Introduction to Deep Learning

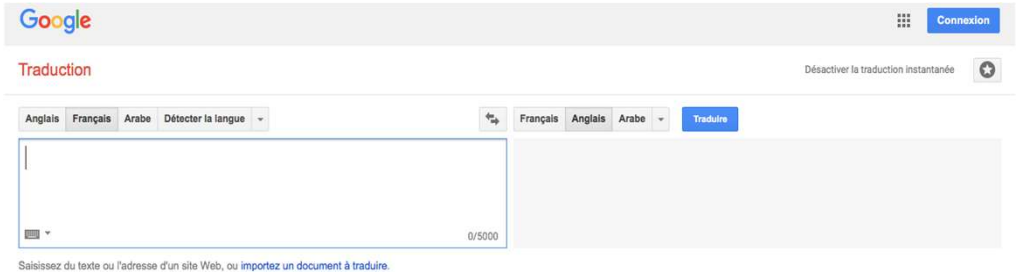
TIN0255/09P9M80/D82 - Aprendizagem Profunda

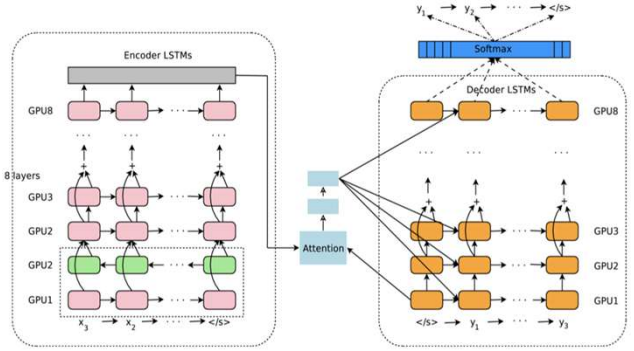
2026.2 - Prof. Pedro Moura

(Baseado nos slides do Prof. Jean-Pierre Briot)

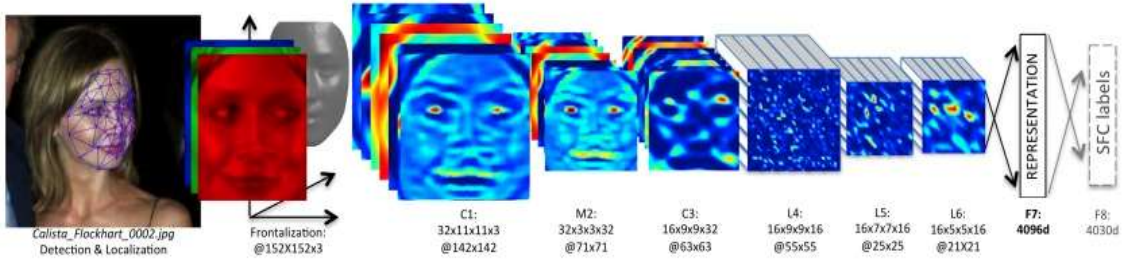
Deep Learning – Already There

- Translation, ex: Google Translate

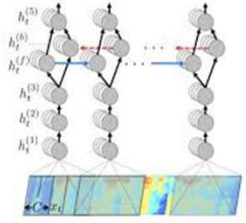




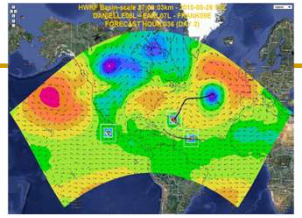
- Image Recognition, ex: Facebook



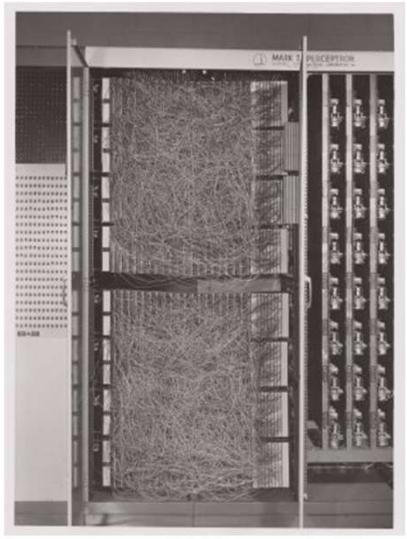
- Speech Recognition & Generation, ex: Apple Siri, Amazon Echo…

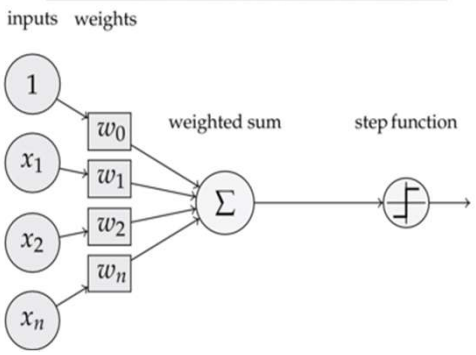


- Weather Prediction

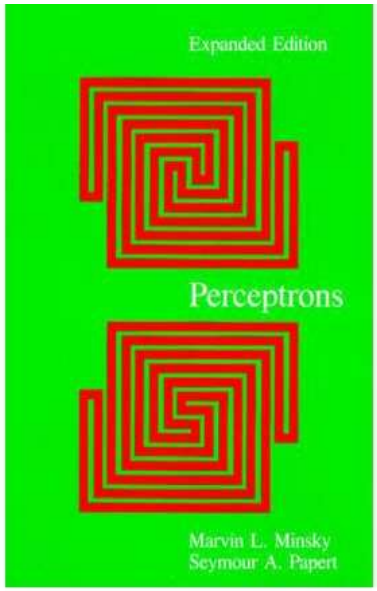


History: From Perceptron to Artificial Neural Networks to Deep Learning (1/4)



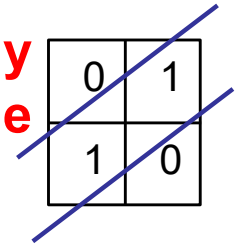


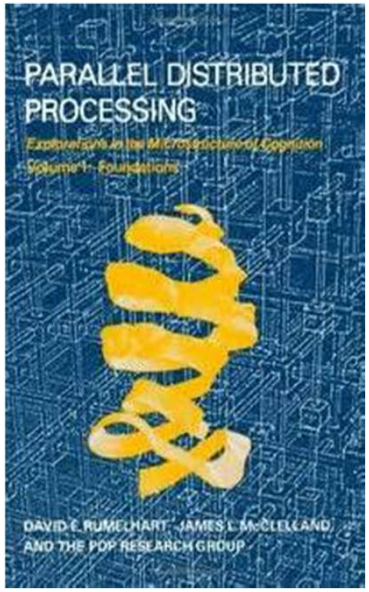
Perceptron
[Rosenblatt 1957]



Perceptrons (Book)
[Minsky & Papert 1969]

Linear Separable only
XOR counter example



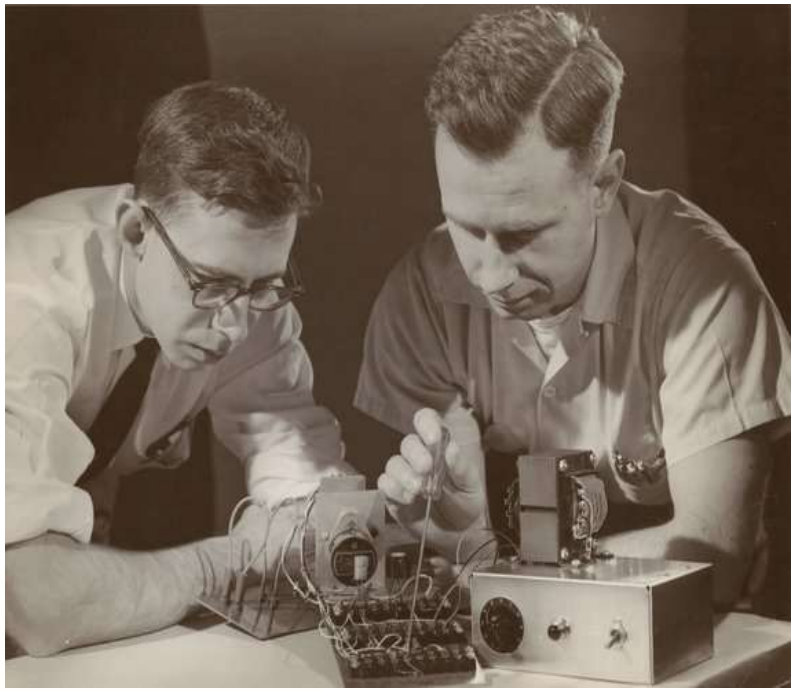


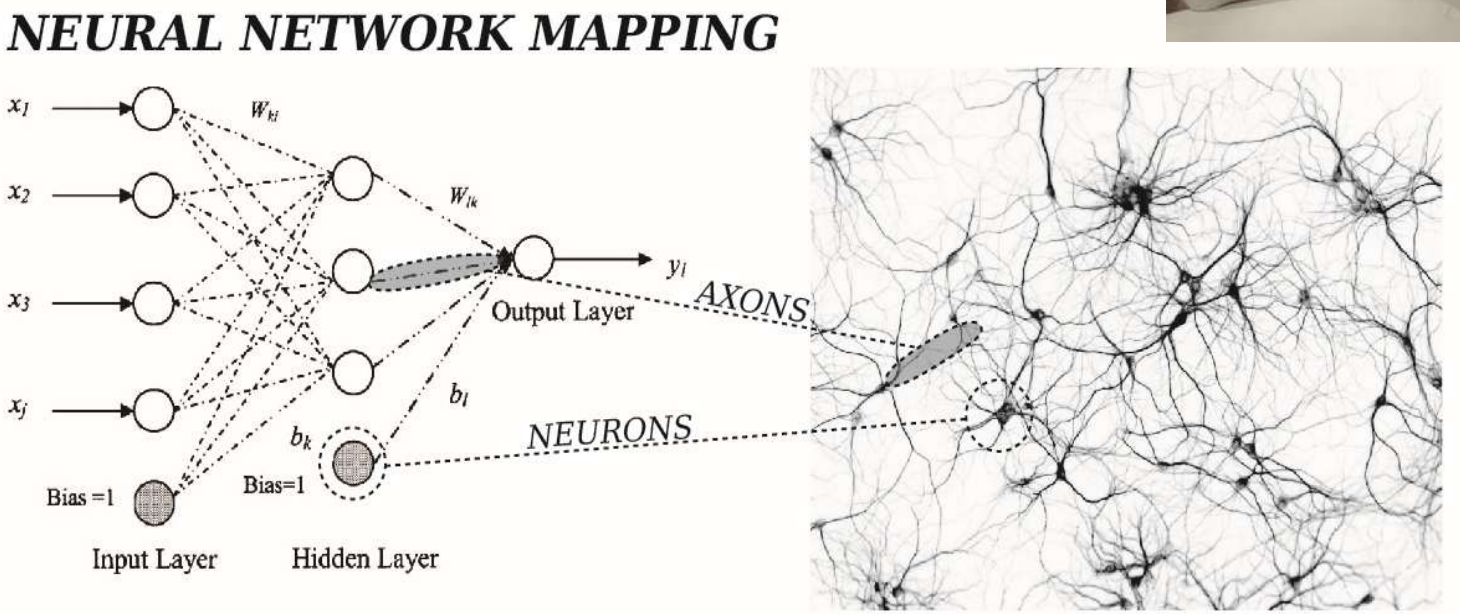
PDP (Books)
[Rumelhart et al. 1986]

Multi-layer networks
Backpropagation

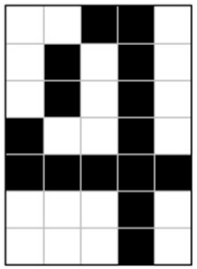
Deep Learning Bio-inspired or/and Regression-based ?

- Historically Perceptron bioinspired
- Aimed at Character Recognition





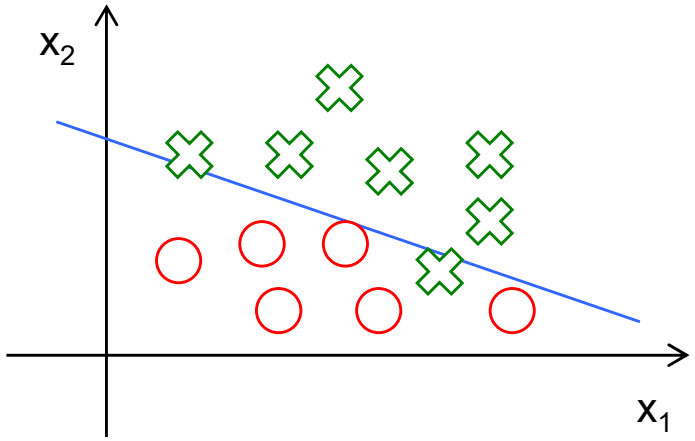
Perceptron [Rosenblatt, 1957]



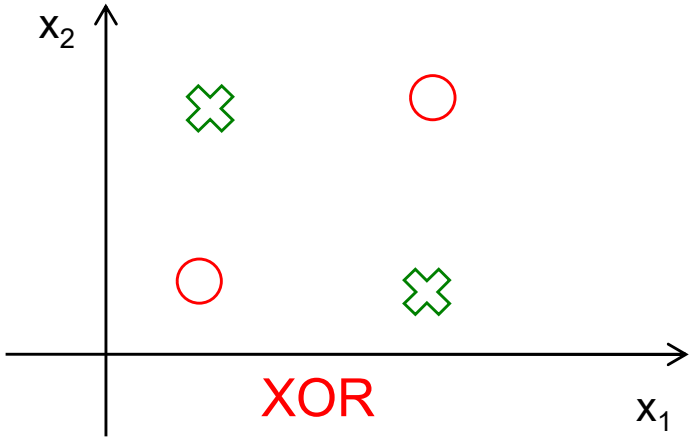
[Knoldus, 2017]

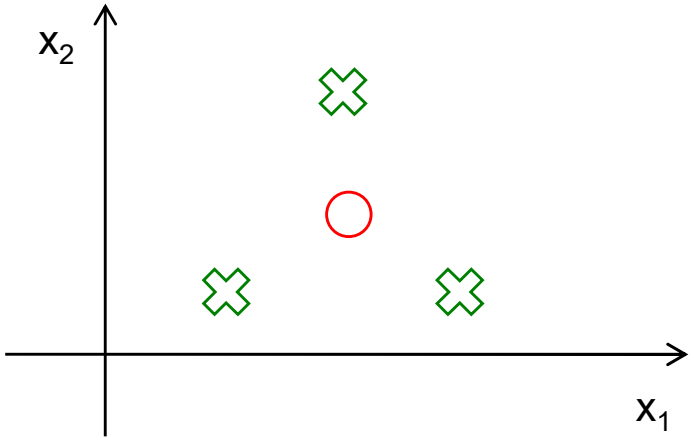
Linear vs Non Linear Decision Boundary

- Linear



- Non Linear

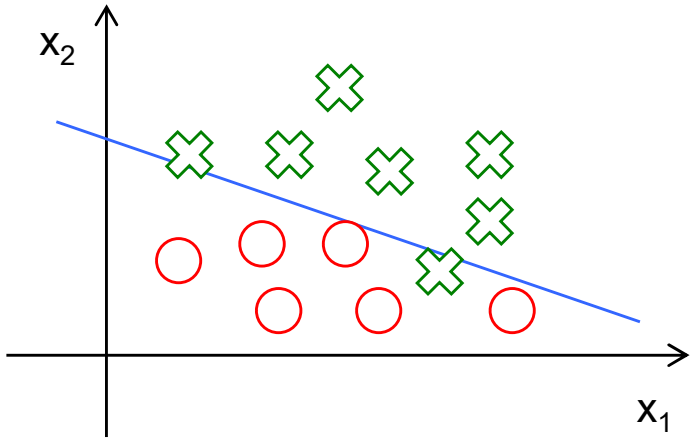




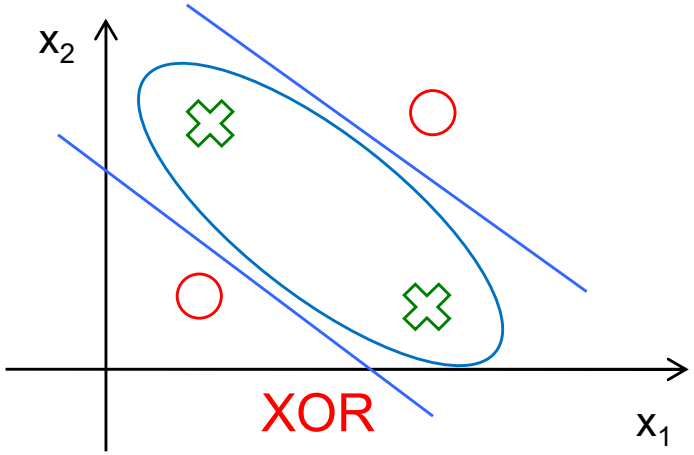
- Argument (XOR) used by [Minsky & Papert 1969] to criticize Perceptrons [Rosenblatt 1957] (and advocate Symbolic Artificial Intelligence)
- This stopped research on Perceptrons/Neural Networks for a long while
  - until Hidden Layers and Backpropagation or/and Kernel Trick (see later)

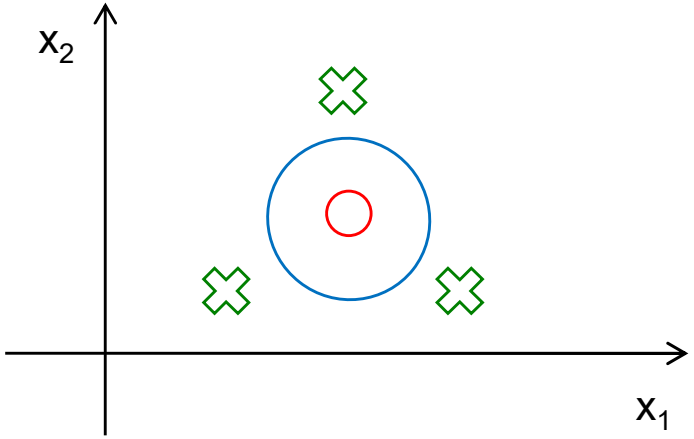
Linear vs Non Linear Decision Boundary

- Linear



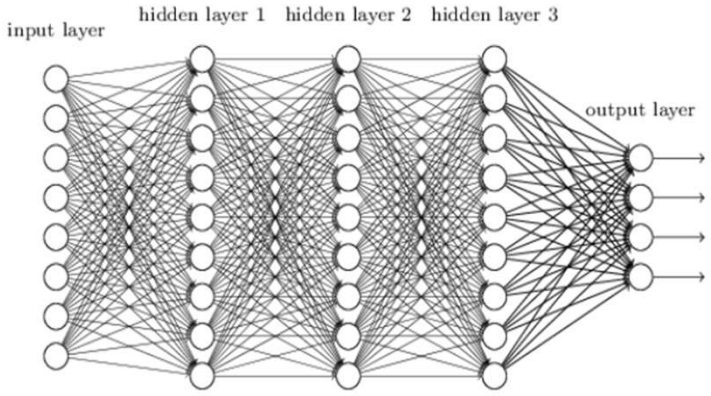
- Non Linear

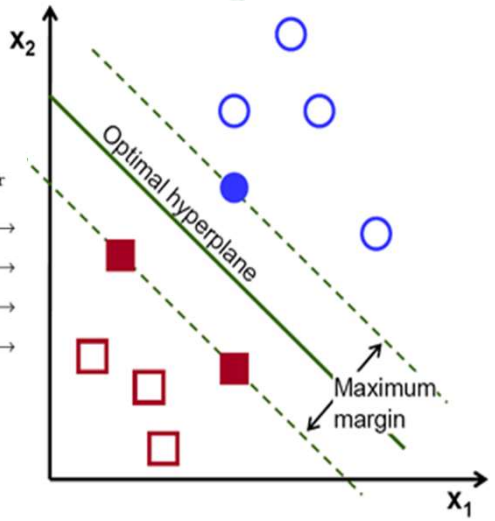


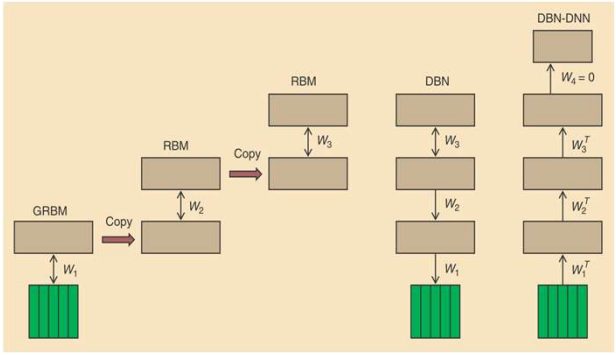


- Argument (XOR) used by [Minsky & Papert 1969] to criticize Perceptrons [Rosenblatt 1957] (and advocate Symbolic Artificial Intelligence)
- This stopped research on Perceptrons/Neural Networks for a long while
  - until Hidden Layers and Backpropagation or/and Kernel Trick (see later)

History: From Perceptron to Artificial Neural Networks to Deep Learning (2/4)







Pre-Training [Hinton et al. 2006]
Layer-Wise Self-Supervised Training/Initialization

SVM [Vapnik 1963]
SVM + Kernel Trick [Vapnik et al. 1992]

Difficulty to Efficiently Train Networks with many Layers

Unstable Gradients

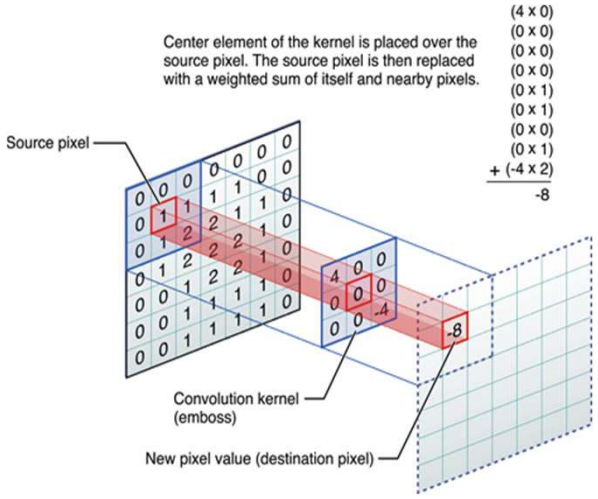
Nice Model and Optimized Implementation

Margin Optimization

Rank	Name	Error rate	Description
1	**U. Toronto**	0.15315	Deep learning
2	U. Tokyo	0.26172	Hand-crafted features and learning models. Bottleneck.
3	U. Oxford	0.26979	
4	Xerox/INRIA	0.27058	

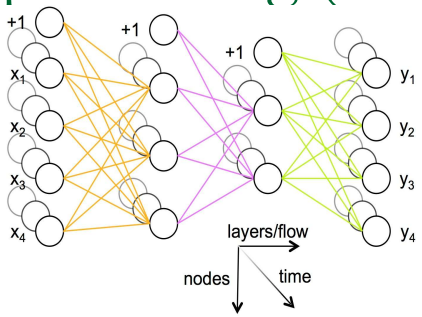
ImageNet 2012 Image Recognition Challenge Breakthrough

History: From Perceptron to Artificial Neural Networks to Deep Learning (3/4)

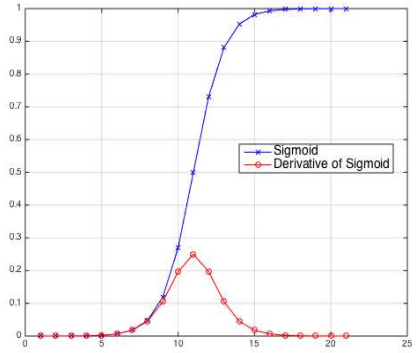


Convolutional Networks [Le Cun et al. 1998]

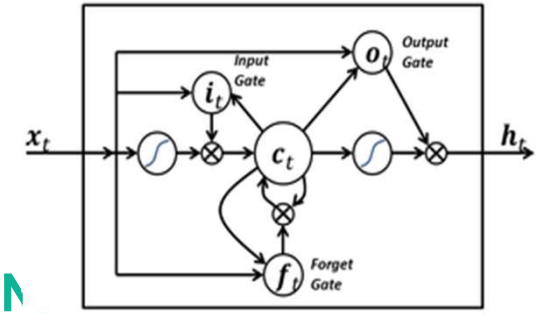
Equivariance (to translation) & Invariance (to small transformations)



Recurrent Neural Networks (RNN) (1986)
Temporal Invariance



Gradient Vanishing or Explosion (1991)



Long Short-Term Memory (LSTM) [Hochreiter & Schmidhüber 1997]

History: From Perceptron to Artificial Neural Networks to Deep Learning (4/4)

+ theano +

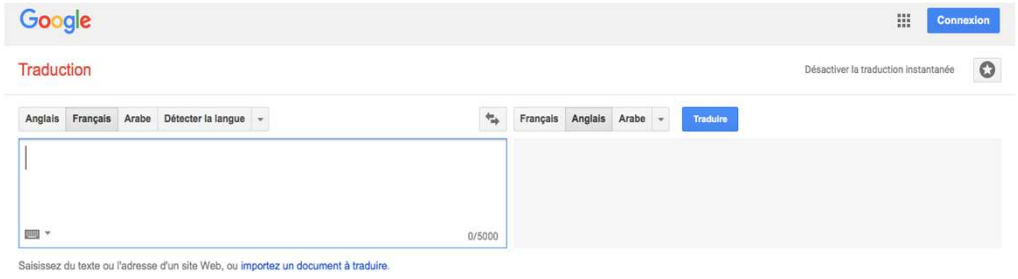
Massive Data Available

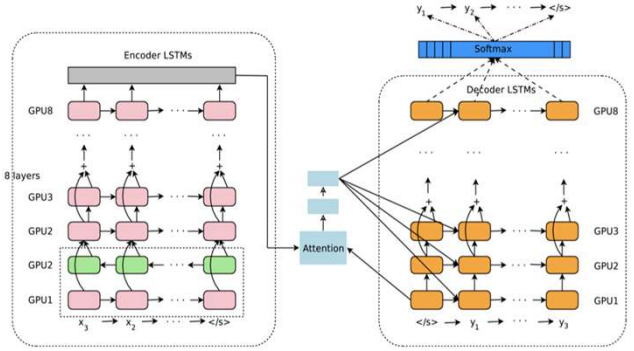
Efficient Implementation Platforms

Affordable Efficient Parallel Processing (Graphic Cards GPU)

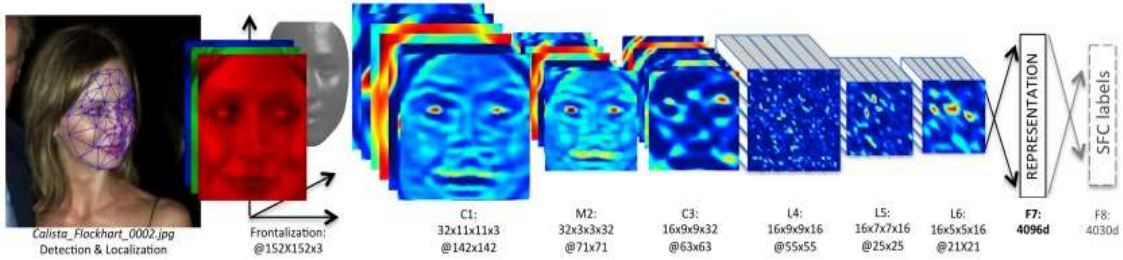
Deep Learning – Already There… And It Will Stay

- Translation, ex: Google Translate

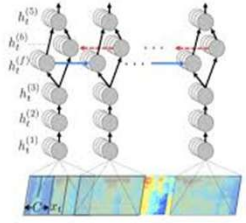




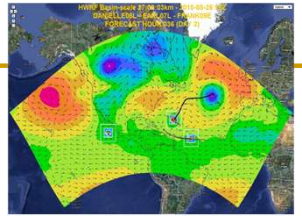
- Image Recognition, ex: Facebook



- Speech Recognition & Generation, ex: Apple Siri, Amazon Echo…

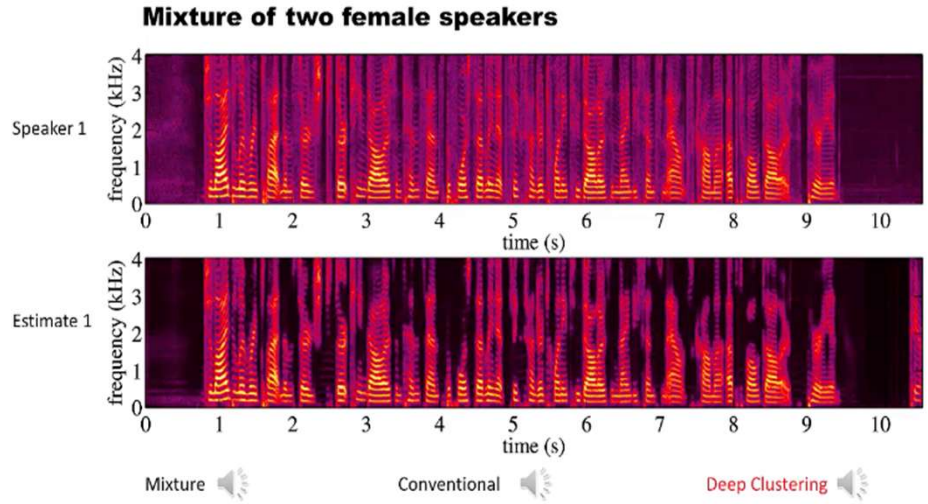


- Weather Prediction



Speech/Music Separation

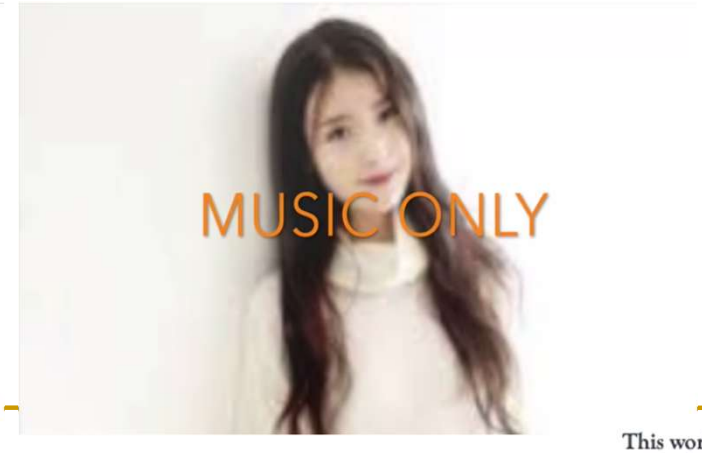
- Cocktail Effect Voice Separation



Mitsubishi Electric Research Labs (MERL)
Publicado em 1 de ago de 2016

https://www.youtube.com/watch?v=vW51cG1Ox98

- Music/Voice Separation



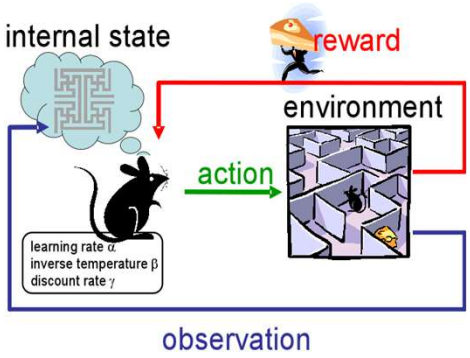
Hexa (Namju) Kim
11 inscritos

This work is from Jeju Machine Learning Camp 2017

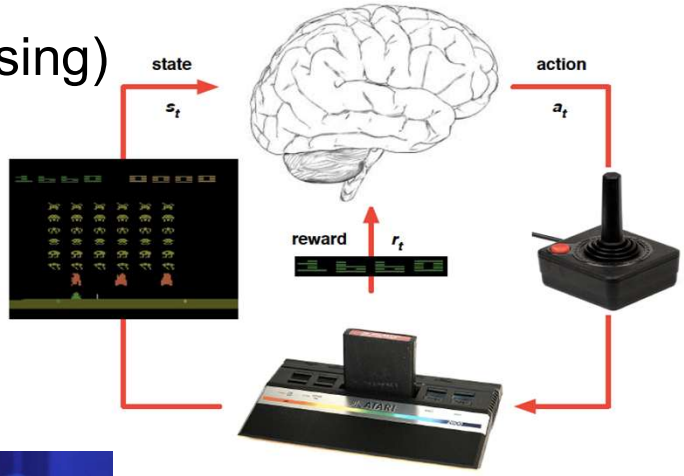
https://www.youtube.com/watch?time_continue=2&v=Cx7Me0Ayz1I

Deep Reinforcement Learning [Silver et al., 2013]

- Deep Learning improves Other Machine Learning Paradigm Implementation:
  - Reinforcement Learning



- Deep Reinforcement Learning
  - Efficient Estimation of Reinforcement Learning
  - Massive Simulation/Evaluation (Massive Processing)
  - Replay Mechanism (Massive Memory)



- First Application: Atari Games

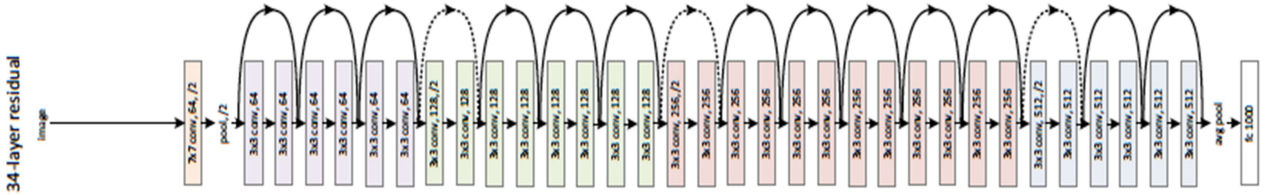
- Second Application: Go
  - Alpha Go, AlphaZero Go

Deep Learning

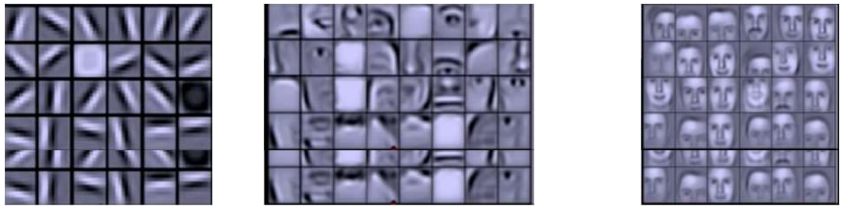
- Many Hiddden Layers
  - Deep Architecture
  - Deep Representation
- Various Types
  - Supervised learning
  - Unsupervised learning
  - Recurrent networks
  - Stochastic Neurons (ex: Restricted Boltzmann Machines)
- Various Objectives
  - *Classification*
  - *Prediction*
  - Feature extraction
  - Generation
    - » Generative Adversarial Network
    - » Variational Autoencoder

Deep?

- Many Layers



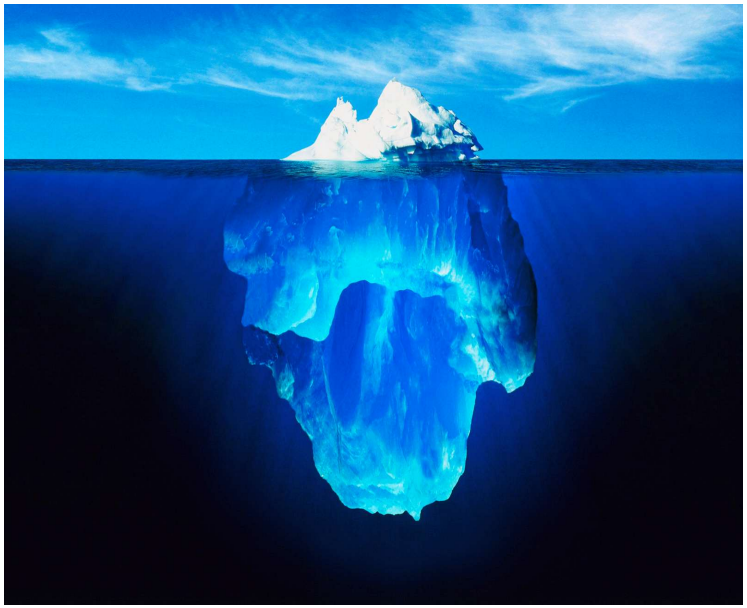
- Many Levels of Abstractions

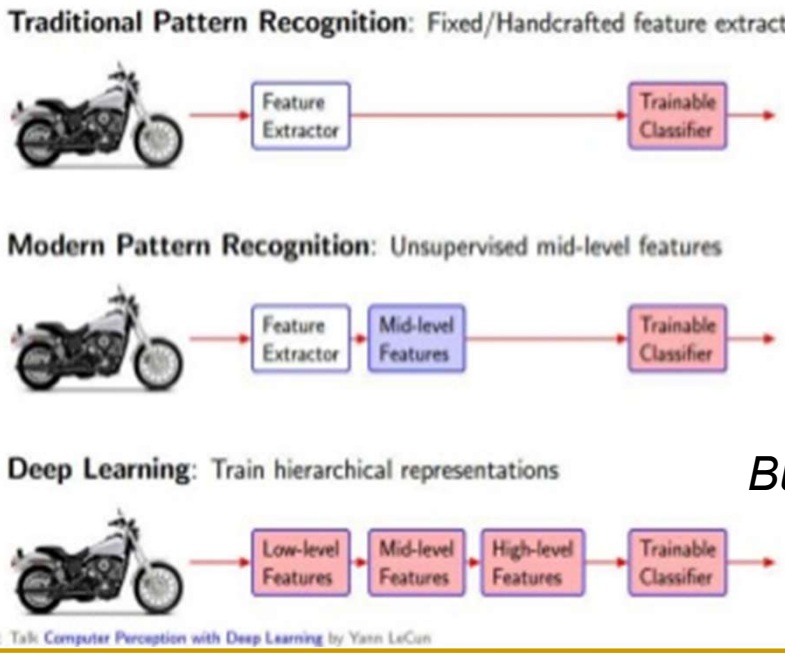


- Many Input Parameters
- Many HyperParameters
- Many Examples (Dataset)
- Many Processing Cores

Deep?

- More Complex Models
- Learns better Complex Functions
- Hierarchical Features/Abstractions
- No Need for Handcrafted Features
  - (Automatically Extracted)





End-to-End Architecture

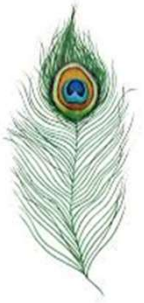
Distributed Representations

But the choice of the input representation is important

- Ex: for Audio:
  - Waveform or Spectrum
- Ex: For Symbolic representation
  - Additional information, ex: beat, fermata, enharmony (Cb =/= B)

Ex. of Handcrafted Features – Bag of Features [Brendel, 2019]

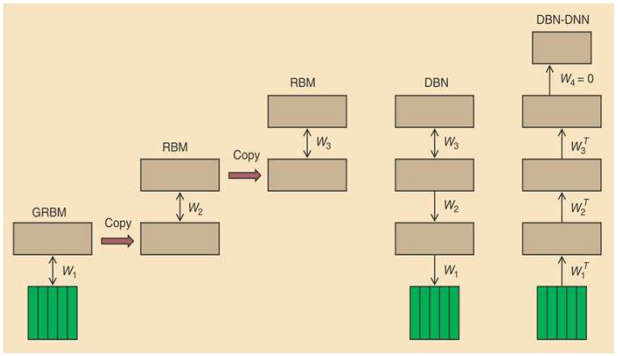
- Bag of Features (BoF) Model
- Define key Visual Features
  - Ex: Human eye, Feather



- Count how often each Feature is Present in the Image

- Classify based on these Numbers
  - Human: Eye = 2
  - Bird: Feather = 18

The Shocking Start of Deep Learning

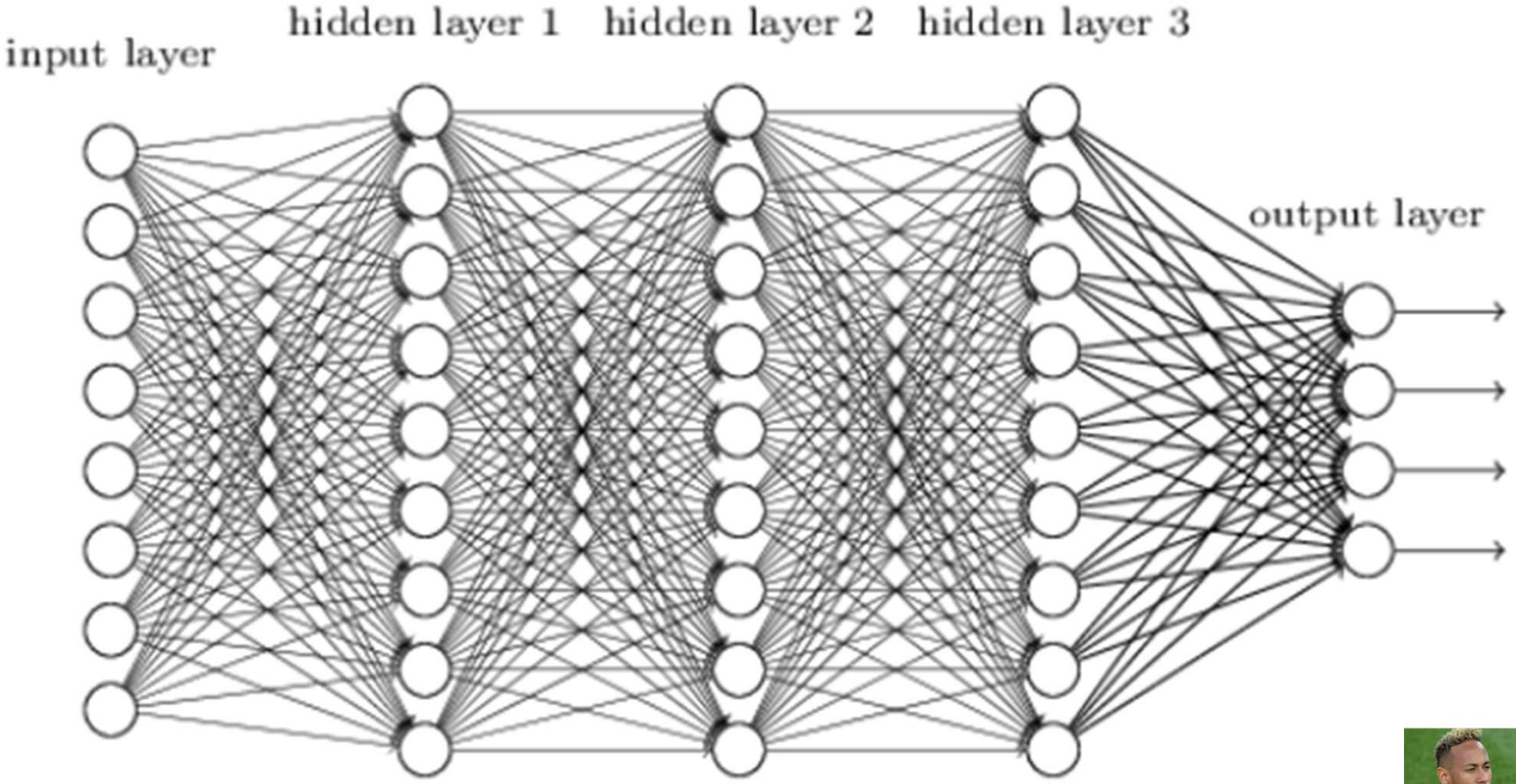


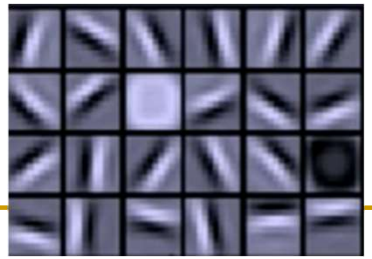
Pre-Training [Hinton et al. 2006]
Layer-Wise Self-Supervised Training/Initialization

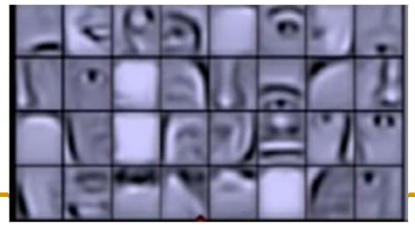
Rank	Name	Error rate	Description
1	**U. Toronto**	0.15315	Deep learning
2	U. Tokyo	0.26172	Hand-crafted features and learning models. Bottleneck.
3	U. Oxford	0.26979	
4	Xerox/INRIA	0.27058	

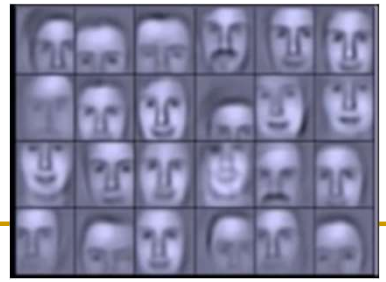
ImageNet 2012 Image Recognition Challenge Breakthrough

Successive Features/Abstractions Constructed

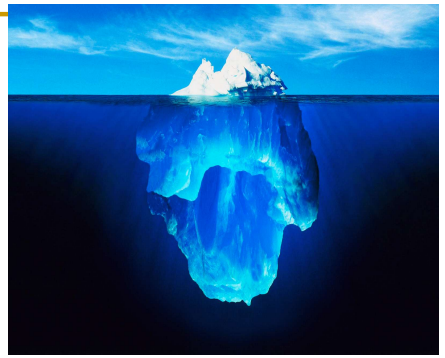






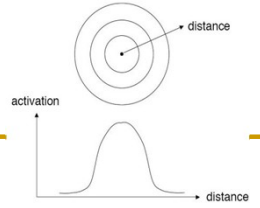


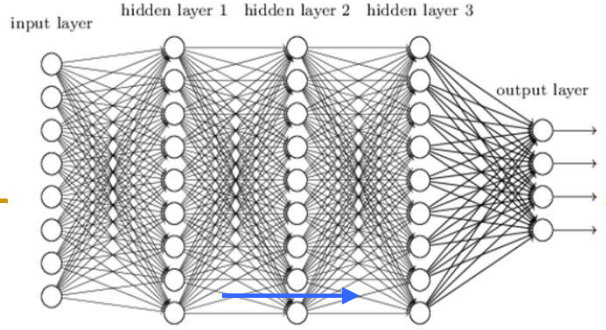
Why Deep?



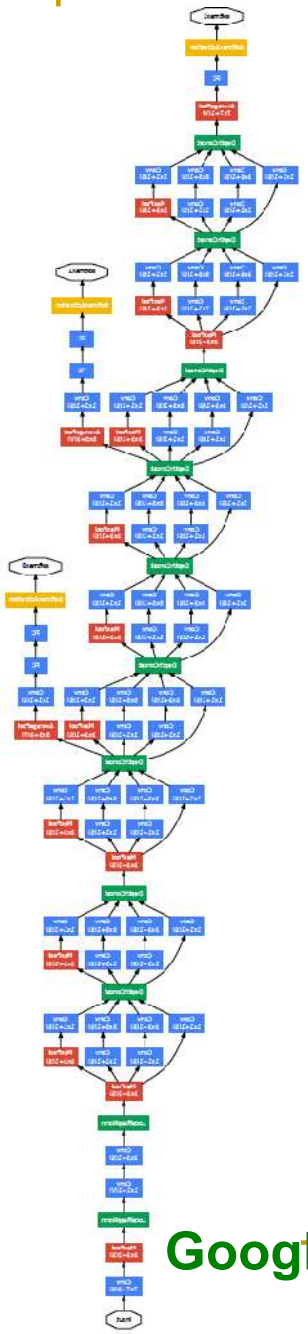
- Need for Deep ?! [Ba & Caruana 2014]
  - Train a Shallow Net to Mimic a Deep Net
  - Mimic = Train on the Logits (predicted values, before softmax/probabilities), not on the Original Output Data
  - Equivalent Accuracy
  - But for Convolutional Nets, Mimic Net Does Not Match the Initial Accuracy ! [Urban et al.& Caruna 2016]

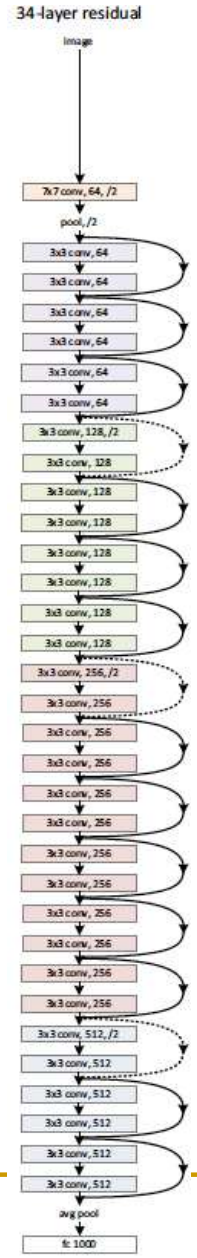
- Counter Argument (Theorem) [Eldan & Shamir 2016]
  - There is a simple radial function on R^d, expressible by a 3-layer net, but which cannot be approximated by any 2-layer net to more than a constant accuracy unless its width is exponential on the dimension d
  - Depth → vs/and Width ↓





Very Deep Learning

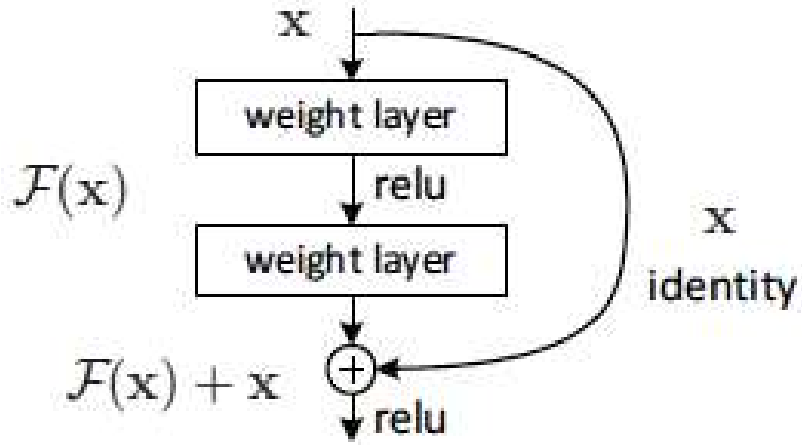




(Very) Deep Networks

Upto 152 Layers !

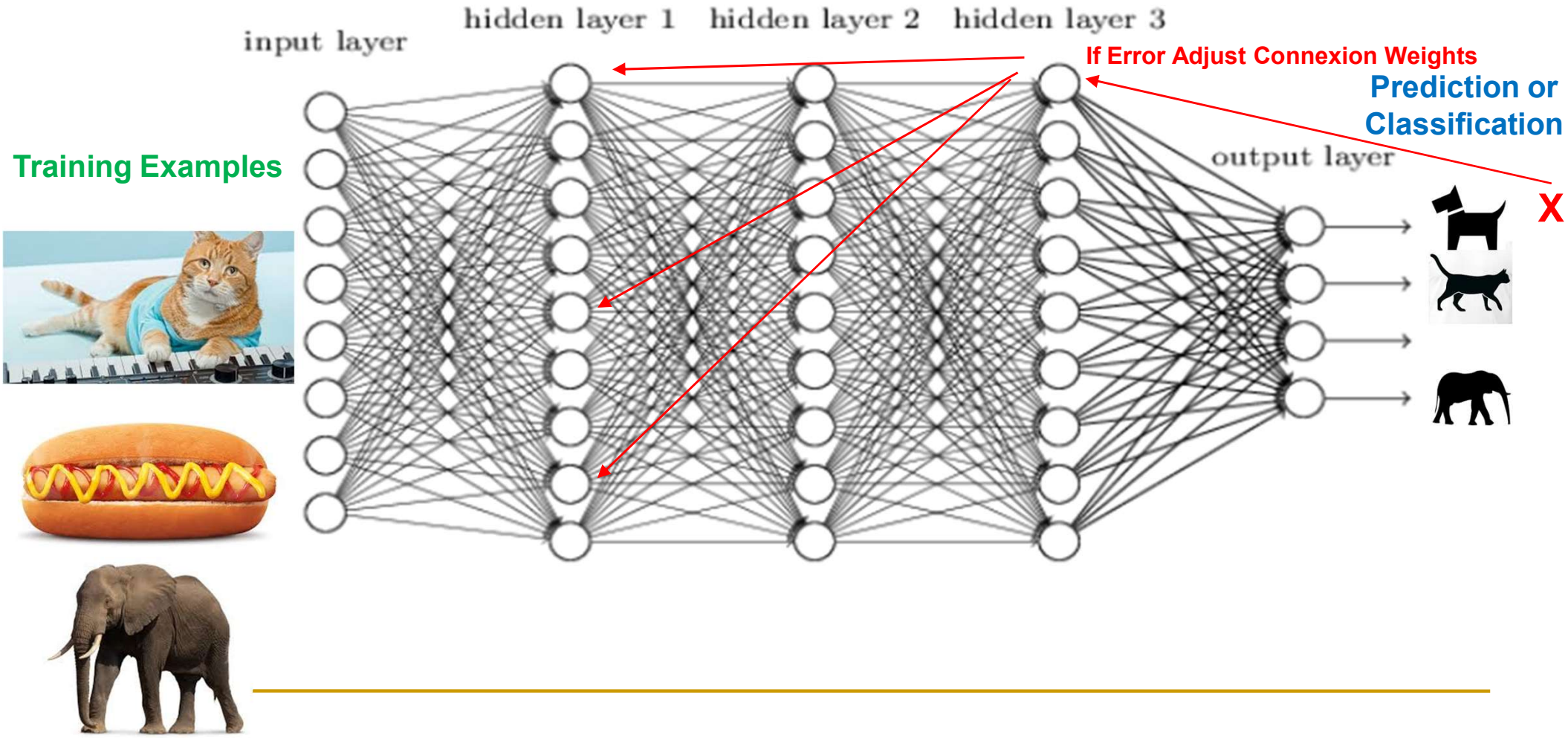
New Techniques (Tricks ?!) e.g.,
Batch Normalization [Ioffe & Szegedy, 2015]
Deep Residual Learning [He et al., 2015]
Replaced Pre-Training (less in vogue)



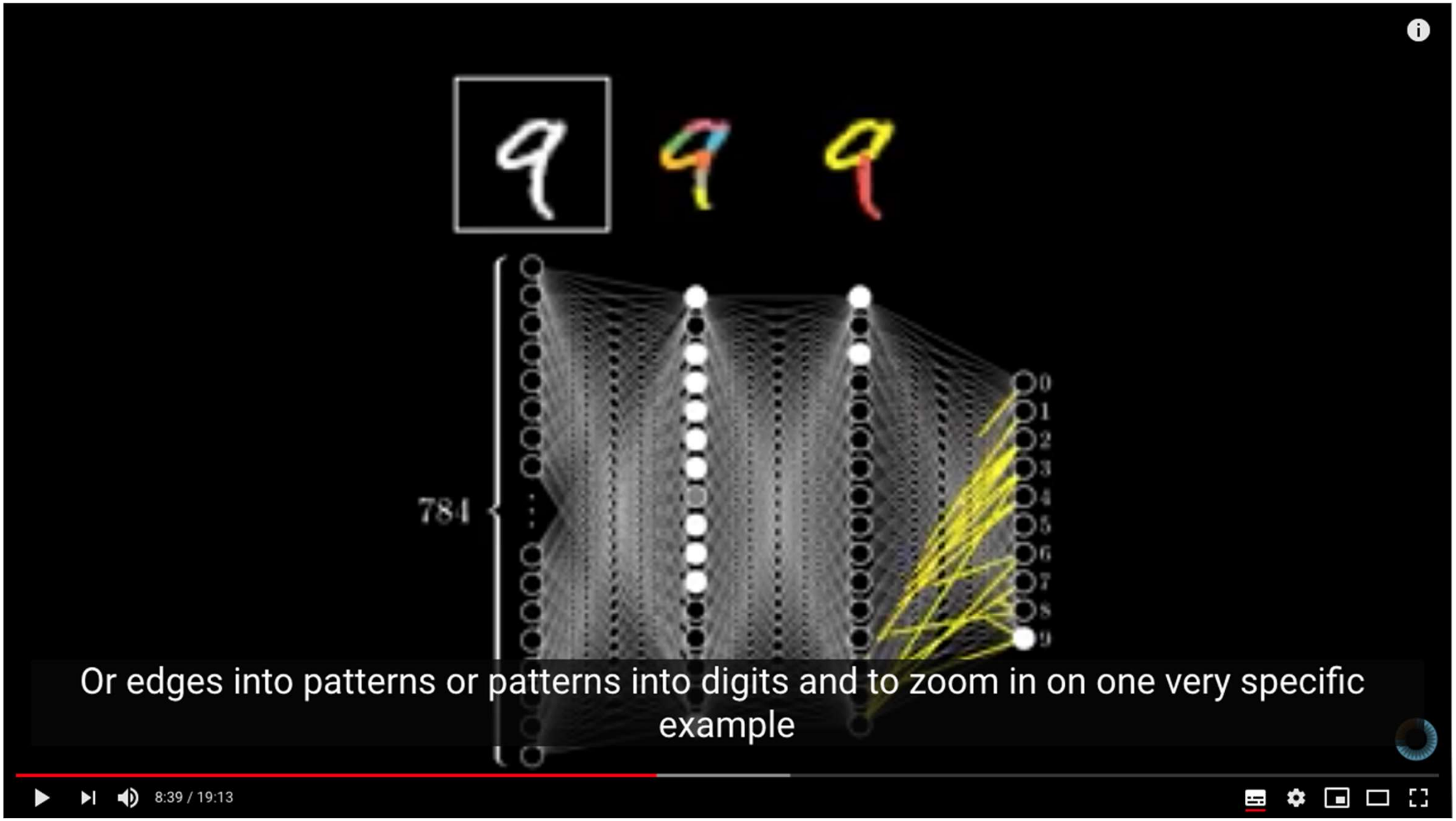
GoogLeNet (2014) **ResNet (2015)**

Neural Networks in One Slide

Principle – Error Prediction/Classification Feedback



Nice Animation [3Blue1Brown]



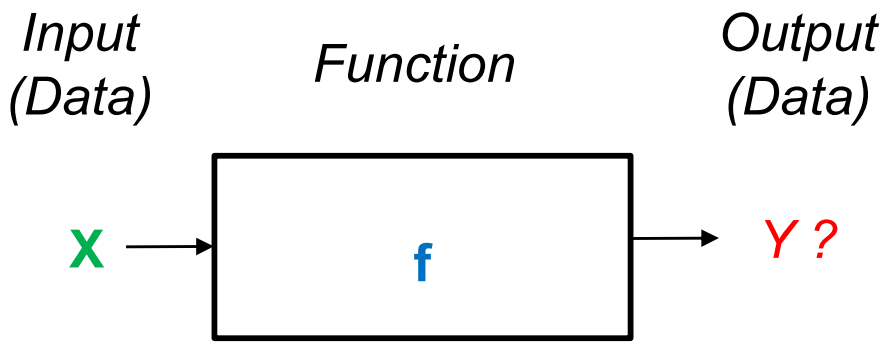
https://www.youtube.com/watch?v=aircAruvnKk

Machine Learning

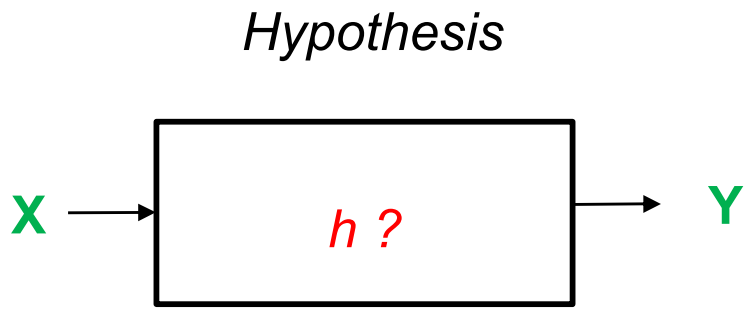
- Why learning ?
- What is machine learning ?
- A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E [Mitchell, 1997]

Machine Learning (Data Oriented/Supervised Learning)

- Traditional Programming



- Machine Learning (Supervised Learning)



Function/Module Configuration (*Pre-Existing Representation Model*)

Machine Learning (Data Oriented)

- 3 Components [Domingos 2015]:

- Representation
  - To Model Predictors/Classifiers
  - e.g., Decision Tree, Rule, Neural Network, Graphical Model…

- Evaluation
  - of Predictors/Classifiers (Accuracy)
  - Cost, e.g., Squared Error, cross-entropy…
  - Precision, Recall

- Optimization
  - To Search among Predictors/Classifiers,
  - e.g., Batch Gradient Descent, Greedy Search, Stochastic Gradient…

Machine Learning Typology

- Learning = Using Experience/Memory
  to Infer Information/Behavior
  and to Improve Decision/Action

- Dynamicity
  - Off-Line (Training Phase vs Using/Production Phase)
    - » Ex: Supervised Learning
      - But also Incremental versions
  - On-Line/Incremental
    - » Ex: Reinforcement Learning
      - But also Off-line versions (ex: Deep Reinf. Learning Replay Mechanism)

- Provided Information
  - Supervised (Correct Output)
  - Unsupervised (Ø)
  - Reward-Based (Quality assessment)

- Various Forms of Machine Learning

- Statistical
- Neural Networks, Bayesian Networks, Clustering…

- Decision
- Reinforcement learning

- Symbolic – Learning Concepts from Examples
- Inductive Logic Programming (ILP)

- Learning and Adapting from Cases
- Case-Based Reasoning

Machine Learning and Artificial Intelligence

- Machine Learning is Part of Artificial Intelligence Techniques

But also:

- Reasoning
- Planning
- Knowledge Representation
- User Modeling and Interaction
- Collaboration (Multi-Agent Systems)

- Natural Language Processing
- Dialogue
- Speech Processing

- Decision
- Game Theory
- Optimization

- Robotics

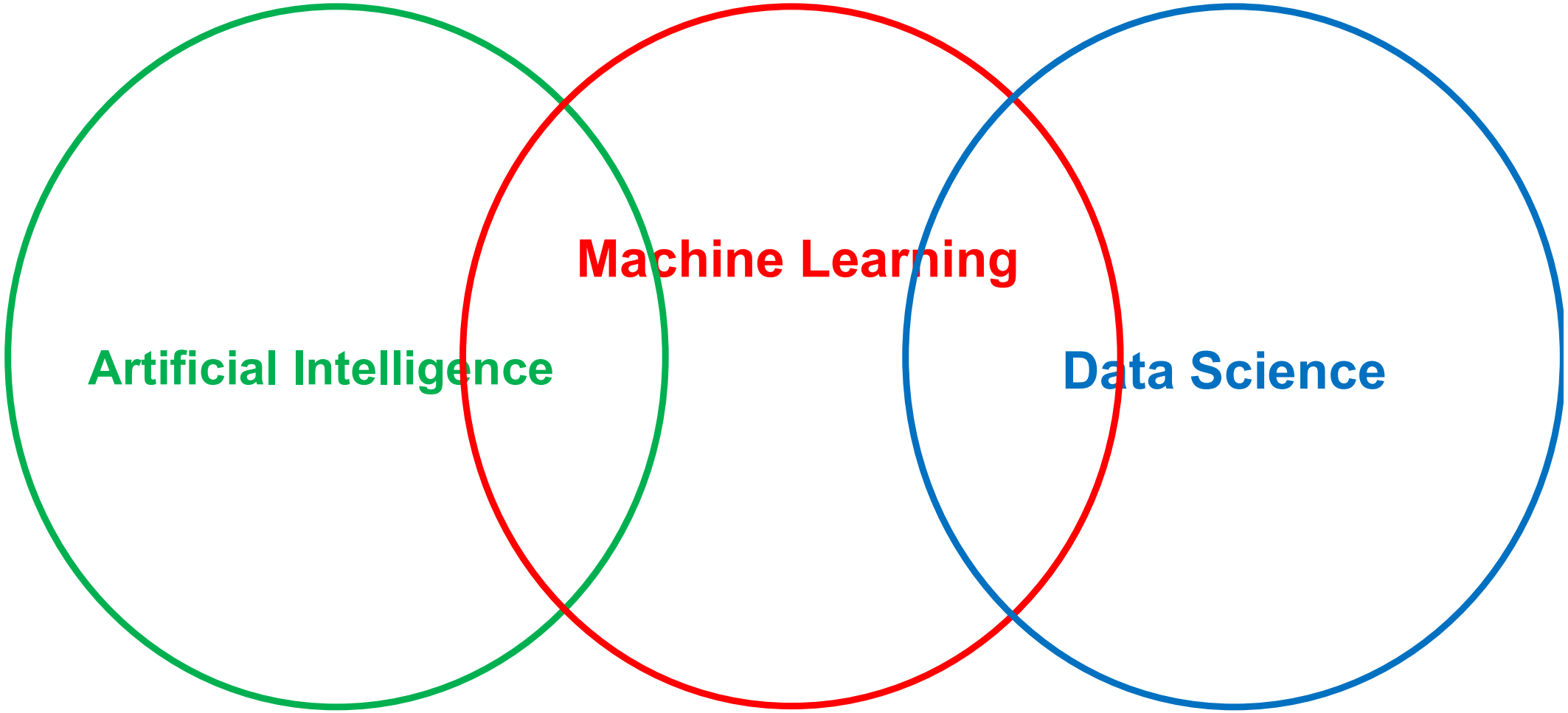
Machine Learning and Artificial Intelligence

- Backfire (Irony) of History
- In 1960, Minsky and Papert founded AI (Artificial Intelligence) based on Concepts, Symbols, Logic, Reasoning…, Against Cybernetics (Feedback) and Connexionism (Neural Networks)
- In 1969, they "Killed" Connexionism/Neural Networks (Sound Critic of Perceptron)

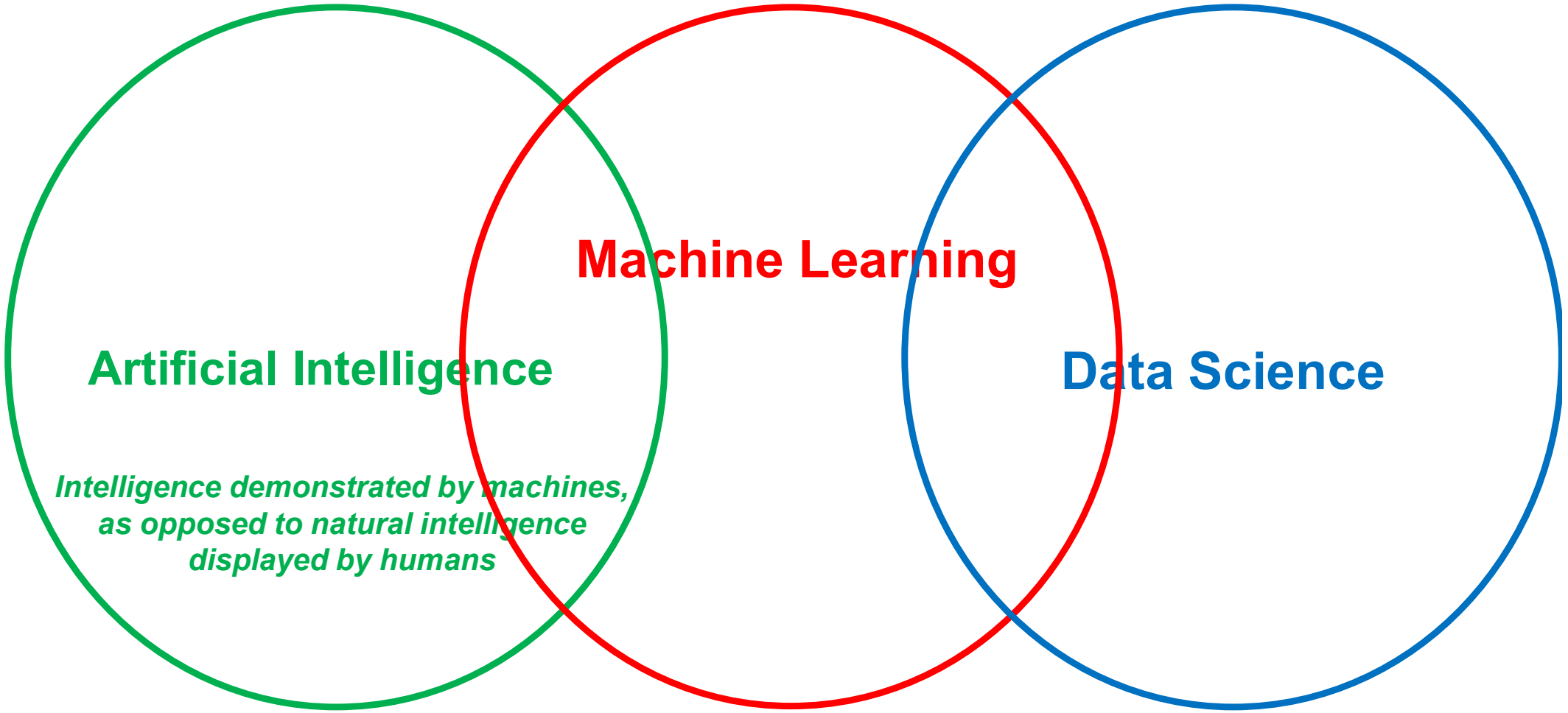
- In **2006**, Start of Deep Learning

- Now, AI is synonym of Deep Learning
- When Actually, Neural Networks are somehow based on Statistical (Correlation) Brute Force

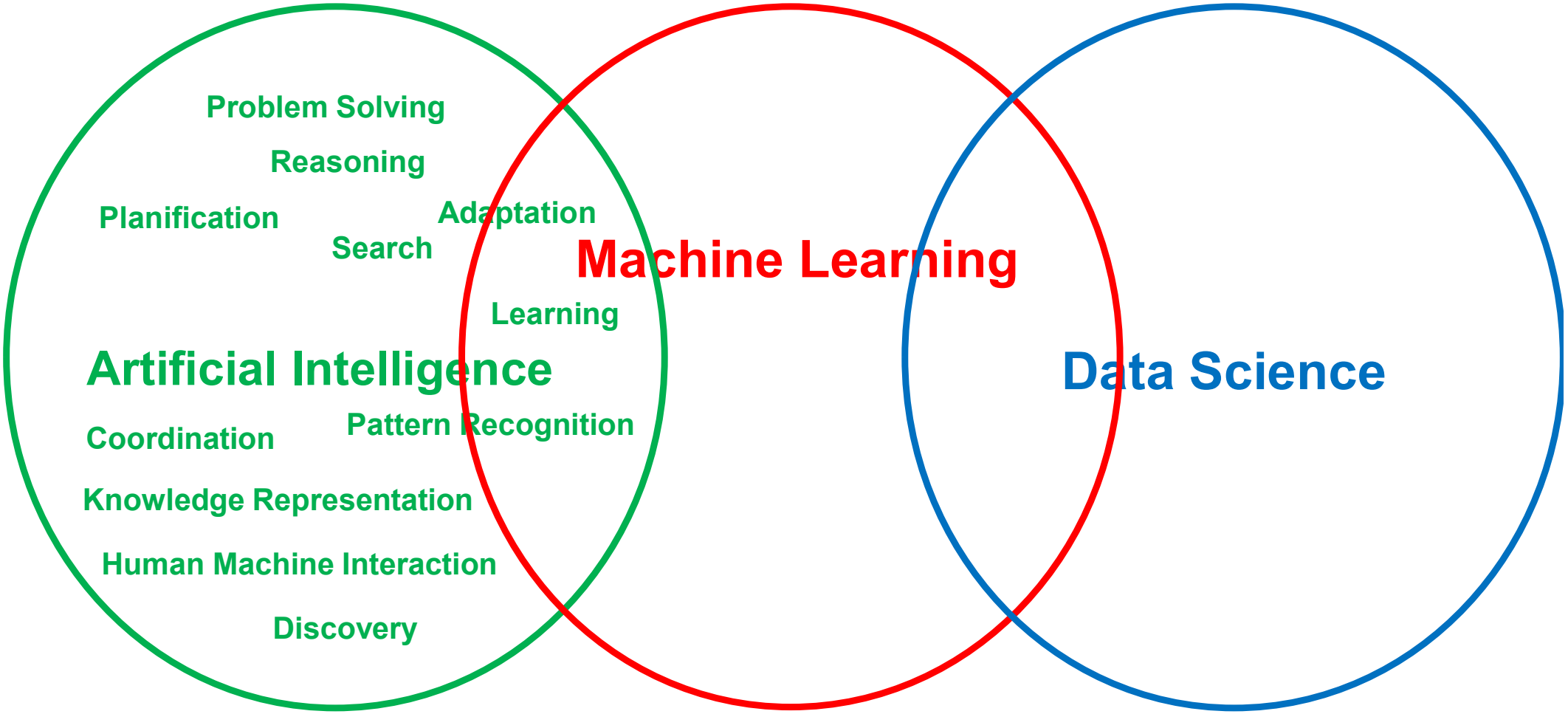
Terminology



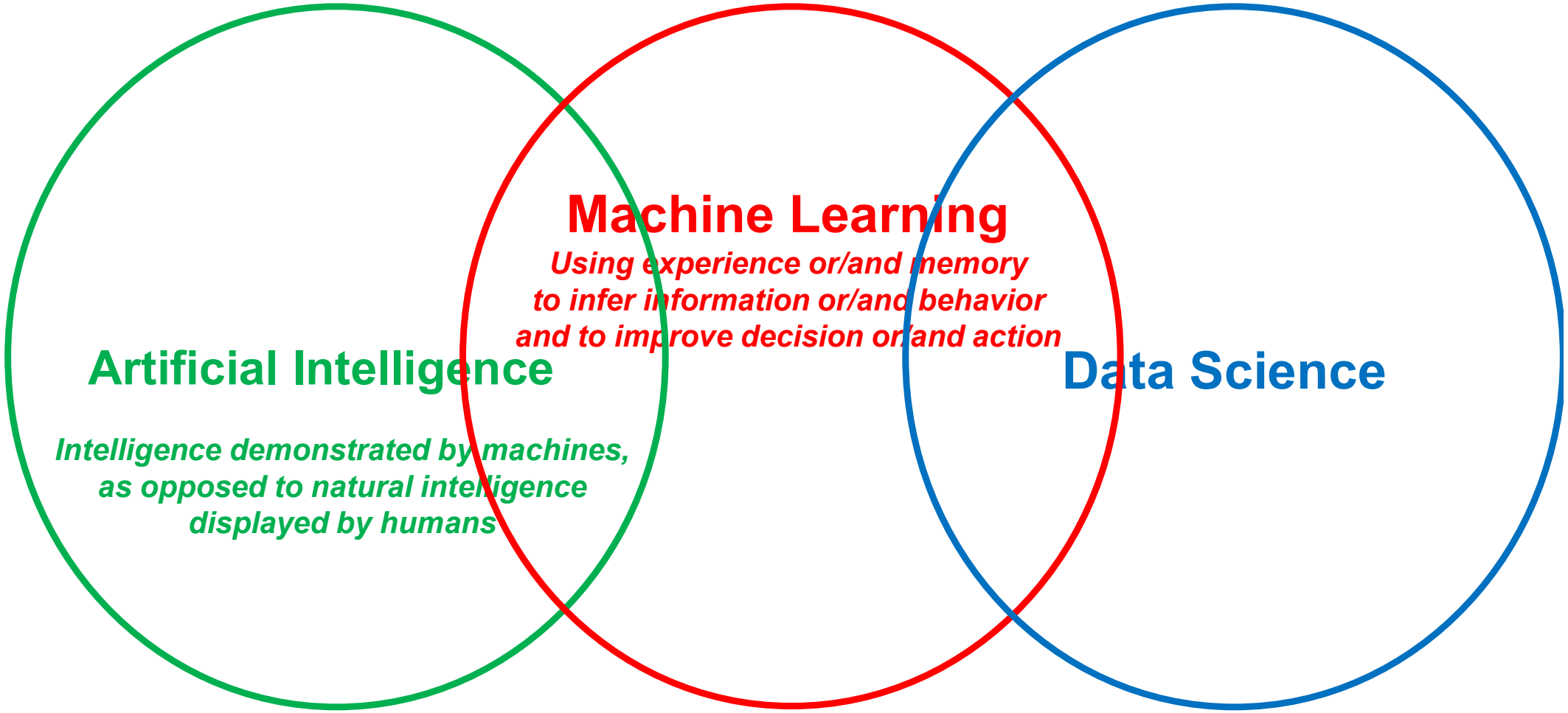
Terminology



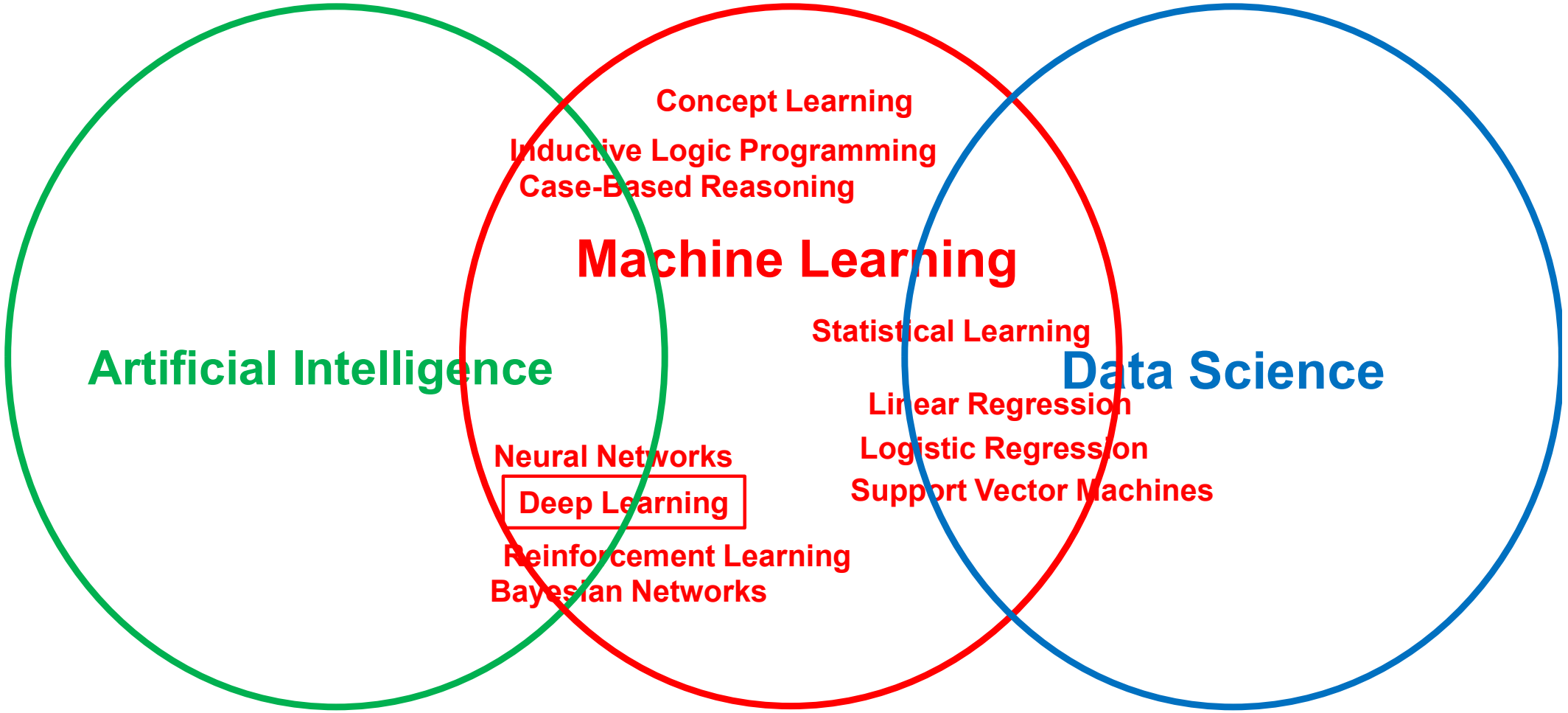
Terminology



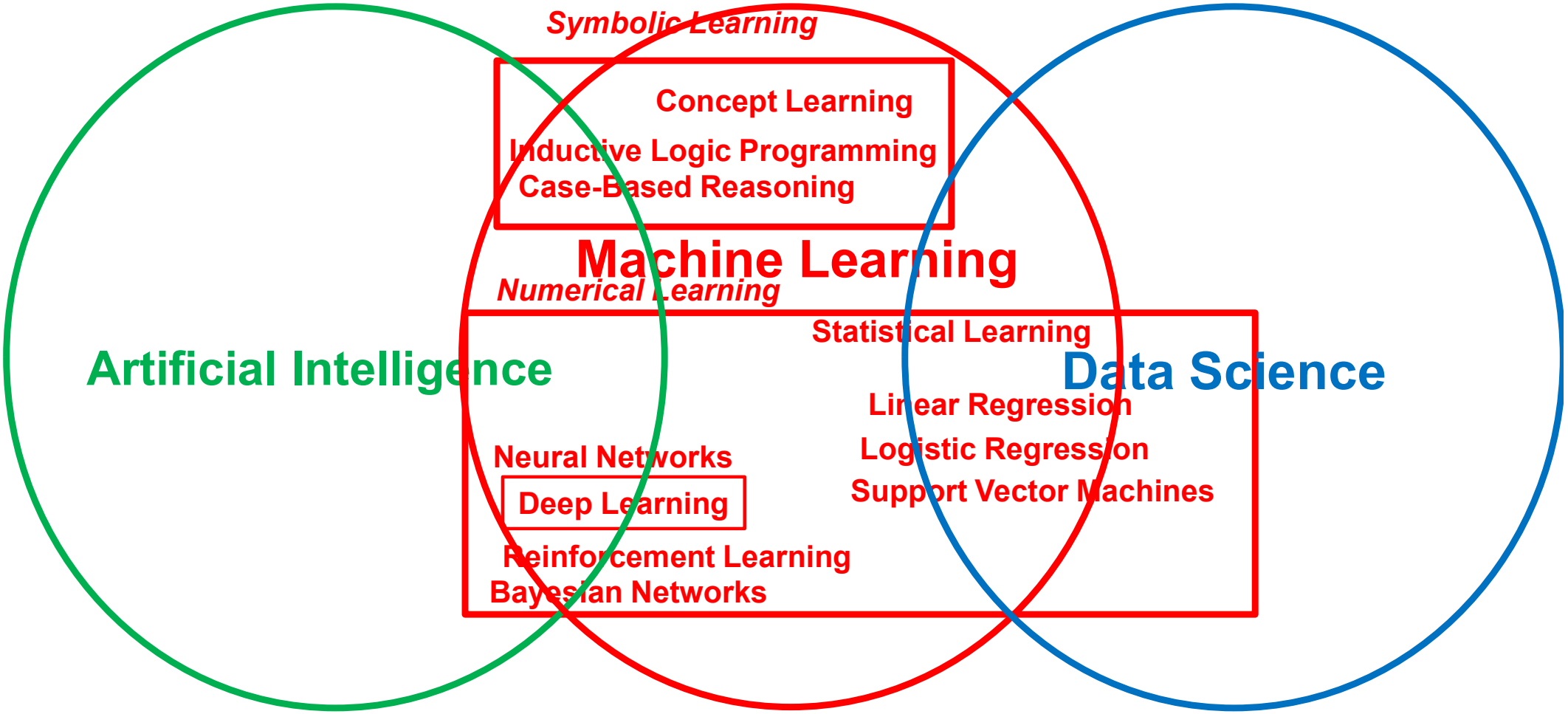
Terminology



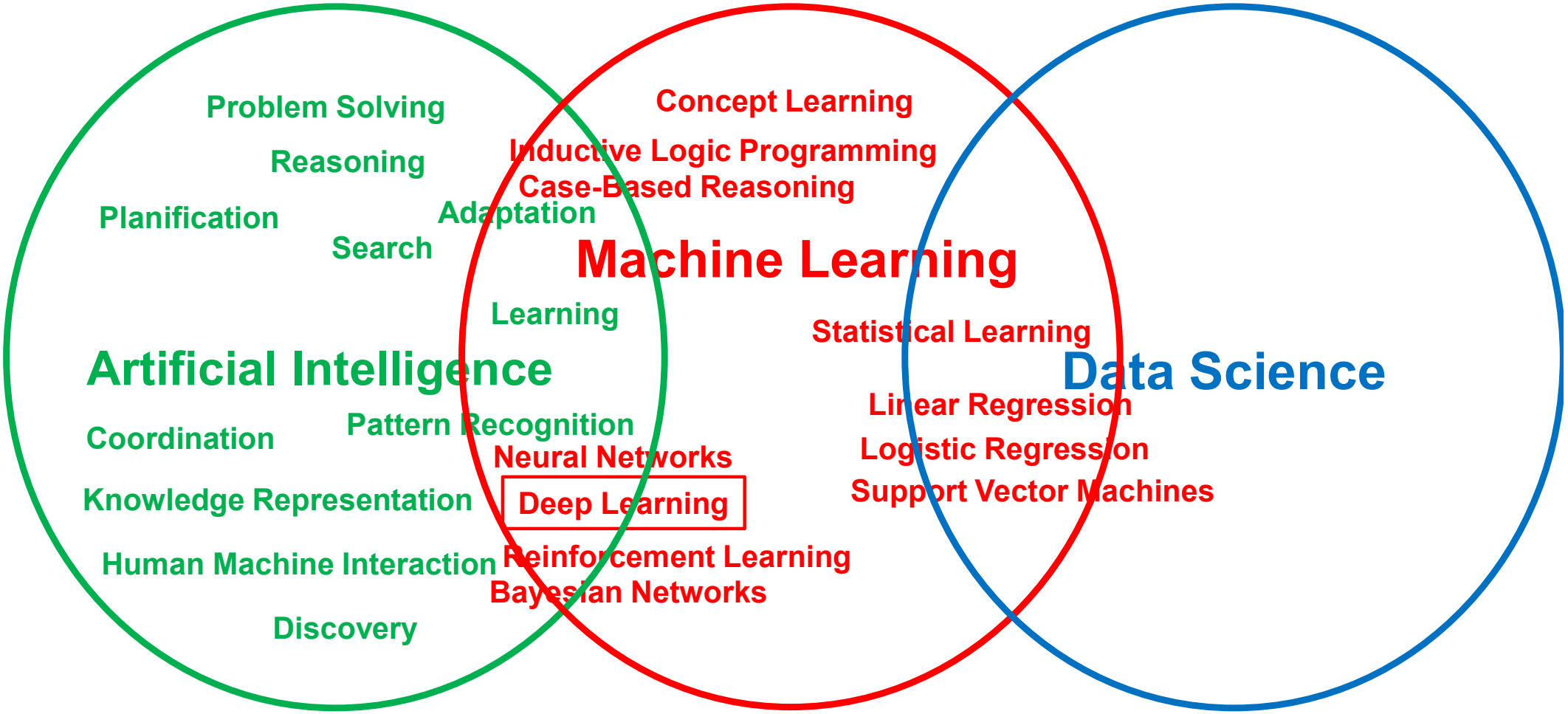
Terminology



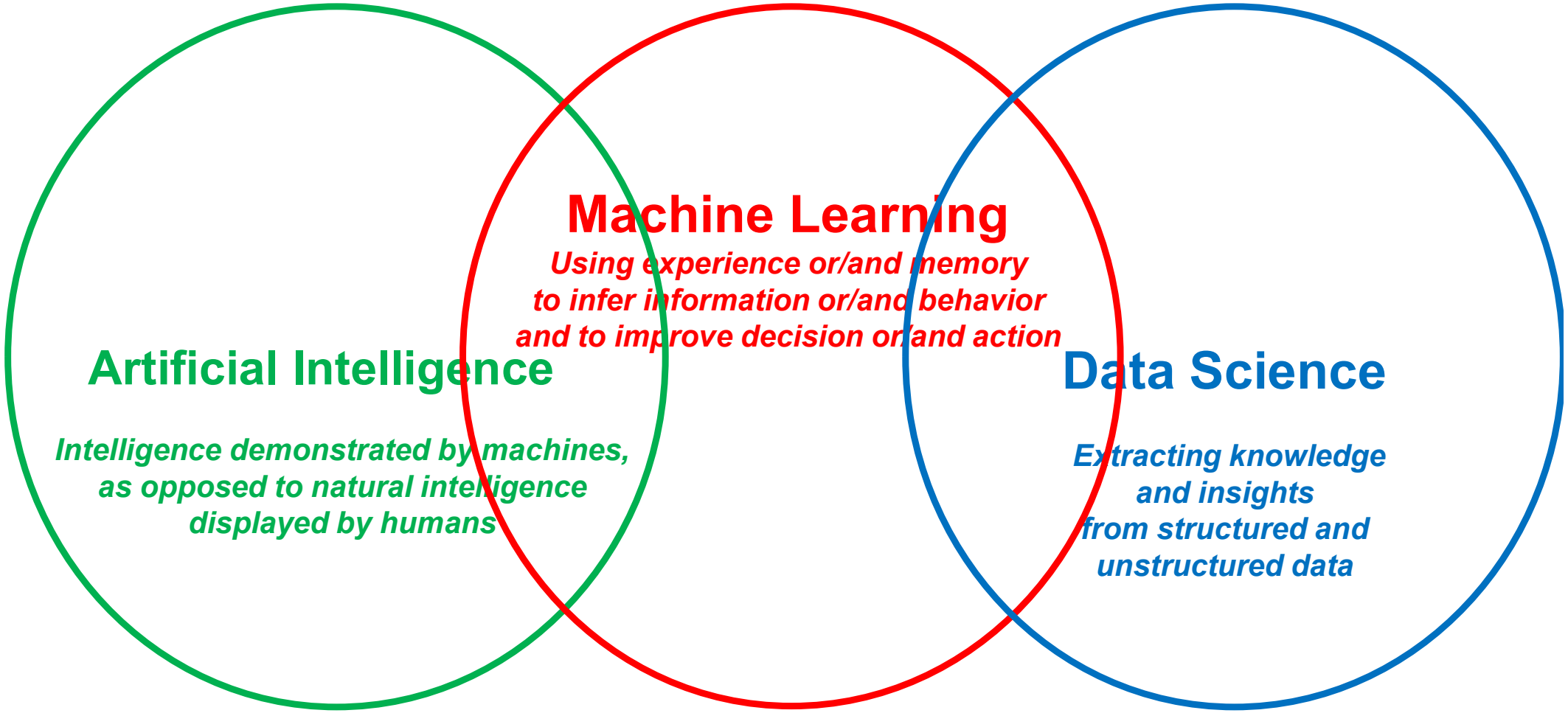
Terminology



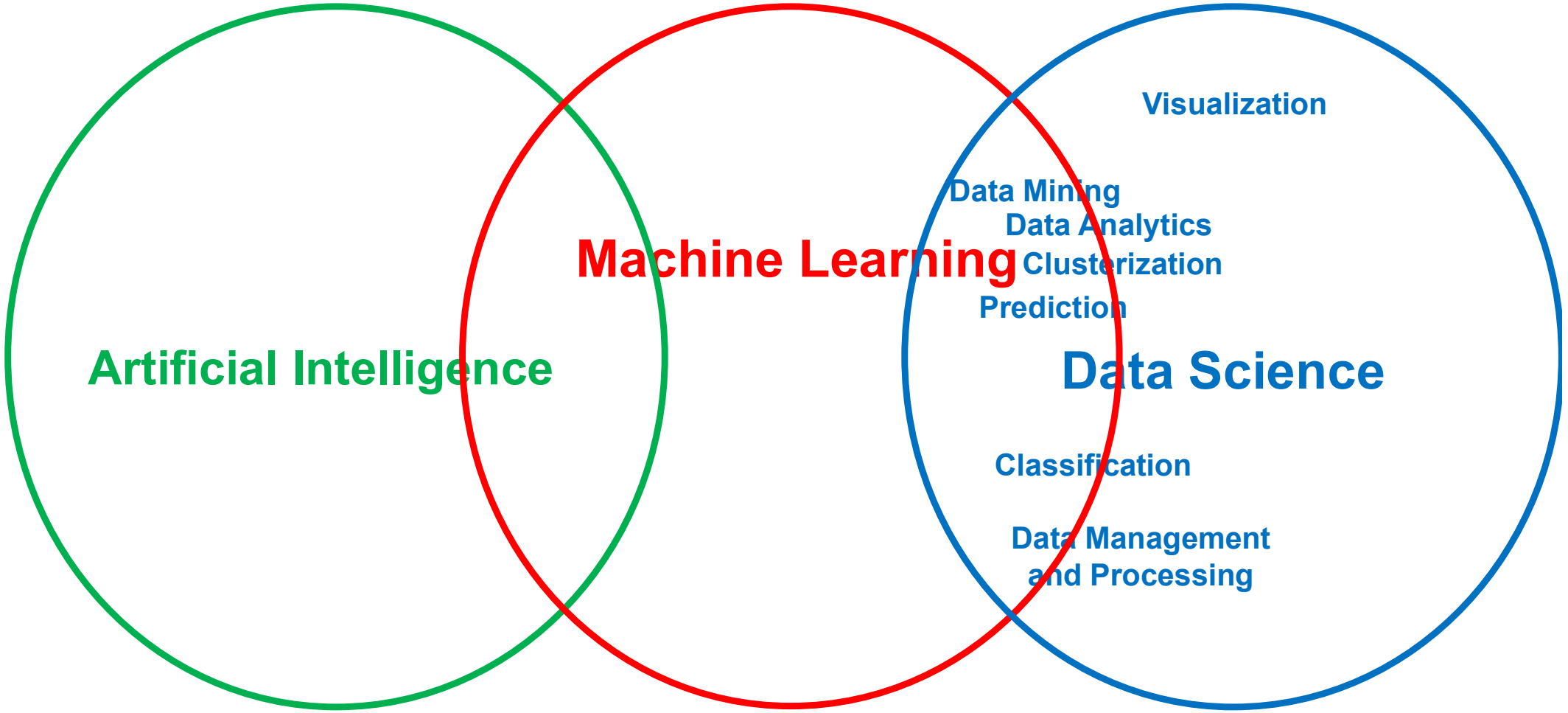
Terminology



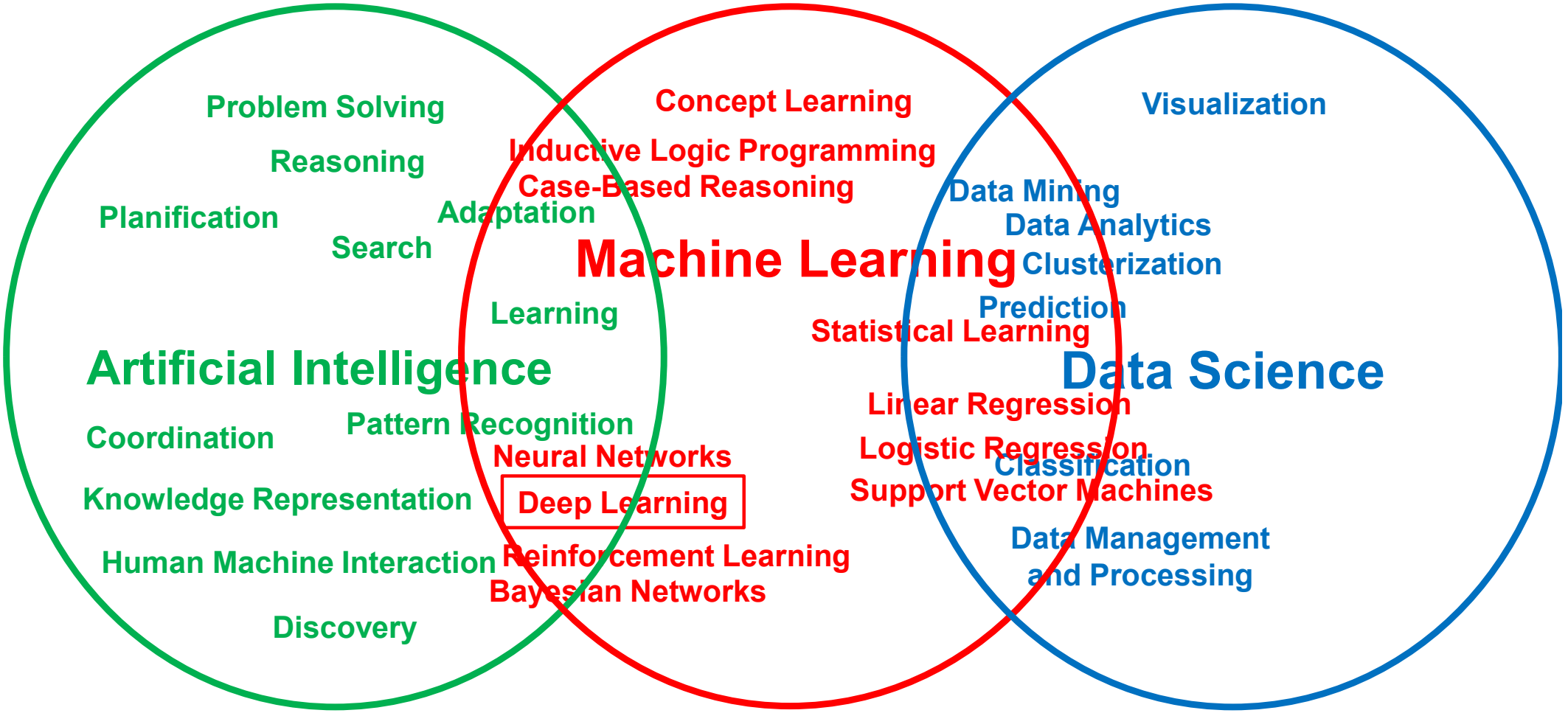
Terminology



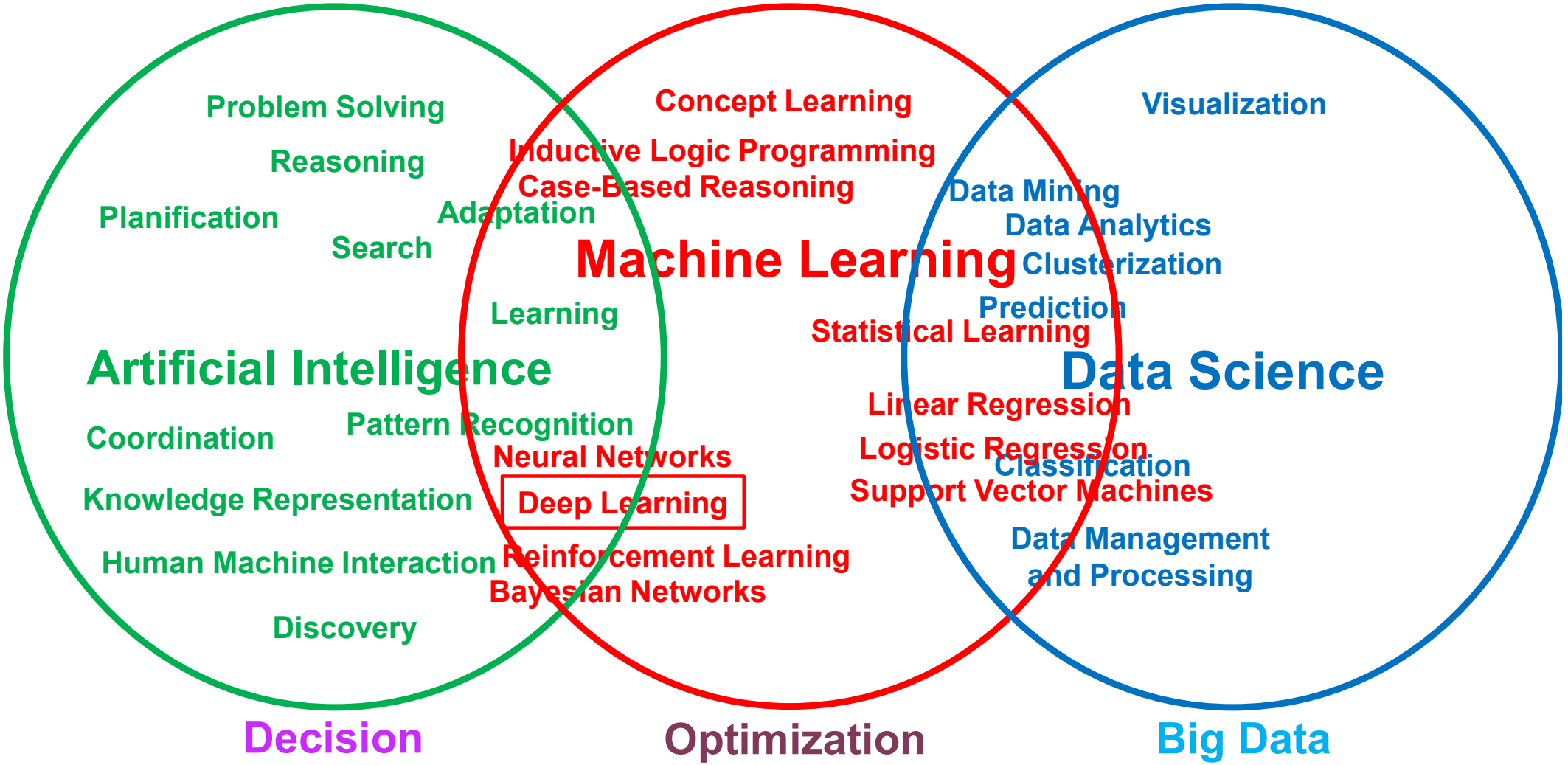
Terminology



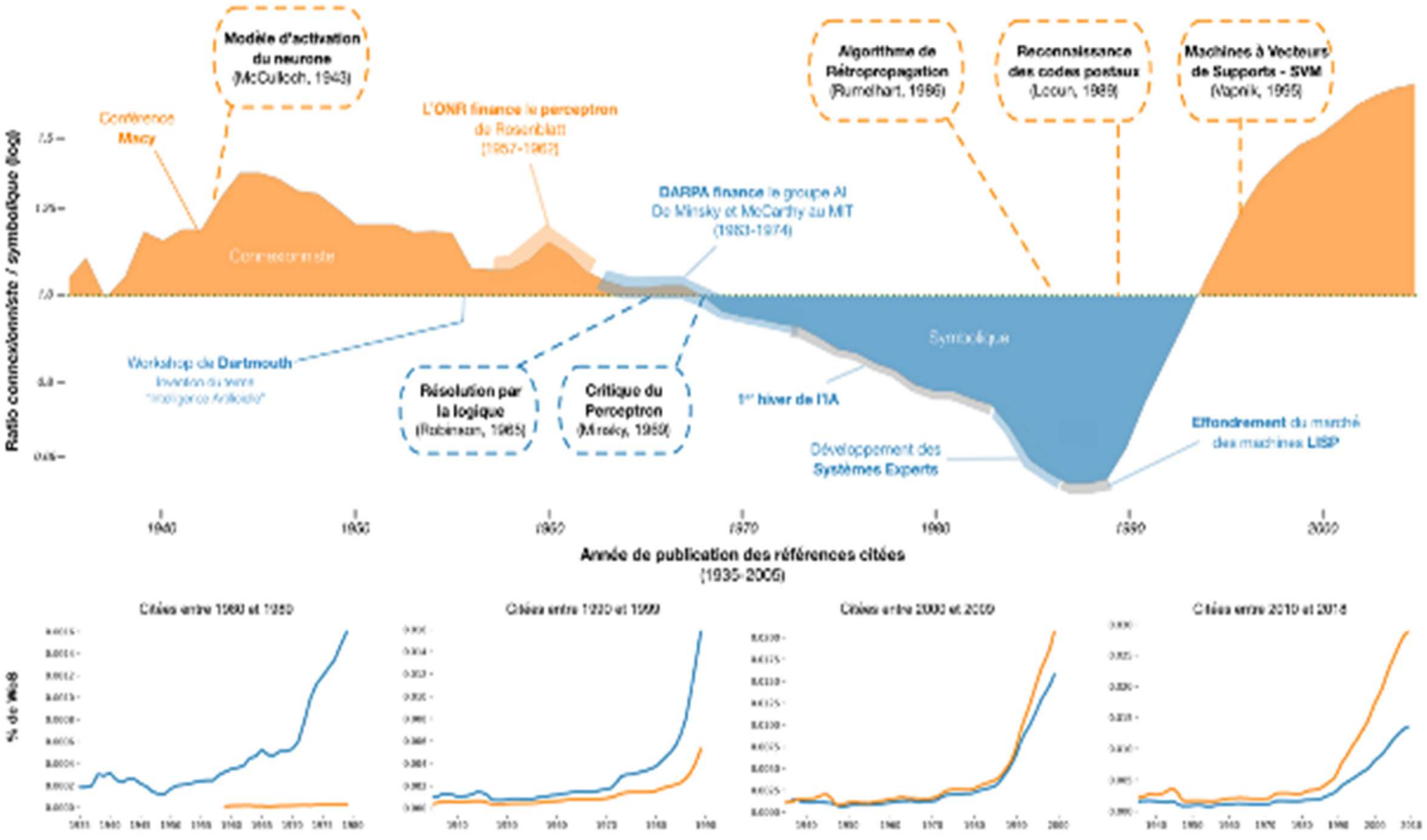
Terminology



Terminology



Symbolic vs Connexionist AI – History



[Cardon et al., 2018]

Machine Learning and Artificial Intelligence

- Machine Learning is Part of Artificial Intelligence Techniques

But also:

- Reasoning
- Planning
- Knowledge Representation
- User Modeling and Interaction
- Collaboration (Multi-Agent Systems)

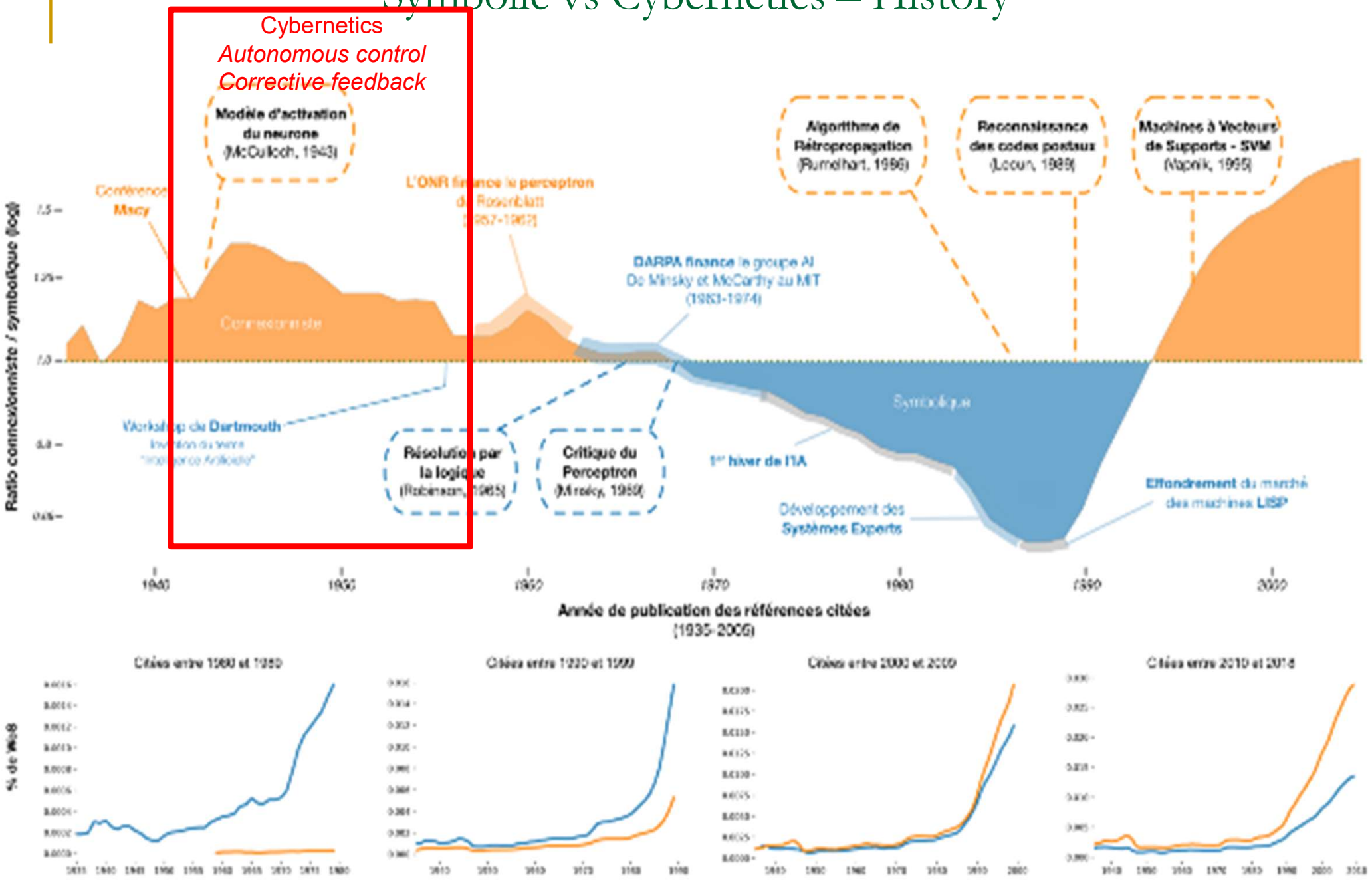
- Natural Language Processing
- Dialogue
- Speech Processing
- Decision
- Game Theory
- Optimization
- Robotics

Backfire of History

In 1960, Minsky and Papert founded AI and in 1969 "killed" Connexionism/Neural Networks

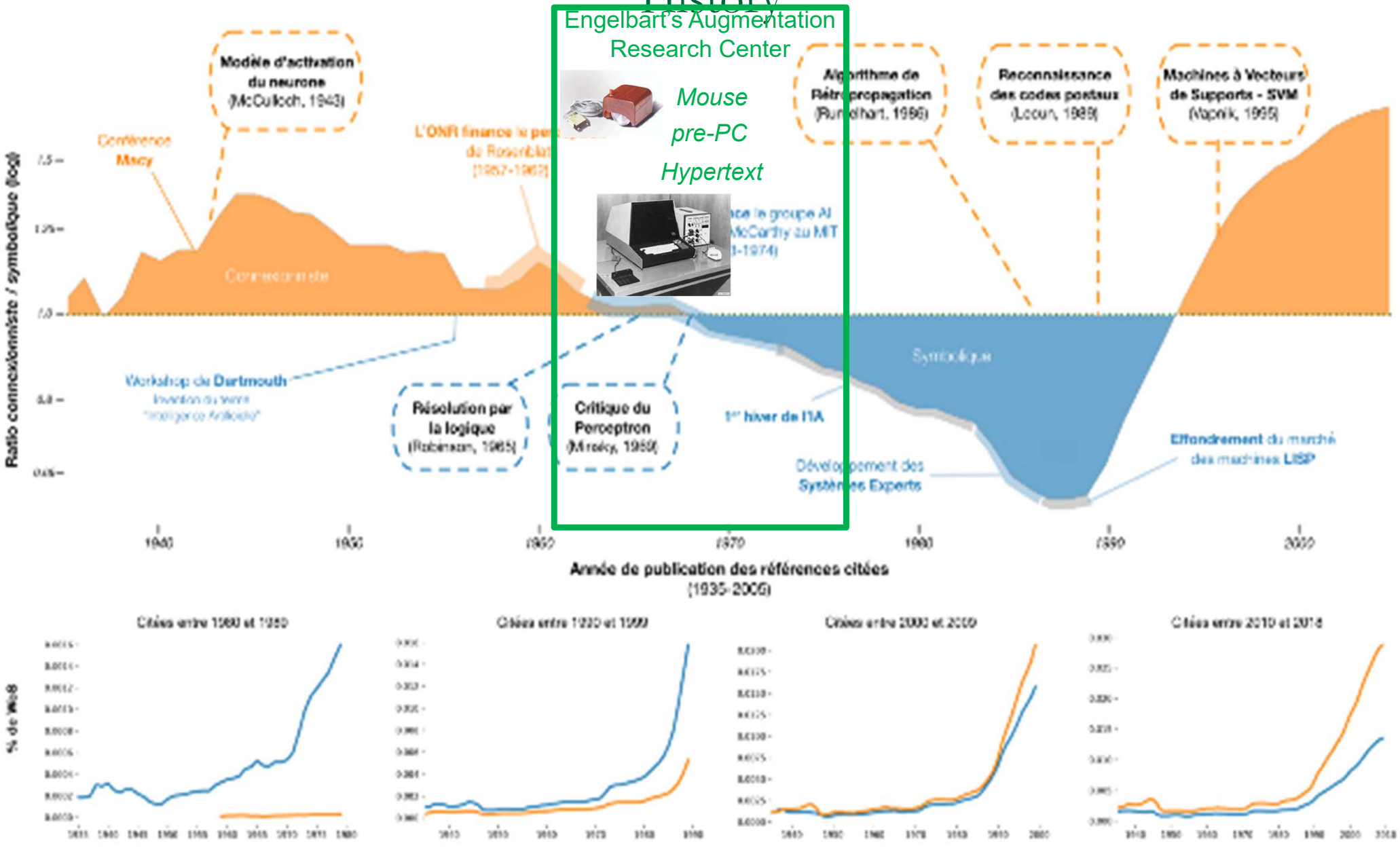
In 2015's, AI is synonym of Deep Learning

Symbolic vs Cybernetics – History



[Cardon et al., 2018]

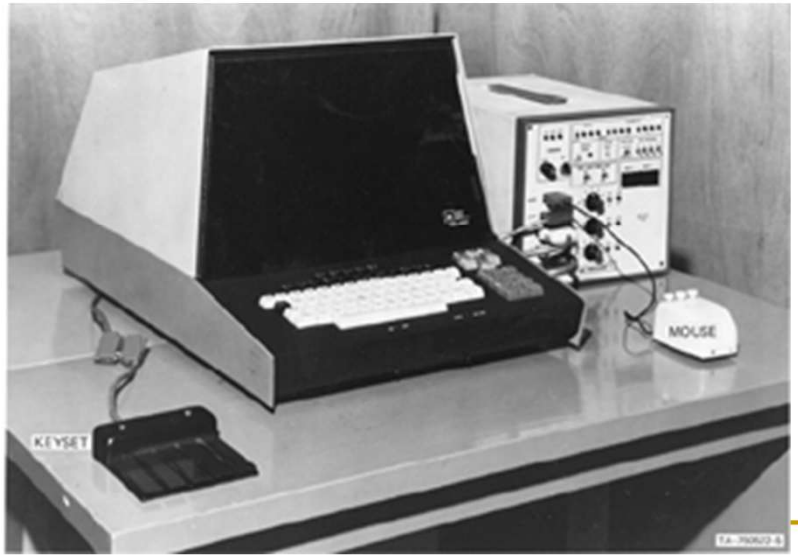
AI (Artificial Intelligence) vs IA (Intelligence Augmentation) – History

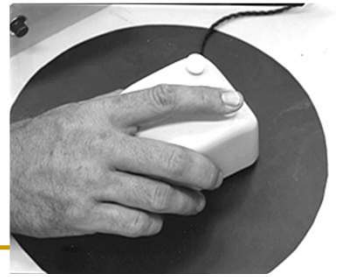


[Cardon et al., 2018]

Engelbart's Augmentation Research Center

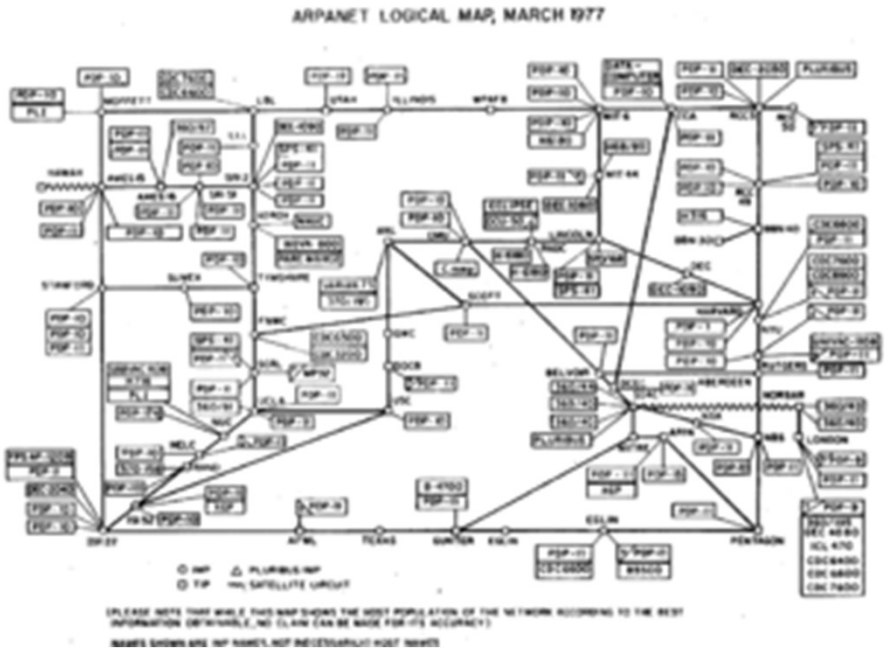
- Augmentation Research Center at Stanford Research Institute (SRI)
- Headed by Douglas Engelbart
- Vision of the Interactive Personal Computer
- Mouse
- Graphical Interfaces
- Hypertext
- Visioconference
- Internet





Engelbart's Augmentation Research Center

Mouse [Engelbart, 1967]



Arpanet [Kleinrock & Engelbart, 1969]

Douglas Engelbart Team Demonstration (9 December 1969)

monday afternoon
december 9
3:45 p.m. / arena

Chairman:
DR. D. C. ENGELBART
Stanford Research Institute
Menlo Park, California

a research center for augmenting human intellect

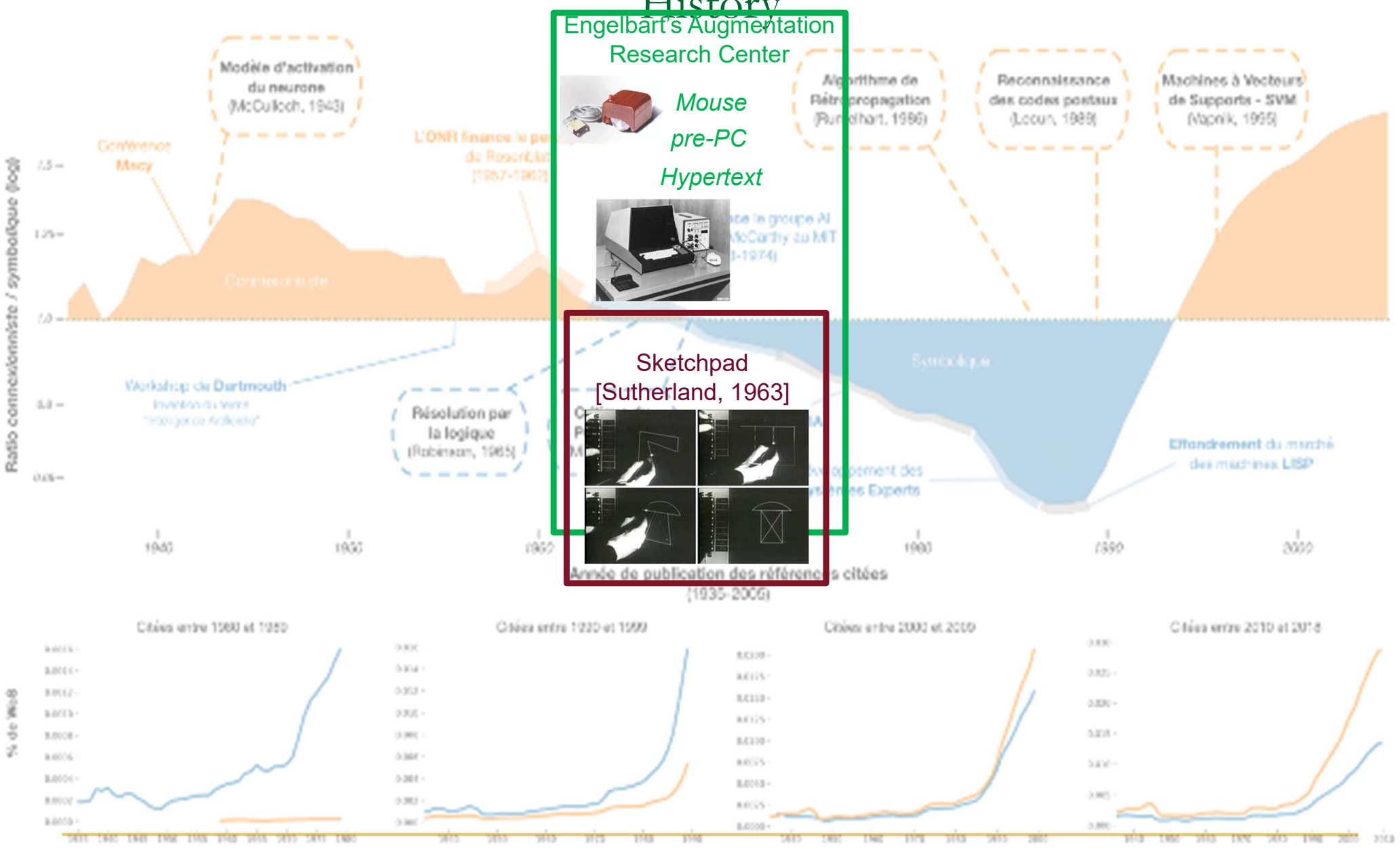
This session is entirely devoted to a presentation by Dr. Engelbart on a computer-based, interactive, multiconsole display system which is being developed at Stanford Research Institute under the sponsorship of ARPA, NASA and RADC. The system is being used as an experimental laboratory for investigating principles by which interactive computer aids can augment intellectual capability. The techniques which are being described will, themselves, be used to augment the presentation.

The session will use an on-line, closed circuit television hook-up to the SRI computing system in Menlo Park.

Following the presentation remote terminals to the system, in operation, may be viewed during the remainder of the conference in a special room set aside for that purpose.

https://web.stanford.edu/dept/SUL/library/extra4/sloan/MouseSite/1968Demo.html

AI (Artificial Intelligence) vs IA (Intelligence Augmentation) – History

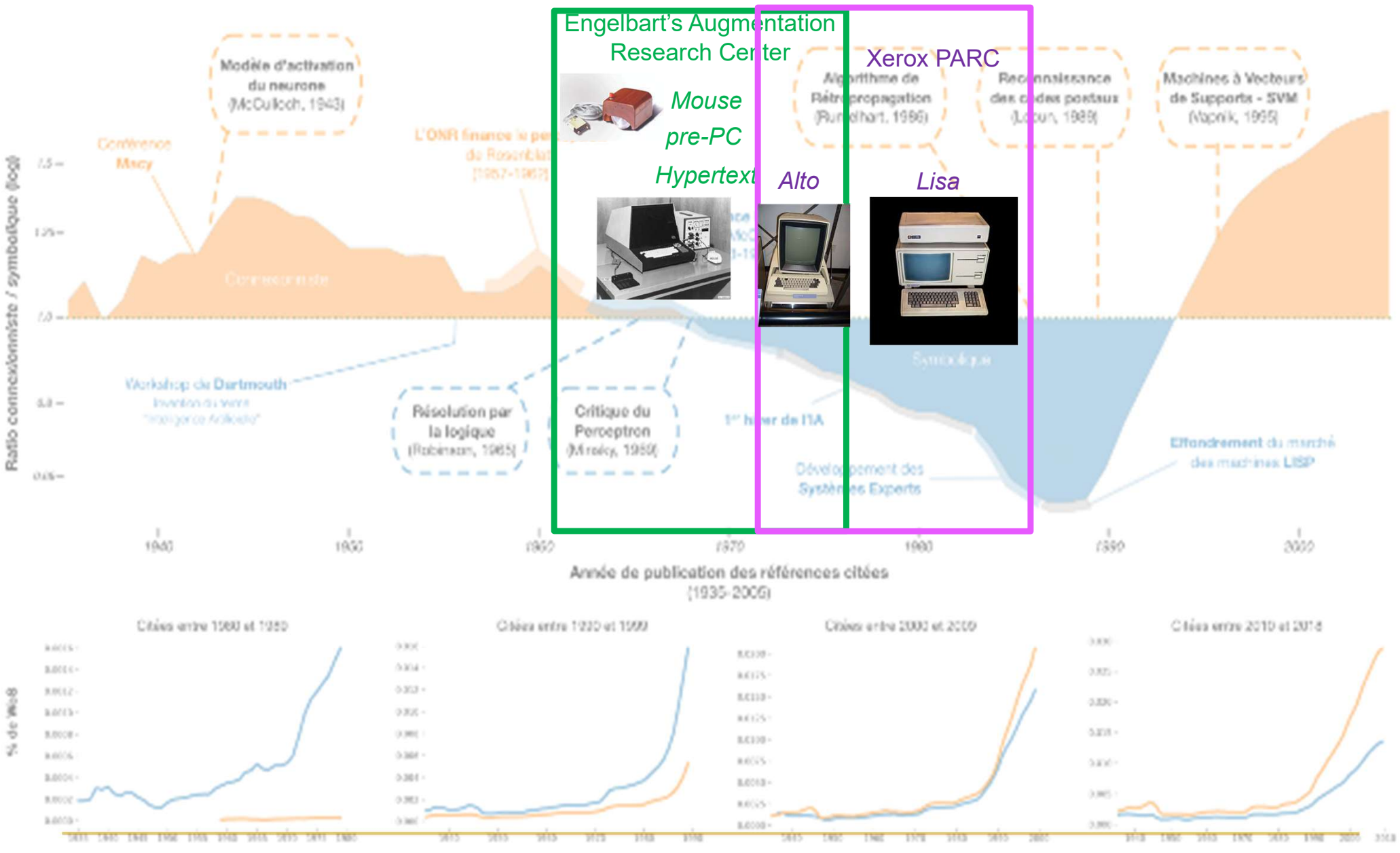


[Cardon et al., 2018]

Sketchpad [Sutherland, 1963]

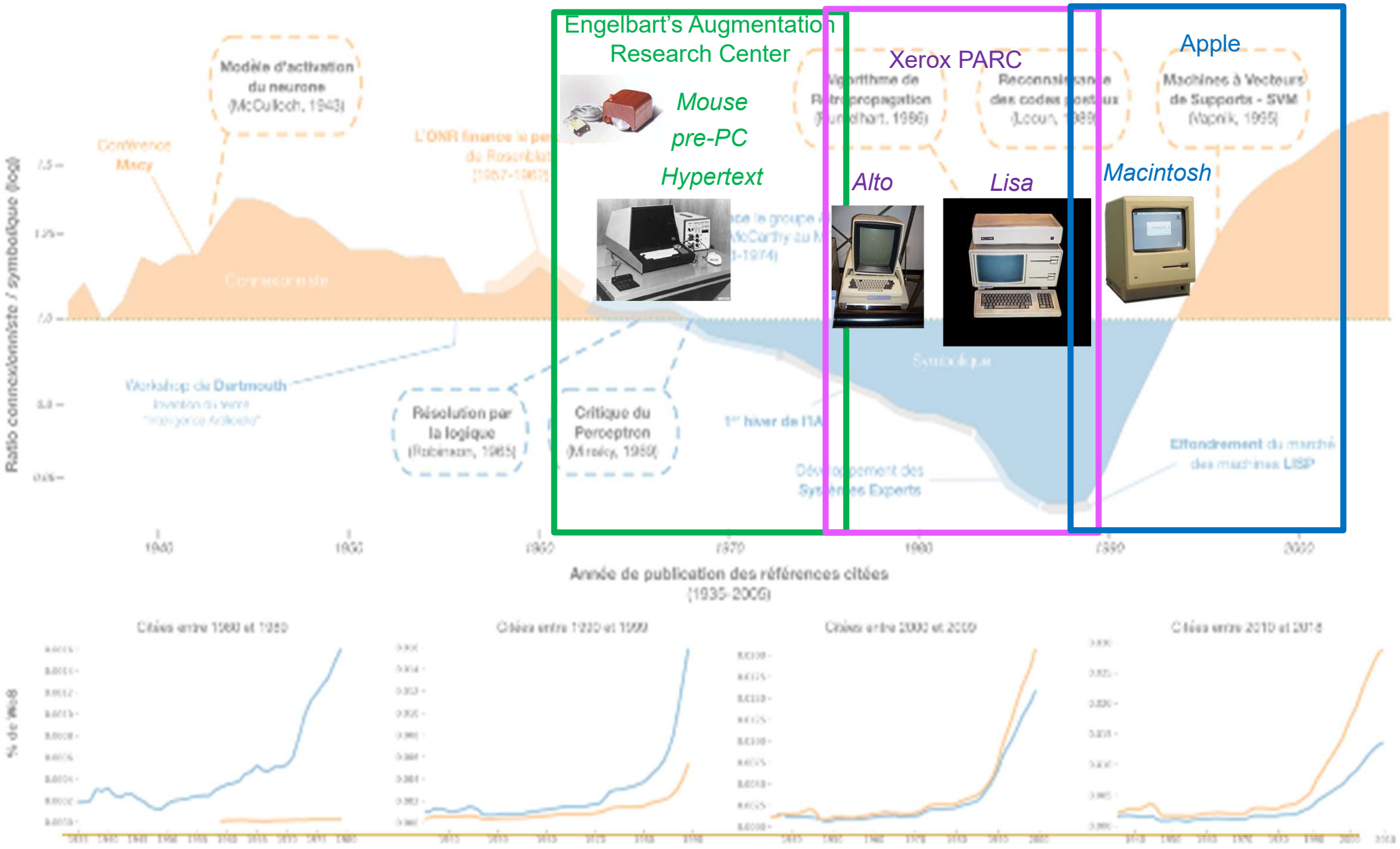
[Cardon et al. 2018]

AI vs IA – History



[Cardon et al., 2018]

AI vs IA – History



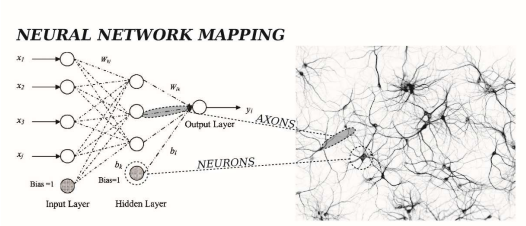
[Cardon et al., 2018]

AI and Connectionism with Progressive Incorporation of IA (Intelligence Augmentation & Interaction)

Symbolic AI [1956-]

Knowledge Representation
Reasoning
Planning

Connectionism [1957-]



Data Induction
Machine Learning
Pattern Recognition

Distributed AI [1975-]

Multi-Agent Systems
Coordination
Theory of Mind

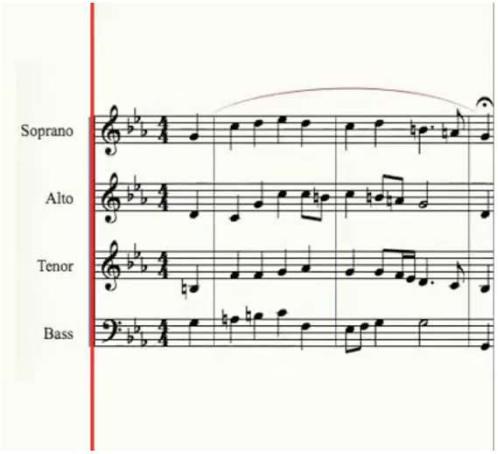
Active Learning [2009-]

Interactive User Query

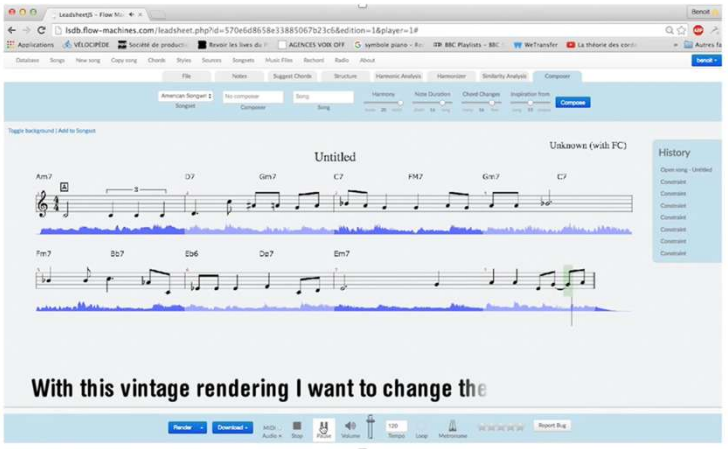
Personal Artificial Assistant [1990-]
Software Agent

Learning over User Shoulder
Chatbot

Autonomous vs Assistance in Music Making

- Autonomous Generation/Interpretation
  - Turing Test
  - Ambient Music
  - 

- Assistance to Human Composers and Musicians
  - Propose
  - Refine
  - Analyze
  - Harmonize
  - Produce
  - …



Referências

1. Capítulo 1 do livro *“Deep Learning Techniques for Music Generation”* de Jean-Pierre Briot, Gaëtan Hadjeres e François-David Pachet.