

Should we use **Machine Learning** to get **Geostatistical realizations**?

Let's try to see what generative ML can and cannot do!

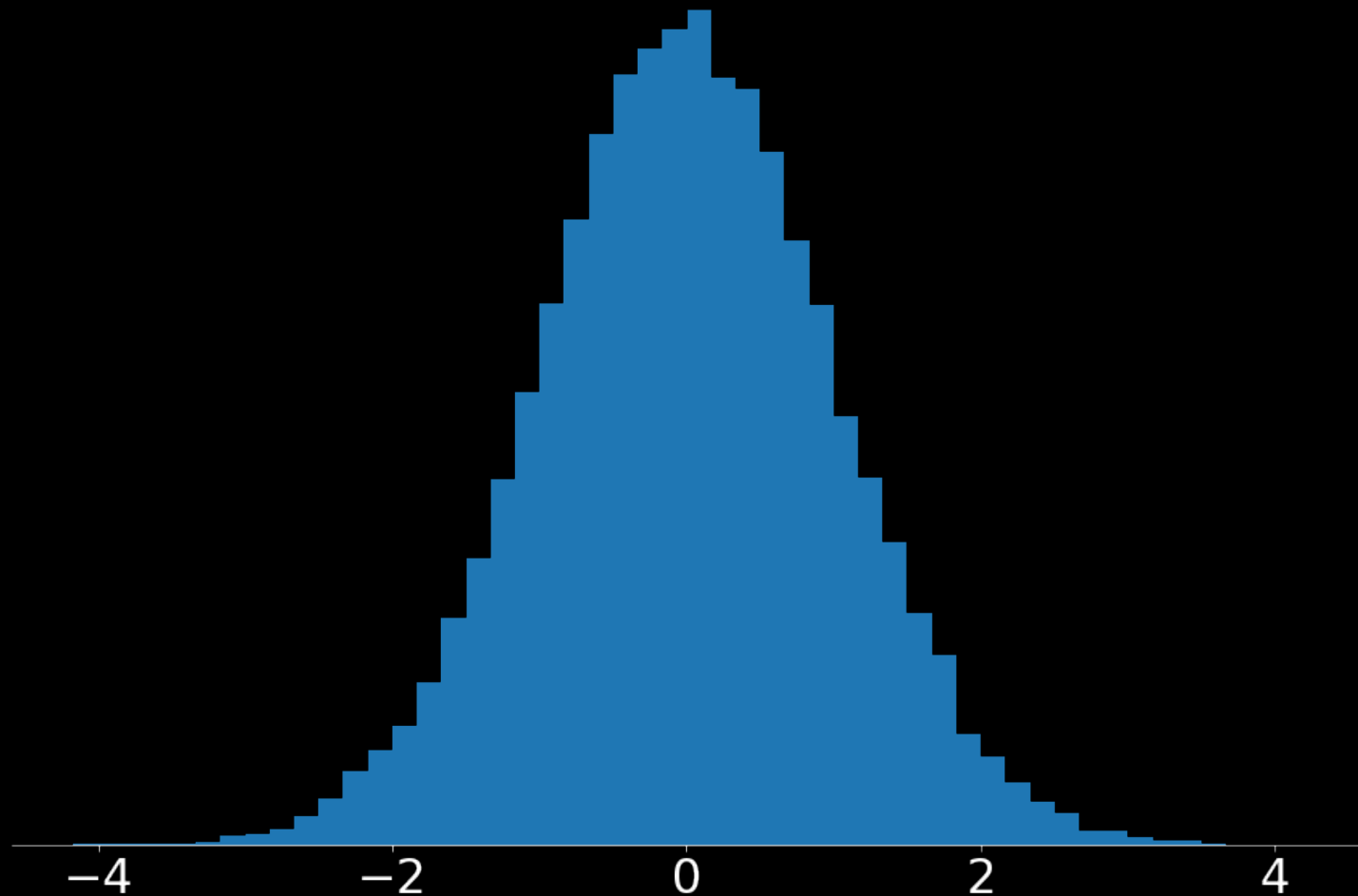


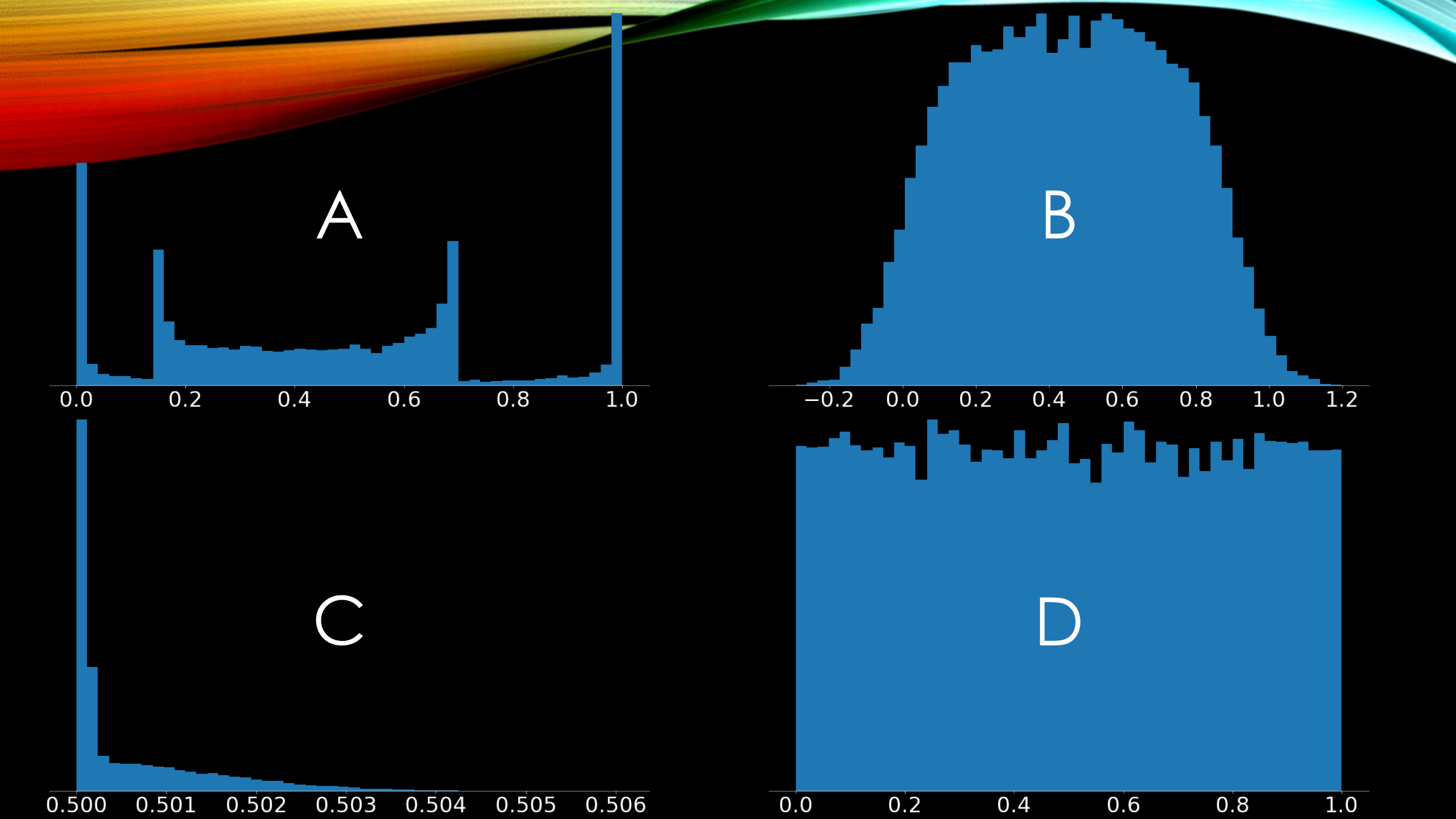
**Utrecht
University**

Mathieu Gravey – Utrecht University - NL
1st September 2022,
Nancy

A simple GAN

$$\mathcal{N}(0,1) \rightarrow \mathcal{U}_{[0,1]}$$





Let's discuss the cornerstones of Geostatistics

Stochastic realizations ?

- Individually
 - Possible
 - Conditioned (if exist)
- As a set
 - Unbiased
 - Equiprobable

Can we do some tests?

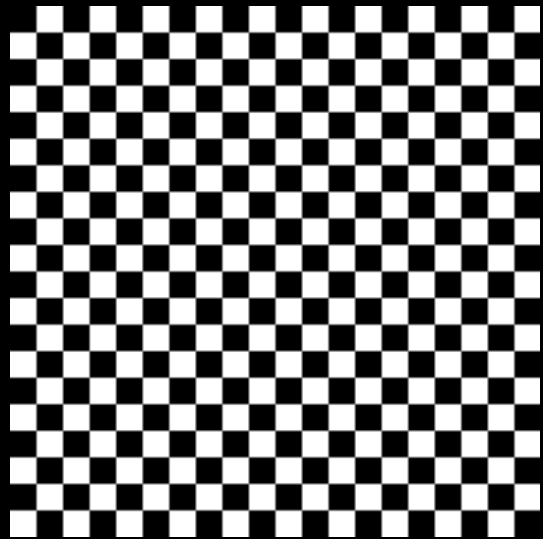
Sure!

➔ exclusively on known cases

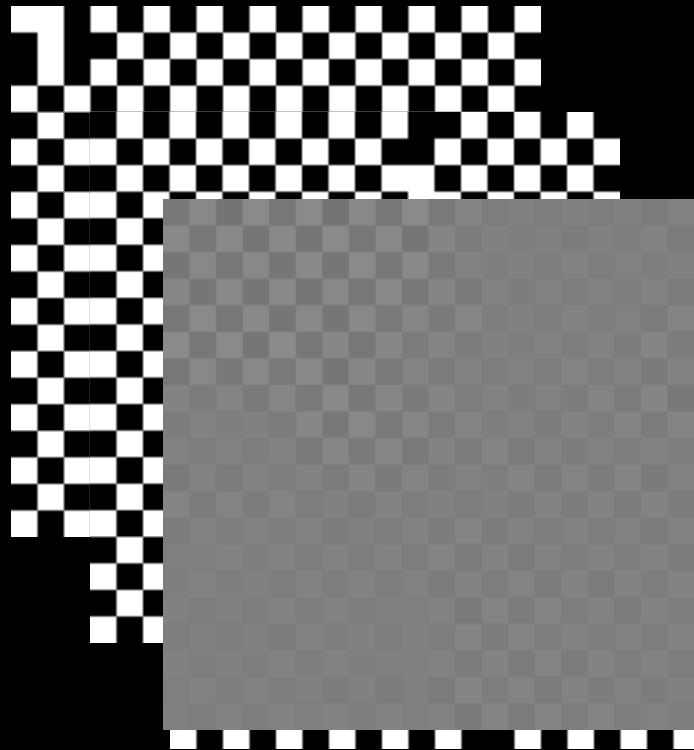
Example:

the marginal should be respected at each pixel in cases of unconditional simulations

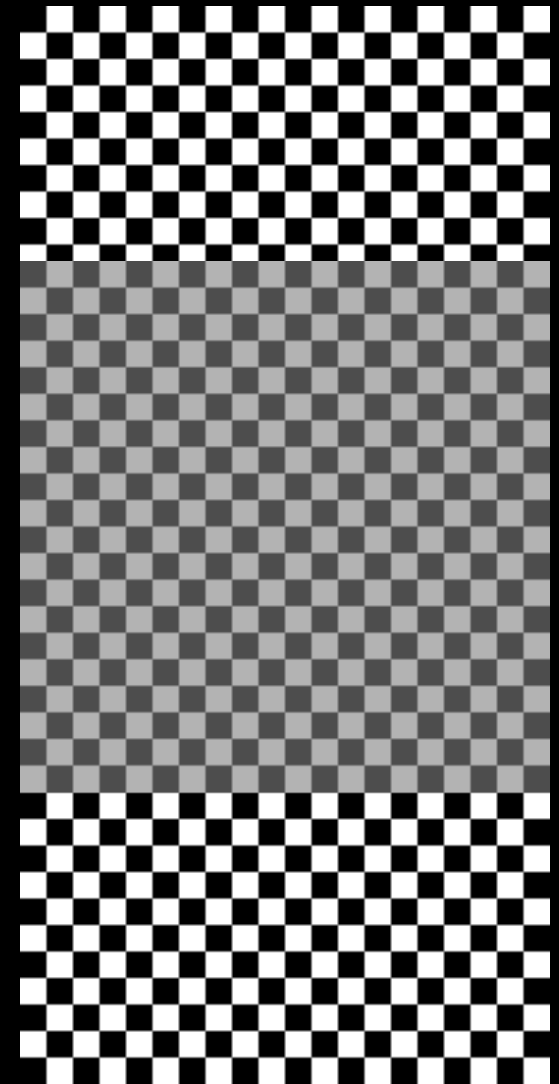
Let's start with a chessboard



TI



MPS



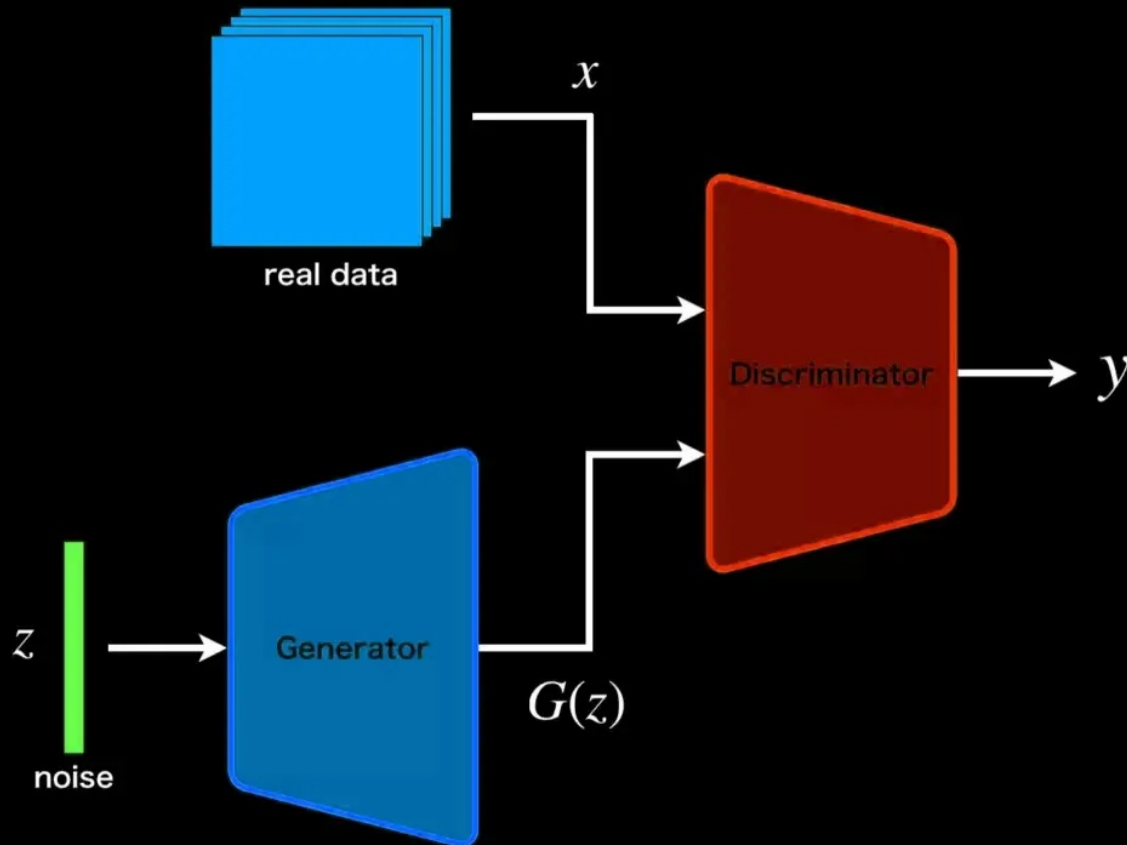
How GAN work

Let's explore constraints on GANs

1. The quality

2.

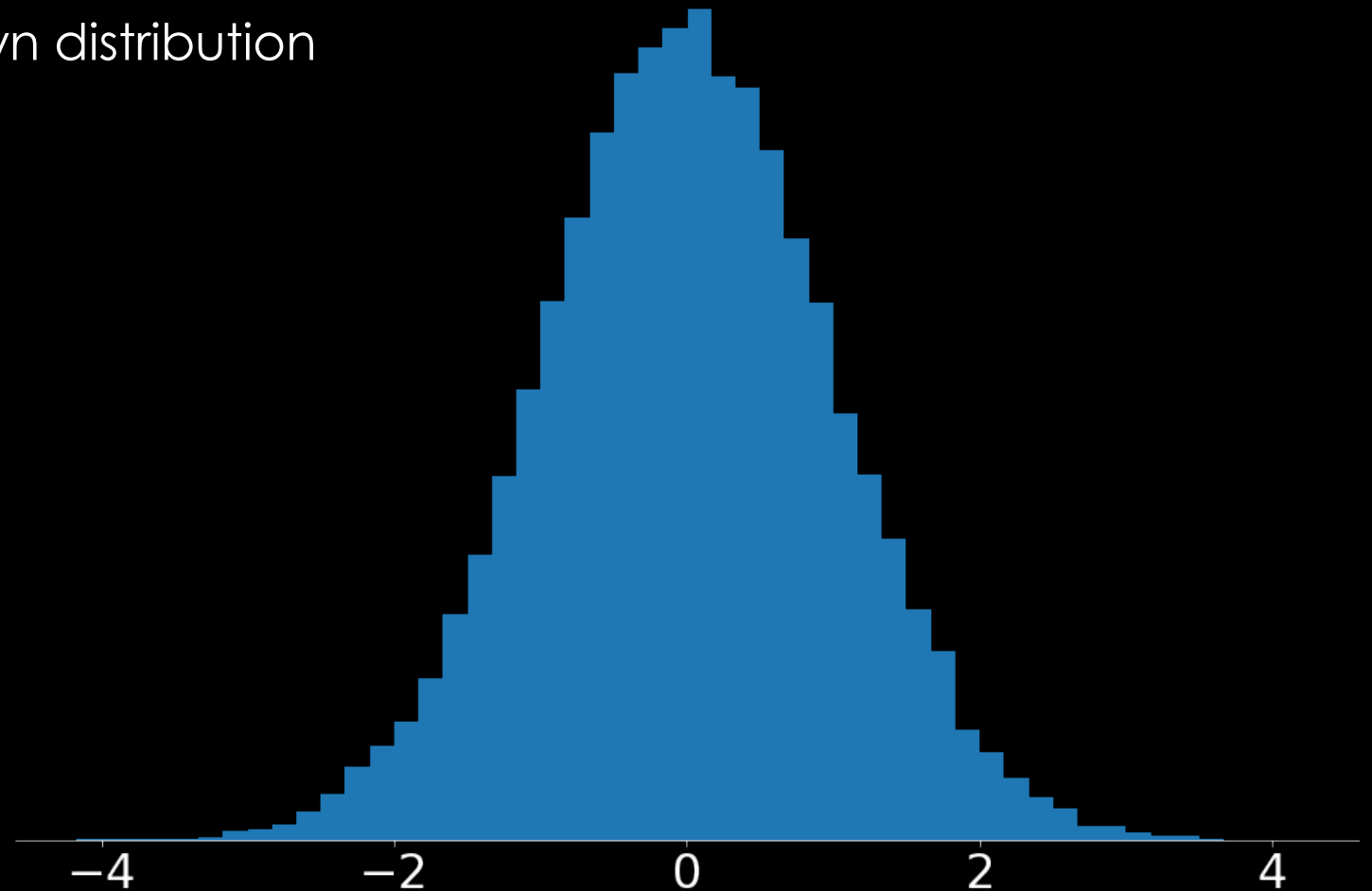
No constraints on the set of images.

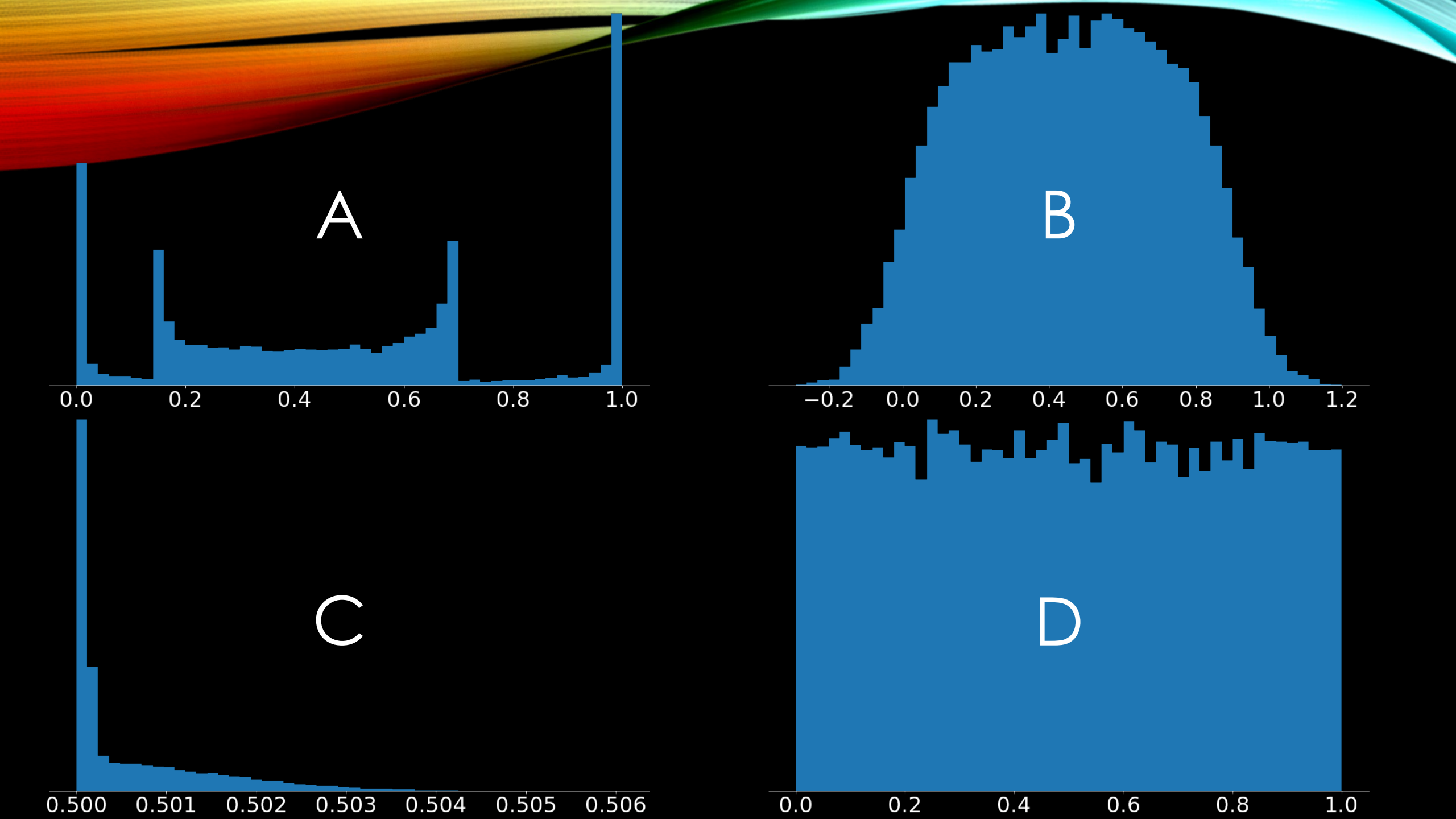


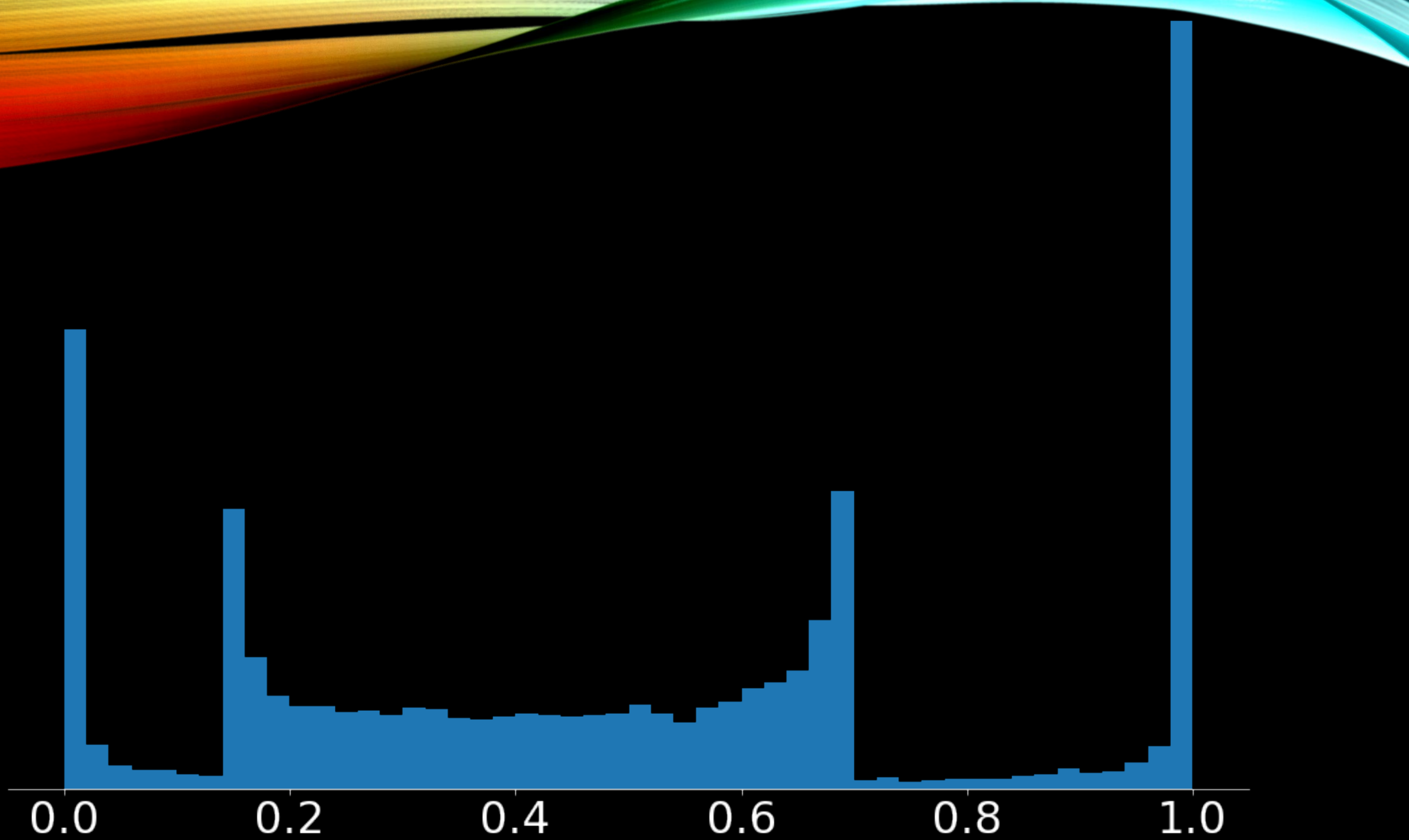
Let's analyze a situation

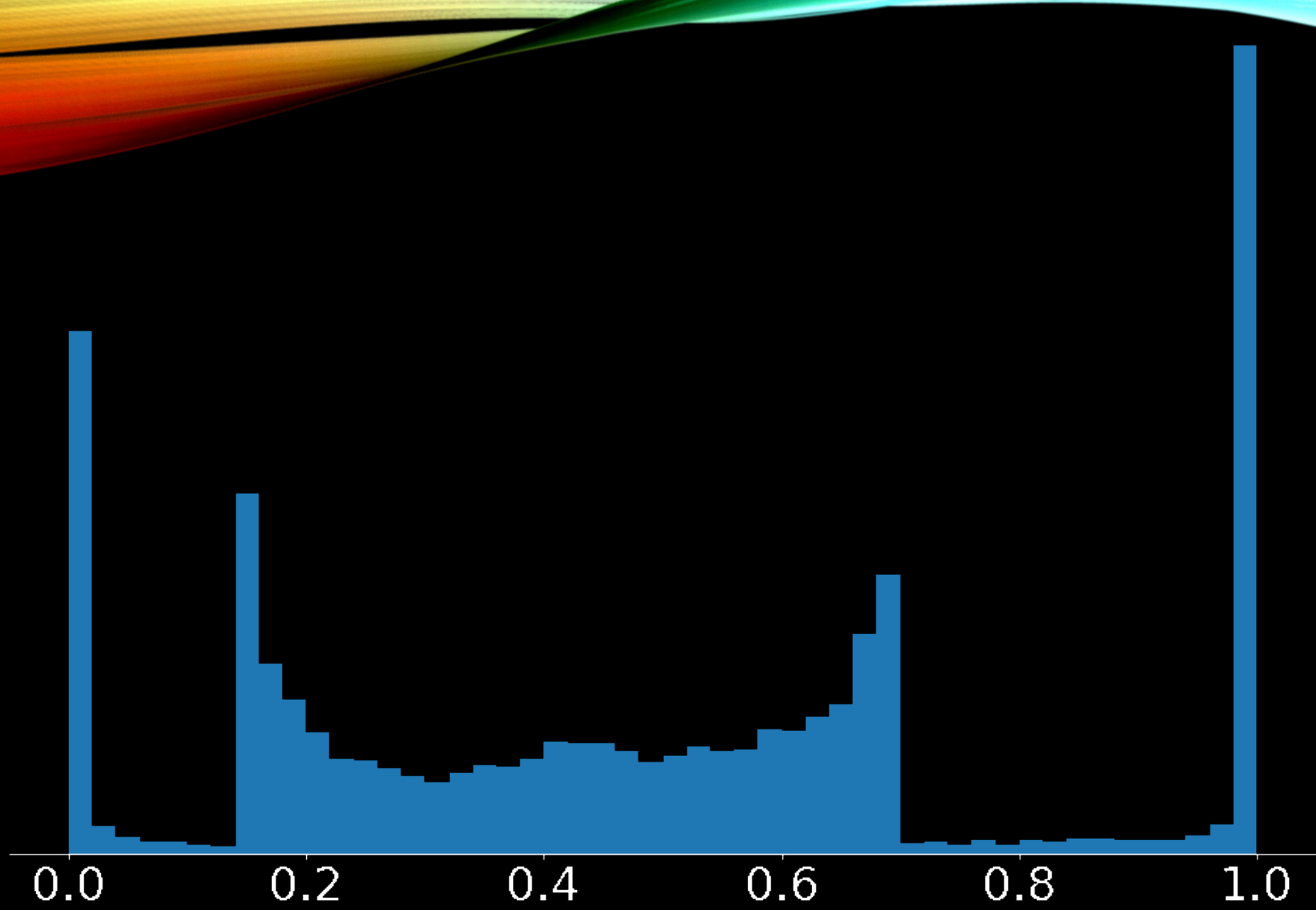
Take a simple example with a known distribution

$$\mathcal{N}(0,1) \rightarrow \mathcal{U}_{[0,1]}$$

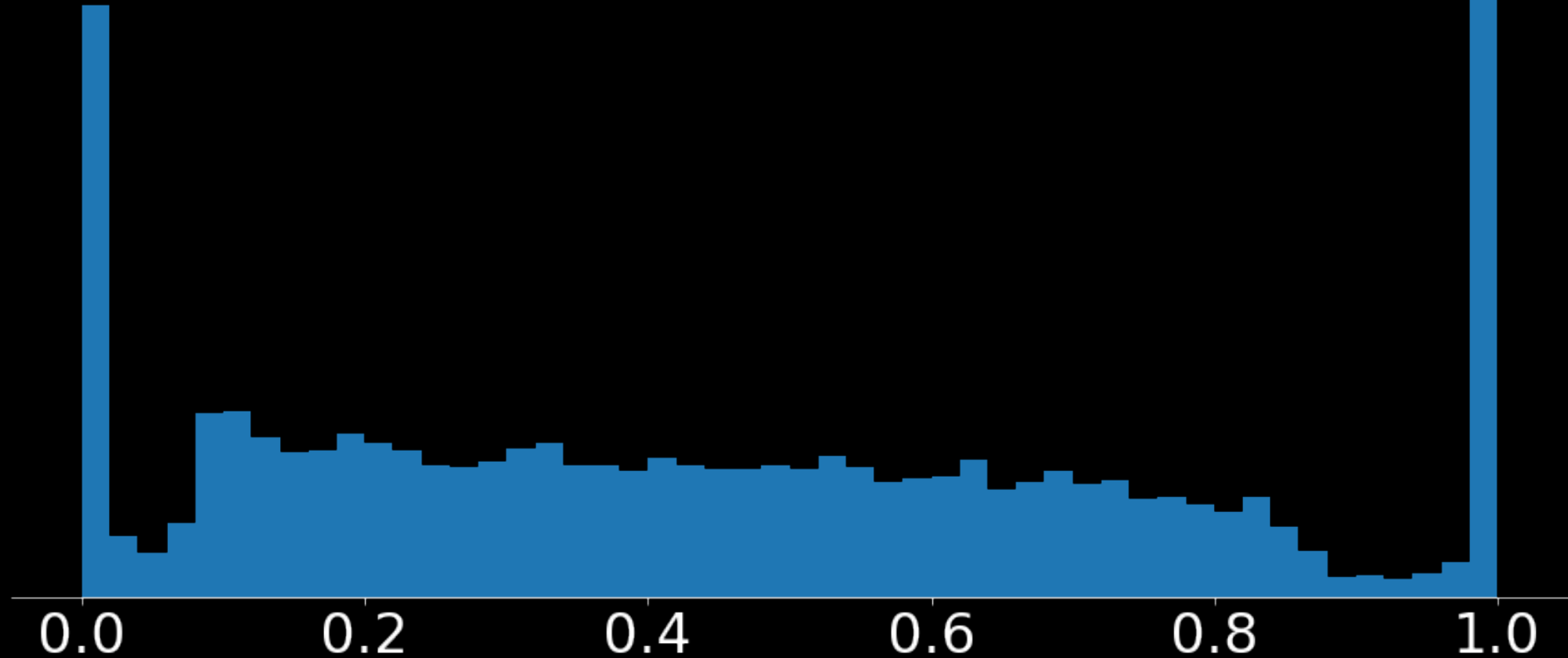






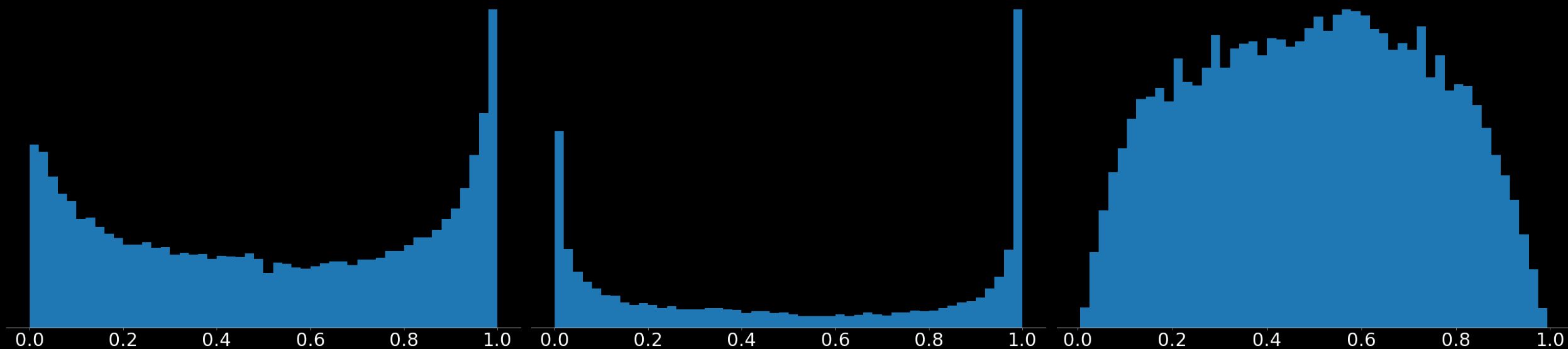


A training == a realization ?

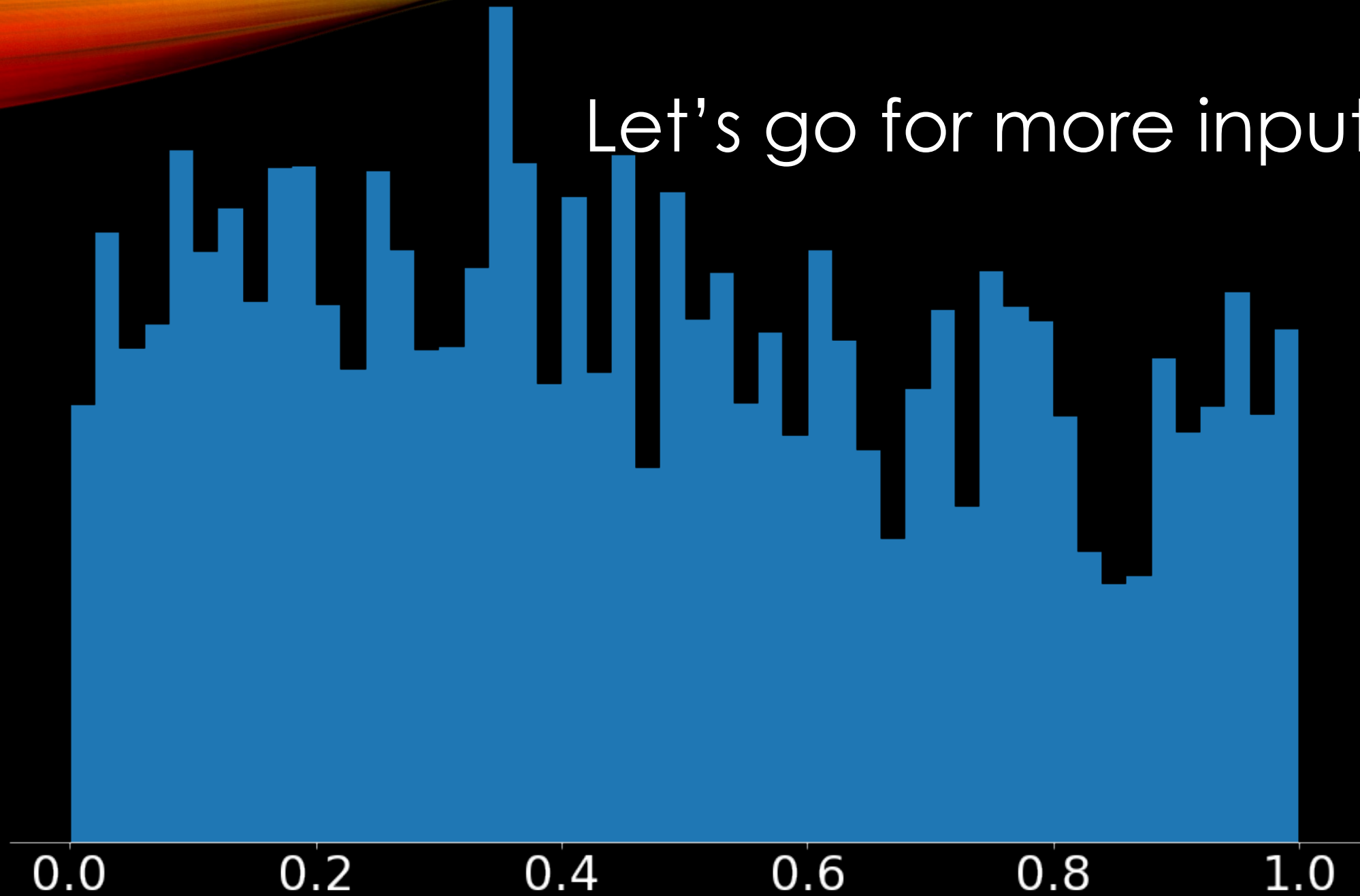


Let's try to fix the situation!

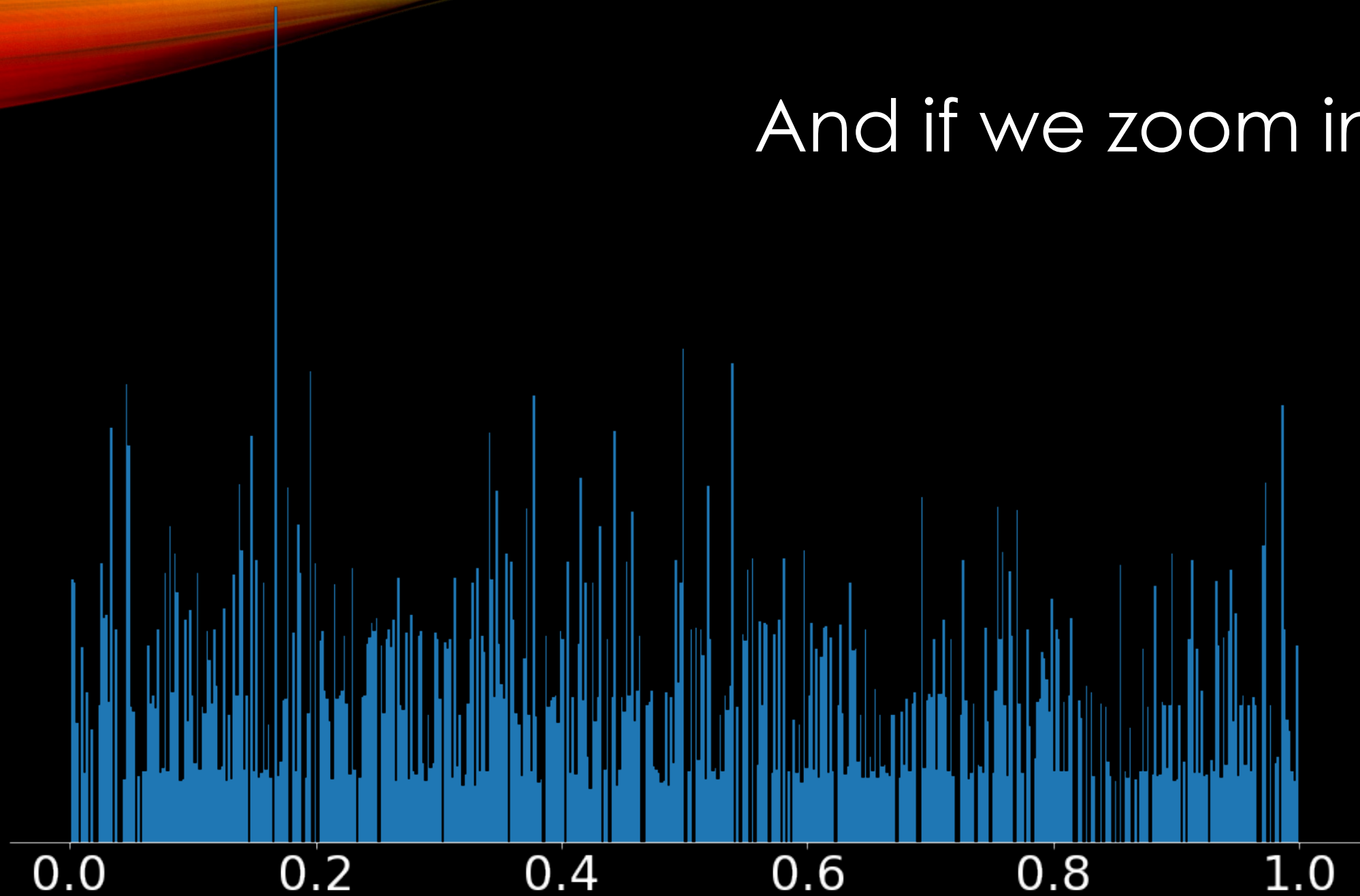
Let's provide a pair of inputs to the discriminator (PACGan)



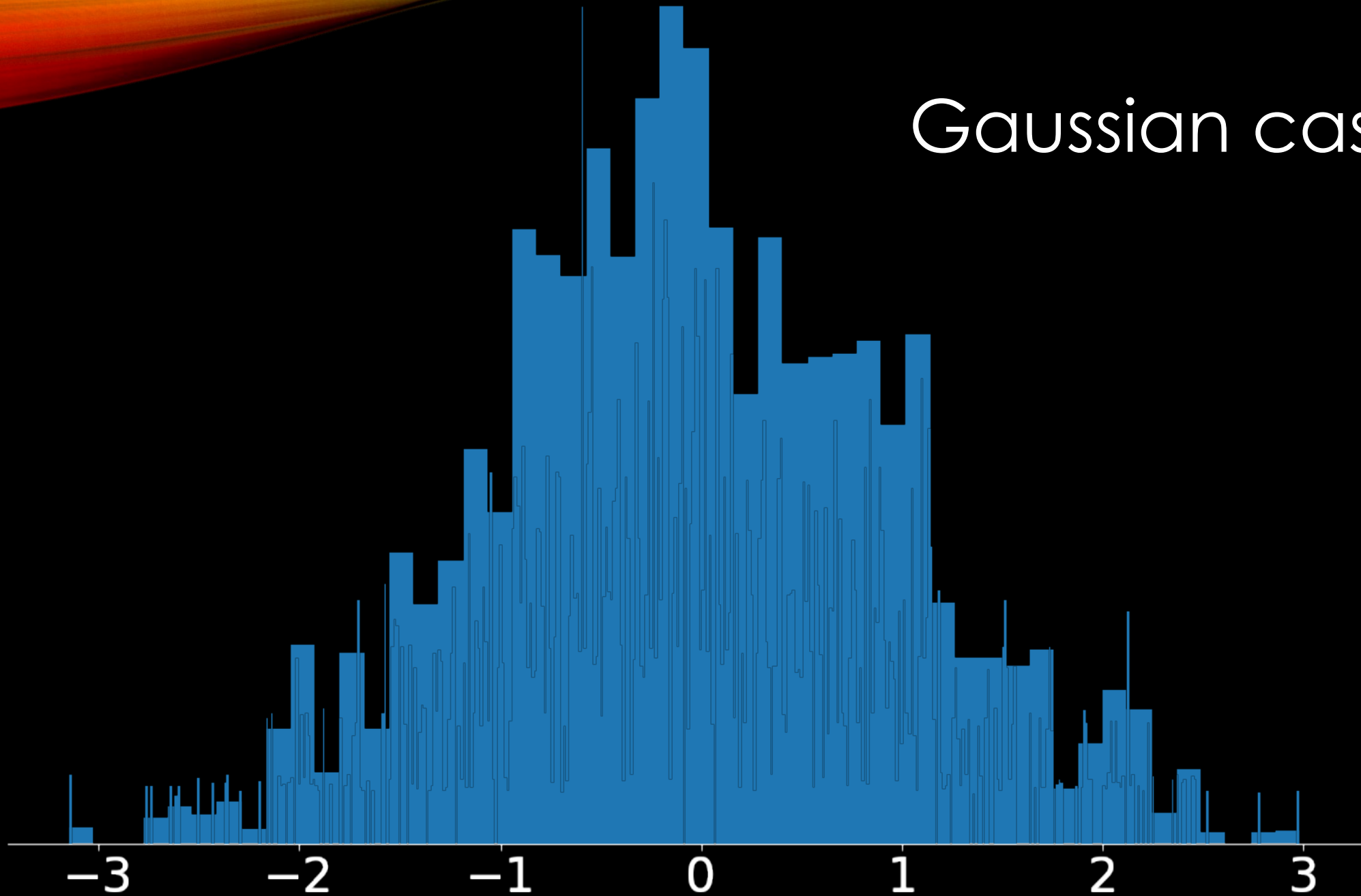
Let's go for more inputs!



And if we zoom in?



Gaussian case



A look back

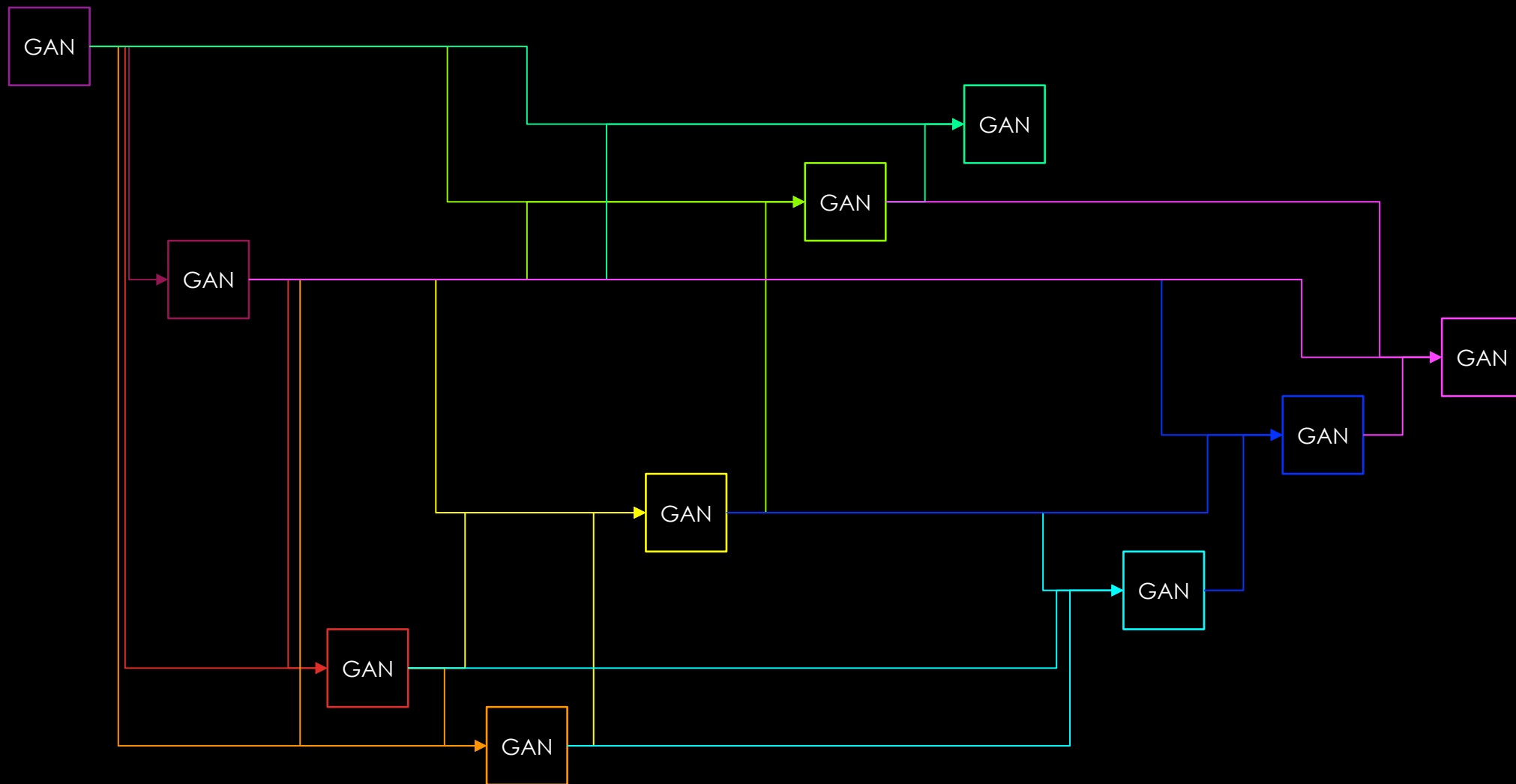
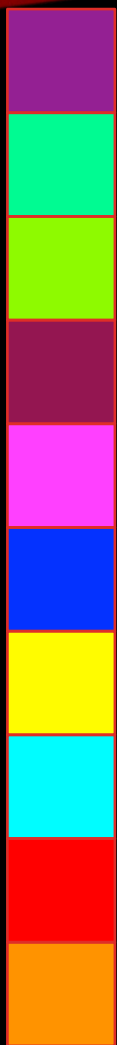
What has been working ? **Sequential simulation!**

Let's construct a sequential simulation using GAN or VAE

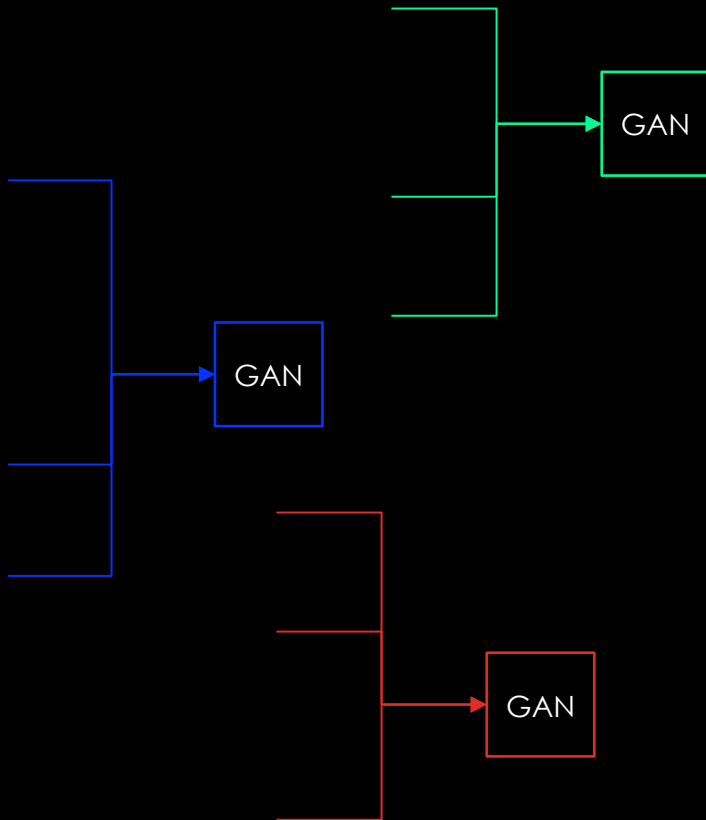
1. We search neighbors, and ask the ML to predict the value for this neighborhood.
2. Then move to the next point

In this framework

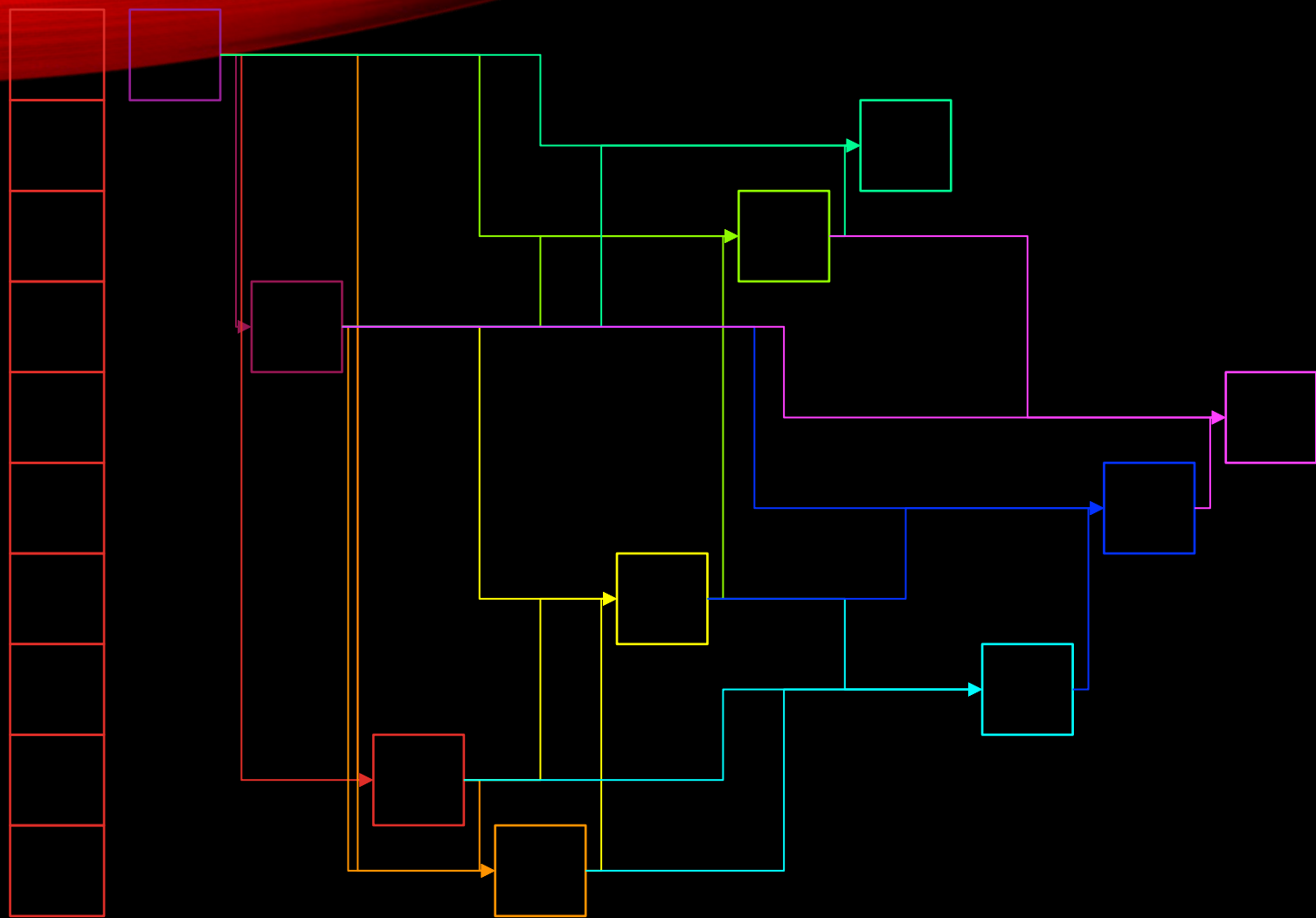
→ MPS is directly comparable to Distributional Random Forest



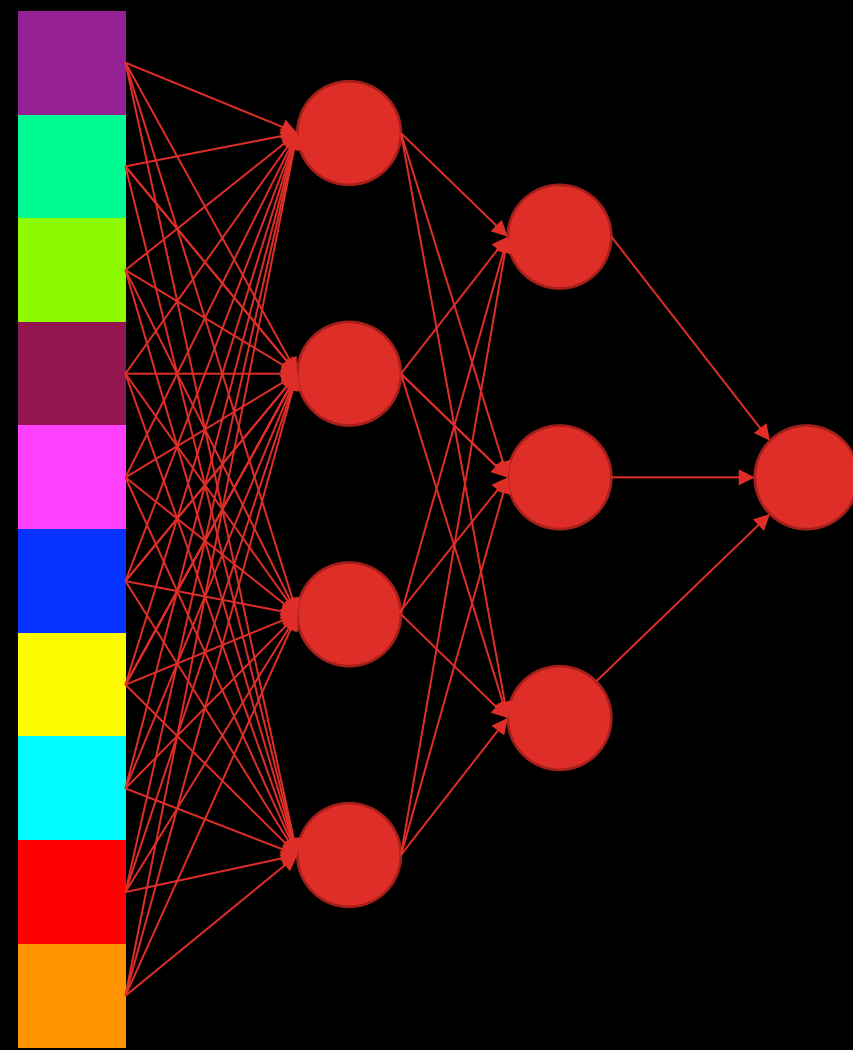
How to train such stuff ?



1. We usually don't have multiple samples
2. We are back to the PDF issue



Generator



Discriminator

Dose it work ?

1D on small example it's relatively ok

2D ? 🙄

64x64 with a very small GAN

Require extremely slow training !

Extremely memory intensive

➔ hardly fit on a GPU

Is it worth it ?

Will it be practical to use it ?

Conclusion

GAN, VAE and other generative model currently suffer from bias in the probability of appearance!

→ single realization is ok, do we need to train a ML for a single sample ?

Solutions?

- New generation of models that do not suffer from such issues
- Can we quantify the true probability of a realization?
Then we could correct the statistics, and partially solve the problem

Do other models really respect this property ?

Is anything Gaussian ?



Questions ?