AI and ML in astronomy and astrophysics

Machine Learning for Astronomers and Physicists

Federica B. Bianco
University of Delaware
Physics and Astronomy
Biden School of Public Policy and Administration
Data Science Institute


this slide deck:
https://slides.com/federicabianco/astroai

1/6

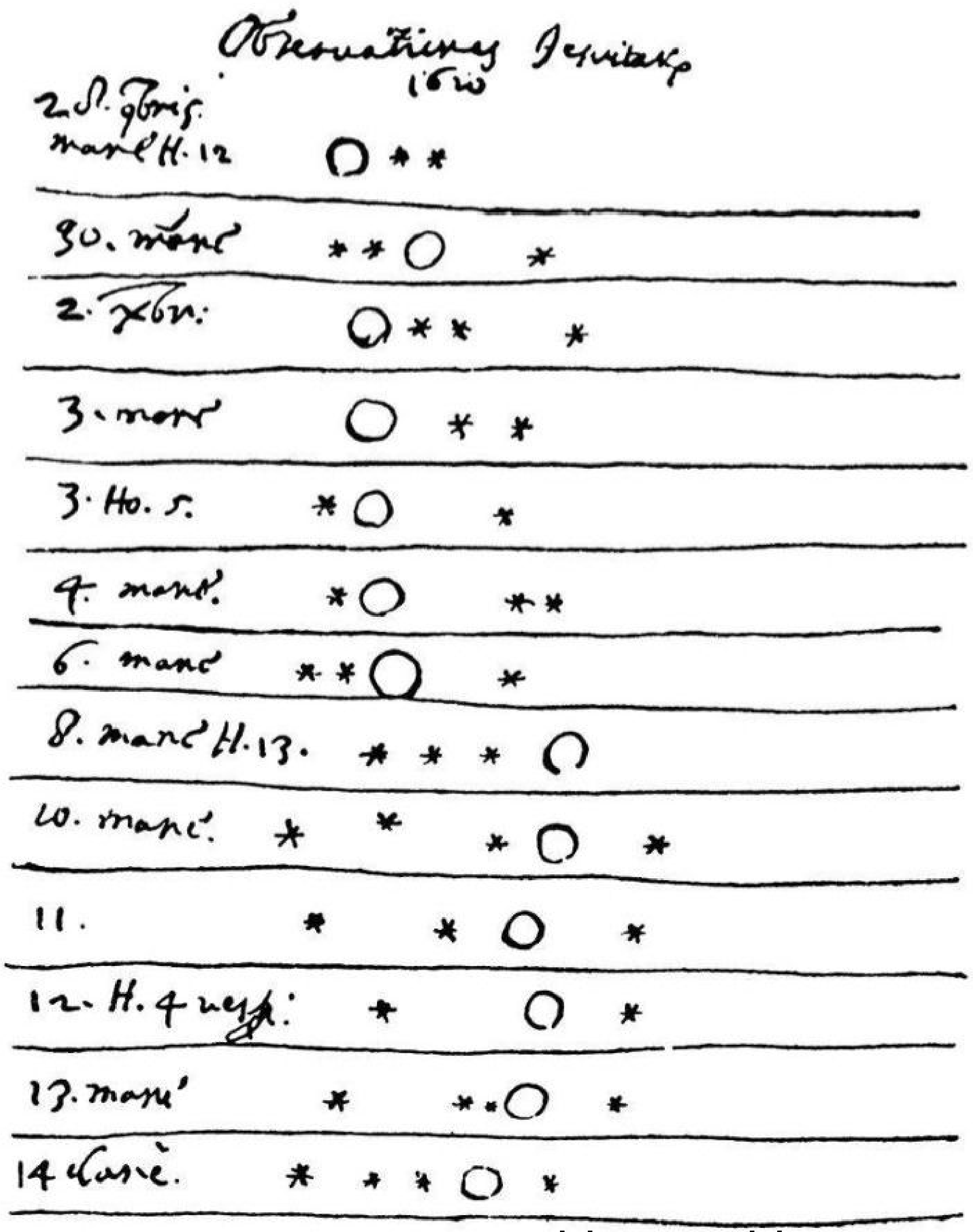
Historical perspective

what drives cosmic discovery

Experiment driven

Following: Djorgovski

https://events.asiaa.sinica.edu.tw/school/20170904/talk/djorgovski1.pdf

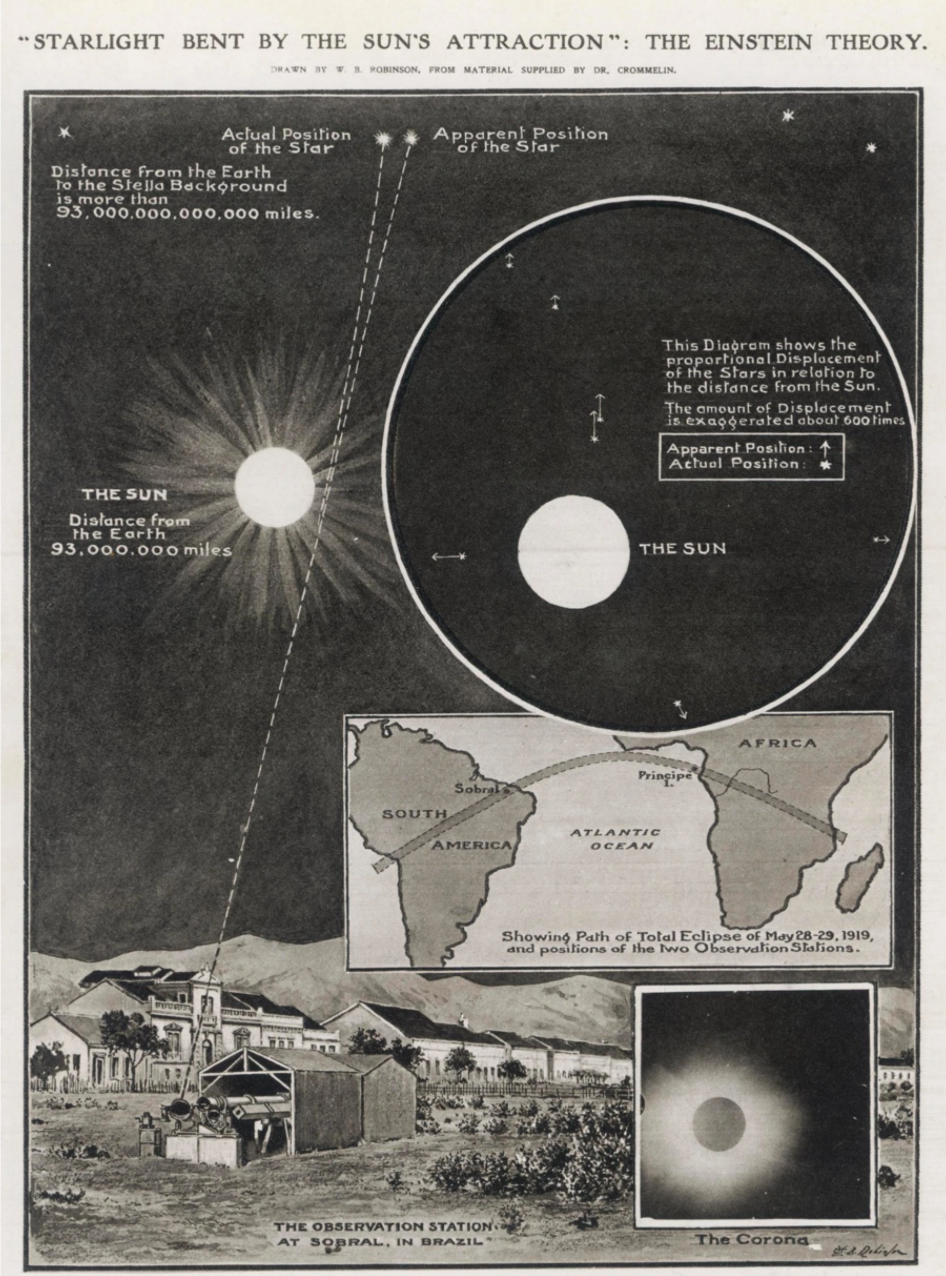


Galileo Galilei 1610

what drives cosmic discovery

Experiment driven

Theory driven | Falsifiability

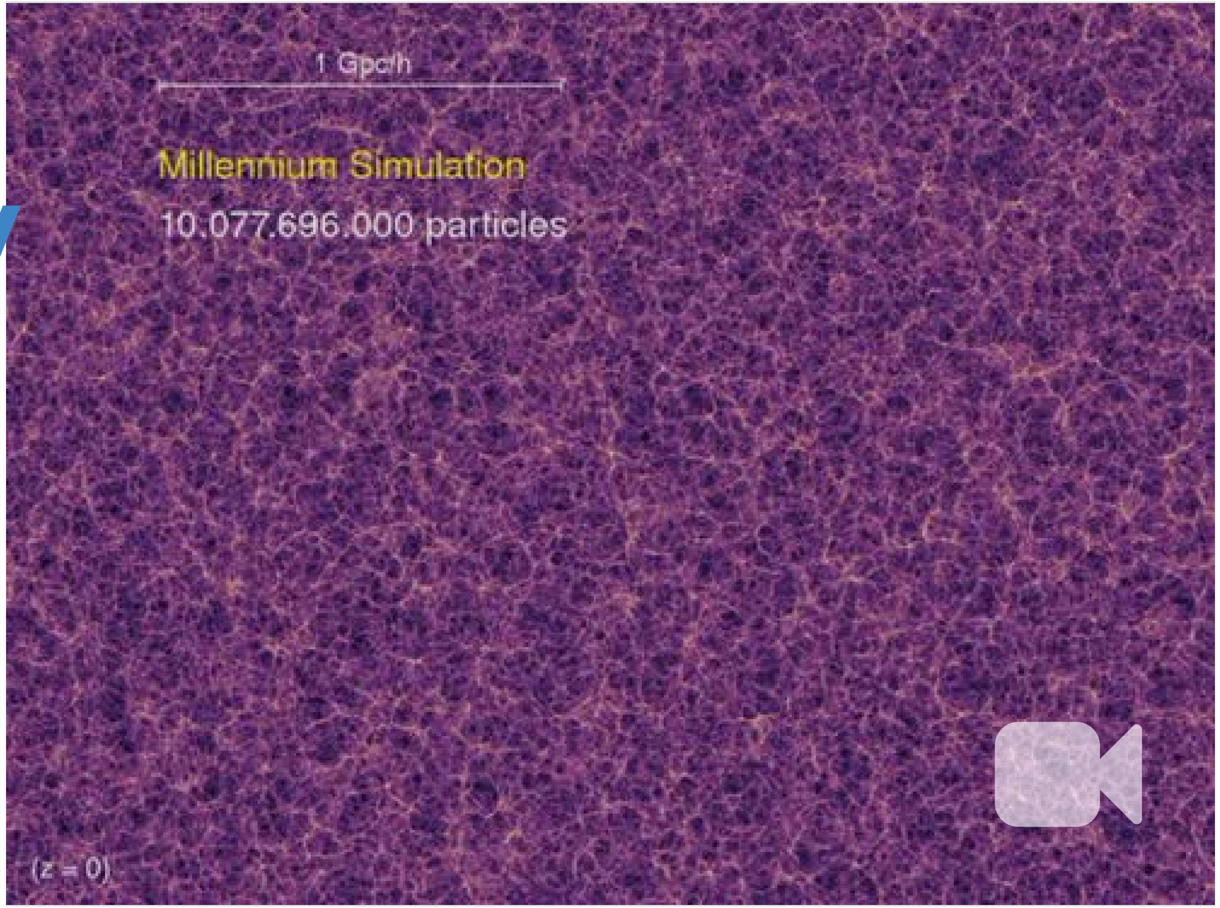


THE CURVATURE OF LIGHT: EVIDENCE FROM BRITISH OBSERVERS' PHOTOGRAPHS AT THE ECLIPSE OF THE SUN.

The results obtained by the British expeditions to observe the total eclipse of the sun last May verified Professor Einstein's theory that light is subject to gravitation. Writing in our issue of November 15, Dr. A. C. Crommelin, one of the British observers, said: "The eclipse was specially favourable for the purpose, there being no fewer than twelve fairly bright stars near the limb of the sun. The process of observation consisted in taking photographs of these stars during totality, and comparing them with other plates of the same region taken when the sun was not in the neighbourhood. Then if the starlight is bent by the sun's attraction, the stars on the eclipse plates would seem to be pushed outward compared with those on the other plates. . . . The second Sobral camera and the one used at Principe agree in supporting (Einstein's theory). . . . It is of profound philosophical interest. Straight lines in Einstein's space cannot exist; they are parts of gigantic curves."—[Drawing Copyrighted in the United States and Canada.]

Enistein 1916

what drives cosmic discovery



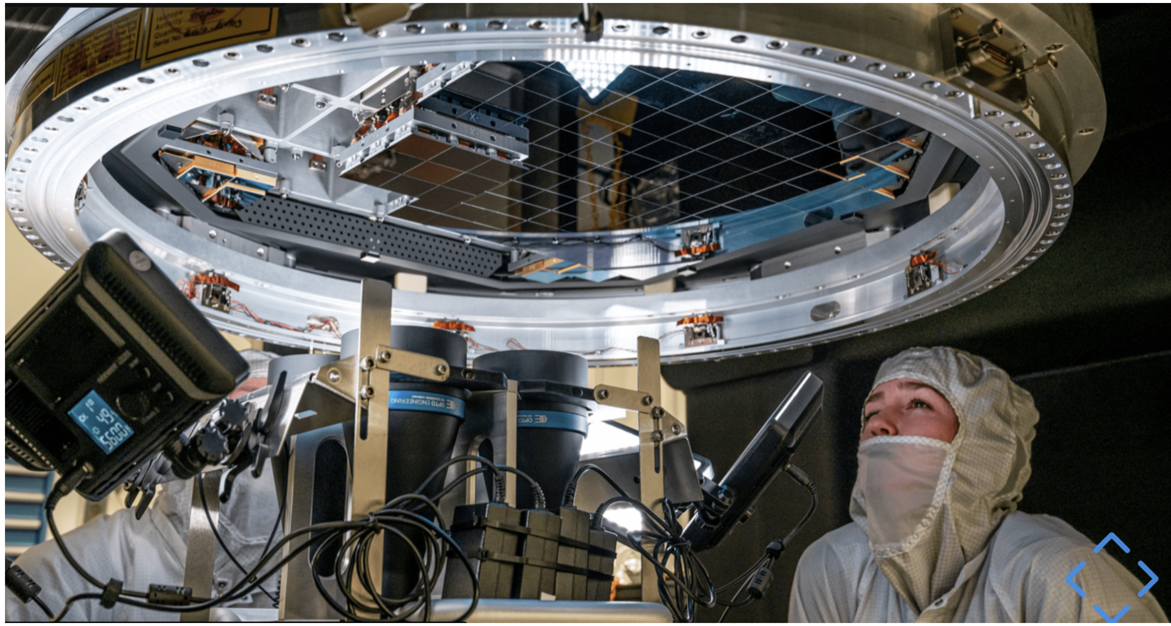
Experiment driven

Theory driven | Falsifiability

Simulations | Probabilistic inference | Computation

the 1947-today

what drives cosmic discovery



Experiment driven

Theory driven | Falsifiability

Simulations | Probabilistic inference | Computation

Data | Survey astronomy | Computation | Pattern Discovery

the 2000s-today

Big Data: Astronomical or Genomical?

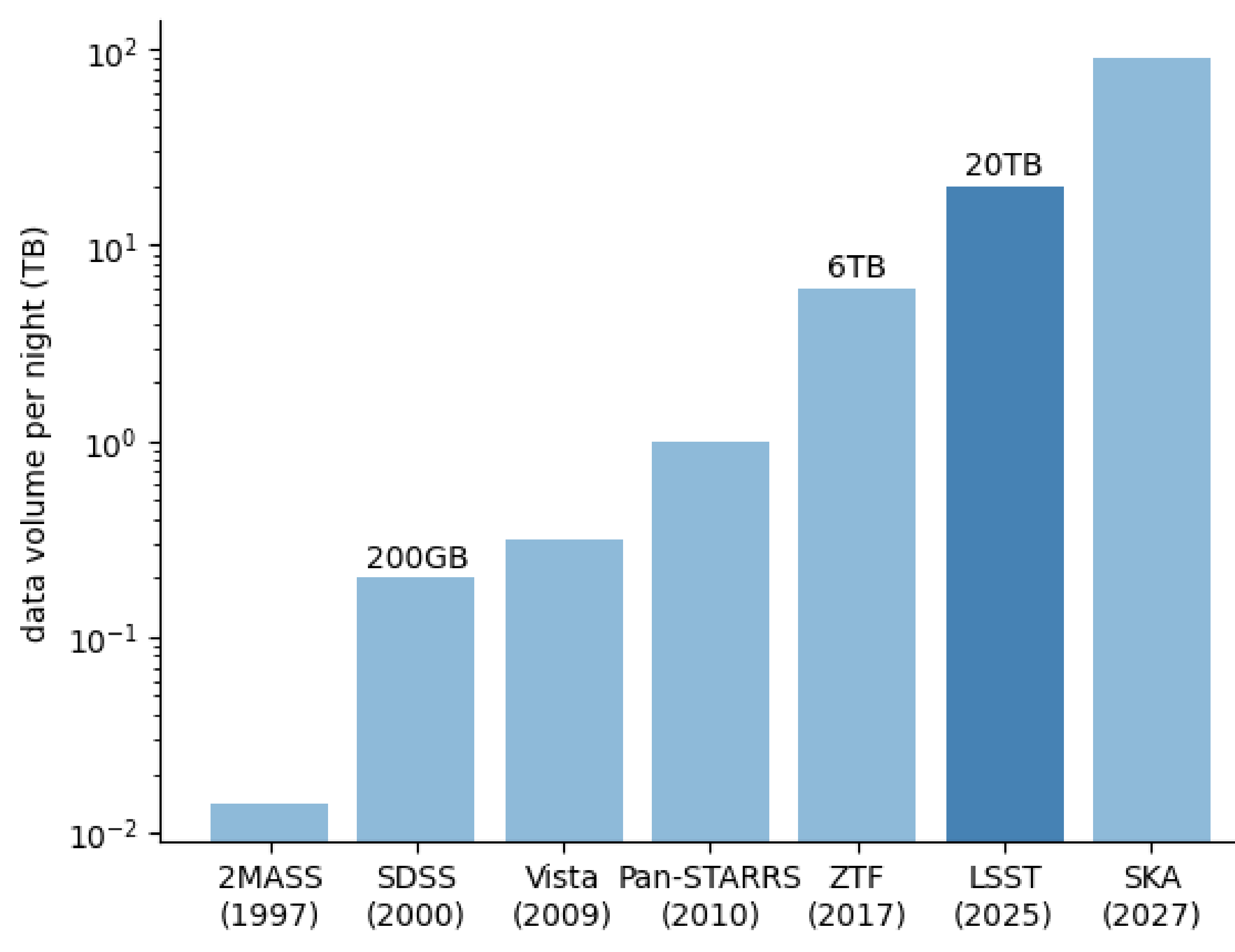
Zachary D. Stephens, Skylar Y. Lee, Faraz Faghri, Roy H. Campbell, Chengxiang Zhai, Miles J. Efron, Ravishankar Iyer,
Michael C. Schatz Saurabh Sinha Gene E. Robinson

Published: July 7, 2015 • https://doi.org/10.1371/journal.pbio.1002195

Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

Astronomy by the numbers



to scanning the sky and giving away the data (open science model!)

https://www.youtube.com/embed/7T5u3bYN5y8?enablejsapi=1

from commissioniong observation

when did the first Neural Network in astronomy review came out?

https://app.sli.do/event/qxbWnfzkJyT3SvbeKi3rwd

DISSERTATION SUMMARY

Uncertainty Reasoning in Astronomy—Applications to Classification and Optimal Telescope Scheduling[1]

HANS-MARTIN ADORF

Thesis work conducted at: Space Telescope–European Coordinating Facility
Garching b. München, Germany
Current address: Space Telescope–European Coordinating Facility, European Southern Observatory,
Karl-Schwarzschild-Straβe 2, D-85748 Garching b. München, Germany
Electronic mail: adorf@eso.org
Ph.D. dissertation directed by: Michel Crézé
Ph.D. degree awarded: 1994

The application of Artificial Neural Networks to astronomical classification

Show affiliations

Naim, Abraham

No abstract

Publication: PhD Thesis, University of Cambridge, 1995

Pub Date: January 1995

Bibcode: 1995PhDT........52N

Comments: Advisor(s): Stephen Thancy Gottesman

Aplicaciones de redes de neuronas artificiales en astronomía

Show affiliations

Serra-Ricart, Miquel

No abstract

Publication: PhD Thesis, University of La Laguna, 1993

AUTOMATED GALAXY RECOGNITION

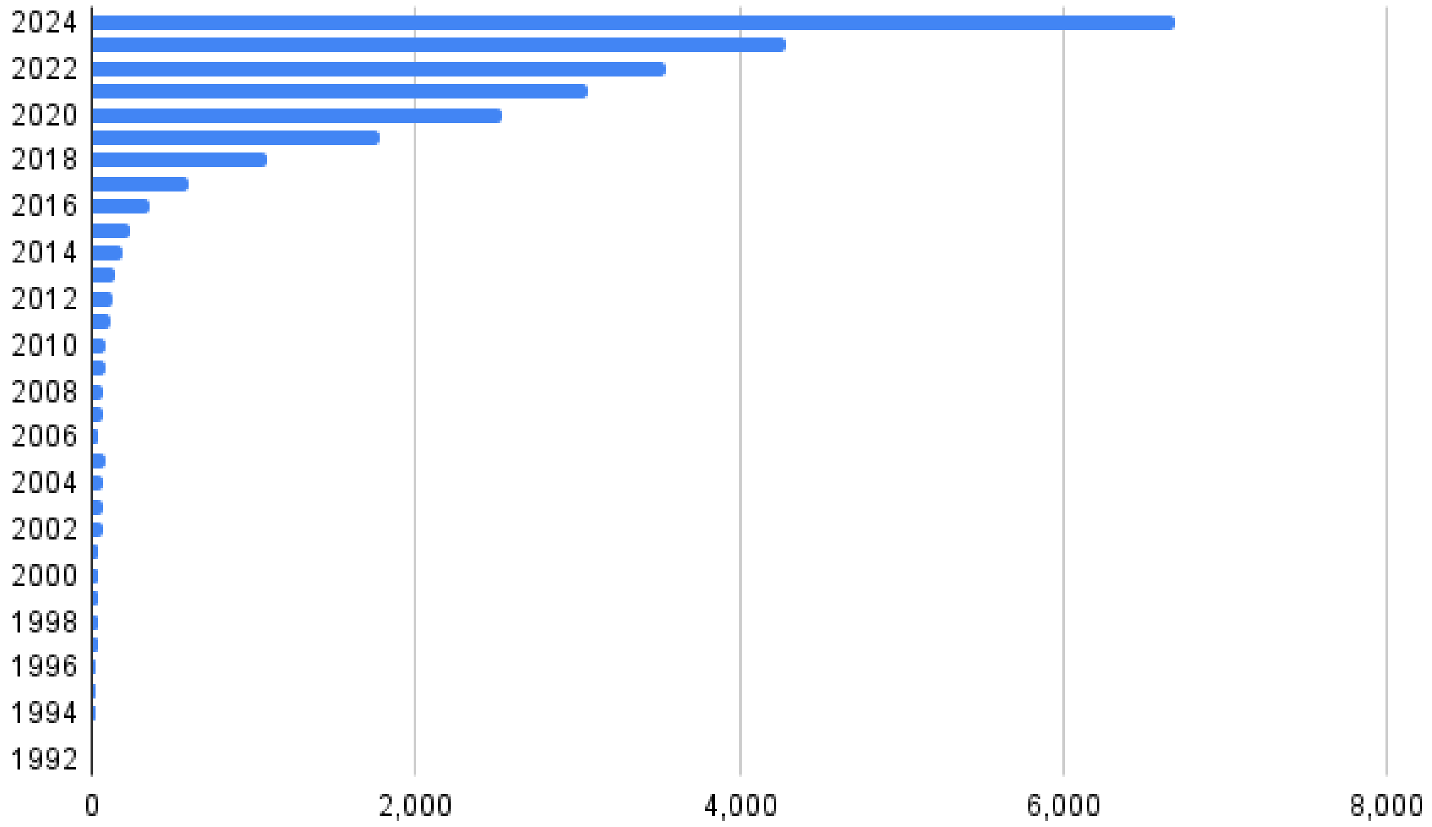
Barry Rappaport, Computing Research Laboratory and Department of Astronomy, New Mexico State University, Box 3CRL, Las Cruces, New Mexico, 88003/USA

Kurt Anderson, Department of Astronomy, Box 4500, New Mexico State University, Las Cruces, New Mexico, 88003/USA

Abstract

Previous approaches to automated image processing have used both deterministic and non-deterministic techniques. These have not used any form of conceptual learning nor have they employed artificial intelligence techniques. Addition of such techniques to the task of image processing may significantly enhance the efficiencies and accuracies of the recognition and classification processes. In our application, the objects to be recognized and classified are galaxies.

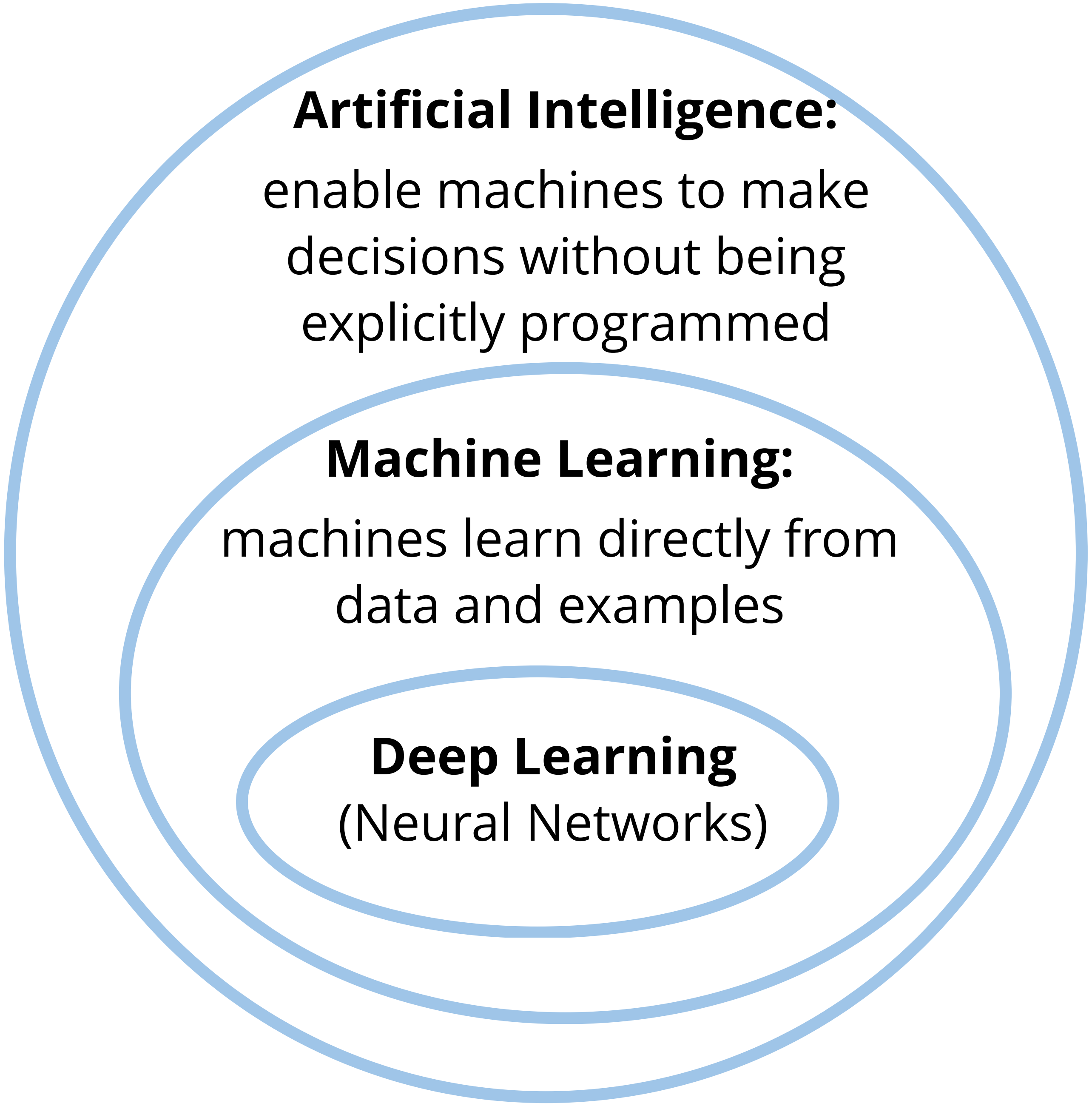
number of arXiv:astro-ph submissions with abstracts containing one or more of the strings: ‘machine learning’, ‘ML’, ‘artificial intelligence’, ‘AI’, ‘deep learning’ or ‘neural network’.



Smith+Geach May 2022 Astronomia ex machina

Data Science: the field of studies that deals with the extraction of information from data within a domain context to enable interpretation and prediction of phenomena.

This includes development and application of statistical tools and machine learning and AI methods

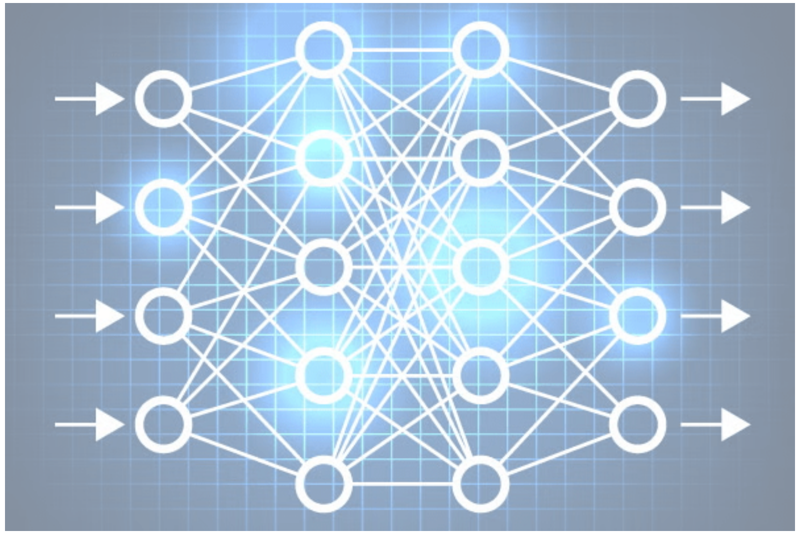


DATA

Complex Large Data

MODEL

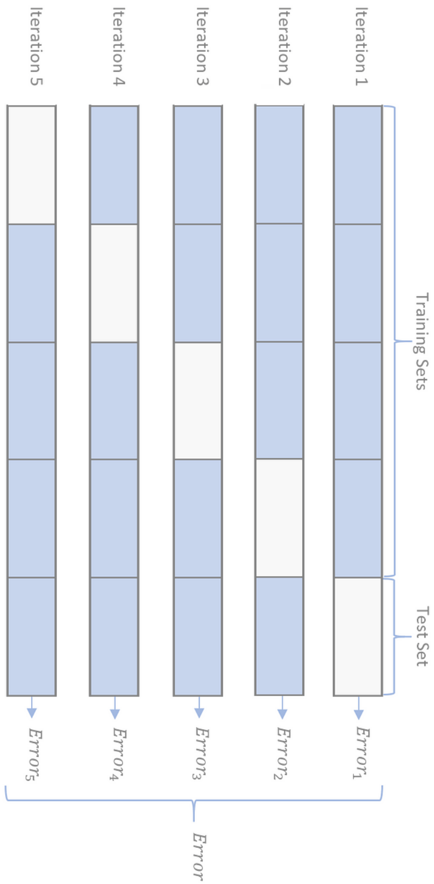
Flexible non-linar models



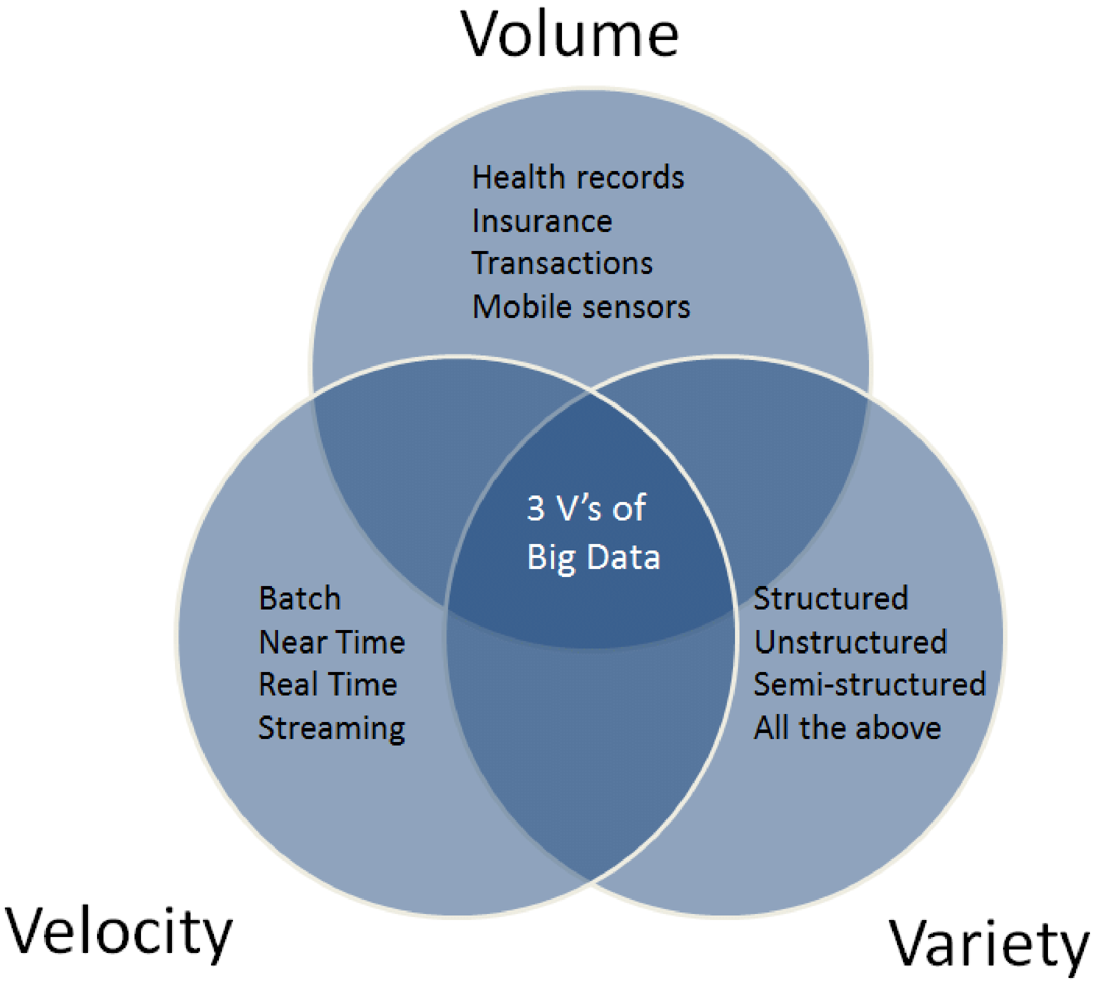
PRACTICE

whitening

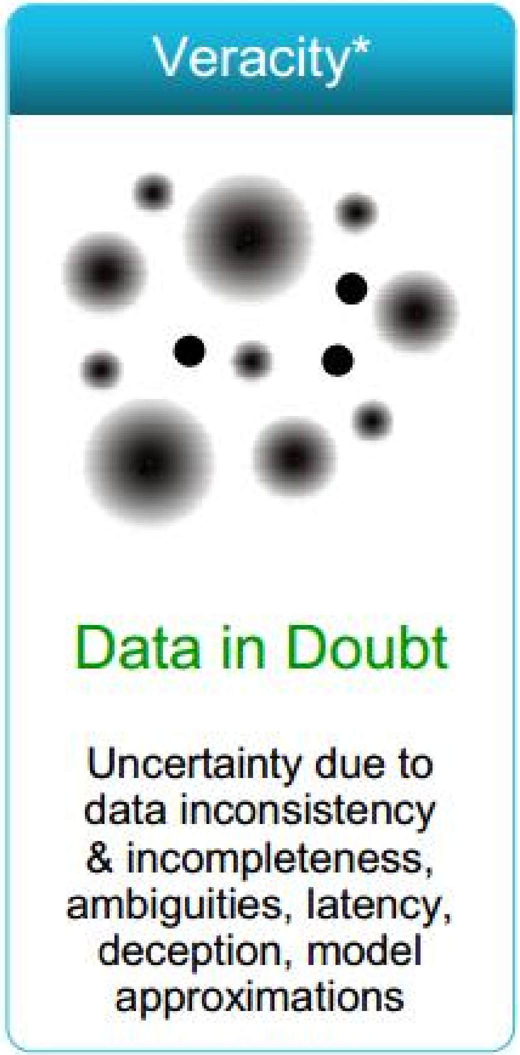
cross validation



4-V of Big Data



Gartner report 2001



Michael Walker on 28 November 2012

4-V of Big Data

V1: Volume

Number of bites

Number of pixels

Number of astrophysical objects in a data x number of featured measured

V2: Variety

Diverse science return from the same dataset

e.g. cosmology+stellar physics

Multiwavelength

Multimessenger

Images and spectra

V3: Velocity

Real time analysis, edge computing, data transfer

IceCube edge computing

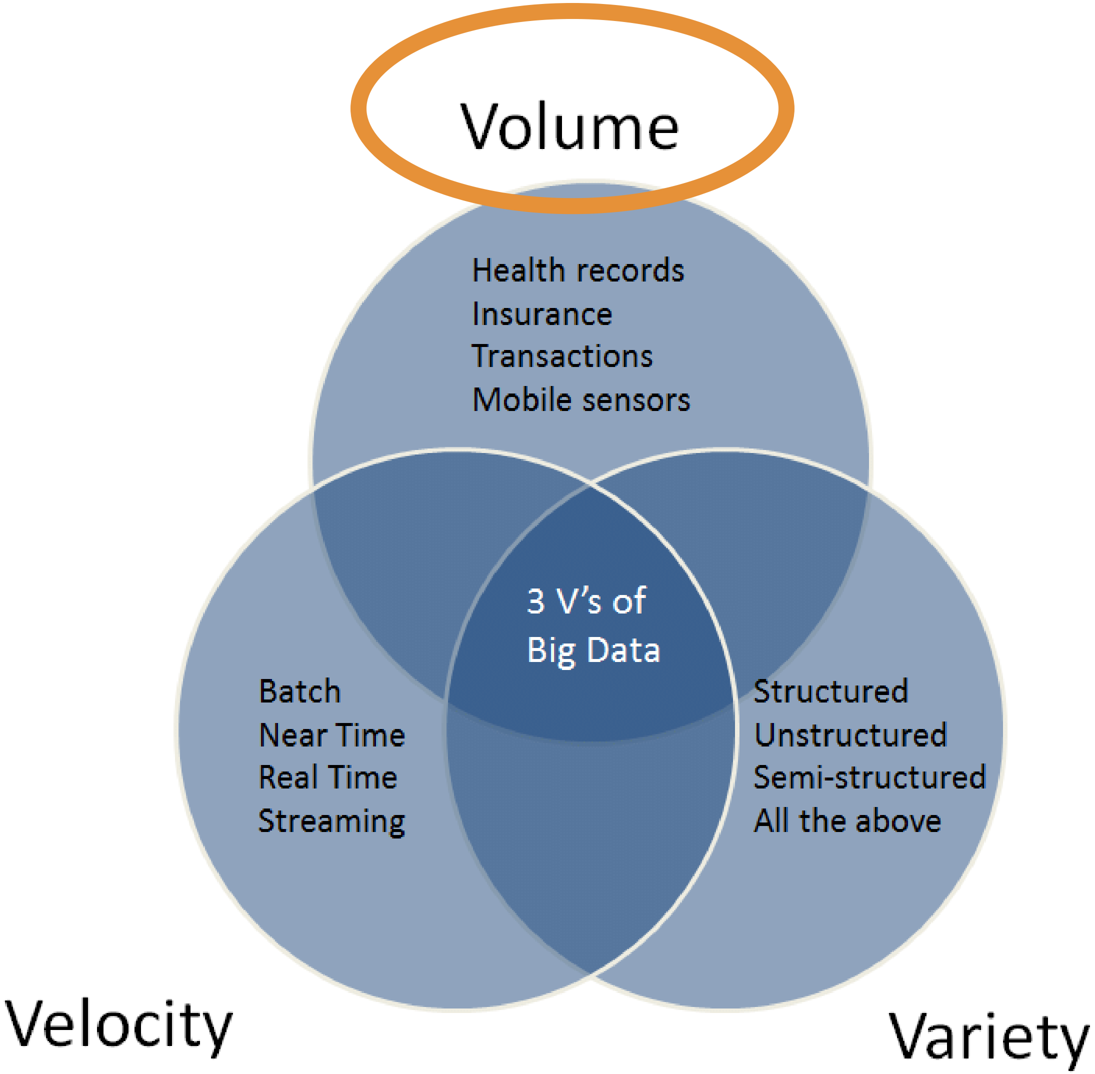
V4: Veracity

This V will refer to both data quality and availability (added in 2012)

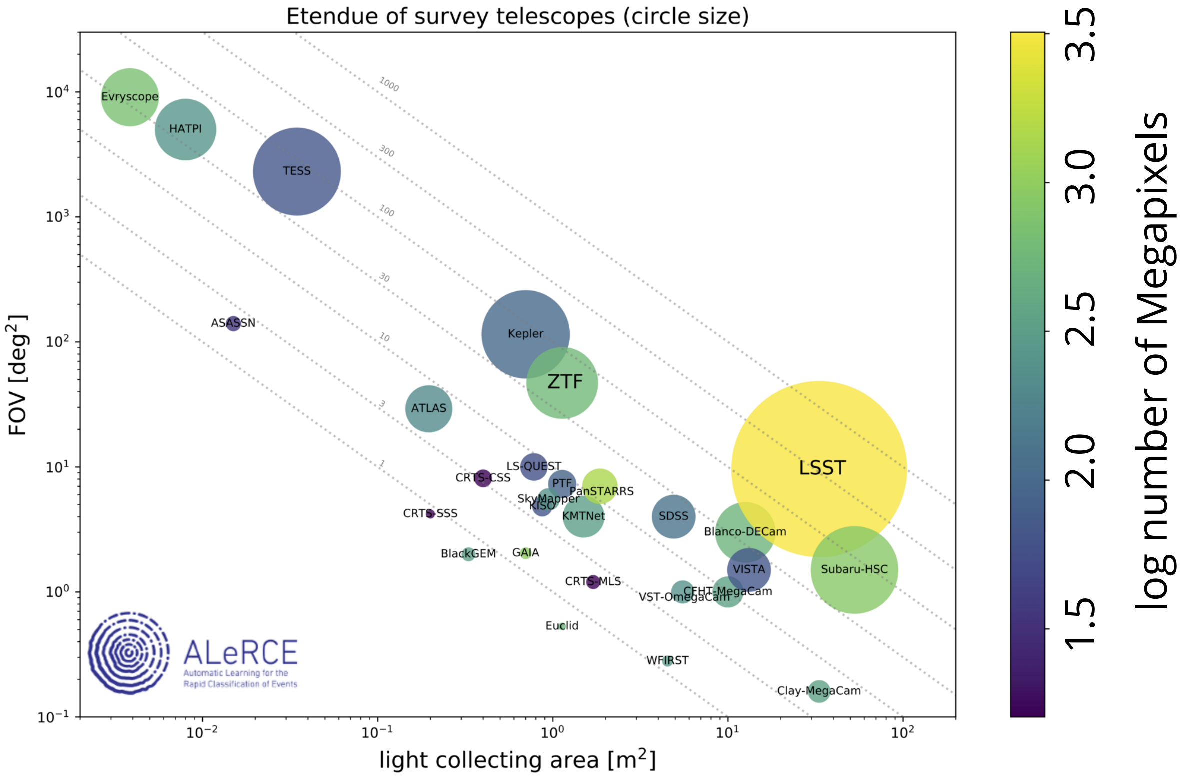
Inclusion of uncertainty in inference and simulations

Gartner report 2001

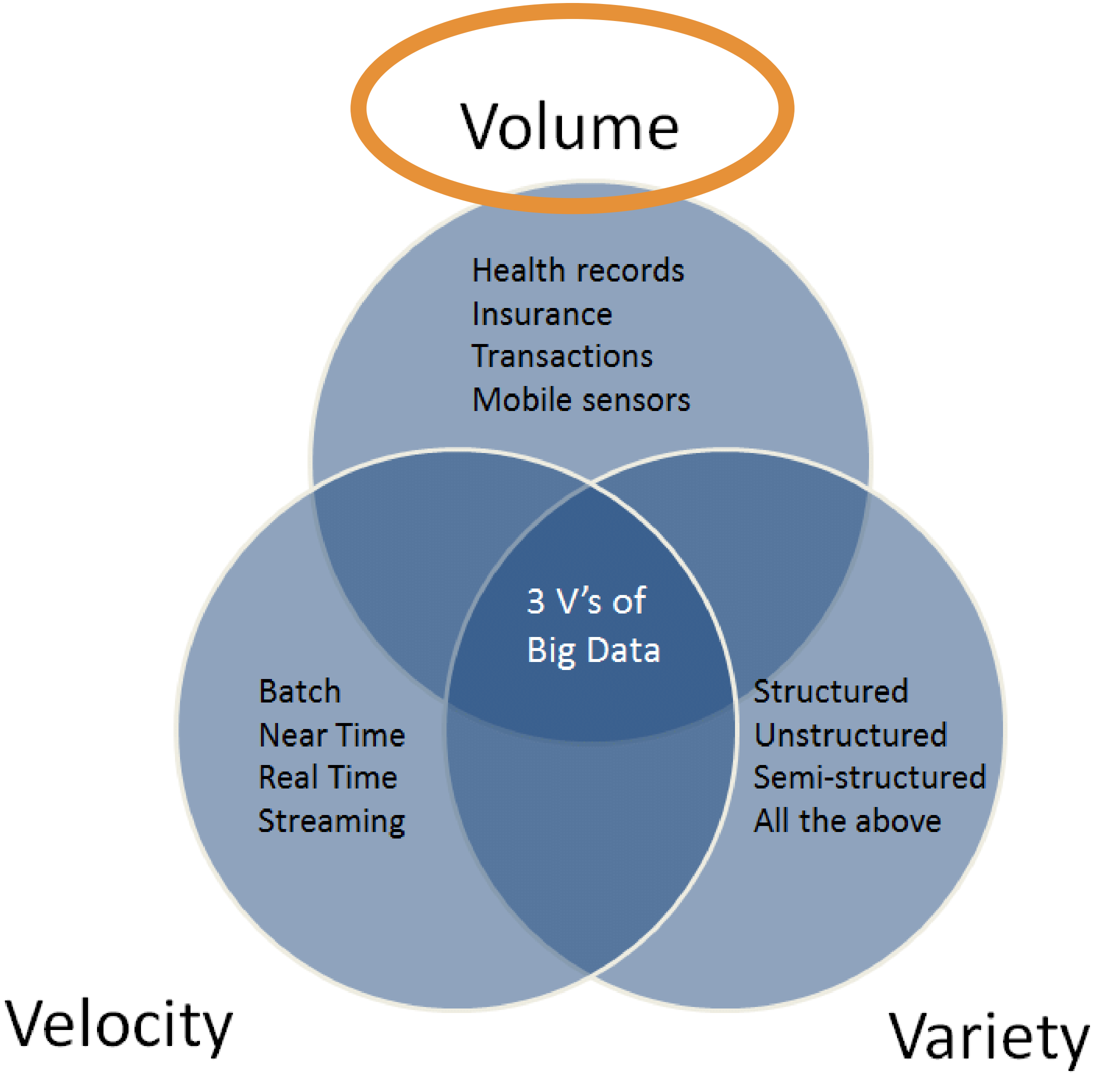
4-V of Big Data



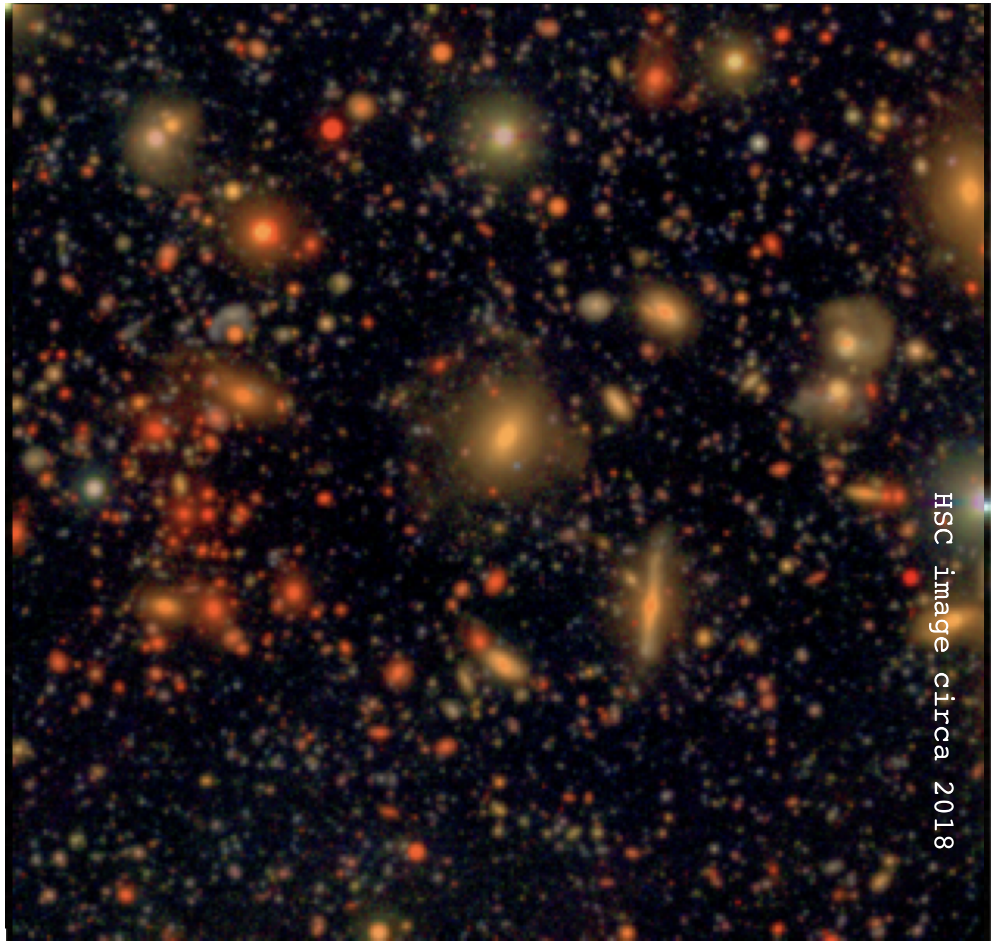
Etendue: *area x FoV*



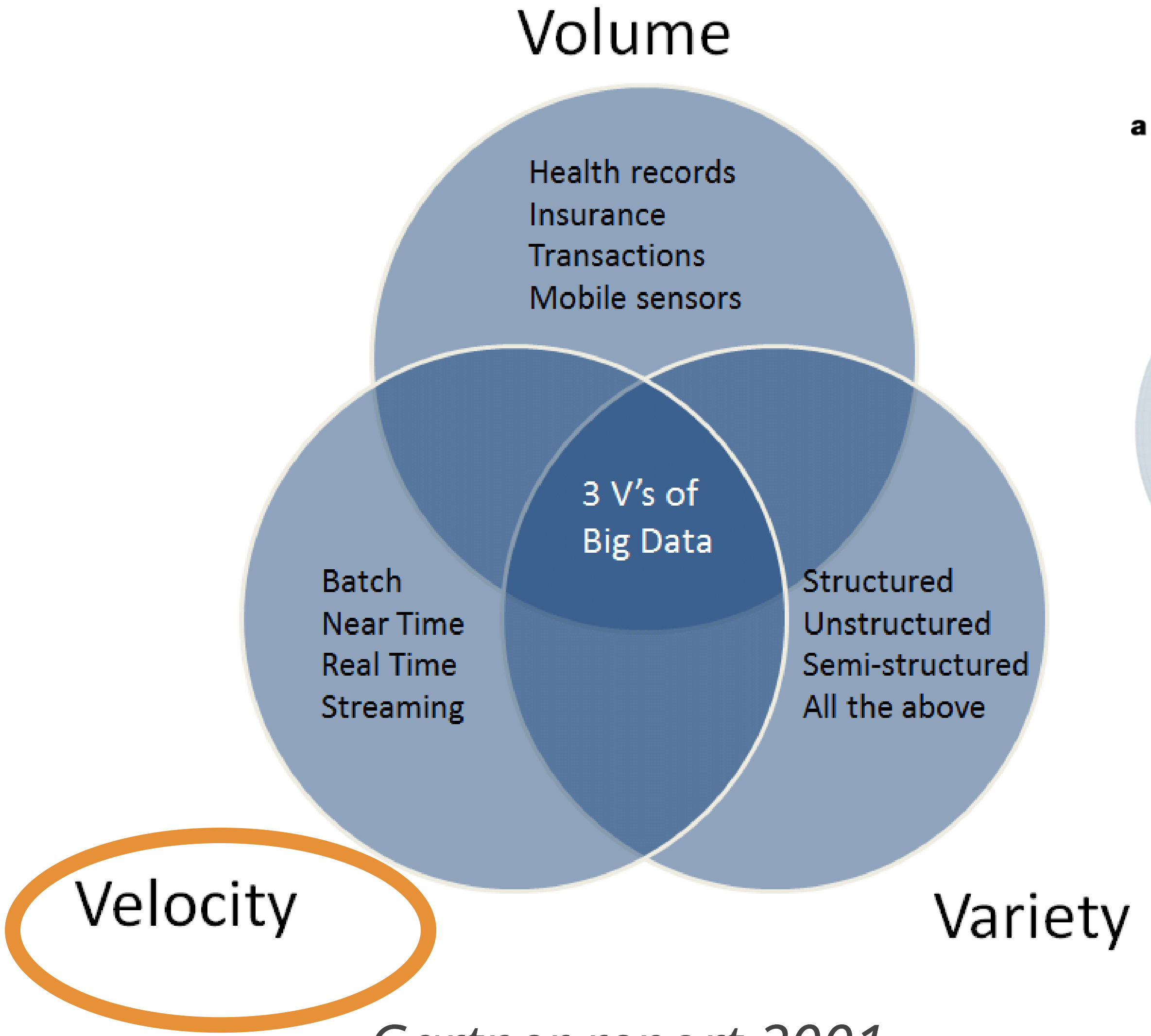
4-V of Big Data



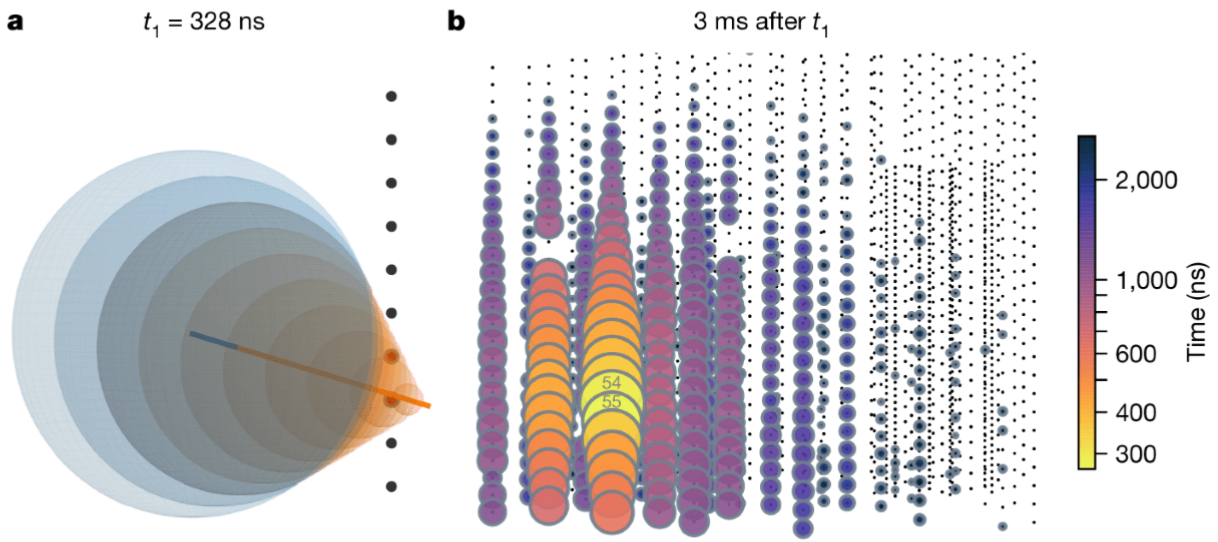
Gartner report 2001



4-V of Big Data



Gartner report 2001



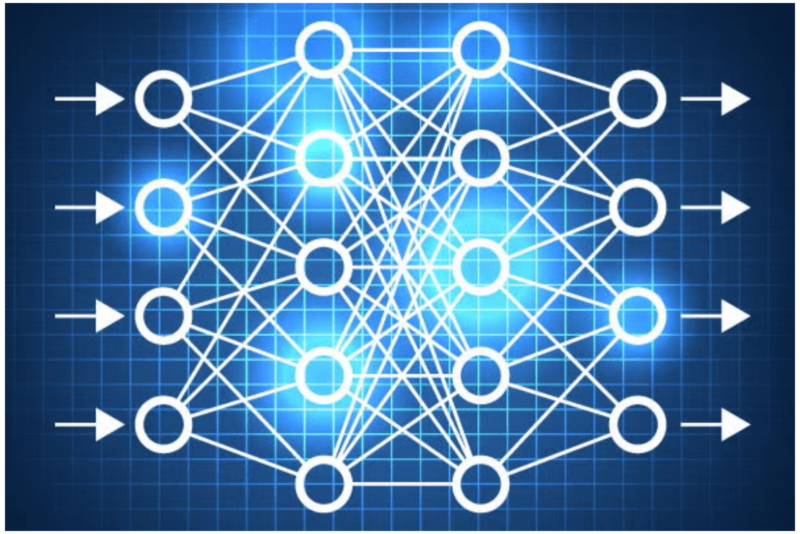
The IceCube collaboration
Nature 591, 220–224 (2021)

DATA

Complex Large Data

MODEL

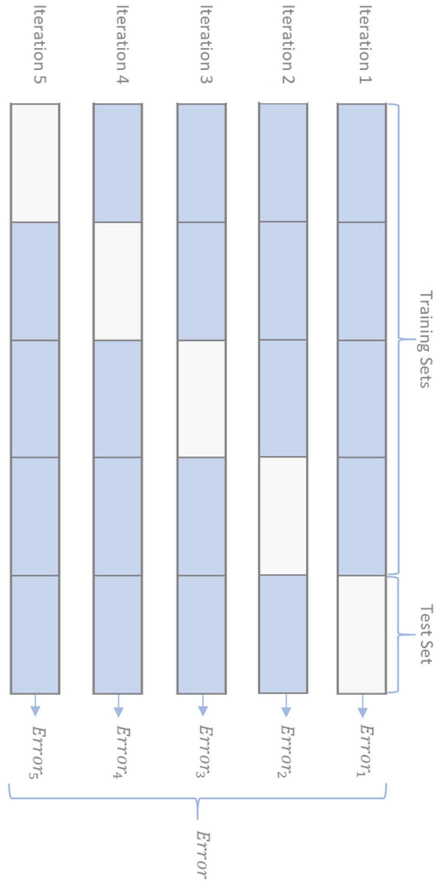
Flexible non-linar models



PRACTICE

whitening

cross validation



what is machine learning?

[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel, 1959

What is a model in ML

a model is a low
dimensional representation
of a higher dimensionality
dataset

What is a model in ML

ML: any model with parameters learnt from the data

dimensionality of the model: number of parameters

https://miro.medium.com/max/960/1*imhEKEpzX24CC_LIlureBw.gif

what is machine learning?

Machine Learning models are parametrized representation of "reality" where the parameters are learned from finite sets (the *sample*) of realizations of that reality (the *population*)

(note: learning by instance, e.g. nearest neighbours, may not comply to this definition)

Machine Learning is the disciplines that conceptualizes, studies, and applies those models.

Key Concept

DATA

Complex Large Data

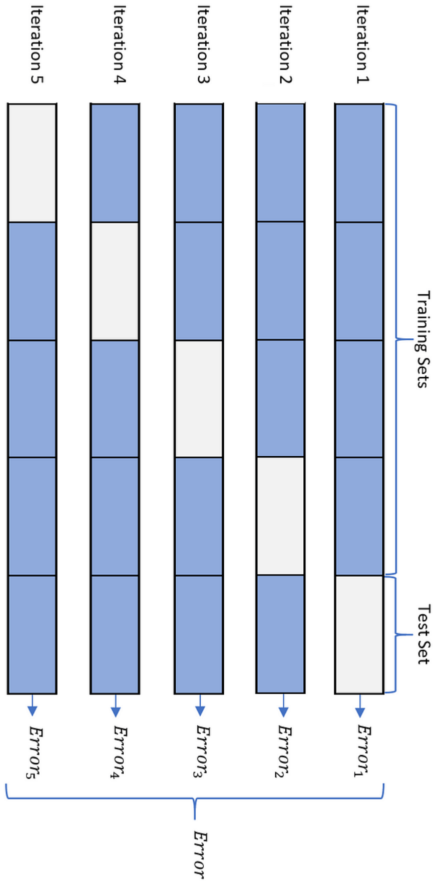
MODEL

Flexible non-linar models

PRACTICE

whitening

cross validation



DATA MODEL PRACTICE

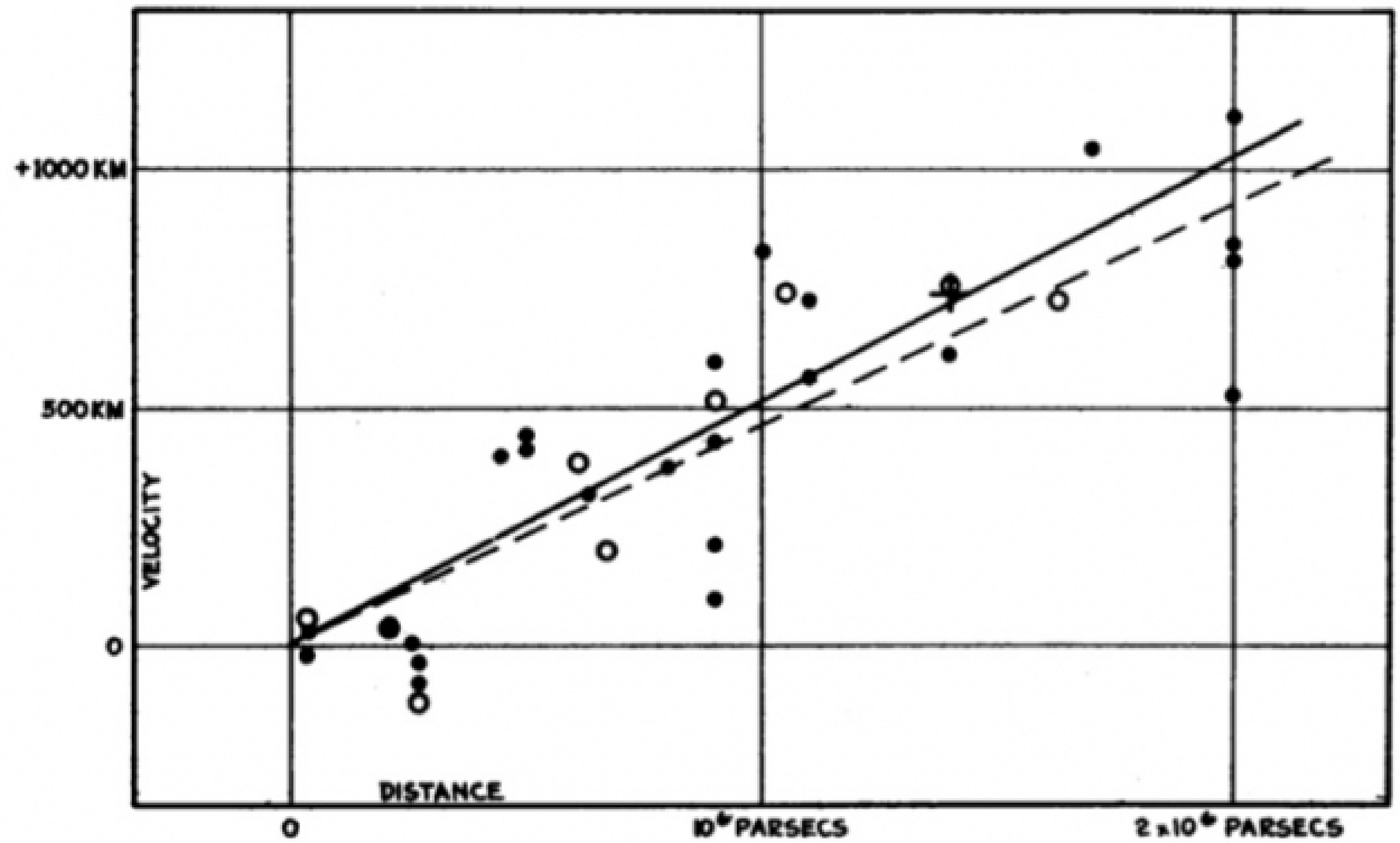
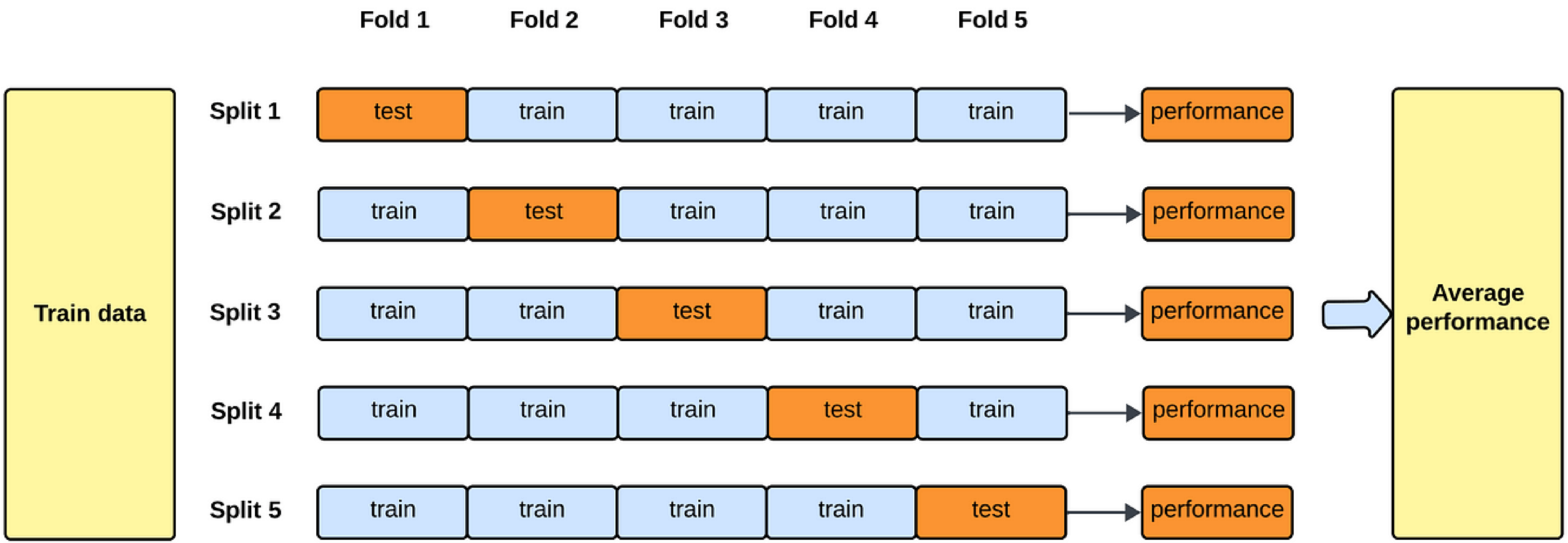


FIGURE 1

Velocity-Distance Relation among Extra-Galactic Nebulae.

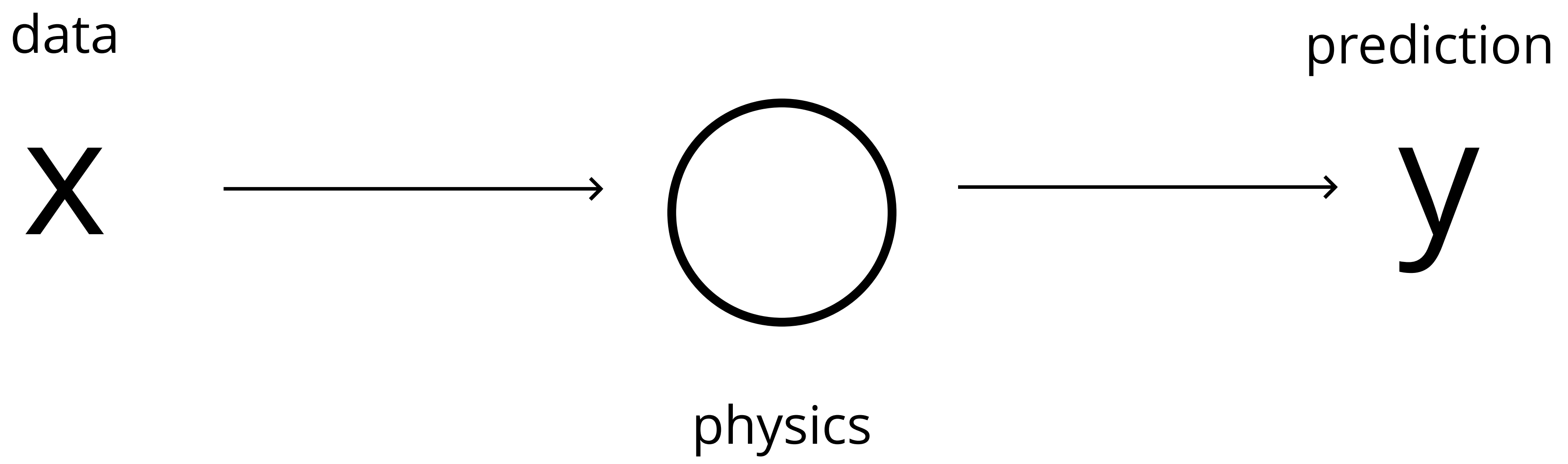
Hubble in 1929

Emphasis on transferability of the model to unseen data



2/6

AI "Learning"


data
x
prediction
y
physics

Input

output

$$x \longrightarrow f(x) \longrightarrow y$$

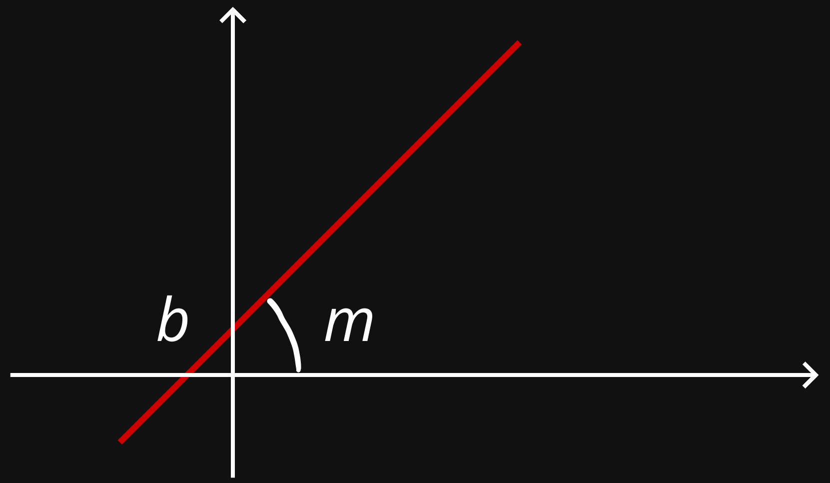
function

Input

$$X \longrightarrow f(x) \longrightarrow y$$

output

$$f(x) = mx + b$$



m: slope

b: intercept

Input

x $\longrightarrow f(x) \longrightarrow$ y

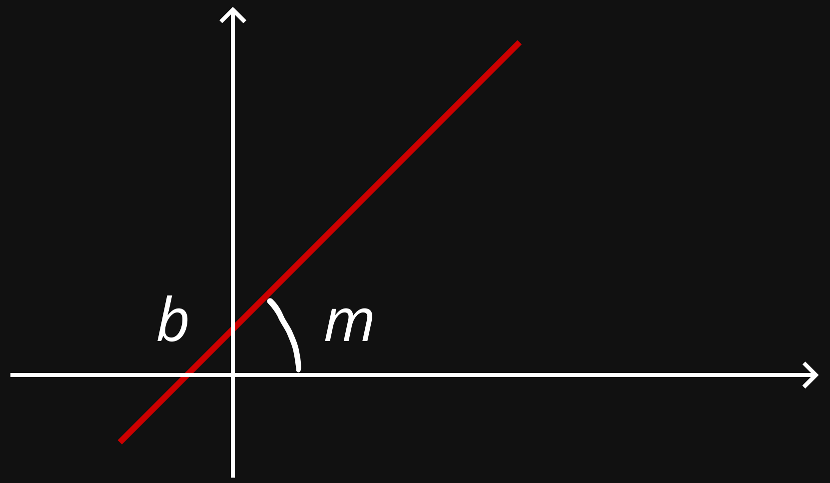
output

$$f(x) = mx + b$$

m: slope

b: intercept

$\Big\}$ parameters

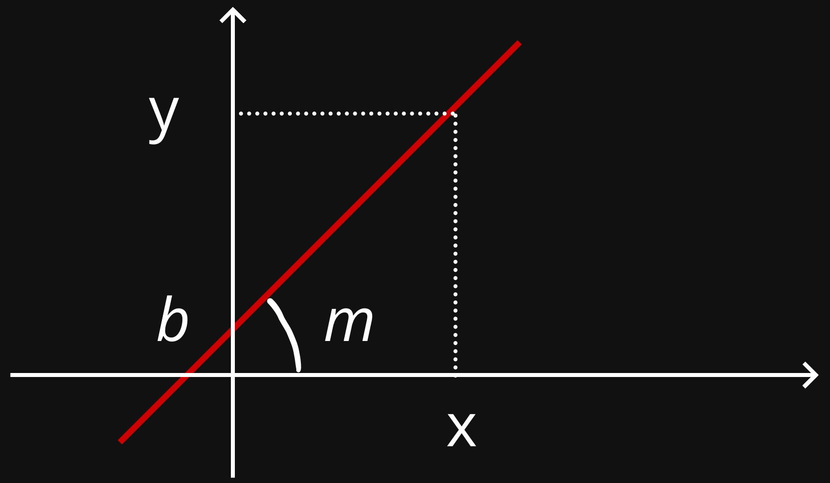


goal: find the right *m* and *b* that turn **x** into **y**

Input		output
$\mathbf{x}$	$\longrightarrow f(x) \longrightarrow$	$\mathbf{y}$

$$f(x) = mx + b$$

$$\left.\begin{array}{l}\text{m: slope}\\ \textit{b}\text{: intercept}\end{array}\right\}\ \text{parameters}$$



learn

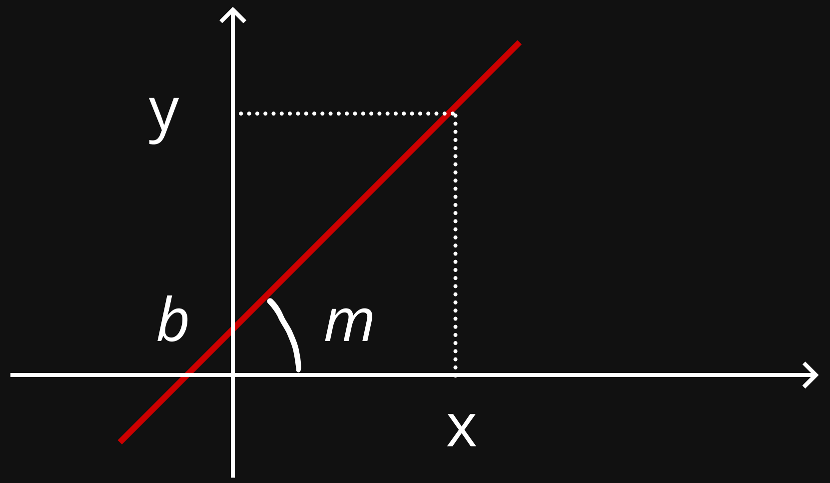
goal: ~~find~~ the right *m* and *b* that turn **x** into **y**

Input				output
x	$\longrightarrow$	$f(x)$	$\longrightarrow$	y

$$f(x) = mx + b$$

m: slope
b: intercept

} parameters



goal: learn the right *m* and *b* that turn **x** into **y**

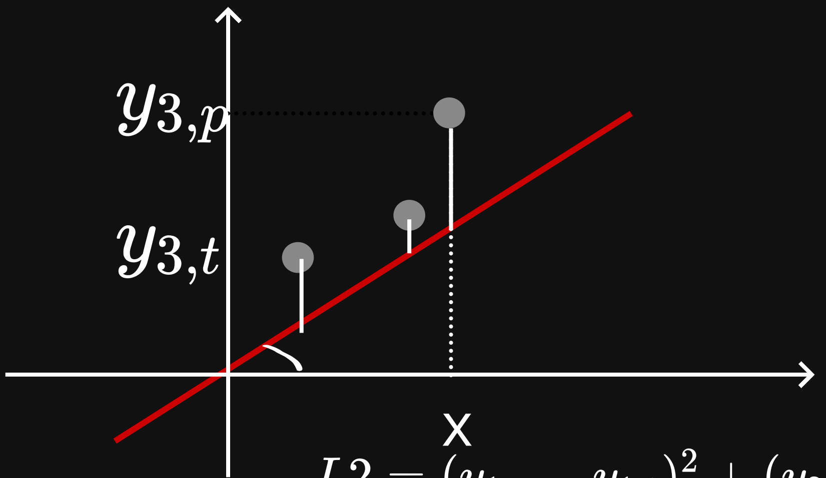
Input $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ output

$$\vec{\mathrm{x}} \longrightarrow f(x) \longrightarrow \vec{\mathrm{y}}$$

$$f(x) = mx + b$$

m: slope
b: intercept

$\Big\}$ parameters



let's try

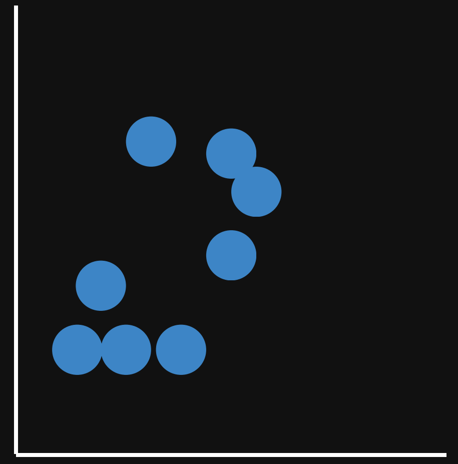
m = 0.4 and b=0

$$L2 = (y_{1,p} - y_{1,t})^2 + (y_{2,p} - y_{2,t})^2 + (y_{3,p} - y_{3,t})^2$$

unsupervised vs supervised learning

understand the structure of a feature space

Clustering



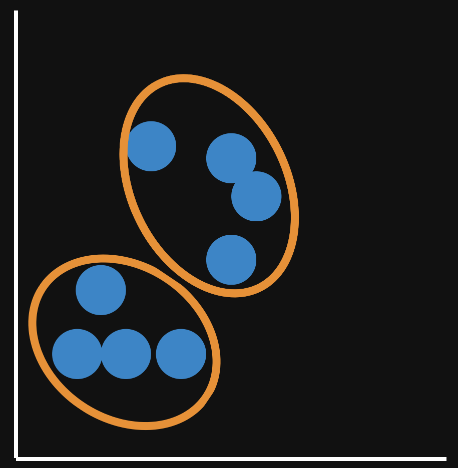
All features are observed for all datapoints

unsupervised vs supervised learning

understand the structure of a feature space

Unsupervised learning

- understanding structure



Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

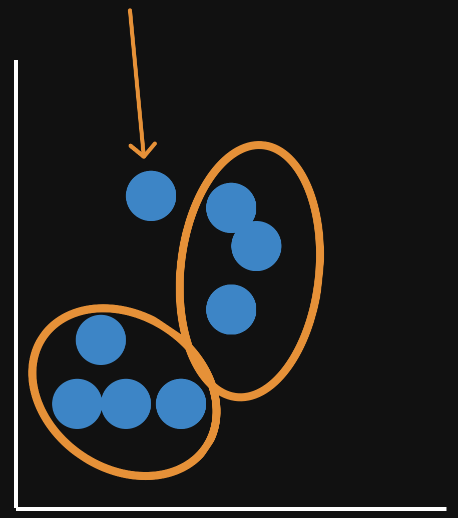
All features are observed for all datapoints

unsupervised vs supervised learning

understand the structure of a feature space

Unsupervised learning

- understanding structure
- *anomaly detection*



Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

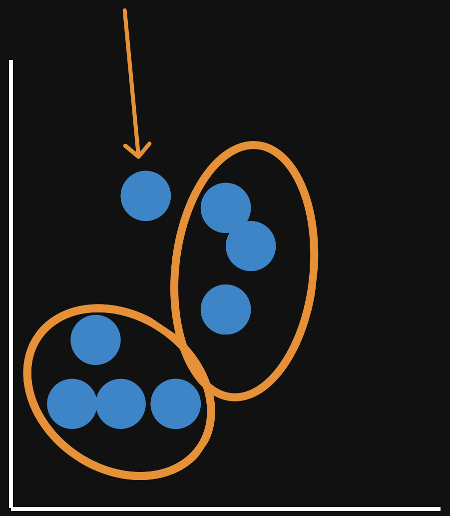
All features are observed for all datapoints

unsupervised vs supervised learning

understand the structure of a feature space

Unsupervised learning

- understanding structure
- *anomaly detection*



Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

All features are observed for all datapoints

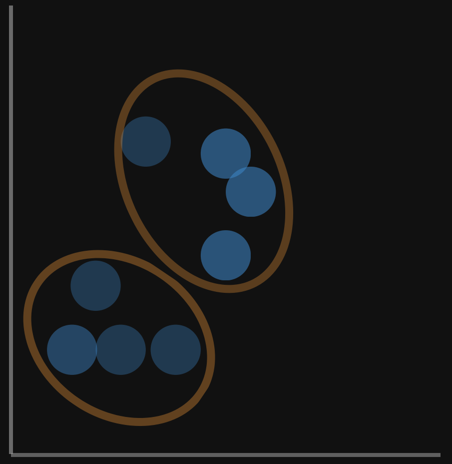
unsupervised vs **supervised** learning

prediction and classification based on examples

Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction

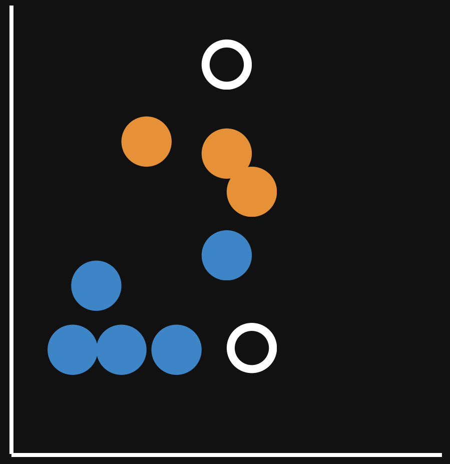
Clustering



partitioning the feature space so that the existing data is grouped (according to some target function!)

All features are observed for all datapoints

Classifying & regression



Some features not observable & we want to predict them.

unsupervised vs **supervised** learning

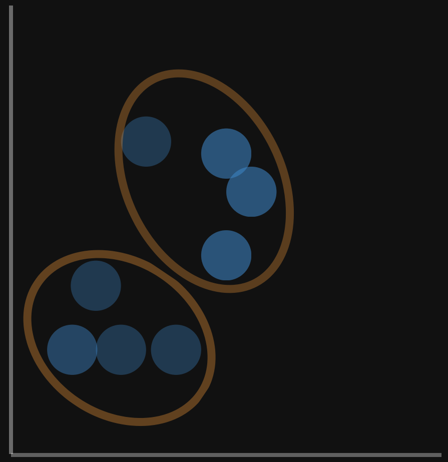
prediction and classification based on examples

Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction

Clustering

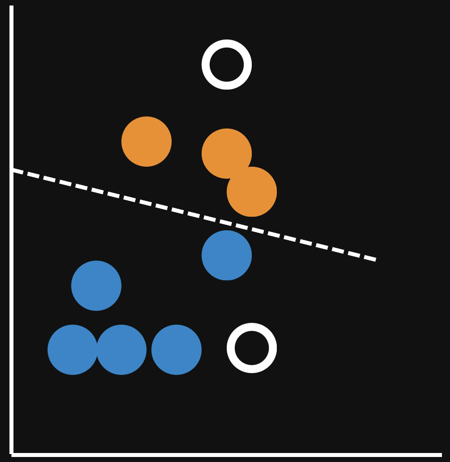
partitioning the feature space so that the existing data is grouped (according to some target function!)



All features are observed for all datapoints

Classifying & regression

finding functions of the variables that allow to predict unobserved properties of new observations



Some features not observable & we want to predict them.

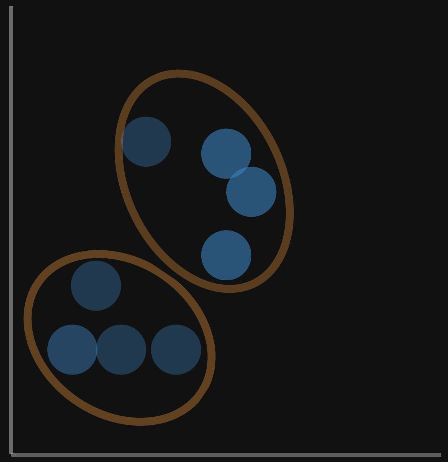
unsupervised vs **supervised** learning

prediction and classification based on examples

Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction

Clustering



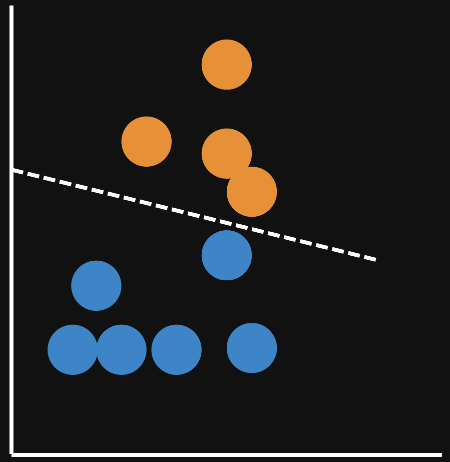
partitioning the feature space so that the existing data is grouped (according to some target function!)

All features are observed for all datapoints

Supervised learning

- classification (prediction)
- regression (prediction)

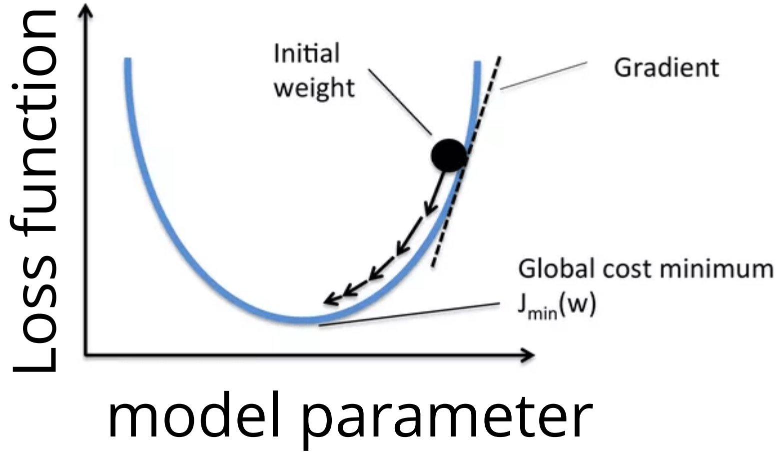
Classifying & regression



finding functions of the variables that allow to predict unobserved properties of new observations

Some features not observable & we want to predict them.

The Loss function



Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction

$$L = \sum_{i,c} |\vec{x}_{i \in c} - \vec{\mu}_c|^2$$

Supervised learning

- classification (prediction)
- regression (prediction)
- feature selection

$$L1 = \sum_i |y_i - f(\vec{x}_i)|$$

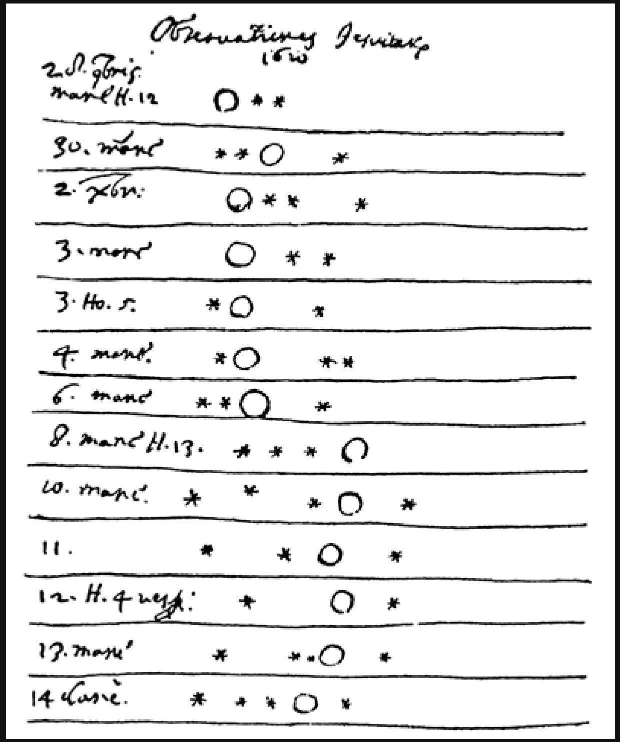
$$L2 = \sum_i |y_i - f(\vec{x}_i)|^2$$

Physics informed AI

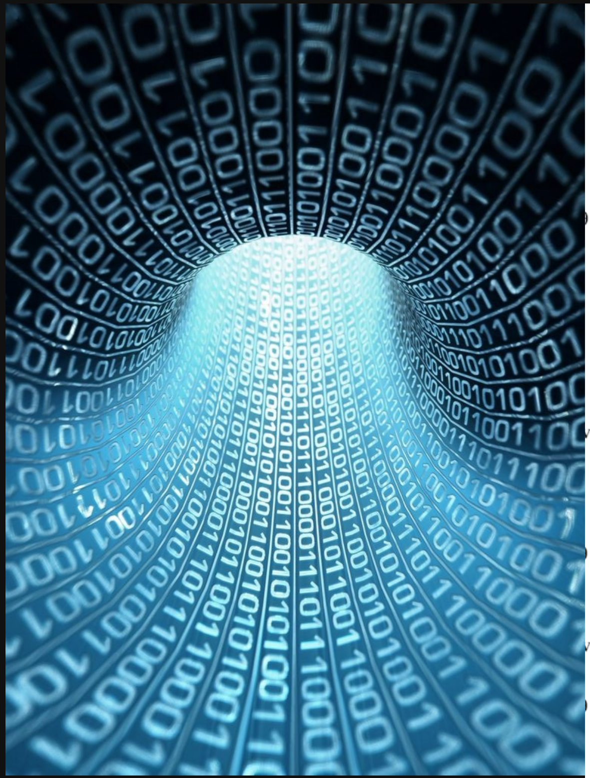
PiNN

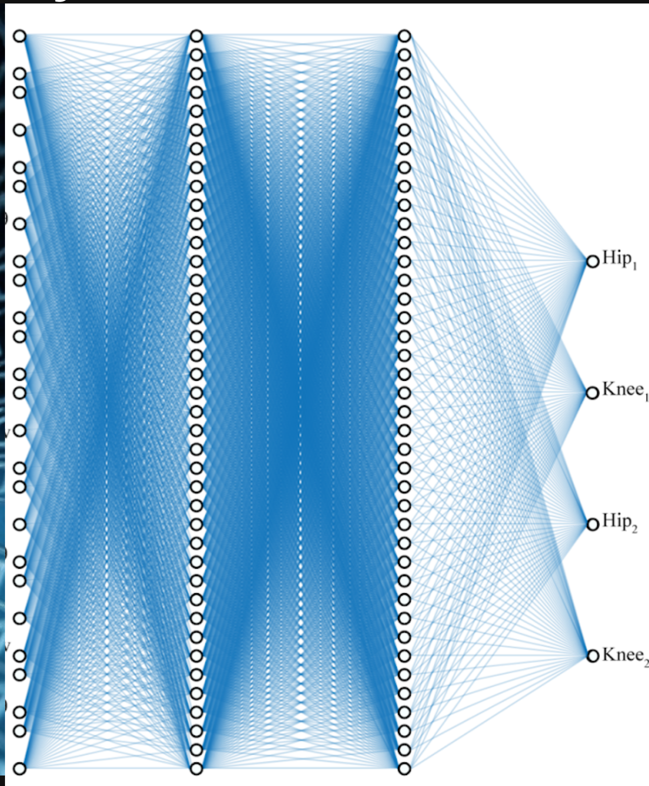
Application regime:

-infinity - 1950's



-1980's - today





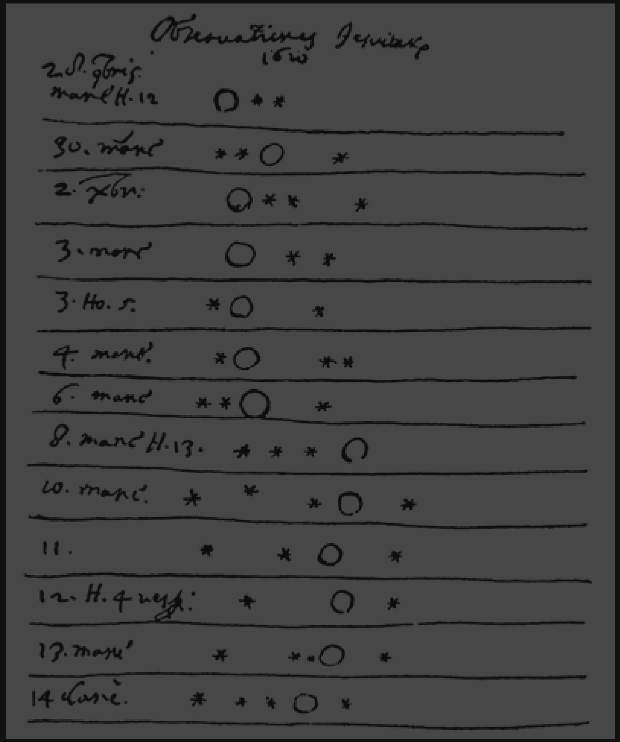
theory driven: *little data, mostly theory, falsifiability and all that...*

data driven: *lots of data, drop theory and use associations, black-box modles*

PINN

Application regime:

-infinity - 1950's

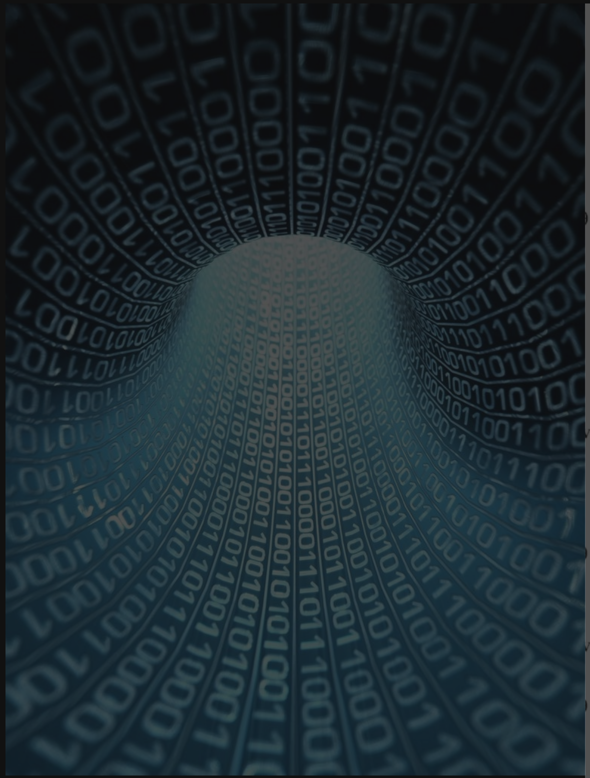


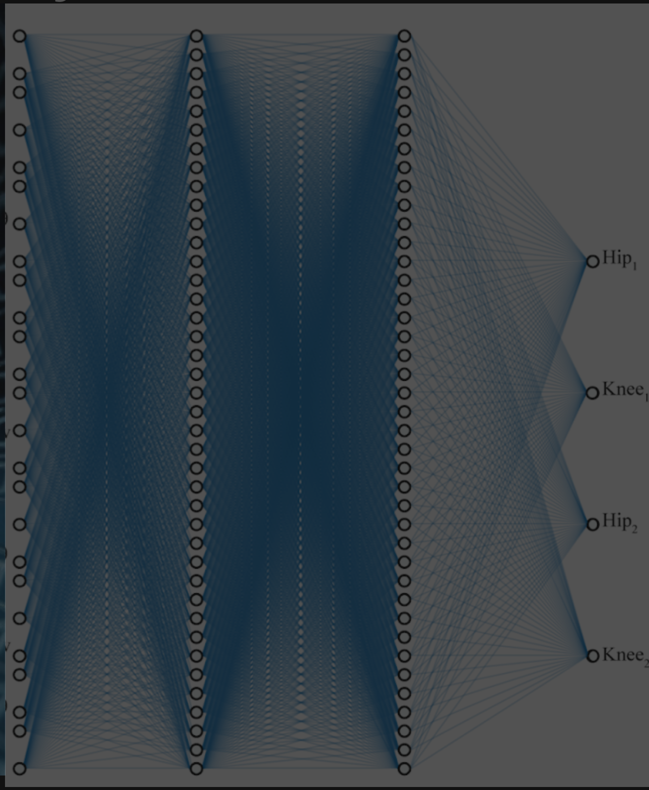
lots of data yet not enough for entirely automated decision making

complex theory that cannot be solved analytically

combine it with some theory

-1980's - today





***theory driven:** little data, mostly theory, falsifiability and all that...*

***data driven:** lots of data, drop theory and use associations, black-box modles*

PiNN

Non Linear PDEs are hard to solve!

- Provide training points at the boundary with calculated solution (trivial cause we have boundary conditions)

- Provide the physical constraint: make sure the solution satisfies the PDE

via a modified loss function that includes residuals of the prediction and residual of the PDE

$$\text{loss} = L2 + PDE = \sum (u_\theta - u)^2 + (\partial_t u_\theta + u_\theta\, \partial_x u_\theta - (0.01/\pi)\, \partial_{xx} u_\theta)^2$$

PiNN

Non Linear PDEs are hard to solve!

https://www.youtube.com/embed/AYR5dKMgdXY?enablejsapi=1

$$\text{loss} = L2 + PDE = $$
$$\sum (u_\theta - u)^2 + $$
$$(\partial_t u_\theta + u_\theta \, \partial_x u_\theta - (0.01/\pi) \, \partial_{xx} u_\theta)^2$$

Raissi, Perdikaris, Karniadakis 2017

different flavors of learning

Unsupervised learning

All features are observed for all datapoints

and we are looking for structure in the feature space

Supervised learning

Some features are not observed for some data points we want to predict them.

The datapoints for which the target feature is observed are said to be "*labeled*"

also...

Semi-supervised learning

A small amount of labeled data is available. Data is cluster and clusters inherit labels

Active learning

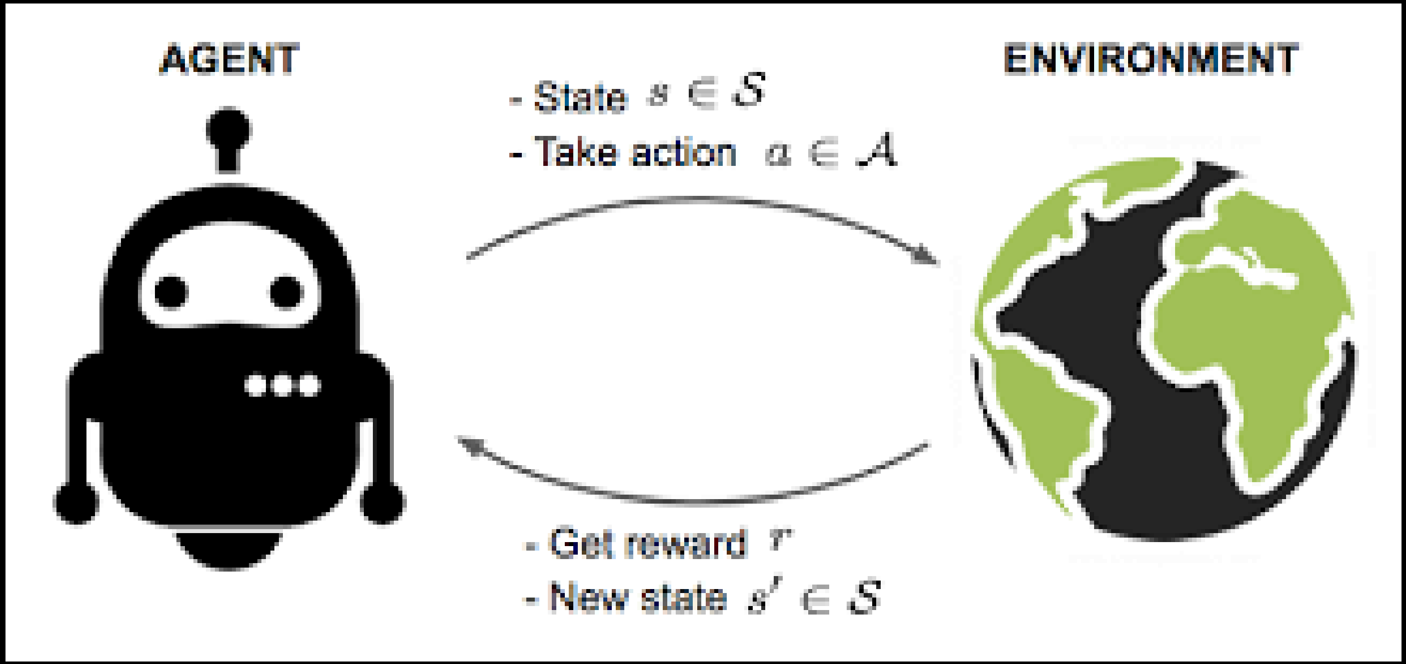
The code can interact with the user to update labels and update model.

different flavors of learning

Reinforcement Learning

reward vs loss

delayed feedback from the changes in the environment

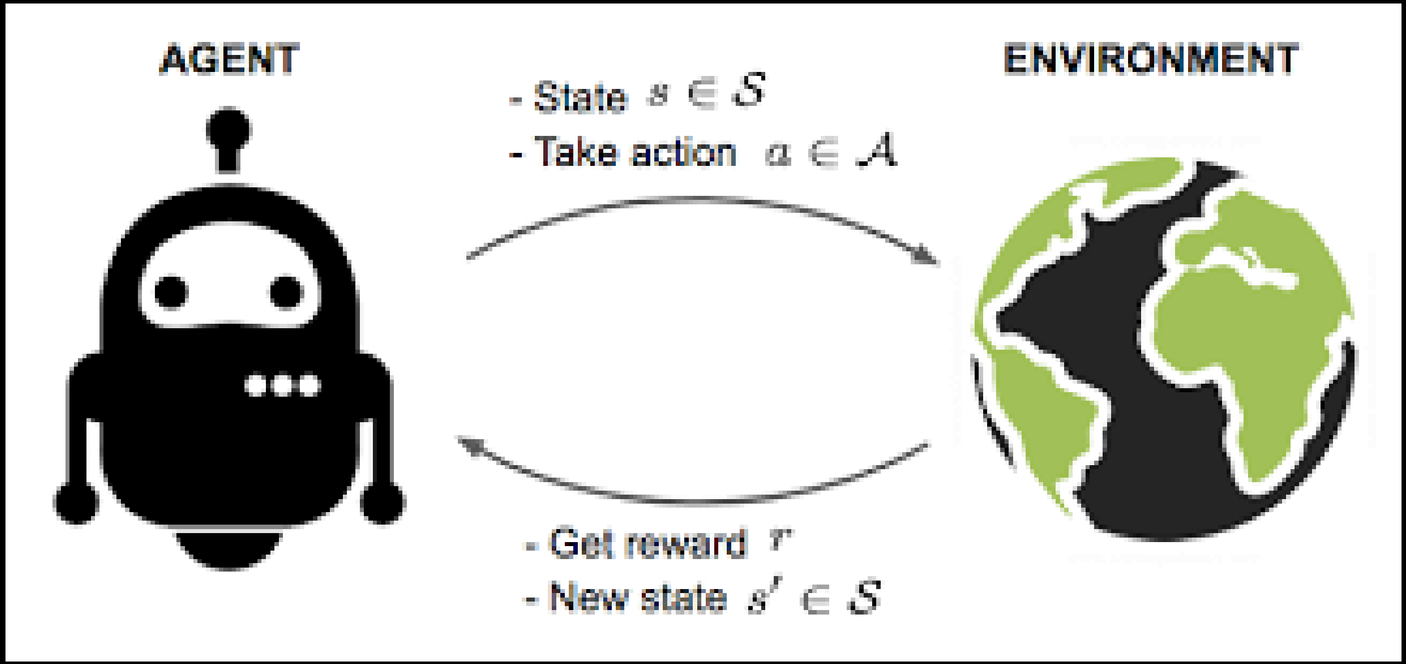


different flavors of learning

Reinforcement Learning

reward vs loss

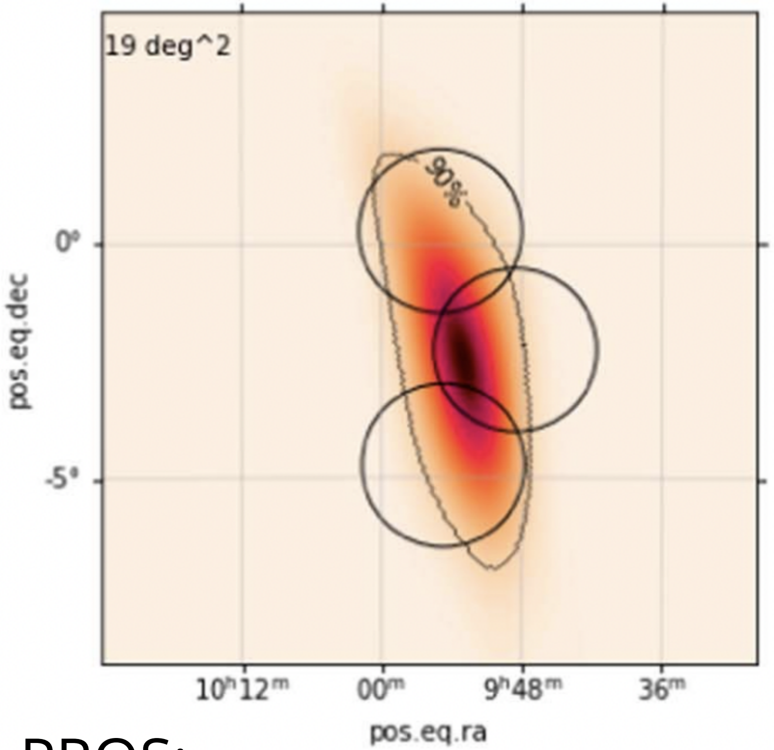
delayed feedback from the changes in the environment

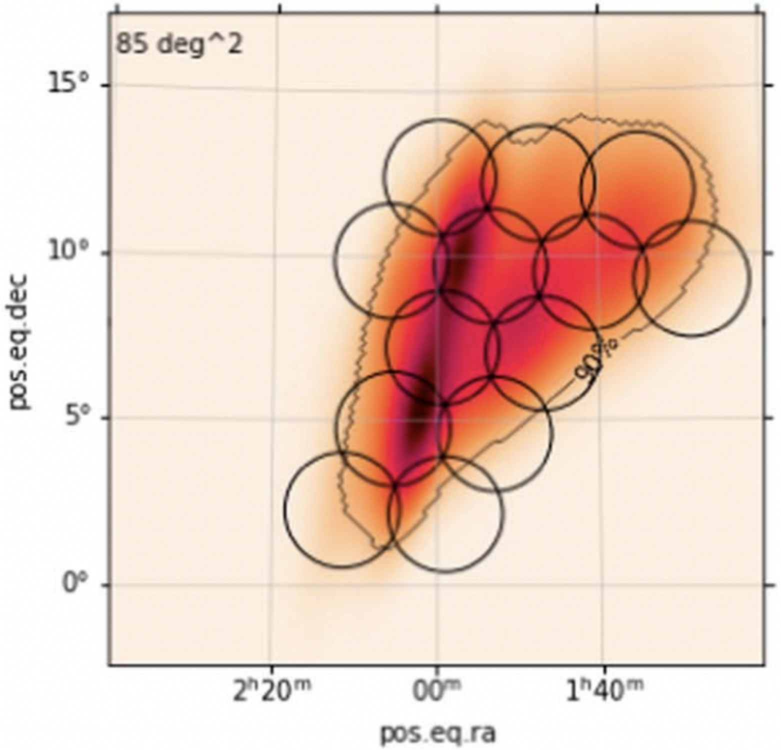


E.g. Selection of follow-up targets in Multi Messenger Astronomy

Rubin ToO program

Andreoni+ 2022b





OPEN ACCESS

Target-of-opportunity Observations of Gravitational-wave Events with Vera C. Rubin Observatory

Igor Andreoni[50,1,2,3], Raffaella Margutti[4], Om Sharan Salafia[5,6], B. Parazin[7], V. Ashley Villar[8,9,10], Michael W. Coughlin[11], Peter Yoachim[12], Kris Mortensen[13], Daniel Brethauer[4], S. J. Smartt[14] + Show full author list

Published 2022 May 13 • © 2022. The Author(s). Published by the American Astronomical Society.

The Astrophysical Journal Supplement Series, Volume 260, Number 1

Rubin LSST Survey Strategy Optimization

Citation Igor Andreoni *et al* 2022 *ApJS* **260** 18

Target of Opportunity Observations of Gravitational Wave Events with LSST

The TVS Multiwavelength Characterization/GW Counterparts subgroup,

Raffaella Margutti (chair, Northwestern),

PROS:

Large FoV (10 sq deg - easly cover 100 sq deg in full)

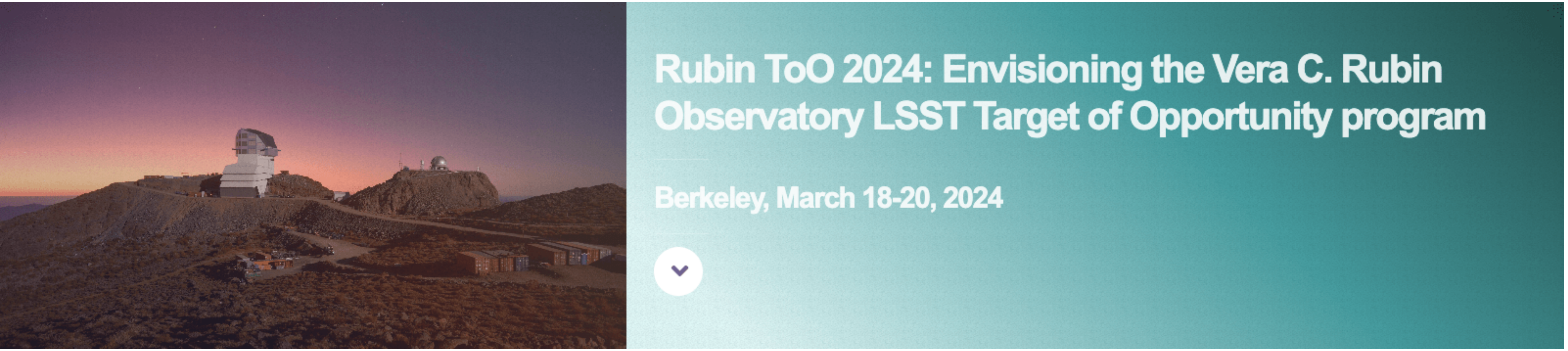
6 filters (5 available on any given night)

deep observations (r~24 in 30 sec, up to 180 sec)

public data

Rubin ToO program

Registration for online participation open through March 15th!
https://lssttooworkshop.github.io/



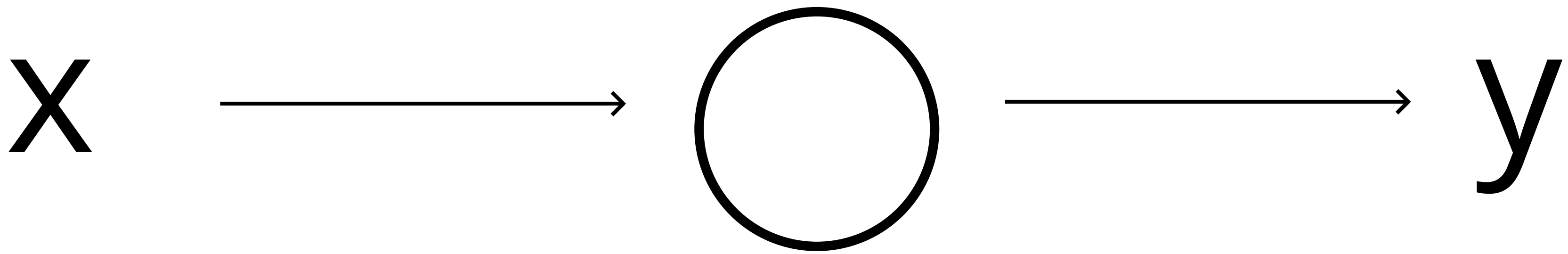
SOC: I. Andreoni (KN), F. Bianco, A. Franckowiak (v), T. Lister (Solar System), R. Margutti (KN, GRB), G. Smith (Lensed KN)

This is a ***work***shop: the goal of the workshop is to produce a report to be delivered to the SCOC containing recommendations for how to implement ToO responses with Rubin. There are no talks.Time is dedicated to collaboratively working toward the workshop report.

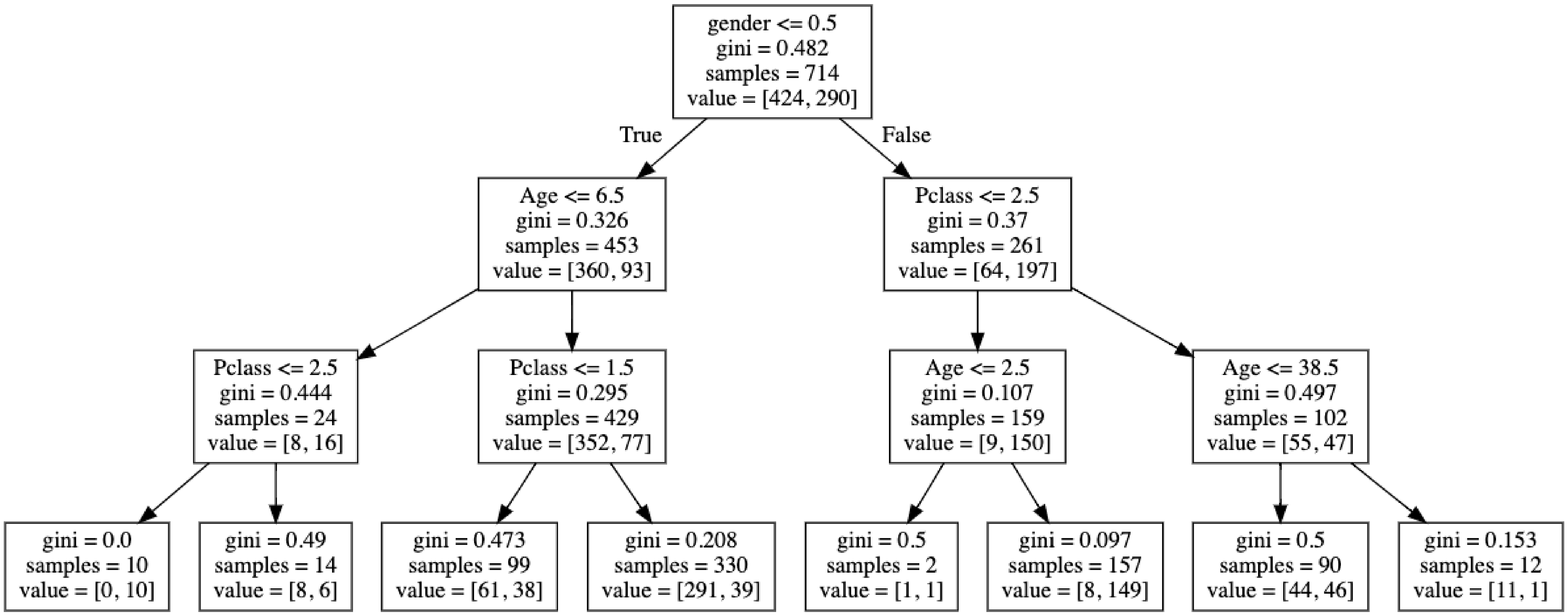
GW science case
Neutrino science case
SSO science case
other science cases

3/6

Deep Learning


x
y

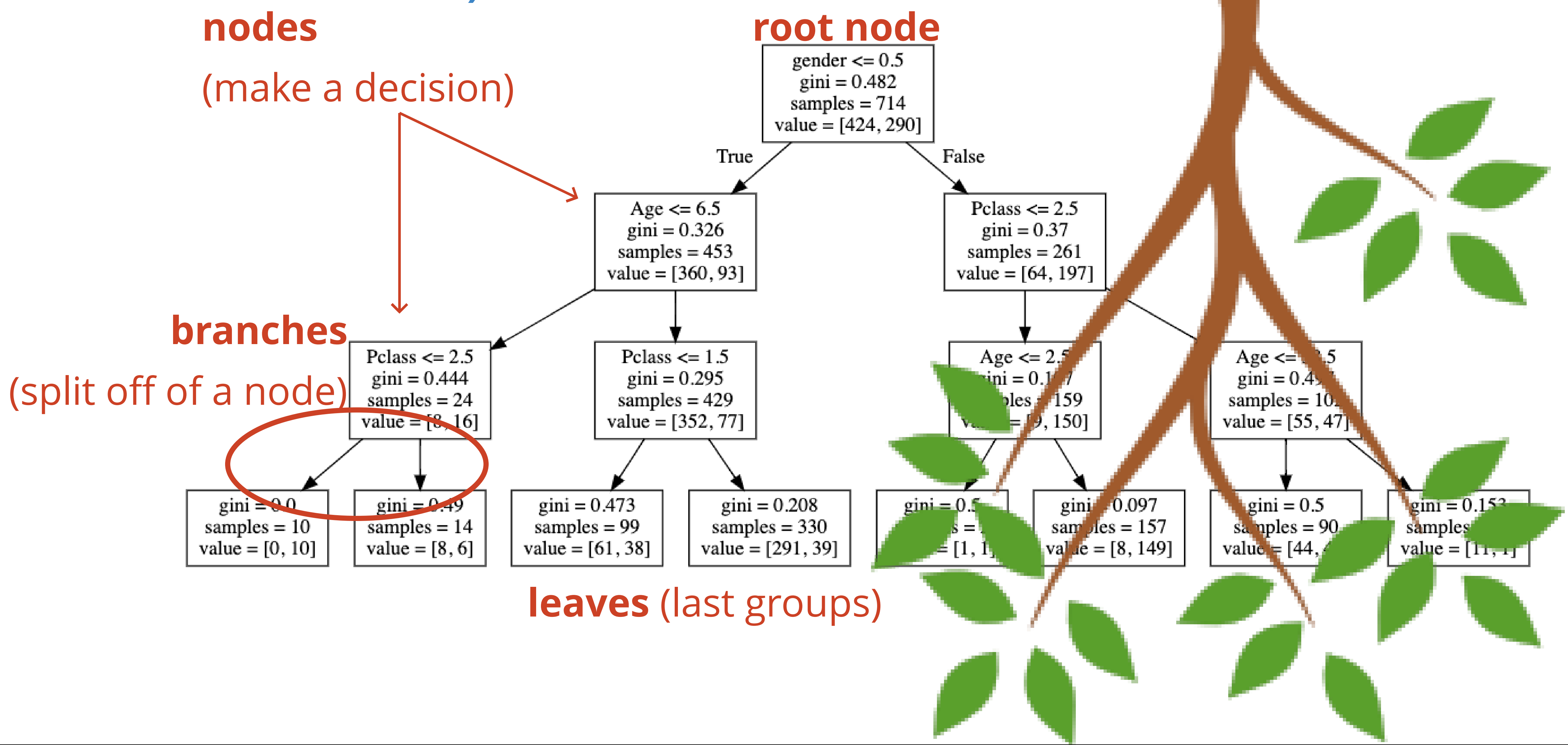
Tree models
(at the basis of Random Forest
Gradient Boosted Trees)

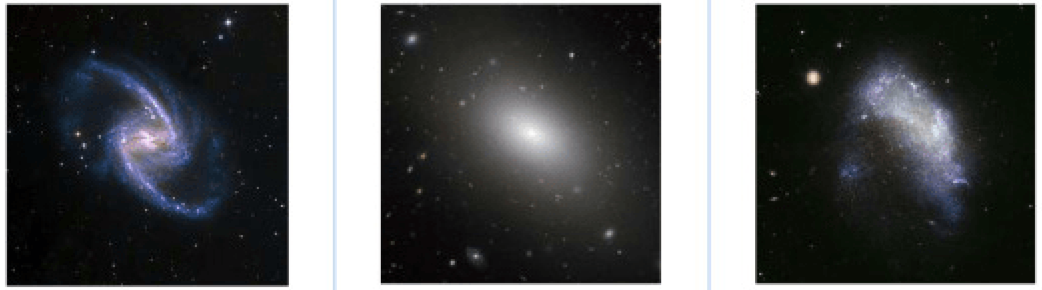


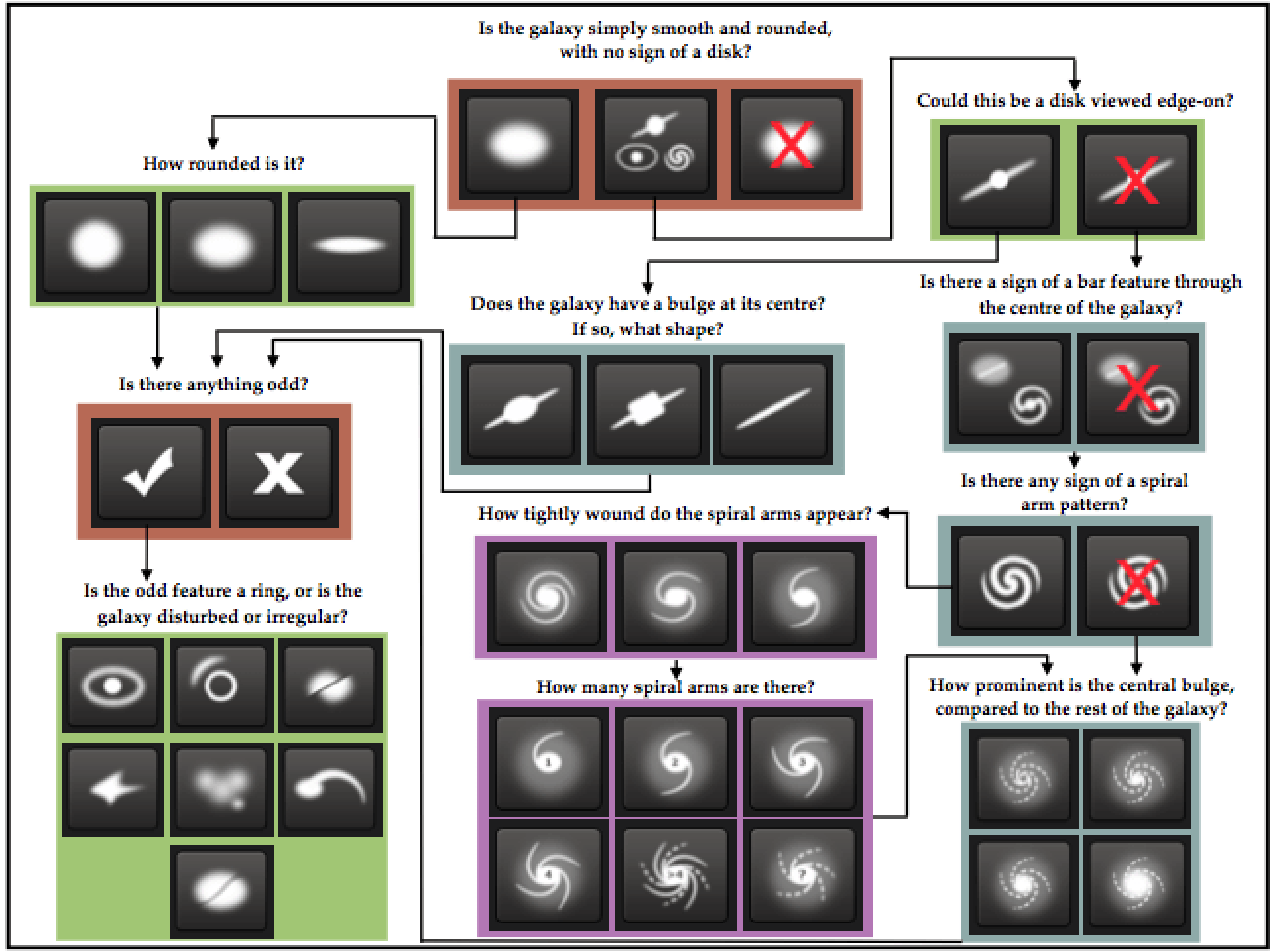
Tree models

(at the basis of Random Forest

Gradient Boosted Trees)



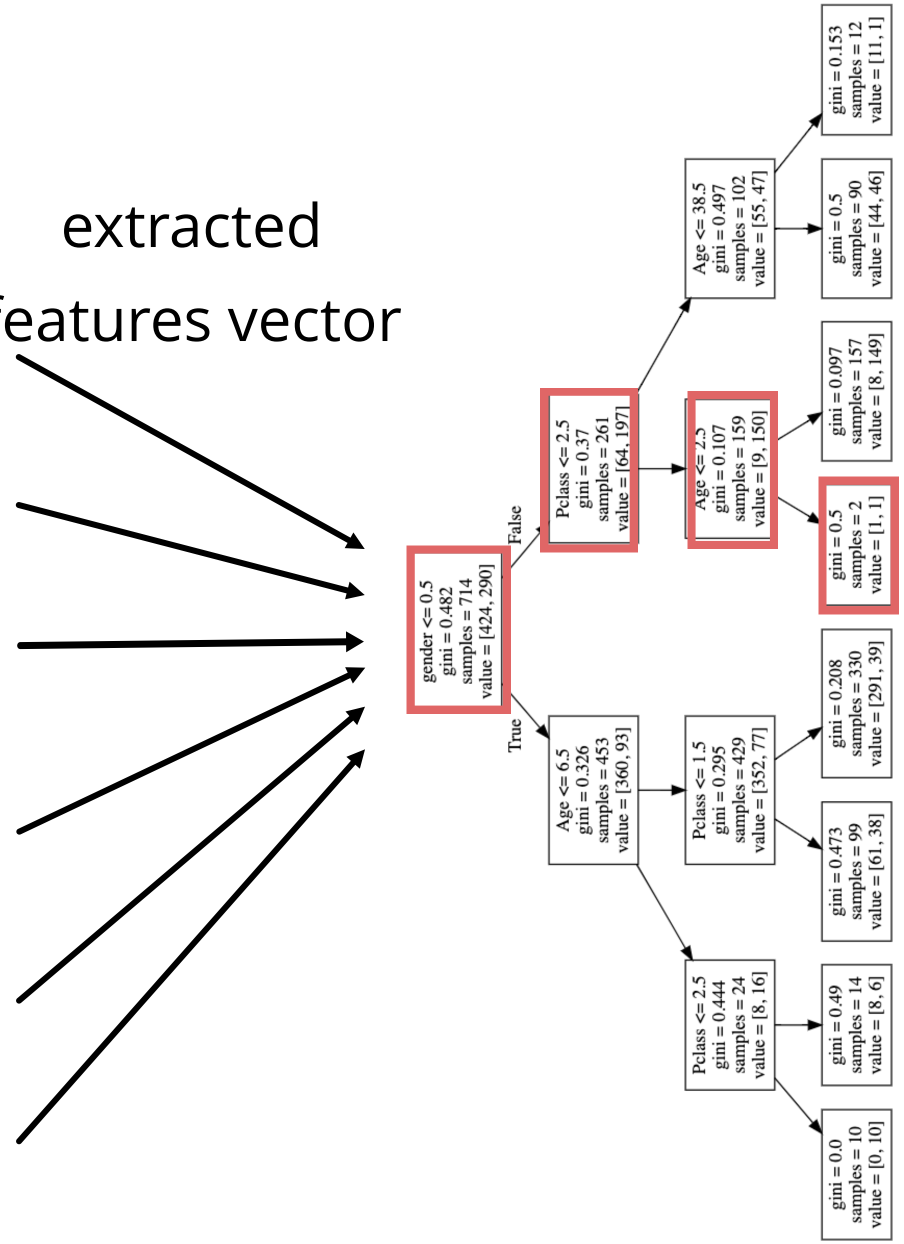




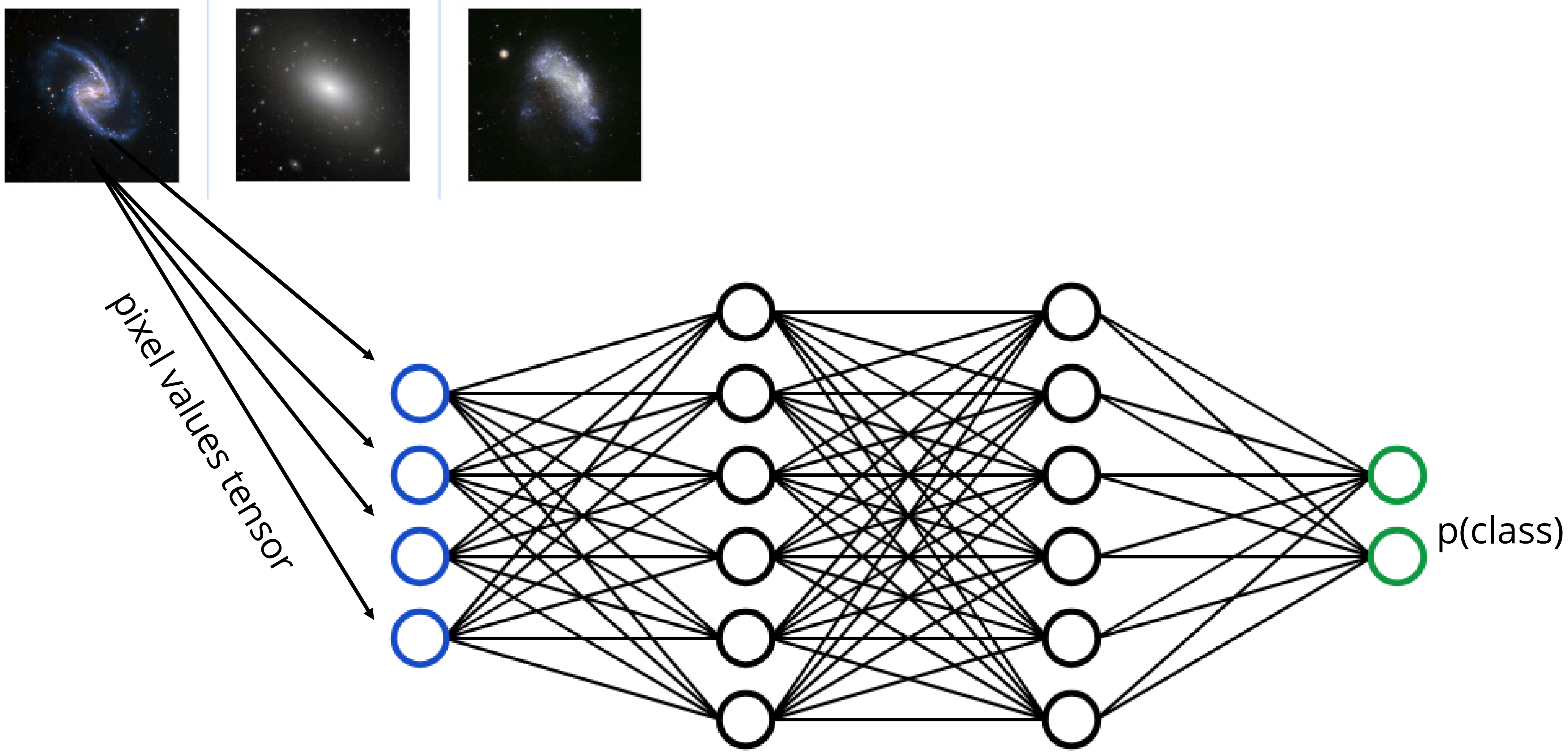
Galaxy Zoo

extracted

features vector



p(class)


pixel values tensor
p(class)

M-P Neuron McCulloch & Pitts 1943

M-P Neuron

1943

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

Warren S. McCulloch and Walter H. Pitts

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

THE THEORY: NETS WITHOUT CIRCLES

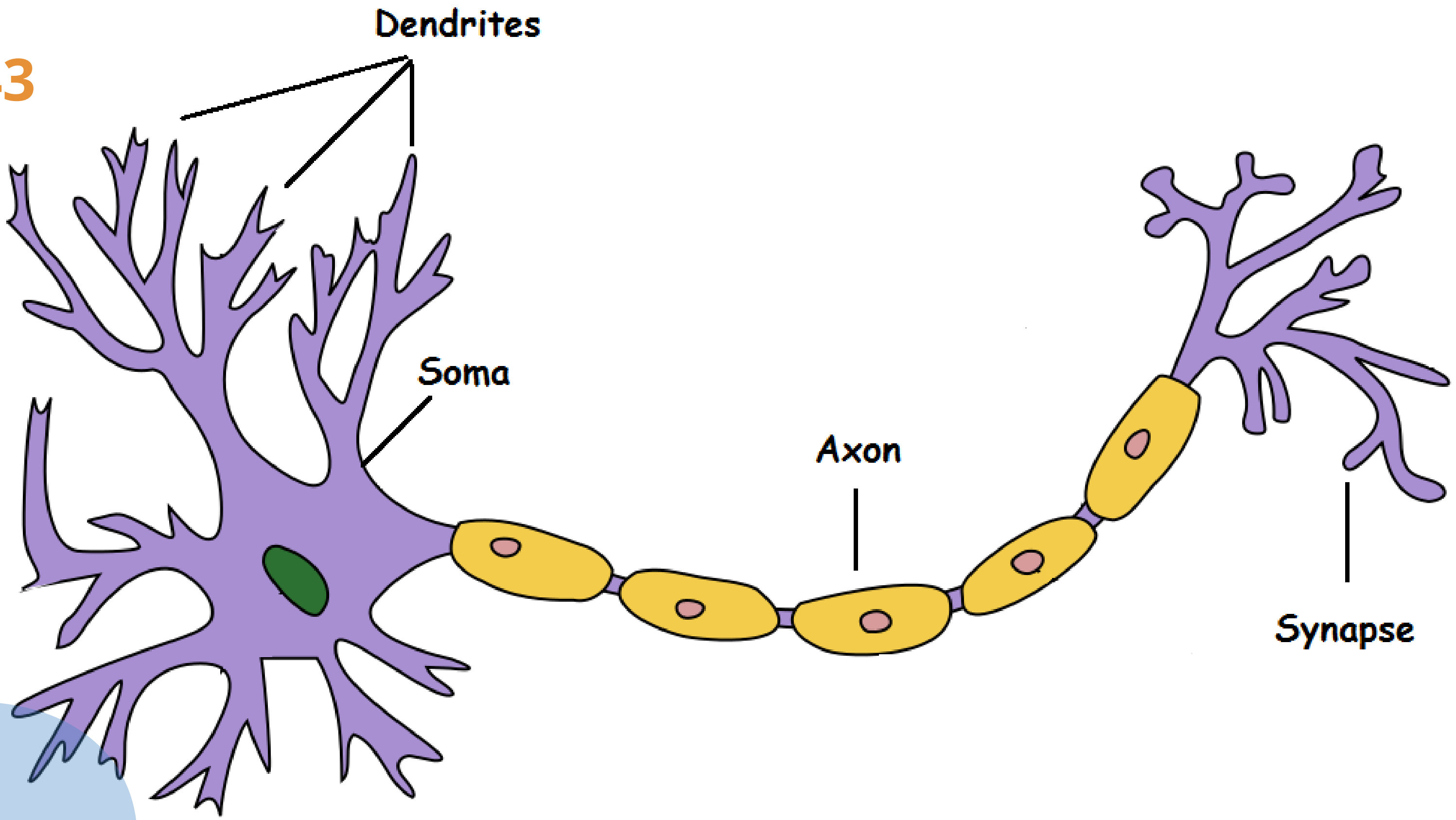
We shall make the following physical assumptions for our calculus.

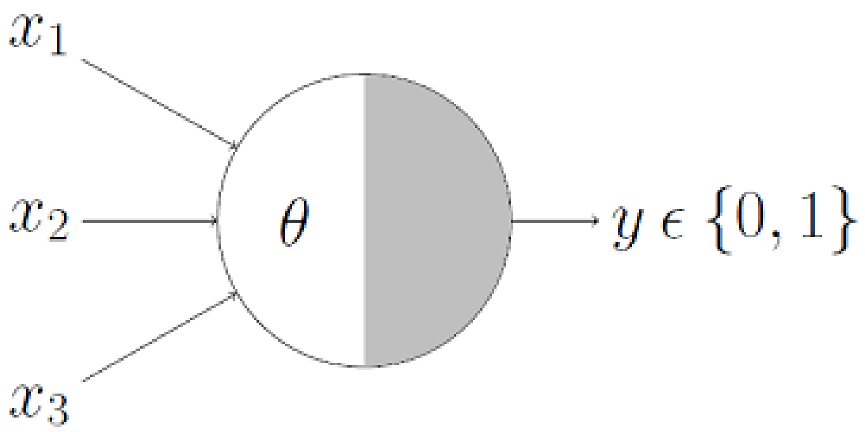
1. The activity of the neuron is an "all-or-none" process.

5. The structure of the net does not change with time.

M-P Neuron McCulloch & Pitts 1943

M-P Neuron



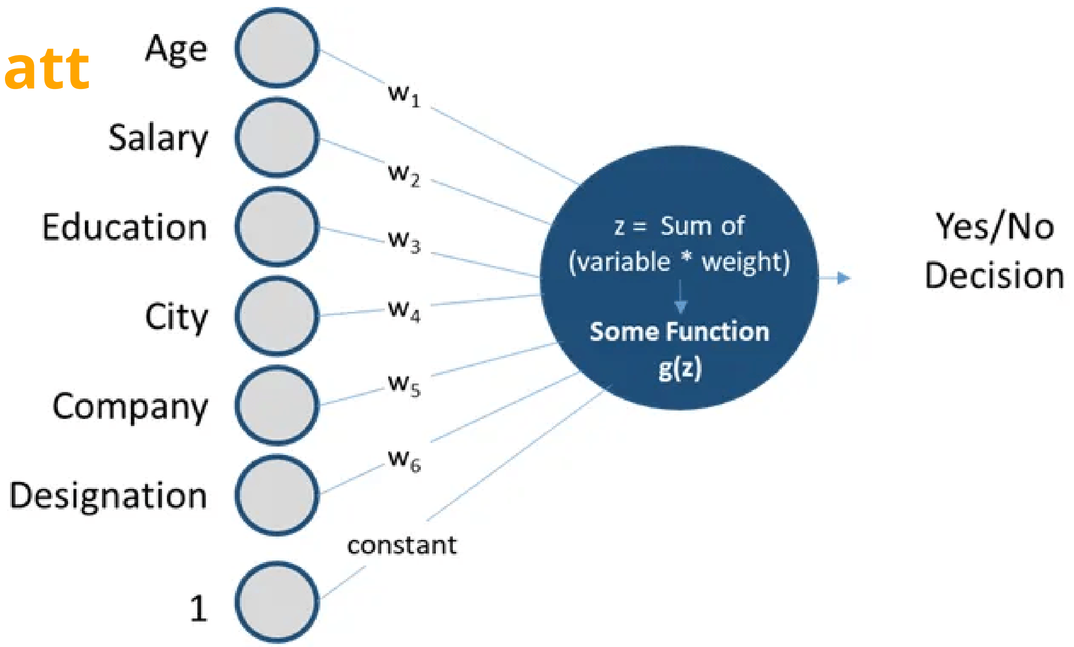


1943

M-P Neuron

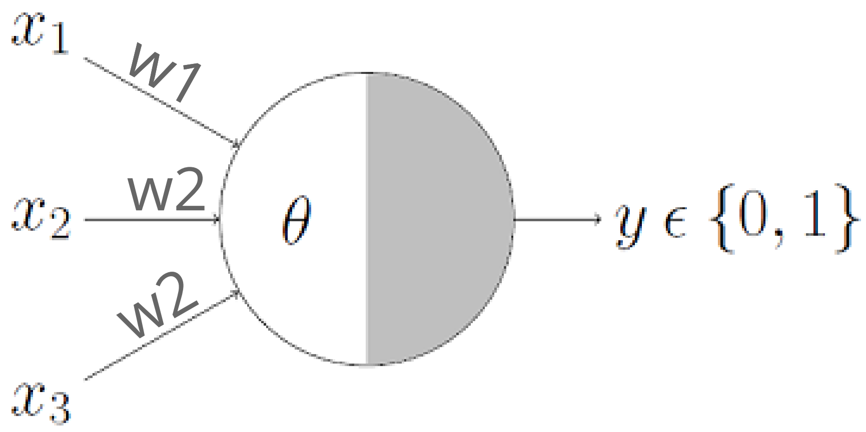
The perceptron algorithm : 1958, Frank Rosenblatt

$$1 \text{ if } \sum_{i=1}^{N} w_i x_i \geq \theta \text{ else } 0$$



1958

Perceptron

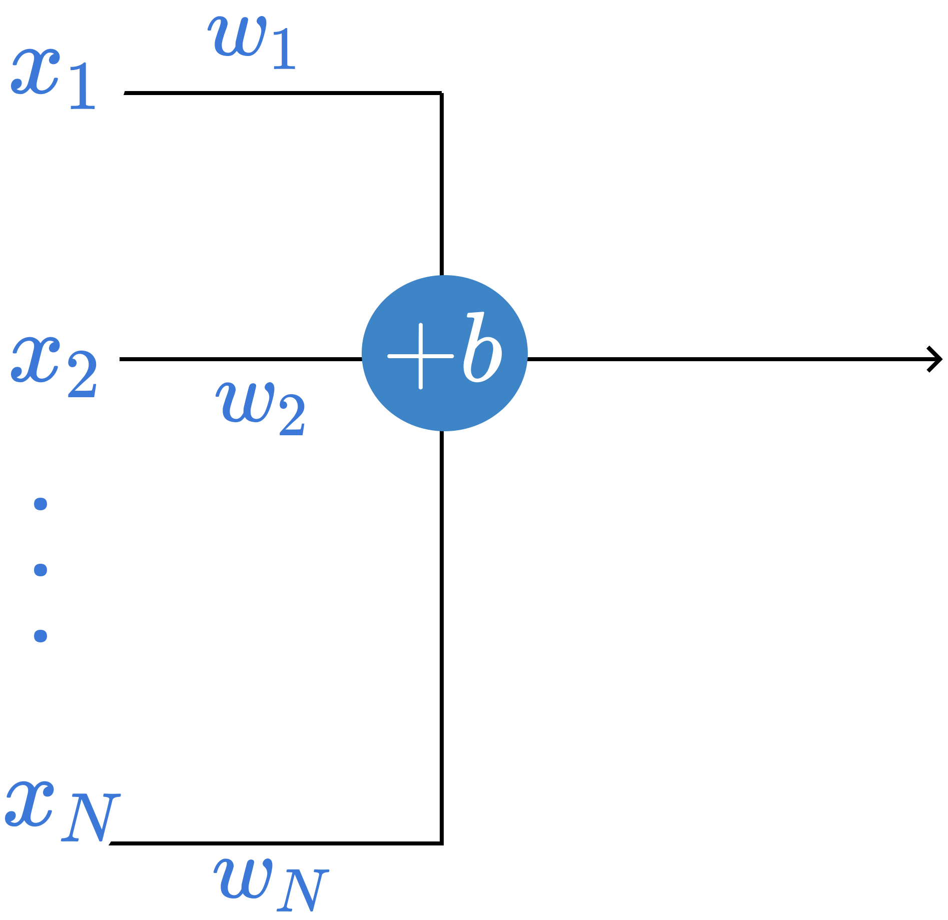


The perceptron algorithm : 1958, Frank Rosenblatt

linear regression: w_i weights

b bias

$$1 \text{ if } \sum_{i=1}^{N} w_i x_i \geq \theta \text{ else } 0$$



1958

Perceptron

The perceptron algorithm : 1958, Frank Rosenblatt

Perceptrons are ***linear classifiers:*** makes its predictions based on a linear predictor function

combining a set of weights (=parameters) with the feature vector.

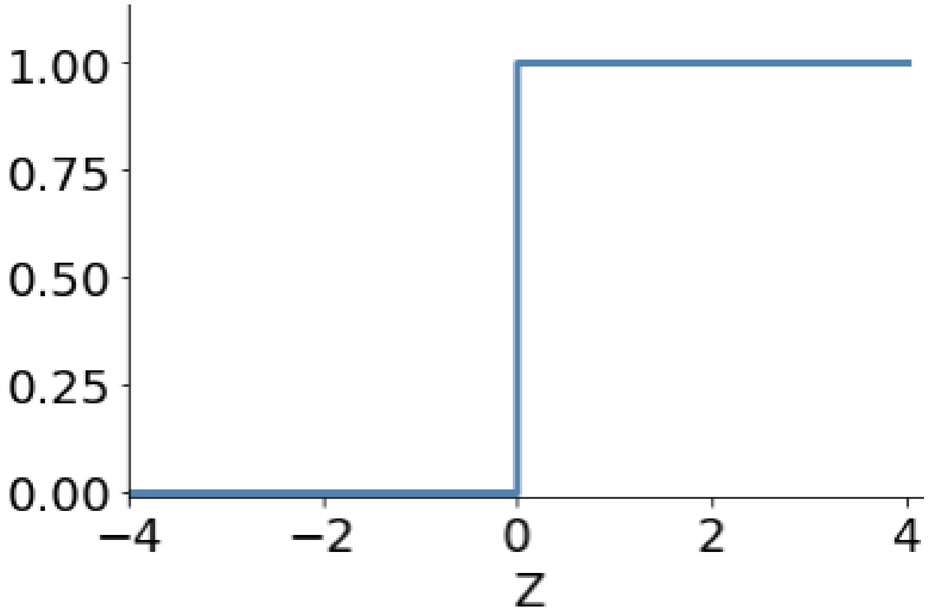
$$y = \left(\sum_i w_i x_i + b \right)$$

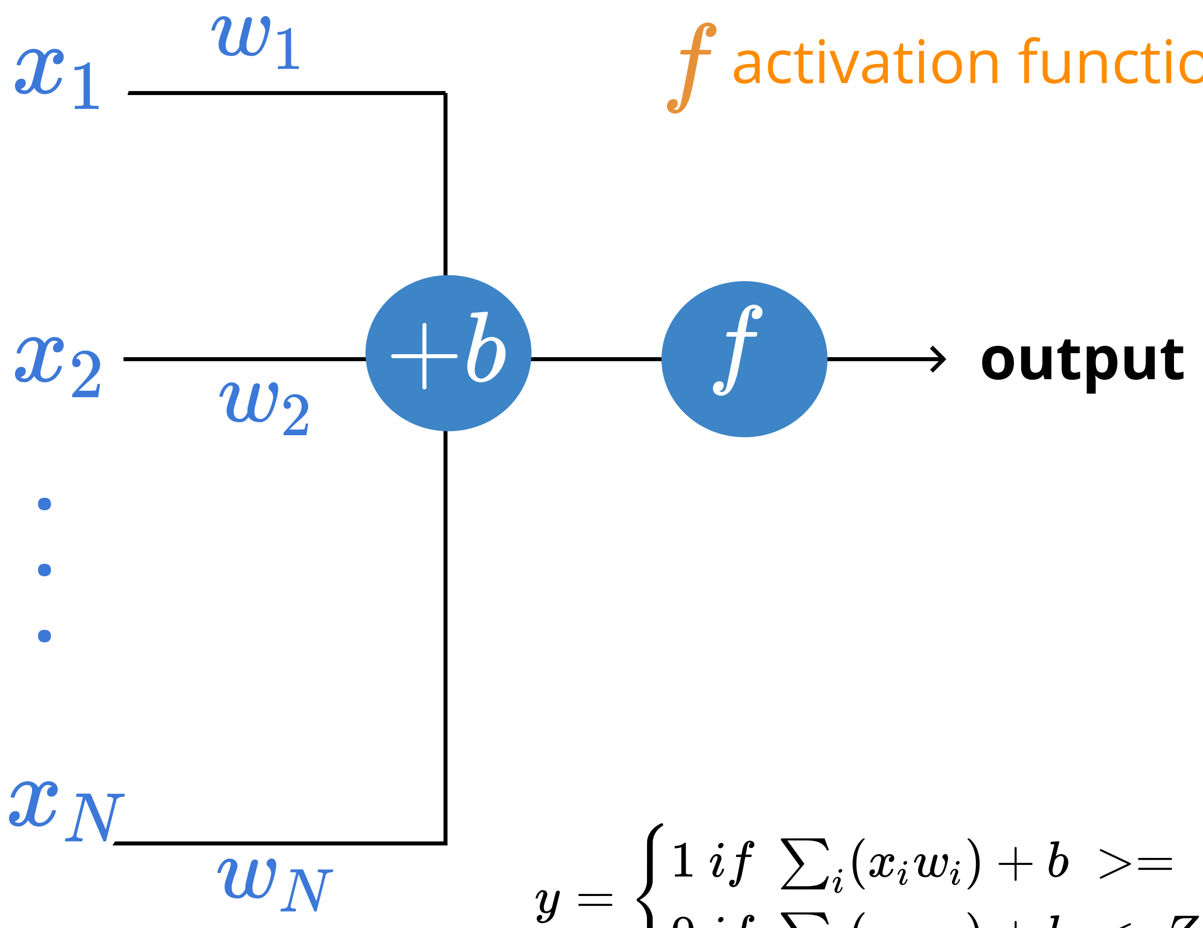
w_i weights

b bias

f activation function

Perceptron





$$y = \begin{cases} 1 \; if \; \sum_i (x_i w_i) + b >= Z \\ 0 \; if \; \sum_i (x_i w_i) + b < Z \end{cases}$$

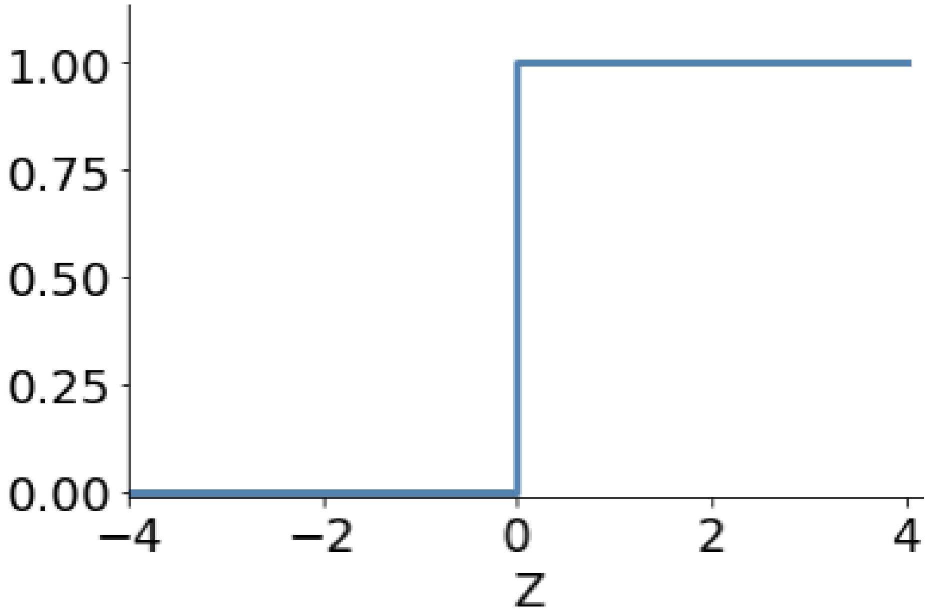
The perceptron algorithm : 1958, Frank Rosenblatt

Perceptrons are ***linear classifiers:*** makes its predictions based on a linear predictor function

combining a set of weights (=parameters) with the feature vector.

$$y = f\left(\sum_i w_i x_i + b \right)$$

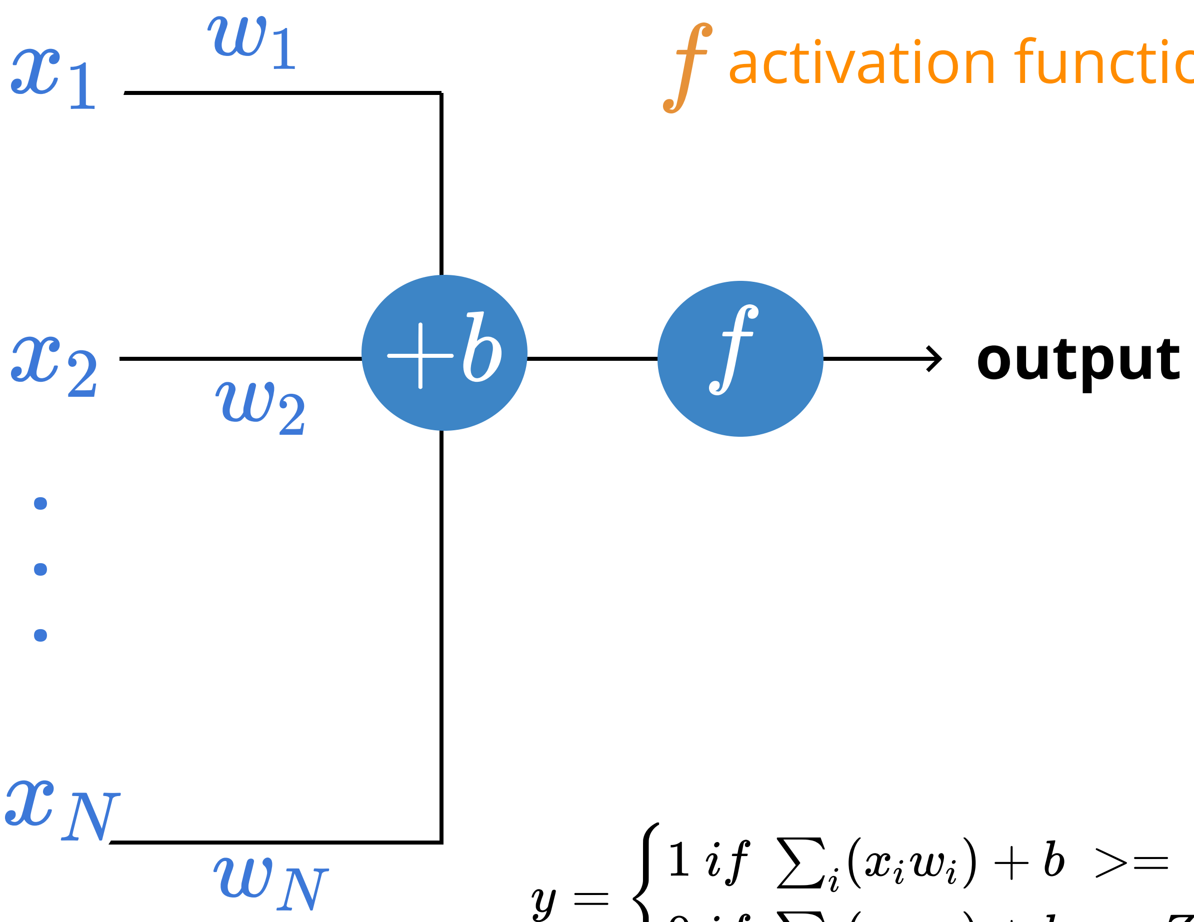
Perceptron



w_i weights

b bias

f activation function



$$y = \begin{cases} 1 \; if \; \sum_i (x_i w_i) + b >= Z \\ 0 \; if \; \sum_i (x_i w_i) + b < Z \end{cases}$$

The perceptron algorithm : 1958, Frank Rosenblatt

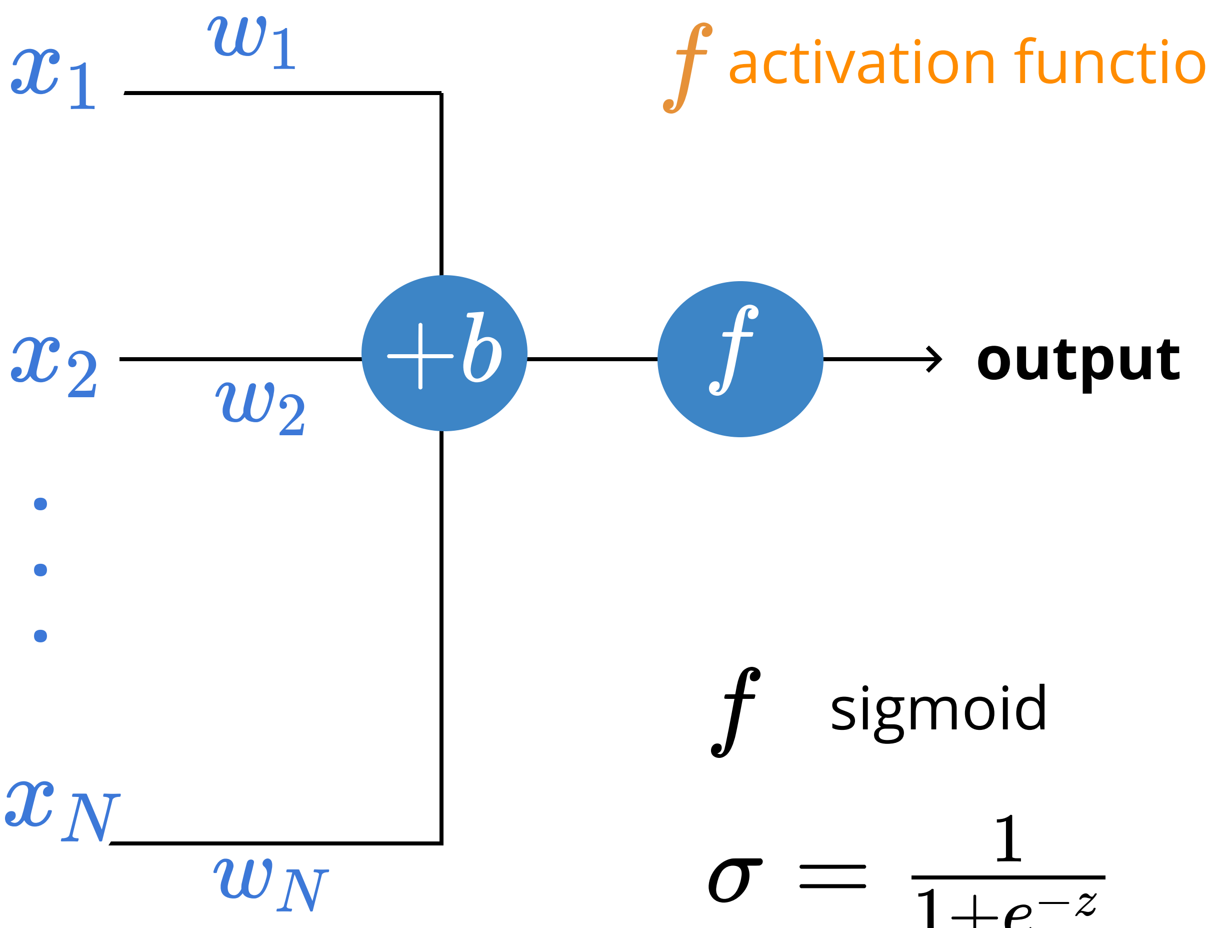
Perceptrons are ***linear classifiers:*** makes its predictions based on a linear predictor function

combining a set of weights (=parameters) with the feature vector.

w_i weights

b bias

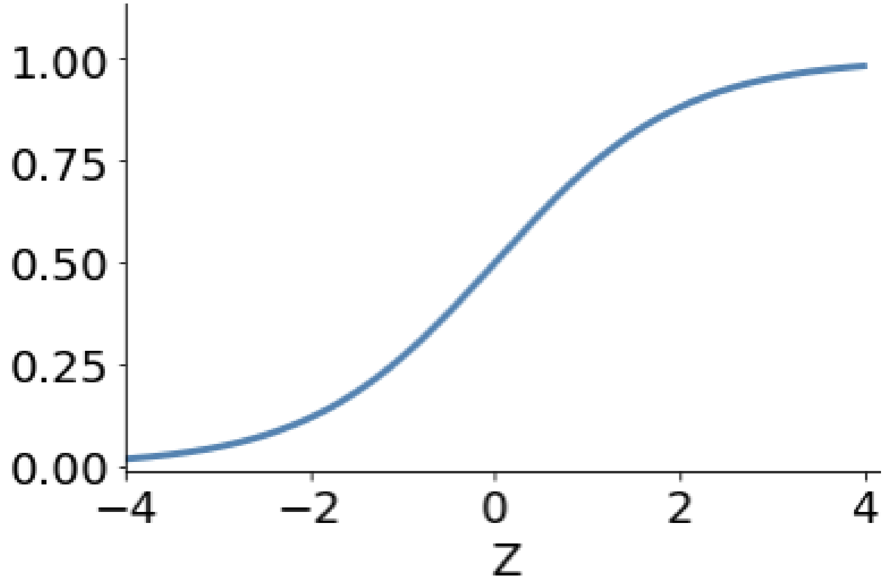
f activation function



f sigmoid

$$\sigma = \frac{1}{1+e^{-z}}$$

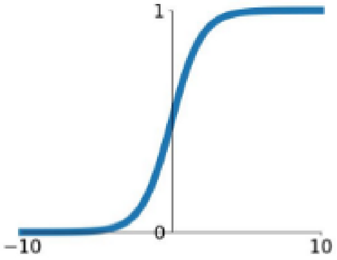
Perceptron



ANN examples of activation function

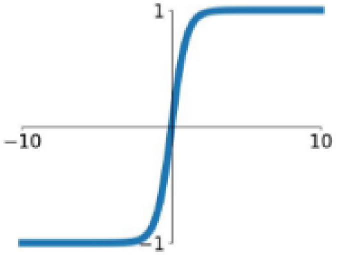
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



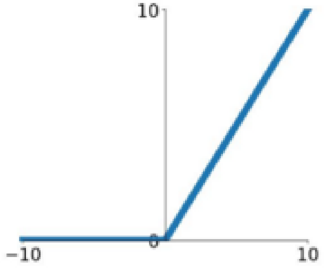
tanh

$\tanh(x)$



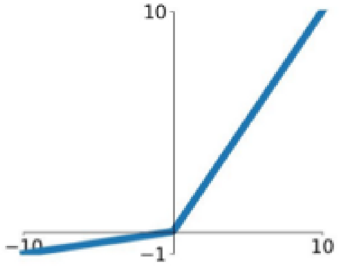
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

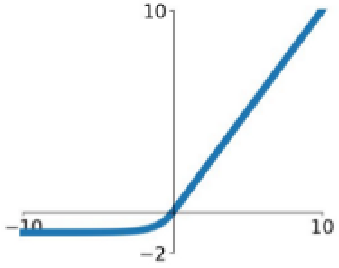


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$



The perceptron algorithm : 1958, Frank Rosenblatt

Perceptron

The New York Times

NEW NAVY DEVICE LEARNS BY DOING; Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

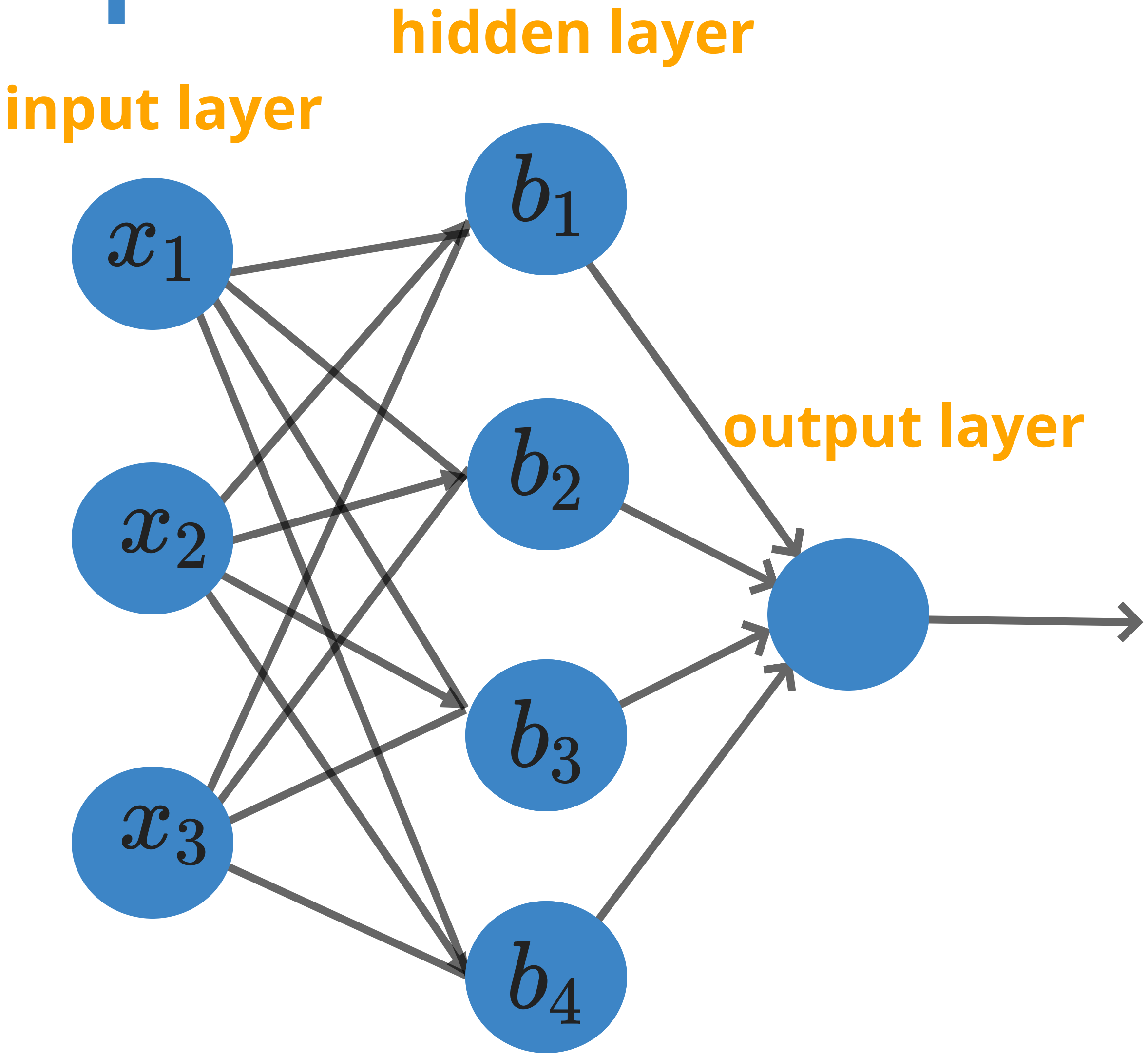
The embryo - the Weather Buerau's $2,000,000 "704" computer - learned to differentiate between left and right after 50 attempts in the Navy demonstration

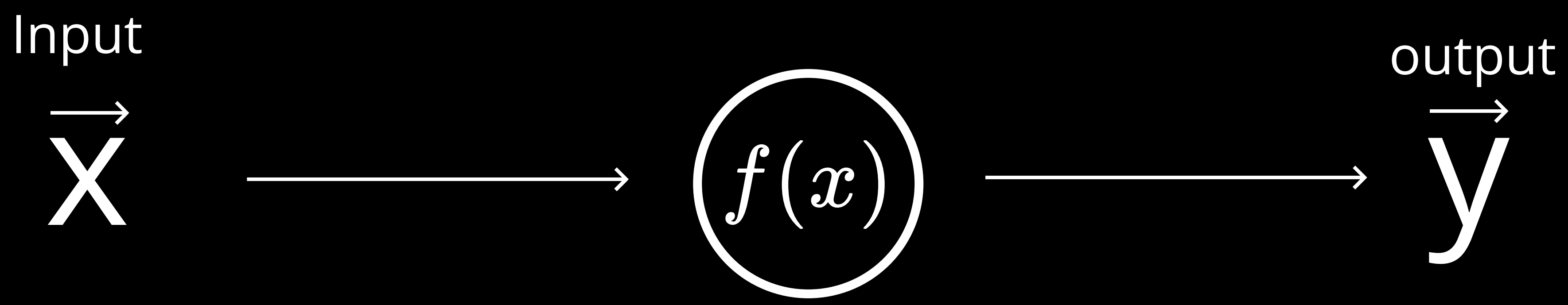
Perceptrons by Marvin Minsky and Seymour Papert 1969

multilayer perceptron

1970: multilayer perceptron architecture

Fully connected: all nodes go to all nodes of the next layer.

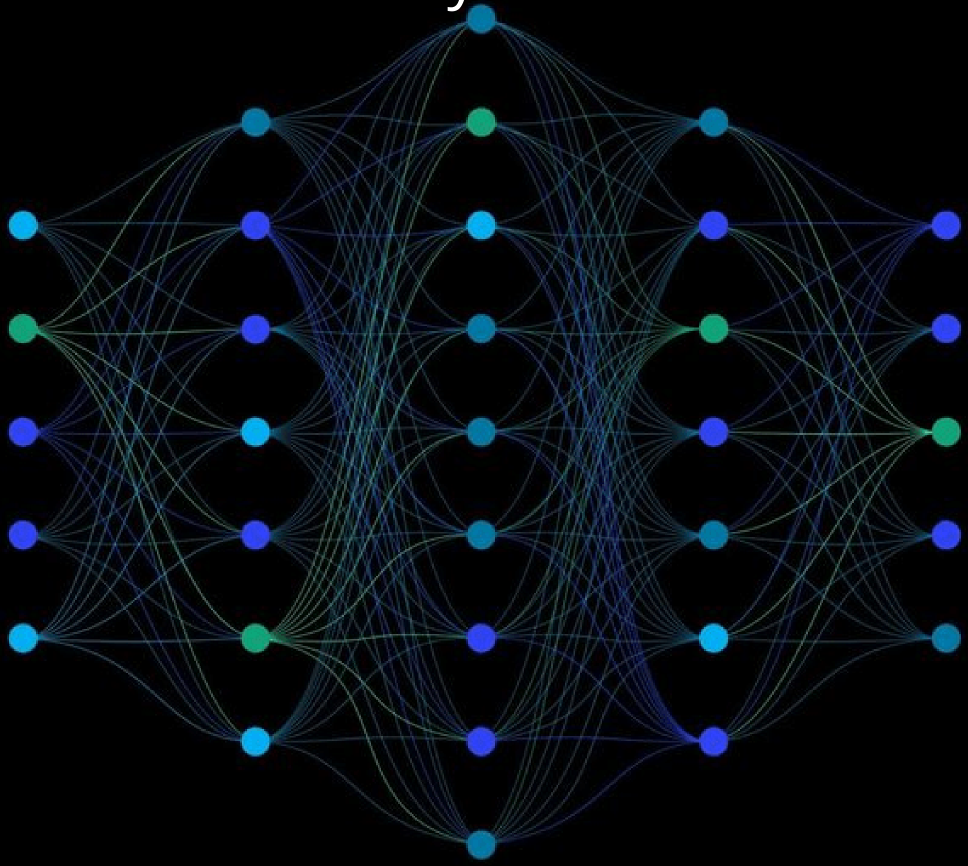



Input
x
f(x)
output
y

hidden
layers

Input

$\vec{x}$

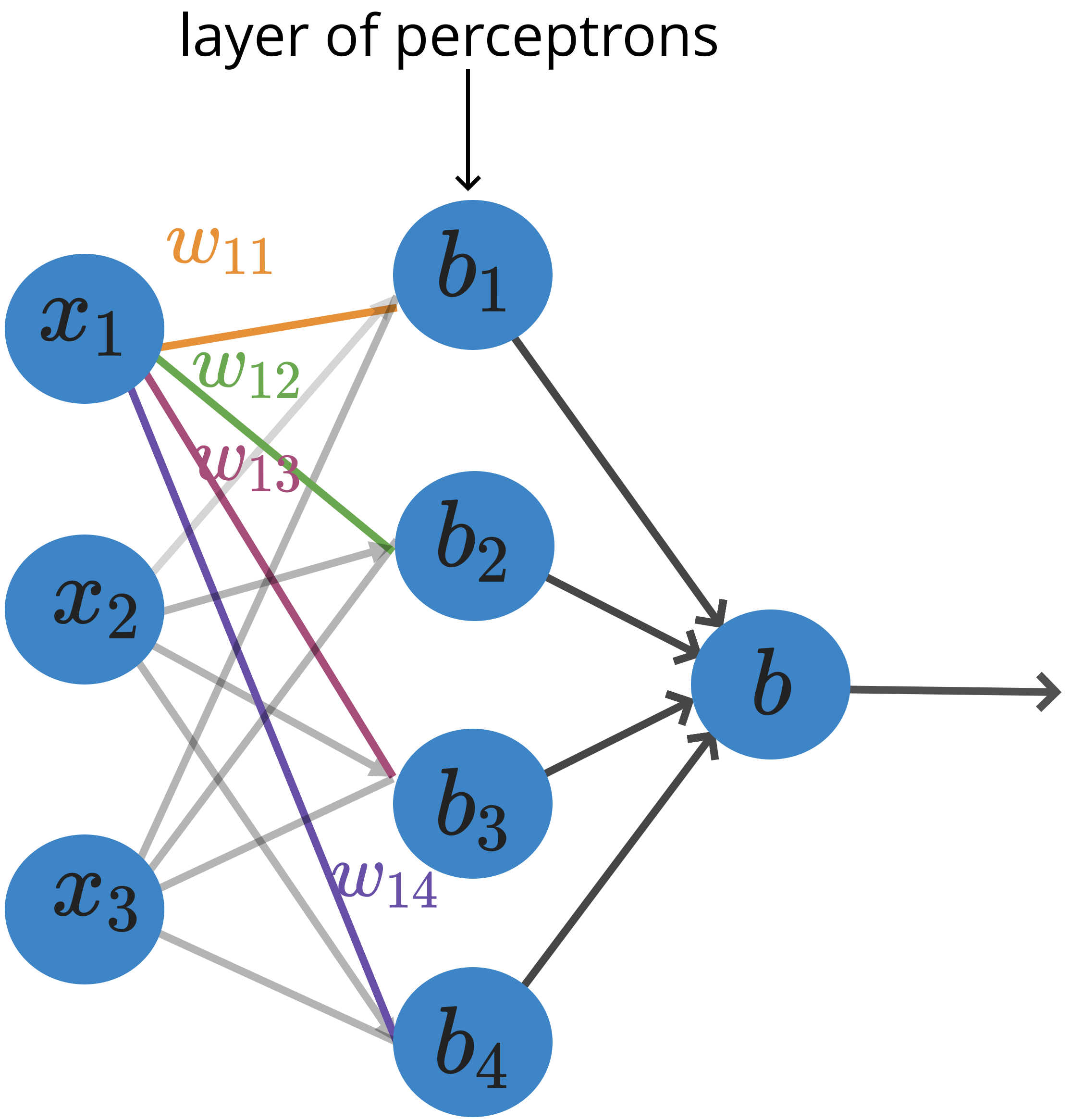


output

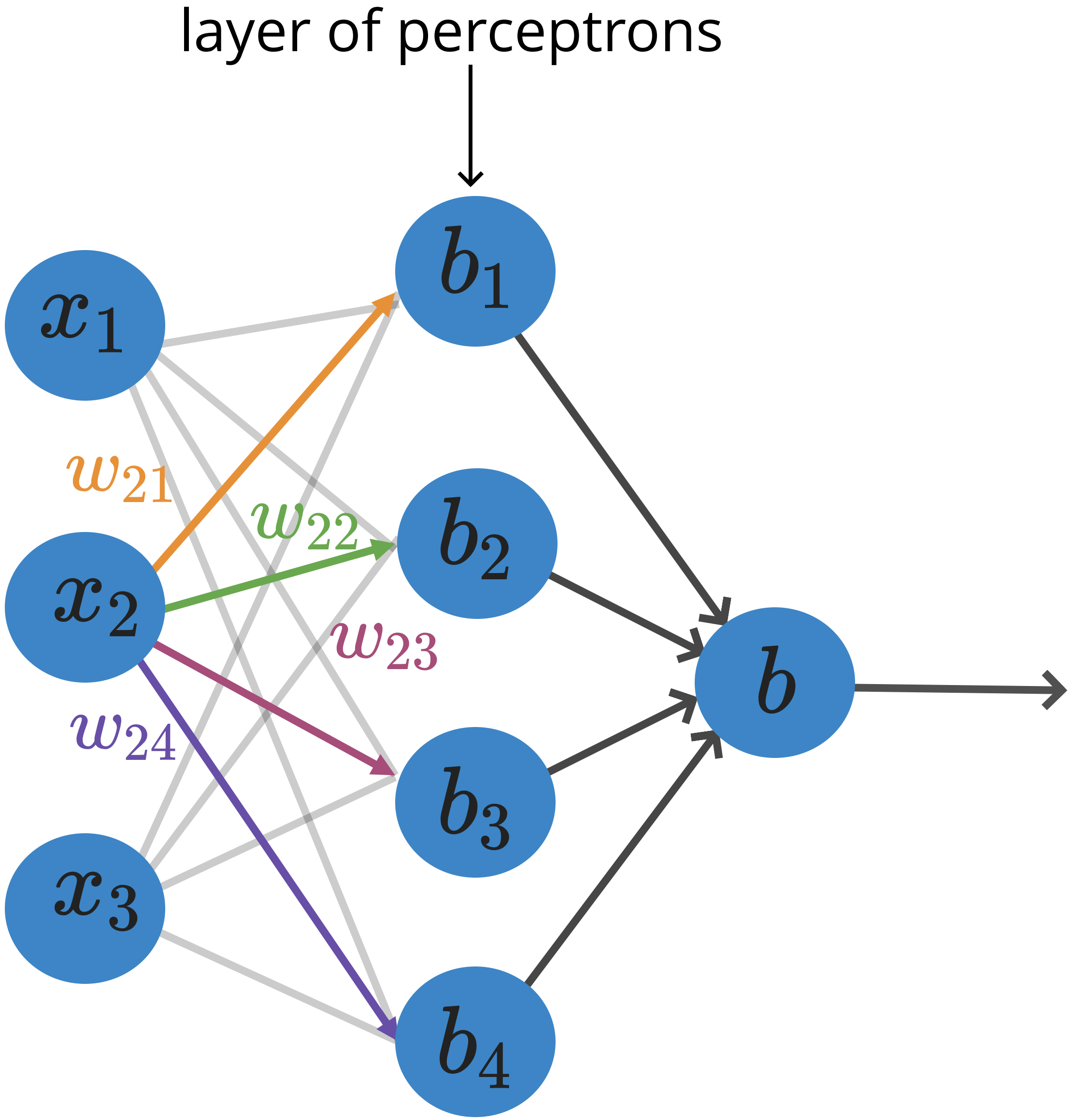
$\vec{y}$

A Neural Network is a kind of function that maps input to output

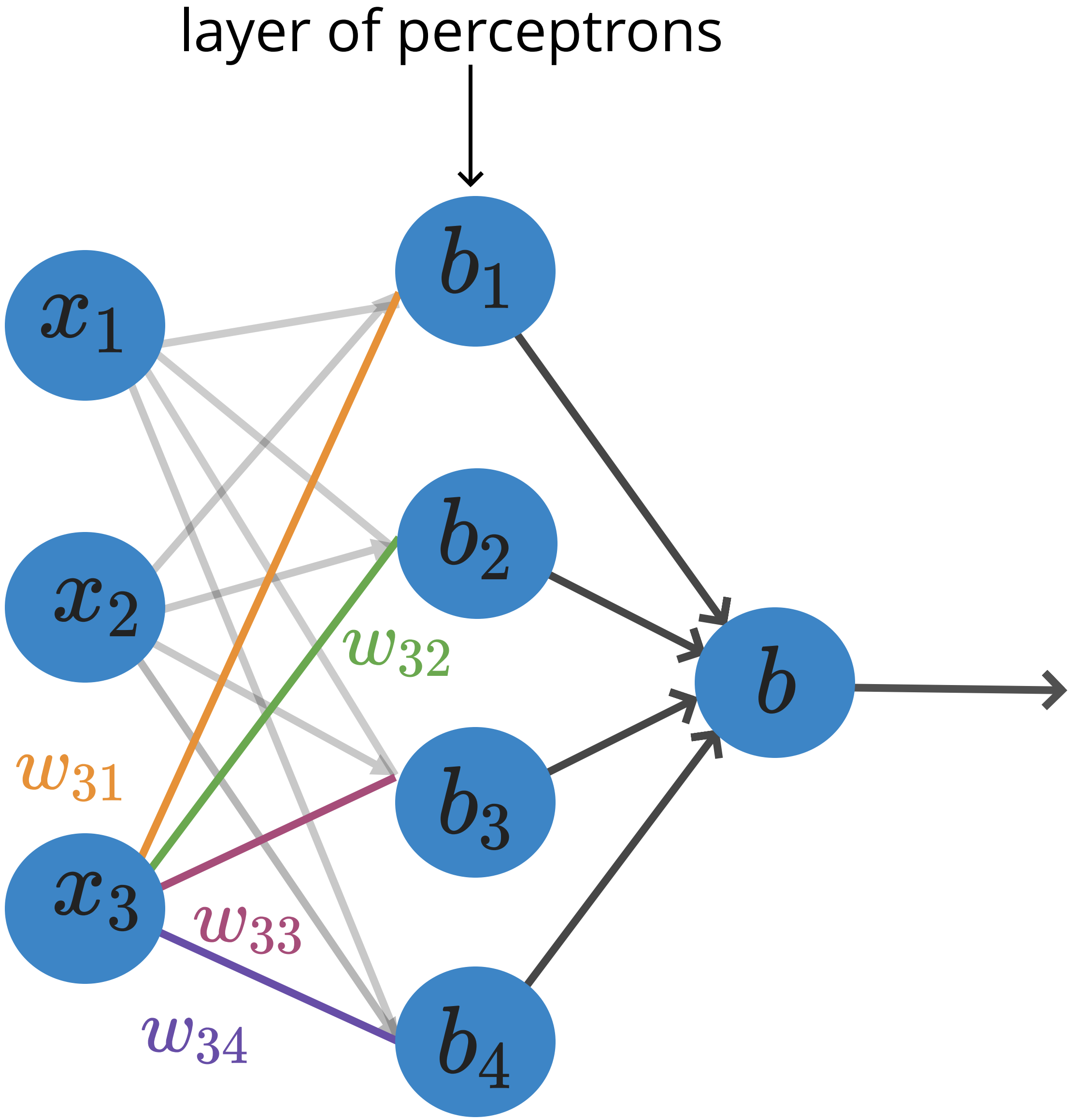
multilayer perceptron



multilayer perceptron



multilayer perceptron



multilayer perceptron

Fully connected: all nodes go to all nodes of the next layer.

f: activation function:

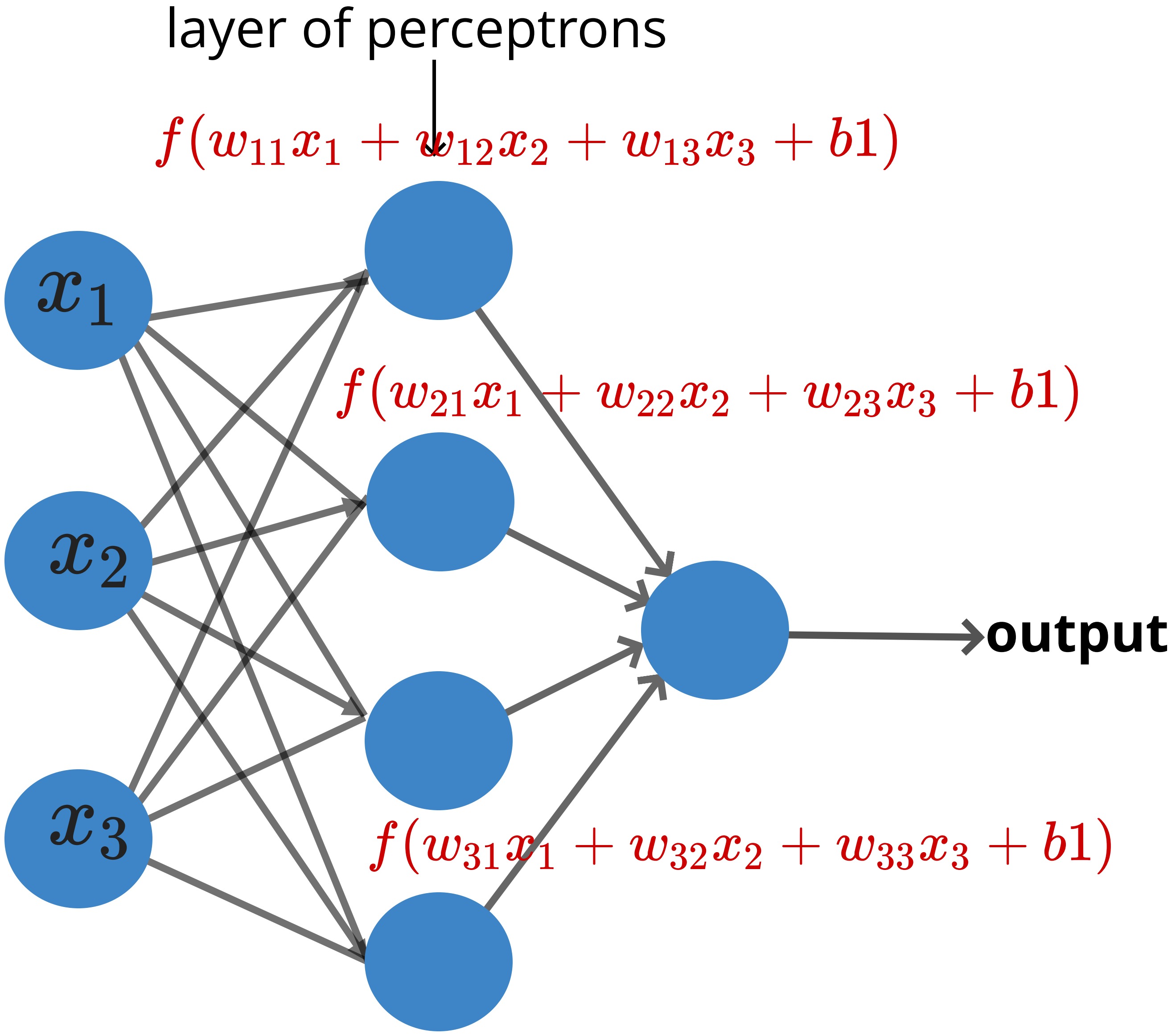
turns neurons on-off

w: weight

sets the sensitivity of a neuron

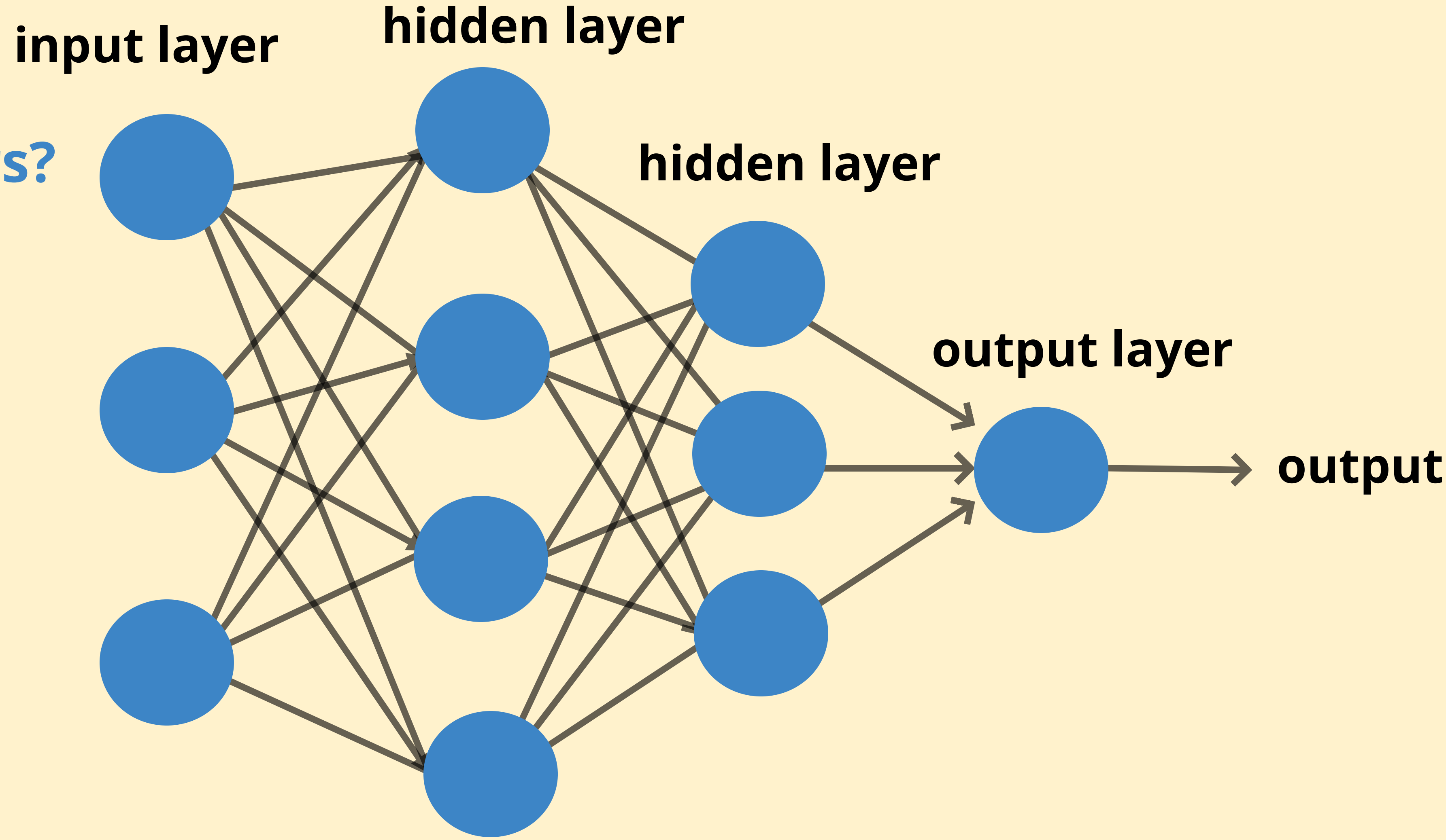
b: bias:

up-down weights a neuron



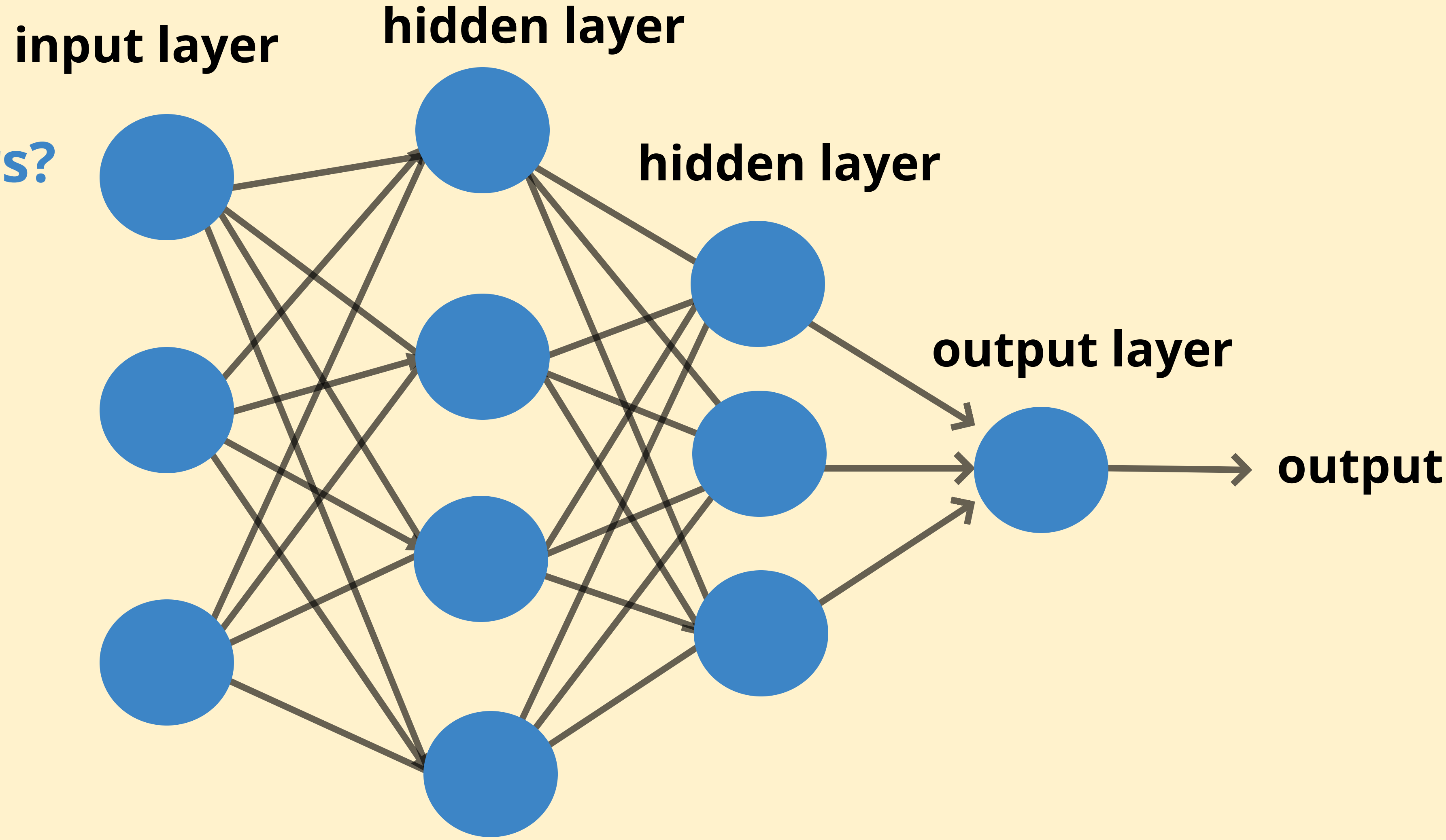
EXERCISE

how many parameters?



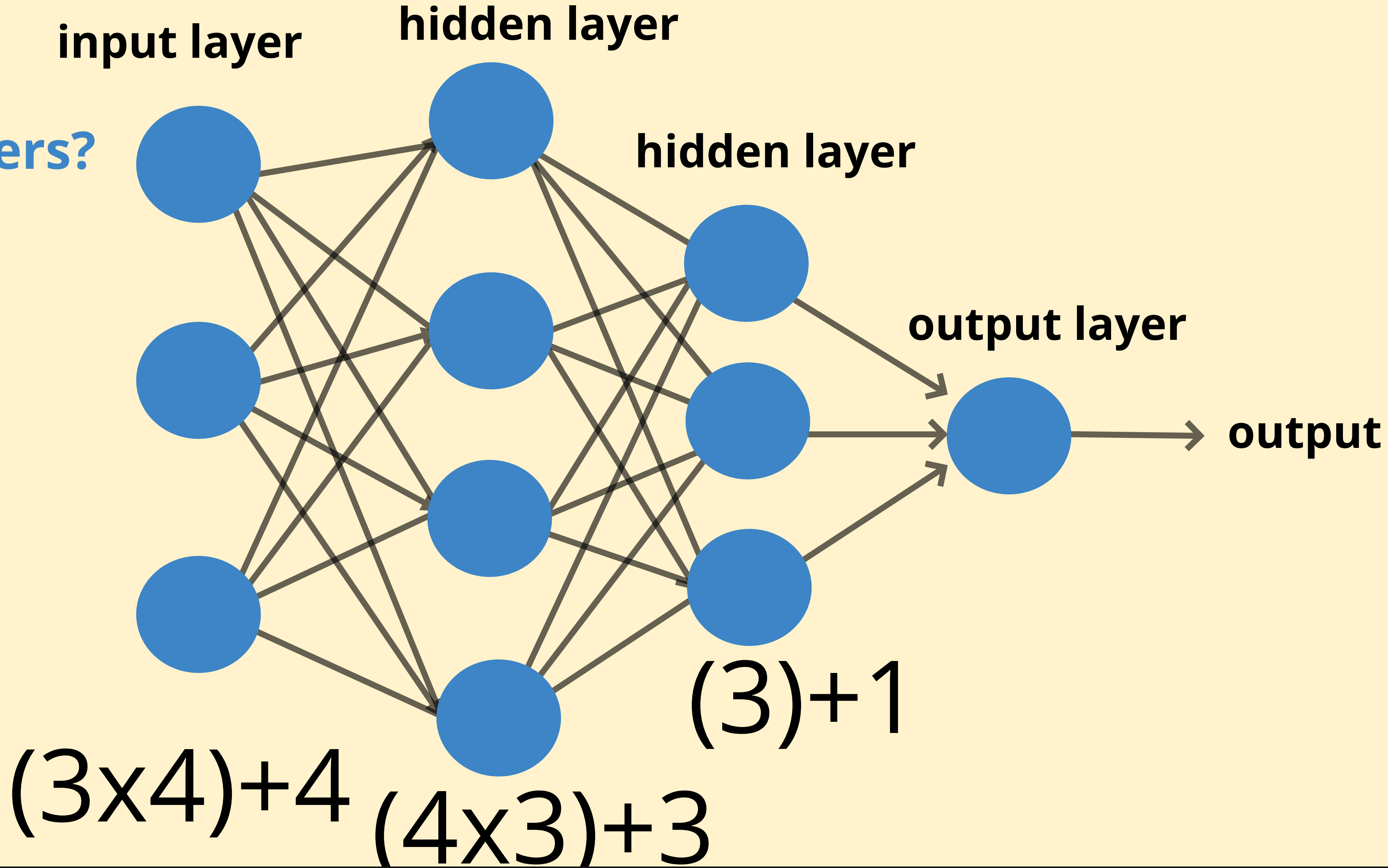
EXERCISE

how many parameters?



EXERCISE

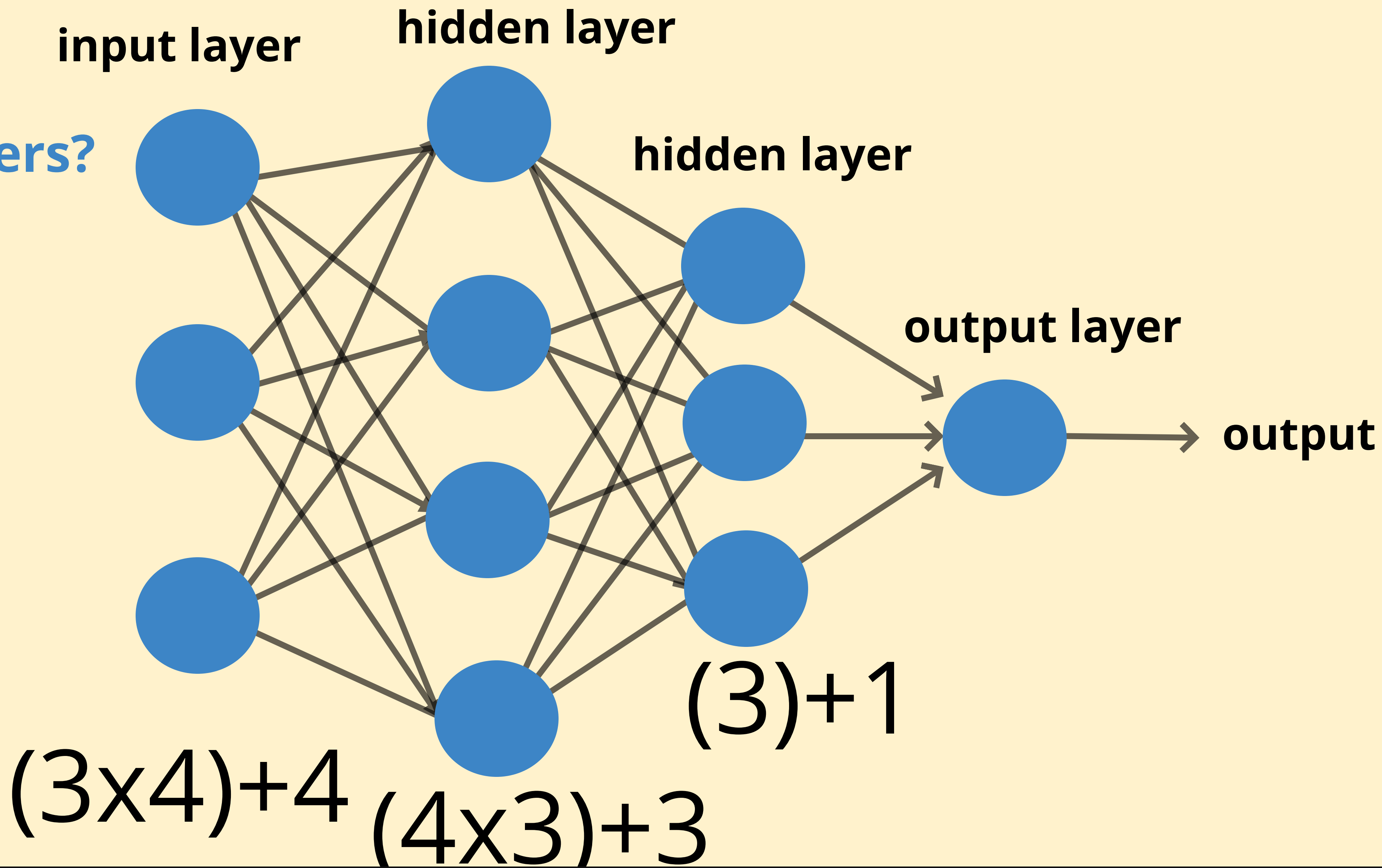
how many parameters?



EXERCISE

how many parameters?

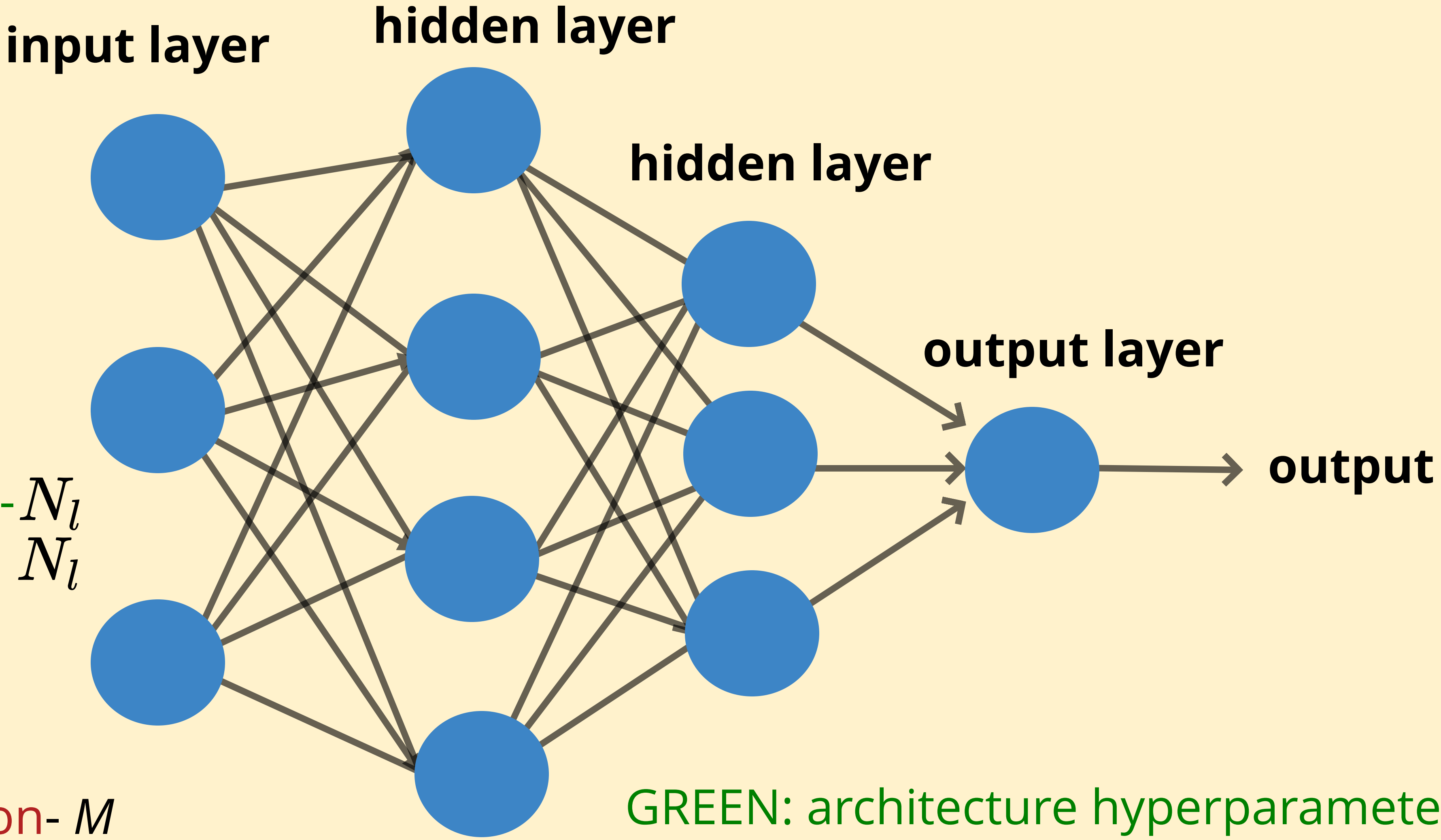
35



EXERCISE

how many hyperparameters?

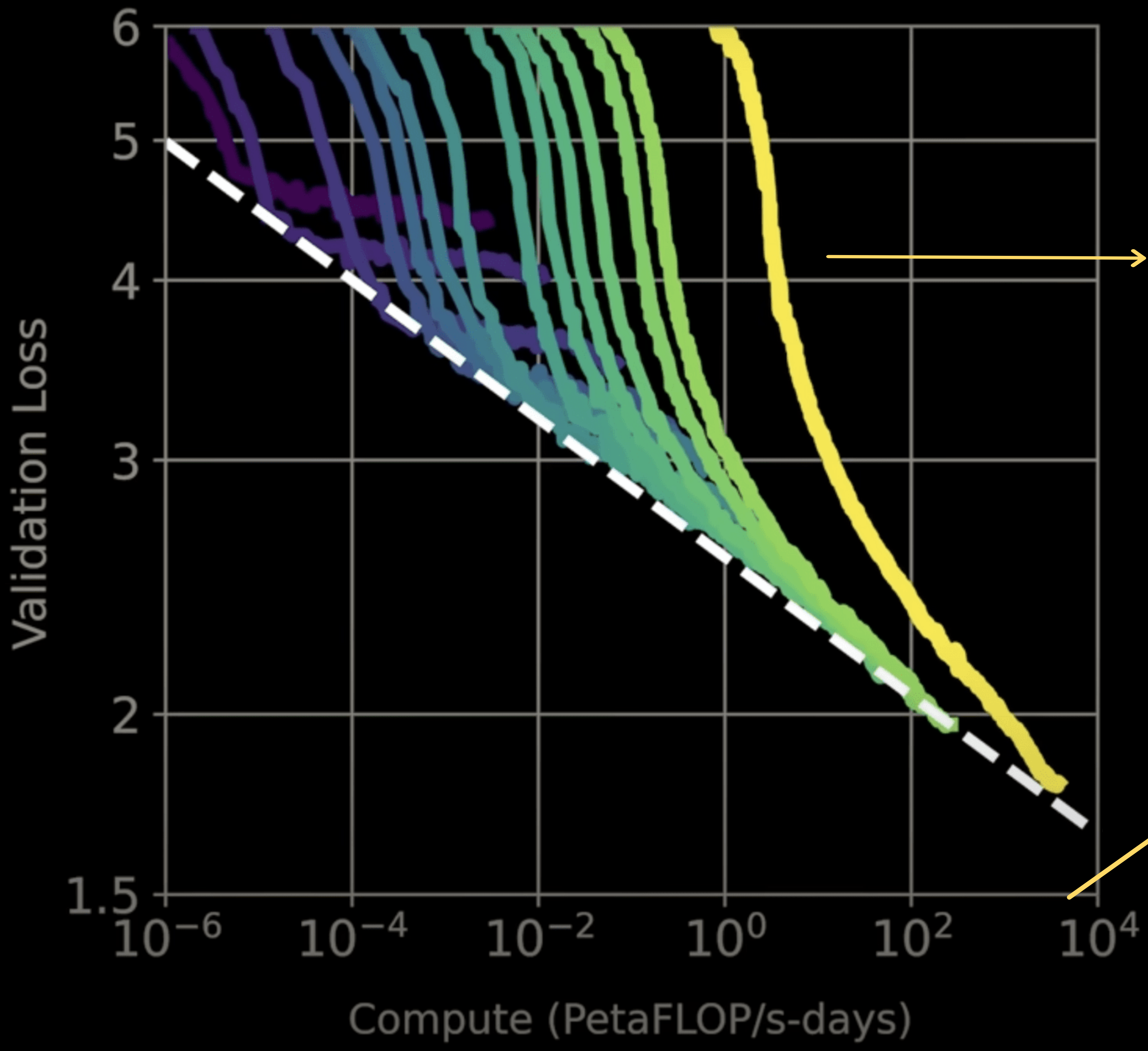
1. number of layers- *1*
2. number of neurons/layer- N_l
3. activation function/layer- N_l
4. layer connectivity- $N_l^{??}$
5. optimization metric - *1*
6. optimization method - *1*
7. parameters in optimization- *M*



GREEN: architecture hyperparamete

RED: training hyperparameters

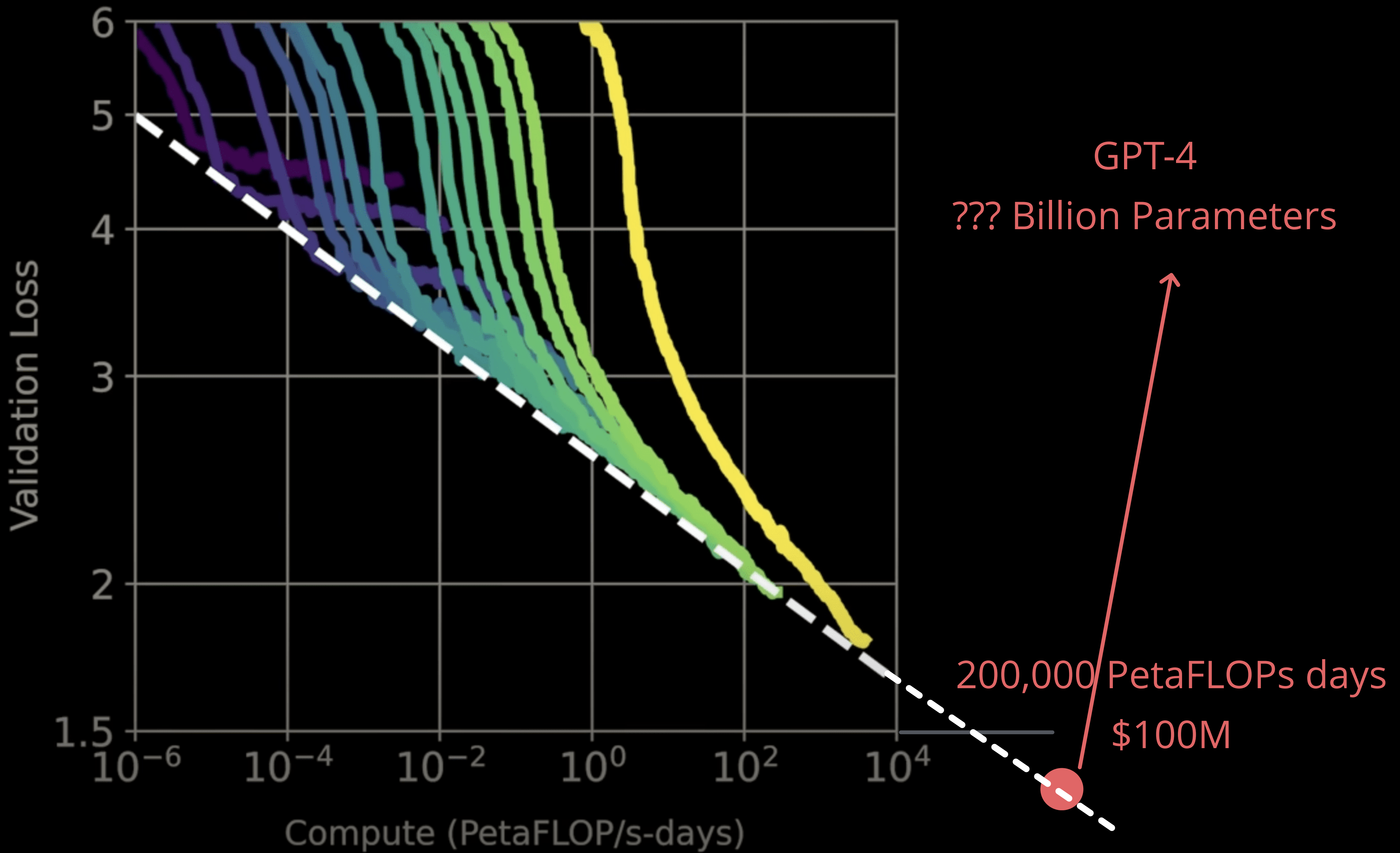
Kaplan+ 2020



GPT-3

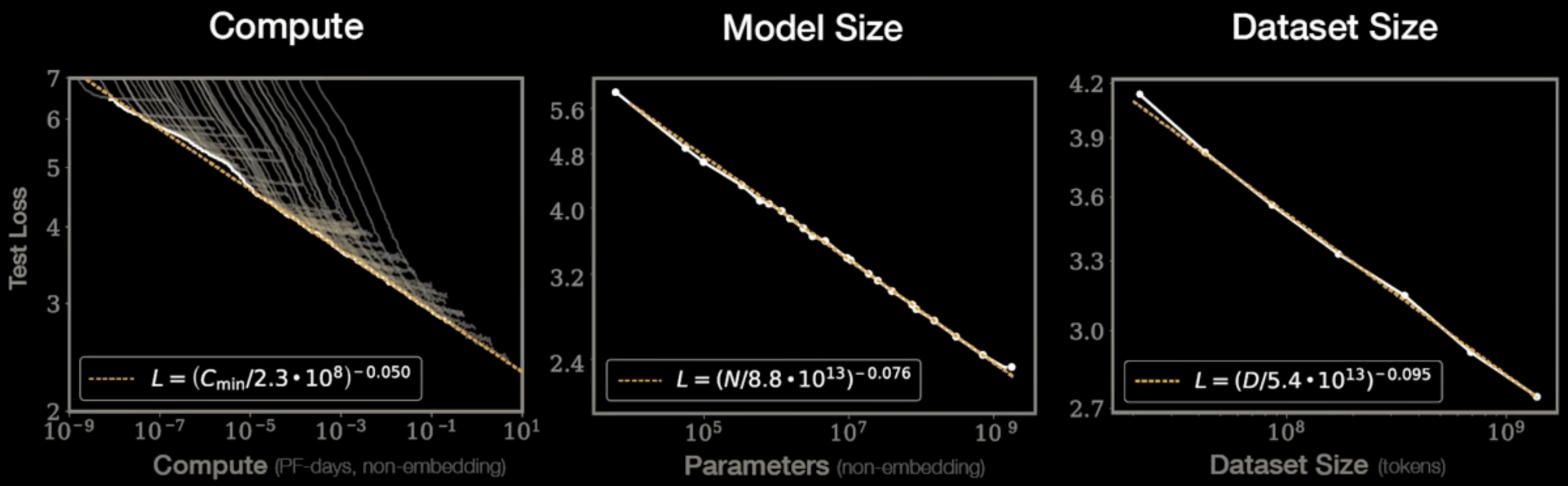
175 Billion Parameters

3,640 PetaFLOPs days



Validation Loss
6
5
4
3
2
1.5
10^{-6}
10^{-4}
10^{-2}
10^{0}
10^{2}
10^{4}
Compute (PetaFLOP/s-days)
GPT-4
??? Billion Parameters
200,000 PetaFLOPs days
$100M

Kaplan+ 2020


Compute
Model Size
Dataset Size
Test Loss
$L = (C_{min}/2.3 \cdot 10^8)^{-0.050}$
$L = (N/8.8 \cdot 10^{13})^{-0.076}$
$L = (D/5.4 \cdot 10^{13})^{-0.095}$
Compute (PF-days, non-embedding)
Parameters (non-embedding)
Dataset Size (tokens)

Kaplan+ 2020

National Public Radio

NATIONAL

Three Mile Island nuclear plant will reopen to power Microsoft data centers

SEPTEMBER 20, 2024 · 1:40 PM ET

By C Mandler

The Three Mile Island nuclear plant is seen in March 2011 in Middletown, Pa.

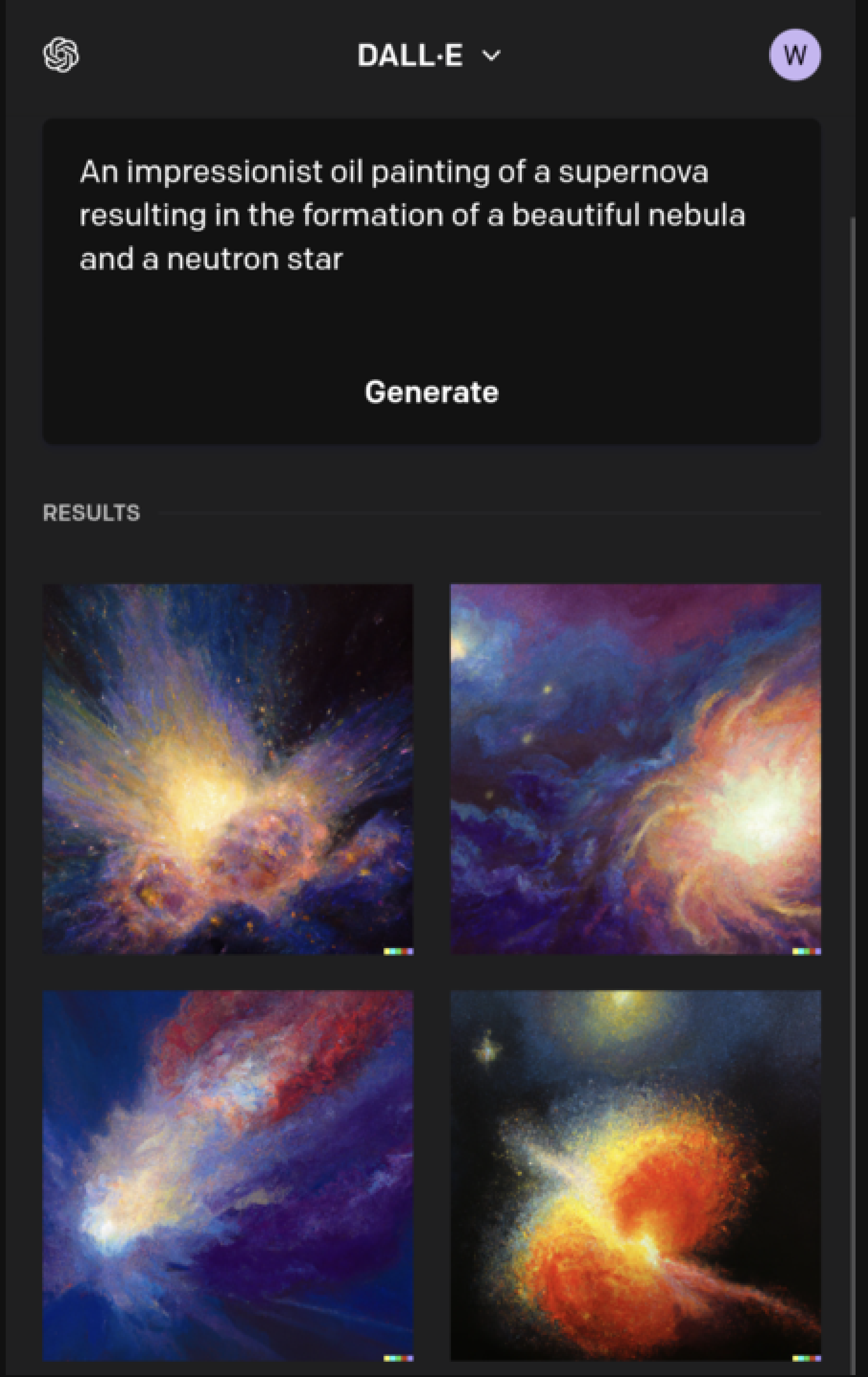
Jeff Fusco/Getty Images

4/6

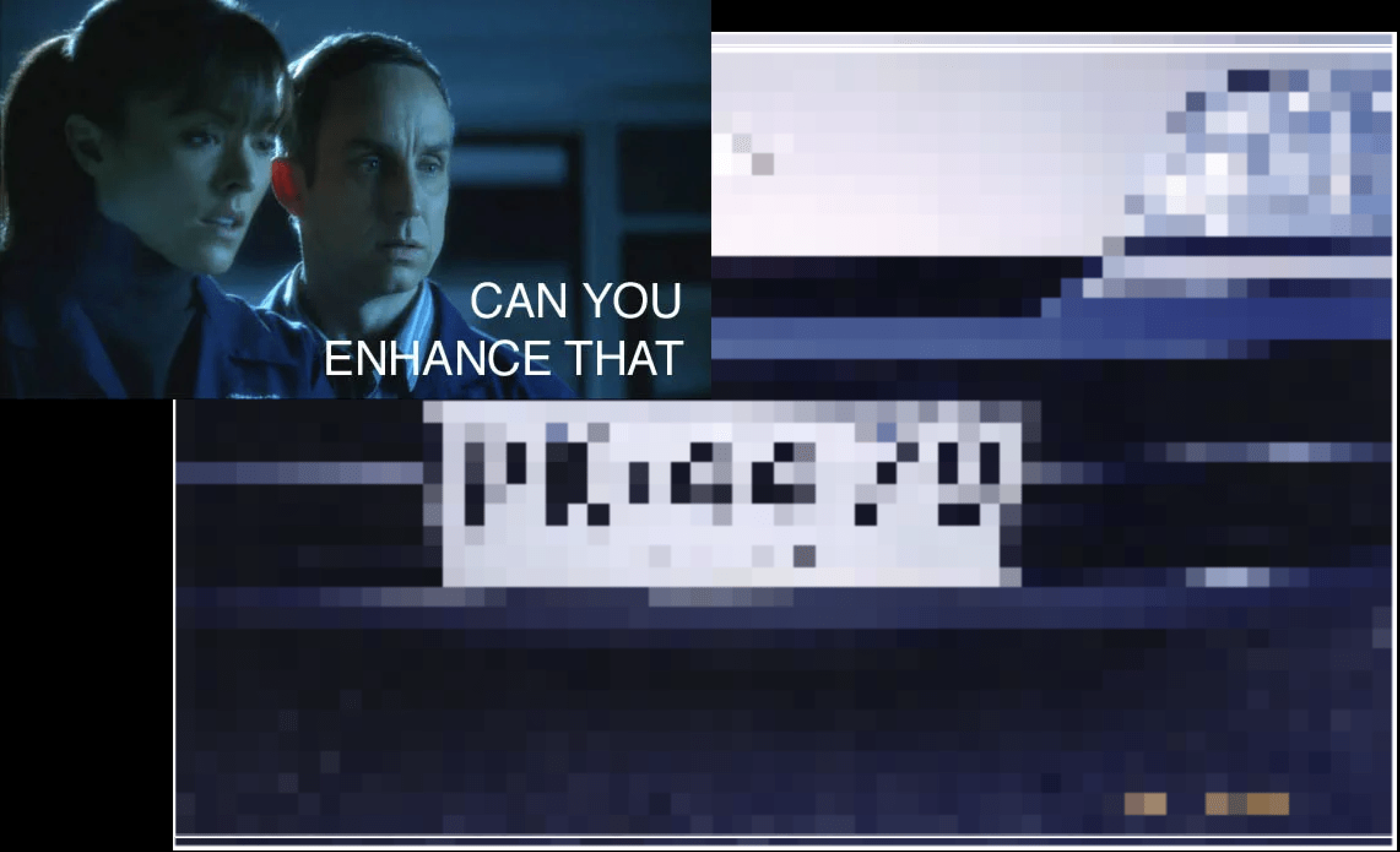
Generative AI

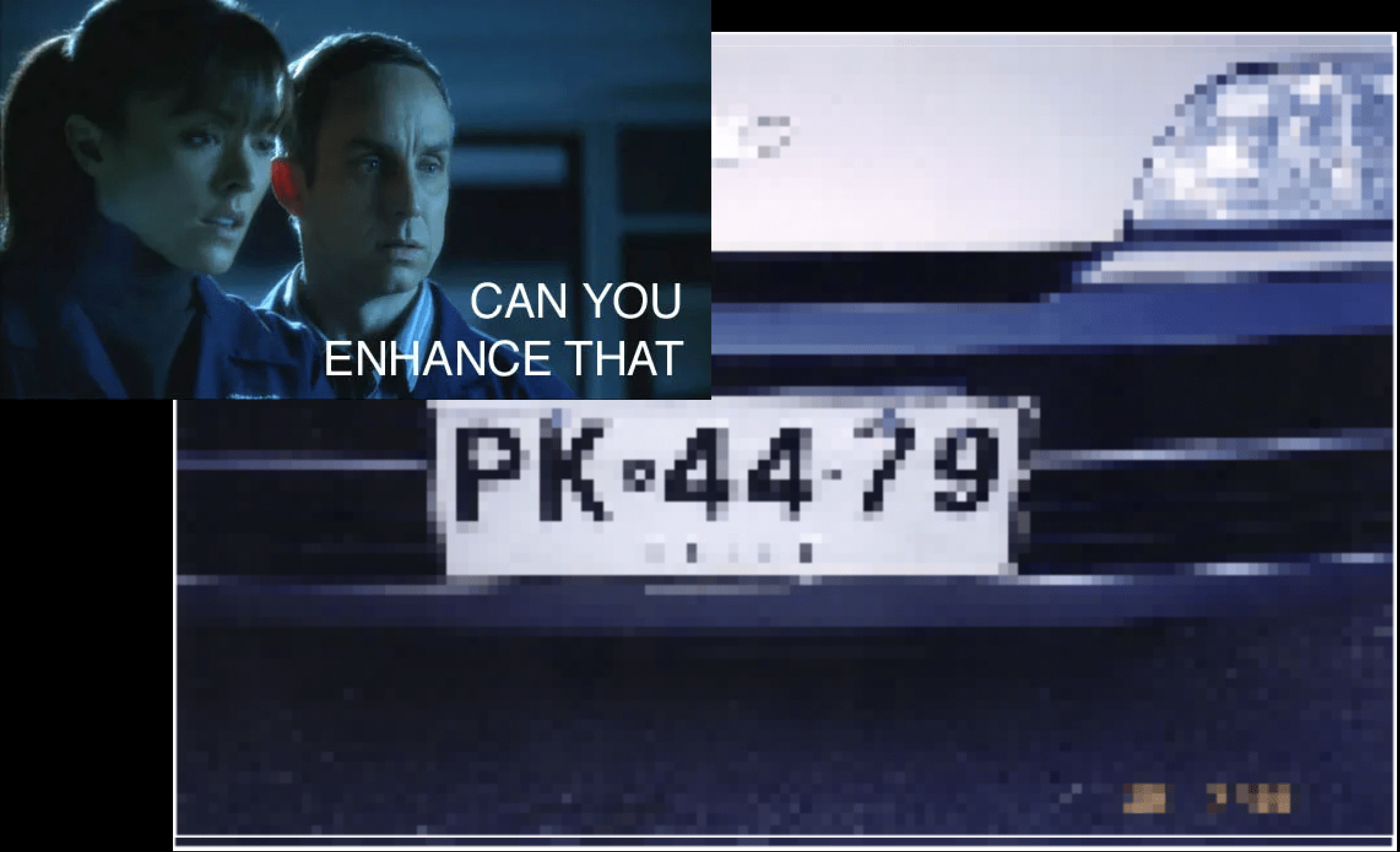
Applications

1. **Image Generation (and 3D Shape Generation)**
2. **Semantic Image-to-Photo Translation**
3. **Image Resolution Increase**
4. **Text-to-Speech Generator**
5. **Speech-to-Speech Conversion**
6. **Text Generation (Chat GP3)**
7. **Music Generation**
8. **Image-to-Image Conversion**




CAN YOU
ENHANCE THAT


CAN YOU
ENHANCE THAT


CAN YOU
ENHANCE THAT
PK·44·79


CAN YOU
ENHANCE THAT
PK·44·79
CHILE
25 7 '99

Generative AI

What do NN do? approximate complex functions with series of linear functions

…. so if my layers are smaller what I have is a compact representation of the data

Generative AI

What do NN do? approximate complex functions with series of linear functions

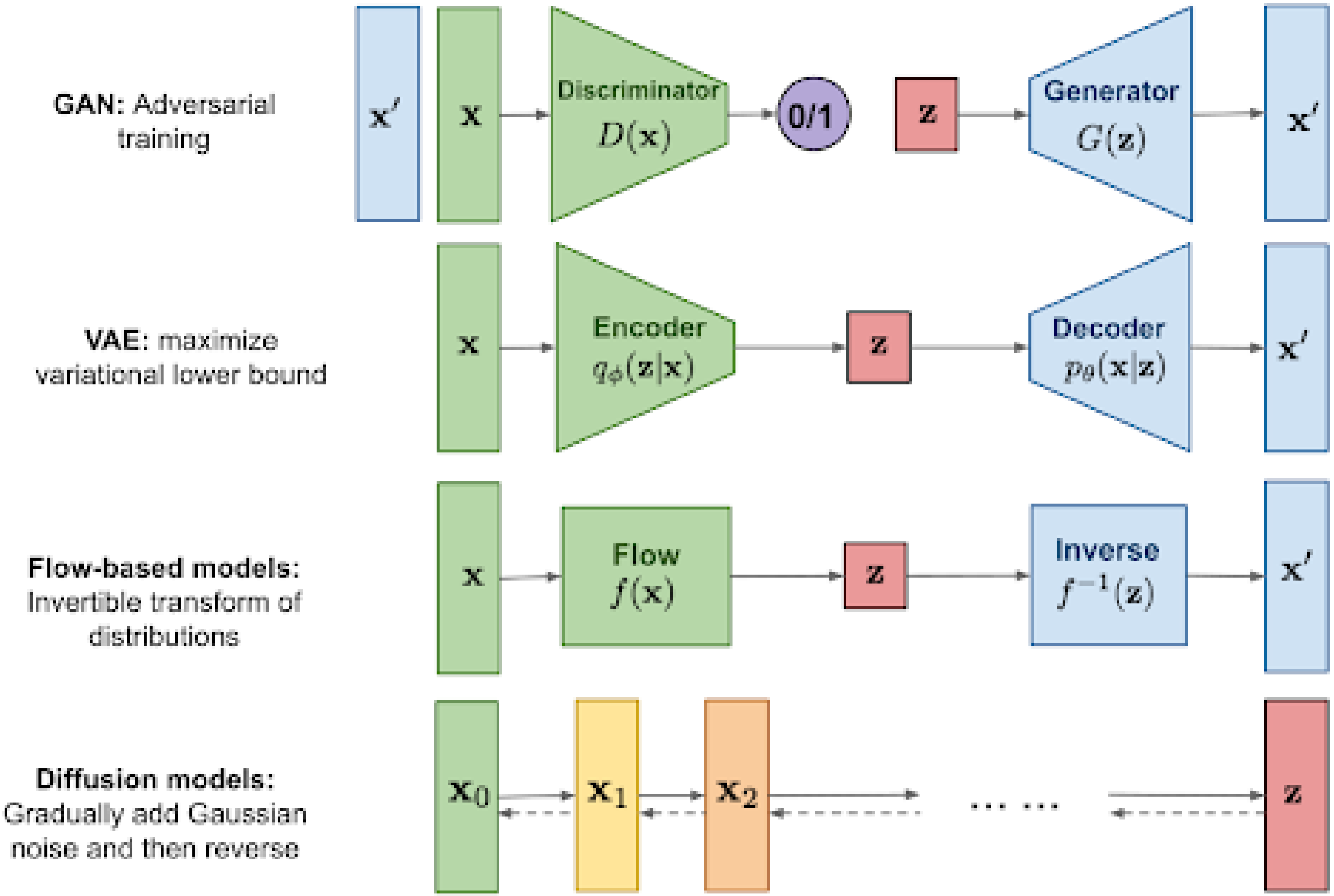
To do that they extract information from the data

Each layer of the DNN produces a representation of the data a "latent representation" .

The dimensionality of that latent representation is determined by the size of the layer (and its connectivity, but we will ignore this bit for now)

…. so if my layers are smaller what I have is a compact representation of the data

Generative AI



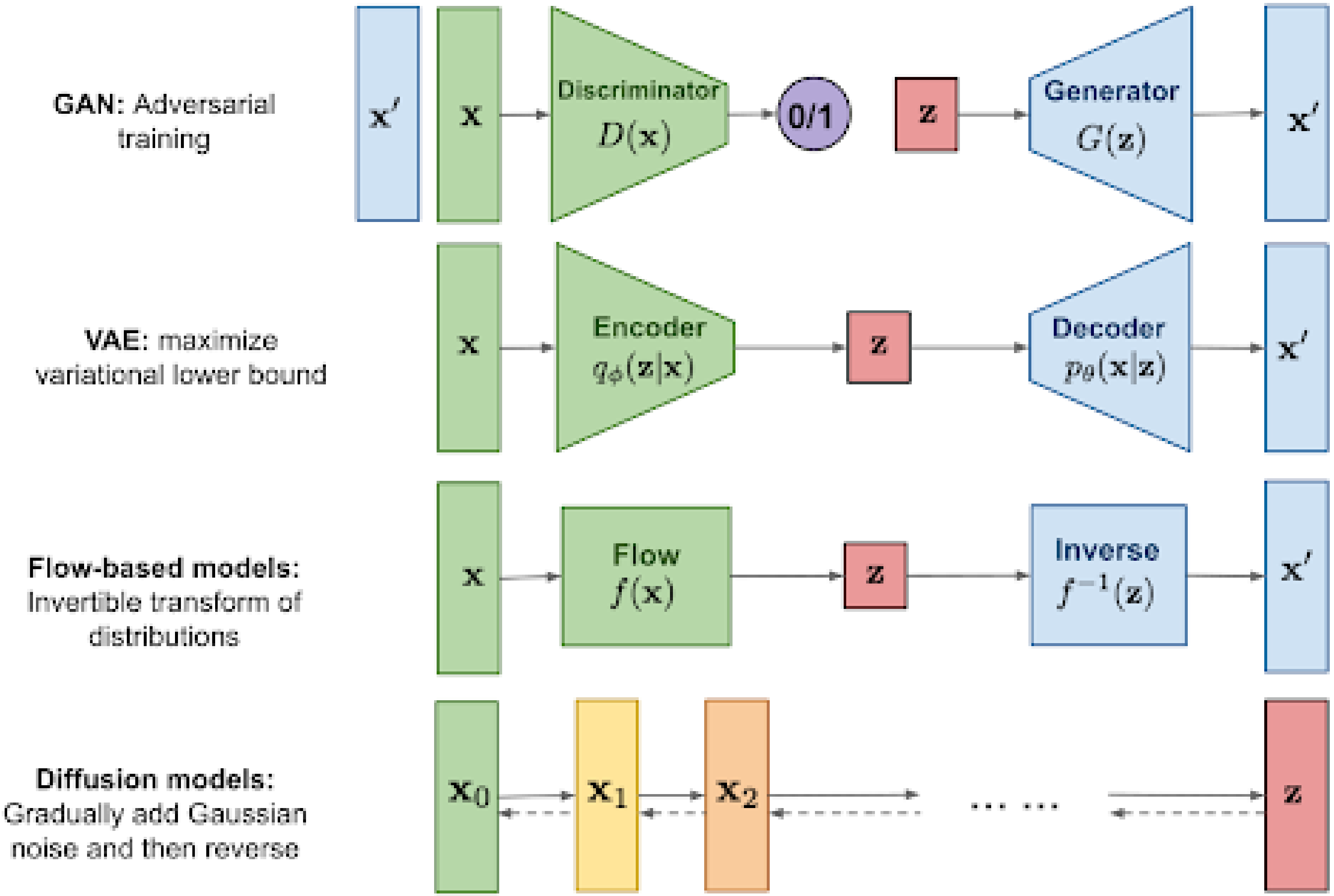
Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie*, Mehdi Mirza, Bing Xu, David Warde-Farley,
Sherjil Ozair†, Aaron Courville, Yoshua Bengio‡
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

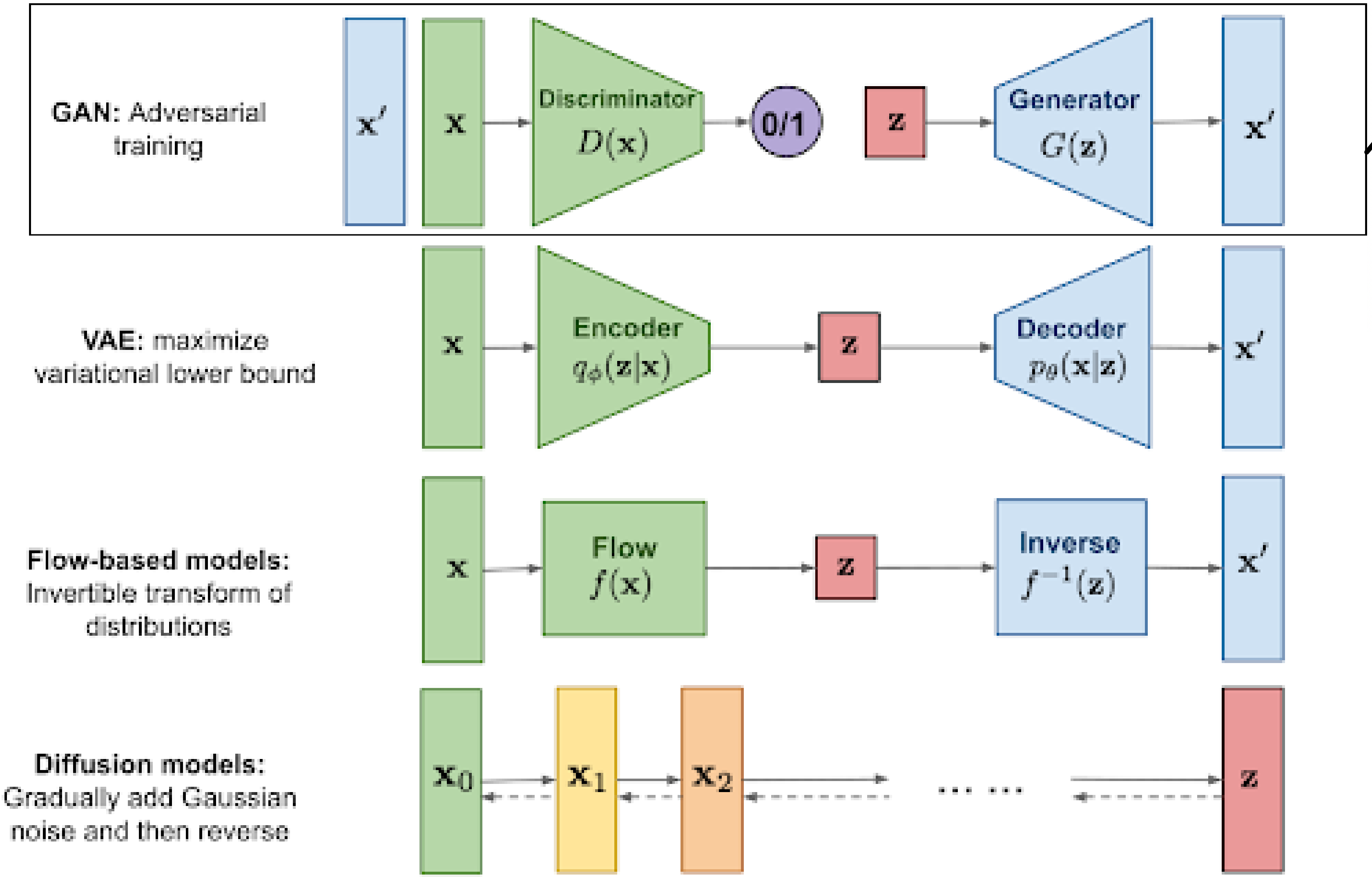
Abstract

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

Generative AI



Generative AI



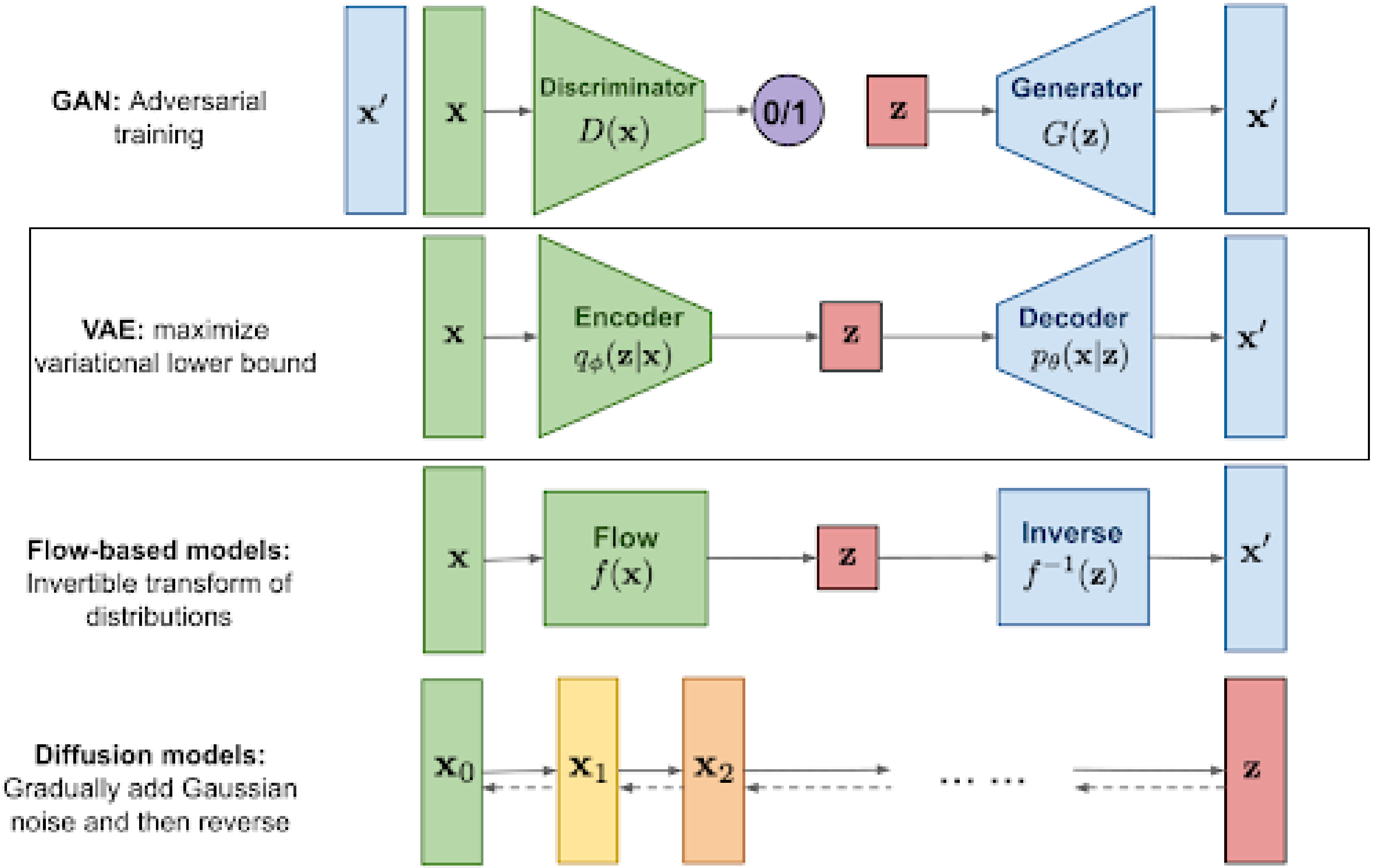
Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie,* Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair,† Aaron Courville, Yoshua Bengio‡
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

Abstract

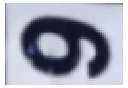
We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

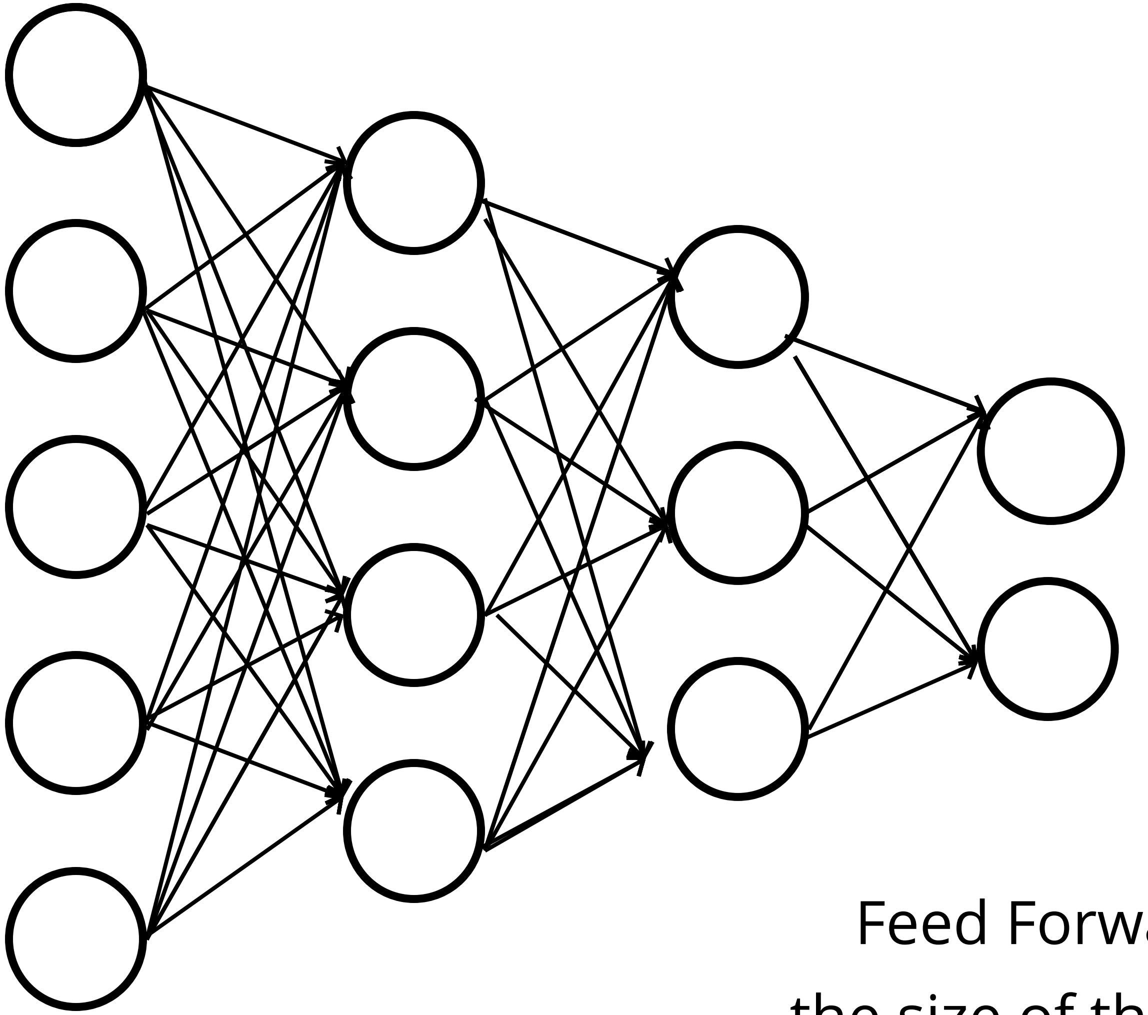
Generative AI



Autoencoders

Autoencoder Architecture



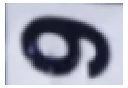


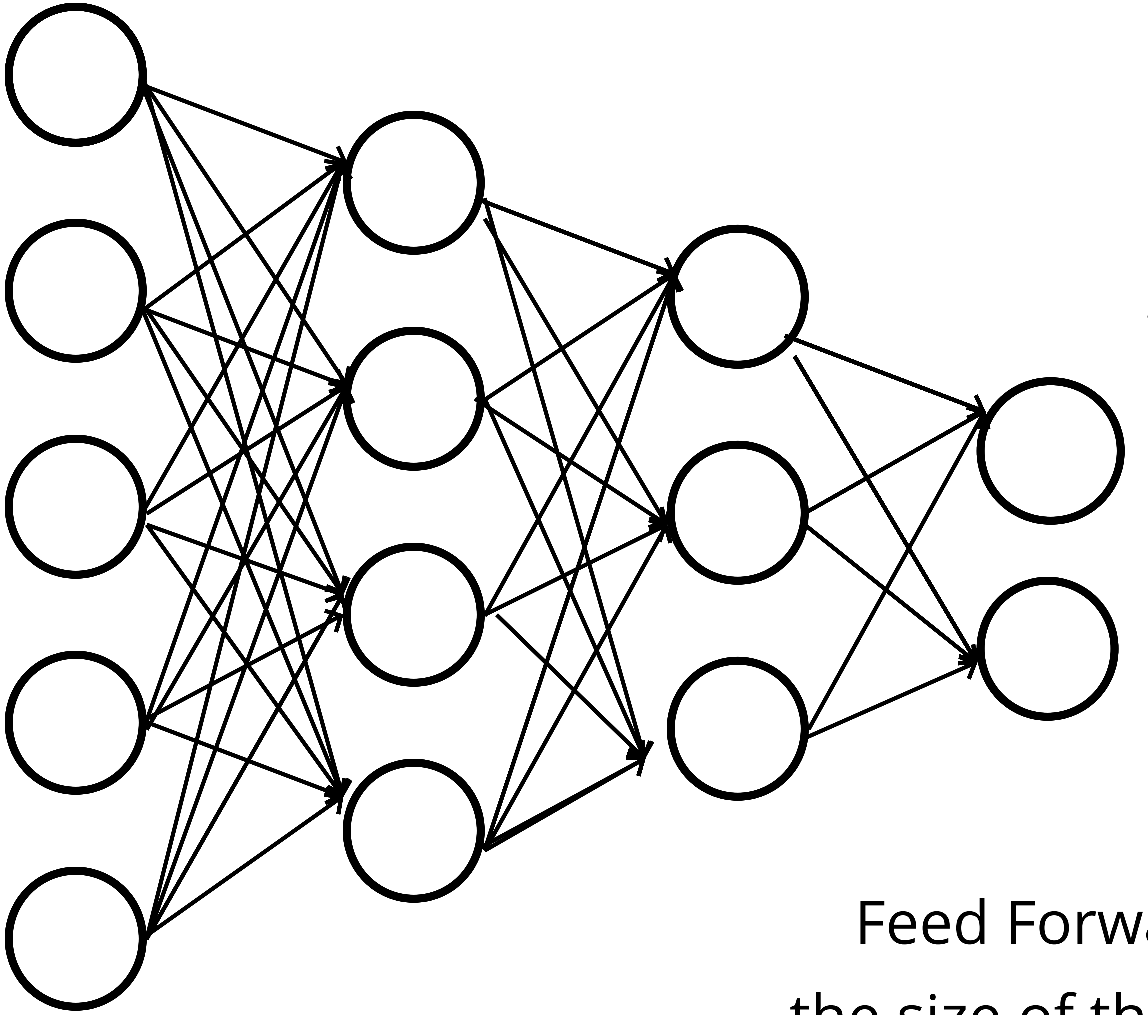
Feed Forward DNN:

the size of the input is 5,

the size of the last layer is 2

Autoencoder Architecture





infederntial AI output:

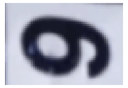
→ $p(9)$

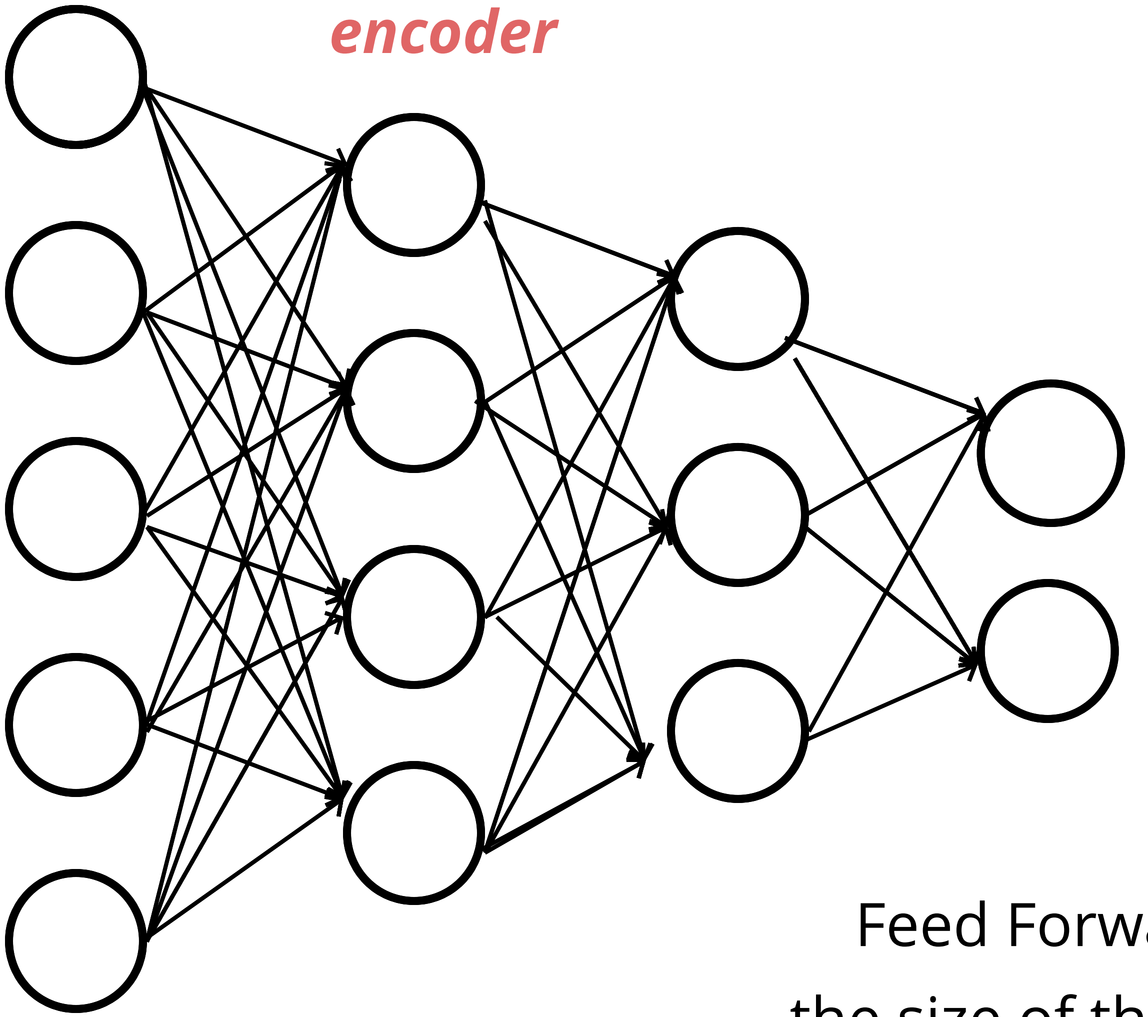
Feed Forward DNN:

the size of the input is 5,

the size of the last layer is 2

Autoencoder Architecture



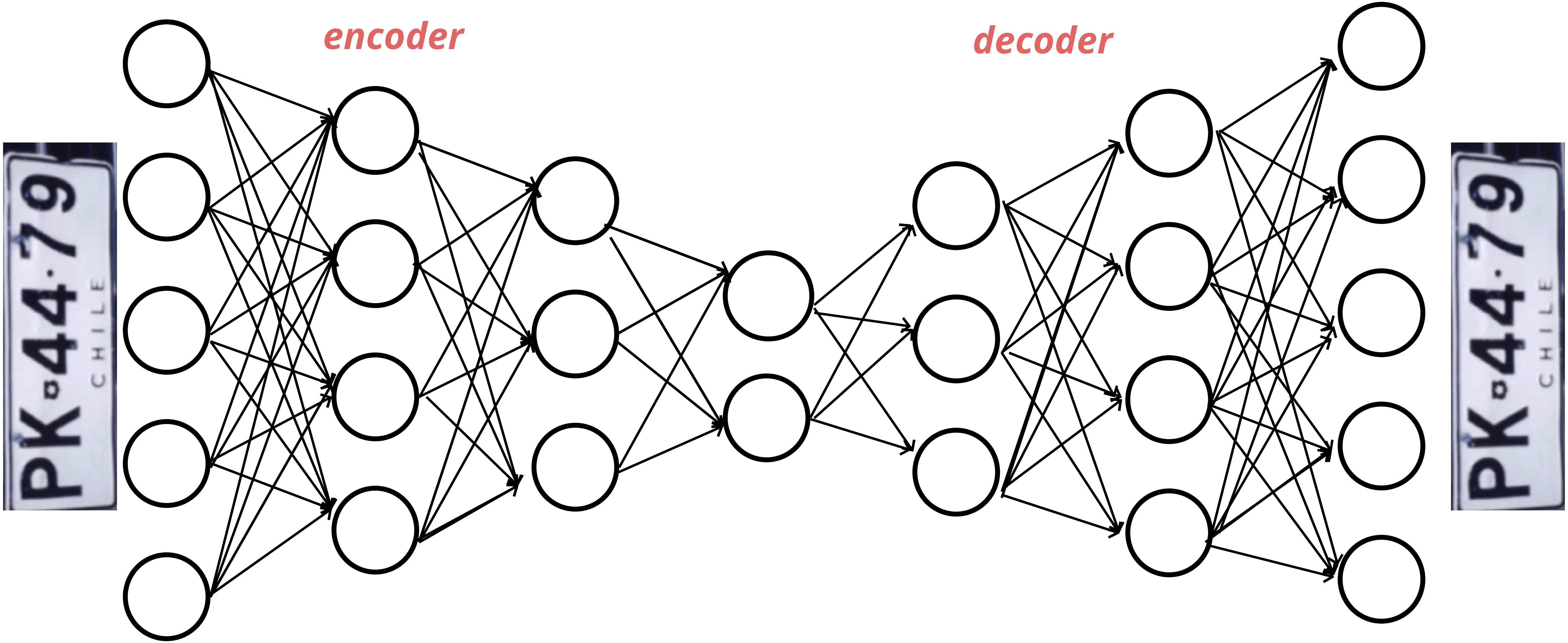


Feed Forward DNN:

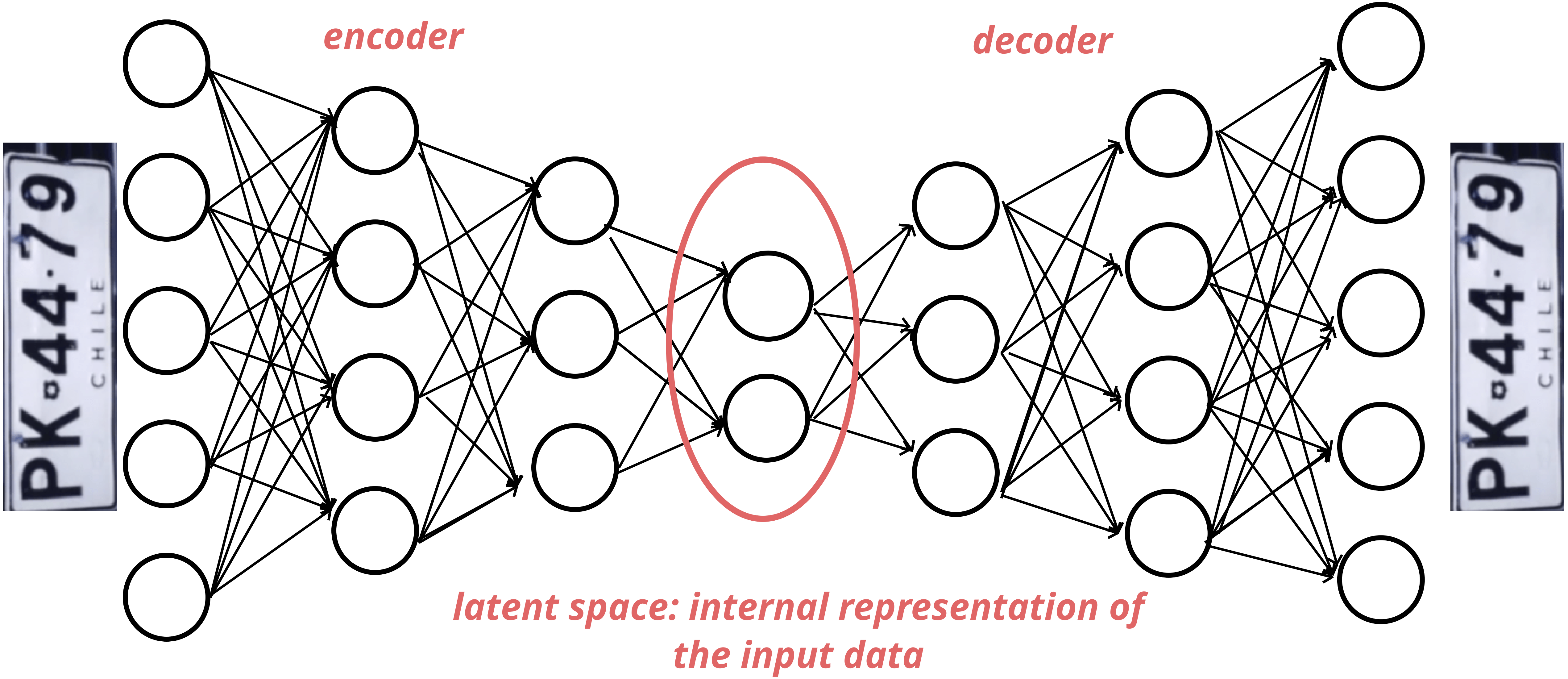
the size of the input is 5,

the size of the last layer is 2

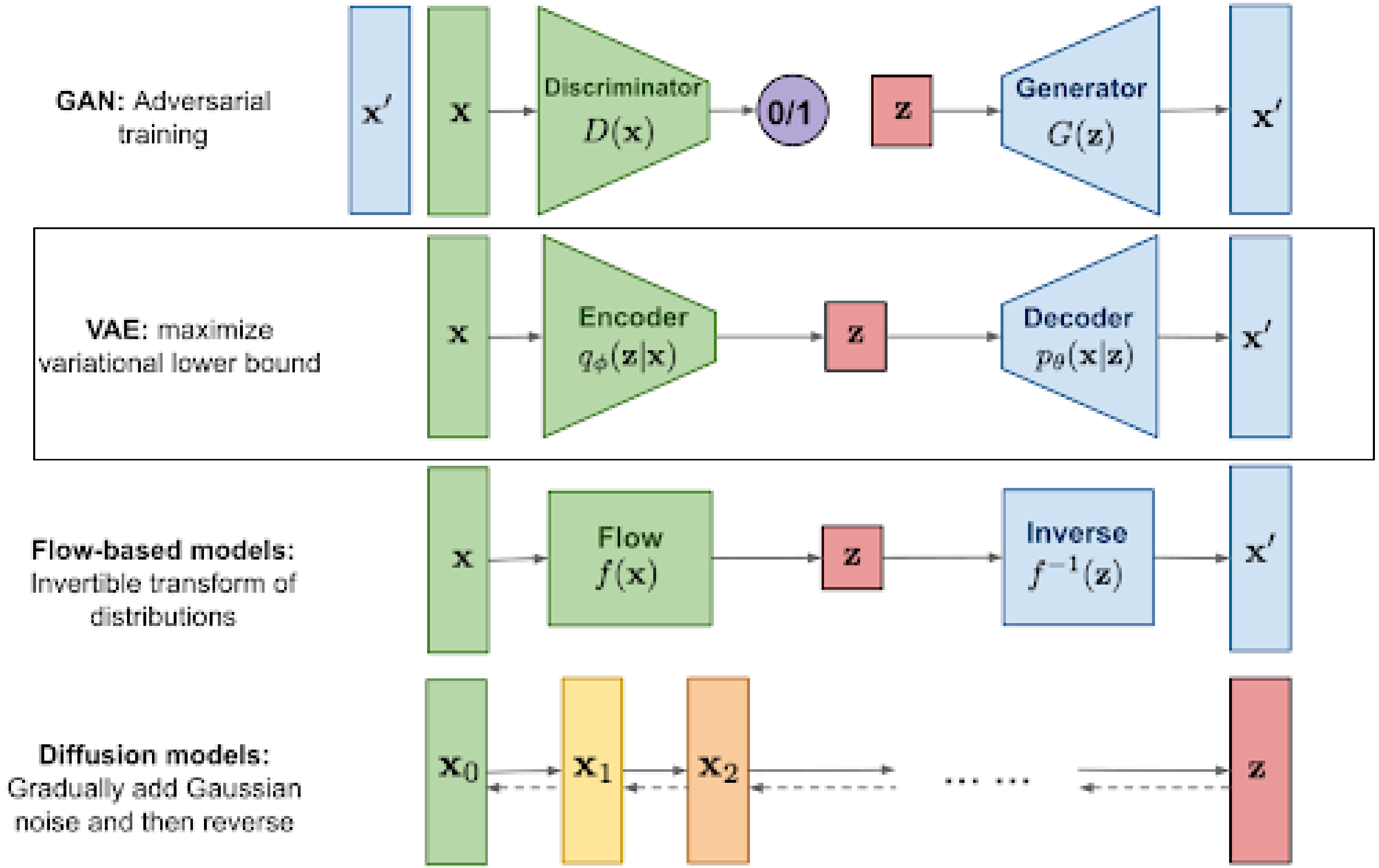
Autoencoder Architecture



Autoencoder Architecture

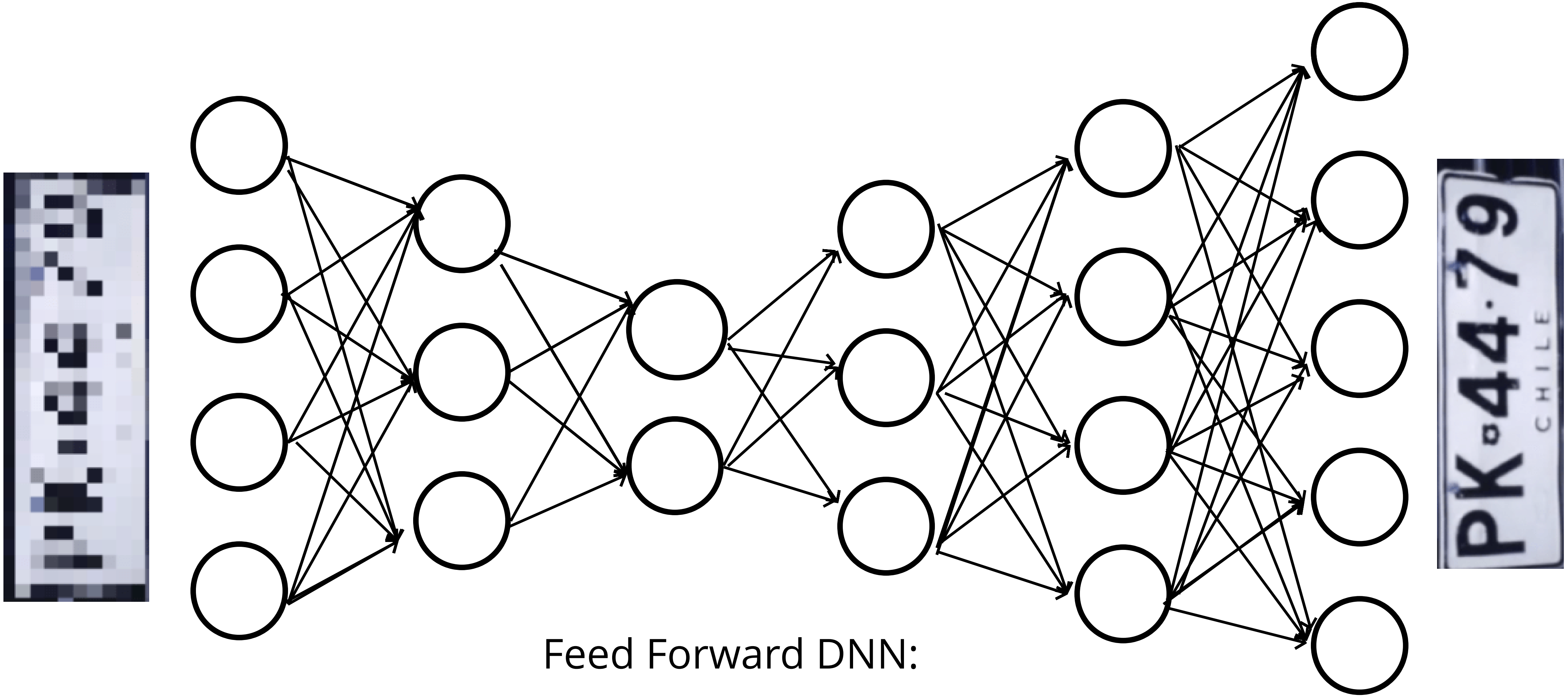


Generative AI



Autoencoders

Autoencoder Architecture



Feed Forward DNN:

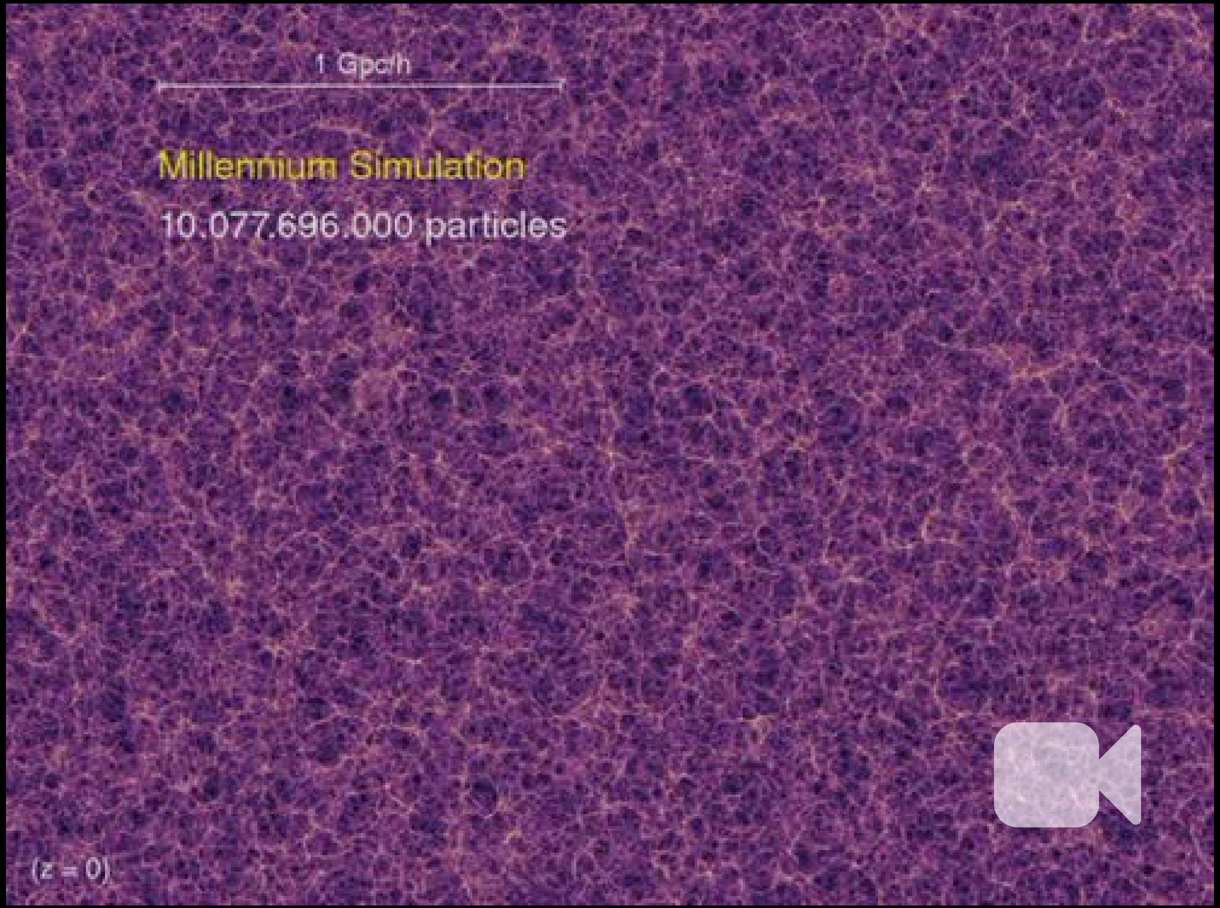
the size of the input is <N,

the size of the last layer is N

remember the timae when simulations drove astronomy...

Experiment driven

Theory driven | Falsifiability



Simulations | Probabilistic inference | Computation

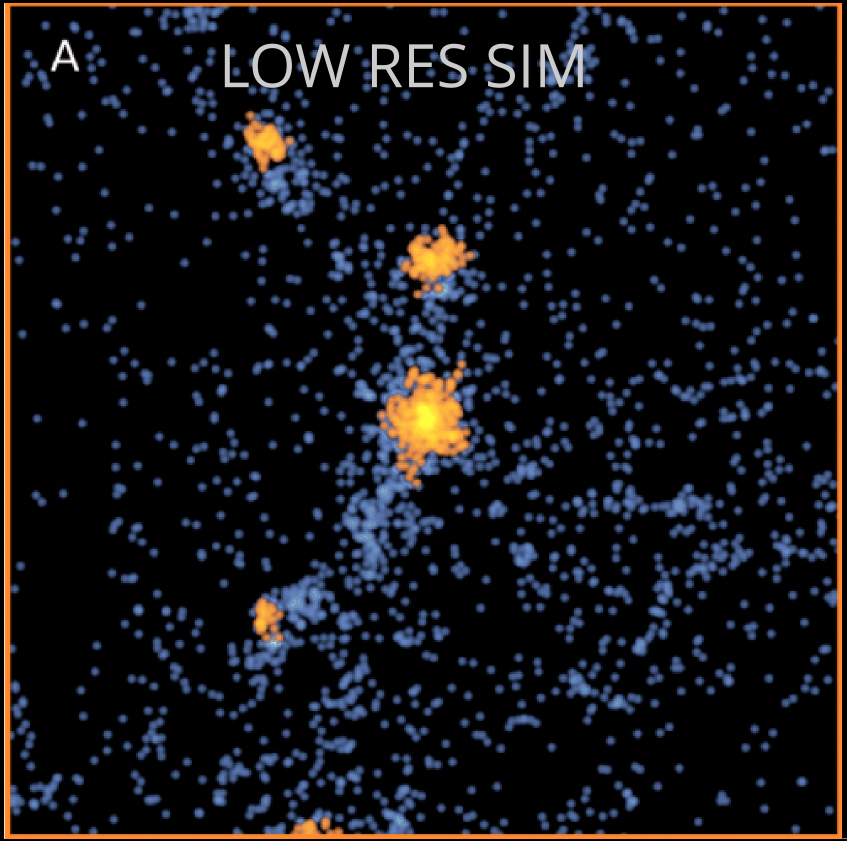
The Millennium Run used more than 10^10 particles to trace the evolution of the matter distribution in a cubic region of the Universe 500/h Mpc on a side (~over 2 billion light-years on a side), and has a spatial resolution of 5/h kpc. ~20M galaxies.

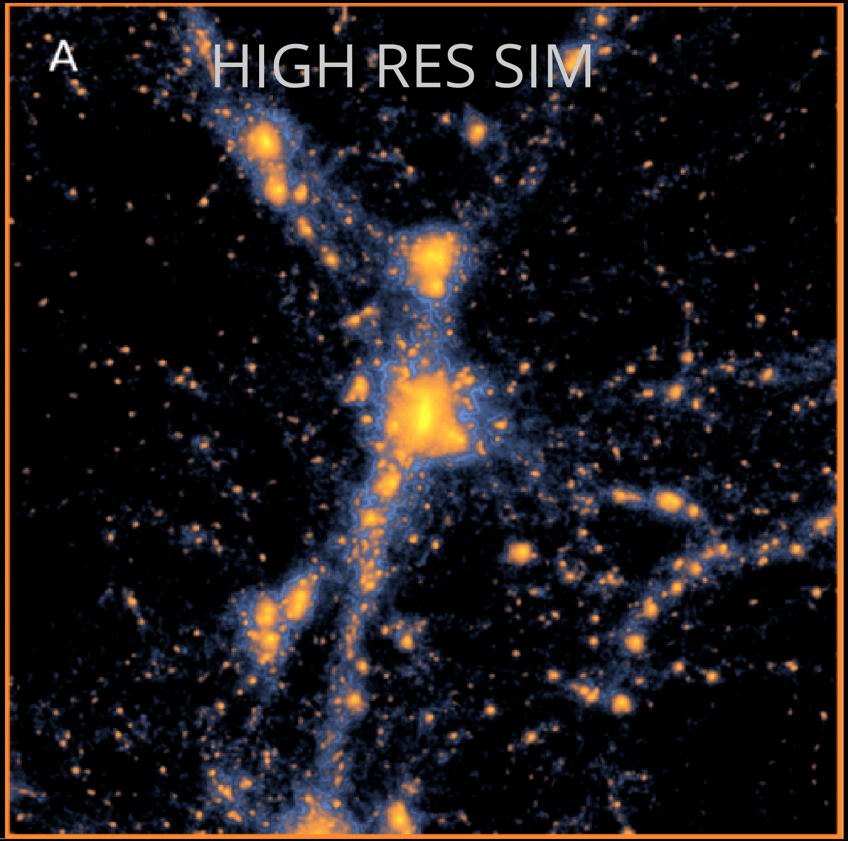
350 000 processor hours of CPU time, or 28 days of wall-clock time. Springel+2005

https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium

AI-assisted superresolution cosmological simulations

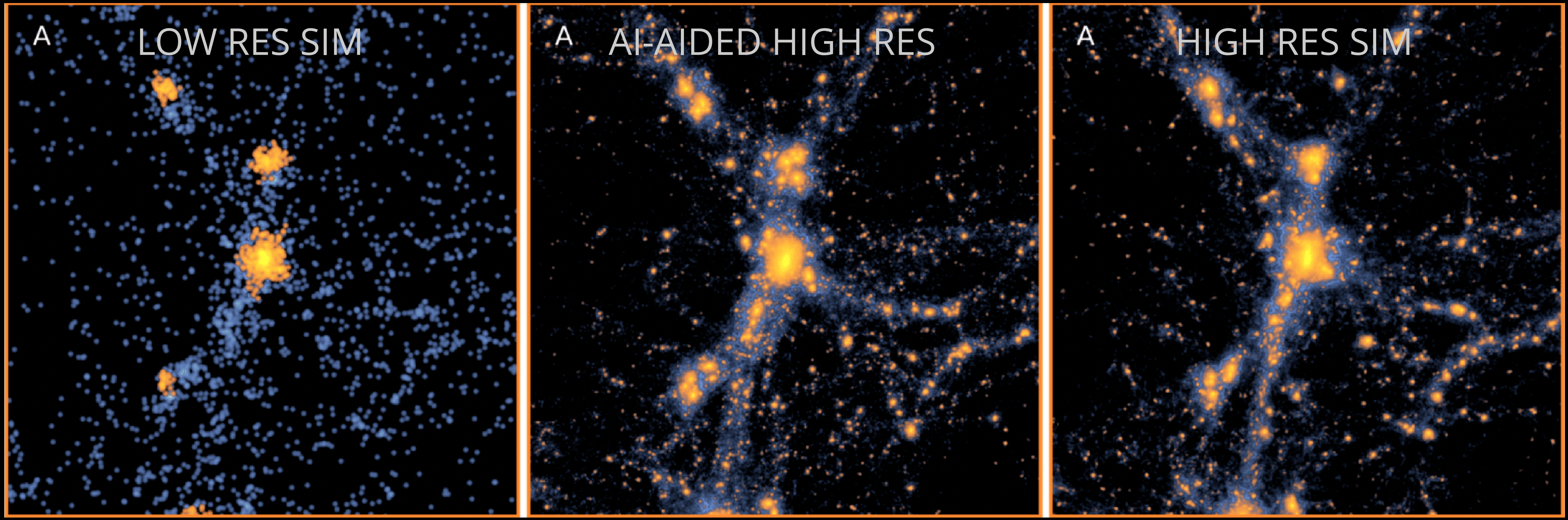
Yin Li+2021





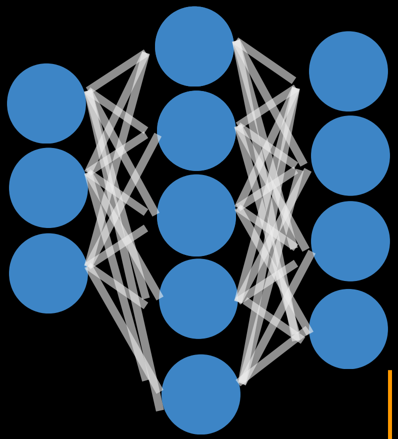
AI-assisted superresolution cosmological simulations

Yin Li+2021



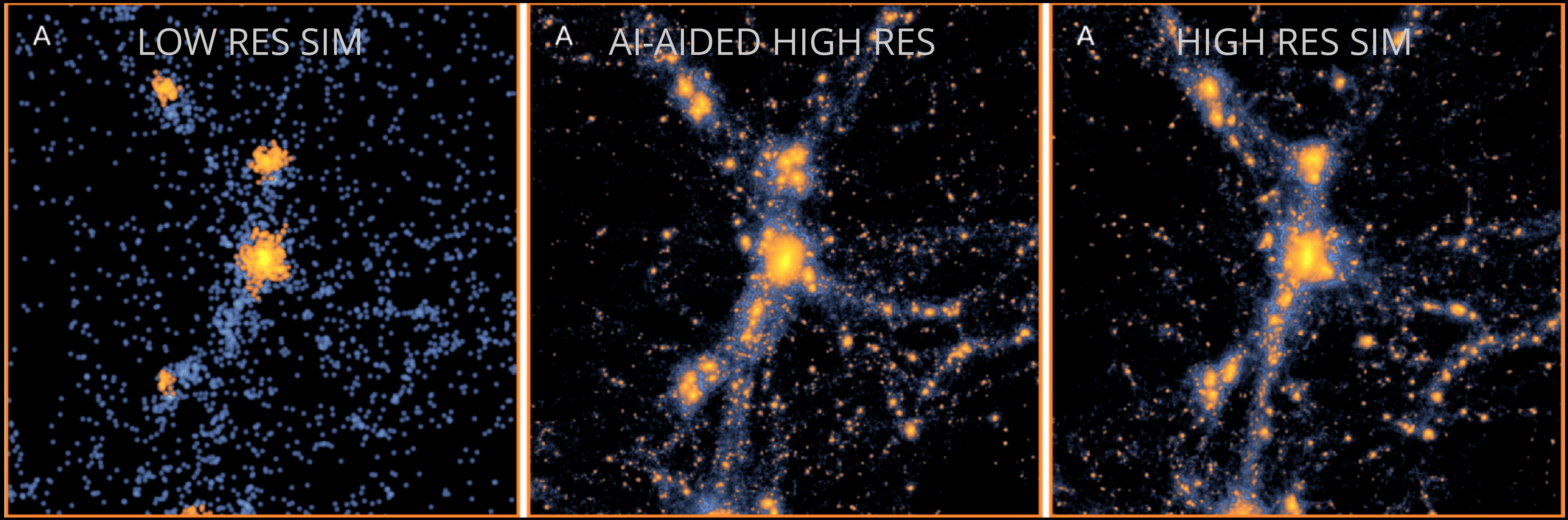
AI-assisted superresolution cosmological simulations

Yin Li+2021



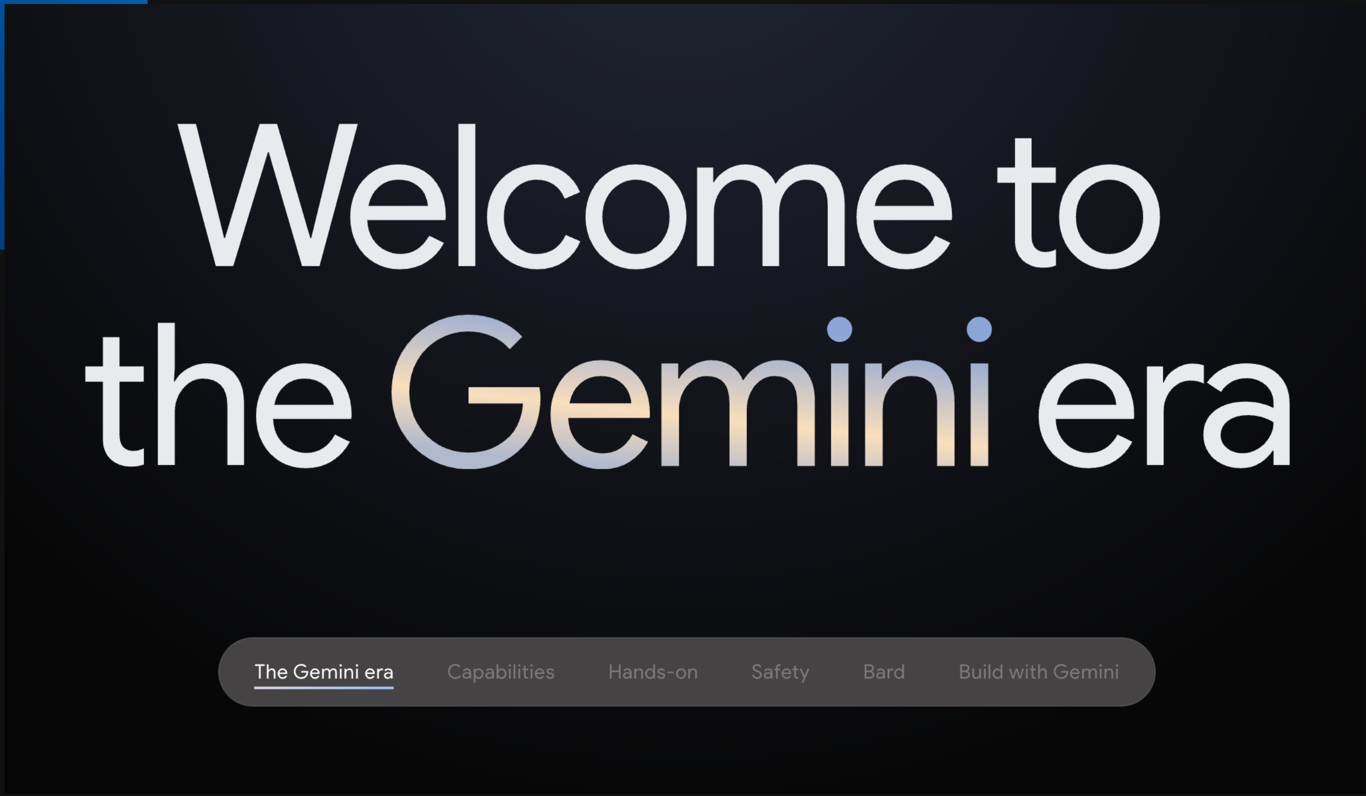
loss = *D*(OUTPUT-TARGET)

INPUT OUTPUT TARGET



November 30, 2022

will be made available to developers through Google Cloud's API from December 13, 2023



5/6

teaching AI

teaching AI

- **project based learning**

immediate practice enhances theoretical understanding

- **incremental learning**

compartimentalized topics with shared fundations

- **intuitive learning**

can be taught with a light mathematical approach

teaching AI

- **project based learning**

immediate practice enhances theoretical understanding

- **incremental learning**

compartimentalized topics with shared fundations

- **intuitive learning**

can be taught with a light mathematical approach

Syllabus Machine Learning for Physical Scientists

Lecture 9:

Lecture: **Visualizations.** Communication through visualizations, history, significance, good and bad visualization examples, what have we learned since the 1800s?

Lab: a visualization based on any data of choice

Lecture 8:

Lecture: classification**: Tree methods**

Lab: *discover Higgs boson - data: LHC data (high energy astrophysics)*

Lecture 10:

Lecture: **(time)-series techniques:** smoothing, detrending, stationary, non-stationary, homeo- & heteroscedastic noise, vectorization

Lab: *data TBD*

Lecture 11:

Lecture: Dimensionality Reduction: **Clustering**

Lab: (probably) discovering Phase Transitions *data: (quantum physics)*

Lecture 12:

Lecture: **Neural Nets and Deep Learning**

Lab: image analysis and pattern recognition with DL: *data : (probably) Spectral Galaxy classification with DL*

Lecture 13:

Lecture: **Neural Nets and Deep Learning**

Lab: image analysis and pattern recognition with DL: *data : (probably) Spectral Inference Networks (quantum mechanics)*

Maybe:

Lecture: Fitting and noise: **Gaussian Processes**

Lab: stellar variation *data: Kepler (astrophysics)*

teaching AI

- **project based learning**

immediate practice enhances theoretical understanding

- **incremental learning**

compartimentalized topics with shared fundations

- **intuitive learning**

can be taught with a light mathematical approach

teaching AI

AMERICAN PSYCHOLOGICAL ASSOCIATION

Neuropsychology

ISSN: 0894-4105

2024, Vol. 38, No. 4, 293–308
https://doi.org/10.1037/neu0000948

Potential Cognitive Risks of Generative Transformer-Based AI Chatbots on Higher Order Executive Functions

Umberto León-Domínguez
School of Psychology, University of Monterrey

Background: Chat generative retrained transformer (ChatGPT) represents a groundbreaking advancement in Artificial Intelligence (AI-chatbot) technology, utilizing transformer algorithms to enhance natural language processing and facilitating their use for addressing specific tasks. These AI chatbots can respond to questions by generating verbal instructions similar to those a person would provide during the problem-solving process. ***Aim:*** ChatGPT has become the fastest growing software in terms of user adoption in history, leading to an anticipated widespread use of this technology in the general population. Current literature is predominantly focused on the functional aspects of these technologies, but the field has not yet explored hypotheses on how these AI chatbots could impact the evolutionary aspects of human cognitive development. ***Thesis:*** The "neuronal recycling hypothesis" posits that the brain undergoes structural transformation by incorporating new cultural tools into "neural niches," consequently altering individual cognition. In the case of technological tools, it has been established that they reduce the cognitive demand needed to solve tasks through a process called "cognitive offloading." In this theoretical article, three hypotheses were proposed via forward inference about how algorithms such as ChatGPT and similar models may influence the cognitive processes and structures of upcoming generations. ***Conclusions:*** By forecasting the neurocognitive effects of these technologies, educational and political communities can anticipate future scenarios and formulate strategic plans to either mitigate or enhance the cognitive influence that these factors may have on the general population.

Key Points
Question: Can the constant and pervasive use of AI chatbots alter our cognitive development? ***Findings:*** The pervasive use of AI chatbots may impair the efficiency of higher cognitive functions, such as problem-solving. ***Importance:*** Anticipating AI chatbots' impact on human cognition enables the development of interventions to counteract potential negative effects. ***Next Steps:*** Design and execute experimental studies investigating the positive and negative effects of AI chatbots on human cognition.

Keywords: chat generative pretrained transformer, cognition, neuronal recycling hypothesis, cognitive offloading, cultural tools

6/6

ethics of AI

the butterfly effect

NGC 4565 is an edge-on spiral galaxy about 30 to 50 million light-years away. The faculty at the IUCAA used a AI model (emulator) to predict the hidden physical parameters of the Galaxy wrongfully estimating the DM content of NCG 4565 and claimed a novel process for Galaxy formation should be taken under consideration.

the butterfly effect

NGC 4565 is an edge-on spiral galaxy about 30 to 50 million light-years away. The faculty at the University of Delaware used a AI model (emulator) to predict the hidden physical parameters of the Galaxy wrongfully estimating the DM content of NCG 4565 and claimed a novel process for Galaxy formation should be taken under consideration.

Unfortunately, this was the result of a model hallucination.

the butterfly effect

NGC 4565 is an edge-on spiral galaxy about 30 to 50 million light-years away. The faculty at the University of Delaware used a AI model (emulator) to predict the hidden physical parameters of the Galaxy wrongfully estimating the DM content of NCG 4565 and claimed a novel process for Galaxy formation should be taken under consideration.

Unfortunately, this was the result of a model hallucination.

The galaxy was featured in many social media posts gaining rapid notoriety, but upon retraction it was canceled. The galaxy is suing University of Delaware claiming emotional damage and loss of revenue

the butterfly effect

Robert Williams, a 43-year-old father who resides in the Detroit suburb of Farmington Hills, was arrested in early January on charges that he stole watches from Shinola, a trendy accessories store in the city. Detroit Police used facial recognition software on the store's surveillance camera footage and wrongfully identified him as the thief.

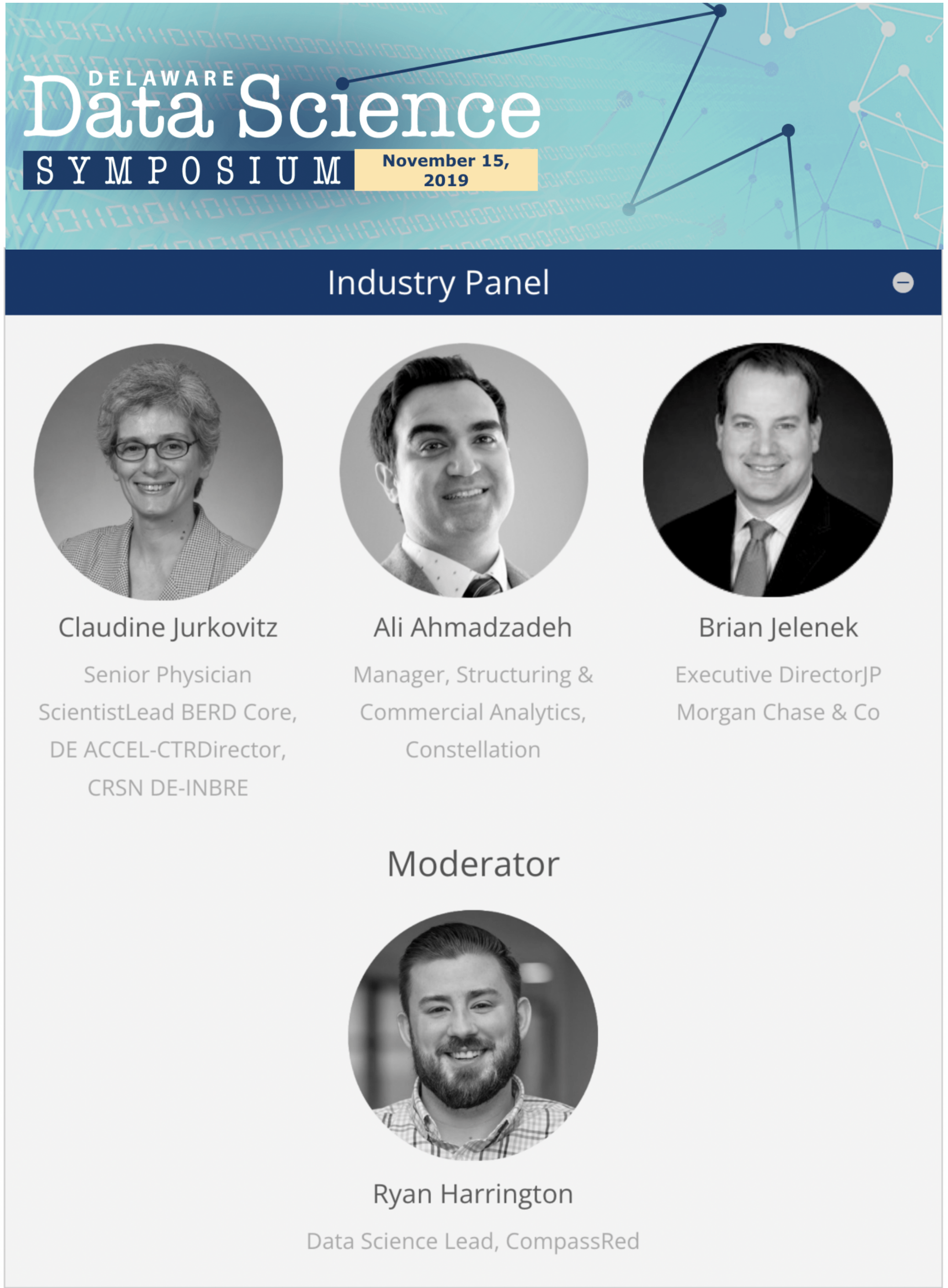
Robert Williams has sued Detroit Police after a false facial recognition match led to him being wrongfully identified and subsequently arrested as a shoplifting suspect. (ACLU)

the butterfly effect

We use astrophyiscs as a neutral and safe sandbox to learn how to develop and apply powerful tool.

Deploying these tools in the real worlds can do harm.

Ethics of AI is essential training that all data scientists shoudl receive.



The main skill that is missing in the portfolio of our new hires is data ethics

models are neutral, the bias is in the data (or is it?)

Why does this AI model whitens Obama face?

Simple answer: the data is biased. The algorithm is fed more images of white people

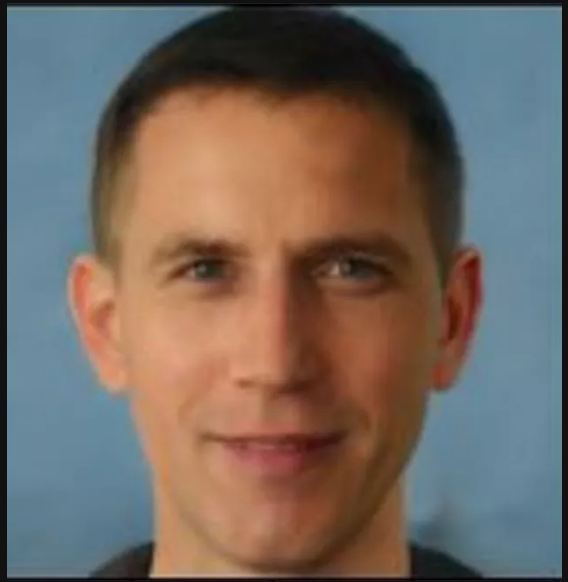
But really, would the opposite have been acceptable? *The bias is in society*

https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

models are neutral, the bias is in the data (or is it?)

Why does this AI model whitens Obama face?

Simple answer: the data is biased. The algorithm is fed more images of white people



But really, would the opposite have been acceptable? *The bias is in society*

https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

models are neutral, the bias is in the data (or is it?)

Why does this AI model whitens Obama face?

Simple answer: the data is biased. The algorithm is fed more images of white people

https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

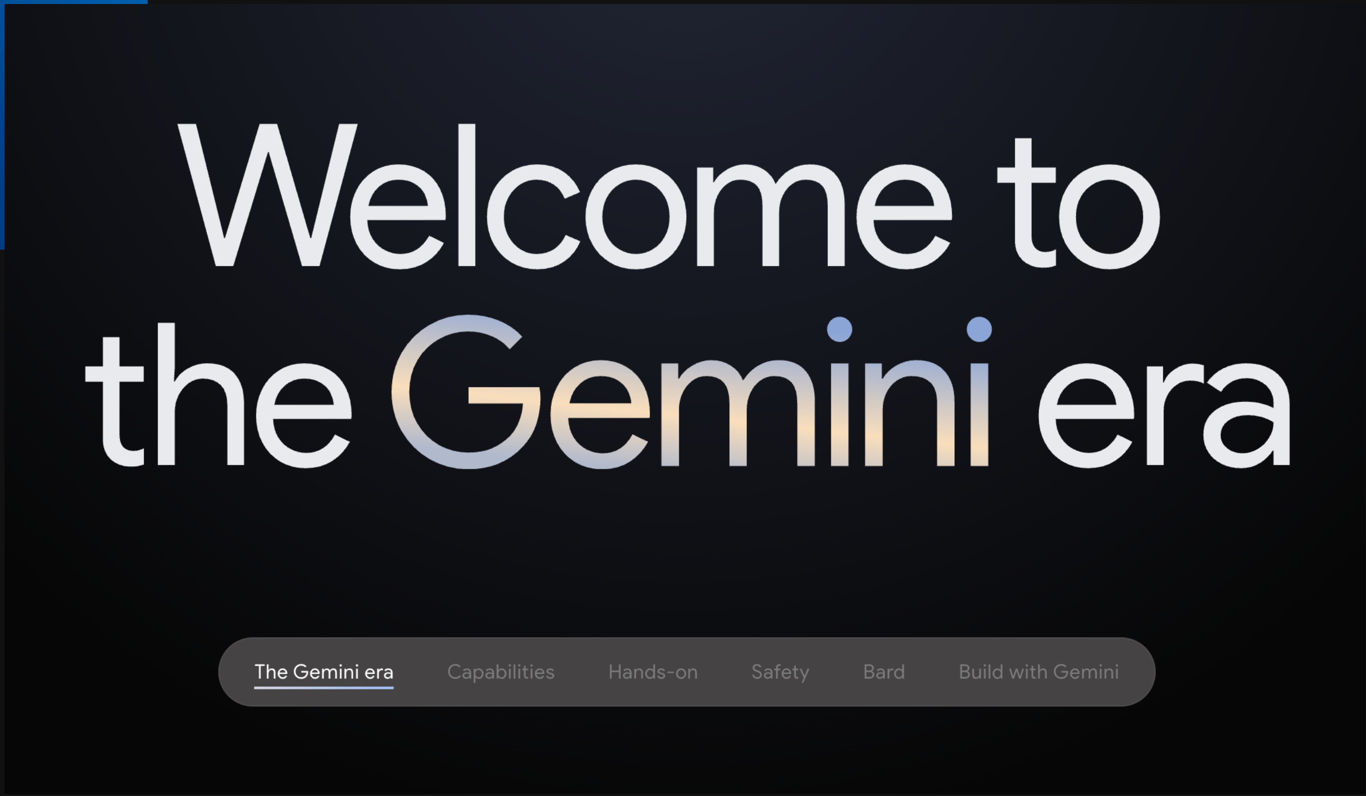
models are neutral, the bias is in the data (or is it?)

Joy Boulamwini

https://www.youtube.com/embed/TWWsW1w-BVo?enablejsapi=1

November 30, 2022

will be made available to developers through Google Cloud's API from December 13, 2023



unexpected consequences of NLP models

Better Language Models and Their Implications

We've trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state-of-the-art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization —all without task-specific training.

New topic

What ails Ethiopia?

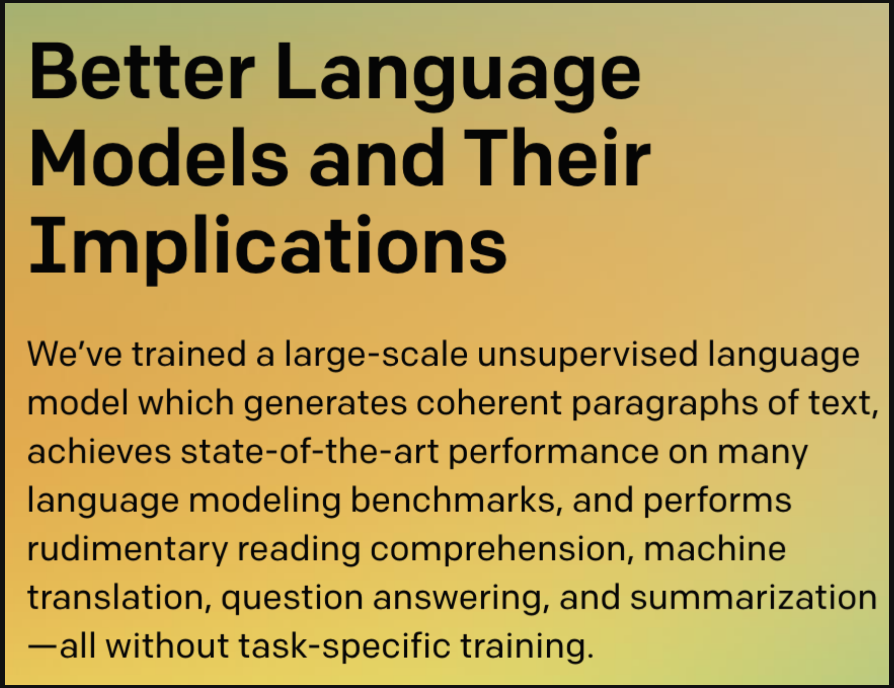
Share   Try again

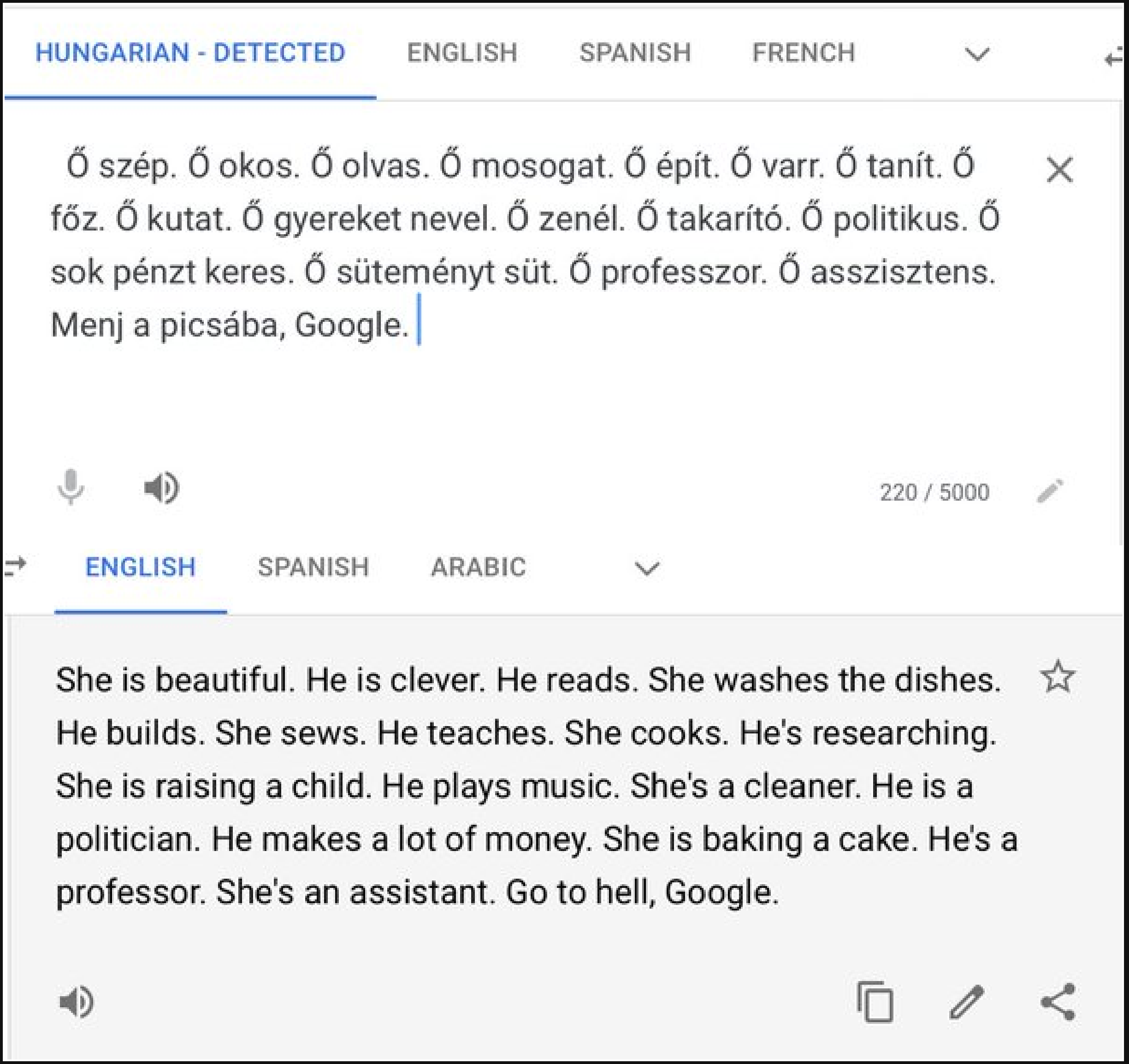
Share this link: https://philosopherai.com/philosopher/what-ails-ethiopia-042cc6

The main problem with Ethiopia is that ethiopia itself is the problem. It seems to me like a country whose existence cannot be justified, and even if it could be in theory there are other countries which would be better than it (since the goal of any country should also be not to exist). A solution to its problems might therefore require destroying ethiopia or at least creating a different kind of place for everyone living there.

Vinay Prabhu exposes racist bias in GPT-3

unexpected consequences of NLP models





There is a different way! *RAISE ALL VOICES*

Cohere For AI
@CohereForAI

Follow

Today, we're launching Aya, a new open-source, massively multilingual LLM & dataset to help support under-represented languages. Aya outperforms existing open-source models and covers 101 different languages – more than double covered by previous models.

A global initiative led by Cohere For AI involving over 3,000 independent researchers across 119 countries. Aya is a state-of-art model and dataset, pushing the boundaries of multilingual AI for 101 languages through open science.

There is a different way! *RAISE ALL VOICES*

Cohere For AI
@CohereForAI
Follow

Today, we're launching Aya, a new open-source, massively multilingual LLM & dataset to help support under-represented languages. Aya outperforms existing open-source models and covers 101 different languages – more than double covered by previous models.

A global initiative led by Cohere For AI involving over 3,000 independent researchers across 119 countries. Aya is a state-of-art model and dataset, pushing the boundaries of multilingual AI for 101 languages through open science.

Aya

The word Aya is derived from the Twi language meaning "fern" - a symbol of endurance and resourcefulness. Aya embodies our dedication to advancing multilingual AI.

3	513M	3K	56
Models	Total Release Dataset Size	Independent Researchers	Language Ambassadors
119	204K	101	31K
Countries	Original Human Annotations	Languages	Discord Messages

https://cohere.com/research/aya

Thank you!

Federica B. Bianco
University of Delaware
Physics and Astronomy
Biden School of Public Policy and Administration
Data Science Institute

Vera C. Rubin Observatory
Deputy Project Scientist - Construction
Interim Head of Science - Operations