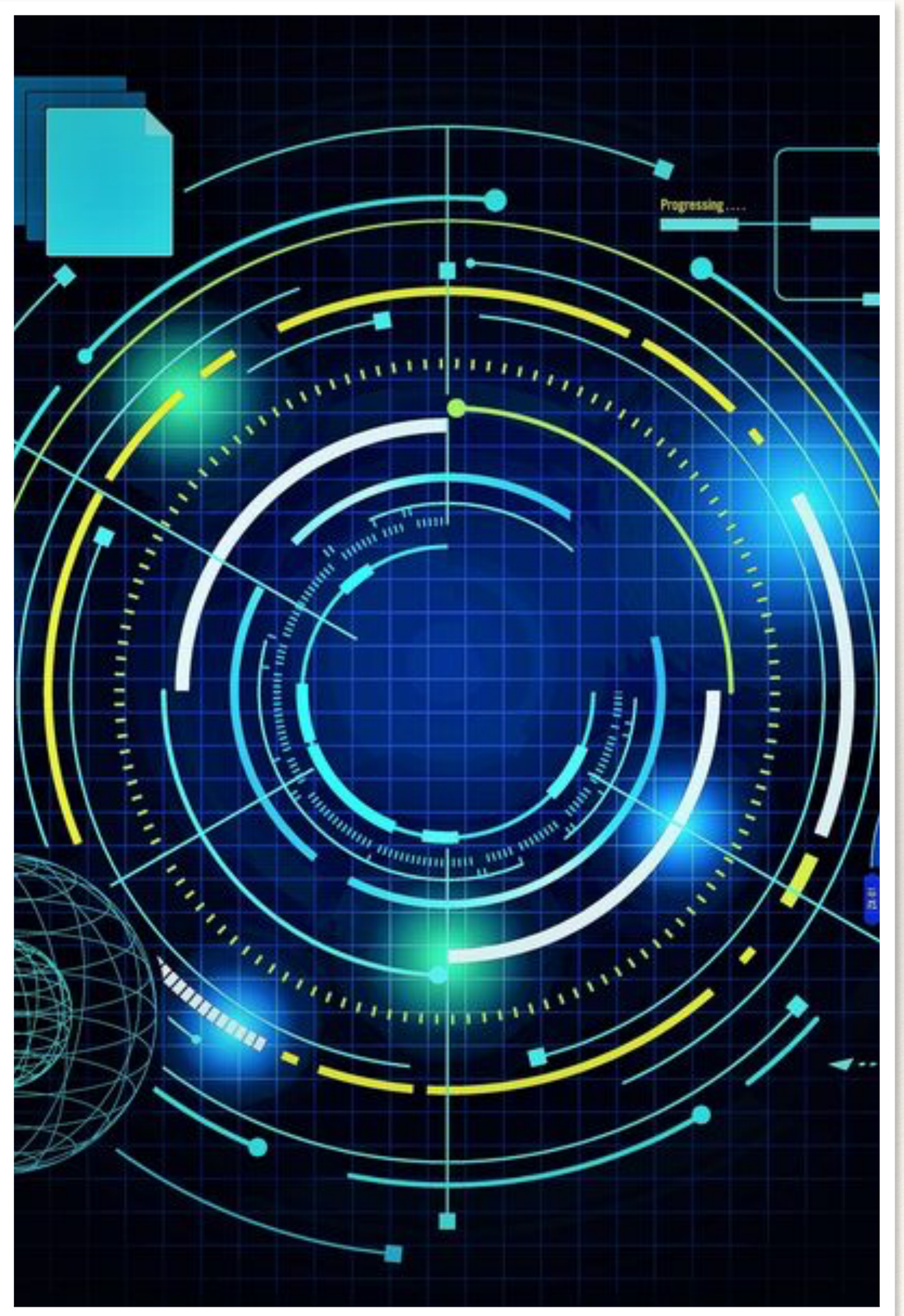
Do machines and people think alike?

Veronica Sanz
Universitat de Valencia - IFIC (Spain)
& Sussex University (UK)

@AI goes MAD '22



Today, we will talk about

Human vs Machine Learning
Learning by example
Imagining new possibilities
What does the AI *really* learn?

My aim is
if you already use ML, make you think a bit differently
if you don't, motivate you to have a closer look

Human vs Machine Learning

Human learning

Previous experience

repeat and improve on a task

Human learning

repeat and improve on a task

predict the evolution of a situation

Previous experience

Human learning

Previous experience

repeat and improve on a task

predict the evolution of a situation

discover unknown relations

Human learning

Previous experience

repeat and improve on a task

predict the evolution of a situation

discover unknown relations

choose the option that maximises return

Human learning

Previous experience

repeat and improve on a task

predict the evolution of a situation

discover unknown relations

choose the option that maximises return

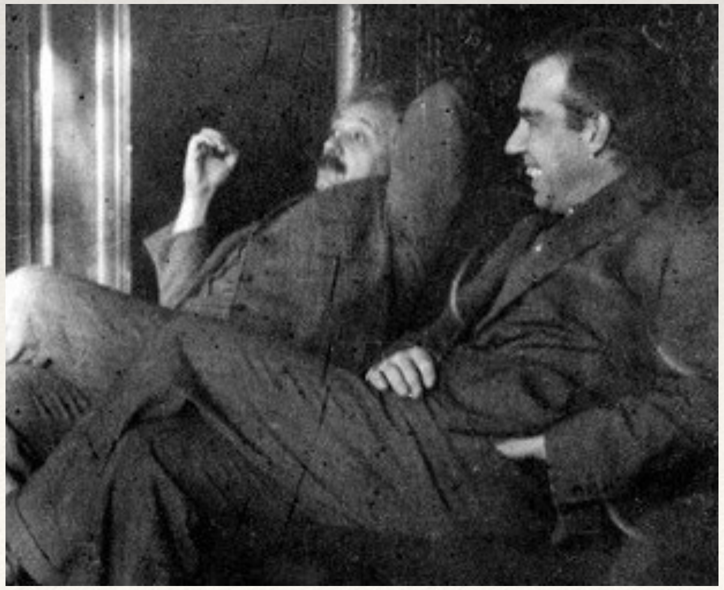
imagine new possibilities

VERY IMPRESSIVE, YET

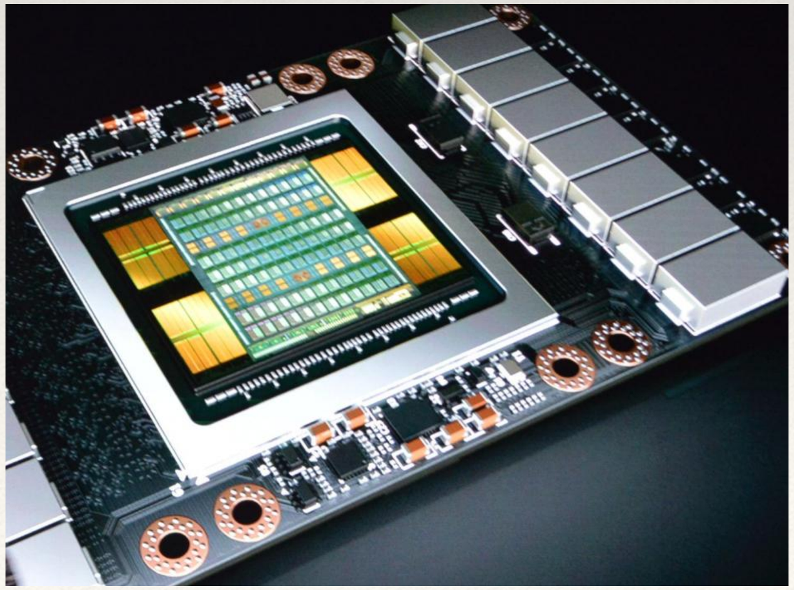
human learning is limited by
our personal viewpoint,
our collective intelligence (*newspeak?)*
& our inherent capacity to process information
(amount , speed, level of detail)

ON THE OTHER HAND

the ultimate limitations of machine learning
are unknown (if they do exist)
CPU-> GPU, TPU, FPGA, IPU -> …
Quantum Computing, Neurophotonics…



VS



Machine learning

Previous experience

repeat and improve on a task

SUPERVISED MACHINE LEARNING

predict the evolution of a situation

RECURRENT LEARNING

discover unknown relations

CLUSTERING/UNSUPERVISED

choose the option that maximises return

REINFORCEMENT LEARNING

imagine new possibilities

GENERATIVE AI

Machine learning

repeat and improve on a task

SUPERVISED MACHINE LEARNING

predict the evolution of a situation

RECURRENT LEARNING

Previous experience

discover unknown relations

CLUSTERING/UNSUPERVISED

choose the option that maximises return

REINFORCEMENT LEARNING

imagine new possibilities

GENERATIVE AI

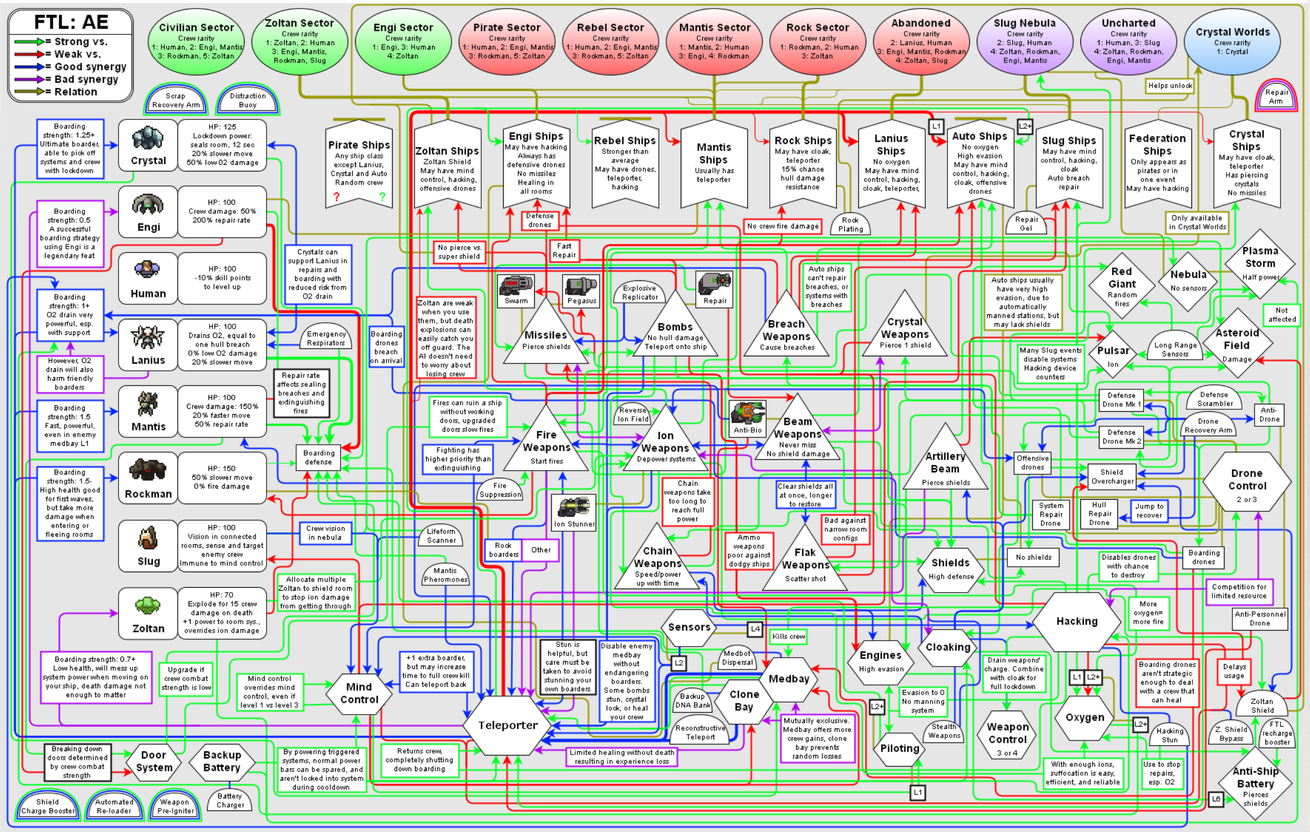
Nowadays, Machine Learning is in the middle of a revolution: processing speed and storing capacity have increased enormously but **more importantly** the *way* machines learn has changed

---

TRADITIONALLY

learning was limited to lines of code we (humans) were writing

```
if something_is_in_the_way is True:
    stop_moving()
else:
    continue_moving()
```



we can write *extremely complex* codes and the machine can improve in performing tasks but the structure of *thought* behind decision making is human

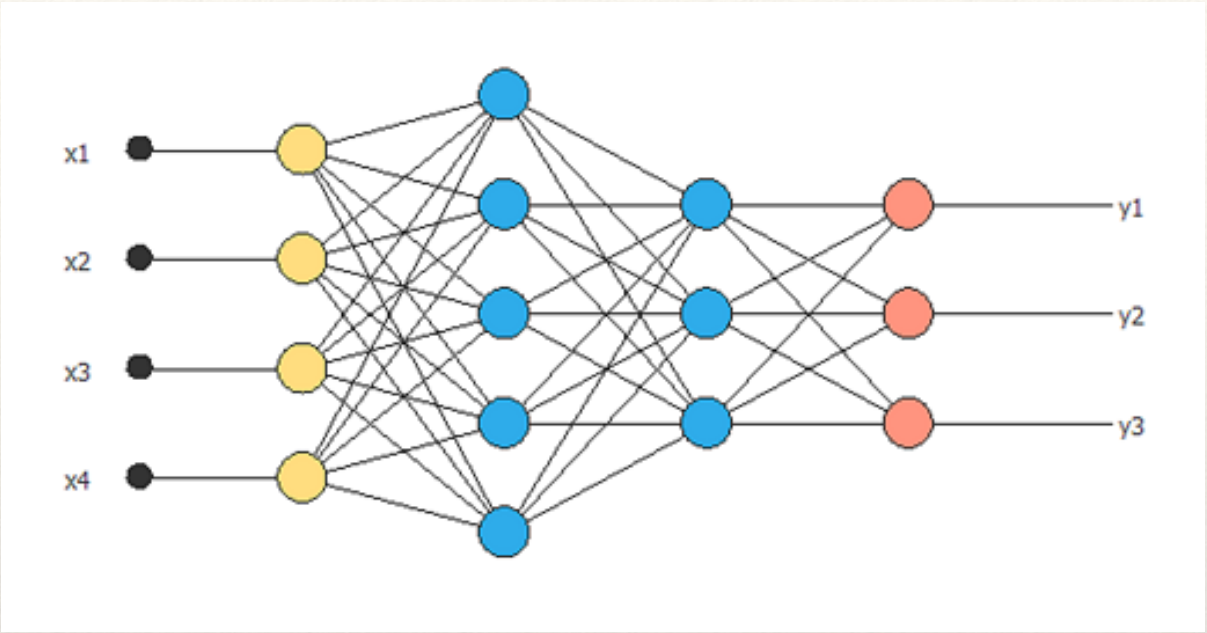
The Machine can't describe relations we haven't coded in

like a born-blind person who is asked to think of ***blue***

A new way of *thinking*: Neural Networks

Structures made of units called *neurons*
and organised by *layers*

inputs



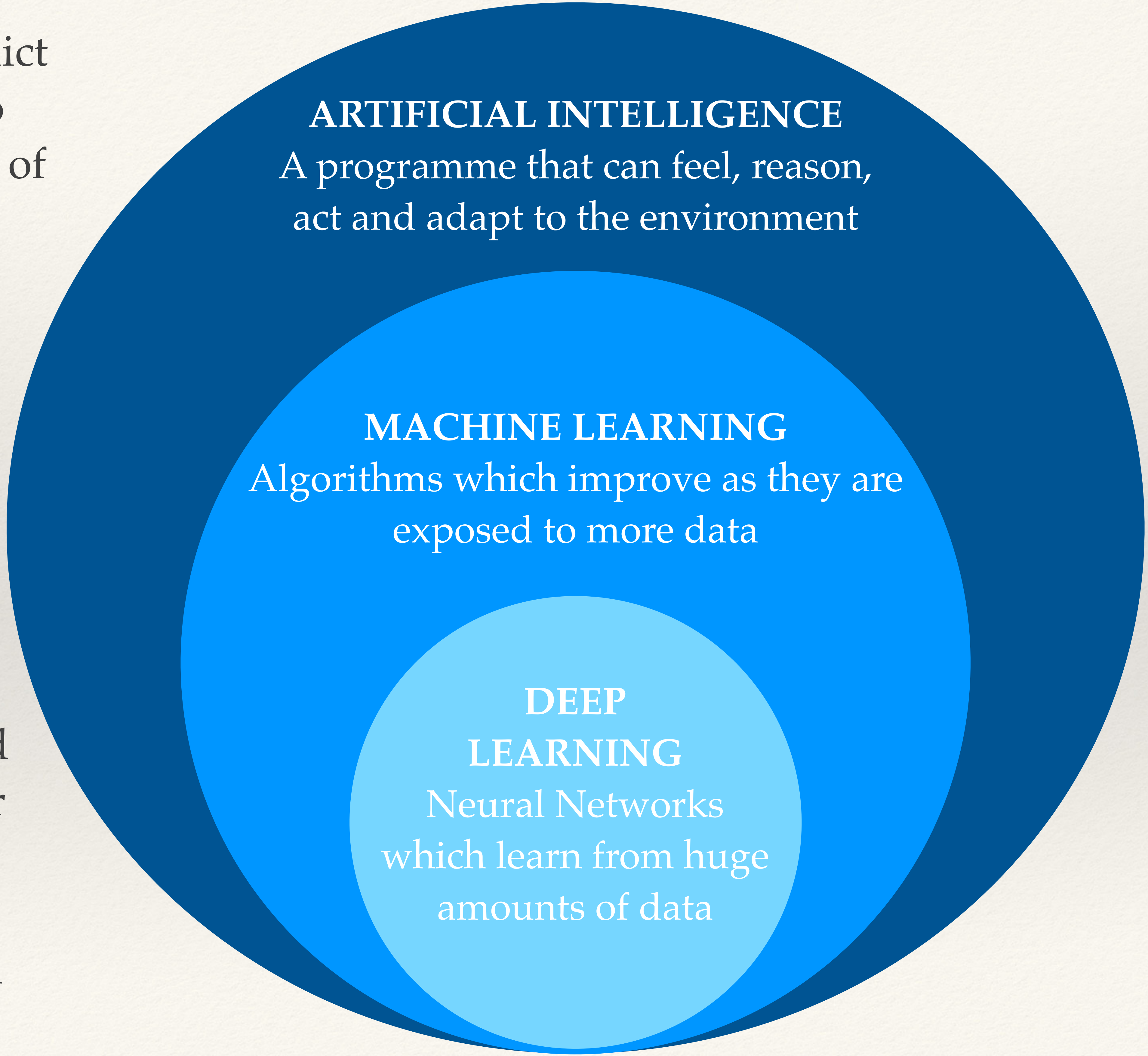
outputs

The network learns from data with no structured instructions

Neural networks are able to explore relations between inputs and outputs which cannot be contained in lines of codes
their degree of expressivity is immense
and it is extremely fast
built from simple units and in a layered architecture

This technology is truly *disruptive*

we are unable to predict how fast is going to evolve and the extent of its applications



new algorithms and applications appear every day, and this tendency does not seem to slow down

Learning by example: Supervised ML

repeat and improve on a task

A basic task: good or bad?

Is it a crocodile?

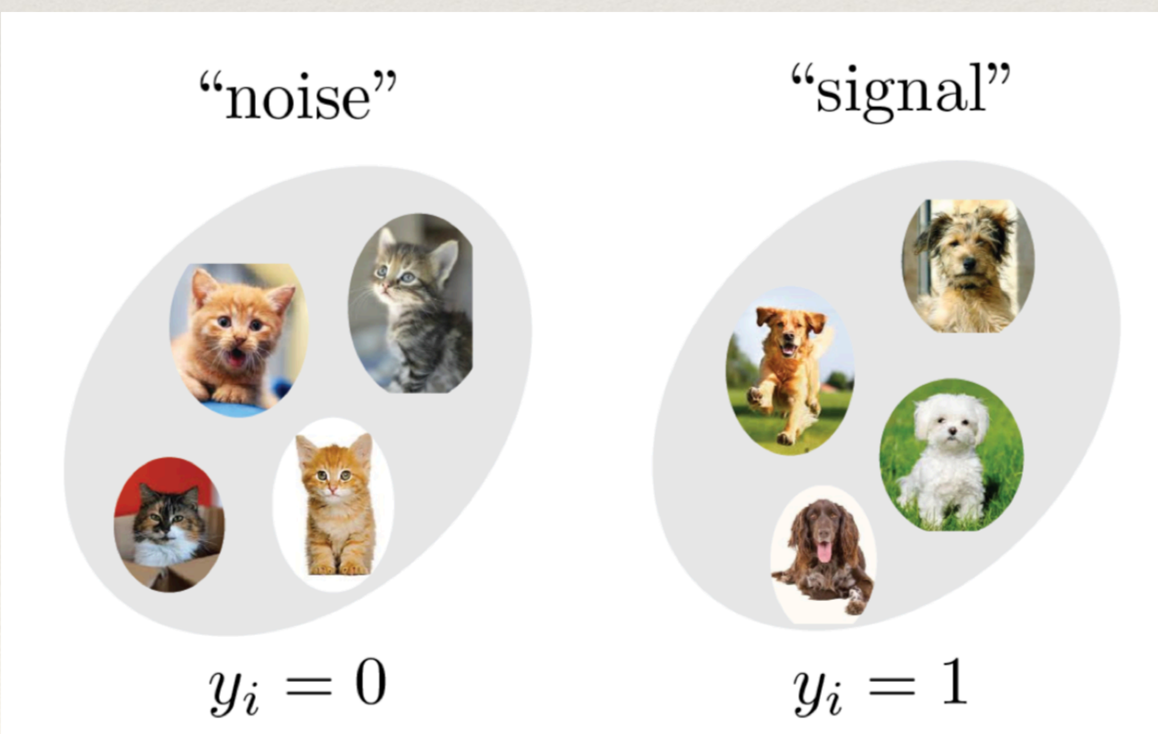
Yes/No answer

A basic task: good or bad?

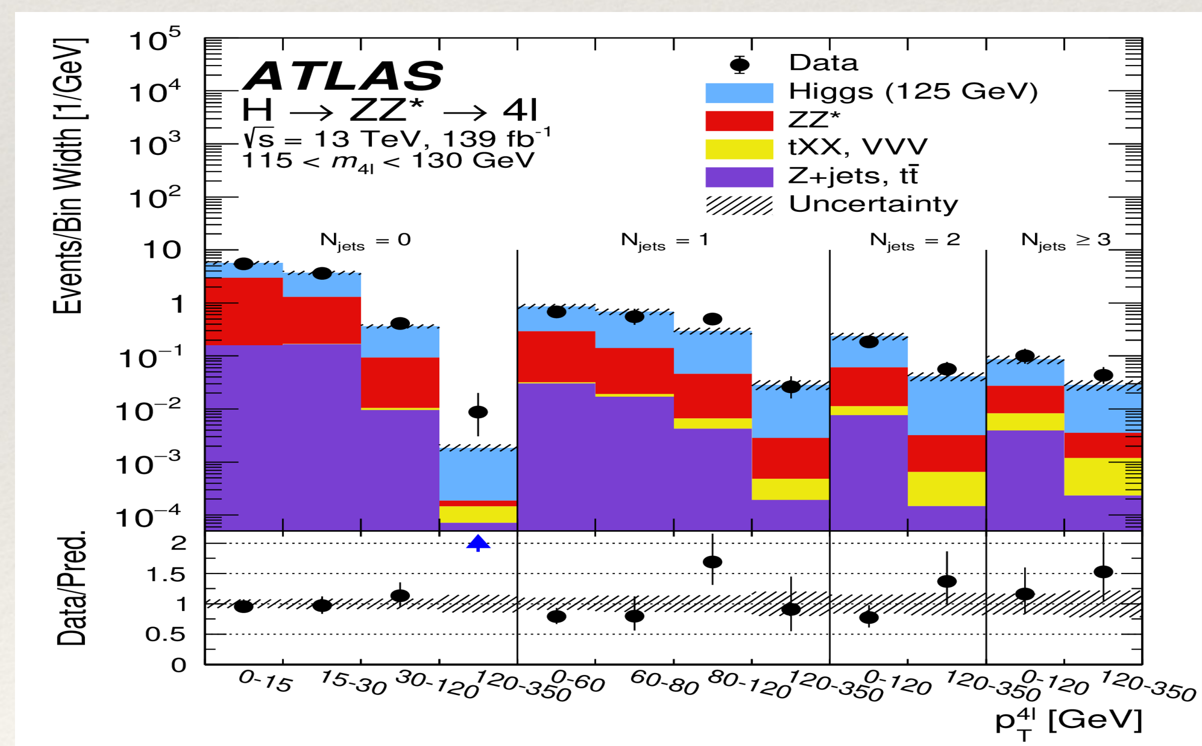
Is it a crocodile?

Yes/No answer

To learn, dataset $\mathcal{D}(x_i, y_i)$ $y \in \{0, 1\}$ with labels



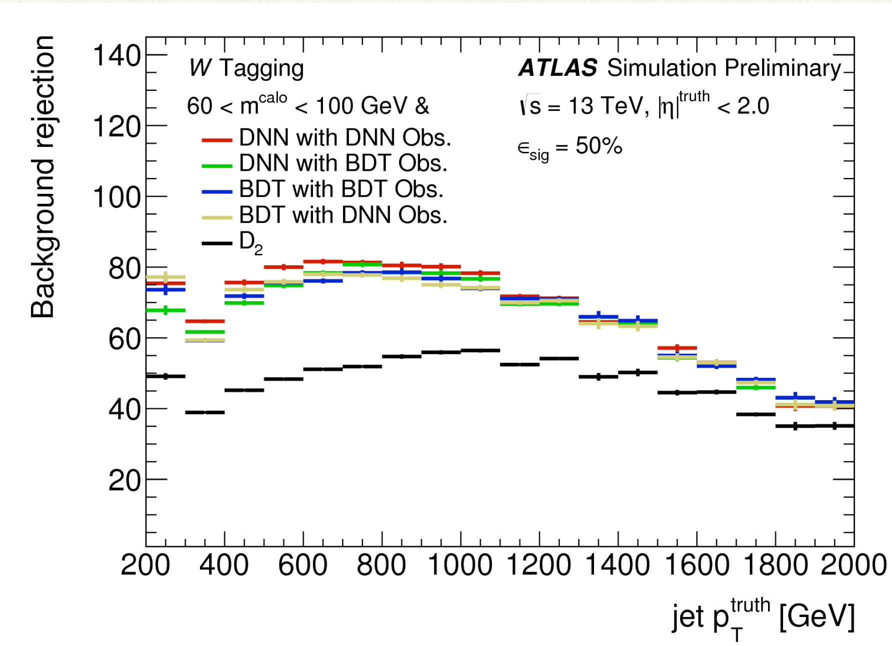
Cat or dog?



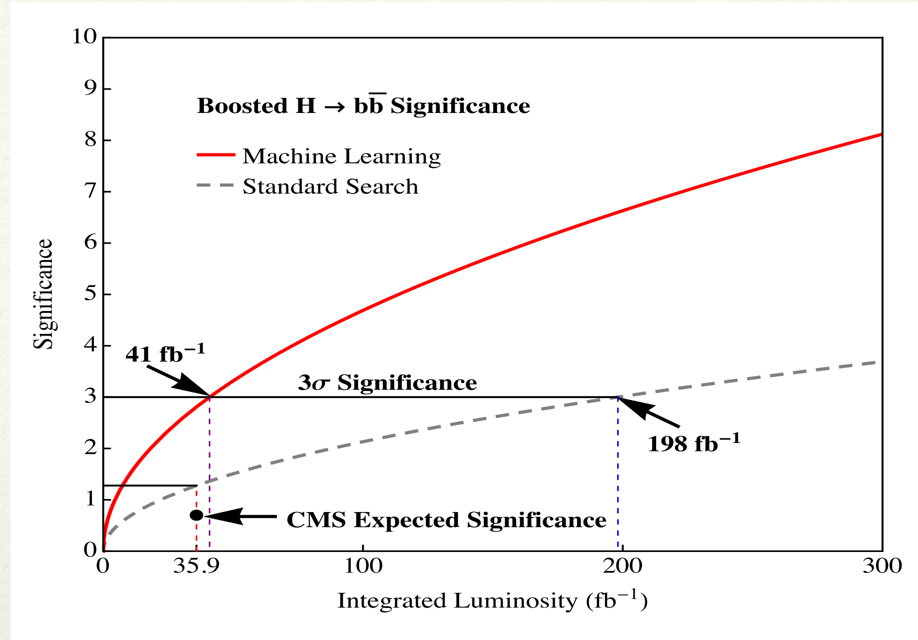
Is this New Physics?

A lot of ML in Particle Physics is answering YES/NO questions

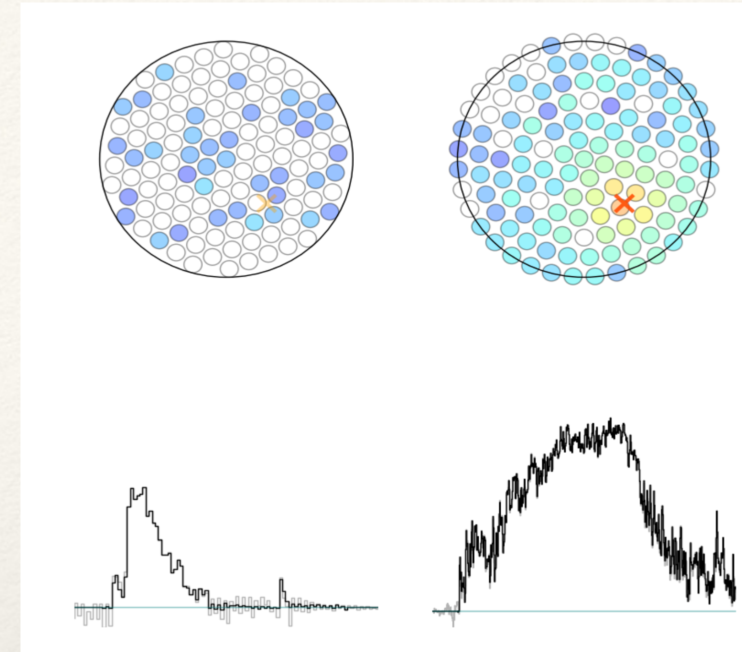
Is it a W?



Is it a Higgs?



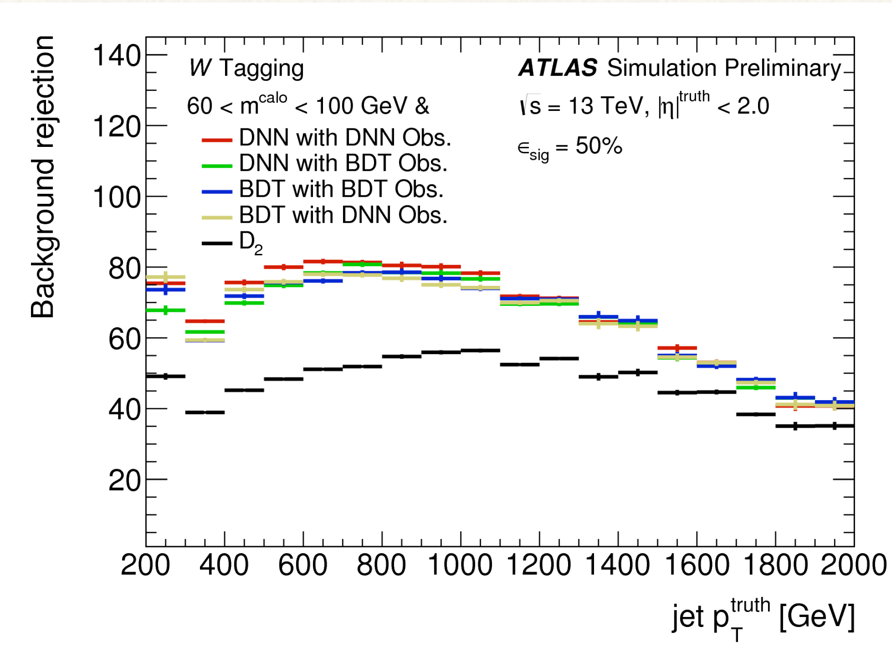
Is it DM?



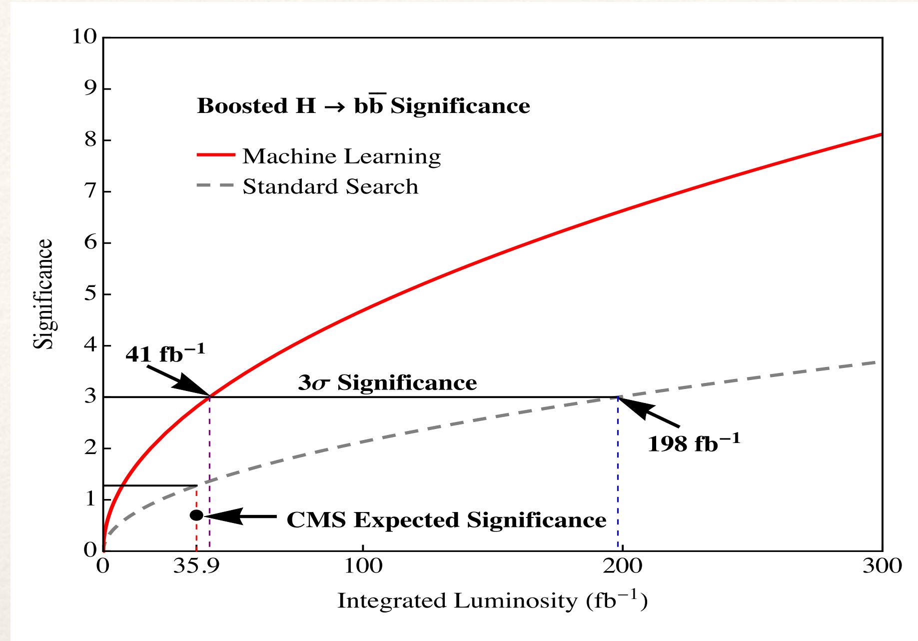
mostly using Neural Networks to deal with images (**CNNs**)

A lot of ML in Particle Physics is answering YES/NO questions

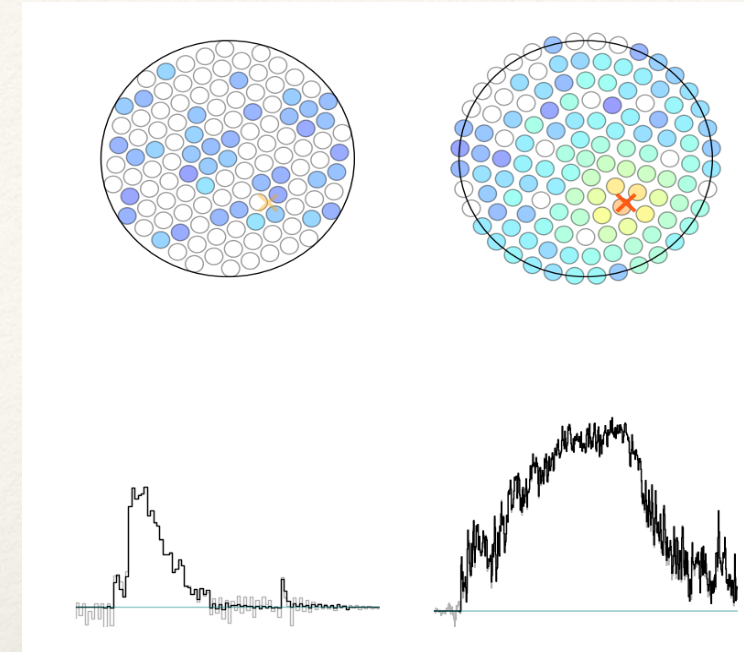
Is it a W?



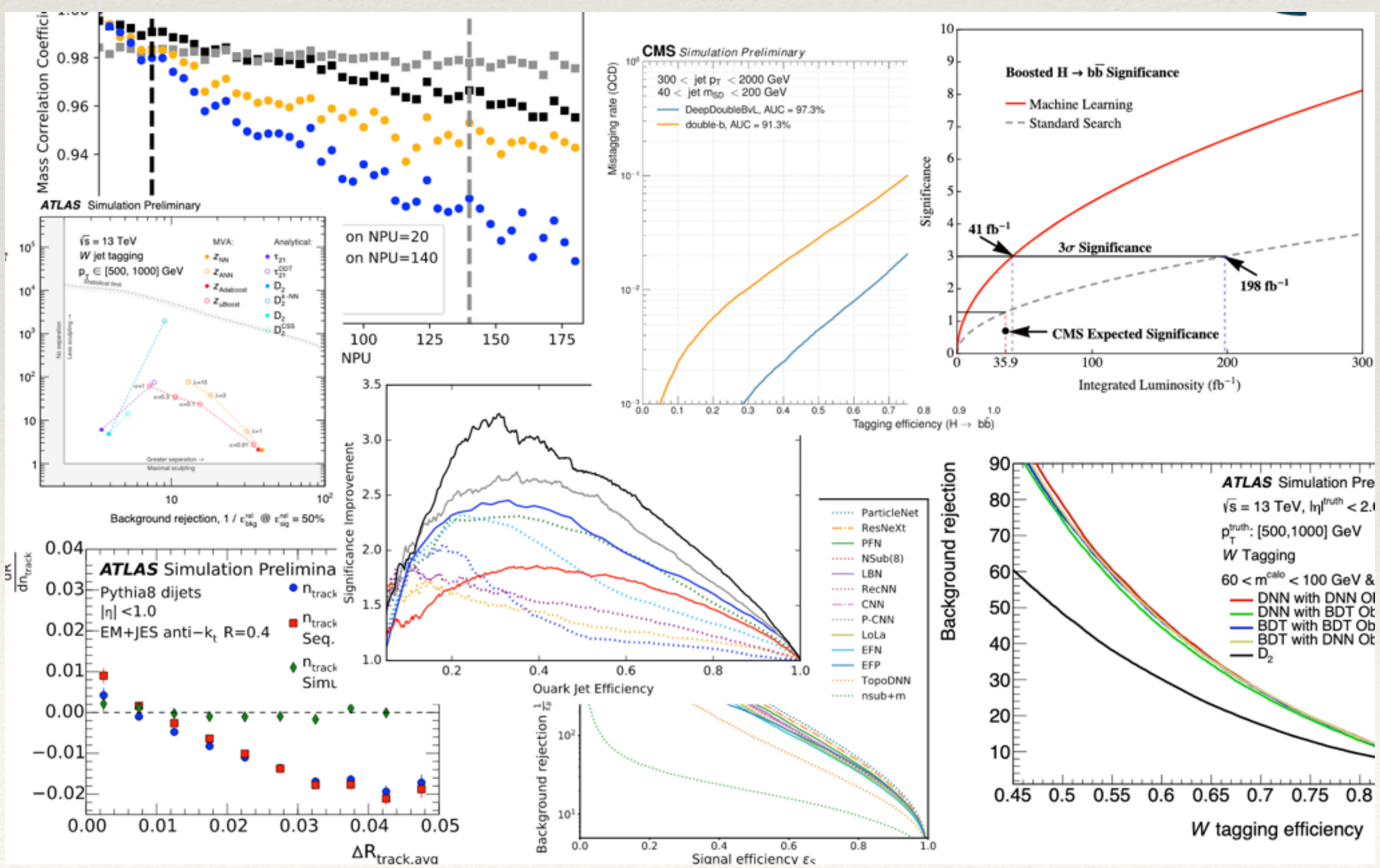
Is it a Higgs?



Is it DM?



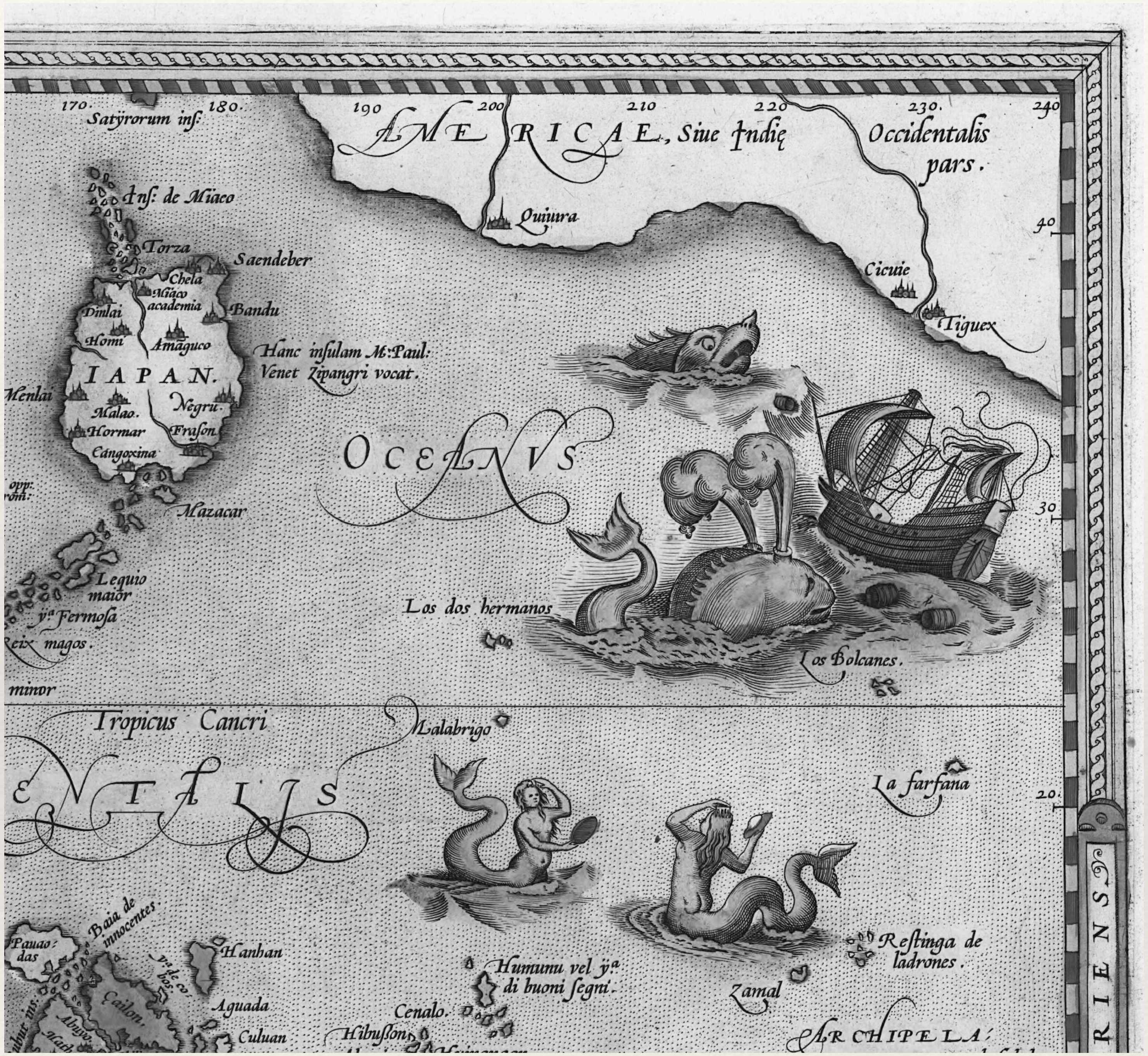
mostly using Neural Networks to deal with images (**CNNs**)



The gains in ID-ing phenomena are typically in the range of 5%-30%

for tricky environments: difference between discovery or not

intellectually, not super-exciting



Here be dragons!

Going further

imagine new possibilities

What if we didn't ask for an outcome?

Supervised learning input-> predict output

what if we just asked 'look at this!' with no determined output?

GANs (Generative Adversarial Networks)

and **VAEs (Variational AutoEncoders)**

In CNNs, benchmarks were cats/dogs and hand-written digits (MNIST)

Here, human faces

What if we didn’t ask for an outcome?

Supervised learning input-> predict output

what if we just asked ‘look at this!’ with no determined output?

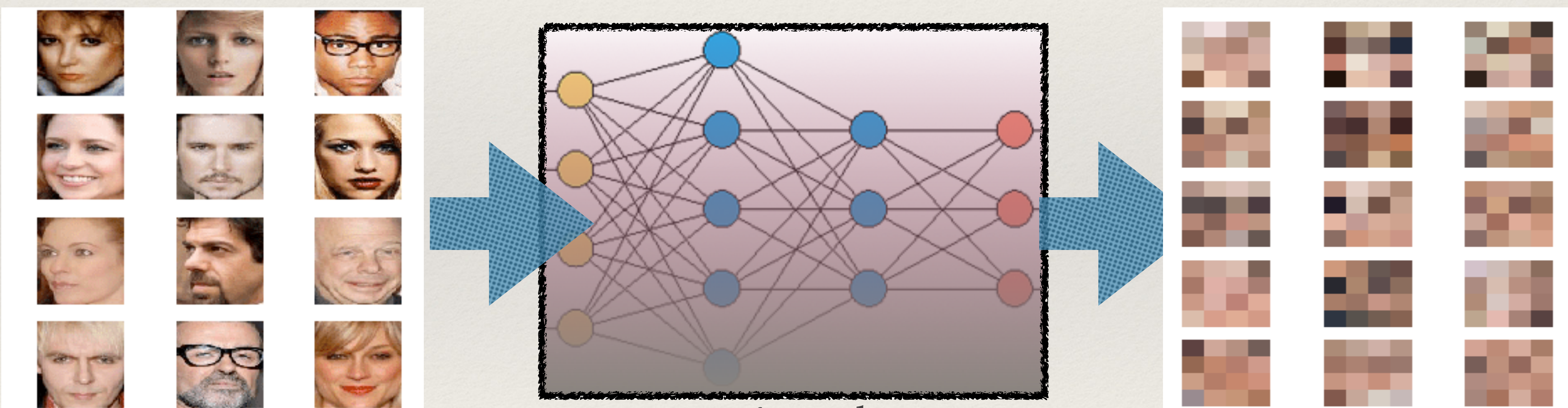
GANs (Generative Adversarial Networks)

and **VAEs (Variational AutoEncoders)**

In CNNs, benchmarks were cats/dogs and hand-written digits (MNIST)

Here, human faces

STEP 1 - ‘LEARN’ what is a human face



Take face images: x

Transform them in complicated ways

Create an avatar: x’

Doing this many times, while the DISCRIMINATOR says:

‘You are going in the right direction’, ‘You are completely lost!’

What if we didn’t ask for an outcome?

Supervised learning input-> predict output

what if we just asked ‘look at this!’ with no determined output?

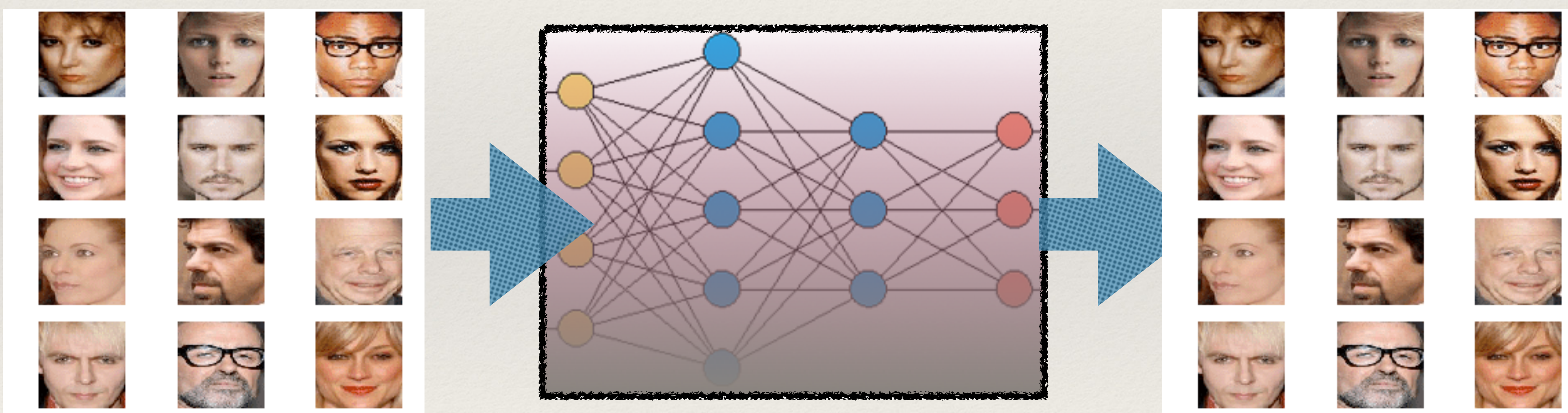
GANs (Generative Adversarial Networks)

and **VAEs (Variational AutoEncoders)**

In CNNs, benchmarks were cats/dogs and hand-written digits (MNIST)

Here, human faces

STEP 2 - AFTER MANY ITERATIONS…



When the avatars are indistinguishable to the DISCRIMINATOR, **game is over**

What if we didn’t ask for an outcome?

Supervised learning input-> predict output

what if we just asked ‘look at this!’ with no determined output?

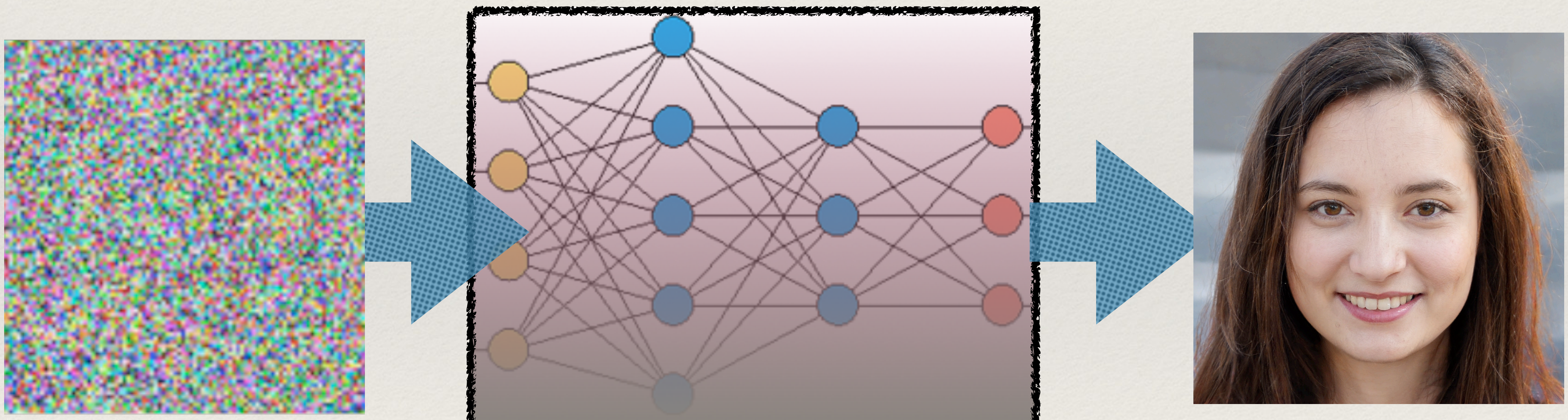
GANs (Generative Adversarial Networks)

and **VAEs (Variational AutoEncoders)**

In CNNs, benchmarks were cats/dogs and hand-written digits (MNIST)

Here, human faces

STEP 3 - CREATE NEW POSSIBILITIES

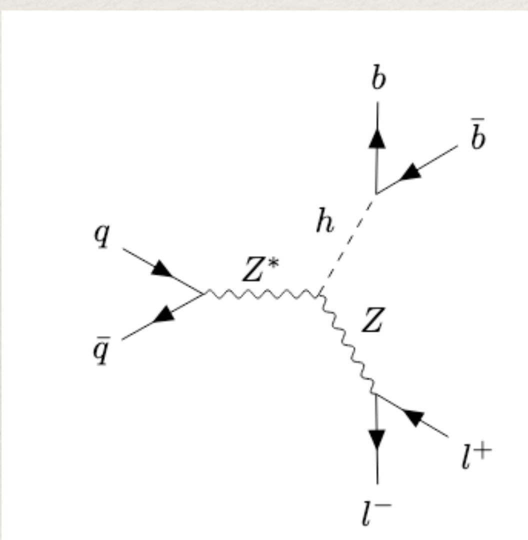


This woman does not exist. It has been generated from noise.
The NN has learnt the concept of ‘human face’ and now can create human faces from noise

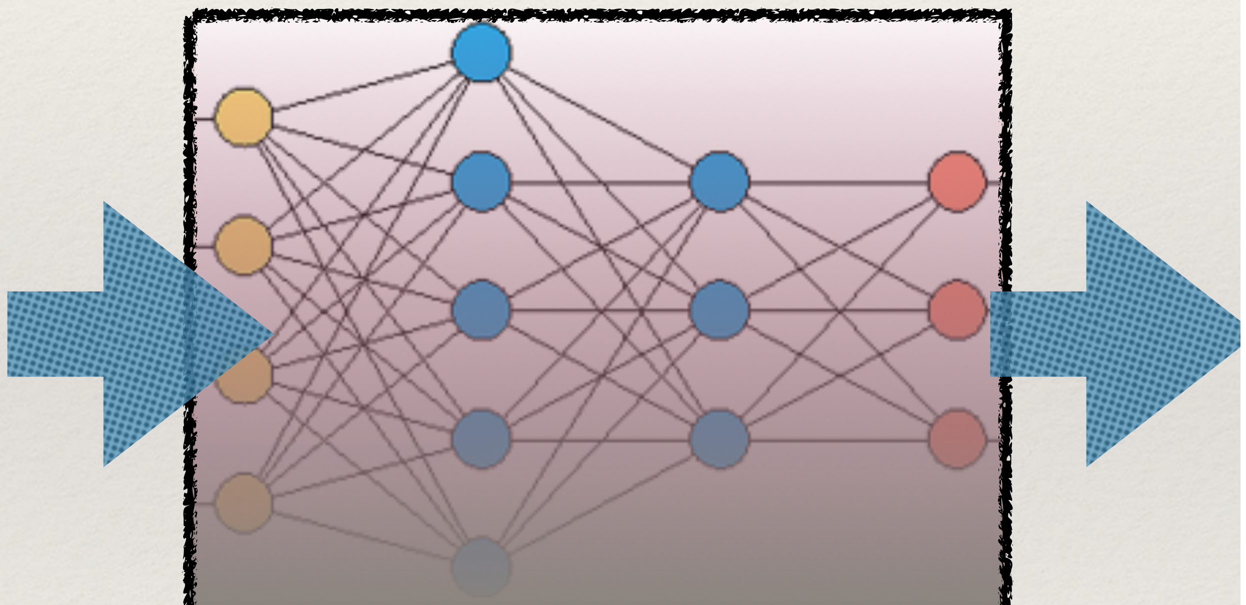
What if we didn't ask for an outcome?

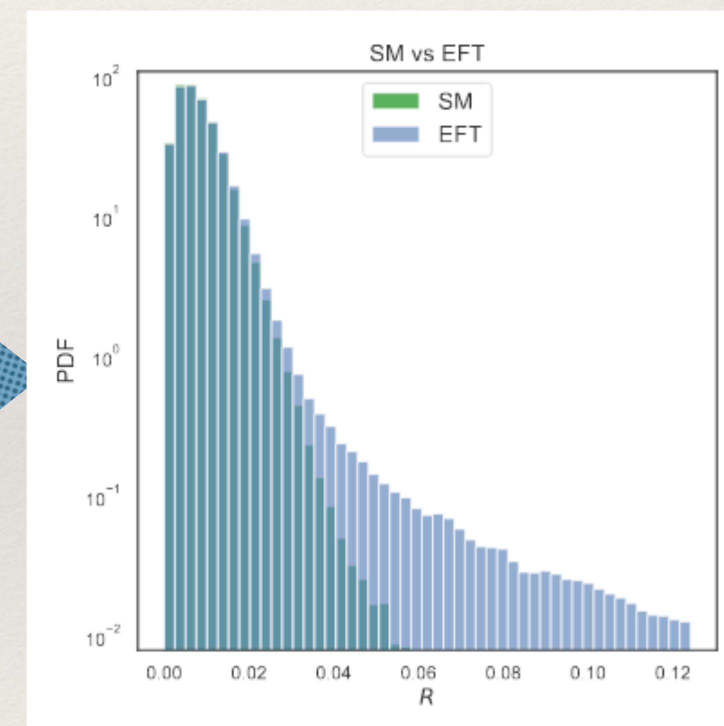
Application I: Once it has learnt a type of phenomena, it will reconstruct well any new similar phenomena
This can be used as a way to detect unknown anomalies

EXAMPLE - ANOMALY DETECTION *with Khosa and Soughton* ***2203.03669***



Ask to look only to Standard Model ('normal') events





Learns to ID outliers ('New Physics')

What if we didn't ask for an outcome?

Application II : Generative AI can be used to produce new situations
To cover the parameter space of possibilities e.g. faces consistent with the laws it has learnt

EXAMPLE - ECOLOGICAL INTERACTIONS

with Ecology experts ***Methods in Ecology and Evolution (2022)***

The landscape where I live is semi-desertic
Among plant species, competition for resources is fierce, and co-existence rules are complex
In our Physics language, higher-order interactions are important

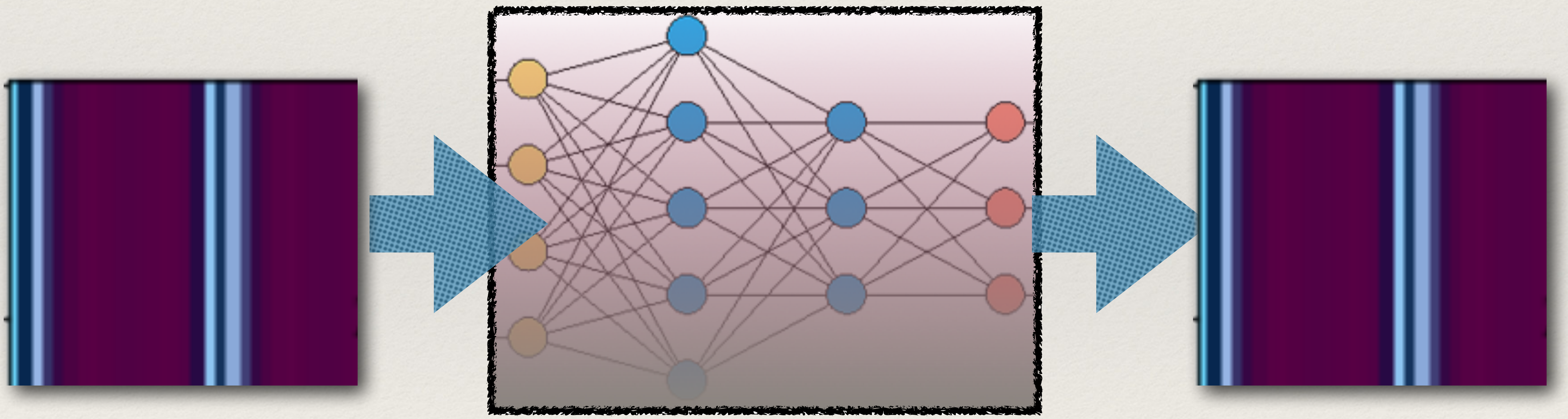
From people at the Research Centre for Desertification, we looked at such an eco-system and use GenAI to guide re-population efforts

What if we didn't ask for an outcome?

Application II : Generative AI can be used to produce new situations
To cover the parameter space of possibilities e.g. faces consistent with the laws it has learnt

EXAMPLE - ECOLOGICAL INTERACTIONS

*with Ecology experts **Methods in Ecology and Evolution (2022)***



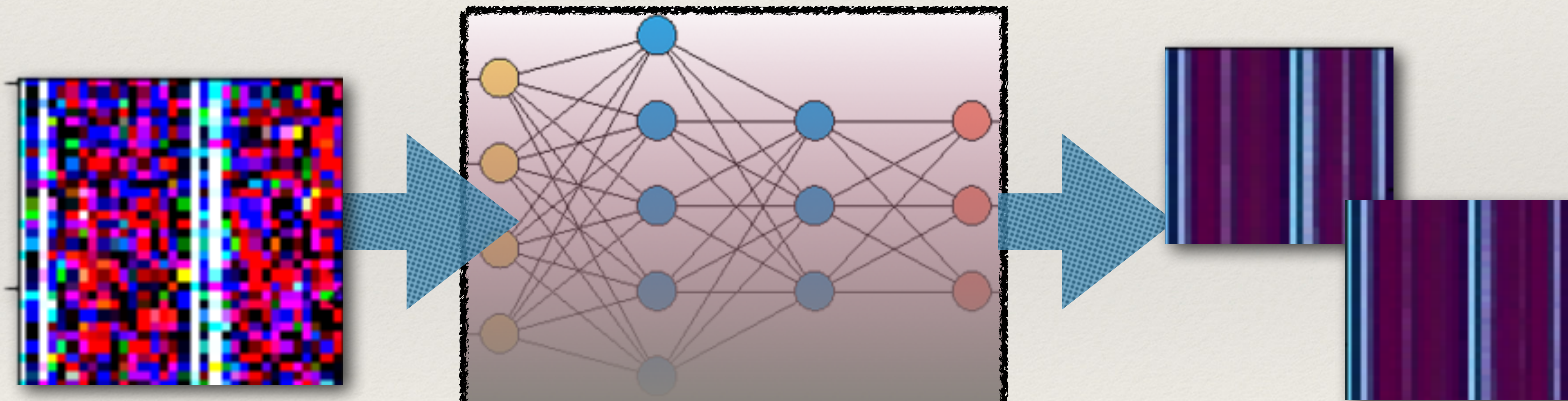
We fed a VAE with the examples of species co-existence until reaching good accuracy

What if we didn't ask for an outcome?

Application II : Generative AI can be used to produce new situations To cover the parameter space of possibilities e.g. faces consistent with the laws it has learnt

EXAMPLE - ECOLOGICAL INTERACTIONS

with Ecology experts ***Methods in Ecology and Evolution (2022)***

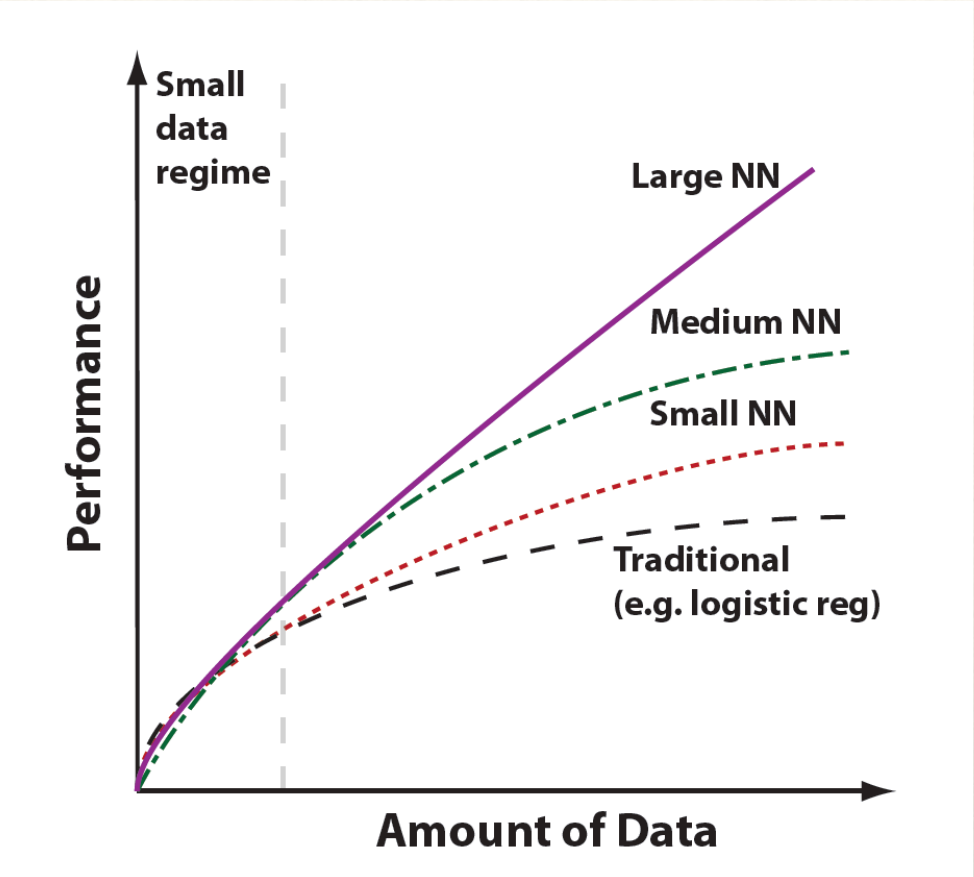


Once trained, we could ask lots of non-trivial questions, e.g. given a patch with species X, what are the most varied and compatible patches

Where is the bunny?

What is the AI *really* learning?

Why are NNs so good at learning?



High-bias low-variance, 1803.08823

Good at handling large amounts of data: needle in a haystack

The NN structure (layers, 0/1 gates) allows a high representation power with moderate computational demands, e.g. allows parallelisation, use of GPUs…

It scales better than other learning methods (like SVMs)

Good at learning: ability to learn with little *domain knowledge*

That's something physicists (as humans) are good at

(Physics -> other things)

DNNs are good at this too, they are able to take large streams of data and learn features with little guidance, work like *black boxes*

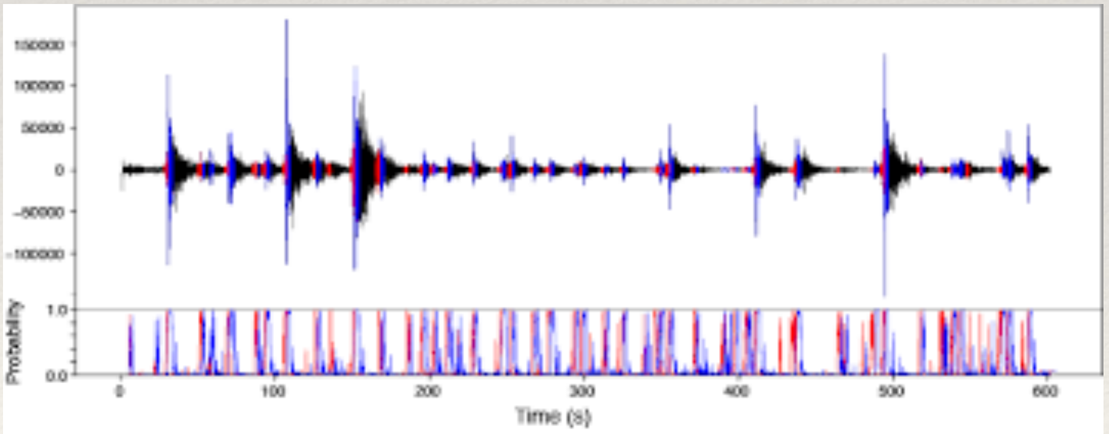
What's wrong with blackboxes?

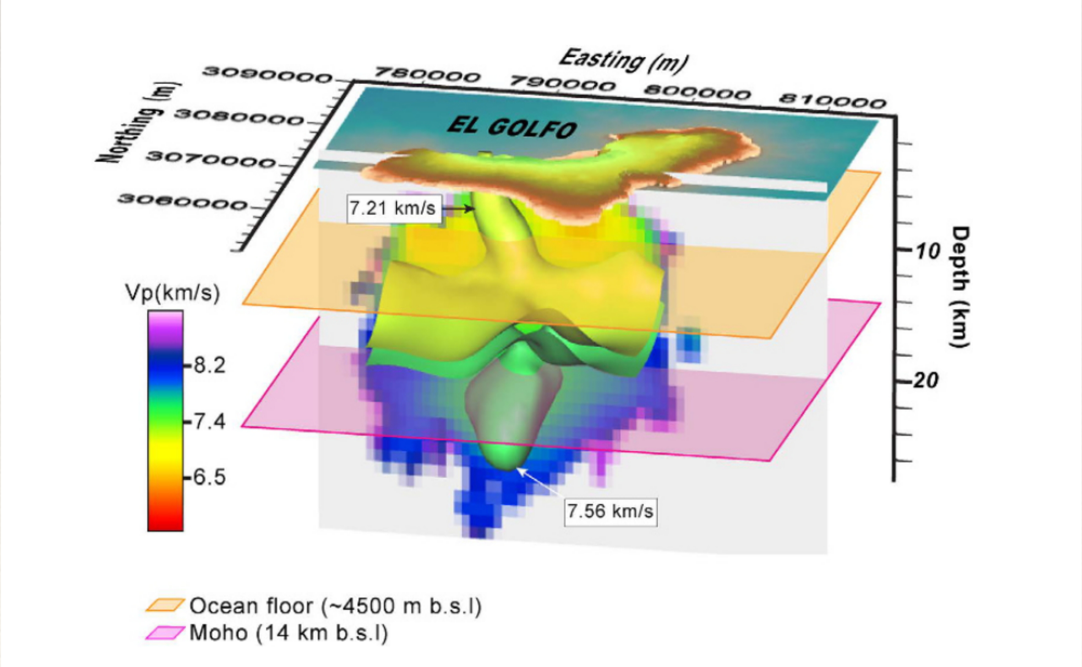
Only open if a disaster happened

If it works, why fix it?
DNN is very powerful, in a way that can be quantified and tensioned against human performance or other techniques

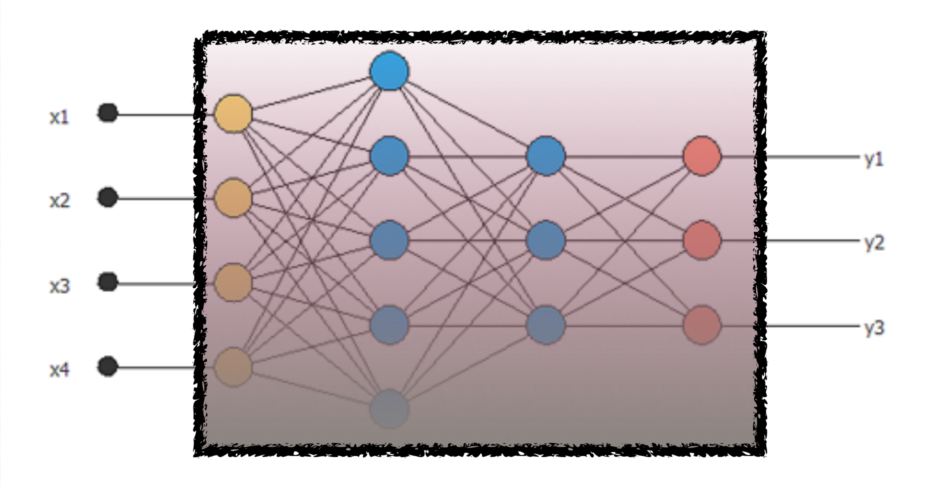
EXAMPLE - AUTOMATIC DETECTION OF SEISMICITY

Tomography





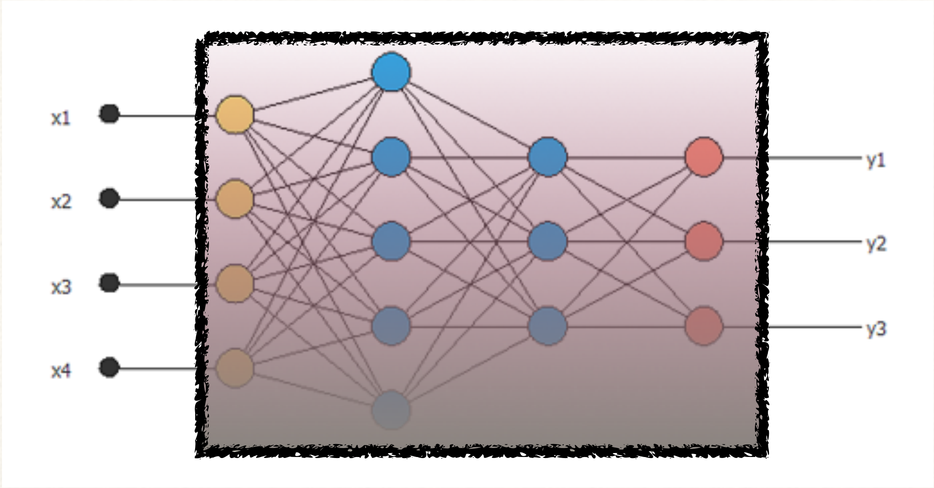
with Seismology experts
Seismological Research Letters (2022)



What's wrong with blackboxes?

If they do work, and help solve problems?

The lack of understanding hurts our pride as scientists
our job is to understand as much as we humanly can
"If you think you understand quantum mechanics, you don't understand quantum mechanics" R. Feynman, *The Character of Physical Law*

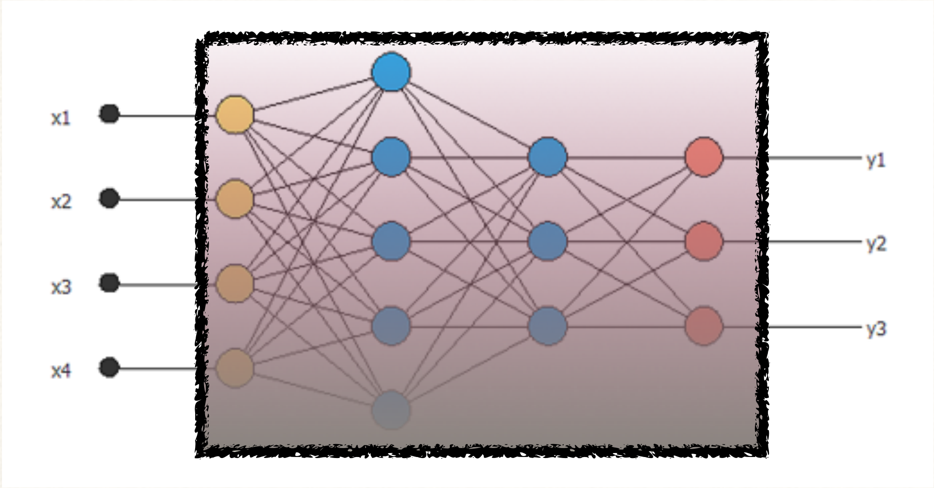


What's wrong with blackboxes?

If they do work, and help solve problems?

The lack of understanding hurts our pride as scientists
our job is to understand as much as we humanly can
"If you think you understand quantum mechanics, you don't understand quantum mechanics" R. Feynman, *The Character of Physical Law*

Any efforts we do to express the workings of NNs from different viewpoints may lead to *new ideas for machine learning*



What's wrong with blackboxes?

If they do work, and help solve problems?

The lack of understanding hurts our pride as scientists
our job is to understand as much as we humanly can
"If you think you understand quantum mechanics, you don't understand quantum mechanics" R. Feynman, *The Character of Physical Law*

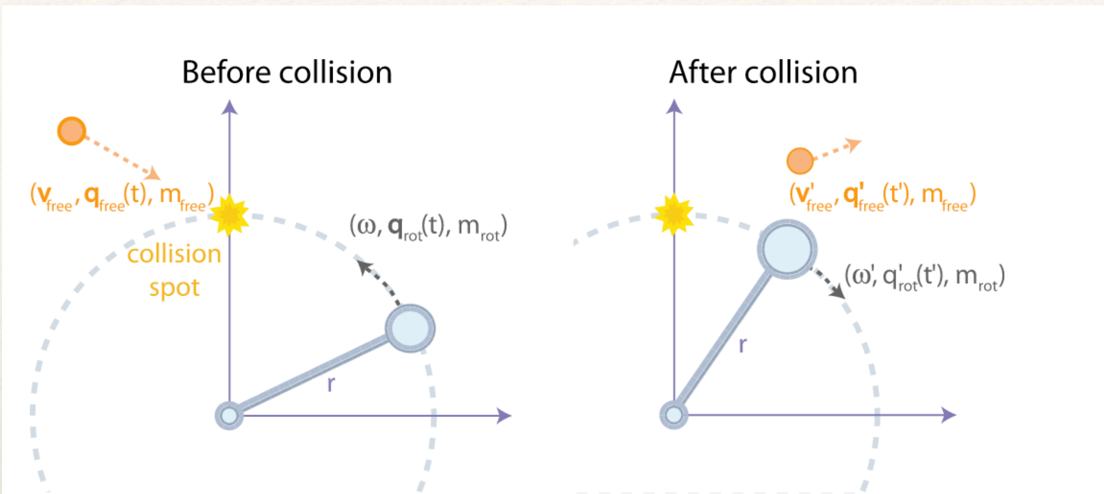
Any efforts we do to express the workings of NNs from different viewpoints may lead to *new ideas for machine learning*

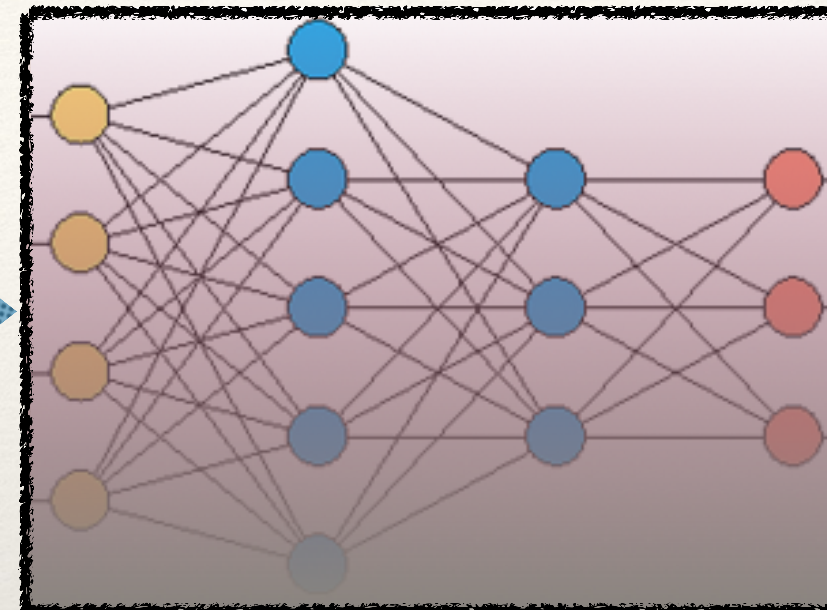
The depth and reach of AI in *decision making* is growing very fast
we should be concerned about our lack of control over this
e.g. see EU's draft on regulating AI, April 21st
XAI, Ethical AI… all these require a better understanding of DNNs

NNs can learn broad concepts, but how?

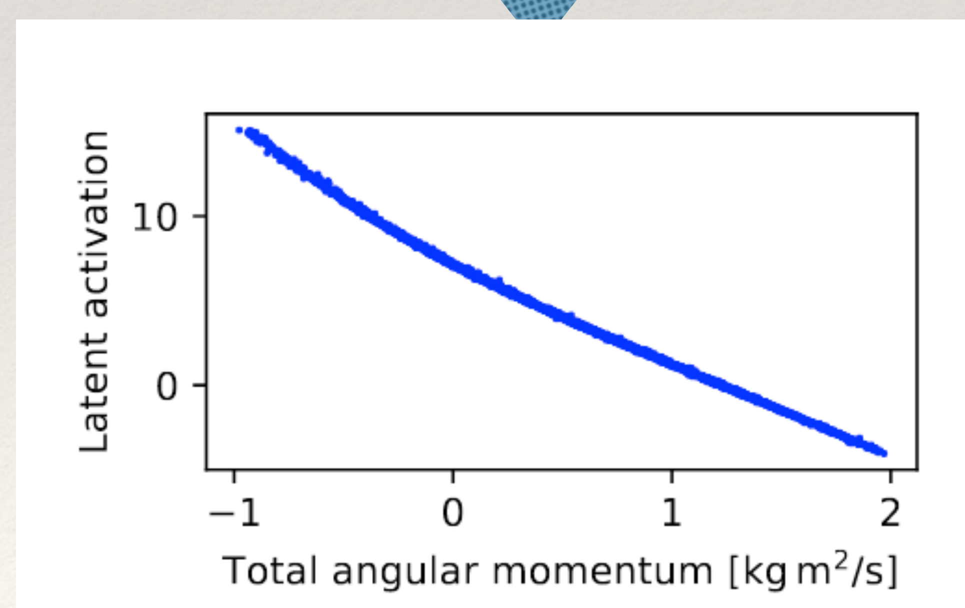
EXAMPLE- CONSERVATION LAWS

Iten et al *PRL*





Trained a VAE with many examples of collisions, no mention of concepts like total angular momentum



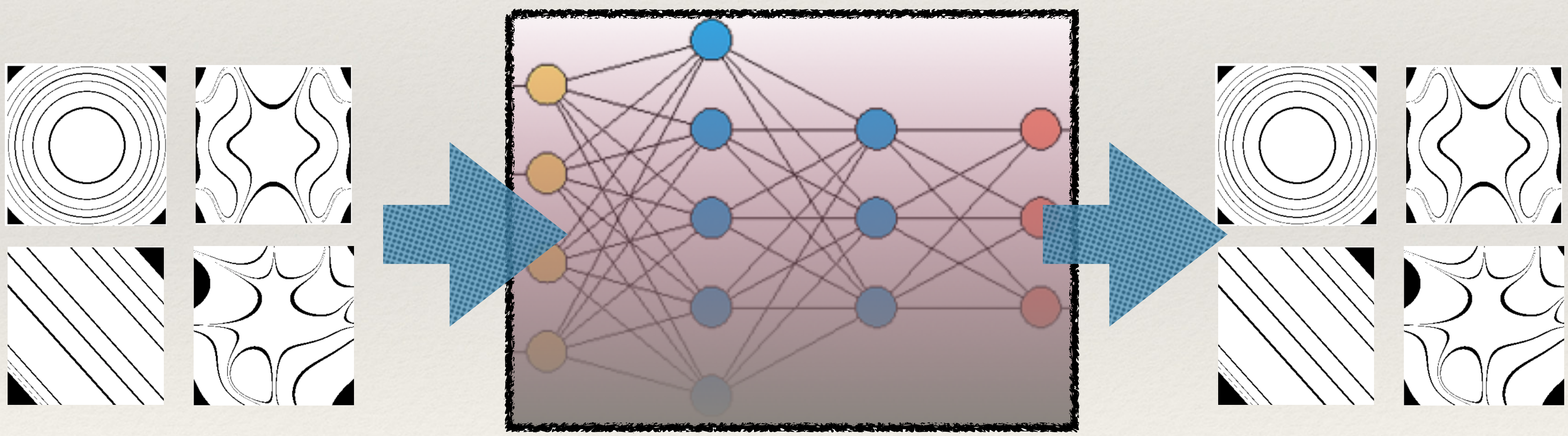
After training,
NNs were storing somehow information of the angular momentum
The size of the latent activation was related to total angular momentum

NNs can learn broad concepts, but how?

We (humans) all know what *is* a human face
but we wouldn't be able to write code to teach a machine
to transform noise in a perfectly realistic face

If we train NNs on physical situations,
could we interrogate the machine and learn what is doing?

EXAMPLE - SYMMETRIES *Symmetry meets AI, **SciPost Physics***



Images of physics potentials

Did the black-box *realise* there is something called Symmetry?

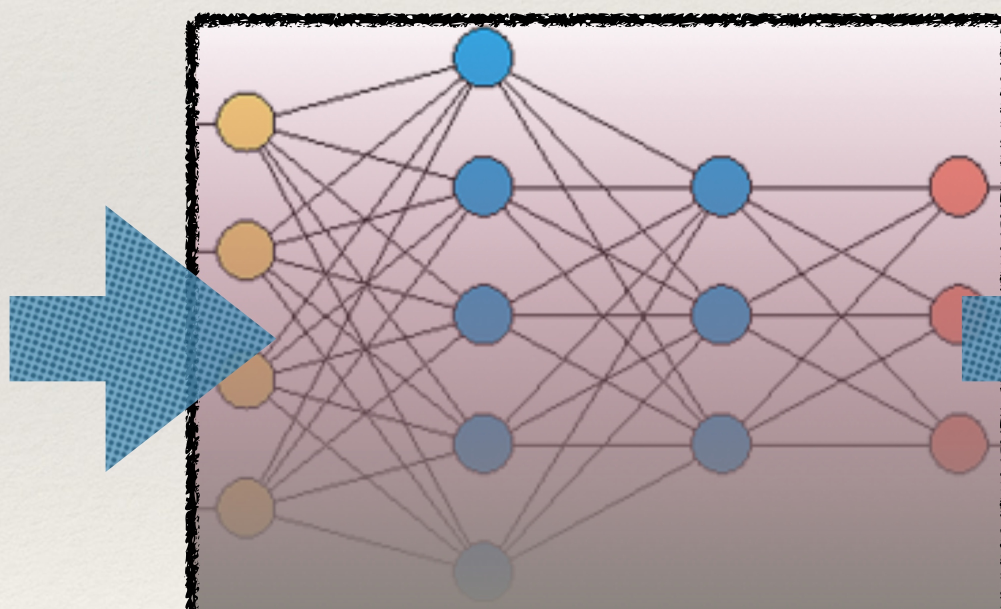
Images of physics potentials

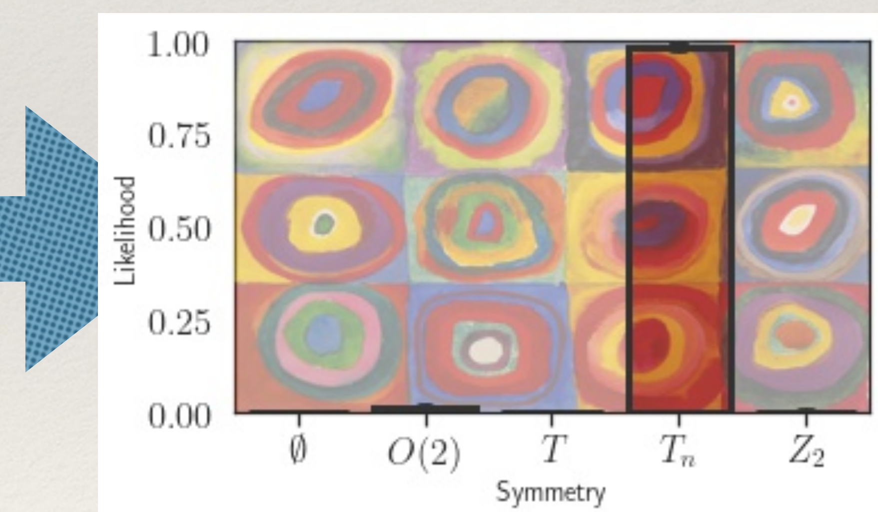
NNs can learn broad concepts, but how?

We (humans) all know what *is* a human face
but we wouldn't be able to write code to teach a machine
to transform noise in a perfectly realistic face

If we train NNs on physical situations,
could we interrogate the machine and learn what is doing?

EXAMPLE - SYMMETRIES *Symmetry meets AI*, ***SciPost Physics***





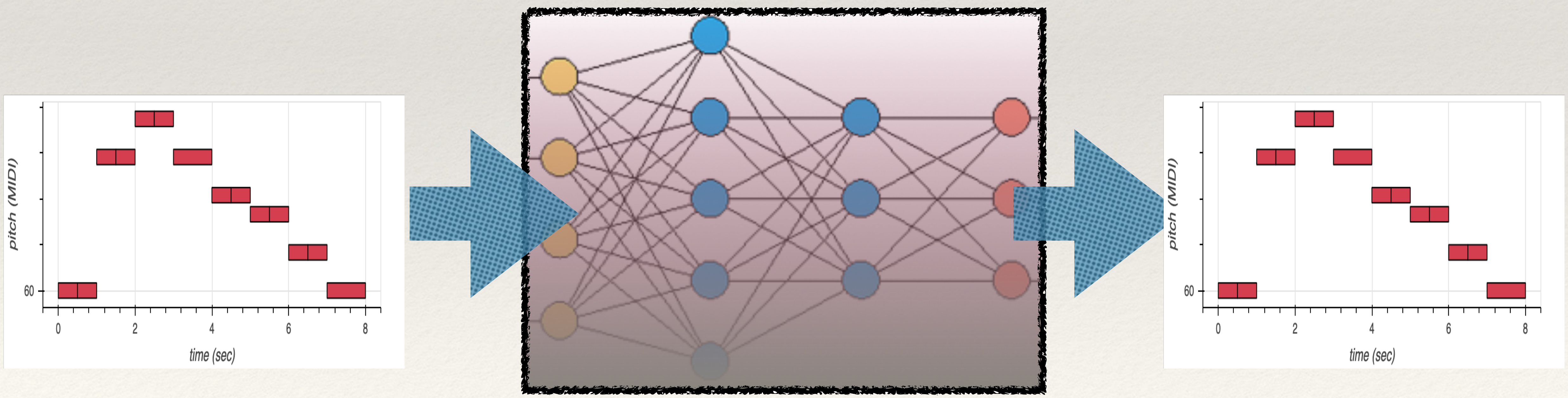
YES, it did realise and we used it to build a SYMMETRY SCORE

NNs can learn broad concepts, but how?

Can we go even further from Physics? What about music?
Does the AI *realise* human concepts?

EXAMPLE - MUSIC *in preparation with Barenboim, Hirn and del Debbio*

We use an open-source VAE from Google's project MAGENTA trained on millions of musical pieces, with the aim to generate new musical pieces, even choosing the style



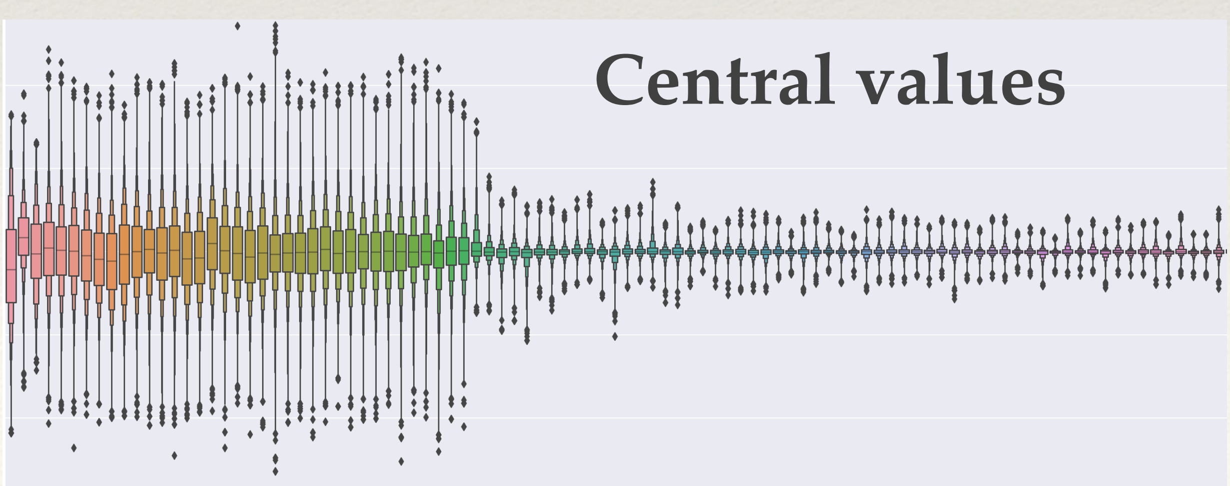
NNs can learn broad concepts, but how?

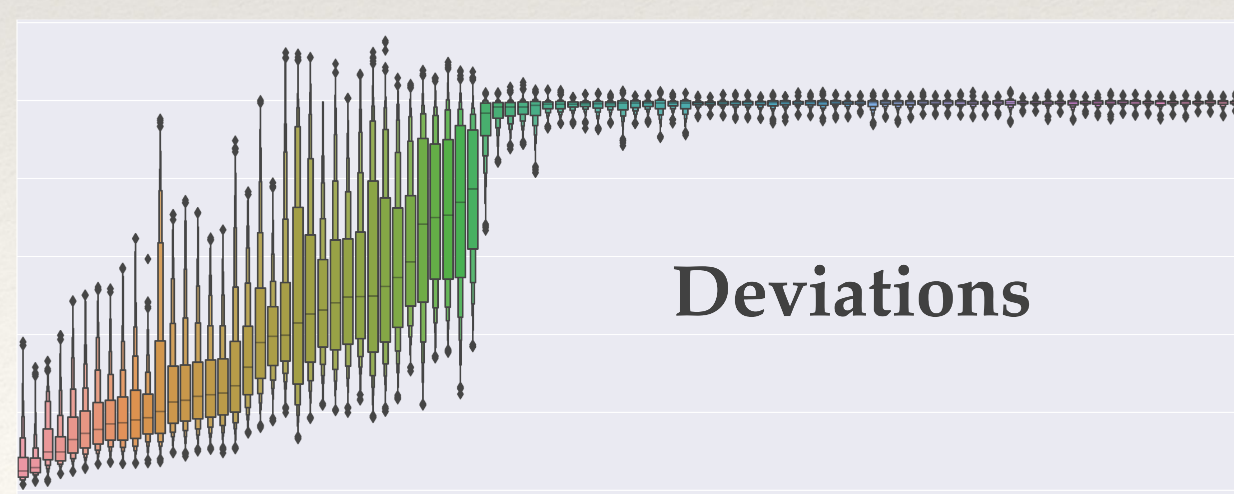
Can we go even further from Physics? What about music?
Does the AI *realise* human concepts?

EXAMPLE - MUSIC *in preparation with Barenboim, Hirn and del Debbio*

The architecture is ginormous, with a latent space of 500 neurons
Did MAGENTA's VAE learn something about the music it was analysing? how do we ask questions?

We discovered the AI is actually **not** mobilising this huge space
most neurons are just noise, waiting to generate diverse new music
only a handful neurons are meaningful, do they carry human information?





NNs can learn broad concepts, but how?

Can we go even further from Physics? What about music?
Does the AI *realise* human concepts?

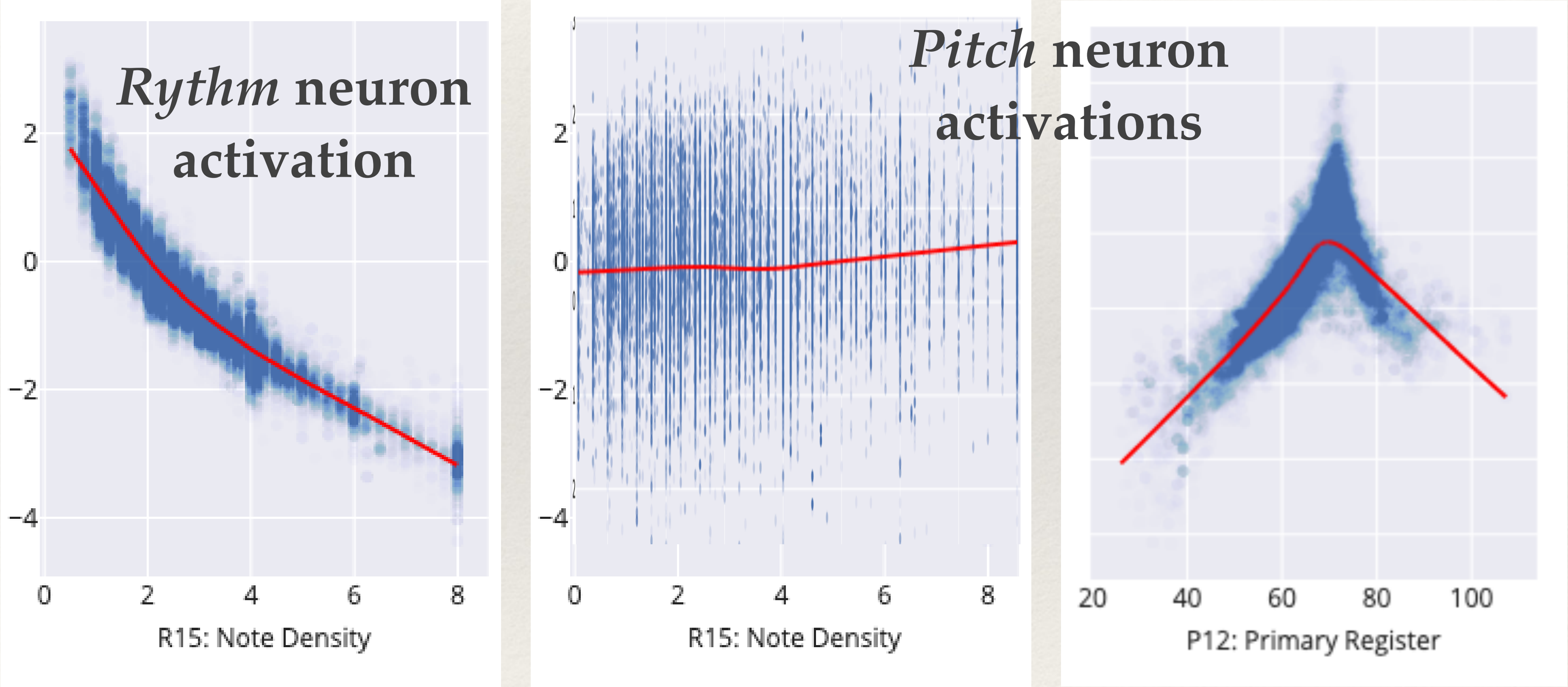
EXAMPLE - MUSIC *in preparation with Barenboim, Hirn and del Debbio*

only a handful neurons are meaningful, do they carry human information?

One neuron for rythm

One neuron for pitch

Similar for melody



The VAE *discovers* the concepts of rythm, pitch and melody
aligns its latent space accordingly

Summing up…

We are just starting to understand the applications of ML in Physics

So far, dominated by the low-hanging fruit: supervised classification
ML brings added value, shortening data taking times

They go beyond a mere iteration of our traditional statistical methods:
unsupervised methods, generative AI, reinforcement learning…

Summing up…

We are just starting to understand the applications of ML in Physics

So far, dominated by the low-hanging fruit: supervised classification
ML brings added value, shortening data taking times

They go beyond a mere iteration of our traditional statistical methods:
unsupervised methods, generative AI, reinforcement learning…

NNs can discover unknown principles (symmetries, conservation laws)
when asked to perform an unsupervised learning task
Opportunity to learn new concepts in physics

Summing up…

We are just starting to understand the applications of ML in Physics

So far, dominated by the low-hanging fruit: supervised classification
ML brings added value, shortening data taking times

They go beyond a mere iteration of our traditional statistical methods:
unsupervised methods, generative AI, reinforcement learning…

NNs can discover unknown principles (symmetries, conservation laws)
when asked to perform an unsupervised learning task
Opportunity to learn new concepts in physics

Through AI methods: interesting cross-pollination between
our area (PP) and others
Opportunity to learn from other areas in Science

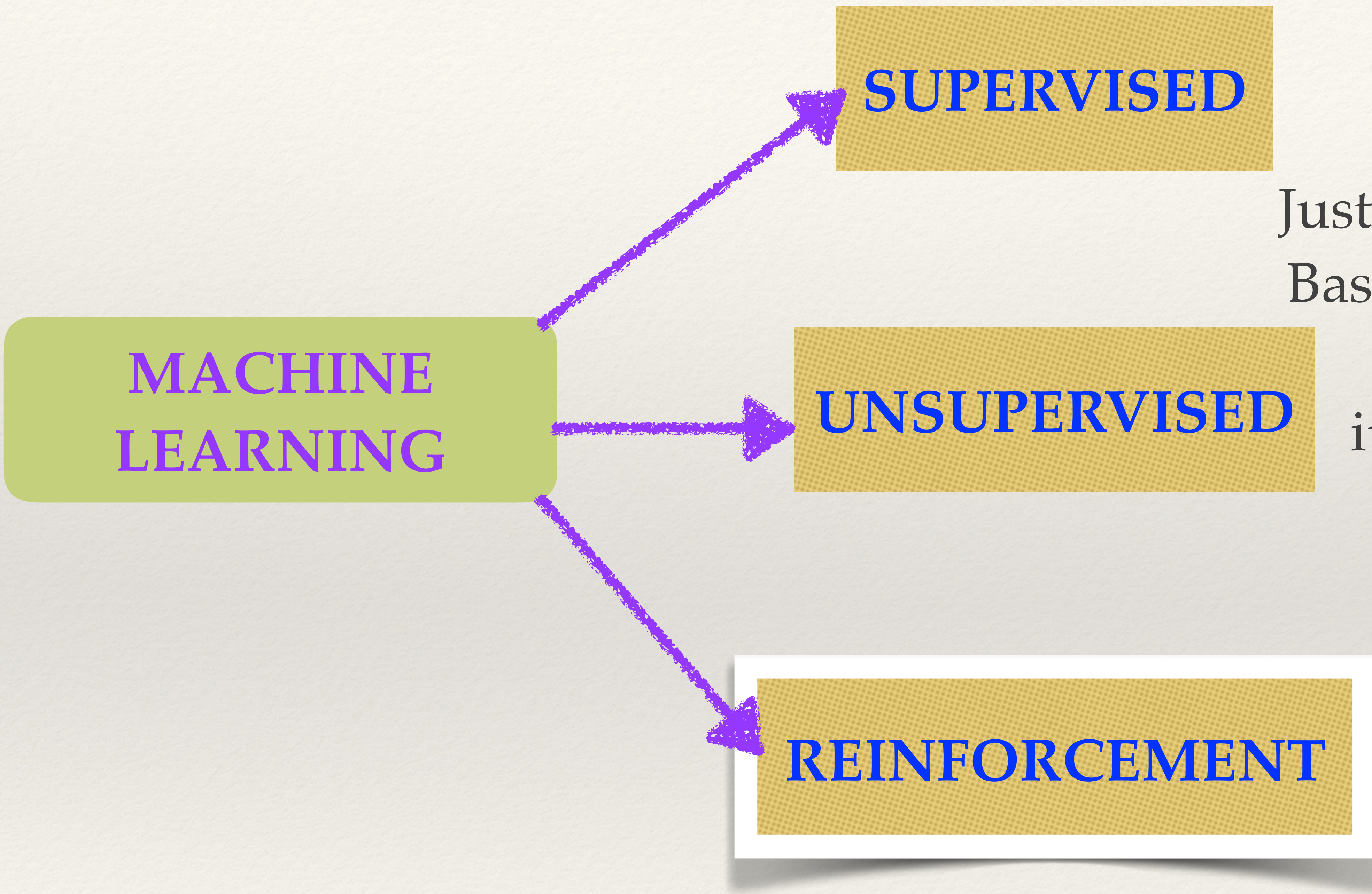
Enjoy the meeting!

Can I have a cookie?

Learning by reward

Additional

Types of learning



Just touched the surface
Basis to explore further
and incorporate
it in your research

Supervised to Reinforced Learning

Cool ways to accelerate learning, capture important aspects of the data, incorporate different types of data

Learn **from** humans to do what humans **already do**, but better and faster, and in more difficult situations

But, what if we wanted a machine to become **better** than a human at completing a high-level task?

* See these lectures

Let's find a DIFFICULT task

A truly human-difficult task
not just a task that a machine can do faster or with lower resolution

Supervised/unsupervised learning identifies *patterns* in data
But this isn't the same as learning to develop a ***strategy***
and to do it better than a human

Chess is a high-level activity
different players develop different strategies
the goal is ***long-term***
important pieces can be sacrificed to achieve
checkmate some moves along the way
and you have an adversary which will oblige
you to ***reassess*** your strategy at each step
combinatorics is ginormous

Human vs Machine

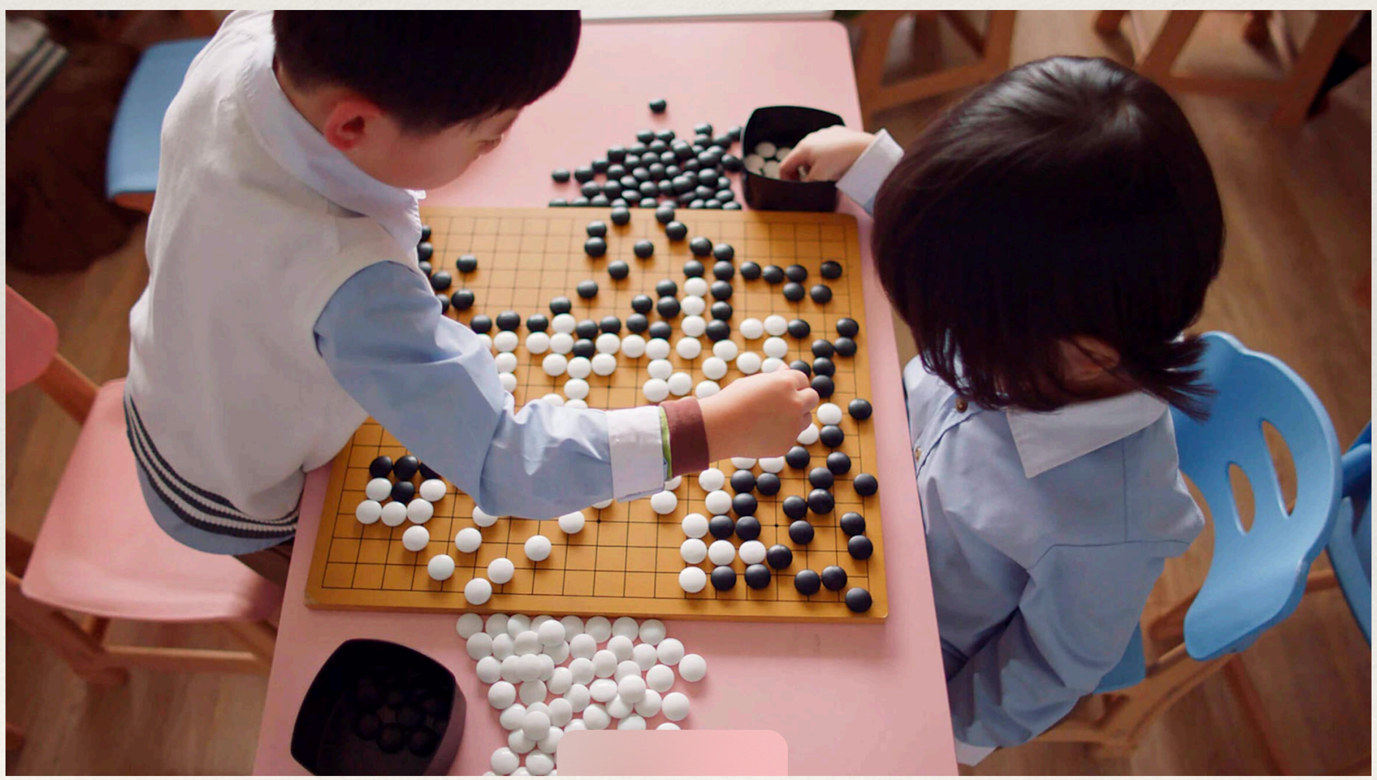
February 1996

Deep Blue (IBM) beat Garry Kasparov (World Champion) and did it again many times after brute-force computing power analysing many hundreds of millions positions / second

Human vs Machine

February 1996

Deep Blue (IBM) beat Garry Kasparov (World Champion) and did it again many times after brute-force computing power analysing many hundreds of millions positions / second



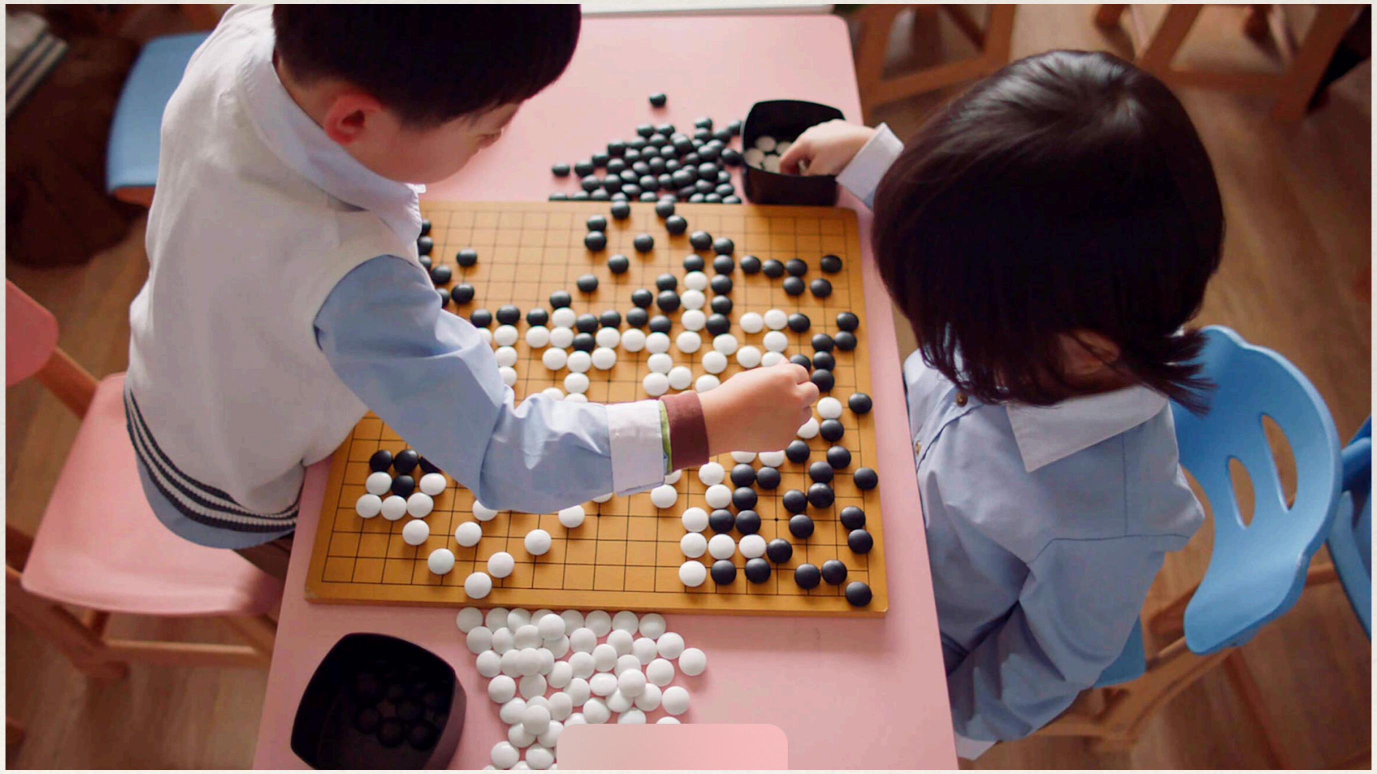
October 2015

AlphaGo Zero beats a professional Go player learned from playing against *itself*

November 2017

AlphaZero builds on DNNs to beat world champions in Go, chess and shogi

Human vs Machine



February 1996

Deep Blue (IBM) beat Garry Kasparov (World Champion) and did it again many times after brute-force computing power analysing many hundreds of millions positions / second

October 2015

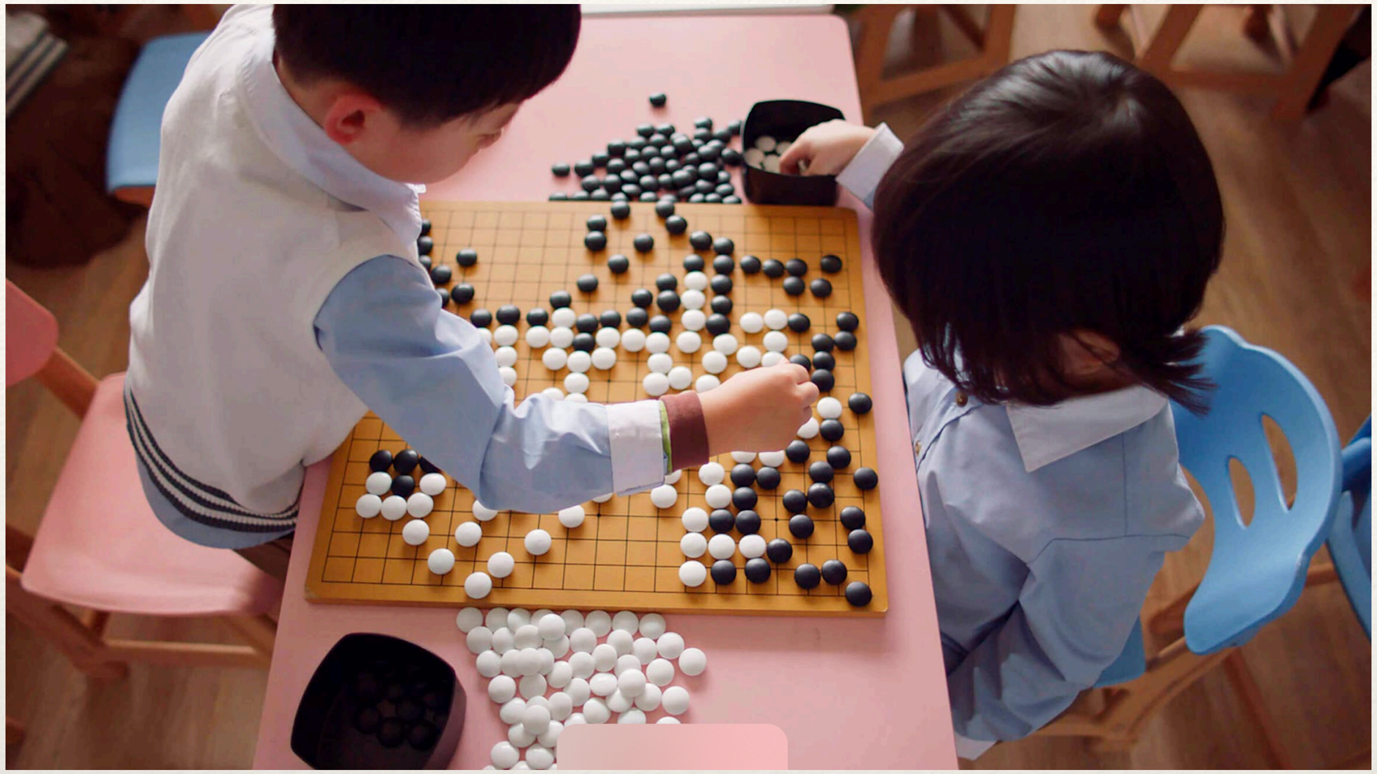
AlphaGo Zero beats a professional Go player learned from playing against *itself*

November 2017

AlphaZero builds on DNNs to beat world champions in Go, chess and shogi

A new paradigm of learning: **REINFORCEMENT**

Go game

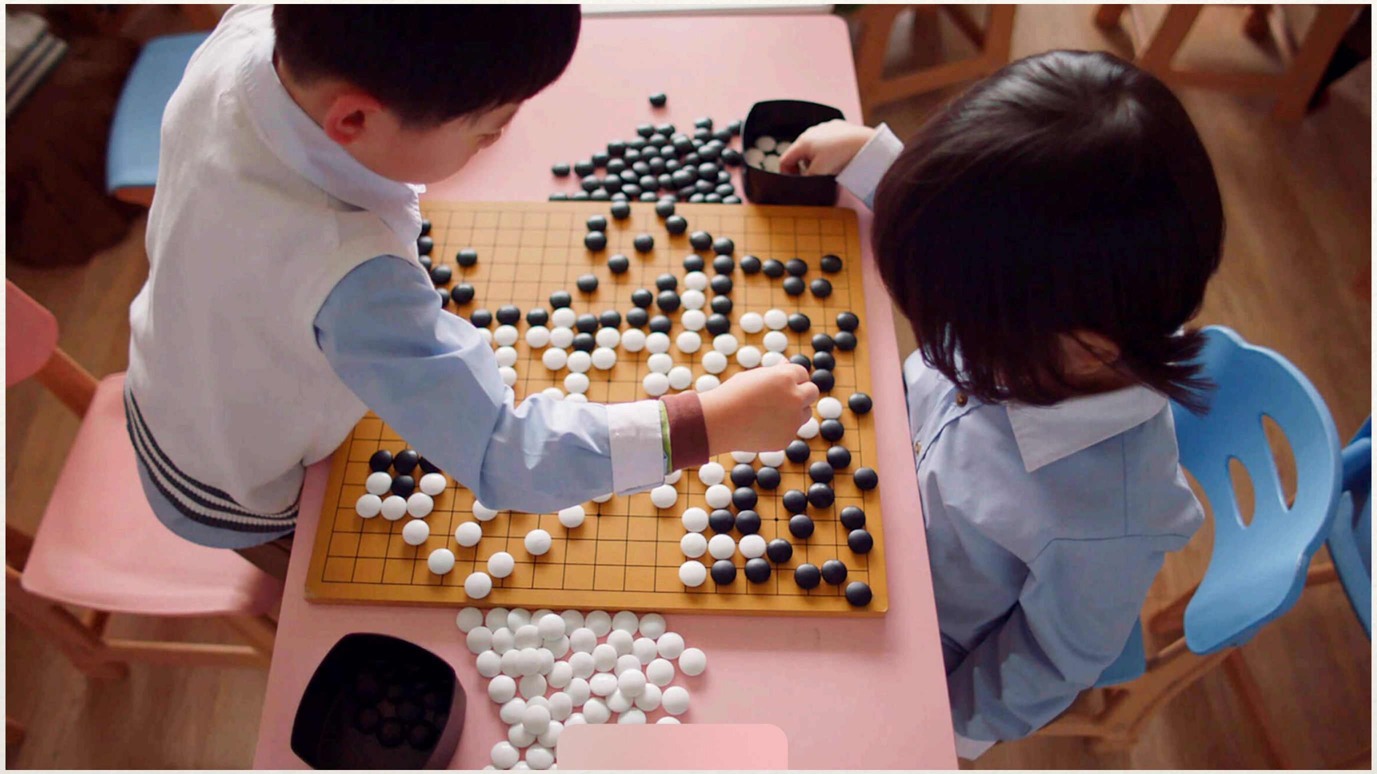


Simple game: moves are simple
no hierarchy like chess
king/queen/bishop/pawn…
goal: surround and capture
opponents' pieces

Simple rules, extreme levels of complexity when building strategies
no machine could beat a Go-master until 2015
Why is it so difficult?
how would you teach a machine to learn this game?

X,y

Go game



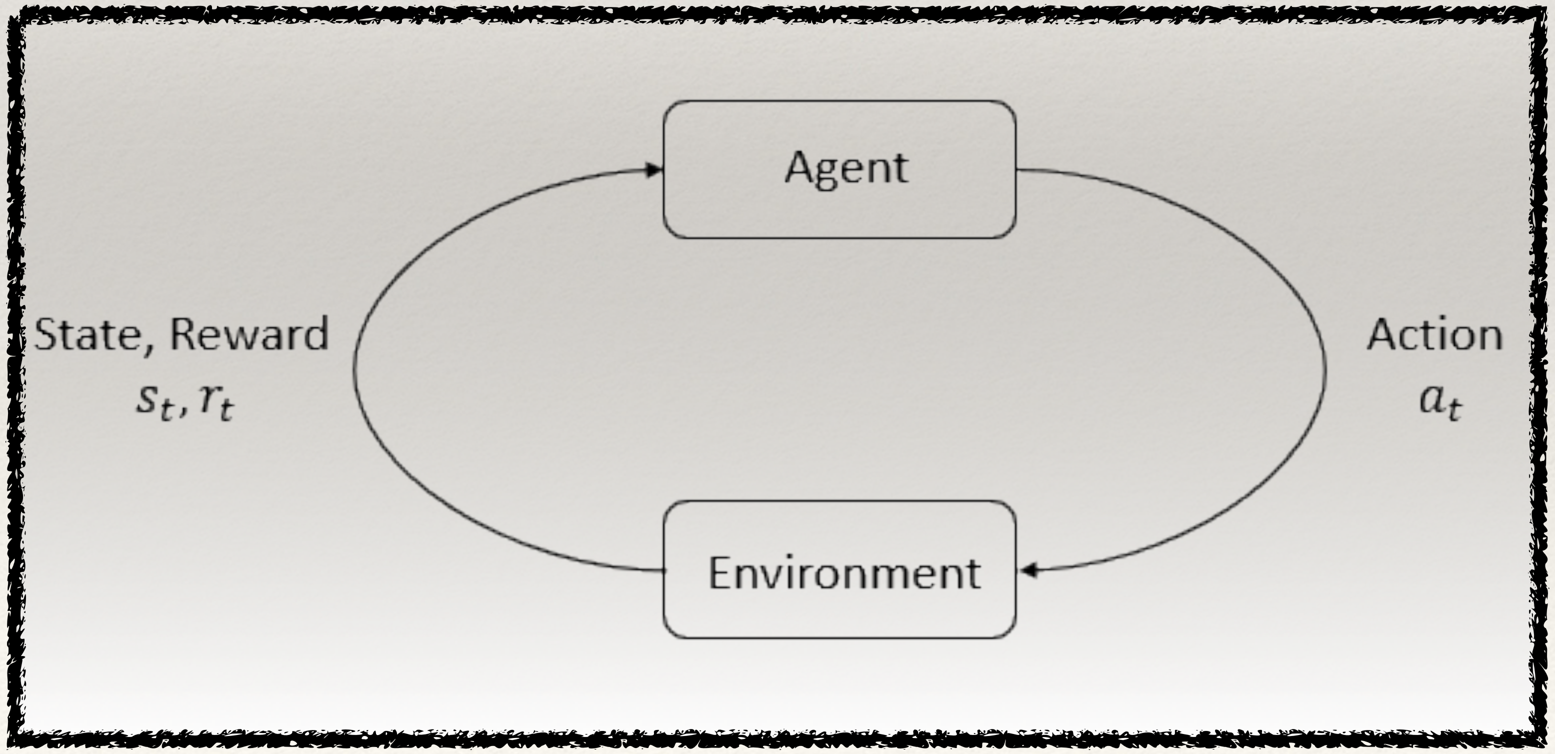
Simple game: moves are simple
no hierarchy like chess
king/queen/bishop/pawn…
goal: surround and capture
opponents' pieces

develop a strategy for long-term winning:
3^(19*19)~10^172 configurations at one step
decision in this one step guided by possible future gains
but opponent's actions change every subsequent move

Reinforcement learning

The task of getting better at Go was too difficult
too many possibilities, no human could teach from example
To beat humans we had to allow machines to learn in a different way

Machine needs to learn to make good sequences of decisions
dealing with delayed labels and developing a long-term strategy
Some form of iterative way of improving strategy
which can examine many steps ahead



agent interacts with
the **environment** in **state** *st*
takes **actions** based on **reward** *rt*
which tells about good current state is
GOAL: maximise total about of
rewards (**return**)
RL help the agent to achieve goal

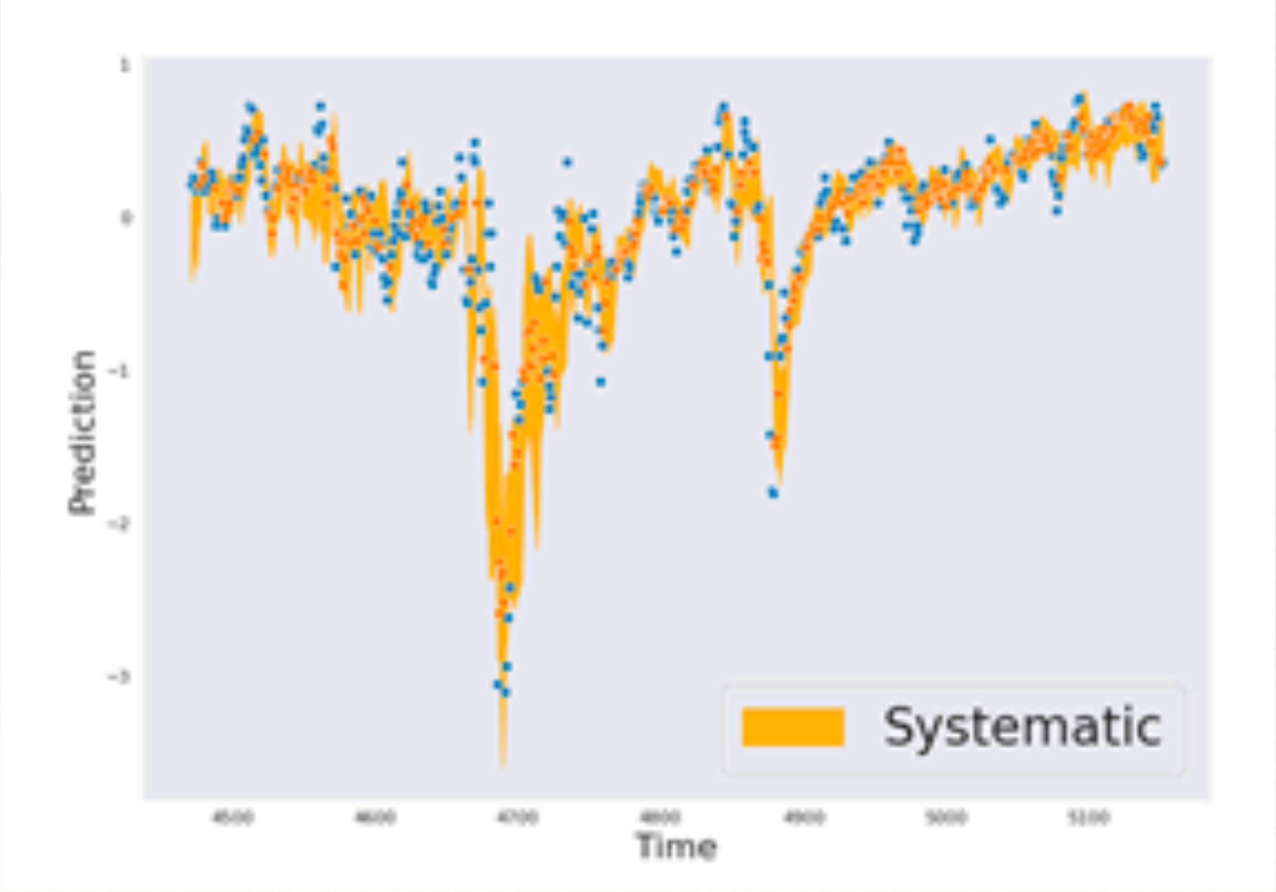
“Experience is a lantern that you carry on your back and that only lights up the path you have traveled.” *Confucius*

Knowing the past, predicting the future

predict the evolution of a situation

Never mind, Confucius!
ML *can* predict the future

By learning from examples of time series
(snapshots of past->future sequences)
and
using **RNNs (recurrent NNs)**
in particular **LSTMs (long short term memory)**



Time evolution of
the solar activity
blue-> reality
orange-> prediction