

<https://attendance.lincoln.ac.uk>

Access Code: **662942**

**CMP9794M**

# **Advanced Artificial Intelligence**

[Heriberto Cuayahuitl](#)



UNIVERSITY OF  
**LINCOLN**

School of Computer Science

# Delivery Team

**Dr. Heriberto Cuayahuitl**  
[hcuayahuitl@lincoln.ac.uk](mailto:hcuayahuitl@lincoln.ac.uk)  
office: INB3108



- British computer scientist (born in Mexico)
- PhD from the University of Edinburgh in 2009
- Areas of interest:
  - Conversational AI
  - Robot Learning

# Recommendations

- Record your attendance in every session:
  - online <https://attendance.lincoln.ac.uk>
  - paper-based register (when online system is off)
- Check Blackboard and your email frequently for updates and notifications
- Mute your mobile phone during lectures
- If you want to ask/say something, raise your hand or just interrupt your lecturer (that's okay)

# Lectures and Workshops

- Lectures: Tuesday 11am-12hrs in NDH0020
- Workshops: Wednesday 11am-1pm in INB2102

← Week 1		→ 25-26	Single	Full	Table (Horizontal)	Export/Print
11am				12pm		
Monday (29th Sep 25)						
Tuesday (30th Sep 25)	CMP9794 NDH0020 11am - 12pm Weeks 1-6, 8-12, 16 Heriberto Cuayahuitl Portilla LECTURE					
Wednesday (1st Oct 25)	CMP9794 INB2102 11am - 1pm Weeks 1-6, 8-12, 16 Heriberto Cuayahuitl Portilla WORKS					
Thursday (2nd Oct 25)						
Friday (3rd Oct 25)	CMP9794M Advanced AI					

# Main Topics in this Module

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

# Agenda

Week	Commencing on	Topic	Delivered by
1	29/09/2024	Introduction to module	Heriberto Cuayahuitl
2	06/10/2024	Bayesian networks w/exact inference	Heriberto Cuayahuitl
3	13/10/2024	Structure learning	Heriberto Cuayahuitl
4	20/10/2024	Gaussian Bayesian networks	Heriberto Cuayahuitl
5	27/10/2024	Gaussian processes I—Exact Inference	Heriberto Cuayahuitl
6	03/11/2024	Gaussian processes II—Approx. Inference	Heriberto Cuayahuitl
7	10/11/2024	Reading week (no lectures/workshops)	
8	17/11/2024	Bayesian networks w/approx. inference	Heriberto Cuayahuitl
9	24/11/2024	Probabilistic reasoning over time	Heriberto Cuayahuitl
10	01/12/2024	Complex decision making	Heriberto Cuayahuitl
11	08/12/2024	Markov decision processes	Heriberto Cuayahuitl
12	15/12/2024	Reinforcement learning	Heriberto Cuayahuitl
13-15		Christmas break	
16	12/01/2025	Ethics and Agentic AI	Heriberto Cuayahuitl

# Learning Objectives

LO1. 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

LO2. Design and develop software algorithms for solving complex (non-trivial) AI problems in an application domain of interest.

# Assessments\*

- Assignment (50%):
  - Bayesian Networks (BNs)
  - Gaussian BNs
  - Gaussian Processes
- In-class test (50%):
  - Mock-test (last workshop)
  - Test (see Hand-in deadlines and your timetable)



\*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

# Thinking Humanly: Cognitive Modelling

- We must have a way of determining how humans think:
  - Introspection (catching our own thoughts as they happen)
  - Psychological experiments (observing a person in action)
  - Brain imaging (observing brain in action)
- Cognitive science brings multiple fields together to try to construct precise and testable theories of the workings of the human mind.

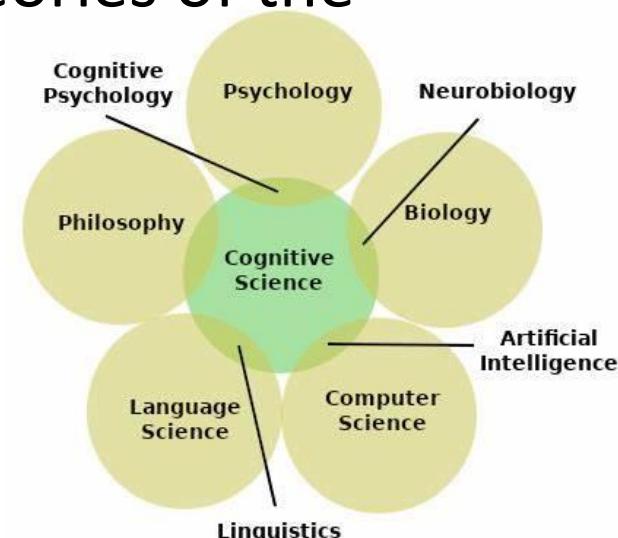
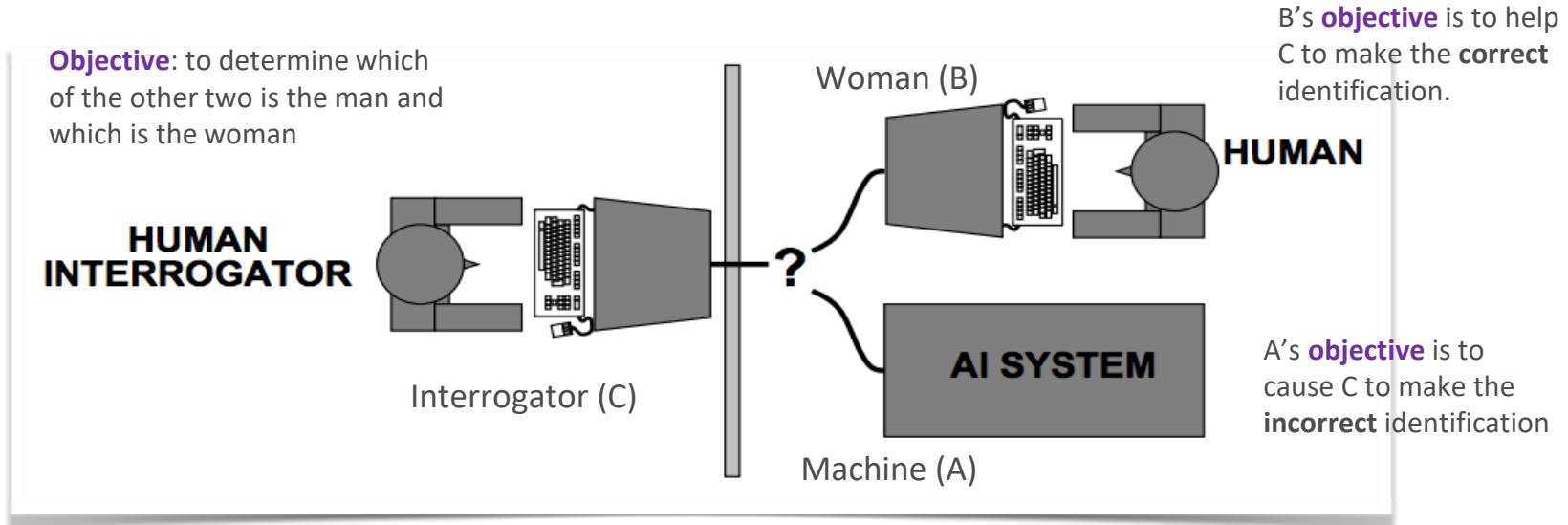


Image from: [Psychological Theory Definition: An Overview \(mantracare.org\)](https://mantracare.org/)

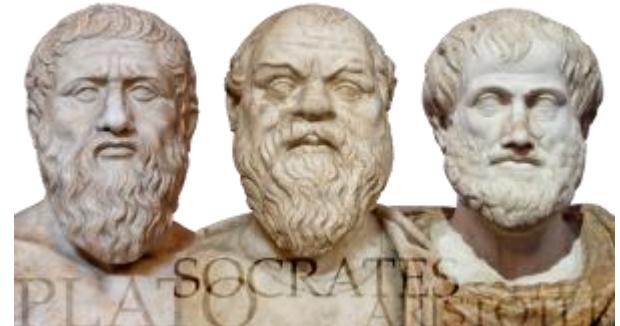
# Acting Humanly: The Turing Test (1950)

- Turing (1950) “Computing machinery and intelligence”
- Can machines behave intelligently?
- Summary of Alan Turing’s paper (1950)



# 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 (e.g., Prolog).
- However, it is not easy to represent informal knowledge using 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 to the AI.
- **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



John McCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



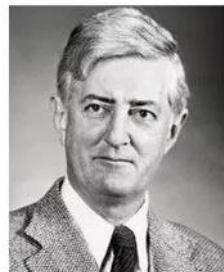
Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

Picture from  
<https://www.scienceabc.com>

# Pillars of AI

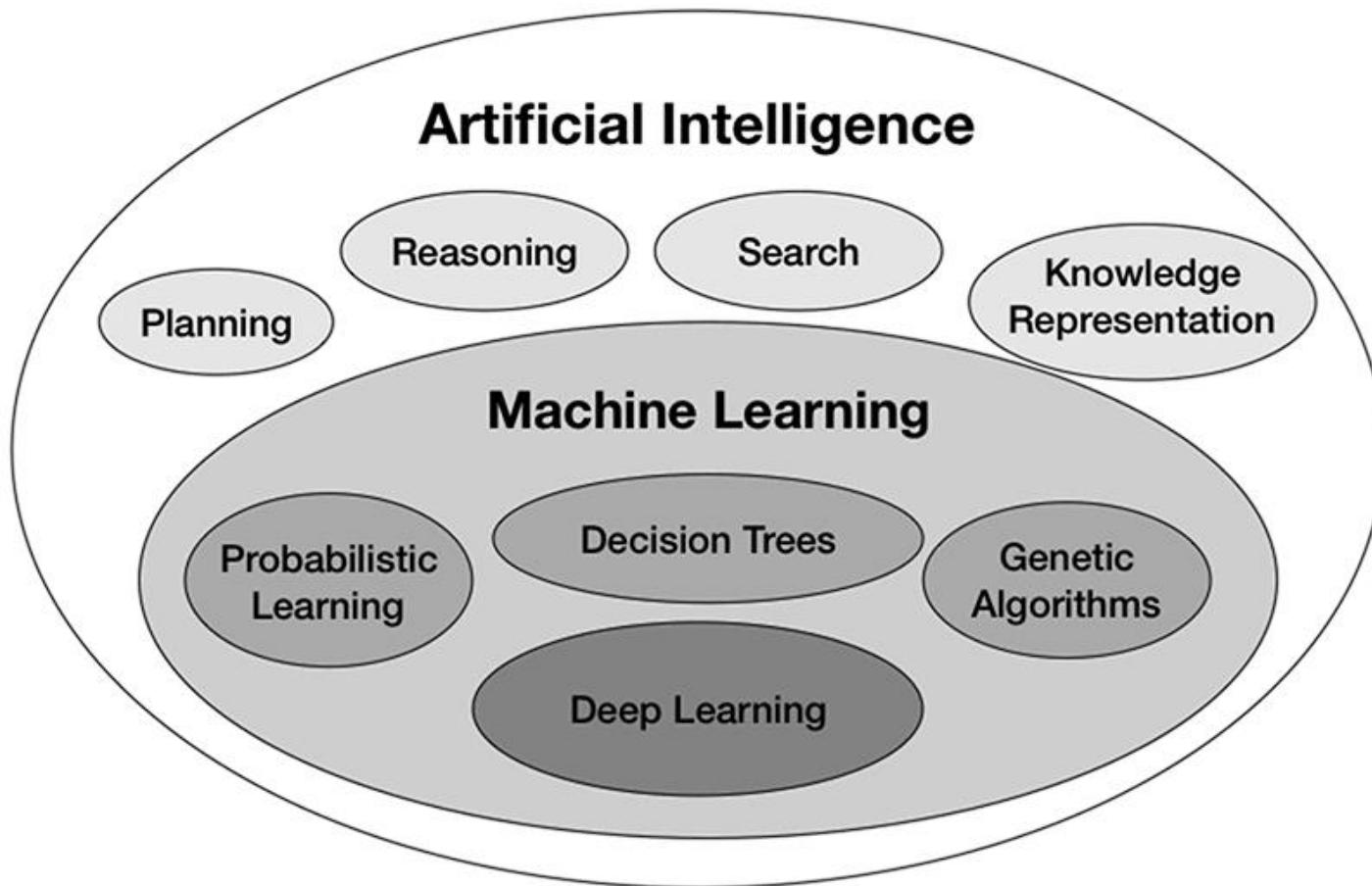


Image from  
<https://behavioralscientist.org>

# Major Areas in AI

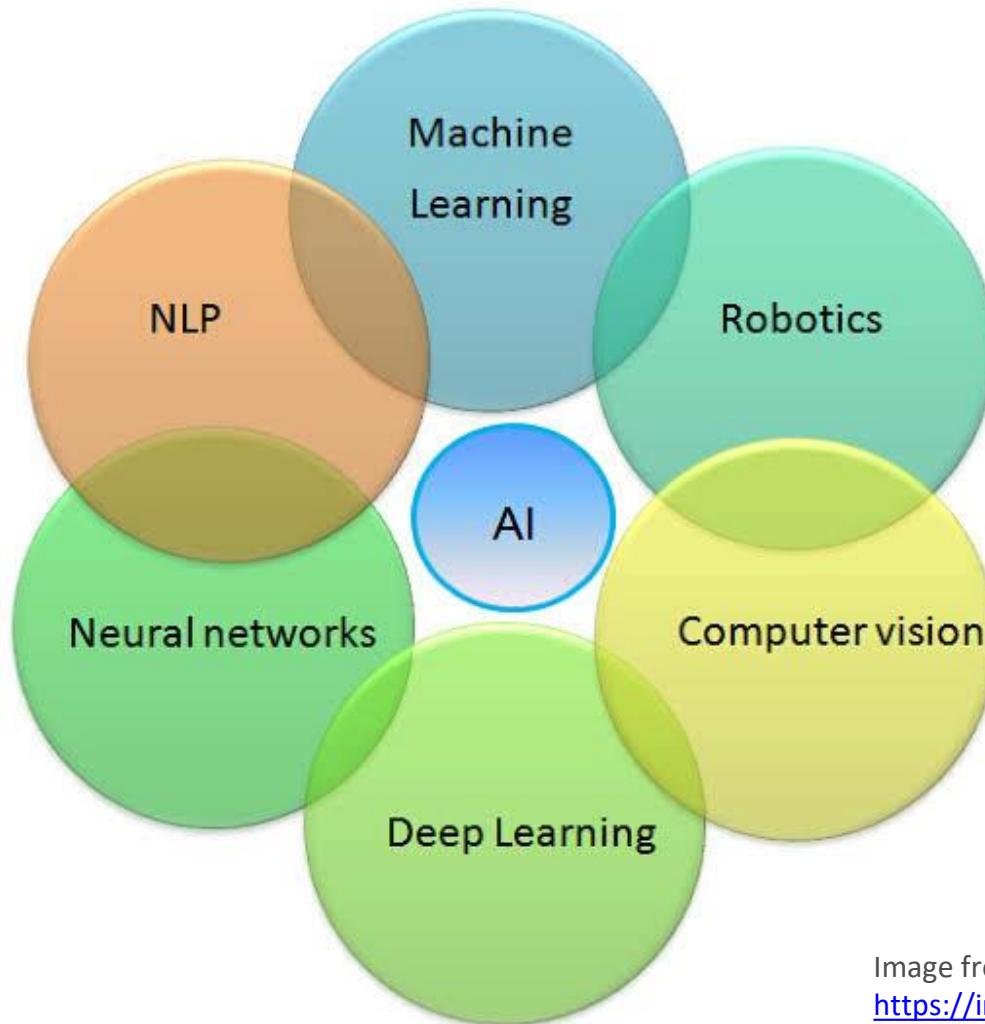


Image from  
<https://infohubns.blogspot.com/2020/07/research-areas-of-ai.html>

# History of AI in a Nutshell – part 1

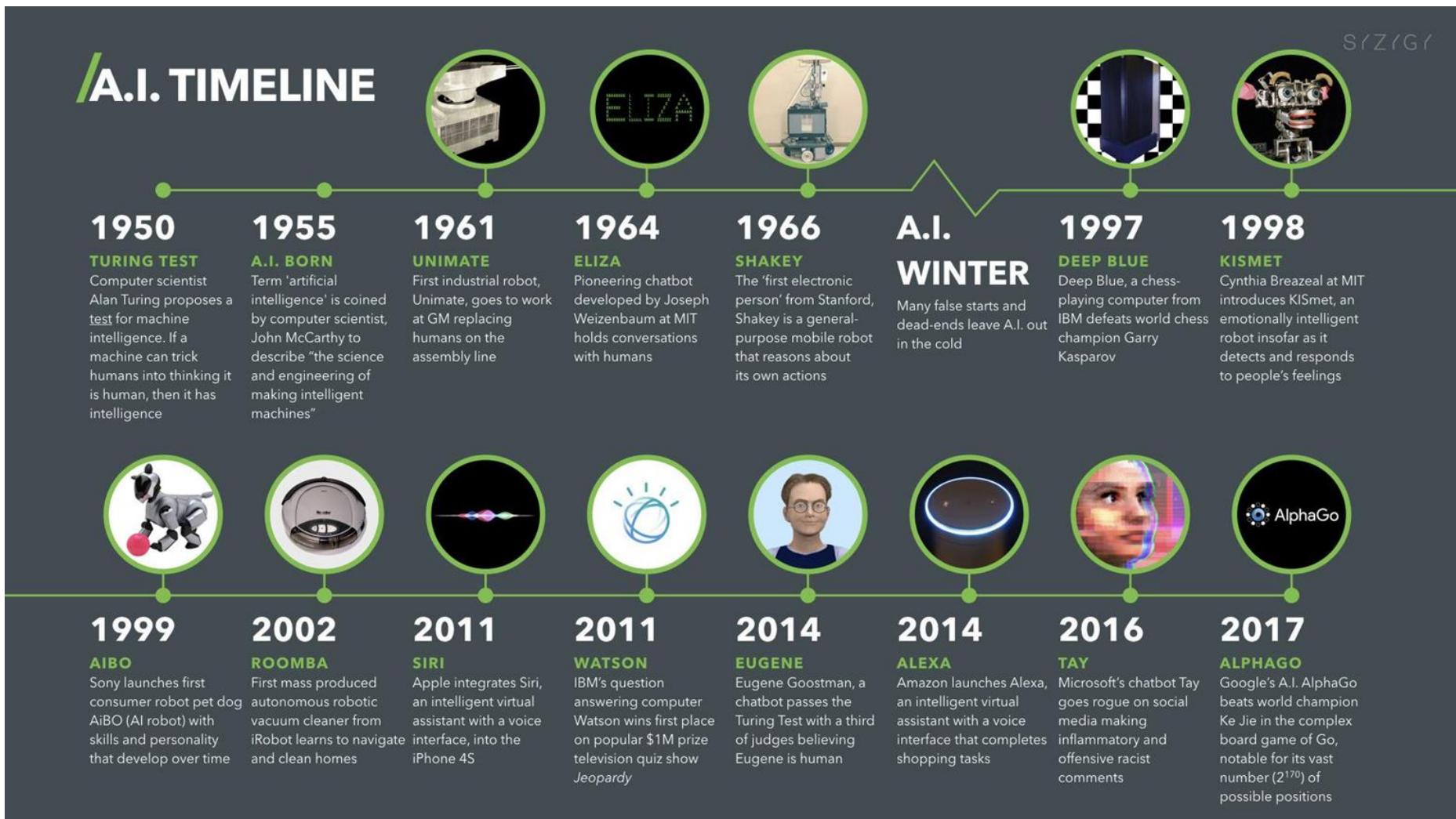
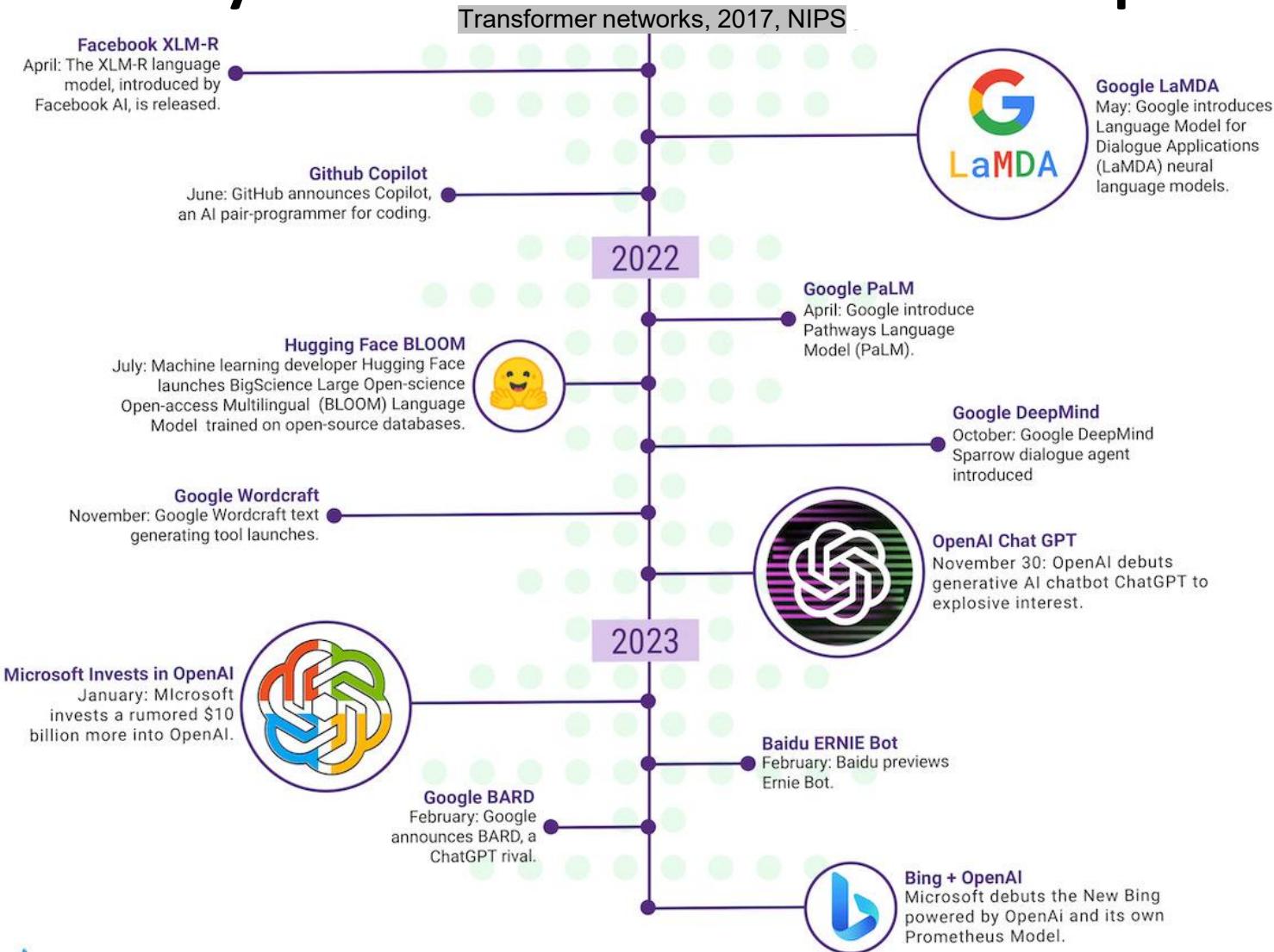
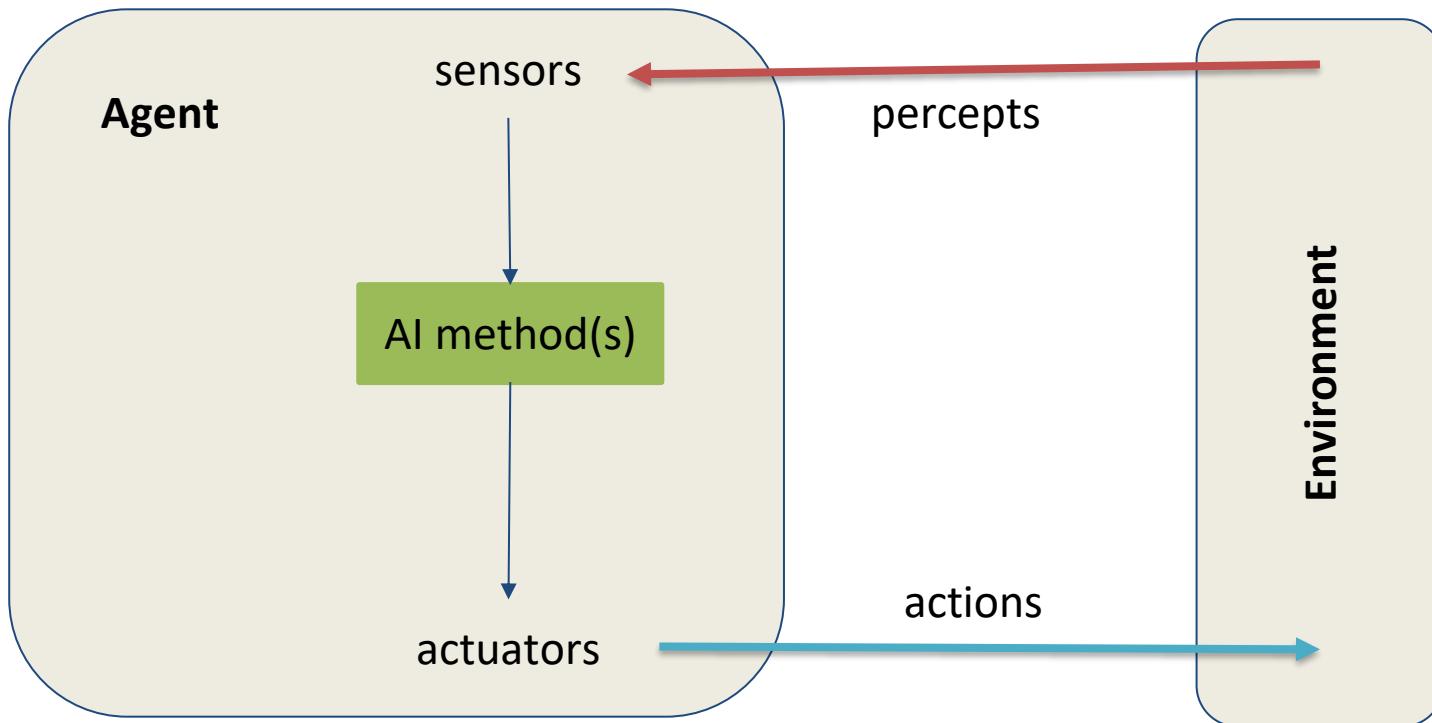


Image from <https://digitalwellbeing.org/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/>

# History of AI in a Nutshell – part 2



# AI 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 }

behaviour

More details of those properties in Russell and Norvig Chapter 2.

Hardest case (in red)

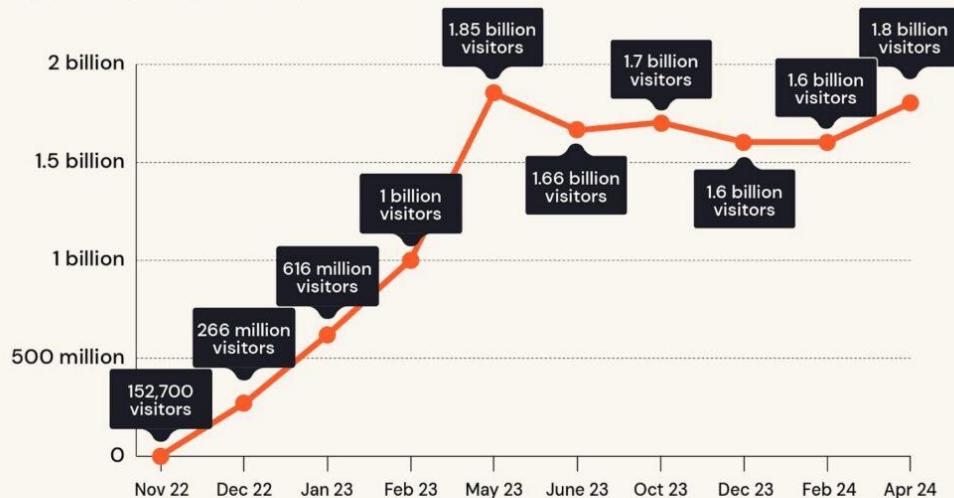
Example1: Autonomous cars driving on unfamiliar roads.

Example2: Robots cooking food in a changing/dynamic kitchen.

# Some Recent Developments in AI

## CHATGPT STATISTICS

### Change in ChatGPT website visitors since launch



Read the full report at [tooltester.com/en/blog/chatgpt-statistics](https://tooltester.com/en/blog/chatgpt-statistics)

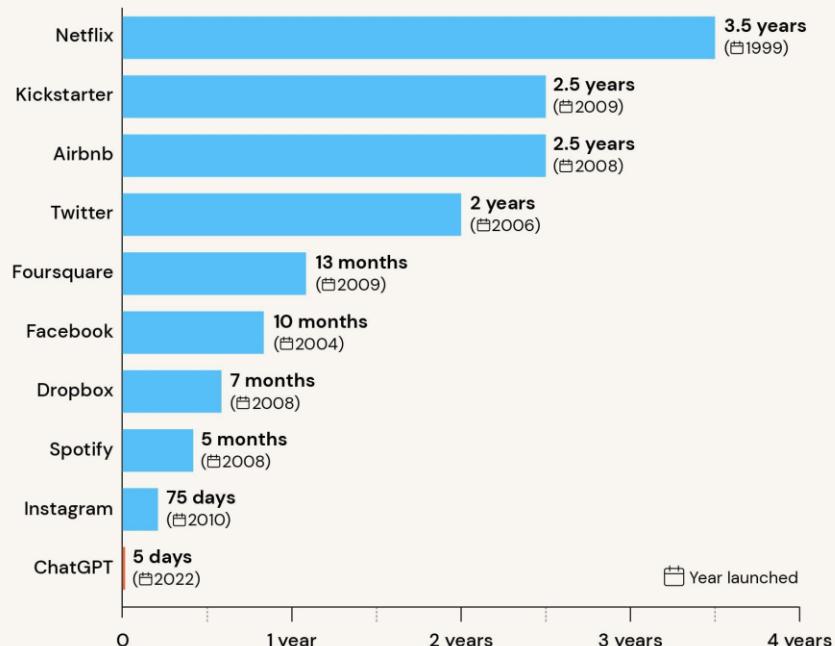
tooltester

ChatGPT received around 1.8 billion visitors in April 2024, with ~100 million active users.

OpenAI's ChatGPT, a 175B parameter neural net, learns to chat on 570GB of text in 34 days, 2022

## CHATGPT STATISTICS

### Time to reach 1 million users



Read the full report at [tooltester.com/en/blog/chatgpt-statistics](https://tooltester.com/en/blog/chatgpt-statistics)

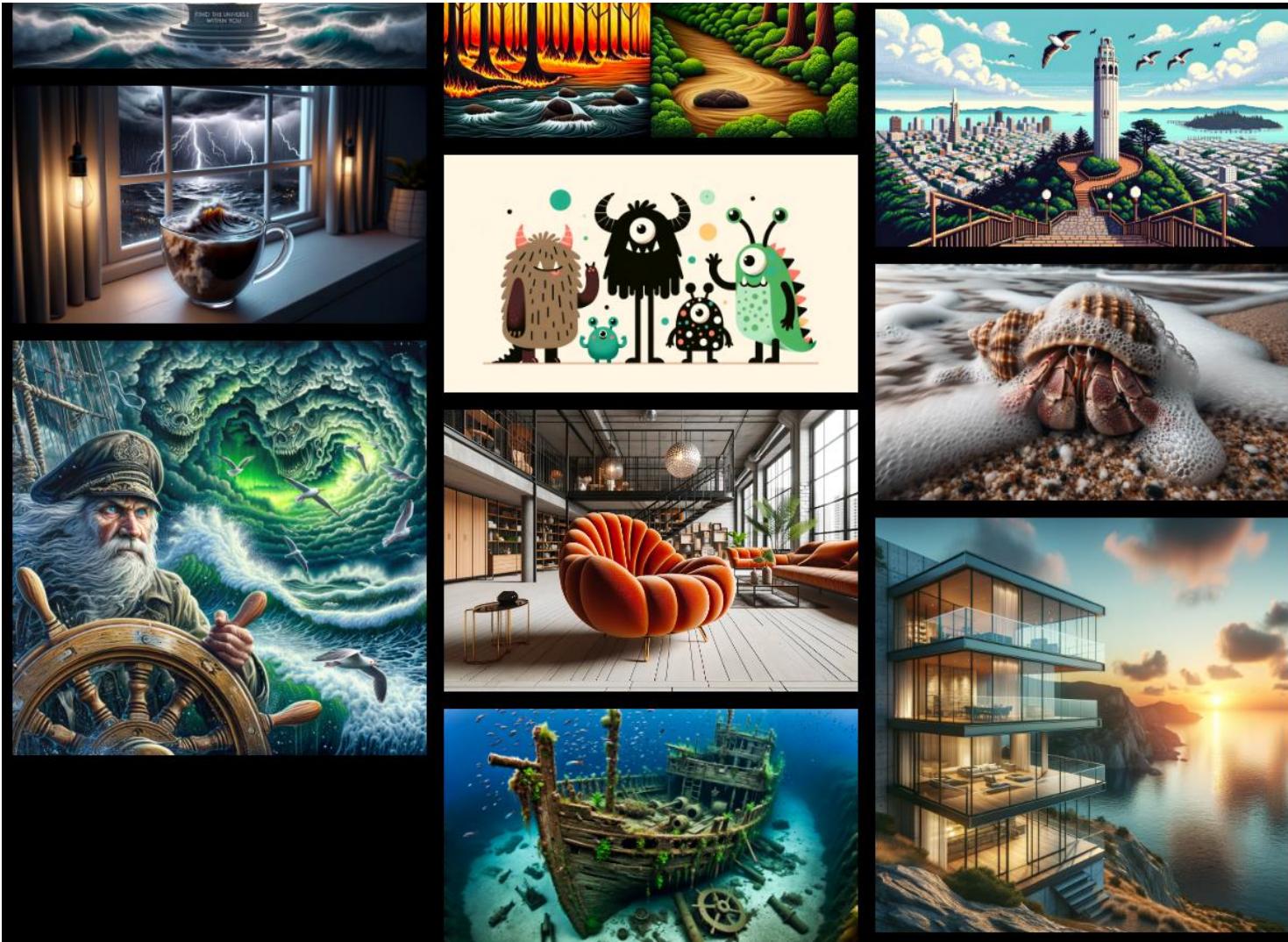
tooltester

# Some Recent Developments in AI



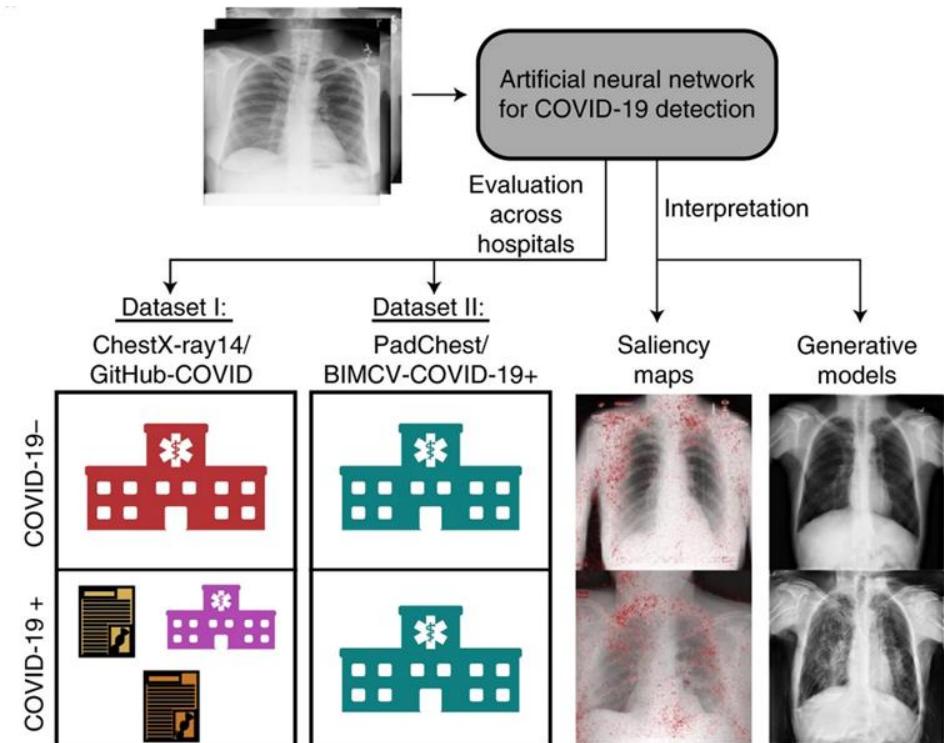
Grey boxes are ignore regions  
(as defined in the original KITTI tracking ground truth)

# Some Recent Developments in AI



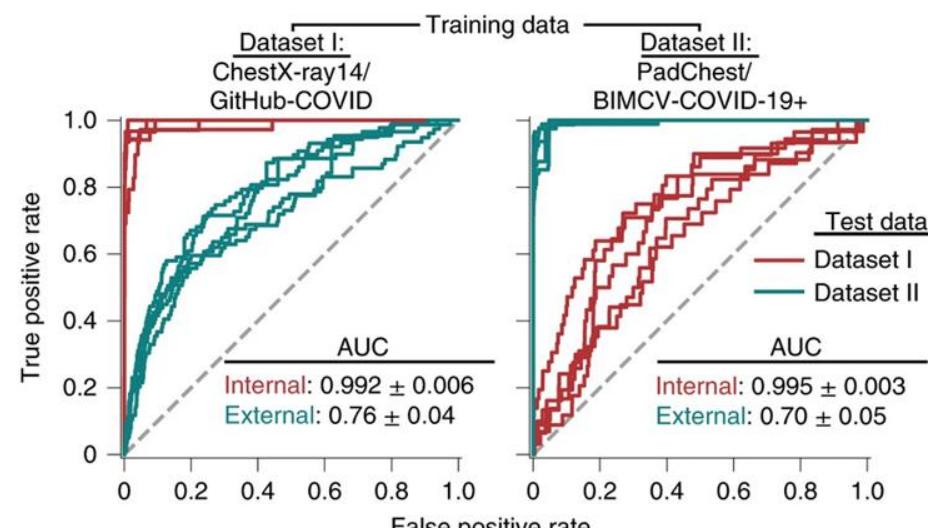
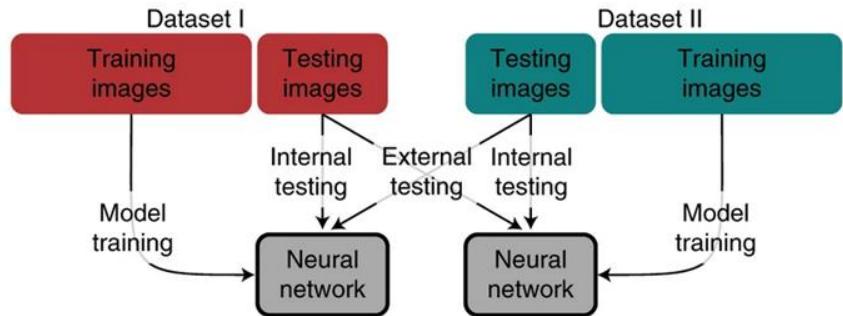
OpenAI's DALL.E 3 is able to generate images that exactly adheres to text provided, 2023

# Some Recent Developments in AI



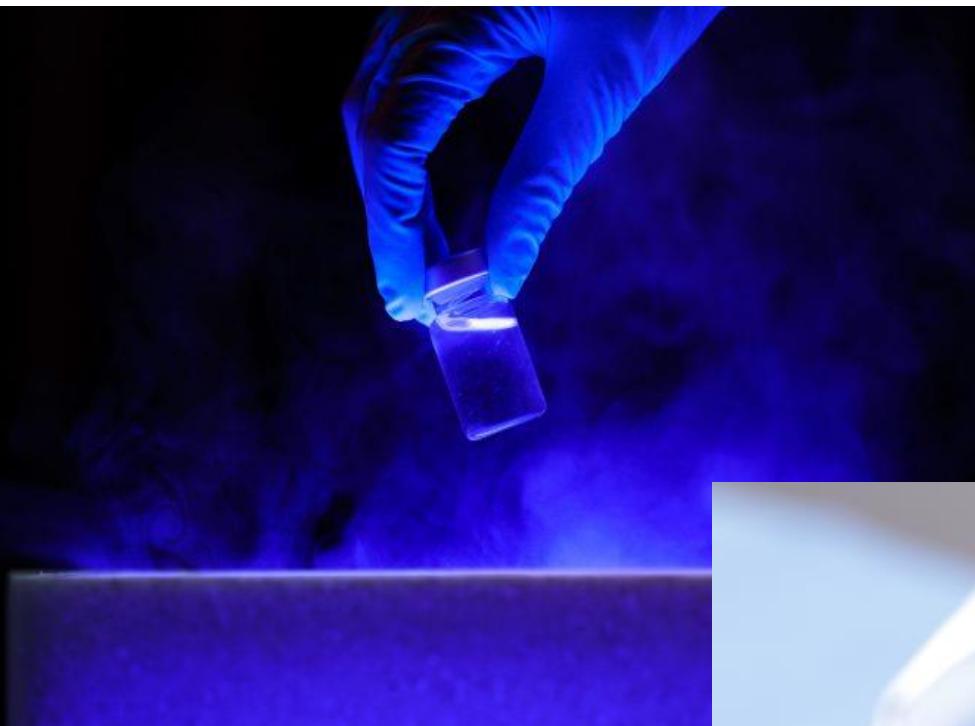
**b**

	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

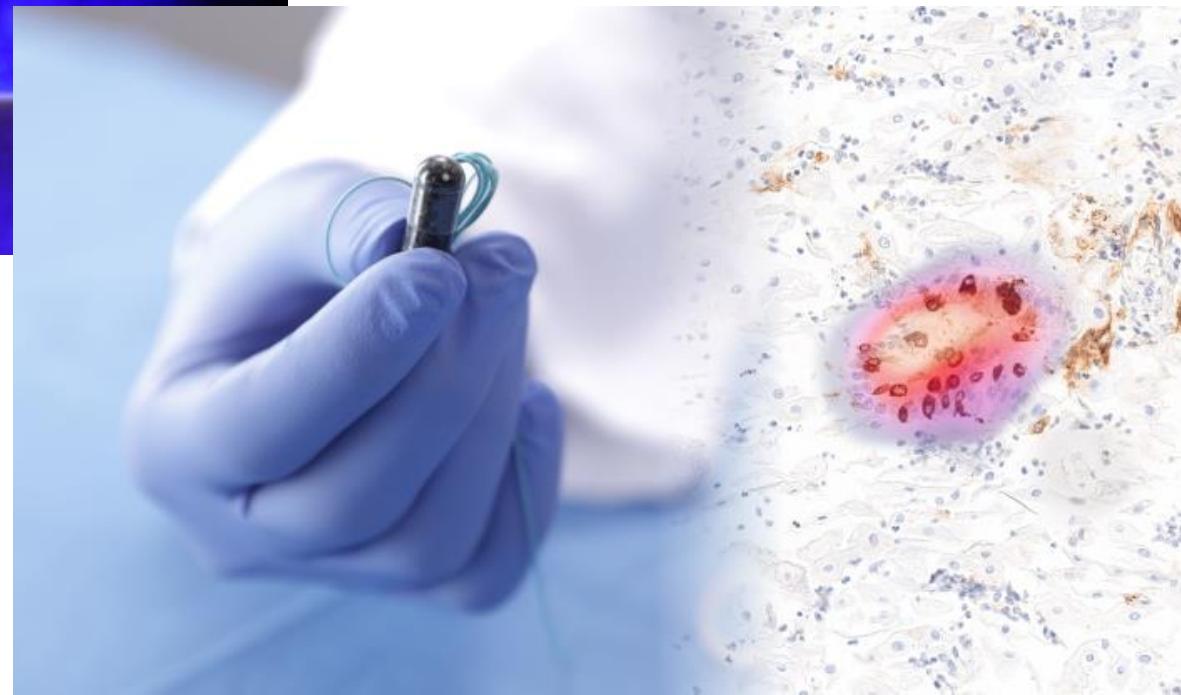


[Medical-imaging AI system learns to detect Covid-19 from chest radiographs, 2021](#)

# Some Recent Developments in AI

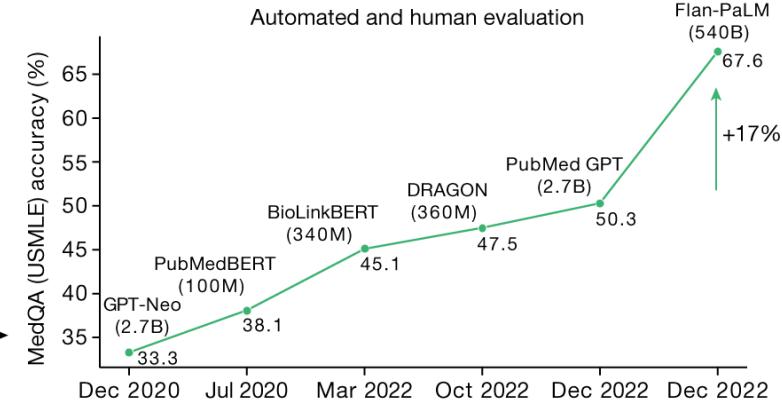
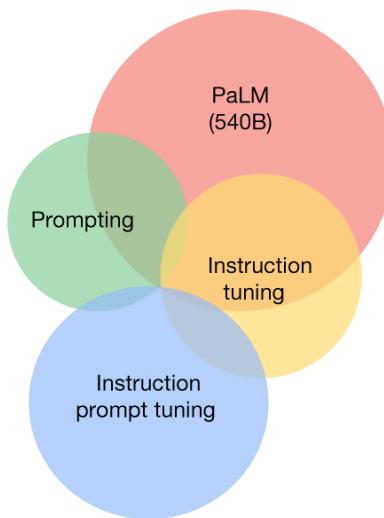
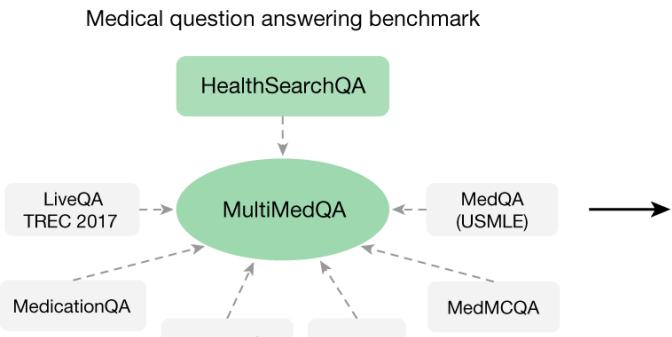


AI system can find COVID vaccine candidates within seconds, 2021



AI for early Cancer detection, 2021

# Some Recent Developments in AI



**Q:** How long does it take for newborn jaundice to go away?

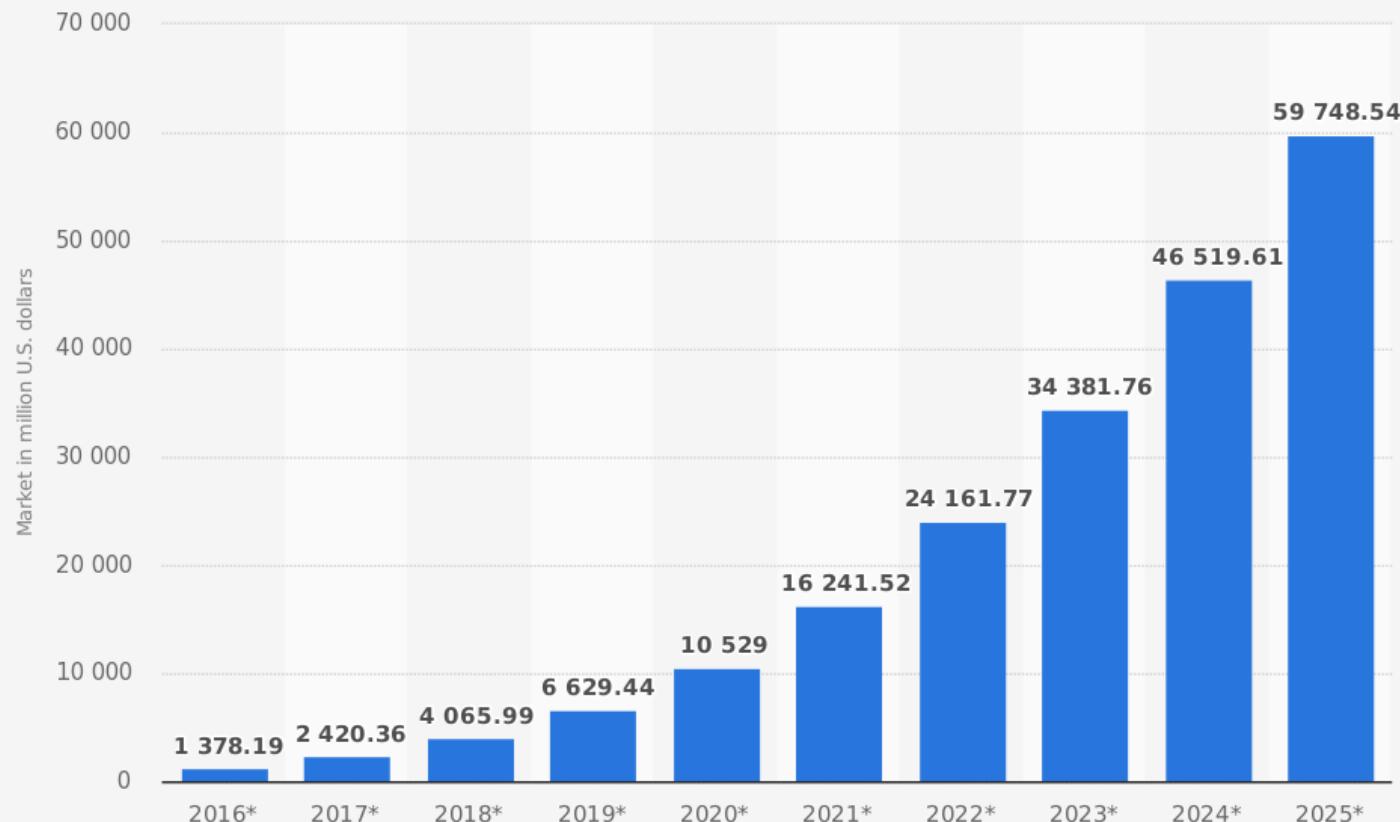
**Med-PaLM:** Newborn jaundice is when a newborn baby's skin and eyes look yellow. It is caused by a buildup of a substance called bilirubin in the baby's blood. Bilirubin is a byproduct of the normal process of red blood cells breaking down. Newborn jaundice is common and typically harmless, but it can be a sign of a more serious condition. The jaundice typically goes away on its own within a few weeks. In some cases, it may last longer or require treatment. The time it takes for the jaundice to go away can vary depending on the cause and the severity of the jaundice. If the jaundice is severe or lasts longer than a few weeks, the doctor may recommend testing or treatment to determine the cause and prevent complications.

[Large language models encode clinical knowledge, 2023](#)

Med-PaLM performs encouragingly on consumer medical question answering

# AI: Economic Impact

**Revenues from the artificial intelligence (AI) market worldwide, from 2016 to 2025 (in million U.S. dollars)**



Source  
Tractica  
© Statista 2017

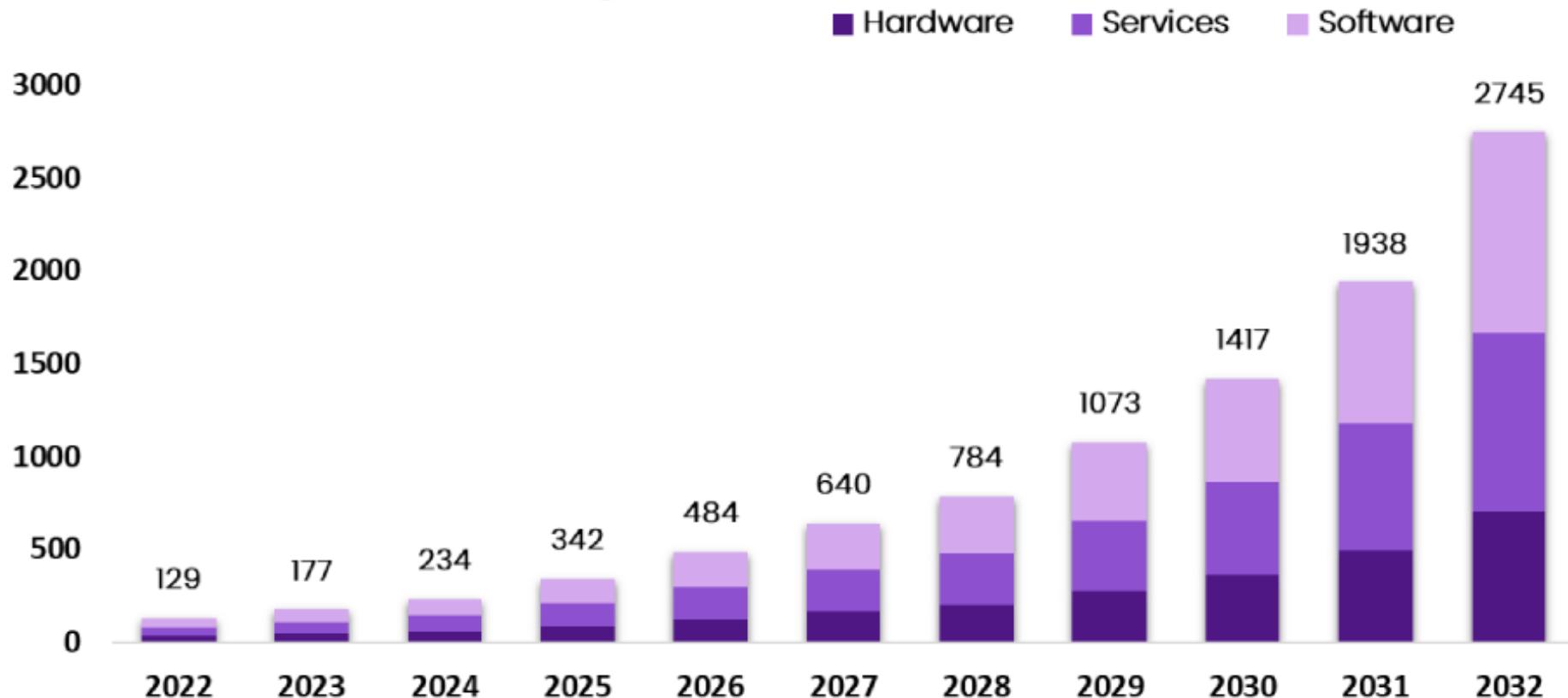
Additional Information:  
Worldwide; 2017

statista

[source](#)

# AI: Economic Impact

## Global Artificial Intelligence Market



The Market will Grow  
At the CAGR of:

36.8%

The forecasted market  
size for 2032 in USD:

\$2745B

 market.us  
ONE STOP SHOP FOR THE REPORTS

[source](#)

# Ethics in AI

## Fairness

All AI systems that process social or demographic data pertaining to features of human subjects must be designed to meet a minimum threshold of discriminatory non-harm. This entails that the datasets they use be equitable; that their model architectures only include reasonable features, processes, and analytical structures; that they do not have inequitable impact; and that they are implemented in an unbiased way.



F

## Accountability

**Accountability By Design:** All AI systems must be designed to facilitate end-to-end answerability and auditability. This requires both responsible humans-in-the-loop across the entire design and implementation chain and activity monitoring protocols that enable end-to-end oversight and review.



## Sustainability

Designers and users of AI systems must remain aware that these technologies have transformative effects on individuals and society. They must thereby proceed with a continuous sensitivity to real-world impacts. They must also keep in mind that the technical sustainability of these systems depends on their safety: their accuracy, reliability, security, and robustness.



S

## Transparency

Designers and implementers of AI systems must be able (1) to explain to affected stakeholders in everyday language how and why a model performed the way it did in a specific context and (2) to justify the ethical permissibility, the discriminatory non-harm, and the public trustworthiness both of its outcome and of the processes behind its design and use.



[Understanding Artificial Intelligence ethics and safety](#)

[Data ethics and AI guidance landscape](#)

# (Discrete) Random Variables

- The events we are interested in have a set of possible values. Examples of random variables:

Coin toss (C): {heads, tails}

Roll a die (D): {1, 2, 3, 4, 5, 6}

Weather (W): {snowy, sunny, rainy, foggy}

Measles (M): {true, false}

- For each event, a random variable takes a value from the associated set. Then we have:

$P(C = \text{tails})$       or       $P(\text{tails})$

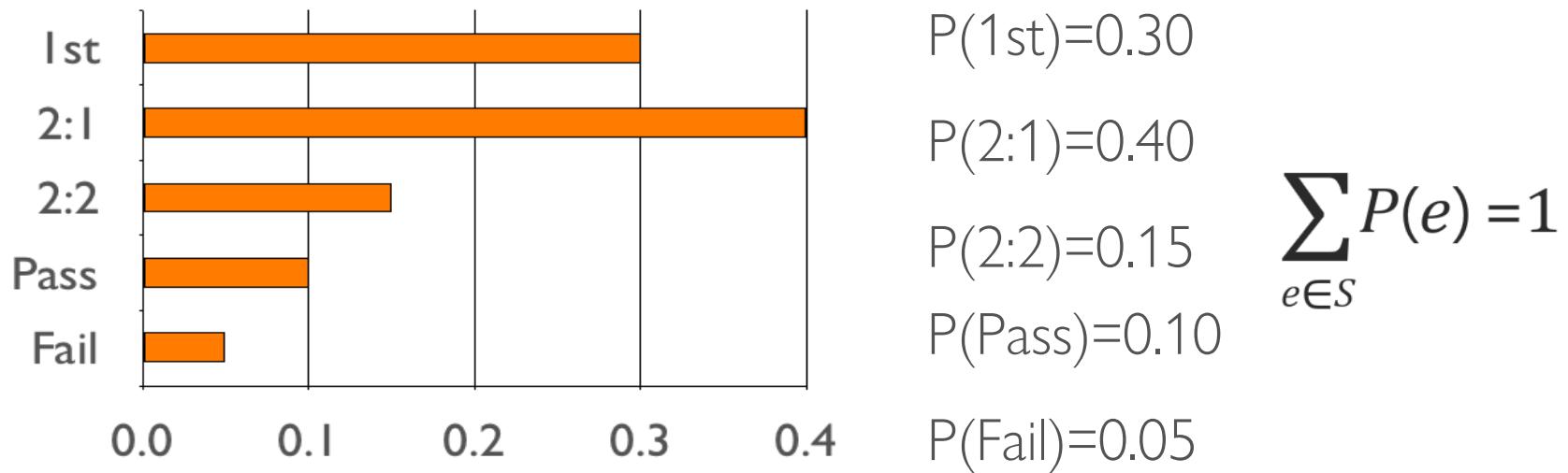
$P(D = 1)$       or       $P(1)$

$P(W = \text{sunny})$       or       $P(\text{sunny})$

$P(M = \text{true})$       or       $P(\text{measles})$

# Discrete Probability Distribution

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



- Probability distributions can be estimated from data

# Joint Probabilities

Joint probabilities represent the whole joint probability distribution

<b>S1=1st</b>	<b>S2=1st</b>	<b>Probability</b>
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.

	$S2 = \text{1st}$	$\neg(S2 = \text{1st})$
$S1 = \text{1st}$	0.2	0.1
$\neg(S1 = \text{1st})$	0.1	0.6

# Marginal Probabilities

	$S2 = 1st$	$\neg(S2 = 1st)$
$S1=1st$	0.2	0.1
$\neg(S1=1st)$	0.1	0.6

**Marginalisation** => summing up the probabilities of the other variables—i.e.  $P(X) = \sum_y P(X, Y = y)$

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

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

# Marginal Probabilities

	<b>S2 =1<sup>st</sup></b>	<b>¬(S2 =1<sup>st</sup>)</b>
<b>S1=1<sup>st</sup></b>	0.2	0.1
<b>¬(S1=1<sup>st</sup>)</b>	0.1	0.6

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

$$\begin{aligned} P(\neg(S1=1\text{st})) &= P(\neg(S1=1\text{st}) \wedge S2=1\text{st}) + P(\neg(S1=1\text{st}) \wedge \neg(S2=1\text{st})) \\ &= 0.1 + 0.6 = 0.7 \end{aligned}$$

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

# Marginal Probabilities

	<b>S2 =1<sup>st</sup></b>	<b>¬(S2 =1<sup>st</sup>)</b>
<b>S1=1<sup>st</sup></b>	0.2	0.1
<b>¬(S1=1<sup>st</sup>)</b>	0.1	0.6

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

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

Note that  $P(S2=1st) + P(\neg(S2=1st)) = 0.3 + 0.7 = 1$

# 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:  
 $P(A | B) = P(A \wedge B) / P(B)$ , where  $P(B) \neq 0$

Product Rule

$$P(A \wedge B) = P(A | B) * P(B)$$

# Chain Rule for Probabilities

- $P(A \wedge B) = P(A | B) * P(B)$
- $$\begin{aligned}P(A \wedge B \wedge C) &= P(A | B \wedge C) * P(B | C) * P(C) \\&= P(C) * P(B | C) * P(A | B \wedge C) \\&= P(A) * P(B | A) * P(C | B \wedge A) \\&= P(C | B \wedge A) * P(B | A) * P(A)\end{aligned}$$
- $$P(A_1 \wedge A_2 \wedge \dots \wedge A_n) = P(A_n | A_{n-1} \wedge \dots \wedge A_1) * P(A_{n-1} | A_{n-2} \wedge \dots \wedge A_1) * \dots * P(A_2 | A_1) * P(A_1)$$

# Conditional Probability

	$S2=1st$	$\neg(S2=1st)$
$S1=1st$	0.2	0.1
$\neg(S1=1st)$	0.1	0.6

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

$$P(A | B) = P(A \wedge B) / P(B)$$

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

	<b>S2=1st</b>	<b><math>\neg(S2=1st)</math></b>
<b>S1=1st</b>	0.2	0.1
<b><math>\neg(S1=1st)</math></b>	0.1	0.6

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

$$P(A | B) = P(A \wedge B) / P(B)$$

Substituting A with  $\neg(S2=1st)$

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

	<b>S2=1st</b>	<b><math>\neg(S2=1^{st})</math></b>
<b>S1=1st</b>	0.2	0.1
<b><math>\neg(S1=1^{st})</math></b>	0.1	0.6

$$\begin{aligned} \text{Since } P(S2=1\text{st} | S1=1\text{st}) &= P(S2=1\text{st}, S1=1\text{st}) / P(S1=1\text{st}) \\ &= 0.2 / 0.3 = 0.6666666 \end{aligned}$$

$$\begin{aligned} \text{and } P(\neg(S2=1\text{st}) | S1=1\text{st}) &= P(\neg(S2=1\text{st}), S1=1\text{st}) / P(S1=1\text{st}) \\ &= 0.1 / 0.3 = 0.3333333 \end{aligned}$$

$$\text{Then } P(S2=1\text{st} | S1=1\text{st}) + P(\neg(S2=1\text{st}) | S1=1\text{st}) = 1$$

# More on Joint Probabilities

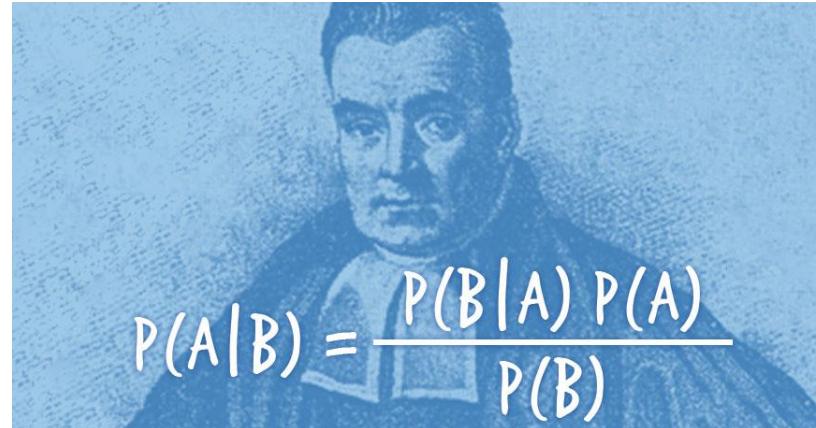
Given a joint probability table, we have all the information we need about the domain. We can calculate the probability of any logical formula.

$$P(S1=\text{first} \vee S2=\text{first}) = 0.2 + 0.1 + 0.1 = 0.4$$

	$S2 = \text{1st}$	$\neg(S2 = \text{1st})$
$S1 = \text{1st}$	0.2	0.1
$\neg(S1 = \text{1st})$	0.1	0.6

# Bayes Rule

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

A blue-toned portrait engraving of Thomas Bayes, an 18th-century English statistician and Presbyterian minister. He is shown from the chest up, wearing a dark coat over a white cravat and a patterned waistcoat.
$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

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  $P(Y)$ .

\* Accurately estimating  $P(X | Y)$  is expensive--  
see Appendix 1.

# 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  $X^{new} = \langle X_1 \dots X_n \rangle$ , 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

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

# Naïve Bayes for Discrete Inputs

- Estimate two sets of parameters:

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

$$\pi_k = P(Y = y_k)$$

- According to

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

$$\pi_k = P(Y = y_k) = \frac{\#D\{Y=y_k\}}{|D|}$$

where  $J$  is the number of unique values in  $X_i$ ,  $l$  (e.g.,  $l = 1$ ) avoids zero estimates, and  $|D|$  is the num. of elements in the training set.

# Example: Training Data

Day	<i>Outlook</i>	<i>Temperature</i>	<i>Humidity</i>	<i>Wind</i>	<i>PlayTennis</i>
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

$$P(\text{PlayTennis} = \text{yes}) = \frac{9}{14} = 0.643$$

$$P(\text{PlayTennis} = \text{no}) = \frac{5}{14} = 0.357$$

$$P(\text{Outlook} = \text{sunny} | \text{PlayTennis} = \text{yes}) = 2/9 = 0.222$$

$$P(\text{Outlook} = \text{sunny} | \text{PlayTennis} = \text{no}) = 3/5 = 0.60$$

$$P(\text{Outlook} = \text{overcast} | \text{PlayTennis} = \text{yes}) = 4/9 = 0.444$$

$$P(\text{Outlook} = \text{overcast} | \text{PlayTennis} = \text{no}) = 0/5 = 0.0$$

$$P(\text{Outlook} = \text{rain} | \text{PlayTennis} = \text{yes}) = 3/9 = 0.333$$

$$P(\text{Outlook} = \text{rain} | \text{PlayTennis} = \text{no}) = 2/5 = 0.4$$

...

We want to avoid  
zero probabilities

# Example: Estimated Parameters

$$P(\text{Temperature} = \text{hot} | \text{PlayTennis} = \text{yes}) = 2/9 = 0.222$$

$$P(\text{Temperature} = \text{hot} | \text{PlayTennis} = \text{no}) = 2/5 = 0.4$$

$$P(\text{Temperature} = \text{mild} | \text{PlayTennis} = \text{yes}) = 4/9 = 0.444$$

$$P(\text{Temperature} = \text{mild} | \text{PlayTennis} = \text{no}) = 2/5 = 0.4$$

$$P(\text{Temperature} = \text{cool} | \text{PlayTennis} = \text{yes}) = 3/9 = 0.333$$

$$P(\text{Temperature} = \text{cool} | \text{PlayTennis} = \text{no}) = 1/5 = 0.2$$

$$P(\text{Humidity} = \text{high} | \text{PlayTennis} = \text{yes}) = 3/9 = 0.333$$

$$P(\text{Humidity} = \text{high} | \text{PlayTennis} = \text{no}) = 4/5 = 0.8$$

$$P(\text{Humidity} = \text{normal} | \text{PlayTennis} = \text{yes}) = 6/9 = 0.666$$

$$P(\text{Humidity} = \text{normal} | \text{PlayTennis} = \text{no}) = 1/5 = 0.2$$

$$P(\text{Wind} = \text{strong} | \text{PlayTennis} = \text{yes}) = 3/9 = 0.333$$

$$P(\text{Wind} = \text{strong} | \text{PlayTennis} = \text{no}) = 3/5 = 0.6$$

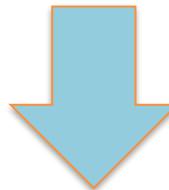
$$P(\text{Wind} = \text{weak} | \text{PlayTennis} = \text{yes}) = 6/9 = 0.666$$

$$P(\text{Wind} = \text{weak} | \text{PlayTennis} = \text{no}) = 2/5 = 0.4$$

# Example: Classifying a New Instance

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

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



$$P(\text{PlayTennis} = \text{yes}|\text{evidence}) = \frac{0.0053}{0.0053 + 0.0206} = 0.205$$
$$P(\text{PlayTennis} = \text{no}|\text{evidence}) = \frac{0.0206}{0.0053 + 0.0206} = 0.795$$

Outlook=sunny  
Temperature=cool  
Humidity=high  
Wind=strong

The probabilities above can be calculated in log space, see Appendix 2.

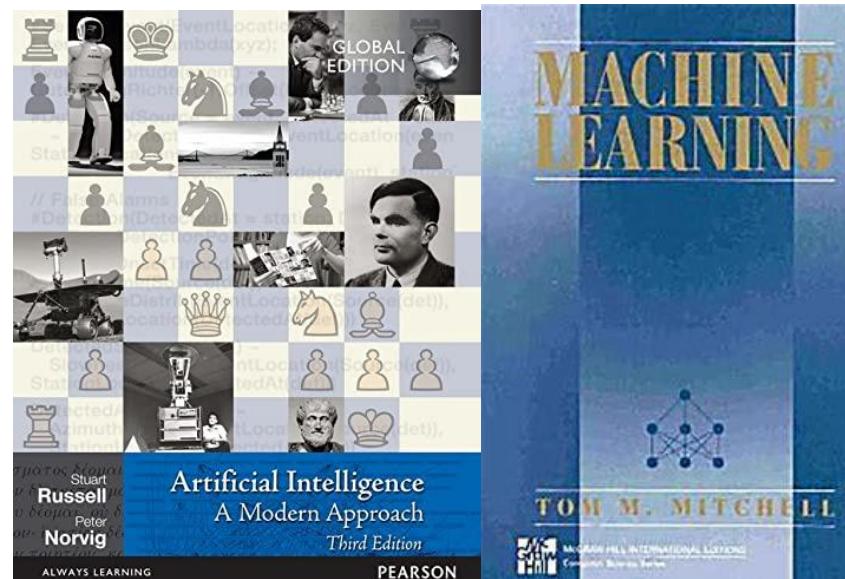
# Today

- 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, 13](#)

Mitchell, T. 2017; 2<sup>nd</sup> Ed. [Chapter 3](#)



## **This and Next Week**

**Workshop (tomorrow):**

Exercises on probability theory

Python programs for Bayesian classifiers

**Lecture (next week):**

Bayesian Networks with Exact Inference

Readings: Russell & Norvig 2016. Chapters 13-14.4

Questions now and during the workshop.

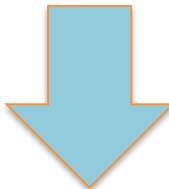
# Appendix 1: Difficulty in Unbiased Bayesian Classifiers

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

# Appendix 2: Example with Log Probabilities

$$\begin{aligned} & P(\text{yes})P(\text{sunny}|\text{yes})P(\text{cool}|\text{yes})P(\text{high}|\text{yes})P(\text{strong}|\text{yes}) \\ &= (-0.442) + (-1.504) + (-1.098) + (-1.098) + (-1.098) \\ &= -5.242 \end{aligned}$$

$$\begin{aligned} & P(\text{no})P(\text{sunny}|\text{no})P(\text{cool}|\text{no})P(\text{high}|\text{no})P(\text{strong}|\text{no}) \\ &= (-1.029) + (-0.51) + (-1.609) + (-0.223) + (-0.51) \\ &= -3.883 \end{aligned}$$



$$\begin{aligned} P(\text{PlayTennis} = \text{yes}|\text{evidence}) &= \frac{e^{-5.242}}{e^{-5.242} + e^{-3.883}} = 0.205 \\ P(\text{PlayTennis} = \text{no}|\text{evidence}) &= \frac{e^{-3.883}}{e^{-5.242} + e^{-3.883}} = 0.795 \end{aligned}$$