Access Code: 188276

CMP9794M
Advanced Artificial Intelligence

Heriberto Cuayahuitl

UNIVERSITY OF LINCOLN

School of Computer Science

Delivery Team

- **Dr. Heriberto Cuayahuitl**
- **Prof. Simon Parsons**
- **Dr. Riccardo Polvara**
- **Miss Kim Bird**

Main Topics in this Module

- **Quantifying uncertainty**
  - Intro to probability theory
- **Probabilistic reasoning**
  - Bayesian Networks (BNs)
- **Reasoning over time**
  - Dynamic BNs, Hidden Markov Models (HMMs)
- **Making complex decisions**
  - Markov Decision Processes (MDPs)
  - Reinforcement learning (without neural nets)

Lectures and Workshops

- Lectures: Wednesday 4-6pm in INB0014
- Workshops: Wednesday 12hrs-4pm in INB1102

CMP9132, CMP9794	CMP9132, CMP9794	CMP9132, CMP9794	CMP9132, CMP9794	CMP9132, CMP9794
Group B	Group D	Group C	Group A	
INB1102	INB1102	INB1102	INB1102	INB0114
12pm - 1pm	1pm - 2pm	2pm - 3pm	3pm - 4pm	4pm - 6pm
Weeks 2-5, 7-11, 15-16	Weeks 2-5, 7-11, 15-16	Weeks 2-5, 7-11, 15-16	Weeks 2-5, 7-11, 15-16	Weeks 1-5, 7-11, 15-16
Heriberto Cuayahuitl	Heriberto Cuayahuitl	Heriberto Cuayahuitl	Heriberto Cuayahuitl	Heriberto Cuayahuitl
Portilla, Simon	Portilla, Simon	Portilla, Simon	Portilla, Simon	Portilla, Simon
Parsons, TBC	Parsons, TBC	Parsons, TBC	Parsons, TBC	Parsons, TBC
Computer Science	Computer Science	Computer Science	Computer Science	Computer Science
WORKS	WORKS	WORKS	WORKS	LECTURE

Note: workshops start from week 2

Agenda

Week	Commencing on	Topic	Delivered by
1	5/10/2022	**Introduction**	Heriberto Cuayahuitl
2	12/10/2022	Bayes nets w/exact inference	Heriberto Cuayahuitl
3	19/10/2022	Structure learning	Heriberto Cuayahuitl
4	26/10/2022	Bayes nets w/approx. inference	Simon Parsons
5	2/11/2022	Gaussian Bayes nets	Simon Parsons
6	9/11/2022	Enhancement week (no AAI sessions)	
7	16/11/2022	Recap and EM Algorithm	Heriberto Cuayahuitl
8	23/11/2022	Strategic reasoning	Simon Parsons
9	30/11/2022	Probabilistic reasoning over time I	Simon Parsons
10	7/12/2022	Probabilistic reasoning over time II	Simon Parsons
11	14/12/2022	Intro to complex decision making	Simon Parsons
12-14		Christmas break	
15	11/01/2023	Markov decision processes	Riccardo or Heriberto
16	18/01/2023	Reinforcement learning	Riccardo or Heriberto

Learning Objectives

Critically appraise a range of AI techniques for knowledge representation, reasoning and decision-making under uncertainty, identifying their strengths and weaknesses, and selecting appropriate methods to serve particular roles

Design and develop software algorithms for solving complex AI problems in an application domain of interest.

Assessments*

- Assignment (50%):
  - Bayesian Networks

- In-class test (50%):
  - Mock-test (last workshop)
  - Test (see Hand-in deadlines)

*You should **READ** the Assessment docs in Blackboard

What is AI?

Many definitions. Example “the science and engineering for equipping machines/robots to acquire their own behaviour”.

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

See Russell & Norving, 2016. Chap. 1

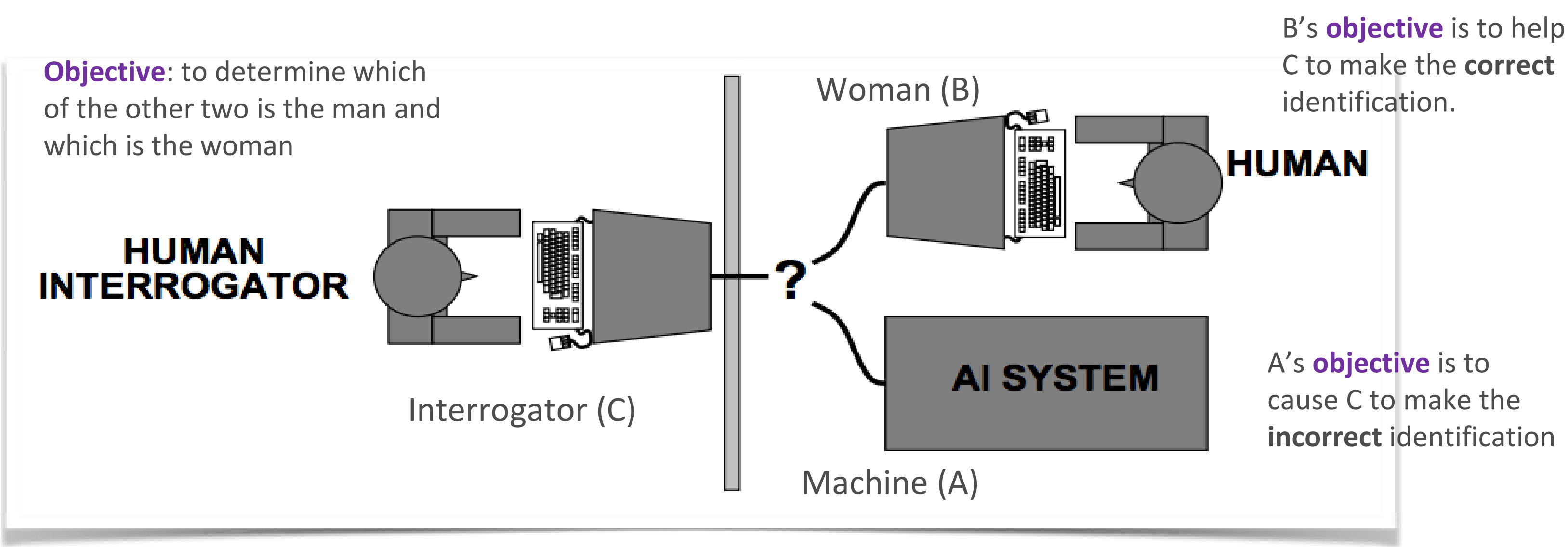
Thinking Humanly: Cognitive Science

- We must have a way of determining how humans think.
- We need to get inside the workings of human minds.
- The field of cognitive science brings together computer models from AI and experimental techniques from psychology to try to construct precise and testable theories of the workings of the human mind.

https://www.coursehero.com/file/p2fog24/Thinking-Humanly-The-cognitive-modeling-approach-For-that-we-must-determine-how/

Acting Humanly: The Turing Test (1950)

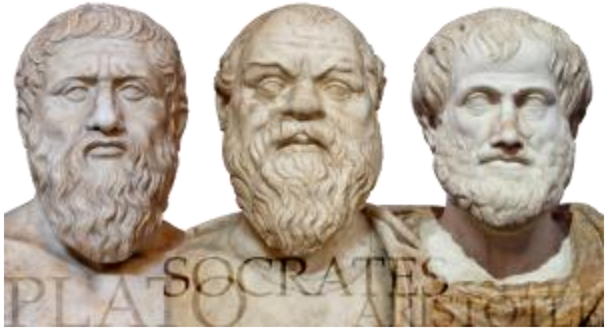
- Turing (1950) “Computing machinery and intelligence”
- Can machines behave intelligently?
- [Summary of Alan Turing’s paper (1950)](https://www.csee.umbc.edu/courses/471/papers/turing.pdf)



https://www.csee.umbc.edu/courses/471/papers/turing.pdf

Thinking Rationally: Laws of Thought

- Several Greek scholars developed various forms of logic: **notation** and **rules** for thoughts



- AI hopes to create intelligent systems using logic programming

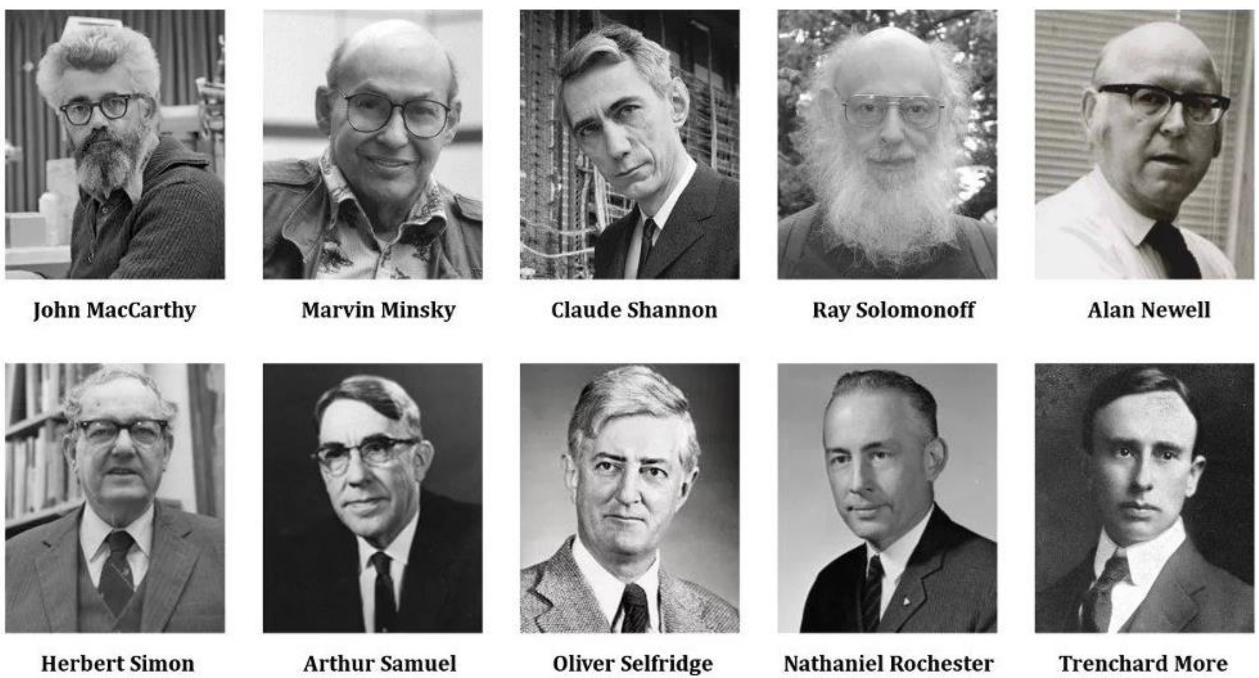
- However, it is not easy to represent informal knowledge by logical notation, particularly when knowledge is **not 100% certain**

Acting Rationally

- **Rational behaviour**: doing the right thing
- **The right thing**: which is expected to maximise goal achievement, given the available information.
- **Rational Agent** is one that achieve the best outcome or, when there is **uncertainty**, the best expected outcome.

Birth of AI

- The Dartmouth Conference (1956) brought together researchers in a variety of topics:
  - complexity theory, language simulation, neuron nets, abstraction of content from sensory inputs, relationship of randomness to creative thinking, learning machines



Picture from https://www.scienceabc.com

Pillars of AI

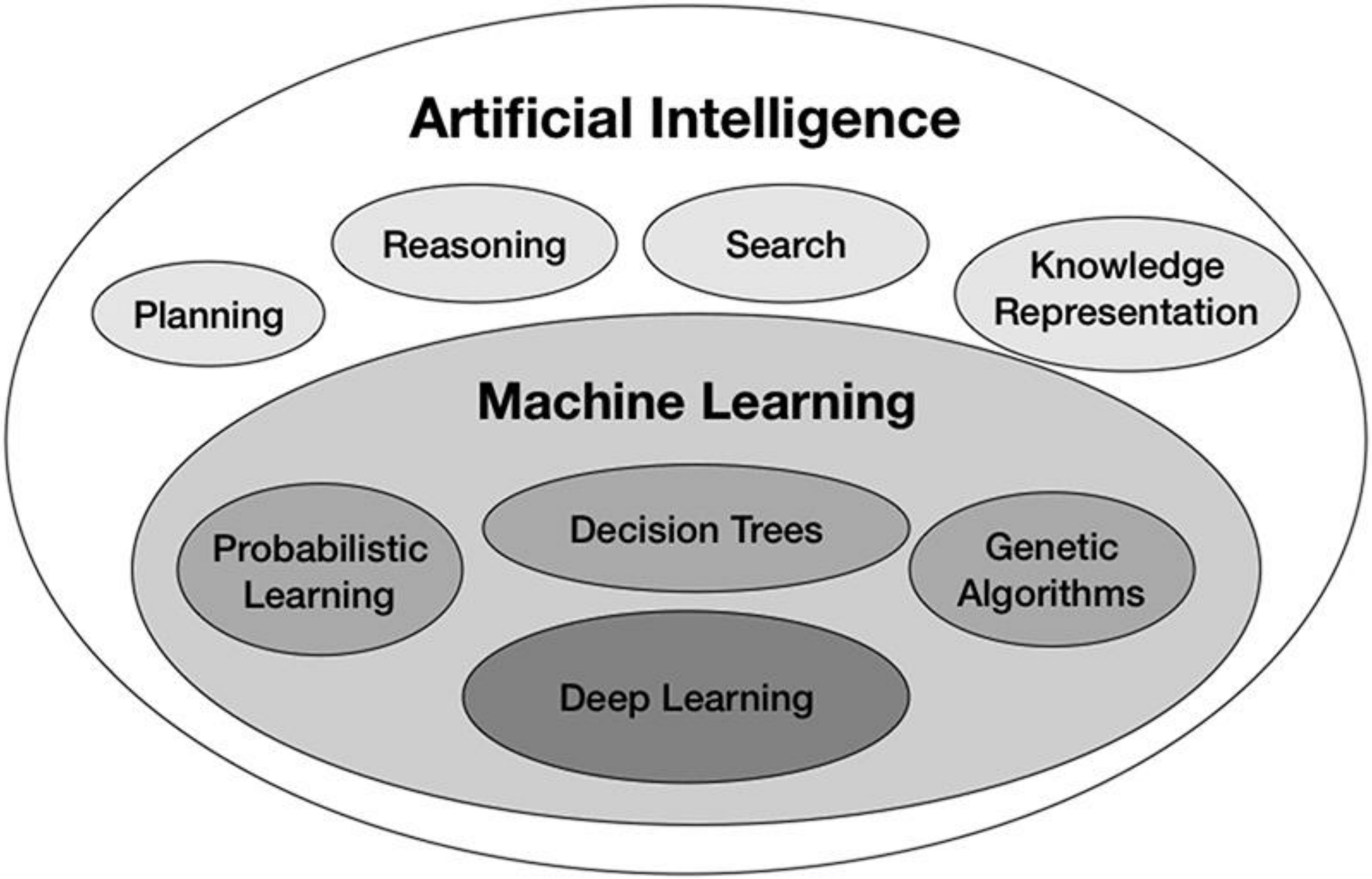


Image from
https://behavioralscientist.org

Major Areas in AI

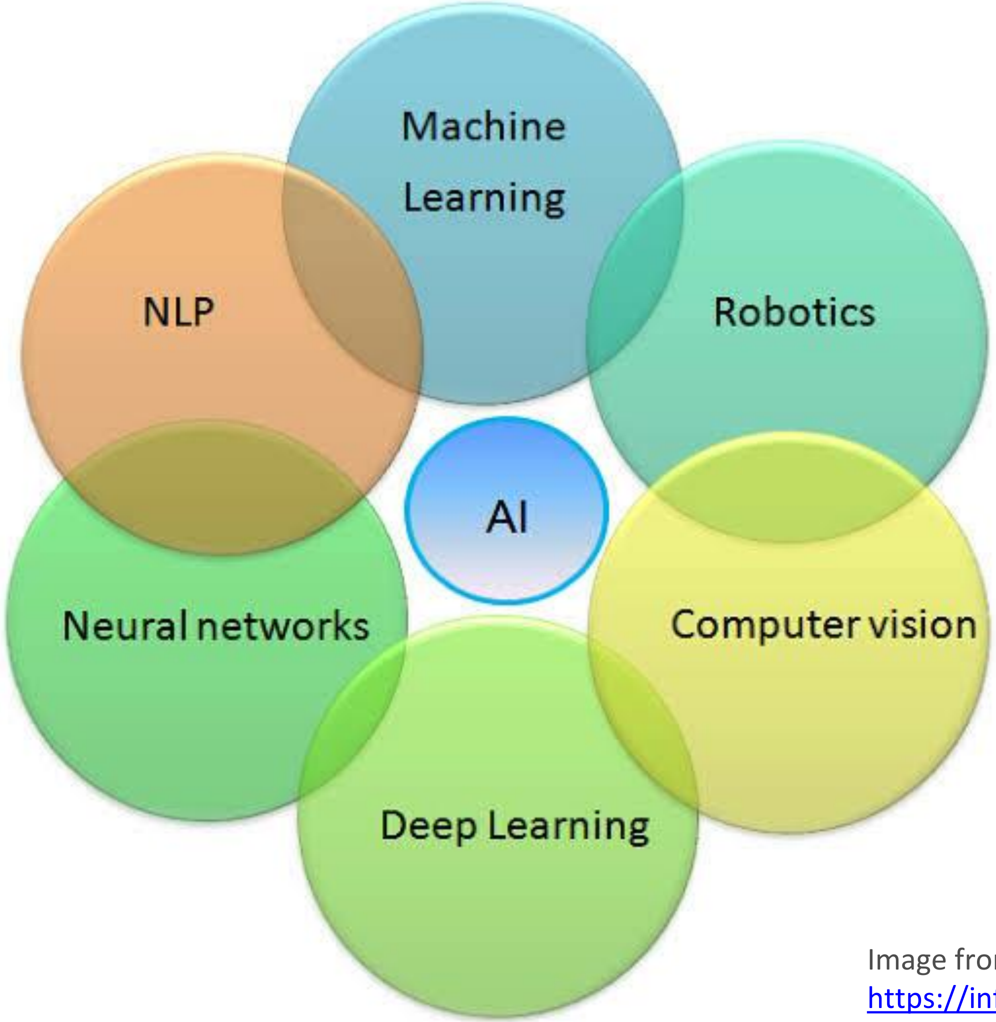


Image from
https://infohubns.blogspot.com/2020/07/research-areas-of-ai.html

History of AI in a Nutshell

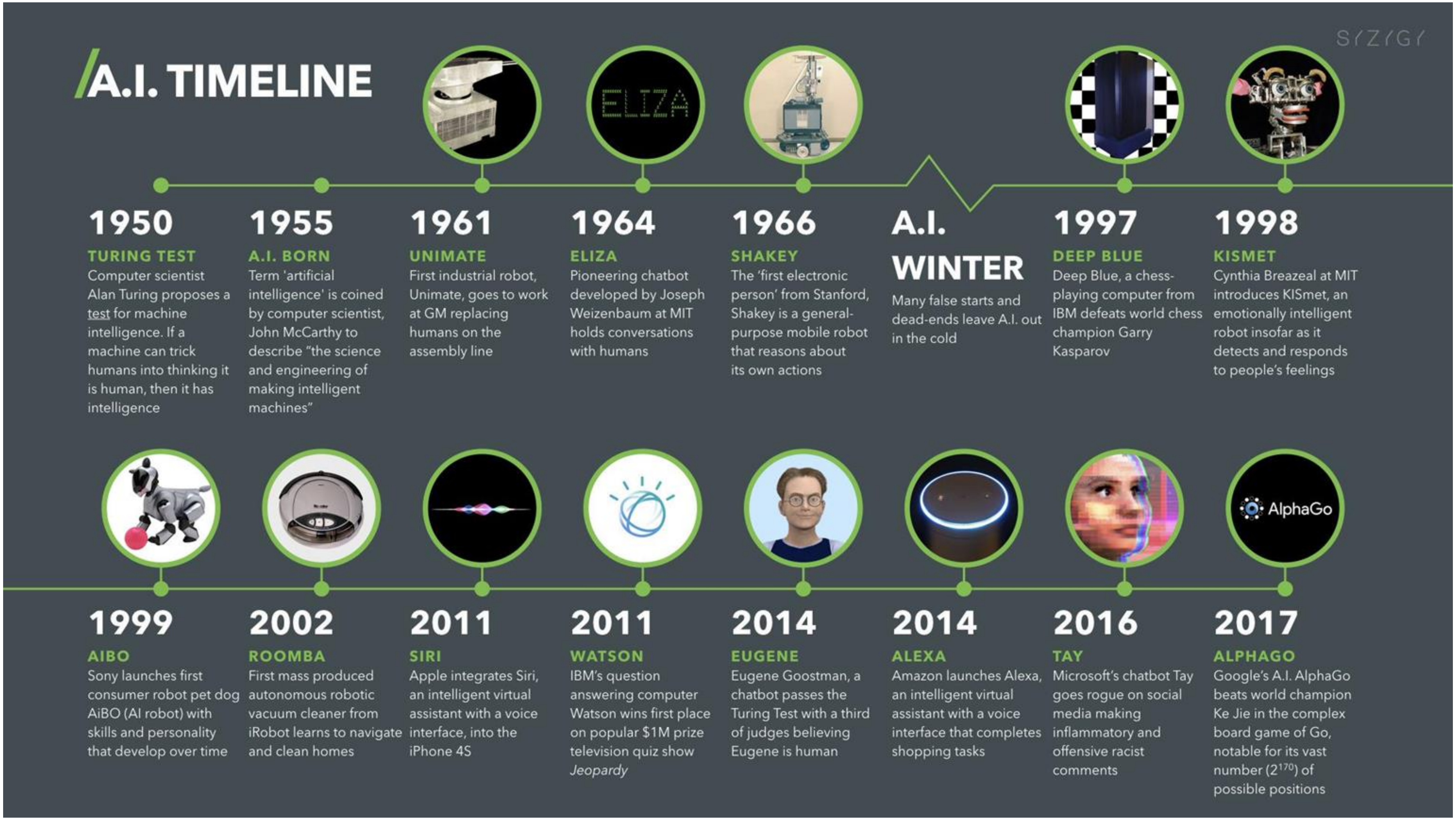
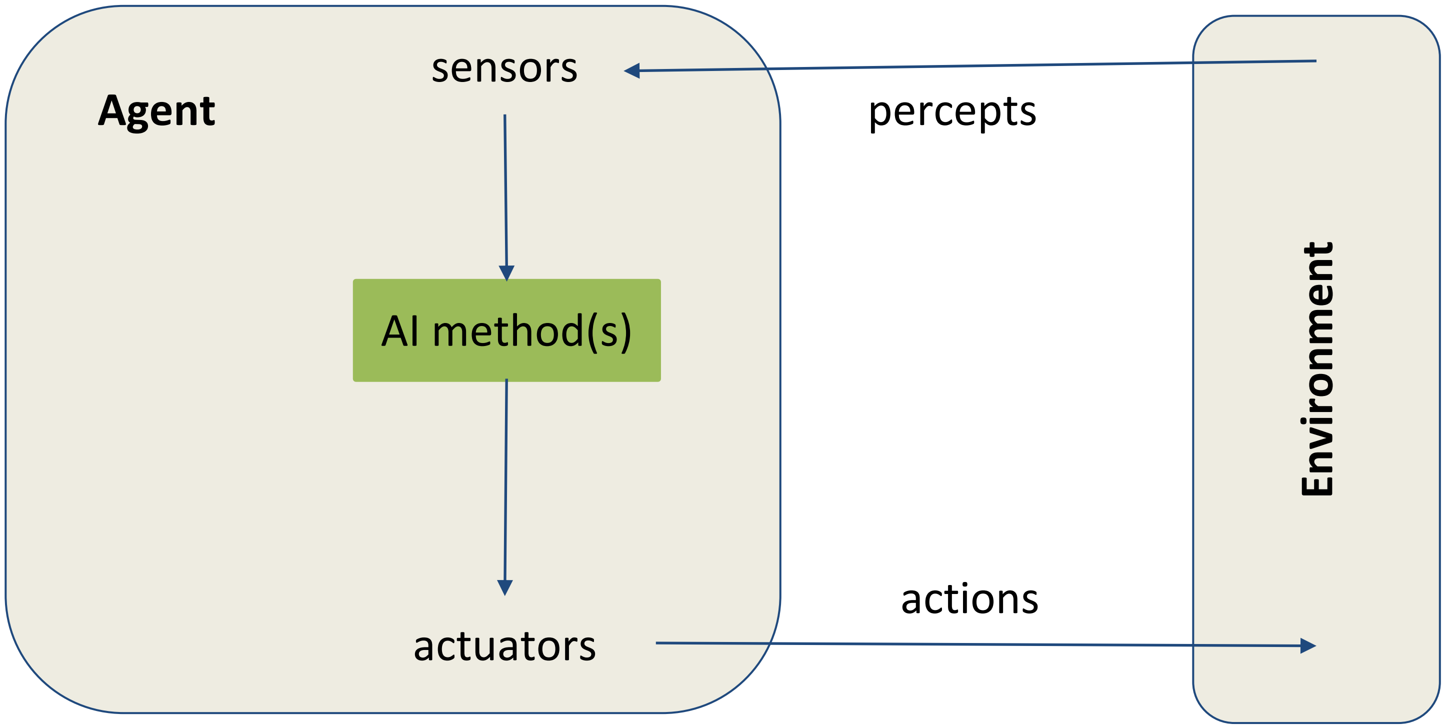


Image from https://digitalwellbeing.org/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/

Logical Agents



AI Agent perceives its environment via its sensors,
makes decisions using AI techniques, &
executes decisions using its actuators.

Properties of Task Environments

- Fully-observable vs. partially observable
- Single-agent vs. multi-agent
- Deterministic vs. stochastic
- Episodic vs. sequential
- Static vs. dynamic
- Discrete vs. continuous
- Known vs. unknown

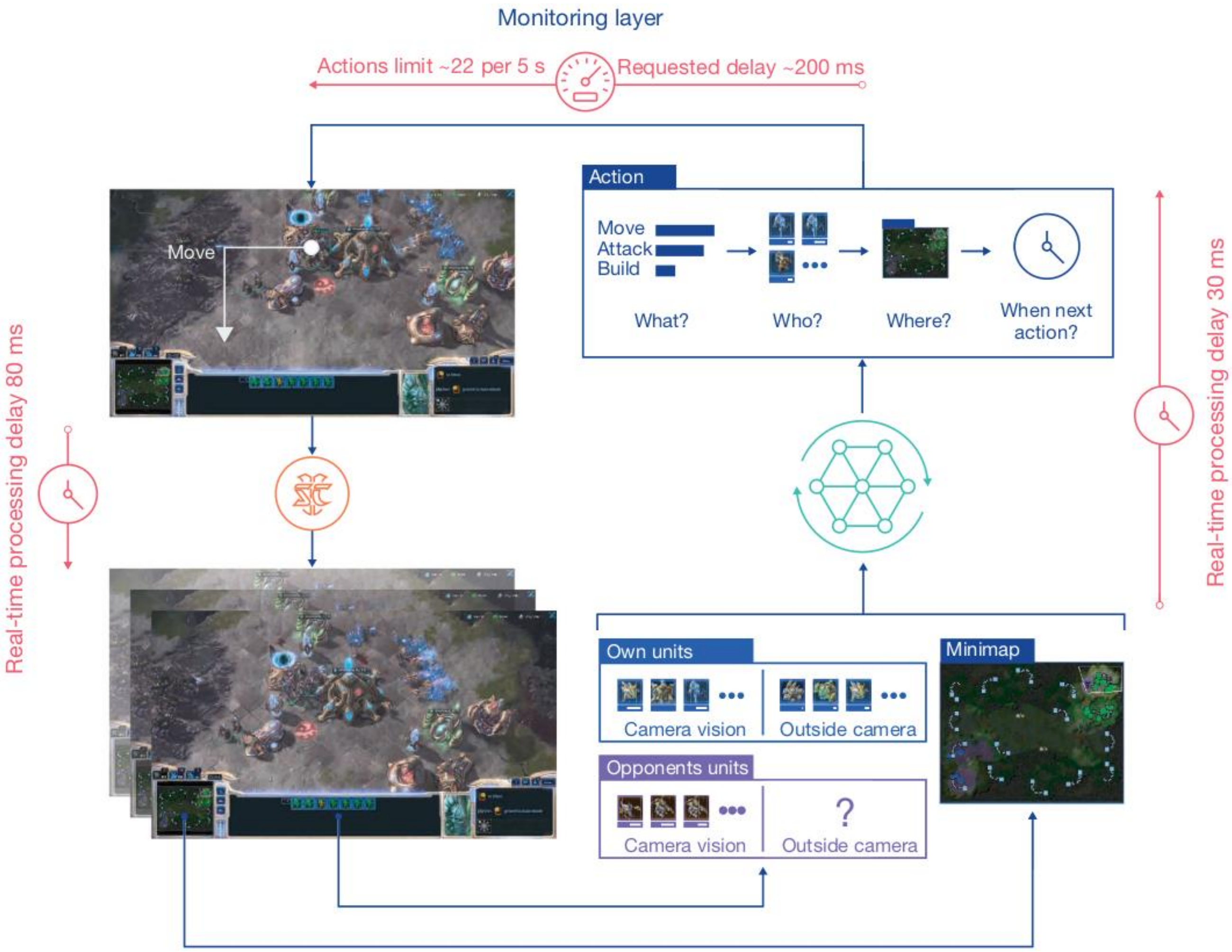
Hardest case

Example: Autonomous cars driving on unfamiliar roads.

Recent Major Developments in AI

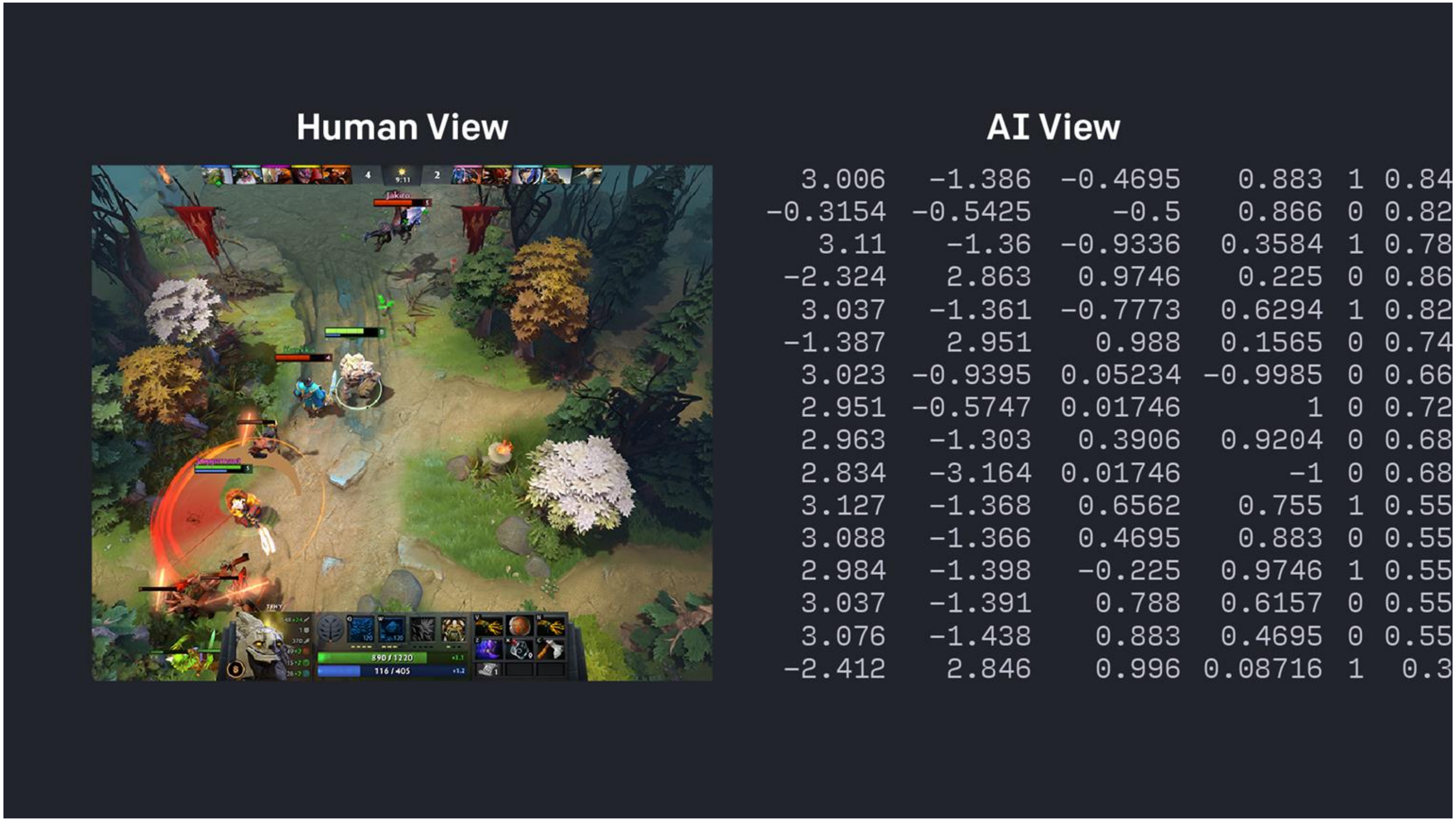
[DeepMind's AlphaGo learned to play the game of Go from scratch]

Recent Major Developments in AI



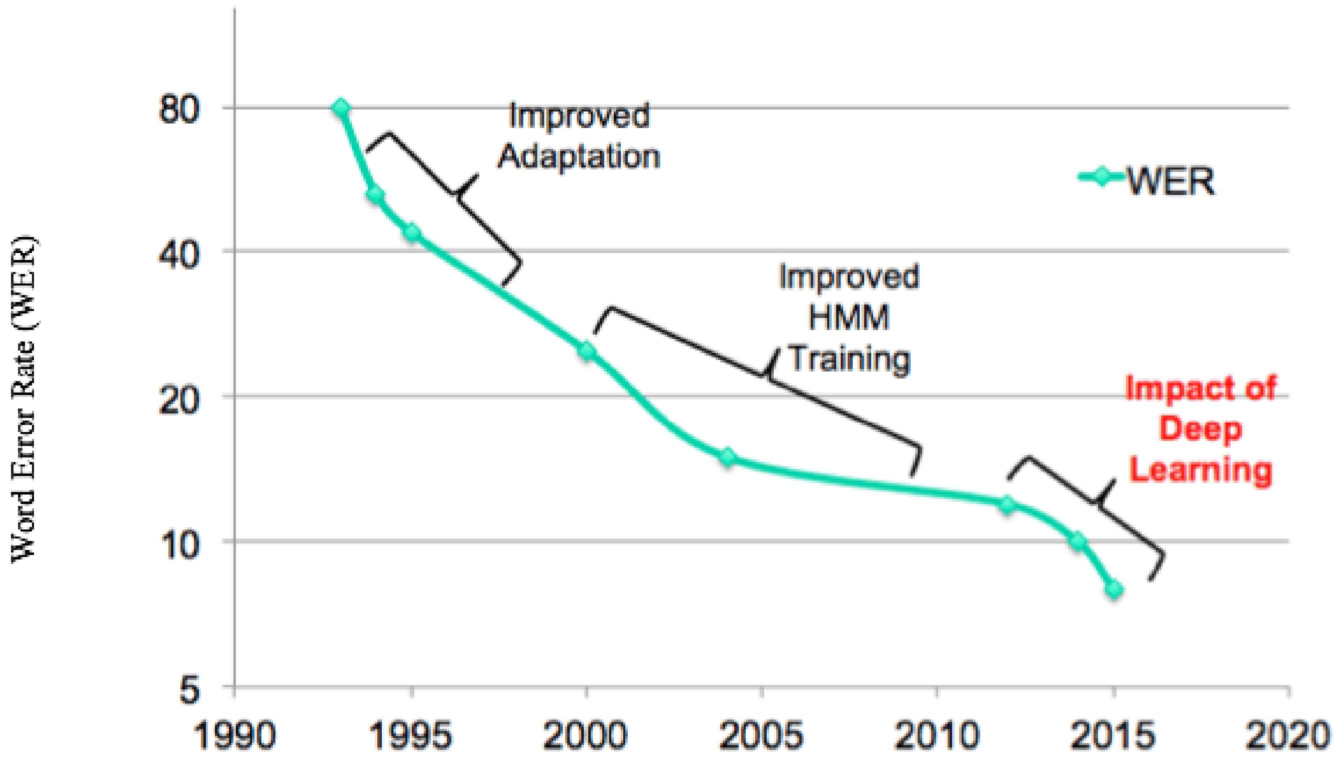
[DeepMind’s AlphaStar learns to play StarCraft II at GrandMaster level]()

Recent Major Developments in AI



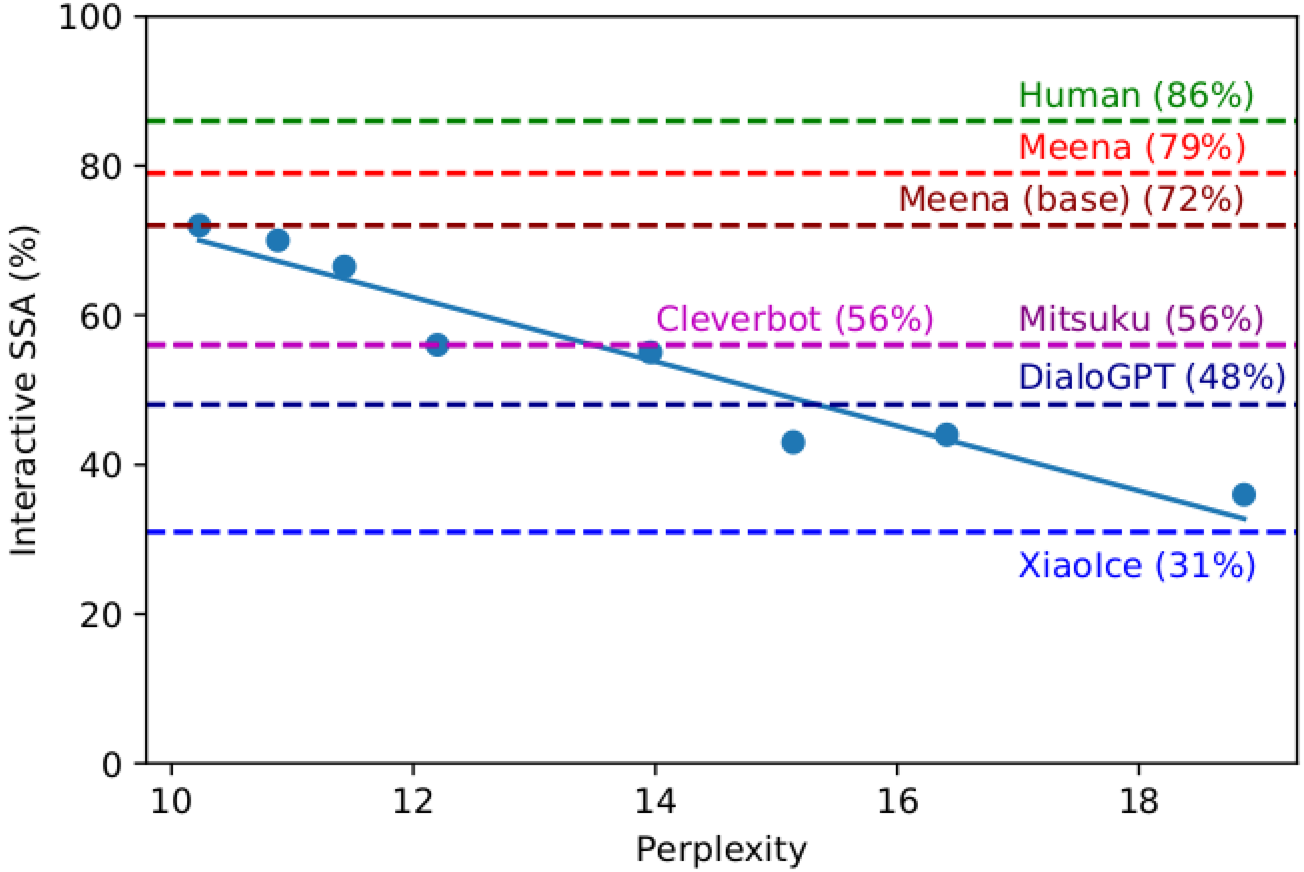
[OpenAI Five is the 1st AI system to defeat world champs at e-sports]()

Recent Major Developments in AI



[Deep learning brought substantial improvements in ASR performance]

Recent Major Developments in AI

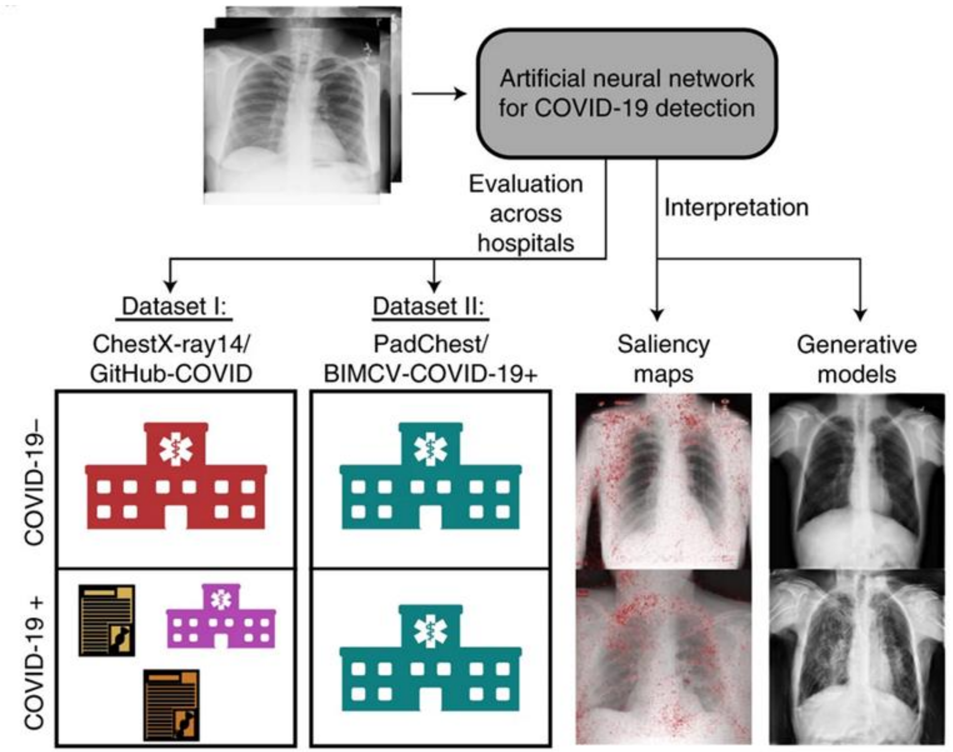


[Google’s Meena, a 2.6B parameter neural net, learns to chat on 341GB of text]

Recent Major Developments in AI

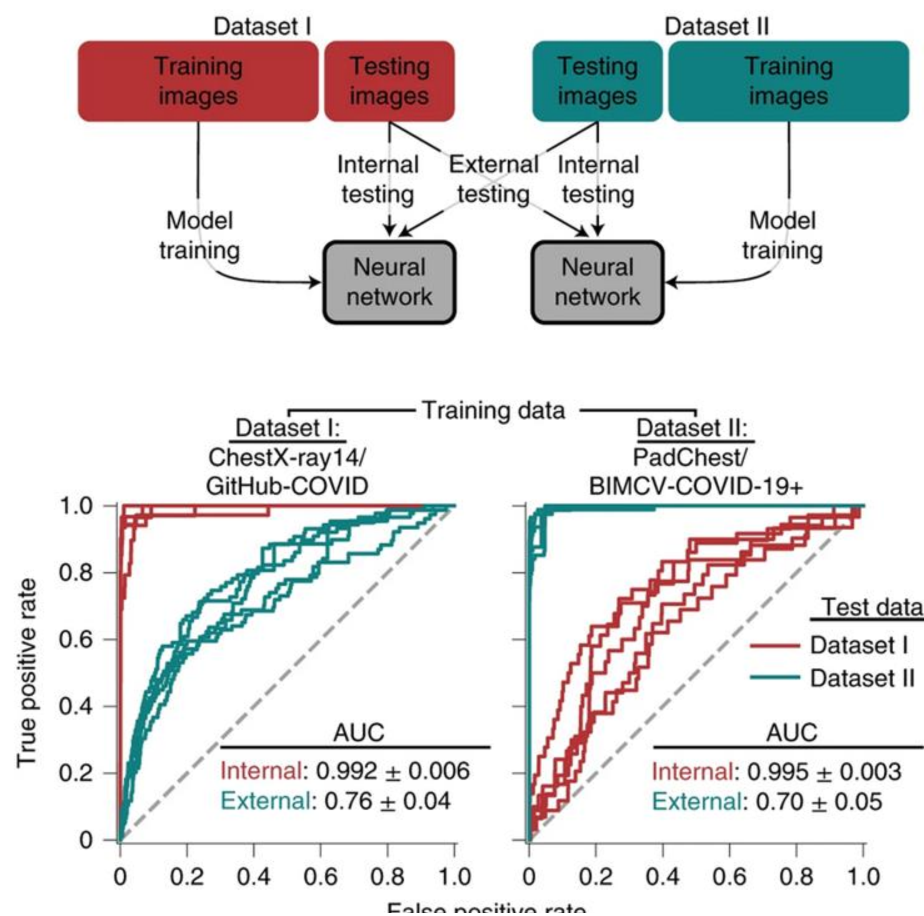
[Aachen+MPI+Tuebingen train neural nets to track multiple objects]

Recent Major Developments in AI



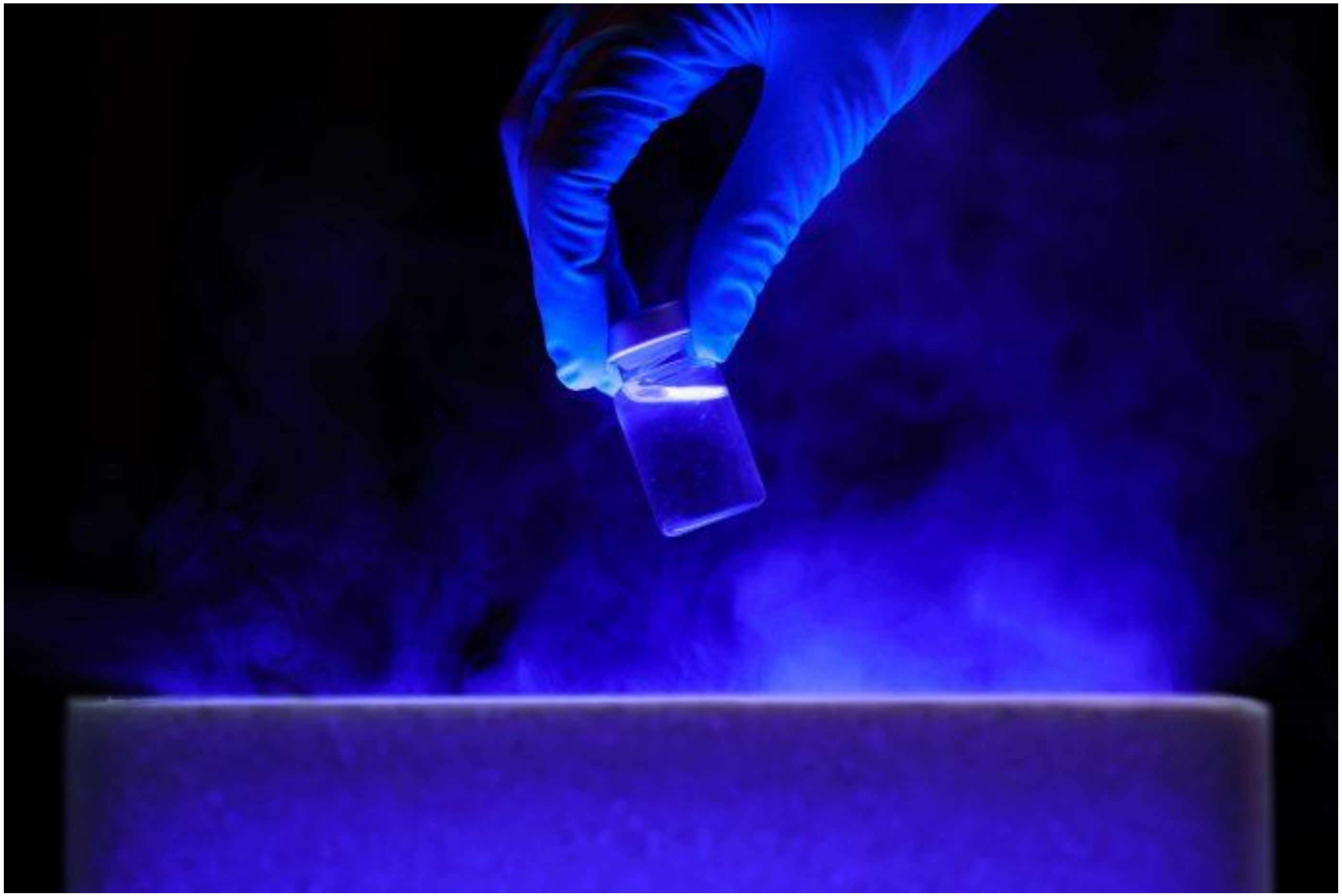
b

	Dataset I			Dataset II		
	Combined	Chest-X-ray14	GitHub-COVID	Combined	PadChest	BIMCV-COVID-19+
No. radiographs	112,528	112,120	408	97,866	96,270	1,596
No. patients	31,067	30,805	262	64,954	63,939	1,105



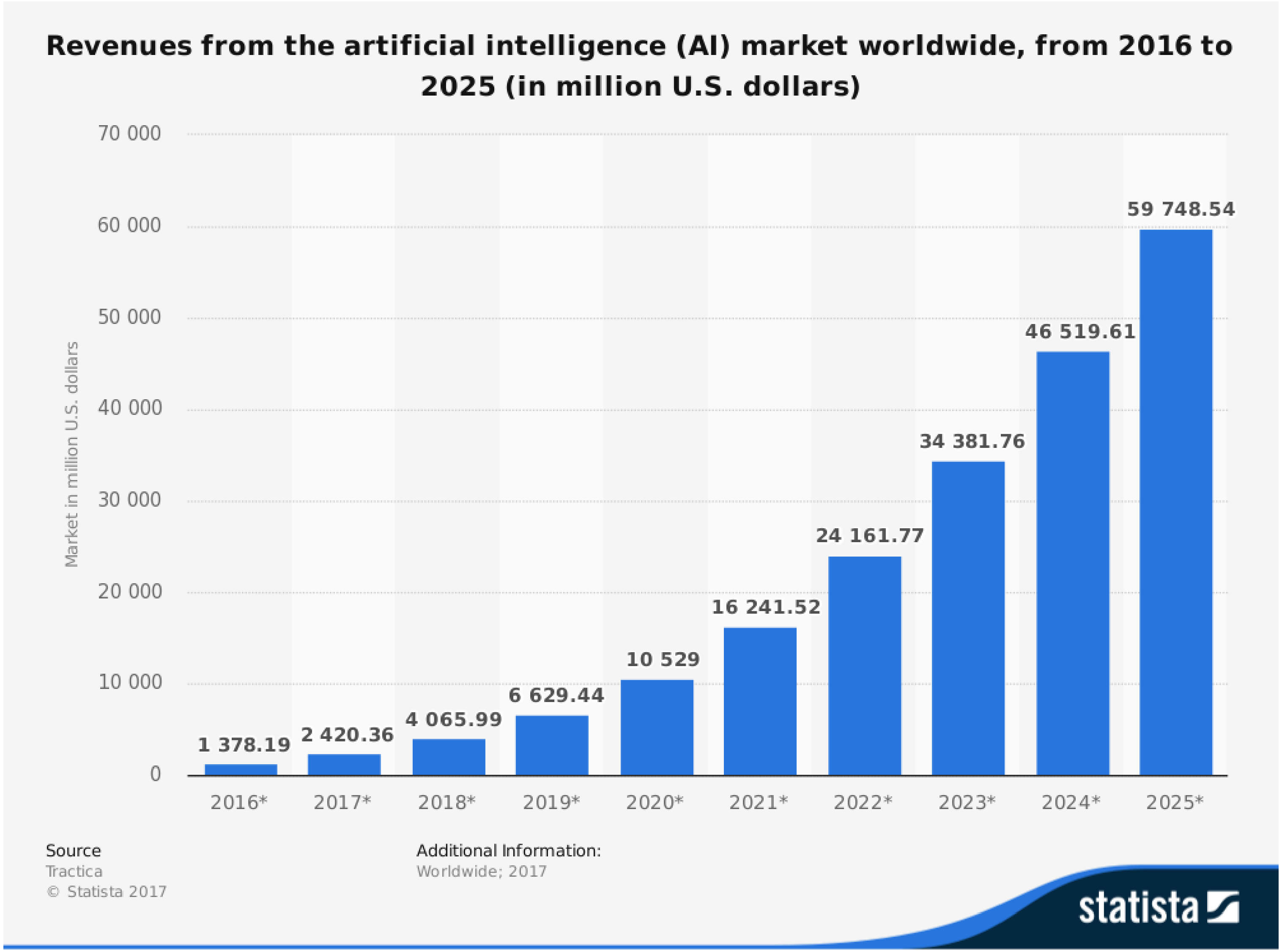
[Medical-imaging AI system learns to detect Covid-19 from chest radiographs]

Recent Major Developments in AI



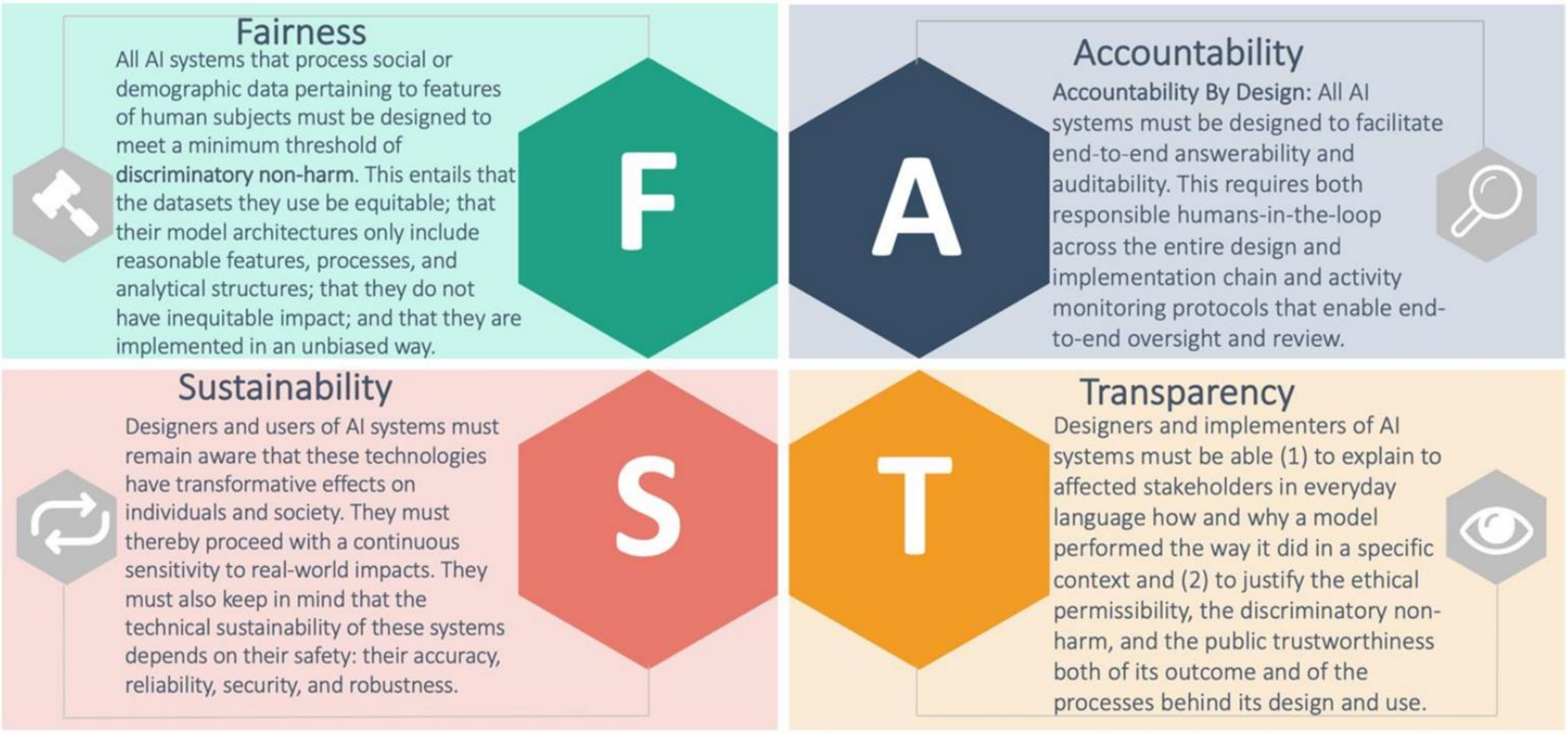
[AI system can find COVID vaccine candidates within seconds]

AI: Economic Impact



source

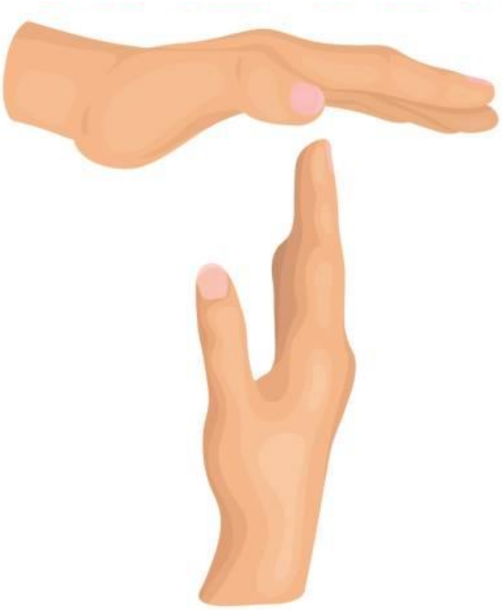
Ethics in AI



[Understanding Artificial Intelligence ethics and safety]

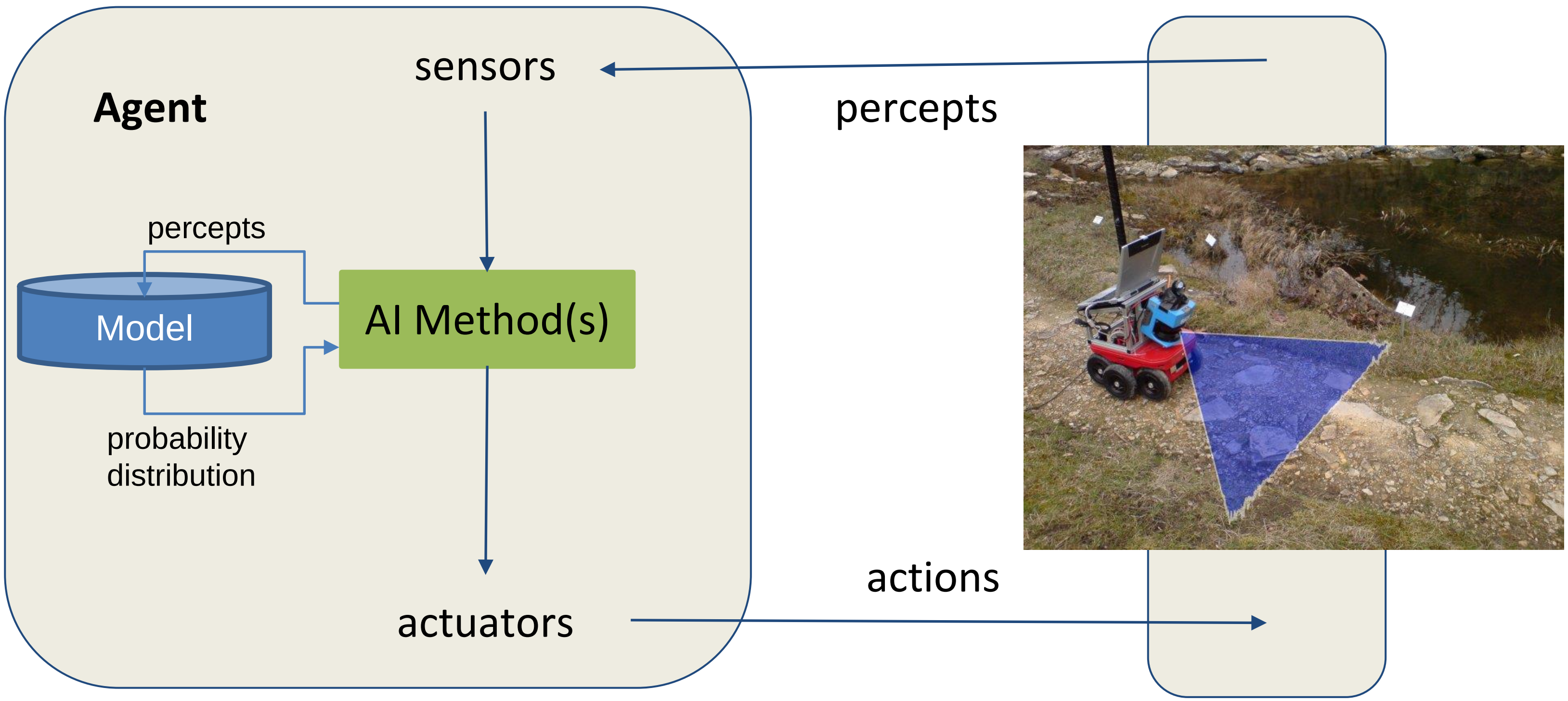
[Data ethics and AI guidance landscape]

Break



Acting Under Uncertainty

- **Perceptions** may be noisy, imprecise and/or incomplete
- **Actions** may not always lead to the intended effects
- **Prior knowledge** (e.g., map for route planning) may not be reliable



Axioms of Probability

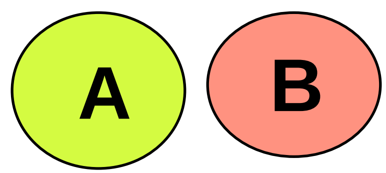
Axioms are basic definitions or rules in maths

Probabilities must satisfy the following:

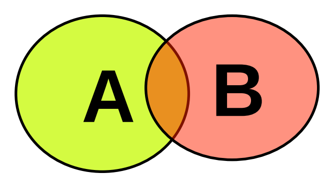
- $0 \leq P(A) \leq 1$

"or"

- $P(A \vee B) = P(A) + P(B)$



- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



- $P(true) = 1$

Total Probability Must Equal 1

- Suppose a set of events is mutually exclusive and collectively exhaustive. This means that one (and only one) of the possible outcomes must occur
- The probabilities for those events must sum to 1:

$$\sum_{e \in S} P(e) = 1$$

P(throw 1) + P(throw 2) + P(throw 3)+
P(throw 4) + P(throw 5) + P(throw 6) **= 1**

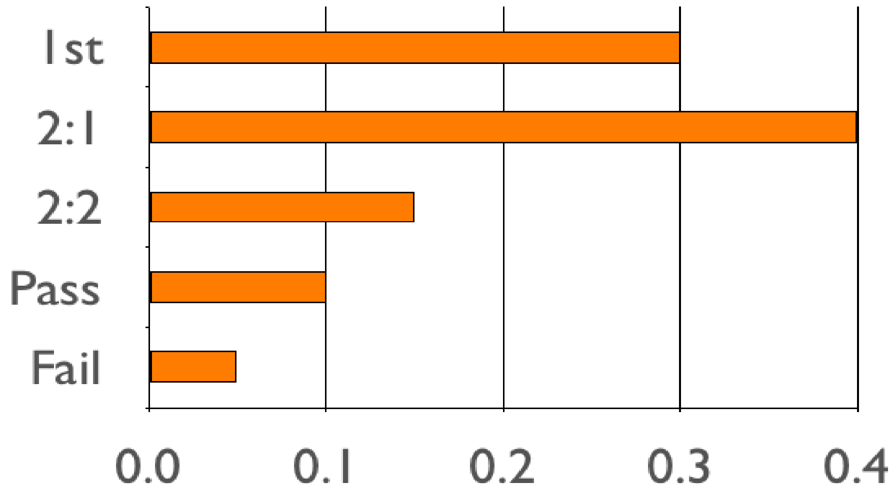
(Discrete) Random Variables

- The events we are interested in have a set of possible values. Examples of random variables:
  - coin toss: {heads, tails}
  - roll a die: {1, 2, 3, 4, 5, 6}
  - weather: {snow, sunny, rain, fog}
  - measles: {true, false}

- For each event, a random variable takes a value from the associated set. Then we have:
  - $P(C = tails)$ rather than $P(tails)$
  - $P(D = 1)$ rather than $P(1)$
  - $P(W = sunny)$ rather than $P(sunny)$
  - $P(M = true)$ rather than $P(measles)$

Discrete Probability Distribution

- A probability distribution is a listing of probabilities for *every possible* value of a single random variable



P(1st)=0.30

P(2:1)=0.40

P(2:2)=0.15

P(Pass)=0.10

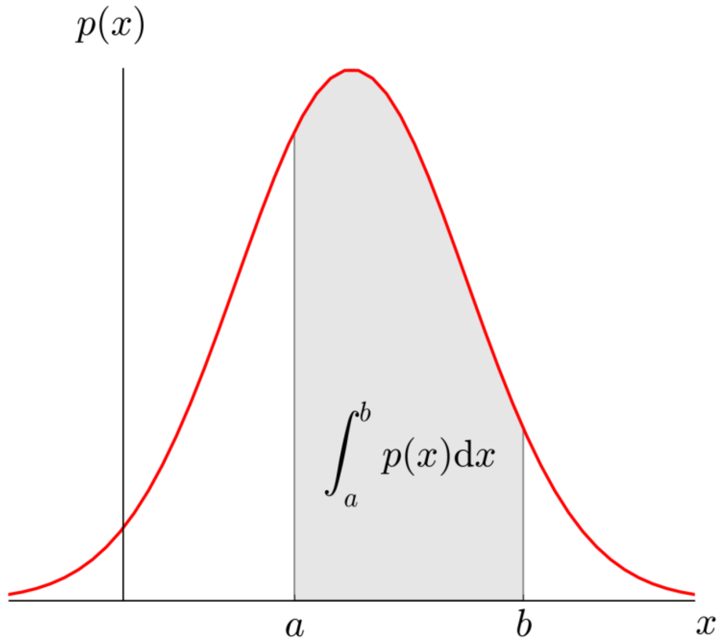
P(Fail)=0.05

$$\sum_{e \in S} P(e) = 1$$

- Probability distributions will be estimated from data

Continuous Probability Distribution

- The distribution of a *continuous random variable* is represented by a **probability density function (PDF)**



- Due to the infinite number of possible values of a continuous variable

Joint Probabilities

Joint probabilities represent the whole joint probability distribution

S1=1st	S2=1st	Probability
TRUE	TRUE	0.2
TRUE	FALSE	0.1
FALSE	TRUE	0.1
FALSE	FALSE	0.6

$$\sum_{e \in S} P(e) = 1$$

Joint Probability Distribution

Sometimes a joint probability distribution looks like the one below, which has the same information as the table on the previous slide.

	S2 =1st	**¬(S2** =1st**)**
S1=1st	0.2	0.1
¬(S1=1st)	0.1	0.6

Marginal Probabilities

	S2 =1st	¬(**S2** =1st)
S1=1st	0.2	0.1
¬(S1=1st)	0.1	0.6

Marginalisation => summing up the probabilities of the other variables—i.e. taking them out of the equation.

Example: What is the probability of S1 getting a 1st?

$$\begin{aligned} P(S1=1st) &= P(S1=1st \land S2=1st) + P(S1=1st \land \neg(S2=1st)) \\ &= 0.2 + 0.1 = 0.3 \end{aligned}$$

Marginal Probabilities

	S2 =1st	¬(**S2** =1st)
S1=1st	0.2	0.1
¬(S1=1st)	0.1	0.6

$P(S1{=}1st) = P(S1{=}1st \wedge S2{=}1st) + P(S1{=}1st \wedge \neg(S2{=}1st))$
$= 0.2 + 0.1 = 0.3$

$P(\neg(S1{=}1st)) = P(\neg(S1{=}1st) \wedge S2{=}1st) + P(\neg(S1{=}1st) \wedge \neg(S2{=}1st))$
$= 0.1 + 0.6 = 0.7$

Note that $P(S1{=}1st) + P(\neg(S1{=}1st)) = 0.3 + 0.7 = 1$

Marginal Probabilities

	S2 =1st	¬(**S2** =1st)
S1=1st	0.2	0.1
¬(S1=1st)	0.1	0.6

P(S2=1st) = P(S1=1st ∧ S2=1st) + P(¬(S1=1st) ∧ S2=1st)
= 0.2 + 0.1 = 0.3

P(¬(S2=1st)) = P(S1=1st ∧ ¬(S2=1st)) + P(¬(S1=1st) ∧ ¬(S2=1st))
= 0.1 + 0.6 = 0.7

Note that P(S2=1st)+ P(¬(S2=1st))=0.3+0.7=1

Conditional Probability

- Conditional probabilities answer the question: *Given that some event B happened, what is the probability of A happening too?*

- **P(A|B)**=?

- A conditional probability is defined as:
  $P(A \mid B) = P(A \land B) / P(B)$, where $P(B) \neq 0$

Conditional Probability

	S2=1st	**¬(S2=1st)**
S1=1st	0.2	0.1
¬(S1=1st)	0.1	0.6

- P(S2=1st | S1=1st)?
- = P(S2=1st, S1=1st) / P(S1=1st)
- = 0.2 / 0.3 = 0.6666666

P(A | B) = P(A ∧ B) / P(B)

Substituting A with S2=1st

Substituting B with S1=1st

If the first student has a 1st, the second has a 66% chance of having a 1st too!

Conditional Probability

	S2=1st	¬(S2=1st)
S1=1st	0.2	0.1
¬(S1=1st)	0.1	0.6

- P(¬(S2=1st) | S1=1st)?
- = P(¬(S2=1st), S1=1st) / P(S1=1st)
- = 0.1 / 0.3 = 0.3333333

P(A | B) = P(A ∧ B) / P(B)

Substituting A with ¬(S2=1st)

Substituting B with S1=1st

If the first student (S1) has a 1st, the second has a 33% chance of not getting a 1st.

Conditional Probability

	S2=1st	**¬(S2=1st)**
S1=1st	0.2	0.1
¬(S1=1st)	0.1	0.6

Since P(S2=1st | S1=1st) = P(S2=1st, S1=1st) / P(S1=1st)

= 0.2 / 0.3 = 0.6666666

and P(¬(S2=1st) | S1=1st) = P(¬(S2=1st), S1=1st) / P(S1=1st)

= 0.1 / 0.3 = 0.3333333

Then P(S2=1st | S1=1st) + P(¬(S2=1st) | S1=1st) = 1

More on Joint Probabilities

Given a joint probability table, we have all the information we need about the domain. We can calculate the probability of any logical formula.

P(S1=first ∨ S2=first) = 0.2 + 0.1 + 0.1 = 0.4

	S2 =1st	¬(**S2** =1st)
S1=1st	$\boxed{0.2}$	$\boxed{0.1}$
¬(S1=1st)	$\boxed{0.1}$	0.6

Example in the Medical Domain

	Measles	**¬Measles**
Rash	0.1	$\boxed{0.8}$
¬Rash	0.01	0.09

PROBABILISTIC QUESTION: What is the probability of not having measles given that a person has a rash?

Example in the Medical Domain

	Measles	**¬Measles**
Rash	0.1	$\boxed{0.8}$
¬Rash	0.01	0.09

PROBABILISTIC QUESTION: What is the probability of not having measles given that a person has a rash?

$P(A \mid B) = P(A \land B) / P(B)$

$P(\neg M \mid R) = P(\neg M \land R) / P(R)$

Example in the Medical Domain

	Measles	**¬Measles**
Rash	0.1 $P(M \wedge R)$	0.8 $P(\neg M \wedge R)$
¬Rash	0.01 $P(M \wedge \neg R)$	0.09 $P(\neg M \wedge \neg R)$

PROBABILISTIC QUESTION: What is the probability of not having measles given that a person has a rash?

$P(\neg M | R) = P(\neg M \wedge R) / P(R)$

Example in the Medical Domain

	Measles	¬Measles
Rash	0.1	$\boxed{0.8}$
¬Rash	0.01	0.09

PROBABILISTIC QUESTION: What is the probability of not having measles given that a person has a rash?

$P(\neg M|R) = P(\neg M \wedge R) / P(R)$

$= 0.8 / 0.9 = \mathbf{0.888}$

Example in the Medical Domain

	Measles	**¬Measles**
Rash	$\boxed{0.1}$	0.8
¬Rash	0.01	0.09

PROBABILISTIC QUESTION: What is the probability of having measles given that a person has a rash?

Example in the Medical Domain

	Measles	¬Measles
Rash	0.1	0.8
¬Rash	0.01	0.09

PROBABILISTIC QUESTION: What is the probability of having measles given that a person has a rash?

$P(M|R) = P(M \wedge R) / P(R)$

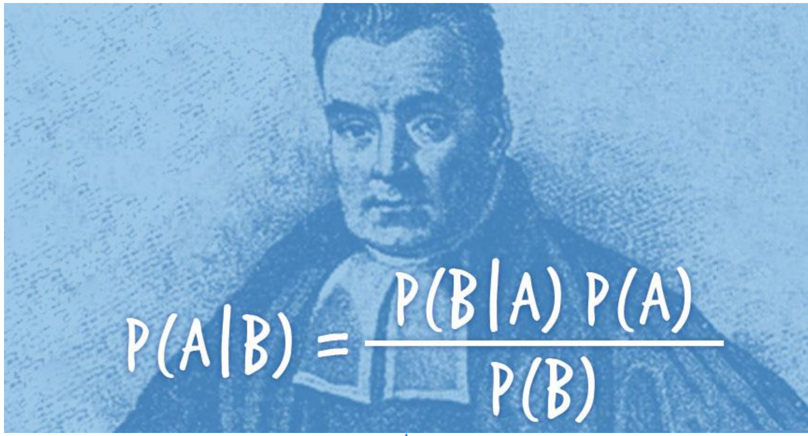
$= 0.1 / 0.9 = \mathbf{0.111}$

Bayes Rule

- **$P(A \wedge B)$ = $P(B \wedge A)$**
- **$P(A \mid B)*P(B) = P(B \mid A)*P(A)$**
- **$P(A \mid B) = (P(B \mid A)*P(A)) / P(B)$**
- **$P(B \mid A) = (P(A \mid B)*P(B)) / P(A)$**

Bayes Rule

- $P(A \land B) = P(B \land A)$
- $P(A \mid B) * P(B) = P(B \mid A) * P(A)$
- $P(A \mid B) = (P(B \mid A) * P(A)) / P(B)$
- $P(B \mid A) = (P(A \mid B) * P(B)) / P(A)$



Thomas Bayes
(1701-1761)

Bayes Rule for Classification

- Given inputs *X* and outputs *Y*, the Bayes rule can be written as

$$P(Y = y_k | X = x_i) = \frac{P(X = x_i | Y = y_k) P(Y = y_k)}{\sum_j P(X = x_i | Y = y_j) P(Y = y_j)}$$

where

y_k is the possible value for Y

x_i is the possible vector value for X

- Use training data to estimate $P(X|Y)$ and $P(Y)$.

Difficulty in Unbiased Bayesian Classifiers

- Accurately estimating $P(X|Y)$ requires a set of parameters such as $\theta_{ij} = P\left(X = x_i \middle| Y = y_j\right)$, where

  index j refers to 2 possible values

  index i refers to 2^n possible values

- This requires 2^{n+1} parameters or probabilities.

- A vector X with 30 Boolean inputs requires 2.15B parameters

Naïve Bayes Classifier

- Conditional independence: $P(X_1 \dots X_n|Y) = \prod_{i=1}^{n} P(X_i|Y)$

- Thus

$$P(Y = y_k|X_1 \dots X_n) = \frac{P(Y = y_k)P(X_1 \dots X_n|Y = y_k)}{\sum_j P(Y = y_j)P(X_1 \dots X_n|Y = y_j)}$$

can be rewritten as

$$P(Y = y_k|X_1 \dots X_n) = \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j)\prod_i P(X_i|Y = y_j)}$$

Naïve Bayes Classifier

- Given a new instance/example $X^{new} =< X_1 \dots X_n >$, the most probable value of Y can be obtained with

$$Y = \arg\max_{y_k} \frac{P(Y = y_k) \prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j) \prod_i P(X_i|Y = y_j)}$$

- Since the denominator does not depend on y_k, the equation can be simplified as

$$Y = \underset{\mathrm{y_k}}{\operatorname{argmax}} P(Y = y_k) \prod_i P(X_i|Y = y_k)$$

Naïve Bayes for Discrete Inputs

- Estimate two sets of parameters:

$$\theta_{ijk} = P\left(X = x_{ij} \middle| Y = y_k\right)$$
$$\pi_k = P(Y = y_k)$$

- According to

$$\theta_{ijk} = P\left(X = x_{ij} \middle| Y = y_k\right) = \frac{\#D\{X_i = x_{ij} \wedge Y = y_k\} + l}{\#D\{Y = yk\} + lJ}$$
$$\pi_k = P(Y = y_k) = \frac{\#D\{Y = y_k\}}{|D|}$$

where J is the number of unique values in X_i, l avoids zero estimates, and $|D|$ is the number of elements in the training set.

Example: Training Data

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Data from: [Mitchell, T. "Machine Learning", McGraw Hill, 1997].

Example: Estimated Parameters

$$P(PlayTennis = yes) = \frac{9}{14} = 0.643$$
$$P(PlayTennis = no) = \frac{5}{14} = 0.357$$

$$P(Outlook = sunny|PlayTennis = yes) = 2/9 = 0.222$$
$$P(Outlook = sunny|PlayTennis = no) = 3/5 = 0.60$$
$$P(Outlook = overcast|PlayTennis = yes) = 4/9 = 0.444$$
$$P(Outlook = overcast|PlayTennis = no) = 0/5 = 0.0$$
$$P(Outlook = rain|PlayTennis = yes) = 3/9 = 0.333$$
$$P(Outlook = rain|PlayTennis = no) = 2/5 = 0.4$$

...

We want to avoid zero probabilities

Example: Estimated Parameters

$$P(Temperature = hot|PlayTennis = yes) = 2/9 = 0.222$$
$$P(Temperature = hot|PlayTennis = no) = 2/5 = 0.4$$
$$P(Temperature = mild|PlayTennis = yes) = 4/9 = 0.444$$
$$P(Temperature = mild|PlayTennis = no) = 2/5 = 0.4$$
$$P(Temperature = cool|PlayTennis = yes) = 3/9 = 0.333$$
$$P(Temperature = cool|PlayTennis = no) = 1/5 = 0.2$$

$$P(Humidity = high|PlayTennis = yes) = 3/9 = 0.333$$
$$P(Humidity = high|PlayTennis = no) = 4/5 = 0.8$$
$$P(Humidity = normal|PlayTennis = yes) = 6/9 = 0.666$$
$$P(Humidity = normal|PlayTennis = no) = 1/5 = 0.2$$

$$P(Wind = strong|PlayTennis = yes) = 3/9 = 0.333$$
$$P(Wind = strong|PlayTennis = no) = 3/5 = 0.6$$
$$P(Wind = weak|PlayTennis = yes) = 6/9 = 0.666$$
$$P(Wind = weak|PlayTennis = no) = 2/5 = 0.4$$

Example: Classifying a New Instance

$$P(yes)P(sunny|yes)P(cool|yes)P(high|yes)P(strong|yes) =$$
$$= 0.643 * 0.222 * 0.333 * 0.333 * 0.333 = 0.0053$$

$$P(no)P(sunny|no)P(cool|no)P(high|no)P(strong|no) =$$
$$= 0.357 * 0.60 * 0.2 * 0.8 * 0.6 = 0.0206$$

$$P(PlayTennis = yes|evidence) = \frac{0.0053}{0.0053 + 0.0206} = 0.205$$

$$P(PlayTennis = no|evidence) = \frac{0.0206}{0.0053 + 0.0206} = 0.795$$

Same example with Log Probabilities

$$P(yes)P(sunny|yes)P(cool|yes)P(high|yes)P(strong|yes)$$
$$= (-0.442) + (-1.504) + (-1.098) + (-1.098) + (-1.098)$$
$$= -5{,}242$$

$$P(no)P(sunny|no)P(cool|no)P(high|no)P(strong|no)$$
$$= (-1.029) + (-0.51) + (-1.609) + (-0.223) + (-0.51)$$
$$= -3.883$$

$$P(PlayTennis = yes|evidence) = \frac{e^{-5{,}242}}{e^{-5{,}242} + e^{-3.883}} = 0.205$$

$$P(PlayTennis = no|evidence) = \frac{e^{-3.883}}{e^{-5{,}242} + e^{-3.883}} = 0.795$$

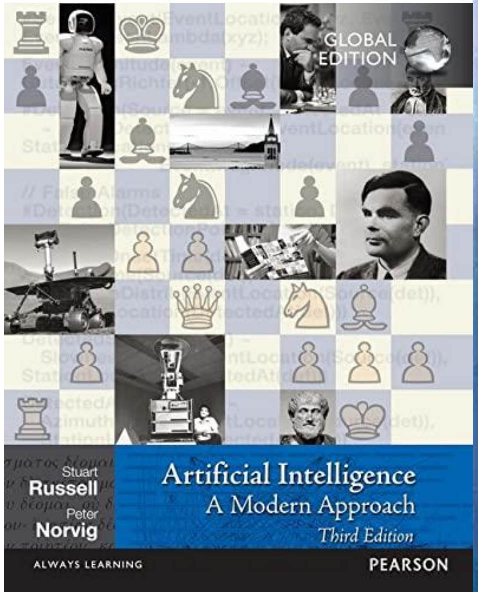
Today

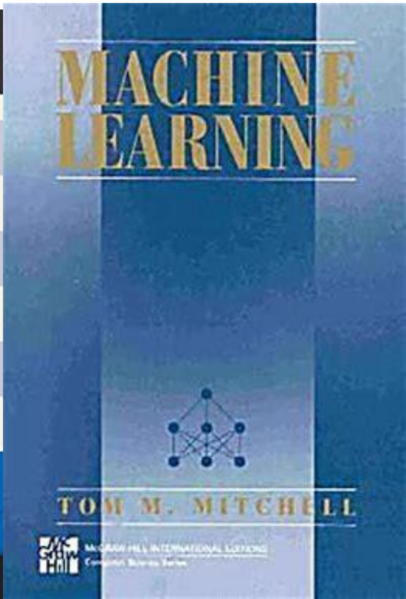
- Elaborated on the term `**Artificial Intelligence**'
- Some recent major developments in AI
- Introduction to probability theory
- Probabilistic reasoning with Naïve Bayes

Readings:

Russell & Norvig 2016. [Chapters 1,2]

Mitchell, T. 2017; 2nd Ed. [Chapter 3]





Next Week

Workshop:

Exercises on probability theory

Python program for Naïve Bayes inference

Lecture:

Bayesian Networks with Exact Inference

Questions?