunication Engineering
Prague CZ

November 13, 2025

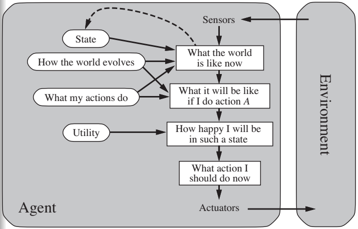
Outline

Outline

Intelligence in General [RN10]

Studies of intelligence in general:

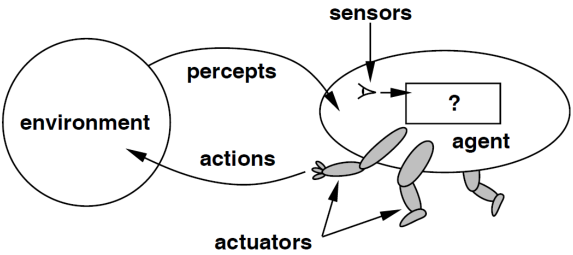
- How do we **perceive** the world?
- How do we **understand** the world?
- How do we **reason** about the world?
- How do we **predict** the consequences of our actions?
- How do we act to **influence** the world?



Artificial Intelligence (AI) not only wants to understand the "intelligence", but also wants to

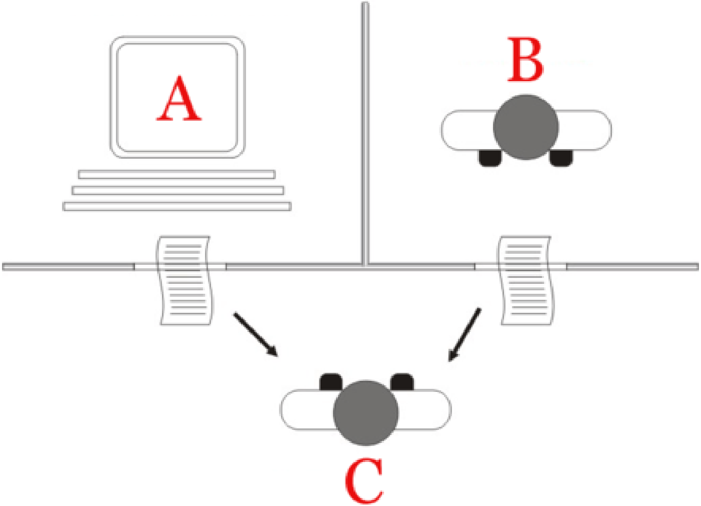
- create an intelligent entity (agent, robot)
- imitating or improving
- the human behavior and effects in the outer world, and/or
- the inner human mind processes and reasoning.

What is AI for us? [RN10]



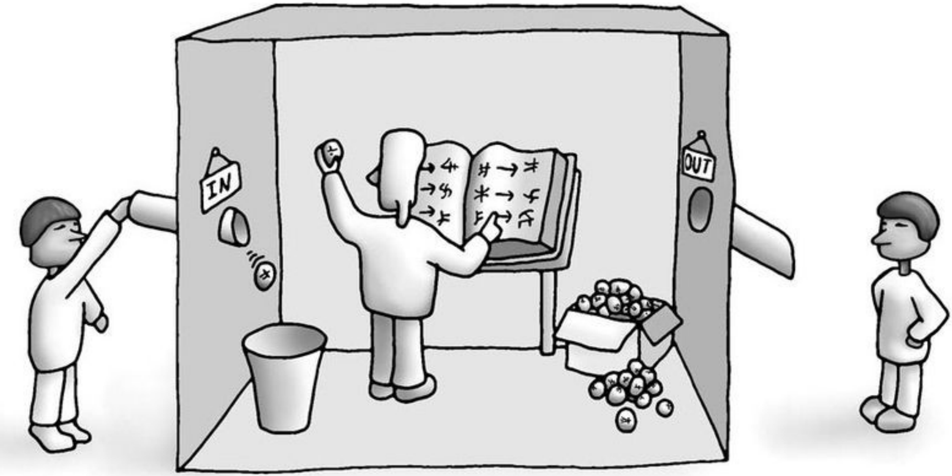
- **Acting rationally:**
  - Care only about what they do and if they achieve their goals optimally.
  - Goals are described in terms of the utility of the outcomes.
  - **Maximize the expected utility of the outcomes of their decisions.**
- Good decisions:
  - Take into account similar situations that happened in the past. **Machine learning**.
  - Simulations using a model of the world. Be aware of the consequences of your actions and plan ahead. **Inference, planning.**

Turing Test



https://en.wikipedia.org/wiki/Turing_test

Chinesse Room



http://america.pink/

Superintelligence? (NO)

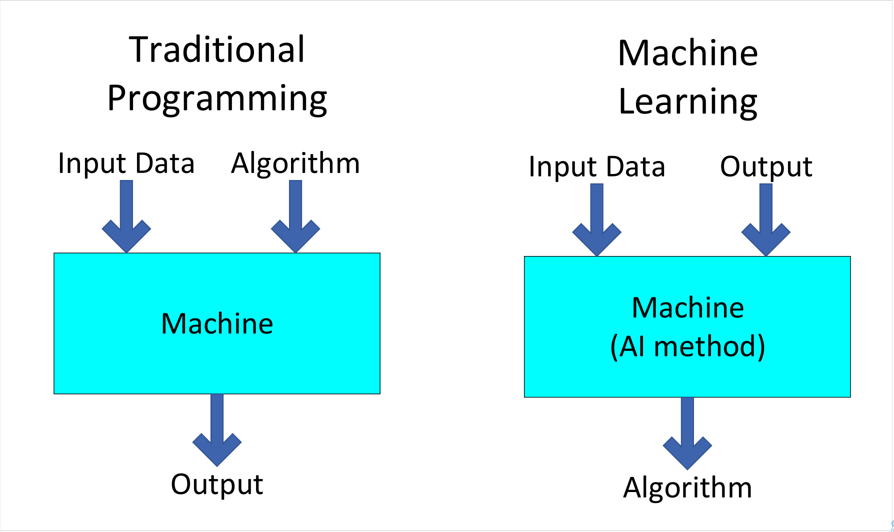
However, even **narrow AI** might be abused by people against people!

AI Science Disciplines [RN10]

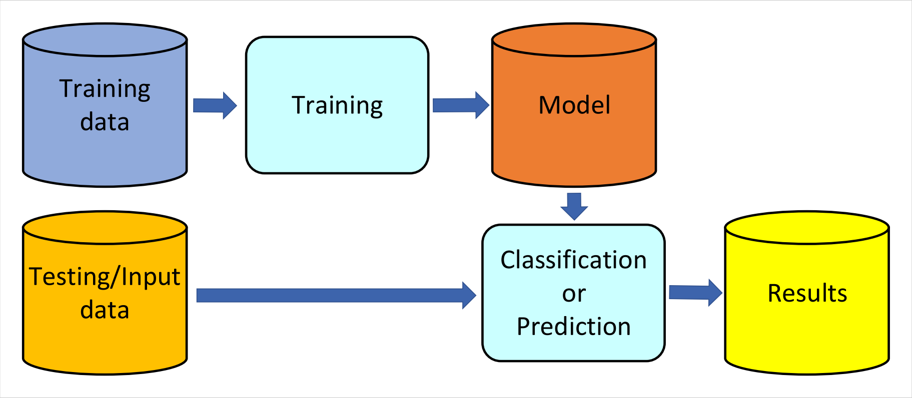
- **Knowledge representation:**
  - how to store the model of the world, the relations between the entities in the world, the rules that are valid in the world, ...
- **Automated reasoning:**
  - how to infer some conclusions from what is known or answer some questions
- **Planning:**
  - how to find an action sequence that puts the world in the desired state
- **Pattern recognition:**
  - how to decide about the state of the world based on observations
- **Machine learning:**
  - how to create/adapt the model of the world using new observations
- **Multiagent systems:**
  - how to coordinate and cooperate in a group of agents to reach the desired goal
- **Natural language processing:**
  - how to understand what people say and how to say something to them
- **Computer vision:**
  - how to understand the observed scene, what is going on in a sequence of pictures
- **Robotics:**
  - how to move, how to manipulate with objects, how to localize and navigate

Outline

AI Machine Learning Principle [RN10]



AI Machine Learning Workflow [RN10]



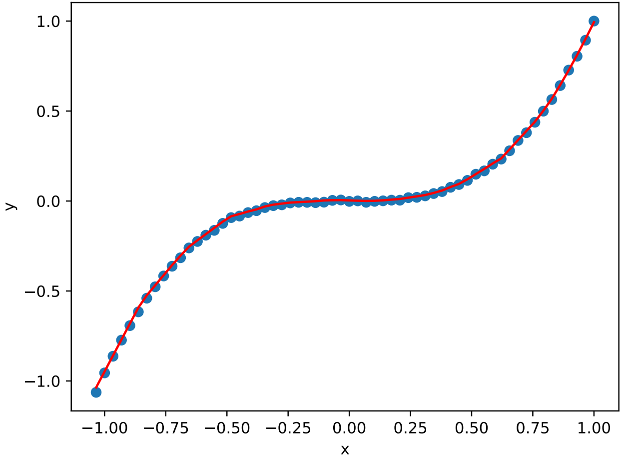
The Universal Approximation Theorem

- Any neural network architecture aims at finding any mathematical function $y = f(x)$ that can map attributes x to output y.
- The function $f(x)$ can be arbitrarily complex.
- The Universal Approximation Theorem tells us that Neural Networks has a kind of **universality** i.e. no matter what $f(x)$ is, there is a network that can approximate the result.

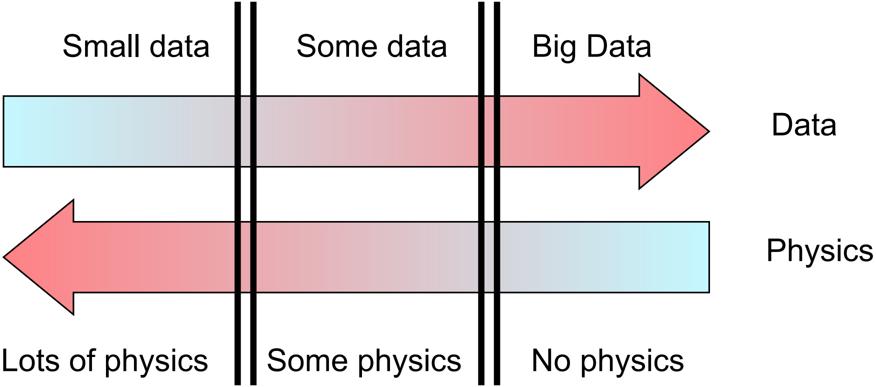
`https://towardsdatascience.com/neural-networks-and-the-universal-approximation-theorem-8a389a33d30a`

Cubic Function - Deep NN

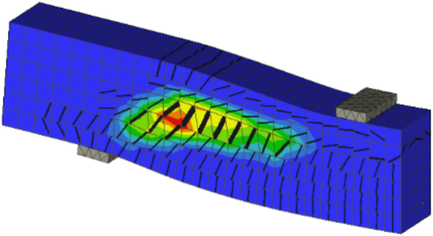
Input(x), Dense(200), Dense(50), Dense(10), Dense(1)



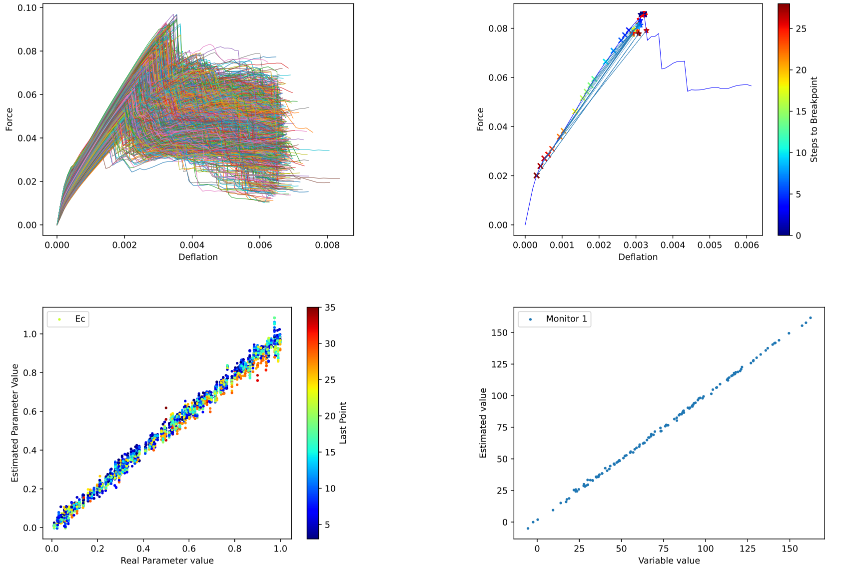
Physics versus Data



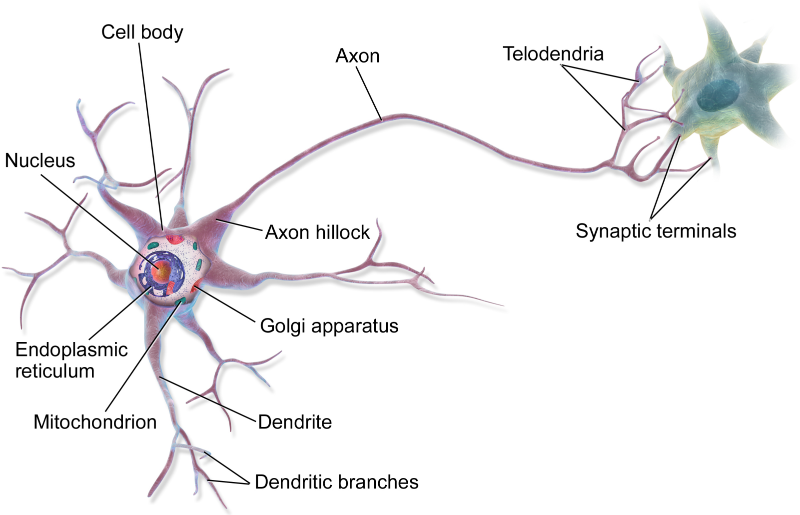
Concrete Bridge Twin (Project TAČR CK03000023 2022-2024)



TwinBridge Results

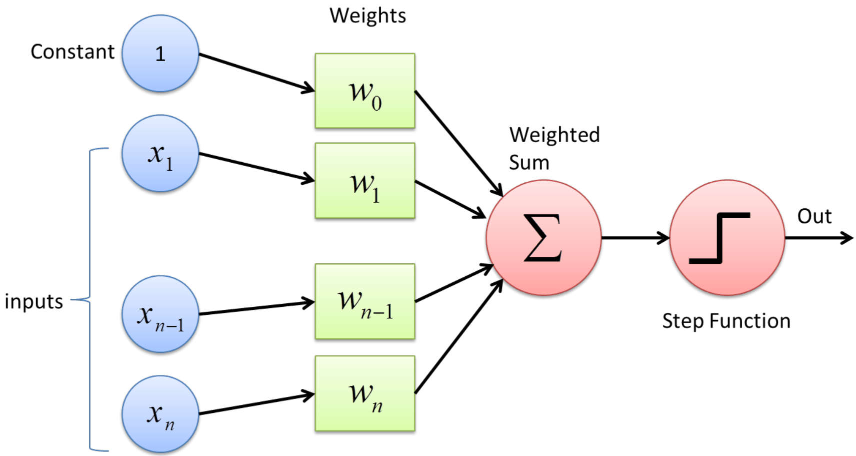


Neuron



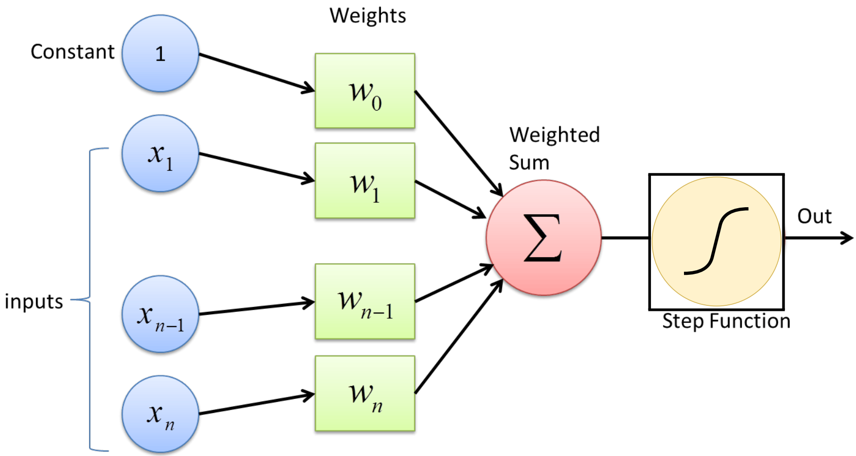
https://en.wikipedia.org/wiki/Multipolar_neuron

Perceptron [Ros58]



https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53

Perceptron [Ros58]



https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53

Binary Classification Task (dichotomy)

Training dataset $T = \{(\mathbf{x}^{(1)}, y^{(1)}), \ldots, (\mathbf{x}^{(|T|)}, y^{(|T|)})\}$:

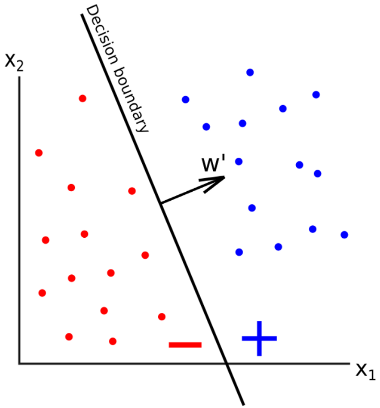
- each sample described by a vector $\mathbf{x} = (x_1, \ldots, x_D)$,
- labeled with the correct class $y \in \{+1, -1\}$.

Discriminant function $f(\mathbf{x})$:

- Used to create a **decision rule** which assignes a class to an observation:

$$f(\mathbf{x}) > 0 \iff \hat{y} = +1, \text{and}$$
$$f(\mathbf{x}) < 0 \iff \hat{y} = -1$$

i.e. $\hat{y} = \mathrm{sign}(f(\mathbf{x}))$

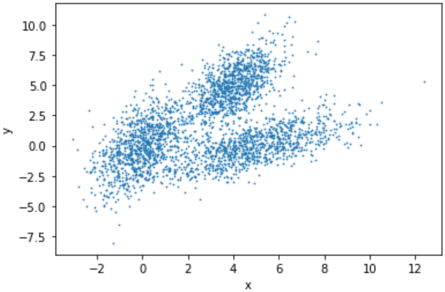


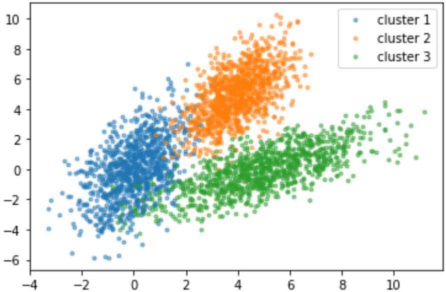
Decision boundary: $\{\mathbf{x} | f(\mathbf{x}) = 0\}$

Linear classification: the decision boundaries must be linear.

Learning: searching (suitable parameters of) function f.

Clustering





- **Unsupervised learning** . . . to discover hidden patterns in any unlabeled data.
- **A cluster** . . . a group containing data points with high similarity and low similarity with data points in other groups.
- **Clustering** . . . to divide a set of data points in such a way that similar items fall into the same cluster, whereas dissimilar data points fall in different clusters.

https://training.galaxyproject.org/training-material/topics/statistics/tutorials/clustering_

Outline

Classification Predictions [Wik19a]

- **The expectation:** the terms *positive* and *negative* refer to the classifier's prediction.
- **The observation:** the terms *true* and *false* refer to whether that prediction corresponds to the external judgment.
- The confusion matrix (CZ kontigenční tabulka)

		Predicted / Classified	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

- **TN / True Negative**
  - the real case is negative
  - and predicted negative
- **FP / False Positive**
  - the real case is negative
  - but predicted positive
  - Type I error
- **TP / True Positive**
  - the real case is positive
  - and predicted as positive
- **FN / False Negative**
  - the real case is positive
  - but predicted negative
  - Type II error

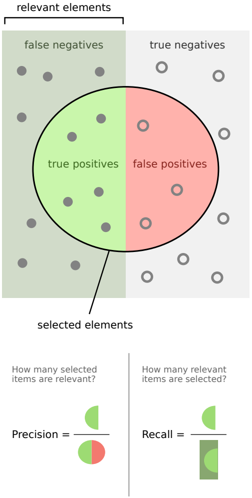
Precision and Recall [Wik19a, ?]

- **Precision**
  - the probability that a (randomly selected) retrieved document is relevant.
  - the probability that a (randomly selected) object is correctly classified.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall**
  - the probability that a (randomly selected) relevant document is retrieved in a search.
  - the probability that a (randomly selected) class object is correctly classified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

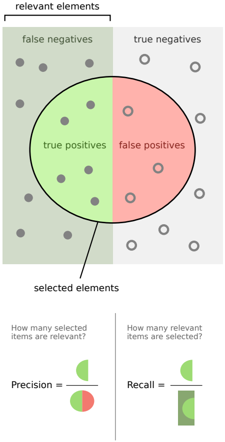


Accuracy and F1-Measure [Wik19a, ?]

- **Accuracy**
  - the proportion of true results (both true positives and true negatives) among the total number of cases examined.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

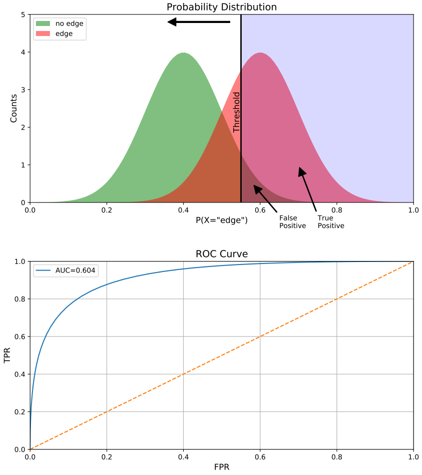
- $F1$-**Measure**
  - the harmonic mean of precision and recall.
  - an $F1$ score reaches its best value at 1 (perfect precision and recall) and worst at 0.



$$\text{F1} = \left(\frac{\text{Precision}^{-1} + \text{Recall}^{-1}}{2}\right)^{-1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

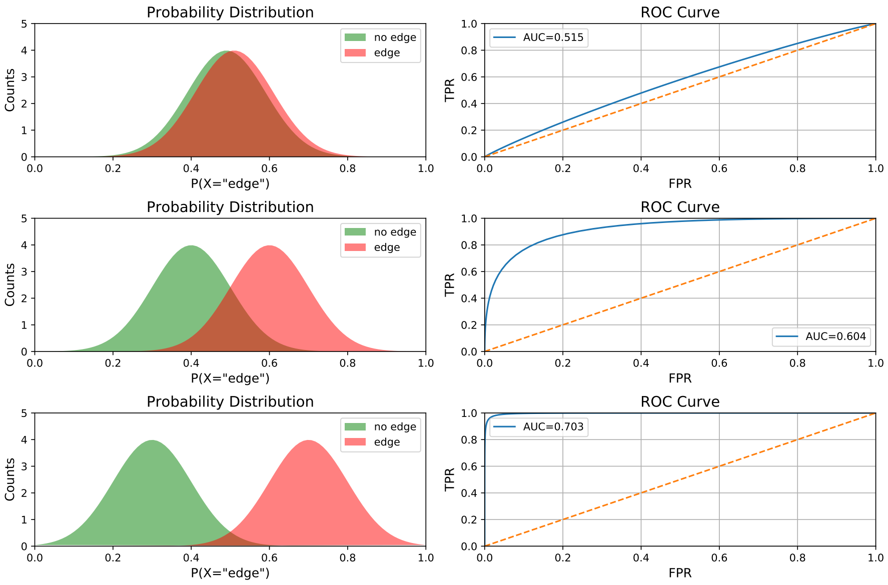
ROC (Receiver Operating Characteristic) Curves [Wik19b]

- Plotting the true positive rate (TPR) against the false positive rate (FPR).
- Dealing with heavy class imbalance.
- The model performance is measured by the area under the ROC curve (AUC).
- The best possible AUC is 1.
- The worst AUC is 0.5 (the 45 degrees random line).
- If the AUC is below 0.5, do the exact opposite of what the model recommends.



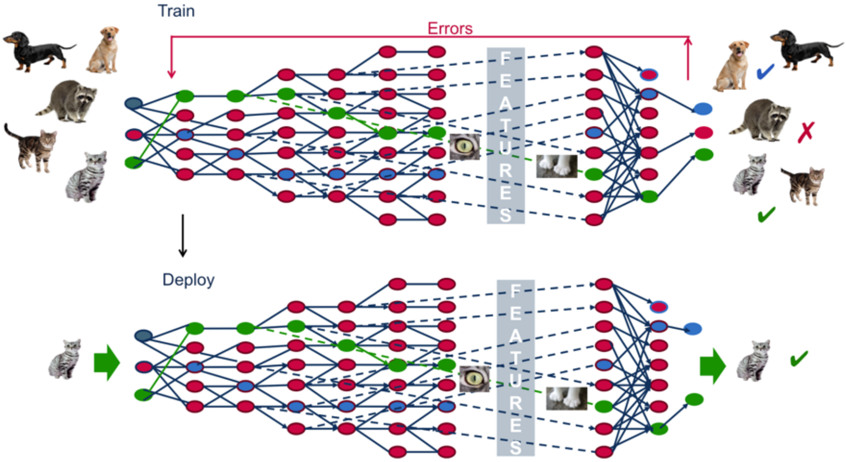
$$\mathrm{TPR} = \frac{TP}{TP+FN}, \ \mathrm{FPR} = \frac{FP}{FP+TN}$$

ROC Performance Assessment [Wik19b]



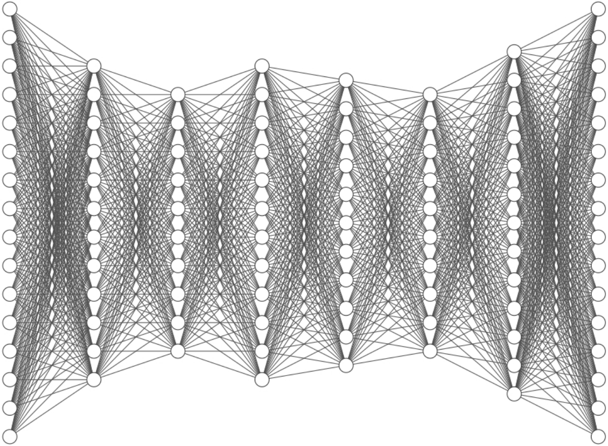
Outline

Deep learning (2006)



https://mapr.com/blog/demystifying-ai-ml-dl/

Multilayer perceptron (MLP)



Fully connected. Used in speech recognition and machine translation.

https://alexlenail.me/NN-SVG/index.html

MLP Structure

- Each neuron is characterized by its **weight**, **bias**, and **activation function**.
- **The input layer** takes raw input from the domain. No computation is performed at this layer. Nodes here just pass on the information (features) to the hidden layer.
- **The hidden layers** perform all kinds of computation on the features entered through the input layer and transfers the result to the output layer.
- **The output layer** brings the information learned through the hidden layer and delivers the final value as a result.

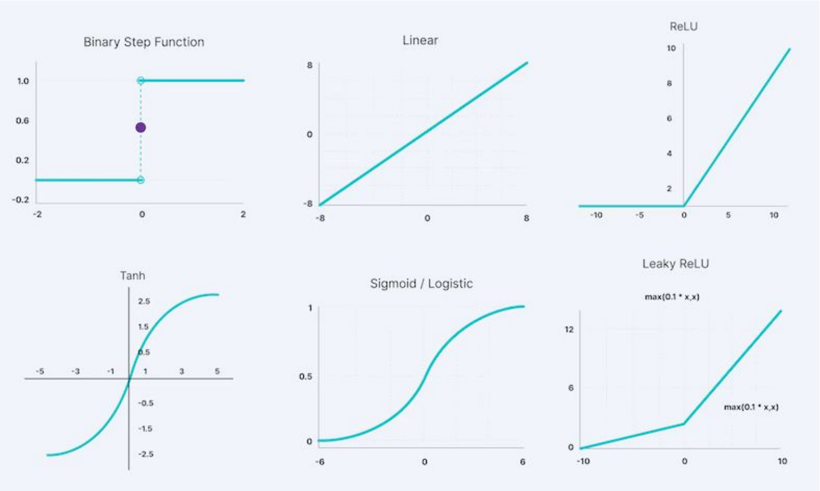
https://www.v7labs.com/blog/neural-networks-activation-functions

Feedforward vs. Backpropagation

- **Feedforward Propagation** - the flow of information occurs in the forward direction. The input is used to calculate some intermediate function in the hidden layer, which is then used to calculate an output.
- **Backpropagation** - the weights of the network connections are repeatedly adjusted to minimize the difference between the actual output vector of the net and the desired output vector.
  - Backpropagation aims to minimize the cost function by adjusting the network's weights and biases.
  - The cost function gradients determine the level of adjustment with respect to parameters like activation function, weights, bias, etc.
- **Training** = Feedforward Propagation and Backpropagation
- **Testing** = Feedforward Propagation

https://www.v7labs.com/blog/neural-networks-activation-functions

Activation Functions



https://www.v7labs.com/blog/neural-networks-activation-functions

Loss/Cost Functions - Regression

- **Regression** - model needs to predict a continuous number
  - *Mean Squared Error (MSE) Loss*

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

  - *Mean Absolute Error (MAE) Loss*

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

https://www.geeksforgeeks.org/deep-learning/loss-functions-in-deep-learning/

Loss/Cost Functions - Classification

- **Classification** - classification loss functions are used to evaluate how well a classification model's predictions match the actual class labels.
  - *Binary Cross-Entropy Loss (Log Loss)*. It is used for binary classification problems. It measures the performance of a classification model whose output is a probability value between 0 and 1.

$$\text{Binary Cross-Entropy} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + [(1 - y_i)\log(1 - \hat{y}_i)]$$

  - *Categorical Cross-Entropy Loss* - is used for multiclass classification problems. It measures the performance of a classification model whose output is a probability distribution over multiple classes.

$$\text{Categorical Cross-Entropy} = -\sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij}\log(\hat{y}_{ij})$$

    where data point i, k is the number of classes, $\hat{y}_{ij}$ is the predicted probability for class j.

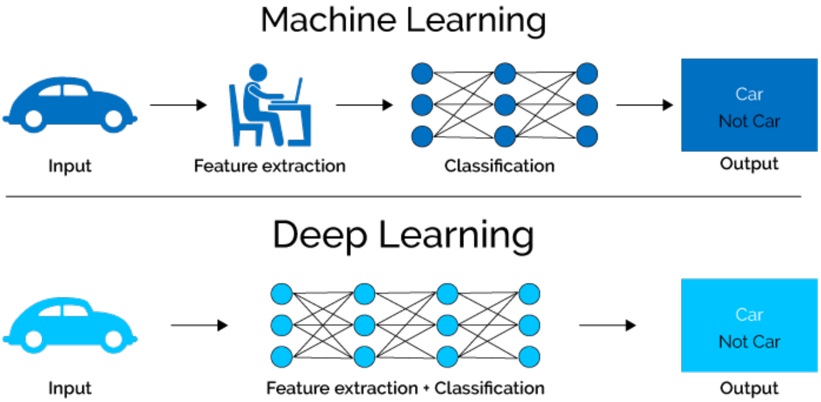
https://www.geeksforgeeks.org/deep-learning/loss-functions-in-deep-learning/

Regularization Functions

- *Regularization* is a technique used in machine learning to prevent overfitting and improve the generalization performance of a model on unseen data.
- *Overfitting* occurs when a model learns to perform well on the training data but fails to generalize to new, unseen data.
- **Dropout** - In Dropout, a random subset of neurons is temporarily excluded or "dropped out" during each iteration.
- *Batch Normalization* involves normalizing the inputs of each layer in a mini-batch by subtracting the mean and dividing by the standard deviation.
- *Adaptive Average Pool* is a pooling layer that automatically calculates the right kernel size, stride, and padding to produce a specific output size, regardless of the input's height and width.

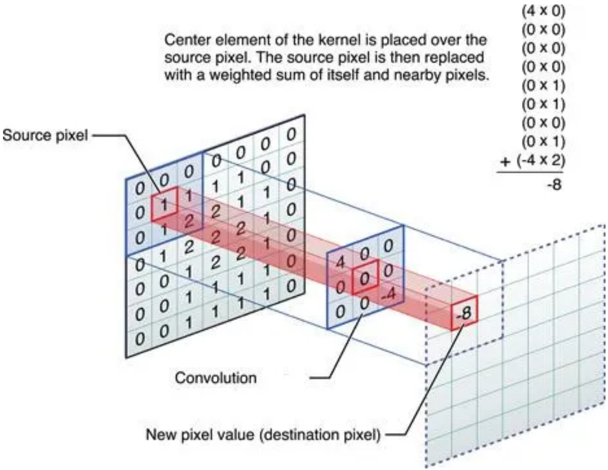
https://medium.com/@datasciencejourney100_83560/regularization-techniques-in-deep-learning-3de958b14fba

Machine Learning vs. Deep Learning



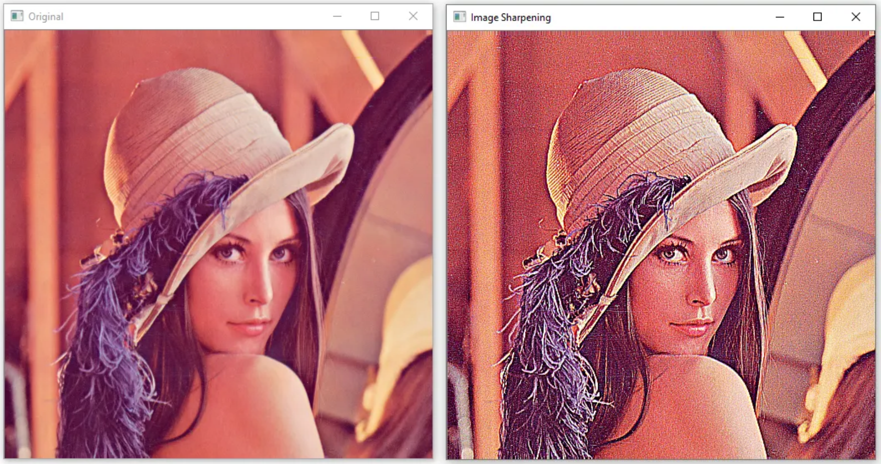
https://www.thecrazyprogrammer.com/2018/03/difference-deep-learning-machine-learning.html

Convolution Operation



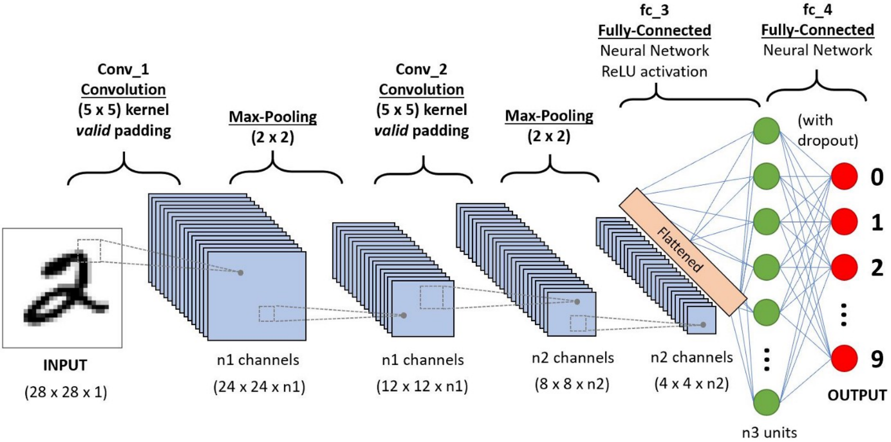
https://medium.com/@bdhuma/6-basic-things-to-know-about-convolution-daef5e1bc411

Feature Enhacement




```
[[-1 -1 -1]
 [-1  9 -1]
 [-1 -1 -1]]
```

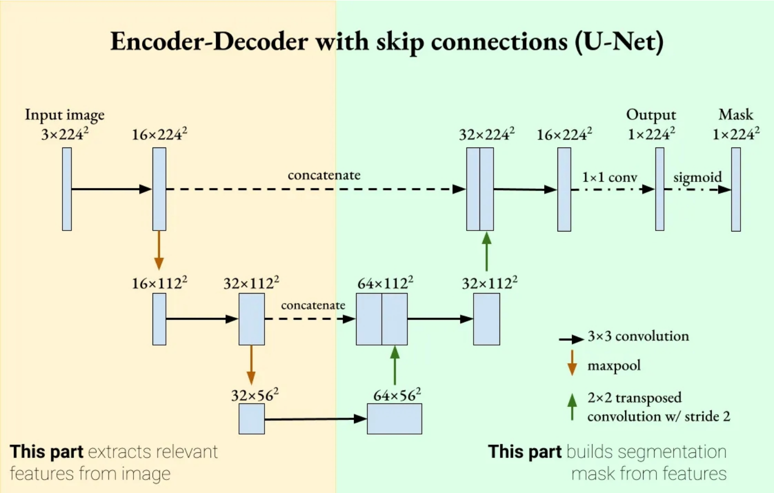
Convolutional Neural Network (ConvNet/CNN)



Used in computer vision to generate features.

https:
//towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

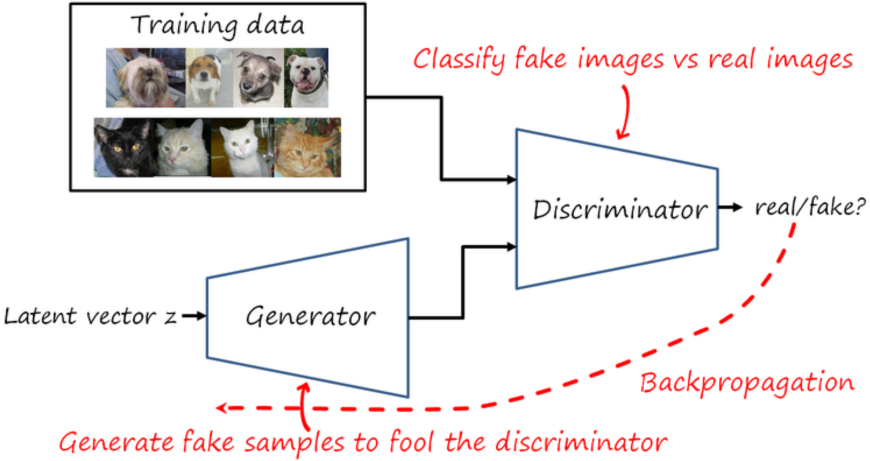
UNet Neural Network (CRAFT [BLH+19])



Used in computer vision in segmentation.

https://towardsdatascience.com/understanding-u-net-61276b10f360

Generative adversarial networks (GANs) (2014)



Wasserstein GAN (2017)

http://www.lherranz.org/2018/08/07/imagetranslation/

Fake Image Generation (2018)

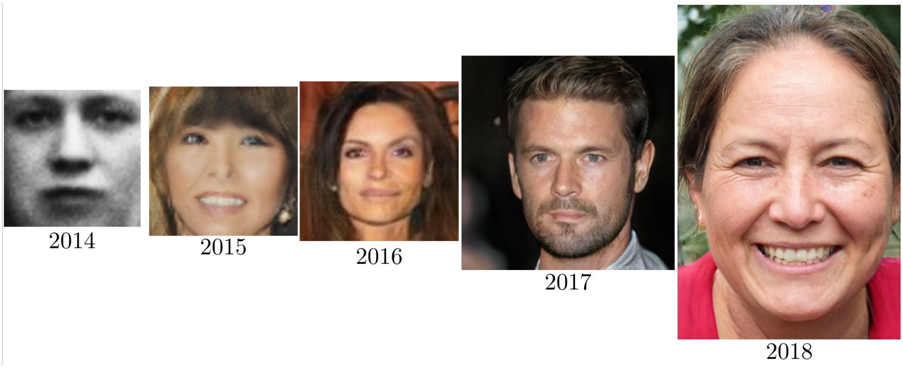


Image Style Conversion - CycleGAN (2017) [ZPIE17]

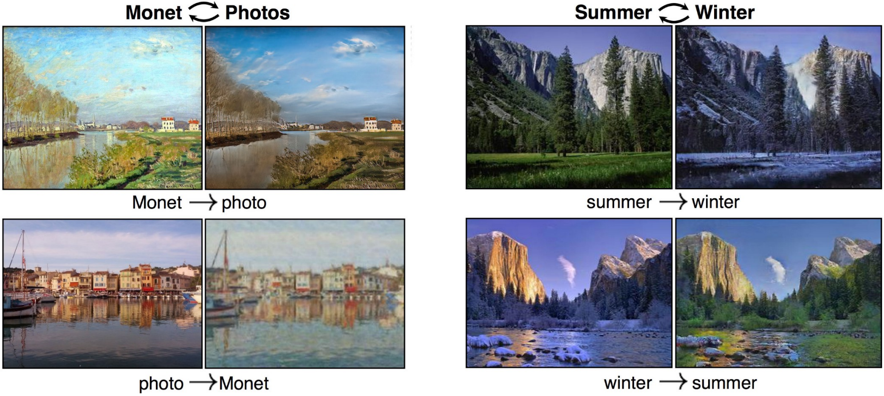
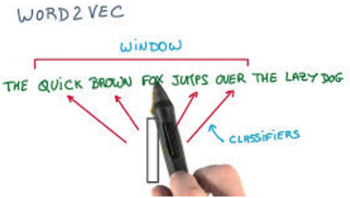
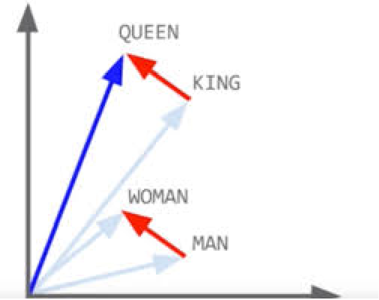


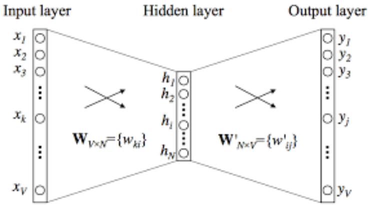
Image Style Conversion Failures (2017) [ZPIE17]

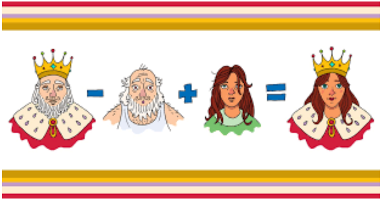
Outline

Embeddings - Word2vec Word Representation (Skipgram)

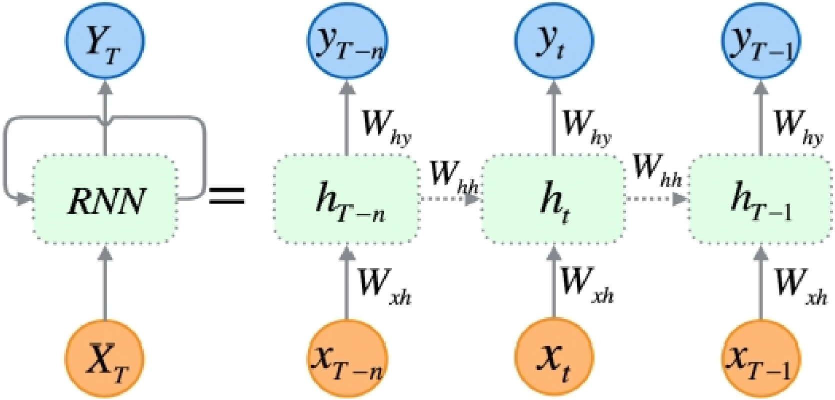








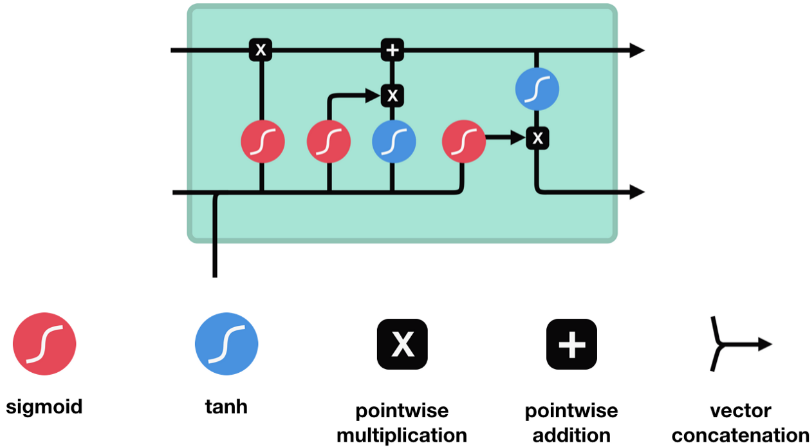
Recurrent Neural Network (RNN)



Used in natural language processing.

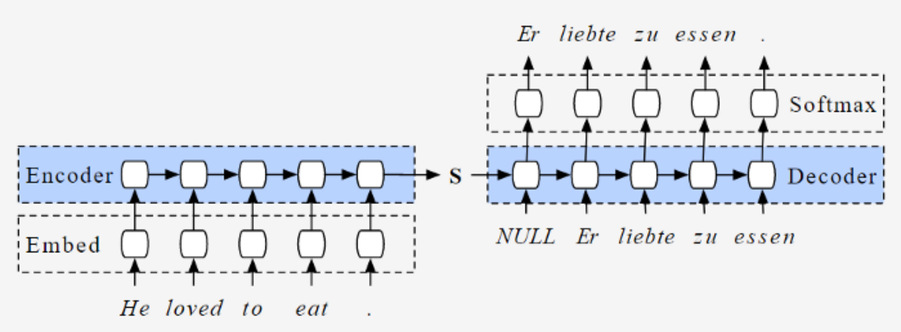
https://www.i2tutorials.com/what-is-the-difference-between-bidirectional-rnn-and-rnn/

Long Short-Term Memory (LSTM) Network [HS97]



Used in natural language processing.

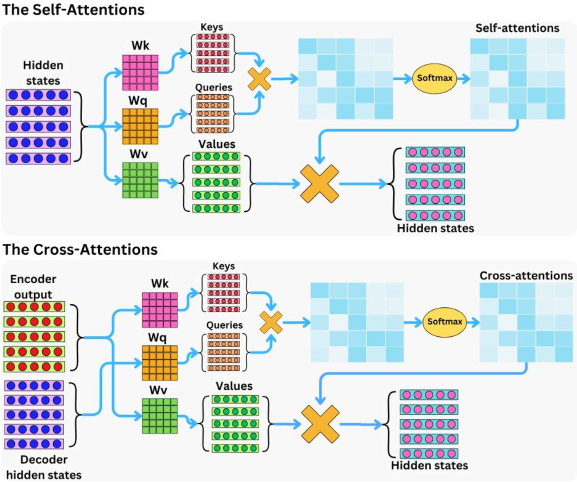
Sequence2Sequence Network [SVL14]



Used in natural language processing.

https://www.analyticsvidhya.com/blog/2020/08/a-simple-introduction-to-sequence-to-sequence-models/

Self and Cross Attention [BCB16]



Used in natural language processing and computer vision.

https://www.linkedin.com/posts/damienbenveniste_what-is-the-difference-between-self-attention-activity-7211029906166624257-m0Wn/

Transformer Network [VSP+17]

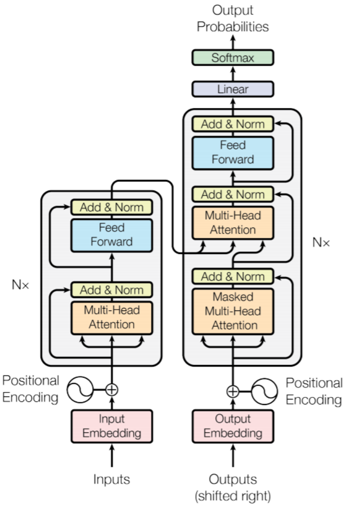


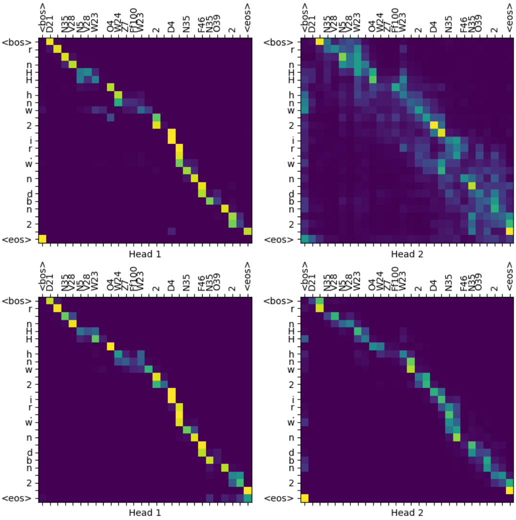
Figure 1: The Transformer - model architecture.

- Used in natural language processing.
- Self-Attention. . . attention applied to a single sequence (position dependencies)
- Multi-head Attention . . . several subspaces of attention
- Language models
  - ELMo . . . March 2018,
  - GPT . . . June 2018, 117Mi pars
  - BERT . . . May 2019, BERTlarge 340M pars,
    - ALBERT(base:12M, xxlarge: 235M), RoBERTa (160GB of text), DistillBERT (66 mil.), DeBERTa (2021.01, 1.5B)
  - GPT2 . . . November 2019, 1.5Bi pars, 8 milion web pages (40GB)
  - XLNet . . . January 2020, Large: 340M pars
  - GPT3 . . . July 2020, 175Bi pars
  - GPT4 . . . March 2023, ≈1.7 trillion pars

https://towardsdatascience.com/transformer-neural-network-step-by-step-breakdown-of-the-beast-b3e096dc857f

Gardiner Code to Transliteration Translation

Model 7 Cross Attention Visualization (2H/1B+2H/2B)



LLM in Telecommunications - Survey

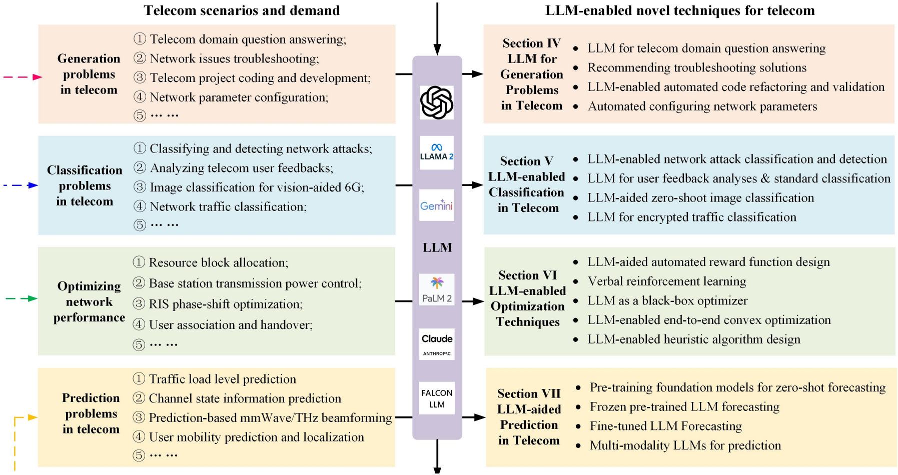
Section I Introduction
- Background
- Motivations
- Contributions

Section II Related Surveys
- Comparison with existing studies
- Main differences

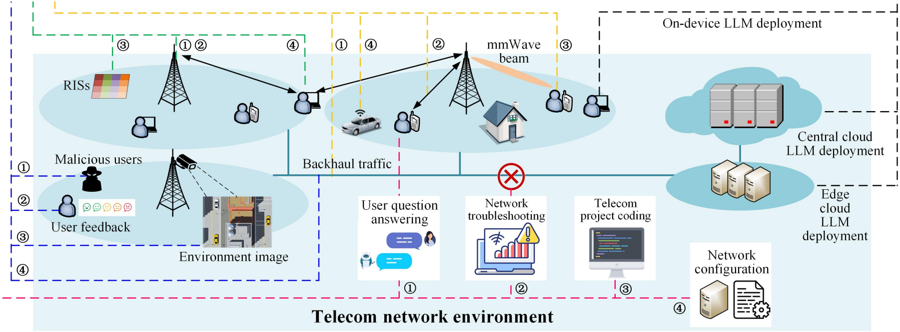
Section III LLM Fundamentals
- Model architecture
- Pre-training
- Fine-tuning
- Inference and utilization
- Evaluation of LLM performance
- Deploying LLMs in telecom networks
- Analyses of LLM fundamentals in the telecom domain

Section VIII Challenges and Future Directions of LLM-empowered Telecom
- Challenges of applying LLMs to telecom
- Future directions

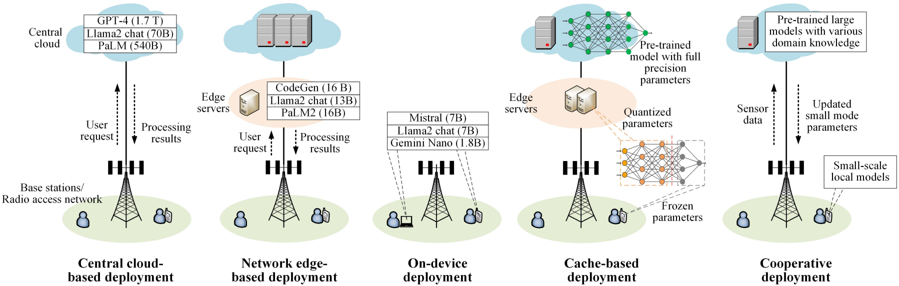
LLM in Telecommunications- Use Cases



LLM in Telecommunications - Environment



LLM in Telecommunications - Deployment



THANK YOU

References I

[BCB16] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2016.

[BLH+19] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, and Hwalsuk Lee. Character region awareness for text detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9357–9366, 2019.

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.

[RN10] Stuart J. Russell and Peter Norvig. *Artificial Intelligence, A Modern Approach*. Pre, third edition, 2010.

[Ros58] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65 6:386–408, 1958.

[SVL14] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks, 2014.

[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[Wik19a] Wikipedia. Precision and recall, 2019.

[Wik19b] Wikipedia. Receiver operating characteristic, 2019.

[ZPIE17] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.