Relative representations
enable zero-shot latent space communication

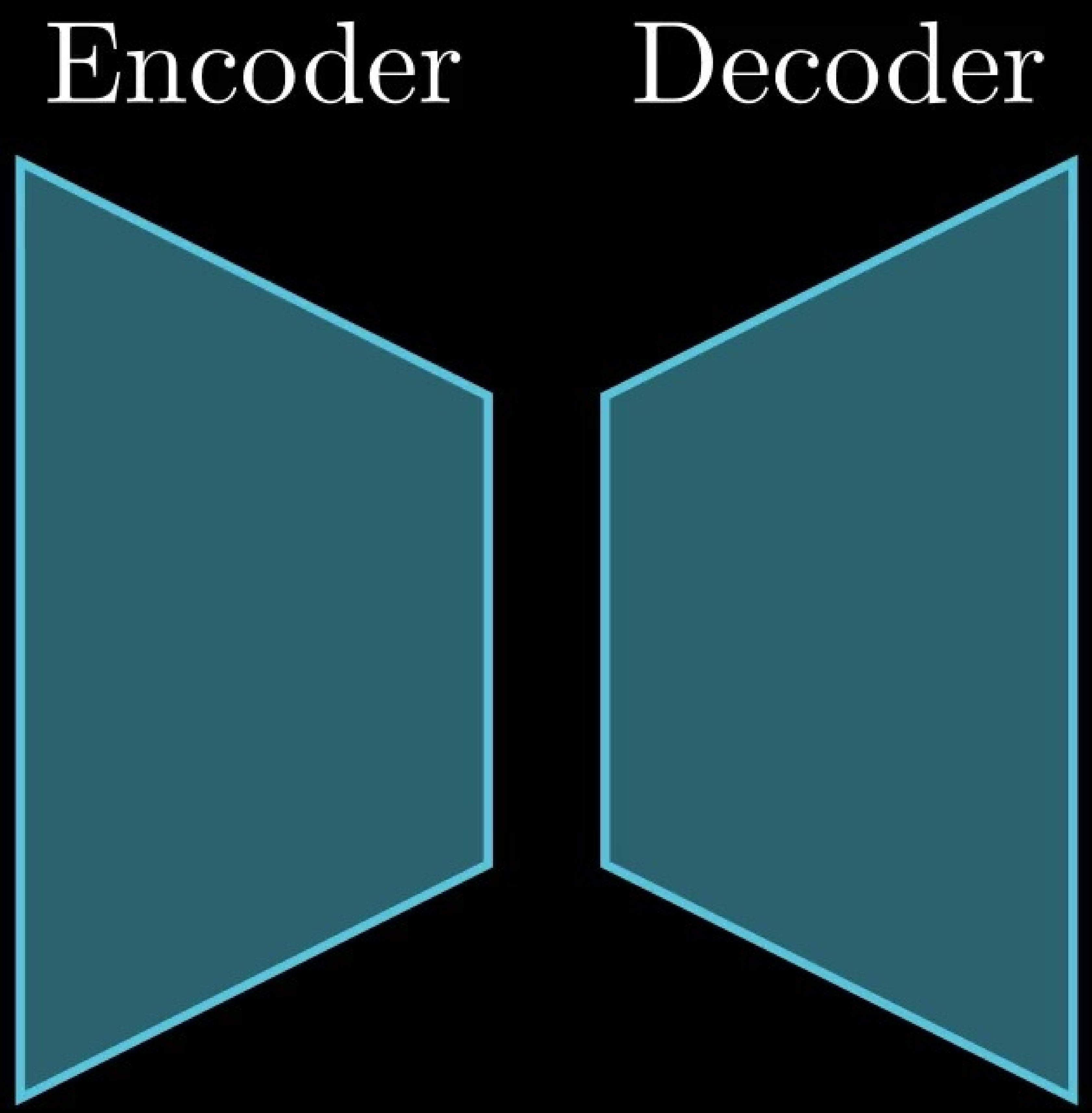
Luca Moschella Valentino Maiorca

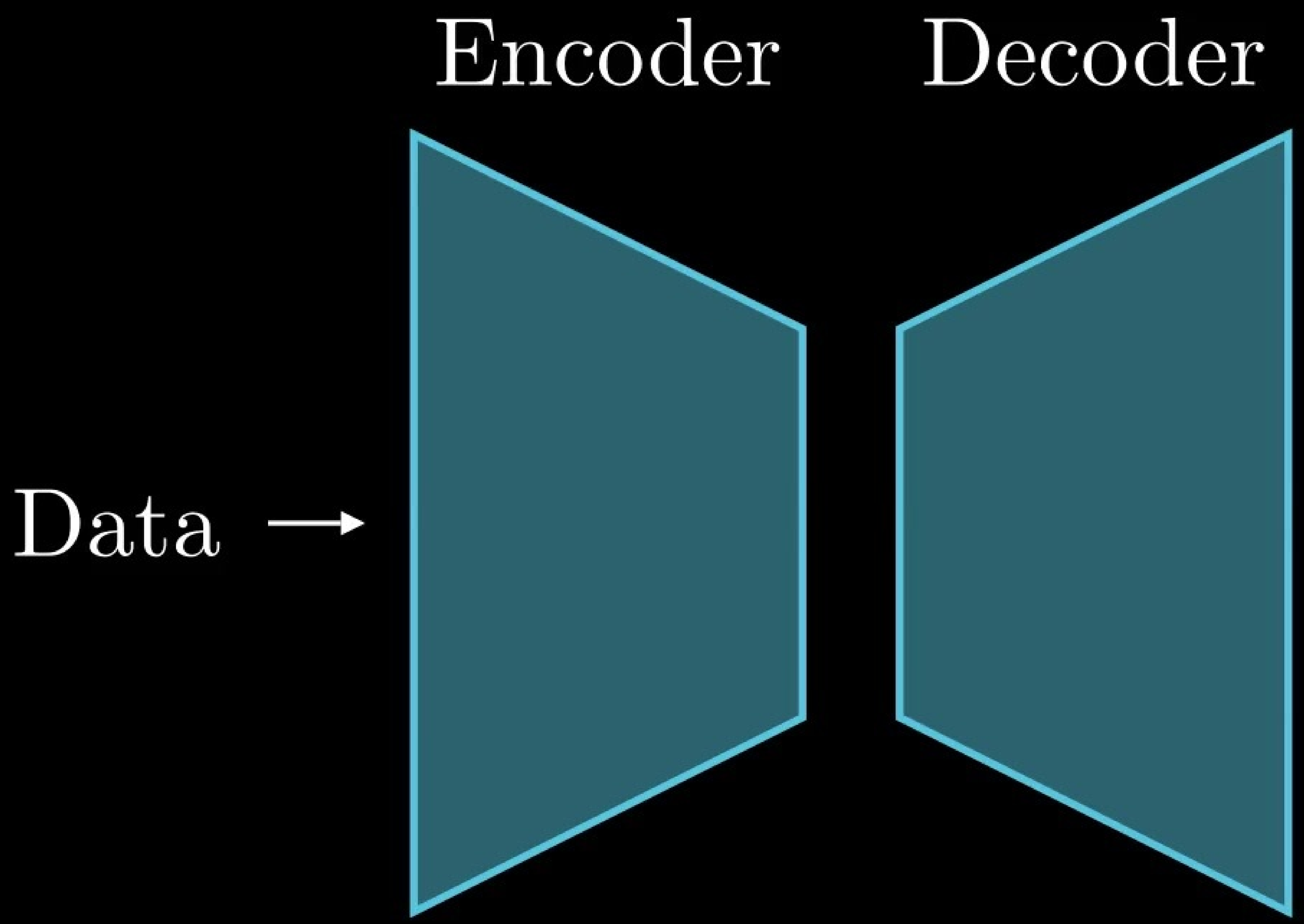
Marco Fumero Antonio Norelli Francesco Locatello Emanuele Rodolà

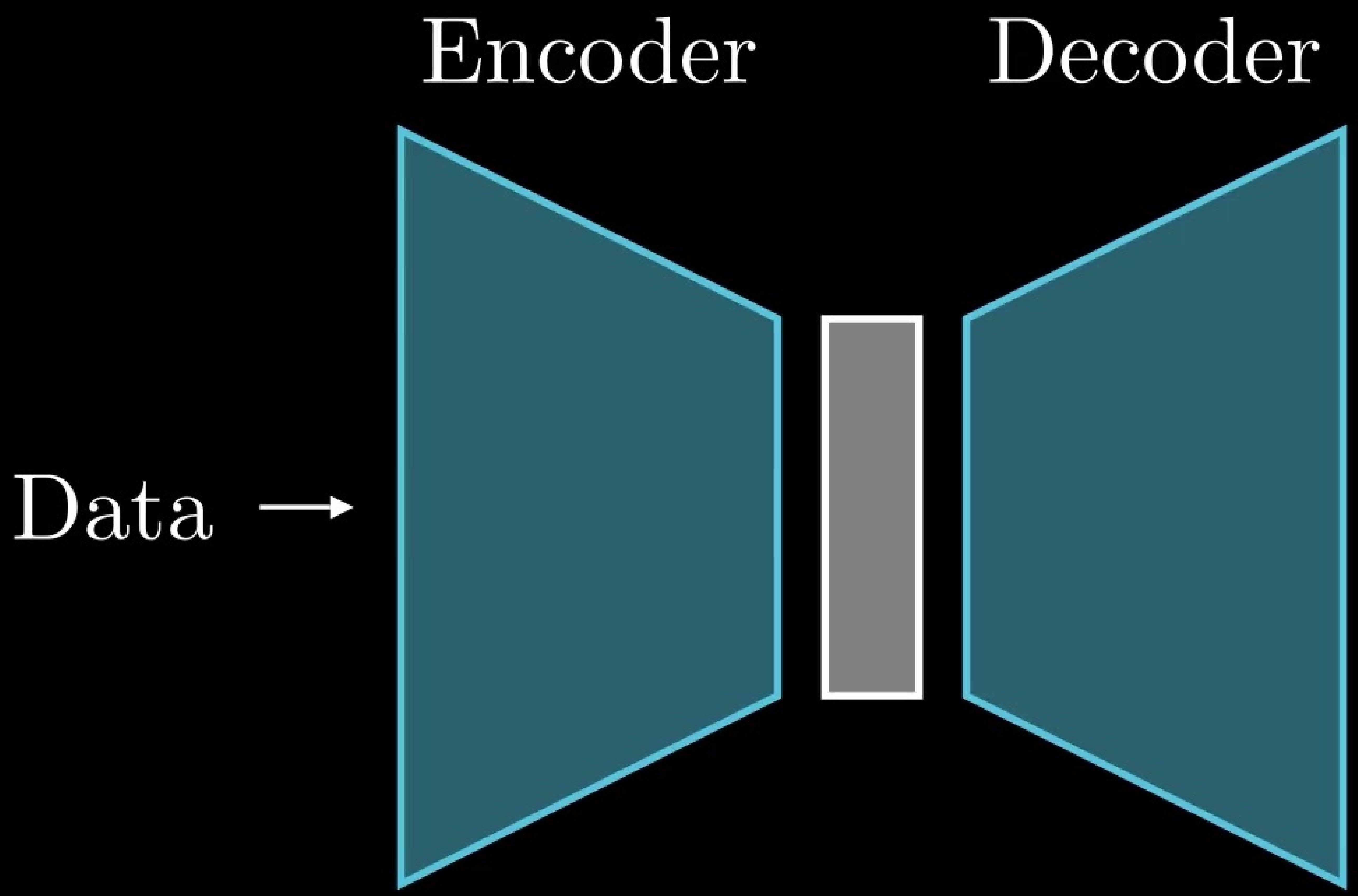
GLADIA

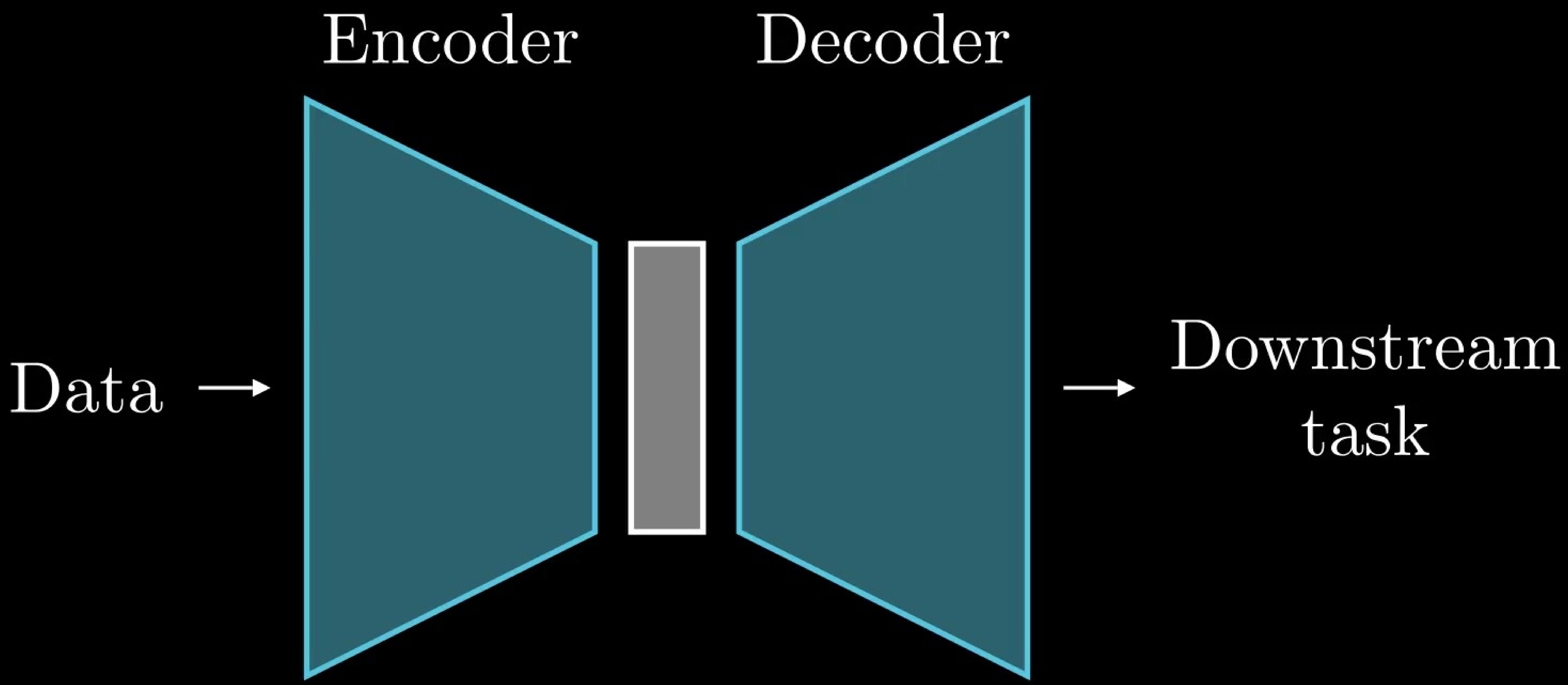
Neural Networks

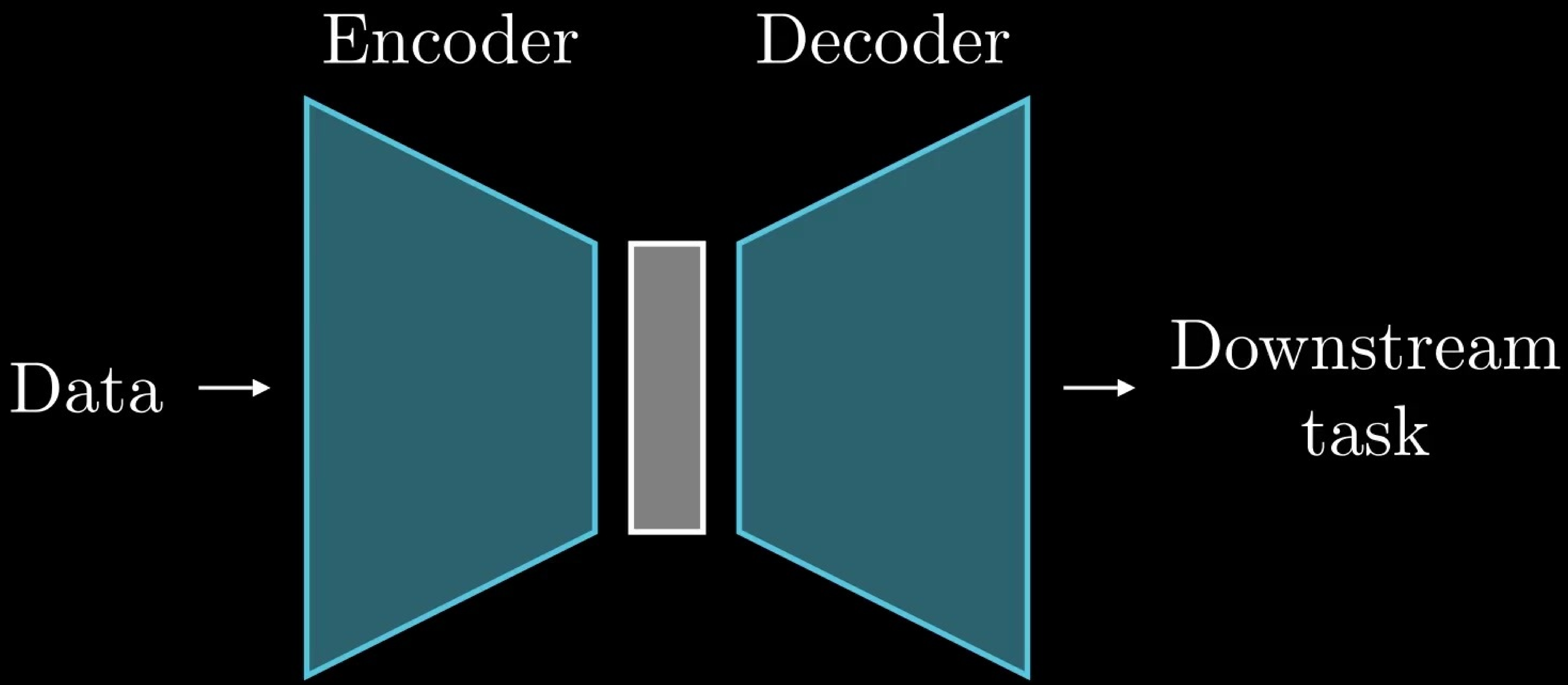
Neural Networks

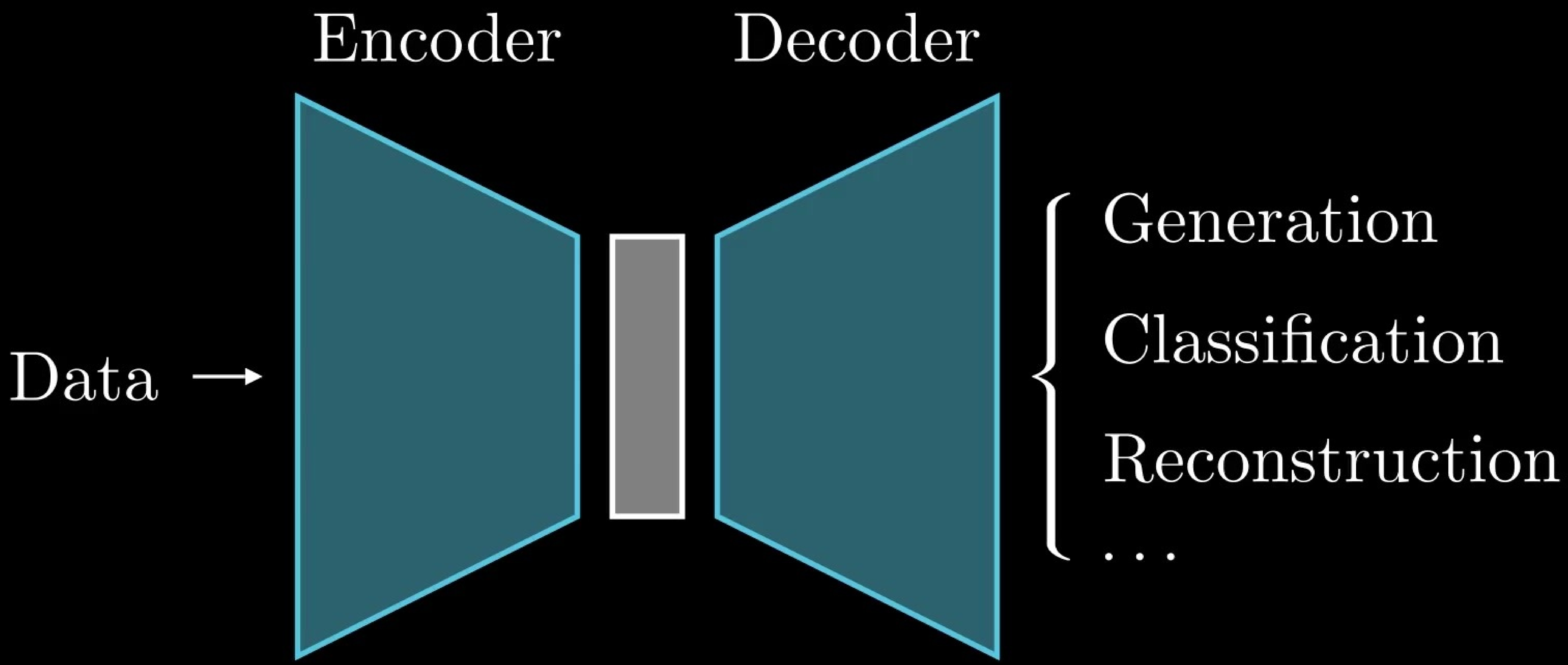

Encoder
Decoder


Encoder
Decoder
Data


Encoder
Decoder
Data

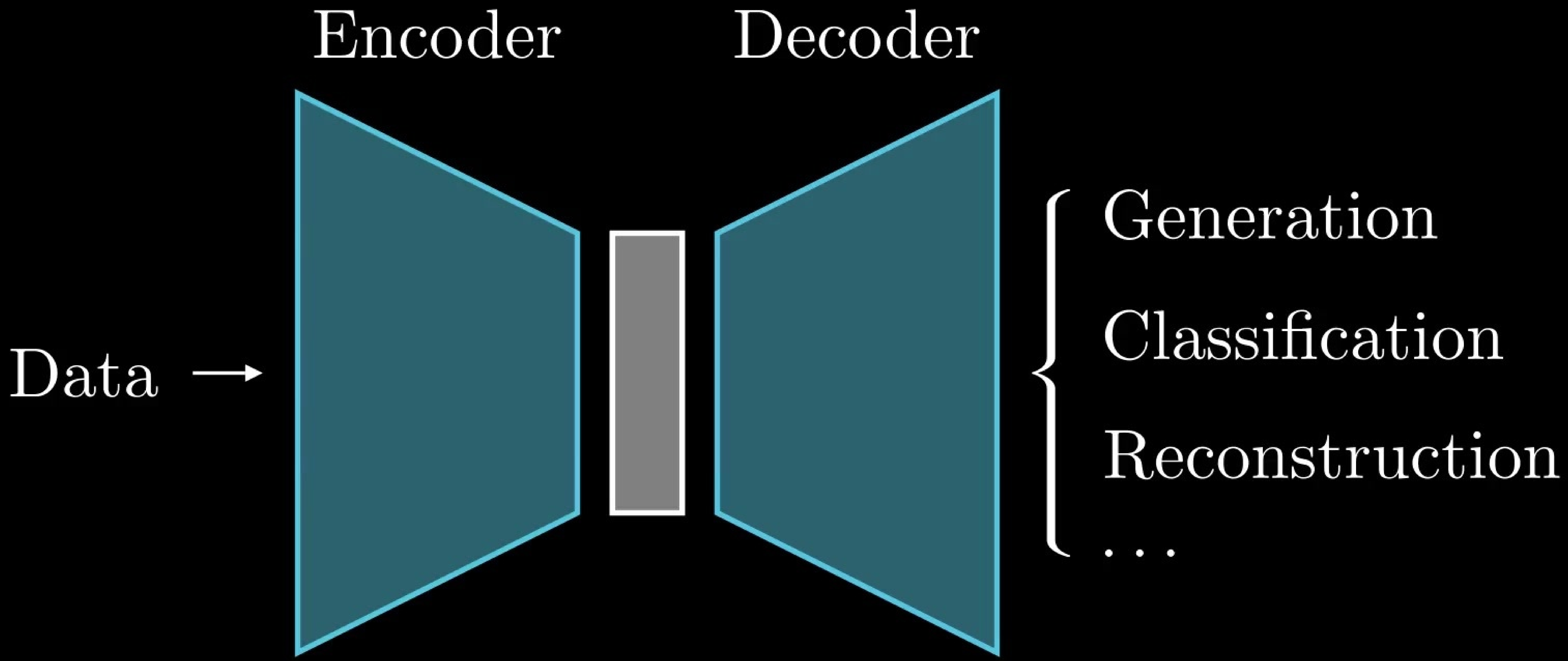

Encoder
Decoder
Data
Downstream
task


Encoder
Decoder
Data
Downstream task

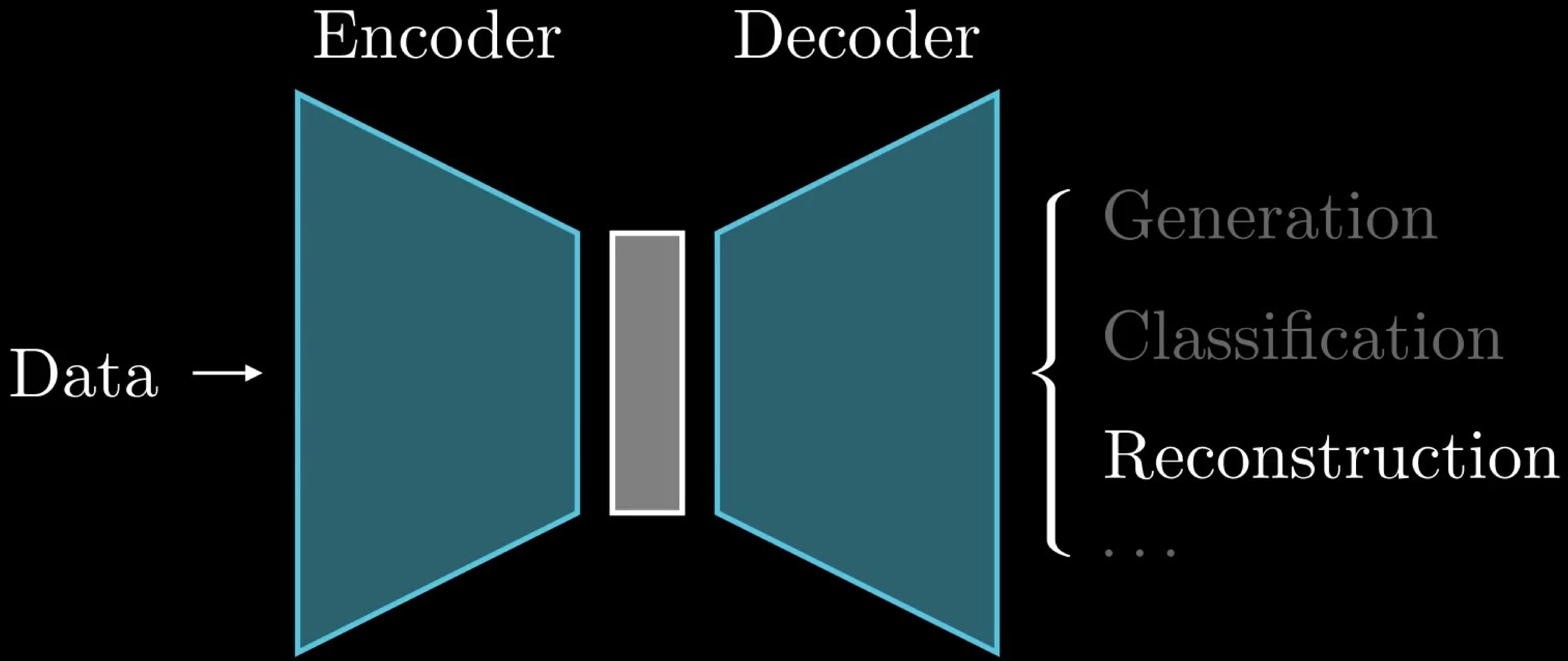

Encoder
Decoder
Data
Generation
Classification
Reconstruction
...

The shape of the latent space

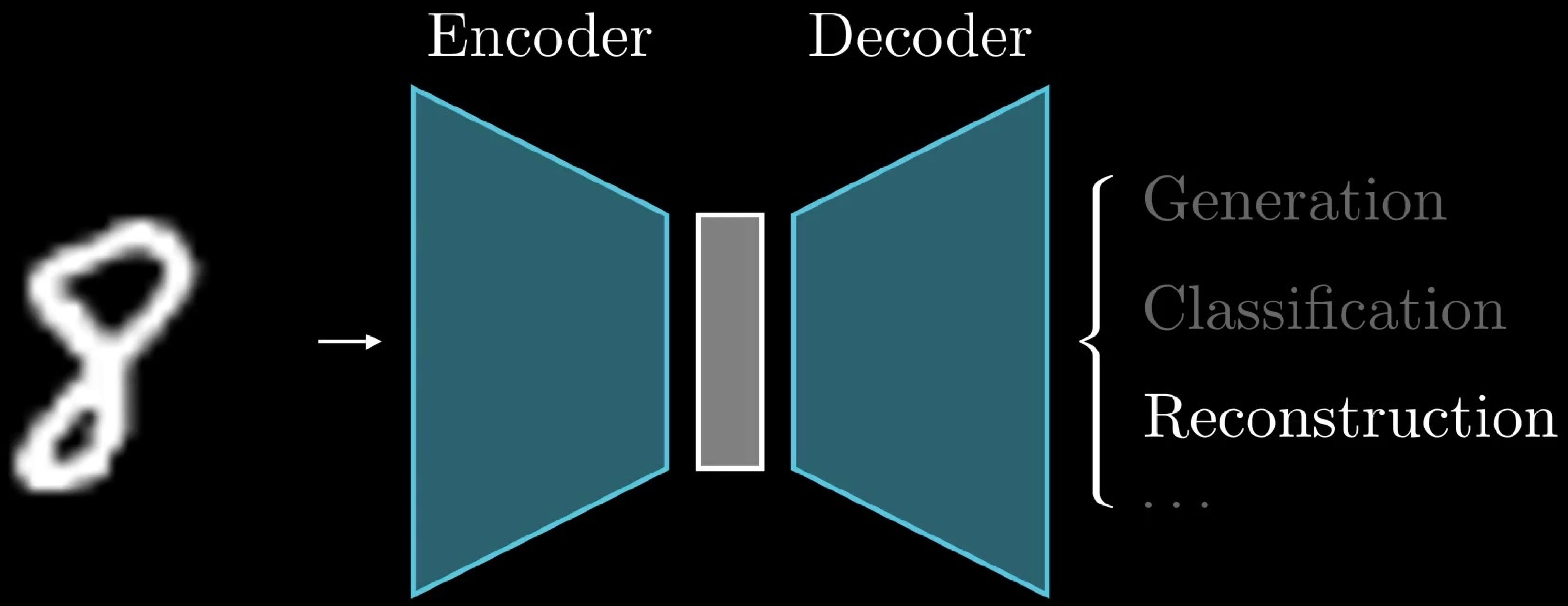
Toy AutoEncoder



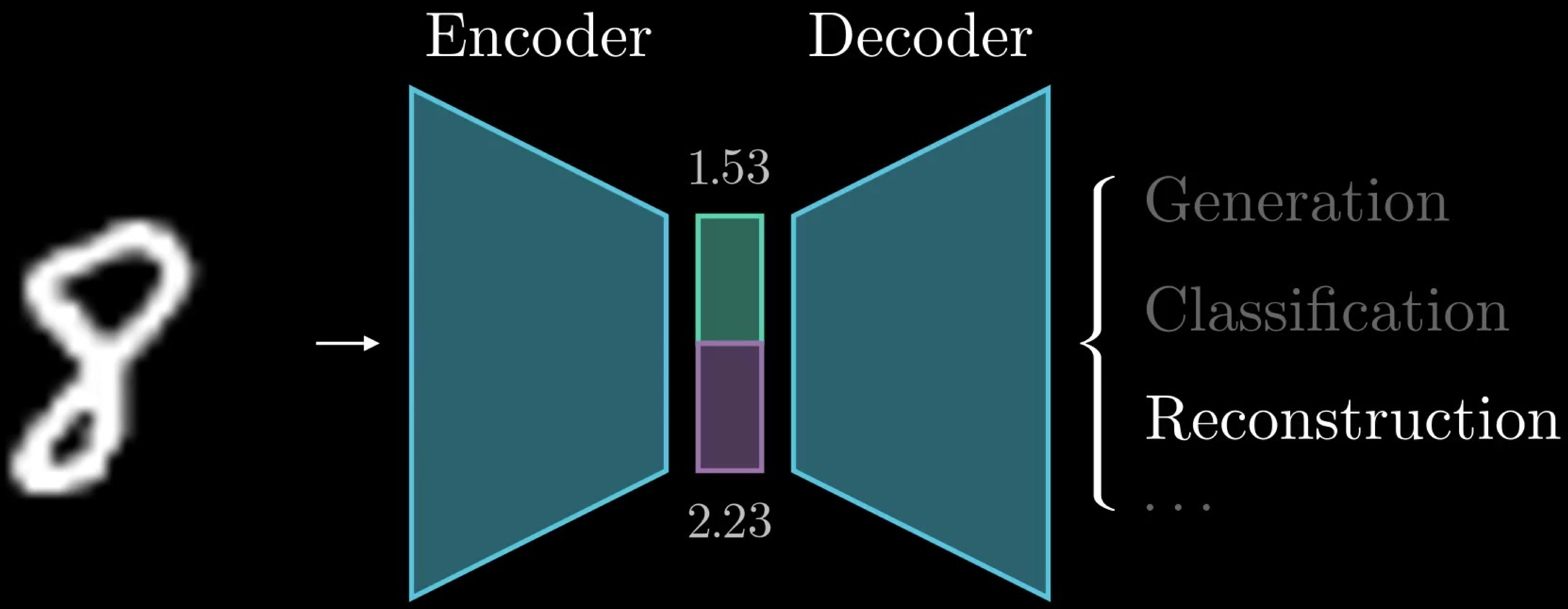
Toy AutoEncoder



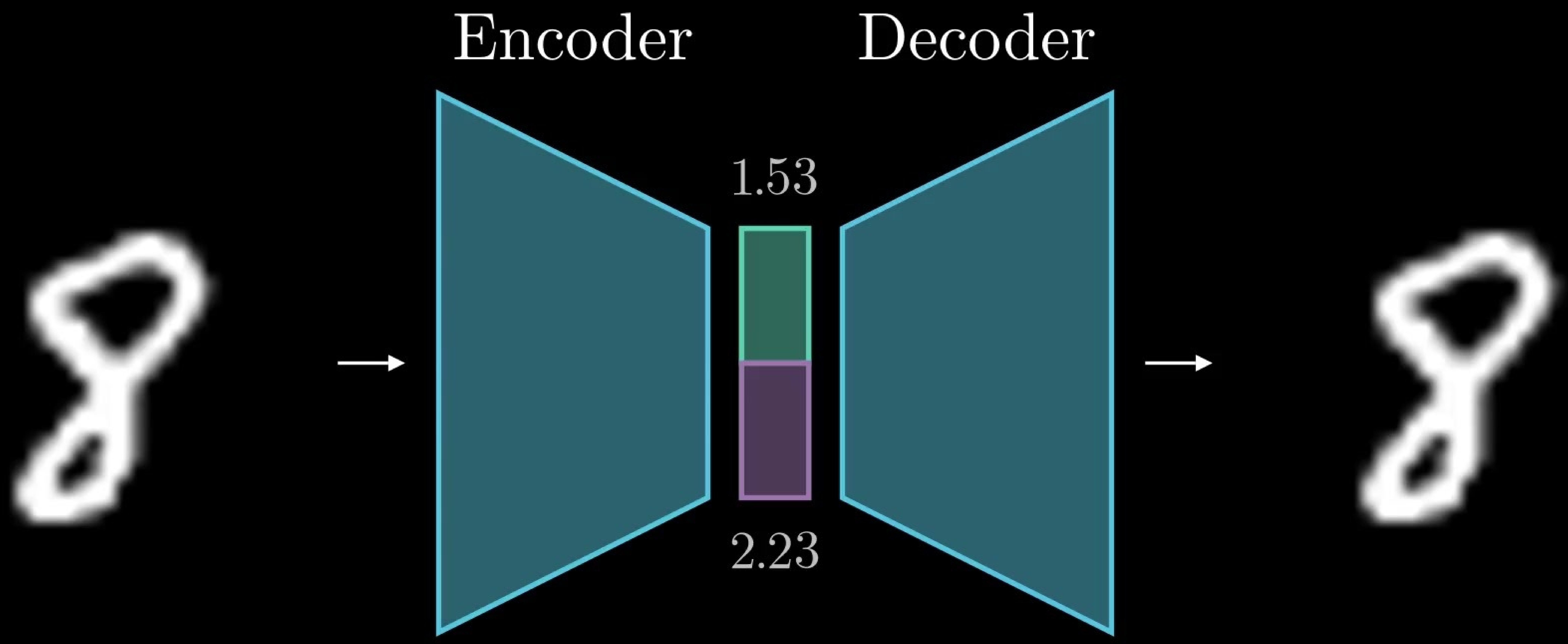
Toy AutoEncoder



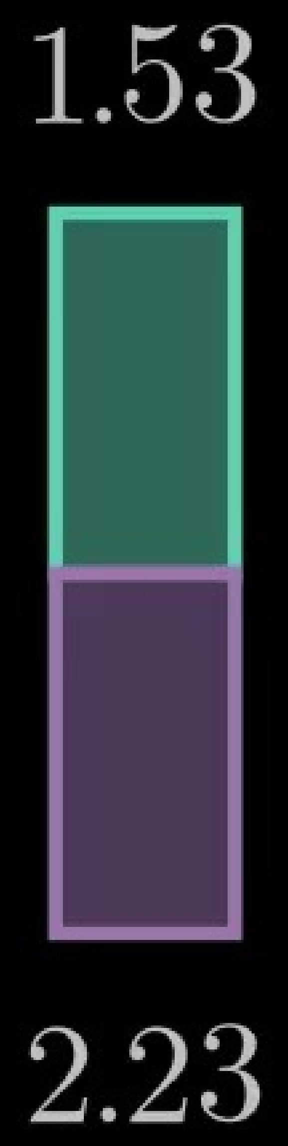
Toy AutoEncoder



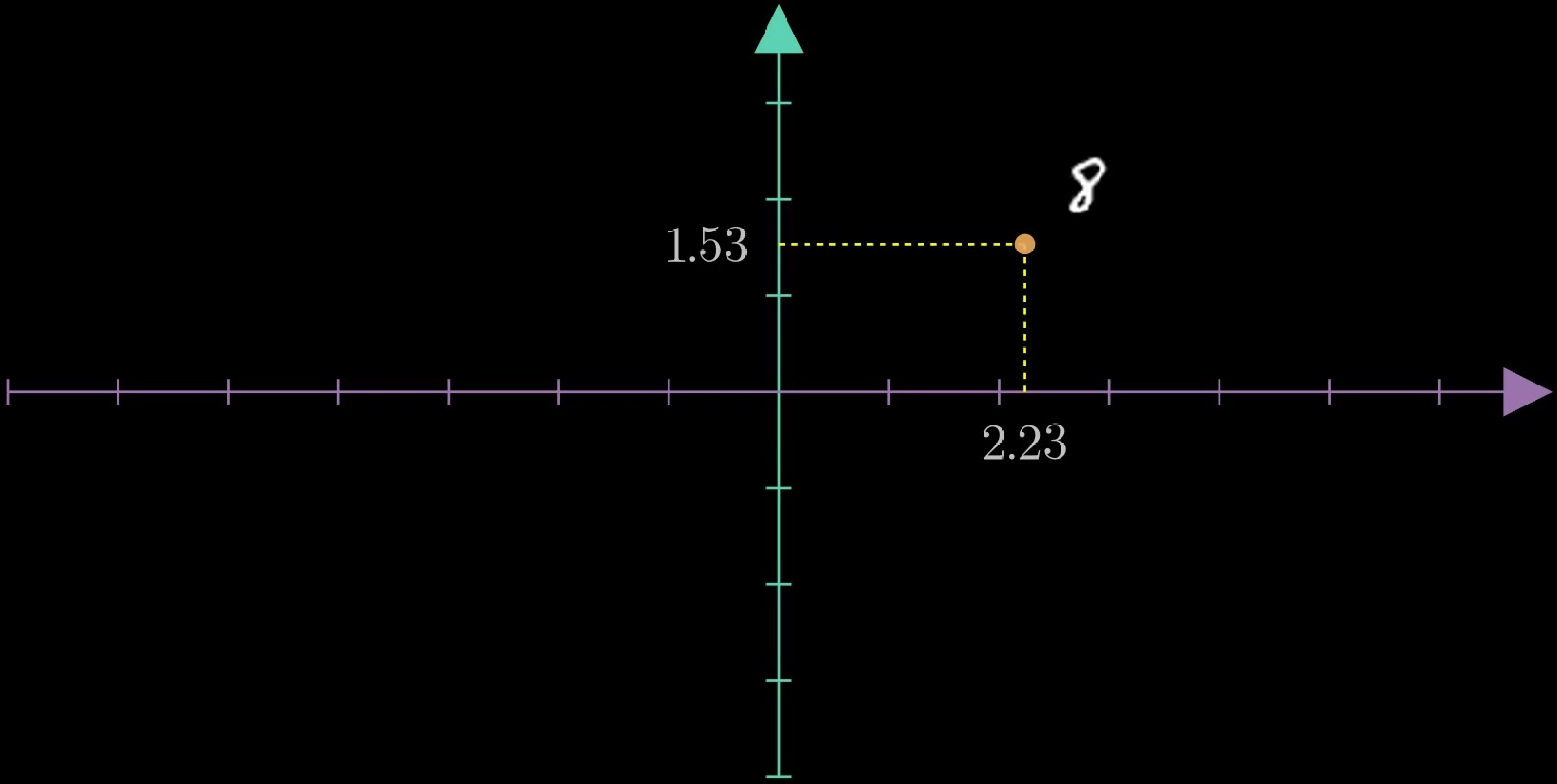
Toy AutoEncoder



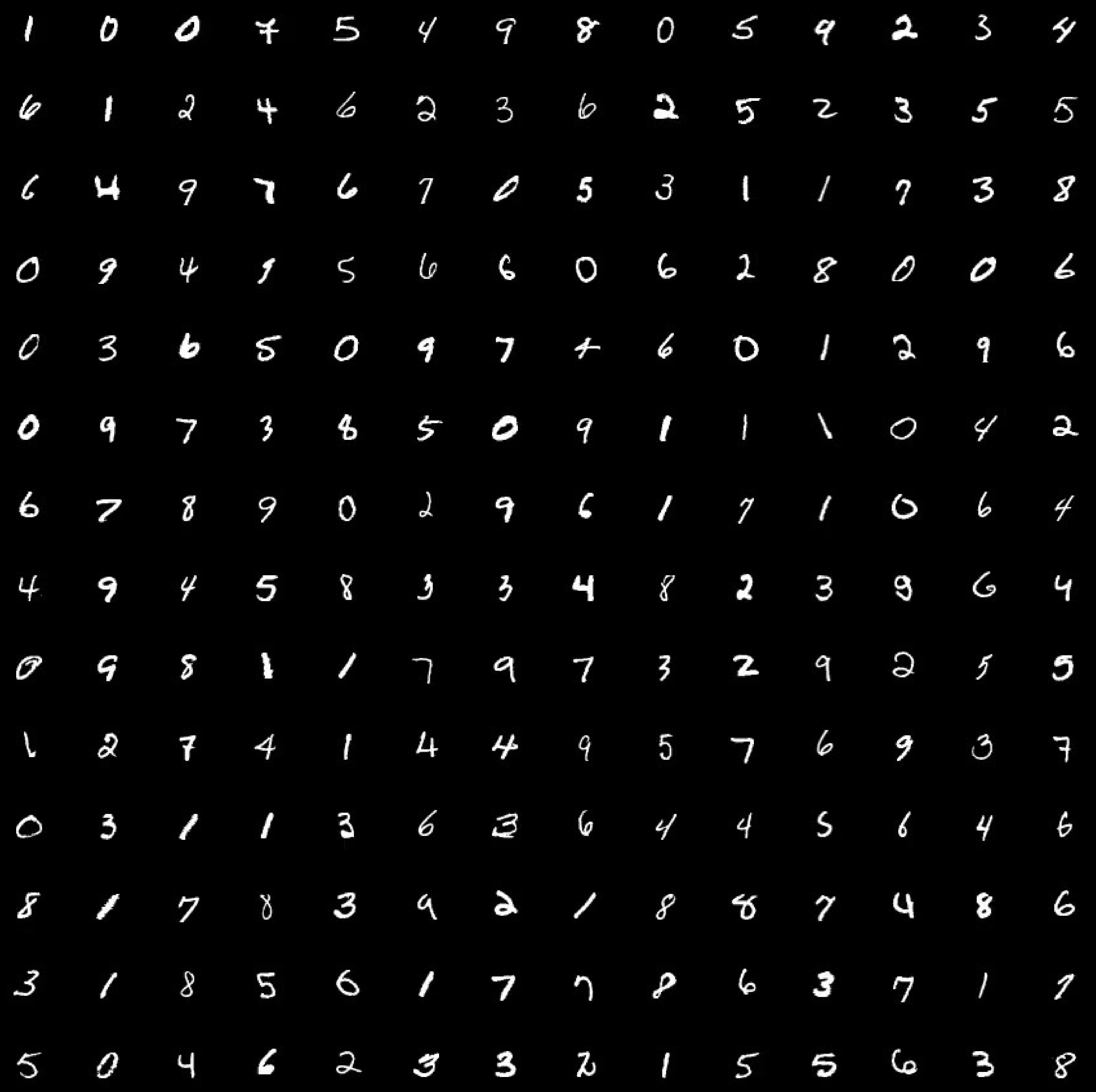
Toy AutoEncoder



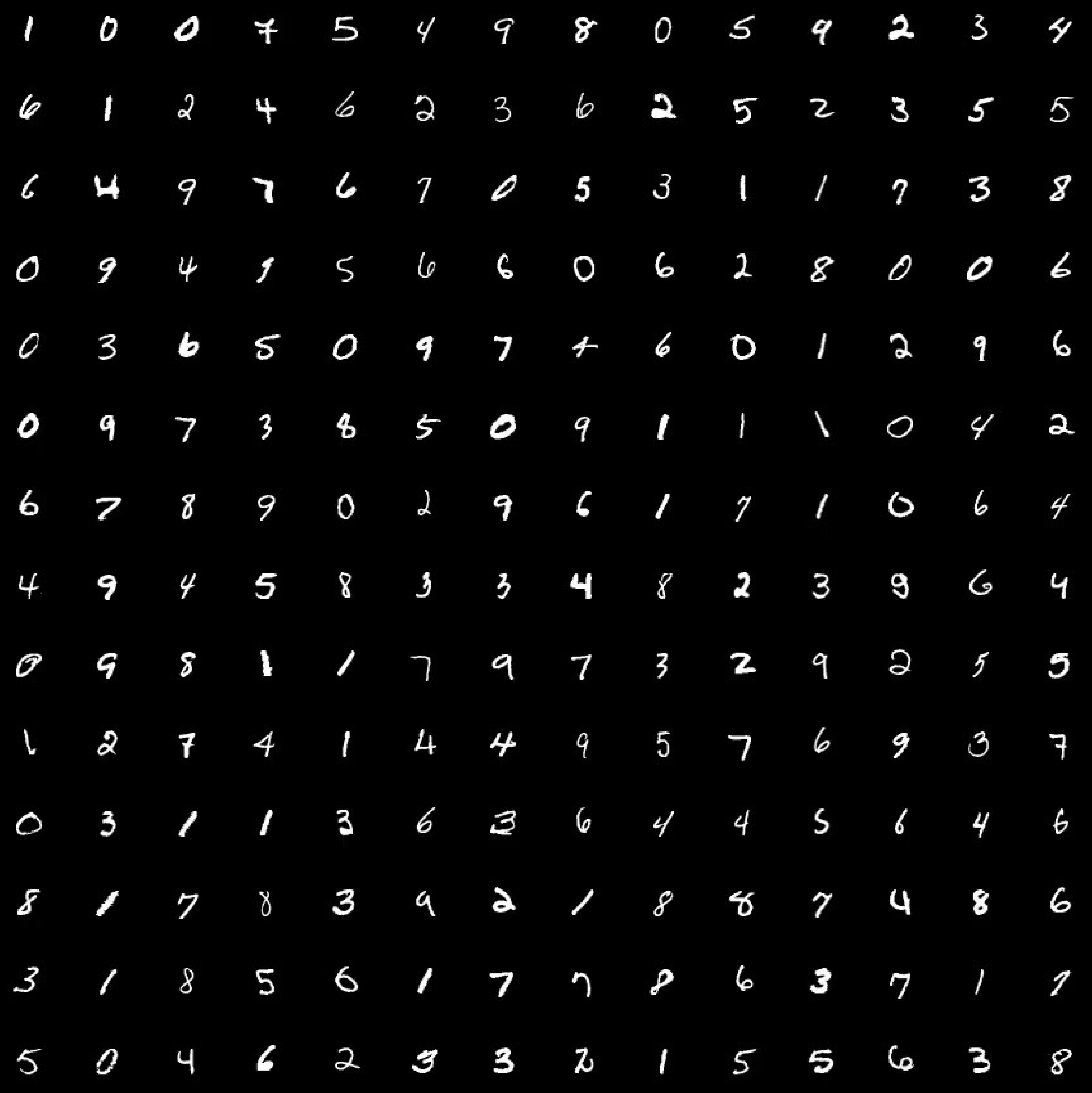
Toy AutoEncoder



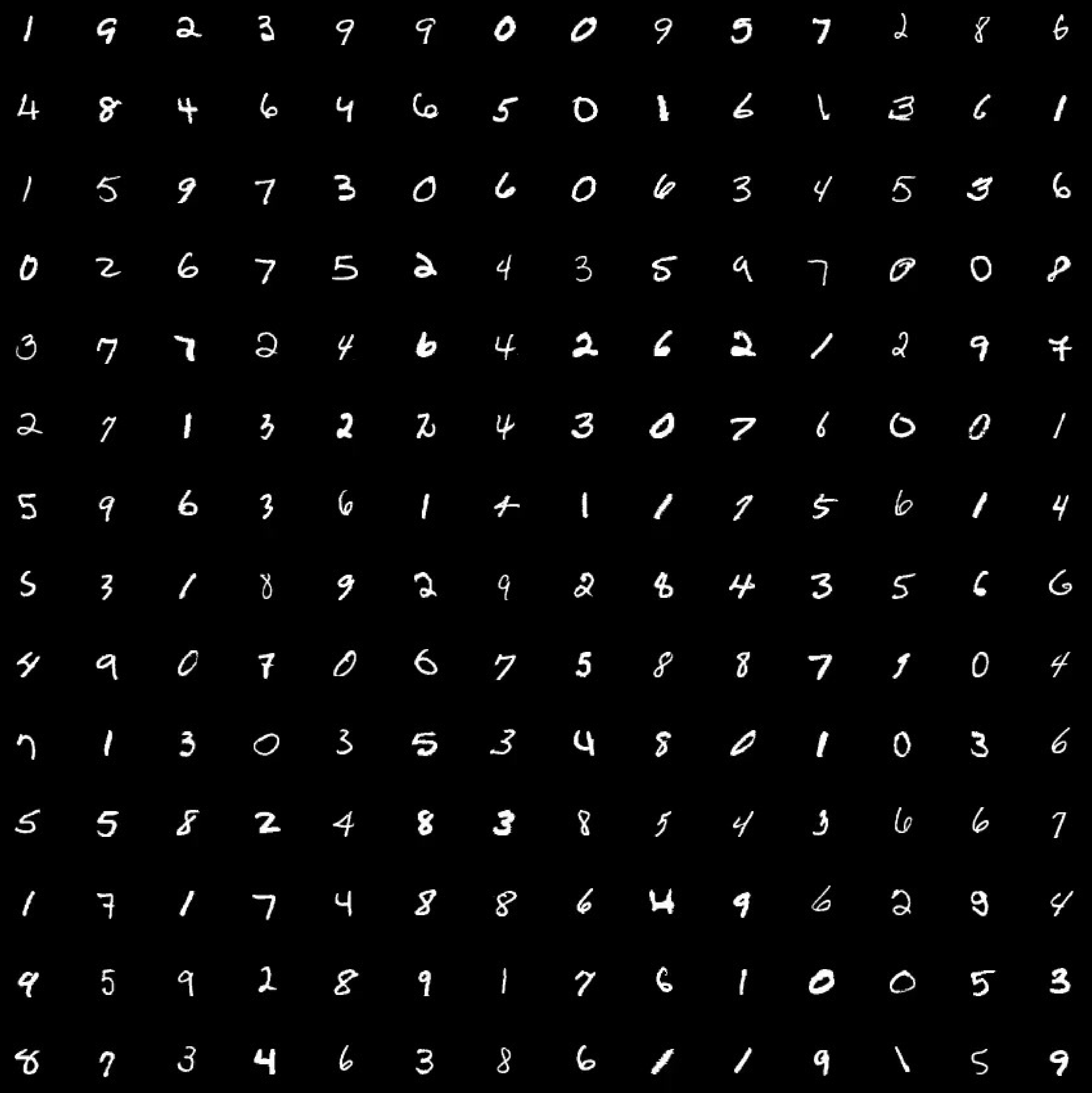
MNIST

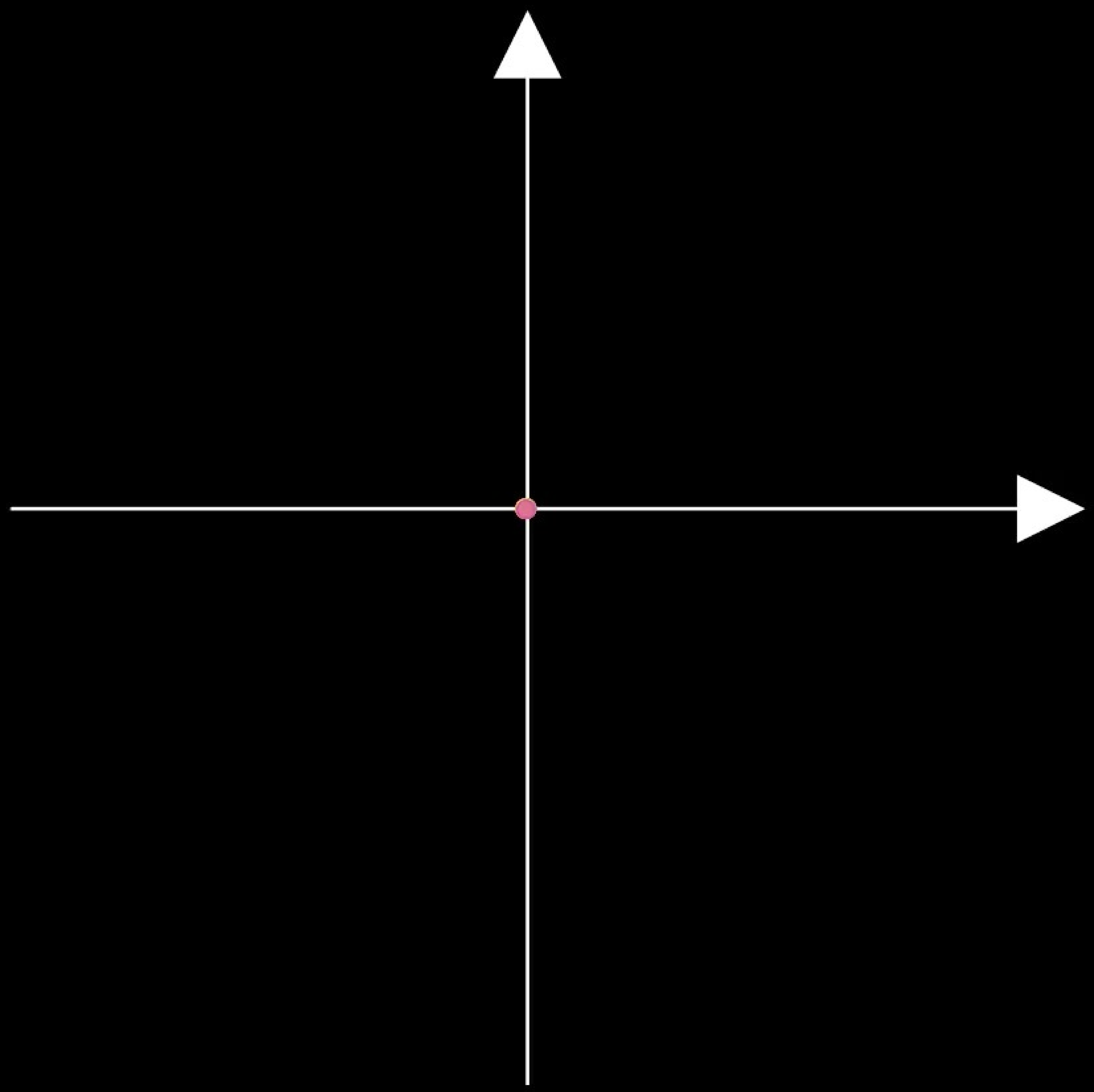


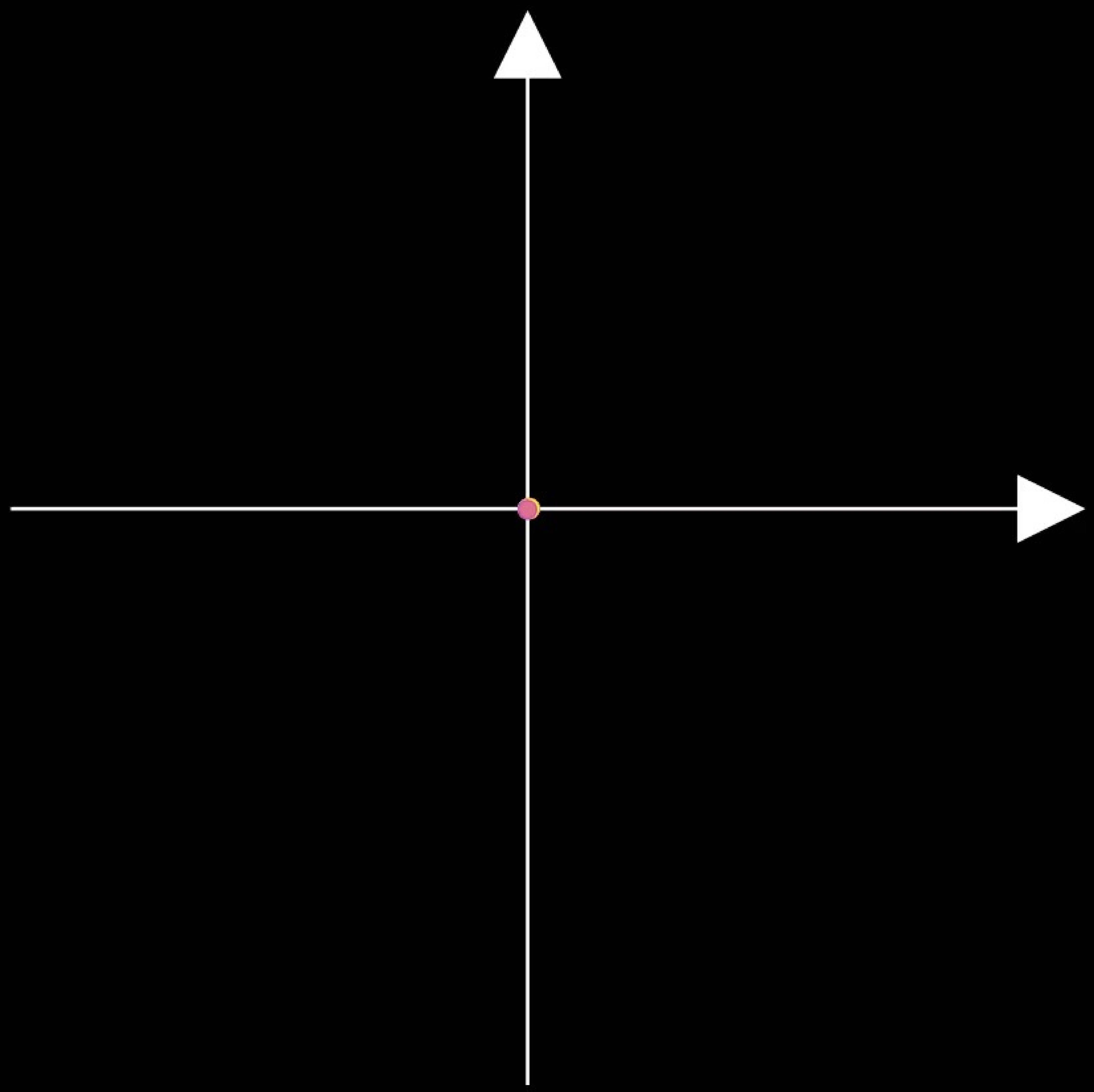
MNIST

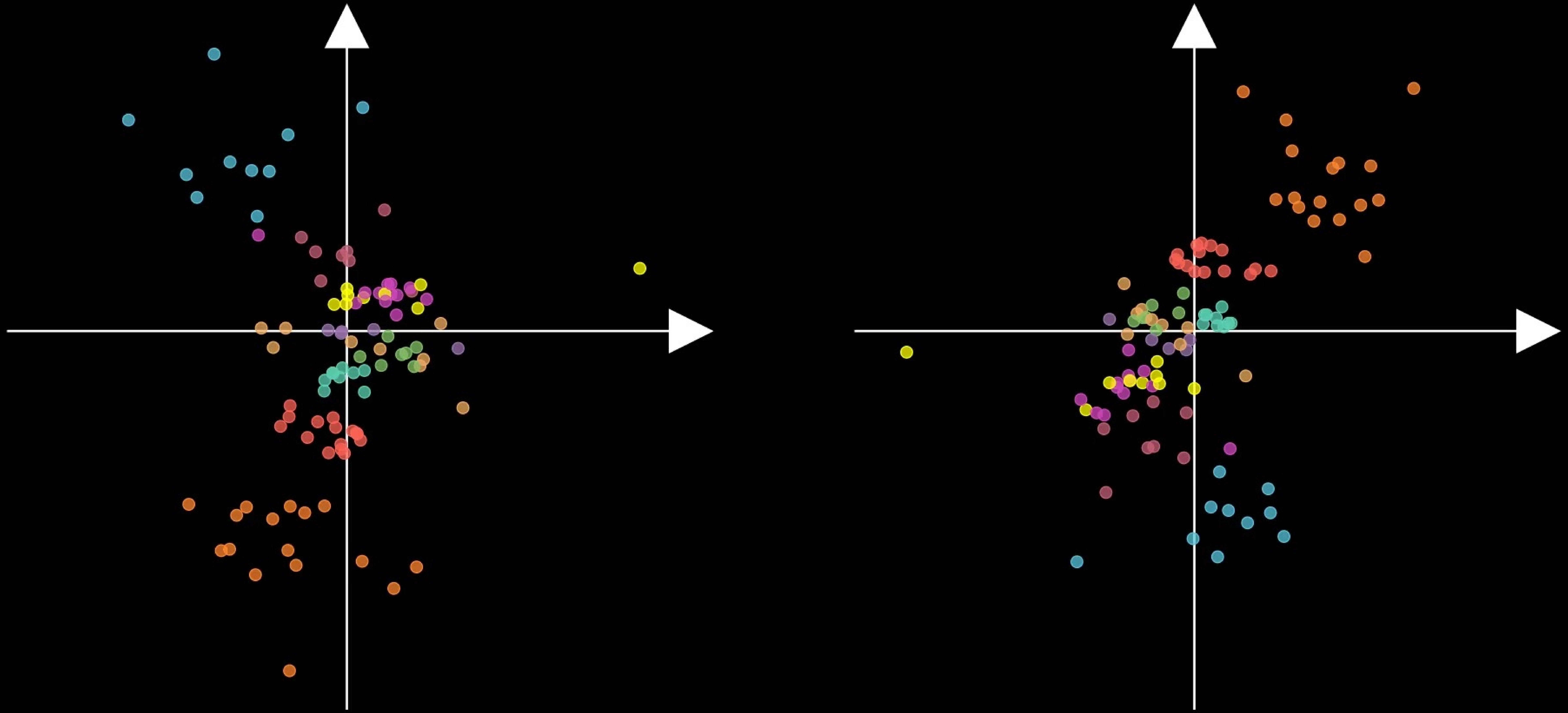


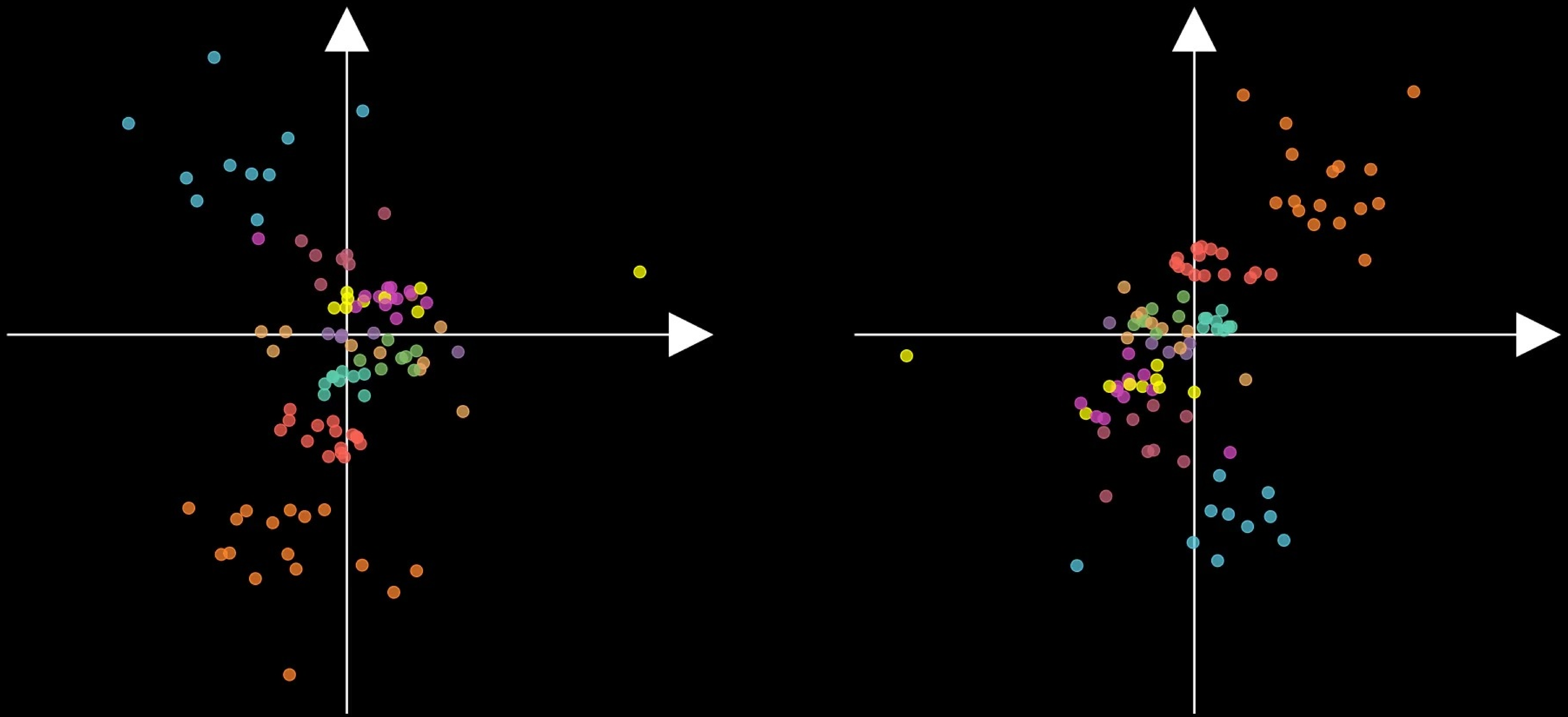
MNIST











...different latent spaces?

The shape of the latent space

Contributing factors...

The shape of the latent space

Contributing factors...

Ideally

- Data Distribution
- The task
- Additional constraints (implicit or explicit)

The shape of the latent space

Contributing factors...

Ideally

- Data Distribution
- The task
- Additional constraints (implicit or explicit)

In practice

- Parameter initialization
- Data shuffling
- Training seed
- Hyperparameters
- ...

The shape of the latent space

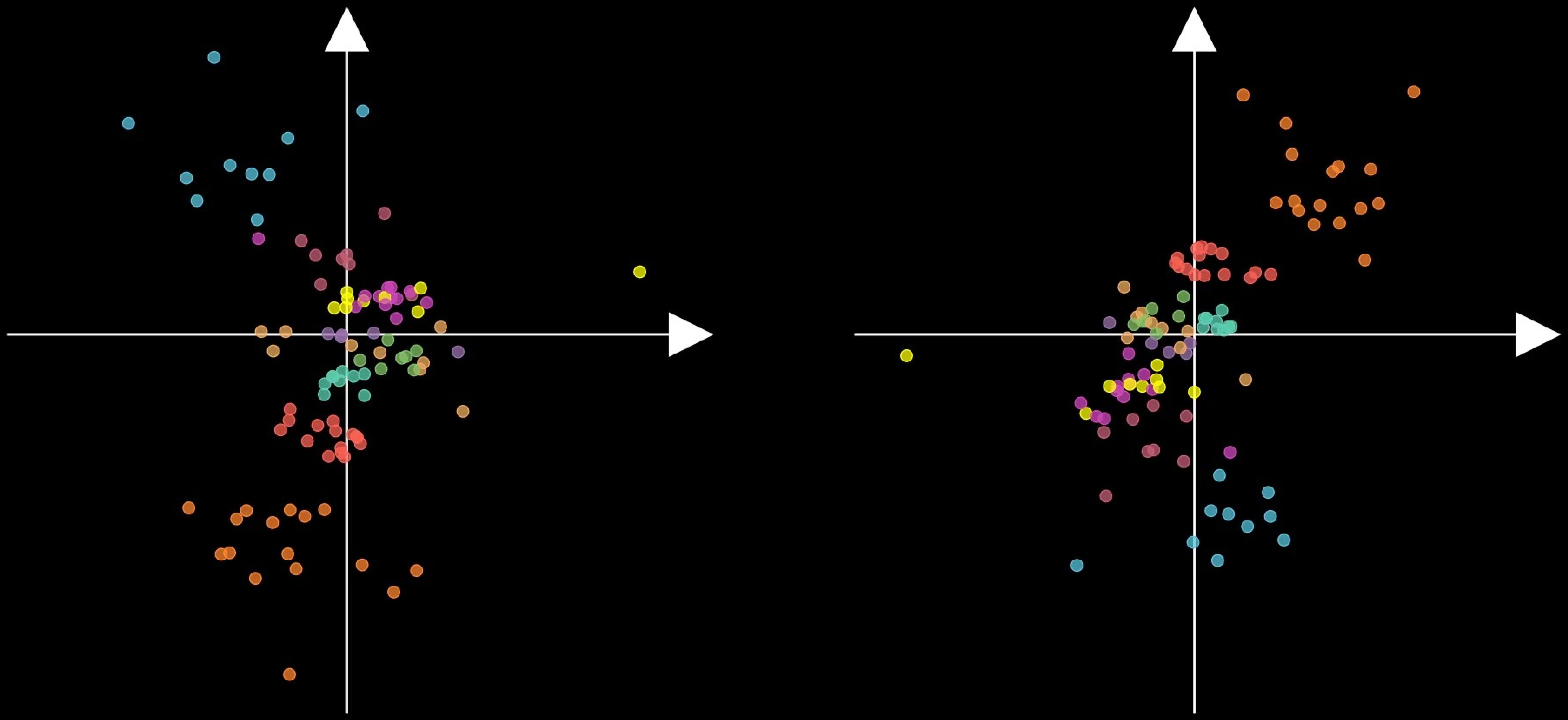
Contributing factors...

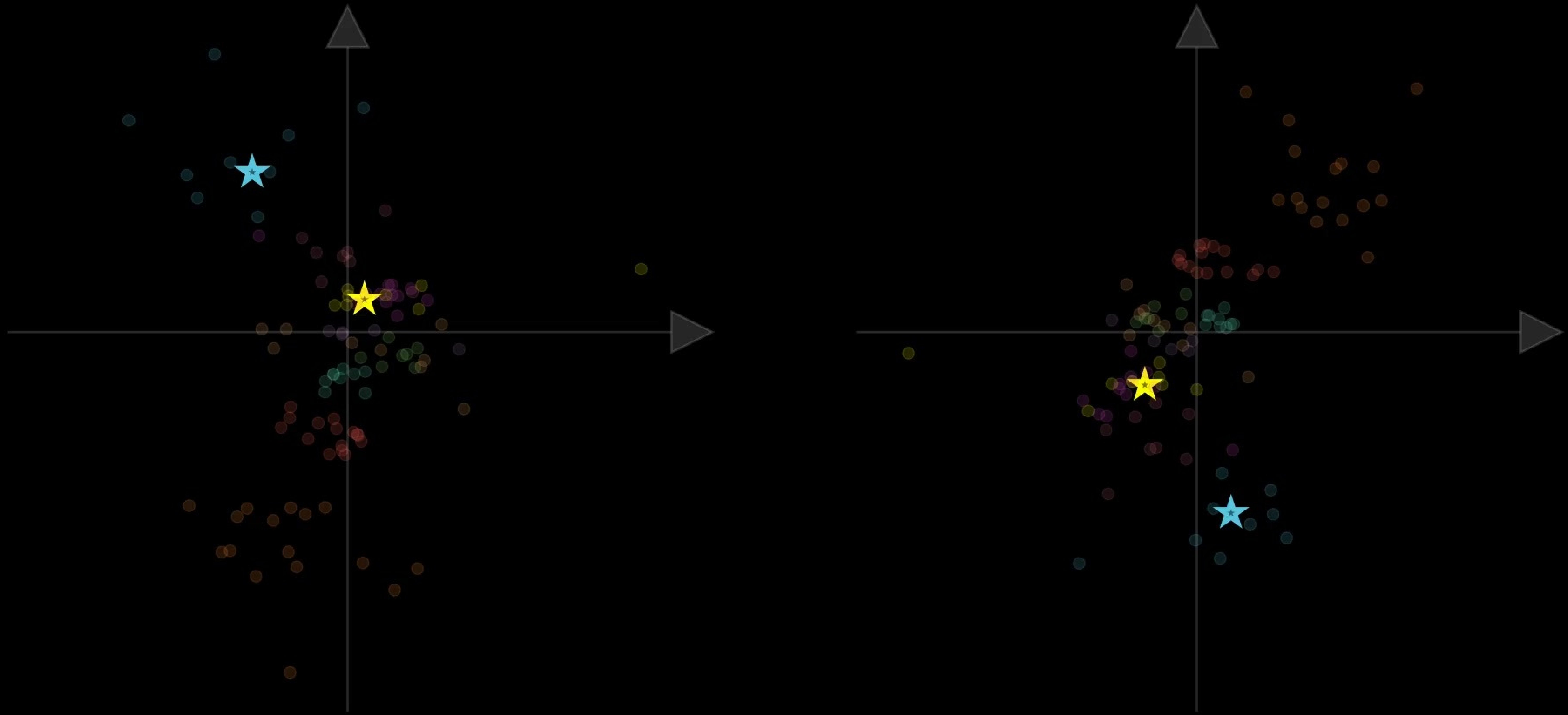
Ideally

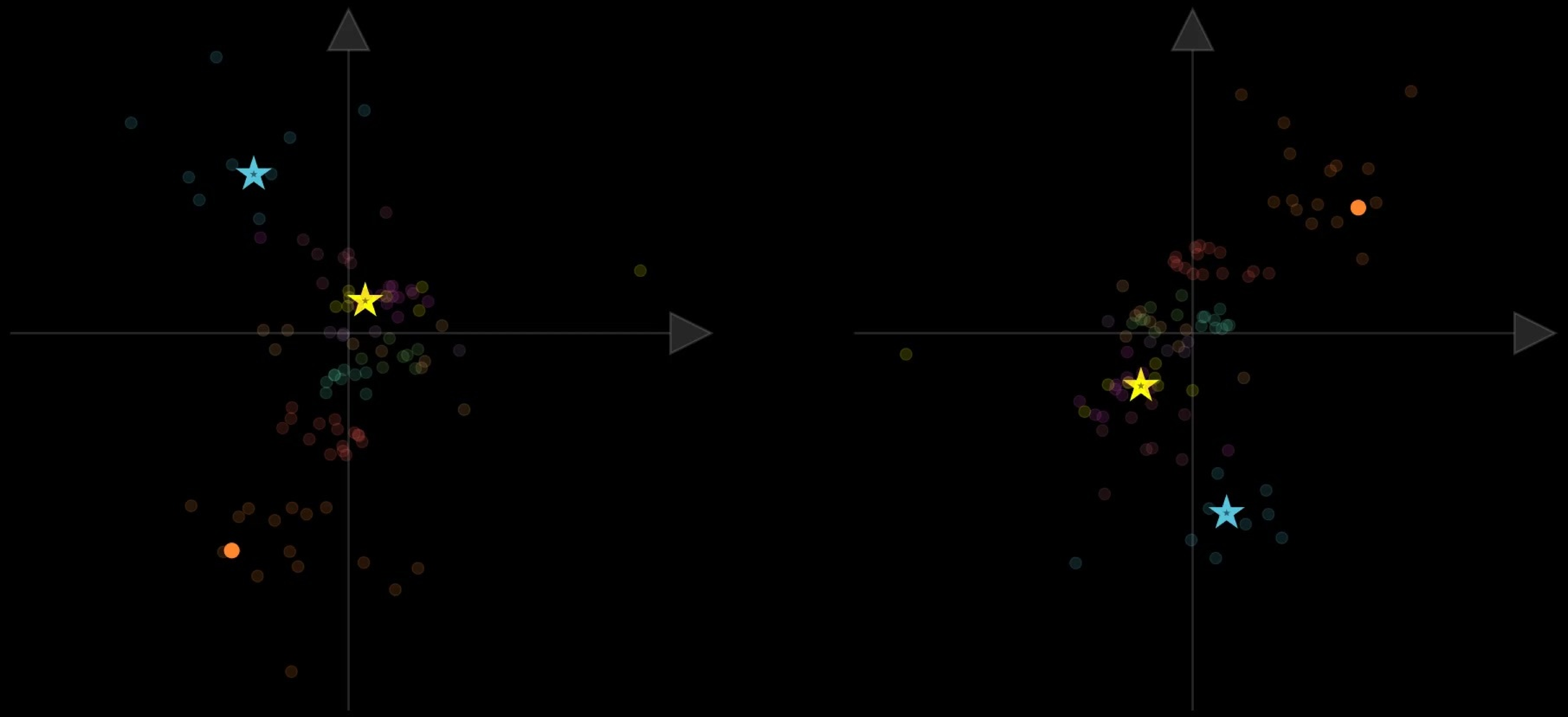
- Data Distribution
- The task
- Additional constraints (implicit or explicit)

In practice

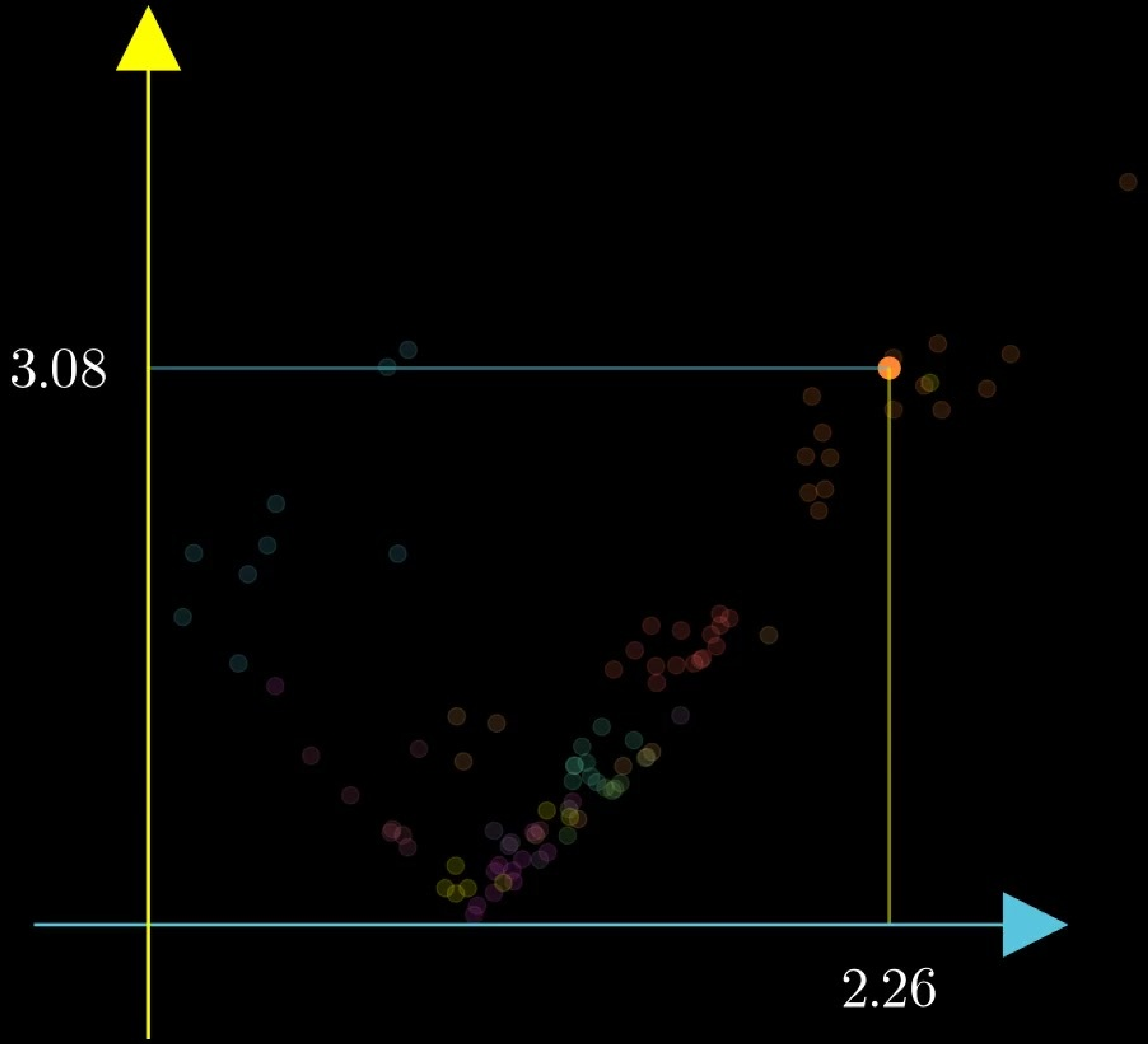
- Parameter initialization
- Data shuffling
- Training seed
- Hyperparameters
- ...

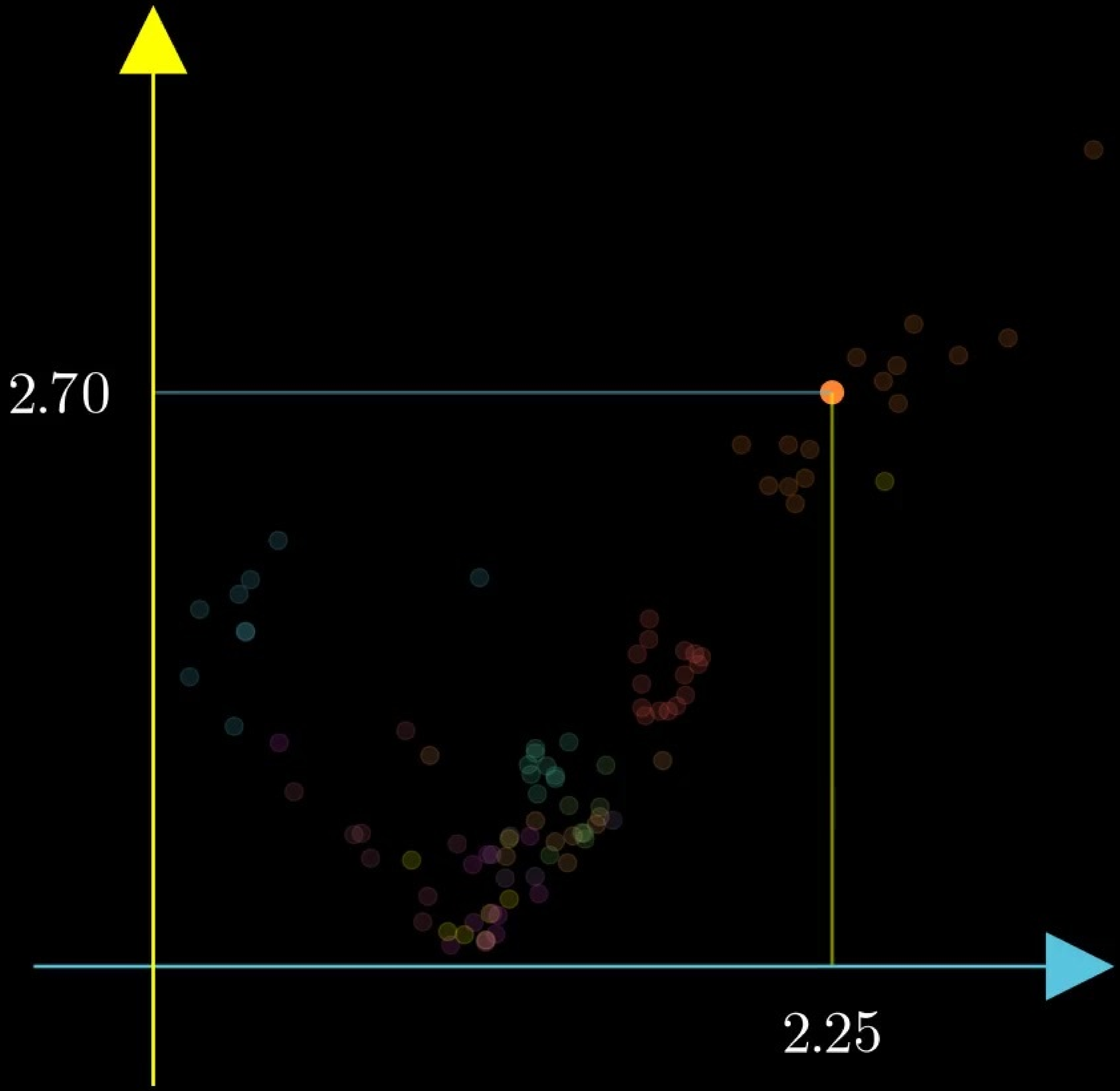


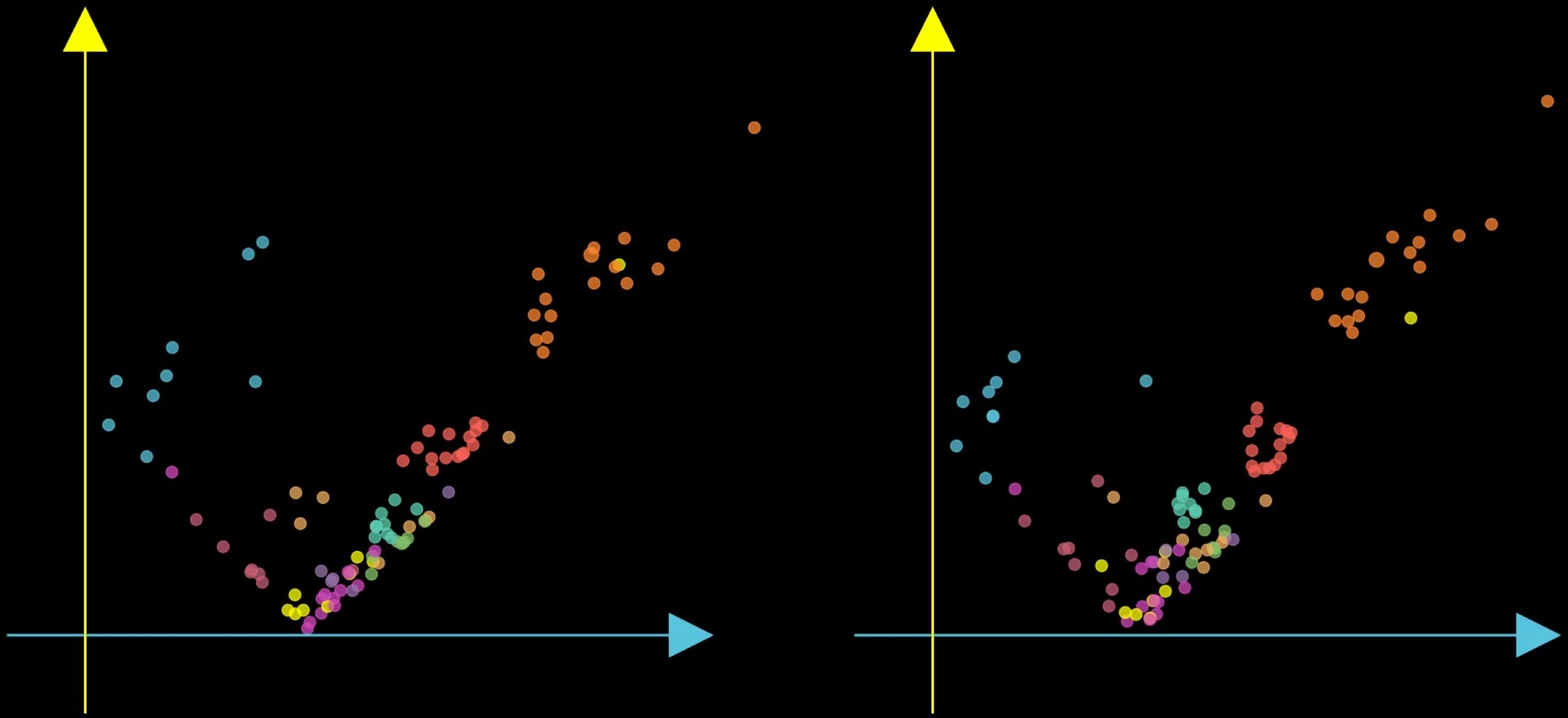





3.08
2.26
2.25
2.70


3.08
2.26


2.70
2.25



Sources of variations in the training process

$\downarrow$

distinct latent spaces

Can we model how the latent spaces are affected?

Yes!

As a near-isometric + rescaling transformation $\mathcal{T}$ of the latent space

Yes!

As a near-isometric + rescaling transformation $\mathcal{T}$ of the latent space

Yes!

As a near-isometric + rescaling transformation $\mathcal{T}$ of the latent space

...we need a representation invariant to $\mathcal{T}$

e

$\mathcal{T}$

Relative Representations

Relative Representations

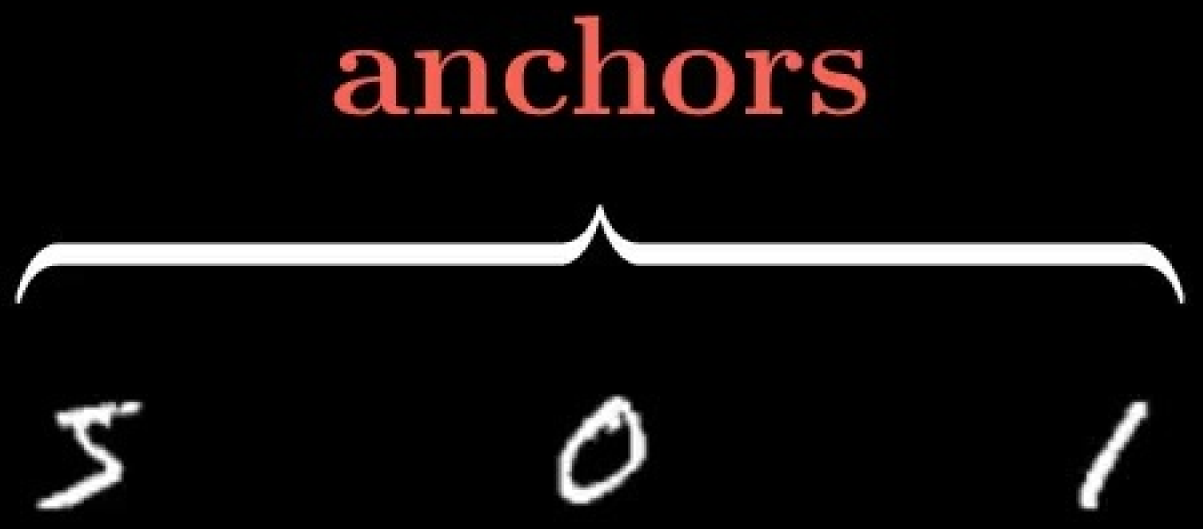
Algorithm

- Select a subset $\mathbb{A}$ of the training set $\mathbb{X}$, denoted **anchors**
- Consider each **sample** x
- Consider an encoding function E
- The **relative representation** of x is:

$$\mathbf{r}_{\mathbf{x}^{(i)}} = (sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(1)}}), sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(2)}}), \ldots, sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(|\mathbb{A}|)}}))$$

Algorithm

- Select a subset $\mathbb{A}$ of the training set $\mathbb{X}$, denoted **anchors**



- Consider each **sample** x

- Consider an encoding function E

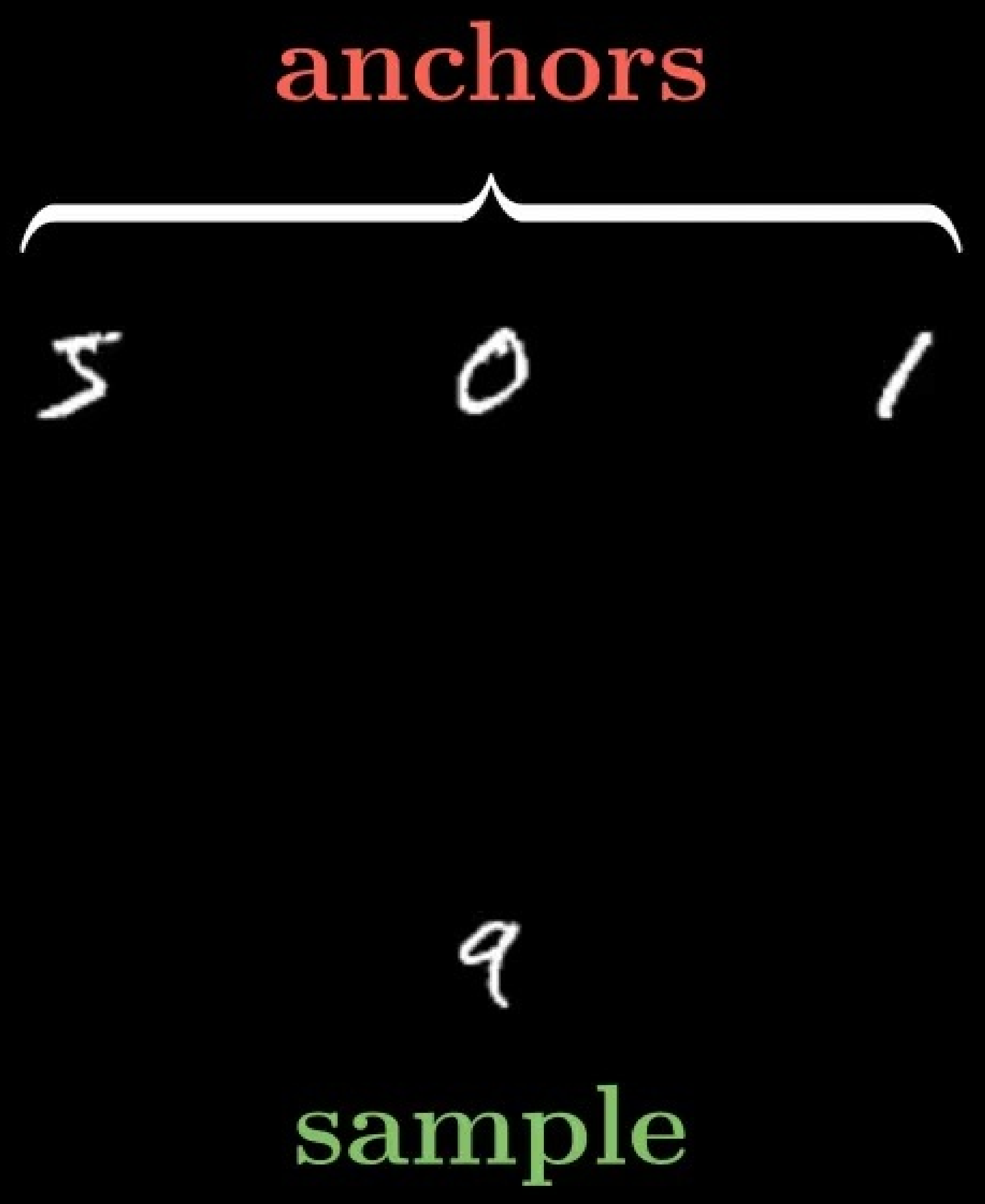
- The **relative representation** of x is:

$$\mathbf{r}_{\mathbf{x}^{(i)}} = \left(sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(1)}}), sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(2)}}), \ldots, sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(|\mathbb{A}|)}})\right)$$

Algorithm

- Select a subset $\mathbb{A}$ of the training set $\mathbb{X}$, denoted **anchors**
- Consider each **sample** x
- Consider an encoding function E
- The **relative representation** of x is:

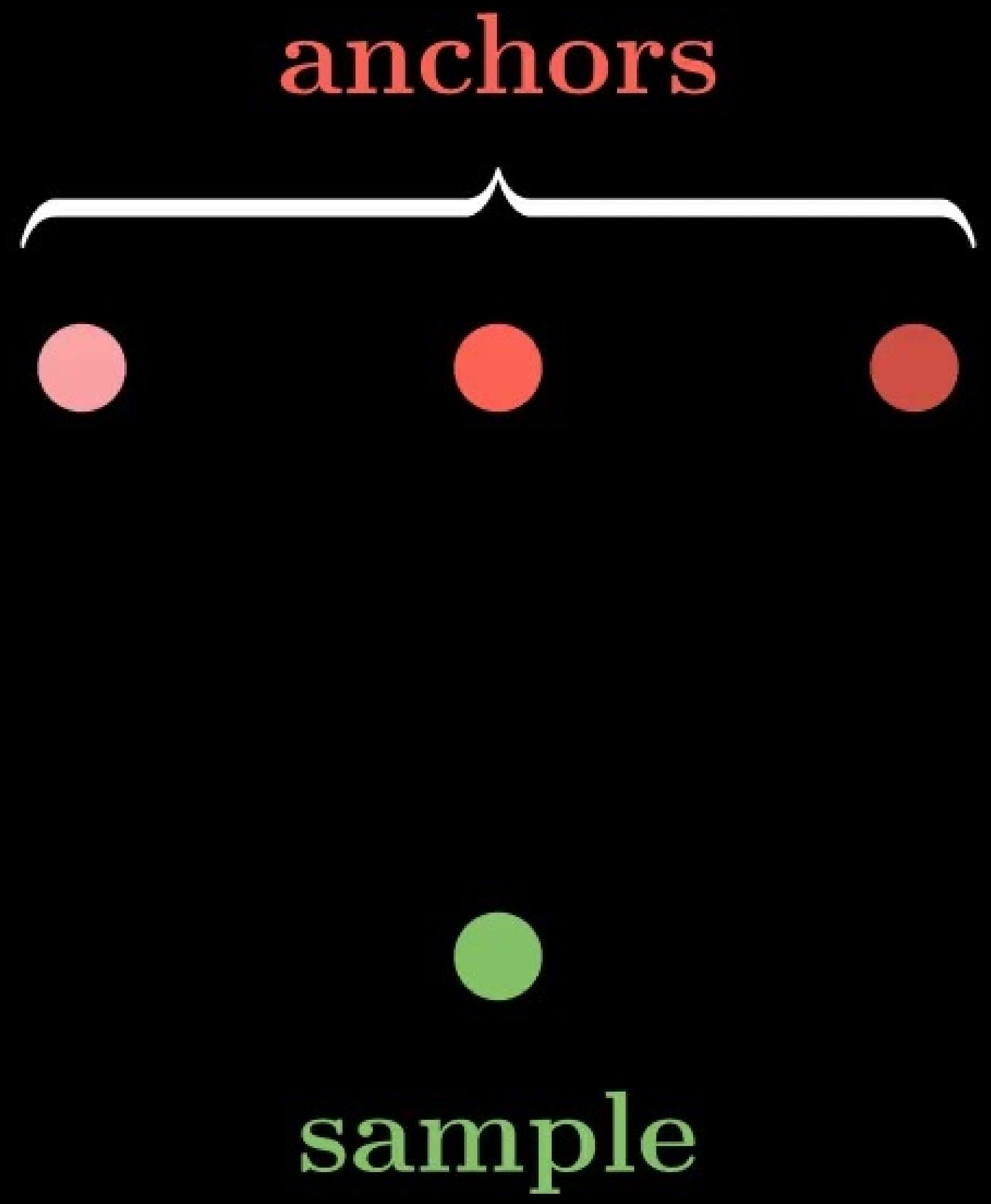
$$\mathbf{r}_{\mathbf{x}^{(i)}} = (sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(1)}}), sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(2)}}), \dots, sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(|\mathbb{A}|)}}))$$



Algorithm

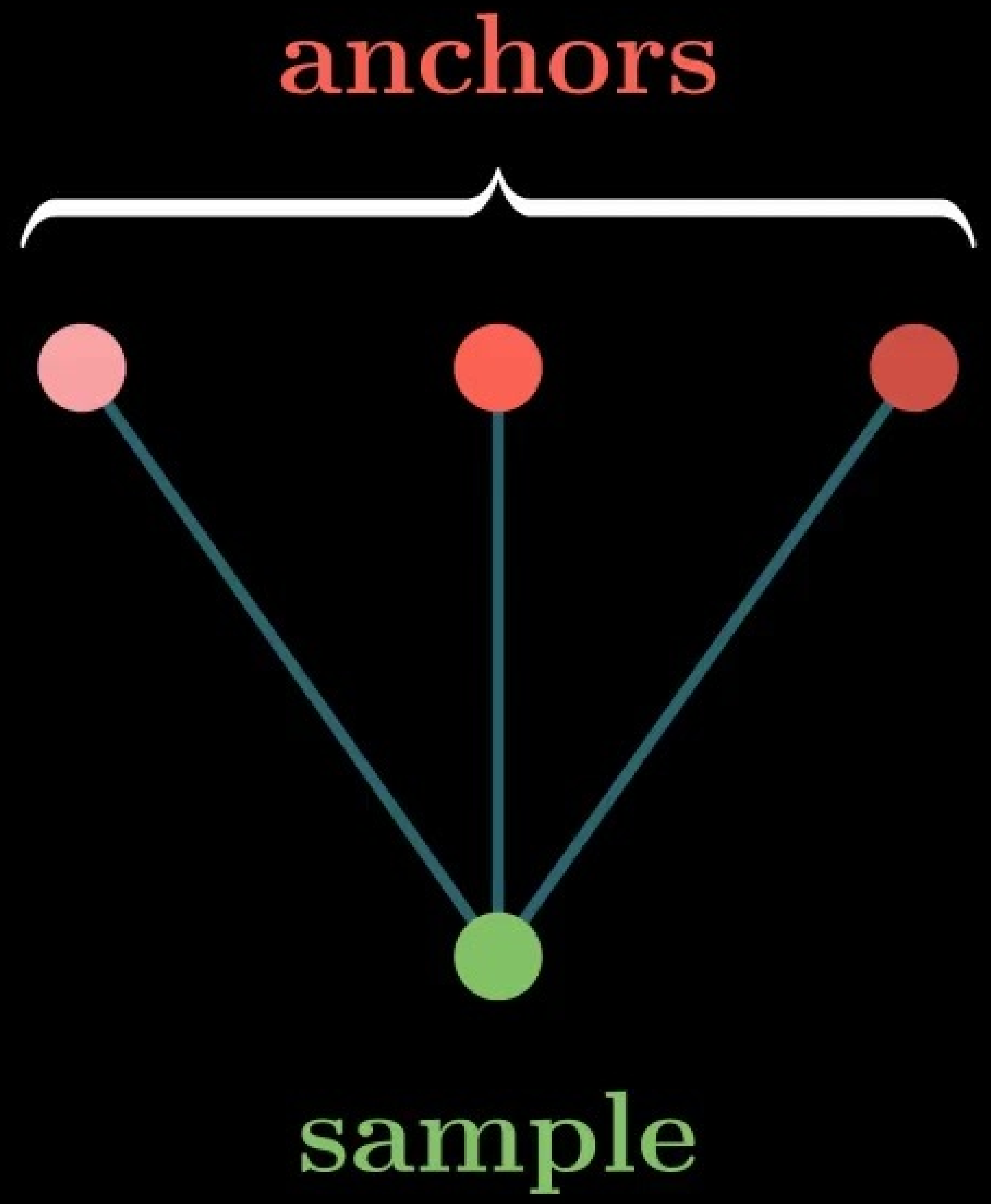
- Select a subset $\mathbb{A}$ of the training set $\mathbb{X}$, denoted **anchors**
- Consider each **sample** x
- Consider an encoding function E
- The **relative representation** of x is:

$$\mathbf{r}_{\mathbf{x}^{(i)}} = (sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(1)}}), sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(2)}}), \ldots, sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(|\mathbb{A}|)}}))$$



Algorithm

- Select a subset $\mathbb{A}$ of the training set $\mathbb{X}$, denoted **anchors**
- Consider each **sample** x
- Consider an encoding function E
- The **relative representation** of x is:



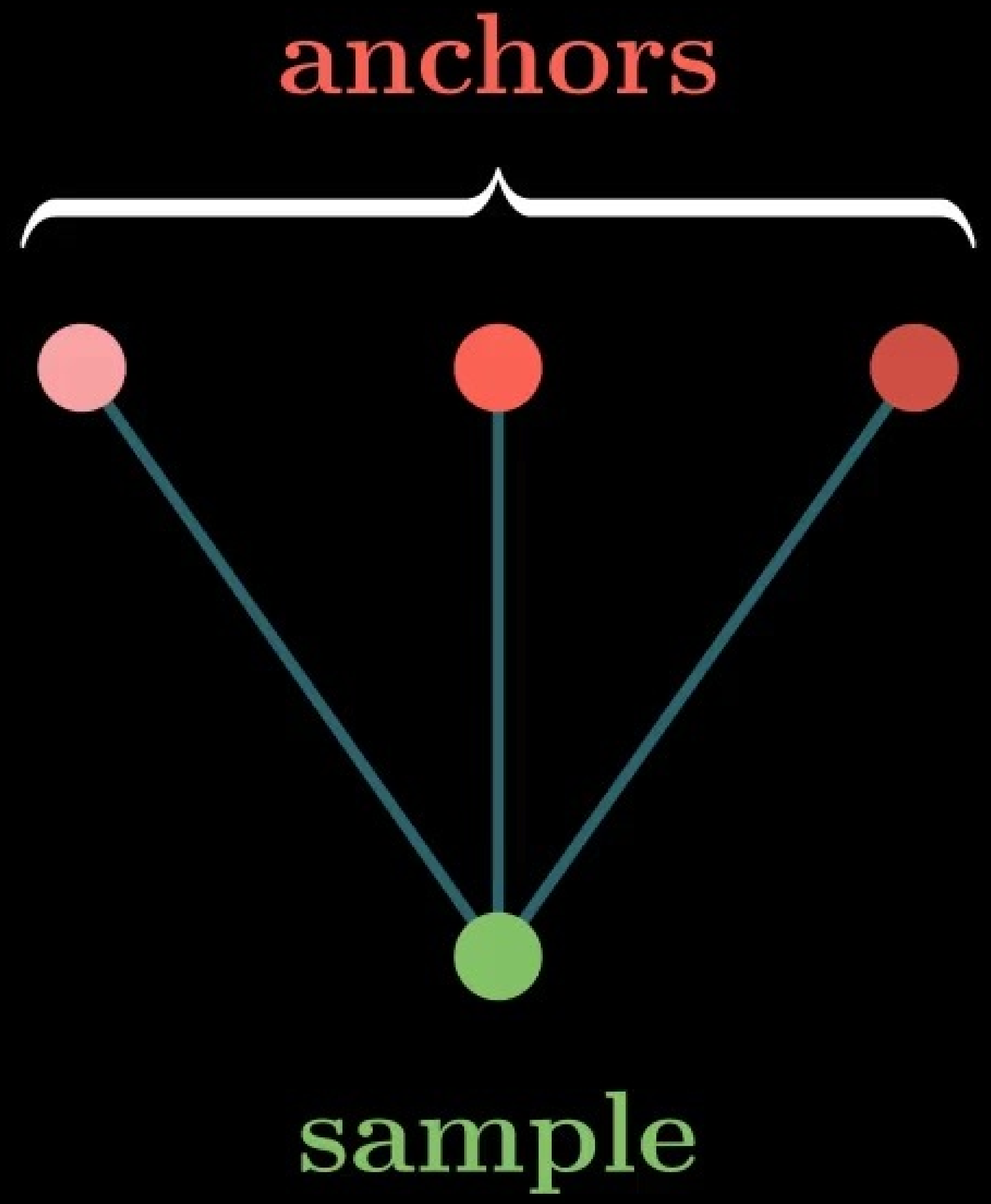
$$\mathbf{r}_{\mathbf{x}^{(i)}} = \left(sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(1)}}), sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(2)}}), \dots, sim(\mathbf{e}_{\mathbf{x}^{(i)}}, \mathbf{e}_{\mathbf{a}^{(|\mathbb{A}|)}})\right)$$

Algorithm

sim = cosine similarity

```
import torch
import torch.nn.functional as F

def relative_projection(x, anchors):
    x = F.normalize(x, p=2, dim=-1)
    anchors = F.normalize(anchors, p=2, dim=-1)
    return torch.einsum("bm, am -> ba", x, anchors)
```



differentiable!

Properties

- The size of $\mathbf{r}_{\mathbf{x}^{(i)}}$ depends on the number of anchors
- The anchors and similarity function choices determine the representation properties

Properties

- The size of $\mathbf{r}_{\mathbf{x}^{(i)}}$ depends on the number of anchors
- The anchors and similarity function choices determine the representation properties

Properties

- The size of $\mathbf{r}_{\mathbf{x}^{(i)}}$ depends on the number of anchors
- The anchors and similarity function choices determine the representation properties

Properties

- The size of $\mathbf{r}_{\mathbf{x}^{(i)}}$ depends on the number of anchors
- The anchors and similarity function choices determine the representation properties

$$\boxed{\textbf{sim} = \text{cosine similarity}}$$

$$\downarrow$$

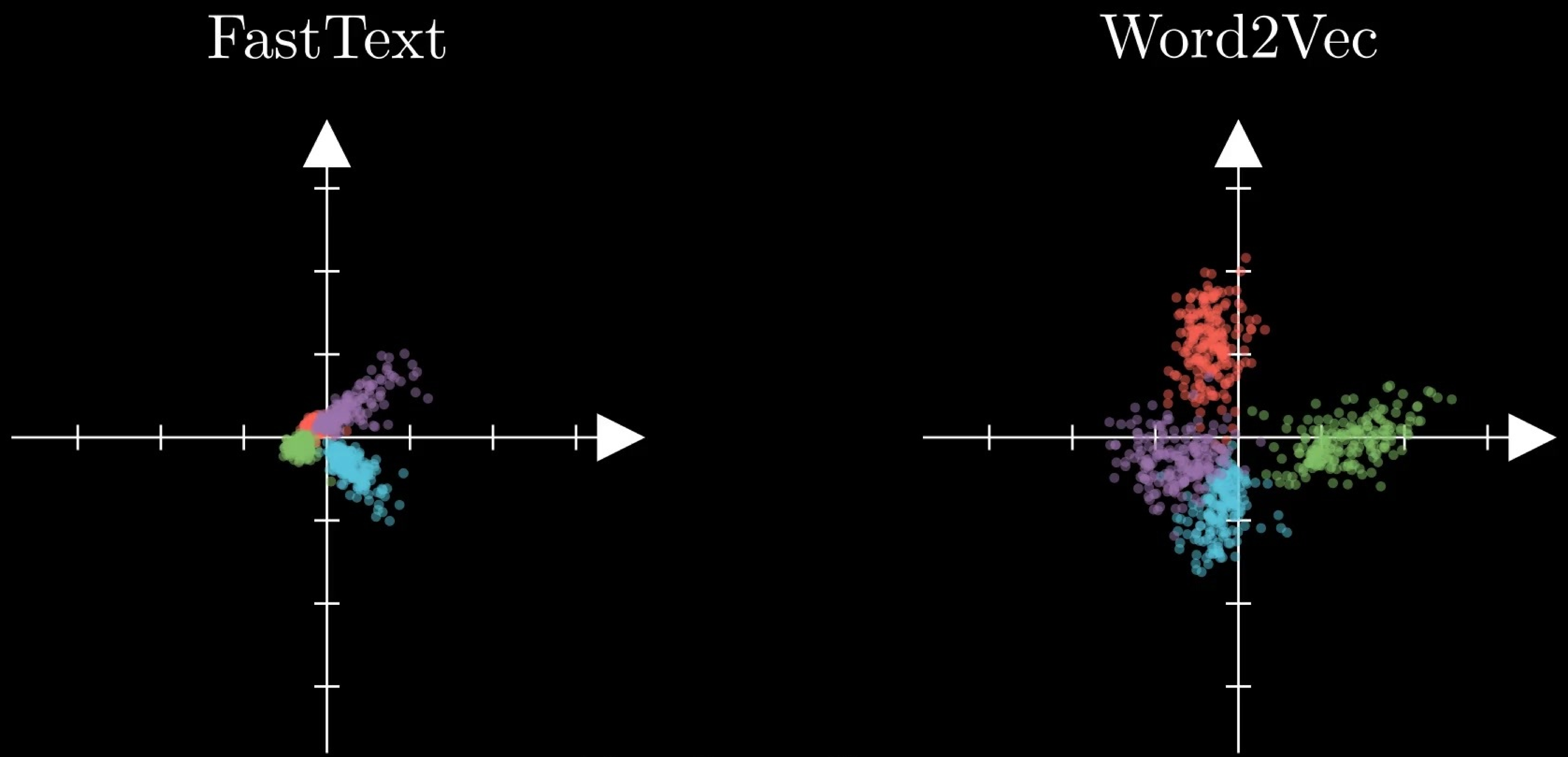
invariant to **rescaled isometric transformations** of the latent space es

e

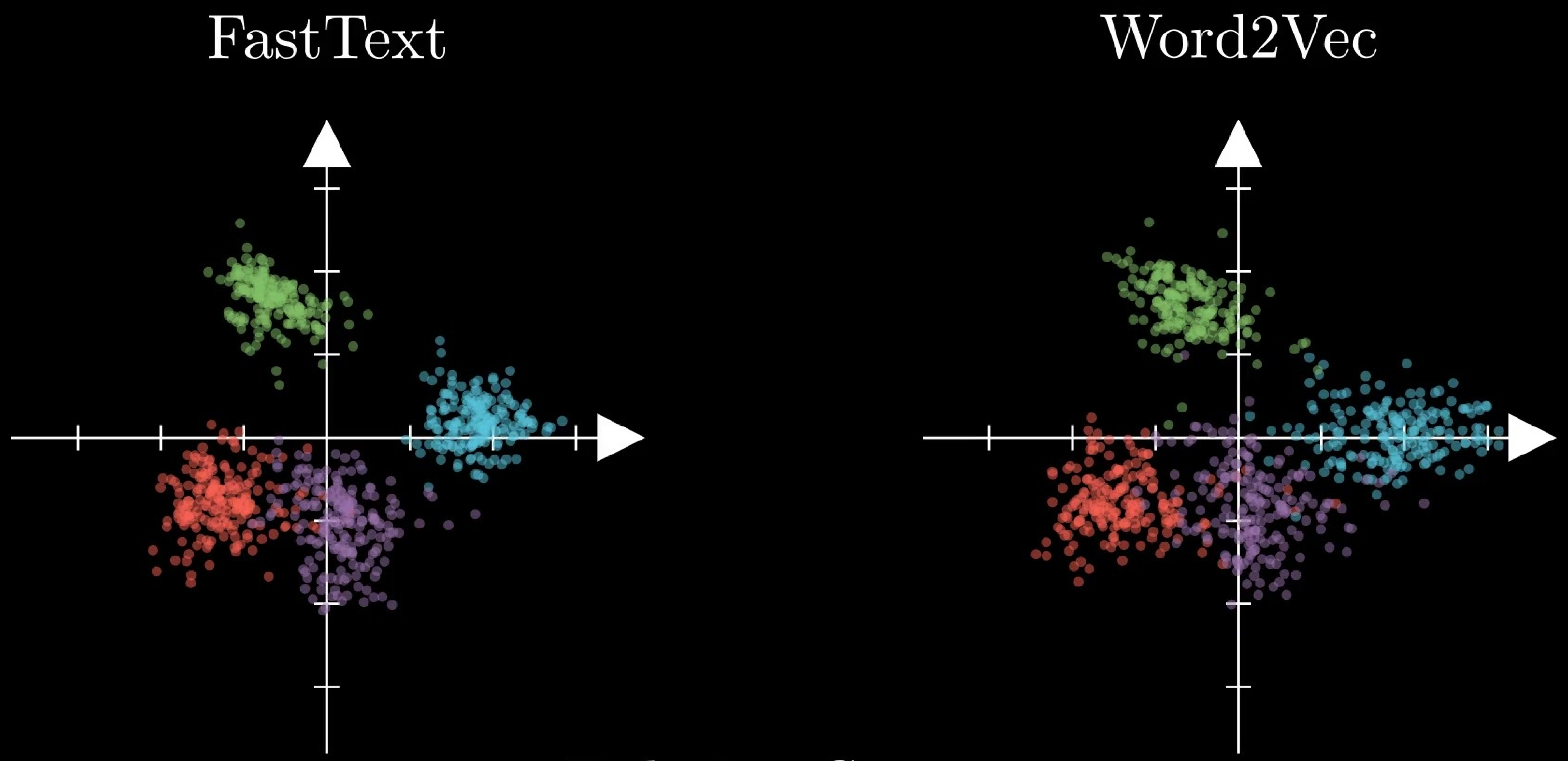
Semantic Invariance

Given some *words*...

...consider their *embeddings* in different spaces



Absolute Spaces



Relative Spaces

latent spaces with the same semantics

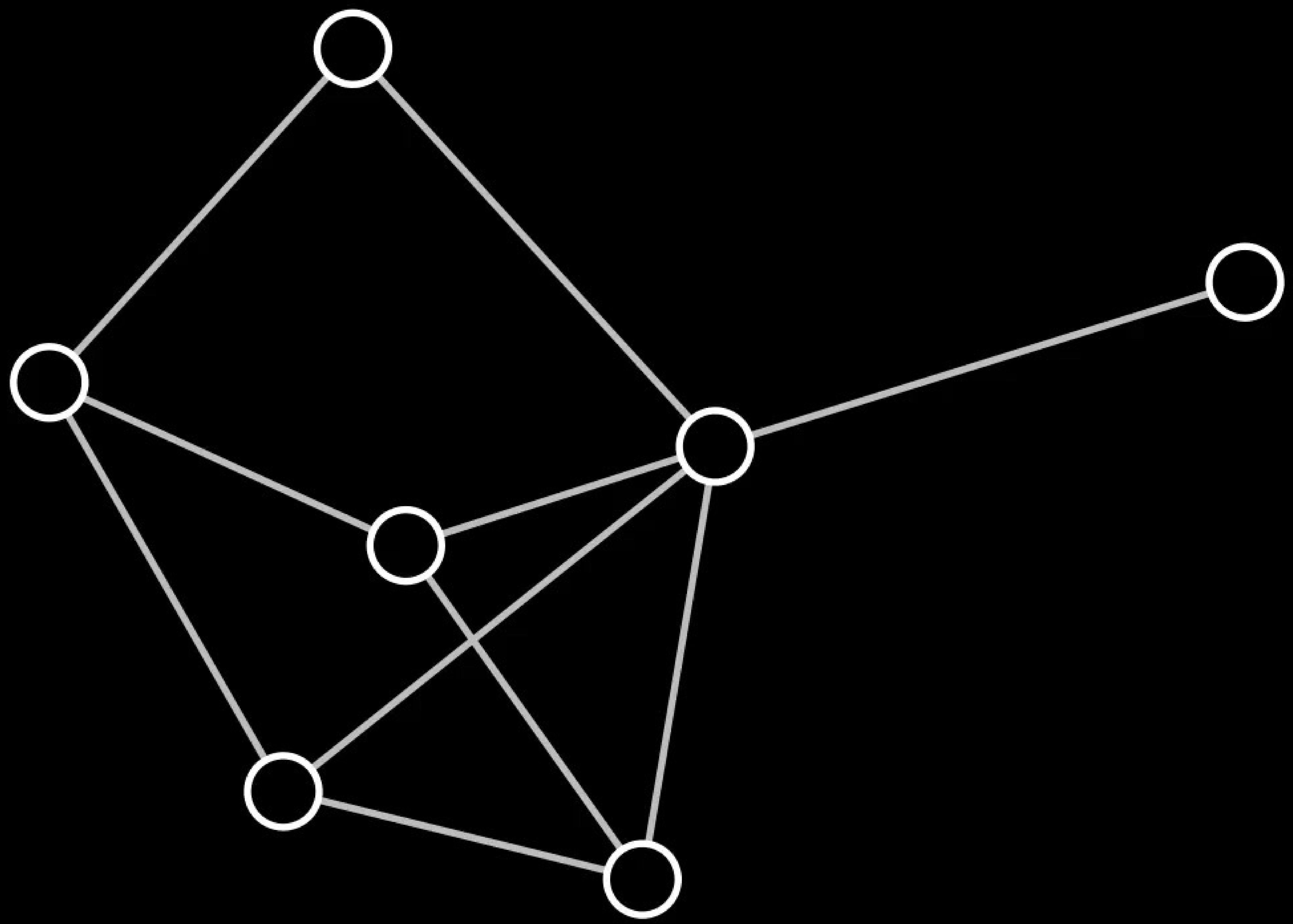
$\downarrow$

represented **similarly** in the relative space

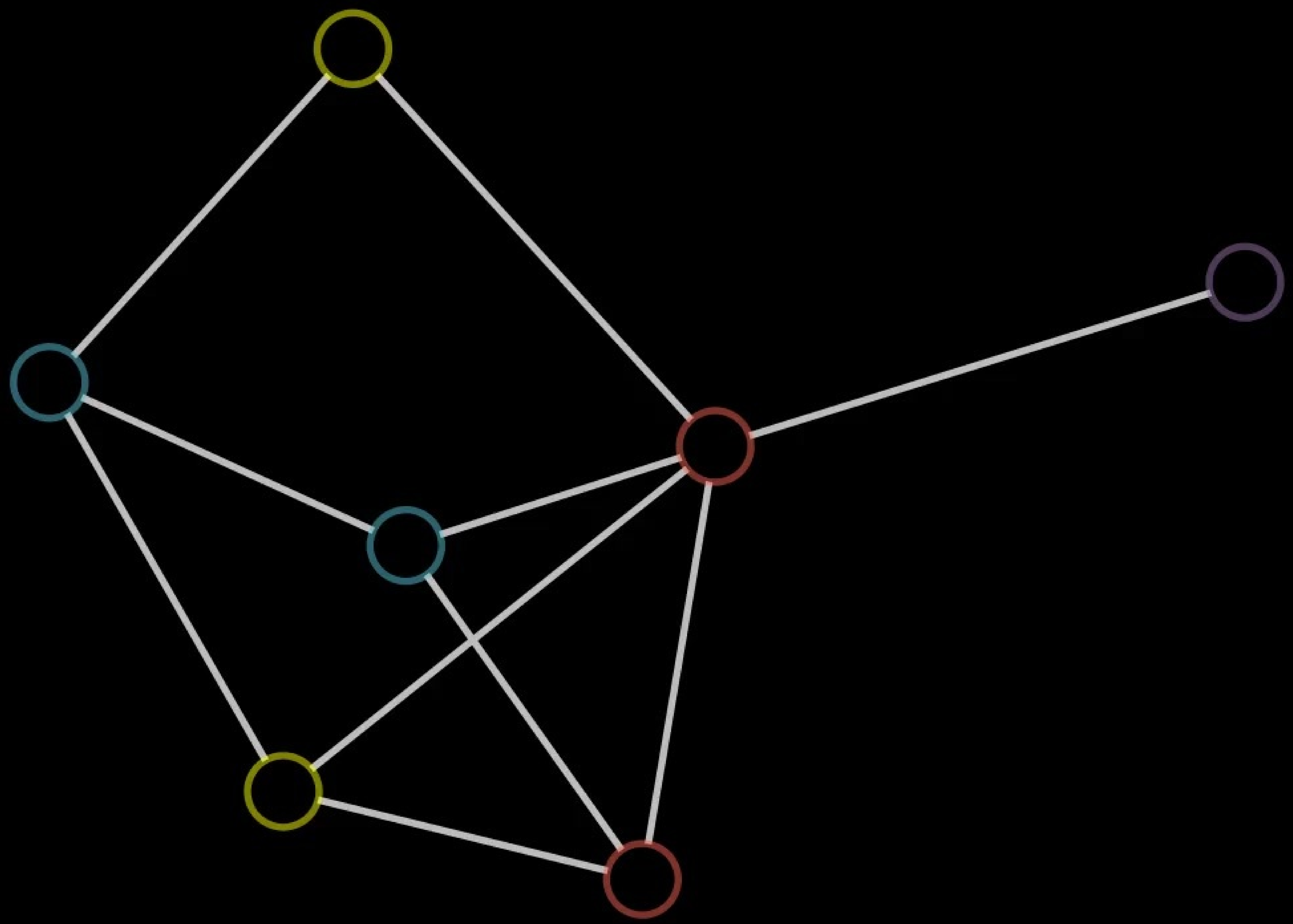
e

Latent Performance Metric

Consider a node classification task...



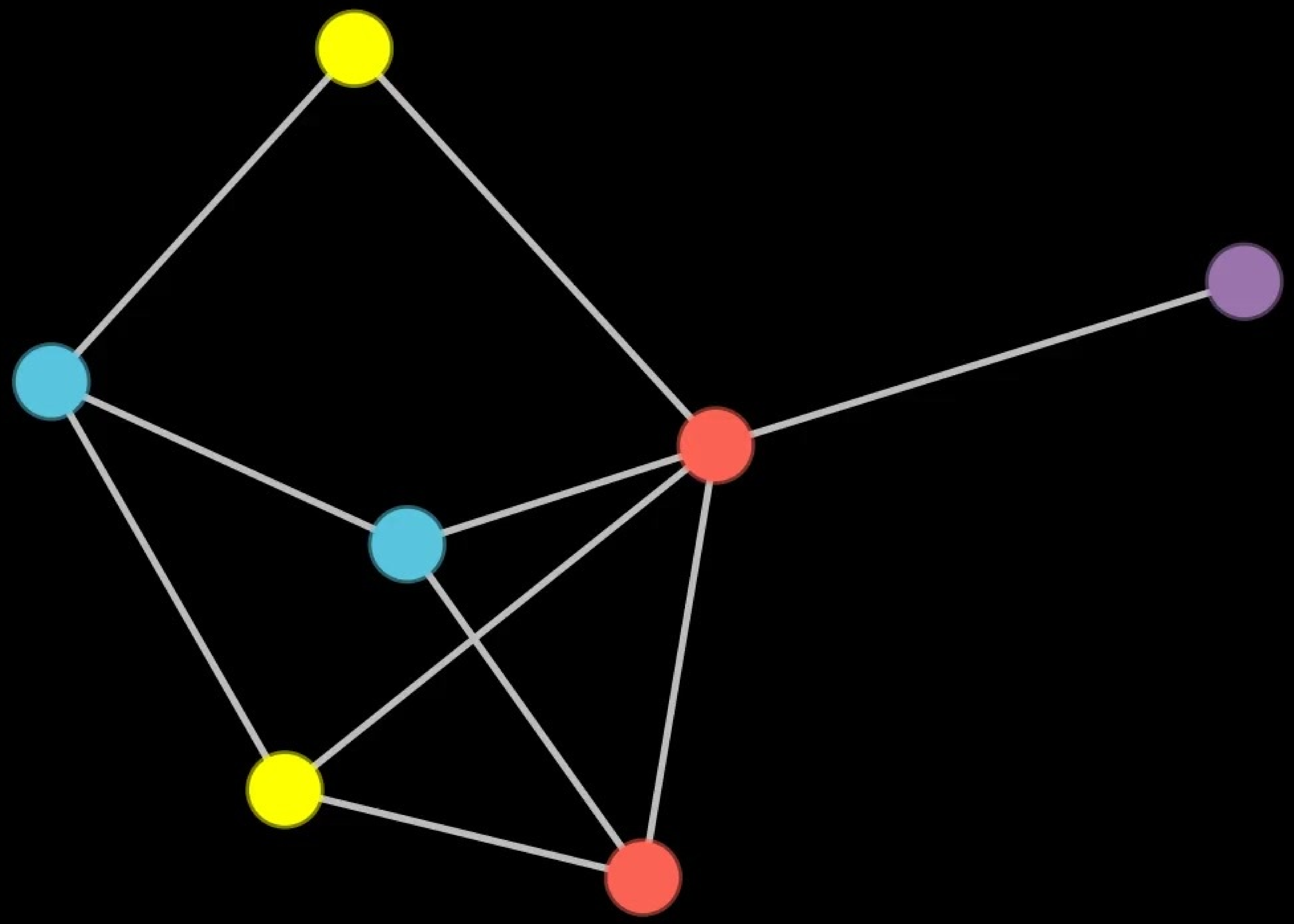
Cora dataset



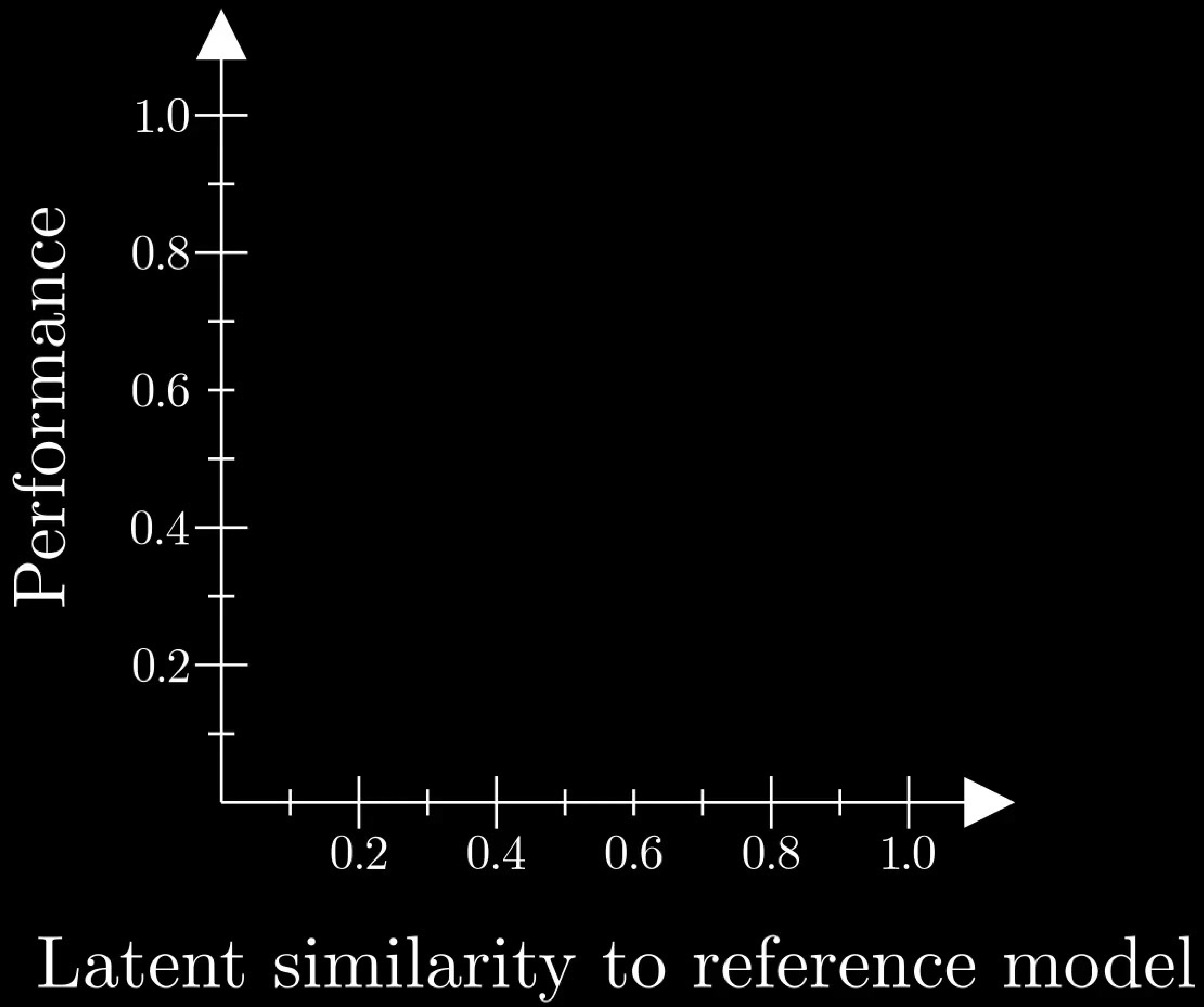
Cora dataset

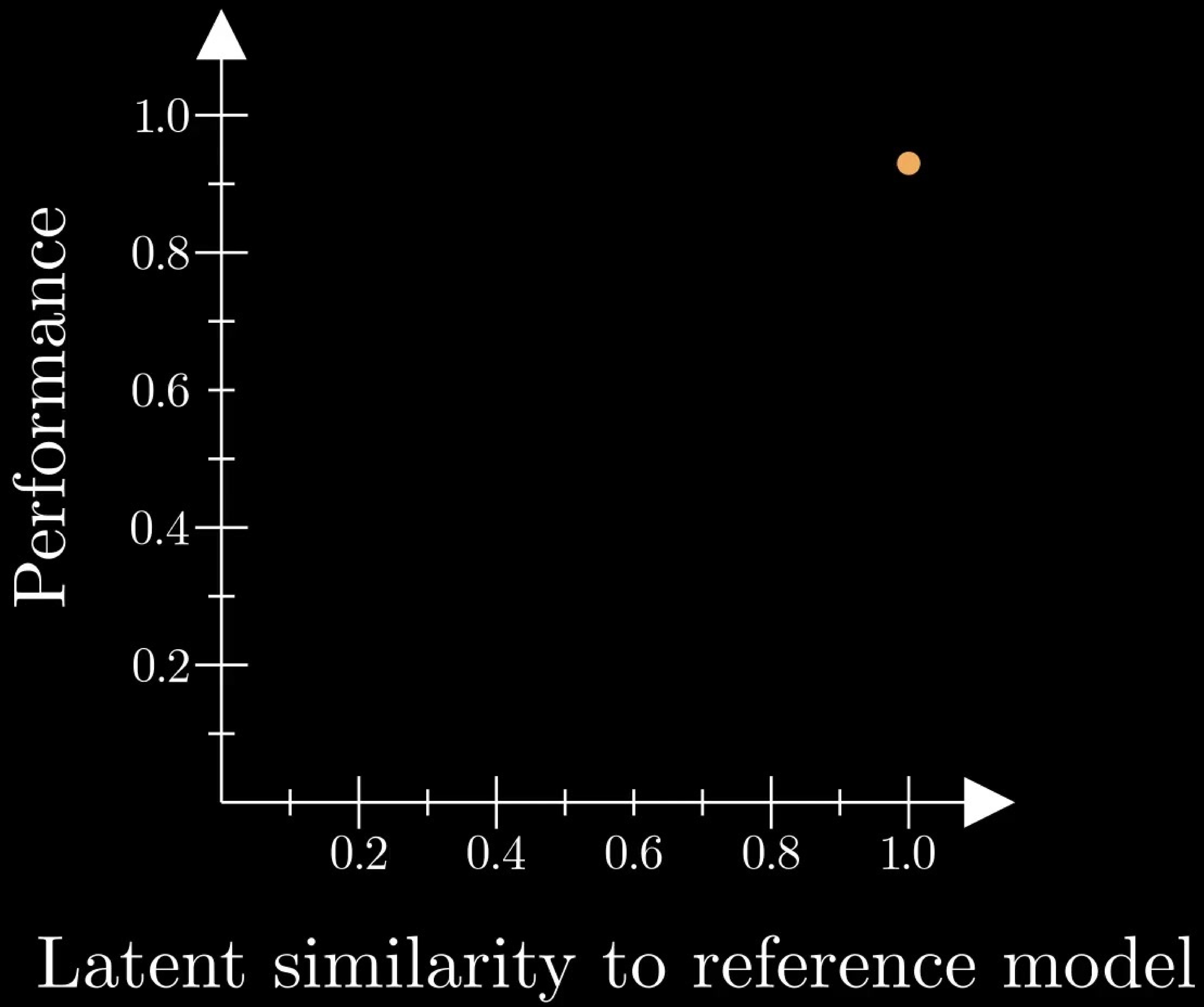
given a good

reference model

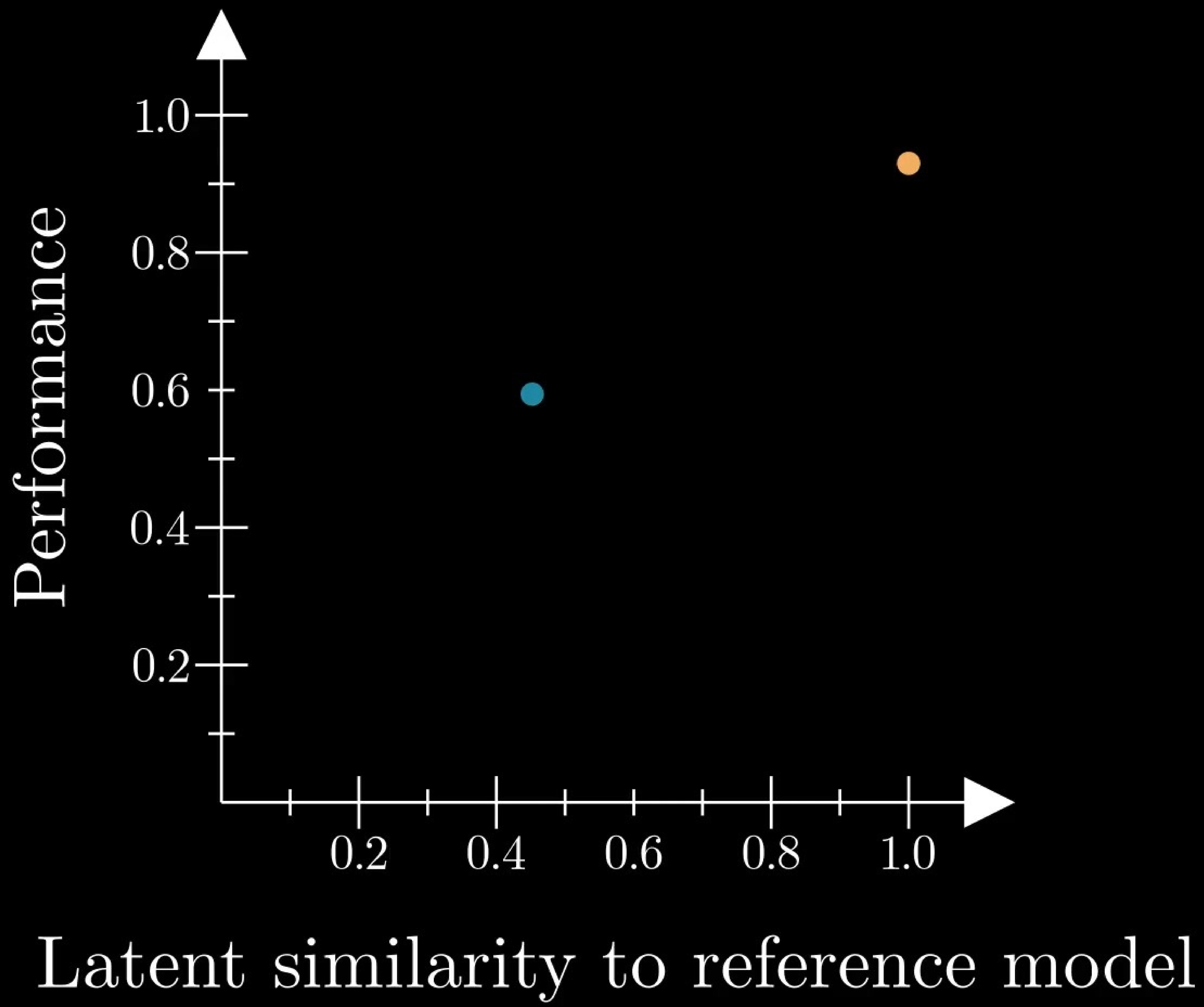


We train other models...

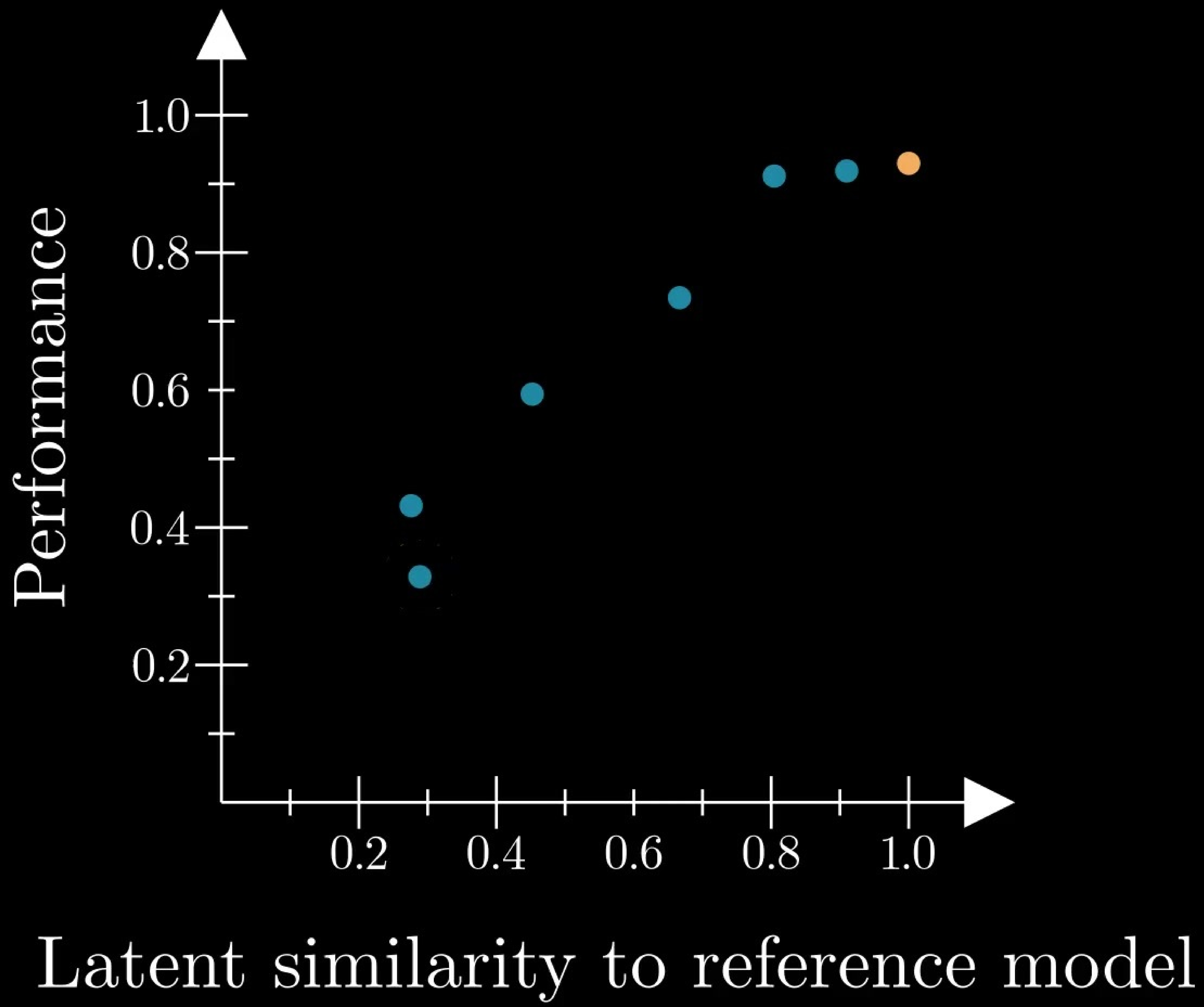

Performance
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
Latent similarity to reference model



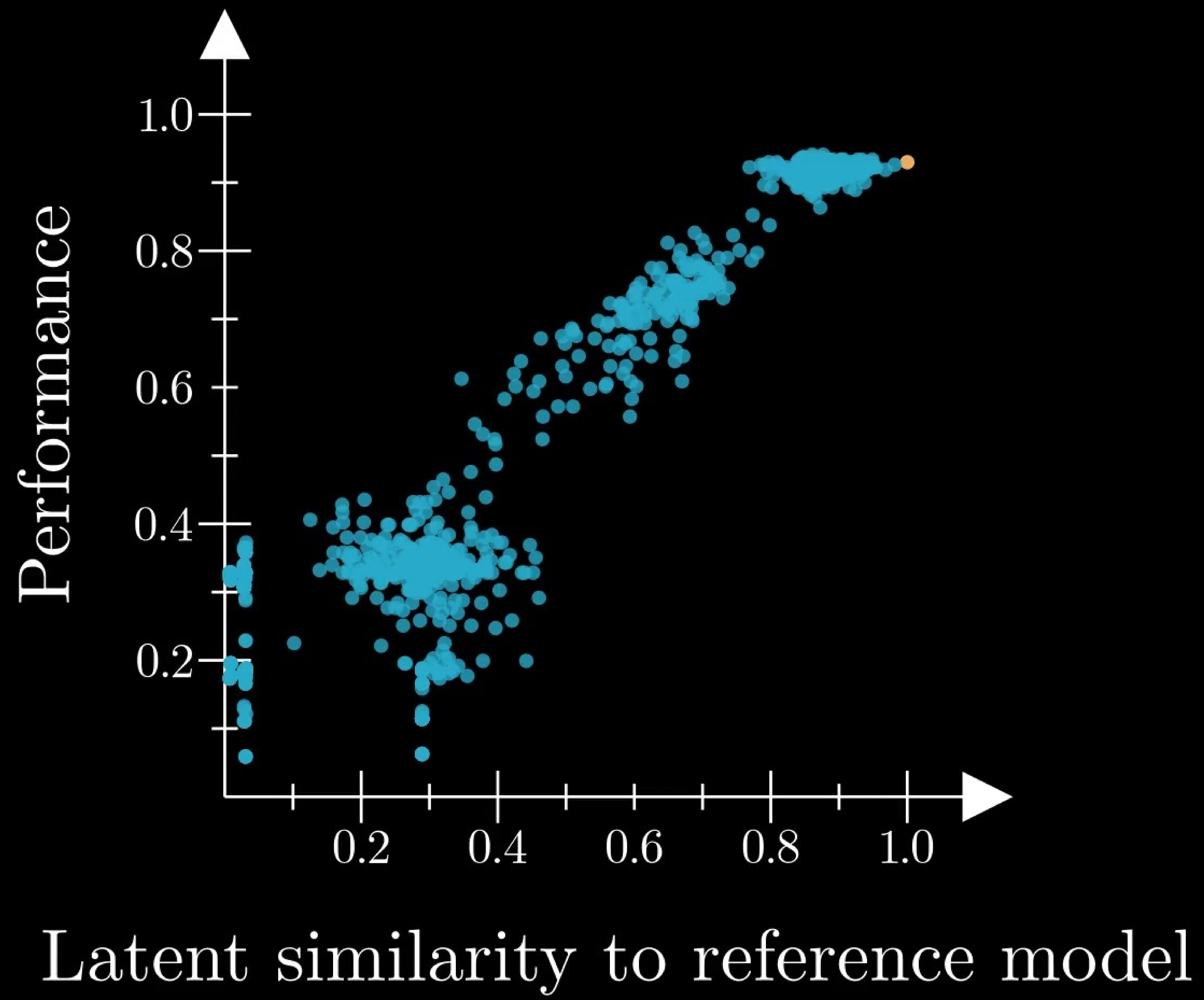
Seed	1
Epochs	500
Layers	32
Dropout	0.5
Hidden act.	ReLU
Conv. act.	ReLU
Optimizer	Adam
Learning rate	0.02
Encoder	GCN2Conv



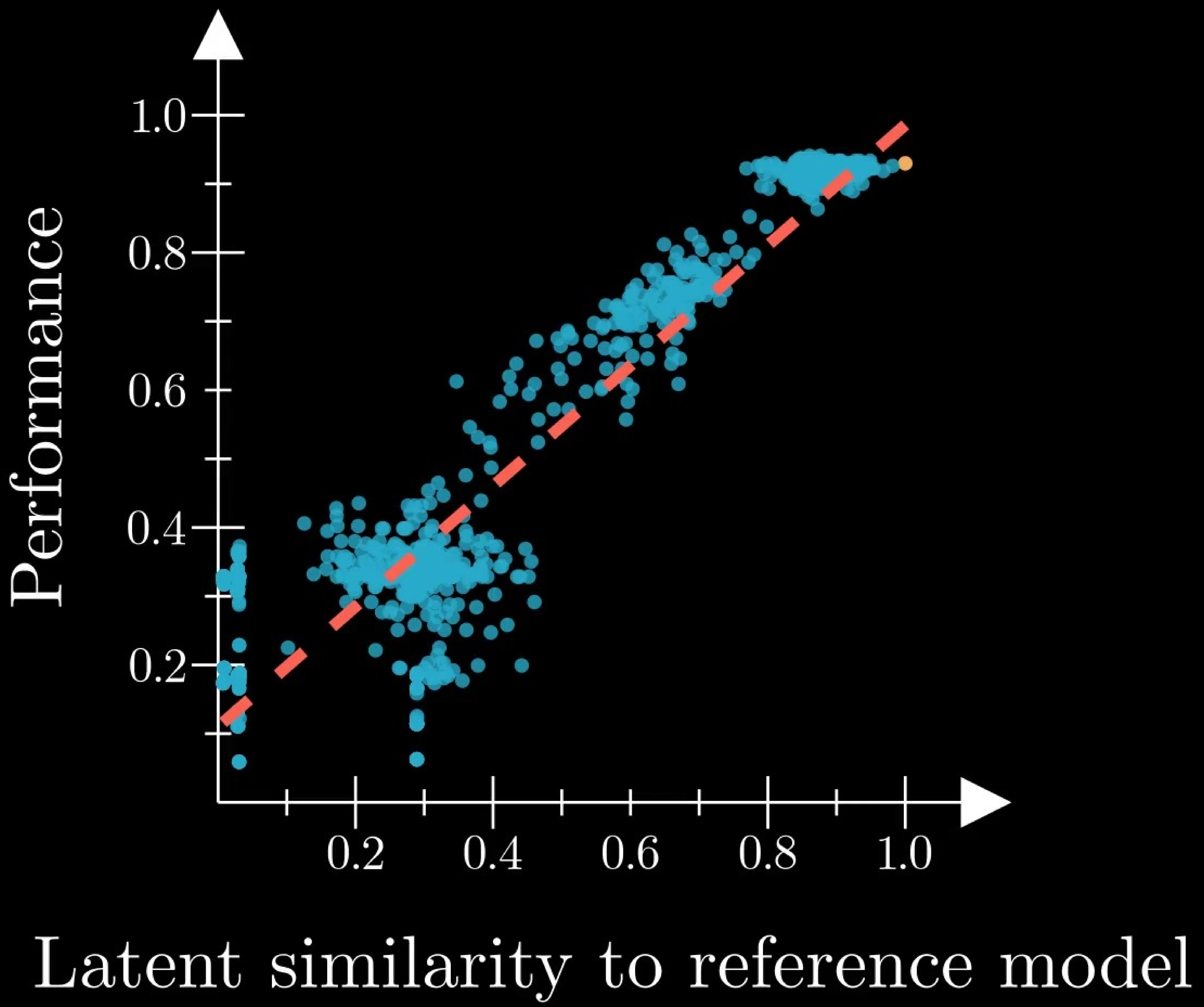
Seed	0
Epochs	10
Layers	32
Dropout	0.1
Hidden act.	ReLU
Conv. act.	ReLU
Optimizer	Adam
Learning rate	0.02
Encoder	GCN2Conv



Seed	2
Epochs	30
Layers	64
Dropout	0.1
Hidden act.	Tanh
Conv. act.	Tanh
Optimizer	Adam
Learning rate	0.02
Encoder	GINConv



Performance
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
Latent similarity to reference model



Performance
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
Latent similarity to reference model

relative space similar to the reference model

$\downarrow$

good performance

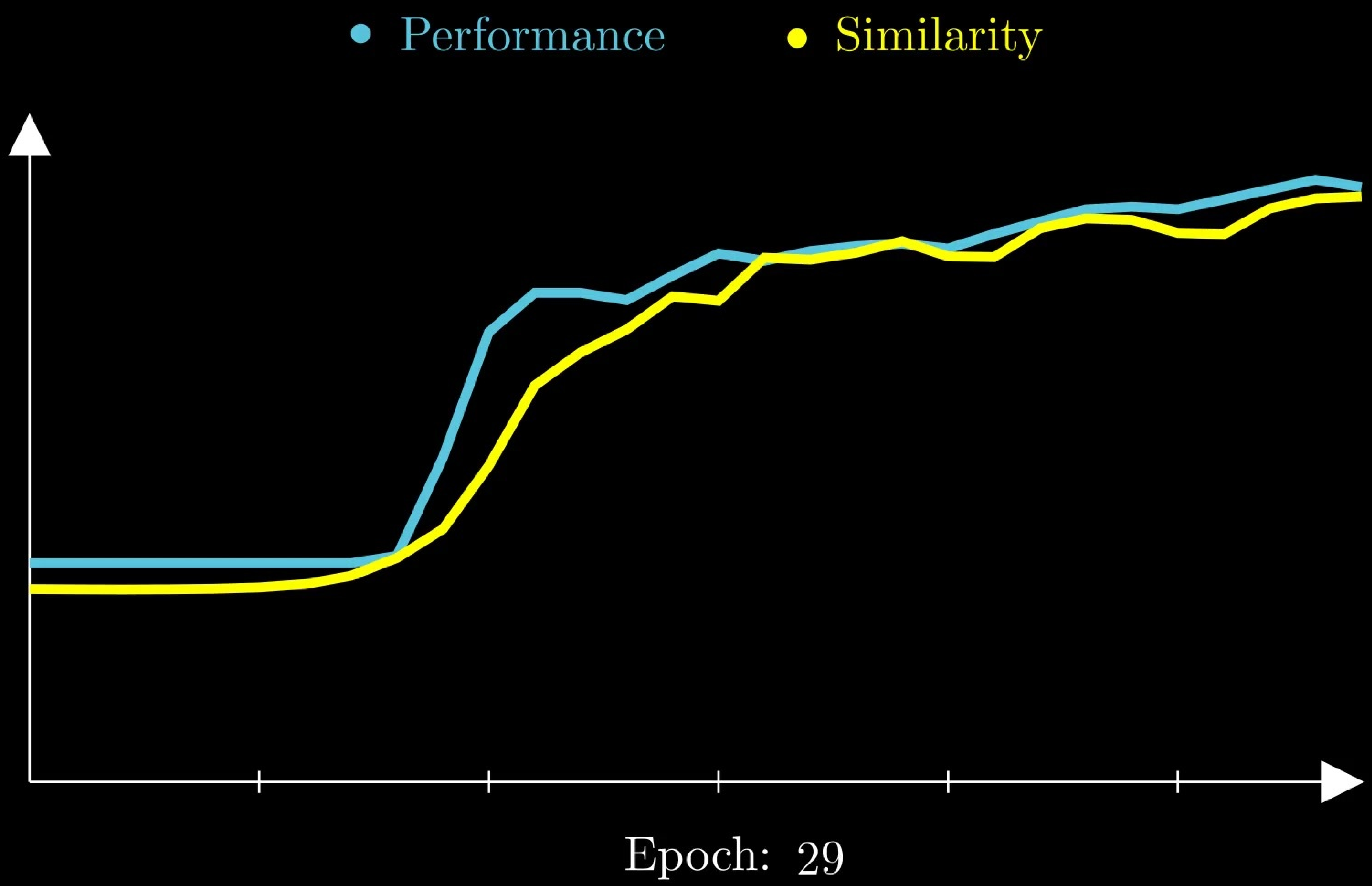
relative space similar to the reference model

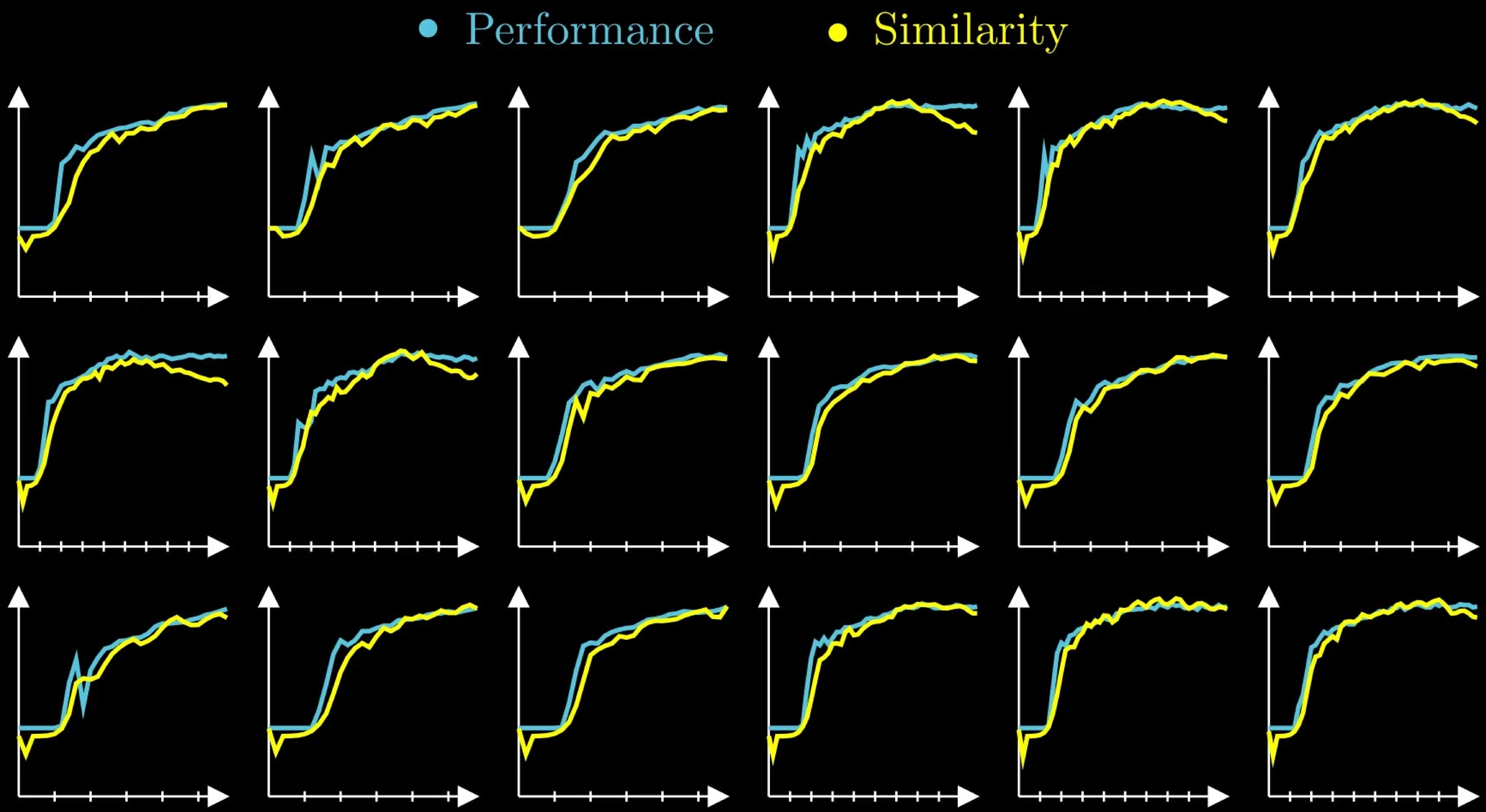
$$\downarrow \qquad \not\uparrow$$

good performance

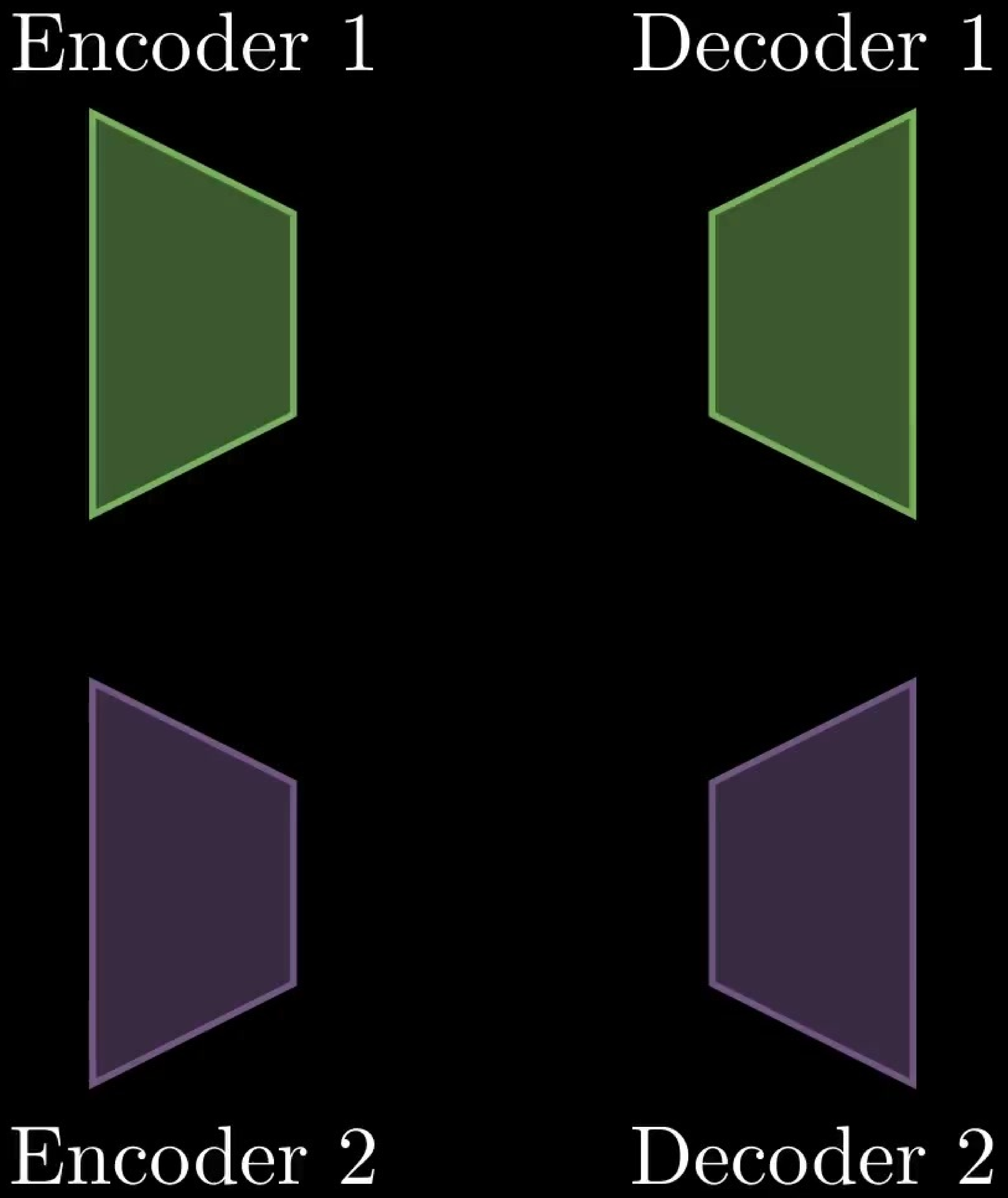
...zooming into

a single model training


Performance
Similarity
Epoch: 29

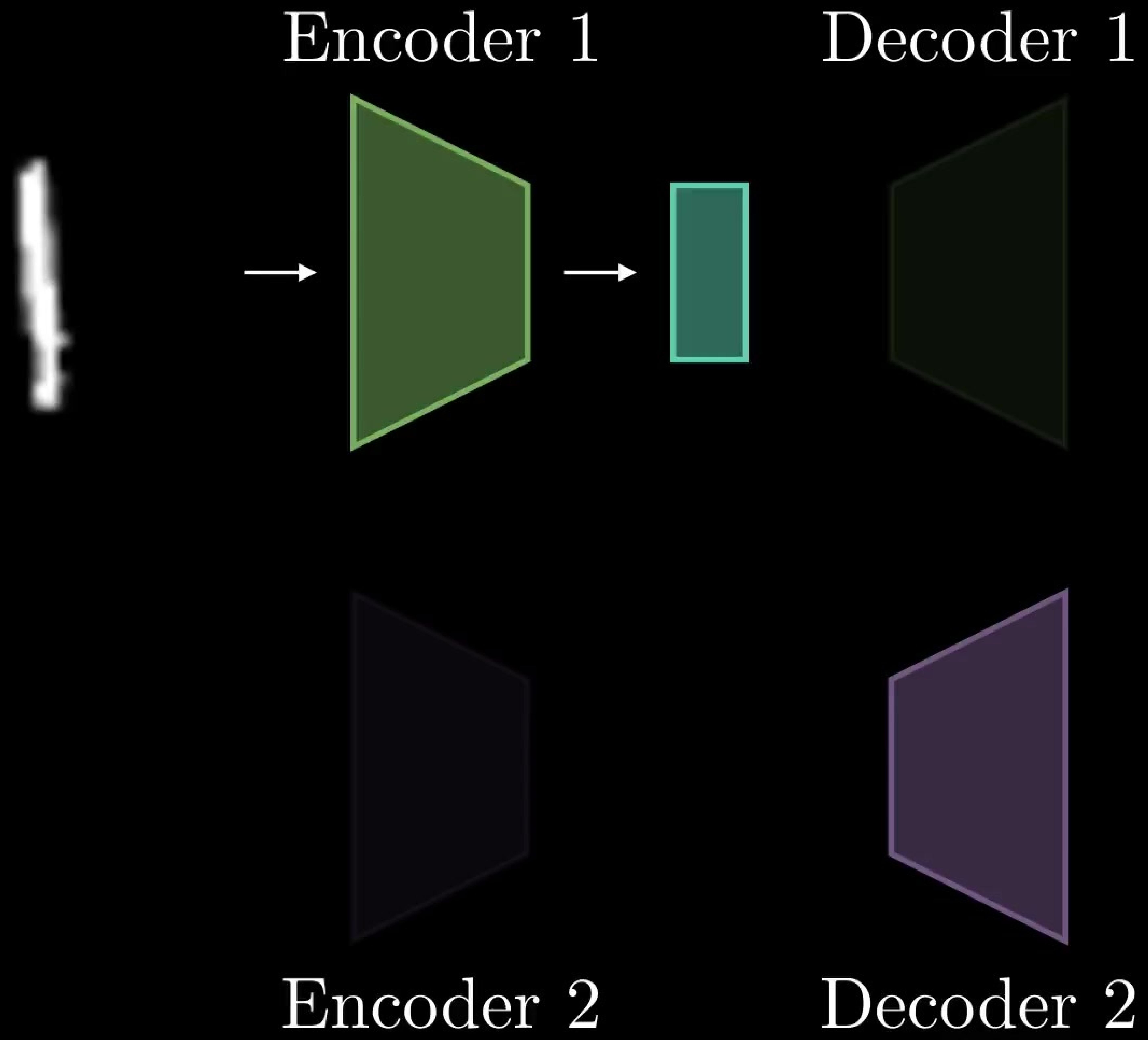

Performance
Similarity

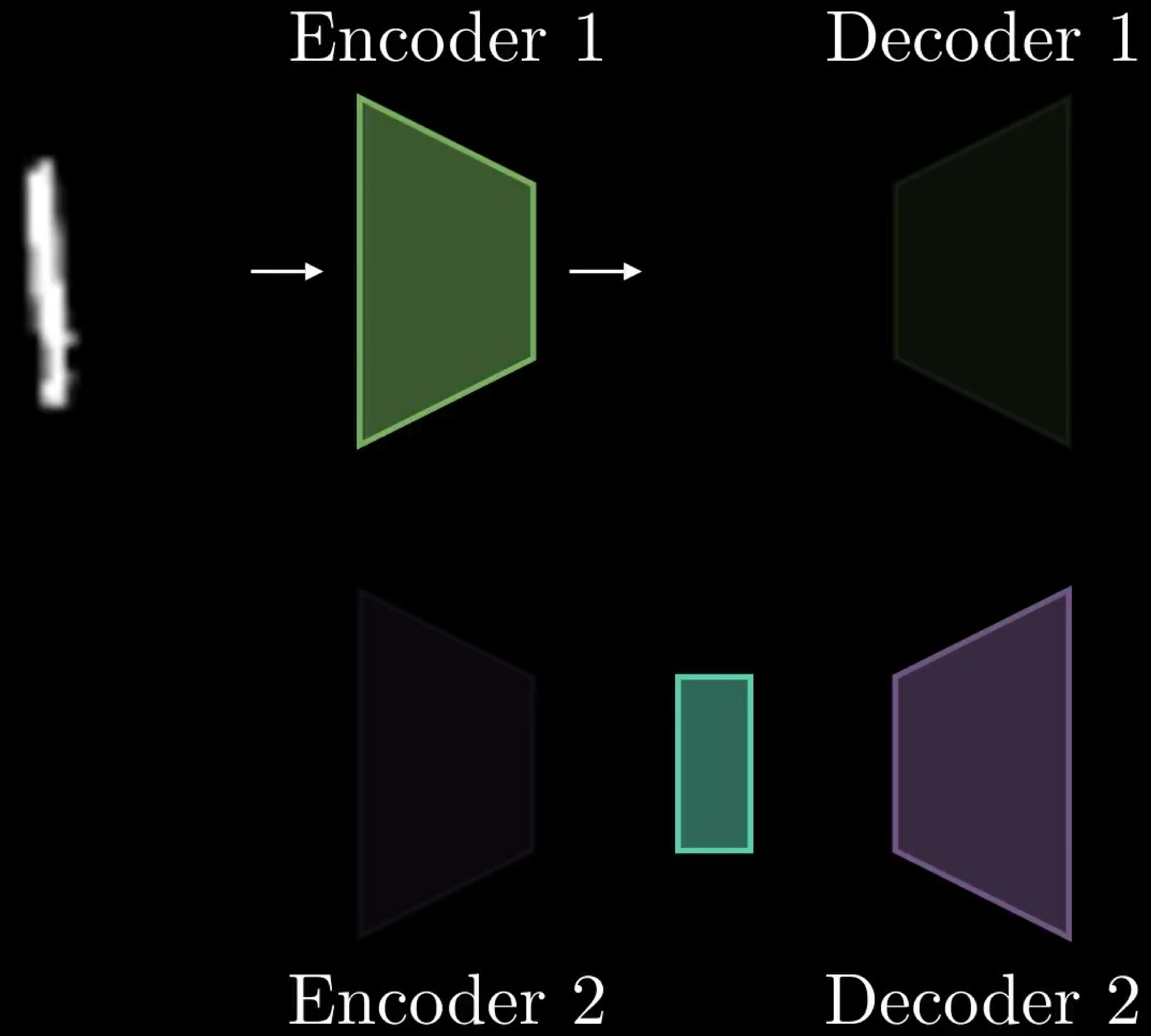
Zero-shot Stitching

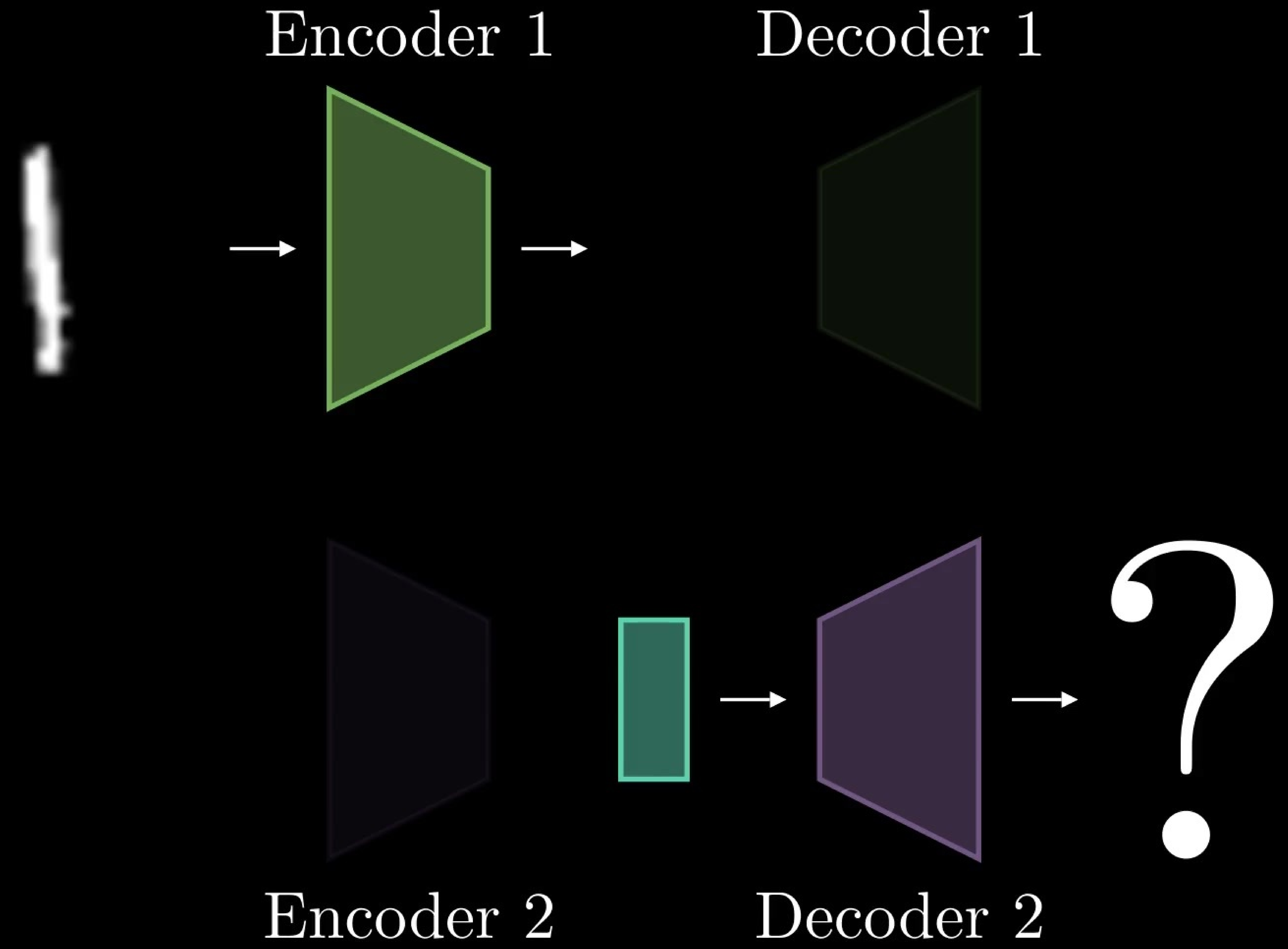


Encoder 1
Decoder 1
Encoder 2
Decoder 2


Encoder 1
Decoder 1
Encoder 2
Decoder 2


Encoder 1
Decoder 1
Encoder 2
Decoder 2


Encoder 1
Decoder 1
Encoder 2
Decoder 2



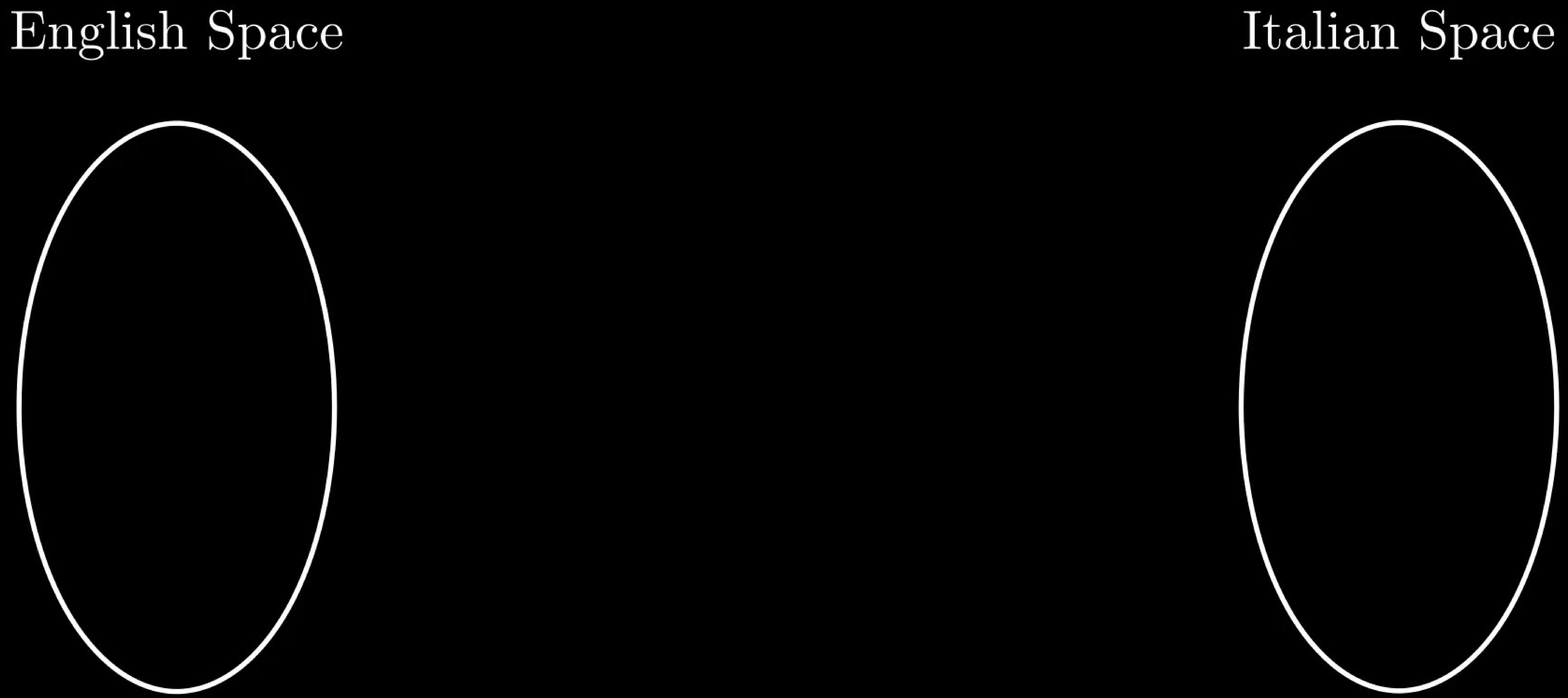
Encoder 1
Decoder 1
?
Encoder 2
Decoder 2

$Fashion - MNIST$
$MNIST$
Identity
AE
Variational AE
Relative AE

E

Parallel Anchors

...stitching between multiple domains?



...solve a downstream task **once for both spaces**

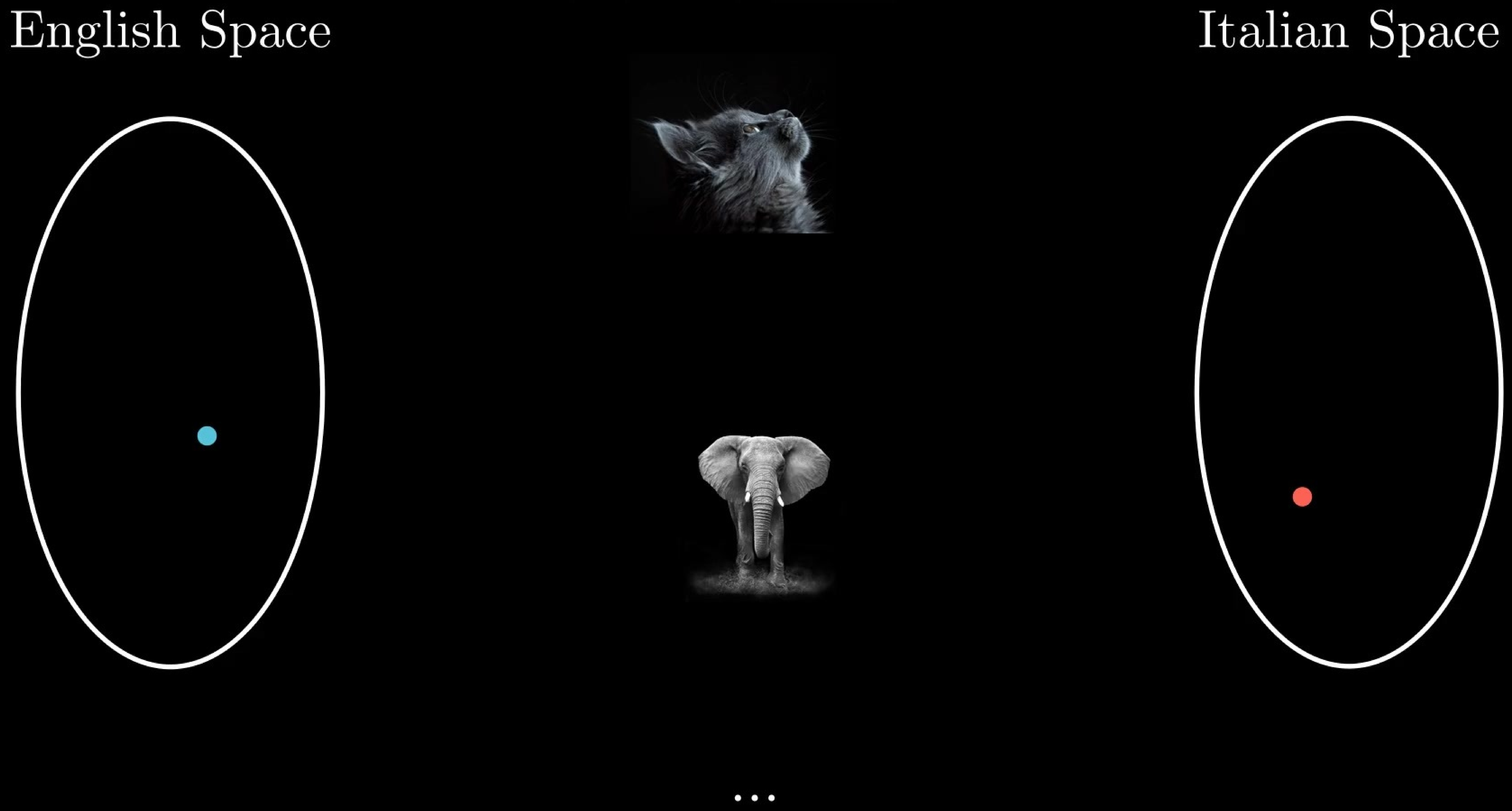
Parallel Anchors

English Space

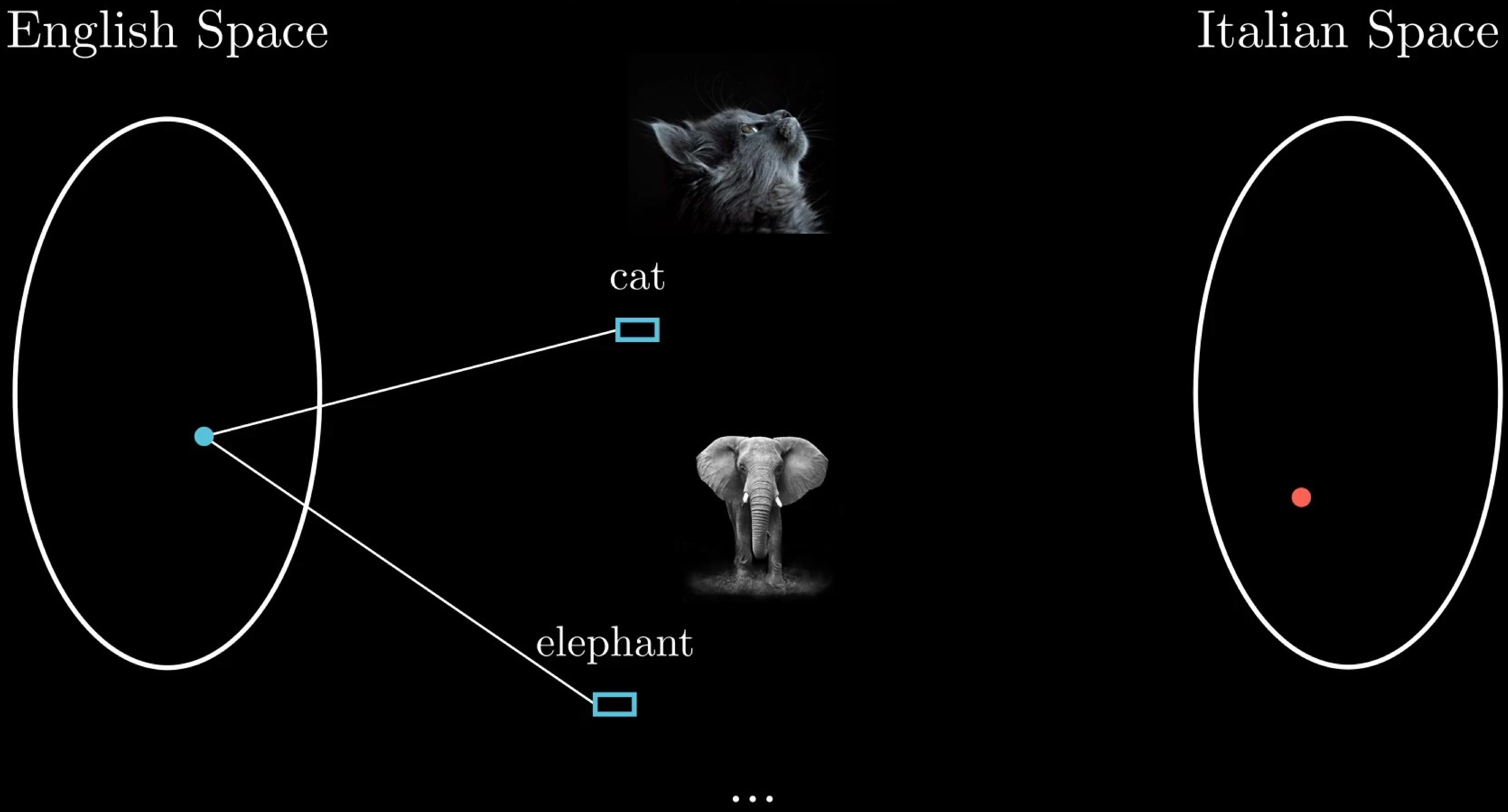
Italian Space

...

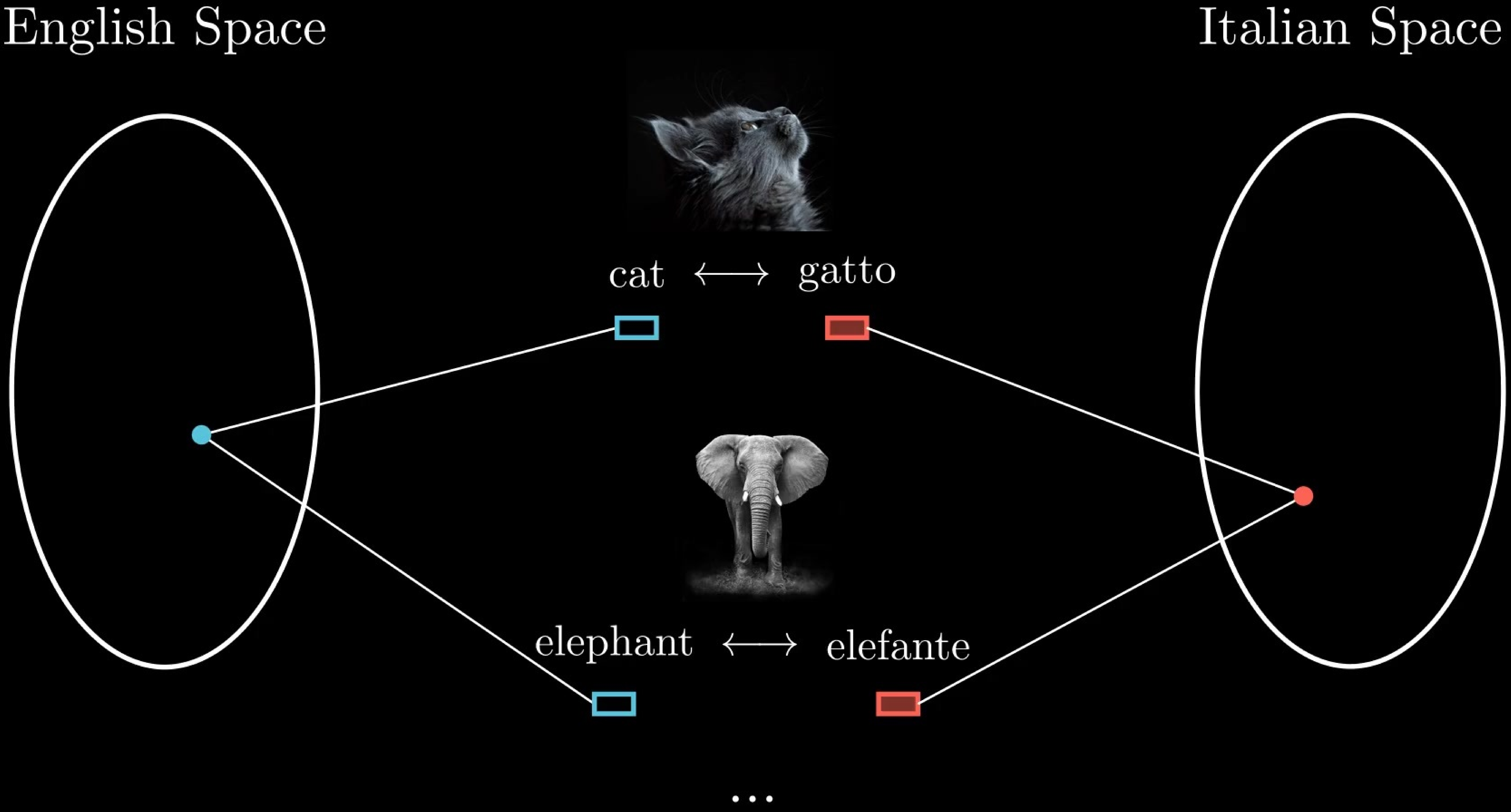
Parallel Anchors



Parallel Anchors



Parallel Anchors



...bridging data modalities

ASIF: Coupled Data Turns Unimodal Models to Multimodal Without Training

Antonio Norelli

Marco Fumero *Valentino Maiorca* *Luca Moschella*

Francesco Locatello *Emanuele Rodolà*

https://arxiv.org/abs/2210.01738

A green bird from New Zealand

Conclusions

Zero-shot communication between different latent spaces

Applications in:

- Model stitching
- Measuring the performance of neural models
- Generalizing unimodal to multimodal models

Future directions:

- Different similarity/anchors choice $\rightarrow$ different invariances/properties
- Relative space similarity as supervision signal

Zero-shot communication between different latent spaces

Applications in:

- Model stitching
- Measuring the performance of neural models
- Generalizing unimodal to multimodal models

Future directions:

- Different similarity/anchors choice $\rightarrow$ different invariances/properties
- Relative space similarity as supervision signal

Zero-shot communication between different latent spaces

Applications in:

- Model stitching
- Measuring the performance of neural models
- Generalizing unimodal to multimodal models

Future directions:

- Different similarity/anchors choice $\rightarrow$ different invariances/properties
- Relative space similarity as supervision signal

Zero-shot communication between different latent spaces

Applications in:

- Model stitching
- Measuring the performance of neural models
- Generalizing unimodal to multimodal models

Future directions:

- Different similarity/anchors choice $\rightarrow$ different invariances/properties
- Relative space similarity as supervision signal

Zero-shot communication between different latent spaces

Applications in:

- Model stitching
- Measuring the performance of neural models
- Generalizing unimodal to multimodal models

Future directions:

- Different similarity/anchors choice $\rightarrow$ different invariances/properties
- Relative space similarity as supervision signal

Zero-shot communication between different latent spaces

Applications in:

- Model stitching
- Measuring the performance of neural models
- Generalizing unimodal to multimodal models

Future directions:

- Different similarity/anchors choice $\rightarrow$ different invariances/properties
- Relative space similarity as supervision signal

s

gnal

Thank you!

...slides by ***Luca Moschella & Valentino Maiorca***