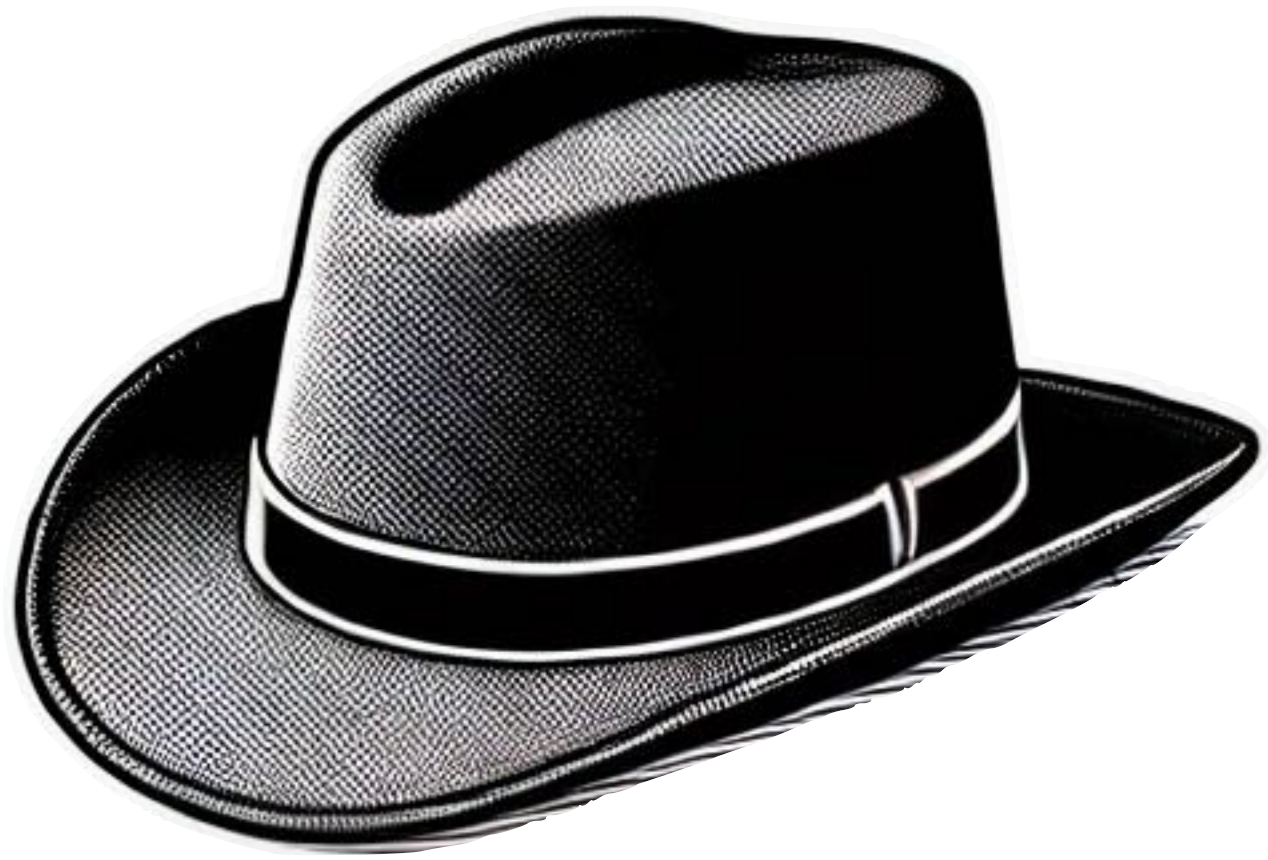
Introduction to Artificial Intelligence and Machine Learning

GRAHAM TAYLOR

SCHOOL OF ENGINEERING
CENTRE FOR ADVANCING RESPONSIBLE AND ETHICAL ARTIFICIAL INTELLIGENCE
UNIVERSITY OF GUELPH

VECTOR INSTITUTE FOR ARTIFICIAL INTELLIGENCE

CANADA CIFAR AI CHAIR

Lecture Outline

1.What is Artificial Intelligence?

2. What is Machine Learning?

3. Components of a ML system

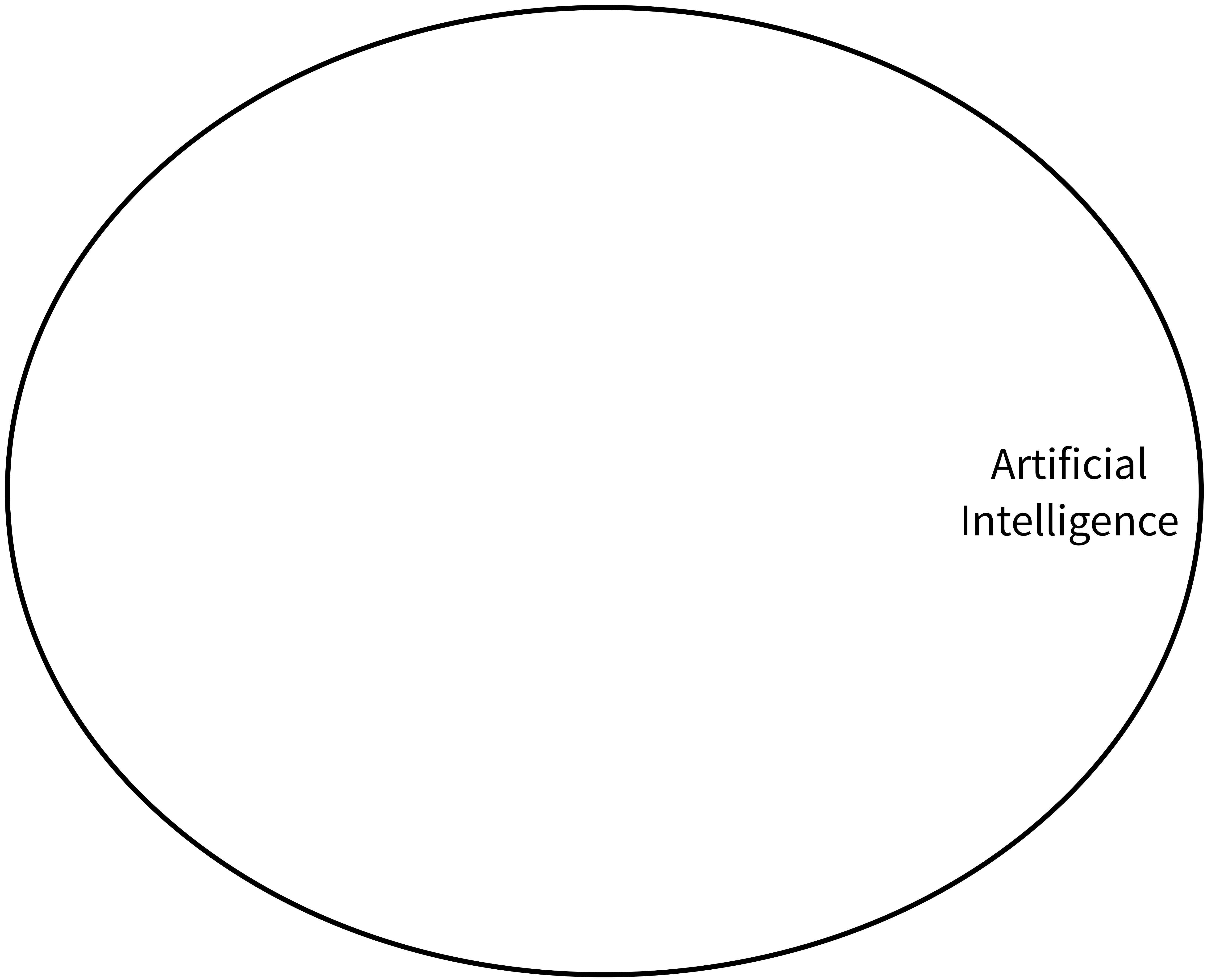
4. ML Systems vs. Traditional Software Systems

5. Machine Learning Tasks

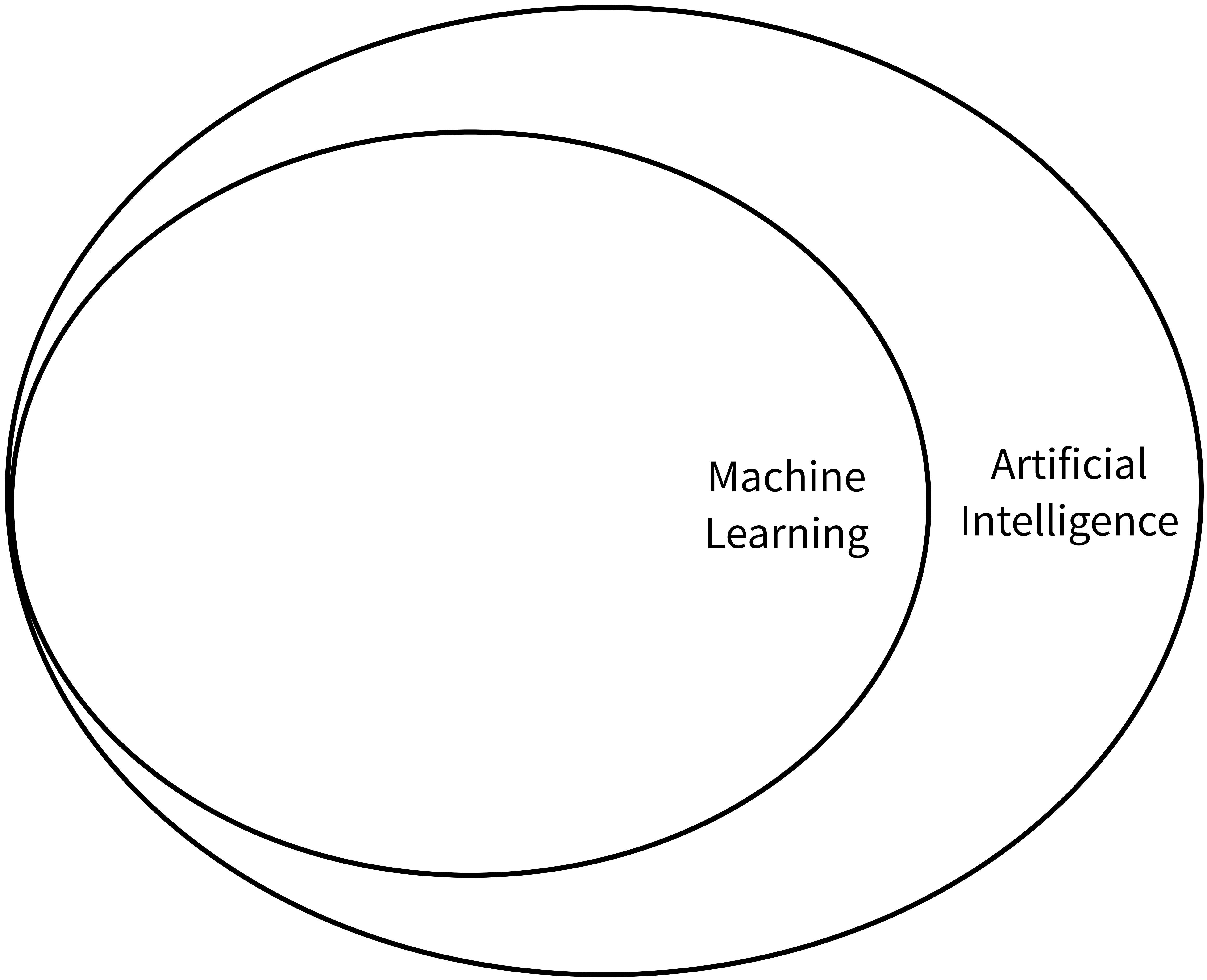
6. The Machine Learning Experience

7. Economics of AI

What is Artificial Intelligence?

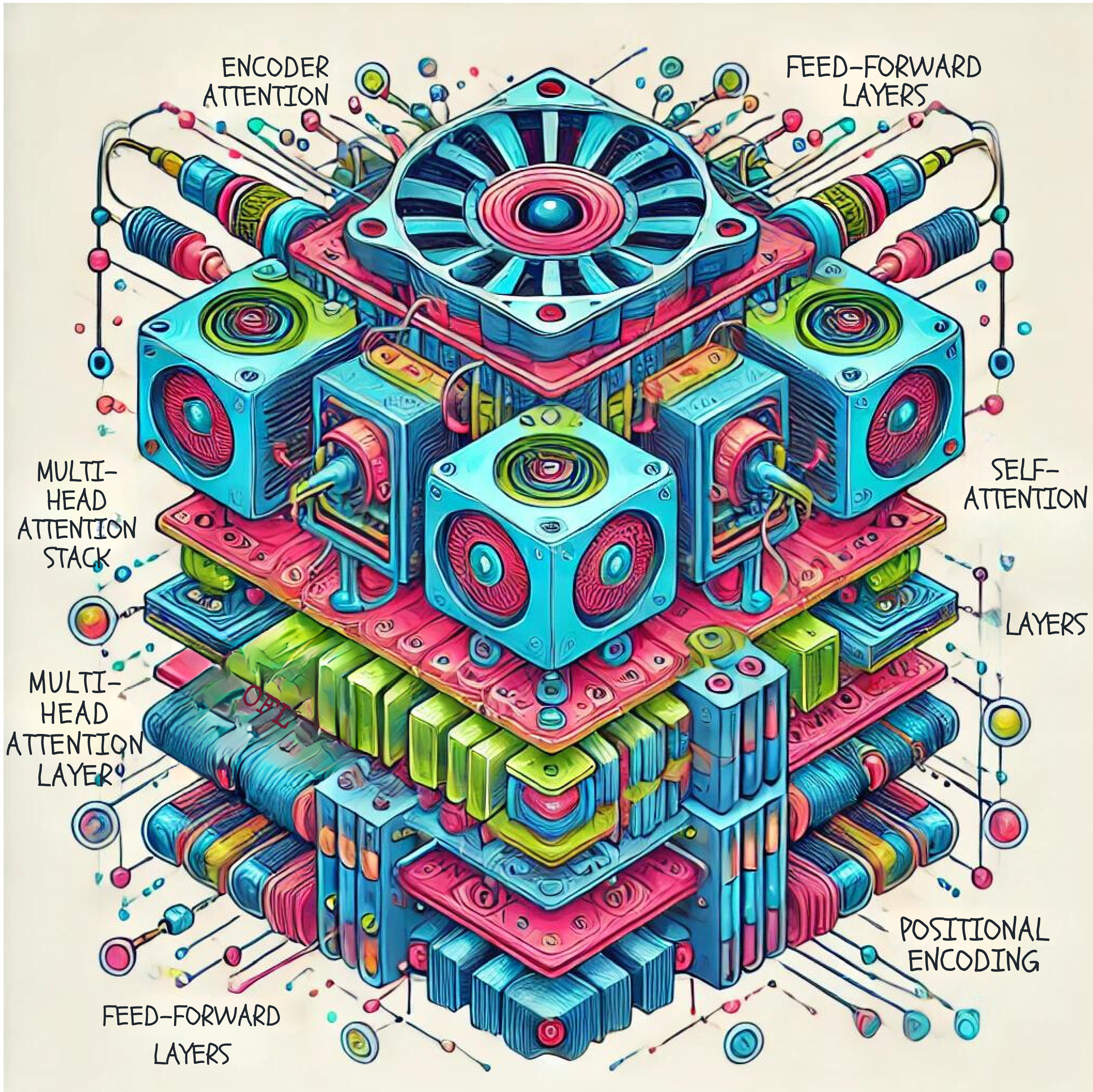


reproduced from Goodfellow et al.



reproduced from Goodfellow et al.

Explaining AI like a ML Professor



Explaining AI like an Economist

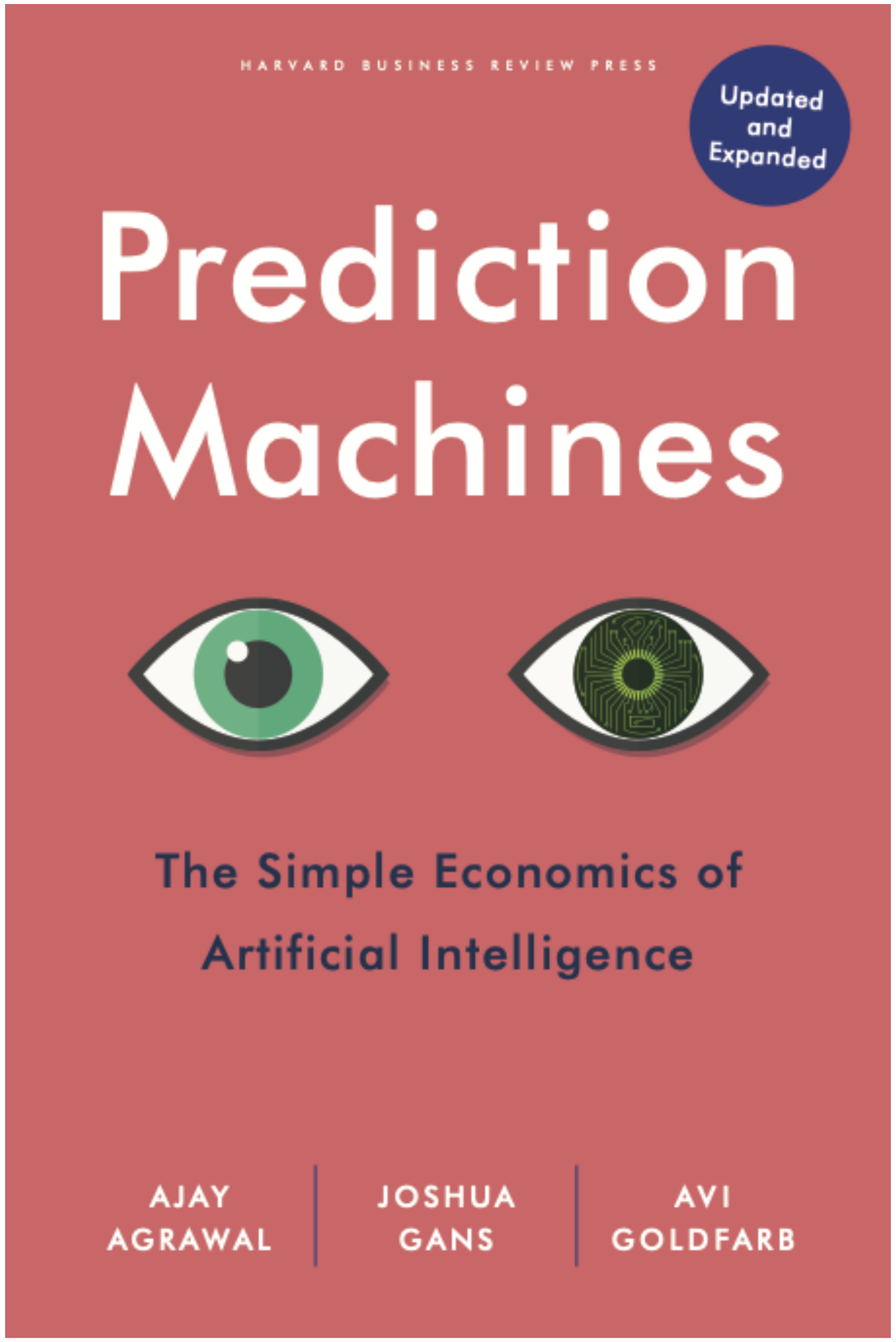
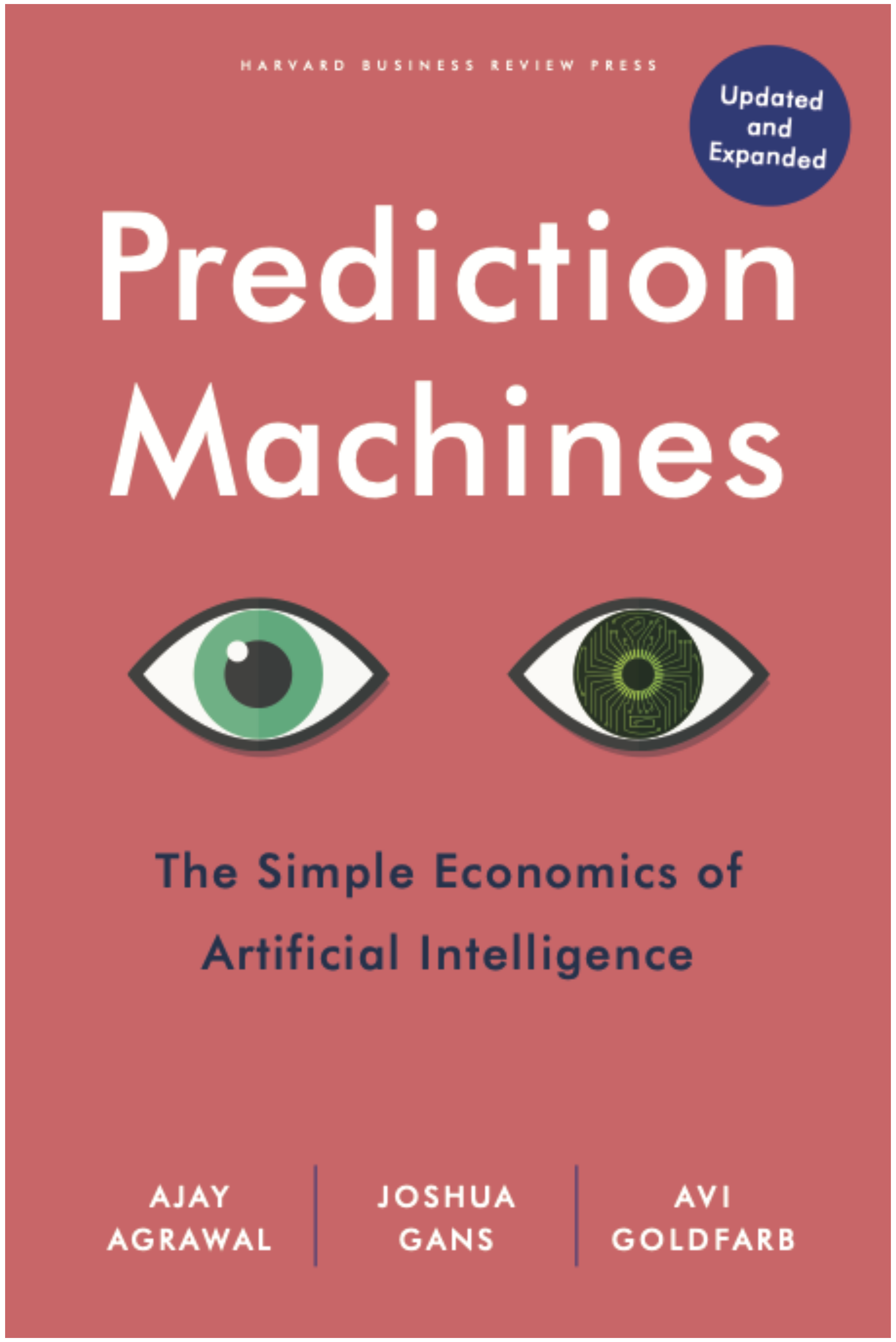


figure reproduced from Ajay Agrawal

Explaining AI like an Economist



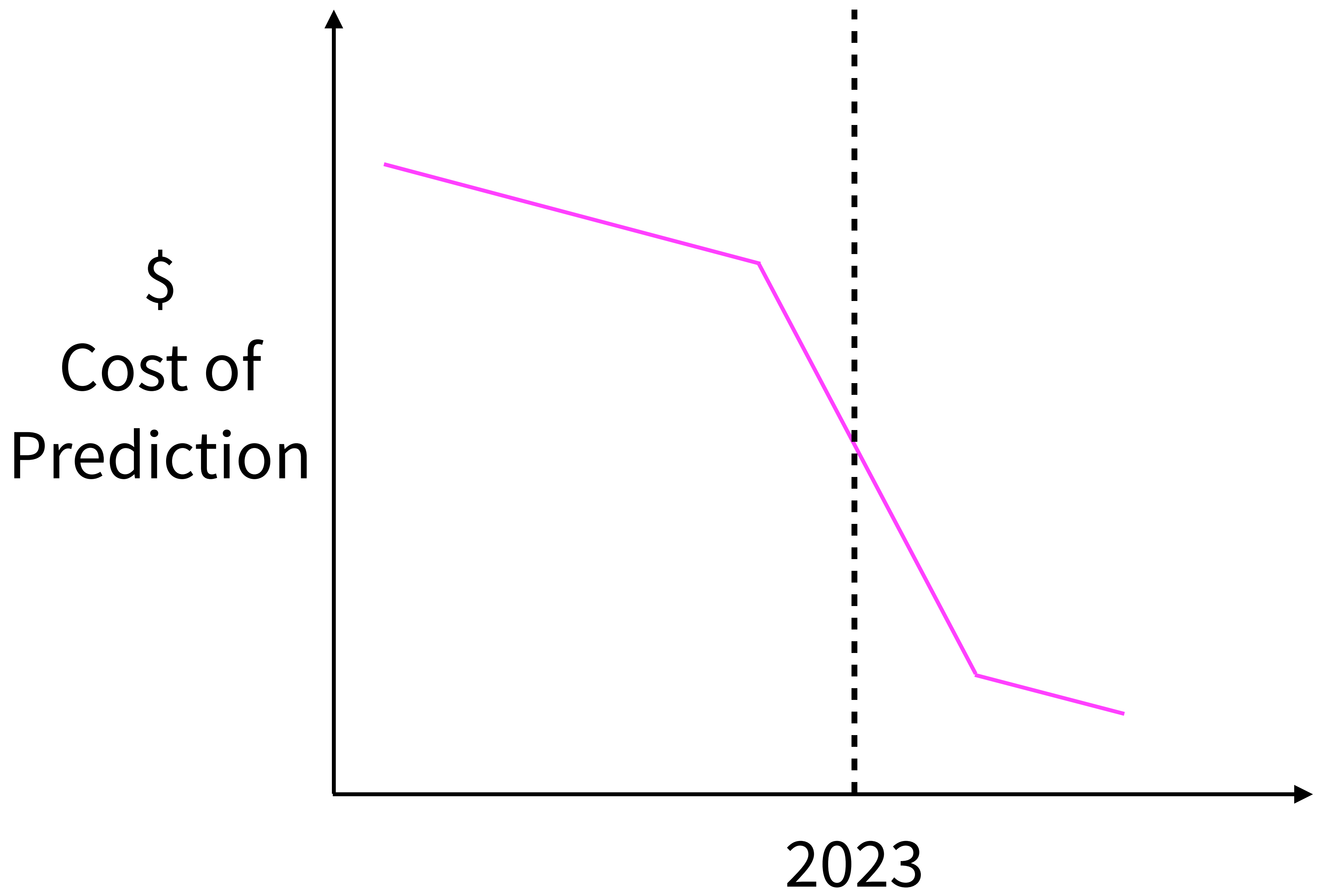
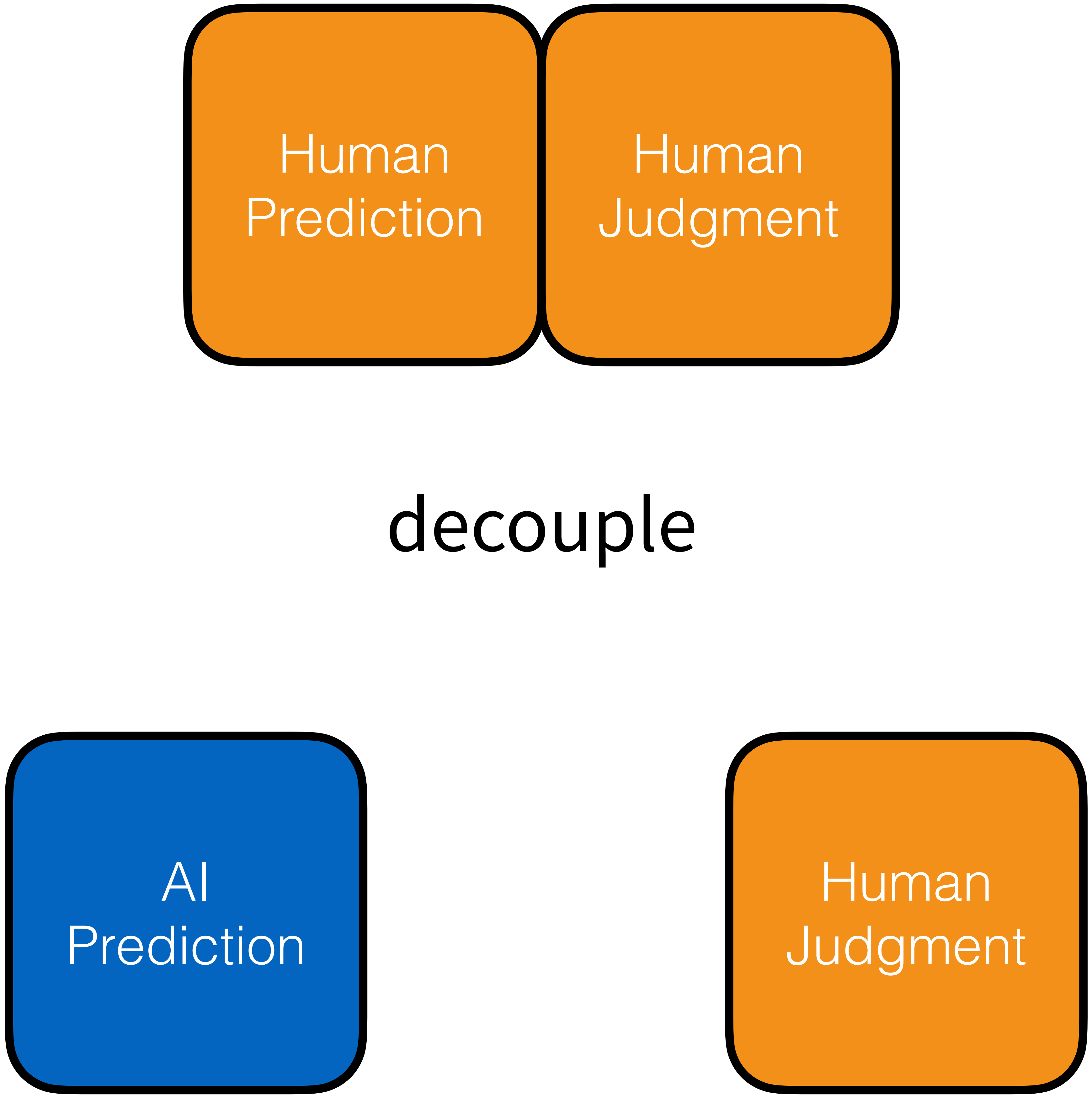
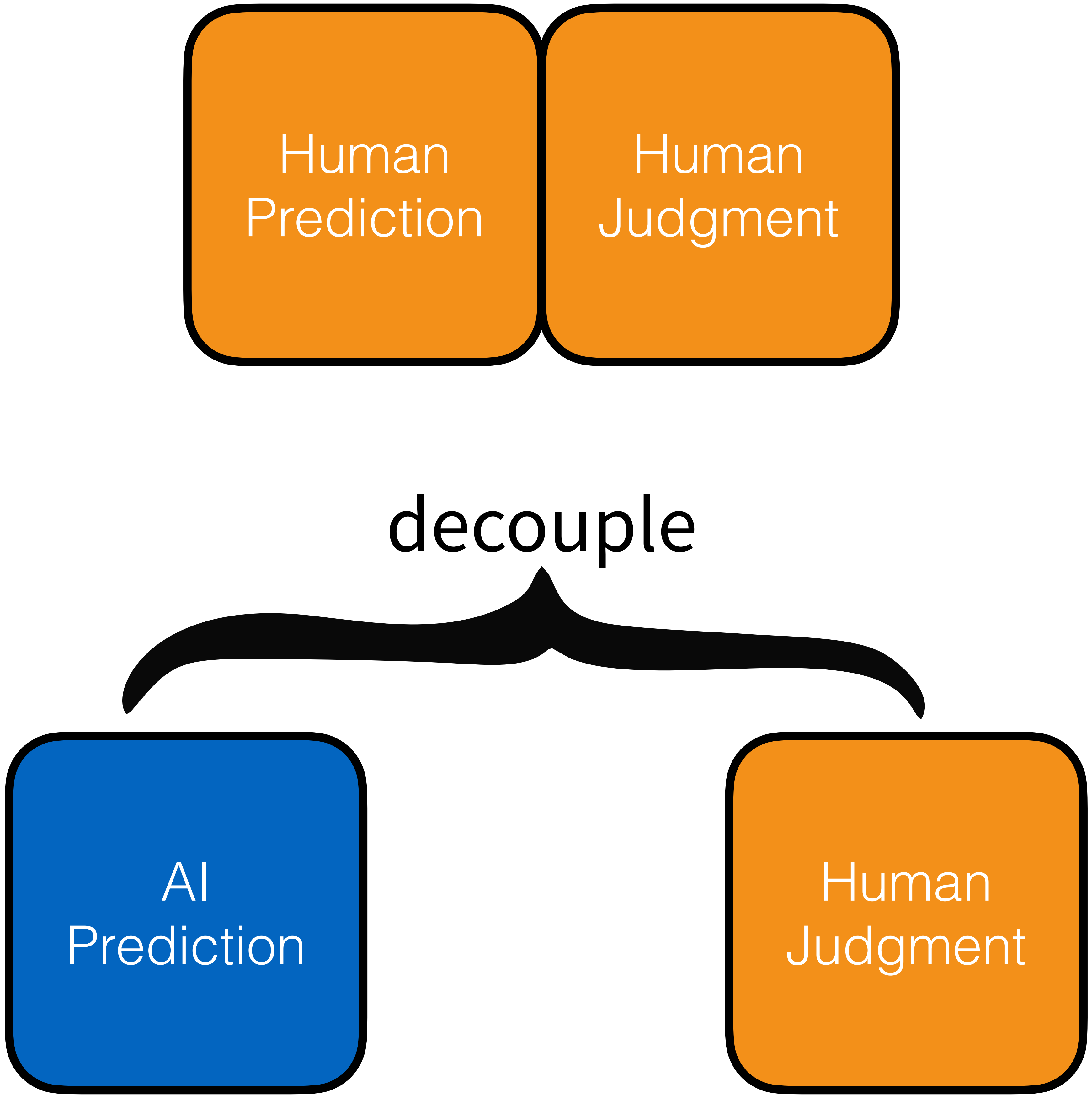


figure reproduced from Ajay Agrawal

Prediction is about using information you have to generate information you don't have.

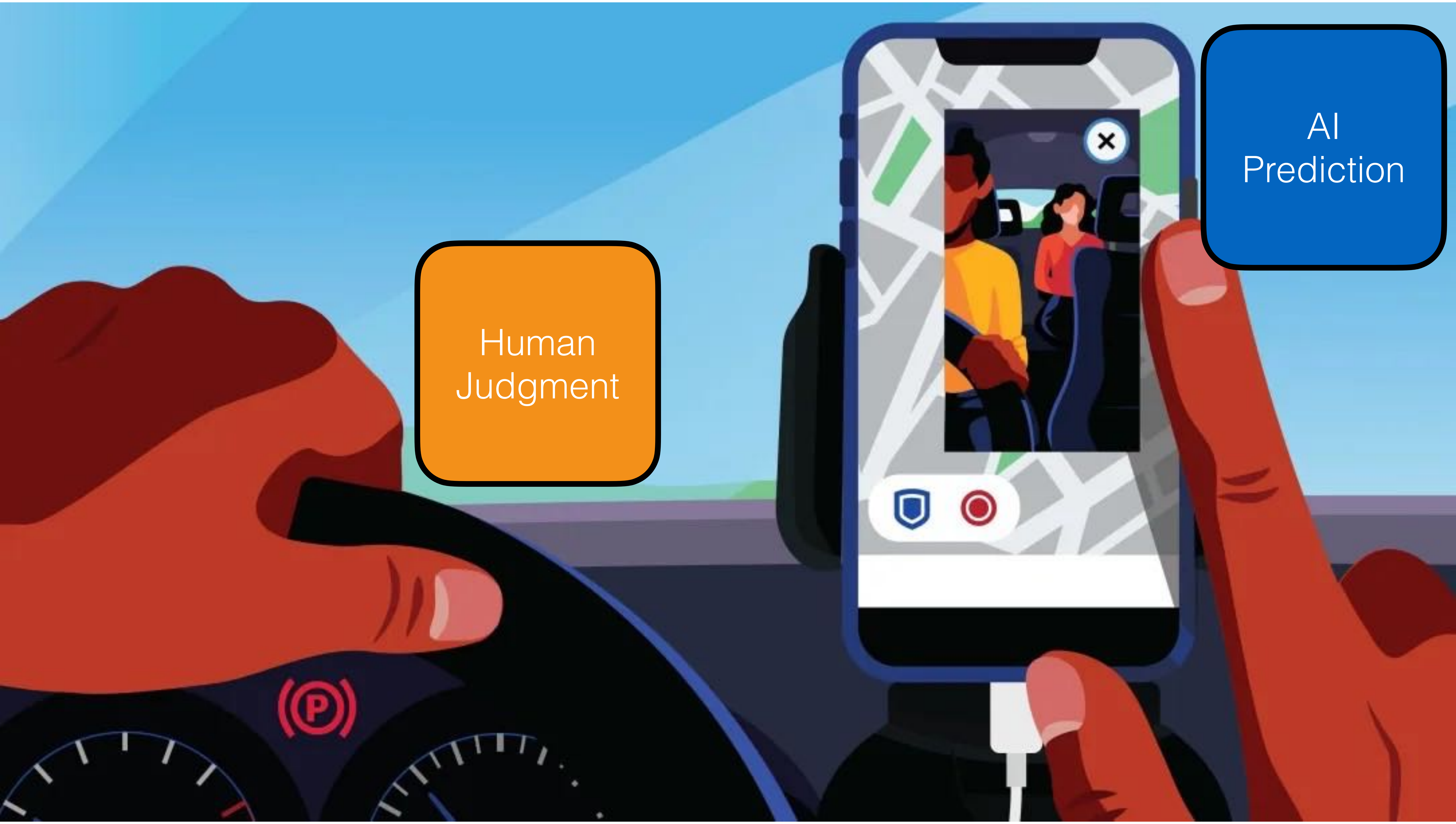


reproduced from Ajay Agrawal

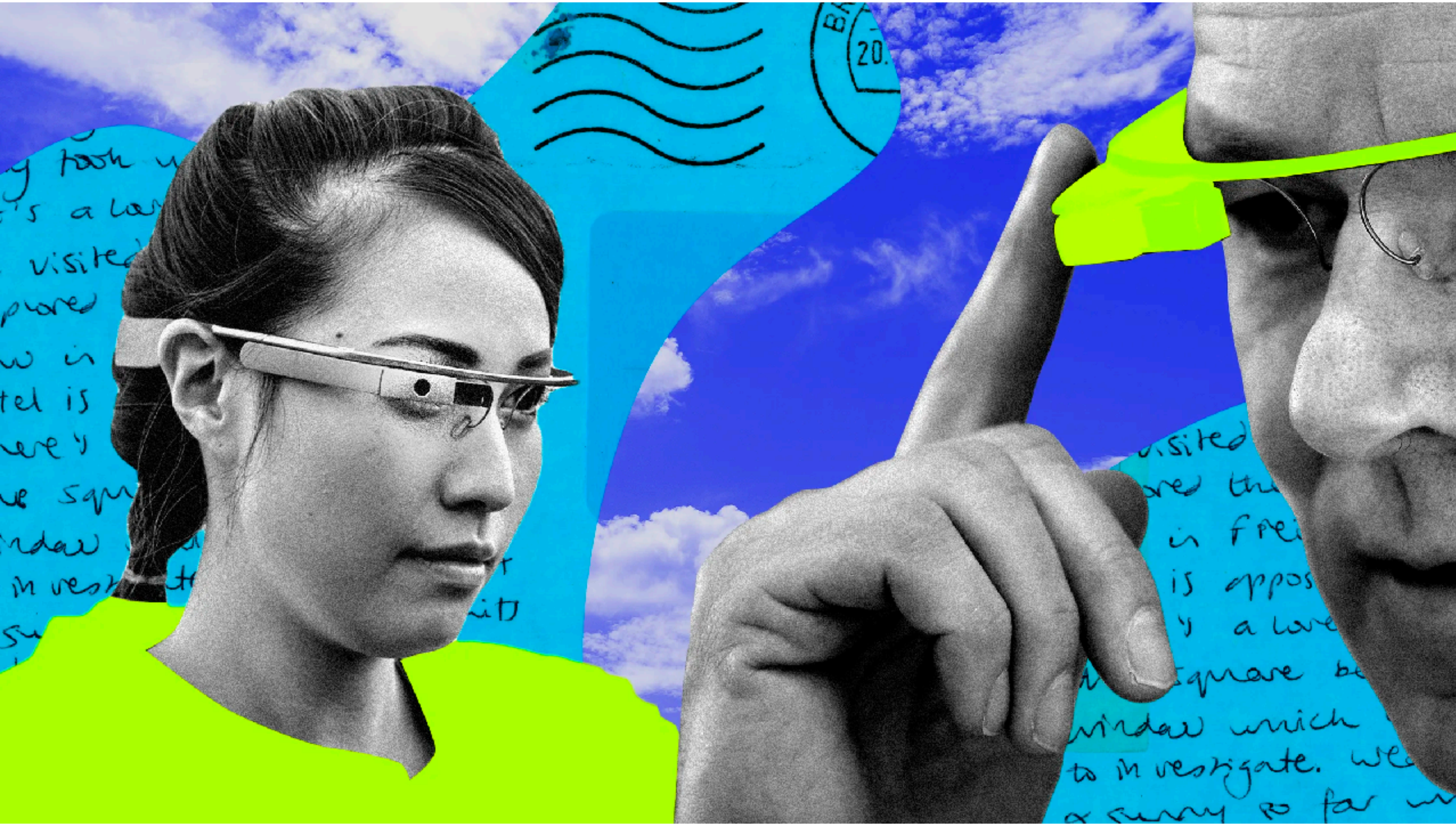


reproduced from Ajay Agrawal

via npr.org with idea from Joshua Gans

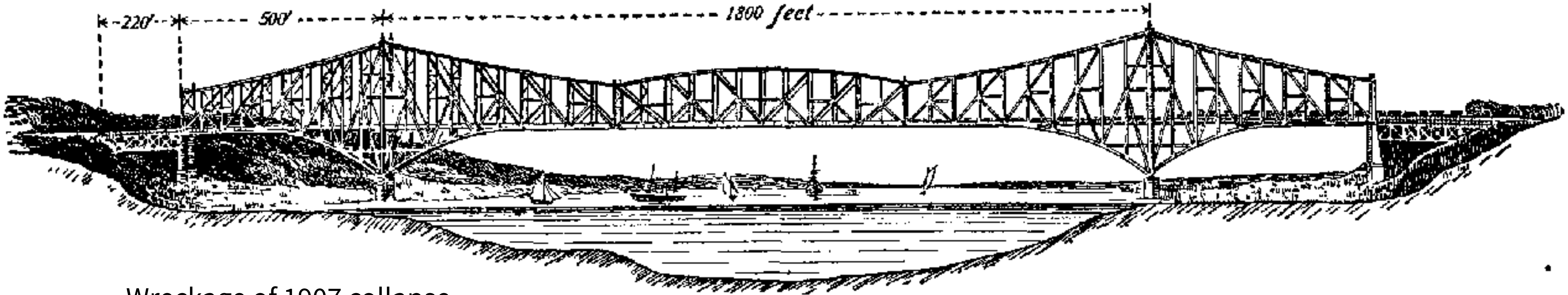


via Uber

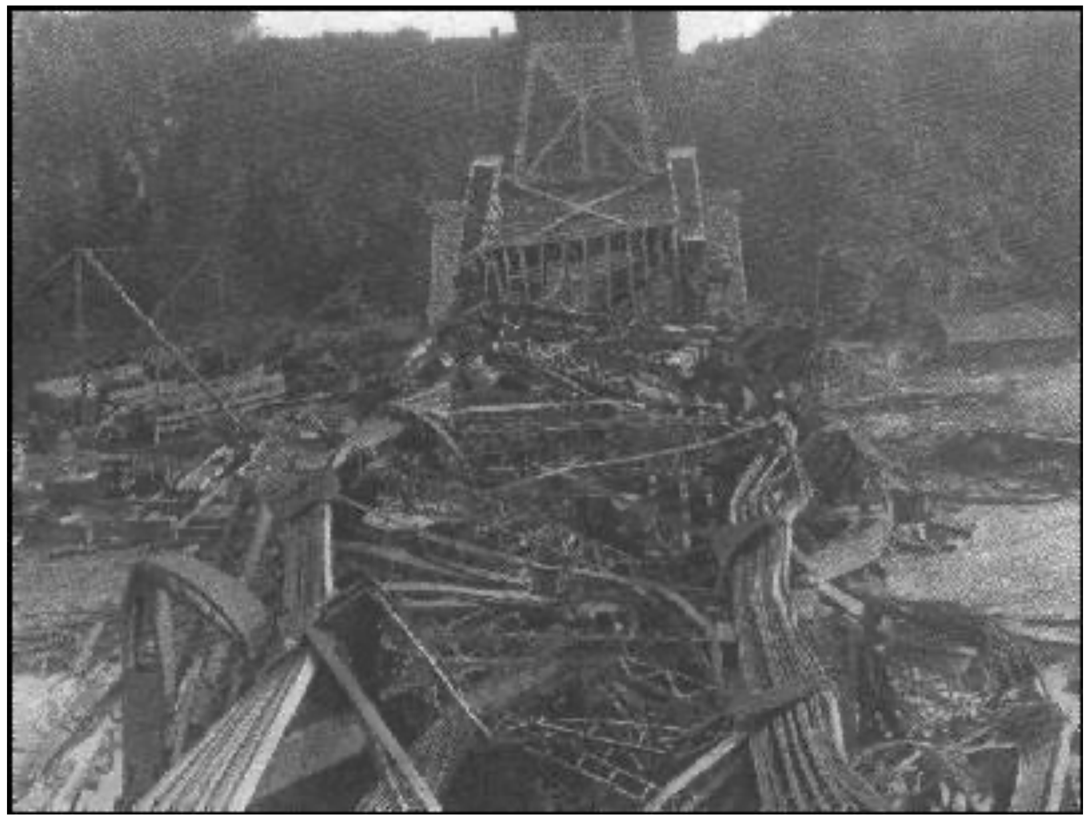


via Wired

Quebec Bridge between Sainte-Foy and Lévis, Quebec



Wreckage of 1907 collapse

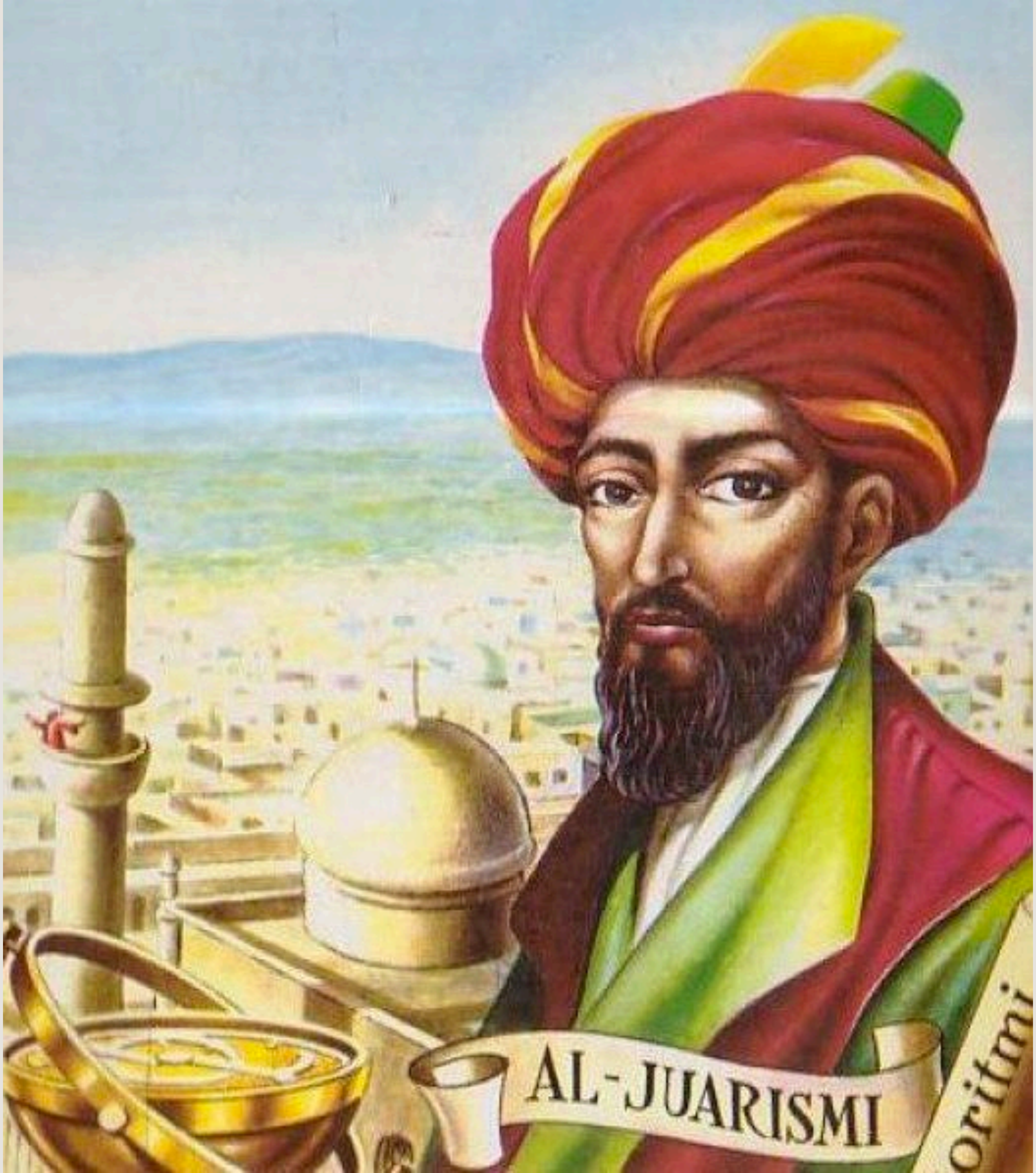


1916 collapse

“We’re seeing collapsing bridges in the machine learning space all the time now” — Deborah Raji

via Public Domain, commons.wikimedia.org

What is Machine Learning?



via Medium

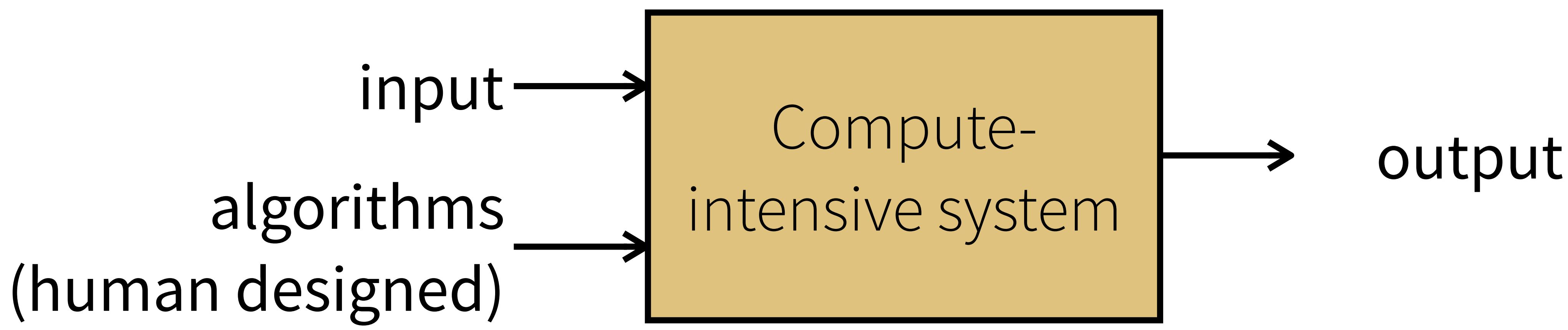
Diagram for the computation by the Engine of the Numbers of Bernoulli. See Note G. (page 722 *et seq.*)

Number of Operation.	Nature of Operation.	Variables acted upon.	Variables receiving results.	Indication of change in the value on any Variable.	Statement of Results.	Data.			Working Variables.										Result Variables.			
						$^1V_1$ ○ 0 0 0 1	$^1V_2$ ○ 0 0 0 2	$^1V_3$ ○ 0 0 0 4	$^0V_4$ ○ 0 0 0 0	$^0V_5$ ○ 0 0 0 0	$^0V_6$ ○ 0 0 0 0	$^0V_7$ ○ 0 0 0 0	$^0V_8$ ○ 0 0 0 0	$^0V_9$ ○ 0 0 0 0	$^0V_{10}$ ○ 0 0 0 0	$^0V_{11}$ ○ 0 0 0 0	$^0V_{12}$ ○ 0 0 0 0	$^0V_{13}$ .... ○ 0 0 0 0	$^1V_{21}$ ○ B_1 in a decimal fraction. 0 0 0	$^1V_{22}$ ○ B_3 in a decimal fraction. 0 0 0	$^1V_{23}$ ○ B_5 in a decimal fraction. 0 0 0	$^0V_{24}$... ○ 0 0 0 0
						$\boxed{1}$	$\boxed{2}$	$\boxed{n}$											$\boxed{B_1}$	$\boxed{B_3}$	$\boxed{B_5}$	$\boxed{B_7}$
1	$\times$	$^1V_2 \times {}^1V_3$	$^1V_4, {}^1V_5, {}^1V_6$	$\left\{\begin{matrix} ^1V_2 = {}^1V_2 \\ ^1V_3 = {}^1V_3 \end{matrix}\right\}$	$= 2n$	...	2	n	$2n$	$2n$	$2n$											
2	$-$	$^1V_4 - {}^1V_1$	$^2V_4$	$\left\{\begin{matrix} ^1V_4 = {}^2V_4 \\ ^1V_1 = {}^1V_1 \end{matrix}\right\}$	$= 2n - 1$	1	...	...	$2n-1$													
3	$+$	$^1V_5 + {}^1V_1$	$^2V_5$	$\left\{\begin{matrix} ^1V_5 = {}^2V_5 \\ ^1V_1 = {}^1V_1 \end{matrix}\right\}$	$= 2n + 1$	1	...	...	...	$2n+1$												
4	$\div$	$^2V_5 \div {}^2V_4$	$^1V_{11}$	$\left\{\begin{matrix} ^2V_5 = {}^0V_5 \\ ^2V_4 = {}^0V_4 \end{matrix}\right\}$	$= \frac{2n-1}{2n+1}$	...	...	...	0	0	...	...	...	...	...	$\frac{2n-1}{2n+1}$						
5	$\div$	$^1V_{11} \div {}^1V_2$	$^2V_{11}$	$\left\{\begin{matrix} ^1V_{11} = {}^2V_{11} \\ ^1V_2 = {}^1V_2 \end{matrix}\right\}$	$= \frac{1}{2} \cdot \frac{2n-1}{2n+1}$	...	2	...	...	...	...	...	...	...	...	$\frac{1}{2} \cdot \frac{2n-1}{2n+1}$						
6	$-$	$^0V_{13} - {}^2V_{11}$	$^1V_{13}$	$\left\{\begin{matrix} ^2V_{11} = {}^0V_{11} \\ ^0V_{13} = {}^1V_{13} \end{matrix}\right\}$	$= -\frac{1}{2} \cdot \frac{2n-1}{2n+1} = A_0$	...	...	...	...	...	...	...	...	...	...	0	....	$-\frac{1}{2} \cdot \frac{2n-1}{2n+1} = A_0$				
7	$-$	$^1V_3 - {}^1V_1$	$^1V_{10}$	$\left\{\begin{matrix} ^1V_3 = {}^1V_3 \\ ^1V_1 = {}^1V_1 \end{matrix}\right\}$	$= n - 1 \ (= 3)$	1	...	n	...	...	...	...	...	...	$n-1$							
8	$+$	$^1V_2 + {}^0V_7$	$^1V_7$	$\left\{\begin{matrix} ^1V_2 = {}^1V_2 \\ ^0V_7 = {}^1V_7 \end{matrix}\right\}$	$= 2 + 0 = 2$	...	2	...	...	...	...	2										
9	$\div$	$^1V_6 \div {}^1V_7$	$^3V_{11}$	$\left\{\begin{matrix} ^1V_6 = {}^1V_6 \\ ^0V_{11} = {}^3V_{11} \end{matrix}\right\}$	$= \frac{2n}{2} = A_1$	...	...	...	...	...	$2n$	2	...	...	...	$\frac{2n}{2} = A_1$						
10	$\times$	$^1V_{21} \times {}^3V_{11}$	$^1V_{12}$	$\left\{\begin{matrix} ^1V_{21} = {}^1V_{21} \\ ^3V_{11} = {}^3V_{11} \end{matrix}\right\}$	$= B_1 \cdot \frac{2n}{2} = B_1 A_1$	...	...	...	...	...	...	...	...	...	...	$\frac{2n}{2} = A_1$	$B_1 \cdot \frac{2n}{2} = B_1 A_1$	....	B_1			
11	$+$	$^1V_{12} + {}^1V_{13}$	$^2V_{13}$	$\left\{\begin{matrix} ^1V_{12} = {}^0V_{12} \\ ^1V_{13} = {}^2V_{13} \end{matrix}\right\}$	$= -\frac{1}{2} \cdot \frac{2n-1}{2n+1} + B_1 \cdot \frac{2n}{2}$ ...	...	...	...	...	...	...	...	...	...	...	....	0	$\left\{-\frac{1}{2} \cdot \frac{2n-1}{2n+1} + B_1 \cdot \frac{2n}{2}\right\}$				
12	$-$	$^1V_{10} - {}^1V_1$	$^2V_{10}$	$\left\{\begin{matrix} ^1V_{10} = {}^2V_{10} \\ ^1V_1 = {}^1V_1 \end{matrix}\right\}$	$= n - 2 \ (= 2)$	1	...	...	...	...	...	...	...	...	$n-2$							
13	$-$	$^1V_6 - {}^1V_1$	$^2V_6$	$\left\{\begin{matrix} ^1V_6 = {}^2V_6 \\ ^1V_1 = {}^1V_1 \end{matrix}\right\}$	$= 2n - 1$	1	...	...	...	...	$2n-1$											
14	$+$	$^1V_1 + {}^1V_7$	$^2V_7$	$\left\{\begin{matrix} ^1V_1 = {}^1V_1 \\ ^1V_7 = {}^2V_7 \end{matrix}\right\}$	$= 2 + 1 = 3$	1	...	...	...	...	...	3										
15	$\div$	$^2V_6 \div {}^2V_7$	$^1V_8$	$\left\{\begin{matrix} ^2V_6 = {}^2V_6 \\ ^2V_7 = {}^2V_7 \end{matrix}\right\}$	$= \frac{2n-1}{3}$	...	...	...	...	...	$2n-1$	3	$\frac{2n-1}{3}$									
16	$\times$	$^1V_8 \times {}^3V_{11}$	$^4V_{11}$	$\left\{\begin{matrix} ^1V_8 = {}^0V_8 \\ ^3V_{11} = {}^4V_{11} \end{matrix}\right\}$	$= \frac{2n}{2} \cdot \frac{2n-1}{3}$	...	...	...	...	...	...	...	0	...	...	$\frac{2n}{2} \cdot \frac{2n-1}{3}$						
17	$-$	$^2V_6 - {}^1V_1$	$^3V_6$	$\left\{\begin{matrix} ^2V_6 = {}^3V_6 \\ ^1V_1 = {}^1V_1 \end{matrix}\right\}$	$= 2n - 2$	1	...	...	...	...	$2n-2$											
18	$+$	$^1V_1 + {}^2V_7$	$^3V_7$	$\left\{\begin{matrix} ^2V_7 = {}^3V_7 \\ ^1V_1 = {}^1V_1 \end{matrix}\right\}$	$= 3 + 1 = 4$	1	...	...	...	...	...	4										
19	$\div$	$^3V_6 \div {}^3V_7$	$^1V_9$	$\left\{\begin{matrix} ^3V_6 = {}^3V_6 \\ ^3V_7 = {}^3V_7 \end{matrix}\right\}$	$= \frac{2n-2}{4}$	...	...	...	...	...	$2n-2$	4	...	$\frac{2n-2}{4}$	...	$\left\{\frac{2n}{2} \cdot \frac{2n-1}{3} \cdot \frac{2n-2}{3} = A_3\right\}$						
20	$\times$	$^1V_9 \times {}^4V_{11}$	$^5V_{11}$	$\left\{\begin{matrix} ^1V_9 = {}^0V_9 \\ ^4V_{11} = {}^5V_{11} \end{matrix}\right\}$	$= \frac{2n}{2} \cdot \frac{2n-1}{3} \cdot \frac{2n-2}{4} = A_3$	...	...	...	...	...	...	...	...	0								
21	$\times$	$^1V_{22} \times {}^5V_{11}$	$^0V_{12}$	$\left\{\begin{matrix} ^1V_{22} = {}^1V_{22} \\ ^0V_{12} = {}^2V_{12} \end{matrix}\right\}$	$= B_3 \cdot \frac{2n}{2} \cdot \frac{2n-1}{3} \cdot \frac{2n-2}{3} = B_3 A_3$	...	...	...	...	...	...	...	...	...	...	0	$B_3 A_3$	....	....	B_3		
22	$+$	$^2V_{12} + {}^2V_{13}$	$^3V_{13}$	$\left\{\begin{matrix} ^2V_{12} = {}^0V_{12} \\ ^2V_{13} = {}^3V_{13} \end{matrix}\right\}$	$= A_0 + B_1 A_1 + B_3 A_3$	...	...	...	...	...	...	...	...	...	...	....	0	$\left\{A_3 + B_1 A_1 + B_3 A_3\right\}$				
23	$-$	$^2V_{10} - {}^1V_1$	$^3V_{10}$	$\left\{\begin{matrix} ^2V_{10} = {}^3V_{10} \\ ^1V_1 = {}^1V_1 \end{matrix}\right\}$	$= n - 3 \ (= 1)$	1	...	...	...	...	...	...	...	...	$n-3$							
					Here follows a repetition of Operations thirteen to twenty-three.																	
24	$+$	$^4V_{13} + {}^0V_{24}$	$^1V_{24}$	$\left\{\begin{matrix} ^4V_{13} = {}^0V_{13} \\ ^0V_{24} = {}^1V_{24} \end{matrix}\right\}$	$= B_7$	...	...	...	...	...	...	...	...	...	...	....	....	....	......	......	......	B_7
25	$+$	$^1V_1 + {}^1V_3$	$^1V_3$	$\left\{\begin{matrix} ^1V_1 = {}^1V_1 \\ ^1V_3 = {}^1V_3 \\ ^5V_6 = {}^0V_6 \\ ^5V_7 = {}^0V_7 \end{matrix}\right\}$	$= n + 1 = 4 + 1 = 5$ .... by a Variable-card. by a Variable card.	1	...	$n+1$	...	...	0	0										

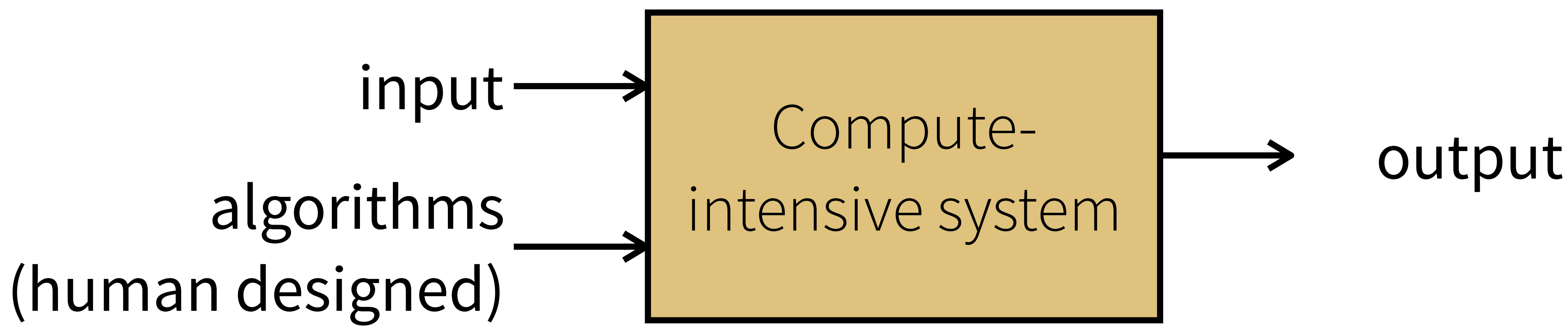
“The algorithm was developed to automate thinking, to remove difficult decisions from the hands of humans, and to solve contentious debates”

Franklin Foer, World Without Mind

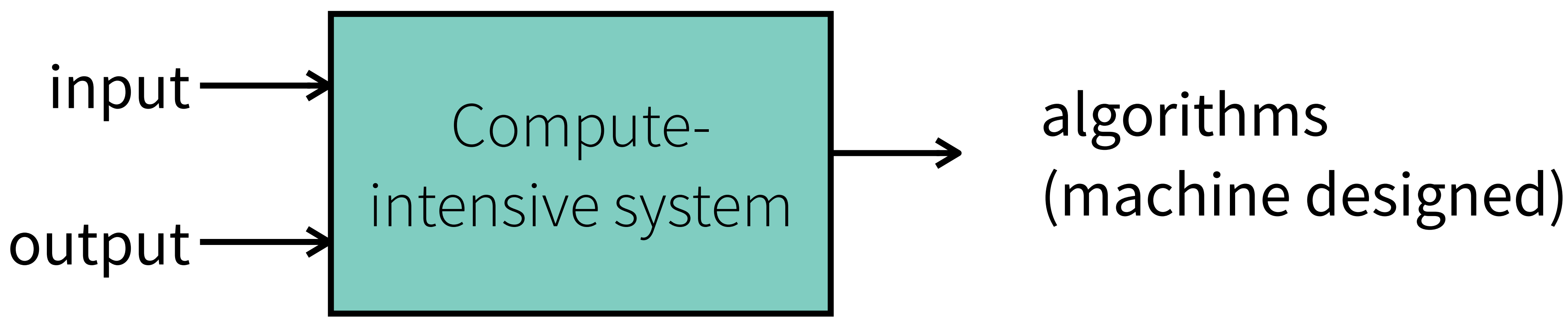
Traditional programming



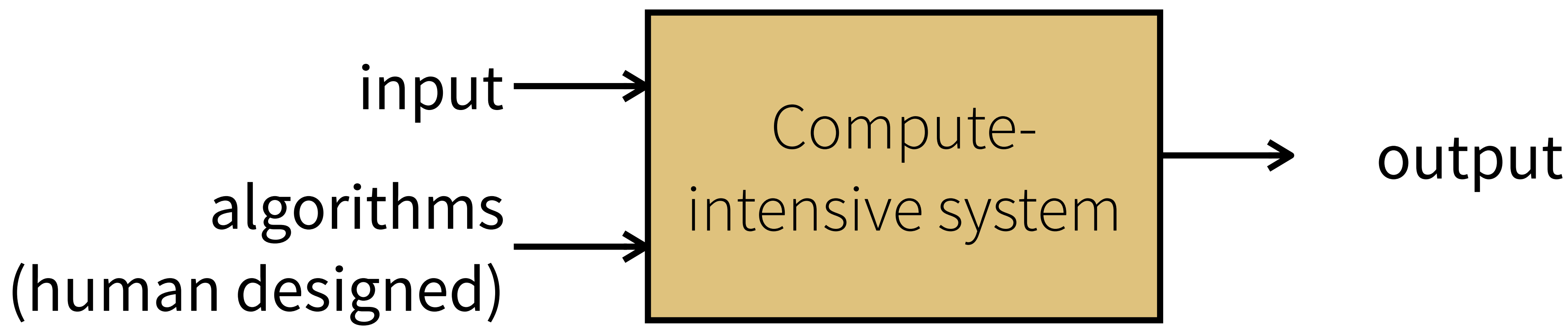
Traditional programming



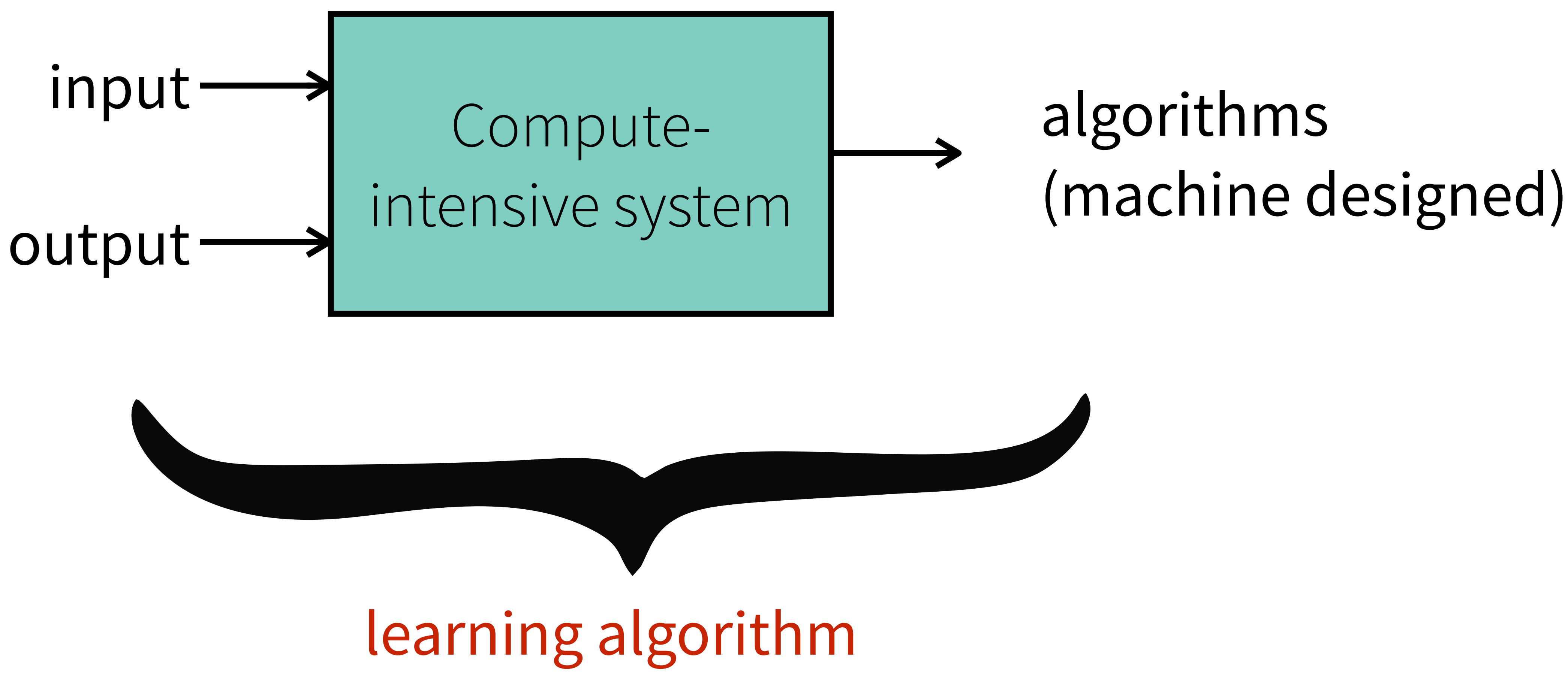
Machine Learning



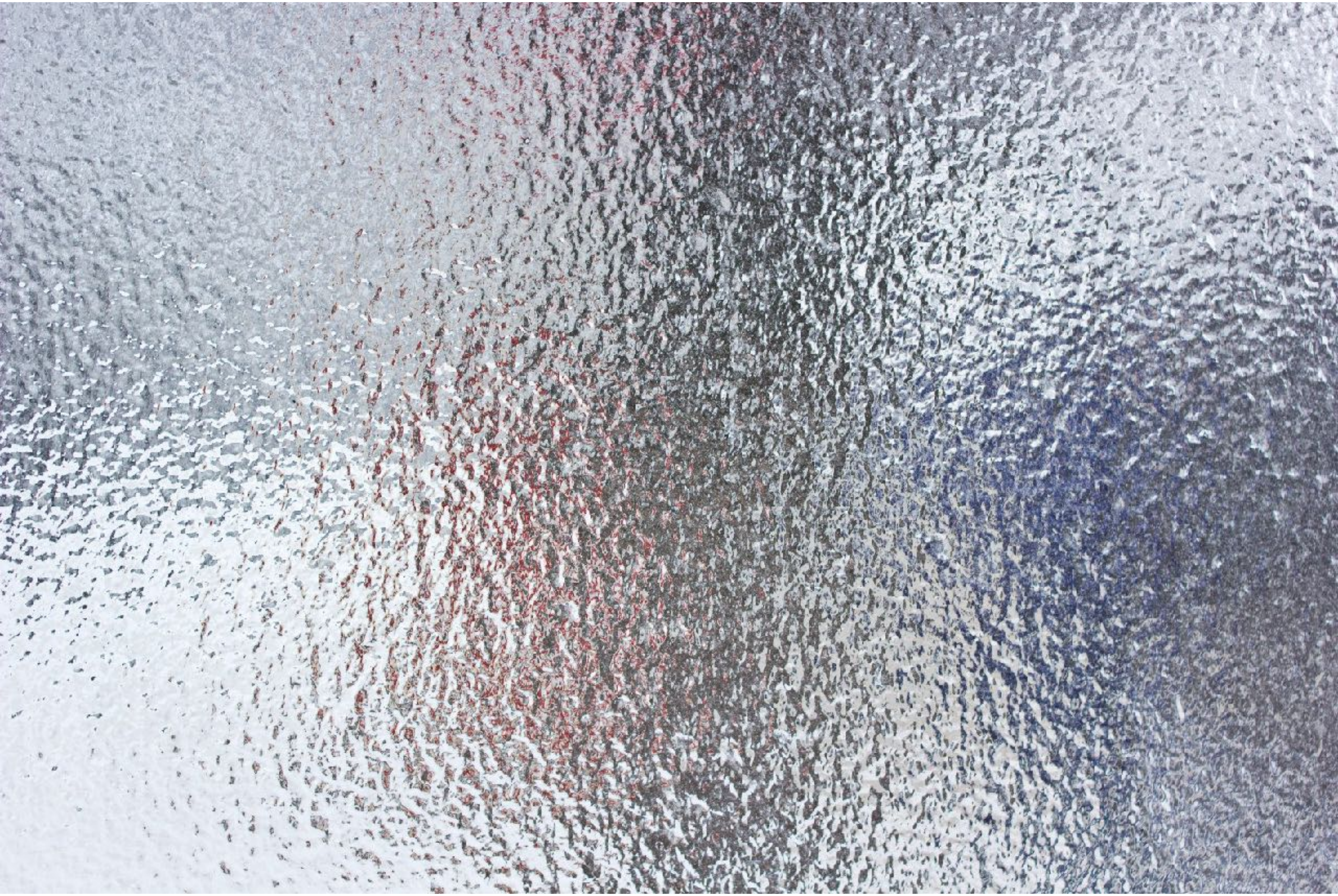
Traditional programming



Machine Learning



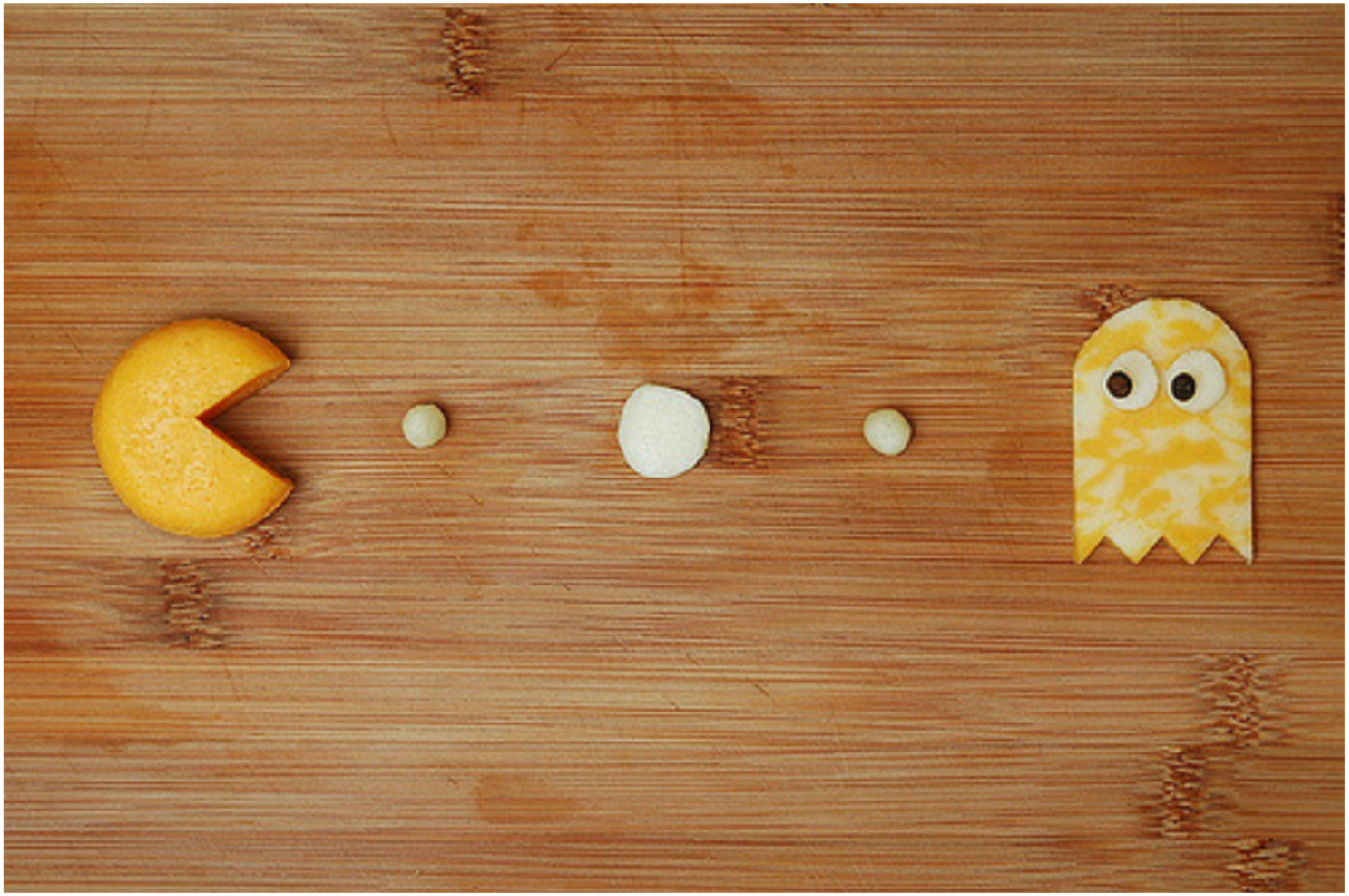
via Medium



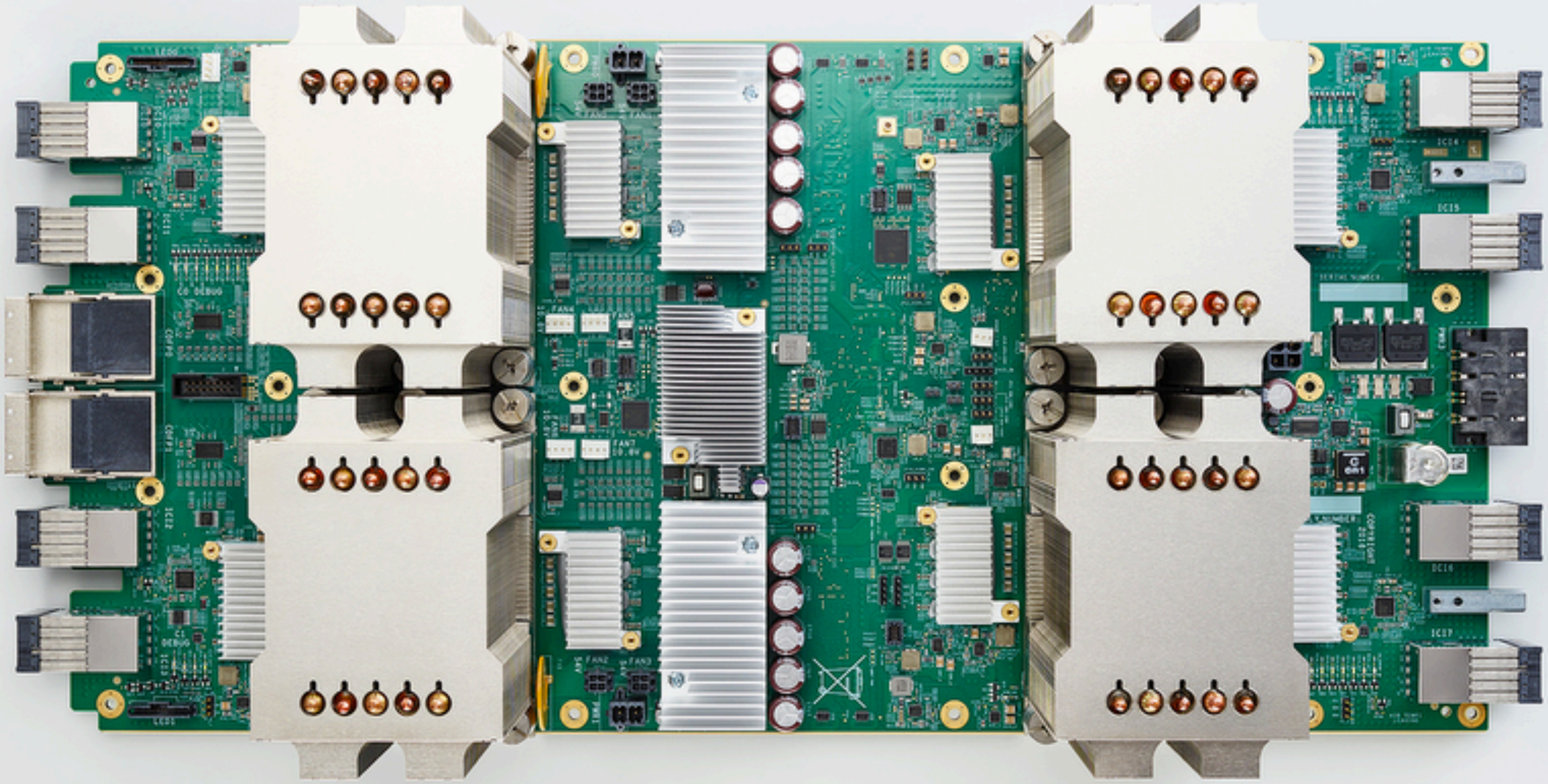
via miro.medium.com

Deep Learning is Eating Software

November 13, 2017
By Pete Warden
in Uncategorized
30 Comments

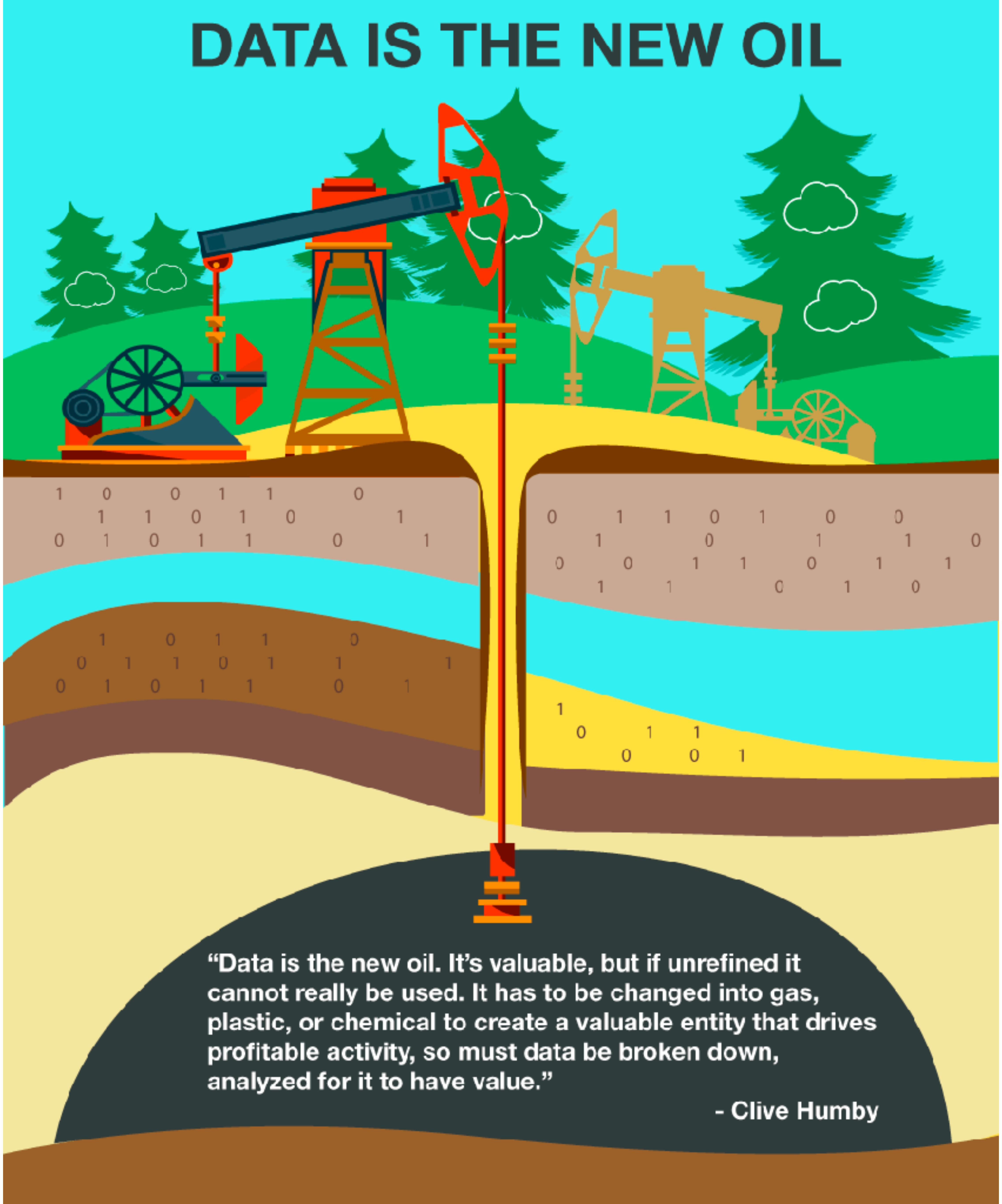


via Pete Warden Blog



via Google AI

Components of a Machine Learning System

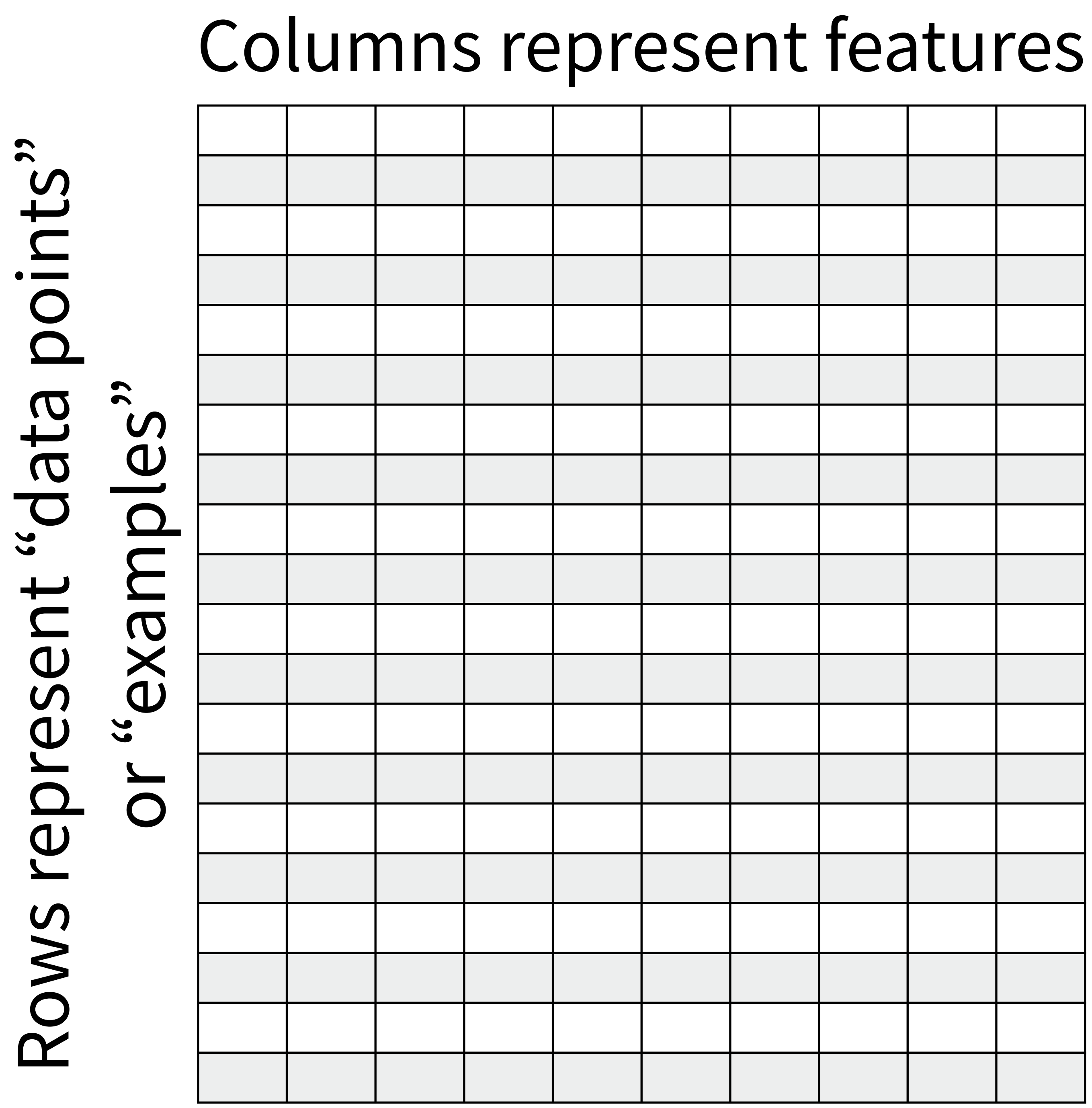


via ZE Perspective

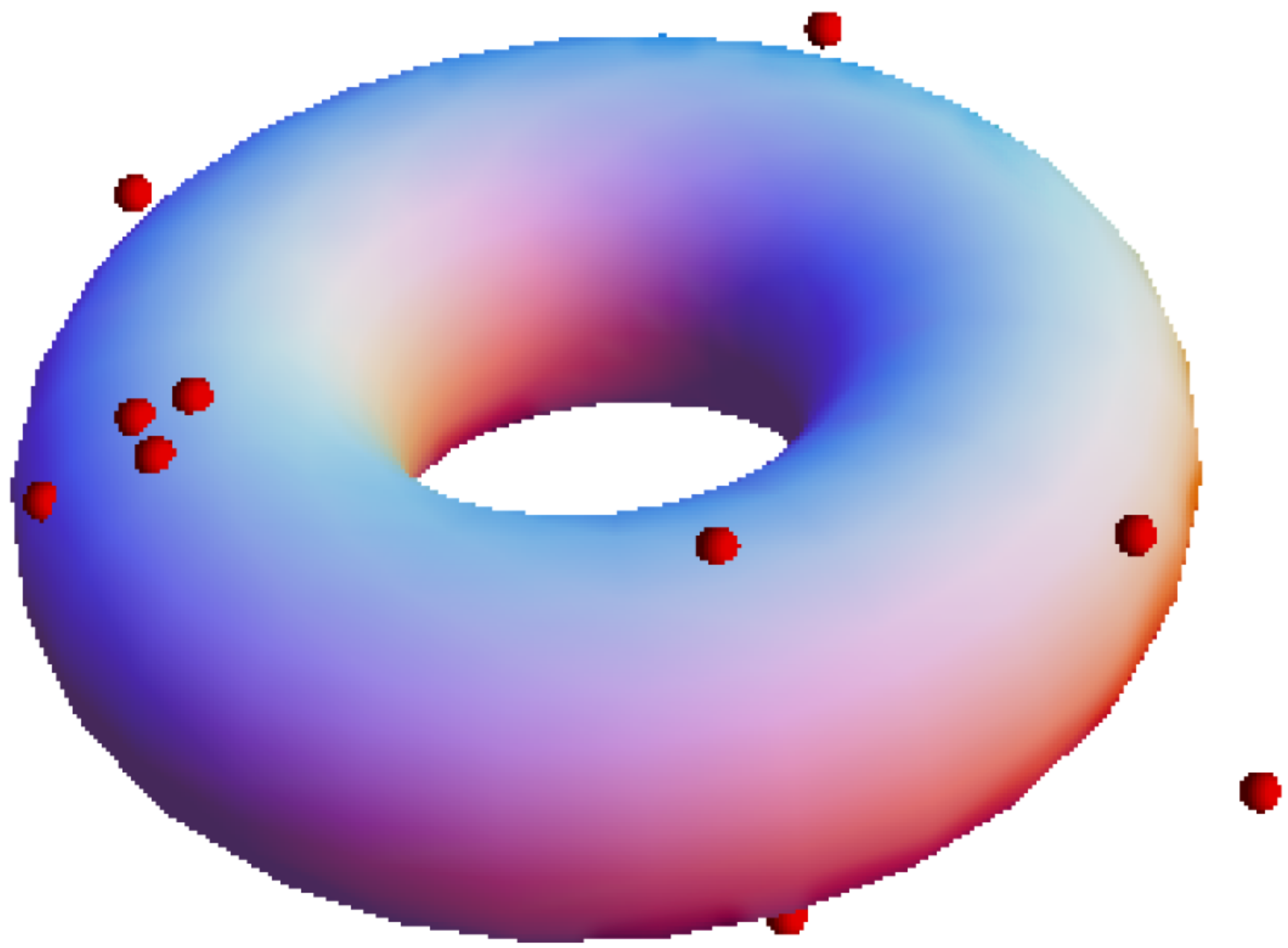
Datasets

Most learning algorithms we discuss will be able to experience an entire dataset

Dataset as a “design matrix”



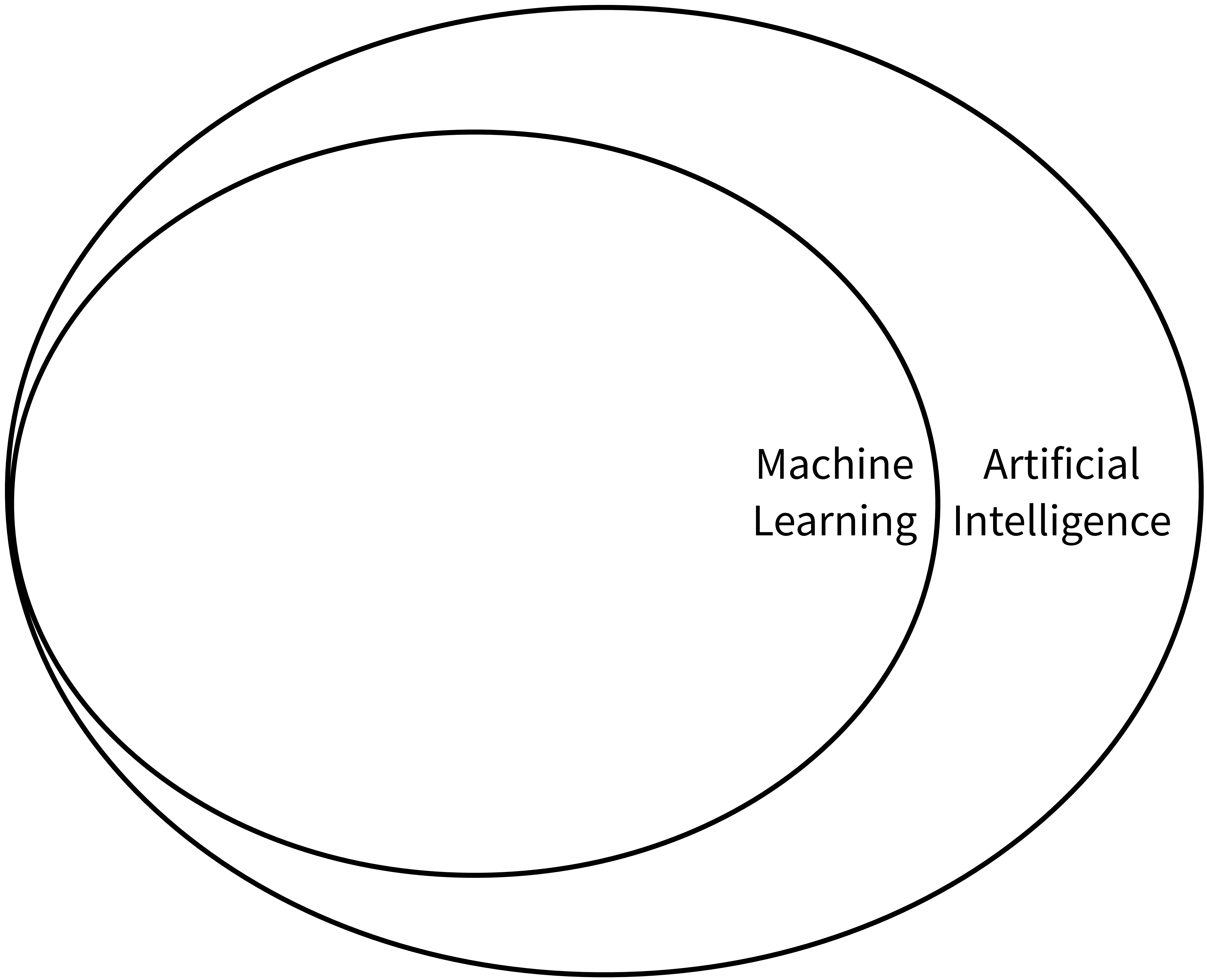
The Geometry of Data



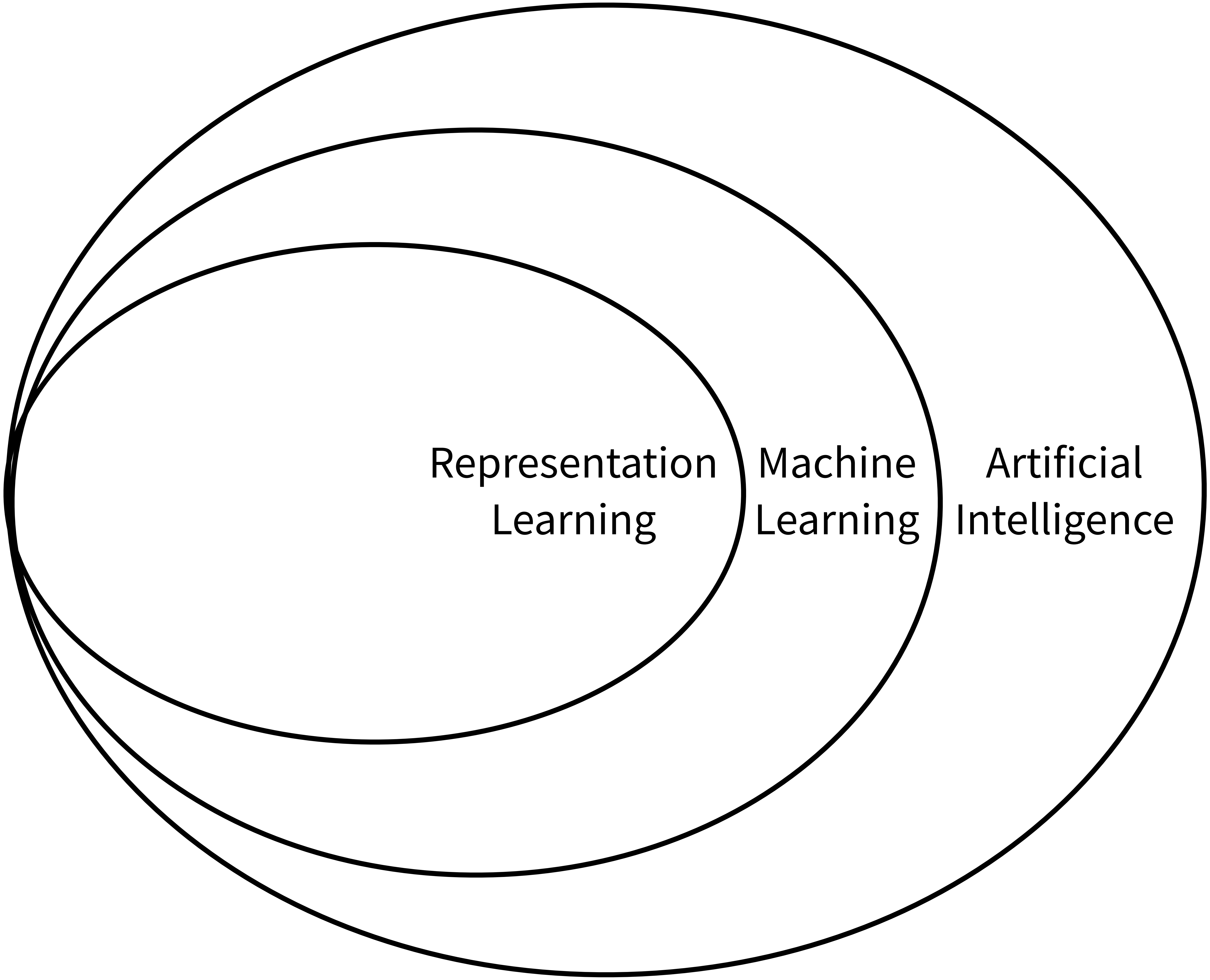
via Fefferman et al.

Disentangling Factors of Variation

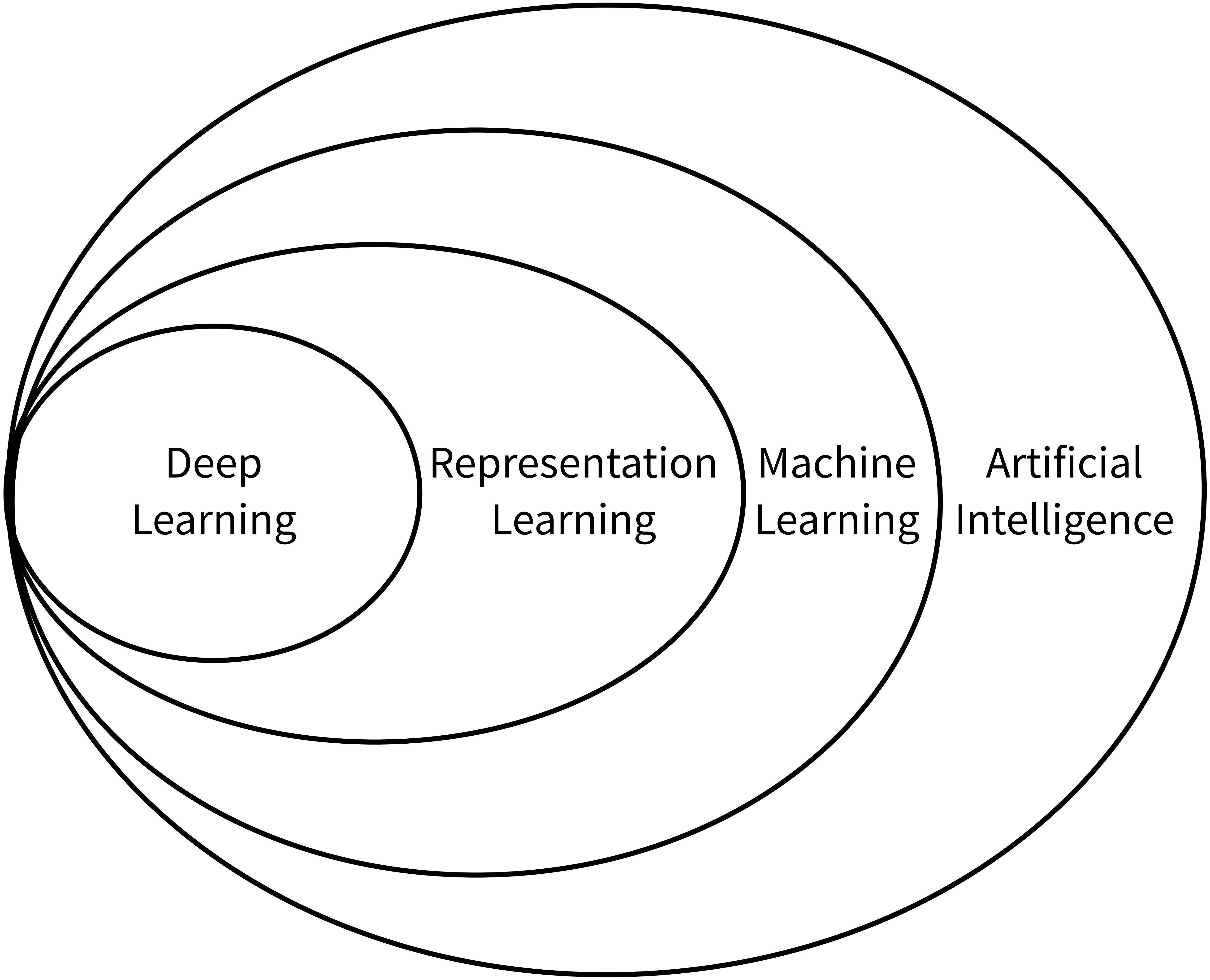
via mmlab.ie.cuhk.edu.hk



reproduced from Goodfellow et al.

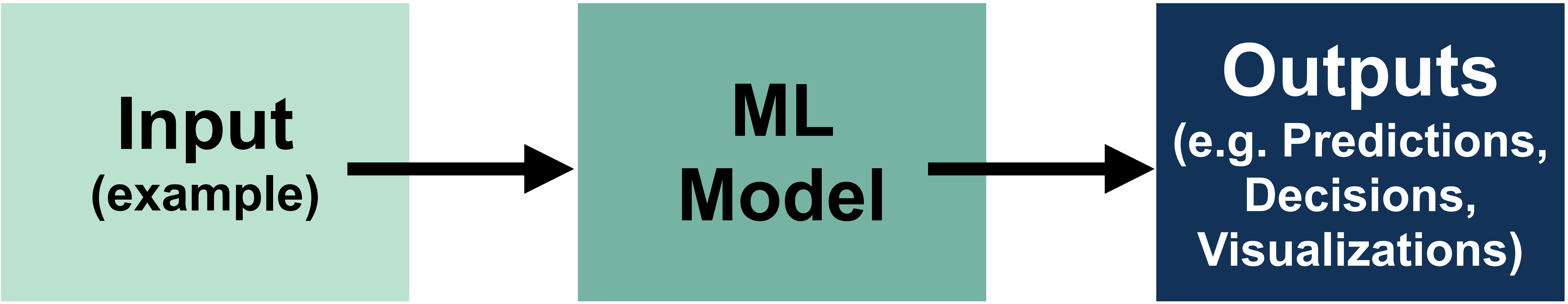


reproduced from Goodfellow et al.

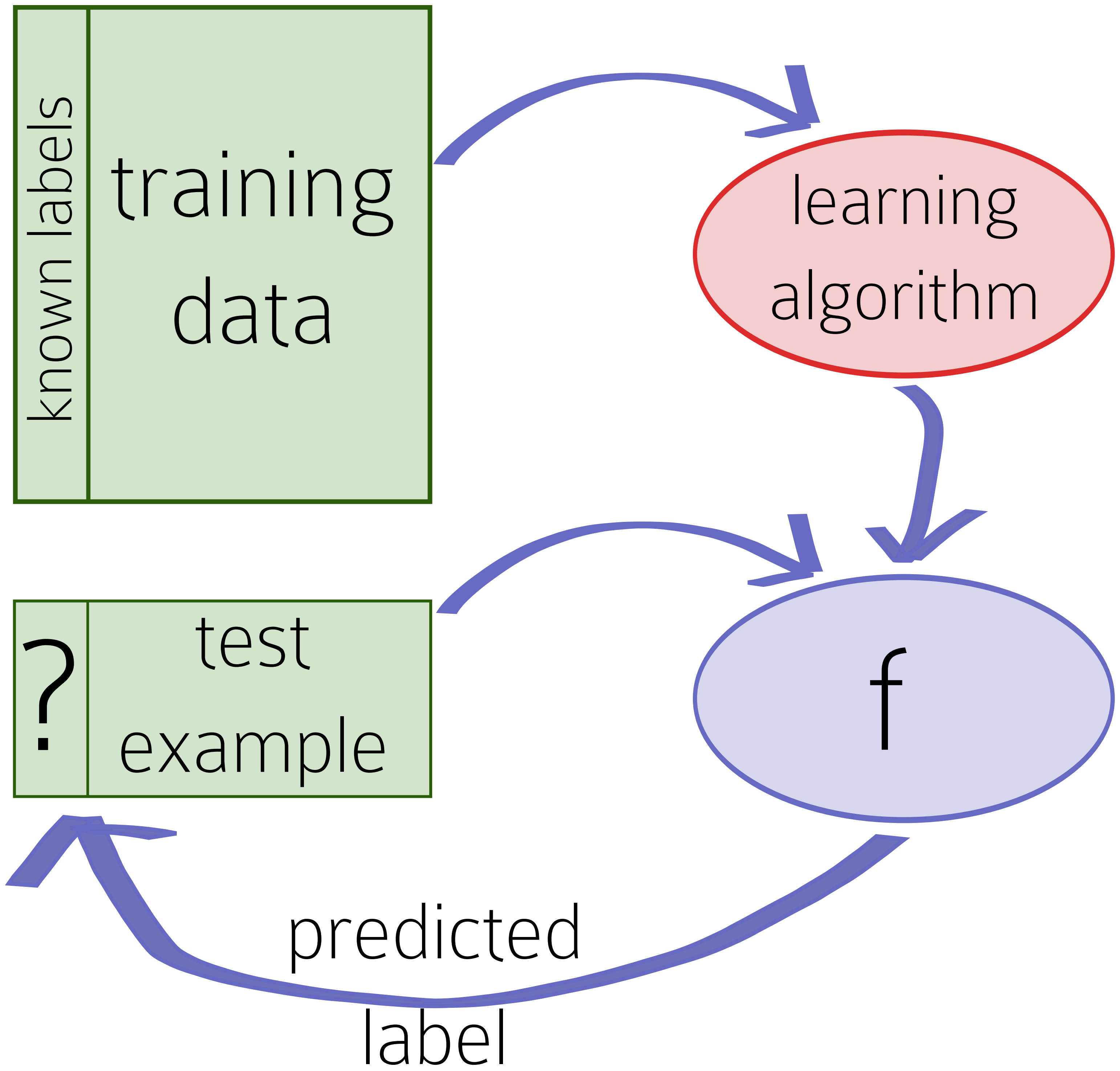


reproduced from Goodfellow et al.

Machine Learning Flow



Induction



via ciml.info

Training vs. Testing

Training vs. Testing

Training Stage (Learning)

Data is fed to model, parameters (degrees of freedom) are updated

Typically takes a long time

Training vs. Testing

Training Stage (Learning)
Data is fed to model, parameters (degrees of freedom) are updated
Typically takes a long time

Testing Stage (Deployment)
Parameters frozen, model consumes data, produces outputs
Typically very fast (e.g. real time!)

Training vs. Testing

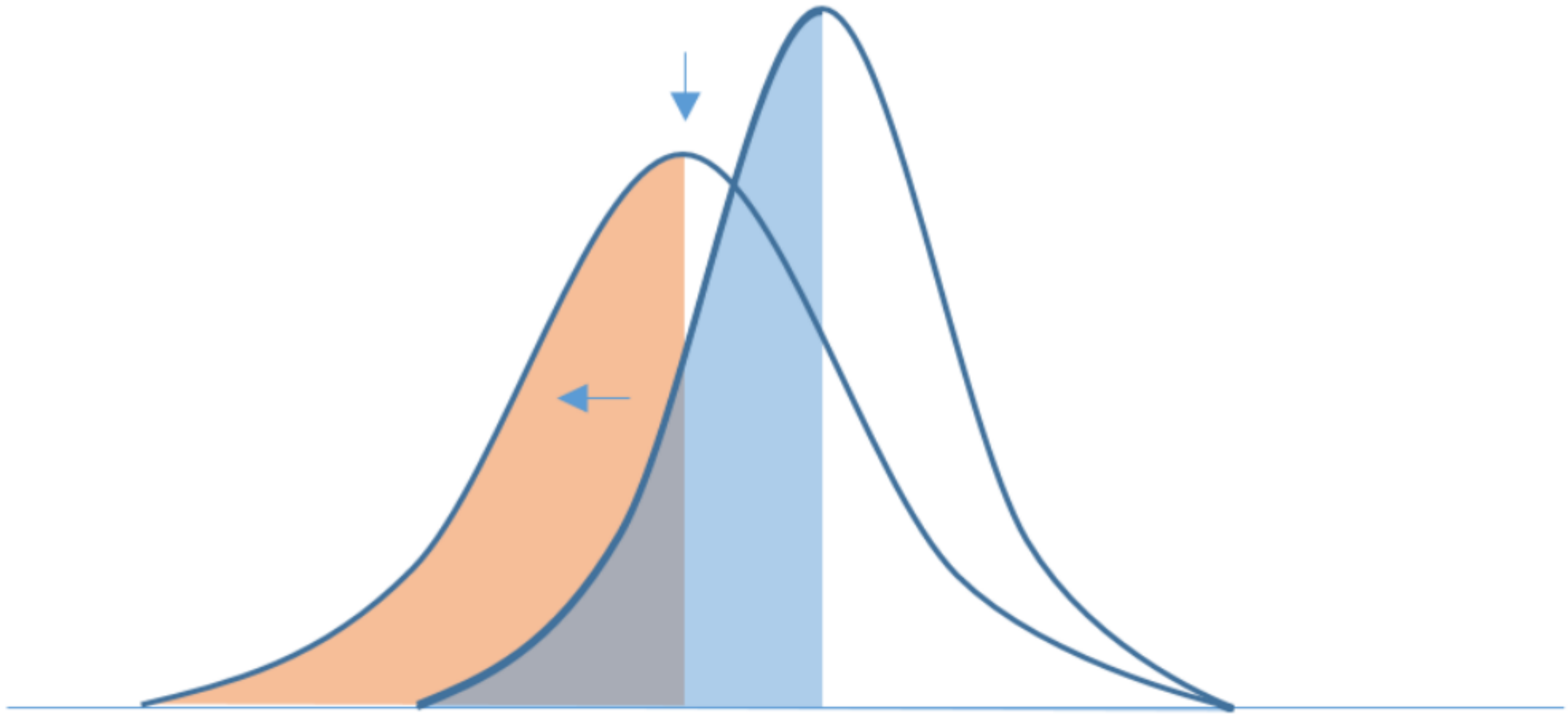
Training Stage (Learning)
Data is fed to model, parameters (degrees of freedom) are updated
Typically takes a long time

Testing Stage (Deployment)
Parameters frozen, model consumes data, produces outputs
Typically very fast (e.g. real time!)

Note that systems may undergo additional learning after deployment

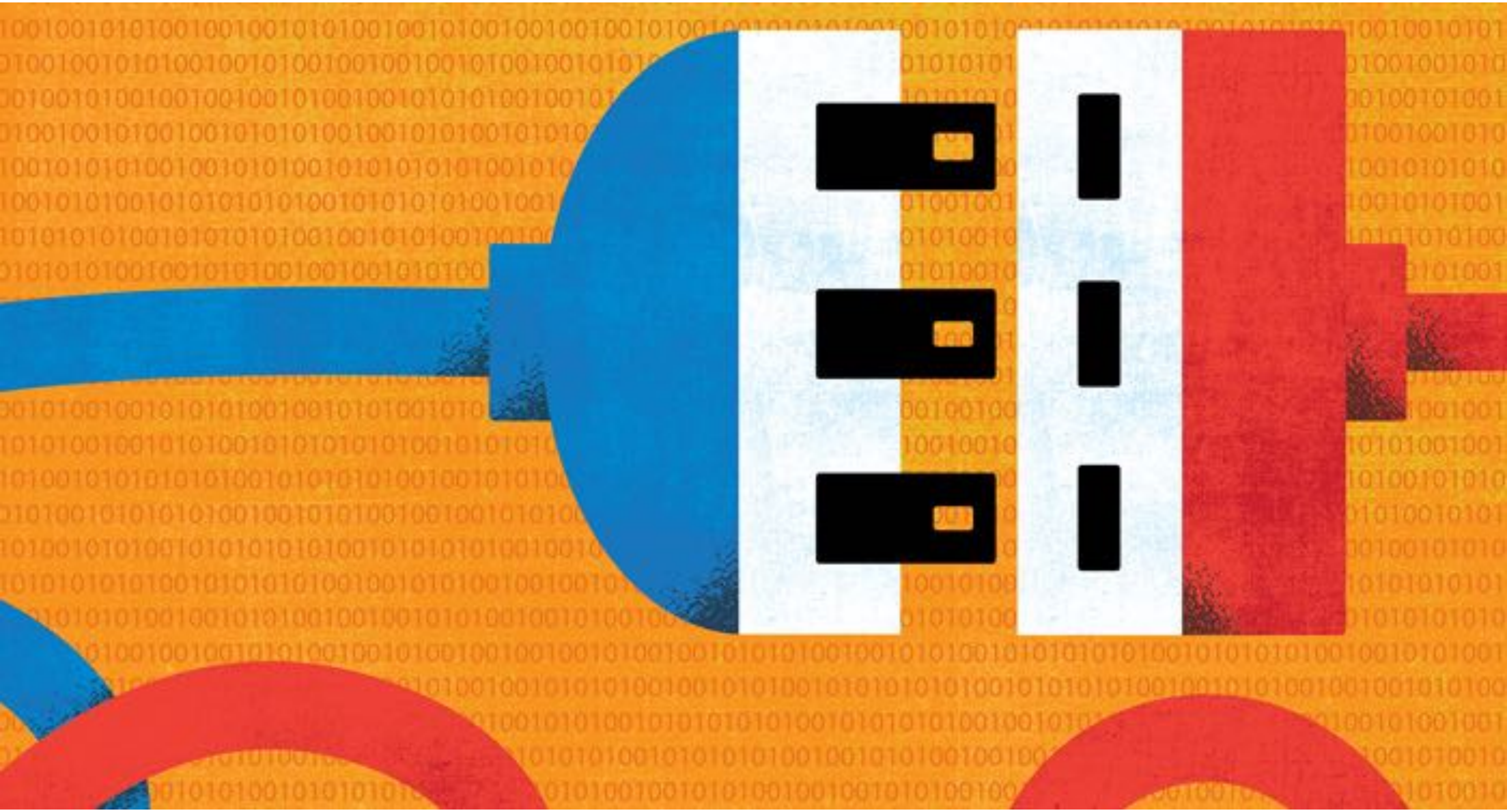
Comparing Traditional Software Systems to Machine Learning Systems

Distributional Shift



via Wikimedia Commons

via MIT Sloan Management Review



via Tax Executive

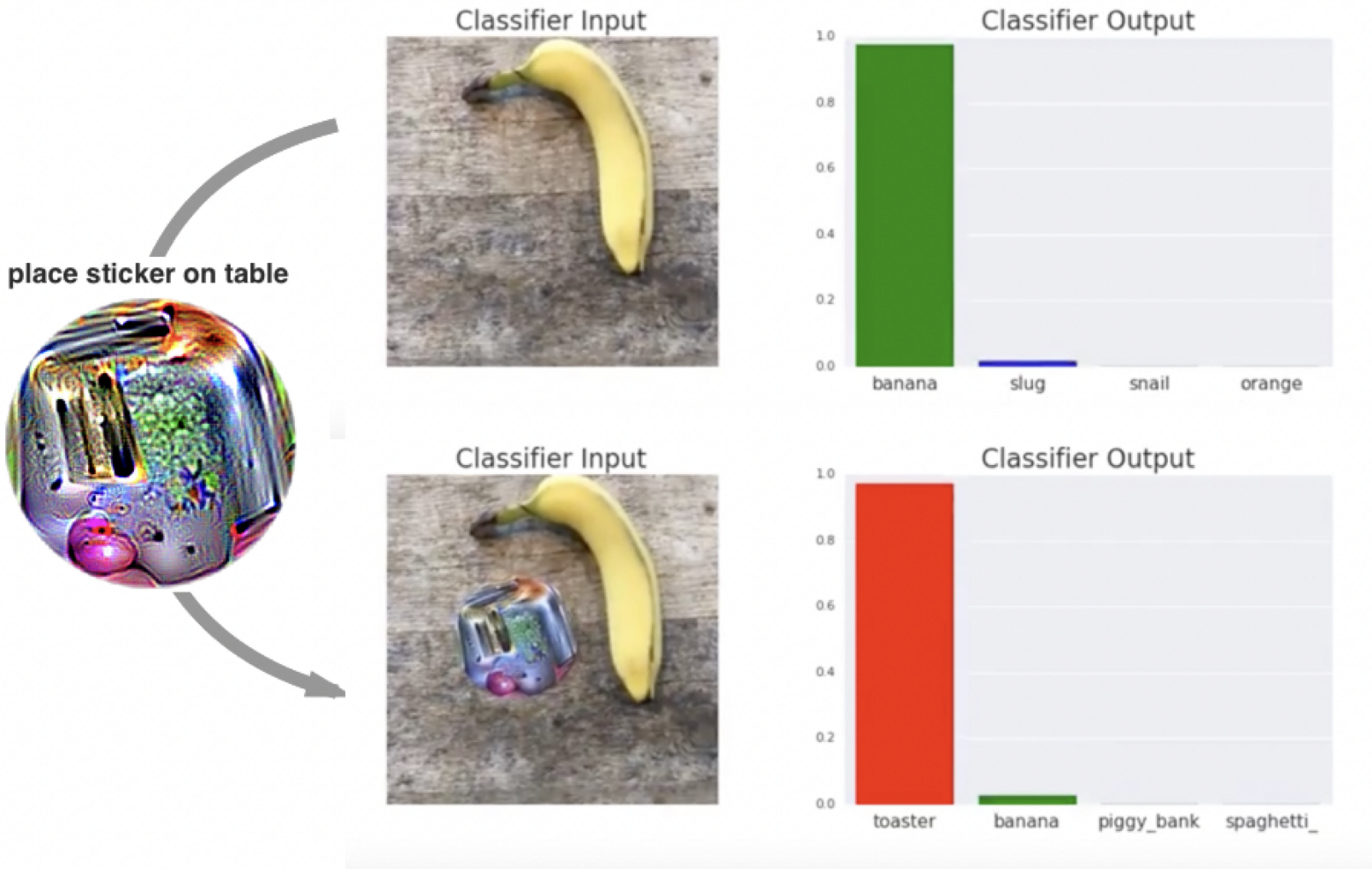
25 Nov 2019 | 14:00 GMT

In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation

The bot learned language from people on Twitter—but it also learned values

By **Oscar Schwartz**

Photo-illustration: Gluekit



via kdnuggets.com

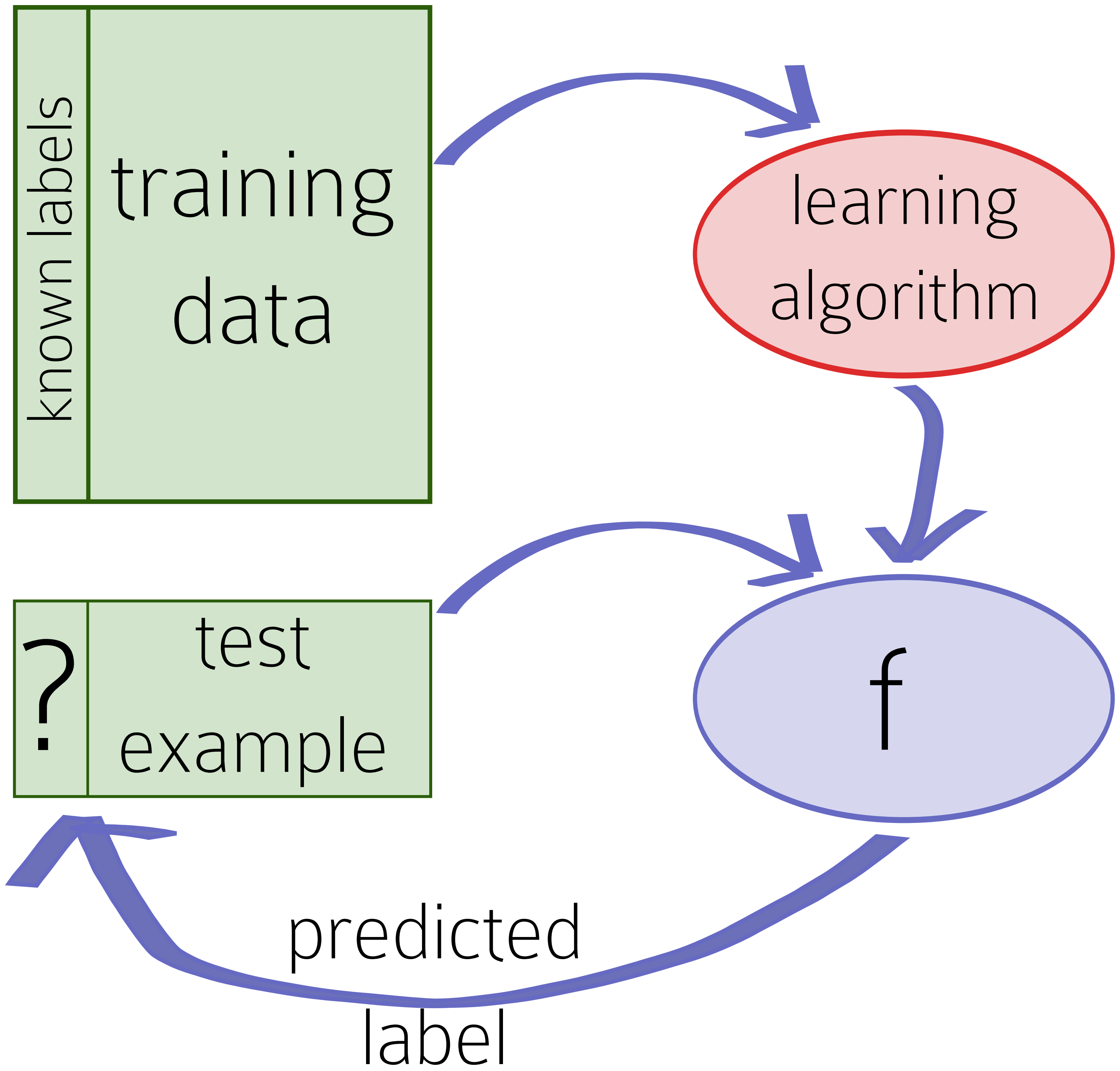
Governance and Liability

Who is responsible if something goes wrong?

- Researcher who authored the learning algorithm?
- ML engineer who implemented the algorithm?
- Customer who supplied the data? Person who collected it?
- Person who acts on the prediction?

Machine Learning Tasks

Learning is not the Task



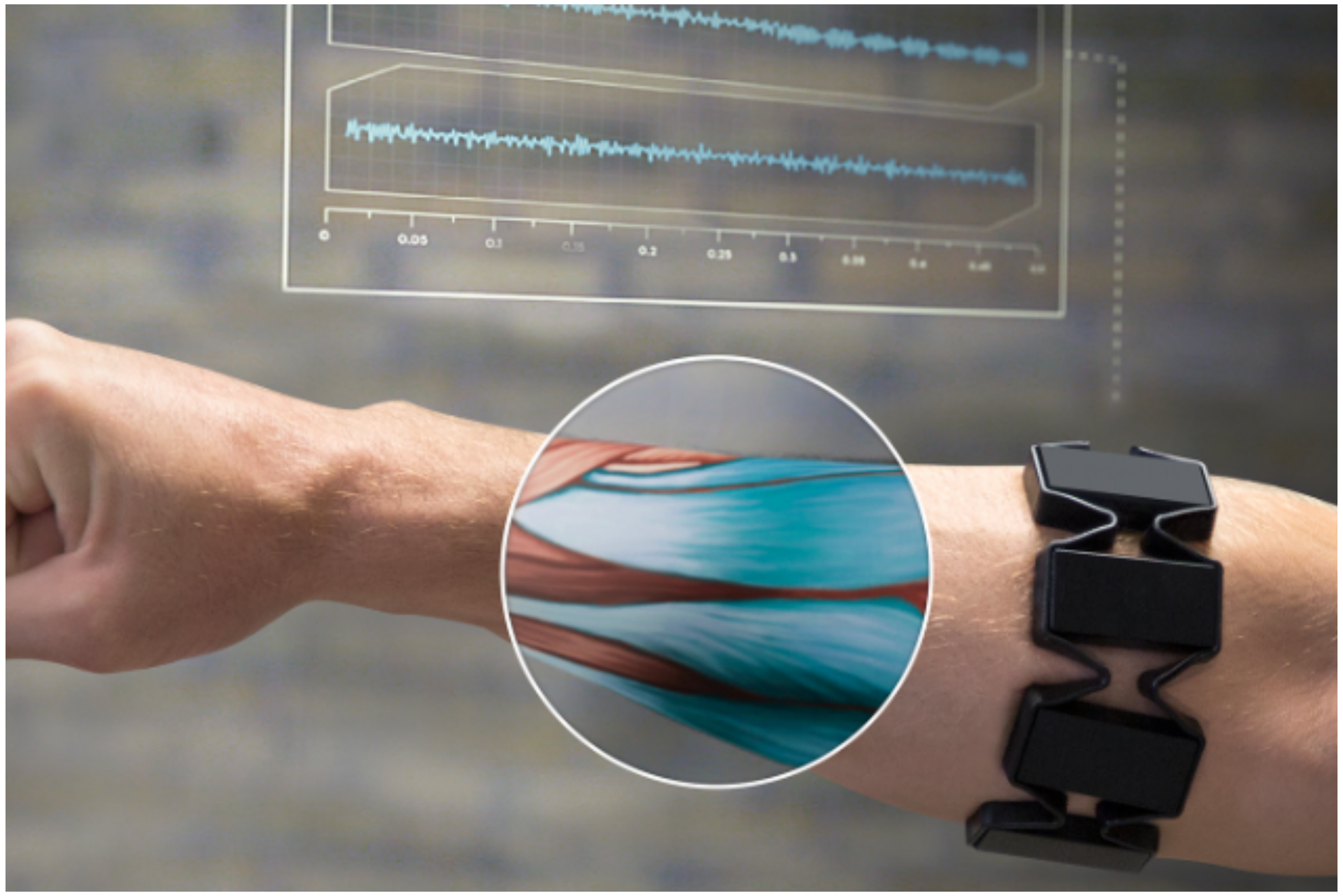
via ciml.info

Classification

Assignment of inputs to one or more known categories.

Examples:

- Object recognition
- Scene labeling
- Medical diagnosis
- Ad click-through prediction
- Tagging news articles
- Spam filtering
- Gesture recognition



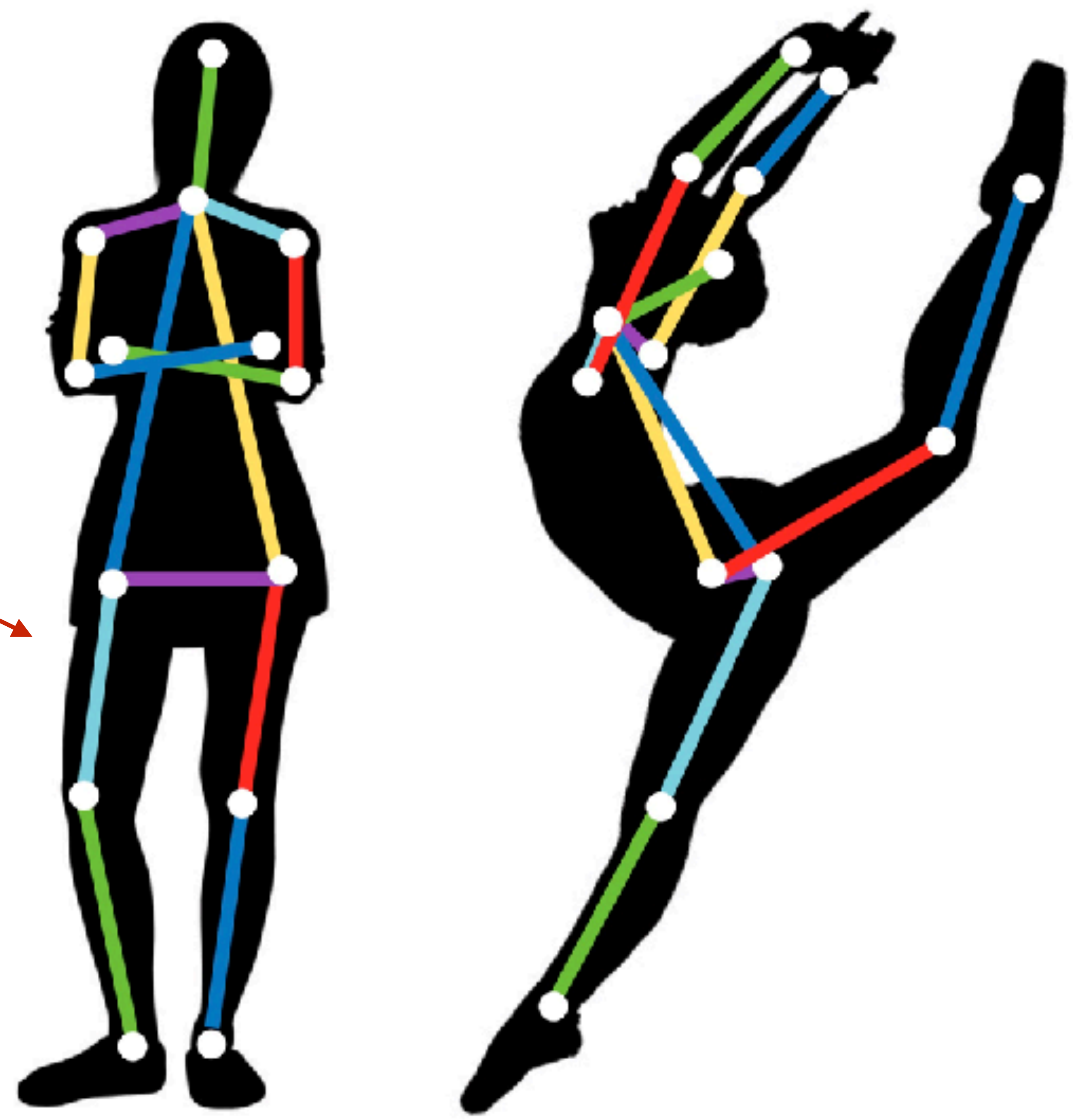
North Myo Armband Technology

Regression

Prediction of one or more real-valued quantities.

Examples:

- Age prediction
- Plant or soil health from aerial imagery
- Forecasting (e.g. weather, financial)
- Pose estimation



via catherinedong.com

Clustering

Assignment of inputs to unnamed groups ("clusters") such that objects in the same group are similar.

Examples:

- Exploratory data mining
- Plant and animal ecology
- Human genetic clustering
- Grouping of shopping items
- Market research
- Semi-automated grading

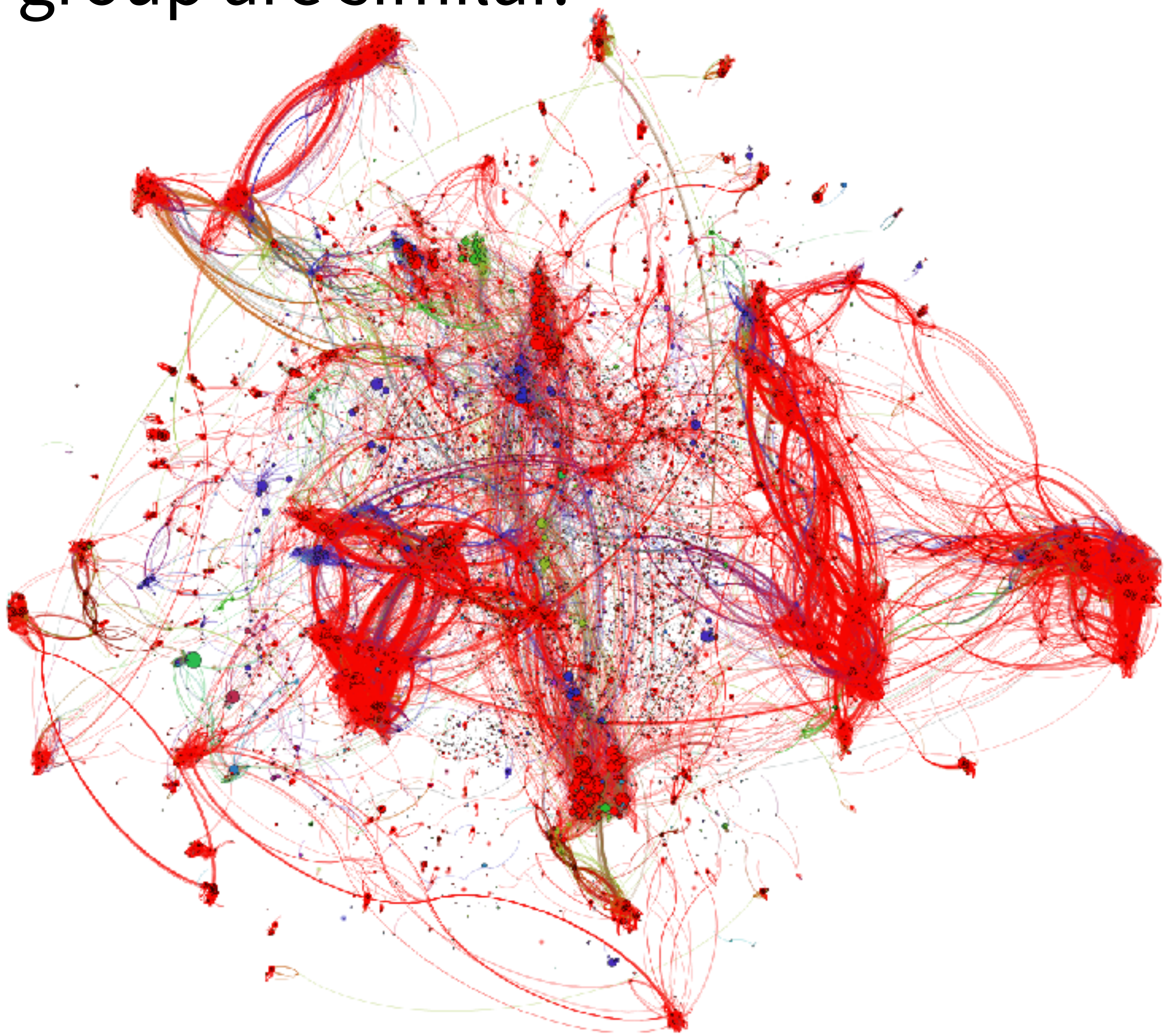


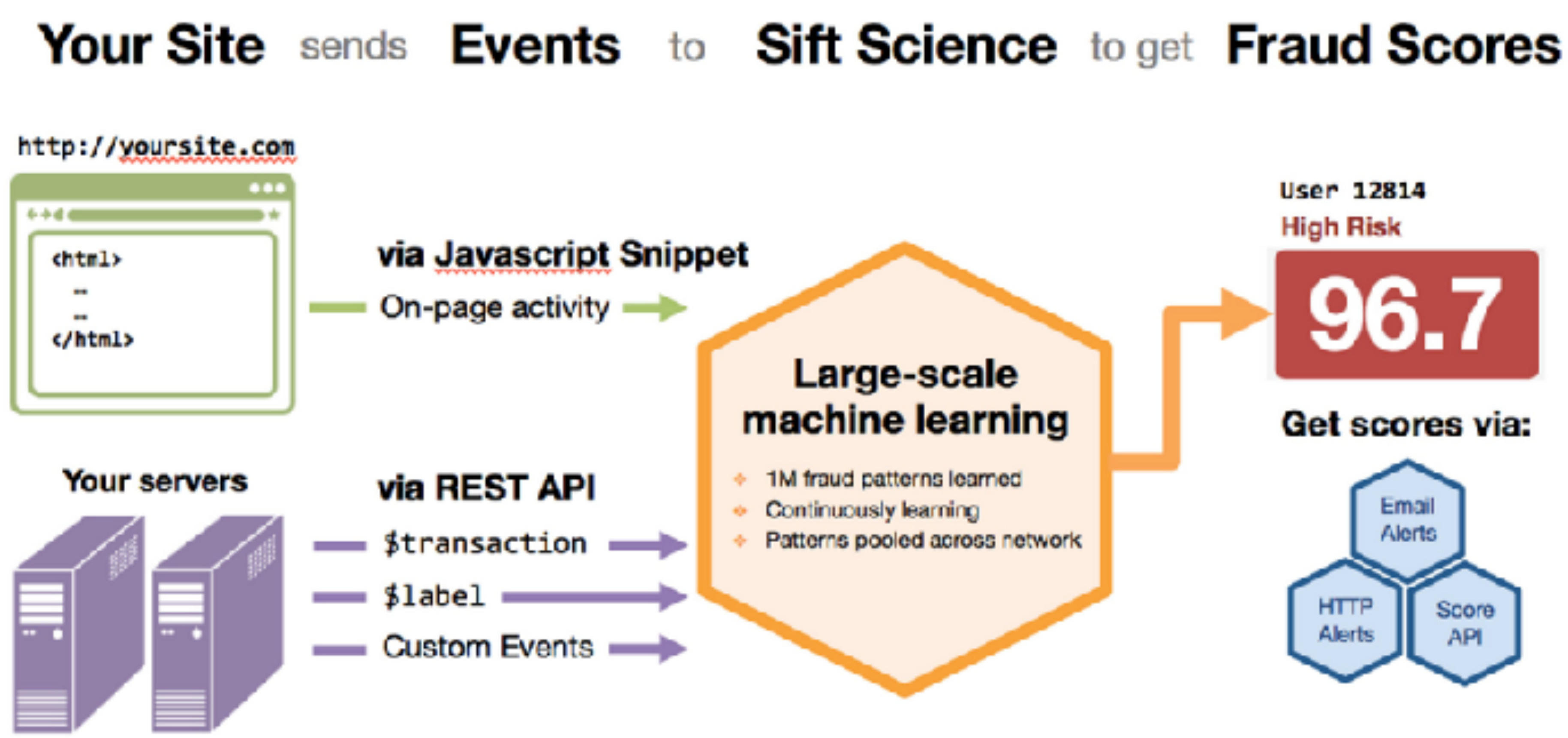
Image credit: Huang et al. 2013 "Codewebs: a Pedagogical Search Engine for Code Submissions to a MOOC"

Anomaly detection

Sift through a set of events or objects and flag some as unusual or atypical.

<u>Examples:</u>

- Manufacturing process inspection
- Cybersecurity (e.g. network attacks)
- Credit card fraud detection



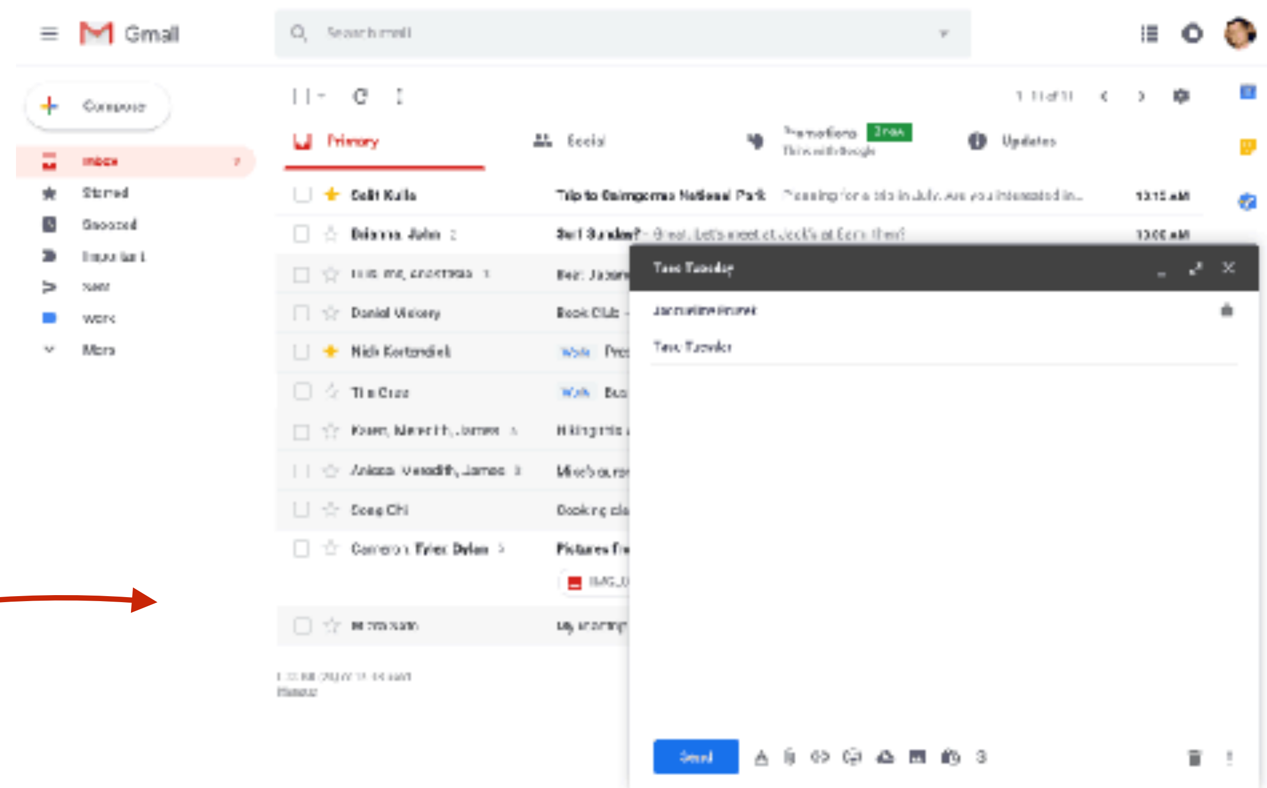
Sift Science Suspicion Score

Generation

Creation of high-dimensional output, often conditional on input.

<u>Examples:</u>

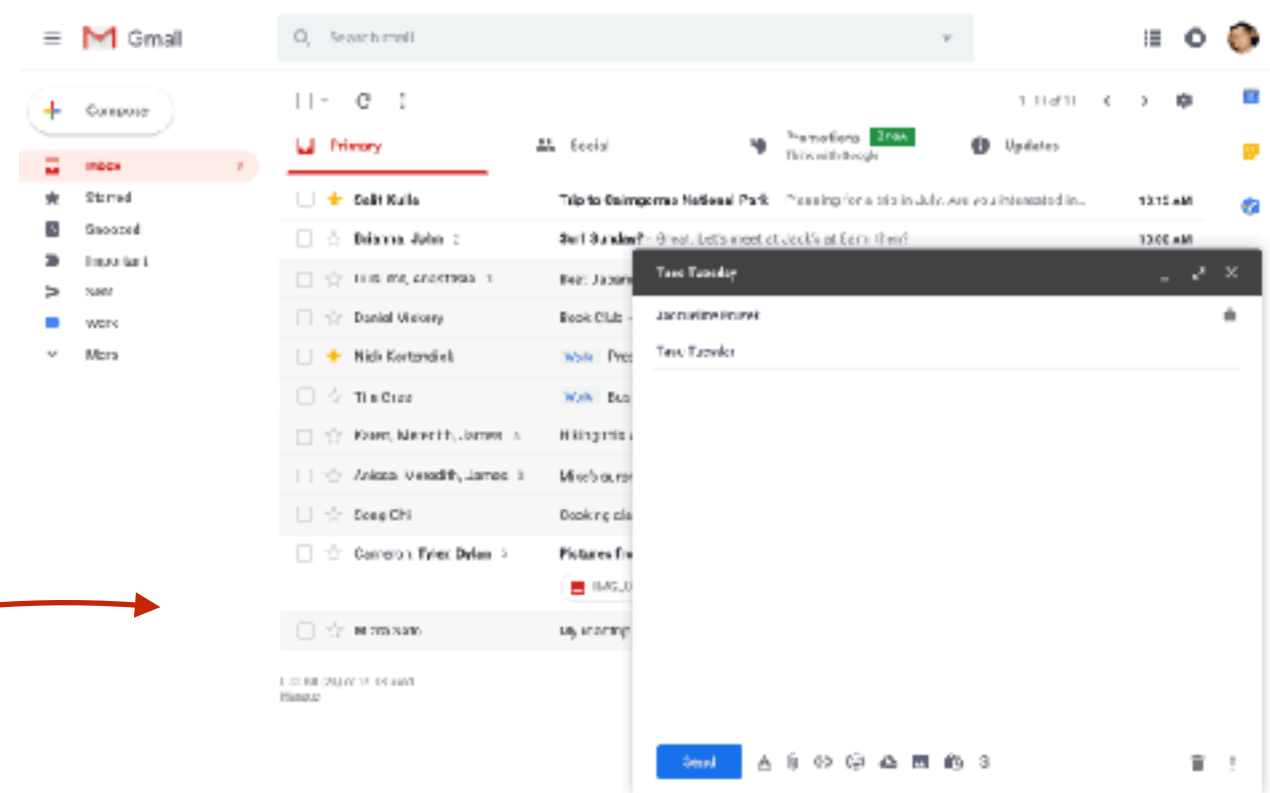
- Image/Speech/Text Synthesis
- Image-to-Text (captioning)
- Text-to-Speech
- Text-to-Text (e.g. smart compose)



Still may require
human-in-the-loop (judgement)

Generation

Creation of high-dimensional output, often conditional on input.



<u>Examples:</u>

- Image/Speech/Text Synthesis
- Image-to-Text (captioning)
- Text-to-Speech
- Text-to-Text (e.g. smart compose)

Still may require
human-in-the-loop (judgement)

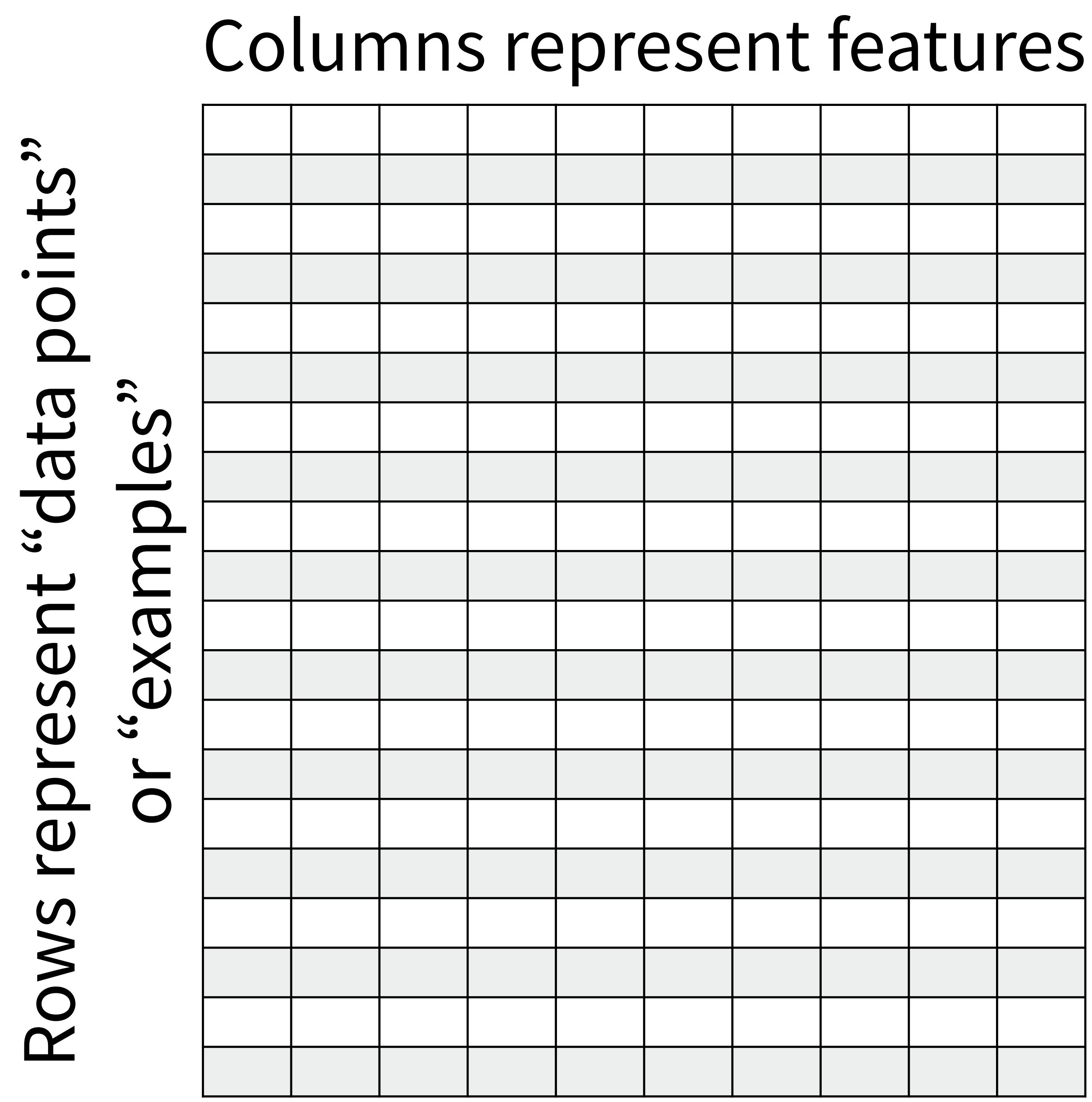
The Machine Learning Experience

Supervised Learning

via parentscafe.org

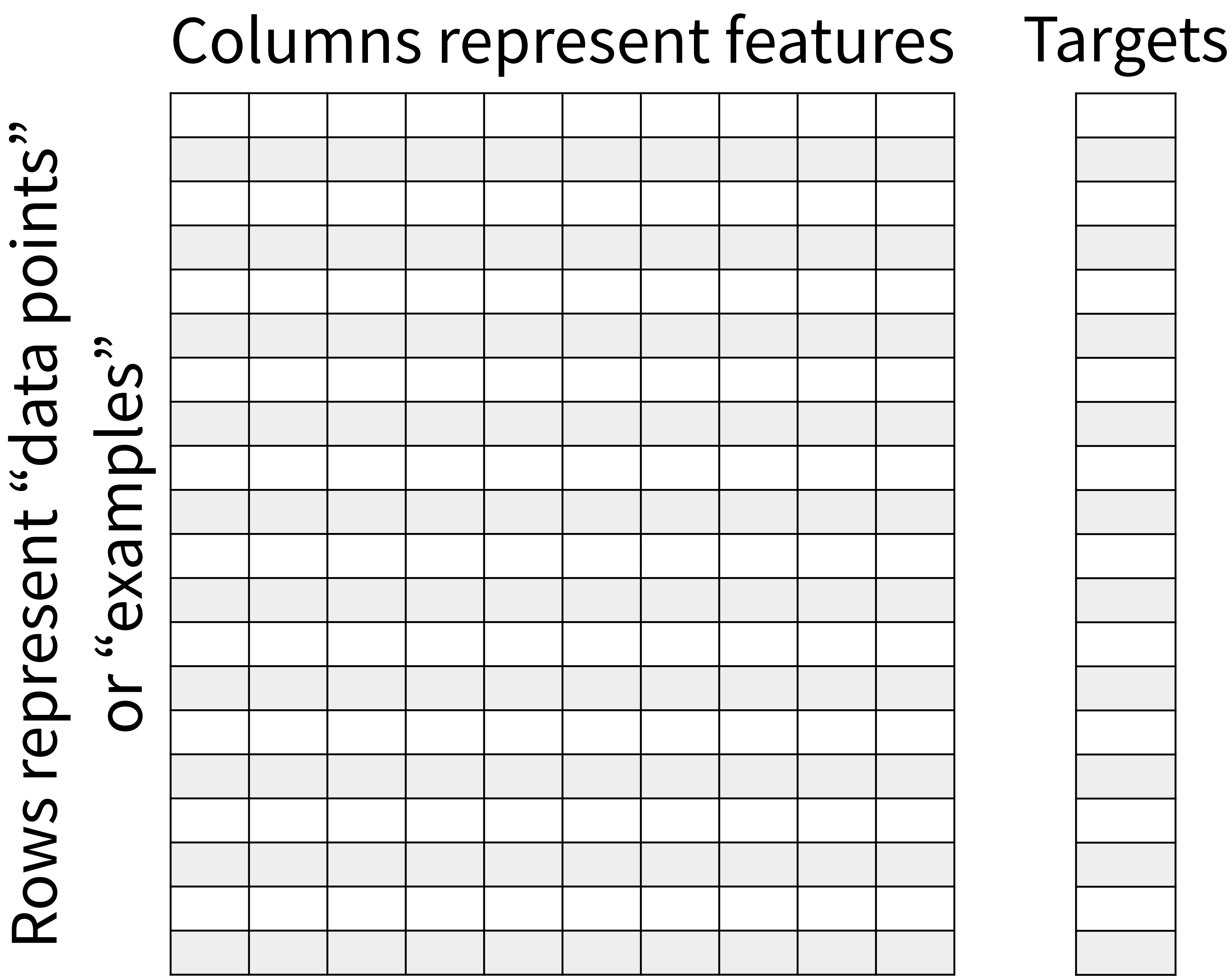
Supervised Learning Algorithms

Experience a dataset containing features, but each example is also associated with a label or target



Supervised Learning Algorithms

Experience a dataset containing features, but each example is also associated with a label or target

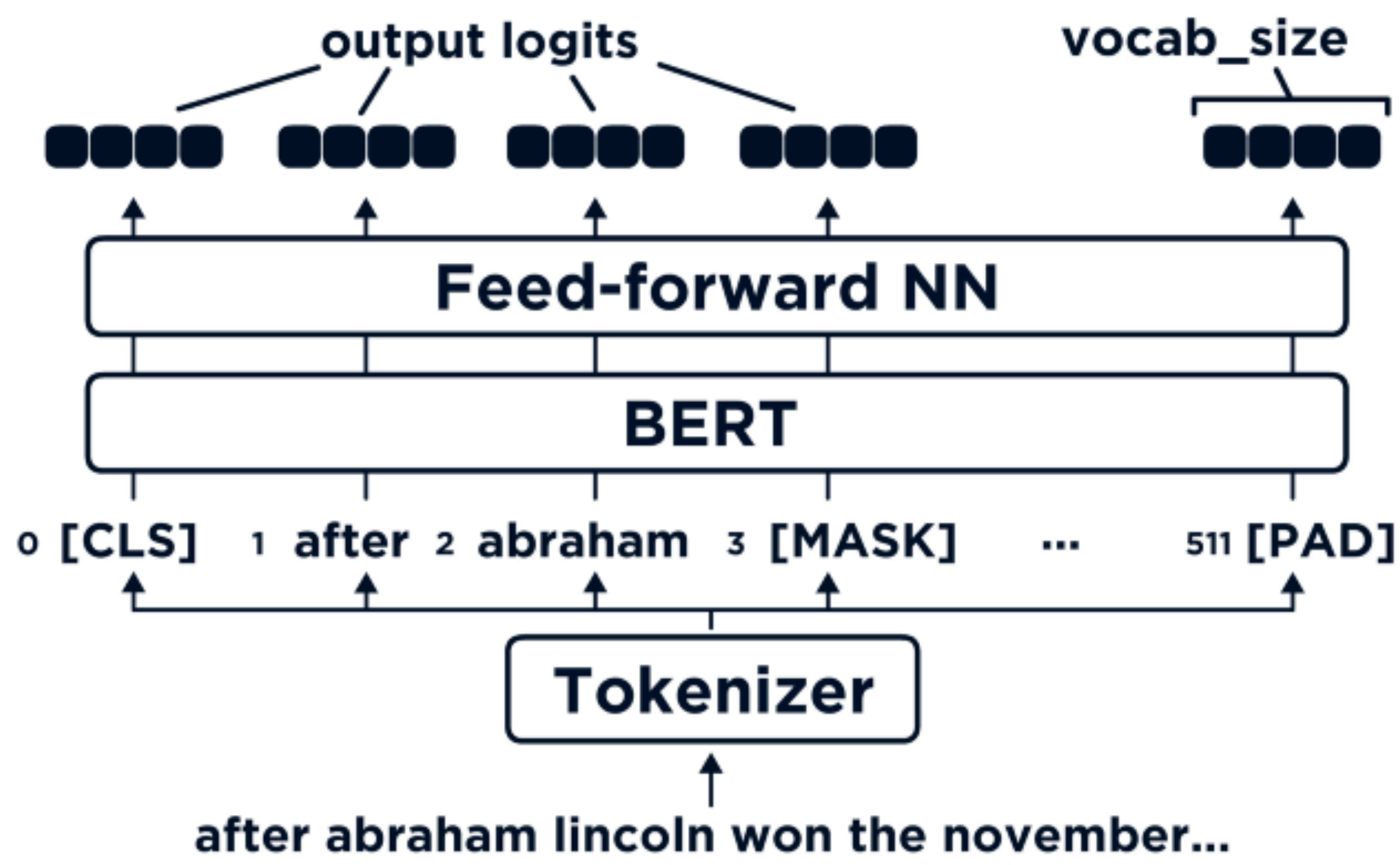


via New York Times

Unsupervised Learning

via thetimes.co.uk

Today's large language models often use "self-supervised" learning in the form of masked language modeling (pictured) or next-token prediction



Reinforcement Learning

via Jamie Campbell from Emsworth (nr Portsmouth), U.K - Falling down

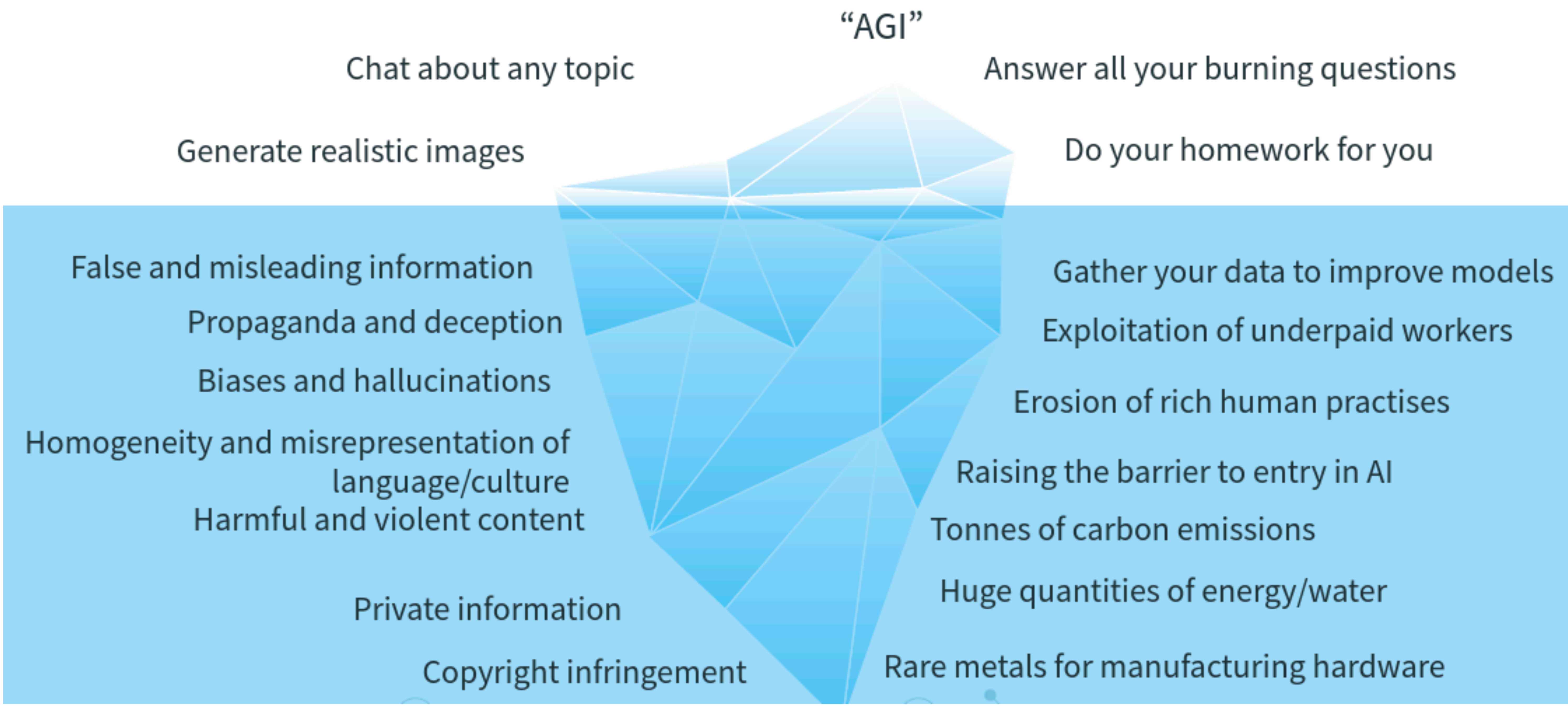
Blurred Paradigms

Unsupervised and supervised learning are not completely distinct or formal concepts.
Other variants of the learning paradigm are possible.

<u>Examples:</u>

- Semi-supervised learning
- Self-supervised learning
- Deep reinforcement learning

Google DeepMind's AlphaGo
A Hybrid of Several Learning Paradigms
+ Some "Brute Force"



via ArsTechnica

Machine Learning Research Group

@ University of Guelph: www.gwtaylor.ca

CARE-AI

@ University of Guelph: www.care-ai.ca

Thank you:

- Vector Institute for Artificial Intelligence
- Gov. Canada through New Frontiers in Research Fund Transformation
- Canada Research Chairs
- NVIDIA
- CIFAR
- Natural Sciences and Engineering Research Council
- Canada Foundation for Innovation
- Ontario Research Fund