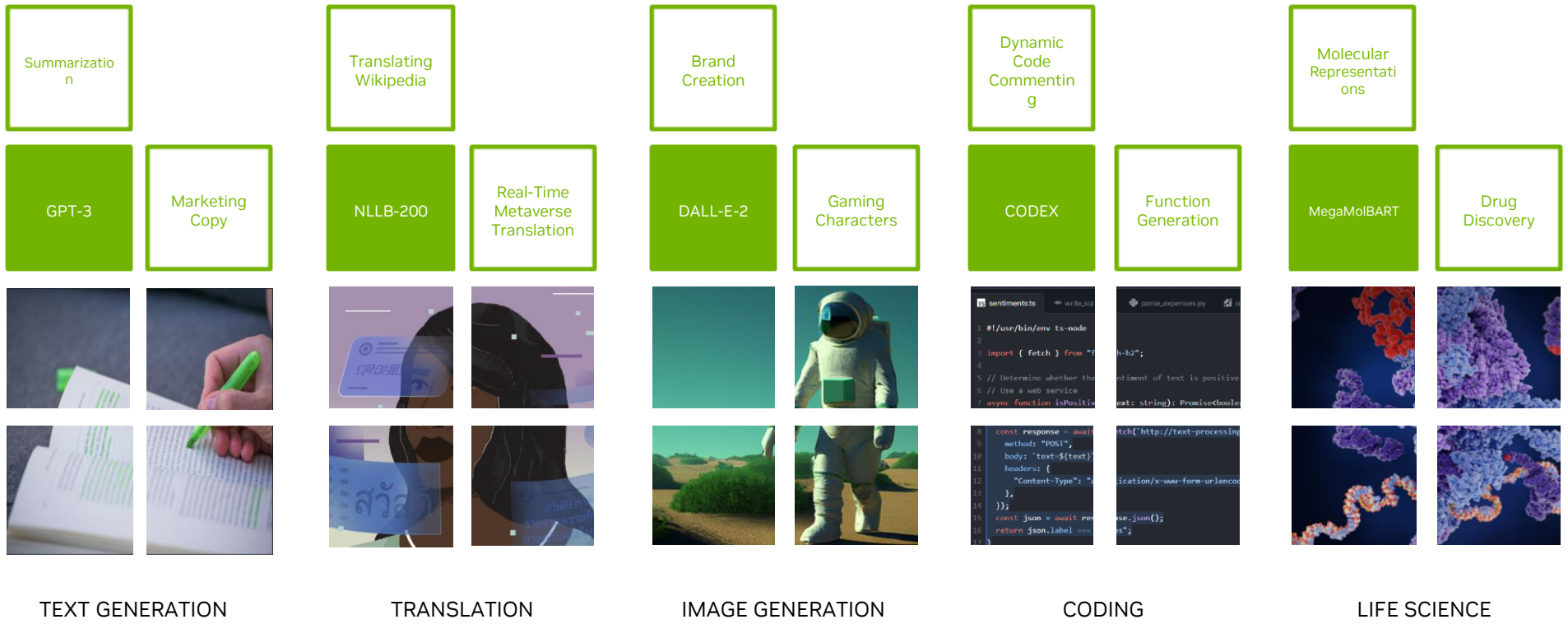
Is Generative AI the path to AGI?

Anima Anandkumar



Generative AI Unlocks New Opportunities

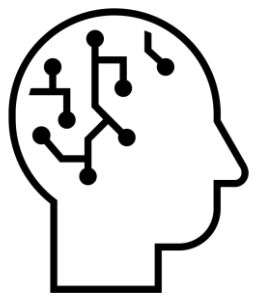
Transcends language and pattern matching

Deep Visual Generative Learning

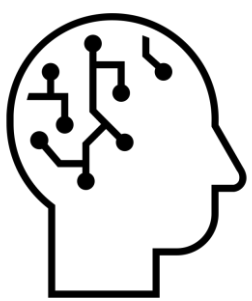
Learning to generate data

Samples from a Data Distribution

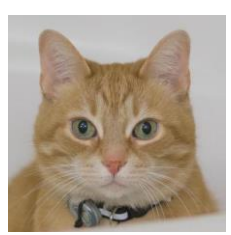
Train



Neural Network



Sample



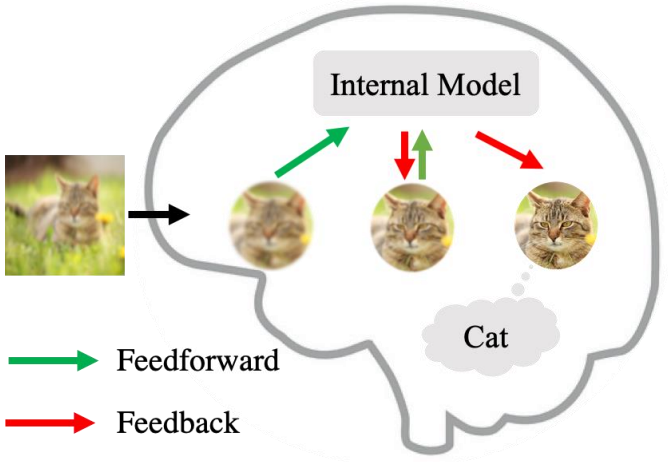
"What I cannot create I do not understand"
Richard Feynman

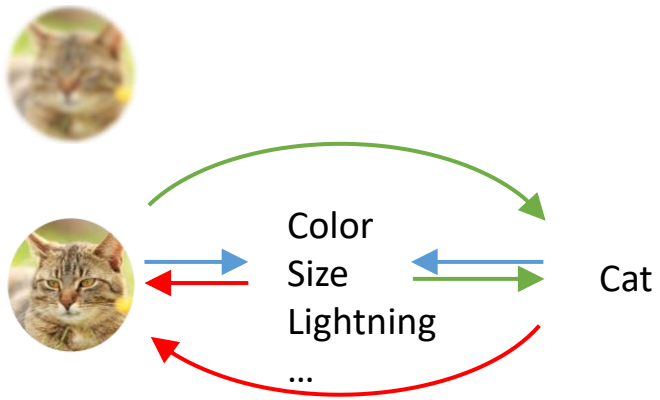
“

Does the brain have a generative model?

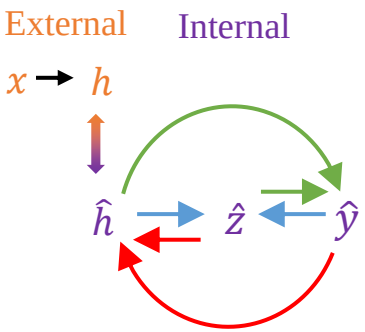
”

BAYESIAN BRAIN THEORY

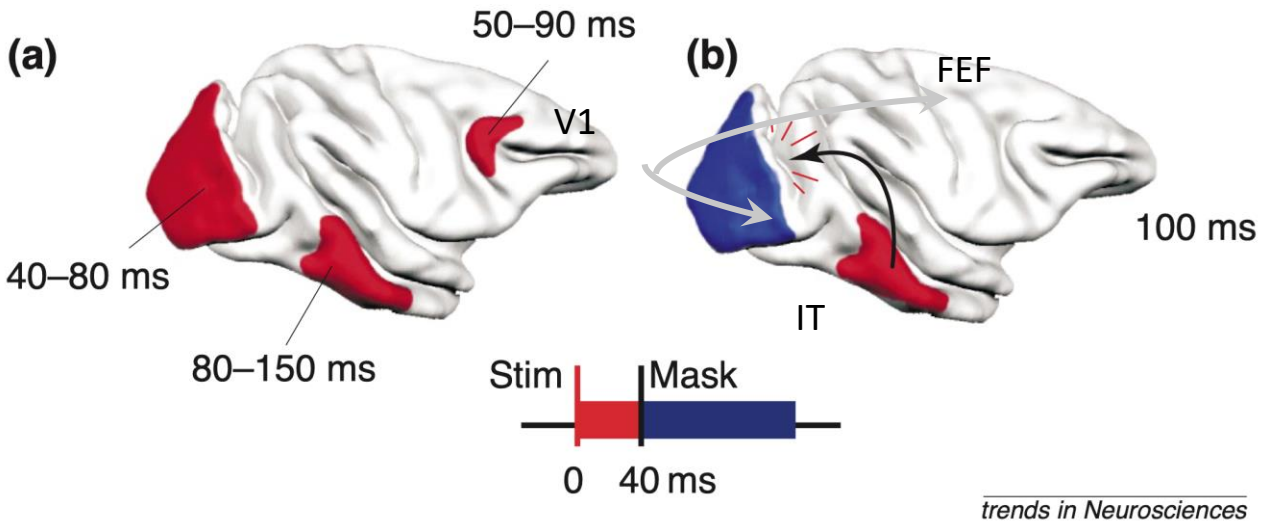




$$\hat{y} = \arg\max_{y} p(y|\hat{h}, \hat{z}),$$
$$\hat{h} = \arg\max_{h} p(h|\hat{y}, \hat{z}),$$
$$\hat{z} = \arg\max_{z} p(z|\hat{h}, \hat{y})$$



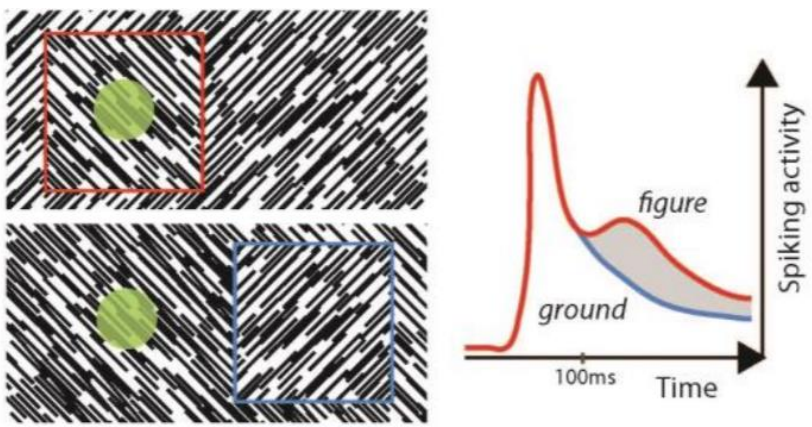
FEEDBACK CONNECTIONS IN THE BRAIN



Experiments indicate recurrent processes lengthen processing time for challenging images.

V.A.F. Lamme et al., The distinct modes of vision offered by feedforward and recurrent processing, Trends in Neurosciences, 2000

SEGMENTATION RELIES ON FEEDBACK IN THE BRAIN

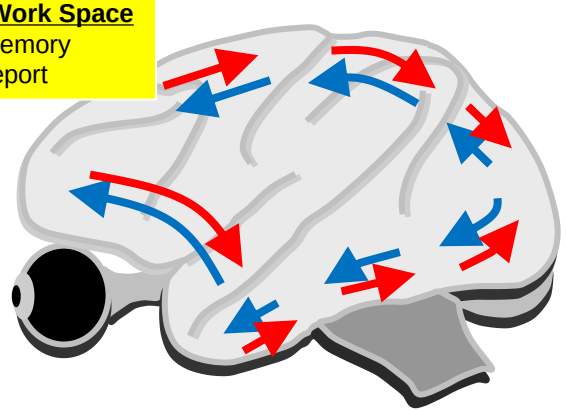


There is a long latency increase in response to a figure that doesn't occur when the same stimulus is part of ground.

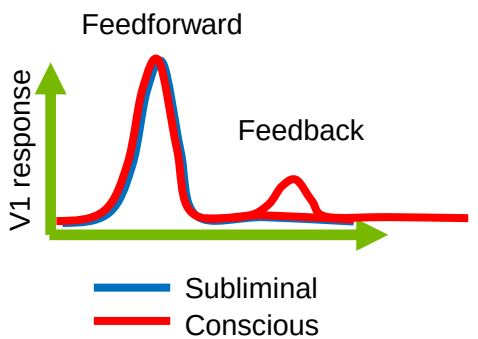
Self and Roelfsema 2016

TOP-DOWN FEEDBACK MAY BE NECESSARY FOR VISUAL CONSCIOUSNESS

Global Neural Work Space
Working memory
Verbal report



Dehaene et al., TICS 2006



Del Cul et al., PLOS Biology 2007

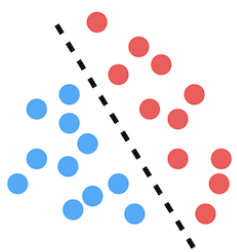
Global neuronal workspace theory: **information only becomes conscious** if it reaches **all the way to the frontal lobe and ignites**, so that it's accessible by the rest of the brain through feedback pathways.

Hallmark of conscious processing is the flexible integration of bottom-up and top-down data streams at the cellular level.


Predictive networks with
recurrent generative feedback

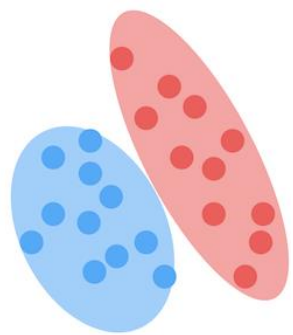
GENERATIVE CLASSIFIER

Logistic Regression



$p(y|x)$

Gaussian mixture model

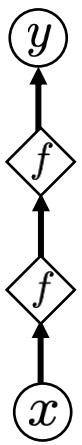


$p(x,y) \rightarrow p(y|x)$

A. Ng, and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In Neurips 2002.

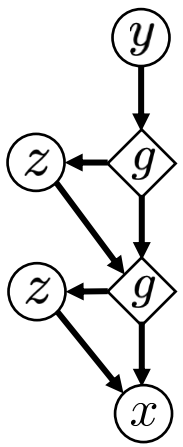
DECONVOLUTIONAL GENERATIVE MODEL (DGM)

CNN



$p(y|x)$

DGM



$p(x, y, z) \rightarrow p(y|x, z)$

$$y \sim p(y)$$

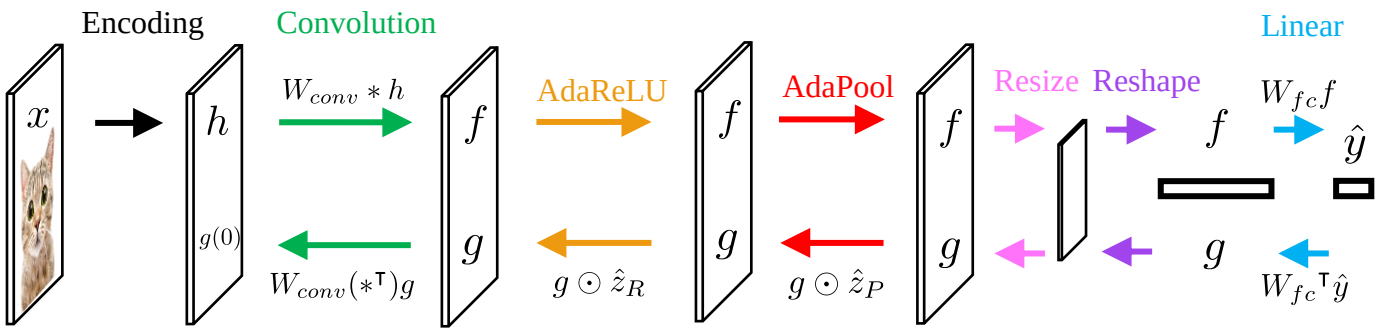
$$z_P^{(i)} \sim \mathrm{Ber}(\frac{e^{b \cdot g^{(i)}}}{e^{b \cdot g^{(i)}} + 1})$$

$$z_R^{(i)} \sim \mathrm{Ber}(\frac{e^{b \cdot g^{(i)}}}{e^{b \cdot g^{(i)}} + 1})$$

$$x \sim \mathcal{N}(g(0), \mathrm{diag}(\sigma^2))$$

T. Nguyen, N. Ho, A. Patel, A. Anandkumar, M. I. Jordan, and R. G. Baraniuk. A bayesian perspective of convolutional neural networks through a deconvolutional generative model. arXiv:1811.02657, 2018.

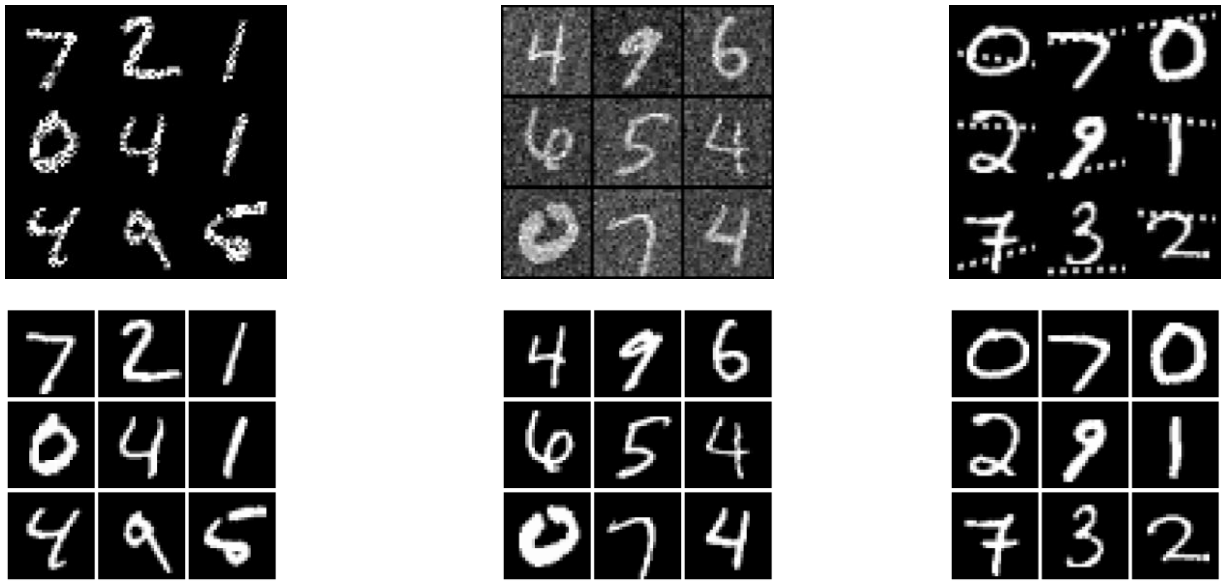
OPERATIONS IN CNN-F



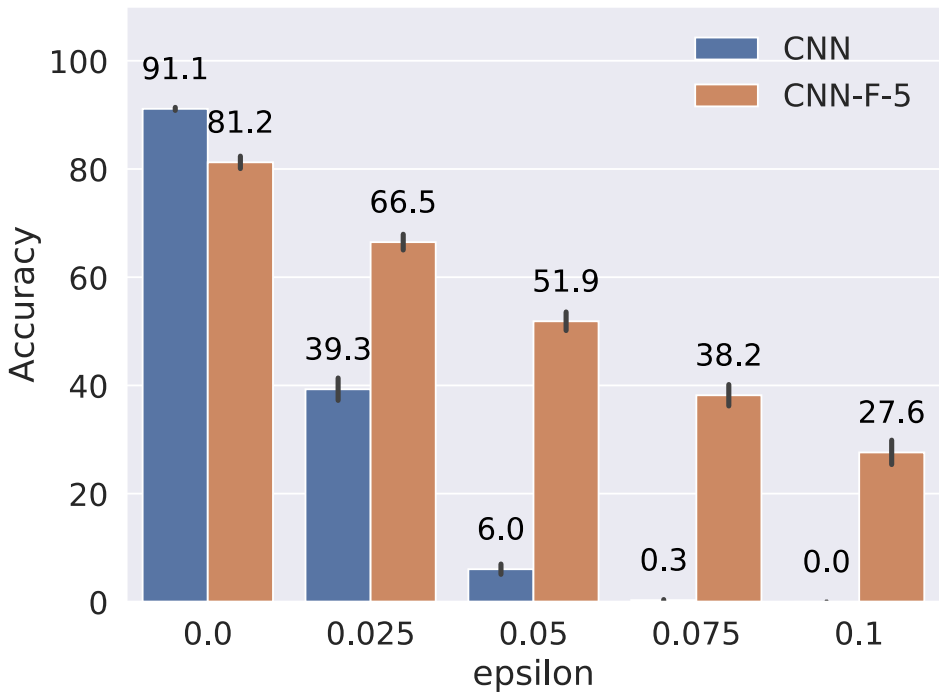
$$\sigma_{\text{AdaReLU}}(f) = \begin{cases} \sigma_{\text{ReLU}}(f), & \text{if } g \geq 0 \\ \sigma_{\text{ReLU}}(-f), & \text{if } g < 0 \end{cases}$$

$$\sigma_{\text{AdaPool}}(f) = \begin{cases} \sigma_{\text{MaxPool}}(f), & \text{if } g \geq 0 \\ -\sigma_{\text{MaxPool}}(-f), & \text{if } g < 0 \end{cases}$$

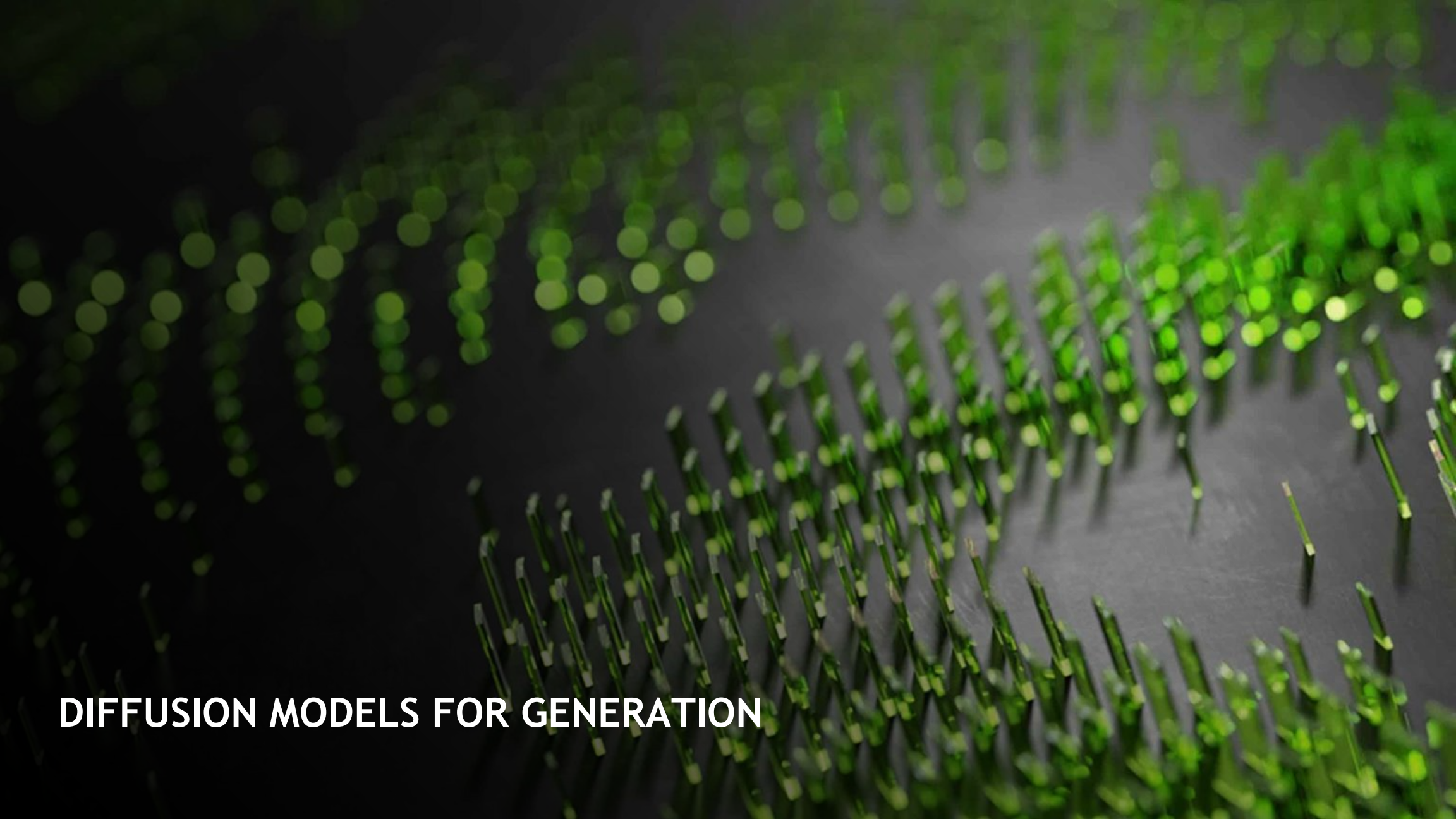
CNN-F REPAIRS DISTORTED IMAGES



CNN-F IMPROVES ADVERSARIAL ROBUSTNESS



- Standard training on Fashion-MNIST.
- Attack with PGD-40.
- CNN-F has higher adversarial robustness than CNN.



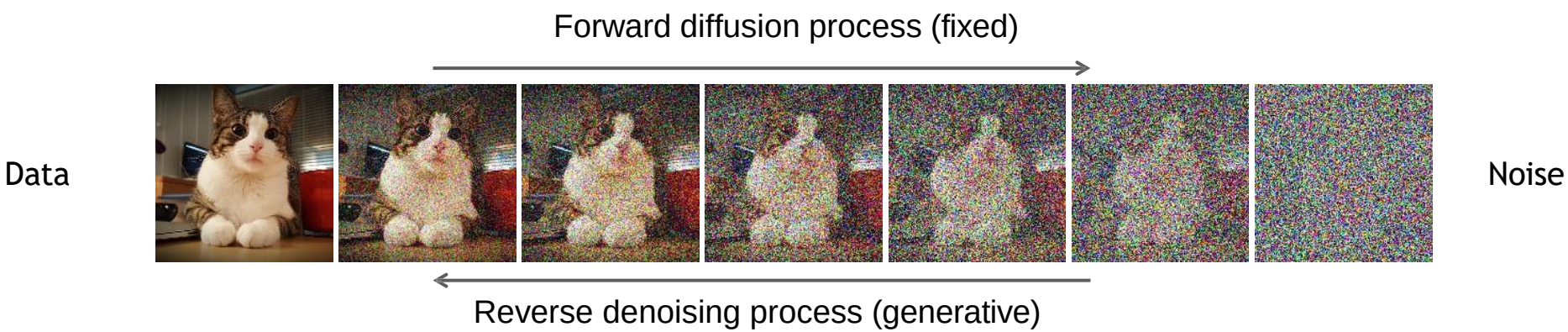
DIFFUSION MODELS FOR GENERATION

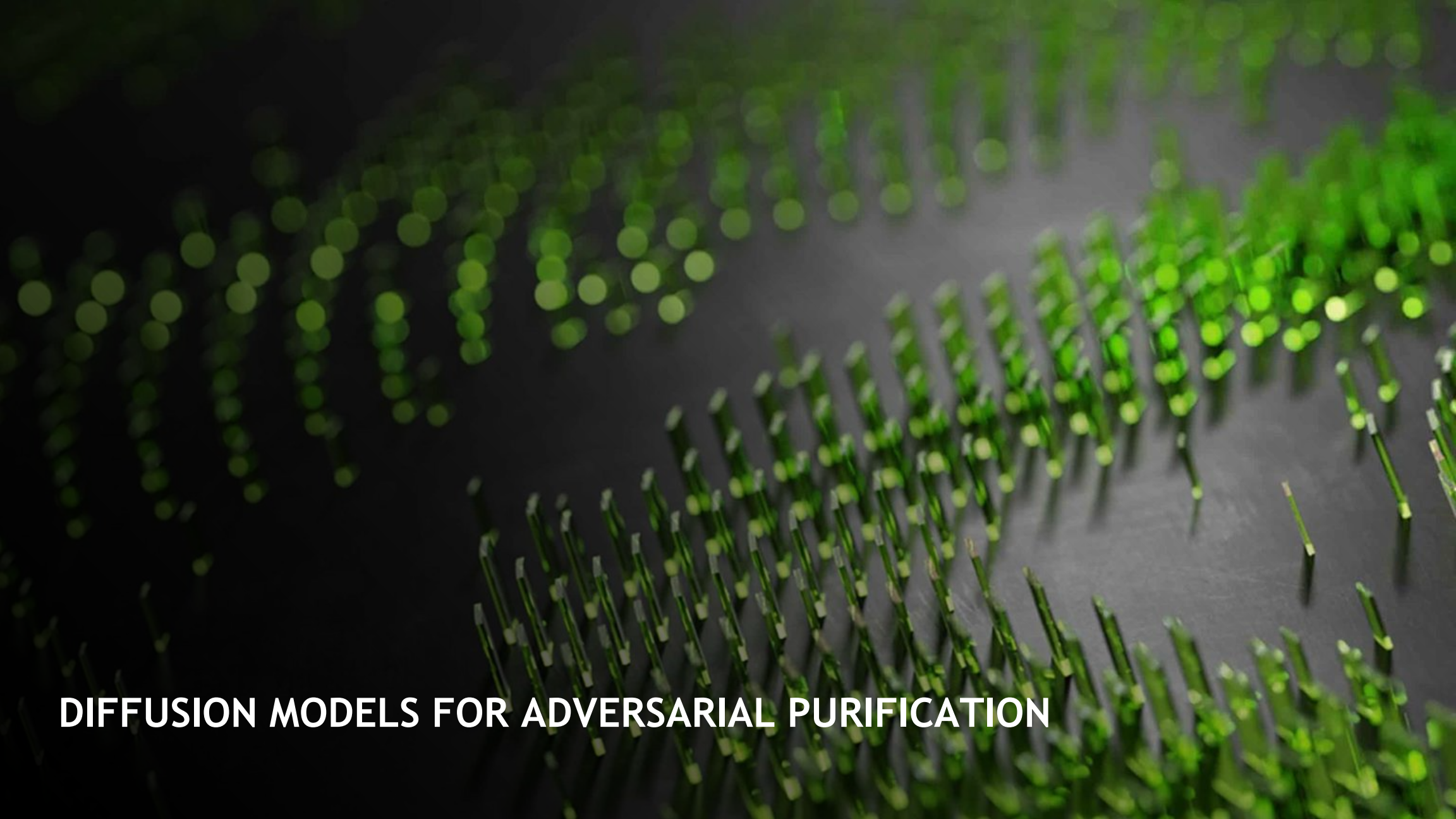
DENOISING DIFFUSION MODELS

Learning to generate by denoising

Denoising diffusion models consist of two processes:

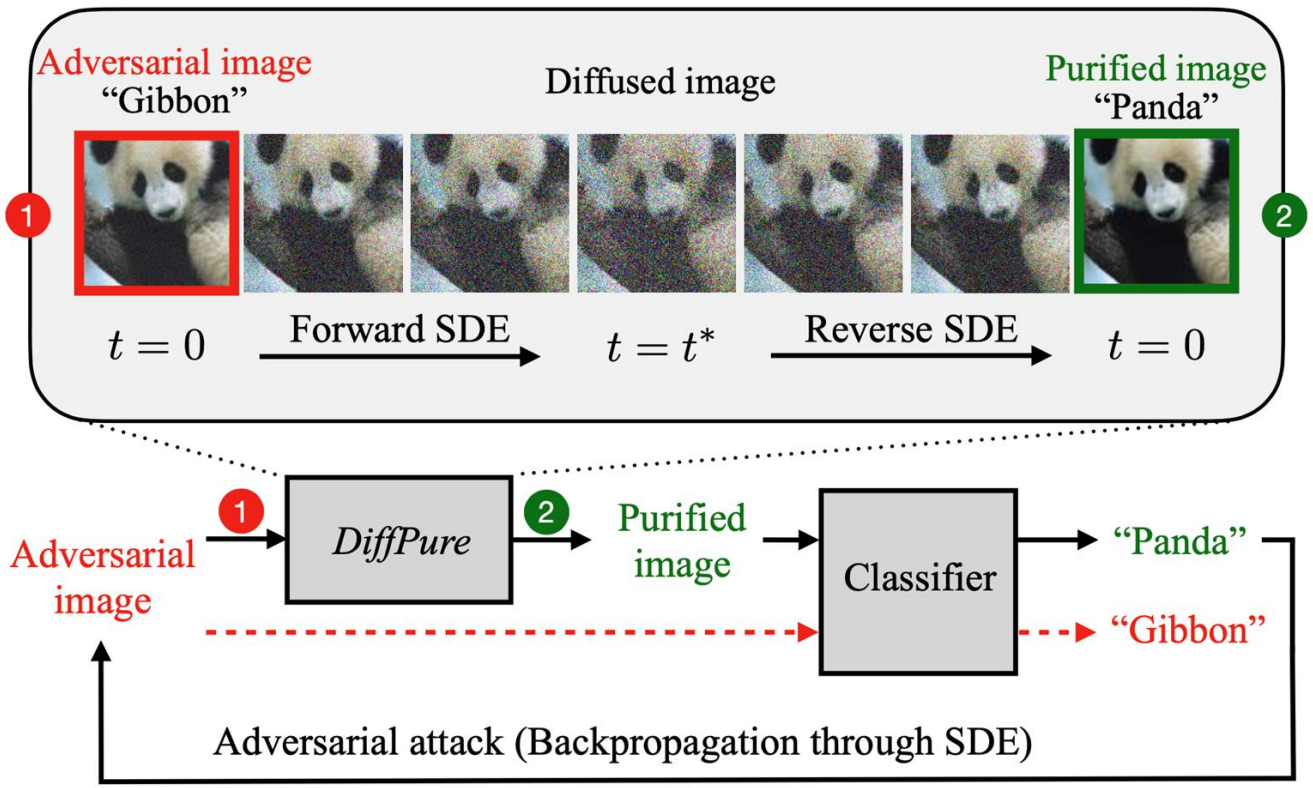
- Forward diffusion process that gradually adds noise to input
- Reverse denoising process that learns to generate data by denoising




DIFFUSION MODELS FOR ADVERSARIAL PURIFICATION

DiffPure : DIFFUSION PURIFICATION

Basic idea: It uses the forward and reverse processes of pre-trained diffusion models to purify adversarial images



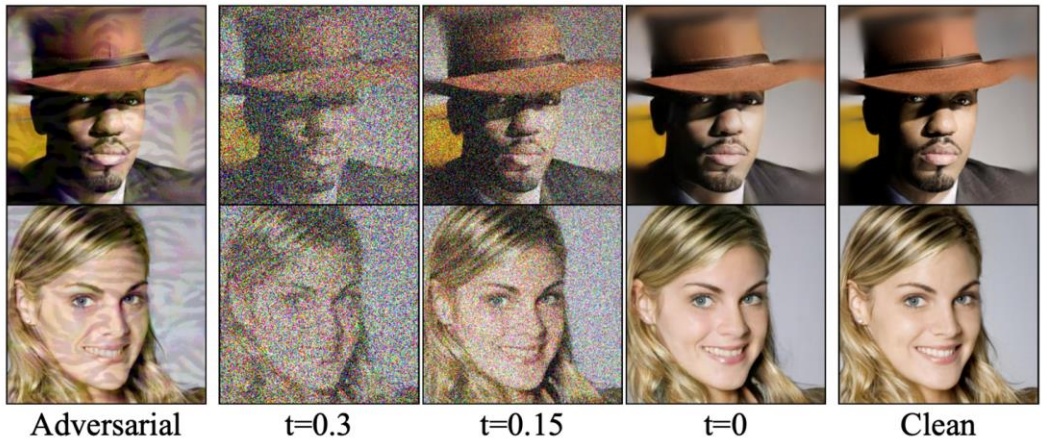
MOTIVATION

- Two classes of methods defending against adversarial attacks
  - *Adversarial training*
    - *It trains neural networks on adversarial examples*
  - *Adversarial purification*
    - *It uses generative models to purify adversarial perturbations*

	Adversarial training	Adversarial purification	DiffPure (Ours)
Defense against seen threats	Y	X	Y
Defense against unseen threats	X	Y	Y
Training complexity	X	Y	Y

QUALITATIVE RESULTS OF DIFFPURE ON CELEBA-HQ

DiffPure removes adversarial perturbations and recover clean images



* Middle three columns show the results of DiffPure at different timesteps

COMPARISON WITH SOTA IN ROBUSTBENCH BENCHMARK

DiffPure outperforms other methods on ImageNet by a large margin

Method	Extra Data	Standard Acc	Robust Acc
ResNet-50			
(Engstrom et al., 2019)	✗	62.56	31.06
(Wong et al., 2020)	✗	55.62	26.95
(Salman et al., 2020)	✗	64.02	37.89
(Bai et al., 2021)†	✗	67.38	35.51
Ours	✗	**67.79±0.43**	**40.93±1.96**
WideResNet-50-2			
(Salman et al., 2020)	✗	68.46	39.25
Ours	✗	**71.16±0.75**	**44.39±0.95**
DeiT-S			
(Bai et al., 2021)†	✗	66.50	35.50
Ours	✗	**73.63±0.62**	**43.18±1.27**

DEFENSE AGAINST UNSEEN THREATS

DiffPure can defend against unseen attacks while baseline performances drop significantly against unseen attacks

Method	Standard Acc	Robust Acc		
		ℓ_∞	ℓ_2	StAdv
Adv. Training with ℓ_∞ (Laidlaw et al., 2021)	86.8	49.0	19.2	4.8
Adv. Training with ℓ_2 (Laidlaw et al., 2021)	85.0	39.5	47.8	7.8
Adv. Training with StAdv (Laidlaw et al., 2021)	86.2	0.1	0.2	53.9
PAT-self (Laidlaw et al., 2021)	82.4	30.2	34.9	46.4
ADV. CRAIG (Dolatabadi et al., 2021)	83.2	40.0	33.9	49.6
ADV. GRADMATCH (Dolatabadi et al., 2021)	83.1	39.2	34.1	48.9
Ours	**88.2±0.8**	**70.0±1.2**	**70.9±0.6**	**55.0±0.7**

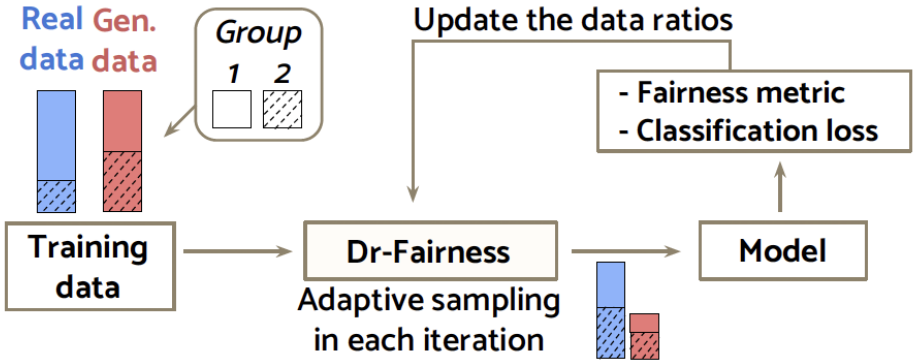
IMPROVING FAIRNESS THROUGH SYNTHETIC DATA GENERATION

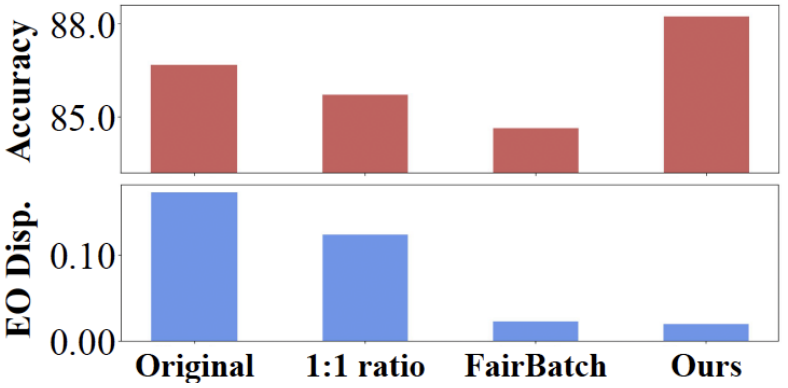
Anchor, gender = female, skin color = medium, age = middle

Astronaut, gender = female, skin color = light, age = adult

Captain, gender = male, skin color = dark, age = middle

Tennis player, gender = male, skin color = light, age = child






EMBODIED GENERALIST AGENTS

BUILDING AI FOR GENERALIST AGENTS

household Robots

virtual agents

self-driving cars

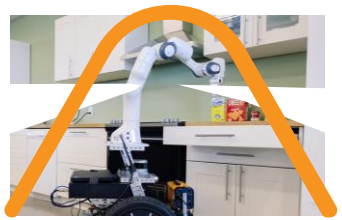
Open-Ended Objectives

Massively Multitask

World Knowledge

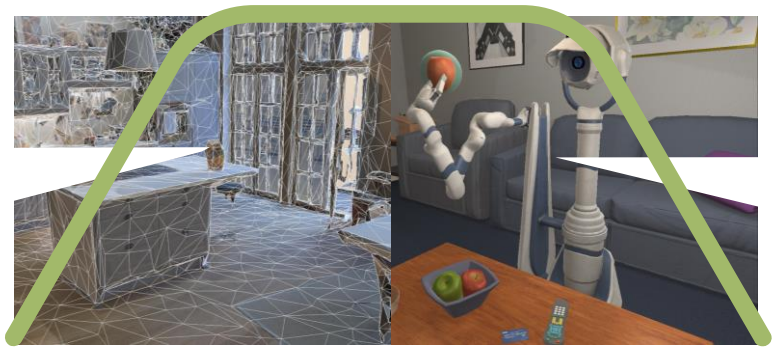
THE DATA PYRAMID FOR GENERALIST AGENTS

Real-World Data



- Small scale and expensive to collect
- Ease of use for agent learning, direct transfer

Synthetic Data



- Unlimited simulated data (in theory)
- Content creation challenge, reality gap, computational burden

Web Data

YouTube

reddit

Common Crawl

WIKIPEDIA
The Free Encyclopedia

- Massive scale and ever-growing
- Multimodal and unstructured
- Human-centered data

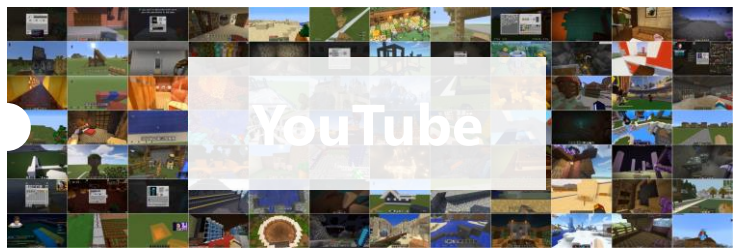

MINEDOJO YouTube
730K videos
300K hours
2.2B words

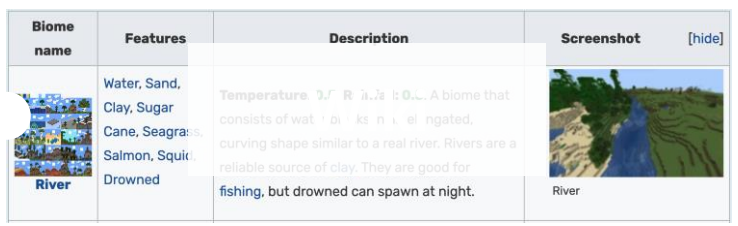
MINEDOJO: FRAMEWORK FOR BUILDING GENERALIST AGENTS

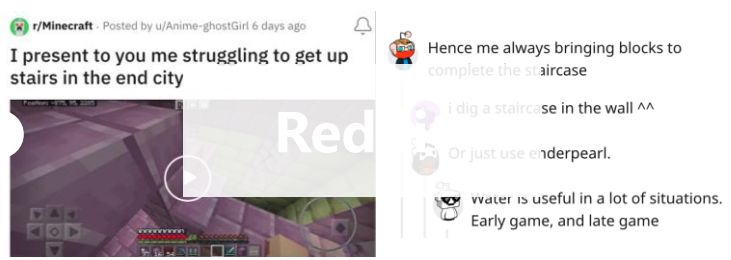
Open-ended Environment

Generalist Agent

Internet-scale Knowledge Base

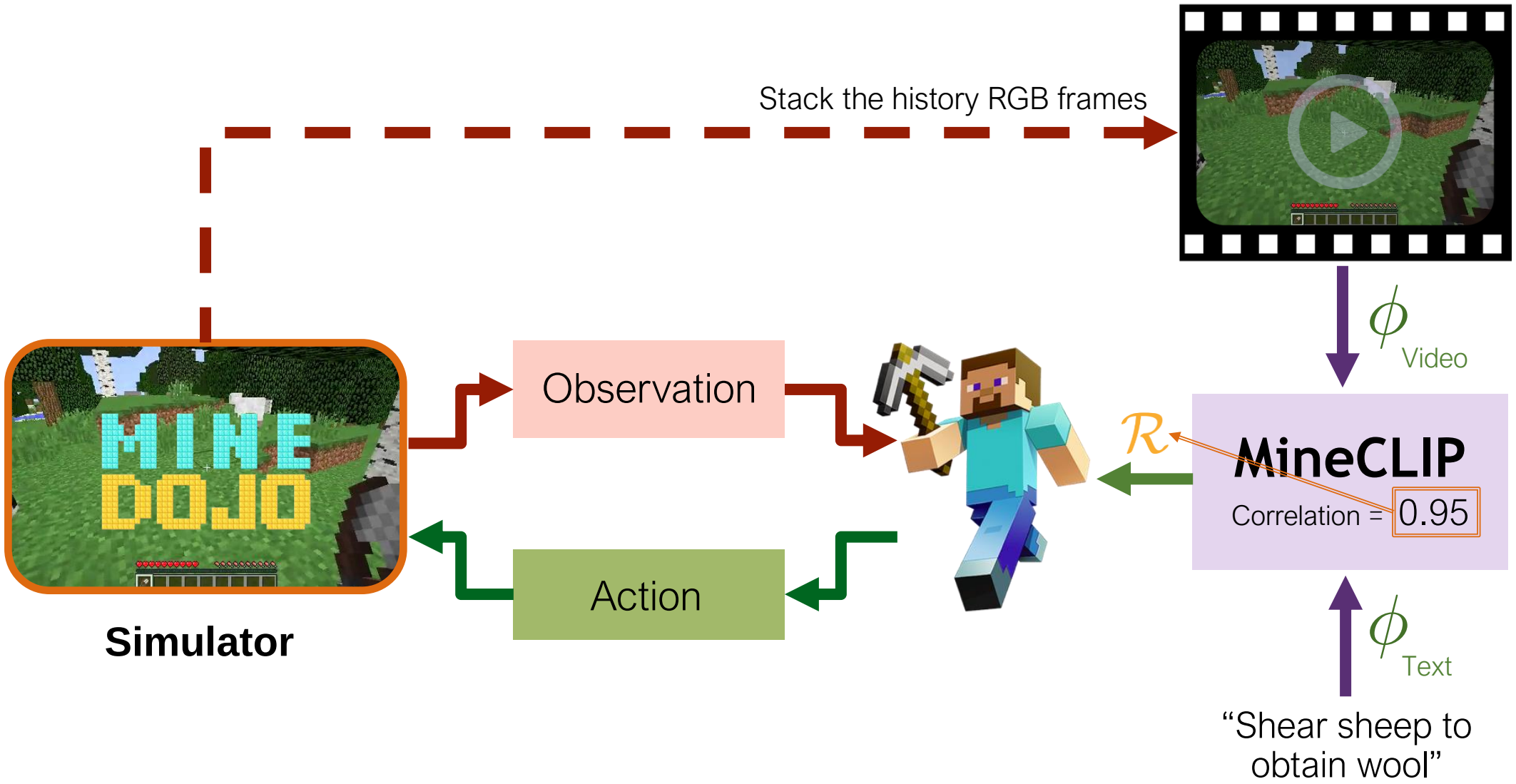






NVIDIA.

MineDojo: Framework for Building Generalist Agents




Milk a co


Hunt a co


Shear a shee


Hunt a shee


Combat a spide


Combat a zombi


Combat a pigman


Combat an enderman

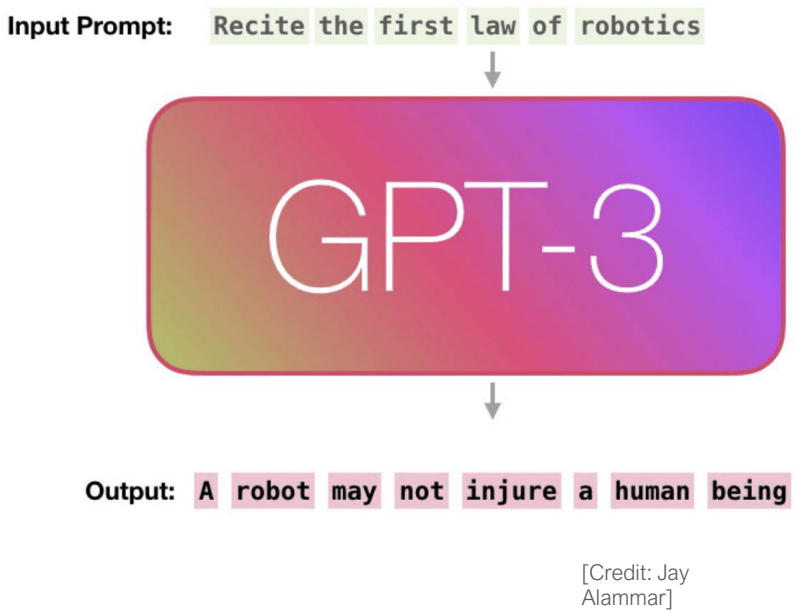

Find a Nether Porta


Find an ocea


Dig a hol


Lay carpe

VIMA: General Robot Manipulation with Multimodal Prompts



[Credit: Jay Alammar]

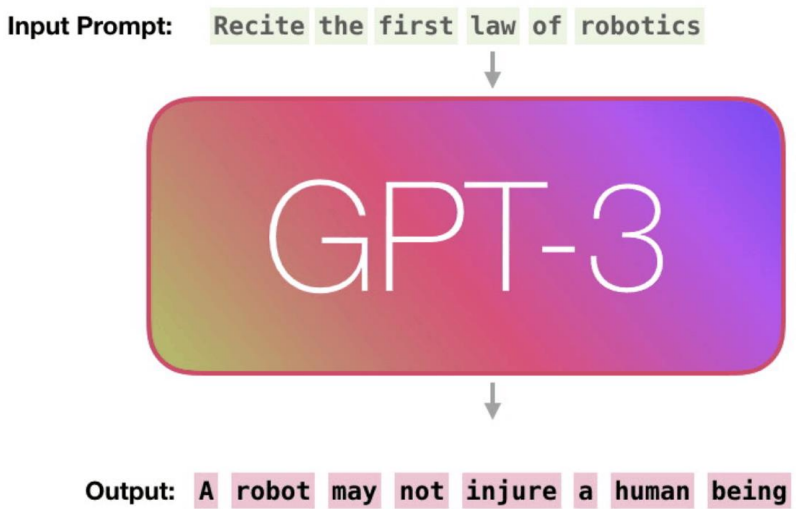
Input Prompt:

How to "find material and craft a gold pickaxe" in Minecraft? Let's think step by step.

GPT-3:

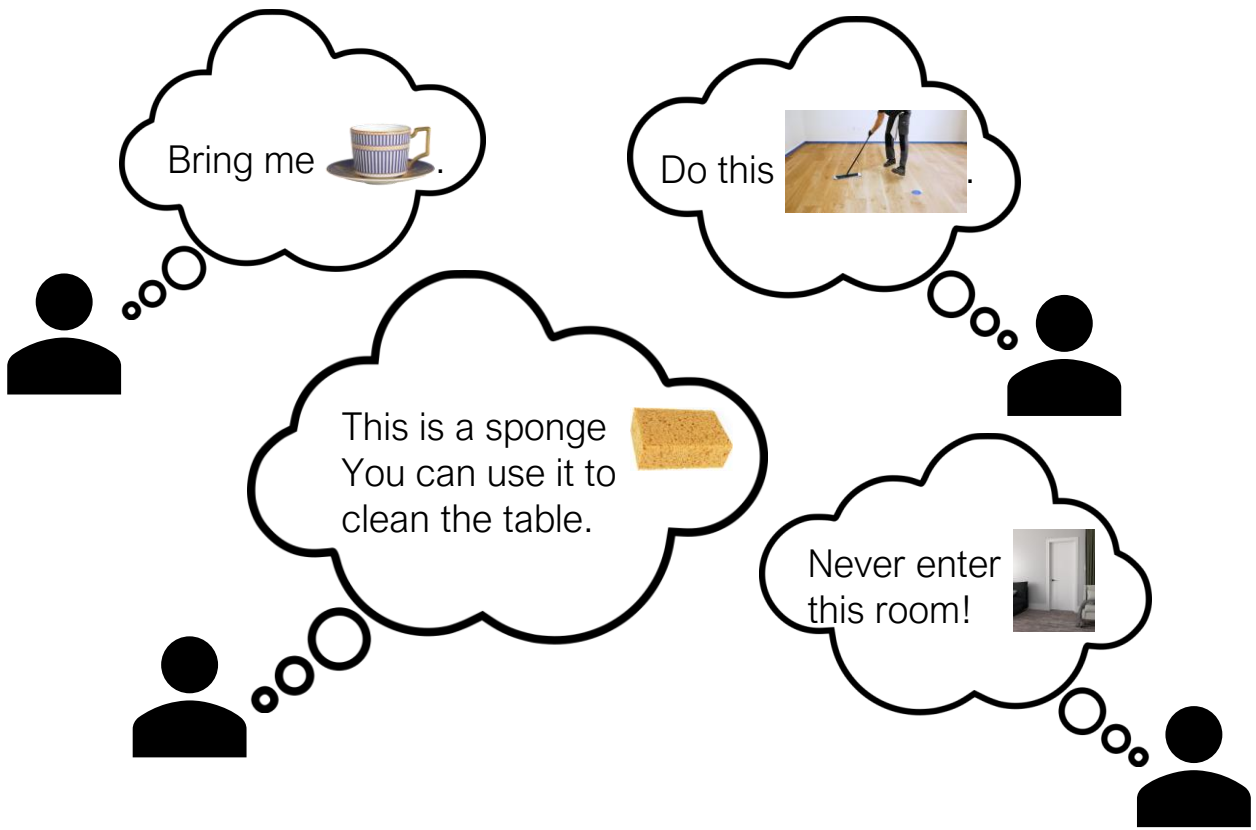
1) Find a place with a lot of trees;
2) Cut down the trees and gather the wood;
3) Find a place with a lot of stone;
4) Mine the stone and gather the cobblestone;
5) Find a place with a lot of iron;
6) Mine the iron and gather the iron ingots;
7) Find a place with a lot of gold;
8) Mine the gold and gather the gold ingots;
9) Craft a gold pickaxe.

VIMA: General Robot Manipulation with Multimodal Prompts

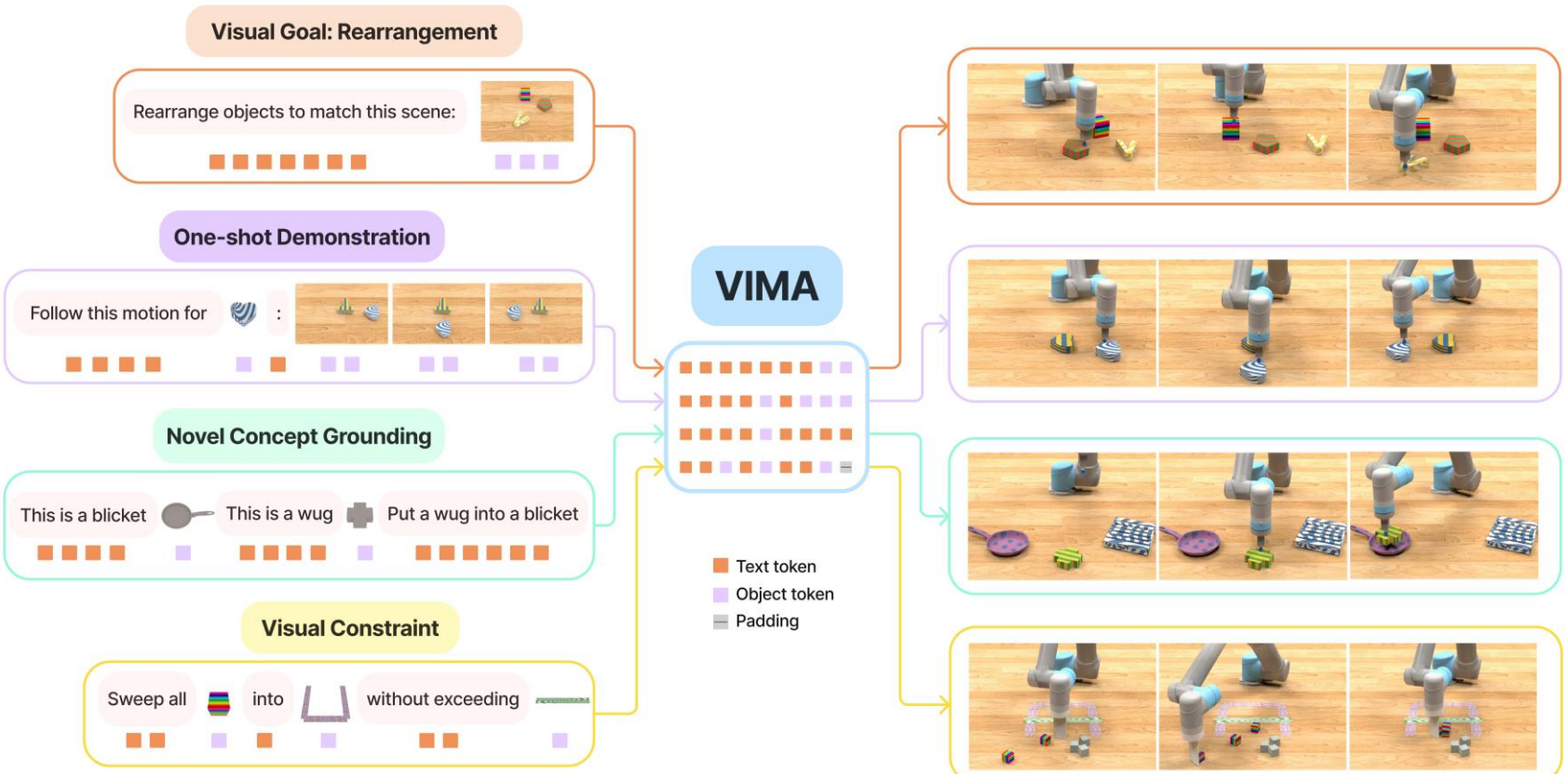


[Credit: Jay Alammar]

What if we can prompt a household robot to …



VIMA: General Robot Manipulation with Multimodal Prompts



generalist robot agent for multi-task learning and zero-shot generalization

"VIMA: General Robot Manipulation with Multimodal Prompts." Jiang, Yunfan et al. ArXiv 2022

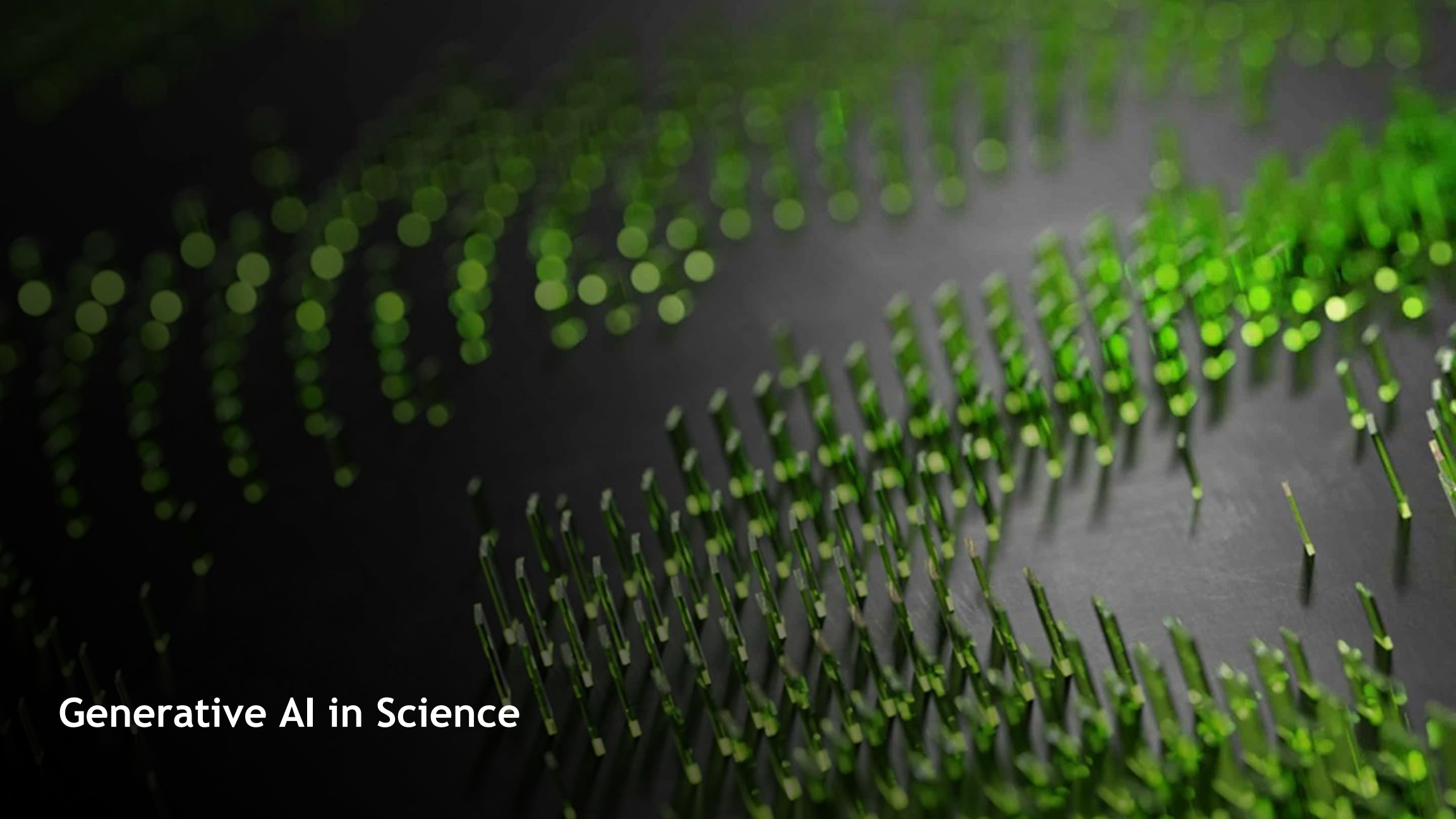
VIMA: General Robot Manipulation with Multimodal Prompts

VIMA: **V**isuo-**M**otor **A**ttention model

- Transformer encoder-decoder;
- Encode multimodal prompts with a frozen LM (Google T5);
- Decode robot arm actions given the prompt and interaction history.

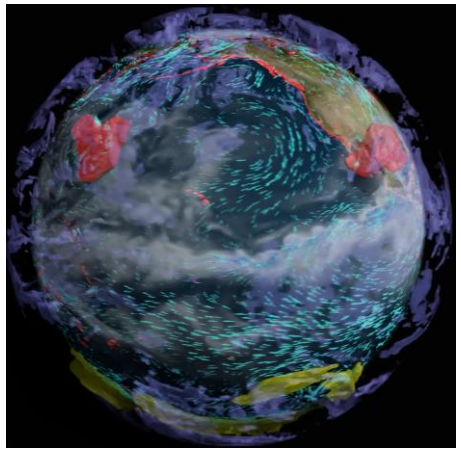
VIMA

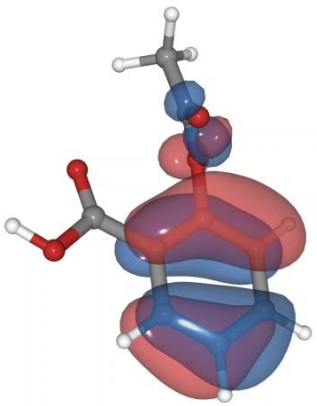
"VIMA: General Robot Manipulation with Multimodal Prompts." Jiang, Yunfan et al. ArXiv 2022

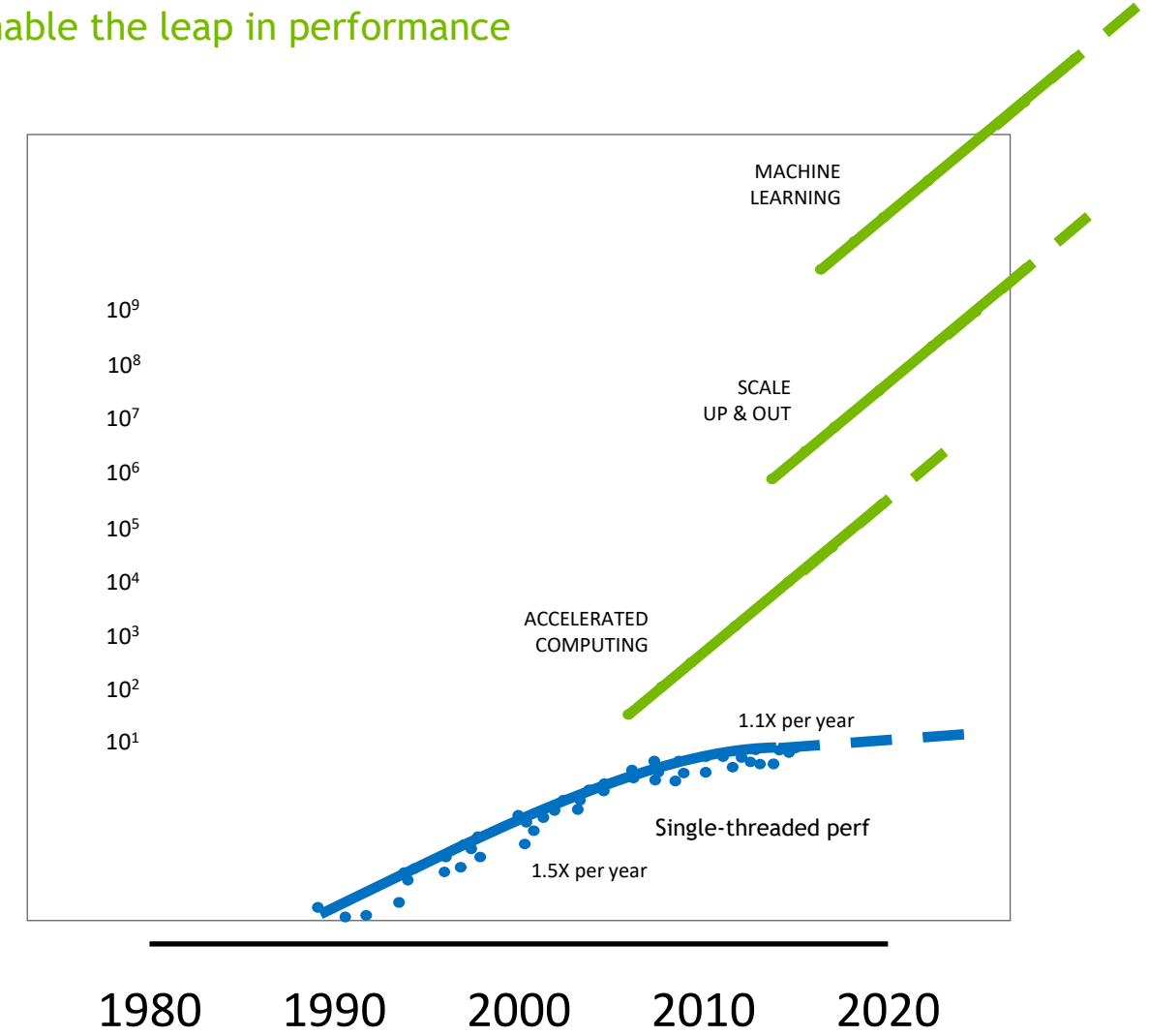

Generative AI in Science

MILLION-X LEAP IN SCIENTIFIC COMPUTING

AI/ML to enable the leap in performance





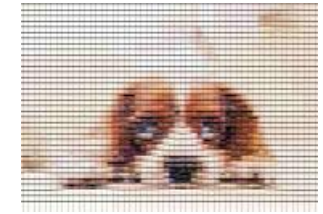


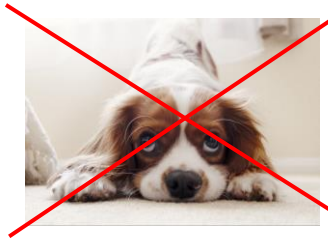
DISCRETIZATION-INVARIANT LEARNING

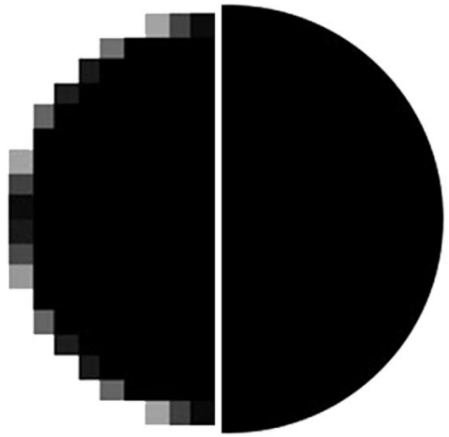
One AI model for any discretization: no re-training

Neural Network

Input and output at fixed resolution

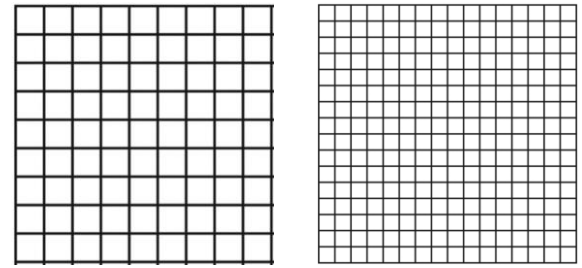


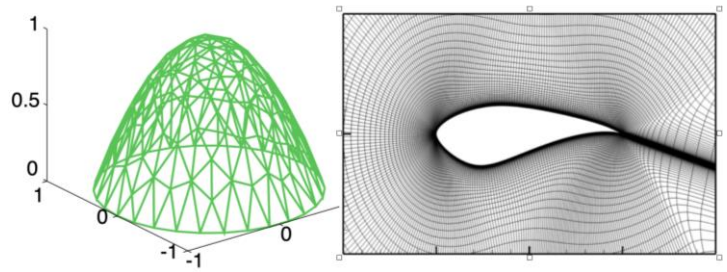




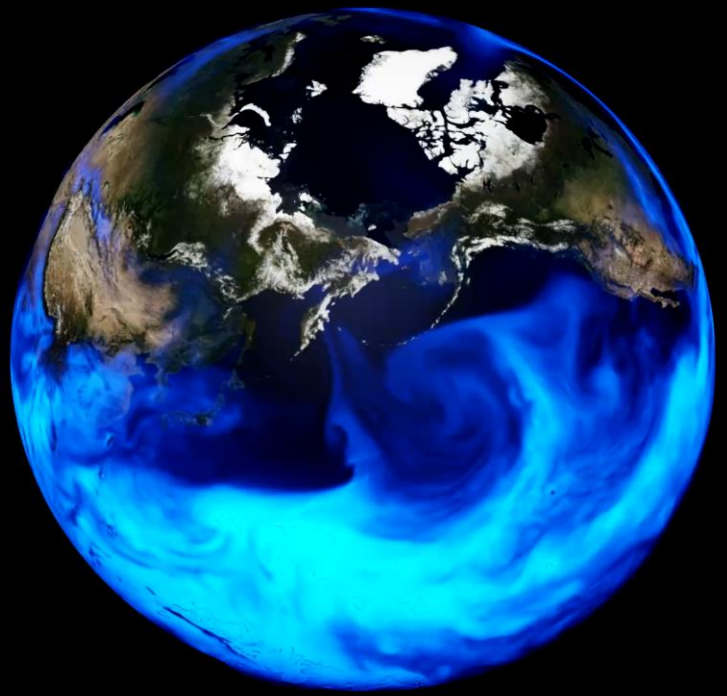
Neural Operator

Input and output at any points in domain

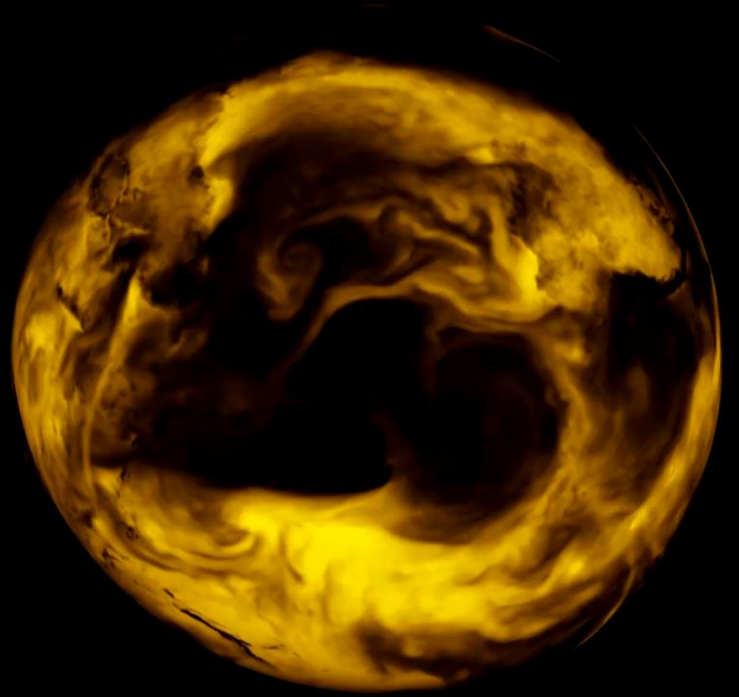




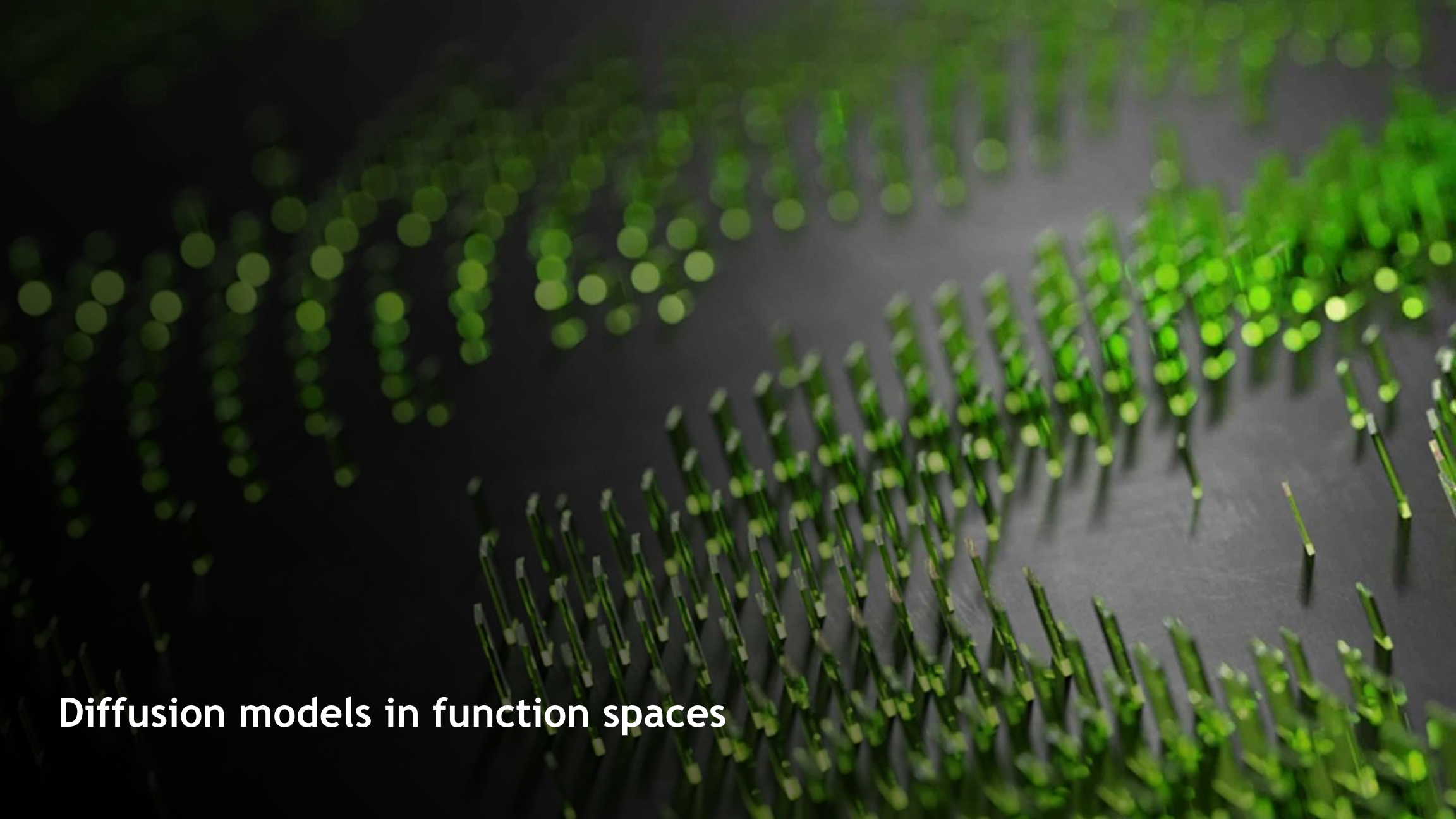
Ground Truth



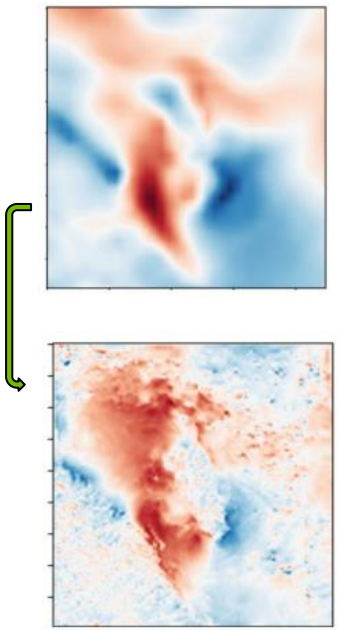
FourCastNet



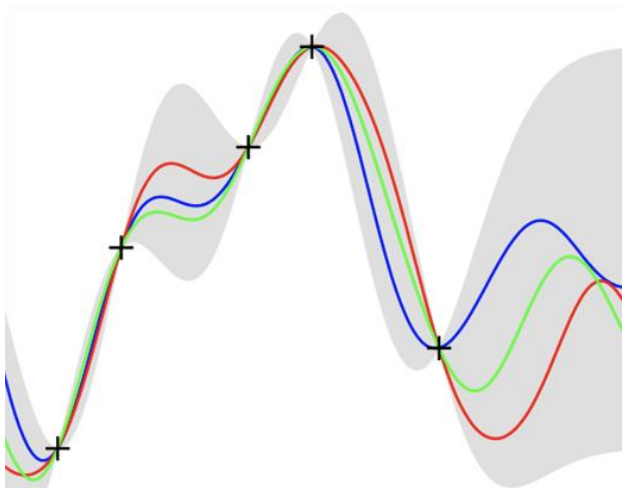
Our AI (FourCastNet) is **45,000 times** faster than current weather models


Diffusion models in function spaces

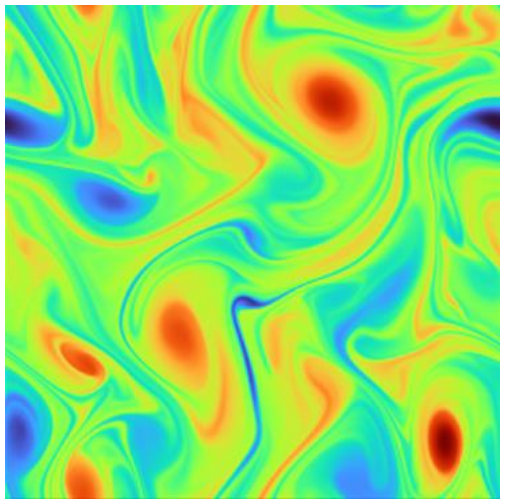
SCIENTIFIC COMPUTING REQUIRES PROBABILISTIC MODELING



INVERSE PROBLEMS

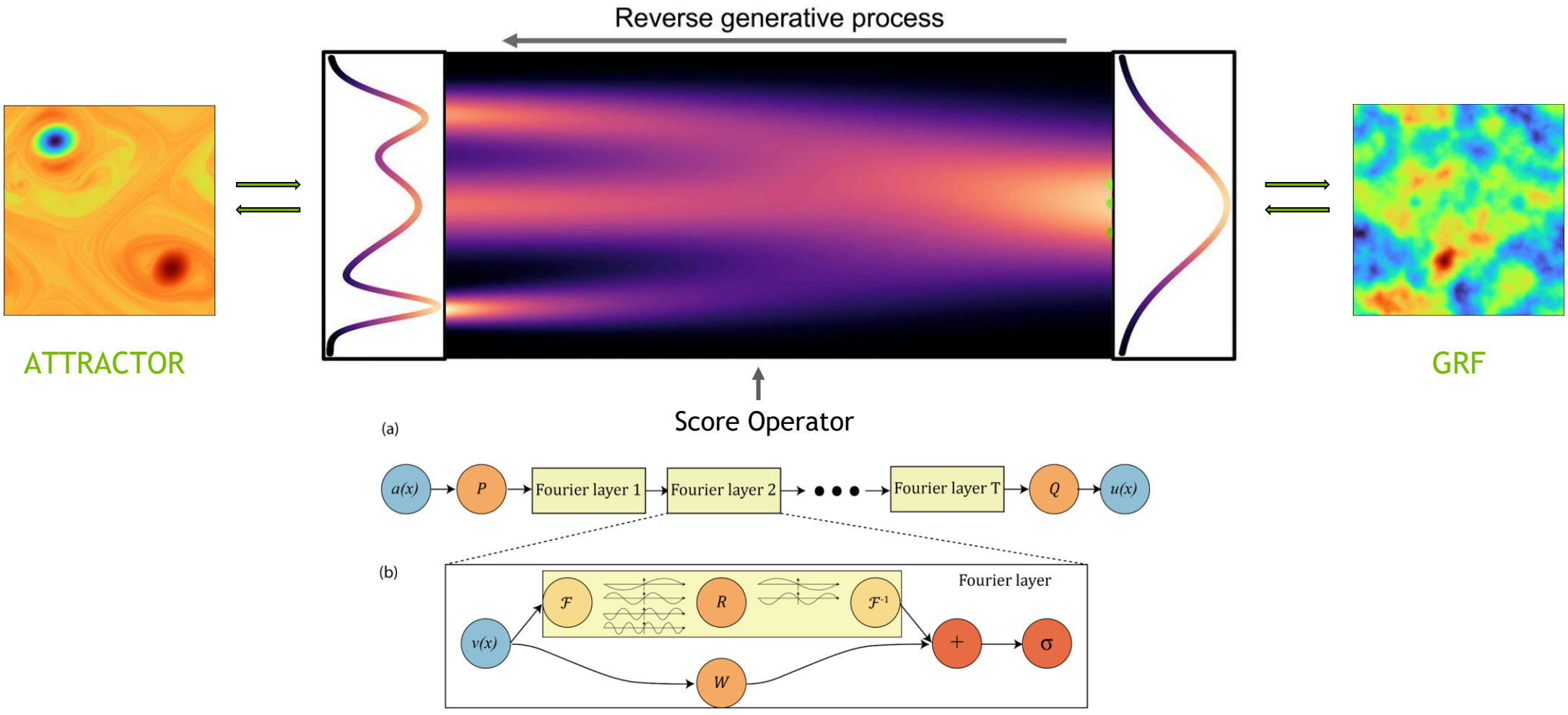


UNCERTAINTY QUANTIFICATION

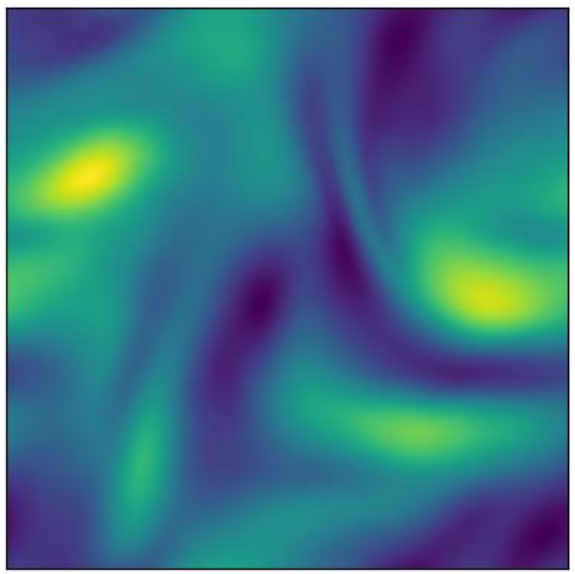


CHAOTIC DYNAMICS

GENERATION BY FUNCTION SPACE SDE

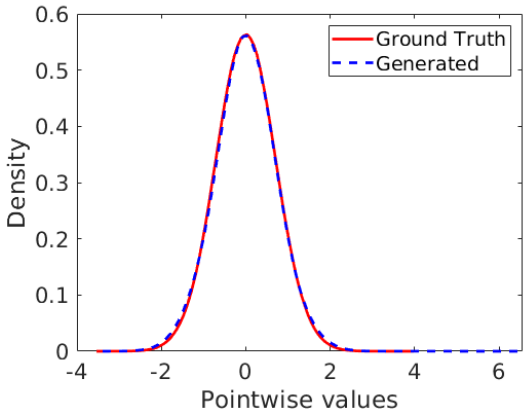


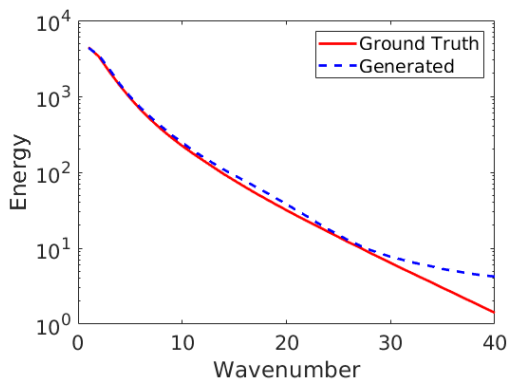
ZERO-SHOT SUPER RESOLUTION SAMPLING

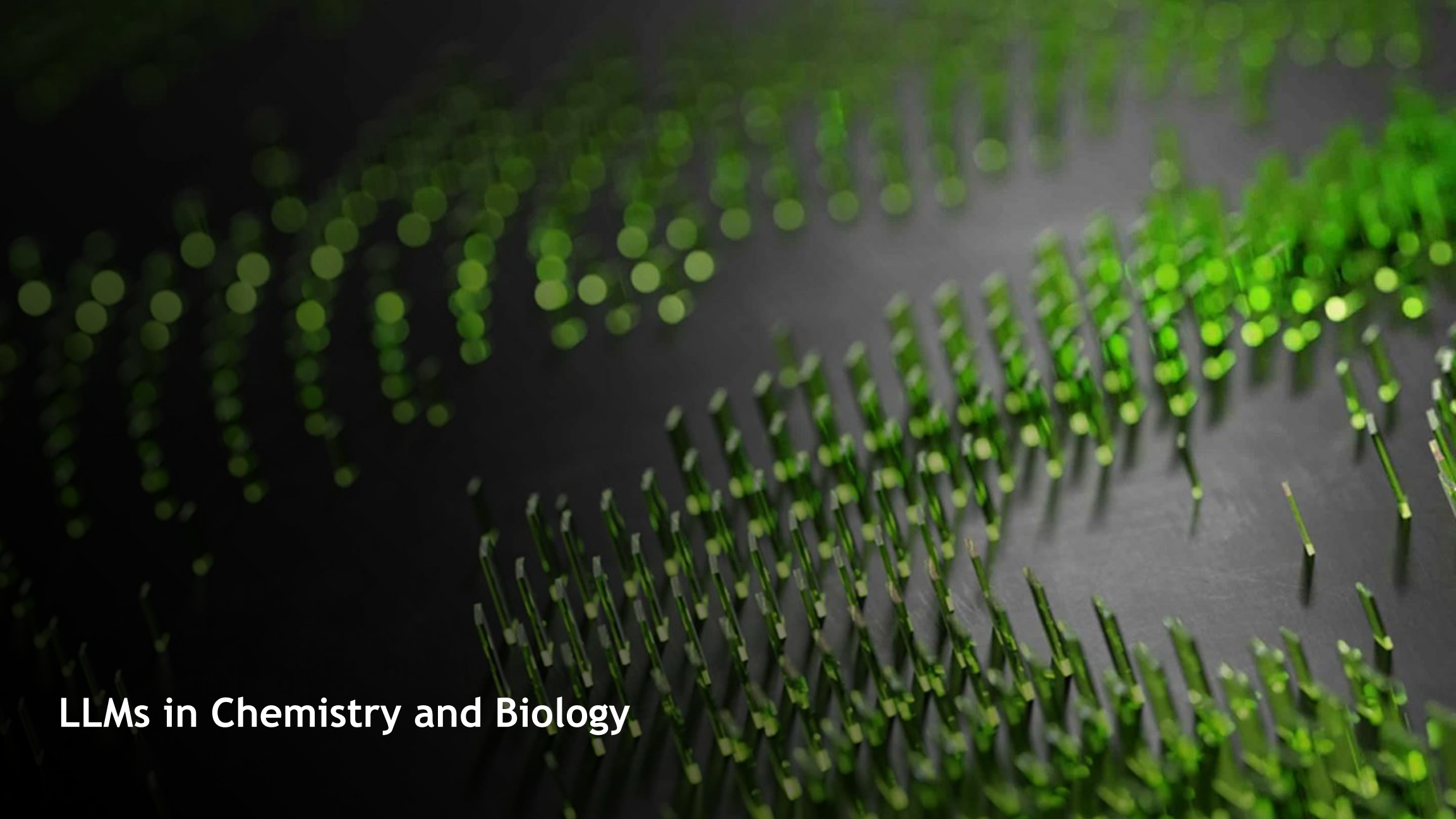


GENERATION: 1024x1024
TRAINING: 128x128

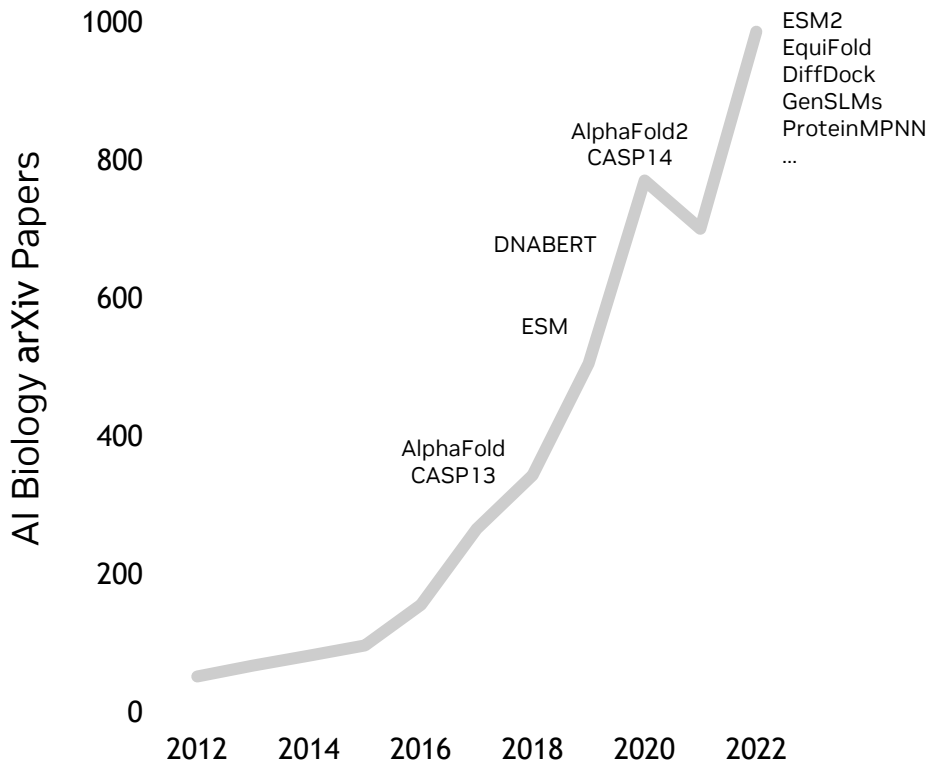
HIGH-RES STATISTICS






LLMs in Chemistry and Biology

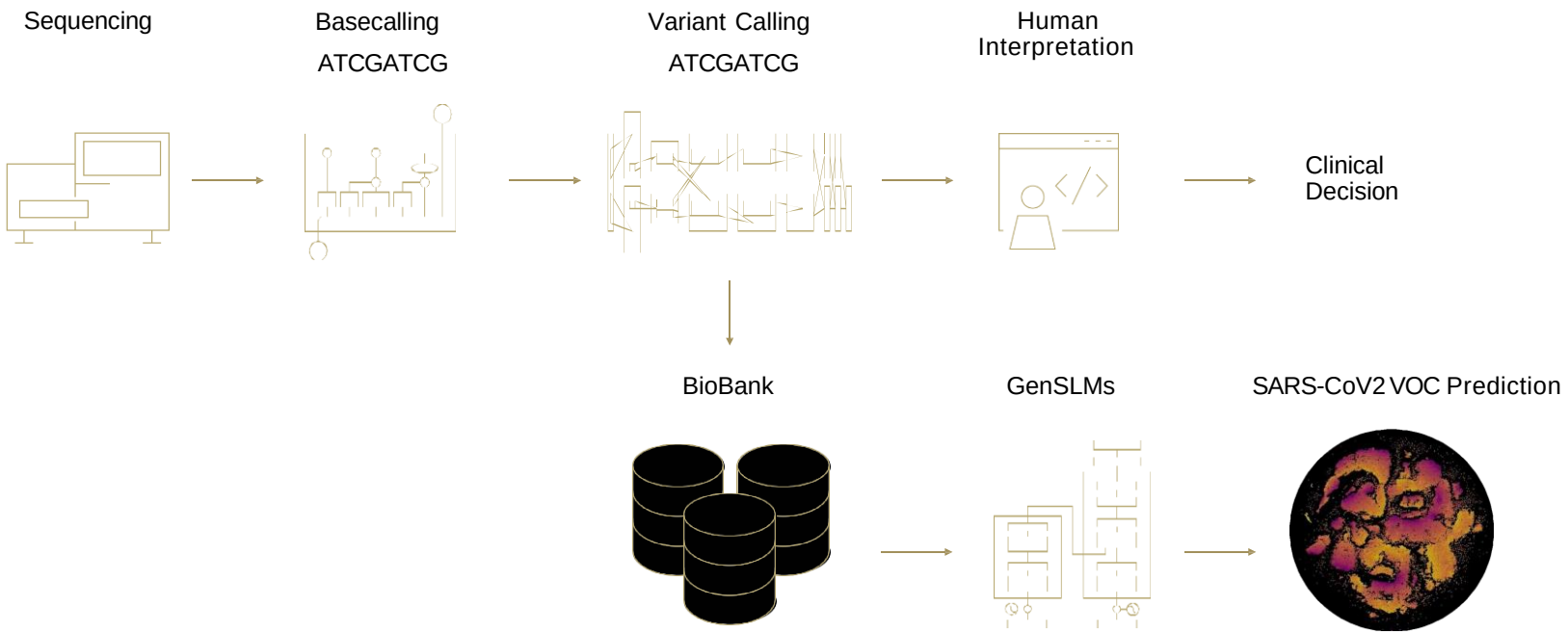
GENERATIVE AI AND LLMS TURN BIOLOGY INTO ENGINEERING

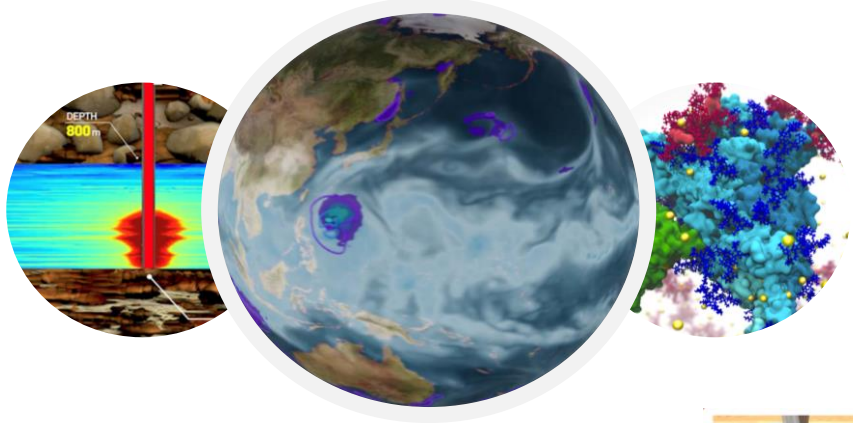


"By 2025, more than 30% of new drugs and materials will be systematically discovered using generative AI techniques, up from zero today."

Gartner, "Innovation Insight for Generative AI"

ARTIFICIAL INTELLIGENCE IS VITAL TO PREDICT VIRUS EVOLUTION GENSLM: WORLDS FIRST GENOME-SCALE AND LARGEST BIOLOGICAL LANGUAGE MODEL WITH 25B PARAMETERS





CONCLUSION

- Generative AI is the golden era of AI
- Broad understanding of domain
- Generalization and robustness
- Generative AI is set to revolutionize science