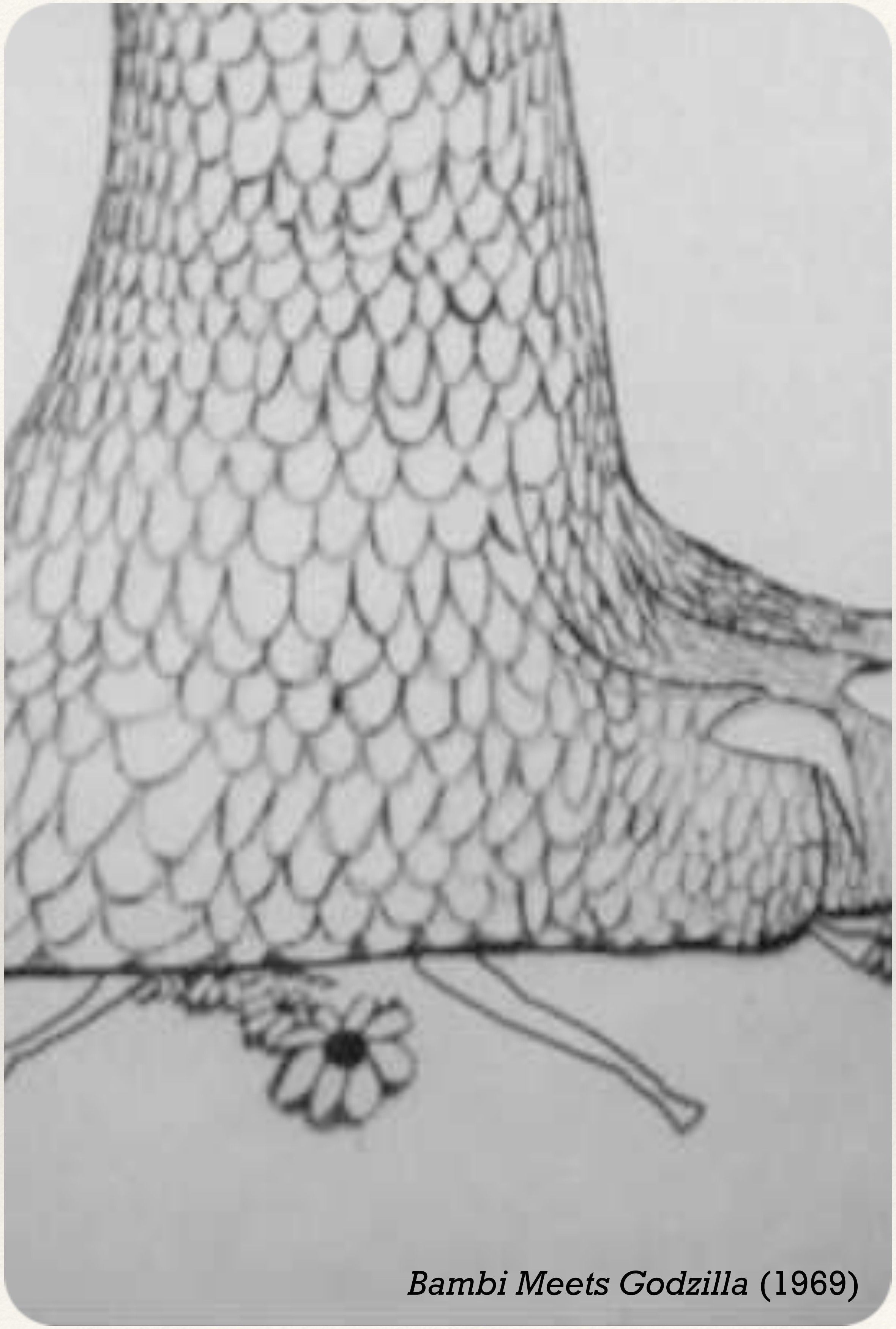
Symmetry meets AI

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

Colloquium@Torino'23



Bambi Meets Godzilla (1969)

Today, we will talk about

*Human vs Machine Learning
*Supervised Learning
*Going further: unsupervised
*Human surrender
*Looking under the hood
*From Physics to Art:
Paintings and music

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

Human vs Machine Learning

Human learning

repeat and improve on a task

Previous experience

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

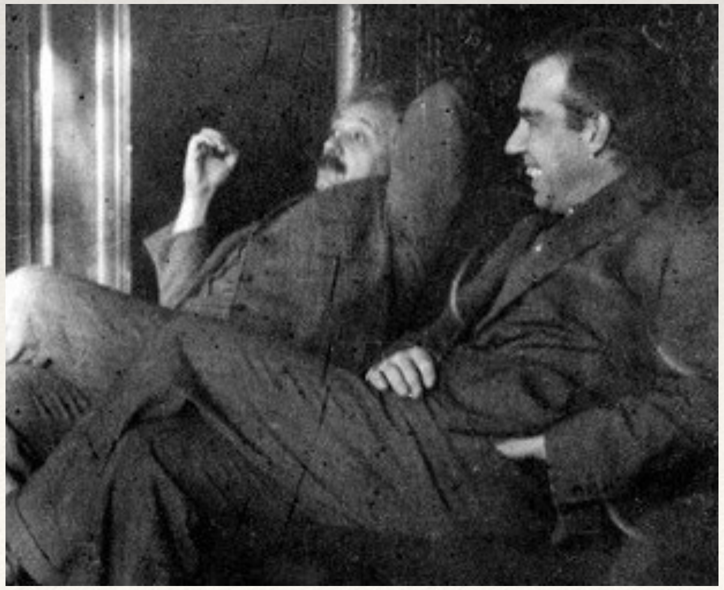
imagine new possibilities

VERY IMPRESSIVE, YET

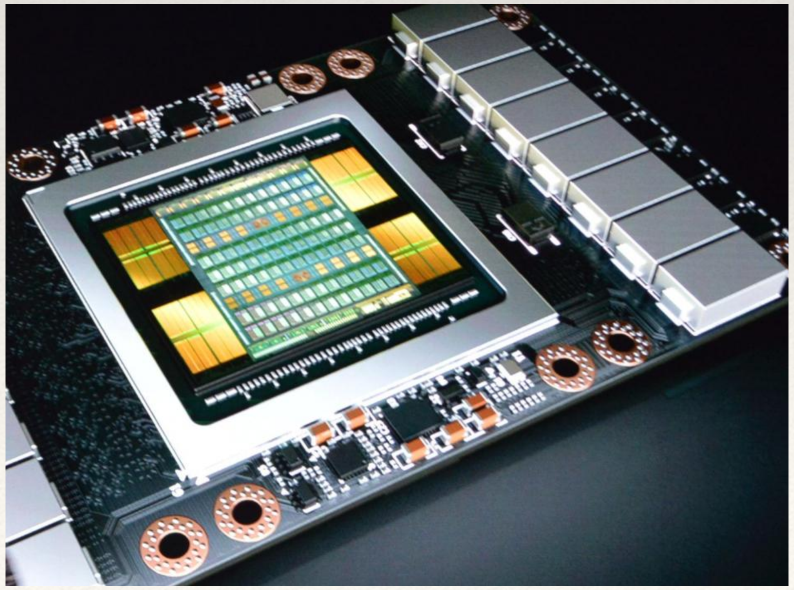
human learning is limited by
our personal viewpoint,
our collective intelligence (*newspeak?)*
& our inherent capacity to process information
(amount , speed, level of detail)

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

TIME-SERIES LEARNING

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

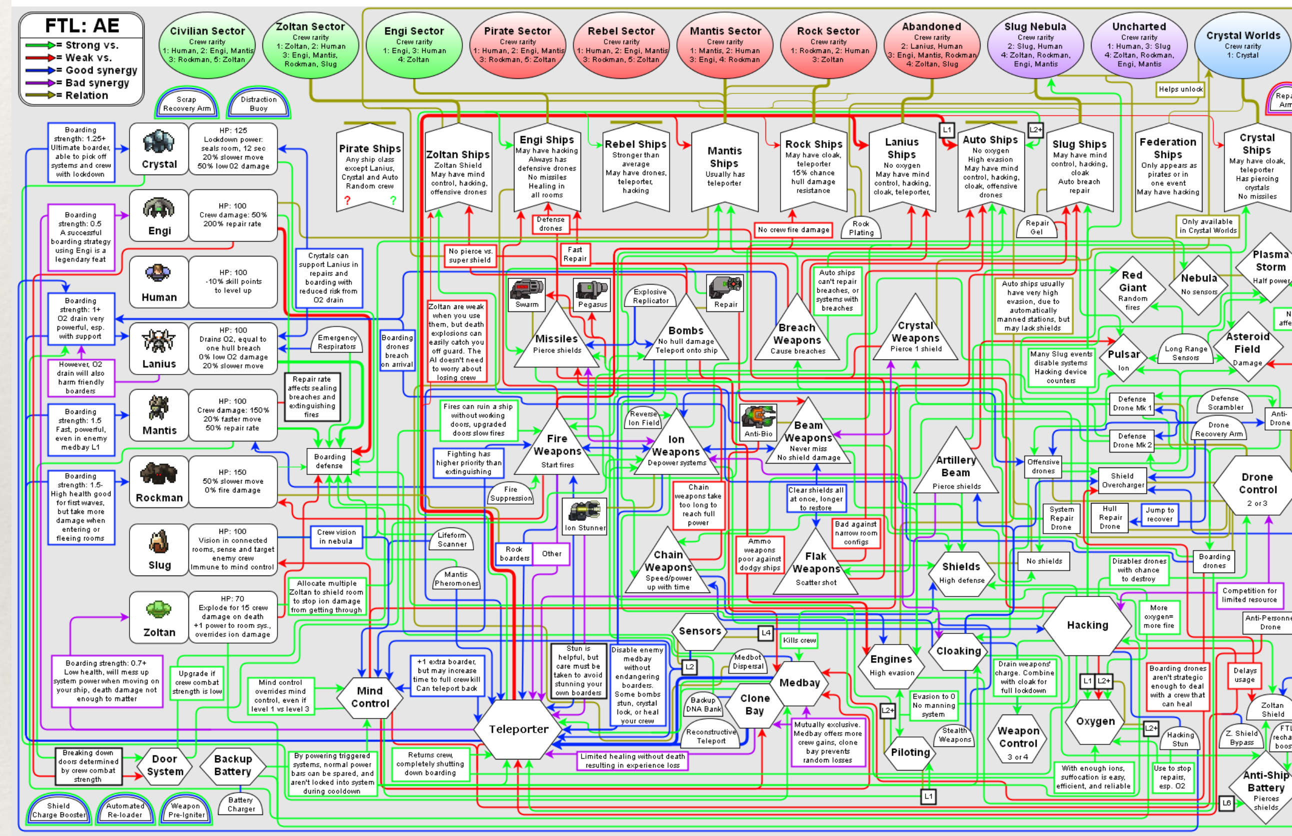
---

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



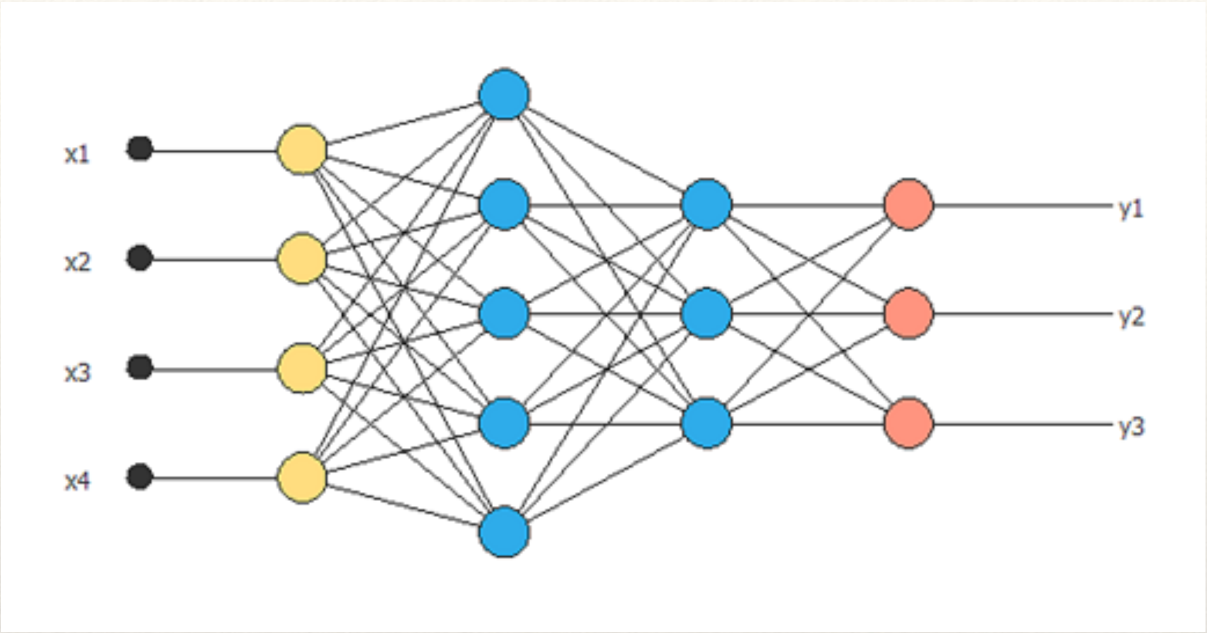
The Machine can't describe relations we haven't coded in

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

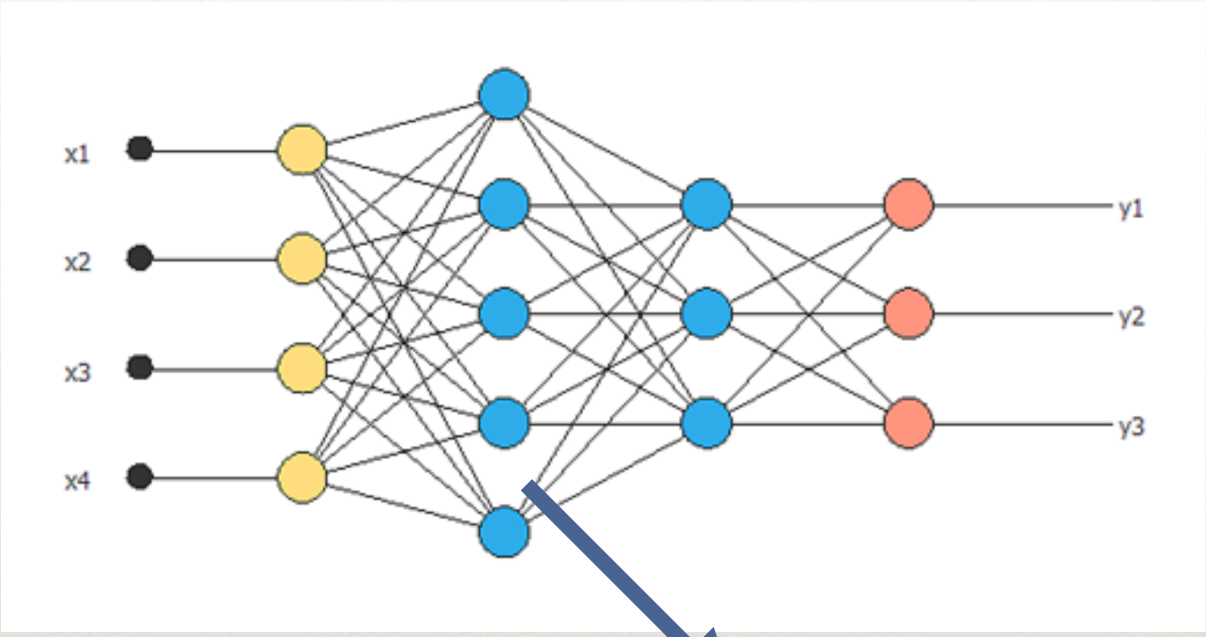
The network learns from data with no structured instructions

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

A new way of *thinking*: Neural Networks

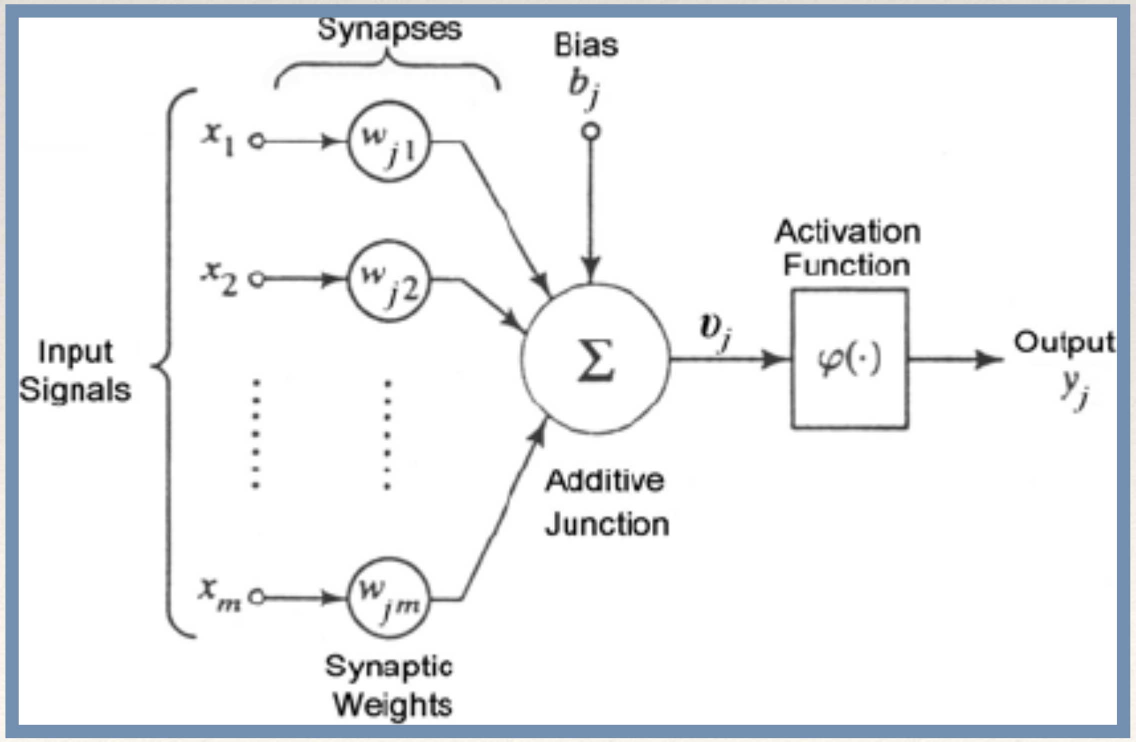
Structures made of units called *neurons*
and organised by *layers*

inputs



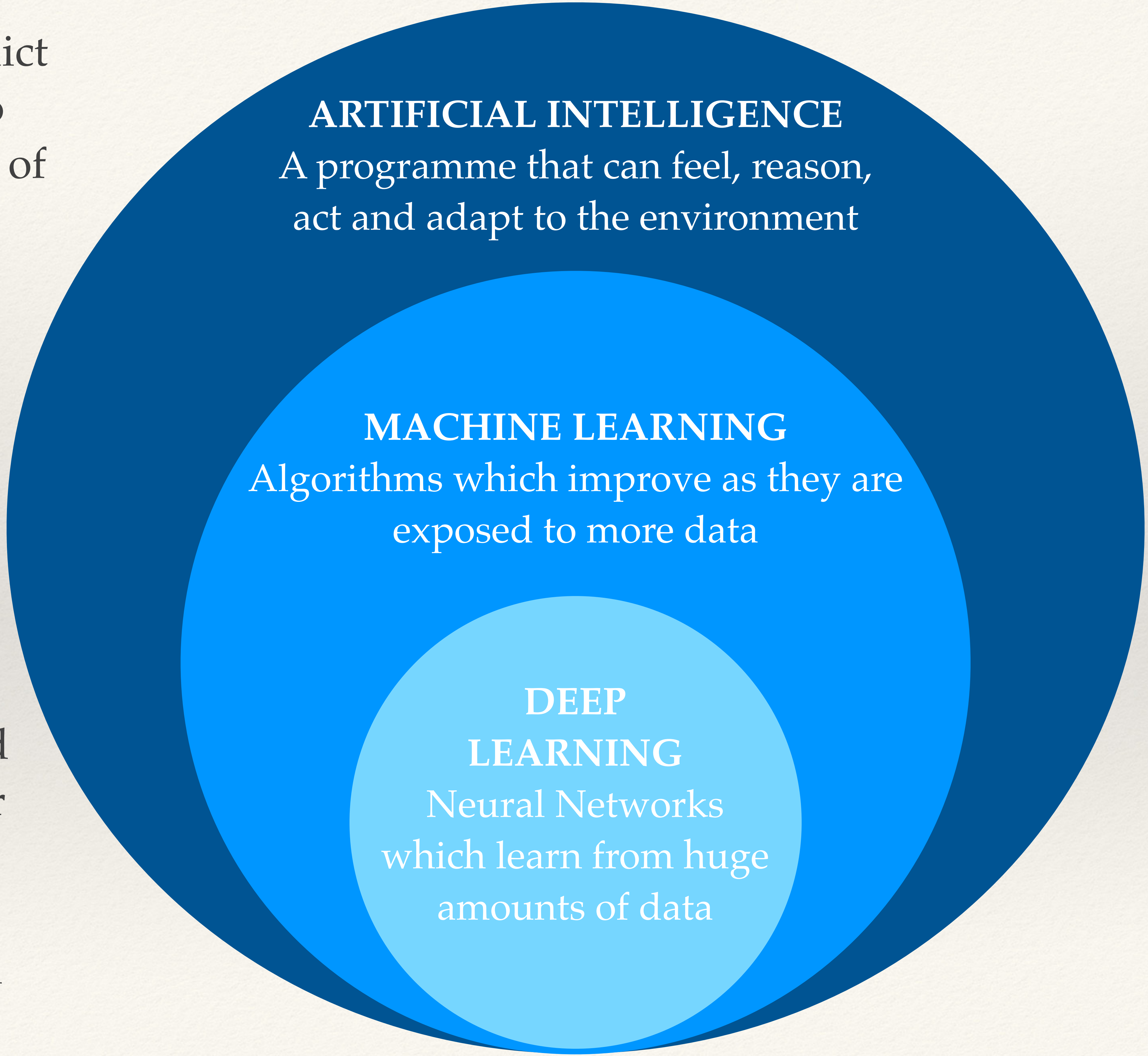
outputs

Inspired in the brain's
neurons and the
mechanism of synapsis



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?

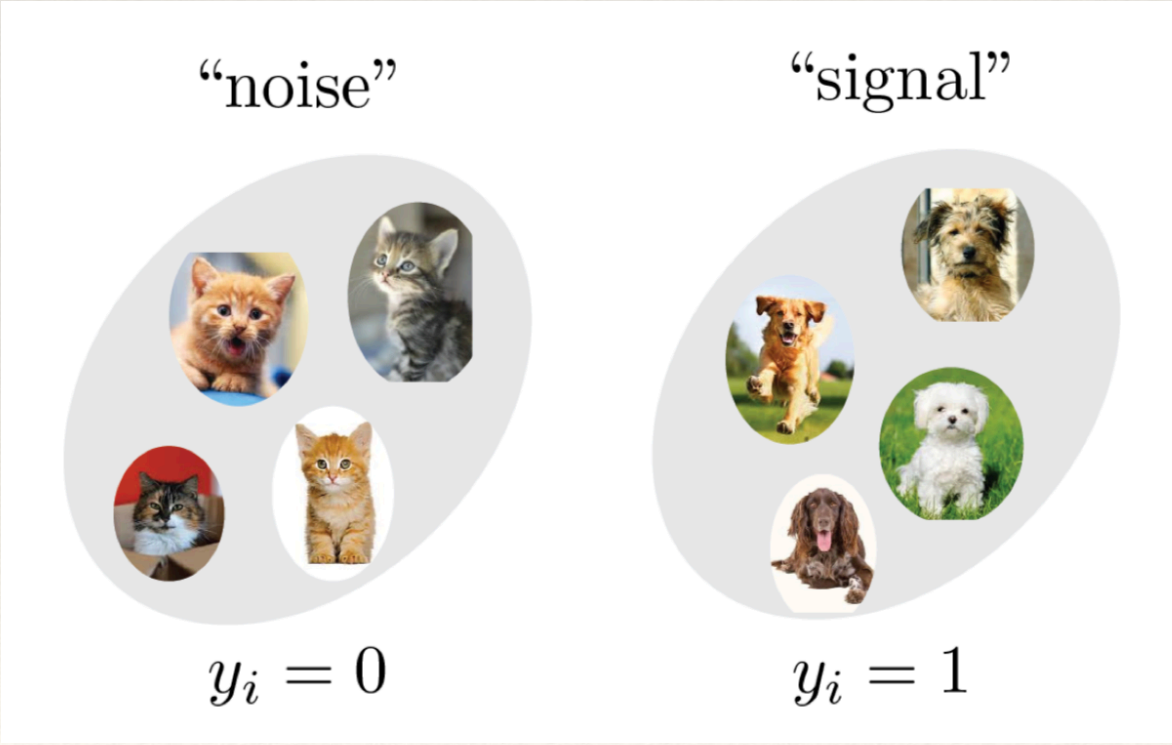
Yes/No answer

A basic task: good or bad?

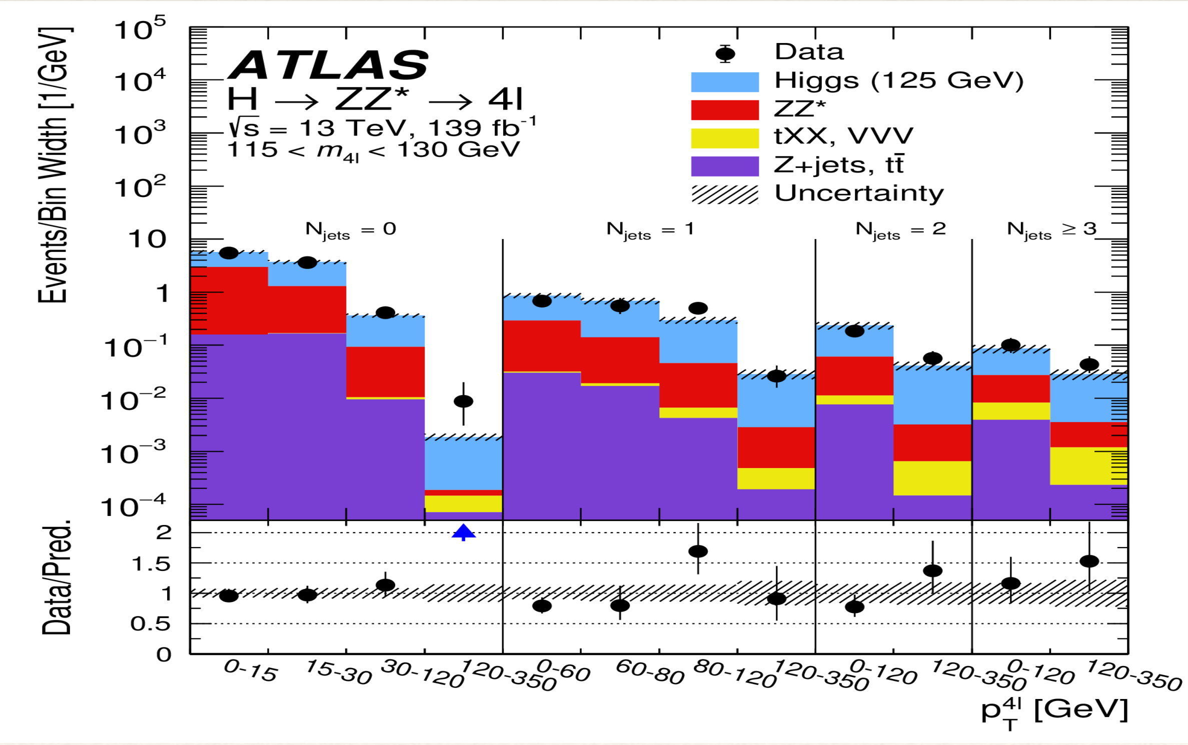
Is it a crocodile?

Yes/No answer

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

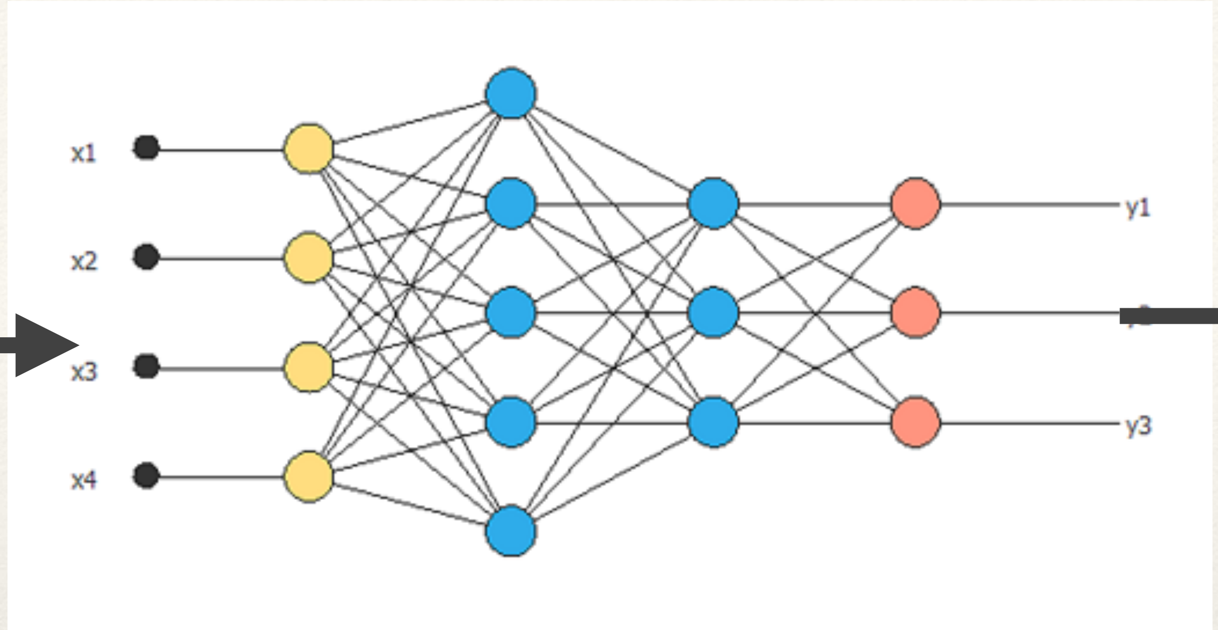


Cat or dog?



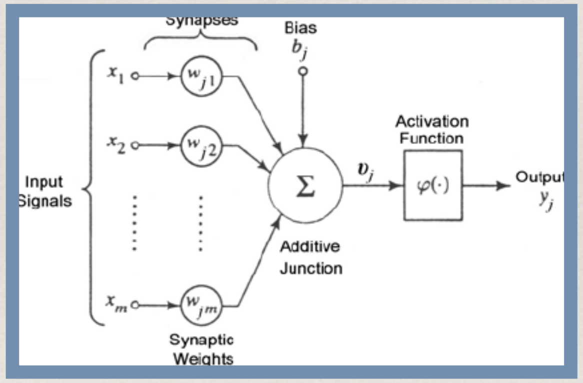
Is this New Physics?

Training a binary network



$\rightarrow y \in \{0, 1\}$

init network → transform all images in numbers *y* the prediction (cat/dog)



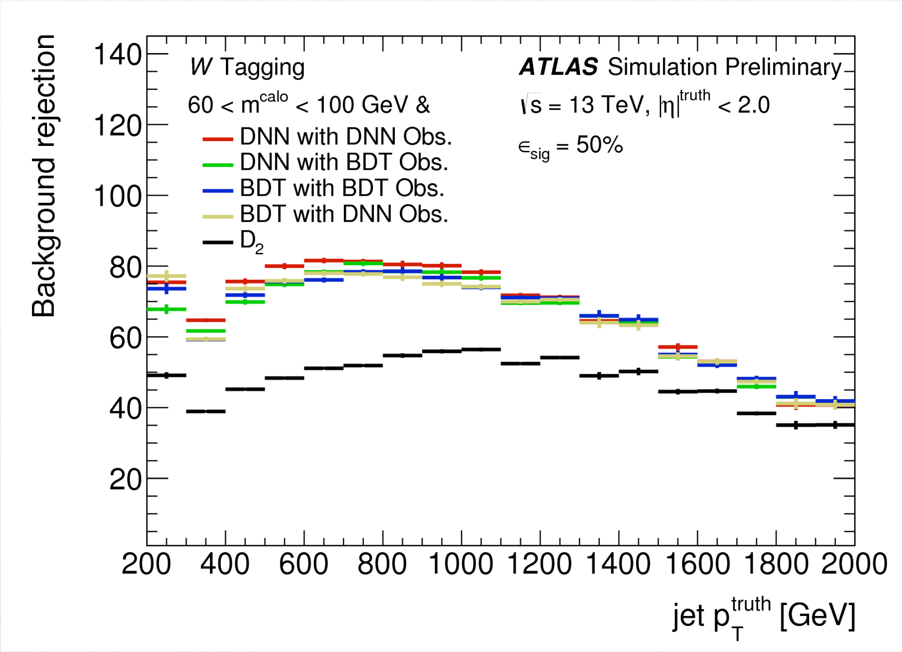
The network applies many non-linear transformations on the input the result $y(x)$ is highly non-linear → check the statistical accuracy of your result

↓

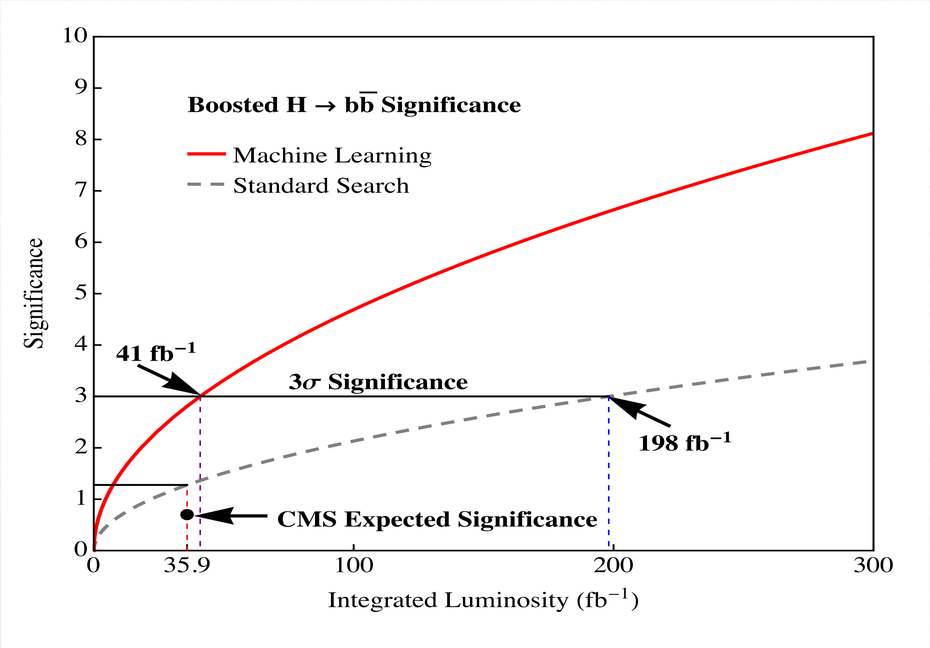
change the init parameters a bit ← we start over, checking the prediction accuracy again ← we adjust the direction depending on learning slope

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

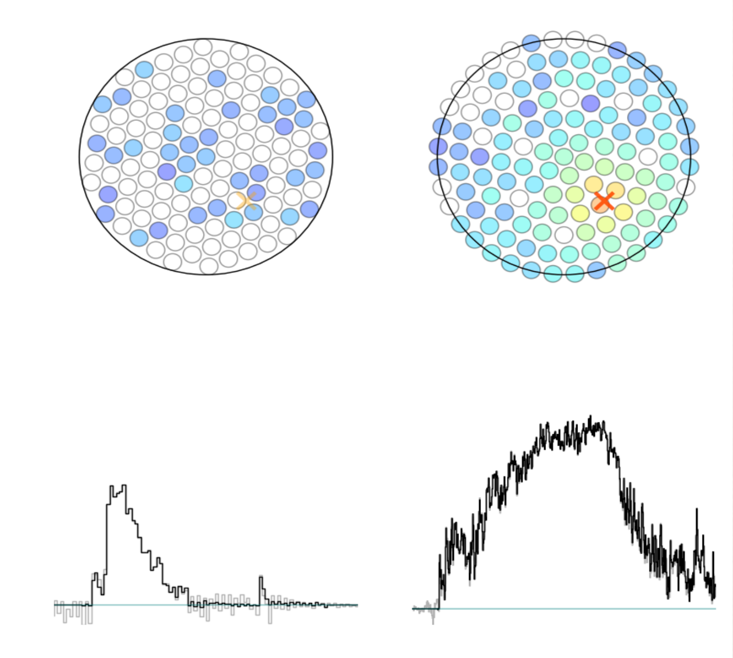
Is it a W?



Is it a Higgs?



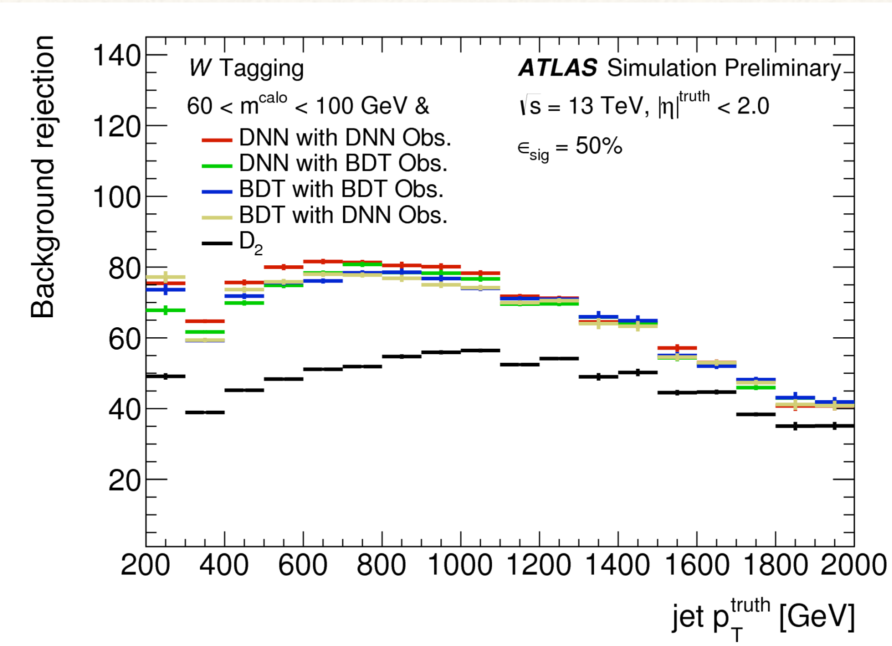
Is it DM?



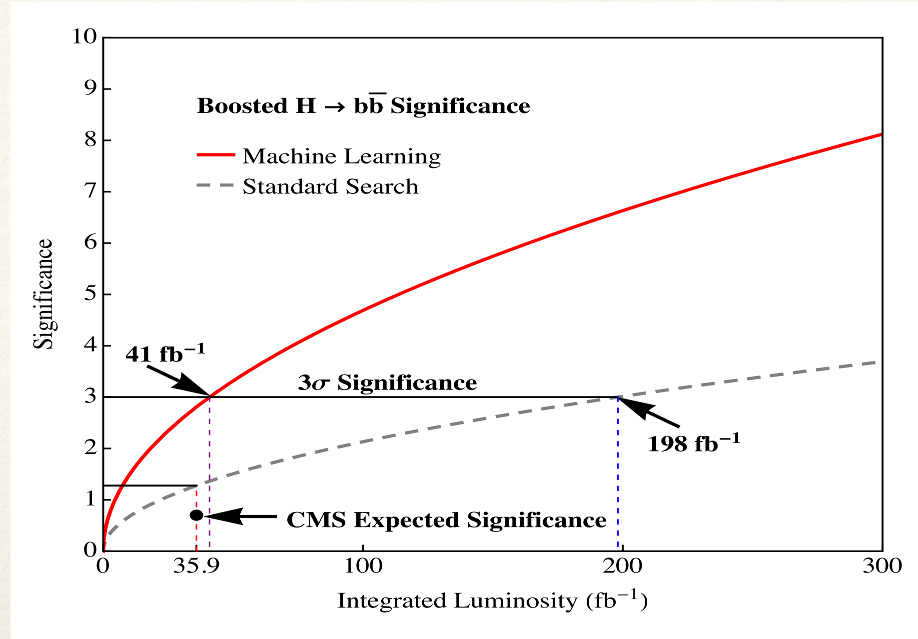
mostly using Neural Networks to deal with images (CNNs)

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

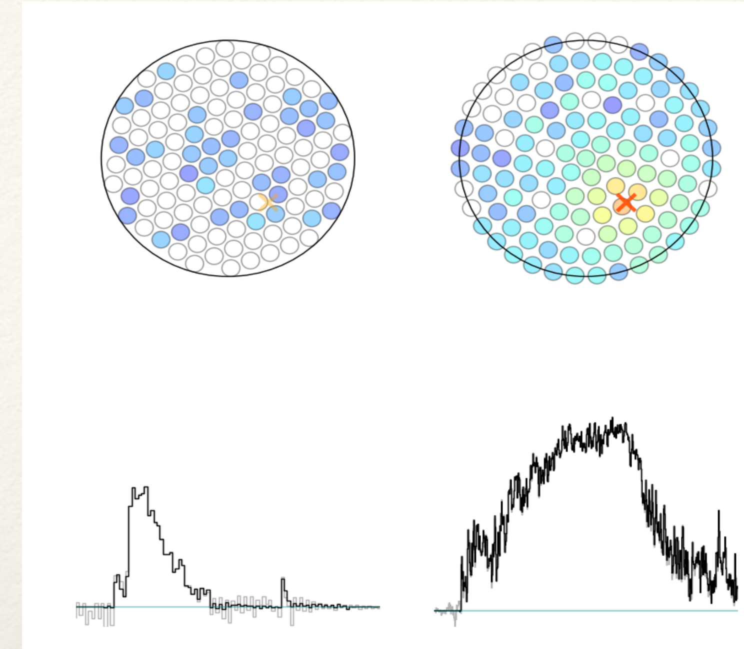
Is it a W?



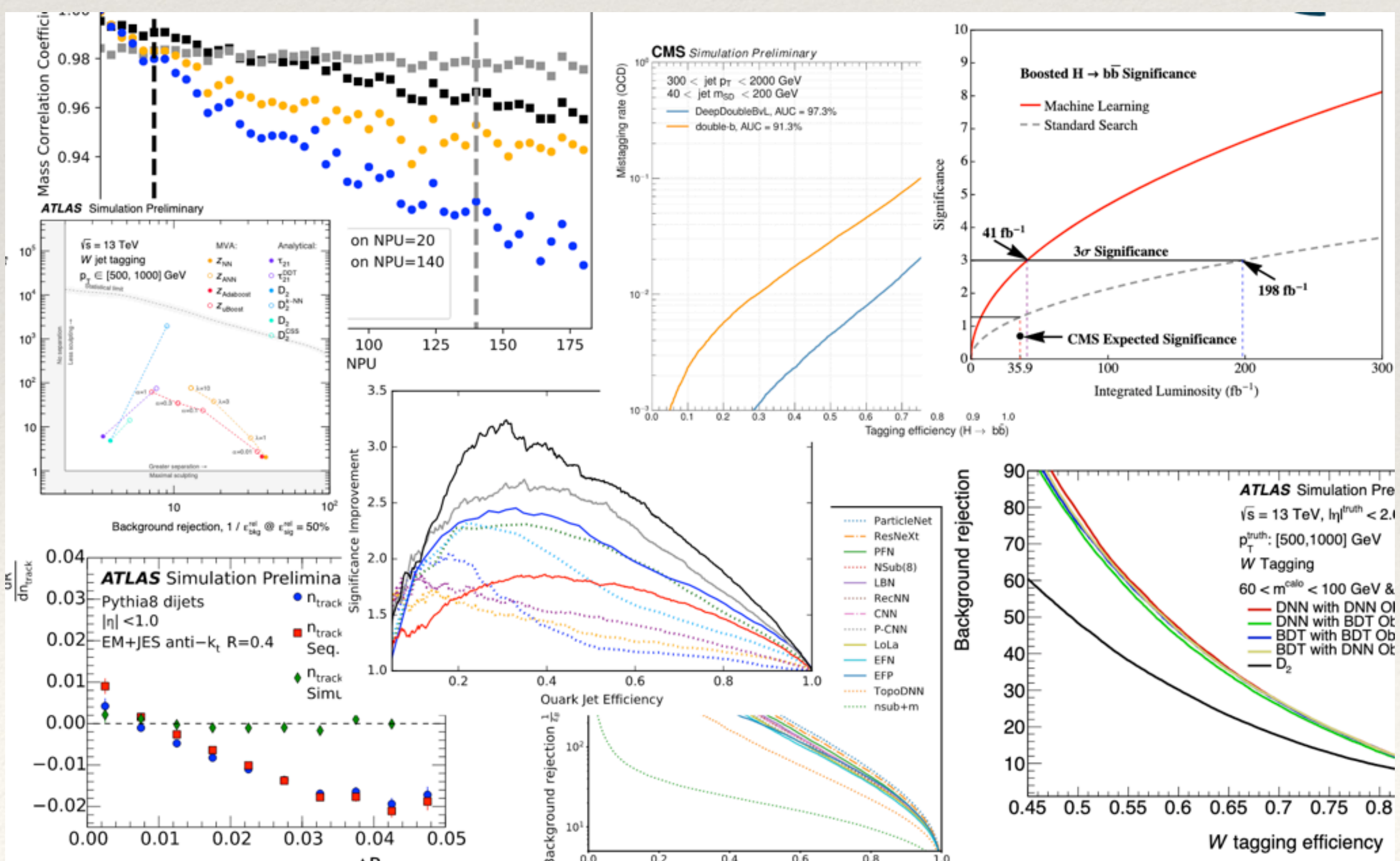
Is it a Higgs?



Is it DM?



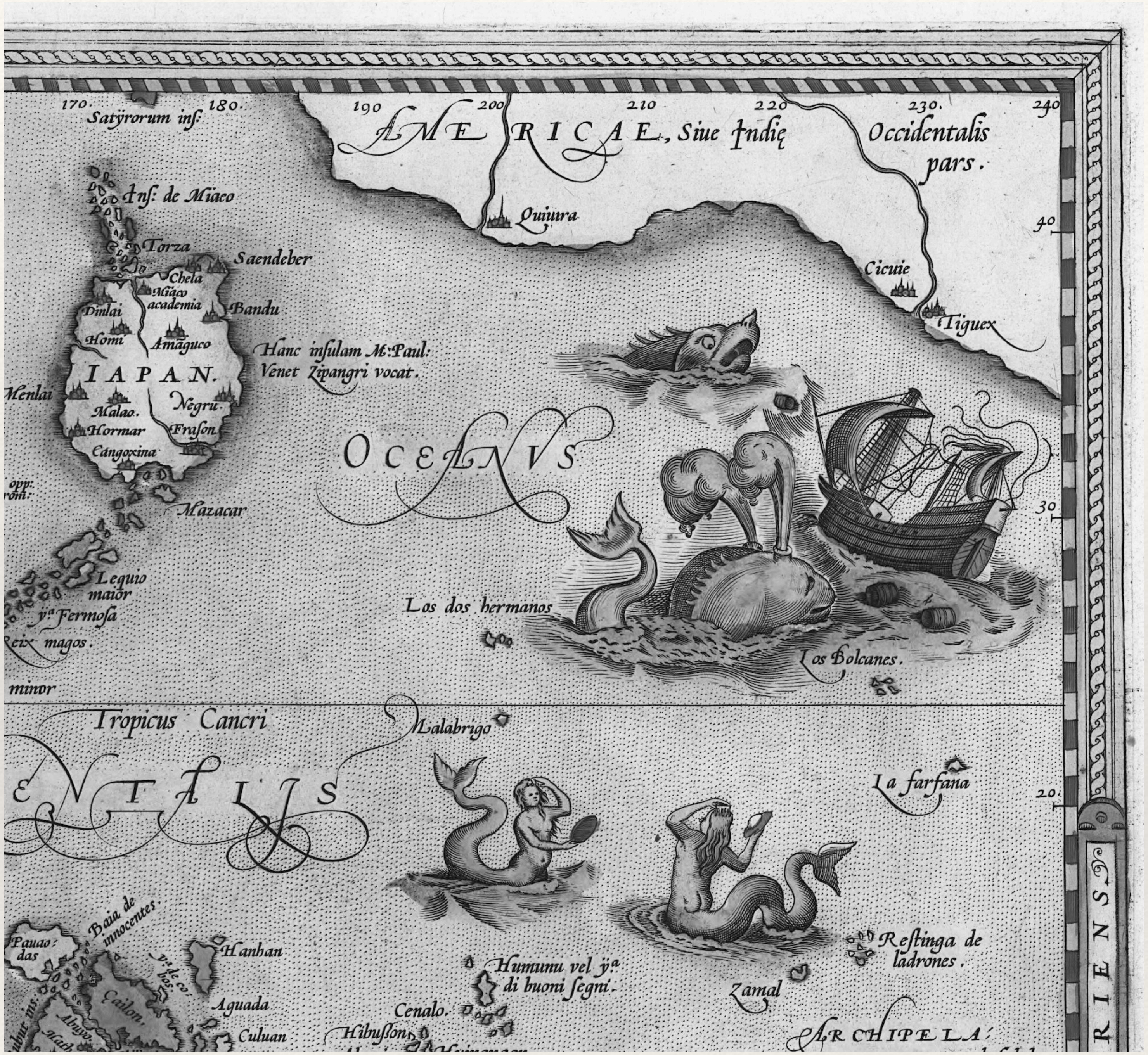
mostly using Neural Networks to deal with images (CNNs)



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)

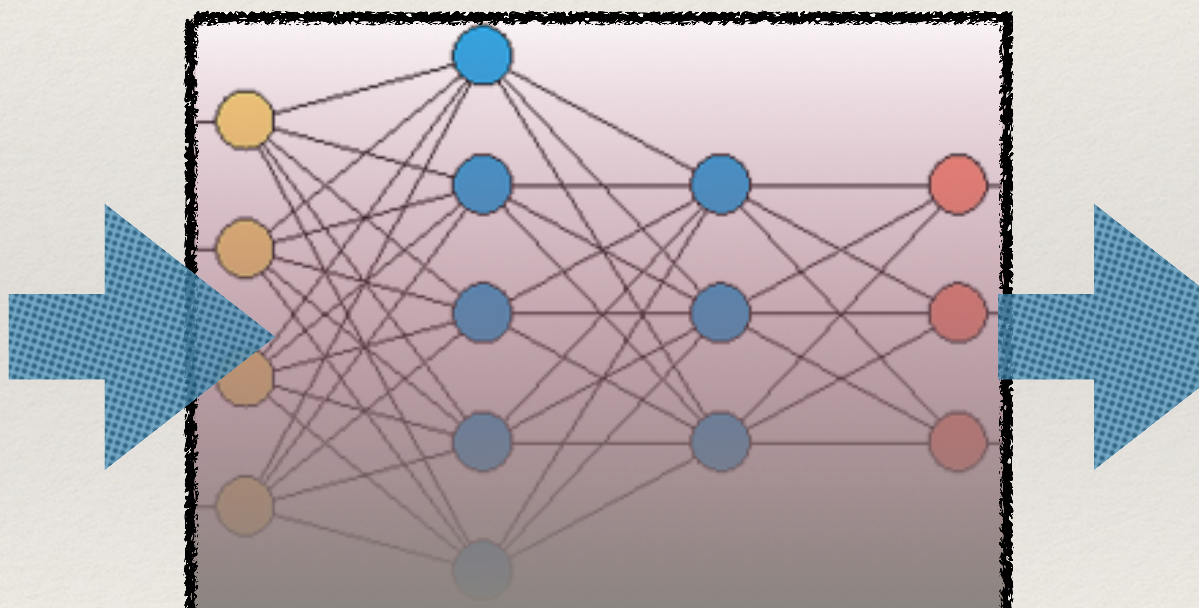
and **VAEs (Variational AutoEncoders)**

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

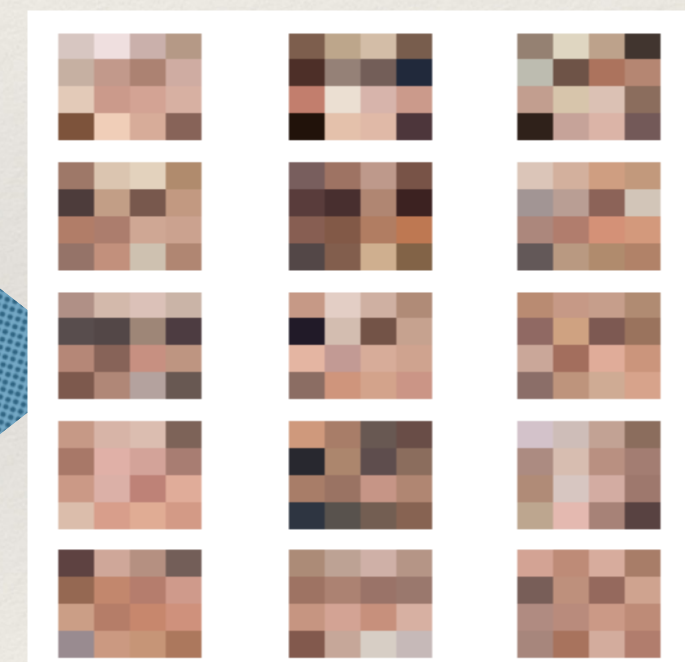
Here, human faces

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?

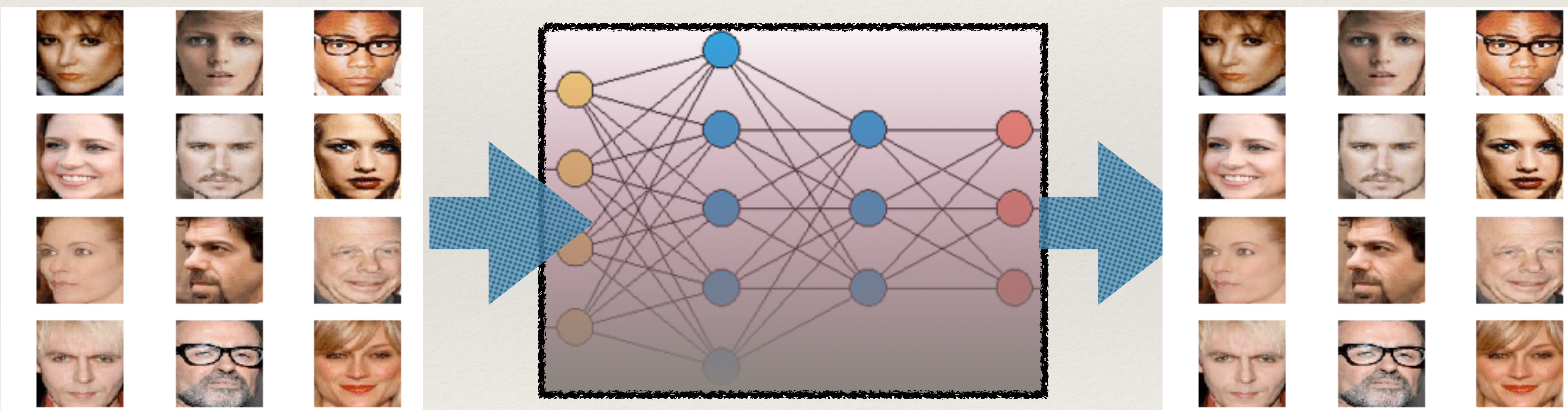
GANs (Generative Adversarial Networks)

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**

Wait a minute!

Aren’t we just programming the identity transformation?

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?

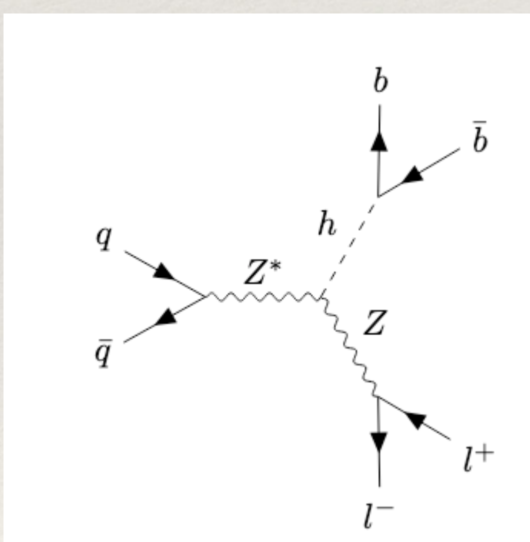
GANs (Generative Adversarial Networks)

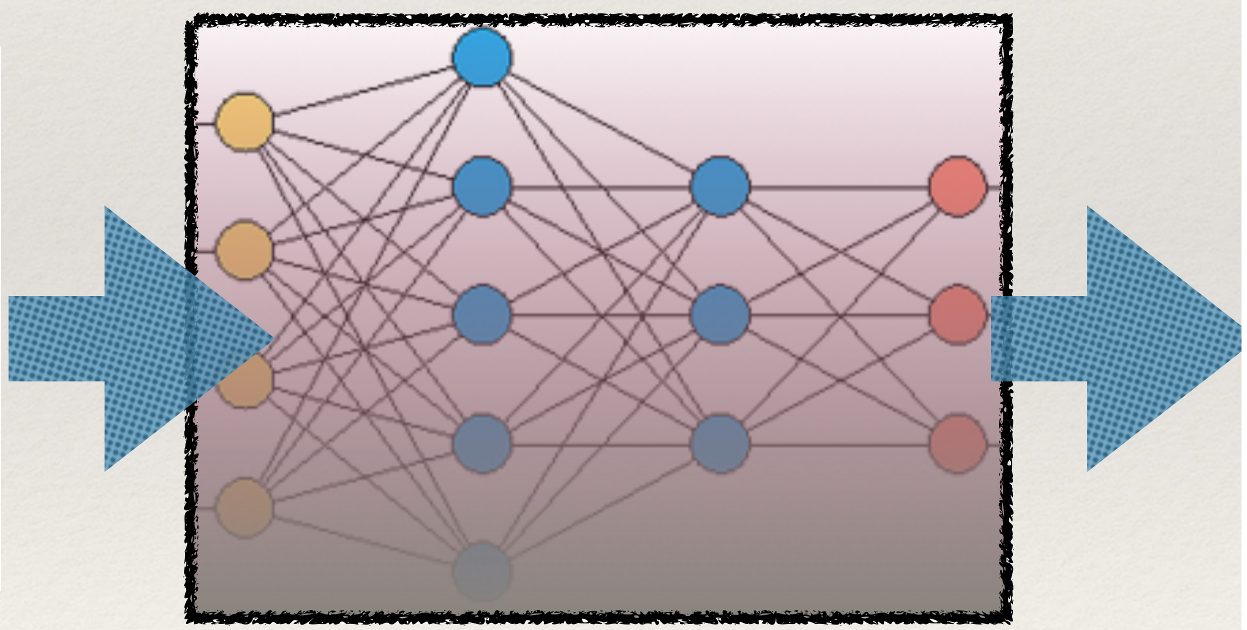
and **VAEs (Variational AutoEncoders)**

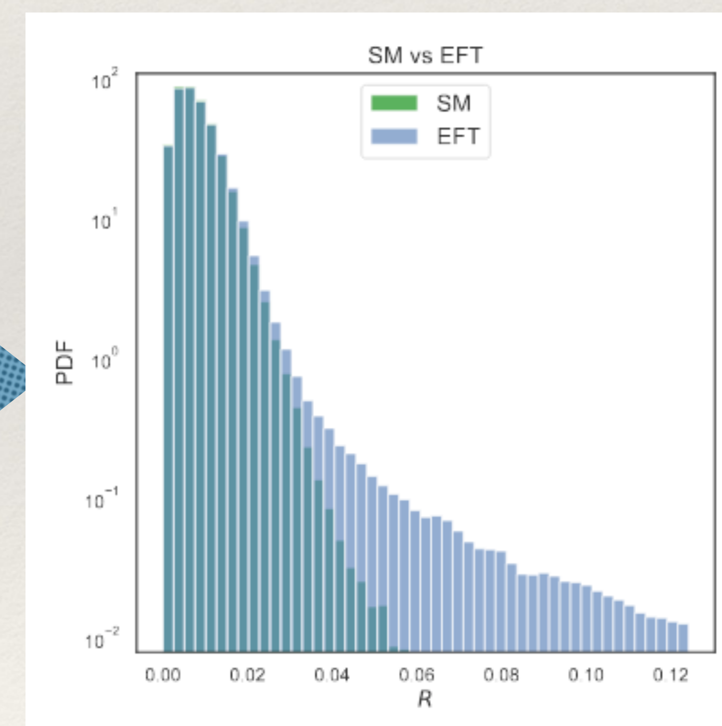
In CNNs, benchmarks were cats/dogs and hand-written digits (MNIST)

Here, human faces

Example in Particle Physics with Mike Soughton and Charanjit Khosa *SciPost*







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

Learns to ID outliers ('New Physics')

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?

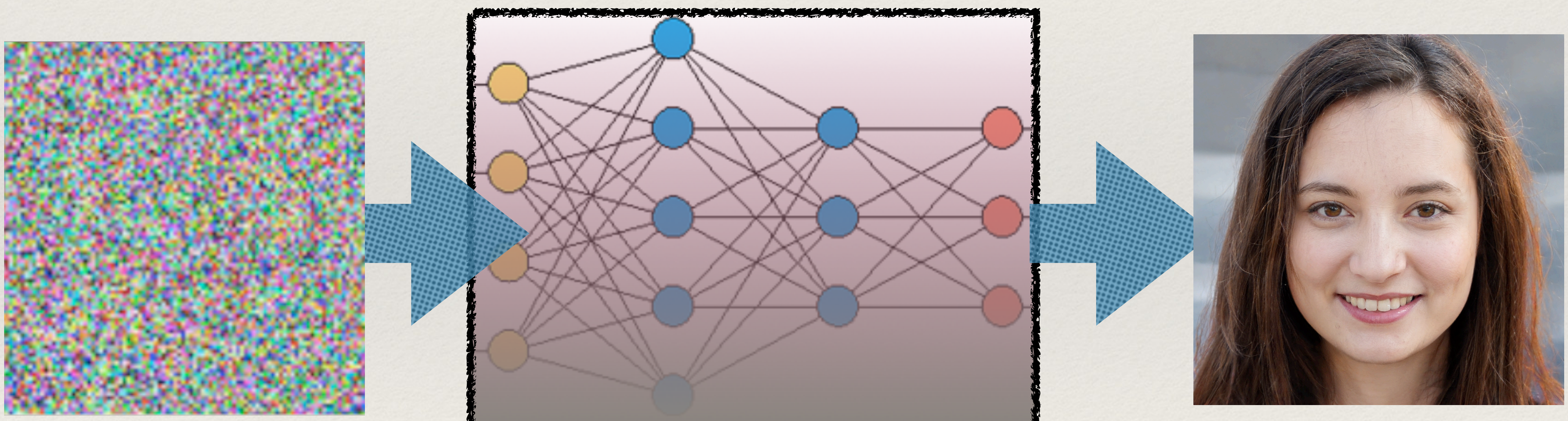
GANs (Generative Adversarial Networks)

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?

Supervised learning input-> predict output

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

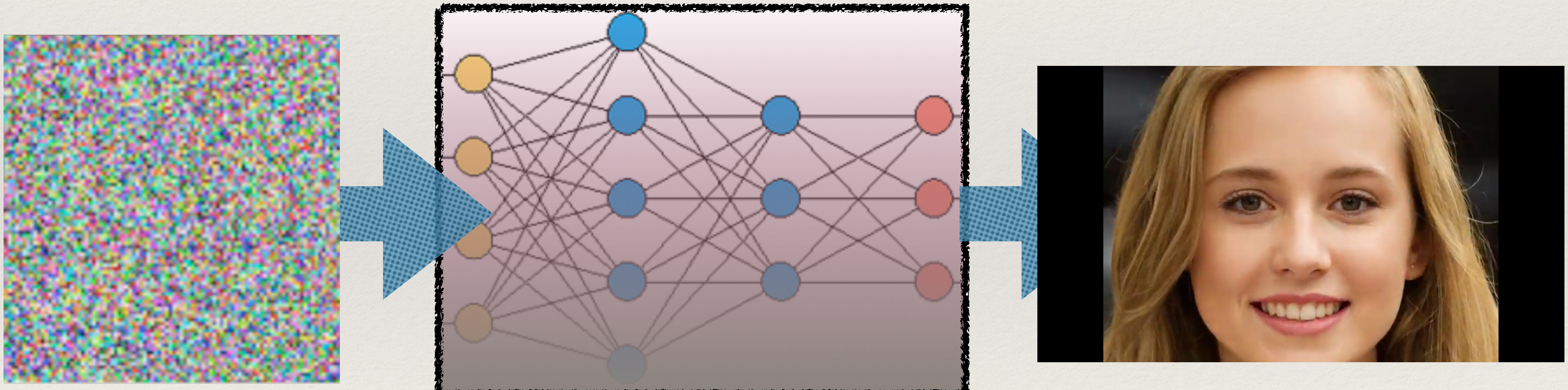
GANs (Generative Adversarial Networks)

and **VAEs (Variational AutoEncoders)**

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

Here, human faces

STEP 3- CREATE NEW POSSIBILITIES



gif con latent space

Random noise generate deformations in the output, leading to new *people*

Human surrender?

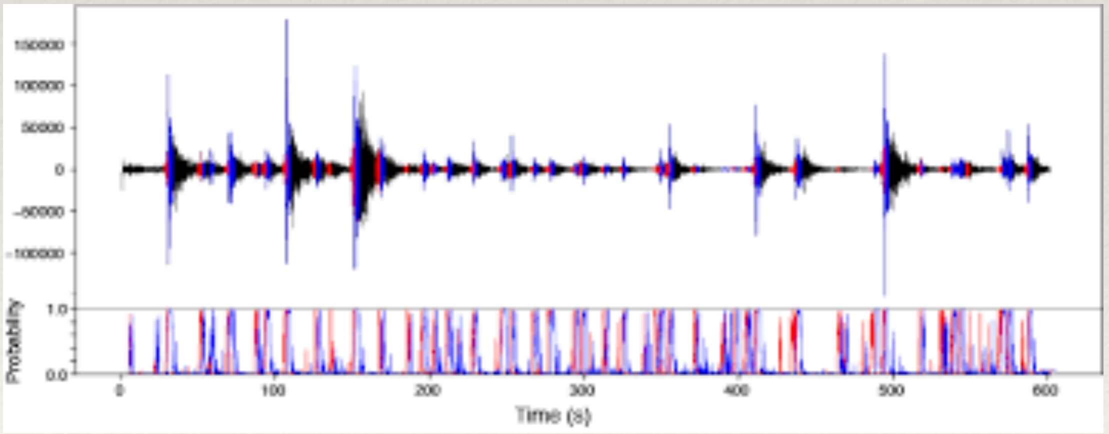
What's wrong with blackboxes?

Only open if a disaster happened

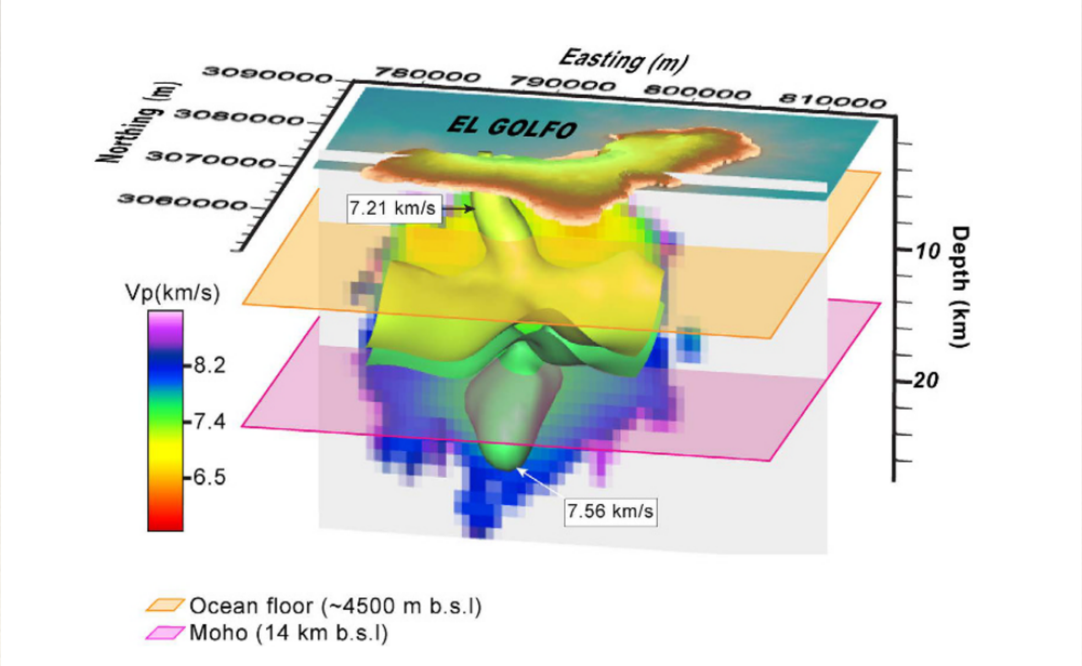
If it works, why fix it?
DNN is very powerful, in a way that can be quantified and tensioned against human performance or other techniques

Example: collaboration with Seismicity experts

Automatic detection of Earthquakes and phase picking

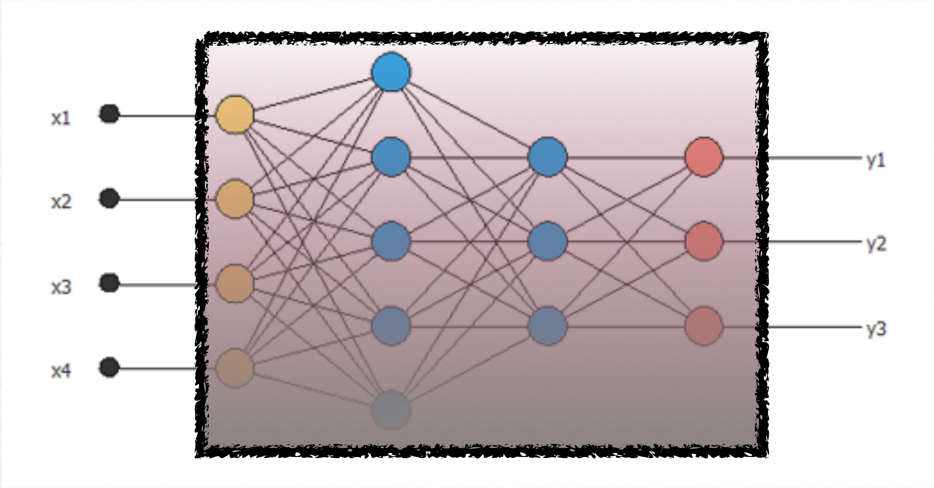


Tomography



Institut de Ciències del Mar + Universitat de València

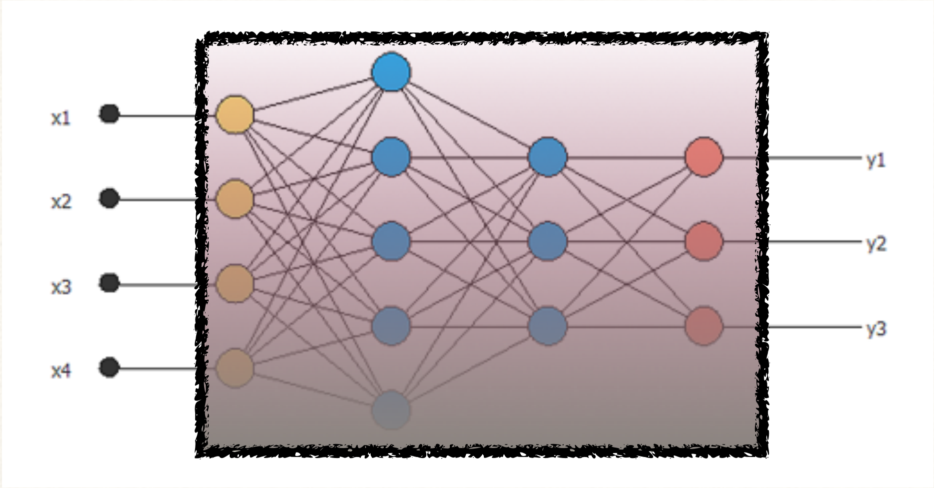
Garcia et al, *Seismological Research Letters*



What's wrong with blackboxes?

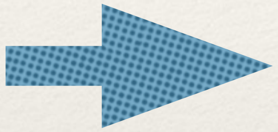
If they do work, and help solve problems?

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*

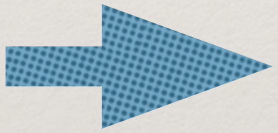


What's wrong with blackboxes?

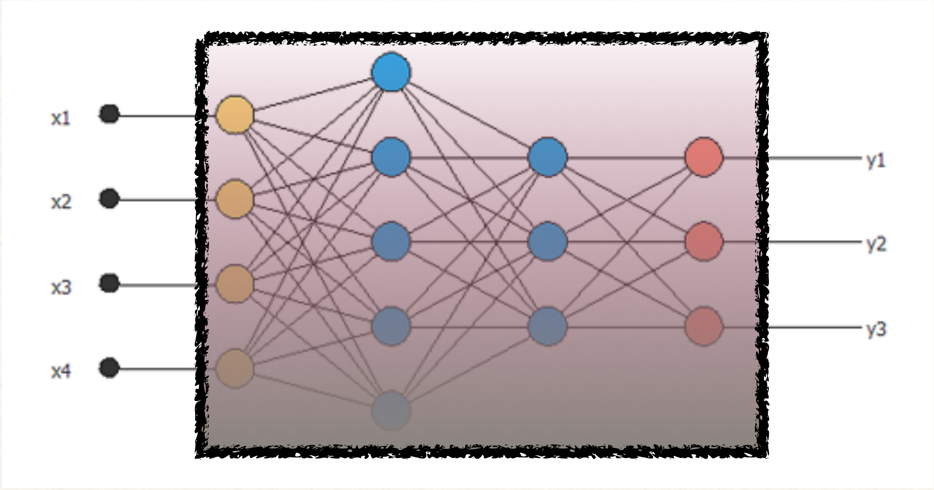
If they do work, and help solve problems?



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

Looking under the hood

with symmetries

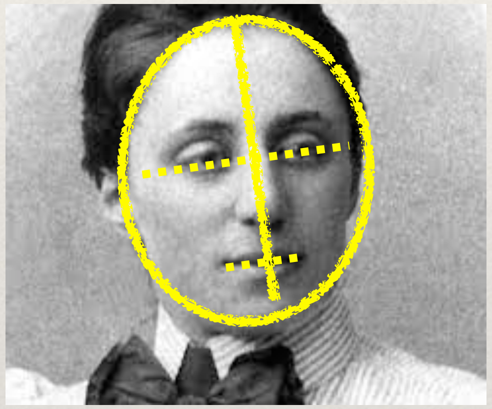
Symmetry is a key concept in Physics

not just as a simplification method or connection with other problems

deeper level: Laws of Physics, understanding of forces, stability…

Symmetries can help with **Machine Learning** problems

e.g., CNNs and data augmentation



The **concept** of symmetry is part of our shared **human appreciation**

We asked ourselves:

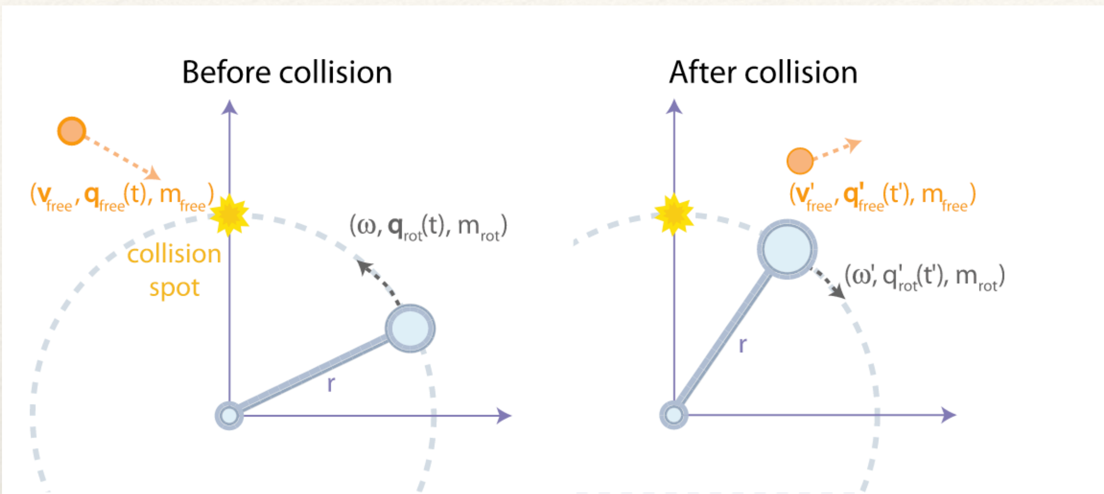
Can Machines Deep-Learn symmetries?

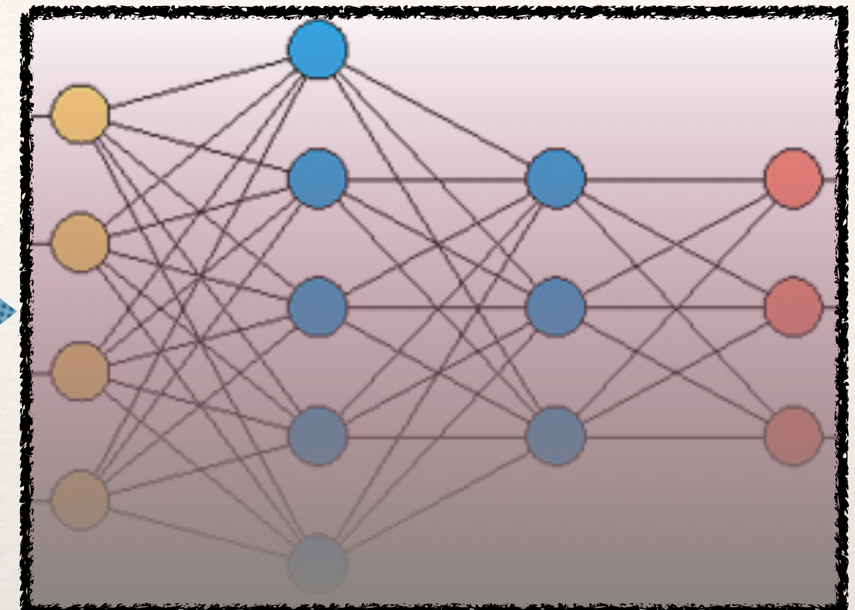
in which ways? and what could we use this for?

Do AI's understand *concepts*?

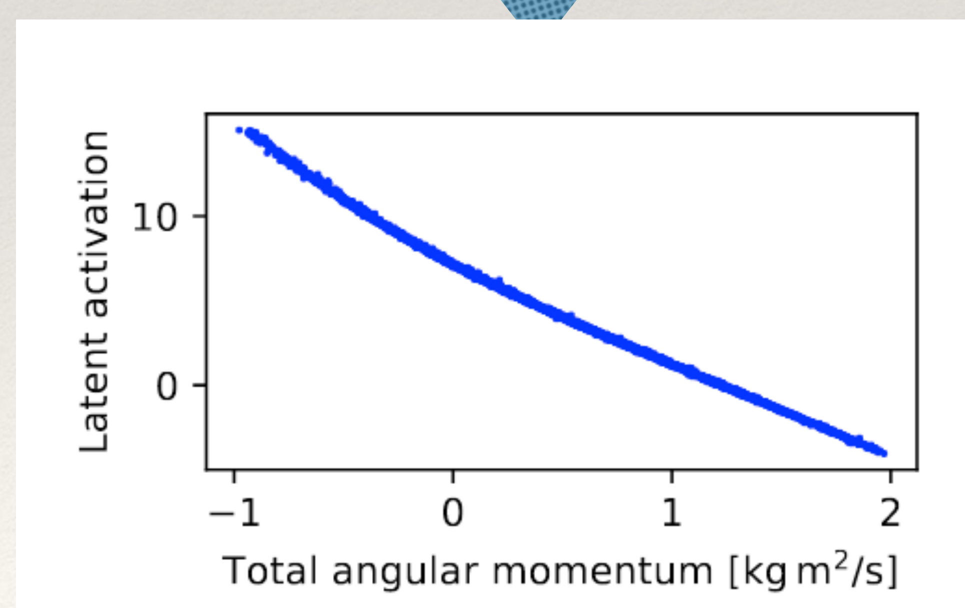
EXAMPLE- CONSERVATION LAWS

Iten et al *Phys Rev Lett*





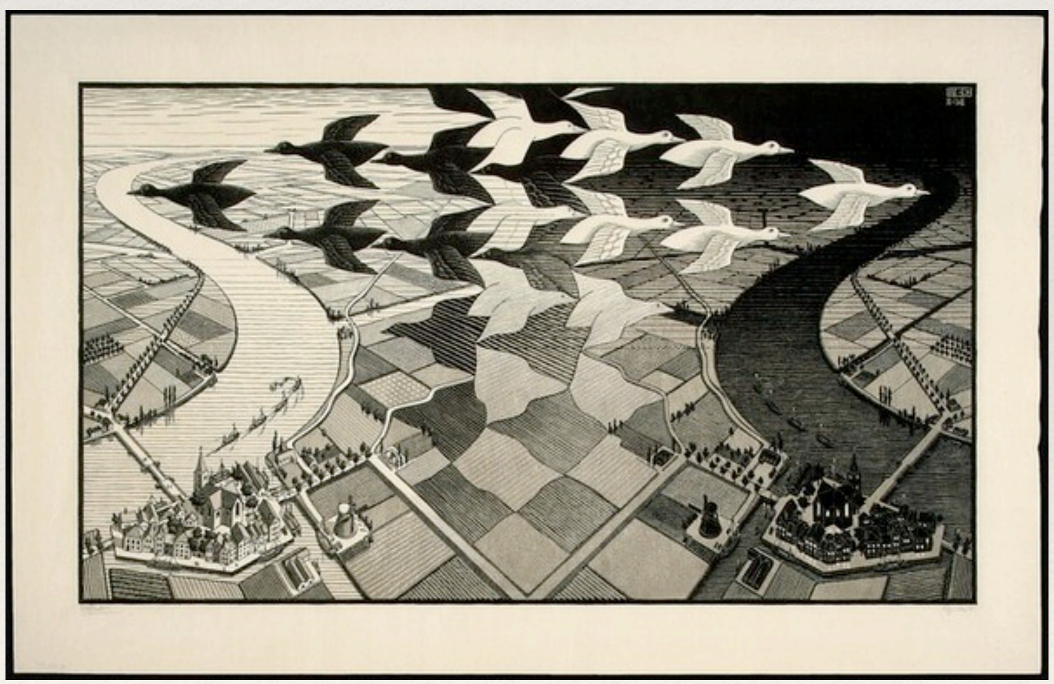
They trained a VAE with many collision examples with input of just kinematic variables



The trained algorithm had somehow *learned* that the concept of angular momentum is important and was storing it in one single neuron

We know the NN is realising higher level features of the data
it seems to *somehow* realise of the presence of a conserved quantities

what about symmetries? and even more,
what if the potential had no symmetry, or was only approximate?
what if we wanted to learn about symmetries in datasets that have no straightforward interpretation?



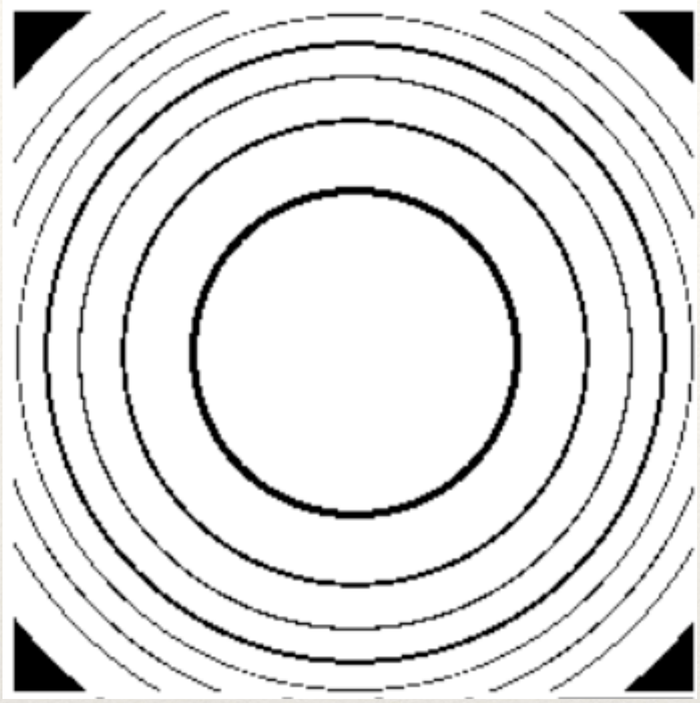
We asked ourselves:

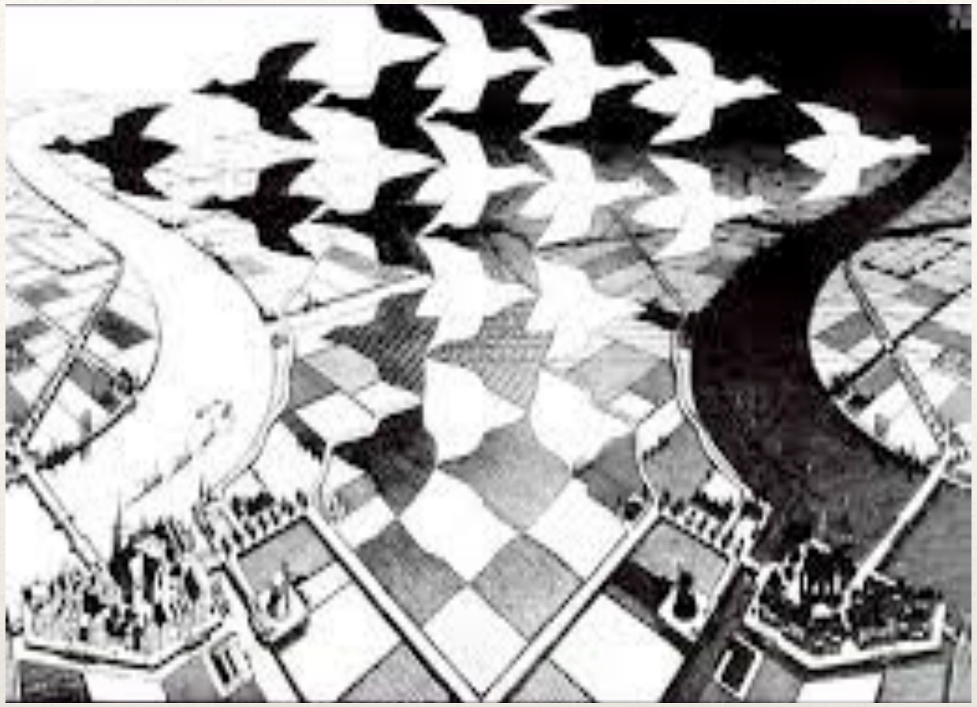
Are there ways to learn about symmetries which detect no symmetry or approximate levels of symmetry?
and that can be applied to a wide range of situations?

We needed a very general procedure

We had to start with something else,
a simpler representation

an image with only two colours





and a universal task:
try to learn as much as possible from this image

dataset = (x , y , 0/1)

and train a FCNN to learn to reproduce the image

then we can ask whether, while learning every detail of the image,
it did realise there was some level of symmetry

To train the FCNN, we build a dataset made of **Physics templates**

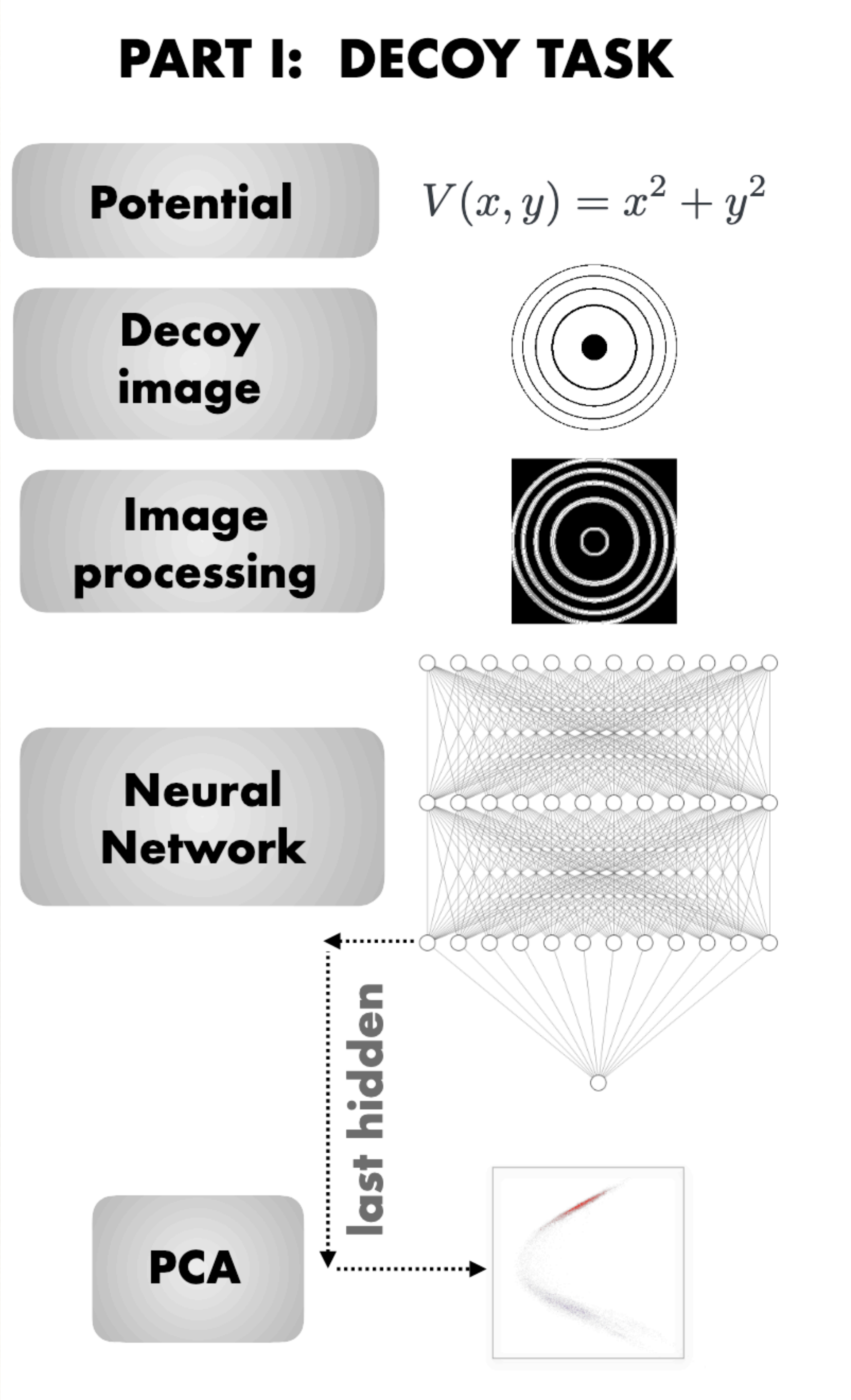
where we know the symmetry properties,
but this information is **not** known to the NN

We pay attention to not **overspecialise** in the physics potentials
FCNN is not allowed to **overfit**, so that it may be more prone to identify the symmetry

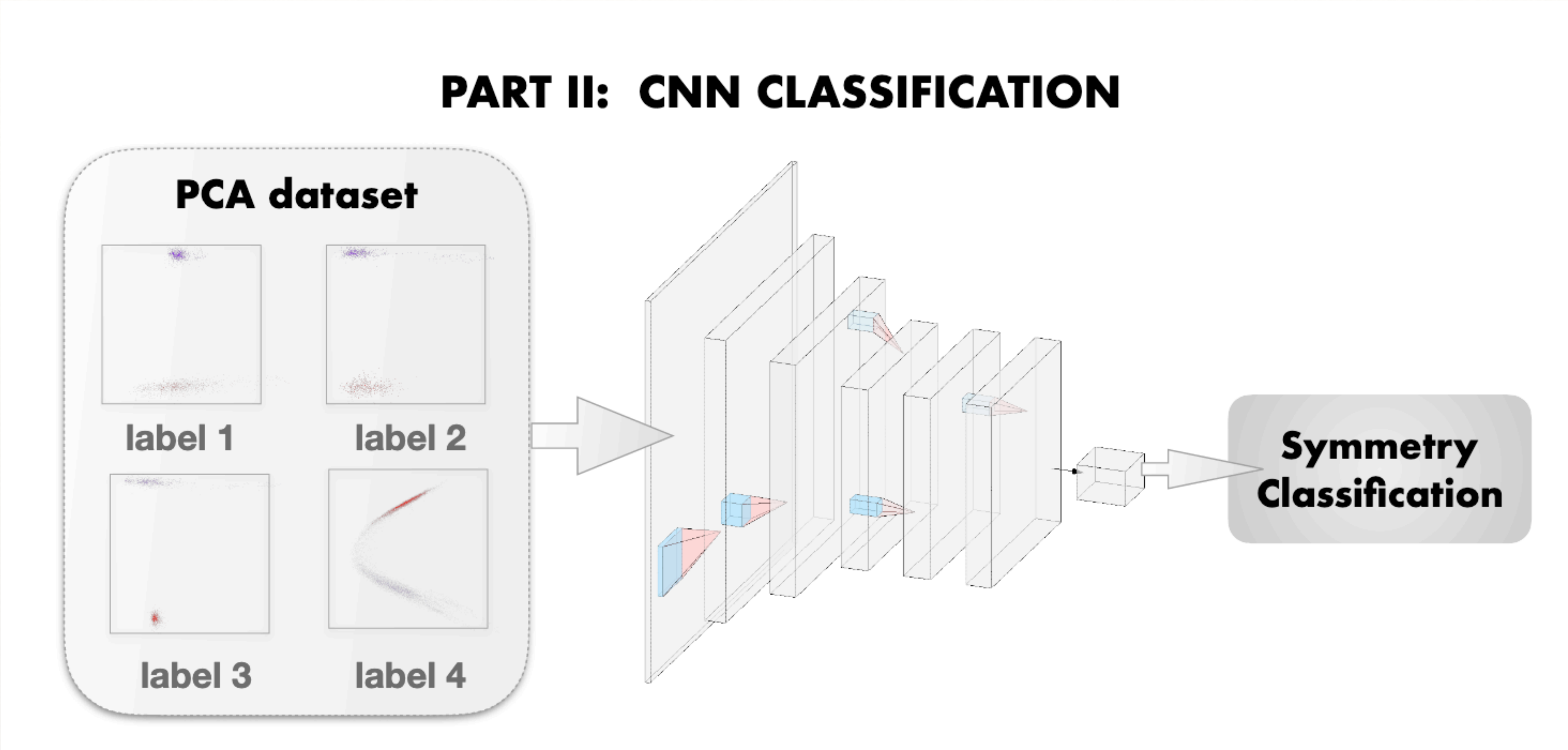
we then get the PCA image from the last hidden layer

at first sight all the PCAs look different, changing
from run to run and from image to image…

Putting it all together



If the FCNN, while paying attention to reproduce the image,
has learnt that there was some symmetry,
the PCAs may encode this learning



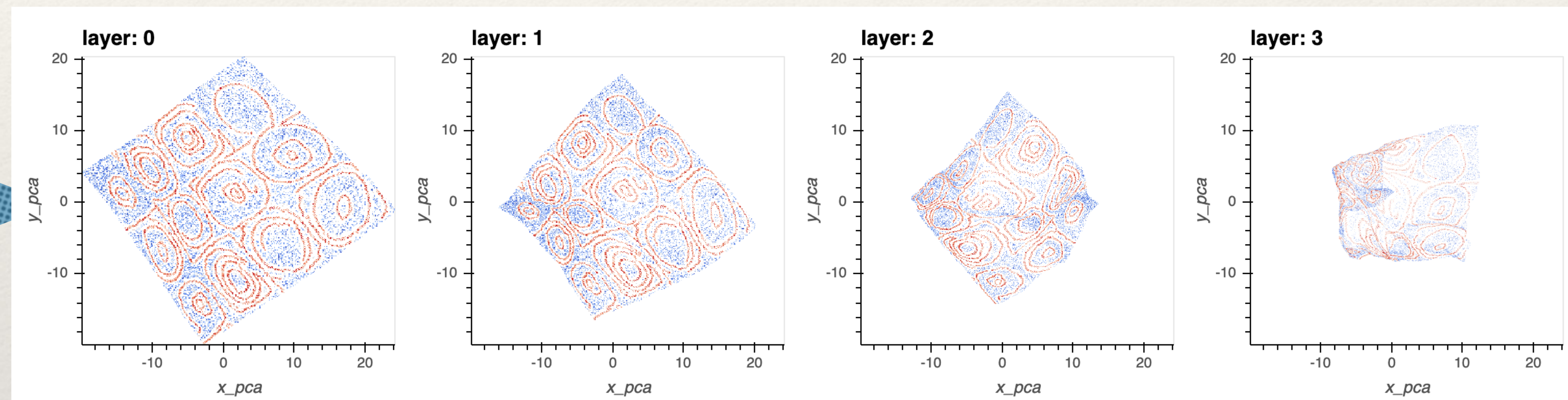
We train a CNN, using the PCAs and the physics labels,
to identify symmetries

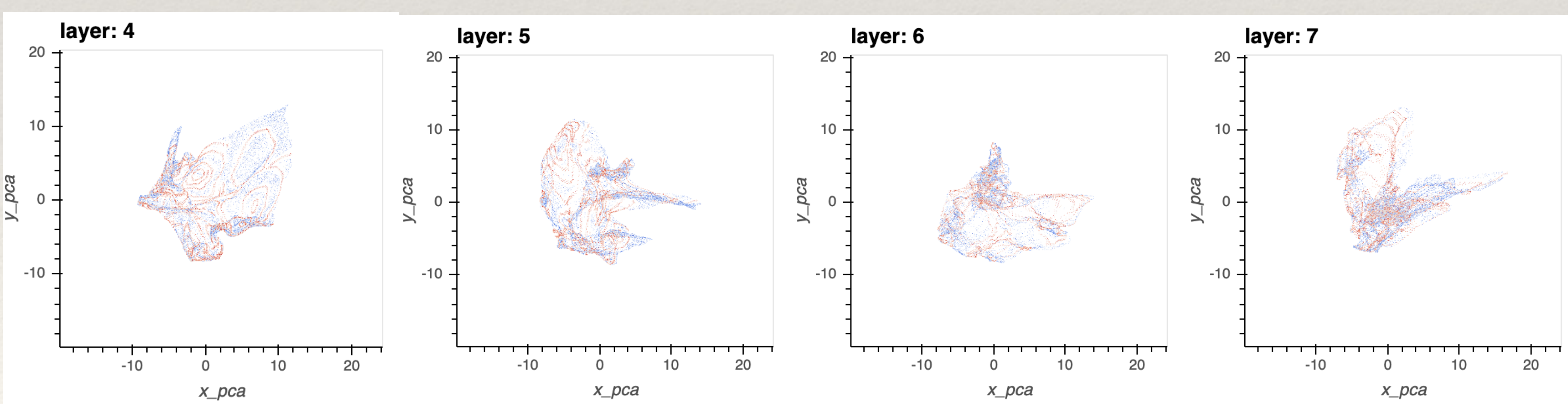
We find that the PCA- >symm classification does work,
the PCA does contain ***some encoding of the symmetry***

From Physics to Art

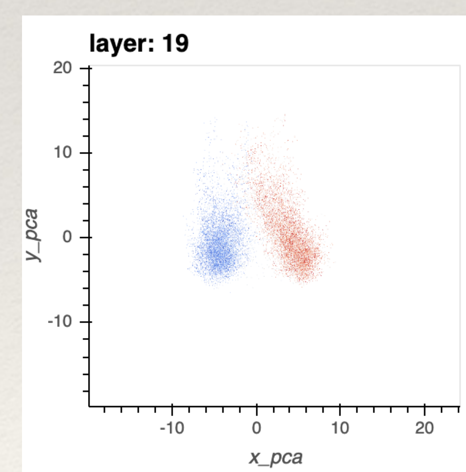
So far have developed an algorithm which takes simple images
and produces a **symmetry score**
It can be used in any type of 2D images:
a children's sketch, a painting, a photograph…

Some examples from Art:
first images-> sketches, then run the algorithm as if they were physics potentials

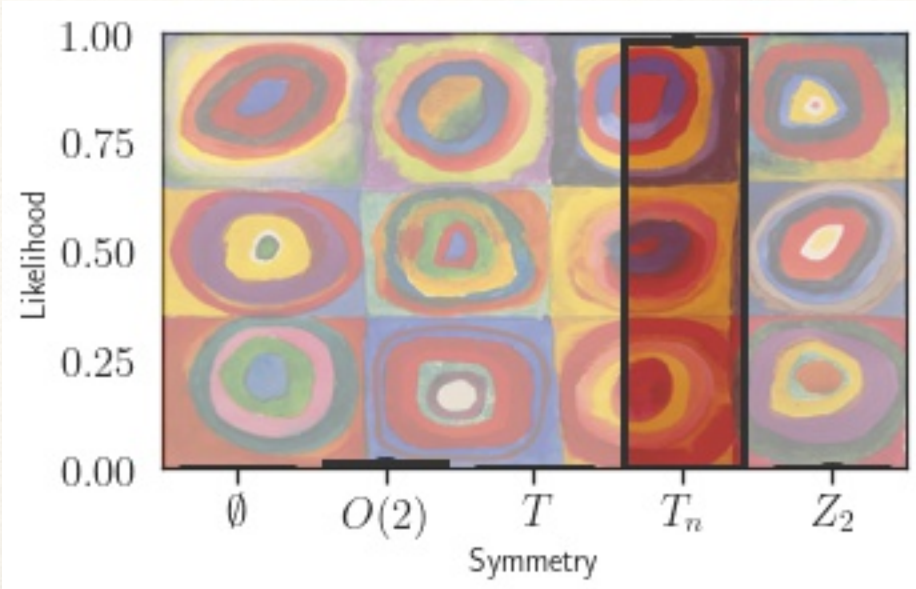


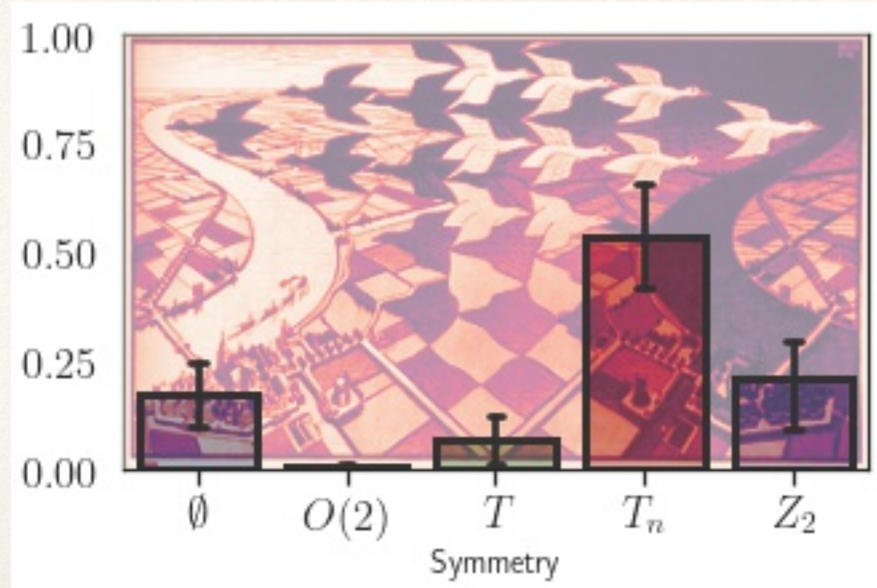


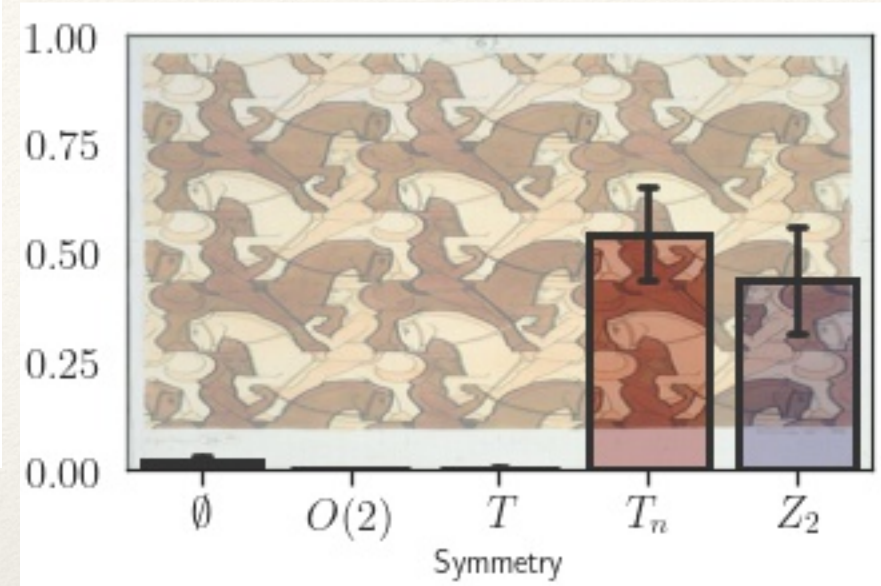
…

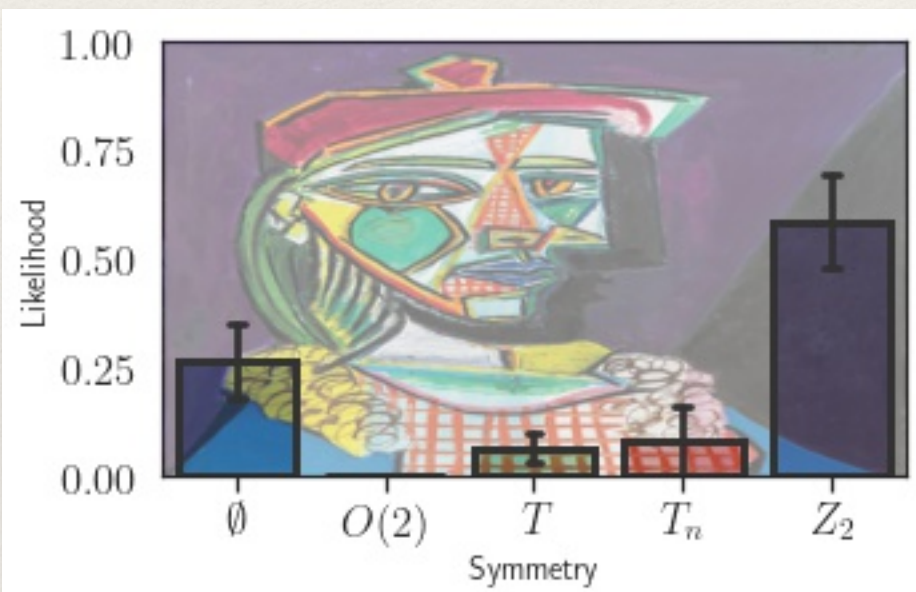


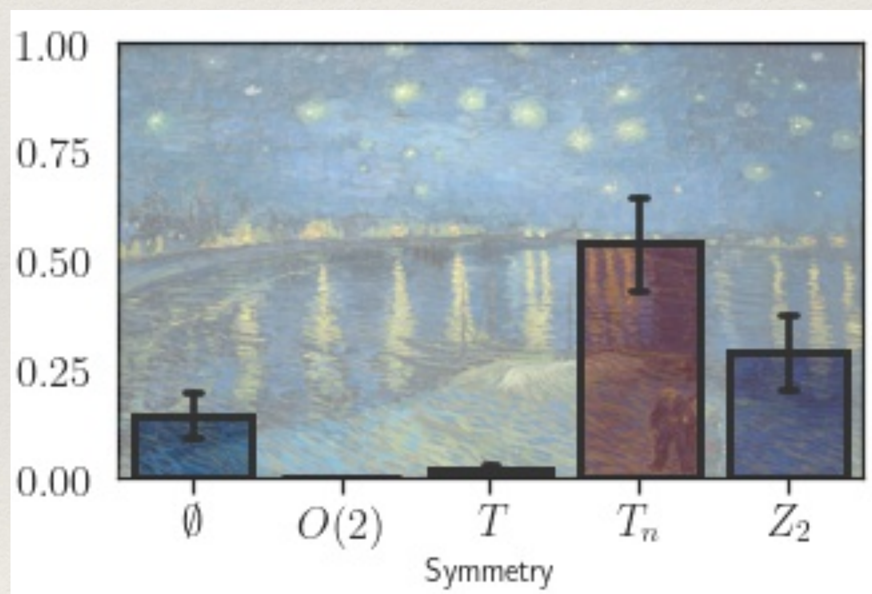
SYMM
SCORE?

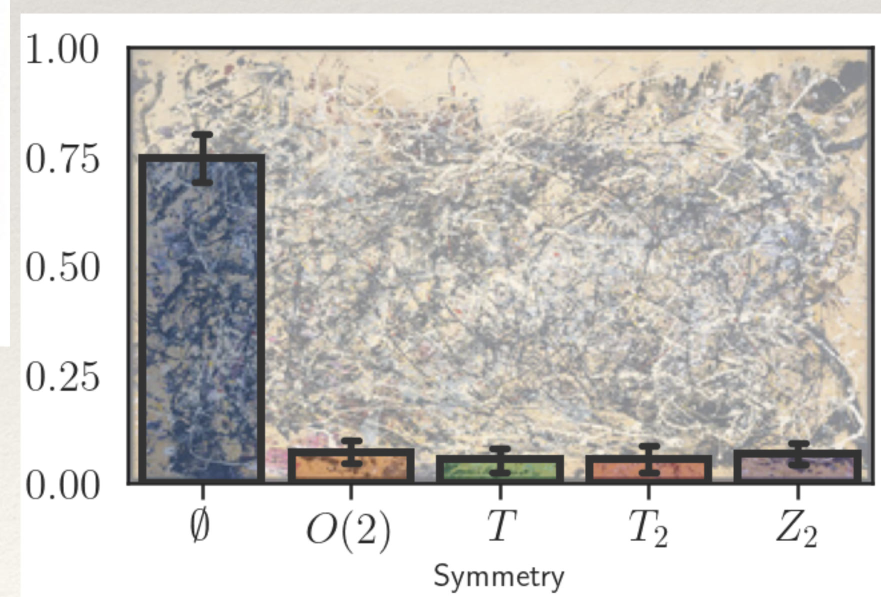










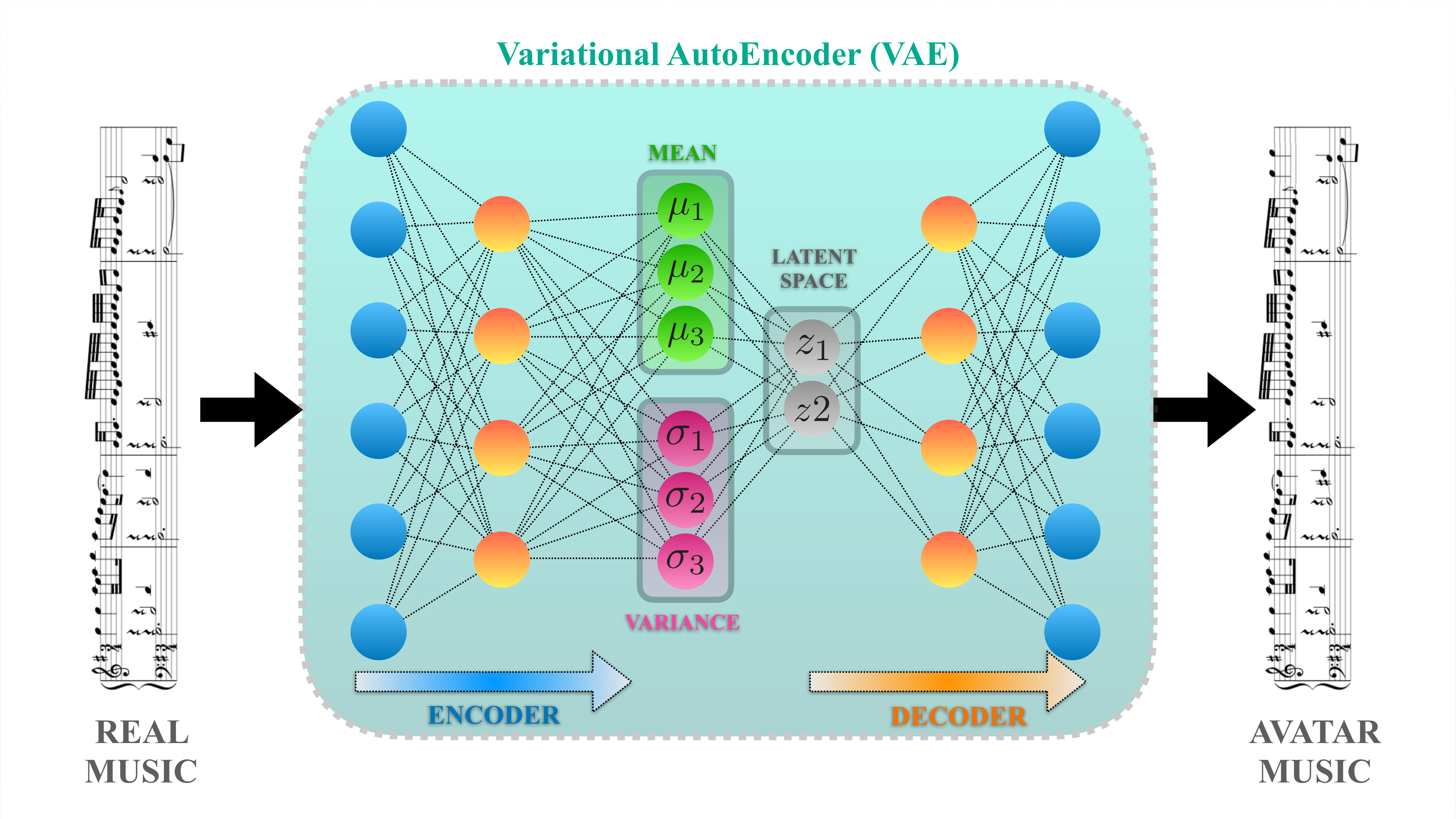


and many more, including children's drawings, fractals, photographs etc

One mor example: the AI *discovers* human concepts in music

coming out soon!

MAGENTA VAE trained on millions of music pieces



latent space of 500 neurons!

Q: did it learn anything non-trivial, or did it just memorize things?

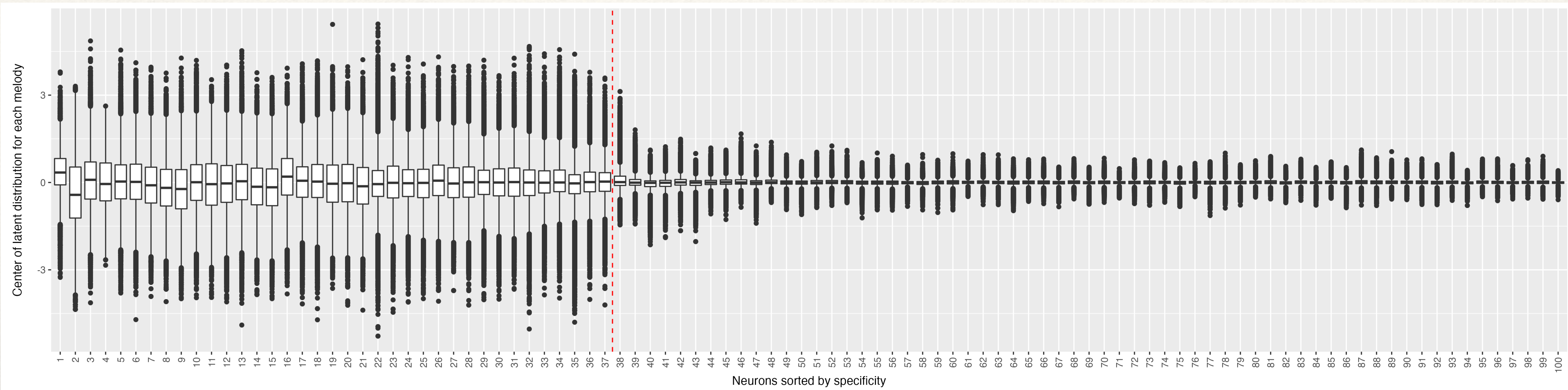
One mor example: the AI *discovers* human concepts in music

coming out soon!

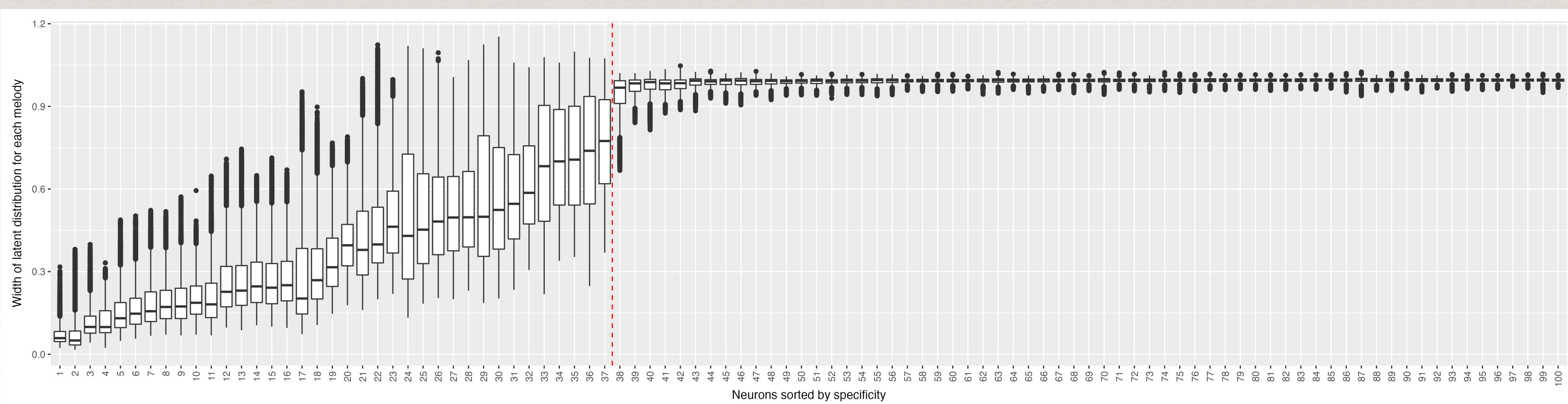
Reordering in the latent space by specificity, distribution

Example of short (two-bar) music

MEAN ACTIVATION



SIGMA

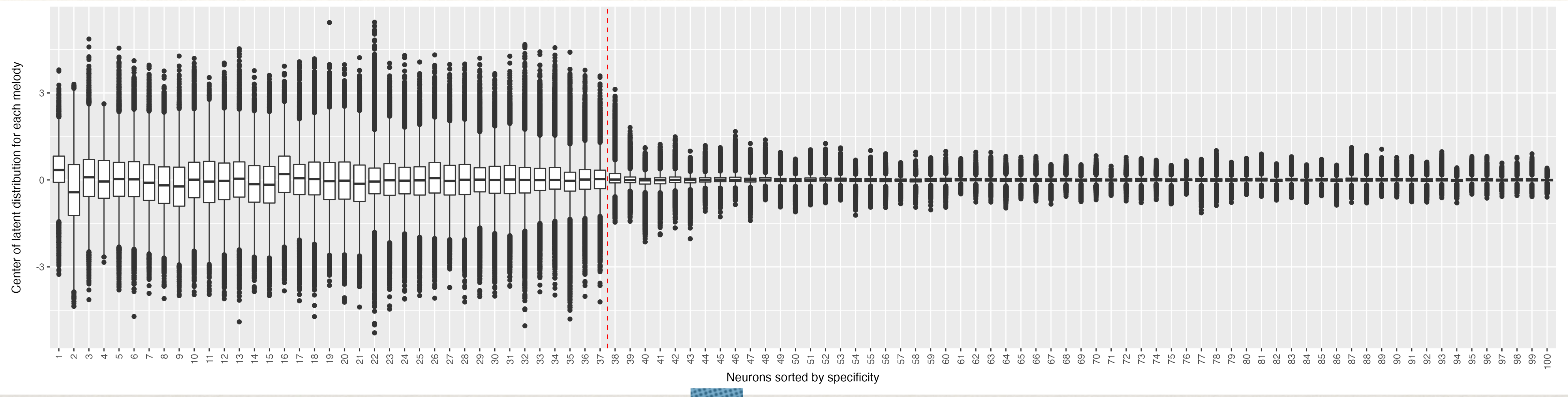


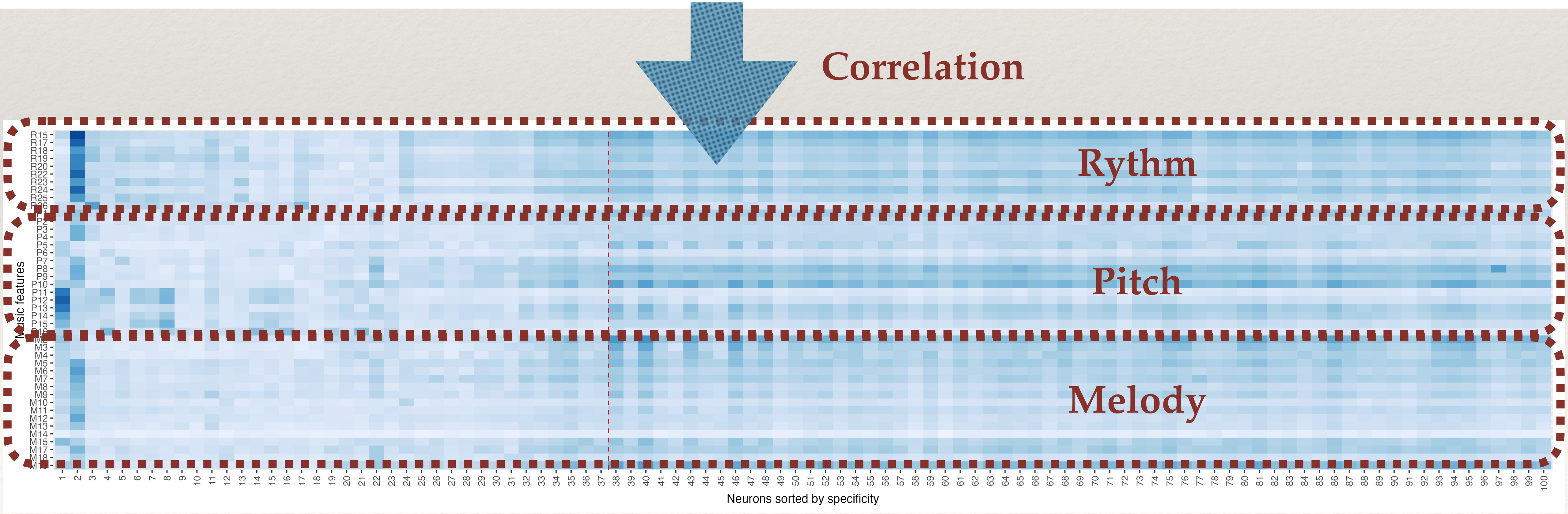
Theme neurons **Variation neurons**

One mor example: the AI *discovers* human concepts in music

coming out soon!

If we then look at how this AI representation relates to human quantities…

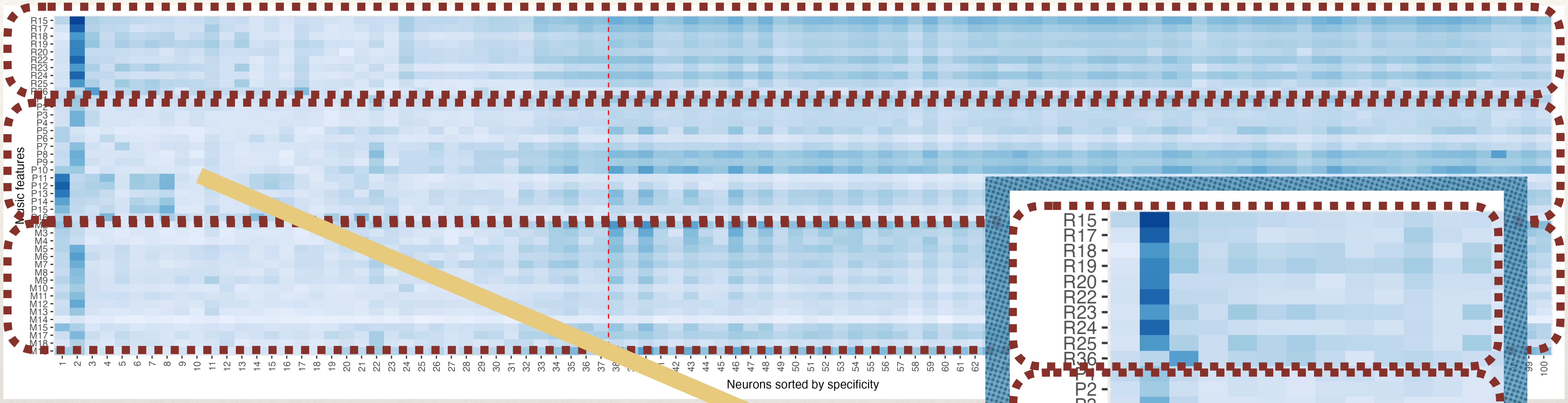




One mor example: the AI *discovers* human concepts in music

coming out soon!

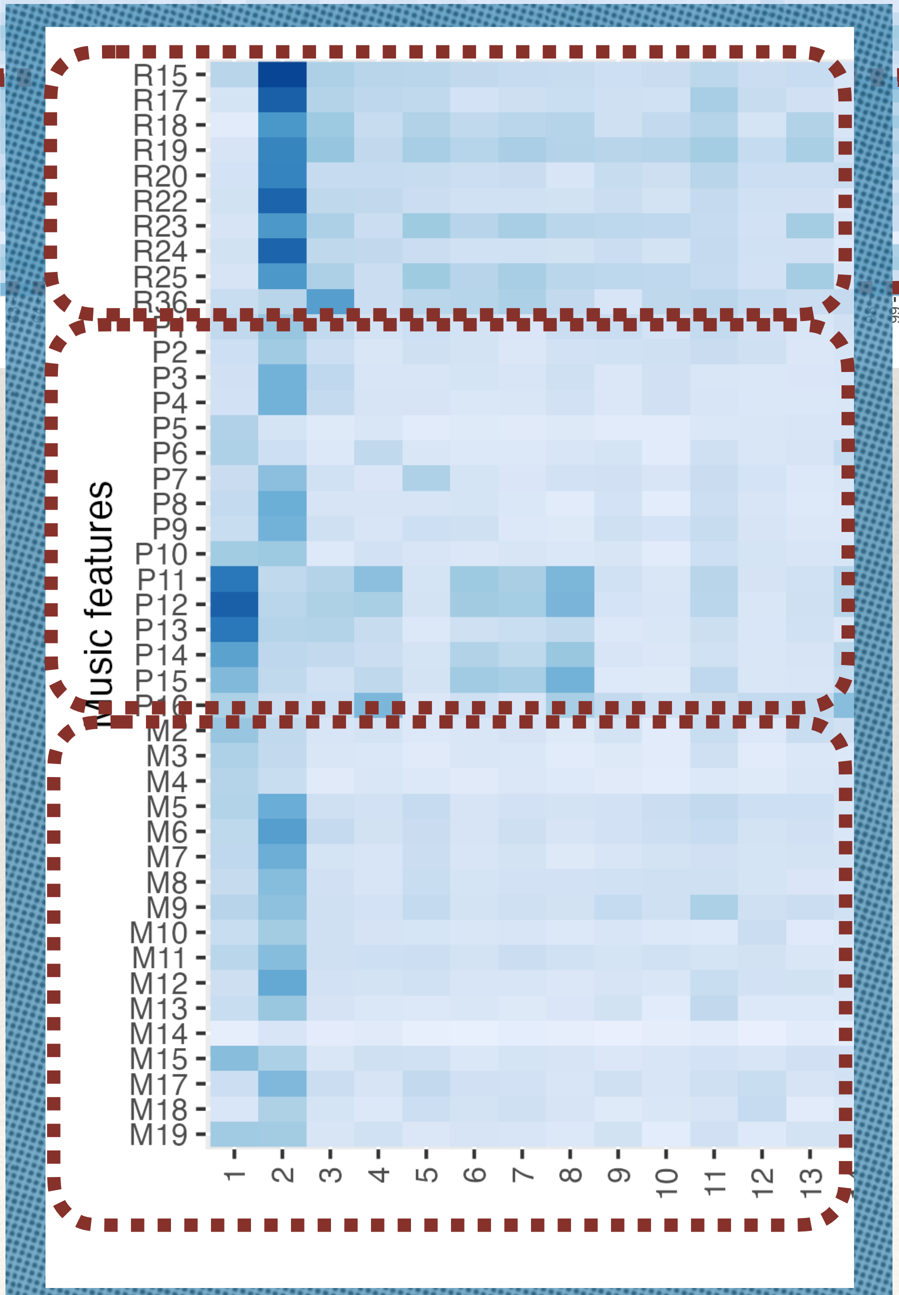
If we then look at how this AI representation relates to human quantities…



Only few neurons carry the meaning of
rythm, pitch
and, to a lesser extent, melody
(only 2-bar)

The AI self-organizes itself
matching those concepts

not just learning by heart



One mor example: the AI *discovers* human concepts in music

coming out soon!

e.g. EXCITATION NEURON 2

Summing up…

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

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

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

Through AI methods, there is interesting cross-pollination between our area and others

Summing up…

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

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

Through AI methods, there is interesting cross-pollination between our area and others

Yet a very efficient blackbox is not good enough for us, we try to *communicate* with the AI, to find ways to understand its inner workings

Summing up…

We are just starting to understand the applications of ML in Physics
They go beyond a mere iteration of our traditional statistical methods:
unsupervised methods, generative AI, reinforcement learning…

Through AI methods, there is interesting cross-pollination between
our area and others

Yet a very efficient blackbox is not good enough for us,
we try to *communicate* with the AI,
to find ways to understand its inner workings

Today we learned that an AI does identify and use symmetries,
even if only approximate, when *inspecting* an image (decoy task)
and this learning can be found in subtle features of the hidden layers

Summing up…

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

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

Through AI methods, there is interesting cross-pollination between our area and others

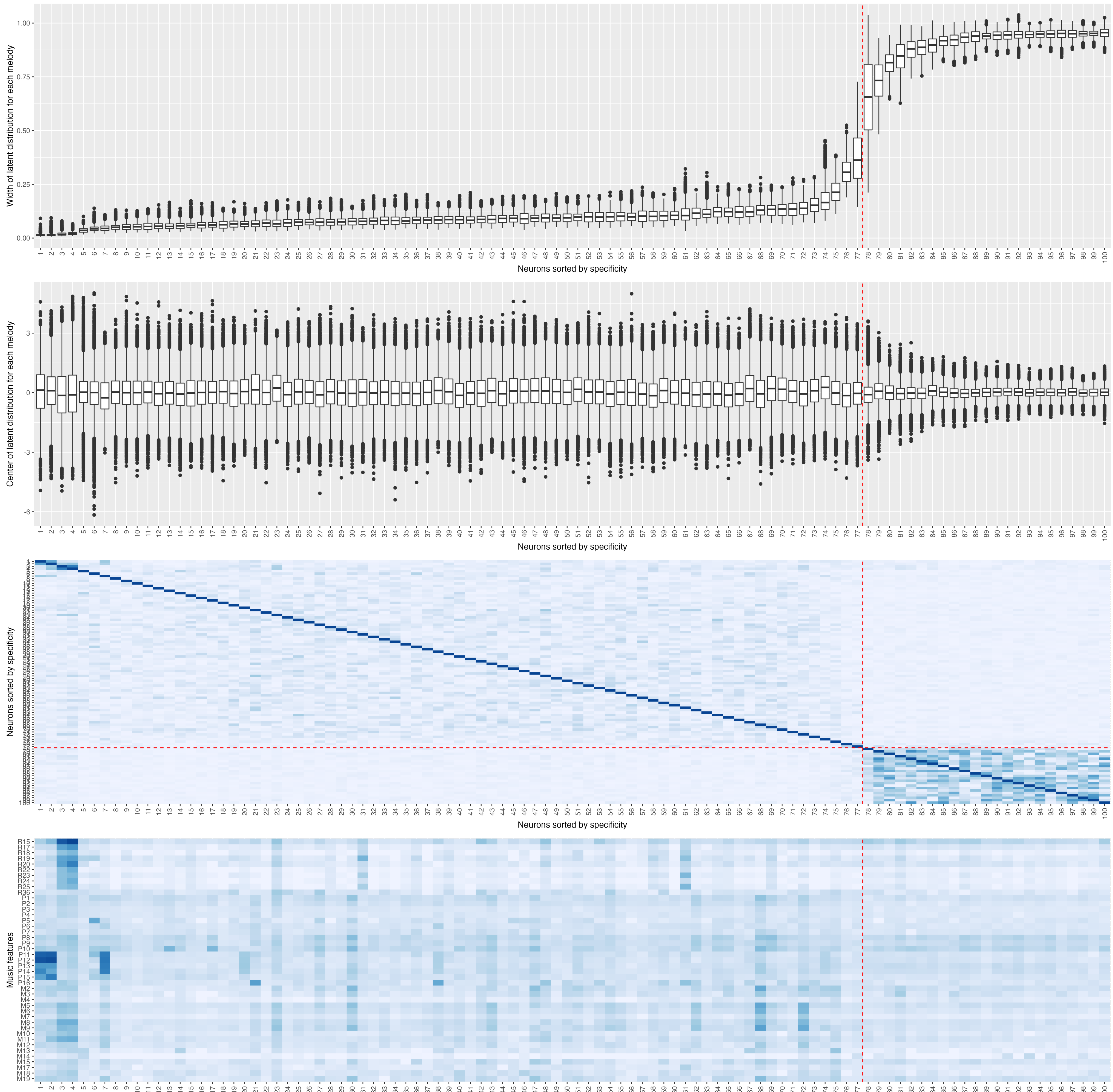
Yet a very efficient blackbox is not good enough for us, we try to *communicate* with the AI, to find ways to understand its inner workings

Today we learned that an AI does identify and use symmetries, even if only approximate, when *inspecting* an image (decoy task) and this learning can be found in subtle features of the hidden layers

We applied this method to Art, finding that it matches human intuition

Without being explicitly told, just by performing a *decoy* task the AI can discover that there is a concept (symmetry, rythm…) which helps better characterising physics potentials, music, collisions

Big Q: Can the AI discover something we haven't thought about?



Width of latent distribution for each melody
1.00
0.75
0.50
0.25
0.00
Neurons sorted by specificity
Center of latent distribution for each melody
3
0
-3
-6
Neurons sorted by specificity
Neurons sorted by specificity
Neurons sorted by specificity
Music features