

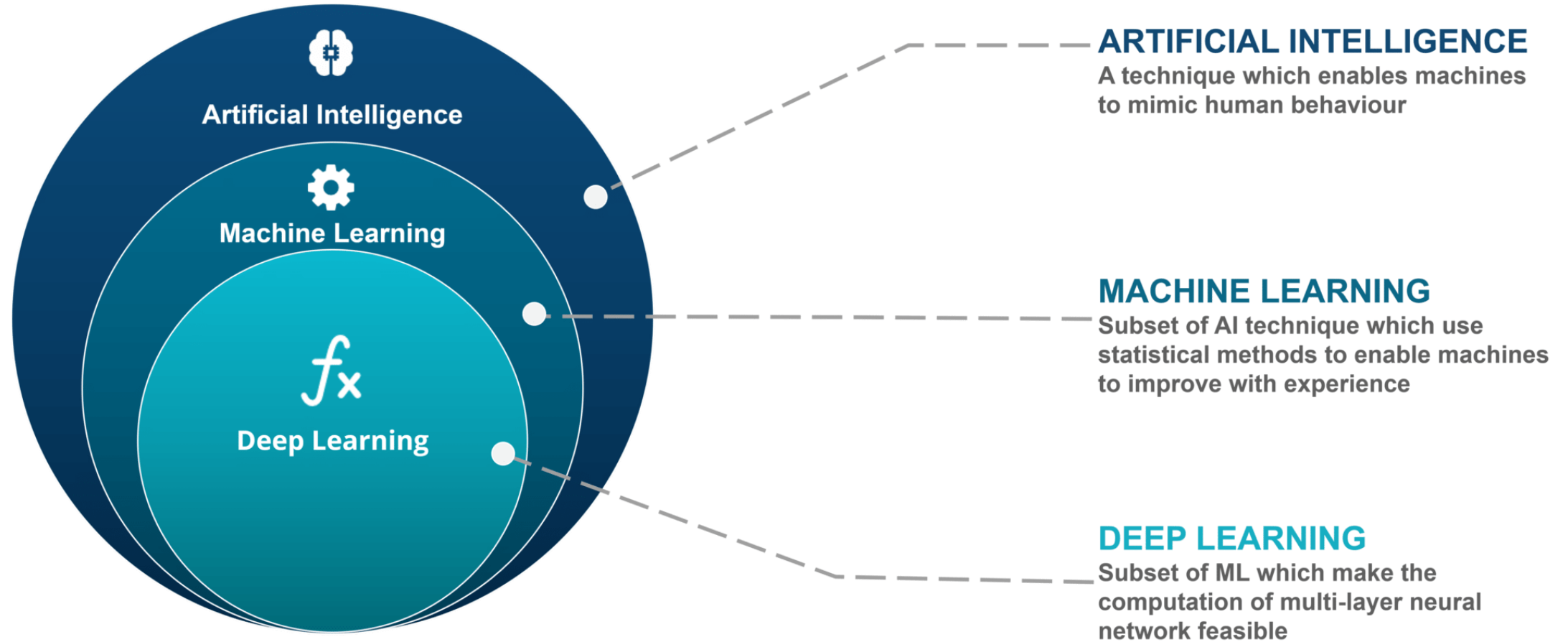
Advanced Artificial Intelligence

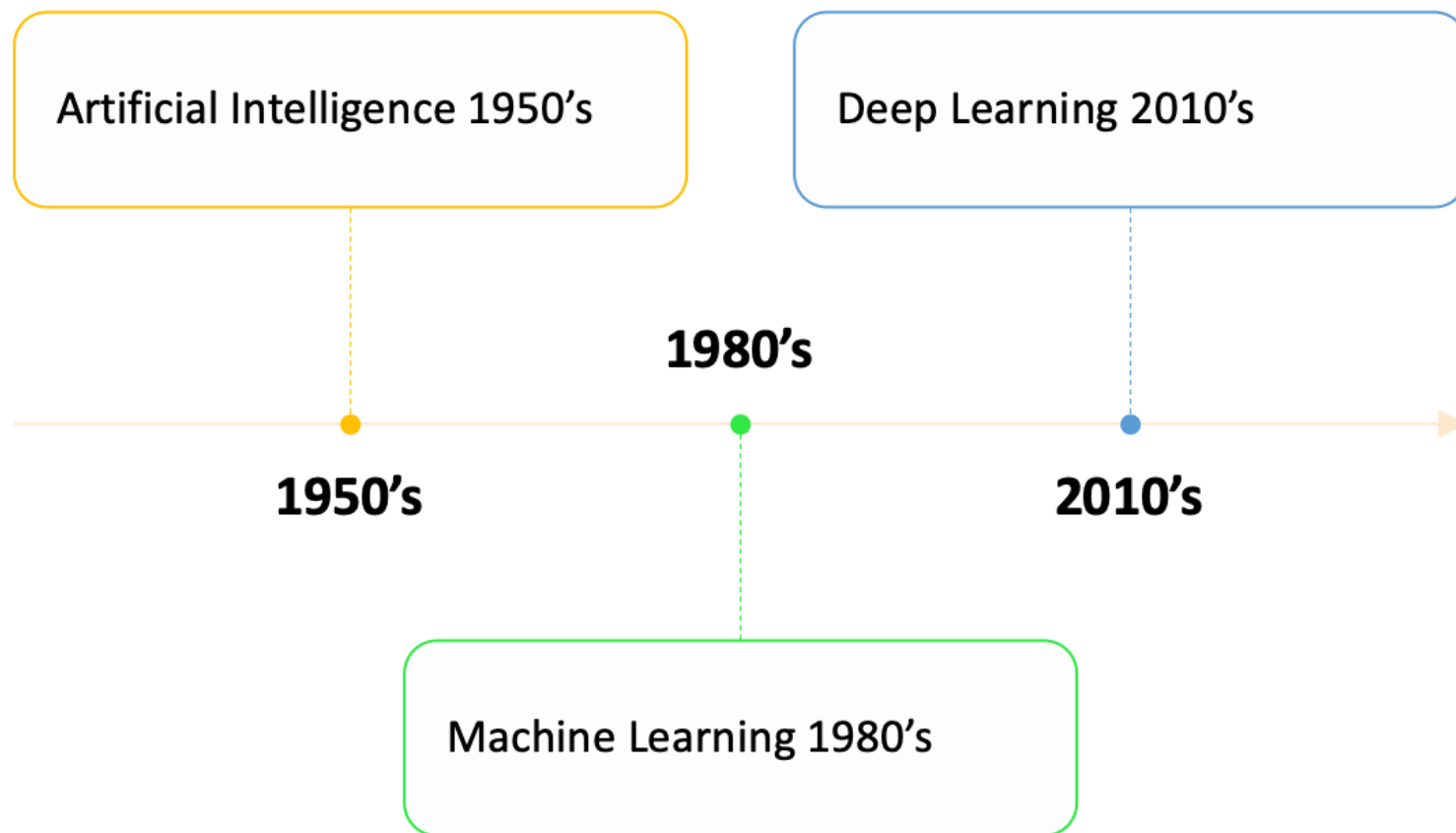


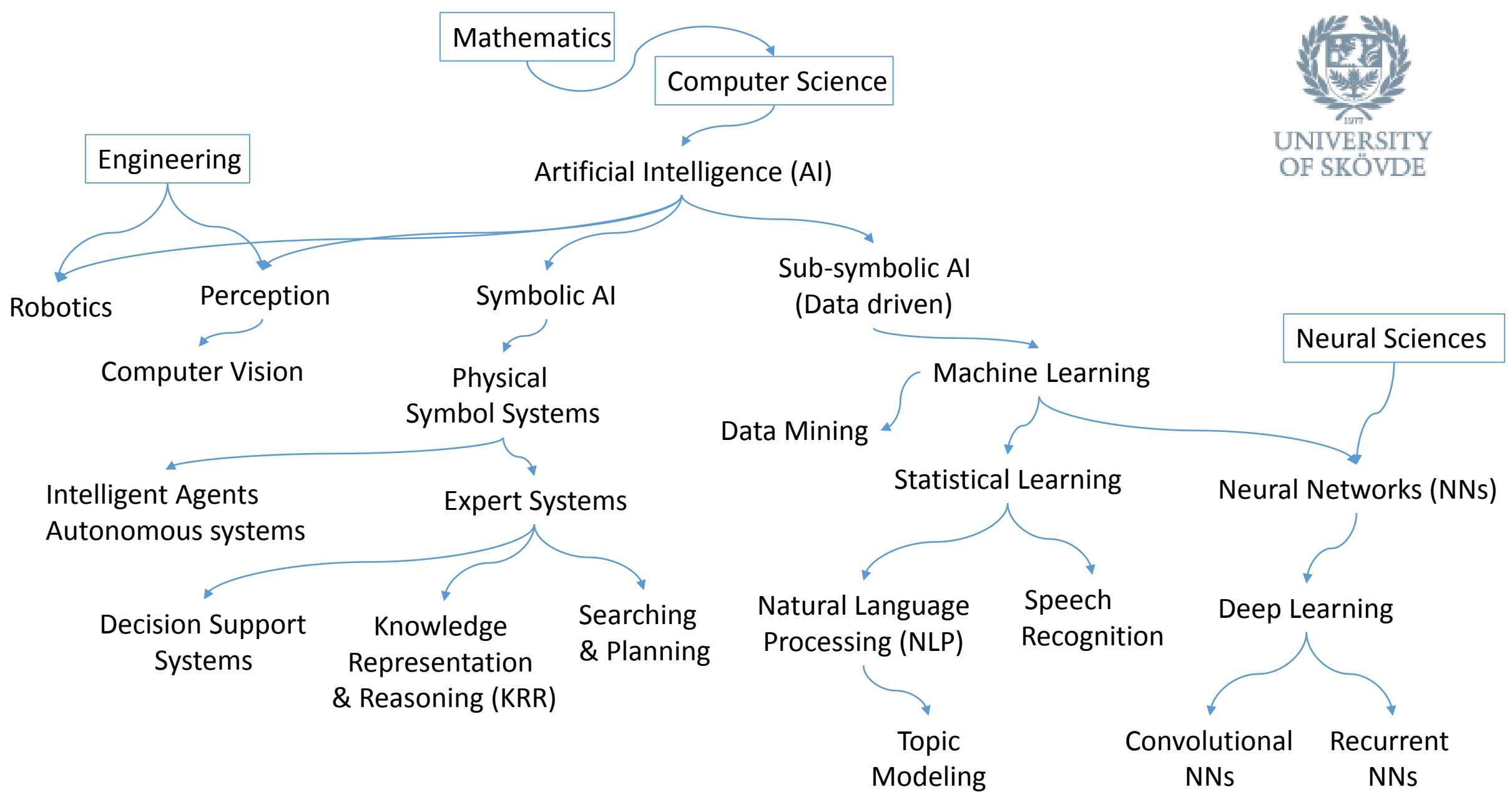
What is AI ?

A tiny bit of history and where we are now

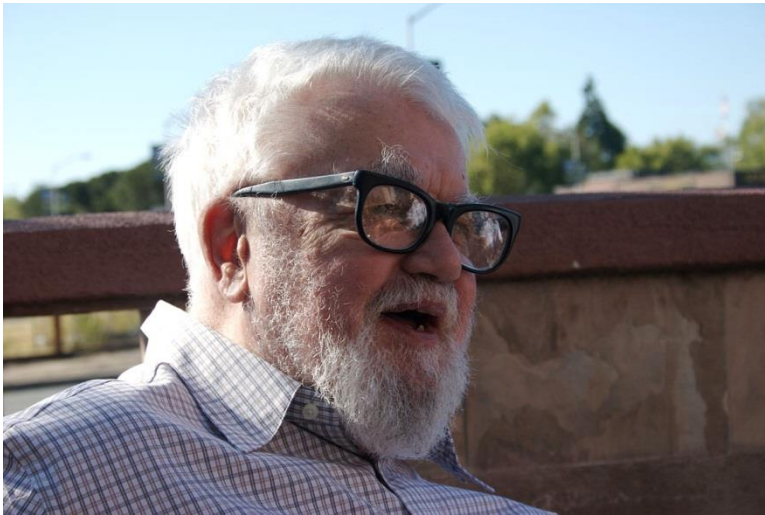
Artificial Intelligence







First definition of Artificial Intelligence by John McCarthy, 1955 at the Dartmouth conference:



4 september 1927 -- 24 oktober 2011

**A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON
ARTIFICIAL INTELLIGENCE**

J. McCarthy, Dartmouth College

M. L. Minsky, Harvard University

N. Rochester, I.B.M. Corporation

C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

“We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”

The 7 aspects of AI addressed by McCarthy et al. (1955)

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

1. Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

“... that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it...”

In other words:

It can be programmed
step by step.

For example playing chess:



Challenges:

- Rules
- Applicable moves
- Reasoning
- Reasoning under uncertainty
- Opponent
- Win/lose/draw
- Plan ahead
- Search for possible moves
 - Own moves/opponent moves
- Decision Making

The 7 aspects of AI addressed by McCarthy et al. (1955)

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

2. How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out.

Formal languages /Computer Languages

Pseudo Code

informal high level
(natural language)
description of what
to do

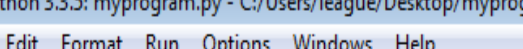
[illegible]

Machine Language / Machine Code

00101110

OPCODE Memory Address

operational code

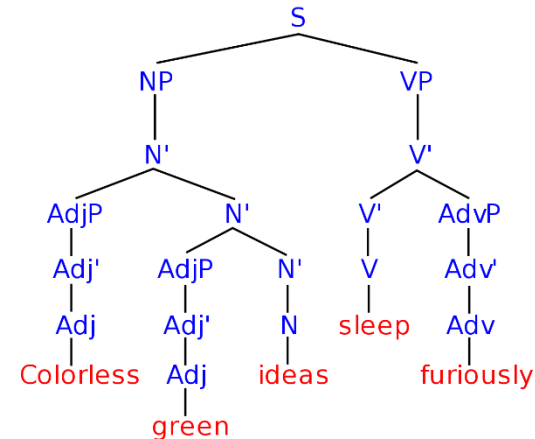
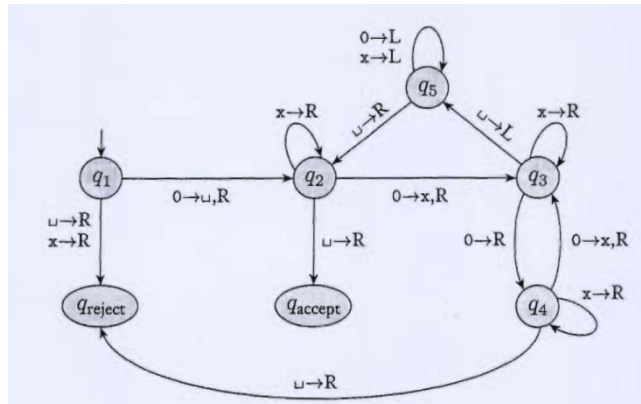


Python 3.3.5: myprogram.py - C:/Users/league/Desktop/myprogram.py

File Edit Format Run Options Windows Help

```
print("Hello!")
print(3+4)
print("3+4")
print("2*7 is" 2*7)
```

Turing machine



Logic, Reasoning

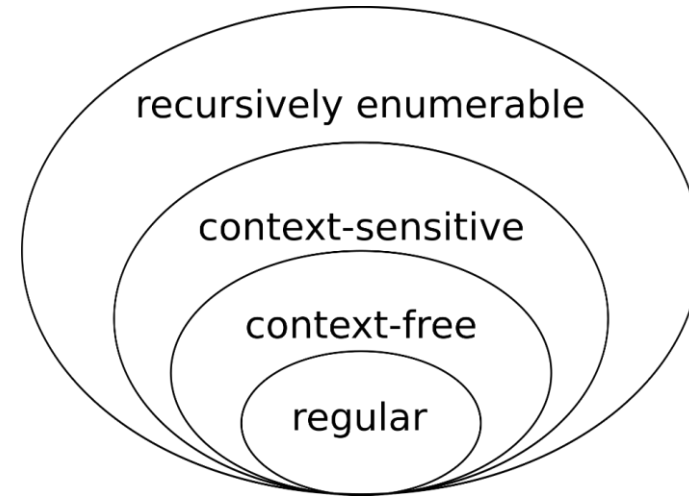
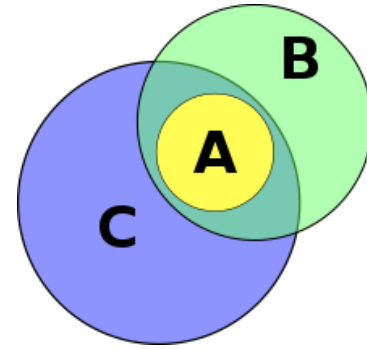
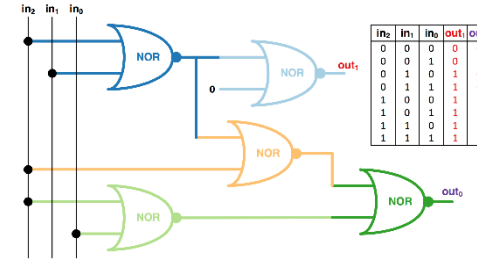
Many different types of logic:

- Propositional logic
- First order logic
- Second order logic
- Default logic
- Non-monotonic logic

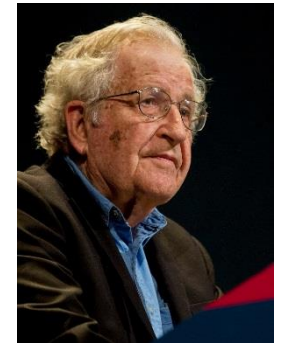
Different types of reasoning:

- Deductive reasoning (logical conclusions)
- Inductive reasoning (generalization based on statistics)
- Abductive reasoning (find best possible explanation)

$$\begin{aligned}\neg A \vee B &\iff \neg A \vee (\neg\neg B) \\ &\iff \neg(\neg B) \vee \neg A \\ &\iff B \vee \neg A \\ &\iff \neg B \rightarrow \neg A\end{aligned}$$



Chomsky hierarchy



Noam Chomsky

Deduction

From premises to logical and certain conclusions

Logic: propositional logic, first order logic, second order logic

Modus Ponens:

$P \rightarrow Q$	(conditional statement, P implies Q)
\underline{P}	(hypothesis stated, P is given)
Q	(conclusion deduced, then Q must be the case)

Modus Tollens:

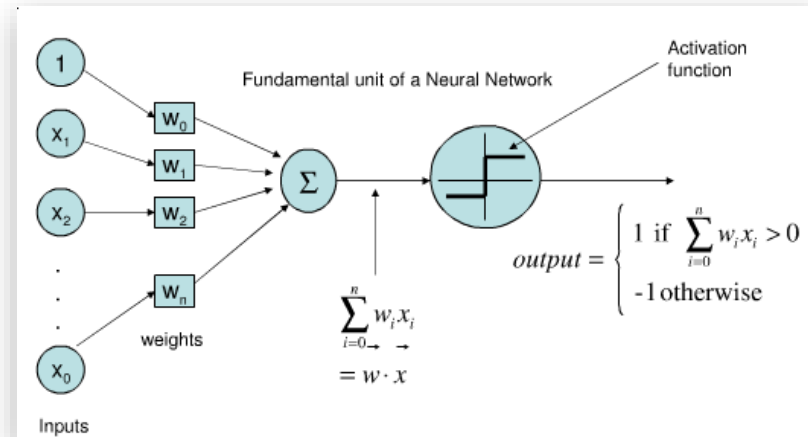
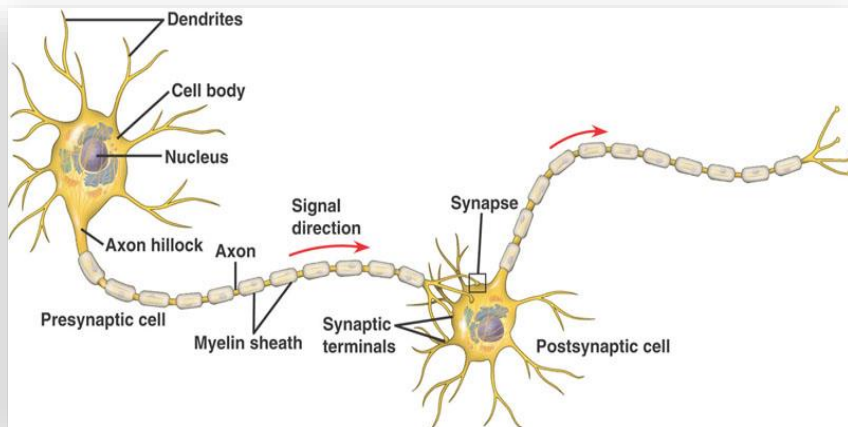
$P \rightarrow Q$	(conditional statement, P implies Q)
$\underline{\sim Q}$	(hypothesis stated, Q is <u>not</u> given)
$\sim P$	(conclusion deduced, P cannot be the case)

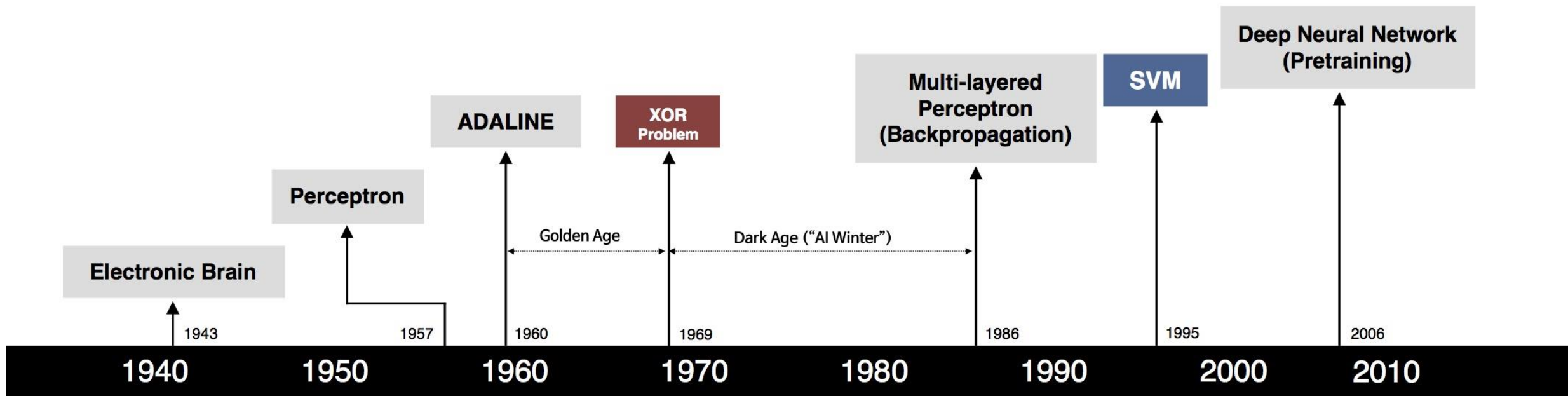
The 7 aspects of AI addressed by McCarthy et al. (1955)

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

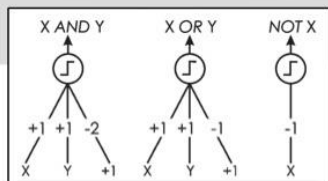
3. Neuron Nets

How can a set of (hypothetical) neurons be arranged so as to form concepts. Considerable theoretical and experimental work has been done on this problem by Uttley, Rashevsky and his group, Farley and Clark, Pitts and McCulloch, Minsky, Rochester and Holland, and others. Partial results have been obtained but the problem needs more theoretical work.





S. McCulloch – W. Pitts



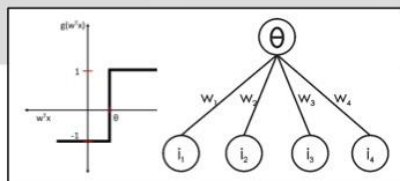
- Adjustable Weights
- Weights are not Learned



F. Rosenblatt



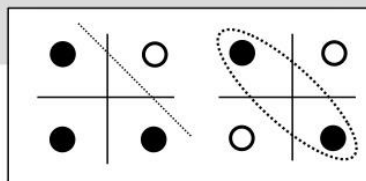
B. Widrow – M. Hoff



- Learnable Weights and Threshold



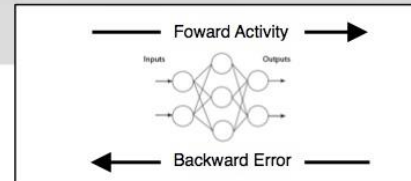
M. Minsky – S. Papert



- XOR Problem



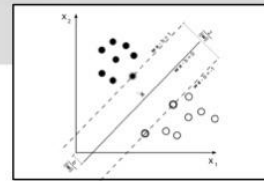
D. Rumelhart – G. Hinton – R. Williams



- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting



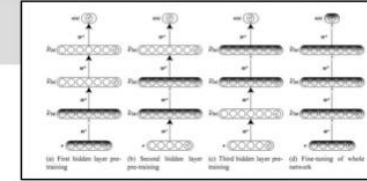
V. Vapnik – C. Cortes



- Limitations of learning prior knowledge
- Kernel function: Human Intervention



G. Hinton – S. Ruslan



- Hierarchical feature Learning

The 7 aspects of AI addressed by McCarthy et al. (1955)

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

4. Theory of the Size of a Calculation

If we are given a well-defined problem (one for which it is possible to test mechanically whether or not a proposed answer is a valid answer) one way of solving it is to try all possible answers in order. This method is inefficient, and to exclude it one must have some criterion for efficiency of calculation. Some consideration will show that to get a measure of the efficiency of a calculation it is necessary to have on hand a method of measuring the complexity of calculating devices which in turn can be done if one has a theory of the complexity of functions. Some partial results on this problem have been obtained by Shannon, and also by McCarthy.

Computational complexity, e.g. Big O notation

measures how well a computer algorithm scales if the amount of input data increases (worst case scenario), n = number of input

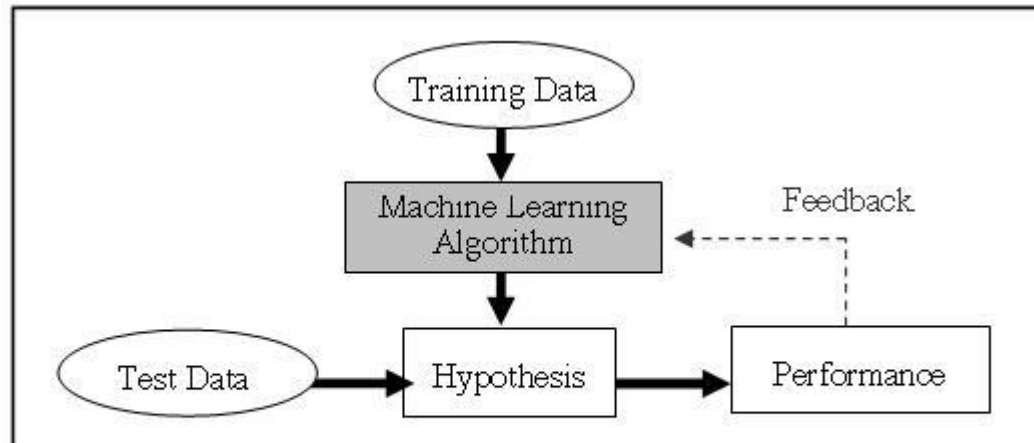
$O(1)$	static	e.g. add item to a list
$O(\log(n))$	logarithmic	e.g. binary search (sorted list)
$O(n)$	linear	e.g. linear search
$O(n \log(n))$	log linear	e.g. quick sort
$O(n^2)$, $O(n^3)$, $O(n^x)$	quadratic, cubic, ...	e.g. bubble sort $O(n^2)$
$O(2^n)$	exponential	
$O(n!)$	factorial	e.g. tic, tac, toe (brute force)
$O(\text{infinity})$	infinite	e.g. tossing a coin

The 7 aspects of AI addressed by McCarthy et al. (1955)

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

5. Self-Improvement

Probably a truly intelligent machine will carry out activities which may best be described as self-improvement. Some schemes for doing this have been proposed and are worth further study. It seems likely that this question can be studied abstractly as well.



AlphaGo

- <https://www.youtube.com/watch?v=TnUYcTuZJpM>
- Deep reinforcement learning
- AGI (Artificial general intelligence)



The 7 aspects of AI addressed by McCarthy et al. (1955)

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

6. **Abstractions**

A number of types of "abstraction" can be distinctly defined and several others less distinctly. A direct attempt to classify these and to describe machine methods of forming abstractions from sensory and other data would seem worthwhile.

Abstraction

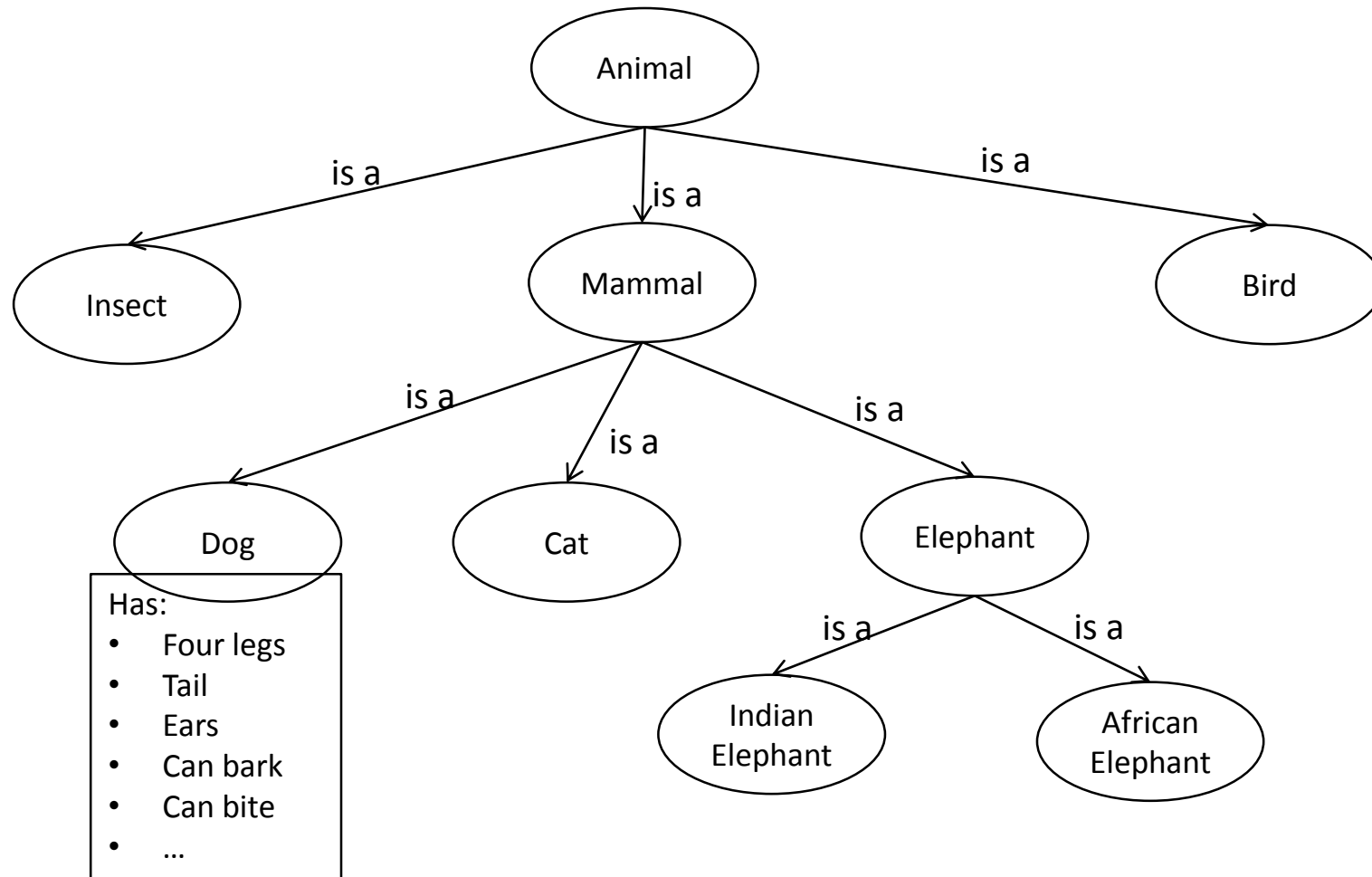
Abstraction is the process of reducing the characteristics of an object/entity/etc. to a set of essential characteristics.

Abstraction is one of the key concepts in object-oriented programming, (the others are encapsulation and inheritance). The objects complexity is reduced by hiding all but the relevant characteristics of the object.

Abstraction

- Humans make abstractions/generalizations/categorizations in daily life
- Cognitive Psychology
 - How is knowledge organized in the human brain?
 - How do people reason?
- What are the relevant/essential characteristics given a task or problem?
- How can they be represented?

Semantic network / object oriented languages



The 7 aspects of AI addressed by McCarthy et al. (1955)

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

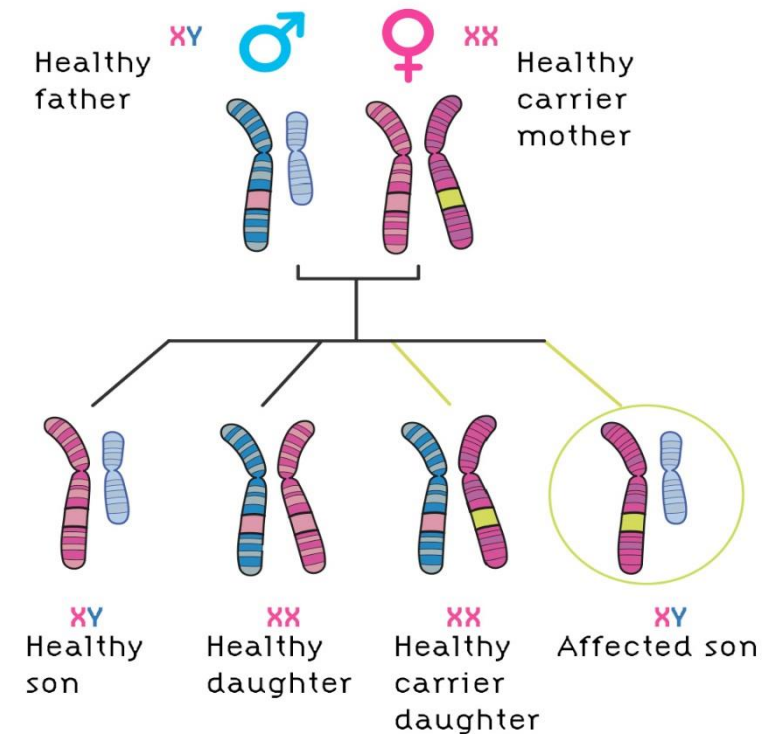
7. Randomness and Creativity

A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.

Inspired by genetics /evolution

- Based on Darwin's principle of natural selection
- Uses three different techniques
 - Selection
 - Mutation
 - Crossover

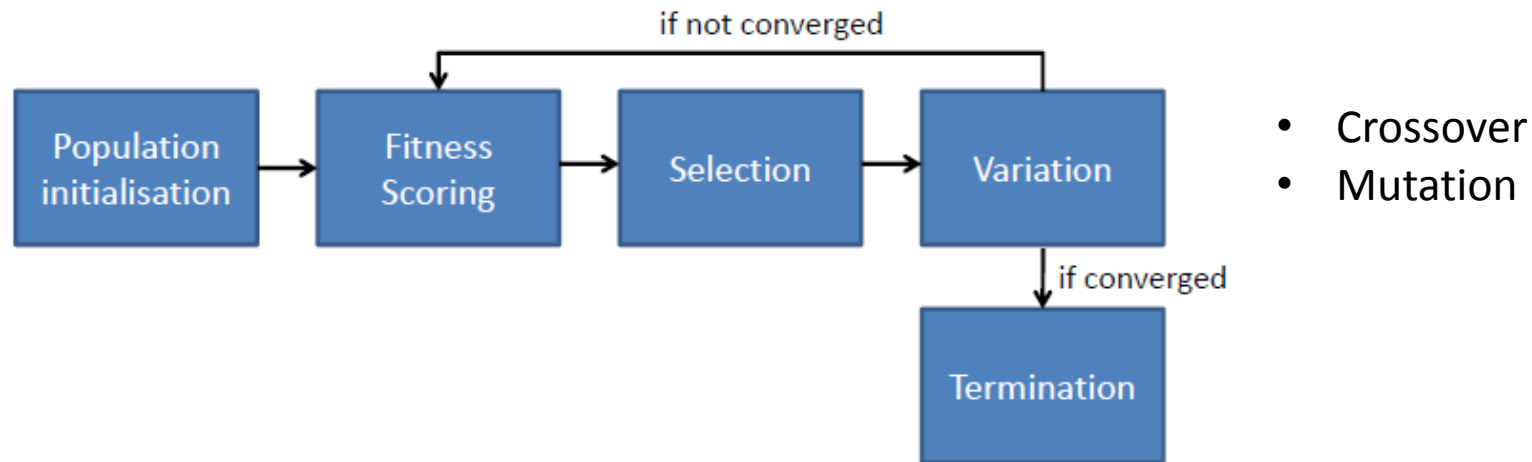
3
X-linked
recessive



Ex: myopathie de Duchenne, hémophilie A

Evolutionary algorithms

- Optimization process of an Evolutionary algorithm



Convergence criteria

- Fixed number of generations
- Fitness function value change $< \epsilon$

Creative AI

AI composing music



- Artificial Intelligence Virtual Artist (AIVA)

Ted talk by Pierre Barreau,
one of the founders of Aiva

<https://business.blogthinkbig.com/how-ai-is-revolutionising-the-classical-music-industry-an-analysis-of-the-musical-ai-by-aiva-technologies/>



10 minutes break!

AI Definition by Marvin Minsky

“Artificial intelligence is the science of making machines do things that would require intelligence if done by men.”



Marvin Minsky, 1927 -- 2016

Intelligence vs. Artificial Intelligence

- Intelligence (Oxford dictionary)
 - “The ability to acquire and apply knowledge and skills.”
- Artificial intelligence (Oxford dictionary)
 - “The theory and development of computer systems able to perform tasks normally requiring human intelligence, **such as** visual perception, speech recognition, decision-making, and translation between languages.”

Natural language communication

Turing Test

- Measuring artificial intelligence
- 1950 proposed by Alan Turing
- <https://www.youtube.com/watch?v=3wLqsRLvV-c>

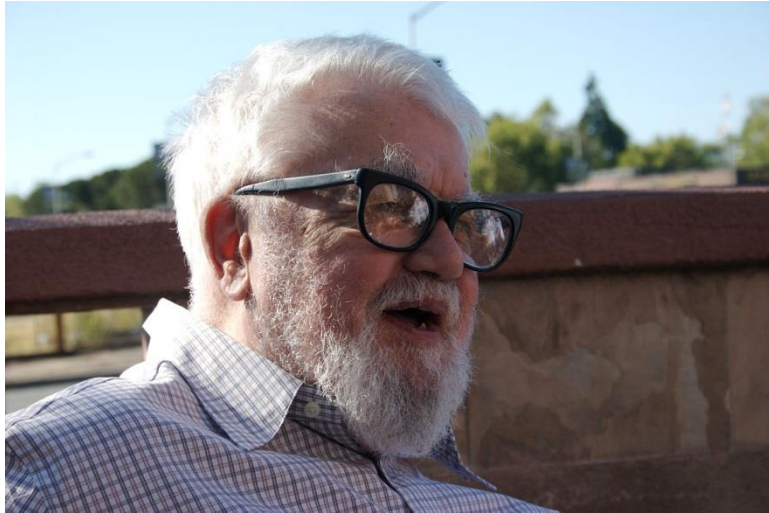
Keynote speaker introduces Siri (2011):

<https://www.youtube.com/watch?v=qlhKlen5gvU>



Alan Turing

First definition of Artificial Intelligence by John McCarthy, 1955 at the Dartmouth conference:



4 september 1927 -- 24 oktober 2011

“We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”

**A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON
ARTIFICIAL INTELLIGENCE**

J. McCarthy, Dartmouth College

M. L. Minsky, Harvard University

N. Rochester, I.B.M. Corporation

C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

1997 IBM's Deep Blue beats human chess master

<https://www.youtube.com/watch?v=NJarxpYyoFI>



Garry Kasparov, chess master

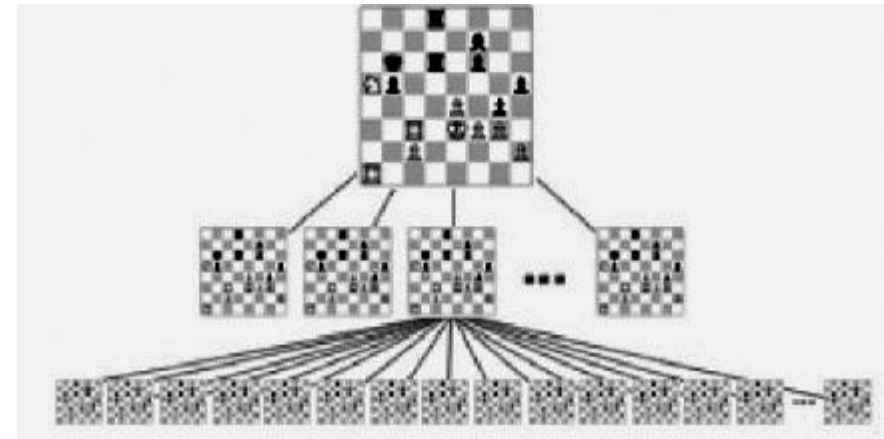


Deep Blue, at the Computer History Museum

Brute Force

Deep Blue

- Computational power
- 256 Processors
- 200 million positions analyzed per second
(Gary Kasparov analysis 3 positions per second)
- Search
searches 6 moves ahead
(can search up to 20 moves ahead)
- Library of good positions
- Evaluation



IBM's Watson (2011)

(cognitive Computing)



[https://www.youtube.com/watch?v= Xcmh1LQB9I](https://www.youtube.com/watch?v=Xcmh1LQB9I)

Cognitive Computing

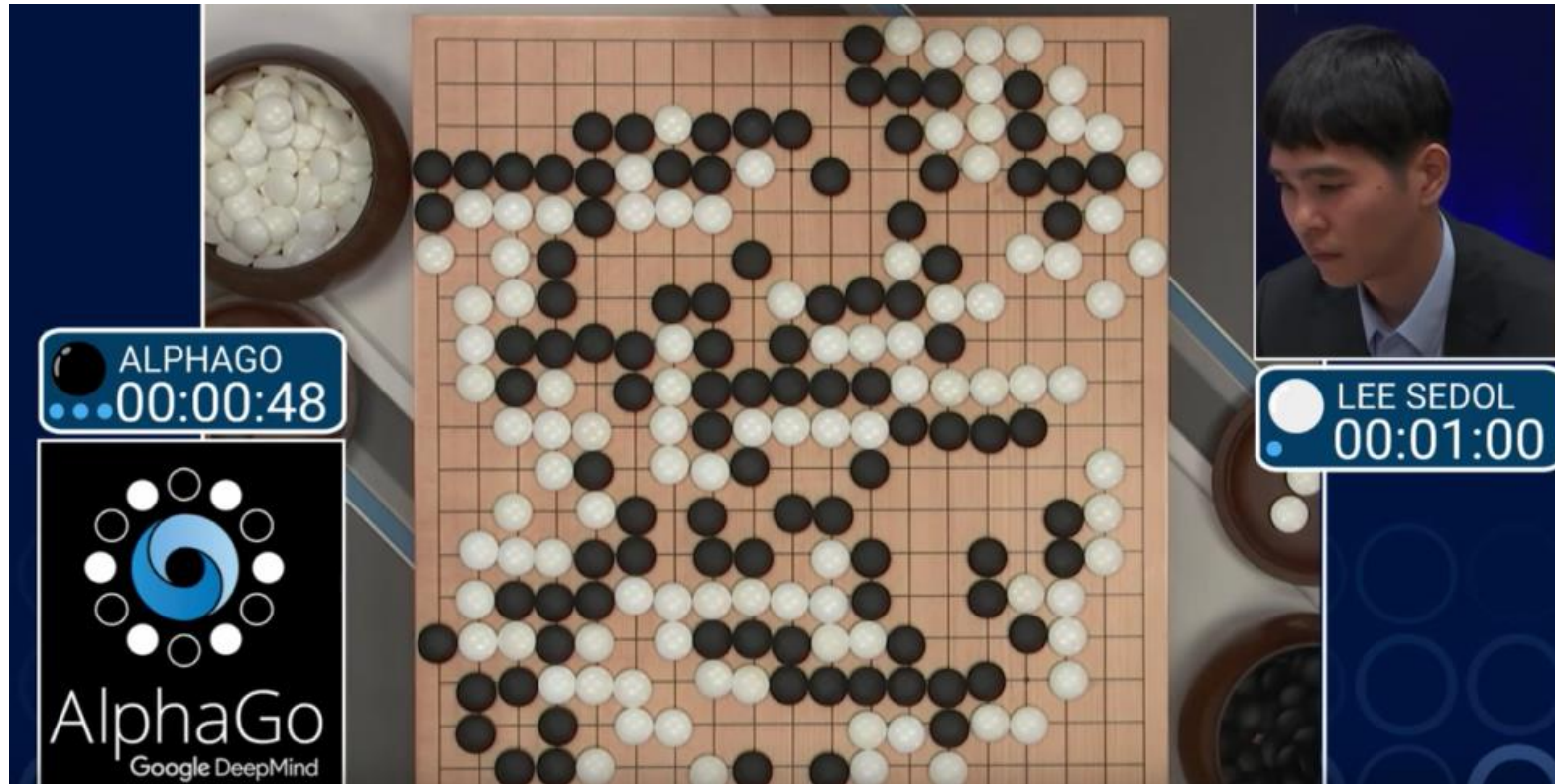
Human

1. Observe (collect evidence)
2. Generate hypothesizes (from the observed, and a body of knowledge)
3. Evaluate hypothesis (eliminate wrong ones)
4. Decide what to do (and act accordingly)

Watson

- "Understands" natural language, can "read" unstructured text
 - Breaks text down according to
 - Grammar, structure,. Culture, etc.
 - Learns the language of the field
 - Terminology, jargon, ambiguities, etc.
- Human guidance by collecting a body of knowledge (what texts to read)
- Creates a knowledge graph (how everything in the field is related)
- Trained by human expert through Question-Answer pairs using Machine Learning
- Updated by human experts

2016 AlphaGo Beats Human Go Master

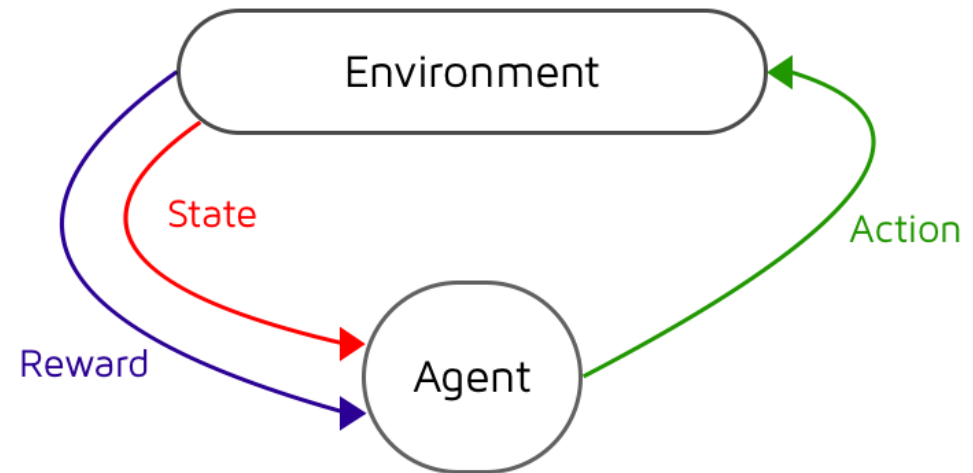


<https://www.youtube.com/watch?v=TnUYcTuZJpM>

Deep Reinforcement Learning

AlphaGo

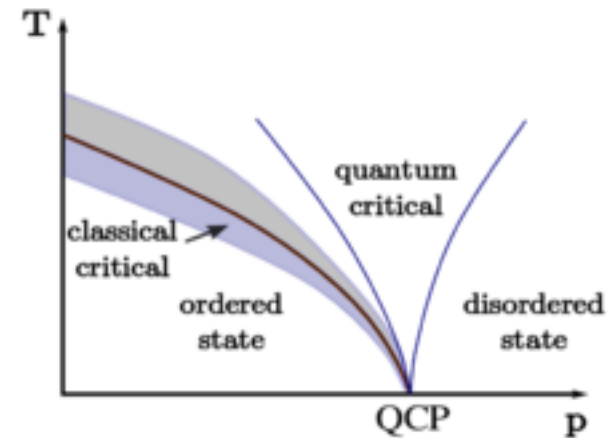
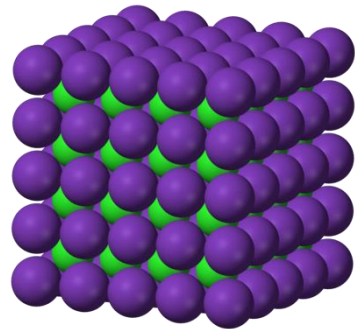
- Can learn on its own
- Raw pixels as input
 - 1) Mimics human player
 - 2) plays itself and improves itself
- Reinforcement learning, ANN
- General Purpose Artificial Intelligence (AGI)



A.I. Learns Nobel Prize Experiment

<https://www.youtube.com/watch?v=IjGzmsLRUo>

Bose–Einstein condensate (BEC)



AlpHaZero

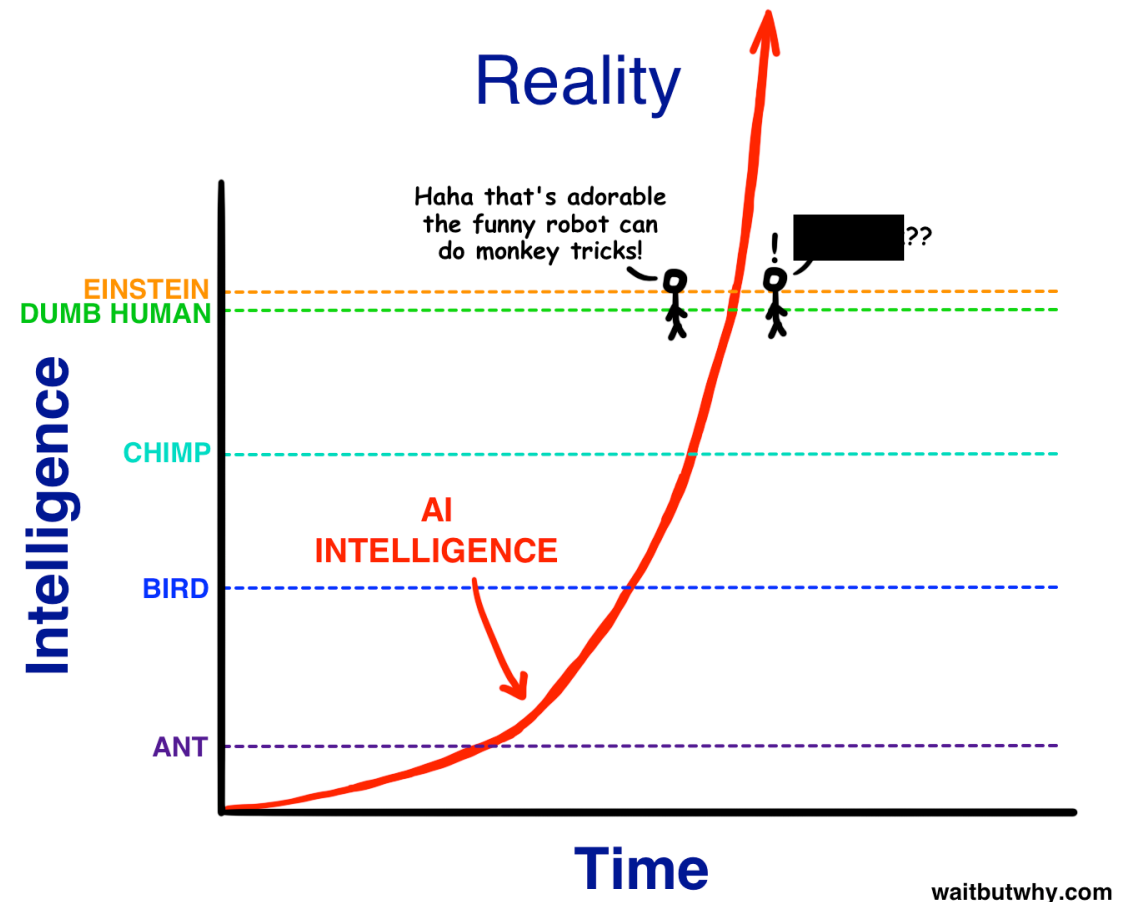


- Reinforcement learning
- Deep neural networks

In 2017 AlphaZero from Google Deep Mind
beats the best playing Chess program Stockfish after just 4 hours of training

Should we be worried ?

- AI is a human creation, however it cannot always be predicted what it does, and also not always how it works. (Black box)
- Method for self-improvement
- In the 1950s AI researchers thought that computers would play chess within a decade → It took until 1997 !
- In 2016 AlphaGo beats human Go master, this occurred at least a decade earlier than expected
- Some people say that humans have an inability to perceive exponential change



ethics

moral

philosophy

concepts

applied

issues

study

immoral

many

social

term

understand

single

based

right

morality

question

seperation

contemporary

behaviour

human

rights

terms

involves

place

philosophical

independent

may

mean

dimension

time

political

example

minds

lessons

benefit

approach

wrong

ethical

Unemployment

- We need money for food, rent, medical care
 - We work to earn money
- What if AI takes over our jobs?
(post work society)
 - How will we afford to live?
 - How will wealth be distributed between people?
 - Will owner of AI earn all the money?
 - What will we do with our time?



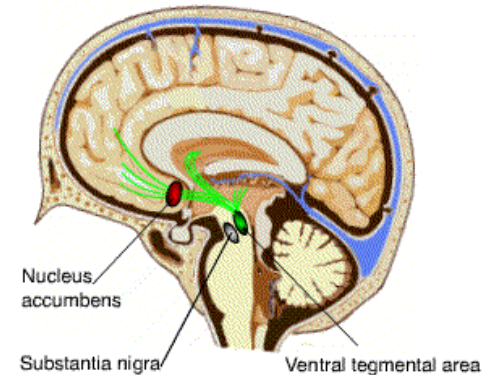
Volker Hirsch Ted Talk: About job loss when AI takes over

<https://www.youtube.com/watch?v=dRw4d2Si8LA>

Technology addiction/dependency

- AI is used to trigger our behavior
 - E.g. click-bait headlines
 - This can be used in a bad way or in a good way
 - To manipulate for someone's/something's own advantage
 - To influence people in a good way, e.g. to achieve a better and healthier lifestyle
- Are we becoming addicted to interaction with AI
 - Games are highly addictive
 - Interaction with technology can be highly addictive
 - Social media, email
 - How long can you be without your cellphone and feel perfectly comfortable?
- Do we always know that we are interacting with AI?
 - How can we find out?
 - Are AI's better company than humans?

Pleasure Reward Pathway



Human safety

- How can we be sure that the AI/machine does not make mistakes?
 - We can not train it for all eventualities.
 - If it is self learning, what will it learn?
 - Can it learn that humans are the reason for all problems?
 - Will the technology singularity become true?
 - Machines improving themselves further and further into super AI that is far more intelligent than any human.
 - Can AI become discriminating towards certain groups of people?
 - It learns from training data/statistics
 - Human prejudices are also often built on statistics experience

Cyber security



- Cyber security becomes essential
 - Very powerful AI systems should not fall into the wrong hands
 - E.g. weapons
 - “Wars” are fought in a different way
 - Not anymore on the battlefield

Ted Talk Jay Tuck: Dark view on AI:

<https://www.youtube.com/watch?v=BrNs0M77Pd4#t=31.9279468>

Robot rights

- AI researchers work on creating AIs that are aware of themselves
 - The AI will be conscious
 - Is that a ground for treating it the same as a human?
 - Will AI in the future be sentient?
 - Would that be ground to treating it as a human?
 - When AI is smarter and more capable as humans and animals what protective rights do they need?

How to deal with it all?

“The challenge now is to make sure everyone benefits from this technology “ – Peter Norvig, director of research at Google

“Without an industry standard for testing AI, it is difficult for these technologies to be widely implemented” – Missy Cummings, director of the Human and Autonomy Lab, Duke University

“There is increasing desire to see these technologies be fair and ethical but these concepts are fuzzy at best “ – Danah Boyd, Microsoft Research

Stuart Russel talks about AI and ethics:

https://www.ted.com/talks/stuart_russell_how_ai_might_make_us_better_people#t-368944

Critics



- Elon Musk (SpaceX, Tesla): “ With artificial intelligence we are summoning the demon.”
 - Founded “Open AI” together with Sam Altman
 - A non-profit AI research company aiming to promote and develop friendly AI



- Bill Gates (co-founder of Microsoft): “I think we do need to worry about Artificial intelligence.”



- Stephen Hawking: “*AI could spell end of the human race*”

Friendly AI



Eliezer Yudkowsky



Stuart Russel

- Term “friendly AI” coined by Eliezer Yudkowsky in 2008
- Friendliness
 - AI will never harm people (It’s most important rule is: not to harm people)
 - AI knows that its own design might be flawed (So it lets a human override its decision, or lets the human switch it of)
 - AI only does useful and good outcomes (measured with human standards of good and useful)
 - https://www.ted.com/talks/stuart_russell_how_ai_might_make_us_better_people#t-368944

Seminar and report on ethical and philosophical questions related to Artificial Intelligence

Assignment 1 / Ethical assignment

Evaluation (assignment 1 / ethical assignment)

- The assignment will be in groups of 3-4 people
- A 10 minutes presentation will be given by each group (the exact date will be announced in the lectures and on the CANVAS page).
- Each group will prepare and hand in a short report (3 pages) to be delivered by email to my urkund-address: joe.steinhauer.his@analys.urkund.se with CC to my usual address: joe.steinhauer@his.se .
- The deadline for the report will be announce in the course and on the CANVAS page.
- The grade for the presentation and report is pass/fail.
- Every group member has had to be contributed to the presentation and the report. If I'm unsure of your performance in the group work, I will call you in for an extra discussion about the subject.
- You will need to be able to explain and discuss different ethical problems in the written exam, so take this assignment as a chance to prepare for that.
- During the presentation, all of the group members need to be present and each of you need to present a part of the presentation.
- You need to be aware of the university's rules on plagiarism and use a very good reference system in your report.

Before you start

- We will assign you to a group of 3-4 students and announce the group and its members on the course's CANVAS page.
- Your group needs to agree on an ethical or philosophical question/problem (your topic) that you propose to your teacher. If accepted you can start your work.
- If the proposed topic is not suitable or already taken by another group, you will have to find a new topic.

The report

- The document needs (at least) two parts.
- The first part should include a discussion or position on what artificial intelligence is (very briefly) and why ethical problem do arise
- The second part should discuss a concrete ethical (or philosophical) aspect on artificial intelligence.
- The report must contain all the names of students that work on the report. Only students that have actively contributed to the report can be named as authors of the report.
- A good research ethic has to be applied in the report. This means that all material that you have read or watched, in order to gather the information presented in the report is clearly referenced. You have gotten information about plagiarism in the lectures. If you are unsure about it, please ask the teacher for clarification before you hand in the report.

The seminar presentation

- The group presentation in the seminar has to focus only on the second part.
- You clearly state an ethical question/dilemma or philosophical question and discuss different views on it. It is important that you cover
- What is the question/dilemma/problem?
- Why is that a problem?
- What are the different points of view on it (pros and cons)
- How could it be solved?
- During the presentation, all of the group members need to be present and each of you need to present a part of the presentation.

Possible topics

This is a list of some possible topics, it is not exhaustive and you do not have to choose from the list. Your own ideas for topics are very welcome.

- AI and employment issues
 - Technology addition
 - Robot rights
 - The singularity
 - Brute force, is that AI?
 - Is strong AI possible?
 - ...
 - ...
- Discriminatory machine learning
 - AI and war (autonomous weapons)
 - How will AI change war strategies?
 - Machine credibility
 - Who is responsible for a machine's actions?
 - ...
 - ...

Study Support Center (Studieverkstan)

- In the university library
 - studieverkstan@his.se
 - You can just drop by or you can make an appointment by email
- Birgit Persson and her colleagues will help you with
 - questions regarding referencing
 - simple formulations when referring to other's text

Refero: a guide to avoid plagiarism

- <https://www.his.se/en/Library/Search-write-and-publish/Writing-references/>

Group work in practice

- Online group meetings e.g. on zoom
- For report writing you can use e.g. google docs.