



DA PALAVRA À IMAGEM

WORKSHOP – CASA FIRJAN – FESTIVAL FUTUROS POSSÍVEIS 26/11 - 10h às 13h

MATTEO MORICONI
VFXRIO

LUIZ VELHO
VISGRAPH

BERNARDO ALEVATO
PUC-RIO

Casa Firjan apresenta:

Festival Futuros Possíveis 2022

IMAGINAÇÃO PARA
CRIAR FUTUROS.

25 e 26/11

On-line e com oficinas presenciais na Casa Firjan



The Art in AI

Luiz Velho
IMPA

Midjourney Teaser





Let's look at the
winning painting

Stable Diffusion Fuss



ARTIFICIAL INTELLIGENCE

This artist is dominating AI-generated art. And he's not happy about it.

Greg Rutkowski is a more popular prompt than Picasso.

Source of Inspiration

- The Secret Pass - Eagle Nest



Personal work featured in Rutkowski's ArtStation portfolio.



MS TECH VIA STABLE DIFFUSION



Rutkowski's name has been used as a prompt around 93,000 times!

The Big Debate

Hot from the Press

WIRED

The Joy and Dread of AI Image Generators Without Limits

AI art looks way too European

Vox

DALL-E and other models keep making art that ignores traditions from the rest

WILL KNIGHT BUSINESS SEP 21, 2022 7:00 AM

WIRED

This Uncensored AI Art Tool Can Generate Fantasies—and Nightmares

The New York Times

<https://www.nytimes.com/2022/10/21/technology/generative-ai.html>

THE SHIFT

A Coming-Out Party for Generative A.I., Silicon Valley's New Craze

A celebration for Stability AI, the start-up behind the controversial Stable Diffusion image generator, re

TECHPERSONAL TECHNOLOGY: JOANNA STERN

THE WALL STREET JOURNAL

Ask an AI Art Generator for Any Image. The Results Are Amazing—and Terrifying.

The New York Times

<https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs->

THE SHIFT

A.I.-Generated Art Is Already Transforming Creative Work

Tech Talk

MIT Technology Review

Publicado por TEC

EDIÇÃO - OUT 2022

Edição Digital | Generative AI

Homem, máquina e criatividade.



MIT
Technology
Review

Featured Topics Newsletters Events Podcasts

ARTIFICIAL INTELLIGENCE

Get ready for the next generation of AI

MIT Technology Review

ARTIFICIAL INTELLIGENCE

The dark secret behind those cute AI-generated animal images

Google Brain has revealed its own image-making AI, called Imagen. But don't expect to see anything that isn't wholesome.

From: MIT Technology Review Brasil

Imagens geradas por IA são vetadas por empresas

O QUE É?

Empresas vetam imagens geradas por inteligência artificial.

Midjourney e DALL-E são alguns dos programas usados.

The Era of Machine Learning

A New Way for Science & Technology

Computer Science > Computer Vision and Pattern Recognition
Submitted on 14 Jun 2022 15:11, last revised 4 May 2022 (this version, v2)
Instant Neural Graphics Primitives with a Multiresolution Hash Encoding
Thomas Müller, Alex Evans, Christoph Schied, Alexander Keller

Neural graphics primitives, parameterized by fully connected neural networks, can be easily to train and evaluate. We reduce this cost with a versatile new input encoding that permits the use of a smaller network without sacrificing quality, thus significantly reducing the number of floating point and memory access operations. A small neural network is augmented by a multiresolution hash table of trainable feature vectors whose values are optimized through stochastic gradient descent. The multiresolution structure allows the network to disambiguate hash collisions, making for a simple architecture that is trivial to parallelize on modern CPUs. We leverage this parallelism by implementing the whole system using fully-fused CUDA kernels with a focus on minimizing wasted bandwidth and compute operations. We achieve a combined speedup of several orders of magnitude, enabling training of high-quality neural graphics primitives in a matter of seconds, and rendering in tens of milliseconds at a resolution of 1920x1080.

Download:
• PDF
• Other formats

Current browse context:
arXiv
< prev | next >
new recent (2022)
Change to browse by:
arXiv
cs.LG
cs.CV

References & Citations
• NASA ADS
• Google Scholar
• Semantic Scholar
2022 - CS Bibliography
Listing | Bibsax
Thomas Müller
Alexander Keller

Export BibTeX citation
Bookmark

Submission history
From: Thomas Müller (new email)
v1 [Tue, 14 Jun 2022 07:22:47 UTC (17.327 KB)]
v2 [Wed, 4 May 2022 07:35:17 UTC (17.278 KB)]

arXiv publication

Instant Neural Graphics Primitives with a Multiresolution Hash Encoding
THOMAS MÜLLER, NVIDIA, Switzerland
ALEX EVANS, NVIDIA, United Kingdom
CHRISTOPH SCHIED, NVIDIA, USA
ALEXANDER KELLER, NVIDIA, Germany
<https://nvlabs.github.io/instant-ngp>

Trained for 1 second 15 seconds 1 minute 15 seconds @ 1024x1024 reference

CGI image
SDF
NeRF
Neural volume

Fig. 1. We demonstrate instant training of neural graphics primitives on a single GPU for multiple tasks. In **gigapixel image** we represent a gigapixel image by a neural network. **SDF** learns a signed distance function in 3D space whose zero level-set represents a 2D surface. **Neural radiance encoding** [Müller et al. 2022] employs a neural network that is trained in real-time to cache costly lighting calculations. Lastly, **NeRF** [Mildenhall et al. 2020] uses 2D images and their camera poses to reconstruct a volumetric radiance and density field that is visualized using ray marching. In all tasks, our encoding and its efficient implementation provide clear benefits: rapid training, high quality, and simplicity. Our encoding is task-agnostic: we use the same implementation and hyperparameters across all tasks and only vary the hash table size which trades off quality and performance. Top row: gigapixel photographs (Three Dubois (CC BY-NC-ND 2.0), Lego building 3D model (Hazard Duden (CC BY-NC 2.0))

Neural graphics primitives, parameterized by fully connected neural networks, can be easily to train and evaluate. We reduce this cost with a versatile new input encoding that permits the use of a smaller network without sacrificing quality, thus significantly reducing the number of floating point and memory access operations. A small neural network is augmented by a multiresolution hash table of trainable feature vectors whose values are optimized through stochastic gradient descent. The multiresolution structure allows the network to disambiguate hash collisions, making for a simple architecture that is trivial to parallelize on modern CPUs. We leverage this parallelism by implementing the whole system using fully-fused CUDA kernels with a focus on minimizing wasted bandwidth and compute operations. We achieve a combined speedup of several orders of magnitude, enabling training of high-quality neural graphics primitives in a matter of seconds, and rendering in tens of milliseconds at a resolution of 1920x1080.

Additional Key Words and Phrases: Image Synthesis, Neural Networks, Encoding, Hashing, GPU, Parallel Computation, Feature Approximation

ACM Reference Format:
Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. ACM Trans. Graph., 41, 4, Article 102 (July 2022), 15 pages. <https://doi.org/10.1145/3528223.3528227>

ACM Trans. Graph., Vol. 41, No. 4, Article 102. Publication date: July 2022.

journal paper

Instant Neural Graphics Primitives with a Multiresolution Hash Encoding

Thomas Müller Alex Evans Christoph Schied Alexander Keller

NVIDIA

ACM Transactions on Graphics (SIGGRAPH 2022)

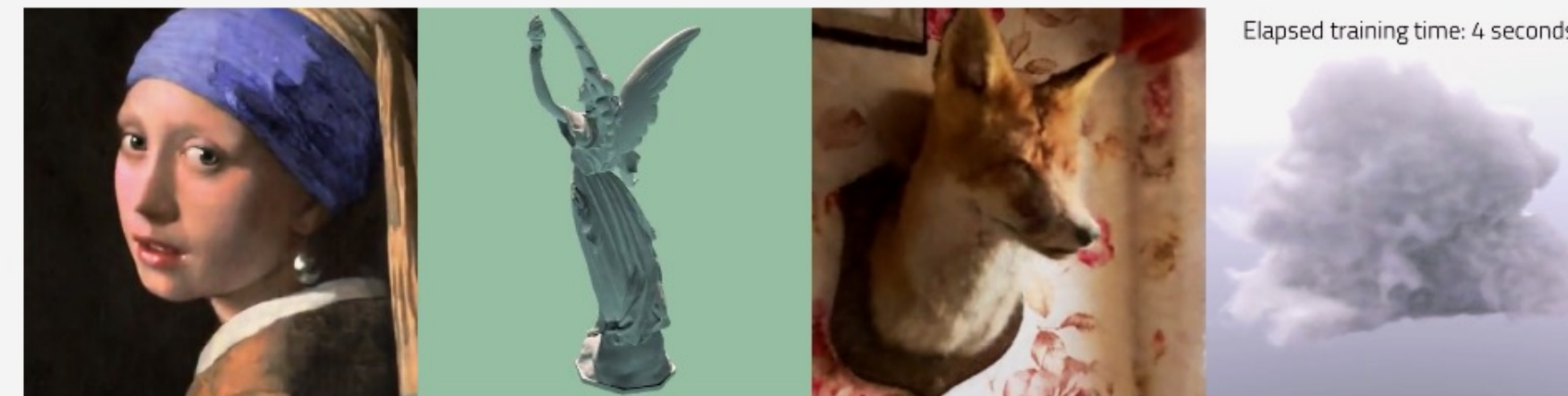
Paper Video Code

Neural gigapixel images

Neural SDF

NeRF

Neural volume



We demonstrate near-instant training of neural graphics primitives on a single GPU for multiple tasks. In **gigapixel image** we represent an image by a neural network. **SDF** learns a signed distance function in 3D space whose zero level-set represents a 2D surface. **Neural radiance encoding** [Müller et al. 2020] uses 2D images and their camera poses to reconstruct a volumetric radiance-and-density field that is visualized using ray marching. In all tasks, our encoding and its efficient implementation provide clear benefits: instant training, high quality, and simplicity. Our encoding is task-agnostic: we use the same implementation and hyperparameters across all tasks and only vary the hash table size which trades off quality and performance. Girl With a Pearl Earring renovation ©Koorosh Orooj (CC BY-SA 4.0)

News

- [July 7th 2022] Paper won the SIGGRAPH Best Paper Award.
- [May 3rd 2022] Paper accepted to ACM Transactions on Graphics (SIGGRAPH 2022).
- [Jan 19th 2022] Paper released on arXiv.
- [Jan 14th 2022] Code released on GitHub.

project page

Search or jump to... Pull requests Issues Marketplace Explore
nvlabs / instant-ngp Public Watch 182

Code Issues 184 Pull requests 6 Discussions 0 Actions Projects Security

master 2 branches 0 tags Go to file Add file Code

Tom94 Update TCNN (to fix compilation on Linux + CUDA 11.3) ✓ ed8ee2 yesterday 472 commits

| | | |
|-------------------------------|---|---------------|
| .devcontainer | Fixed for new docker + ignore envv | 8 months ago |
| github/workflows | Fix Ubuntu CI | 11 days ago |
| cmake | Bump default OptiX search location to 7.5 | 3 months ago |
| configs | NeRF base: use identity encoding for extra dimensions by... | 6 months ago |
| data | Smaller fox dataset (should fit into 8 GB of VRAM now) | 10 months ago |
| dependencies | Update TCNN (to fix compilation on Linux + CUDA 11.3) | yesterday |
| docs | Add 250x250px png of fox as representative image for re... | 2 months ago |
| include/neural-graphics-pr... | Merge pull request #1022 from JamesPerlan/master | 6 days ago |
| notebooks | add notebook with guide for execution in colab | 3 months ago |
| scripts | Remove implicit tonemapping from SSM computation | 11 days ago |
| src | Merge pull request #1022 from JamesPerlan/master | 6 days ago |
| editorconfig | Initial commit | 10 months ago |
| gitattributes | Ignore notebooks for language statistics (logs inflate the L... | 2 months ago |
| gitignore | Update TCNN; preparation for depth supervision; f-theta L... | 8 months ago |
| gitmodules | DLSS through Vulkan (+ depth/motion vector code) | 6 months ago |
| CMakeLists.txt | On linux, link to libl even without GUI support | 2 months ago |
| LICENSE.txt | Initial commit | 10 months ago |
| README.md | List compute capabilities of Ada and Hopper | 12 days ago |
| requirements.txt | Relax requirements.txt | 8 months ago |

README.md

GitHub repository

Technical Blog NVIDIA DEVELOPER
Subscribe

Technical Walkthrough May 12, 2022 English

Getting Started with NVIDIA Instant NeRFs

By Jonathan Stephens

Discuss (1) +7 Like

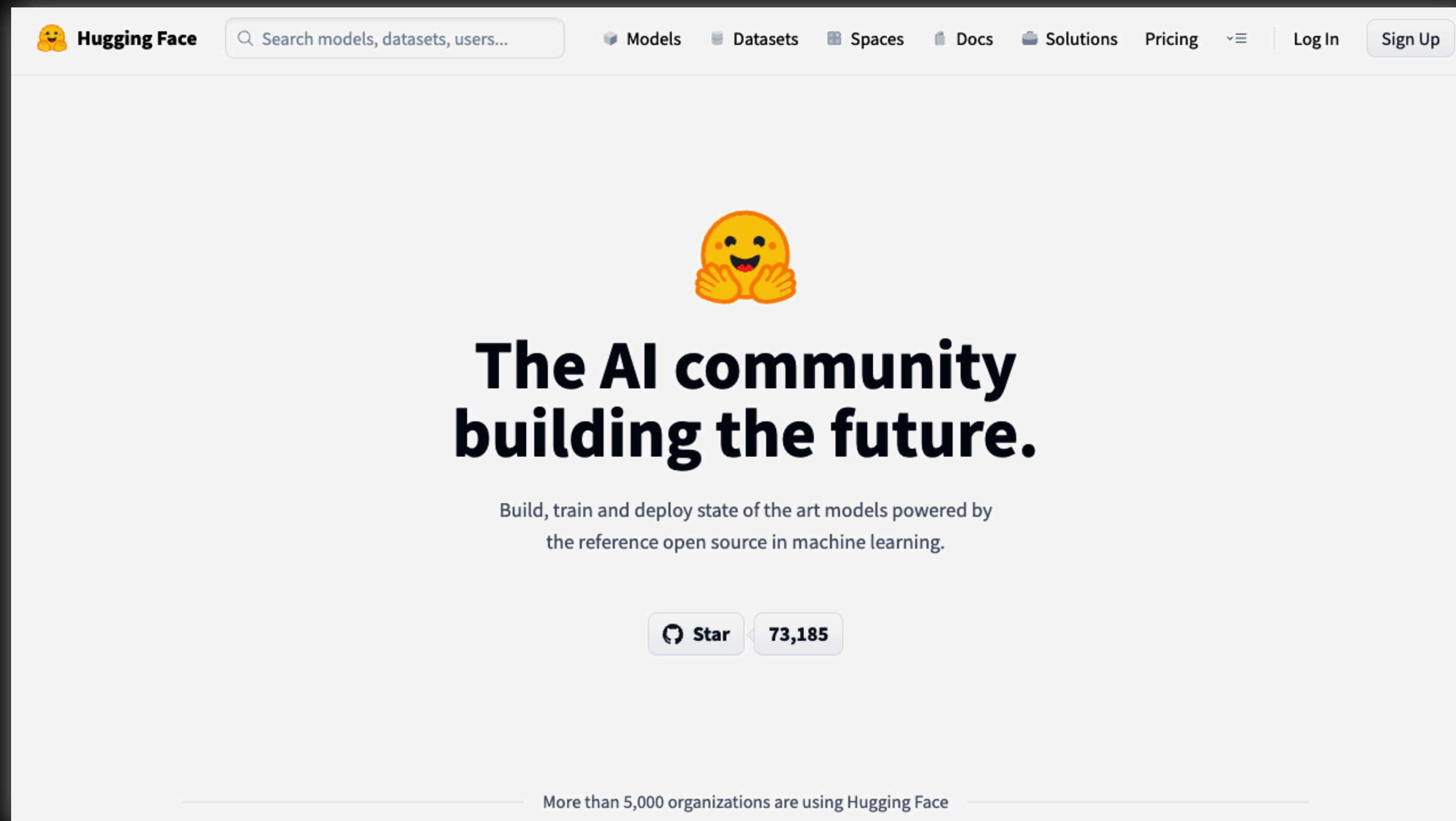
Tags: featured, NVIDIA Research, Technical Walkthrough

Input Images

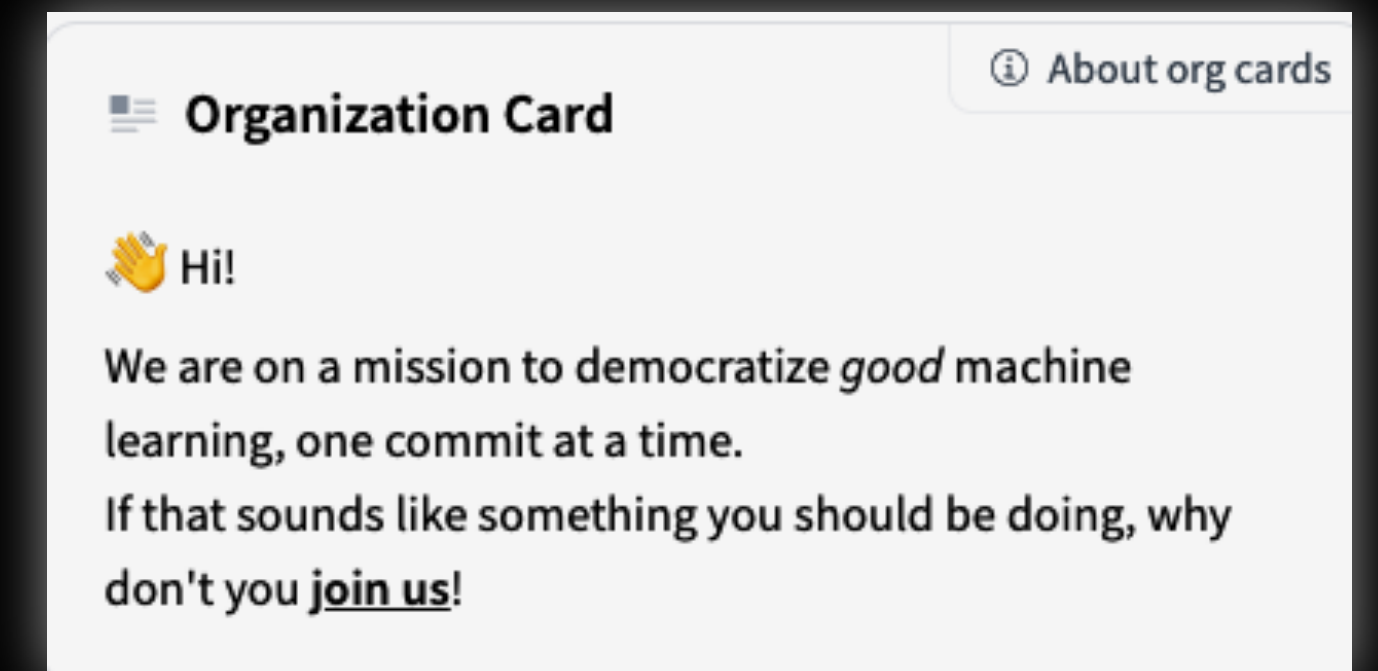
blog posts

Open Source Community

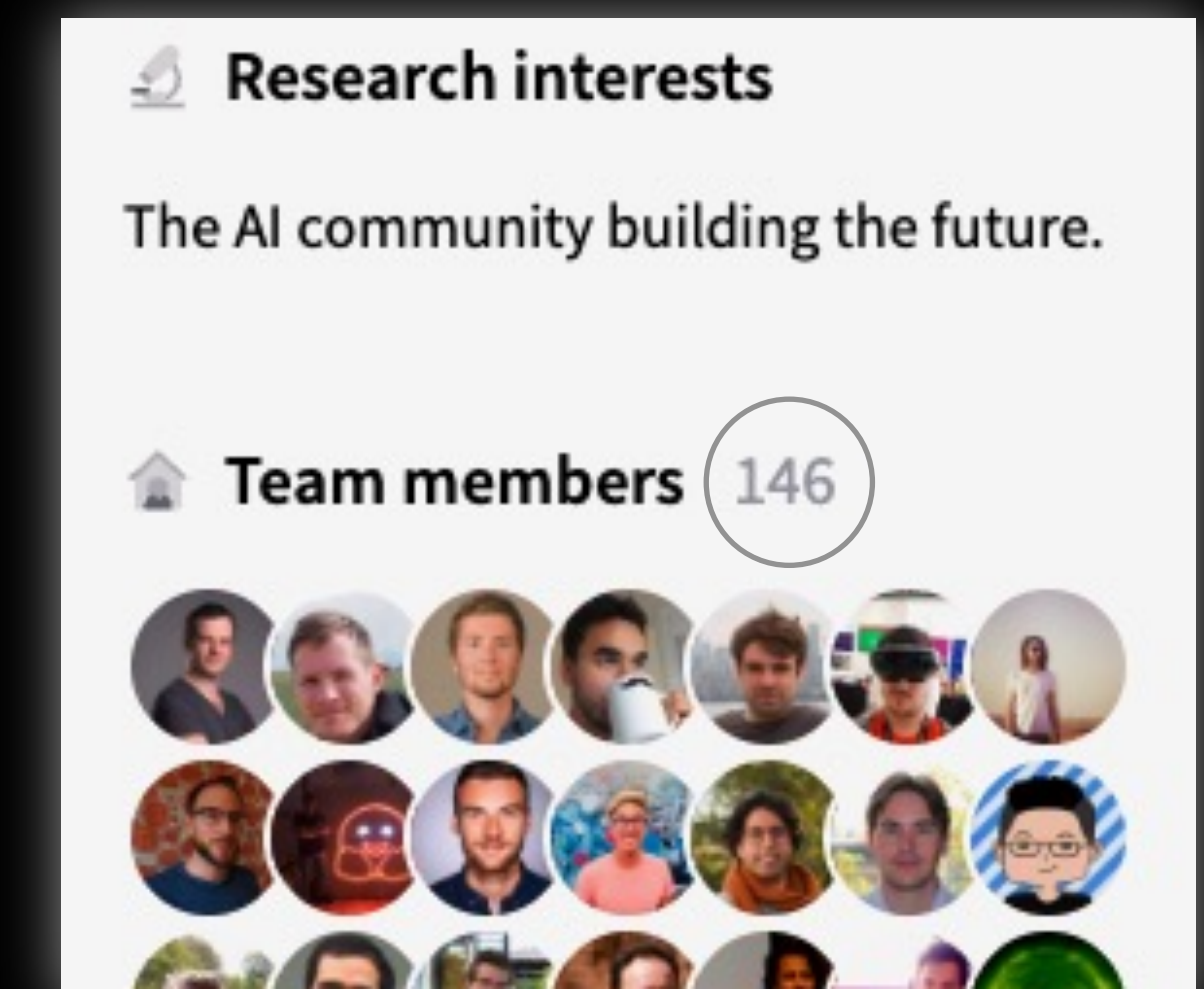
- One Good Example



The screenshot shows the Hugging Face homepage. At the top left is the Hugging Face logo and a search bar with the text "Search models, datasets, users...". To the right of the search bar are navigation links for "Models", "Datasets", "Spaces", "Docs", "Solutions", "Pricing", "Log In", and "Sign Up". The main content area features a large yellow emoji with its hands clasped in front of its face. Below the emoji is the text "The AI community building the future." in a large, bold font. Underneath this is a smaller line of text: "Build, train and deploy state of the art models powered by the reference open source in machine learning." At the bottom of the main content area, there is a "Star" button with a star icon and the number "73,185". At the very bottom of the page, a horizontal line is followed by the text "More than 5,000 organizations are using Hugging Face".



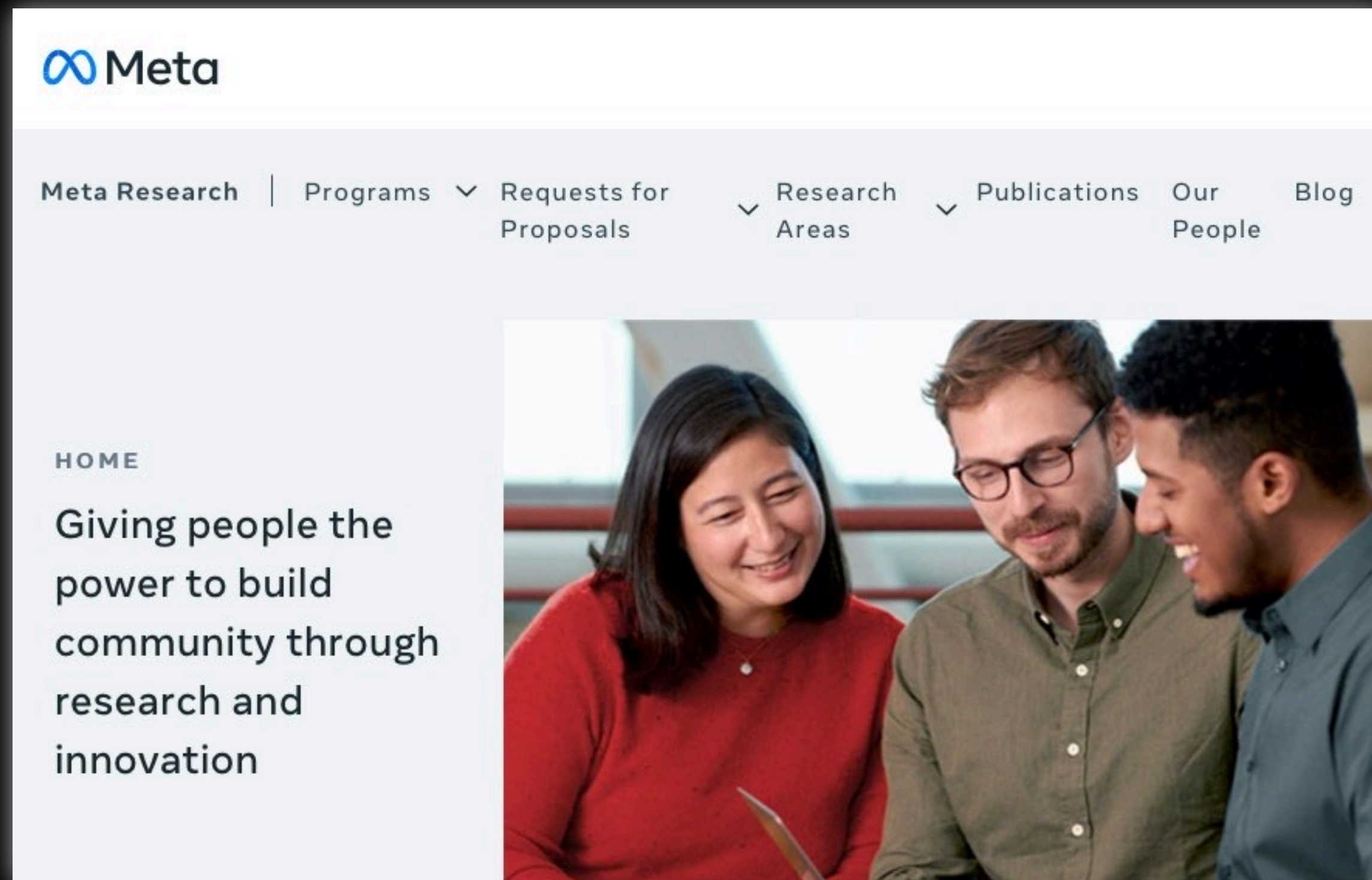
This screenshot shows an "Organization Card" for Hugging Face. It features a header with a hamburger menu icon, the text "Organization Card", and a link "About org cards". Below the header is a yellow emoji with its hands clasped, followed by the text "Hi!". The main body of the card contains the text: "We are on a mission to democratize *good* machine learning, one commit at a time. If that sounds like something you should be doing, why don't you [join us!](#)".



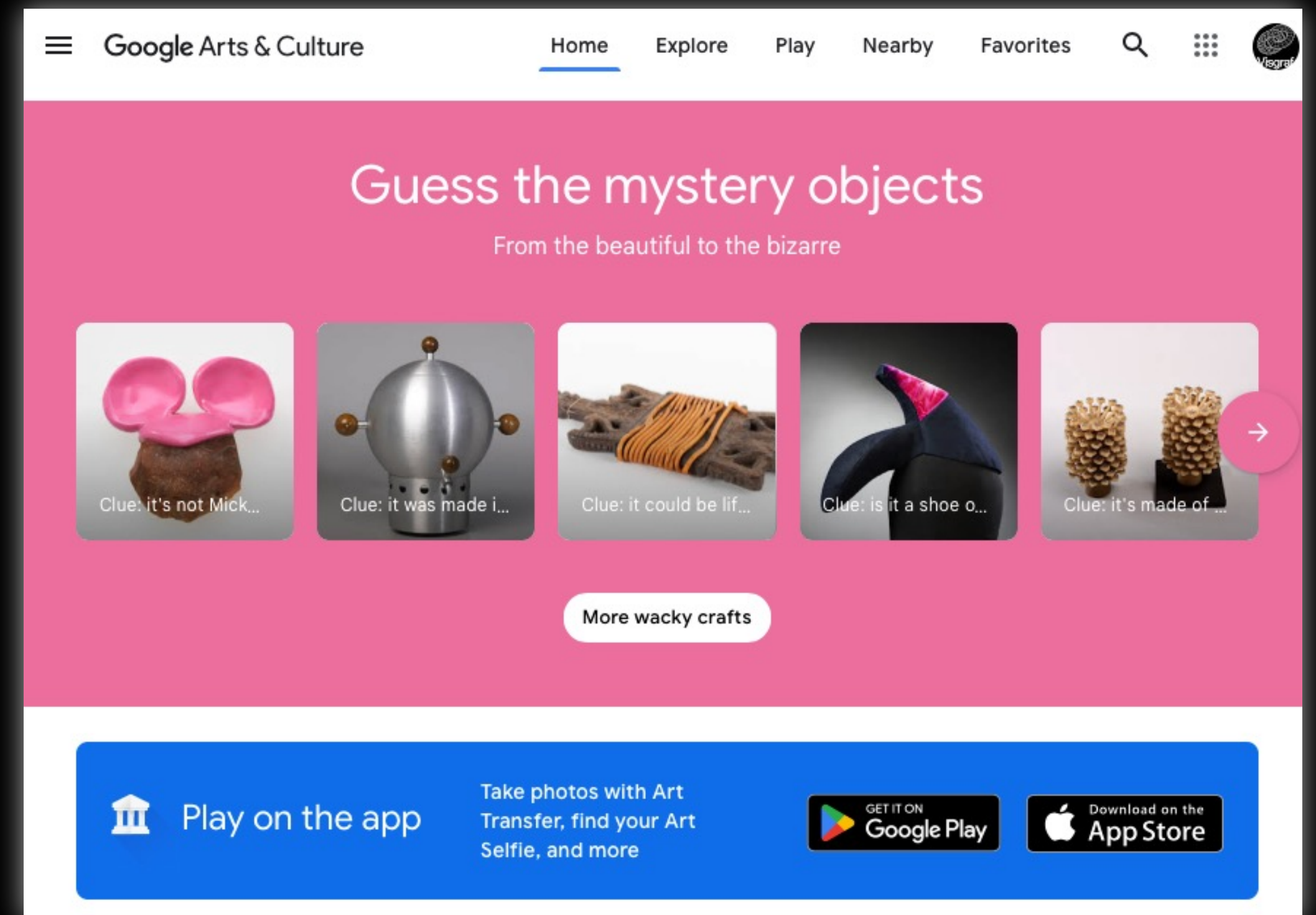
This screenshot shows two sections. The top section is titled "Research interests" and contains the text "The AI community building the future." The bottom section is titled "Team members" and shows a count of "146" members. Below the count is a grid of 18 circular profile pictures of team members.

Outsourcing to Society

- Google, Meta and many more...



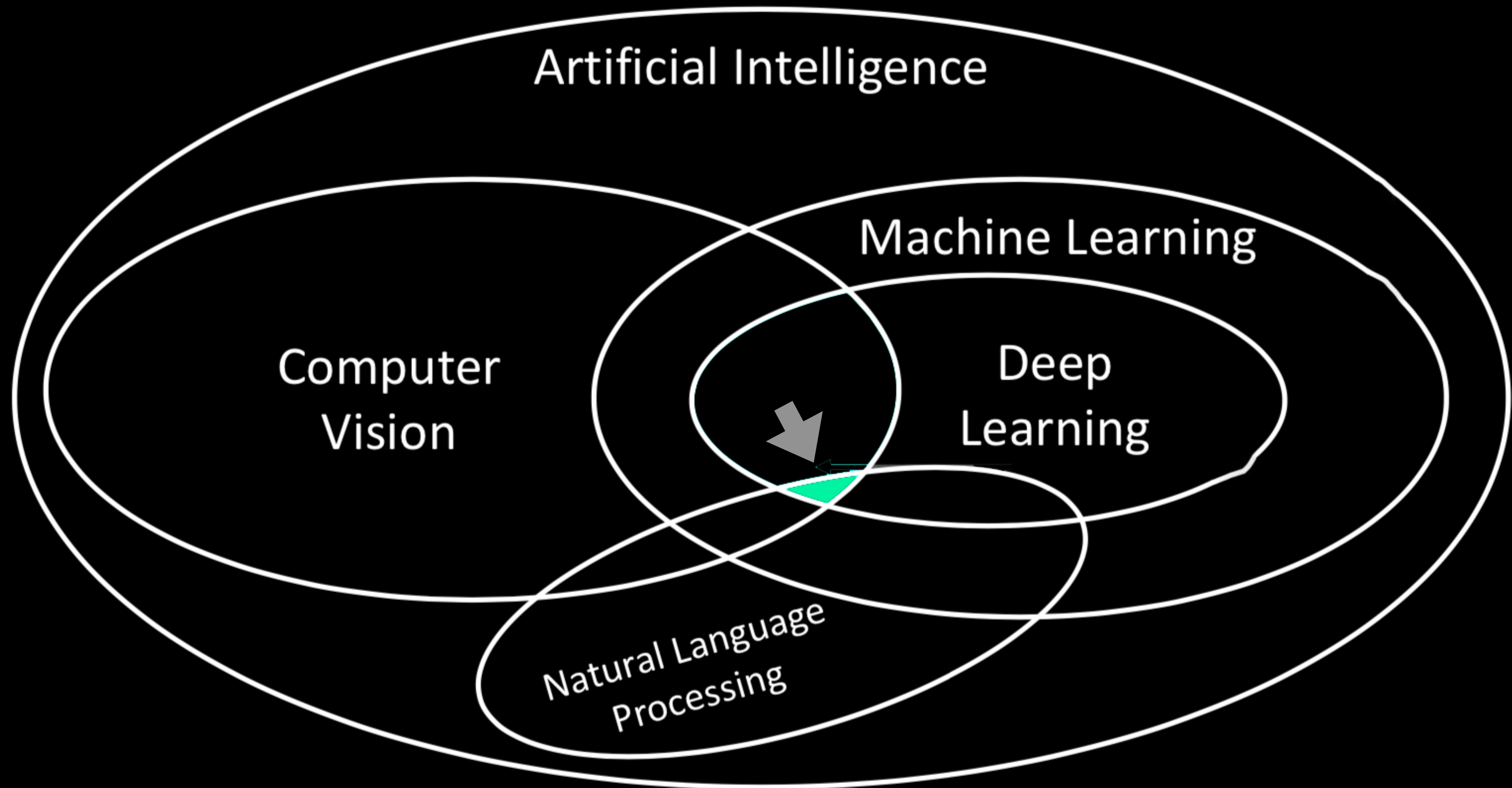
The screenshot shows the Meta website homepage. At the top left is the Meta logo. Below it is a navigation menu with links for Meta Research, Programs, Requests for Proposals, Research Areas, Publications, Our People, and Blog. The main content area features a large image of three people (two men and one woman) looking at a laptop together. To the left of the image, the text reads: "HOME Giving people the power to build community through research and innovation".



The screenshot shows the Google Arts & Culture website homepage. At the top is the Google Arts & Culture logo and a navigation menu with links for Home, Explore, Play, Nearby, and Favorites. The main content area features a large pink banner with the text "Guess the mystery objects" and "From the beautiful to the bizarre". Below the banner are five mystery objects with clues: a pink object (Clue: it's not Mick...), a silver object (Clue: it was made i...), a wooden object (Clue: it could be lif...), a black object (Clue: is it a shoe o...), and a golden object (Clue: it's made of...). A "More wacky crafts" button is located below the objects. At the bottom, there is a blue banner with the text "Play on the app" and "Take photos with Art Transfer, find your Art Selfie, and more", along with "GET IT ON Google Play" and "Download on the App Store" buttons.

Deep Learning (R)evolution

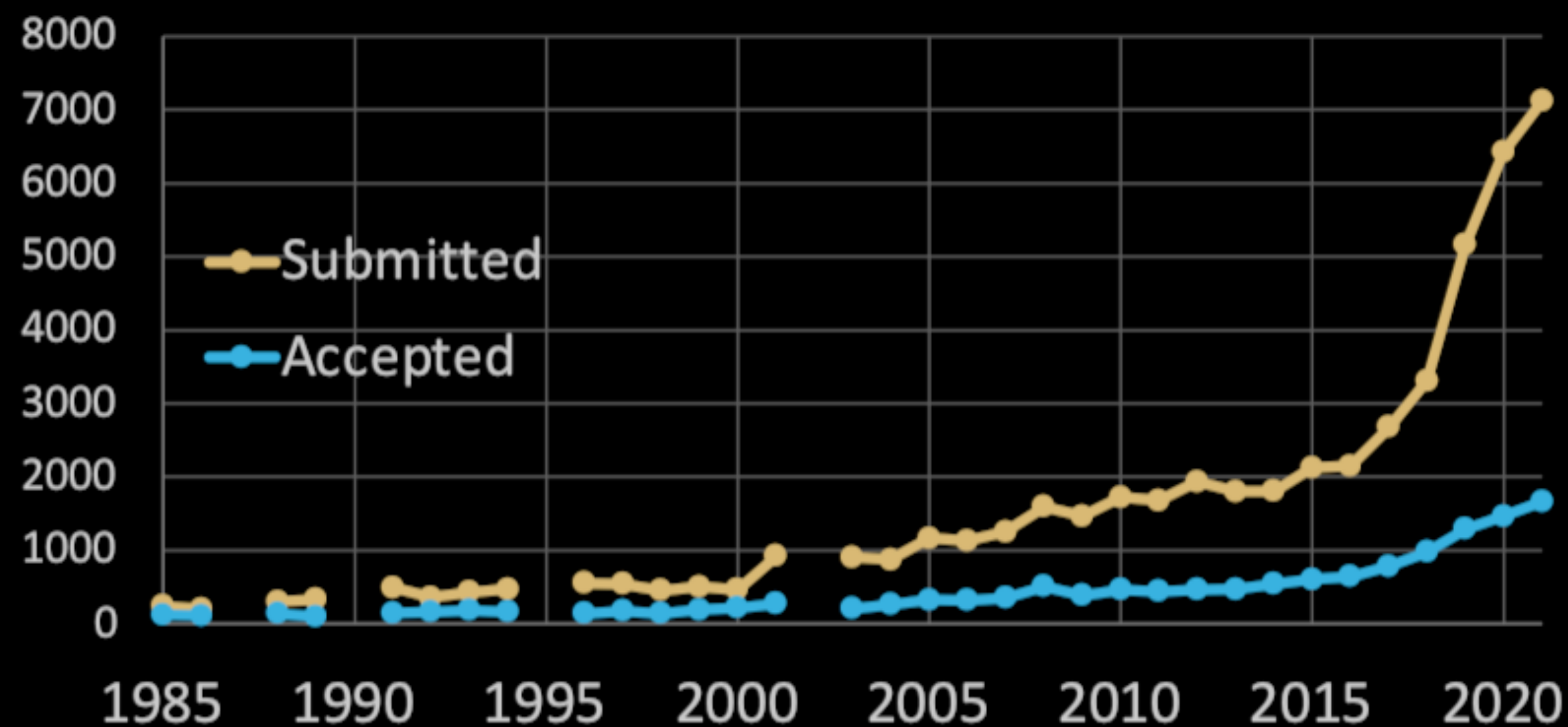
Historical Context



Exponential Growth

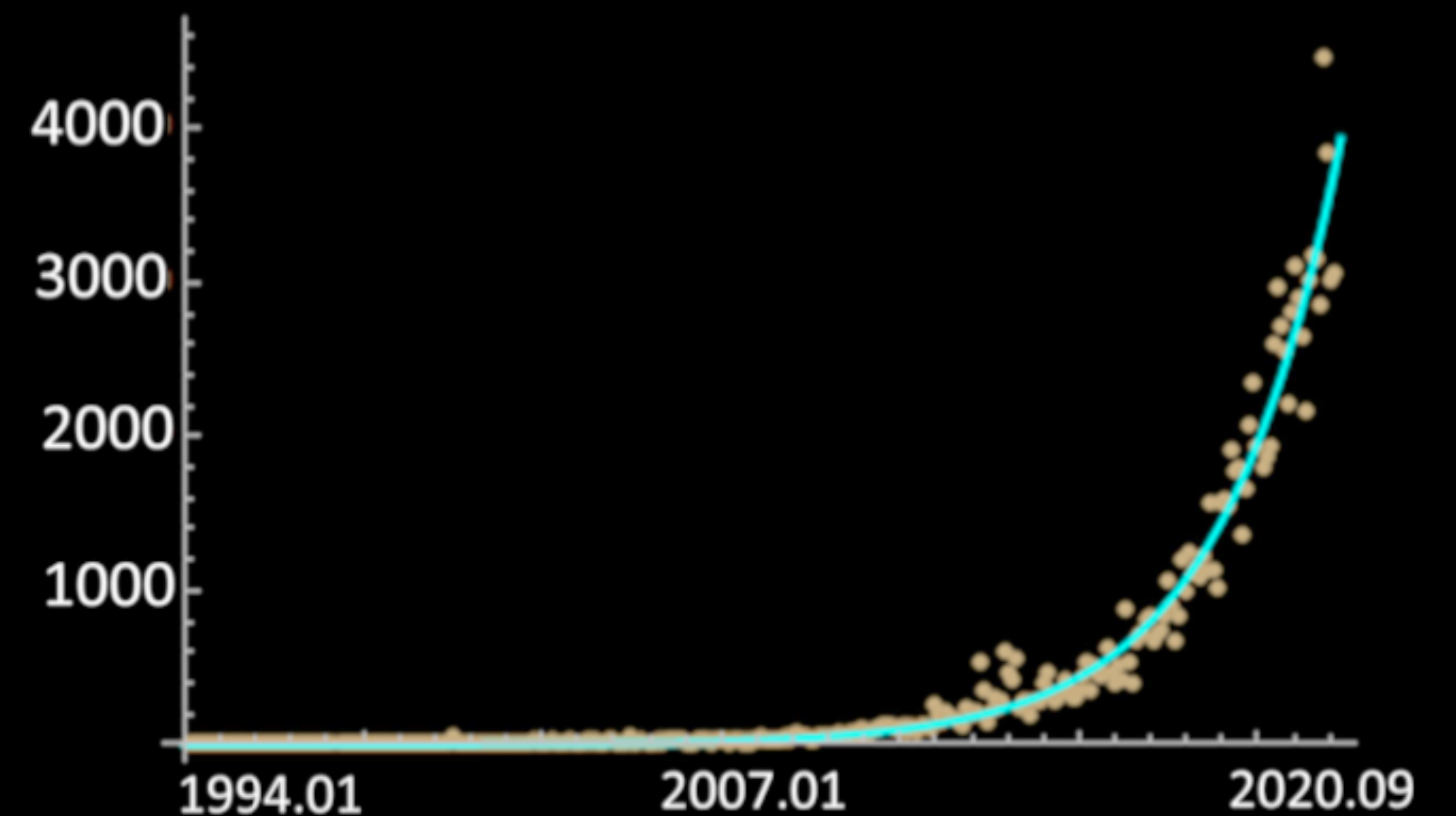
2012 to Present: Deep Learning Explosion

CVPR Papers



Publications at top Computer Vision conference

ML+AI arXiv papers per month



arXiv papers per month ([source](#))

Image Serendipity

Metadata from Images

- Image Analysis

Image Classification

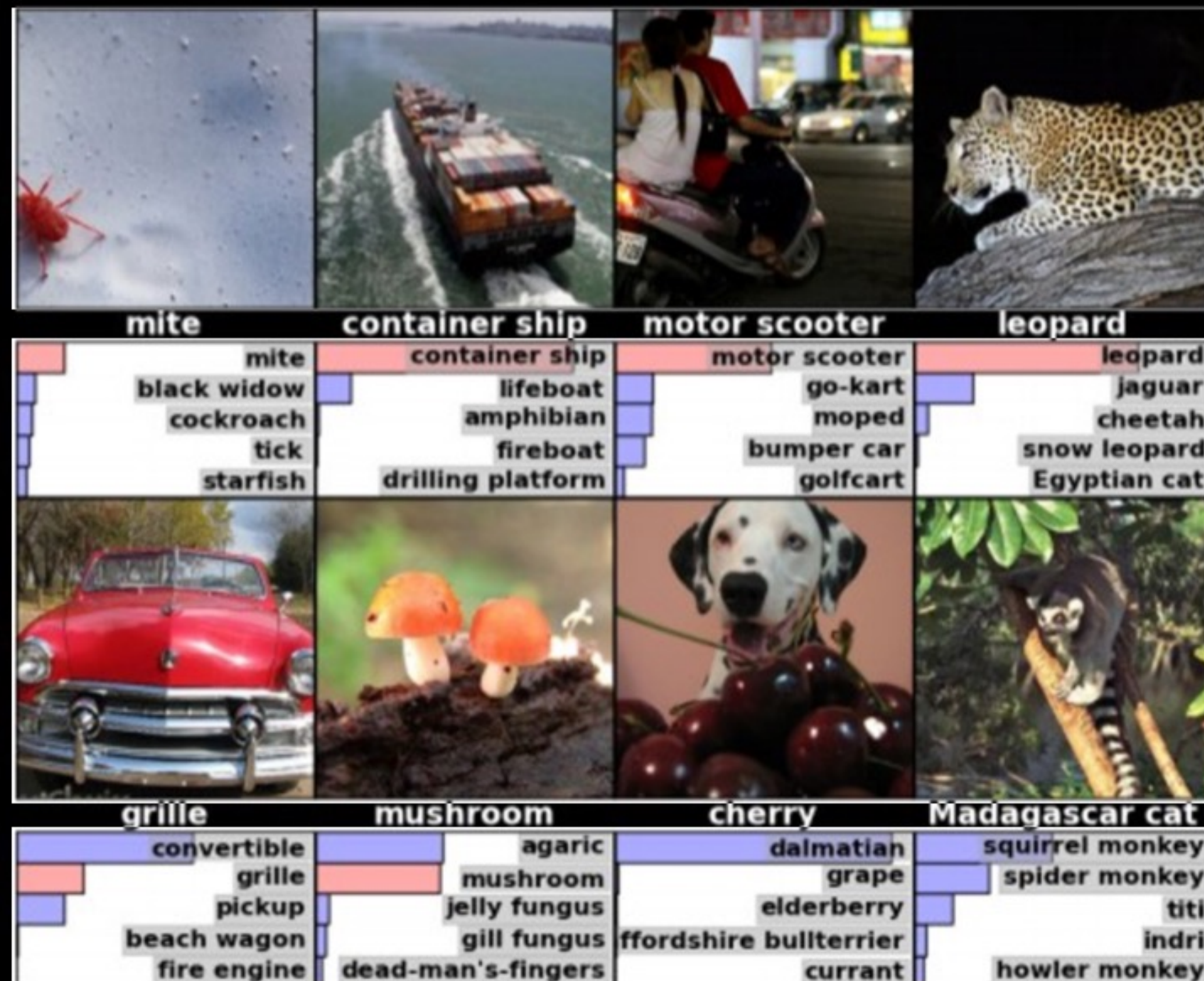


Image Retrieval

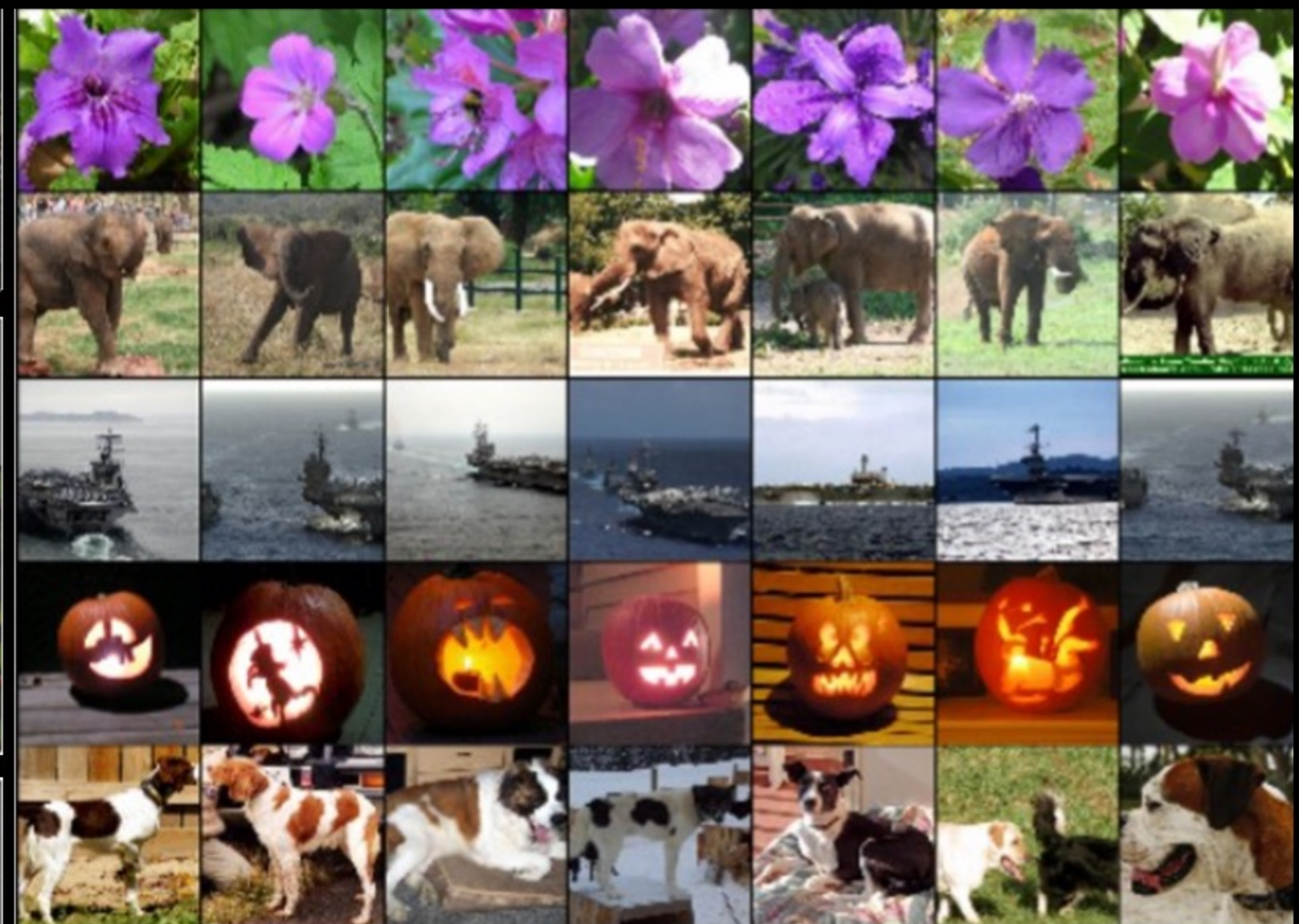


Image from Images

- Style Transfer



Original Image



Style A



Style B

Image-to-Text

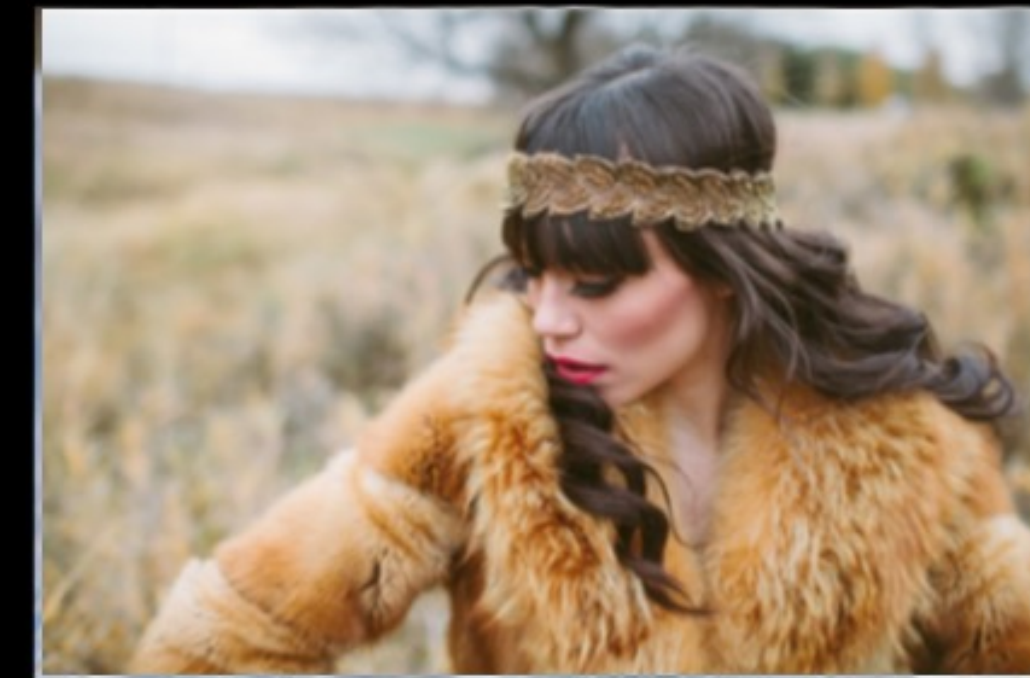
- Image Captioning



A white teddy bear sitting in the grass



A man in a baseball uniform throwing a ball



A woman is holding a cat in her hand

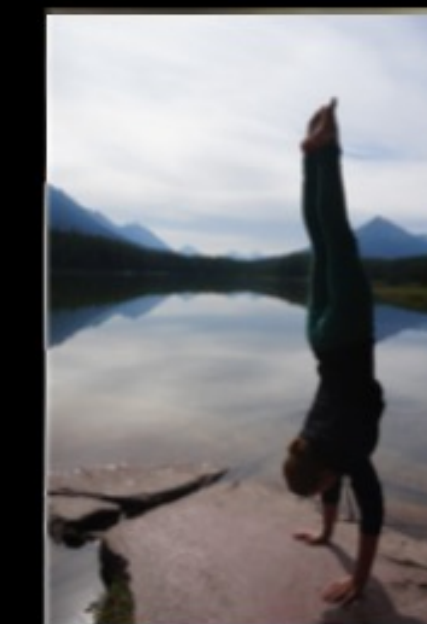
Vinyals et al, 2015
Karpathy and Fei-Fei, 2015



A man riding a wave on top of a surfboard



A cat sitting on a suitcase on the floor



A woman standing on a beach holding a surfboard

All images are CC0 Public domain:
<https://pixabay.com/en/usage-antique-cat-1643010/>
<https://pixabay.com/en/teddy-plush-bear-cute-teddy-bear-1623436/>
<https://pixabay.com/en/surf-wave-summer-sport-litoral-1668716/>
<https://pixabay.com/en/woman-female-model-portrait-adult-983967/>
<https://pixabay.com/en/handstand-lake-meditation-496008/>
<https://pixabay.com/en/base-ball-player-shortstop-infield-1045263/>

Captions generated by Justin Johnson using [NeuralTalk2](#)

Text-to-Image

- Image Synthesis

TEXT PROMPT

an armchair in the shape of an avocado. an armchair imitating an avocado.

AI-GENERATED IMAGES



The Mathematics of Machine Learning

Computational Applied Mathematics

- Computation Model

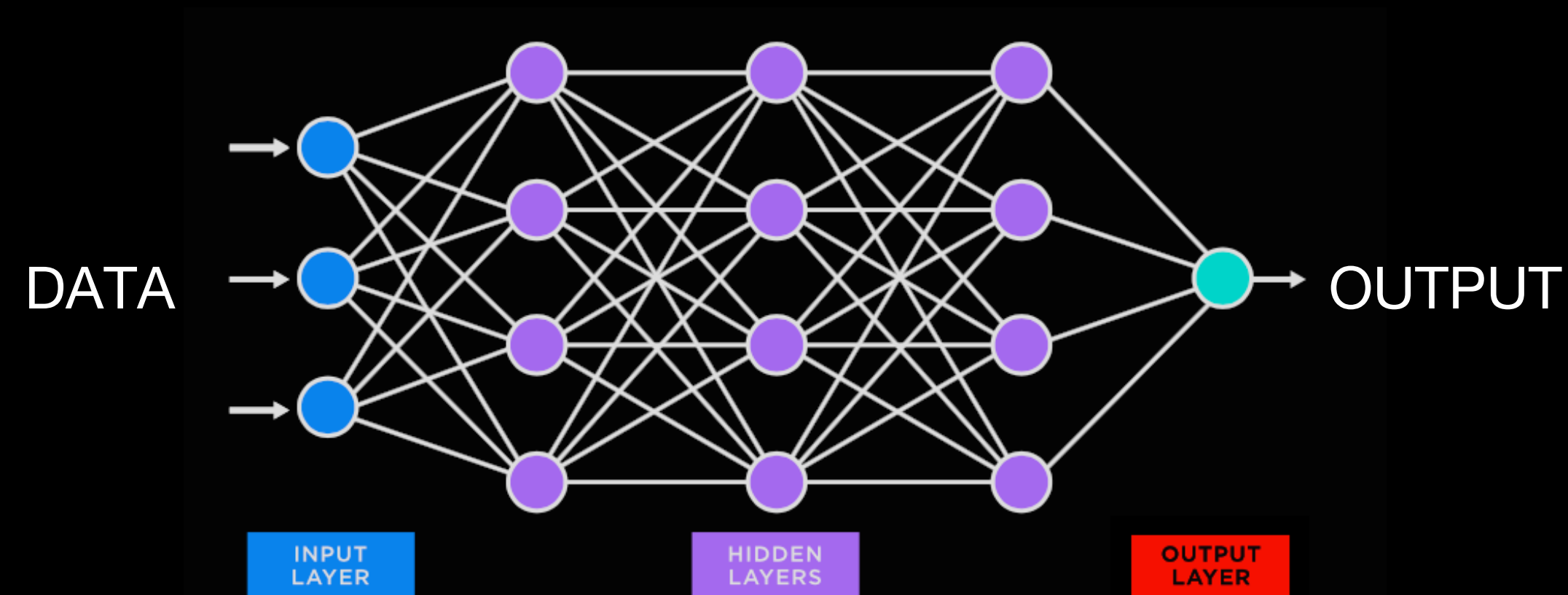


Old, New Way

- Algorithm Code

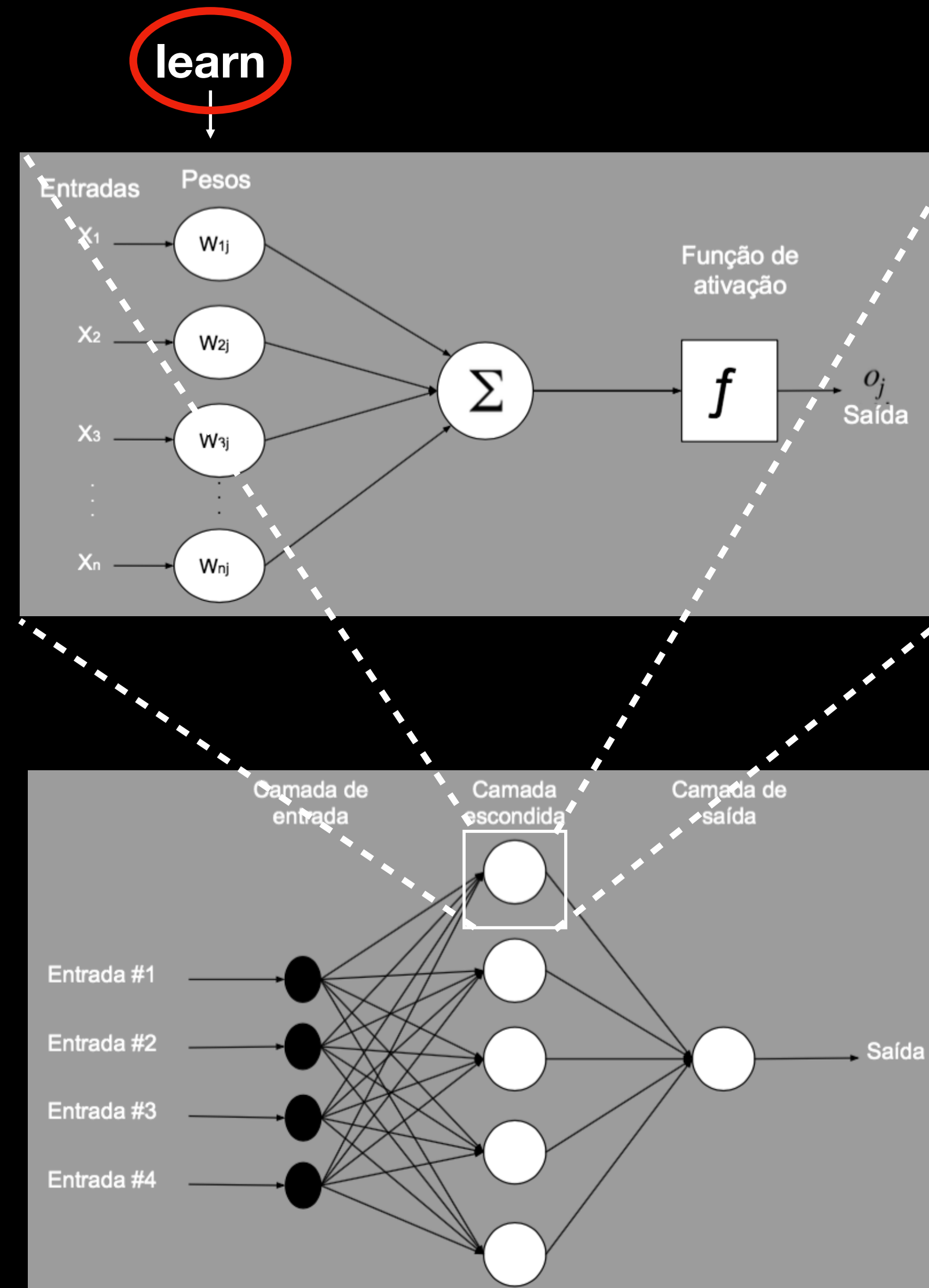
```
Step 1: Input M1,M2,M3,M4
Step 2: GRADE ← (M1+M2+M3+M4)/4
Step 3: if (GRADE <50) then
        Print "FAIL"
        else
        Print "PASS"
        endif
```

- Network Architecture



Neural Networks

- Perceptron (operator)
- Neural Nets (composition)



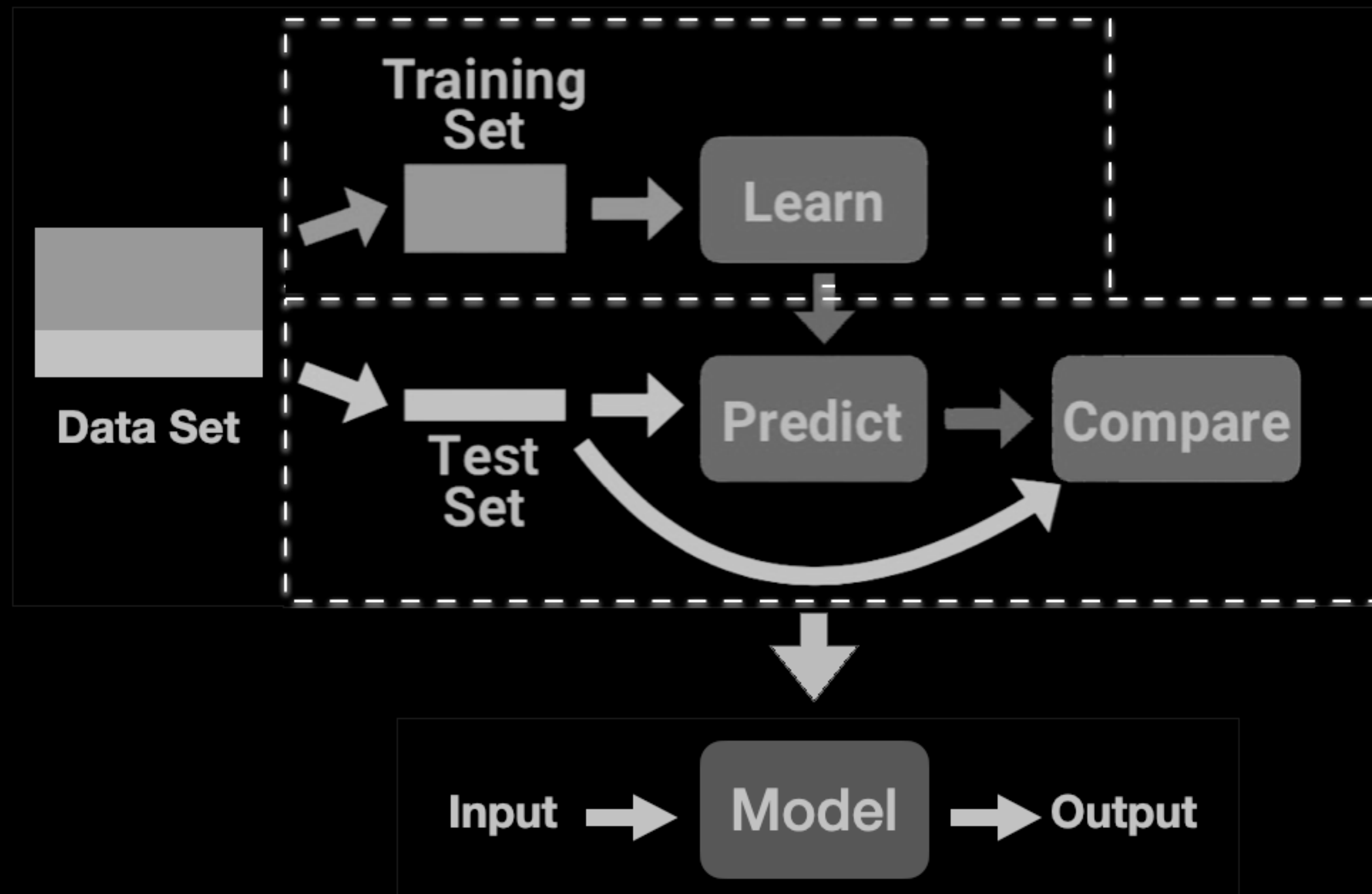
Learning from Data

- Stages

1. Train

2. Validate

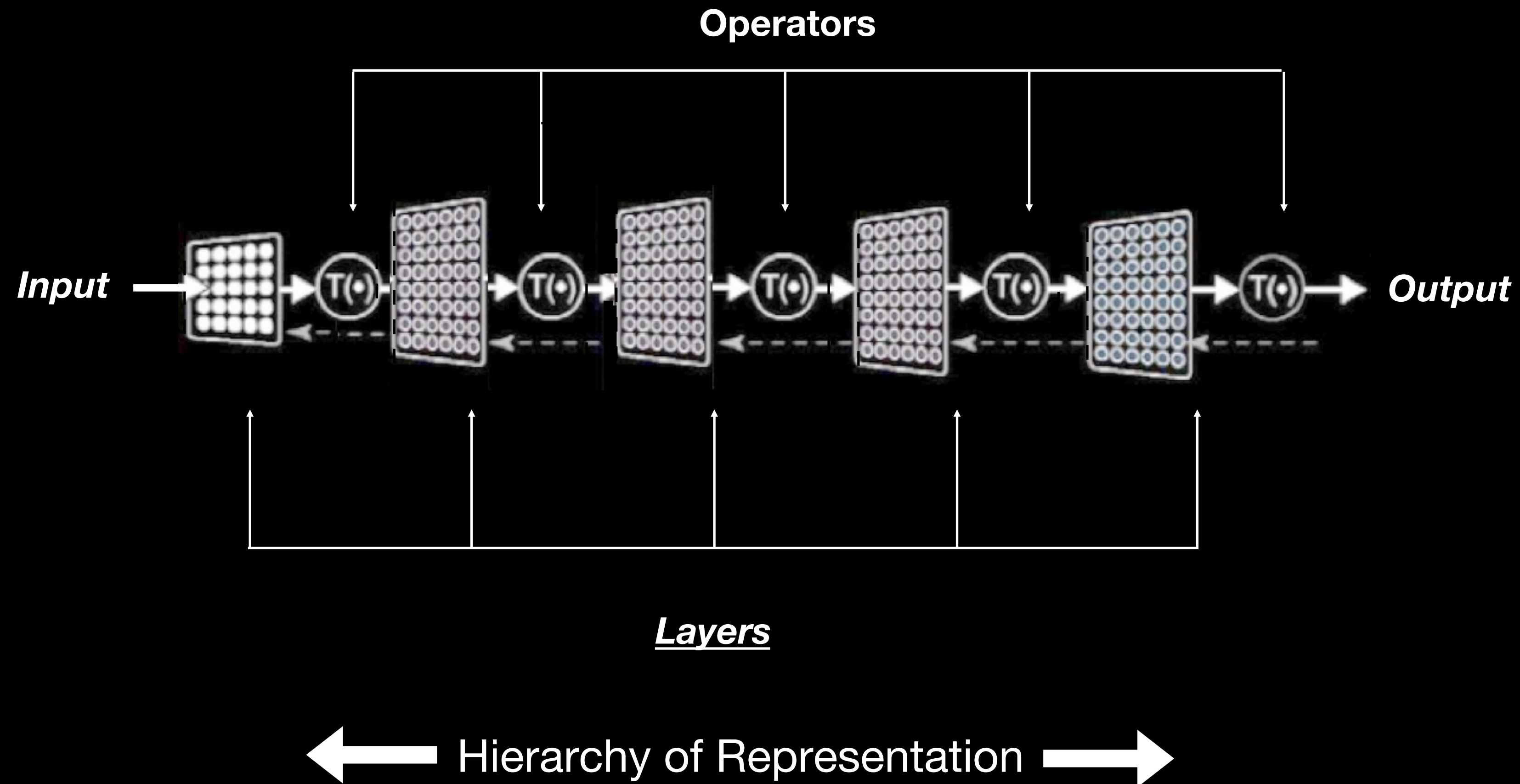
3. Inference



Deep Neural Network Architectures

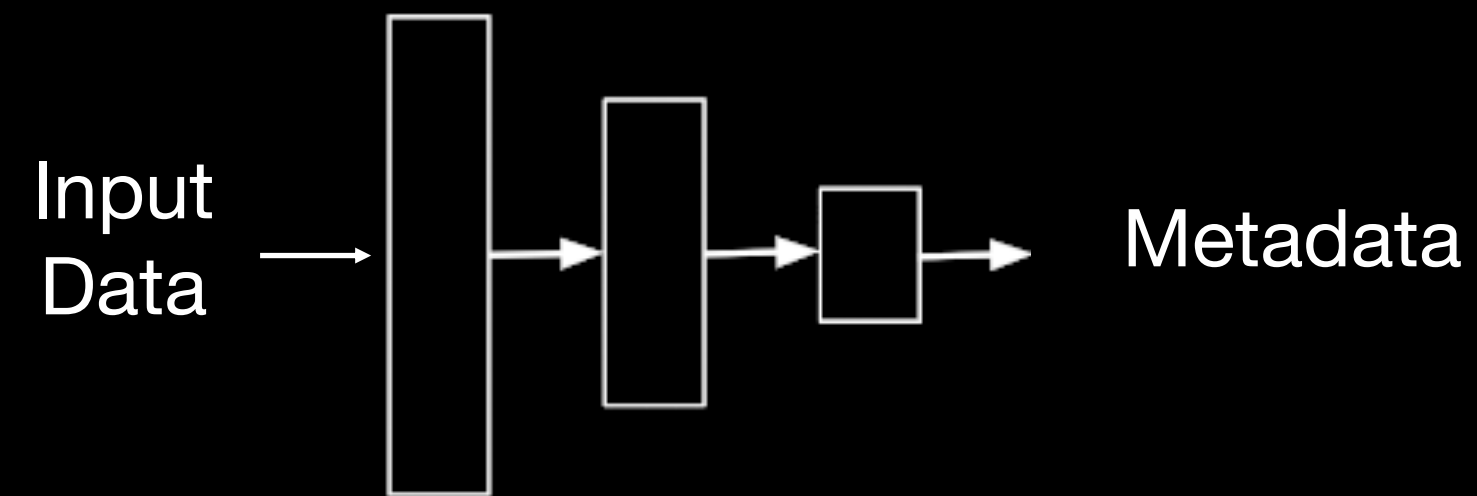
Deep Neural Networks

- Multi Layer Architecture

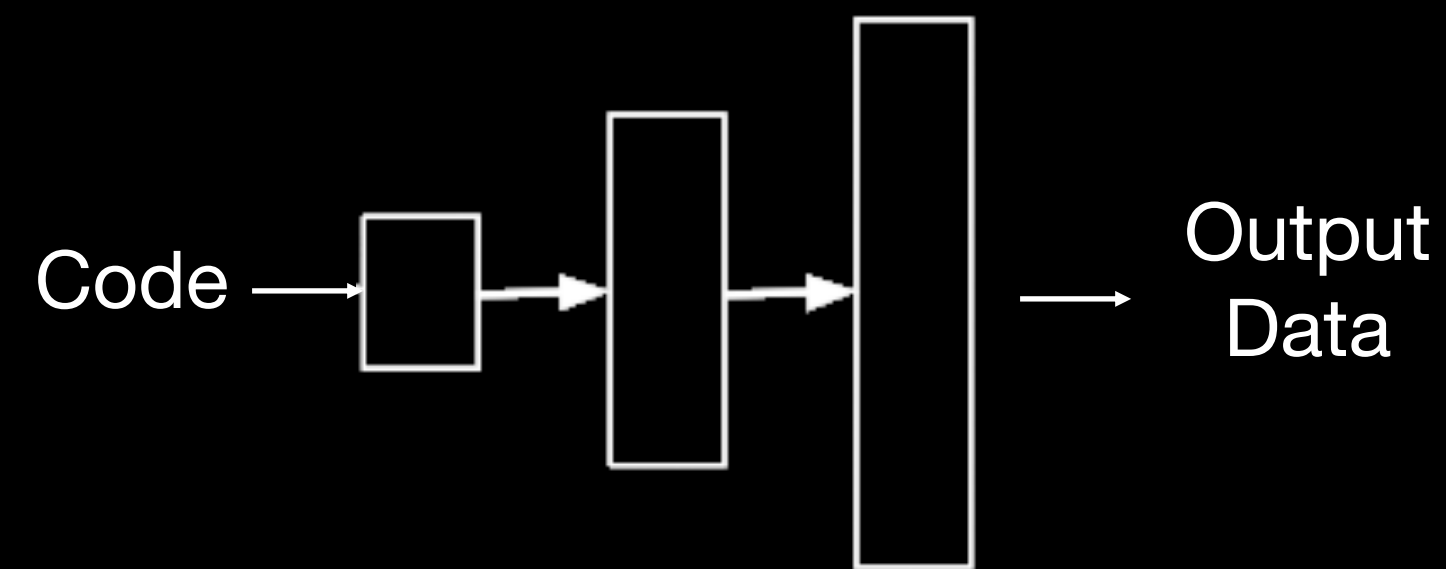


Types of Neural Networks

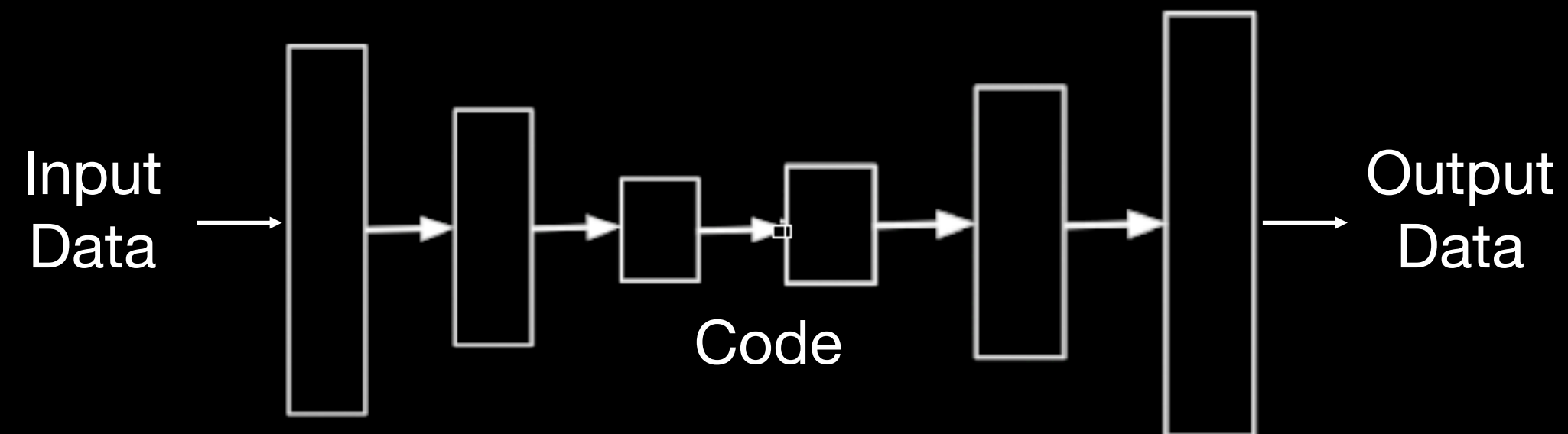
- Analysis



- Synthesis

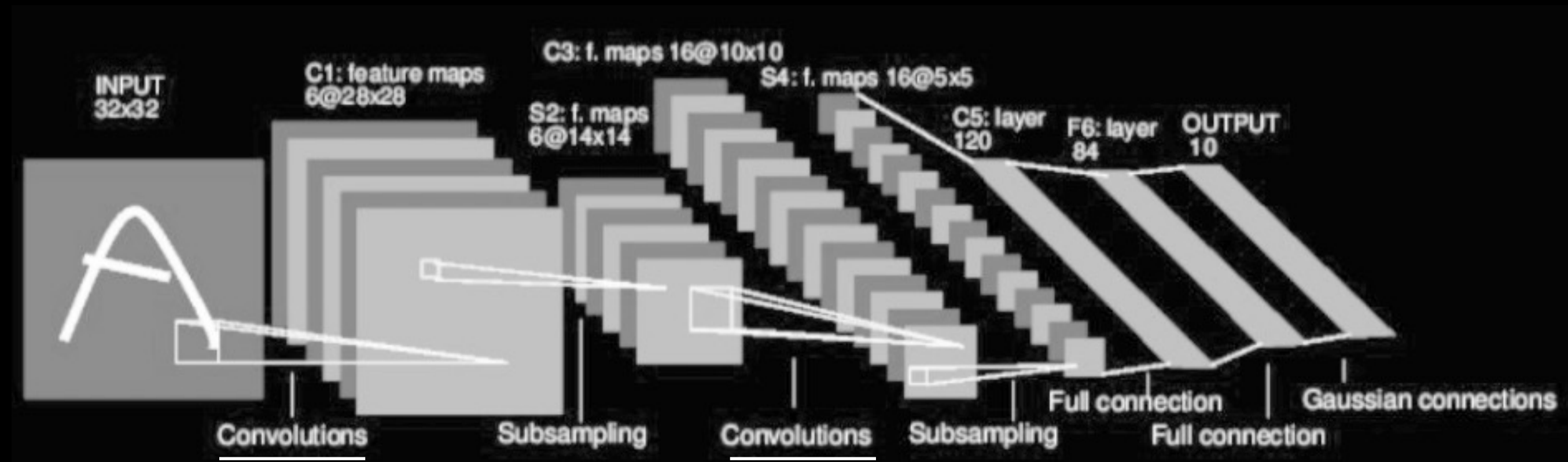


- Full

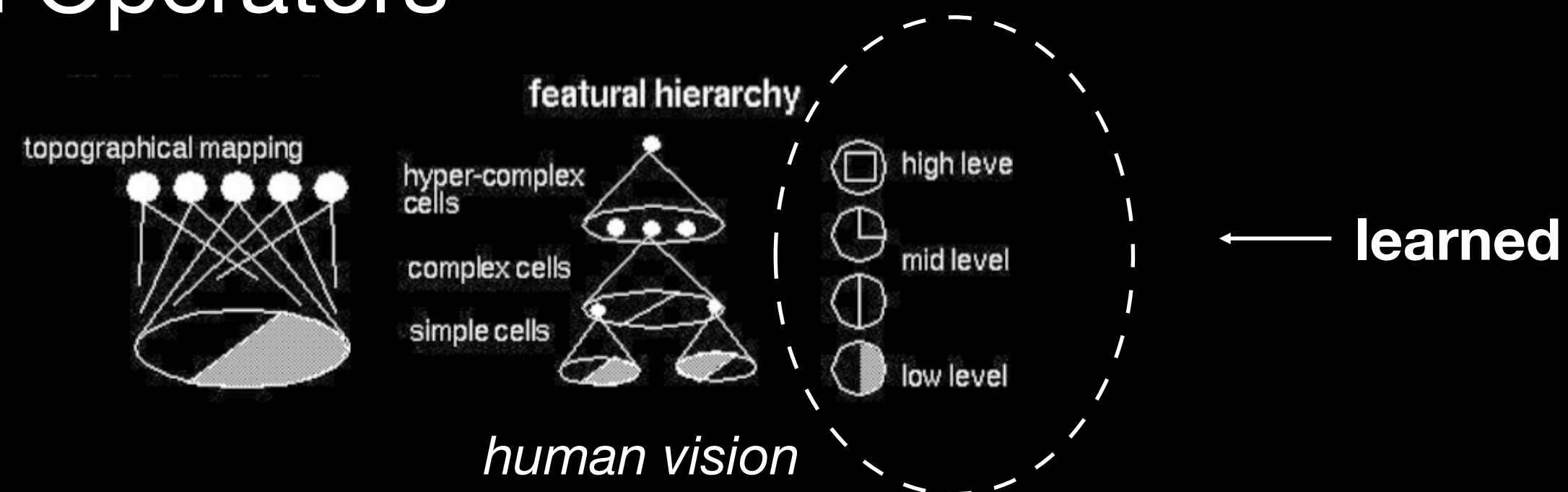


Convolutional Networks

- ➔ • Learn Features (Image)
- Deep Convolutional Neural Networks

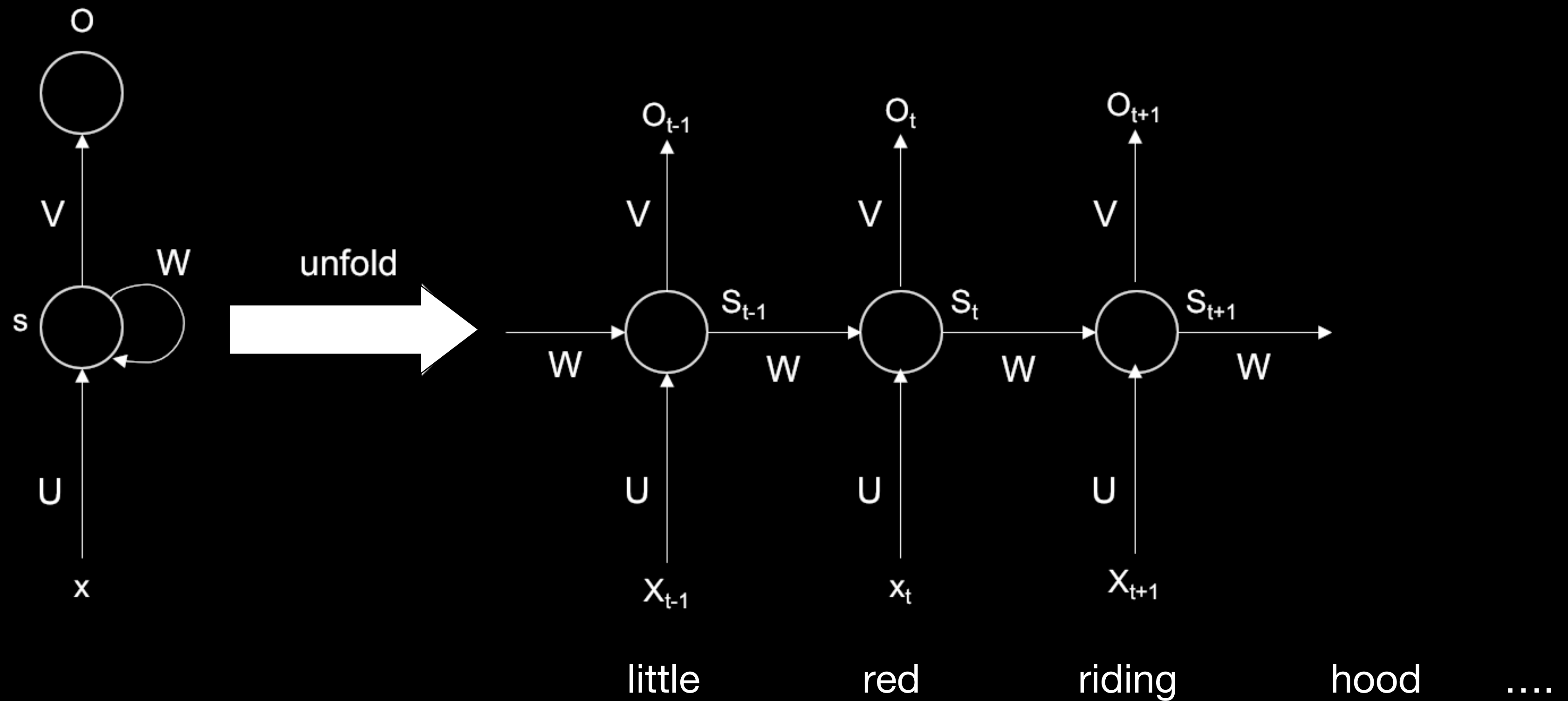


- Convolution Operators



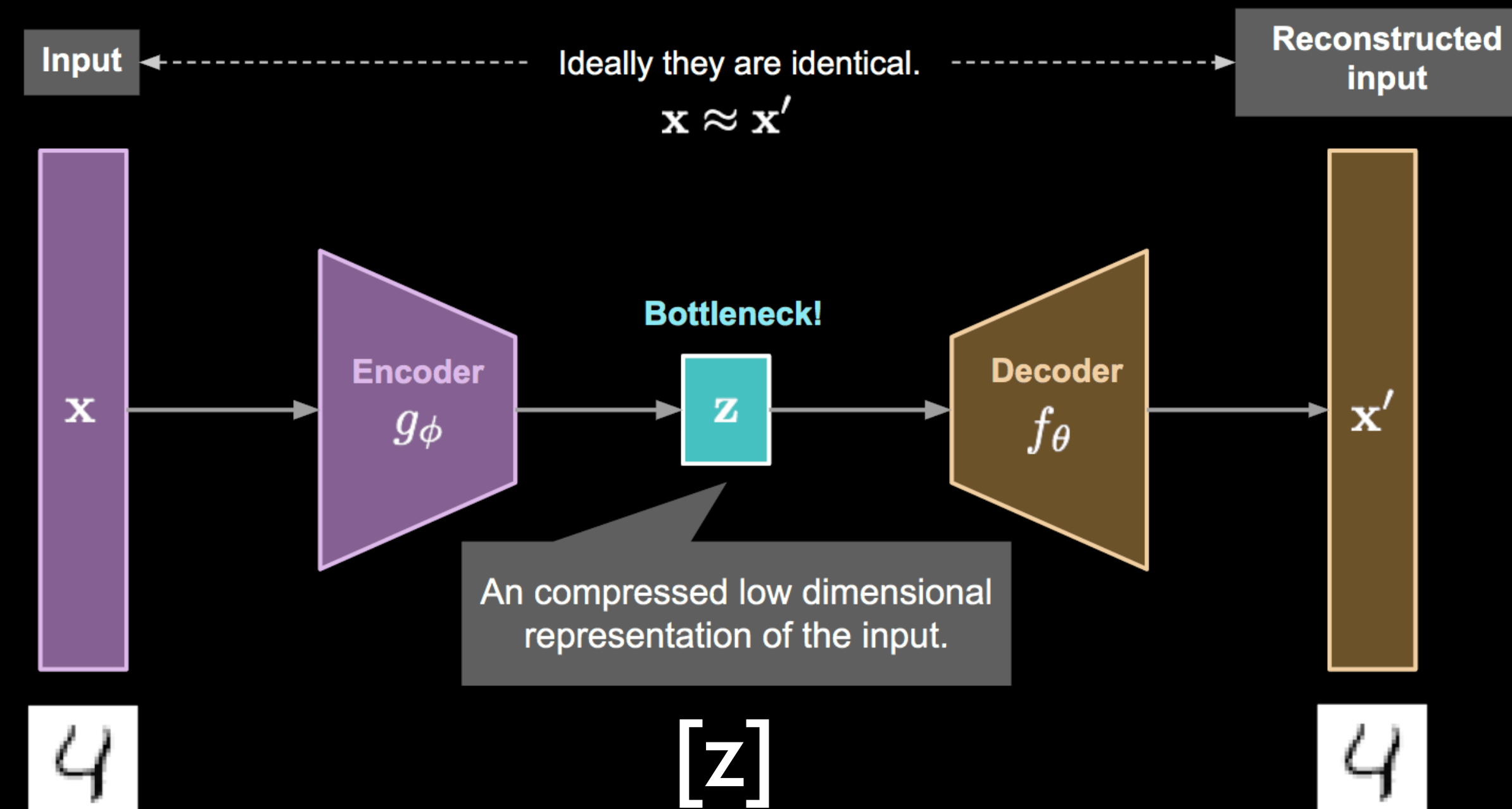
Recurrent Networks

- ➔ • Learn a Sequence Process (*Text*)



(Variational) Auto-Encoders

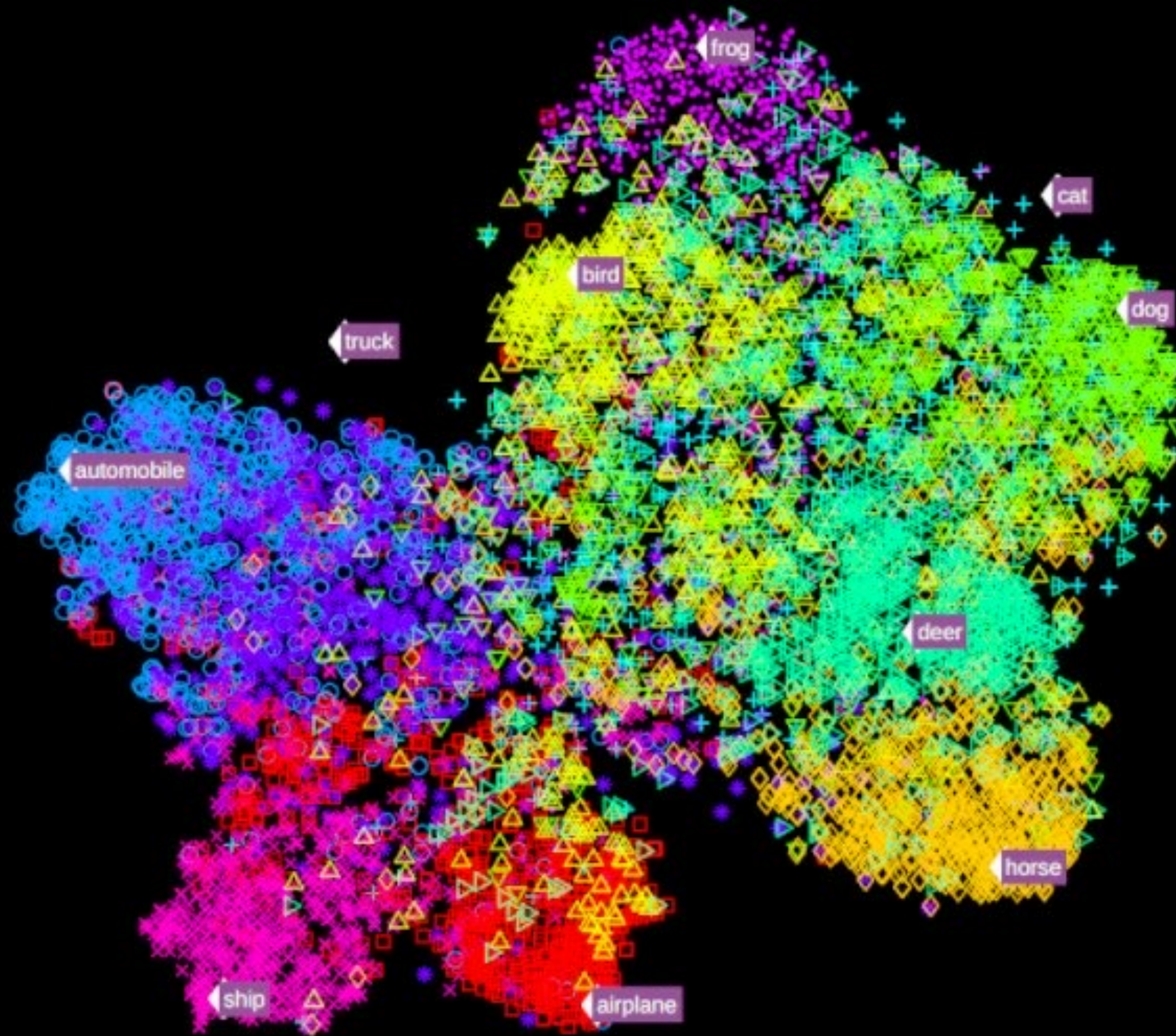
- ➔ • Learn Representations (*General Probability Distributions*)



Natural Language Models

Model

- + cat
- automobile
- truck
- frog
- × ship
- airplane
- ◇ horse
- △ bird
- ▽ dog
- ▷ deer

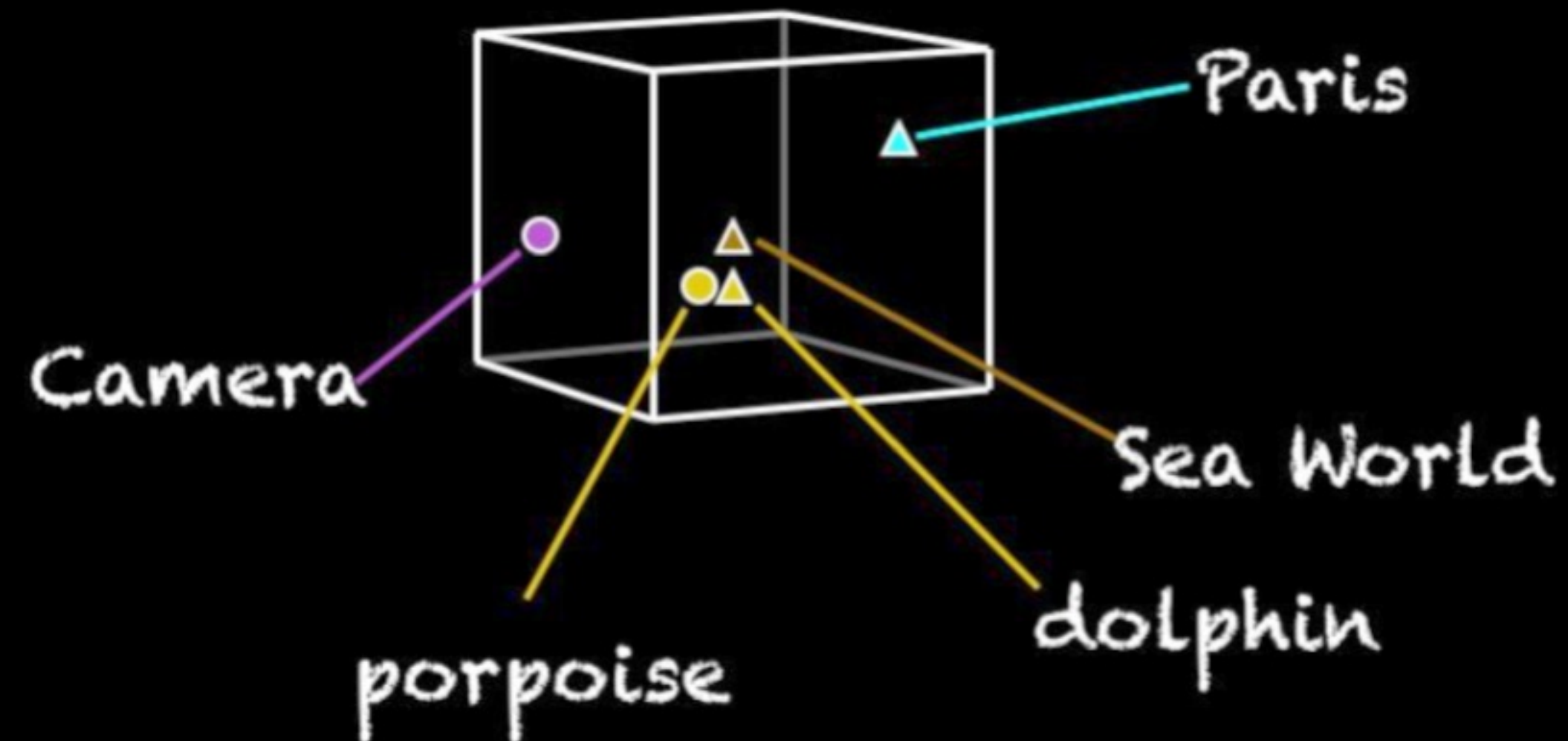


Semantic Structure

Model is a Subspace in a High Dimensional Space

Embeddings of Vector Spaces

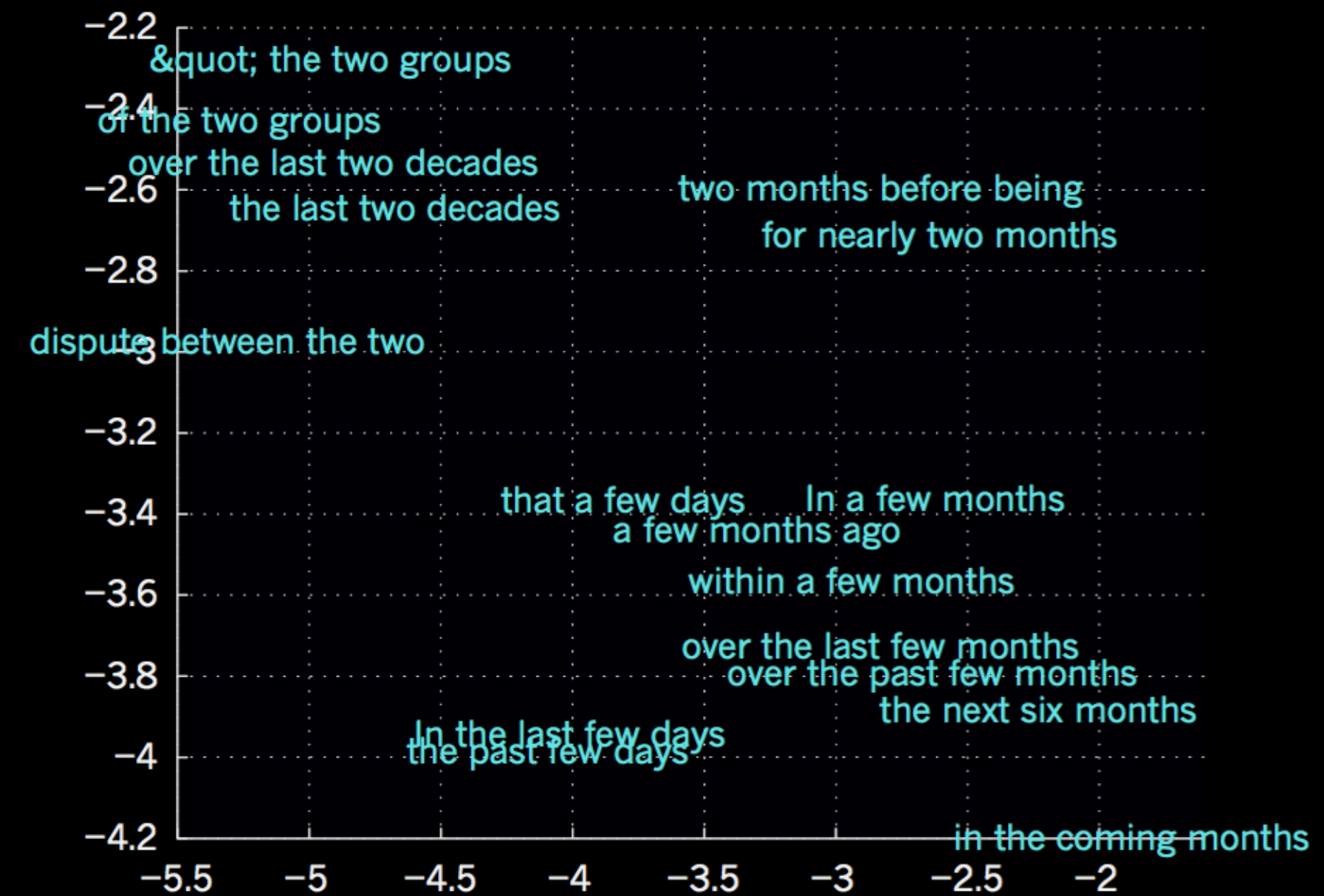
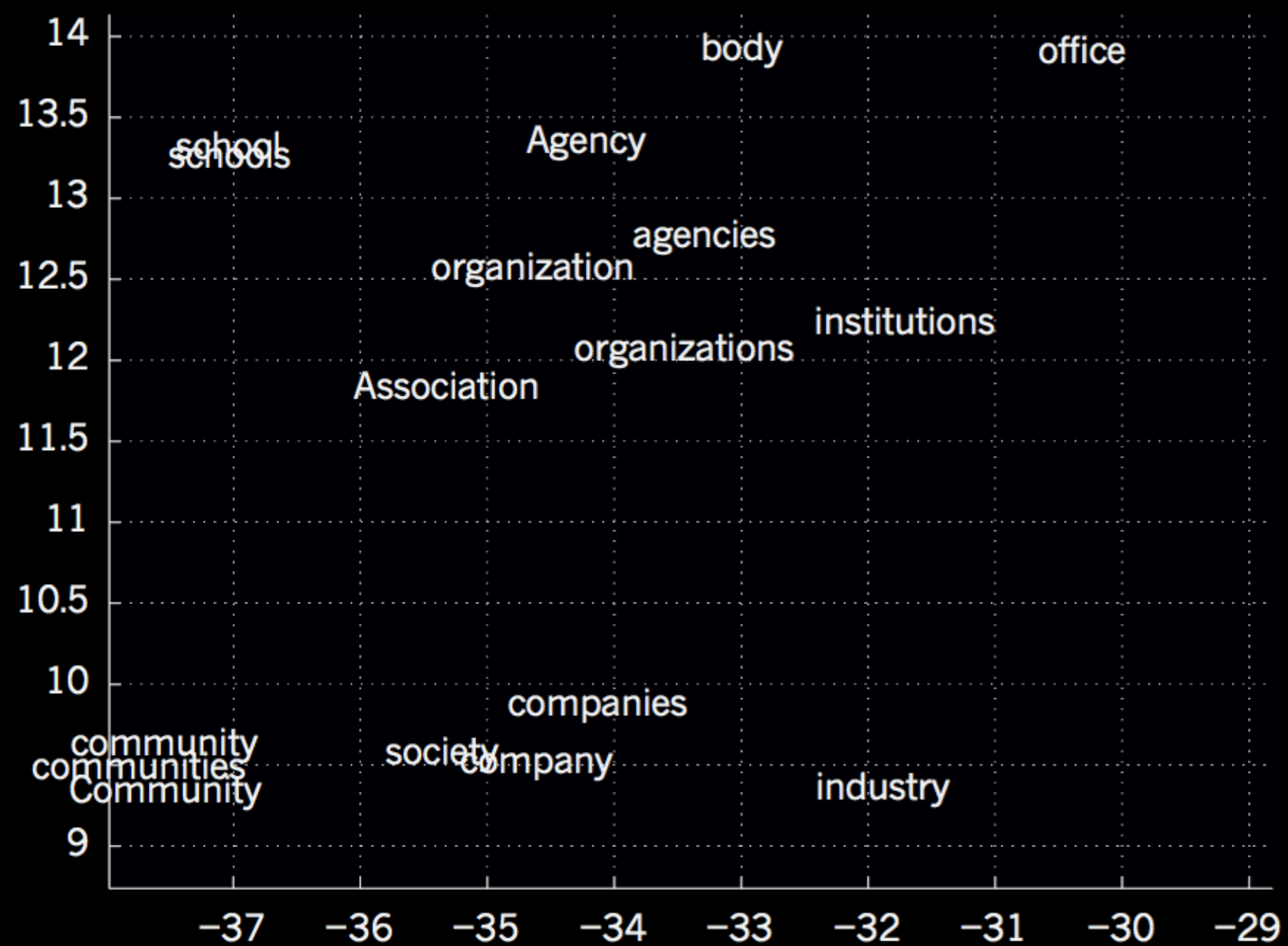
- Embedding Space (100D - 1000D)



Embedding Function: A look-up-table that maps sparse features into dense floating point vectors.

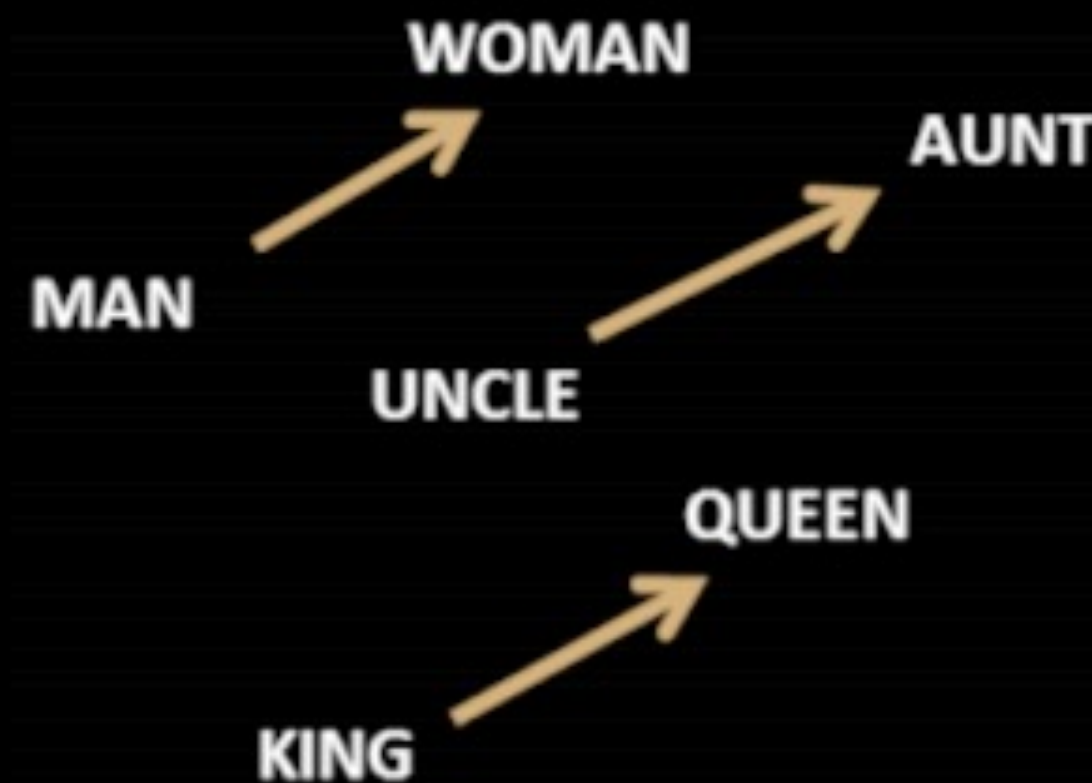
Word - 2 - Vect

- Semantically related words are close in vector space



Analogies / Operations

- Differences between words ~ Structure



$$W(\text{"woman"}) - W(\text{"man"}) \simeq W(\text{"aunt"}) - W(\text{"uncle"})$$

$$W(\text{"woman"}) - W(\text{"man"}) \simeq W(\text{"queen"}) - W(\text{"king"})$$

Generative Models

The First Paper (2014)

- Ian Goodfellow

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

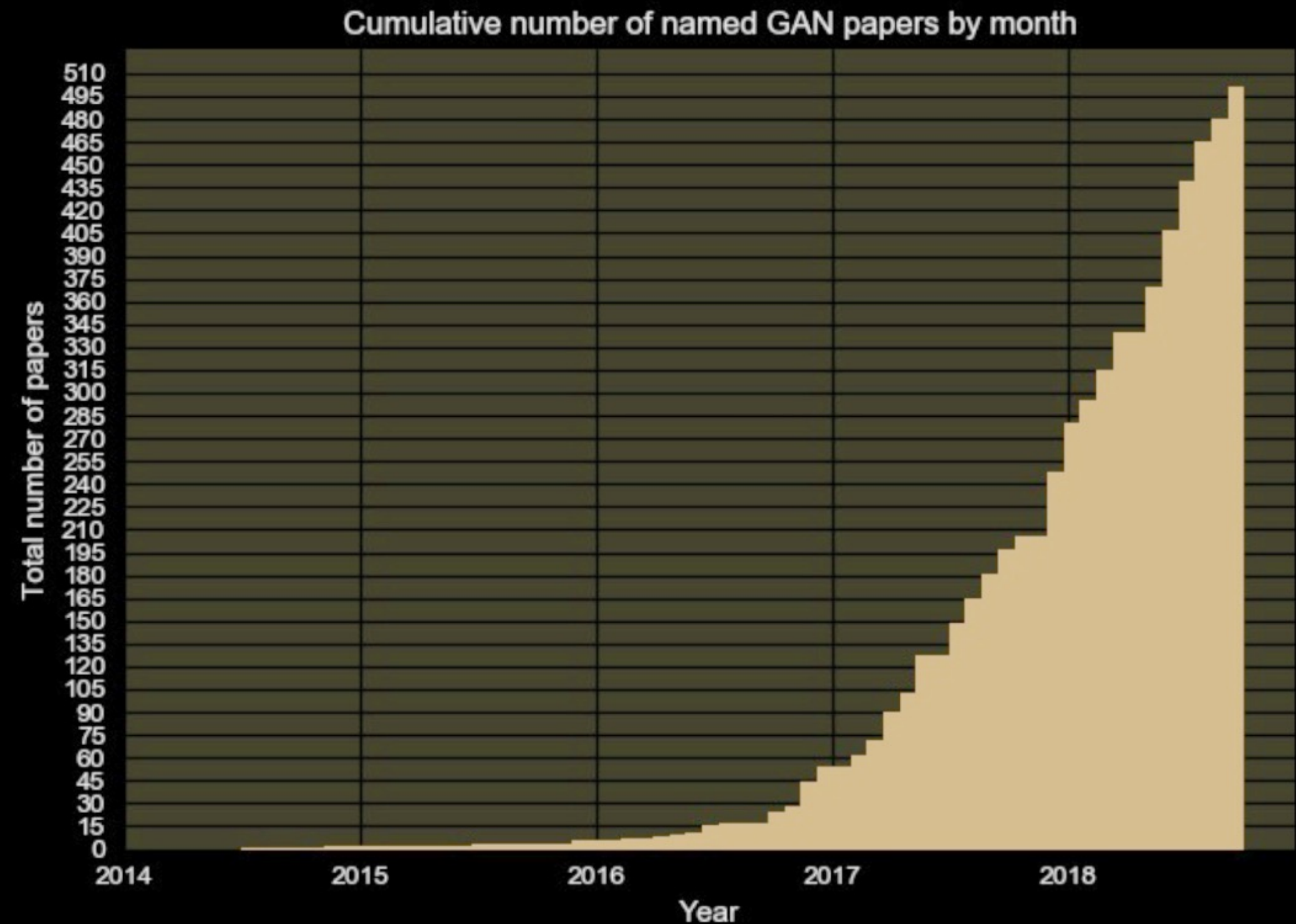
We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than 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.

arXiv:1406.10 Jun 2014

GANs Explosion (2017)

“The GAN Zoo”

- GAN - Generative Adversarial Networks
- 3D-GAN - Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling
- acGAN - Face Aging With Conditional Generative Adversarial Networks
- AC-GAN - Conditional Image Synthesis With Auxiliary Classifier GANs
- AdaGAN - AdaGAN: Boosting Generative Models
- AEGAN - Learning Inverse Mapping by Autoencoder based Generative Adversarial Nets
- AffGAN - Amortised MAP Inference for Image Super-resolution
- AL-CGAN - Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts
- ALI - Adversarially Learned Inference
- AM-GAN - Generative Adversarial Nets with Labeled Data by Activation Maximization
- AnoGAN - Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery
- ArtGAN - ArtGAN: Artwork Synthesis with Conditional Categorical GANs
- b-GAN - b-GAN: Unified Framework of Generative Adversarial Networks
- Bayesian GAN - Deep and Hierarchical Implicit Models
- BEGAN - BEGAN: Boundary Equilibrium Generative Adversarial Networks
- BiGAN - Adversarial Feature Learning
- BS-GAN - Boundary-Seeking Generative Adversarial Networks
- CGAN - Conditional Generative Adversarial Nets
- CaloGAN - CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks
- CCGAN - Semi-Supervised Learning with Context-Conditional Generative Adversarial Networks
- CatGAN - Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks
- CoGAN - Coupled Generative Adversarial Networks



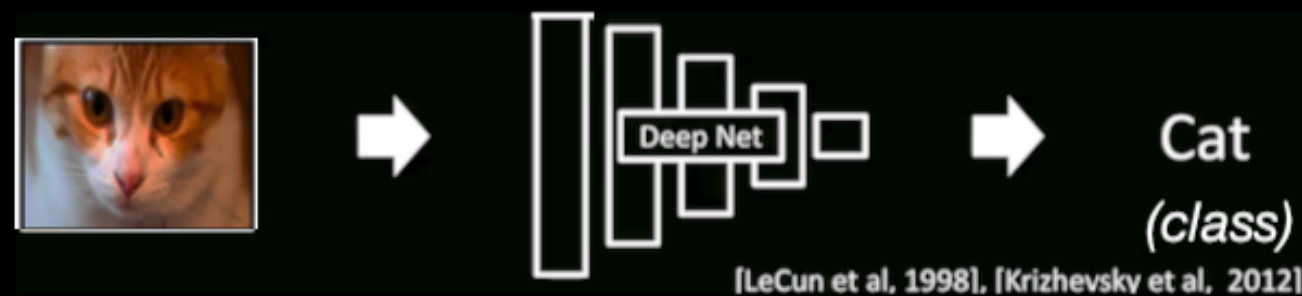
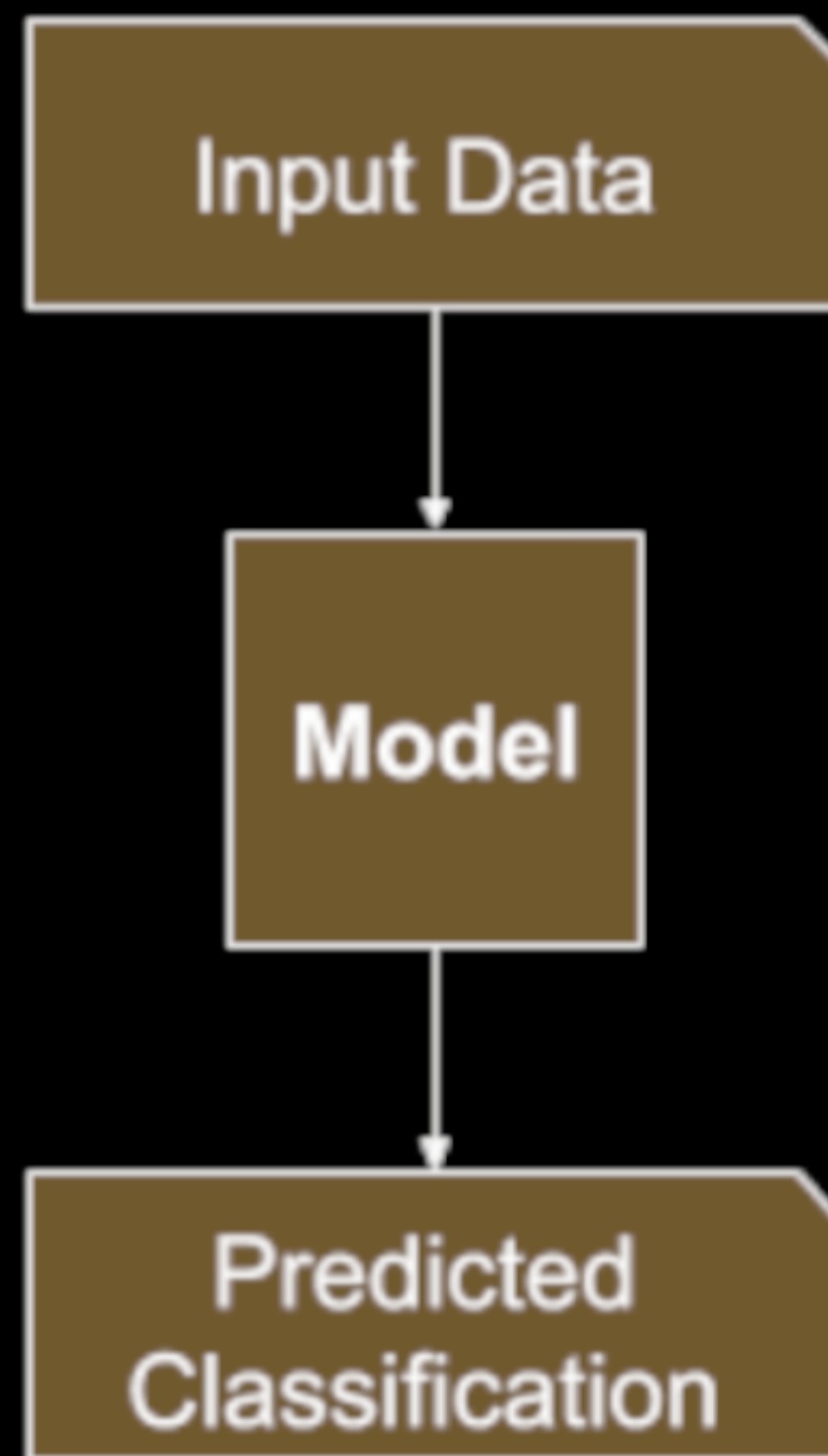
<https://github.com/hindupuravinash/the-gan-zoo>

(600+ papers)

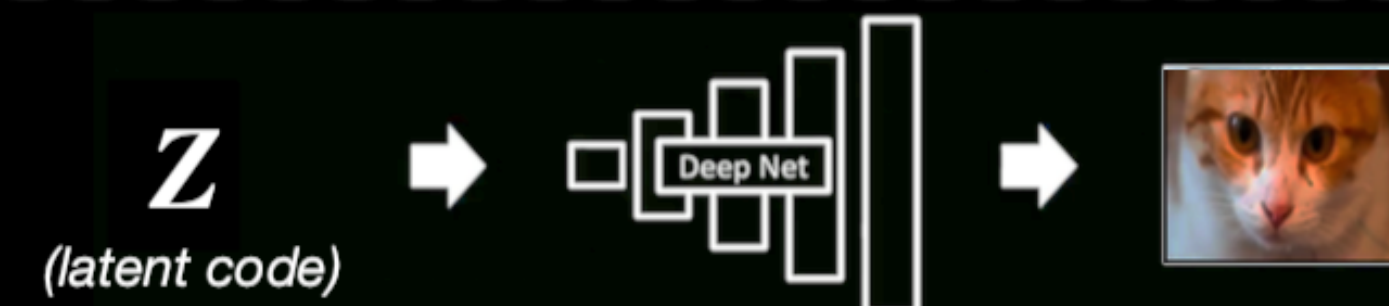
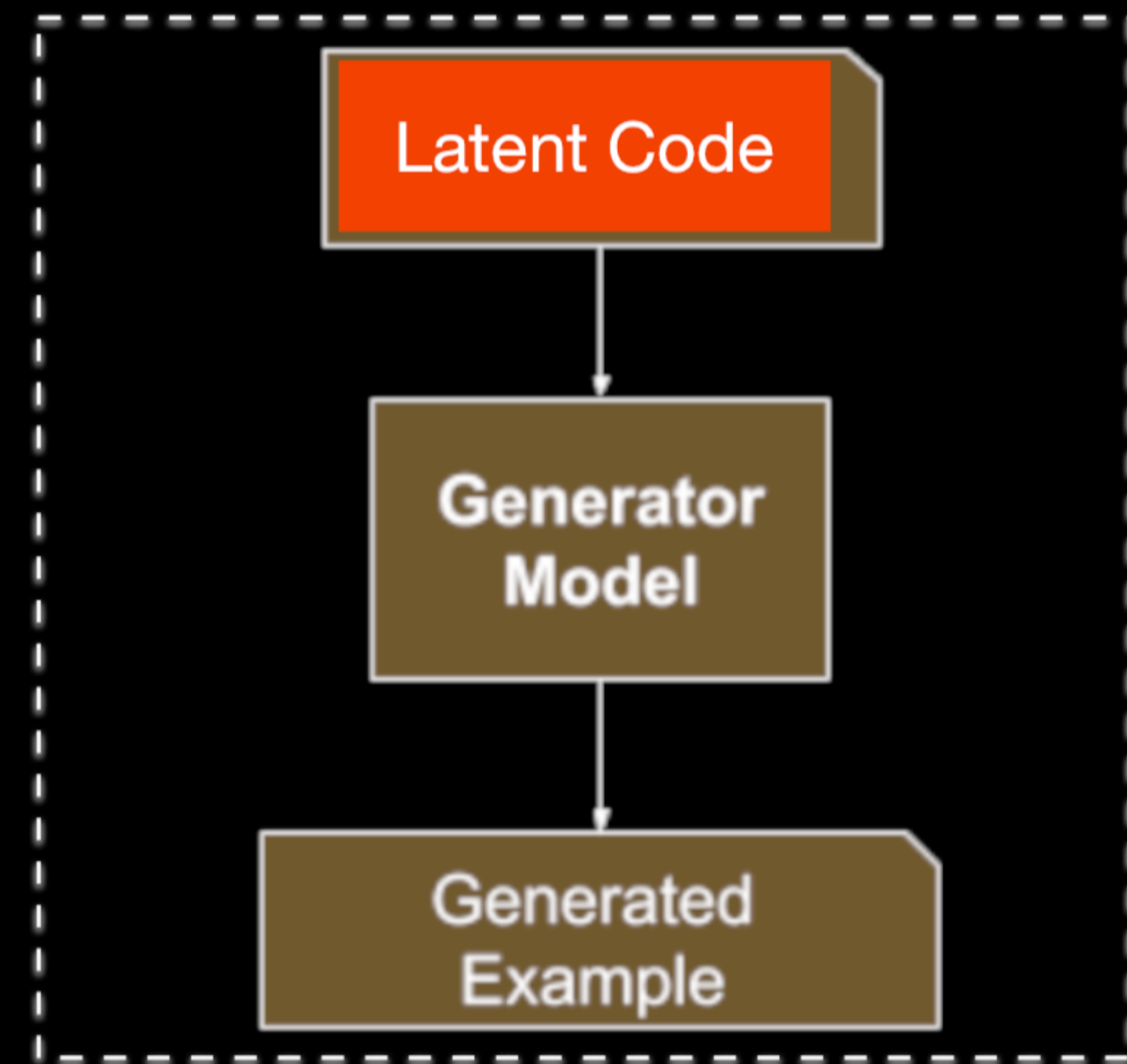


Analysis vs. Synthesis

- Discriminative Model

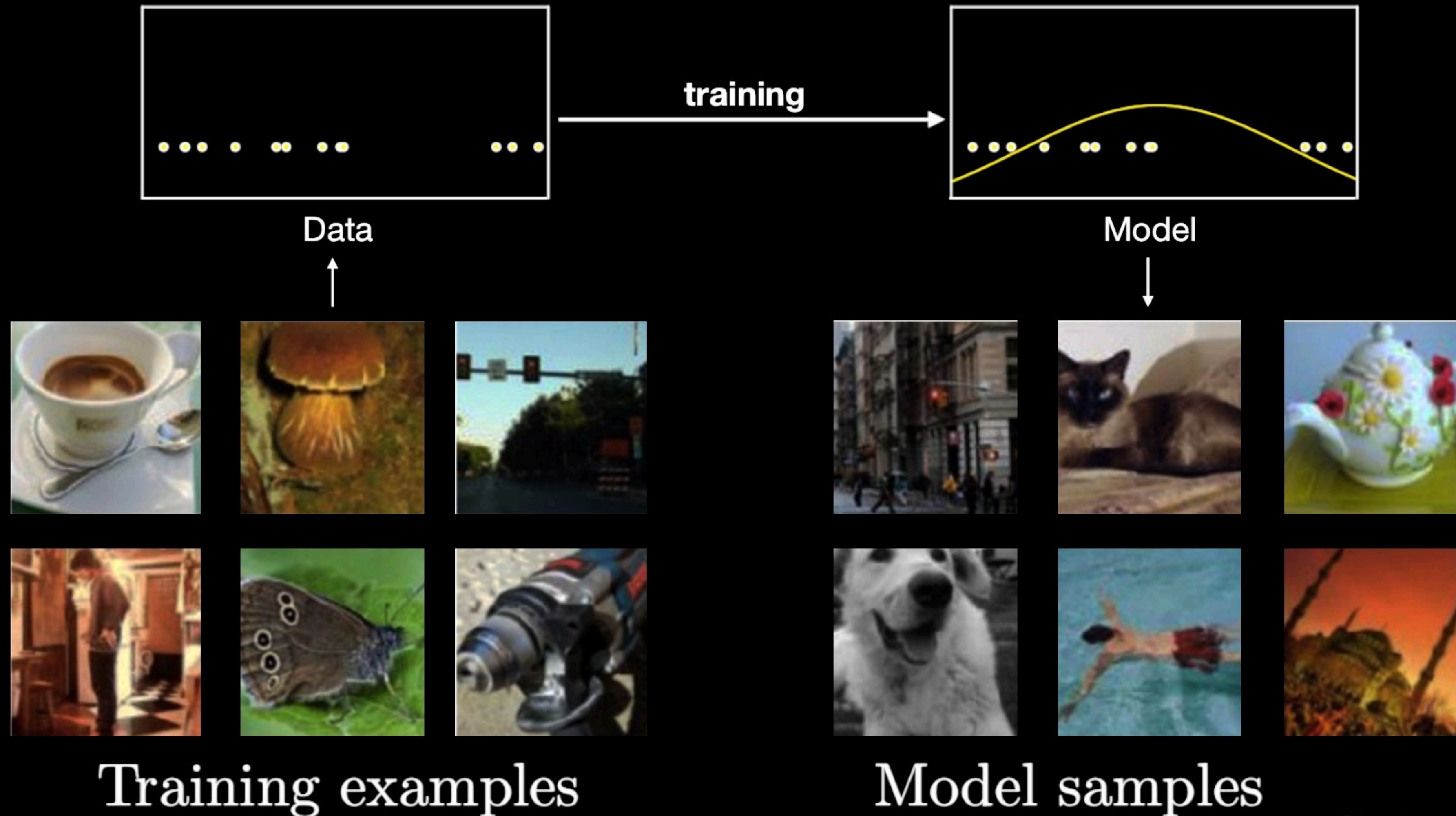


- Generative Model



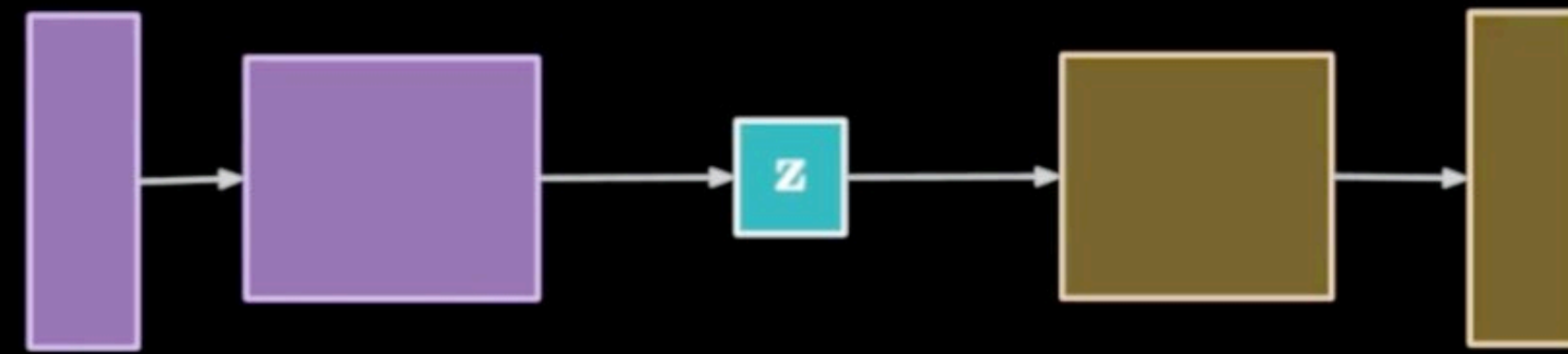
Modeling a Data Distribution

- Density estimation



Training and Inference

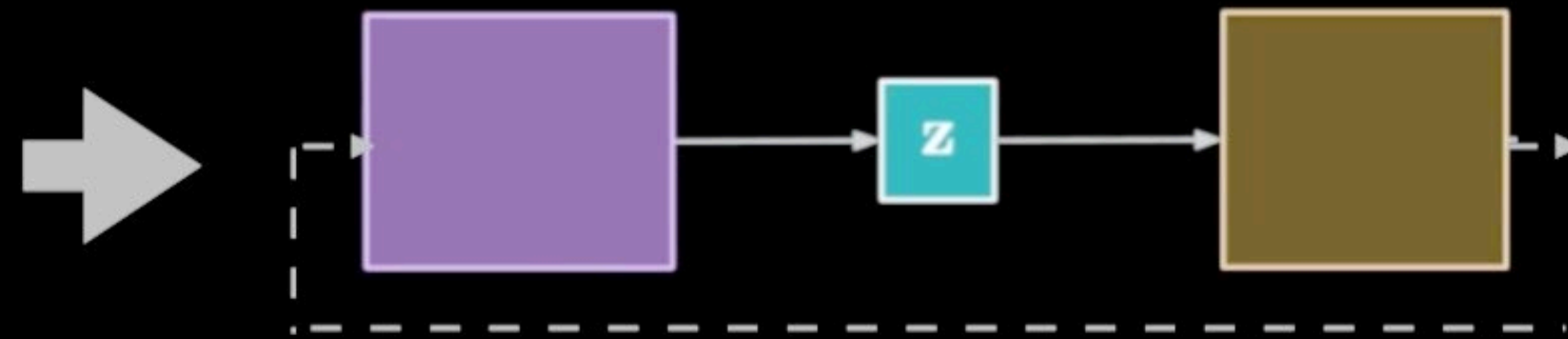
- Network (Model)



- *Training the Model*

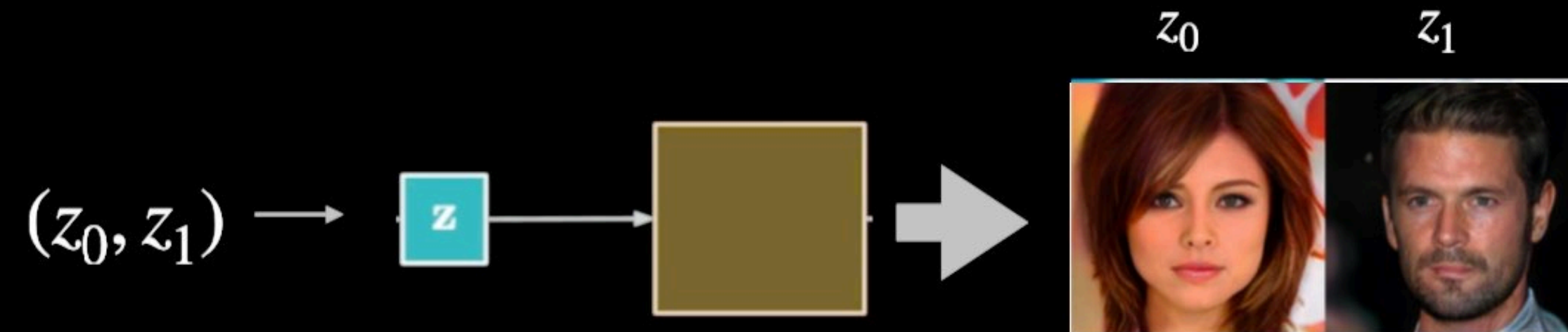


Training Data
(CelebA)



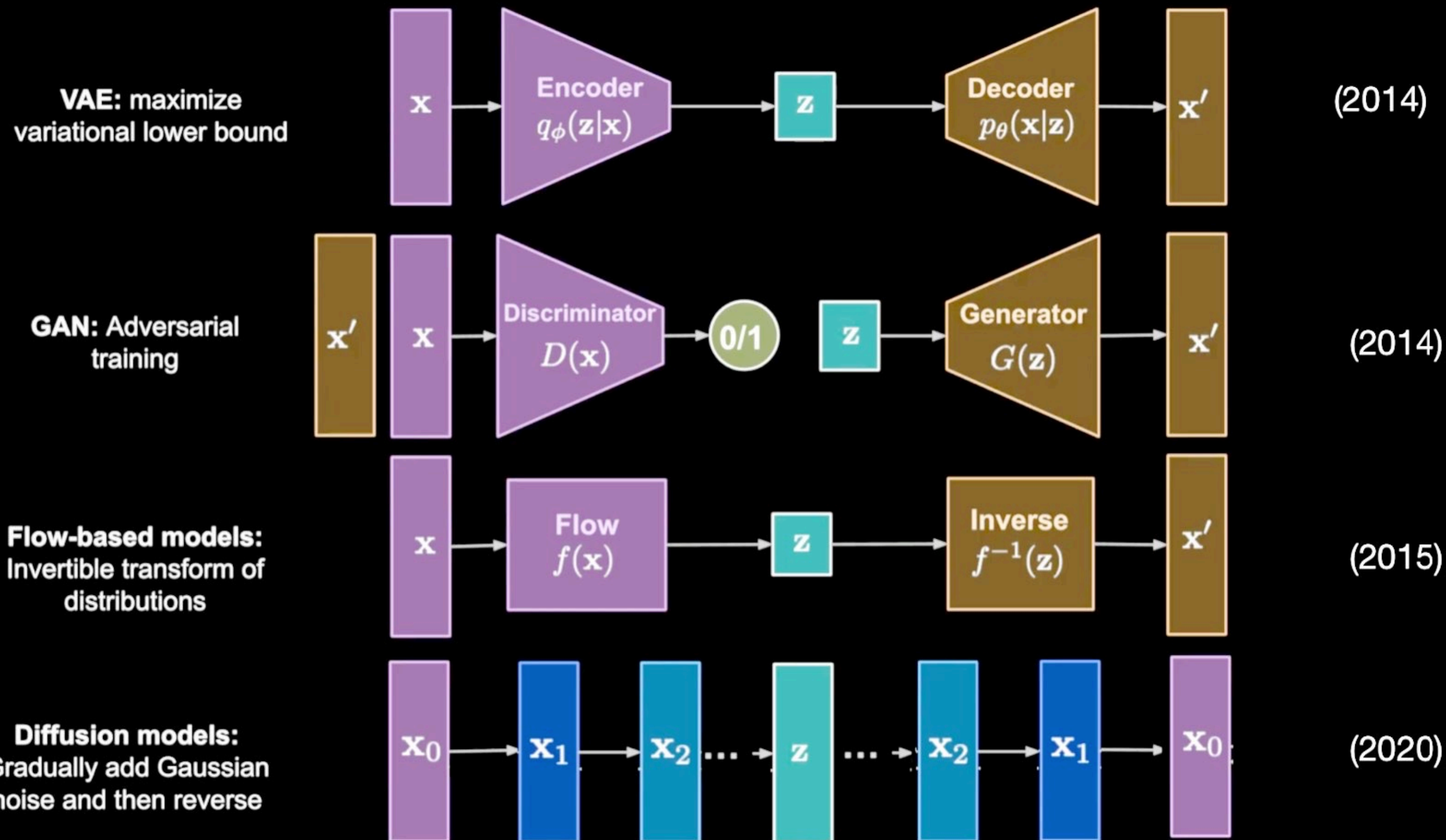
(self-supervised training)

- *Sample Generation*



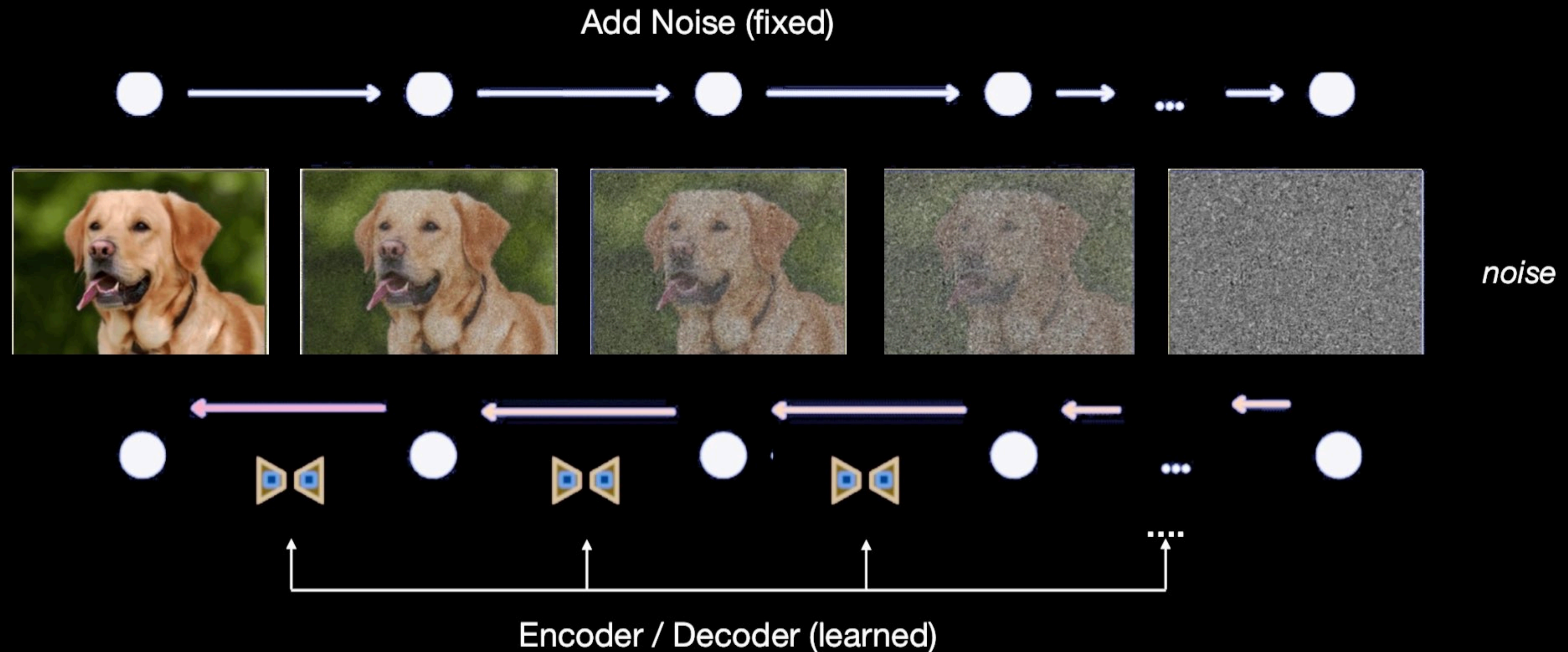
Sample Generator
(Karras et al, 2017)

Types of Generative Models



Diffusion Models

- Sequential Application of Denoising Auto-Encoders



Text & Images

Clip - unClip

- CLIP : *image to text*



**"a photo of
guacamole, a
type of food"**

(Image Caption)

- unCLIP : *text to image*

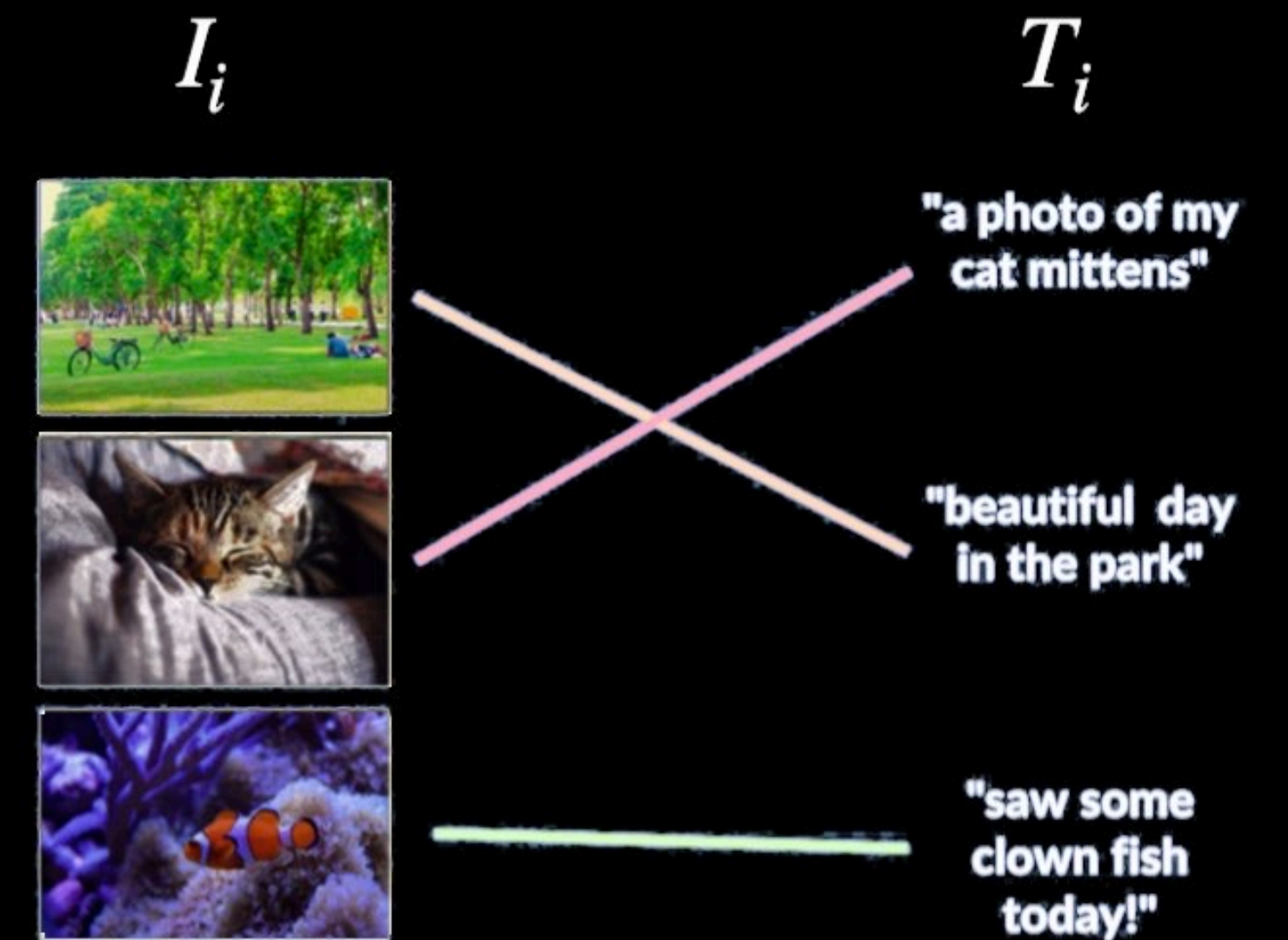
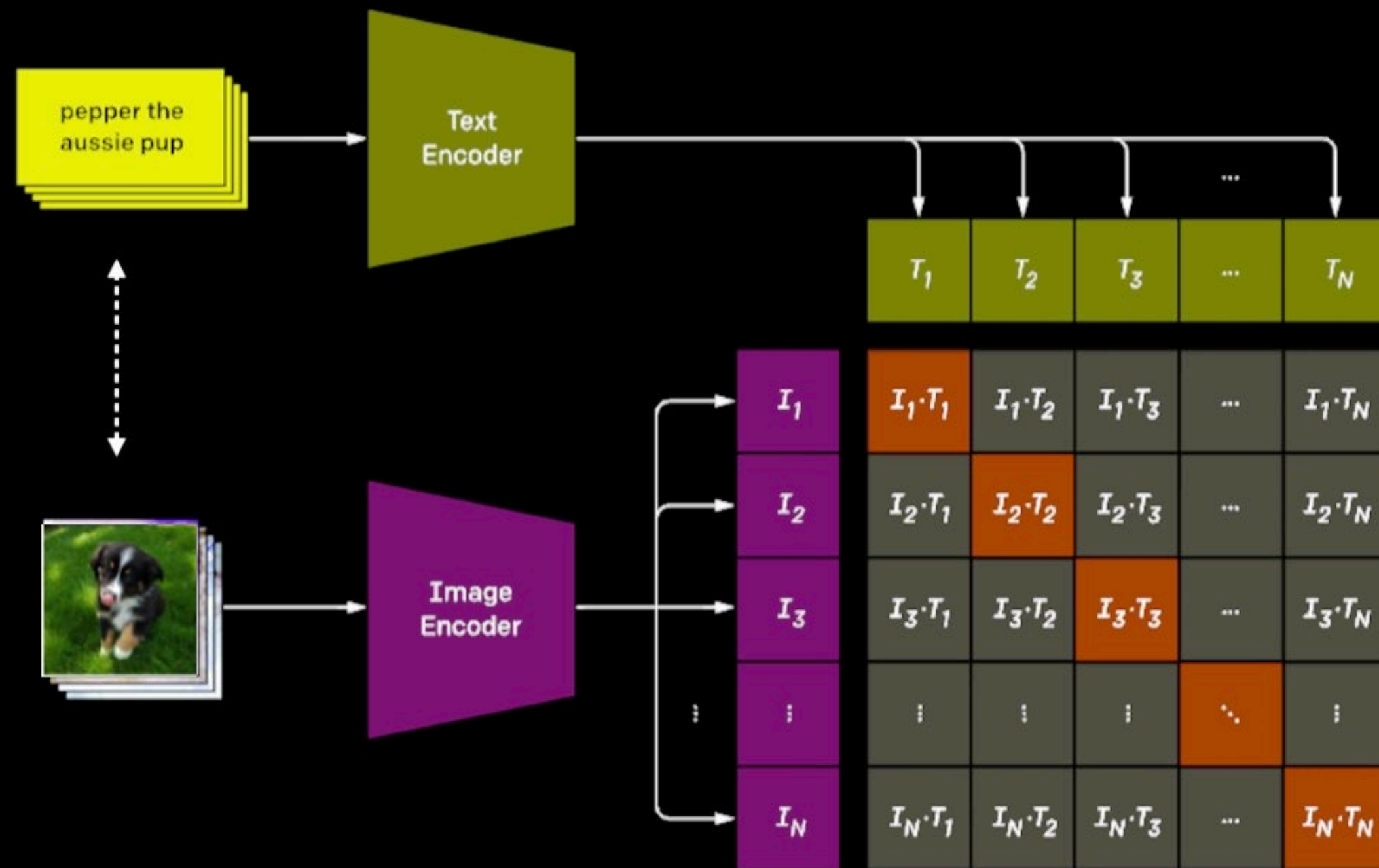
**"An astronaut
lounging in a
tropical resort
in space in a
photorealistic
style"**



(Text-Image Generation) ★

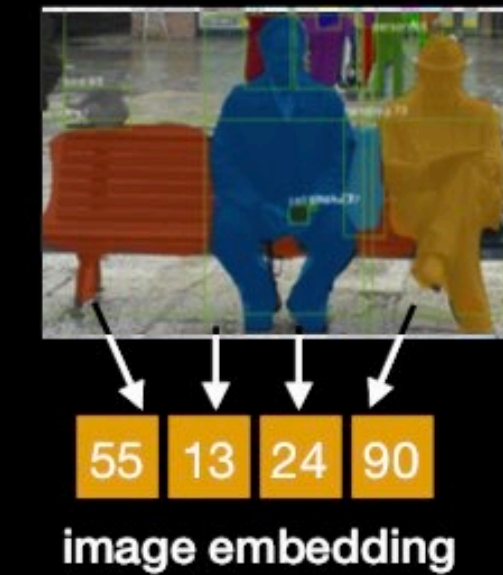
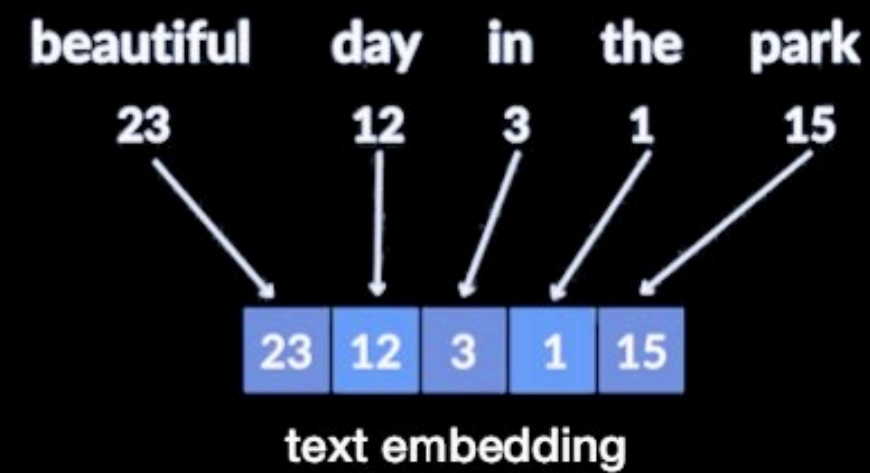
Contrastive Language-Image Pre-training

- Learn Text-Picture Correlations

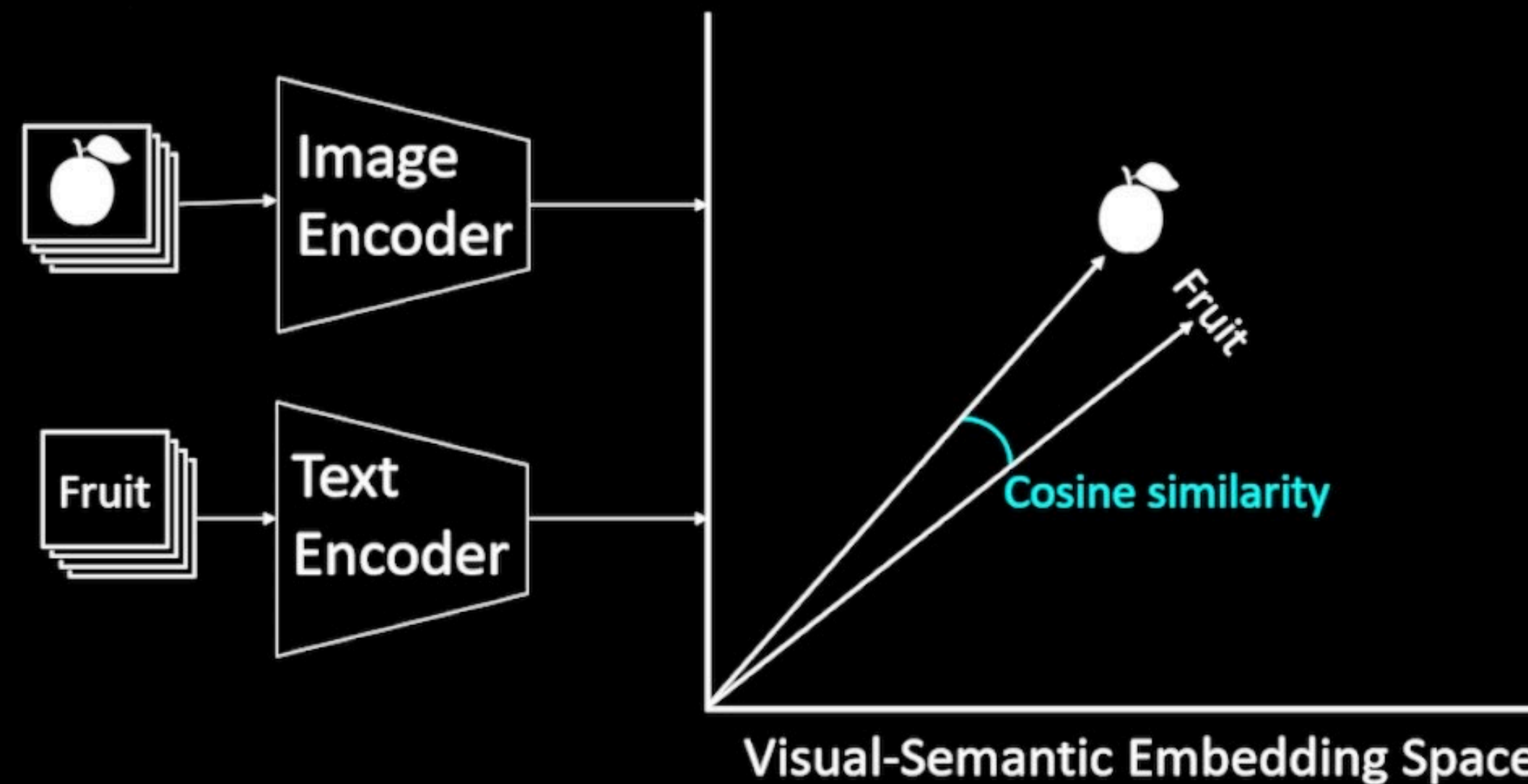


Visual Concept Prediction

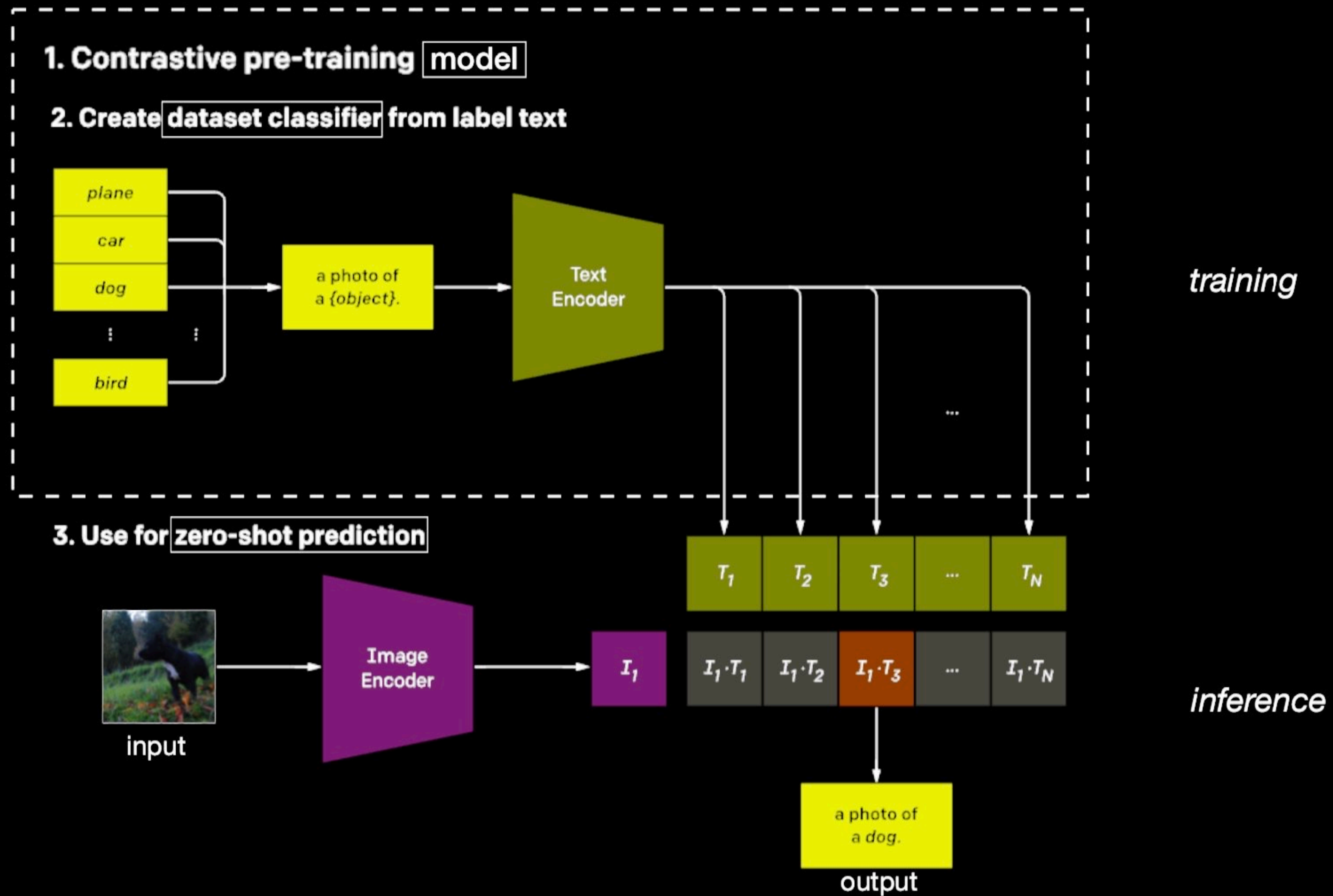
- Embeddings



- Visual-Semantic Similarity



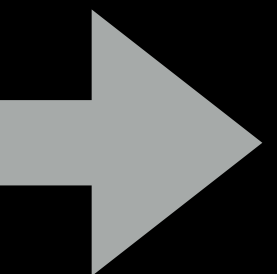
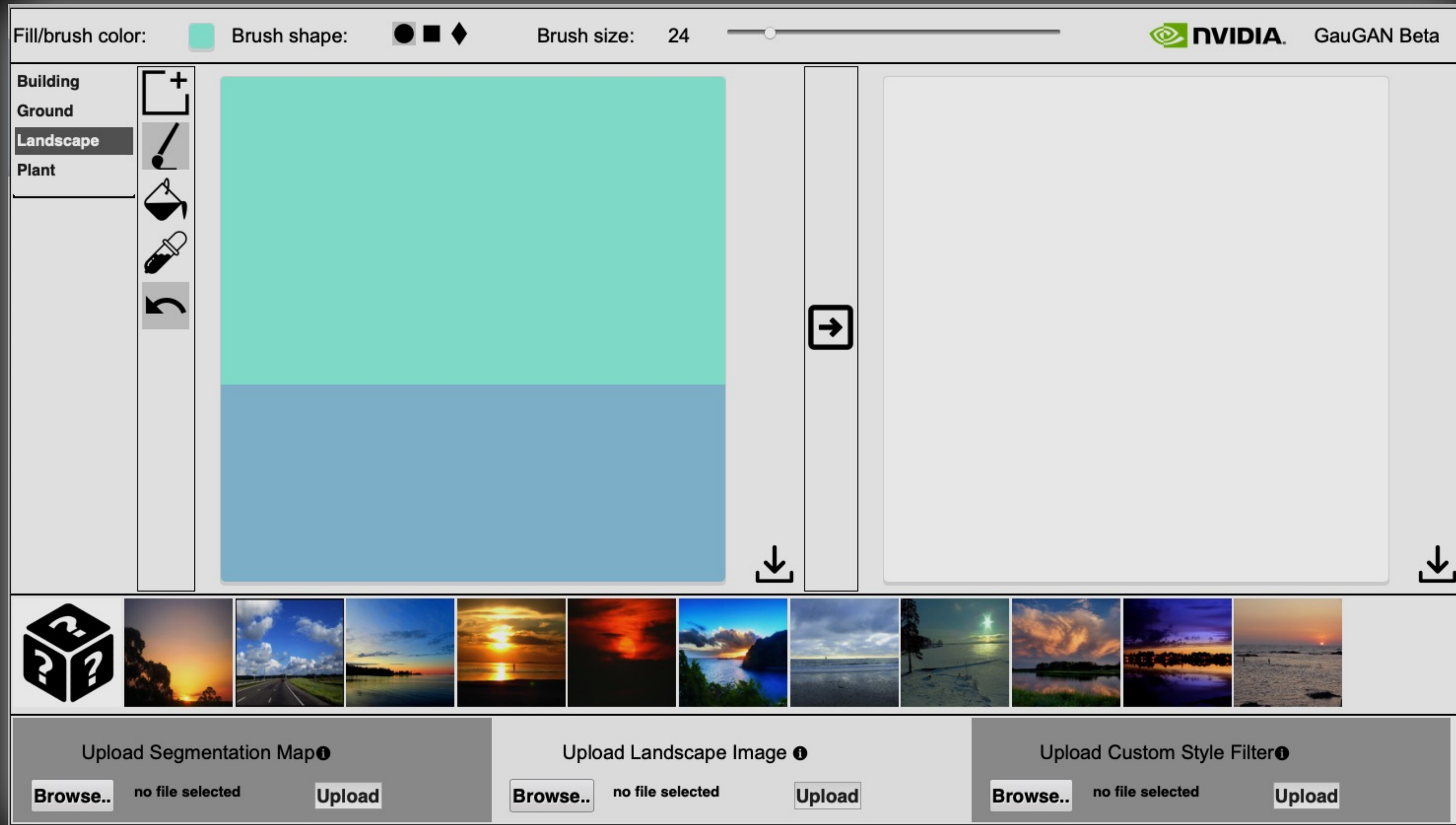
Clip

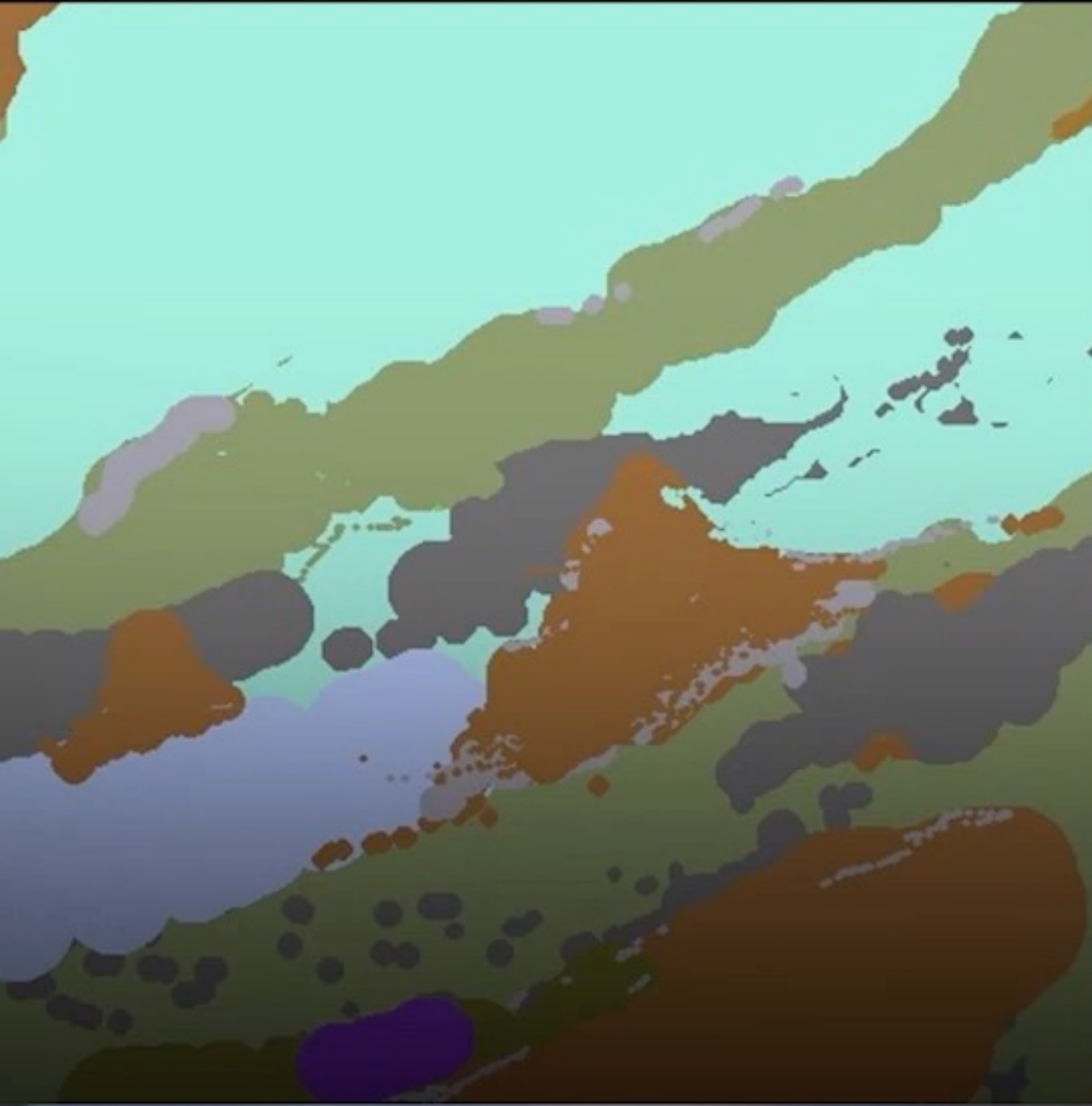


Creative Tools

GauGAN

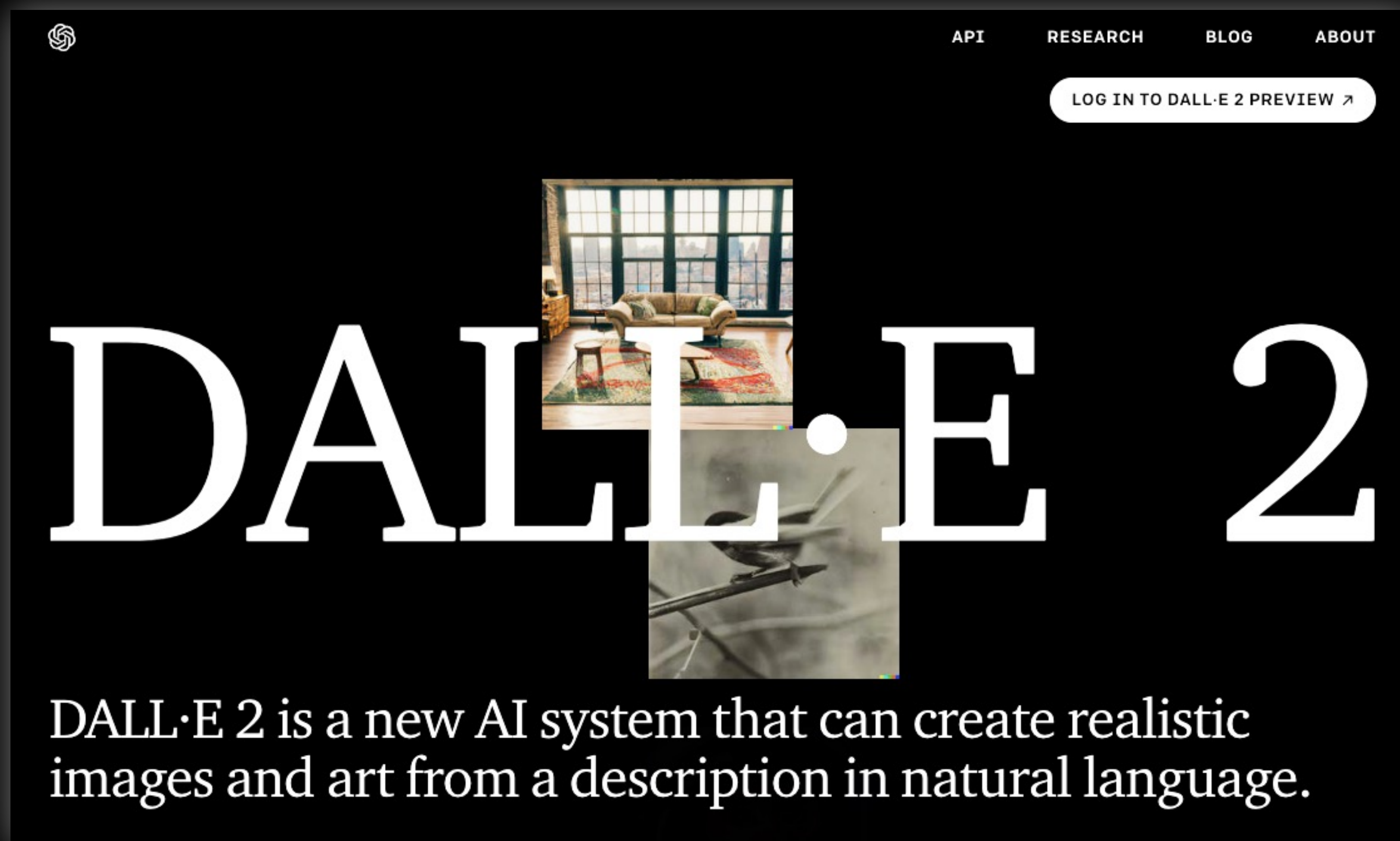
- Visual Semantics



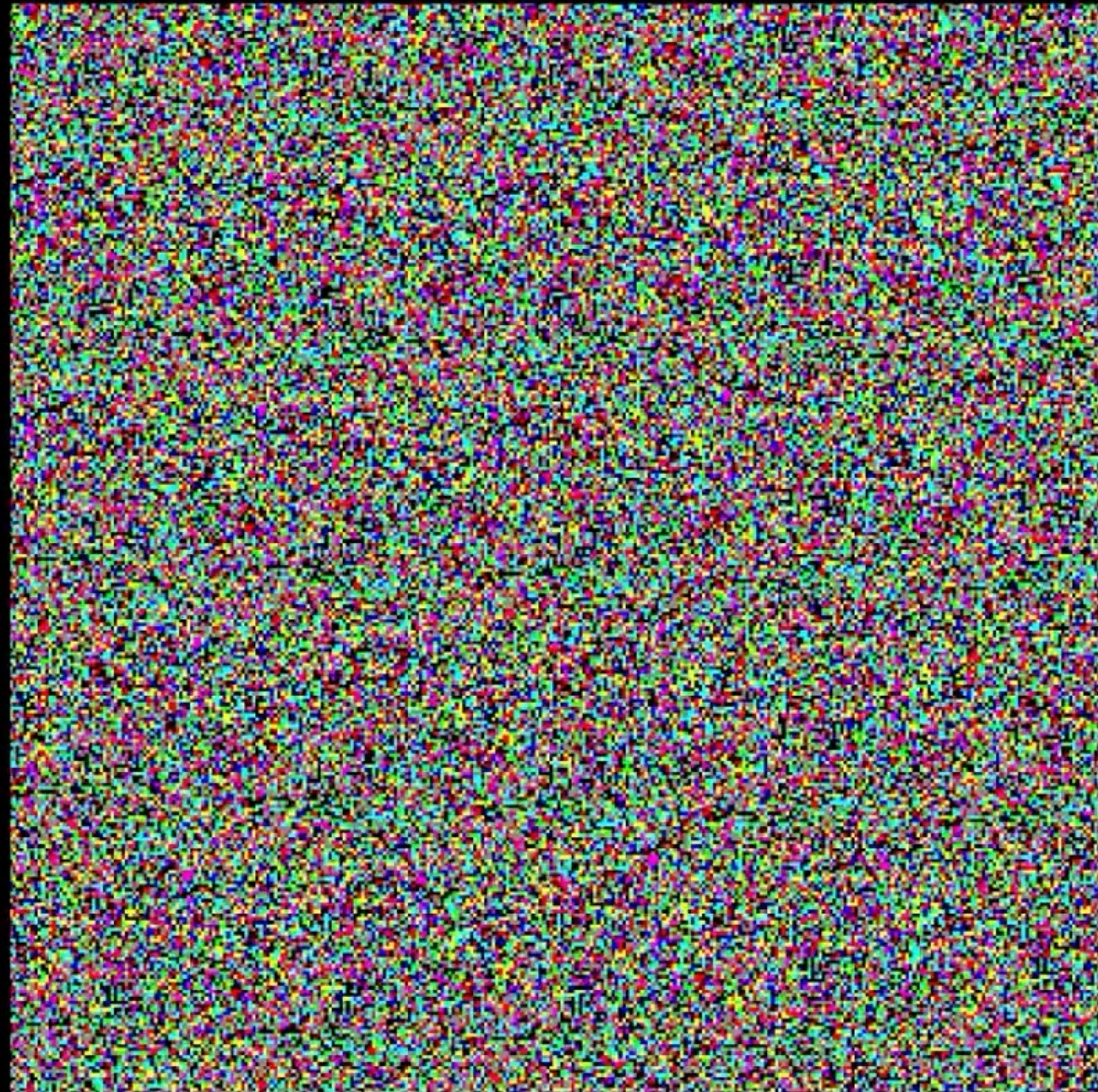


DALL-E 2

- Lexical-Visual Semantics



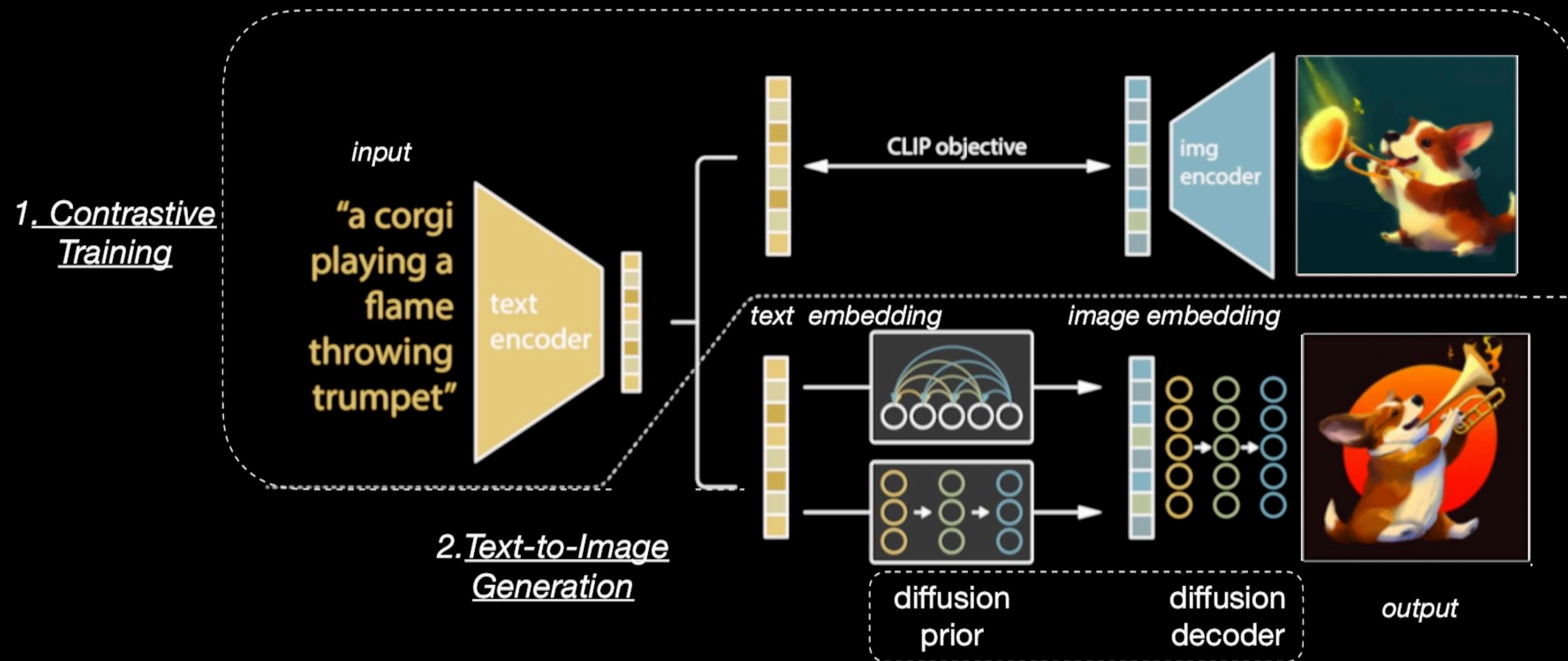
The image shows a screenshot of the DALL-E 2 website homepage. At the top left is the OpenAI logo. The top right navigation bar includes links for API, RESEARCH, BLOG, and ABOUT. A prominent button in the top right corner reads "LOG IN TO DALL-E 2 PREVIEW" with an external link icon. The main visual is the large text "DALL·E 2" in a white serif font, centered on a dark background. Two small images are overlaid on the text: a living room with a large window and a colorful rug, and a bird perched on a branch. Below the title, a paragraph of text reads: "DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language."



**A Look Deep Inside:
DALL-E, Stable Diffusion, et. al.**

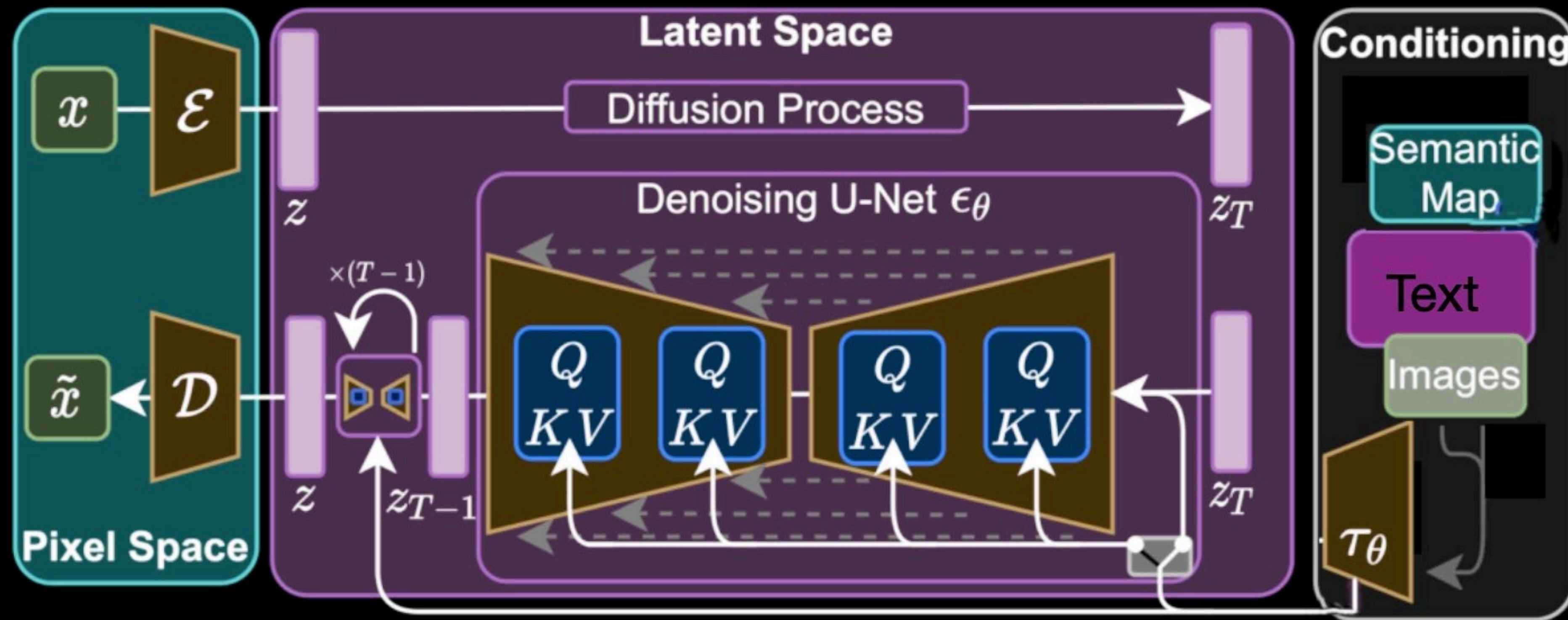
DALL-E 2 (unClip)

- OpenAI (April 2022)
 - Clip + Diffusion



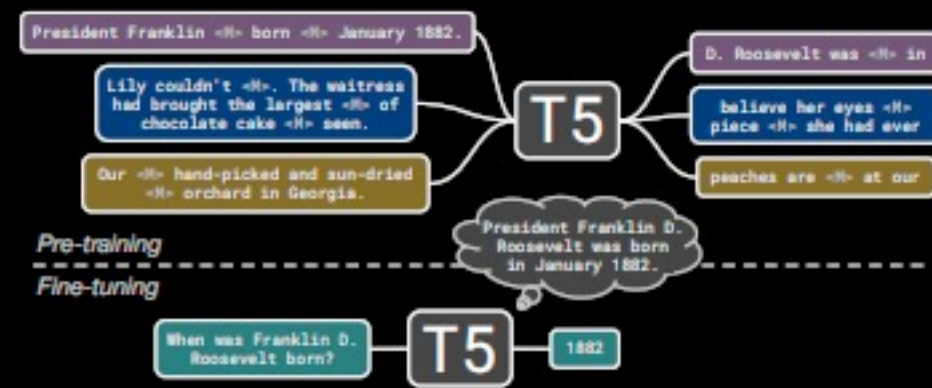
Stable Diffusion

- LMU Munich (April 2022)
 - Latent Diffusion Model

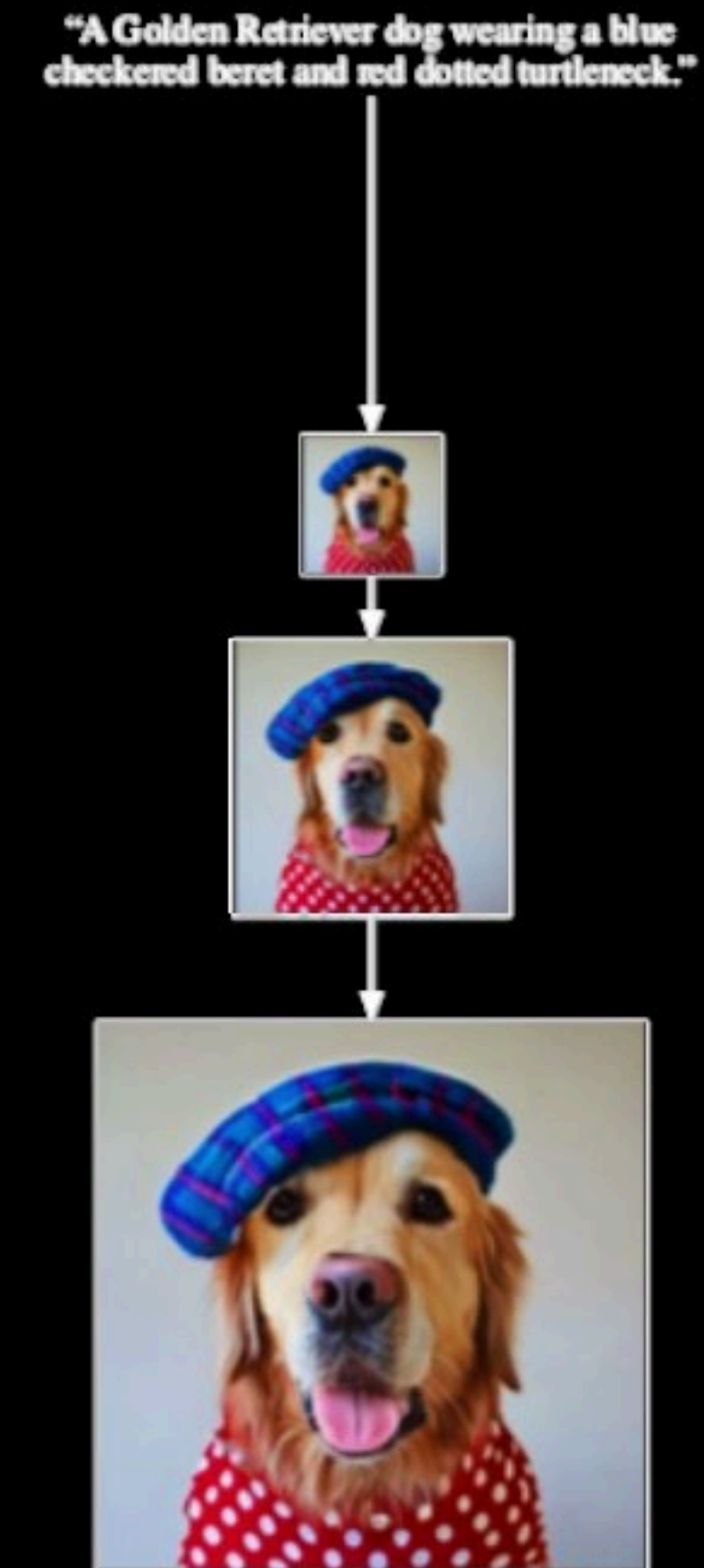
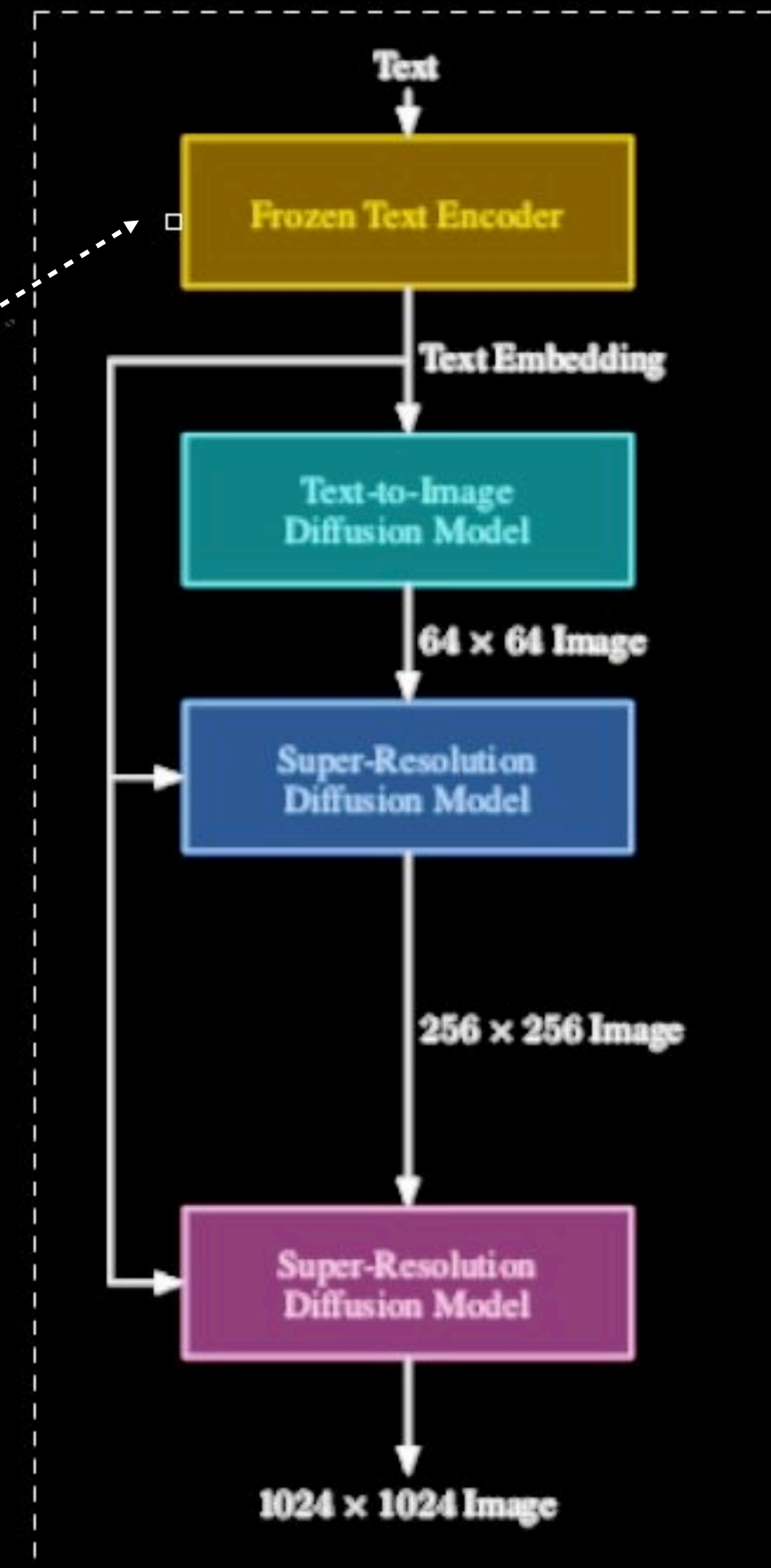


Imagen

- Google Brain (June, 2022)
 - ▶ Text Encoder (Transformer, T5-XXL)
 - maps text to sequence of embeddings



- ▶ Cascade of Conditional Diffusion Models
 - map embeddings to images of increasing resolutions



“That’s All Folks!”

"Stay Tuned for the Next Episodes."

-L.V.