國立臺灣科技大學
NATIONAL TAIWAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

電資產業導論
Introduction to Elect-Computer **Industries**

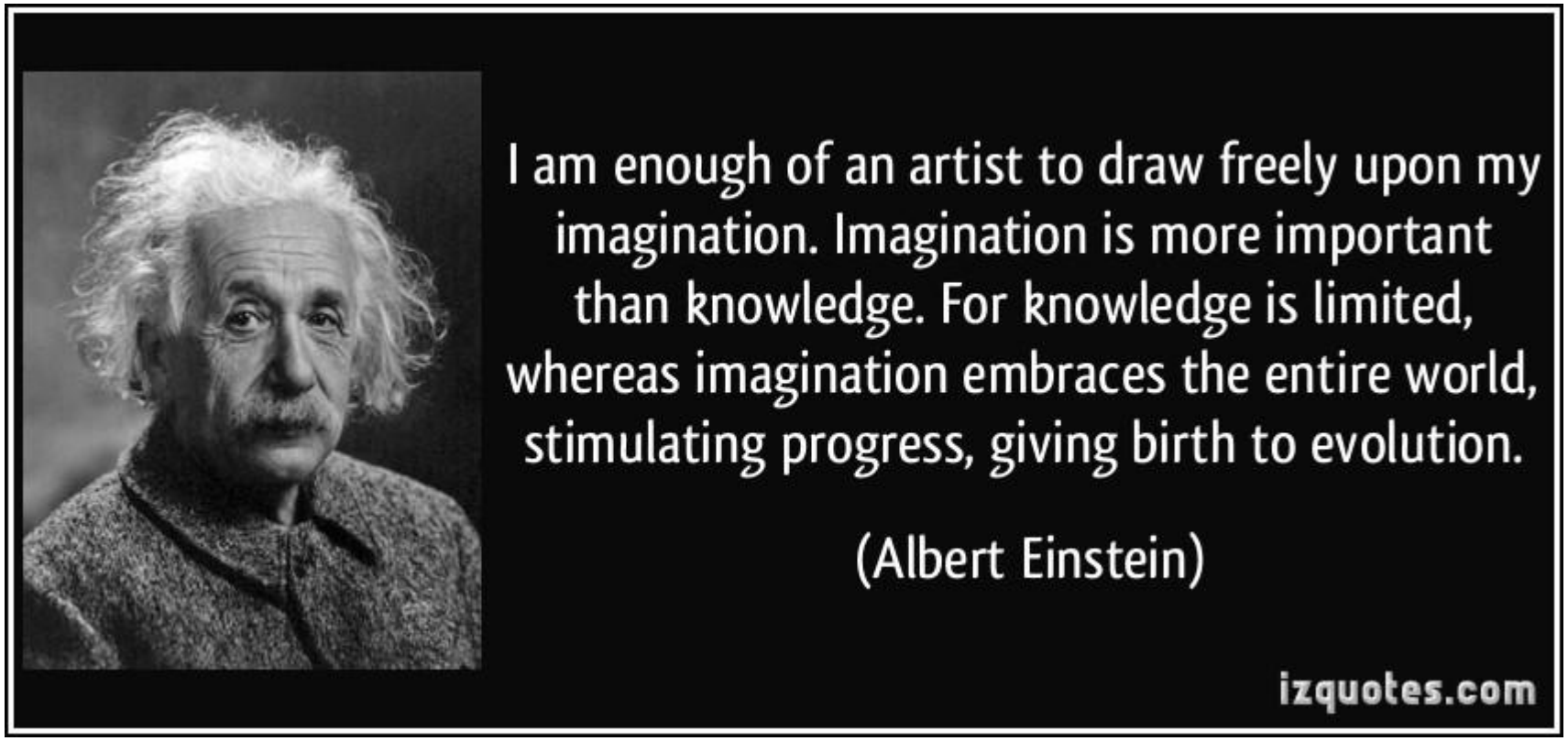
An Introduction to eXplainable Artificial Intelligence (XAI) for Data Sciences

Kevin Kam Fung YUEN, PhD

Outline

- Understand the basic concepts of AI for Data sciences in industries
  - Nexus among Artificial Intelligence, Machine Learning and Deep Learning.
  - AI & Data sciences
  - Data analytics vs. Data science
  - DIKW and KDD
  - Understand Big Data: how big is big?
- eXplainable AI
- Coding Lab in Python for Demo: GradCam for XAI

Imagination is more important than knowledge



Imagination gives birth to Evolution.

First introduced in 1963 in comic books

Iron man

First movie introduced in 2008 https://www.youtube.com/watch?v=PZlL5gZeGDI

https://youtu.be/-rT8TVv6JKA?si=XTfOEa37-kh9UOCV&t=29

https://www.youtube.com/watch?v=ZHk13Xa8dqI

https://youtu.be/U1wEO-pHizQ?si=cUncyMs7QU0f6OYu&t=455

Can it be real from movie?
What will the future war look like???

Move Forward: Can a robot replace human and is with emotion?

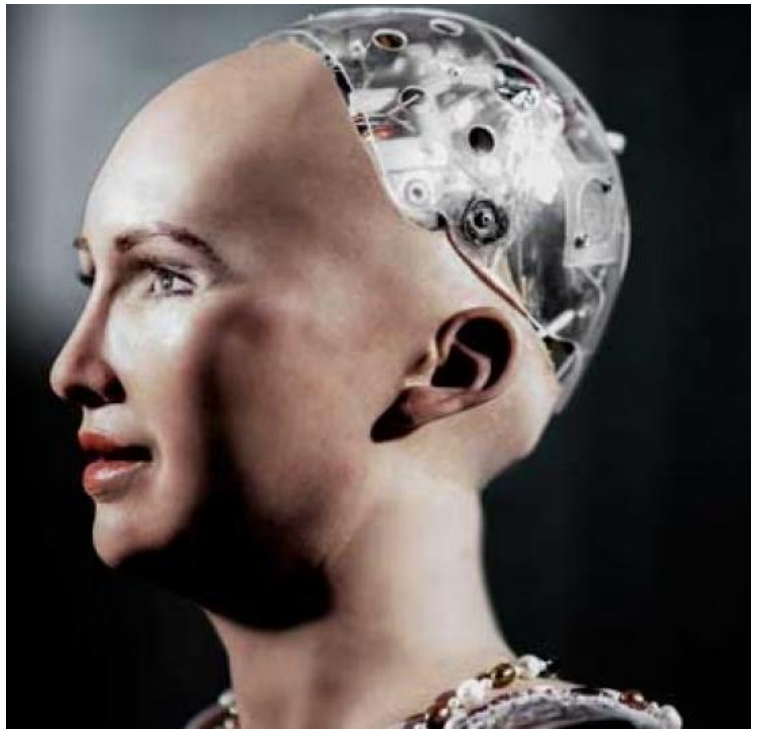
2001

https://www.warnerbros.com/movies/ai-artificial-intelligence
https://www.imdb.com/title/tt0212720/
https://www.youtube.com/watch?v=_N784W3at8c

World's first AI robot citizen "Sophia"

https://www.itc.gov.hk/enewsletter/180801/en/worlds_first_AI_robot_citizen_sophia.html

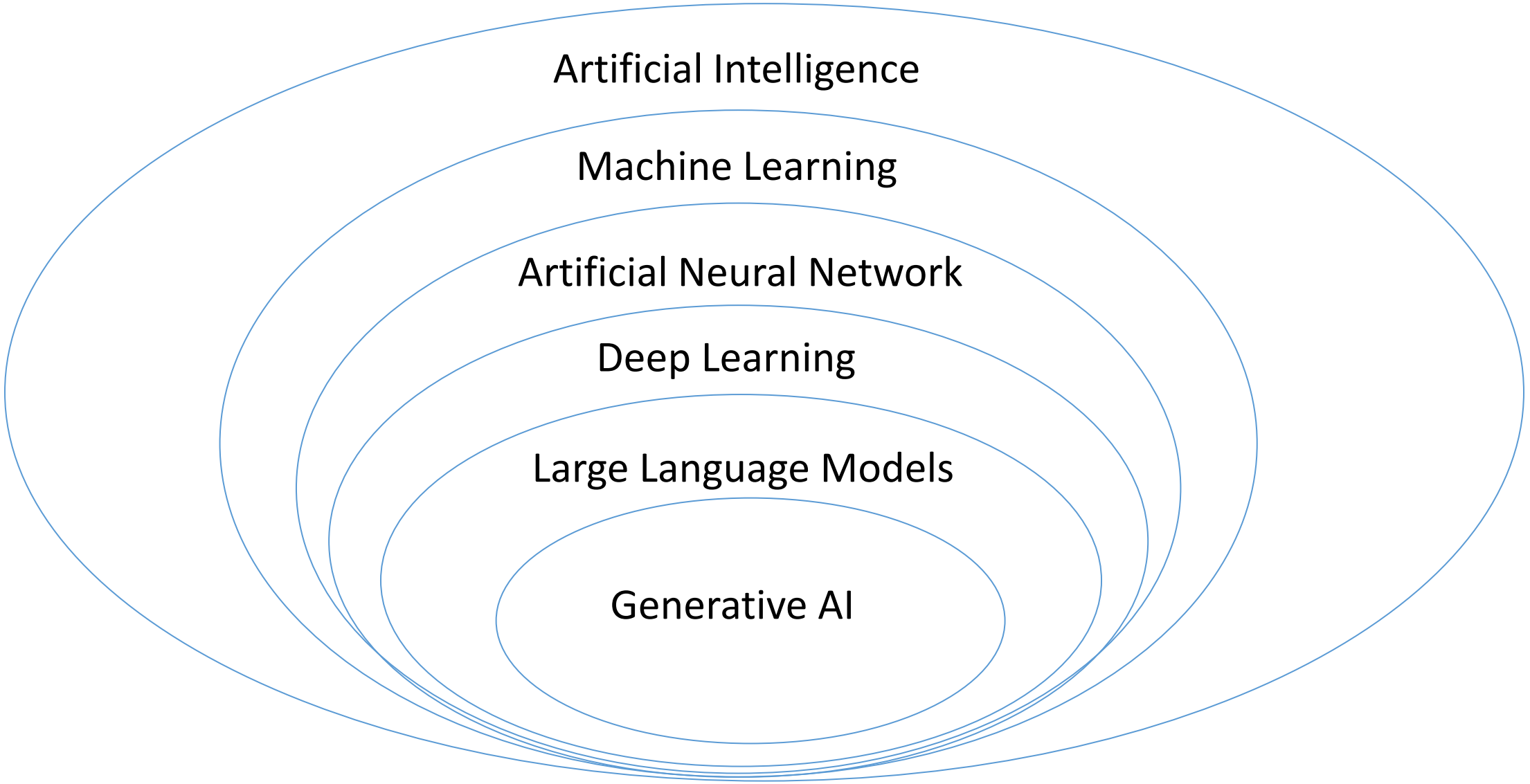
- Artificially intelligent robot "Sophia" created and programmed by Hanson Robotics, a Hong Kong-based humanoid robotics company, is the first robot in the world to be recognised with a citizenship of Saudi Arabia.
- Sophia was marketed as a "social robot" who can mimic social behaviour and induce feelings of love in humans.
- A lot of critics again this: Should robots be citizens? infamous threat to 'destroy humans. https://www.britishcouncil.org/anyone-anywhere/explore/digital-identities/robots-citizens



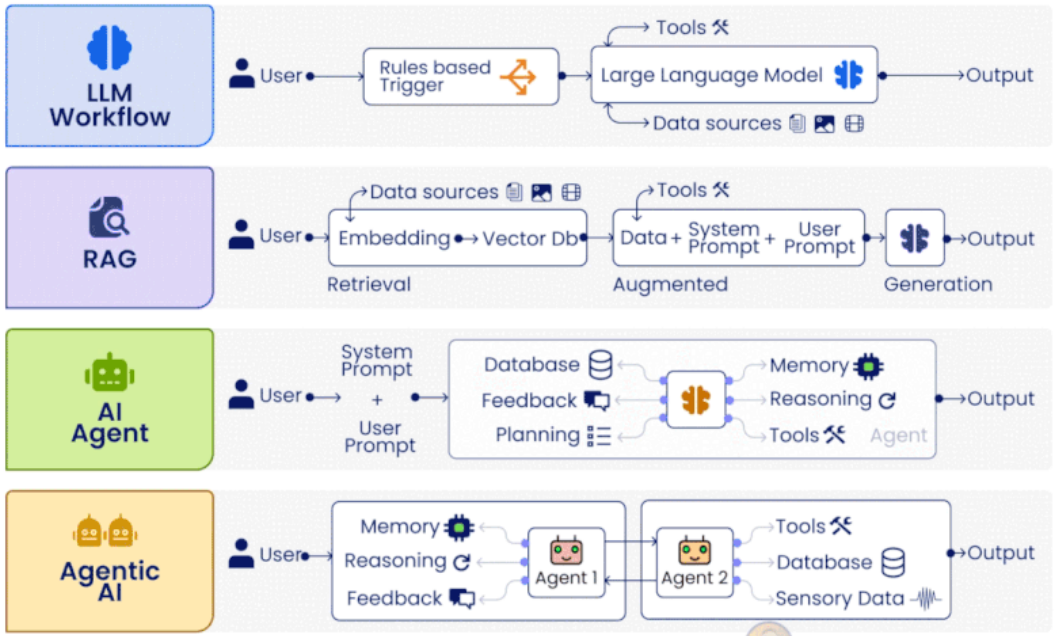
https://www.youtube.com/watch?v=78-1MlkxyqI

https://en.wikipedia.org/wiki/Sophia_(robot)

Nexus among AI, ML, ANN, DL, LLM and GAI



AI Agent VS
LLM vs RAG vs Agentic AI



Topic	LLM Workflow	RAG	AI Agent	Agentic AI
Functionality	Next token prediction based on input	Smart Knowledge retrieval sources	Autonomous action using components	Multi-Agent system to work autonomously
Best Use case	Text generation and summarization	Accurate Q&A from various sources	Workflows requiring tool and reasoning	Large scale tasks needing collaboration
Strength	Fast, low complexity, easy to deploy	Enhanced accuracy with external data	Task automation with planning + reasoning	Flexible, can split work to specialized agents
Weakness	Limited context understanding	Sensitive to data quality	Needs well-defined goals and tool access	Harder to design+ control agents
Examples	Chatbots, email drafting bots	Graph RAG, Advanced RAG, Modular RAG	ReACT Agent, Rewoo Agent	CUA, Embodied Agents

From linkedin

Artificial Intelligence

"A branch of computer science dealing with the simulation of intelligent behavior in computers." (Merriam-Webster)

"A program that can sense, reason, act, and adapt." (Intel)

"Colloquially, the term 'artificial intelligence' is applied when a machine mimics 'cognitive' functions that humans associate with other human minds, such as 'learning' and 'problem solving'." (Wikipedia)

More Definitions

- "Activities that we associate with human thinking, activities , as such decision-making, problem solving, learning" (Bellman, 1978).
- "The study of how to make computers do things, at the moment, people are better" (Rich and Knight, 1991).
- "The study of the computations that make it possible to perceive, reason, and act" (Winston, 1992).
- "Computational Intelligence is the study of the design of intelligent agents" (Poole *et al.,* 1998).
- "AI …. is concerned with intelligent behavior in artifacts" (Nilsson, 1998) .

Four Views on AI Definitions

Systems that think like humans	Systems that think rationally
"The exciting new effort to make computers thinks … *machine with minds,* in the full and literal sense" (Haugeland 1985)	"The study of mental faculties through the use of computational models" (Charniak et al. 1985)
"The art of creating machines that perform functions that require intelligence when performed by people" (Kurzweil, 1990)	"A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes" (Schalkol, 1990)
Systems that act like humans	**Systems that act rationally**

Acting Humanly

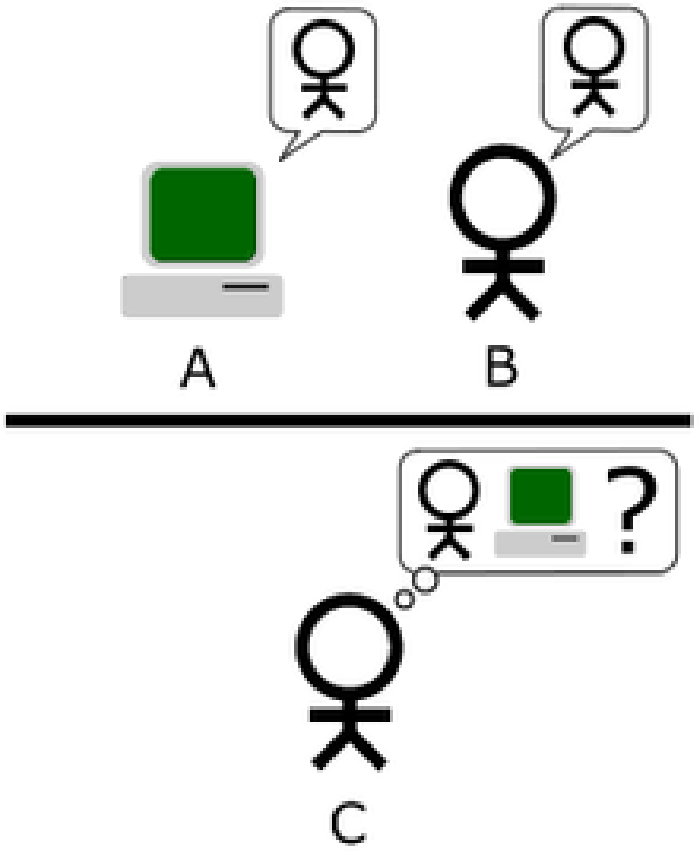
- Emphasis on how to *tell* that a machine is intelligent, not on how to *make it* intelligent

- *when can we count a machine as being intelligent*?

  "Can machines think?"
  – "Can machines behave intelligently?"

- Most famous response due to **Alan Turing**, British mathematician and computing pioneer:

Turing Test (1950)

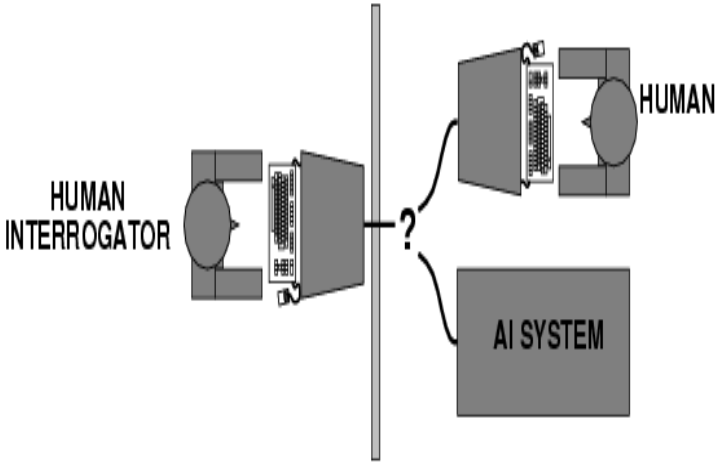


The "standard interpretation" of the Turing Test, in which player C, the interrogator (審問員), is tasked with trying to determine which player - A or B - is a computer and which is a human. The interrogator is limited to using the responses to written questions in order to make the determination.

Turing Test (1950)

- The Turing test is a proposal for a test of a machine's ability to demonstrate intelligence.
- Described by Alan Turing in the 1950 paper "Computing Machinery and Intelligence", it proceeds as follows: a human judge engages in a natural language conversation with one human and one machine, each of which try to appear human.
- All participants are placed in isolated locations. If the judge cannot reliably tell the machine from the human, the machine is said to have passed the test.
- In order to test the machine's intelligence rather than its ability to render words into audio, the conversation is limited to a text-only channel such as a computer keyboard and screen.

Turing test (1950)



System passes if the questioner cannot tell the difference

Up to 2019?

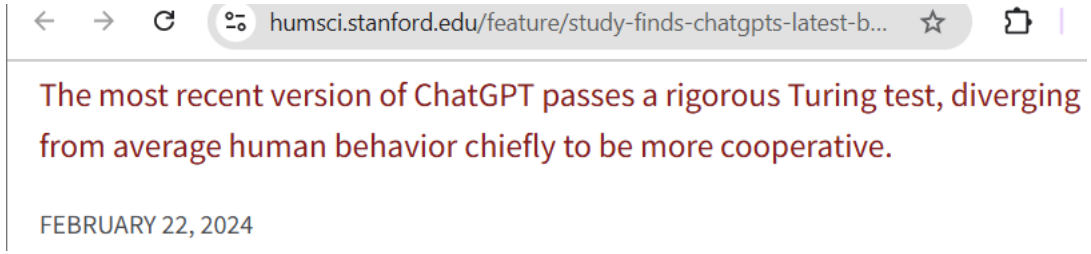
- No program has yet passed Turing test!
  (Annual Loebner competition & prize.)

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

- A program that succeeded would need to be capable of:
  - natural language understanding & generation;
  - knowledge representation;
  - learning;
  - automated reasoning.
- Note: no *visual* or *aural* component to basic Turing test
  — augmented test involves video & audio feed to entity.

Discussion

- Generative AI applications, such as Chatgpt and Deepseek, are very popular.
- Do you think if any chatbots can pass the Turing test now?



https://humsci.stanford.edu/feature/study-finds-chatgpts-latest-bot-behaves-humans-only-better

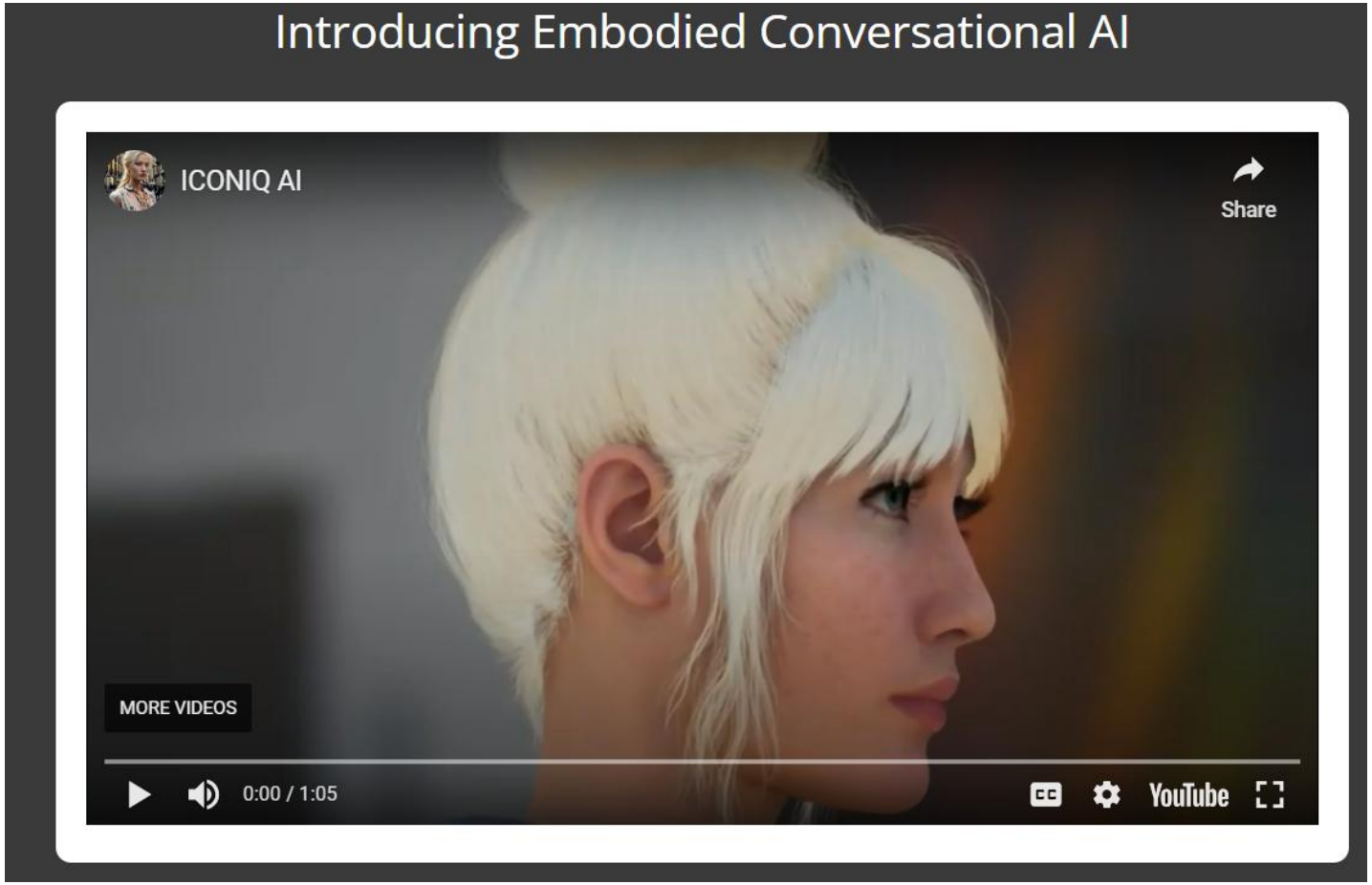
https://www.pnas.org/doi/10.1073/pnas.2313925121

Significance

As AI interacts with humans on an increasing array of tasks, it is important to understand how it behaves. Since much of AI programming is proprietary, developing methods of assessing AI by observing its behaviors is essential.

We develop a Turing test to assess the behavioral and personality traits exhibited by AI. Beyond administering a personality test, we have ChatGPT variants play games that are benchmarks for assessing traits: trust, fairness, risk-aversion, altruism, and cooperation. Their behaviors fall within the distribution of behaviors of humans and exhibit patterns consistent with learning. When deviating from mean and modal human behaviors, they are more cooperative and altruistic. This is a step in developing assessments of AI as it increasingly influences human experiences.

Examples of AI Systems: a version of Alice

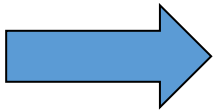


https://chat.kuki.ai/chat

Two Views of the AI Goal

- AI is about duplicating what the (human) brain **DOES**.

  Cognitive science, Turing test

   AI is about duplicating what the (human) brain **SHOULD** do.

  RATIONALITY

Thinking rationally: "laws of thought"

What are the rules (laws) of thought?

Aristole → Gorge Boole→ David Hilbert →..... = Logic

(Aristotelian logic) (Boolean Logic) (Hilbert's axioms)

Thinking rationally "laws of thought"

- Aristotle: what are correct arguments/thought processes?
- Several Greek schools developed various forms of *logic*: *notation* and *rules of derivation* for thoughts; may or may not have proceeded to the idea of mechanization.
- Direct line through mathematics and philosophy to modern AI (logic-based agents).
- Problems:
  1. Not all intelligent behavior is mediated by logic.
  2. What is the purpose of thinking? What thoughts should I have?

Acting Rationally

- *Rational behavior:* doing the right thing
- The right thing = what is expected to maximize goal achievement, given the available information
- Does not necessarily involve thinking— e.g., blinking reflex — but thinking should be in the service of rational action

Aristotle:

Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good

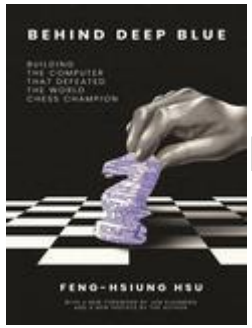
Example AI System: Chess Playing

Example: Deep Blue (IBM)

- Perception: advanced features of the board
- Actions: choose a move
- Reasoning: heuristics to evaluate board positions, search

https://www.ibm.com/history/deep-blue

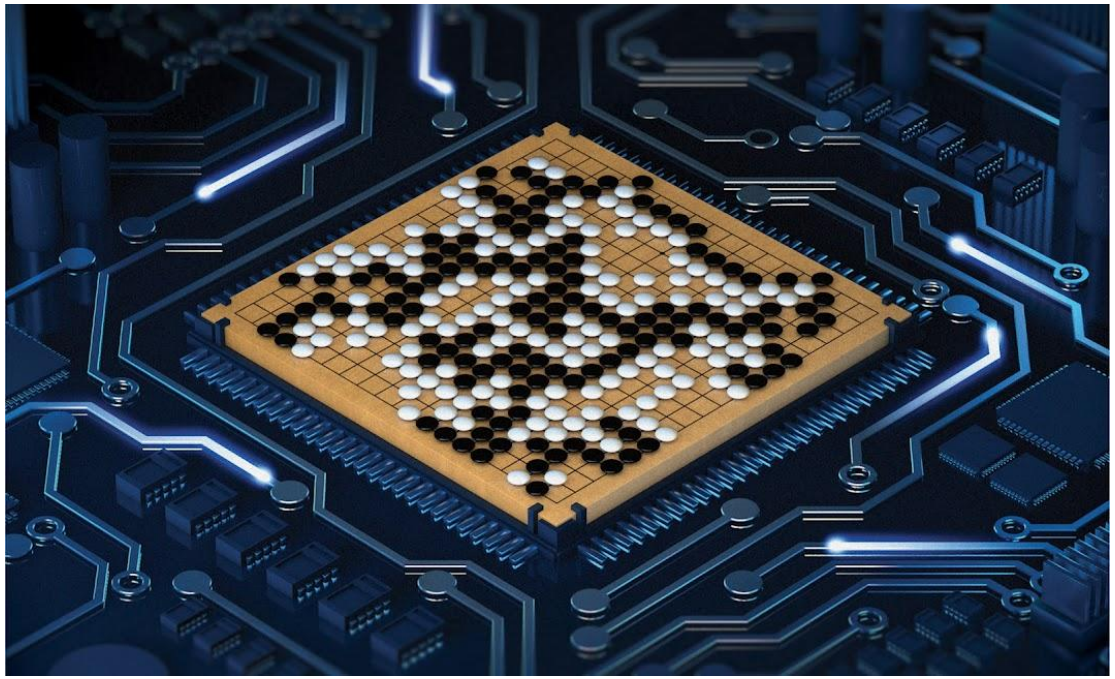
IBM's computer checkmated a human chess champion in a computing tour de force



Feng-hsiung Hsu; Jon Kleinberg, *Behind Deep Blue: Building the Computer That Defeated the World Chess Champion* , Princeton University Press, 2022.

Google deepmind: AlphaGo

- **AlphaGo** mastered the ancient game of Go, defeated a Go world champion, and inspired a new era of AI systems.



In October 2015, AlphaGo played its first game against the reigning three-time European Champion, Fan Hui. AlphaGo won the first ever match between an AI system and Go professional, scoring 5-0.

https://deepmind.google/research/breakthroughs/alphago/

https://www.youtube.com/watch?v=WXuK6gekU1Y

Examples of AI Systems: The Internet

Today's Internet uses AI technologies to create a faster, safer, easier-to-use environment.

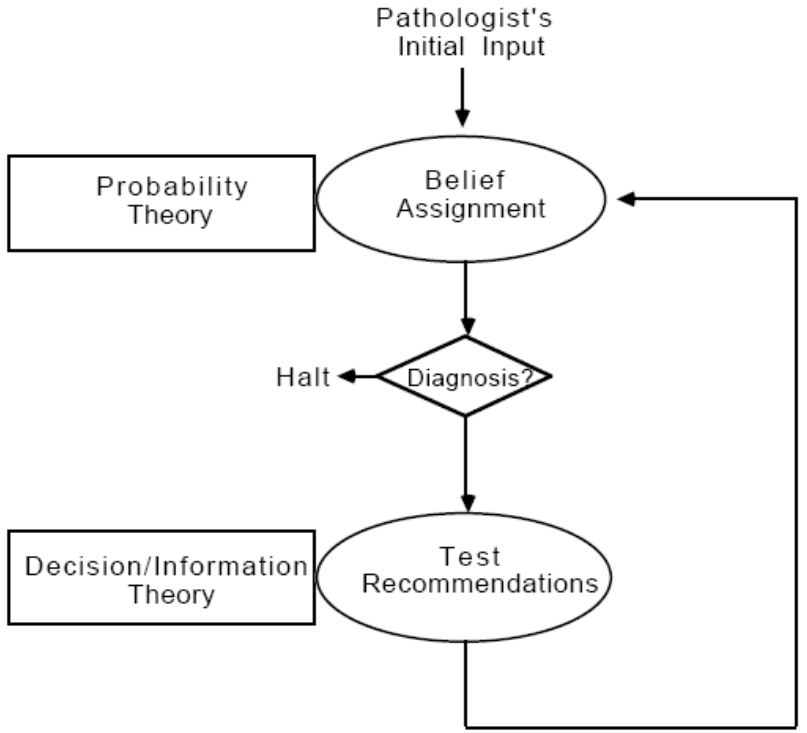
- Perception: users' information (statistics)
- Actions:
  - `guess' your preferences and make recommendations about other products you may like;
  - scour the web for new information;
  - create site ratings and make sure the most popular sites are presented first;
  - predict the things you're interested in and to guess advertisers you're most likely to click onto.
- Reasoning: inference, machine learning, …

Example AI System: Medical Diagnosis

Example: Pathfinder

(D.Heckerman, Microsoft Research)

- Perception: symptoms, test results
- Actions: suggest tests, make diagnosis
- Reasoning: Bayesian inference, machine learning, Monte Carlo simulation



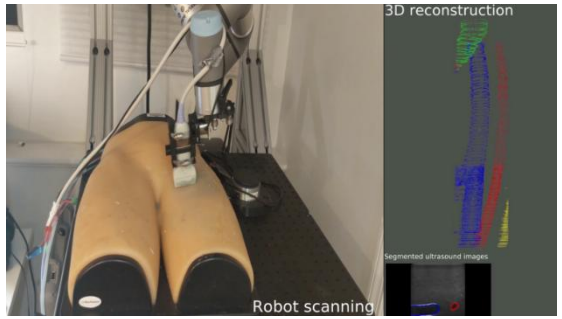
Pathfinder architecture. Initial evidence presented to the system is used to form a list of hypotheses. Next, information analysis identifies the next best tests.

heurabs.pdf

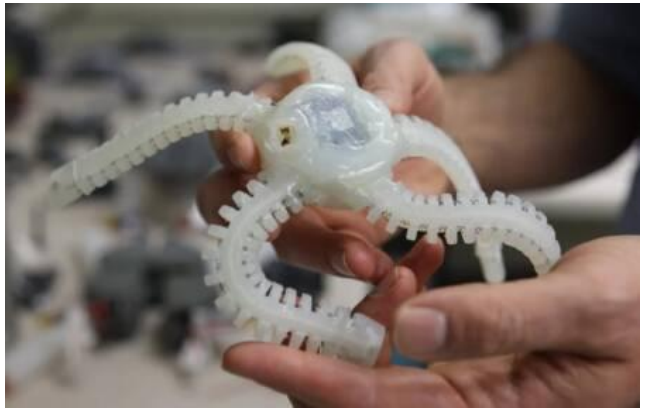
Examples of AI Systems

Field Robotics

Computer Vision



Healthcare Robotics



Robot Embodiment

Human Robot Interaction

https://www.ri.cmu.edu/research-overview/ , captured at 2025

Examples of AI Systems: Robot

Honda Avatar Robot

An avatar is the representation of the user. By offering avatar robots, Honda wants to create a society where everyone, no matter how far away they are, where they are and who they are, can do what they want to do without needing to be “at the spot” in person.

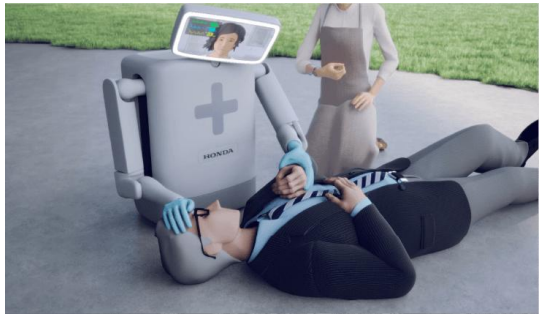
Picking up a coin

Opening a can using a pull tab

PET bottle cap (required torque of 1.5N·m)

Glass jar lid (required torque of 5.0N·m)

Holding heavy item (15kg)



https://global.honda/en/tech/Avatar_robot/

More: https://www.youtube.com/watch?v=wgthZ30kkLk

Examples of AI Systems: Sony AIBO

https://us.aibo.com/

Discussion

- Should we use GAI in education? State your pros and cons.

Discussion

- How should we use AI in an Industry?
- Has Artificial Intelligence been overhyped?

Discussion: AI transformation

1. Execute pilot projects to gain momentum
2. Build an in-house AI team
3. Provide broad AI training
4. Develop an AI strategy
5. Develop internal and external communications

https://landing.ai/ai-transformation-playbook/

Captured in 2021

- Familiar with the following software/tools:
  - Coding knowledge and experience with several languages: C, C++, Java, JavaScript, etc.
  - Knowledge and experience in statistical and data mining techniques: GLM/Regression, Random Forest, Boosting, Trees, text mining, social network analysis, etc.
  - Experience querying databases and using statistical computer languages: R, Python, SLQ, etc.
  - Experience using web services: Redshift, S3, Spark, DigitalOcean, etc.
  - Experience creating and using advanced machine learning algorithms and statistics: regression, simulation, scenarios analysis, modeling, clustering, decision trees, neural networks, etc.
  - Experience analyzing data from 3rd party providers: Google Analytics, Site Catalyst, Coremetrics, Adwords, Crimson Hexagon, Facebook Insights, etc.
  - Experience with distributed data/computing tools: Map/Reduct, Hadoop, Hive, Spark, Gurobi, MySQL, etc.
  - Experience visualizing/presenting data for stakeholders using: Periscope, Business Objects, D3, ggplot, etc.

From the advertisement, it looks that the company is not AI ready, and no clear objectives.

Source: LinkedIn Job – Data Scientist

FORTUNE

MIT report: 95% of generative AI pilots at companies are failing

August 18, 2025

https://finance.yahoo.com/news/mit-report-95-generative-ai-105412686.html

"Some large companies' pilots and younger startups are really excelling with generative AI," Challapally said. Startups led by 19- or 20-year-olds, for example, "have seen revenues jump from zero to $20 million in a year," he said. "It's because they pick one pain point, execute well, and partner smartly with companies who use their tools," he added.

But for 95% of companies in the dataset, generative AI implementation is falling short. "The 95% failure rate for enterprise AI solutions represents the clearest manifestation of the GenAI Divide," the report states. The core issue? Not the quality of the AI models, but the "learning gap" for both tools and organizations. While executives often blame regulation or model performance, MIT's research points to flawed enterprise integration. Generic tools like ChatGPT excel for individuals because of their flexibility, but they stall in enterprise use since they don't learn from or adapt to workflows, Challapally explained.

Source: linkedin

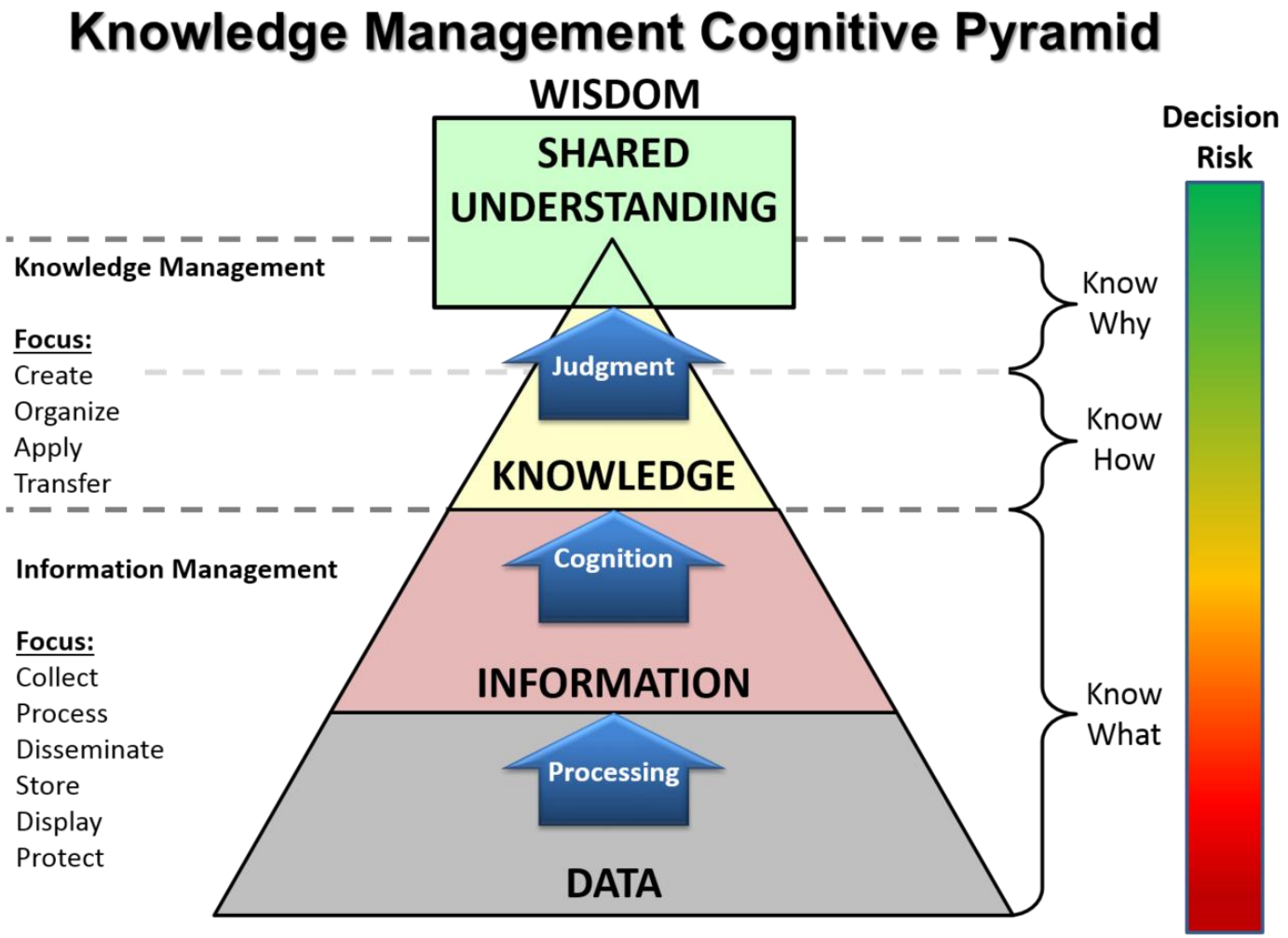
Data analytics vs. Data science

- **Data analytics** refers to the process and practice of analyzing data to answer questions, extract insights, and identify trends.
- This is done **using** an array of tools, techniques, and frameworks that vary depending on the type of analysis being conducted.

- **Data science** is centered on building, cleaning, and organizing datasets.
- Data scientists **create** and leverage algorithms, statistical models, and their **own custom analyses** to collect and shape raw data into something that can be more easily understood.

What's the Difference Between Data Analytics & Data Science? (hbs.edu)

DIKW pyramid

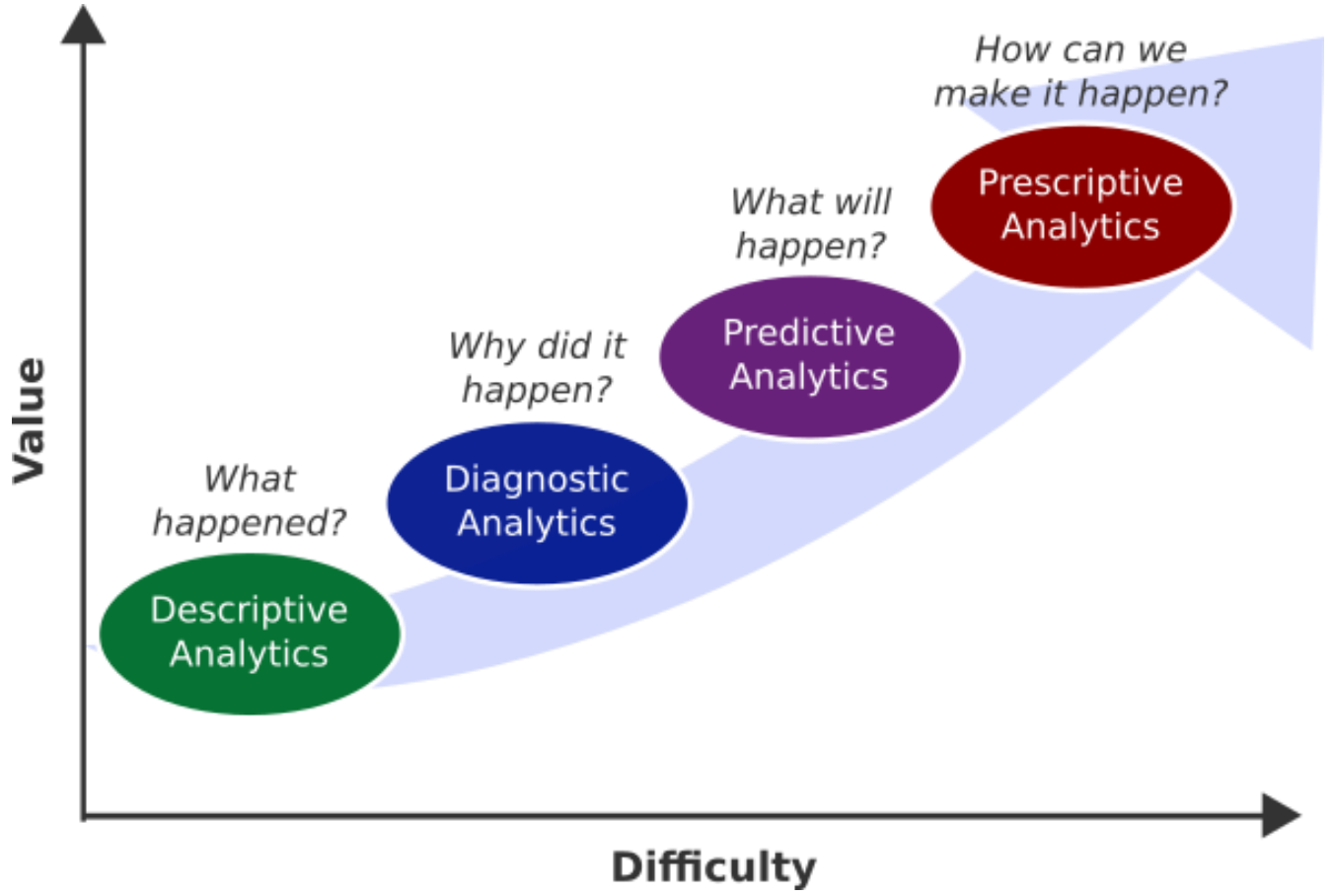


https://en.wikipedia.org/wiki/DIKW_pyramid

Also watch:
https://www.youtube.com/watch?v=sIjSY05JE9Q
https://www.youtube.com/watch?v=u9DoQ9gY4z4

Types of Analytics

Descriptive, Diagnostic, Predictive, and Prescriptive Analytics

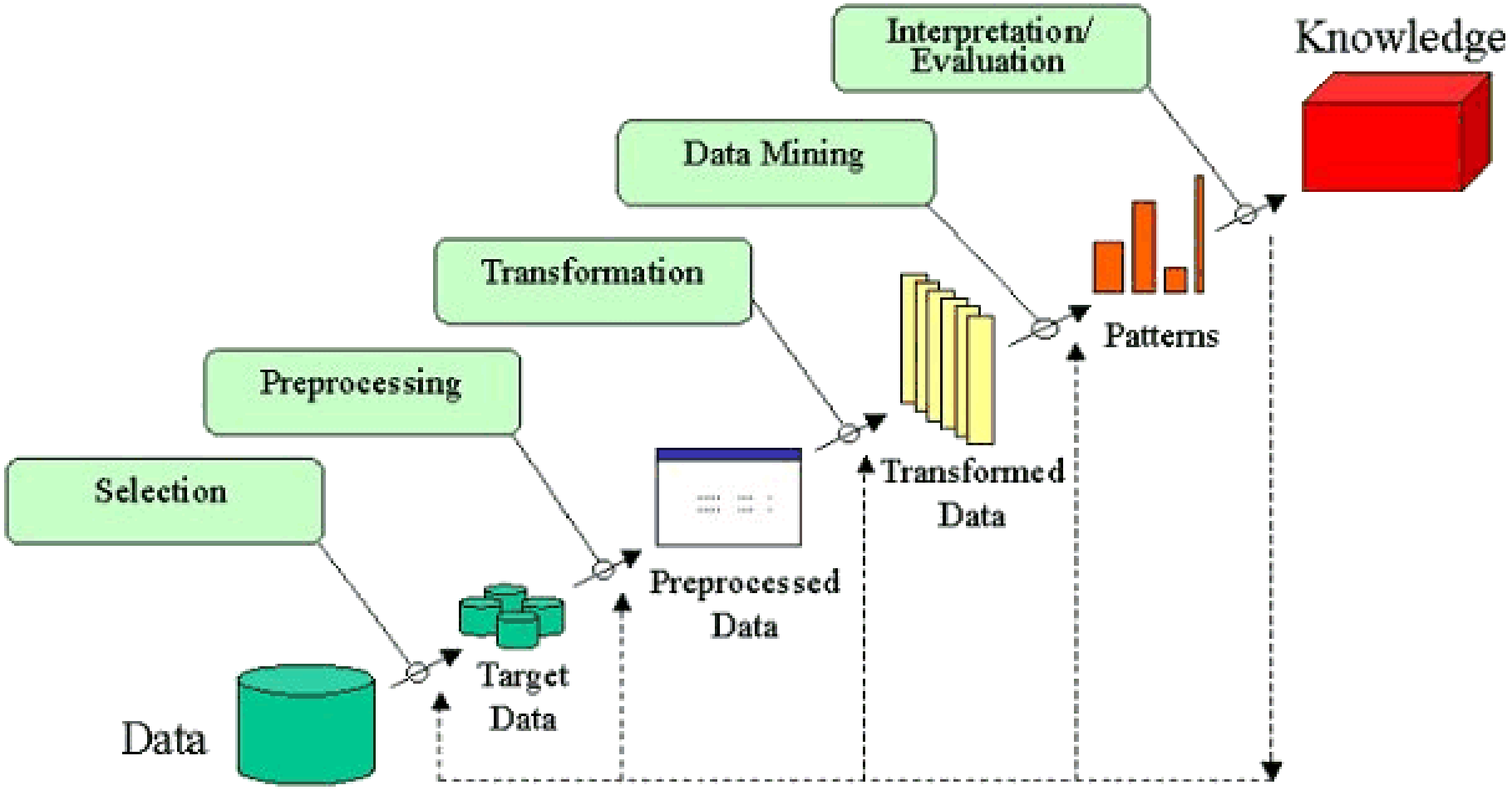


[Data classification and analysis - Governance Analytics Knowledge Base]

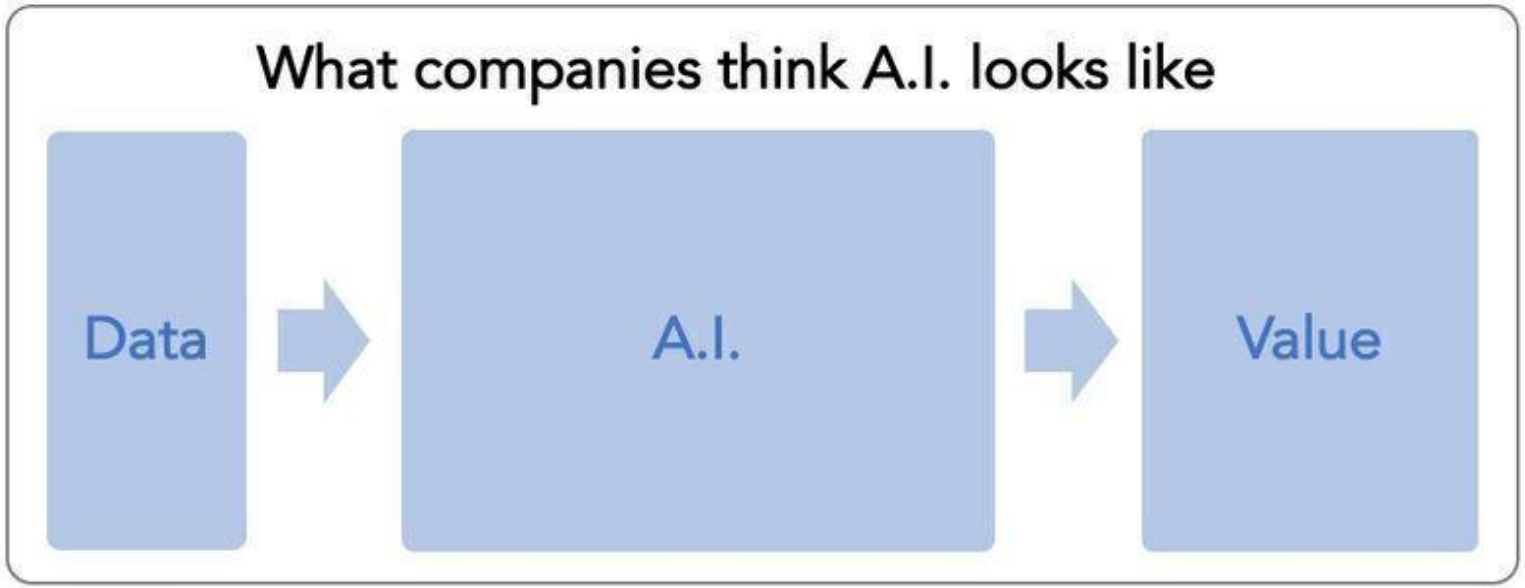
[Analytics Building Blocks | SpringerLink]

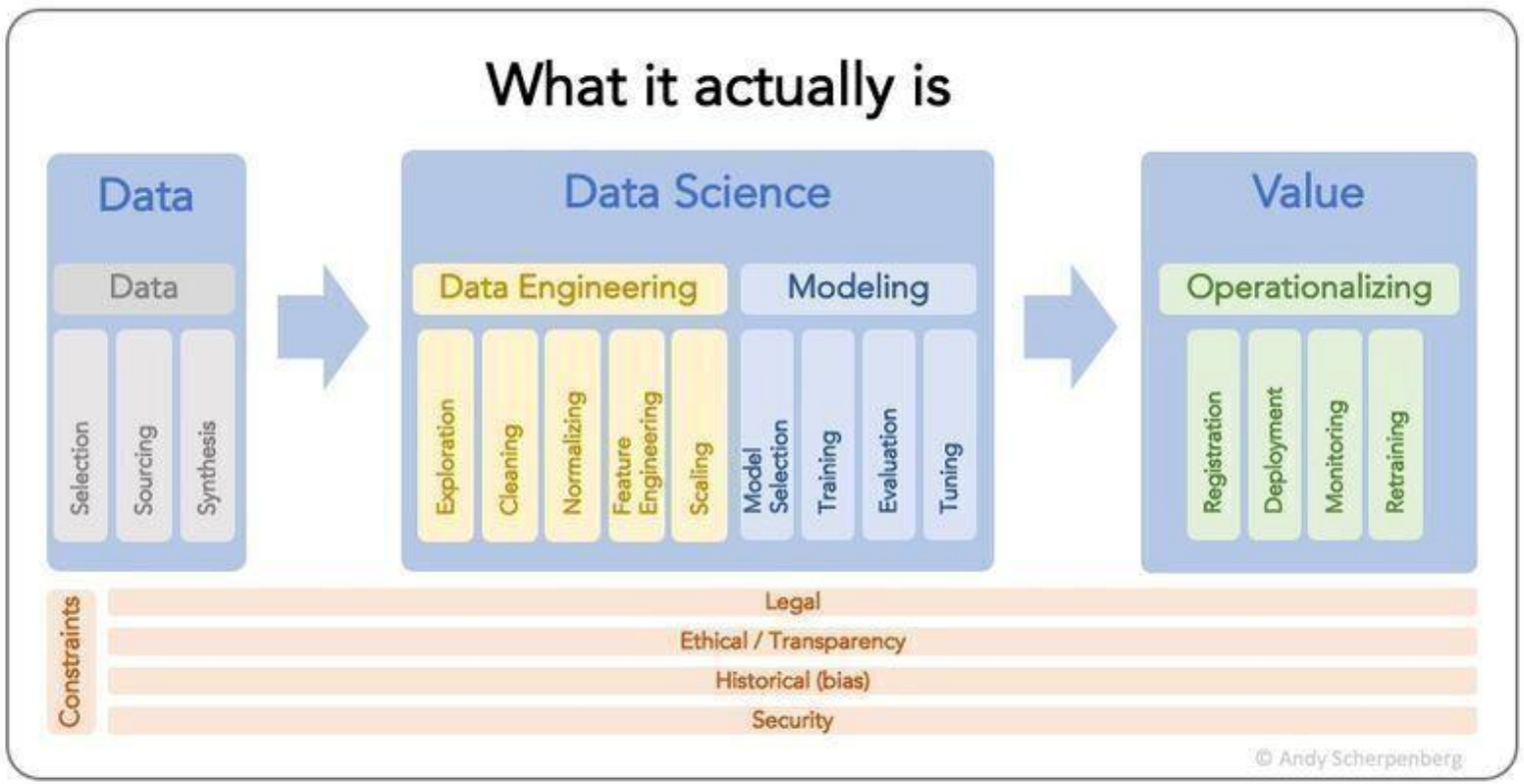
Also watch:
https://www.youtube.com/watch?v=aCiOCPrTH-E

Overview of the Knowledge Discovery in Databases (KDD) Process



KDD Process/Overview (uregina.ca)





Source: Linkedin

Big Data Analytics

- The rapid growth of data we experience today has led to the urgent need to develop effective and efficient techniques for big data analytics, which are much required by industries and academic communities in order to be able to discover useful information, or knowledge in big data.
- Big data analytics concerns the use of modern statistical and other machine learning techniques to analyze huge amounts of data.

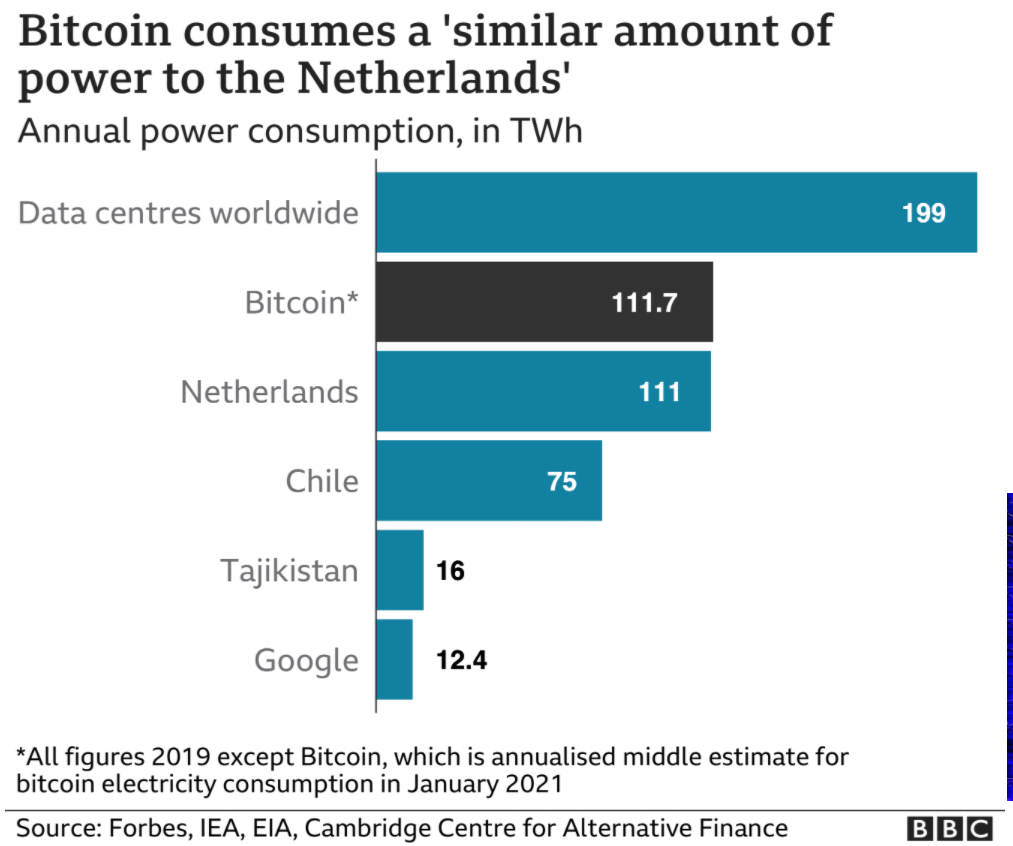
(Yuen et al, 2018)

Challenging issues in Big Data Analytics

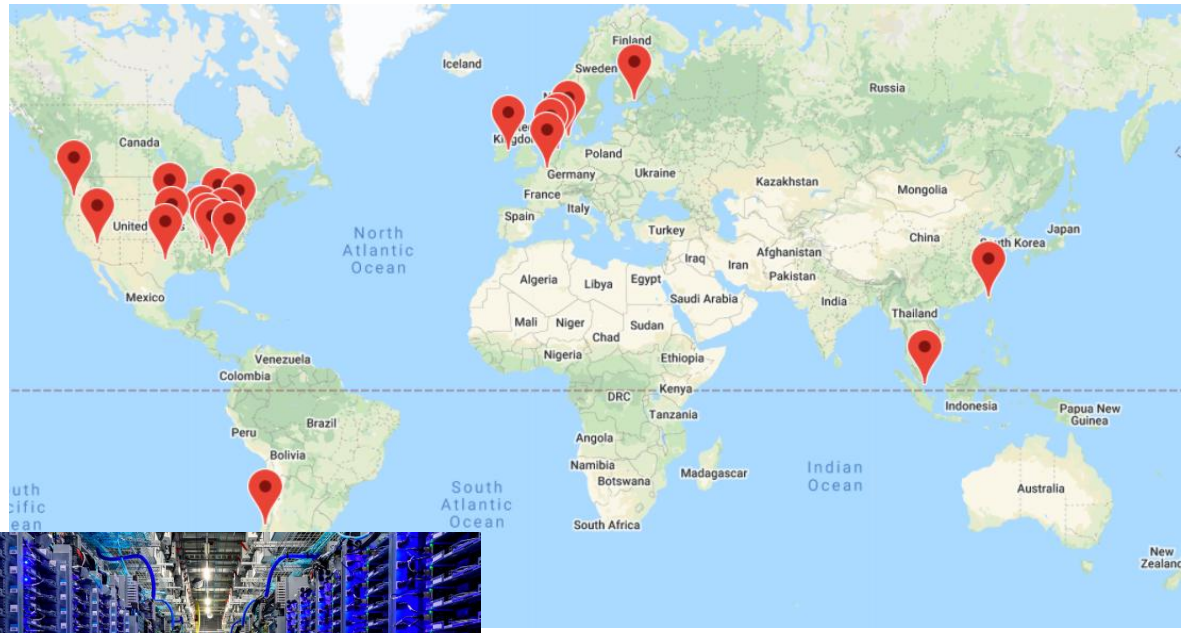
- High dimensionality of data
- Multiple objectives of the problems under study
- Conventional 5Vs,
  - large scale of data (Volume),
  - multiple sources of data (Variety),
  - rapid growth of data (Velocity),
  - quality of data (Veracity),
  - usefulness of data (Value).

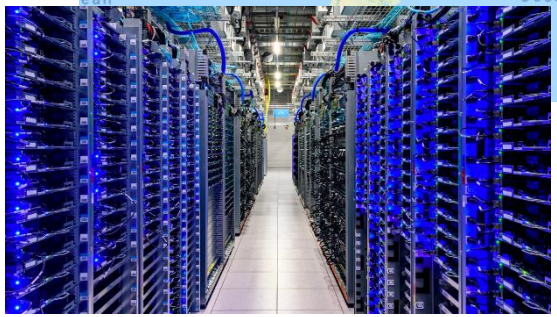
(Yuen et al, 2018)

Data centers



https://www.bbc.com/news/science-environment-56215787





Google - Our Secure Data Centers - YouTube

https://www.google.com/about/datacenters/innovations/

Discussion

For big data, how big is big? Which unit below is most likely correct for minimal size of big data today?

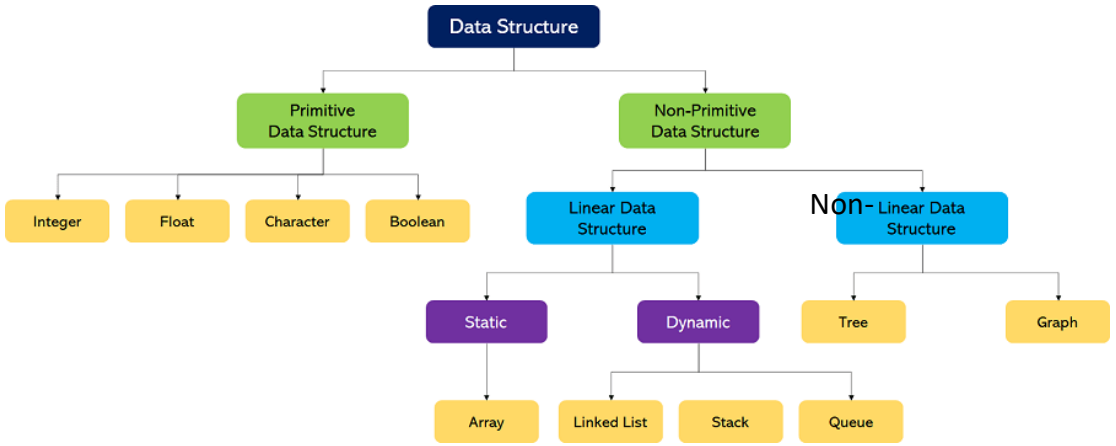
a) kilo bytes (one thousand (1,000) byes)
b) mega bytes (one million (1,000,000) byes).
c) giga bytes (one billion (1,000,000,000) byes).
d) tera bytes (one trillion (1,000,000,000,000) byes).
e) peta bytes (one quadrillion (1,000,000,000,000,000) byes).
f) exa bytes (one quintillion (1,000,000,000,000,000,000) byes).

??"640KB ought to be enough for anybody“, Bill Gate, 1981??

Why Big Data is now such a big deal | Internet | The Guardian

Two Aspects

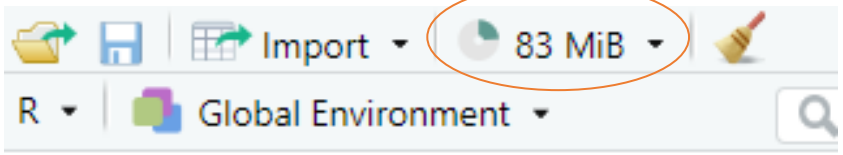
- Software

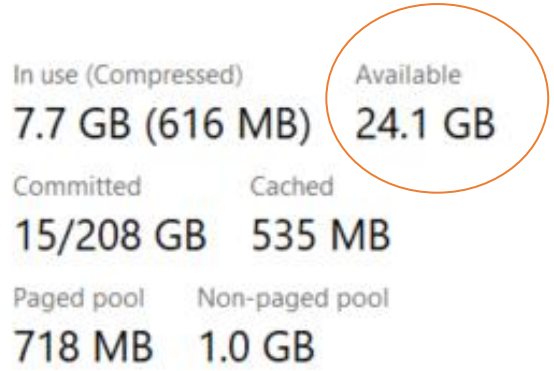


- Hardware

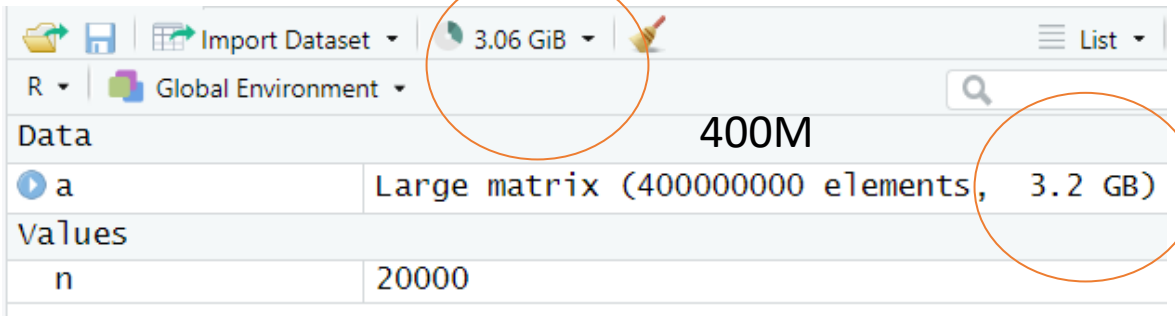
Lab: Simple example for Volume
How big is big?
Problem of Matrix in R

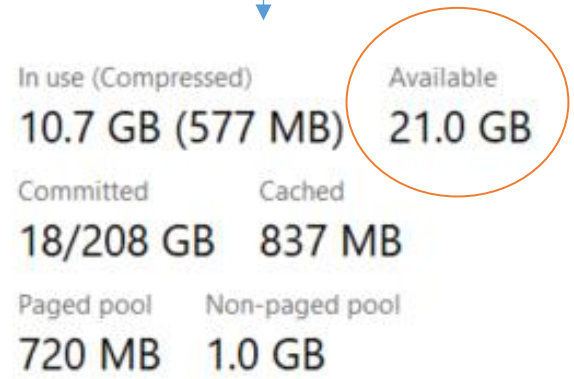
```
> # 3 GB
> n = 20000
> system.time (
+    (a = matrix(0, n, n))
+ )
   user  system elapsed
   0.13    0.03    0.42
>
> print(
+    object.size(a),
+    units = "GB"
+ )
3 Gb
>
```



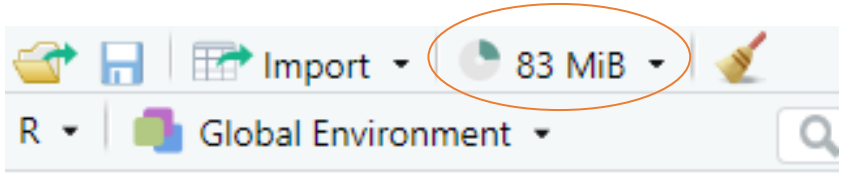


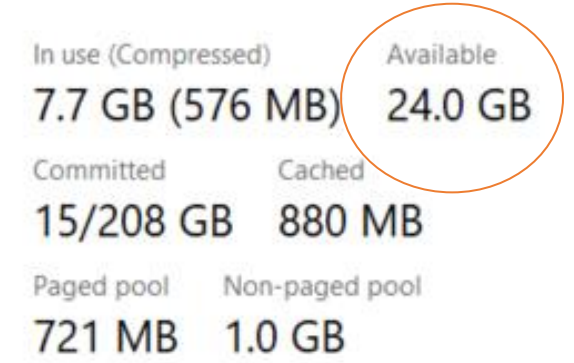
Environment is empty



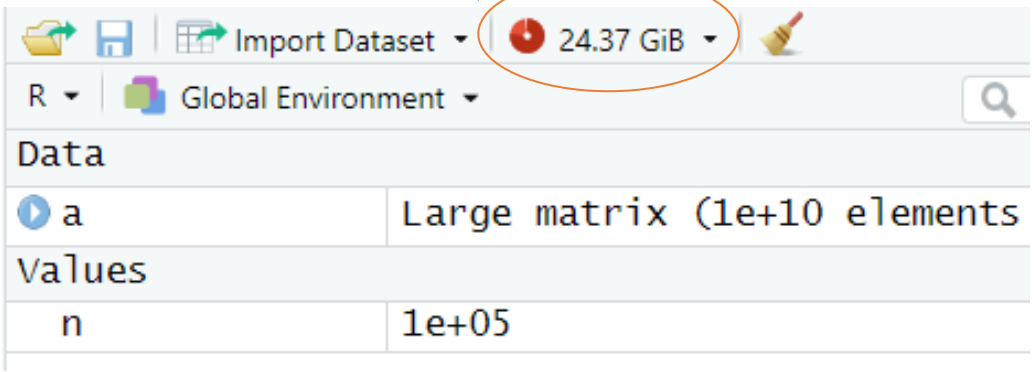


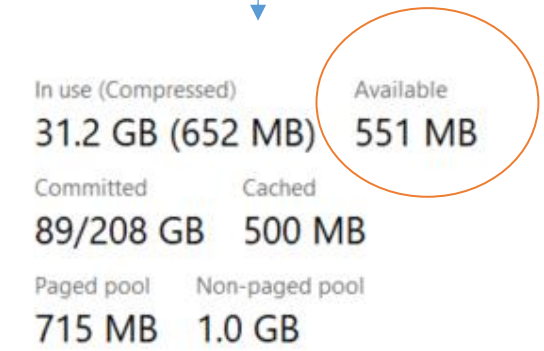
```
> n = 100000
> system.time (
+    (a = matrix(0, n, n))
+ )
   user  system elapsed
   1.75    5.78   17.92
>
> print(
+    object.size(a),
+    units = "GB"
+ )
74.5 Gb
```

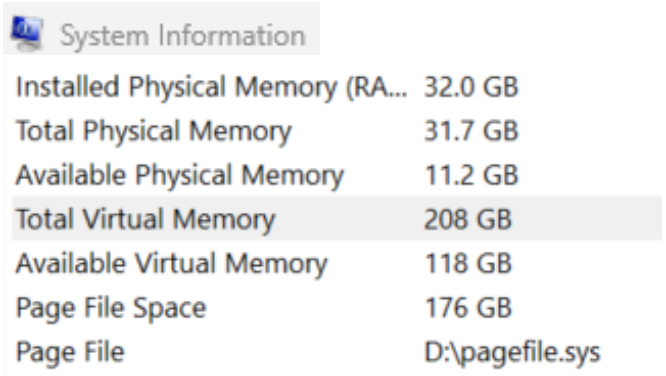




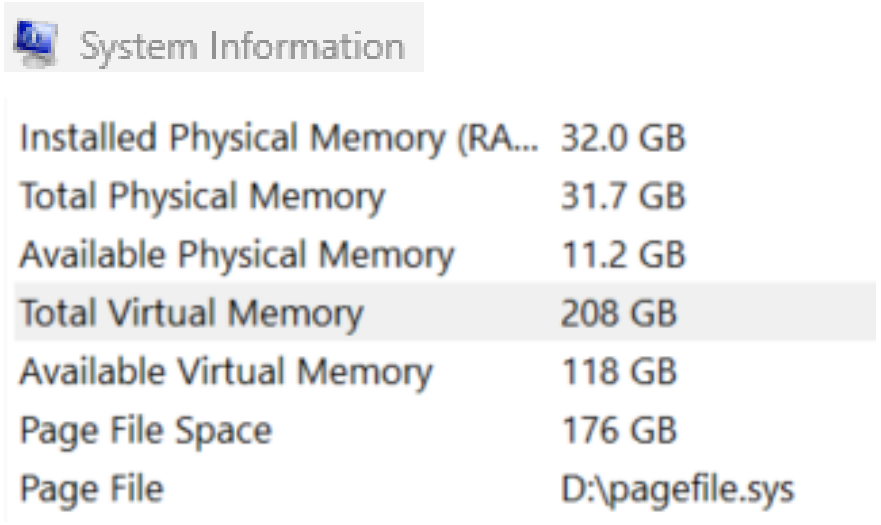
Environment is empty







```
> n = 200000
>
> system.time (
+    (a = matrix(0, n, n))
+ )
Error: cannot allocate vector of size 298.0 Gb
Timing stopped at: 0.01 0 0.06
```



- Unlike the classical programming language, in the latest version of R, the variable size depends on the physical and virtual memory size in the computer.
- Microsoft recommends that you set virtual memory to be no less than 1.5 times and no more than 3 times the amount of RAM on your computer.

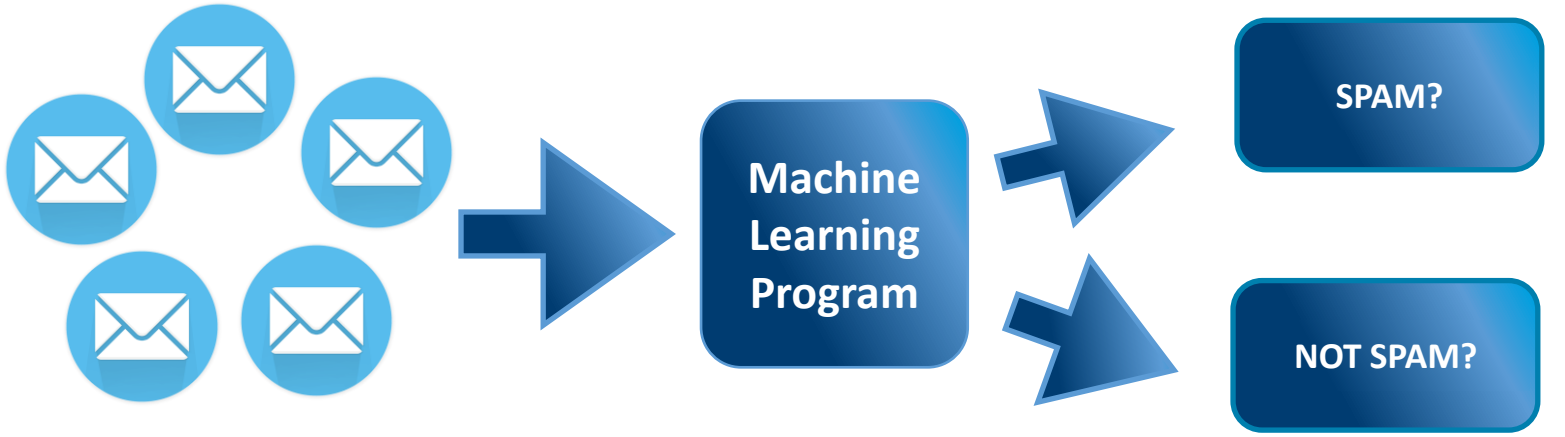
Discussion

- Search internet, what is the largest file size which MS Excel can handle?

Machine Learning

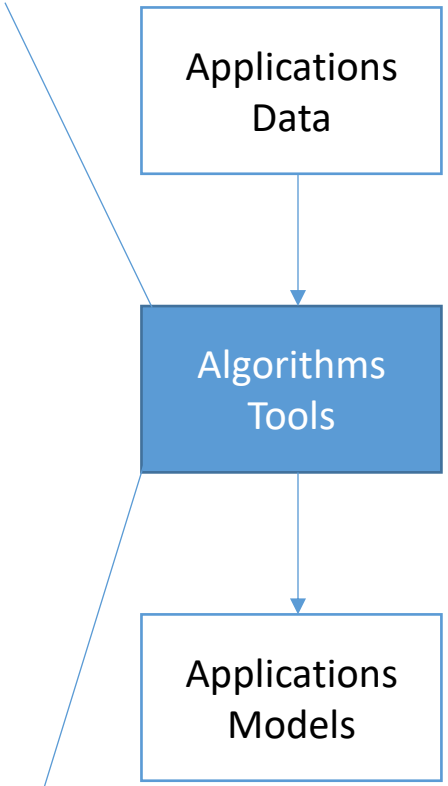
"The study and construction of programs that are *not explicitly programmed*, but learn patterns as they are exposed to more data over time." (Intel)

These programs learn from repeatedly seeing data, rather than being explicitly programmed by humans.



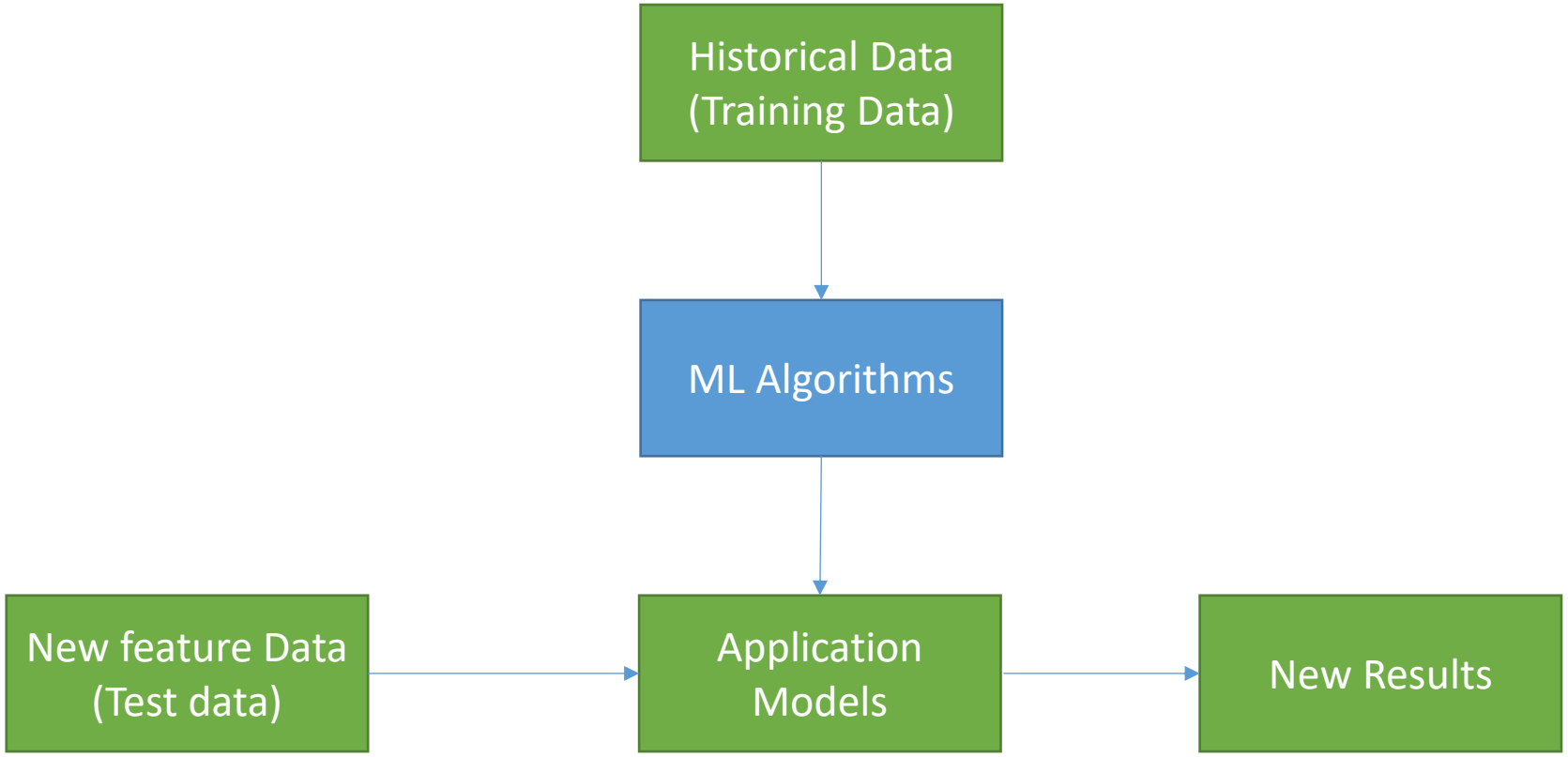
Source: Intel AI Academy

Data Analytics with Machine Learning



Source:
http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Elements of ML for Data Analytics



Scale of measurement

Attribute type	Scale of measurement	Description	Examples
Qualitative / Categorical	Nominal	Without order	Gender, colour
	Ordinal	With order	Grades
Quantitative / Metric	Interval	Arbitrary zero	Temperature in C, Intelligence Quotient
	Ratio	Natural zero	Temperature in Kelvin, Height, weight

Supervised and Unsupervised Learning

	Datasets	Goals	Examples
Supervised Learning	With Target Variable	Make predictions	Credit card Fraud detection
Unsupervised Learning	No Target Variable	Find (unknown) structure /insight in the data	Customer segmentation

Supervised and Unsupervised Learning

- Classification: training with Target variable
- Clustering: training without Target variable

- Regression: training with Target

Features:
Independent variables

Target:
Attribute to be predicted.
dependent variables

Sepal.Length <dbl>	Sepal.Width <dbl>	Petal.Length <dbl>	Petal.Width <dbl>	Species <fct>
4.9	2.5	4.5	1.7	virginica
5.1	3.8	1.9	0.4	setosa
5.6	2.5	3.9	1.1	versicolor
6.5	3.2	5.1	2.0	virginica
5.4	3.4	1.5	0.4	setosa
4.9	3.1	1.5	0.1	setosa
5.6	3.0	4.5	1.5	versicolor
4.7	3.2	1.6	0.2	setosa
6.2	3.4	5.4	2.3	virginica
6.7	3.3	5.7	2.5	virginica
5.8	4.0	1.2	0.2	setosa

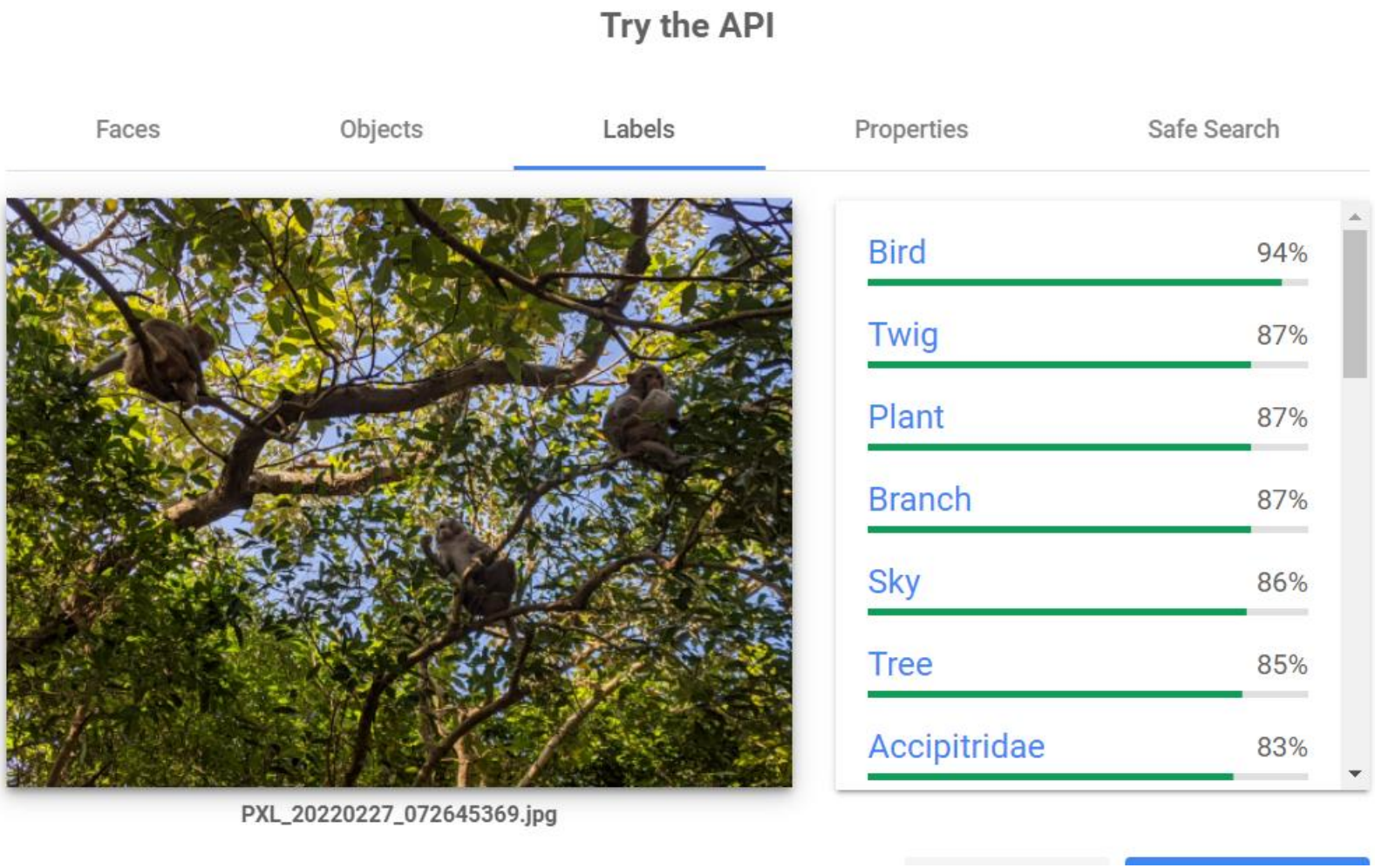
	age <int>	sex <fct>	bmi <dbl>	children <int>	smoker <fct>	region <fct>	charges <dbl>
1	19	female	27.900	0	yes	southwest	16884.924
2	18	male	33.770	1	no	southeast	1725.552
3	28	male	33.000	3	no	southeast	4449.462
4	33	male	22.705	0	no	northwest	21984.471
5	32	male	28.880	0	no	northwest	3866.855
6	31	female	25.740	0	no	southeast	3756.622

Discussion: Algorithm Trust

- Algorithm aversion exists: people don't choose algorithms over humans.
- They may choose the algorithm if they haven't seen any results (even if those results are better than theirs!)
- **If people see algorithms make mistakes, they strongly avoid it.**
- **If people could change the algorithm, even a bit, after seeing its performance, they were more likely to use it.**
- **Combining their judgement with their algorithm helped them feel confident.**

Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey: "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err", Journal of Experimental Psychology: General, 2014.

https://cloud.google.com/vision



Why did the computer produce such results?
Can we trust the result?

Captured in 2023

Why eXplainable AI?

- As AI becomes more advanced, humans are challenged to comprehend and retrace how the algorithm came to a result. The whole calculation process is turned into what is commonly referred to as a “black box" that is impossible to interpret. These black box models are created directly from the data. And, not even the engineers or data scientists who create the algorithm can understand or explain what exactly is happening inside them or how the AI algorithm arrived at a specific result.

> Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms.

https://www.ibm.com/think/topics/explainable-ai

A case: Woman wrongly accused of carjacking loses lawsuit against Detroit police who used facial tech

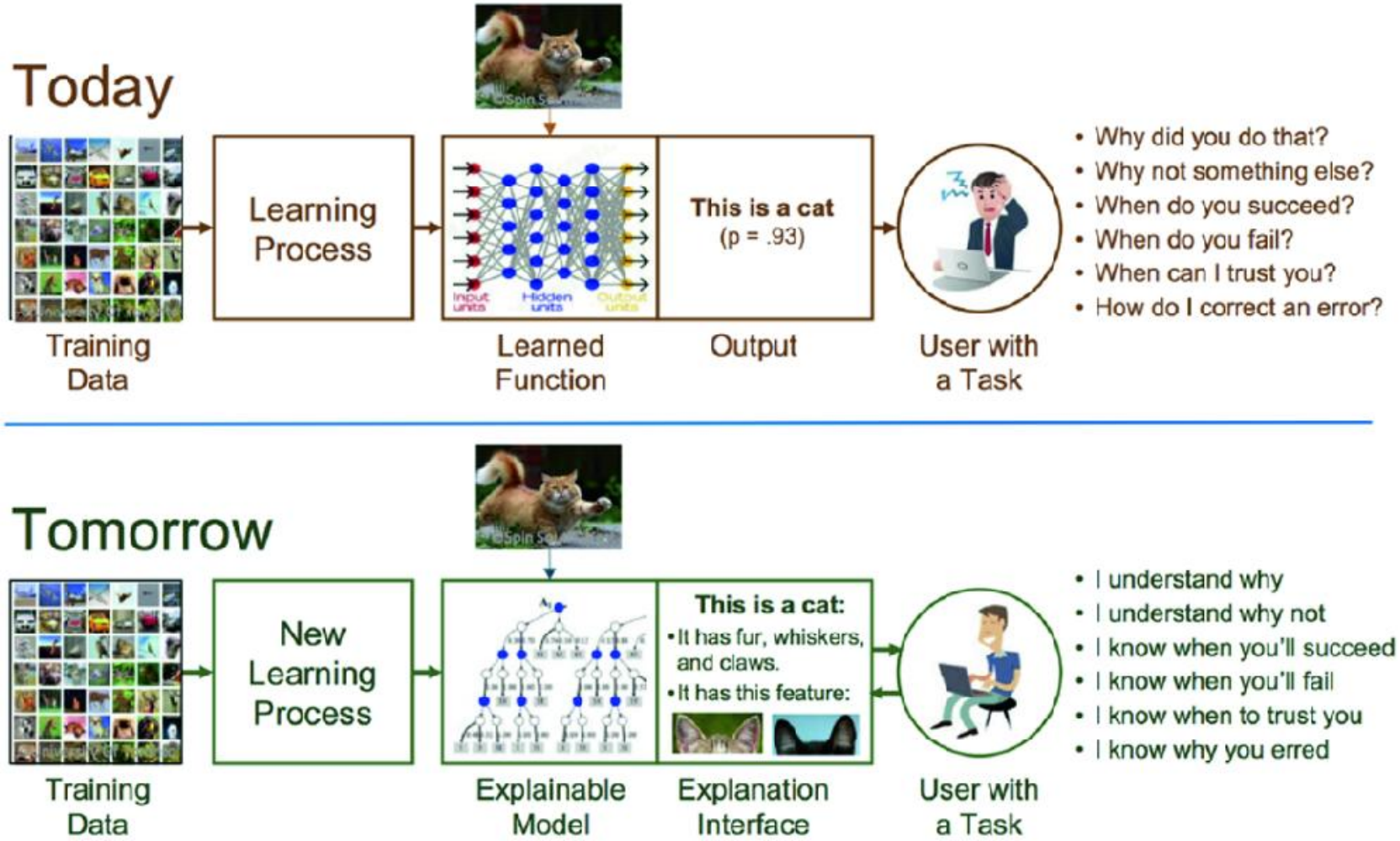
- A judge has dismissed a lawsuit against Detroit police in the wrongful arrest of an eight months pregnant woman who was charged in a carjacking partly because of facial recognition technology.
- Police put a file photo of Woodruff in a photo lineup after gas station video from the scene was run through facial recognition technology. The carjacking victim picked Woodruff, who was among other women in the lineup.
- Why did the AI algorithms detect Porcha as a suspect? The AI system was unable to explain the reasons for the model results.

https://abcnews.go.com/US/wireStory/woman-wrongly-accused-carjacking-loses-lawsuit-detroit-police-125273580

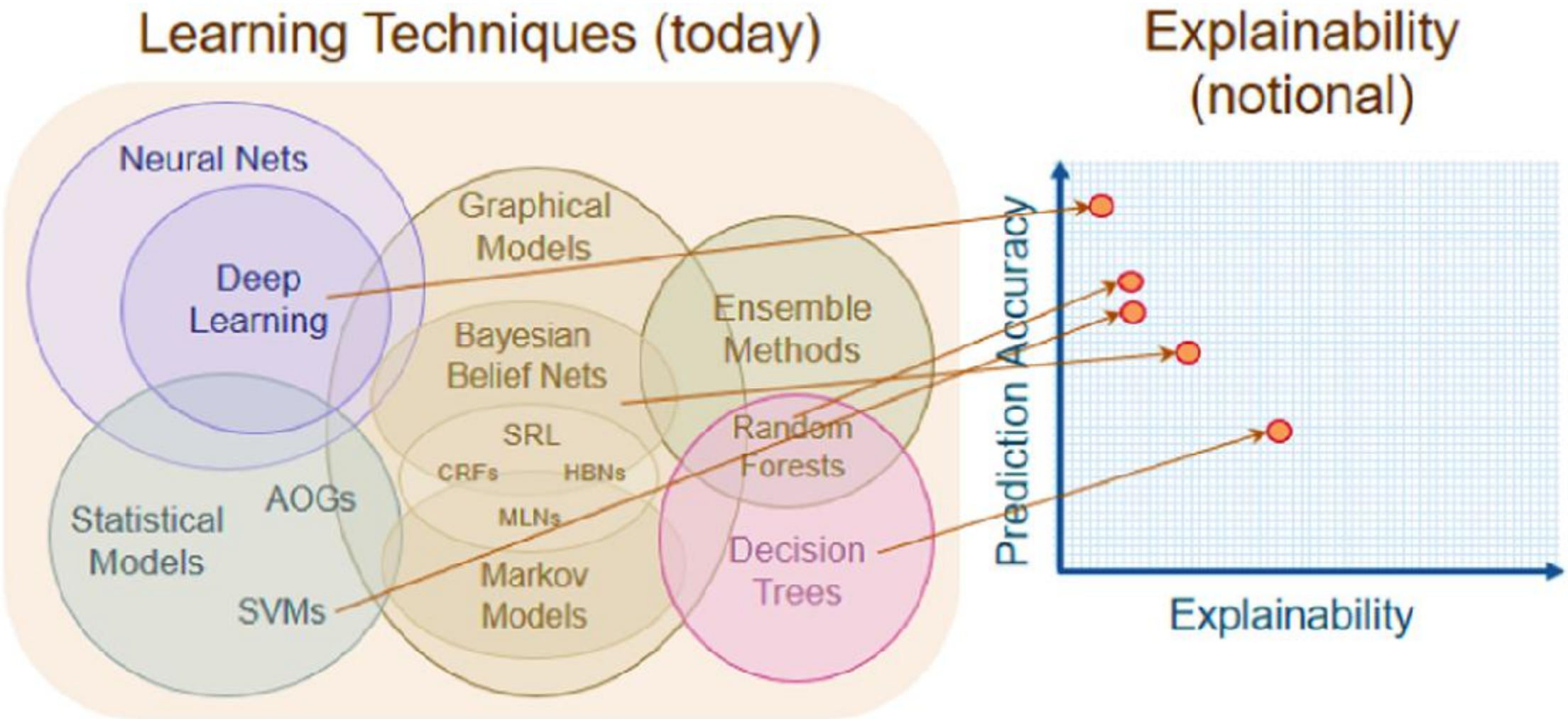
Discussion: AI shortlisting for job interview.

- Supposed that you are a HR manager. You received a lot of job applications for a post in your company. Some candidates are not selected for the interview and ask you why, as they think they fit the job requirements perfectly. Shortlisting for job interview is merely based on AI decisions which cannot explain the exact reasons.
- You also find that they are the excellent candidates not selected by the AI system. What should you do?

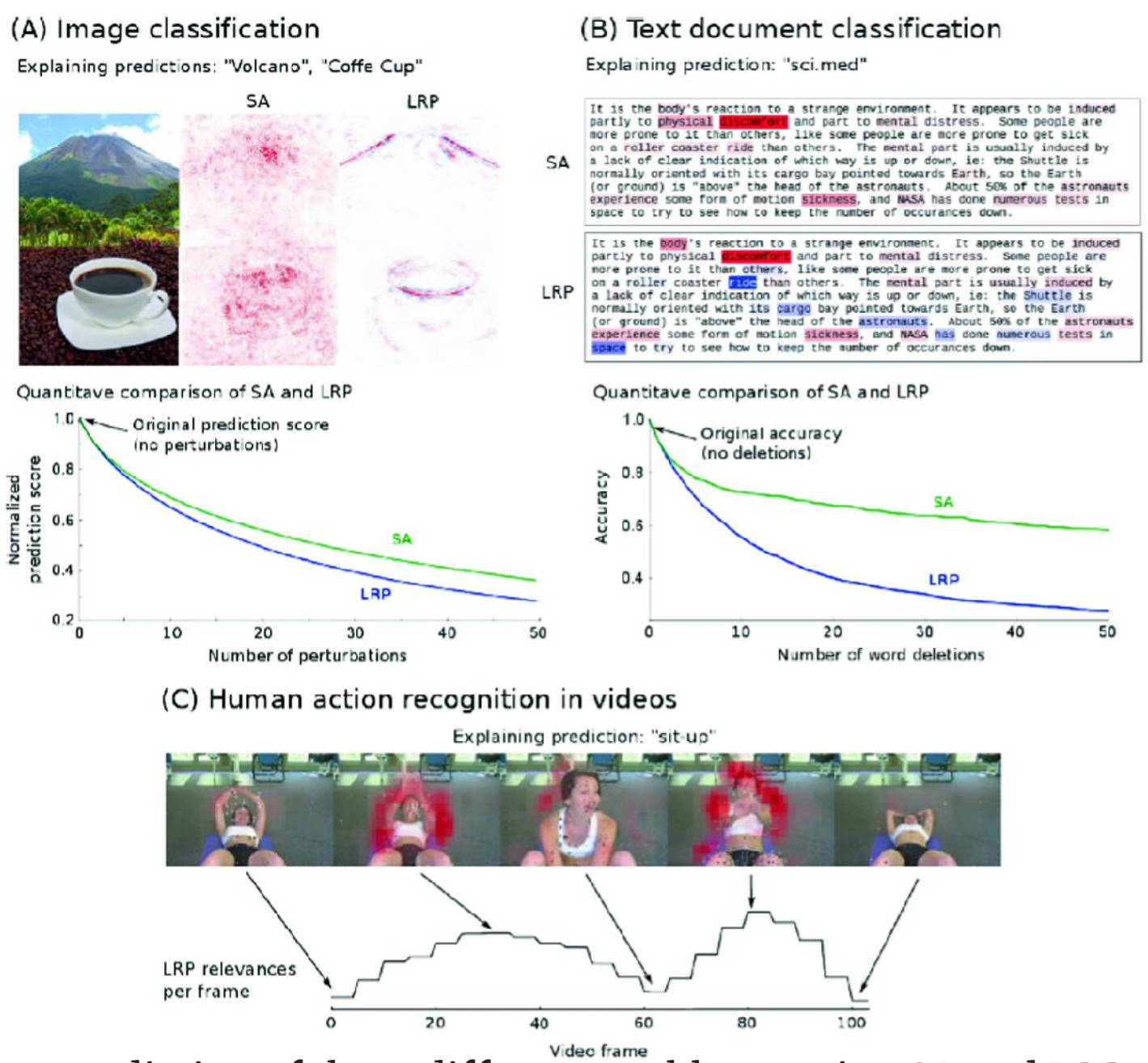
What is XAI?



(DARPA, 2017; Xu, et al. 2019)



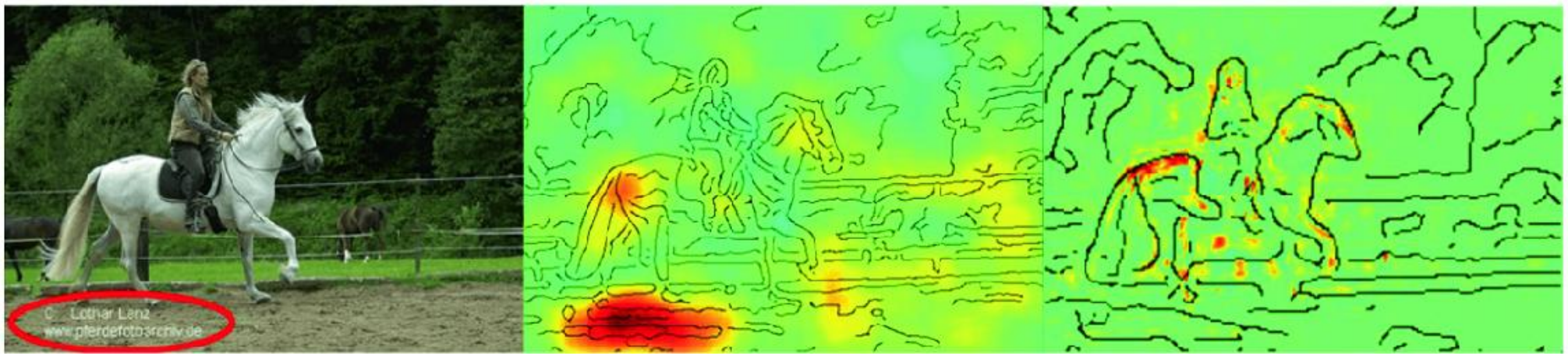
(DARPA, 2017; Xu, et al. 2019)



(Samek et al, 2017)

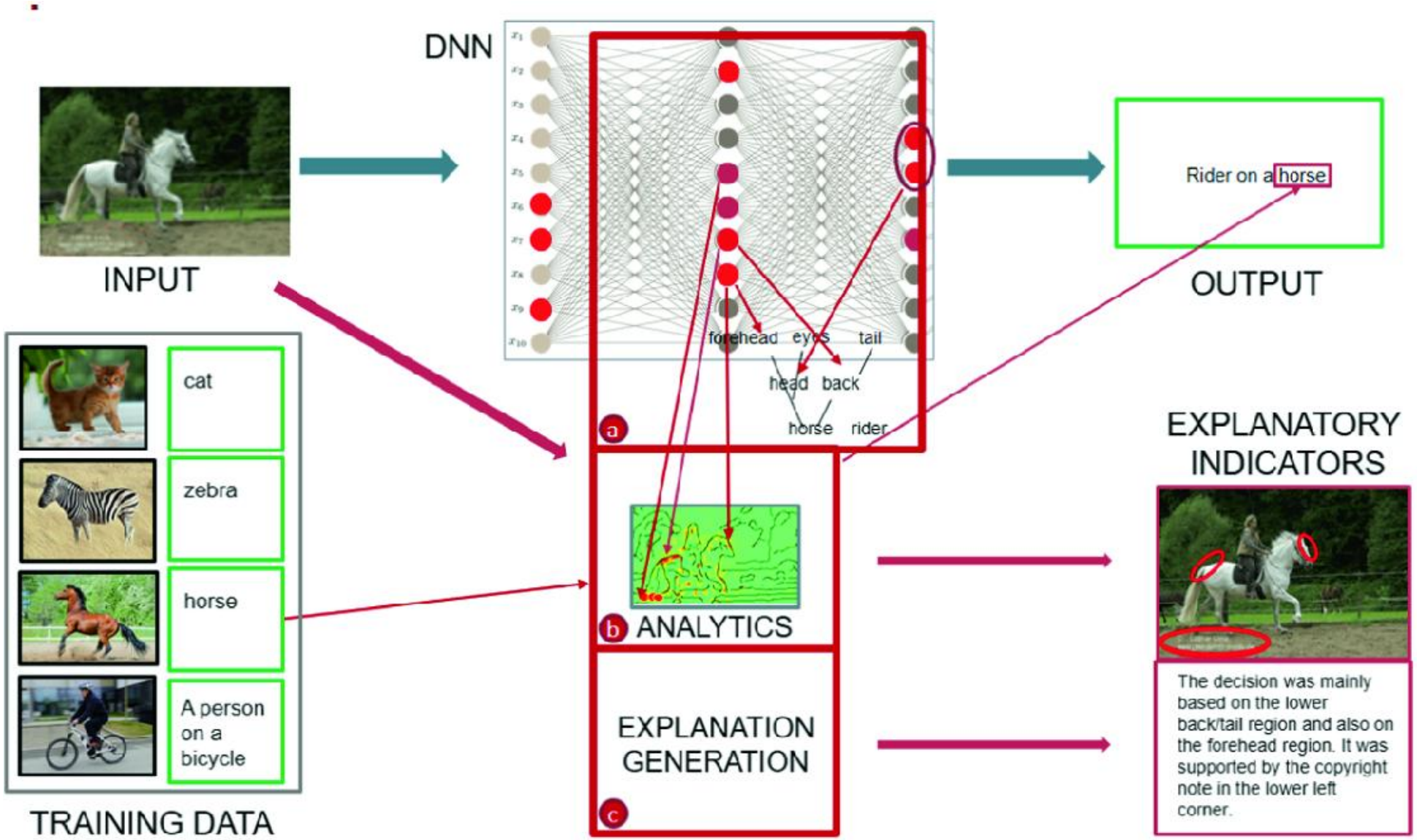
Explaining prediction of three different problems using SA and LRP.

	aeroplane	bicycle	bird	boat	bottle	bus	car
Fisher	79.08%	66.44%	45.90%	70.88%	27.64%	69.67%	80.96%
DeepNet	88.08%	79.69%	80.77%	77.20%	35.48%	72.71%	86.30%
	cat	chair	cow	diningtable	dog	horse	motorbike
Fisher	59.92%	51.92%	47.60%	58.06%	42.28%	80.45%	69.34%
DeepNet	81.10%	51.04%	61.10%	64.62%	76.17%	81.60%	79.33%
	person	pottedplant	sheep	sofa	train	tvmonitor	mAP
Fisher	85.10%	28.62%	49.58%	49.31%	82.71%	54.33%	59.99%
DeepNet	92.43%	49.99%	74.04%	49.48%	87.07%	67.08%	72.12%



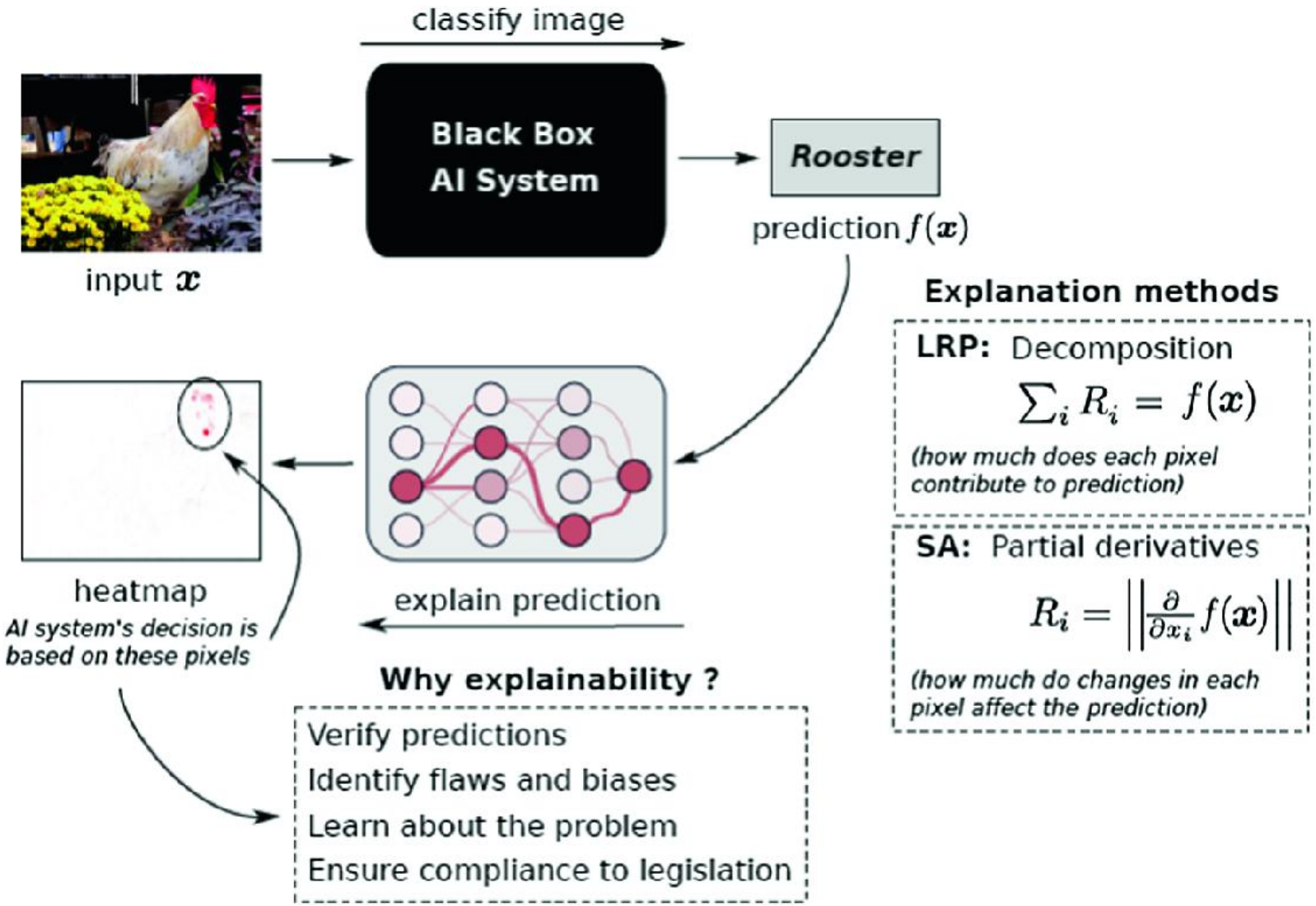
Upper: the prediction accuracy of Fisher Vector and Deep Neural Network in tasks of object recognition; Lower: model diagnosis using saliency map method.

(Lapuschkin et al. , 2016; Xu, et al. 2019)



Three approaches for understanding a neural network, indicated by red-boxes (a), (b) and (c)

(Xu, et al. 2019)



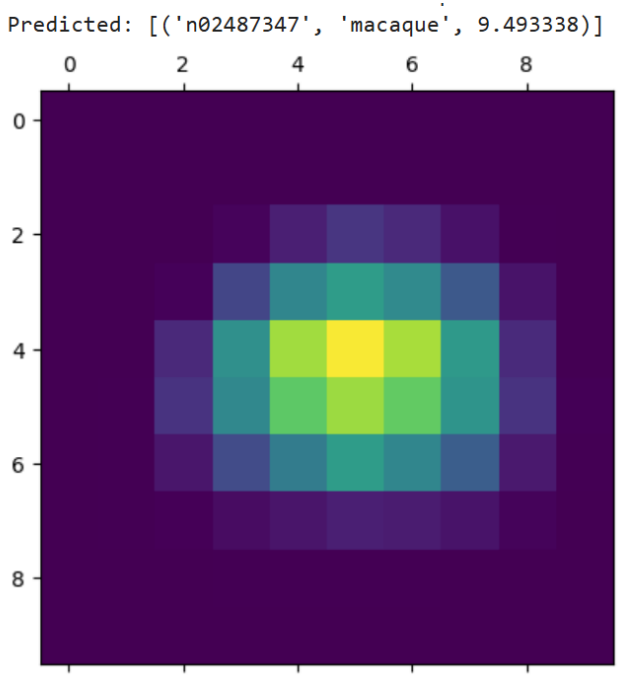
Explaining predictions of an AI system using SA (sensitivity analysis) and LRP (layer-wise relevance propagation) .

(Samek et al, 2017)

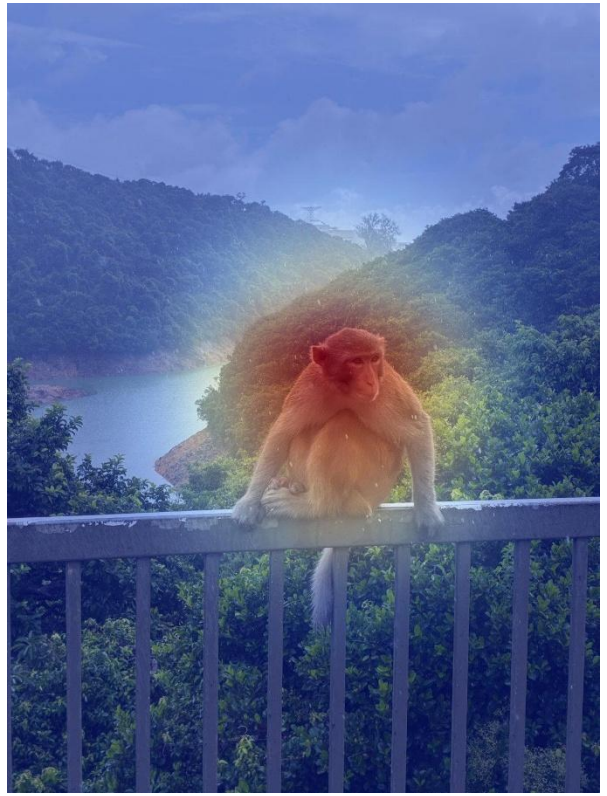
Lab: Gradcam for XAI

A HK monkey

+



=



https://keras.io/examples/vision/grad_cam/

My recent example

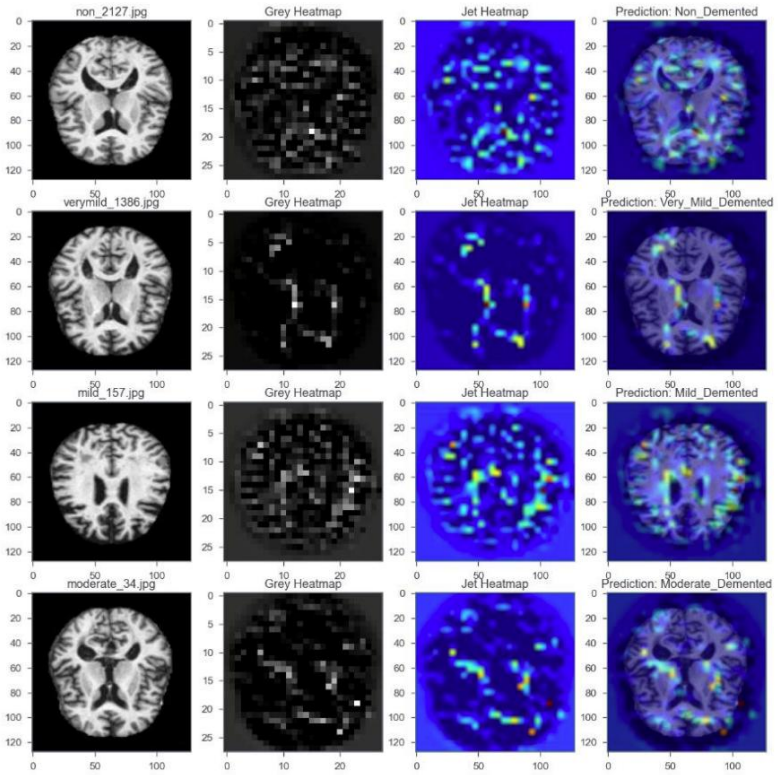


Figure 10. Grad-CAM images of correct classification

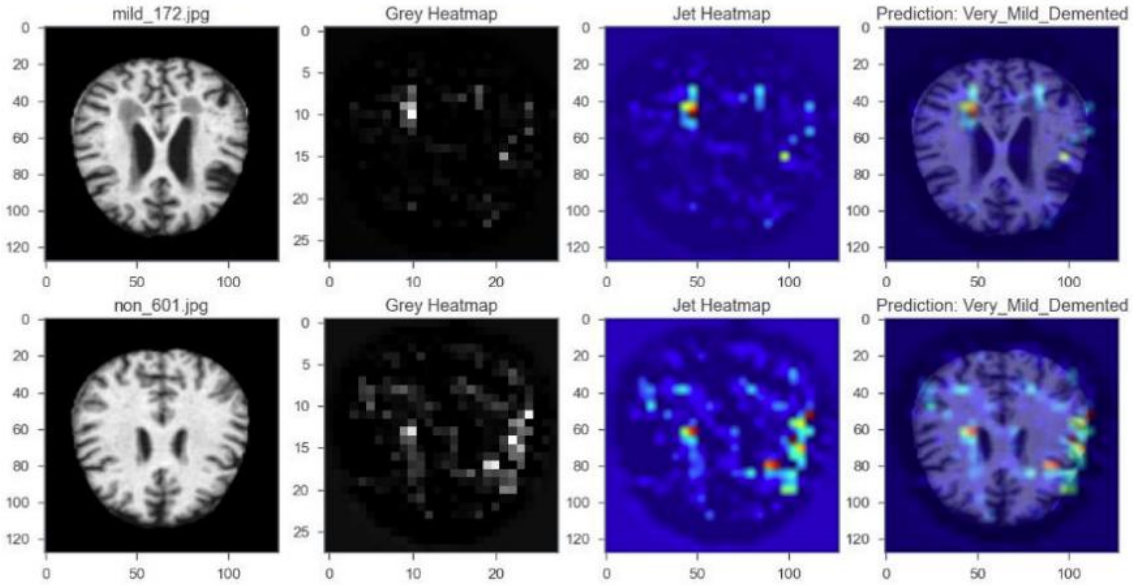


Figure 11. Grad-CAM images of incorrect classification

Yuen, K.K.F. (2025), A Tutorial on Explainable Image Classification for Dementia Stages Using Convolutional Neural Network and Gradient-weighted Class Activation Mapping, ***Studies in Health Technology and Informatics***, accepted, preprint https://arxiv.org/pdf/2408.10572

References

- Yuen, K.K.F. (2025), A Tutorial on Explainable Image Classification for Dementia Stages Using Convolutional Neural Network and Gradient-weighted Class Activation Mapping, ***Studies in Health Technology and Informatics***, accepted, preprint https://arxiv.org/pdf/2408.10572
- Russell, S. and Norvig, P., 2020, Artificial Intelligence - A Modern Approach, 4th edition, Pearson
- Luger, G.F., 2009, Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 6th edition, Addison-Wesley.
- Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J. (2019). Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In: Tang, J., Kan, MY., Zhao, D., Li, S., Zan, H. (eds) Natural Language Processing and Chinese Computing. NLPCC 2019. Lecture Notes in Computer Science(), vol 11839. Springer, Cham. https://doi.org/10.1007/978-3-030-32236-6_51
- Samek, W., Wiegand, T., Müller, K.R.: Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296 (2017)
- https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models
- Hsieh, W., Bi, Z., Jiang, C., Liu, J., Peng, B., Zhang, S., ... & Liang, C. X. (2024). A comprehensive guide to explainable ai: From classical models to llms. *arXiv preprint arXiv:2412.00800*.

Q&A
Thank you