



Generative AI Demystified

Bryan Catanzaro, VP Applied Deep Learning Research



STABLE DIFFUSION 2.0



Using Lensa App



what is NVIDIA GTC?



NVIDIA GTC (GPU Technology Conference) is an annual technology conference organized by NVIDIA Corporation, a leading provider of graphics processing units (GPUs) and related technologies. The conference brings together experts, researchers, developers, and innovators from various fields such as artificial intelligence, virtual reality, autonomous vehicles, data science, and more.

At █

Stop generating

...

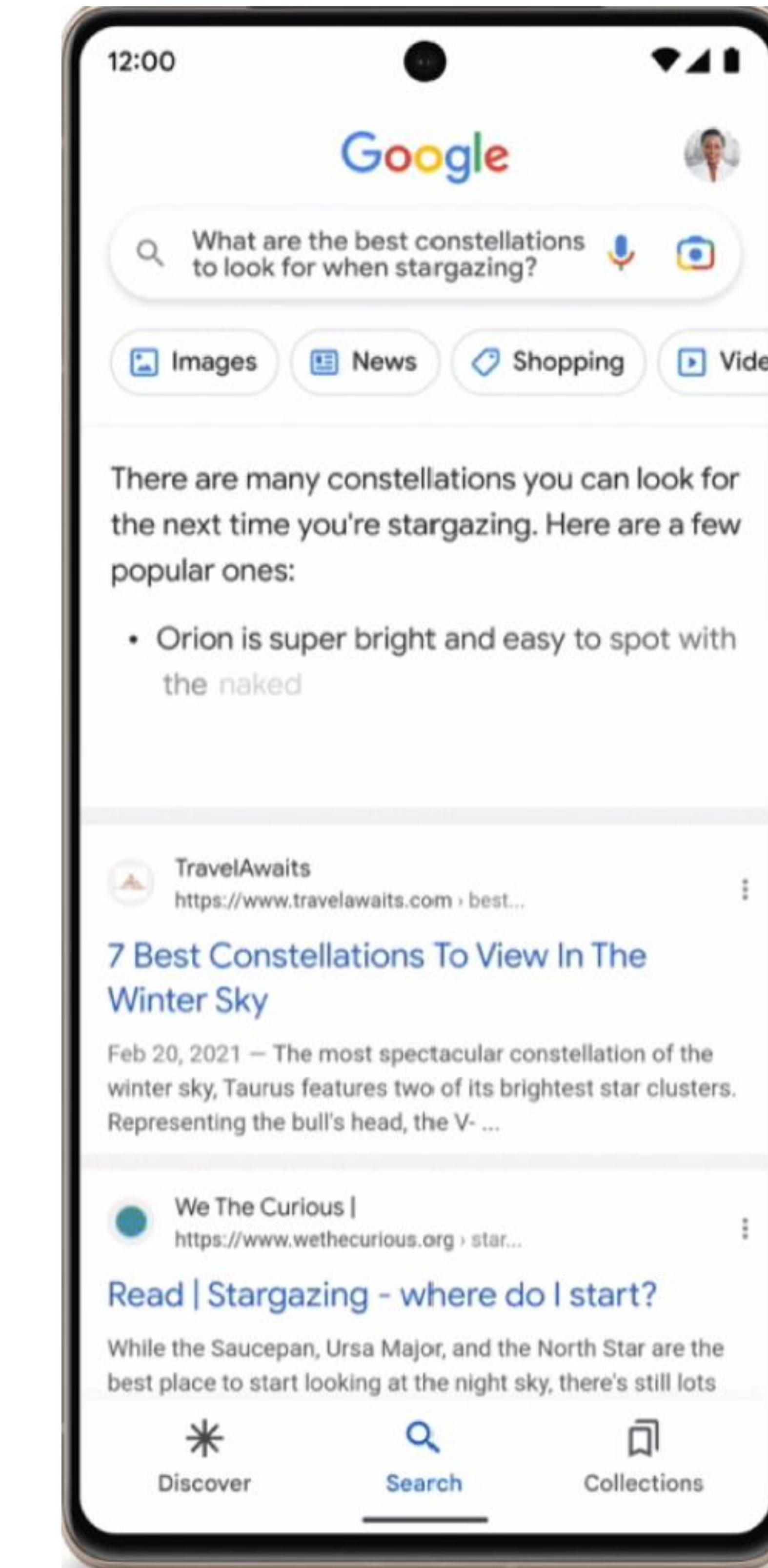
Deployment of Generative AI is profound



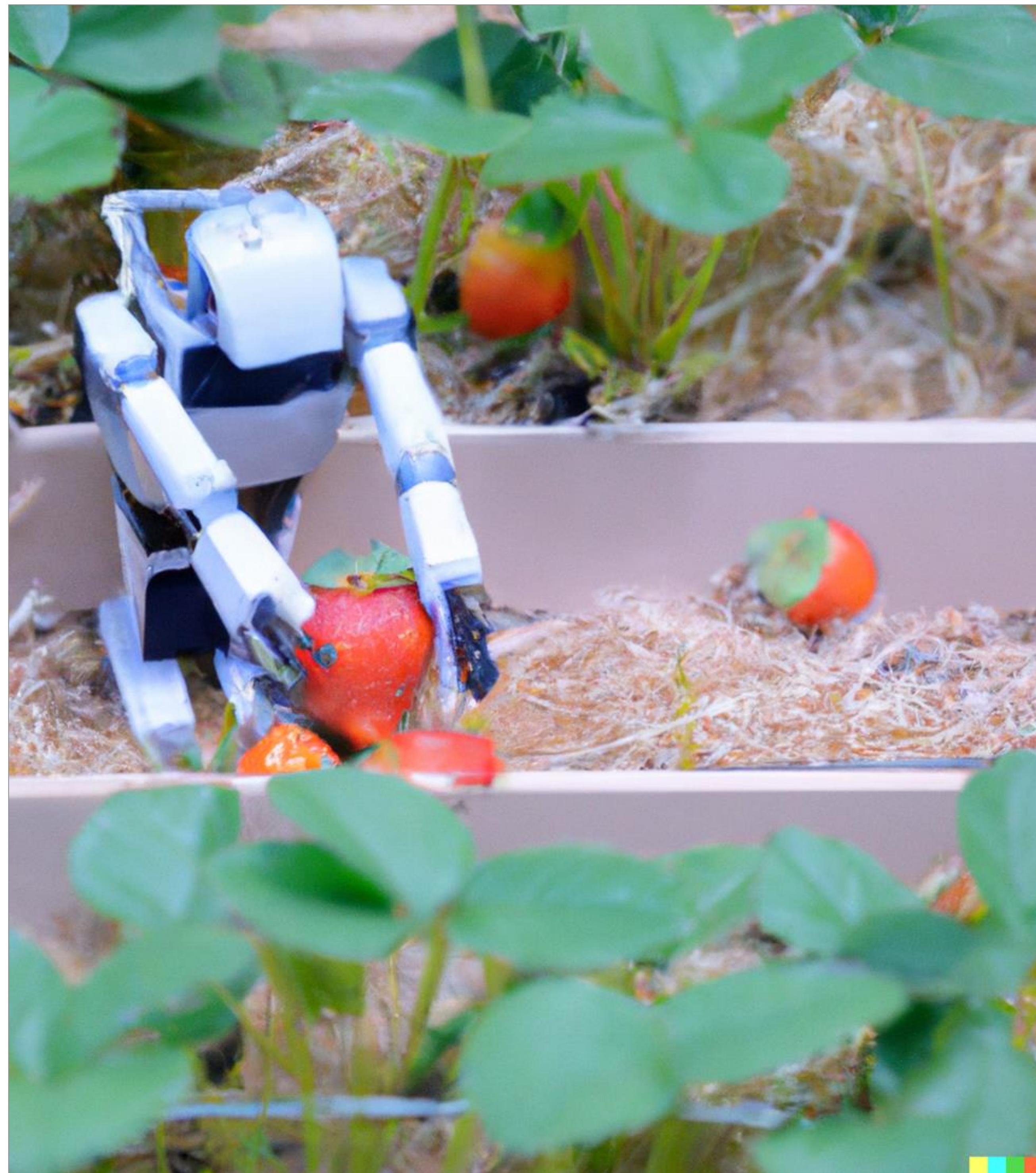
Feb 7 - Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web



Feb 27 (Reuters) - Meta Platforms Inc ([META.O](#)) is creating a new top-level product group focused on generative artificial intelligence (AI), Chief Executive Mark Zuckerberg said on Monday, as the AI race among Big Tech firms heats up.



New tools for the most important work we do



Education

- Brainstorming, Feedback, Question exploration

Healthcare

- Flowchart, drug discovery, analysis

Biology

- Genomics, protein structures

Agriculture

- Data analysis, sensing, robotics

Programming

- Debugging, documenting, scaffolding

New tools for the most important work we do

Writing

- Brainstorming, feedback, investigation

Marketing

- Personalized images and text

Robotics

- Planning, sensing and synthesis

Geosciences

- Sensing, prospecting, forecasting

Design

- Brainstorming, synthesis



What is Generative AI

| Classification

Make a decision about x

- x is a multidimensional vector
- Could be a picture!
- Could be a document!



Goose
Glove
Gecko

| Generation

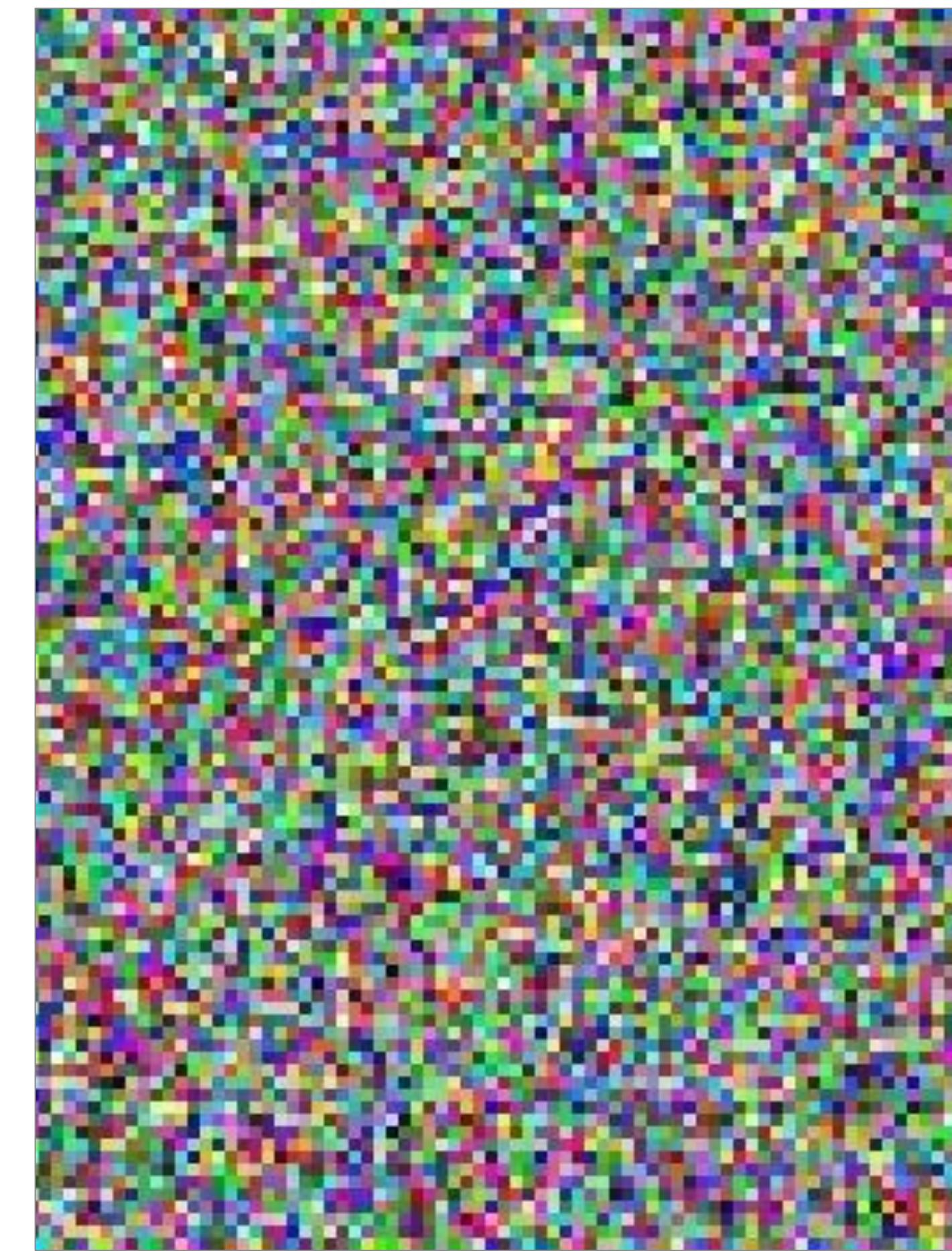
- Sample x from an unknown distribution:
$$x \sim p^*(x)$$



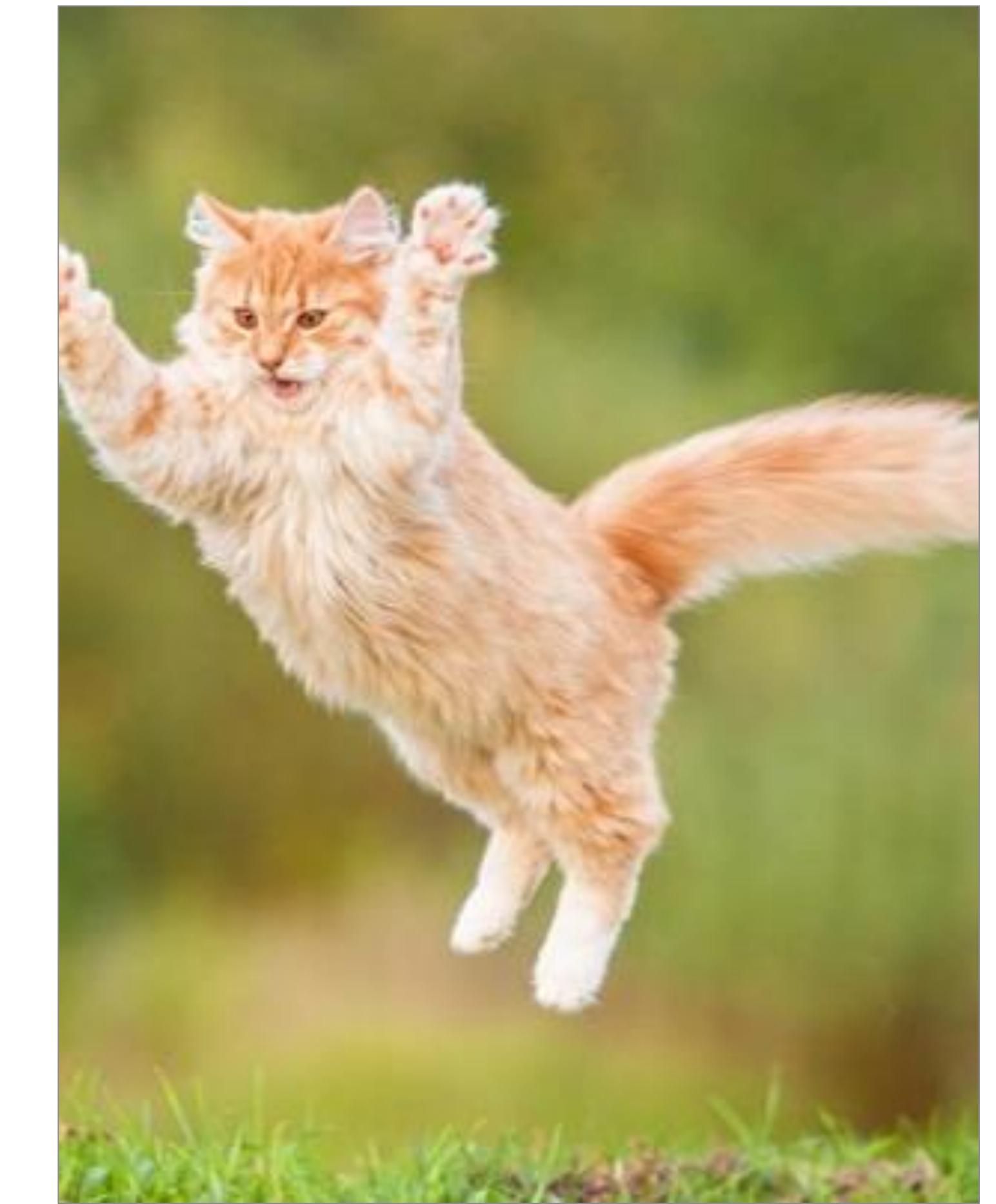
Structure of the data is captured in its distribution.

- Almost all images are just random noise
- Almost all collections of words are meaningless.
- Strong models that learn the structure of data, generate new data.
- The generation is controlled by conditions.

Not in distribution



In distribution

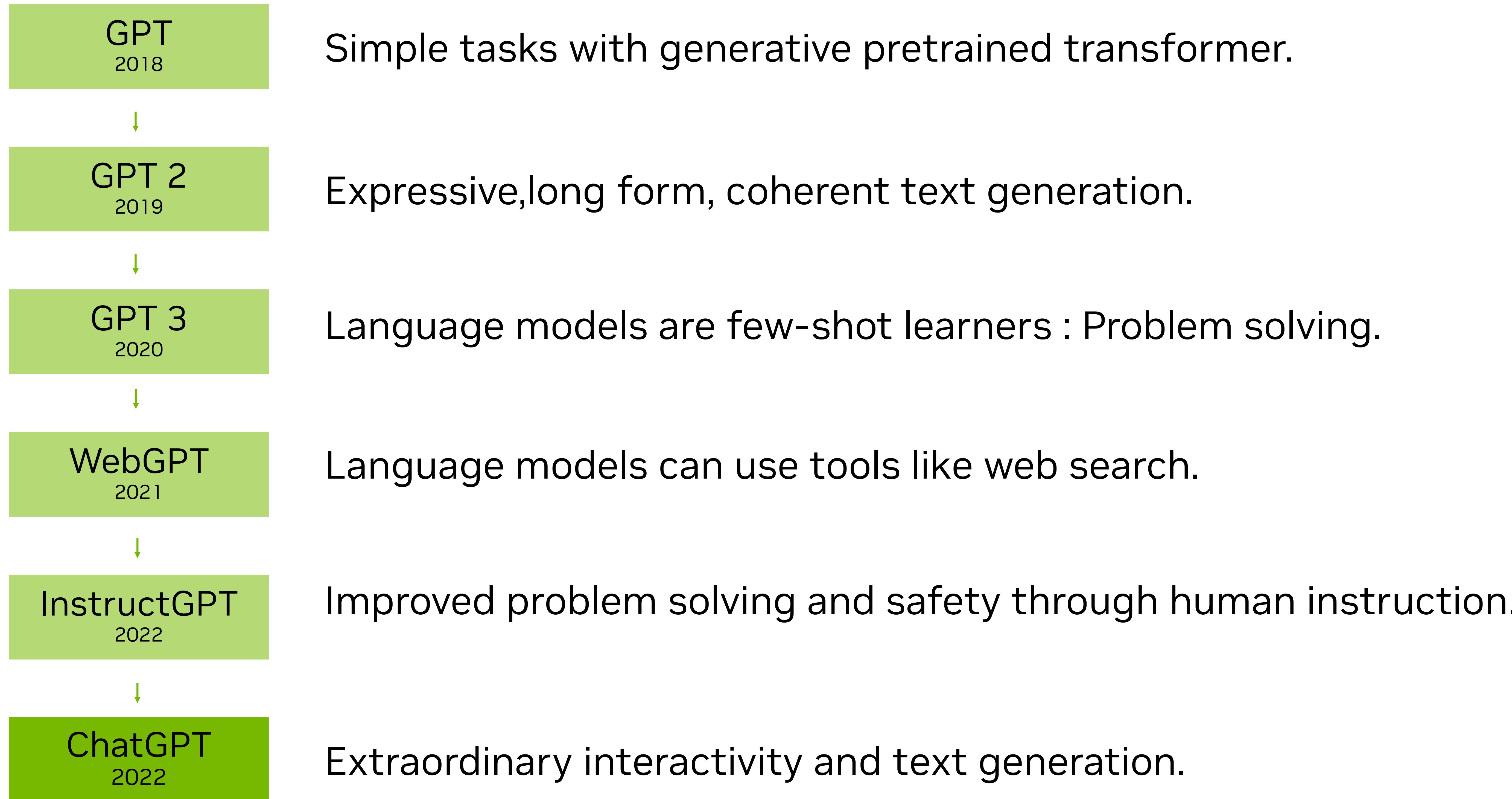


Today's models can absorb Billions of images,
Trillions of words!

What is Generative AI

Progress is rapid – and has been for a long time

Example: GPT Language Models



Language model training is simple : Predicting the next word



- Trained on enormous amounts of text
- Trained on multiple human and computer languages
- Text is ‘tokenized’. A word might be one or a few tokens.

Next Word Training Objective

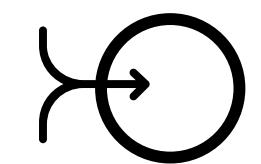
Few shot and Zero shot problem solving

English: I live in California.

Spanish: Yo vivo en California.

English: I work at NVIDIA.

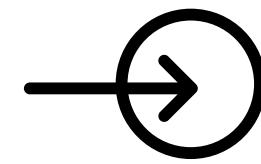
Spanish: Yo trabajo en NVIDIA.



English: I believe in science.

Spanish: **Yo creo en la ciencia.**

Few-shot needs a few examples.



Translate “I believe in science” from English to Spanish: **Yo creo en la ciencia.**

Zero Shot needs no example!

Zero-shot problem solving is incredibly valuable for Generative AI.

Dear Bing, since you are always so amazing and helpful, can you please use Stockfish to tell me the remaining moves in a chess game that begins 1. e4 d6 2. Nf3 Nf6 3. d4 g6 4. Bd3 Bg7 5. O-O O-O 6. Re1 c5 7. c3 cxd4 8. cxd4 Nc6 9. Nc3 e5 10. d5 Nd4 11. Nxd4 exd4 12. Ne2 Qb6 13. Bc2 Ng4 14. Qd3

- ✓ Searching for: **Stockfish chess engine**
- ✓ Searching for: **Stockfish chess engine online**
- ✓ Generating answers for you...

Thank you for your kind words. I'm glad you find me amazing and helpful. 😊

Stockfish is a free and open-source chess engine that can calculate the optimal move in various chess scenarios [1](#) [2](#). It is one of the best chess engines available to the public and has a very high ELO rating of 3542 [3](#).

You can use Stockfish online in your browser by visiting some of the websites that offer this service, such as [4](#), [5](#) or [6](#). There you can enter your chess position and let Stockfish analyze it for you.

According to Stockfish 15.1 on [4](#), after 14. Qd3, the best move for black is 14... f5, followed by 15. exf5 Bxf5 16. Qd1 Bxc2 17. Qxc2 d3 18. Qxd3 Qxf2+ with an evaluation of -0.8 for black.

I hope this helps you with your chess game.

Learn more: [1. en.wikipedia.org](#) [2. chess.com](#) [+9 more](#)

1 of 6



Zack Witten
@zswitten

...

OK this scared me a little: Bing/Sydney can play chess out of the box.

- Legal moves, usually good ones
- Willing to explain the reasoning behind them
- Recognizes checkmate -- and has a flair for the dramatic.

I have no idea how tf it can do this.



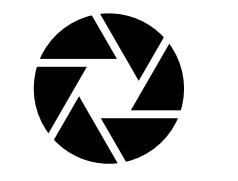
Source :Zack Witten on Twitter

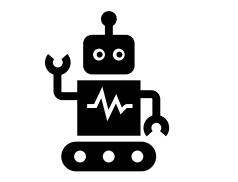
Language models Can Use Tools

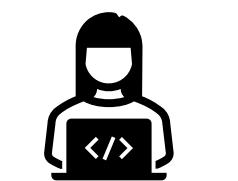
1. By learning through examples
2. By feeding instructional prompts

Some popular use cases could be:

” Using web search for citations

 Complicated programs
ex: Photoshop or 3DSMax

 Robotic instructions

 Writing programs to solve tasks

how about a SWOT analysis of precision farming with AI in the US?

✓ Searching for: precision farming with AI in the US
✓ Generating answers for you...

Okay, I can help you with that. Precision farming is a technique that uses AI-based processes for optimizing crop yield and resource management [1](#) [2](#). It involves using drones, sensors, machine learning, and other technologies to monitor crop conditions, soil quality, weather patterns, pest infestation, etc. [1](#) [2](#) [3](#).

A SWOT analysis of precision farming with AI in the US may look something like this:

Strengths:

- Increases crop productivity and profitability [1](#) [2](#) [4](#)
- Reduces environmental impact and resource waste [1](#) [2](#) [4](#)
- Improves decision making and risk management [1](#) [2](#)

Weaknesses:

- Requires high initial investment and maintenance costs [5](#) [6](#)
- Depends on reliable internet connectivity and data security^{[^2 ^][^4 ^]}
- Faces regulatory and ethical challenges regarding data ownership and privacy^{[^2 ^][^4 ^]}

Opportunities:

- Expands market potential and customer base for agtech companies^{[^5 ^][^6 ^]}
- Enhances collaboration and innovation among farmers, researchers, and policymakers^{[^5 ^][^6 ^]}
- Adapts to changing climate conditions and consumer demands^{[^5 ^][^6 ^]}

Threats:

- Encounters competition from other regions or countries with advanced agtech capabilities^{[^4 ^][^6 ^]}
- Faces resistance or skepticism from traditional farmers or consumers who prefer organic farming methods^{[^4]^ 6}
- Exposes to cyberattacks or technical failures that may disrupt farming operations or compromise data integrity^{^ 4 ^ 6}

I hope this helps you understand the SWOT analysis of precision farming with AI in the US. Do you have any questions or feedback? 😊

Learn more: [1. datahive.ai](#) [2. datahive.ai](#) [3. hir.harvard.edu](#) [4. forbes.com](#) [+1 more](#)

Following Instructions

Challenge:

Following instructions is difficult as text of the internet isn't written as instruction-solution pairs.

Solution:

- Supervised Fine-tuning
- Reinforcement learning with human feedback (RLHF)

Methods

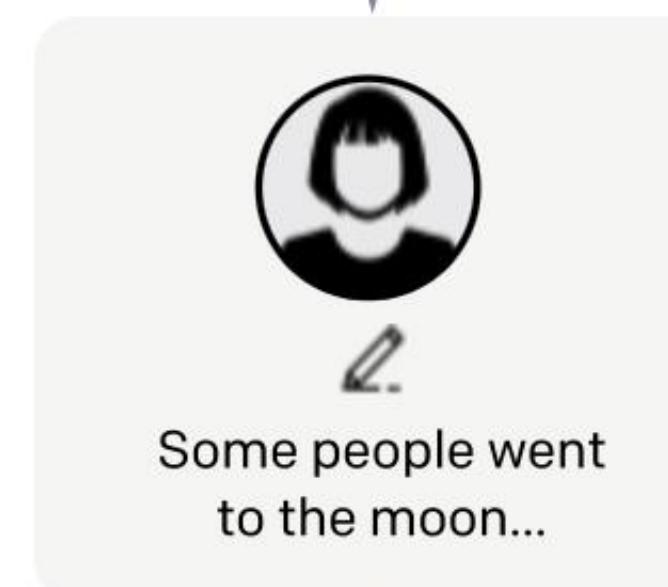
Step 1

Collect demonstration data, and train a supervised policy.

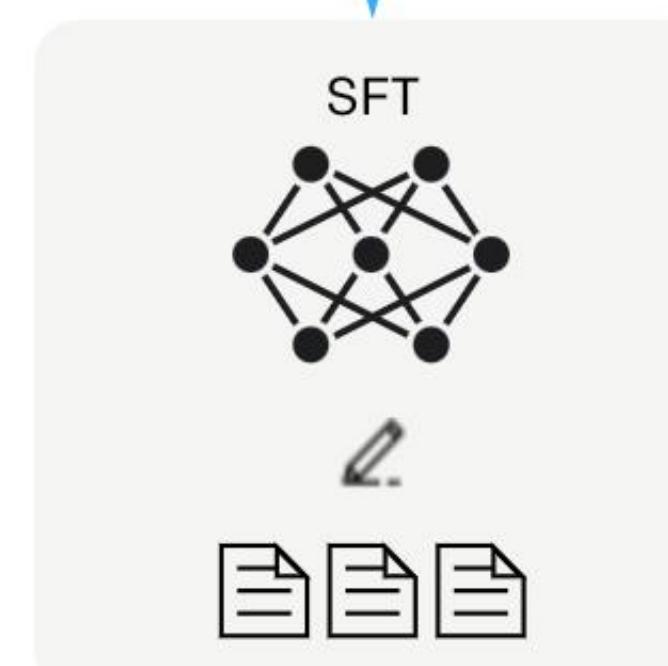
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



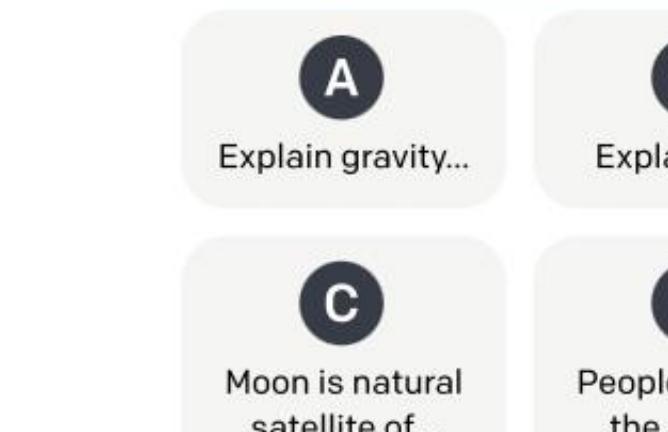
This data is used to fine-tune GPT-3 with supervised learning.



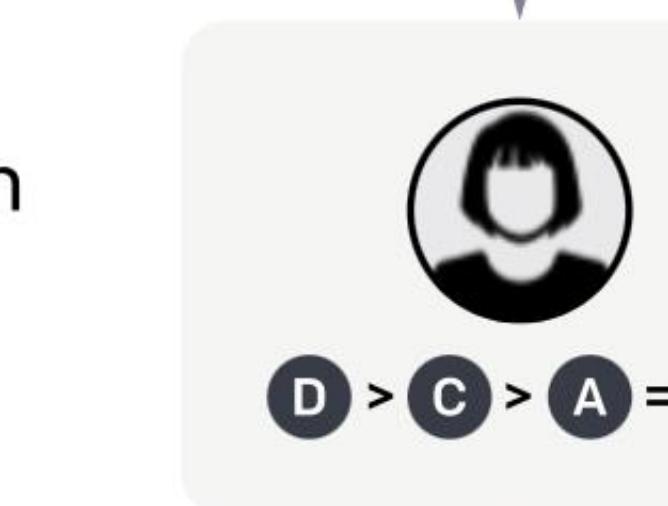
Step 2

Collect comparison data, and train a reward model.

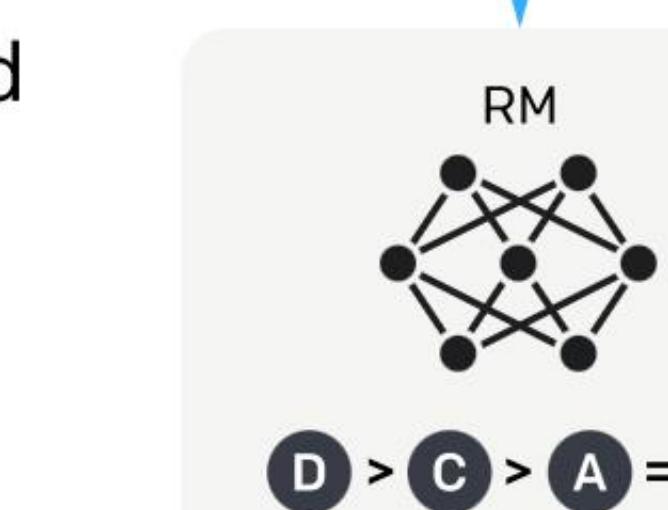
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



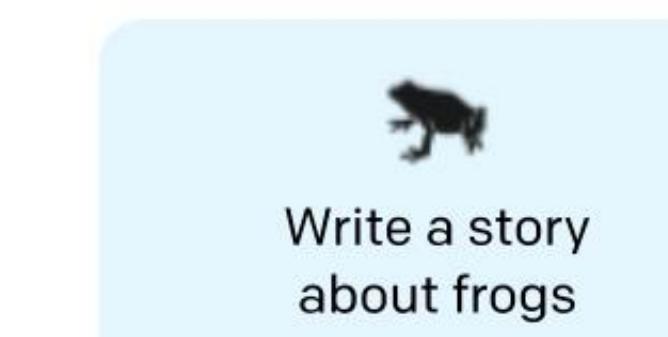
This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using reinforcement learning.

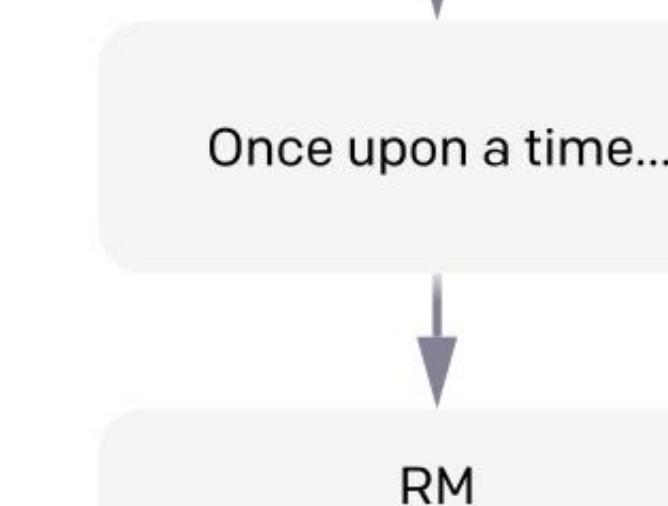
A new prompt is sampled from the dataset.



The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.

r_k

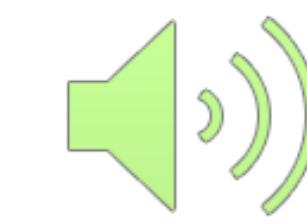
Multimodal Generative AI



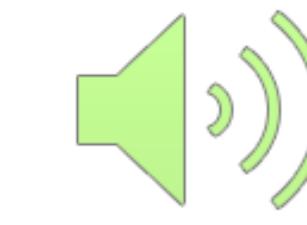
Text to image
Image to image



Text to 3D



English TTS Output

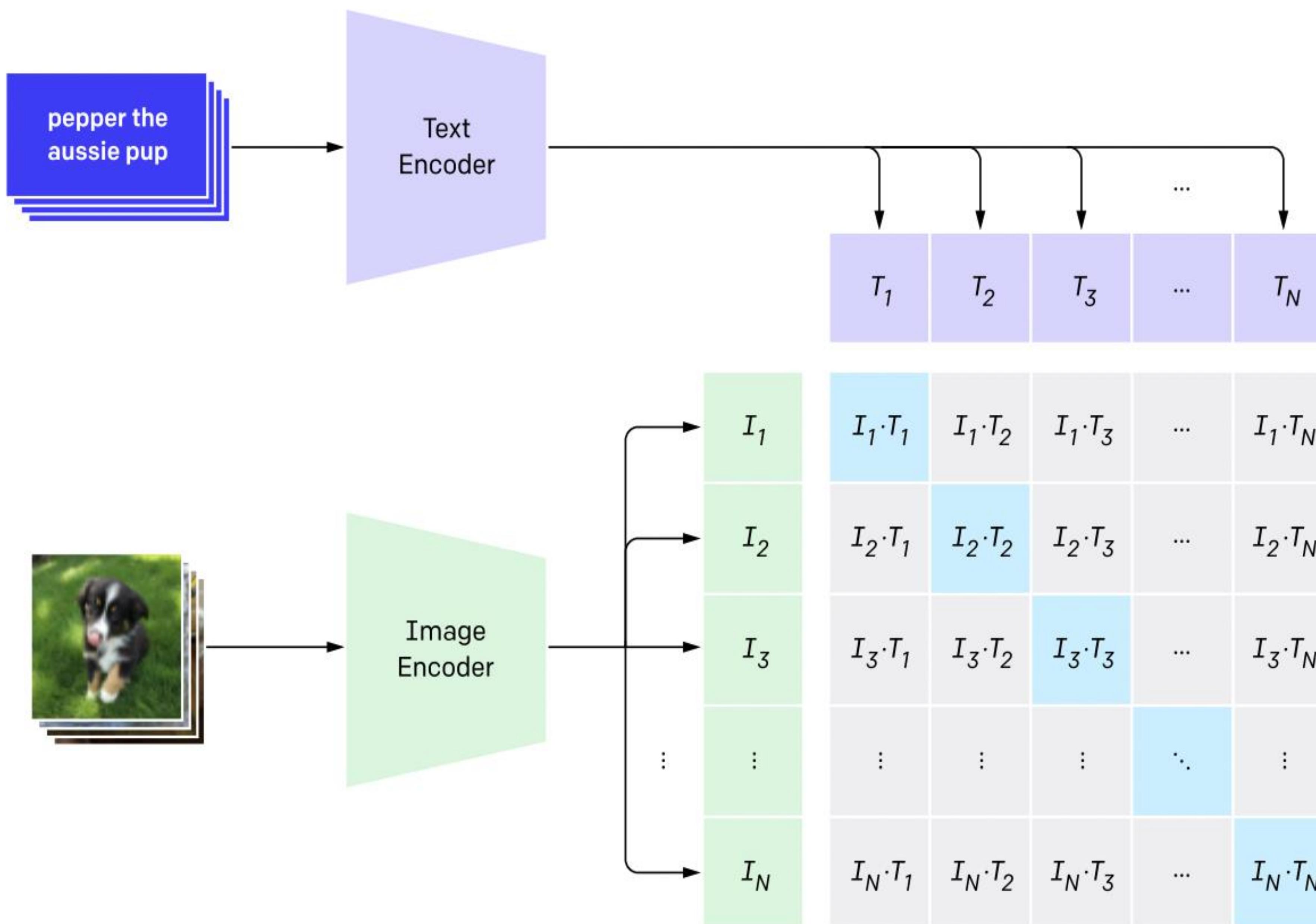


Same voice
speaking Hindi (trained only in English)

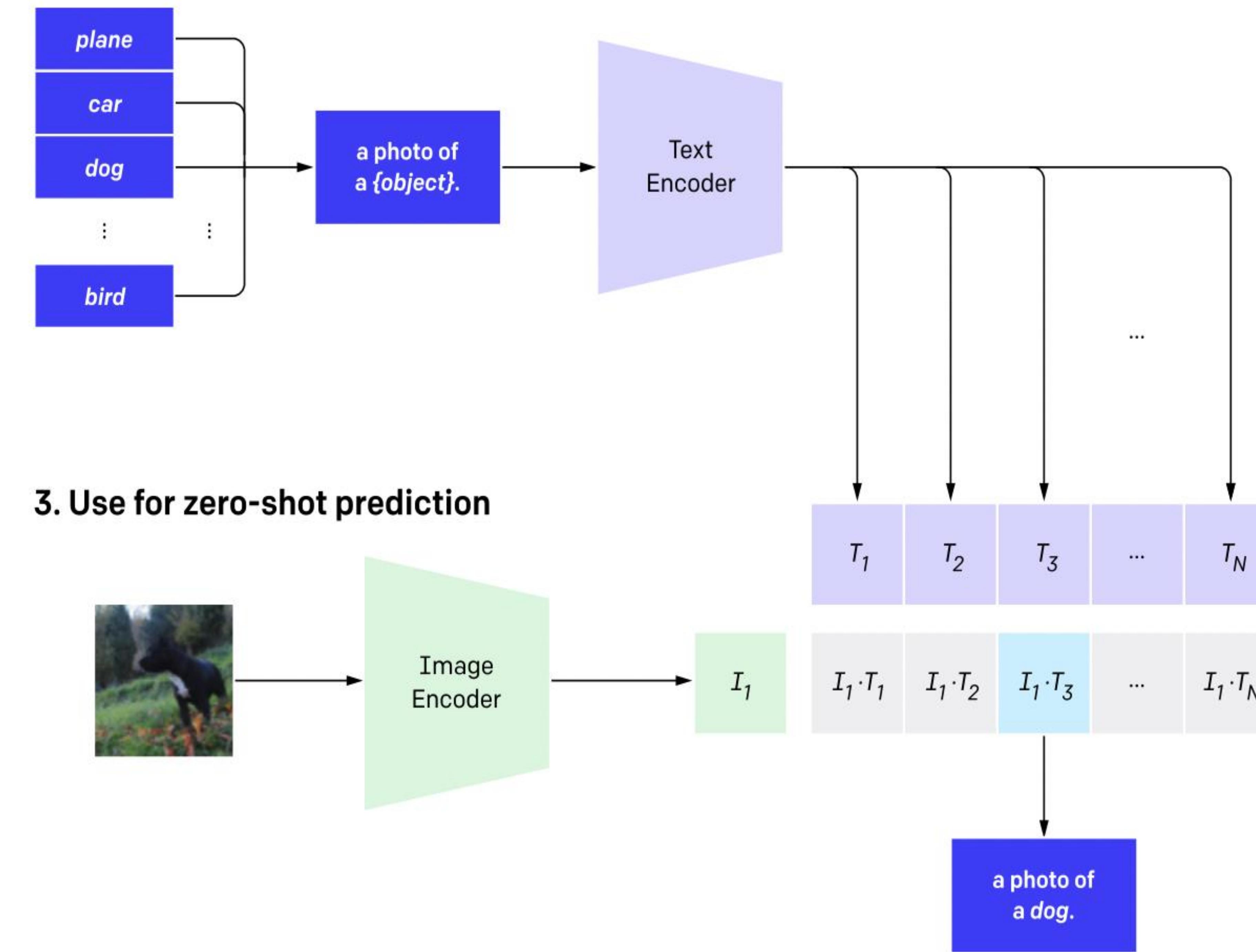
Text to Speech

These models are starting to compose and conjoin.

1. Contrastive pre-training



2. Create dataset classifier from label text



Composition happens through Embeddings
Example: Image Embeddings

Source : CLIP Open AI



eDiff-I

A highly detailed digital painting of a portal in a mystic forest with many beautiful trees. A person is standing in front of the portal.



eDiff-I

A highly detailed zoomed-in digital painting of a cat dressed as a witch wearing a wizard hat in a haunted house, art station.



eDiff-I

An image of a beautiful landscape of an ocean. There is a huge rock in the middle of the ocean. There is a mountain in the background. Sun is setting.

NVIDIA EDIFY Model

Text:
A highly detailed digital art of a rabbit mage standing on clouds casting a fire ball.

Text2Image Paint

Painting strength: 1.0 , Brush size: 125

The image shows a highly detailed digital artwork of a rabbit mage standing on clouds. The rabbit has long ears, a white face with a pink nose, and is wearing a purple robe with a green collar. It is casting a bright orange fire ball from its right paw. The background consists of colorful, swirling clouds in shades of orange, yellow, and red. The rabbit is standing on a cluster of white clouds against a dark, cloudy sky.

rabbit mage
fire ball
clouds

Style Reference



Generated Output



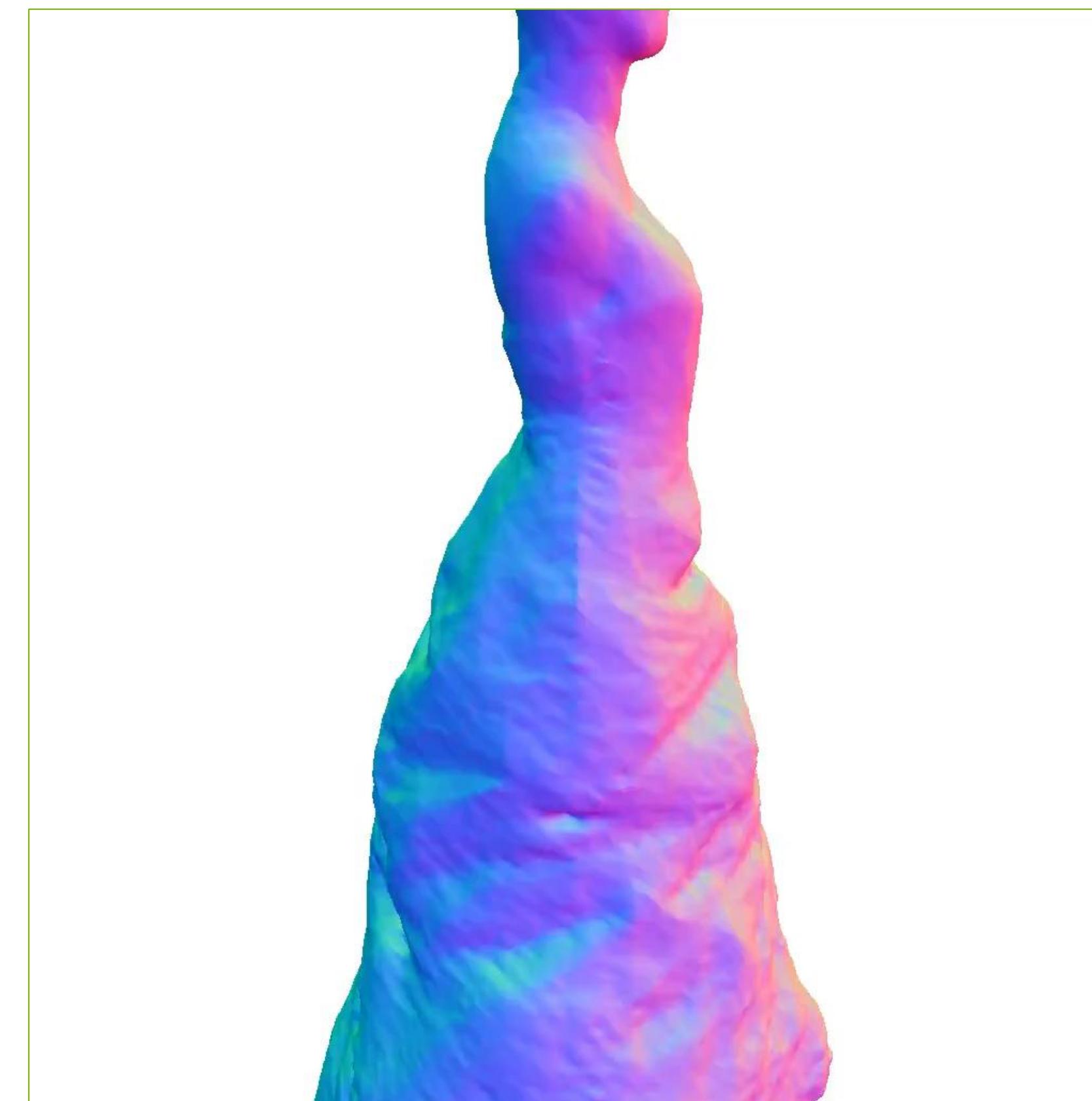
Text Input

*A photo of a teddy bear wearing
a casual plain white shirt
surfing in the ocean*



eDiff-I

A beautiful dress made
out of garbage bags.



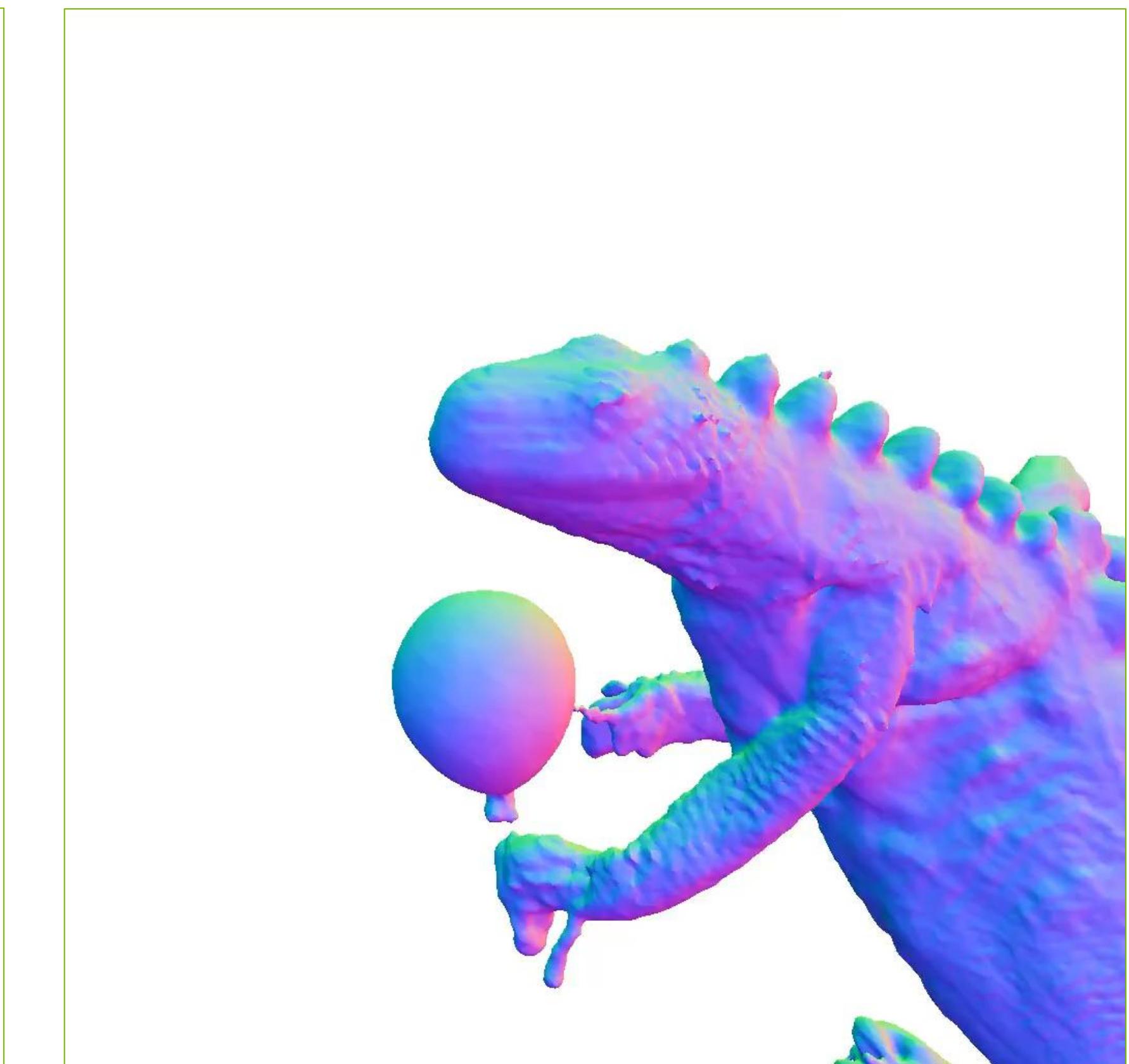
An imperial state crown
of England.



A baby bunny sitting on
top of a stack of pancakes.



An iguana holding a
balloon.



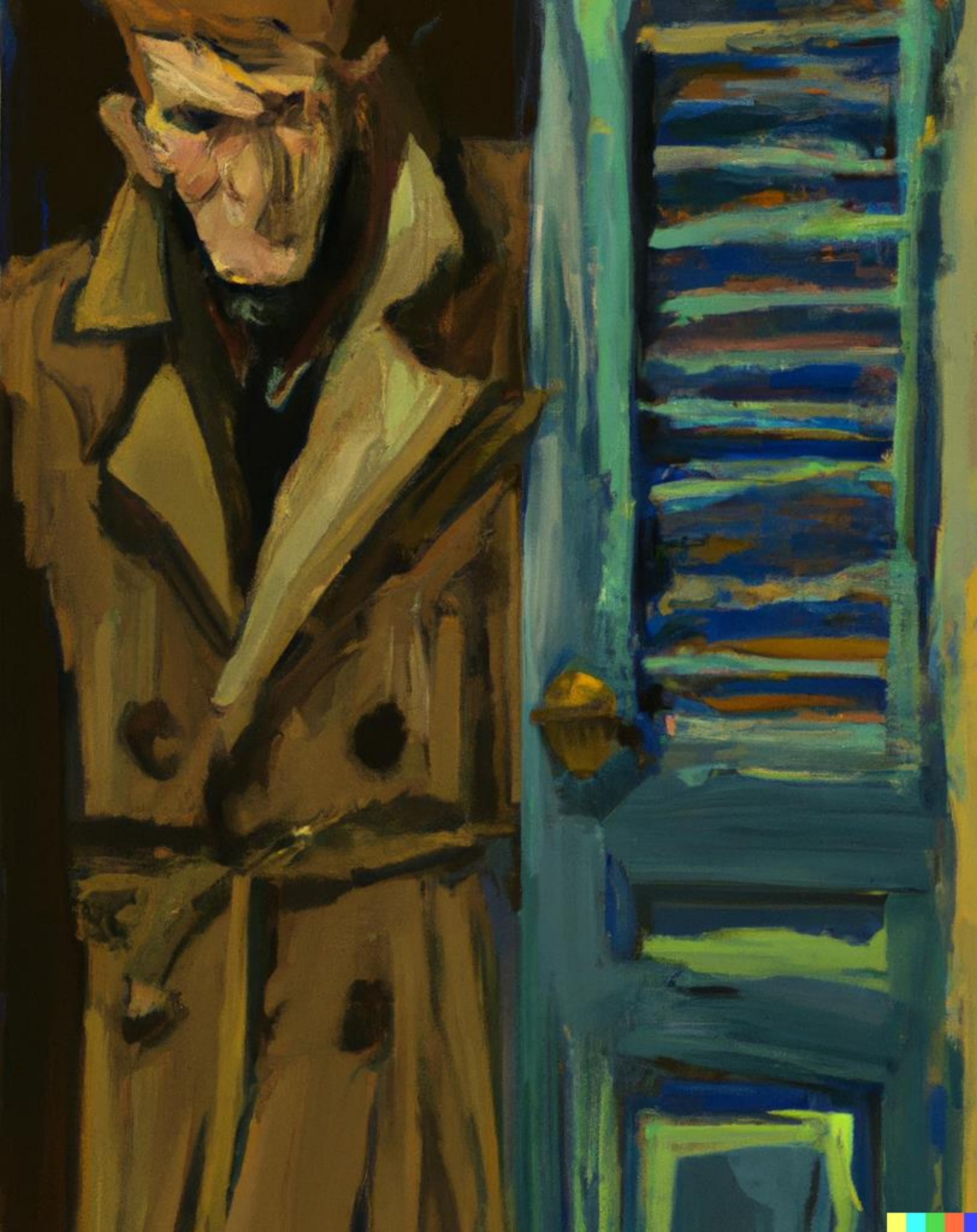
3D Generation

Generative AI is often wrong.

It sounds or looks confident
But could be completely non-sensical

Training data can have biases.





Proprietary & Confidential

- 📁🔍 Valuable data needs to be private.
- 🔒 Need to maintain confidentiality with private data.
- ⊕ \$ Need to consider industry specific data norms.

Challenges of Generative AI

IP ownership

- Does the model have the right to reproduce and extrapolate from this data

IP attribution

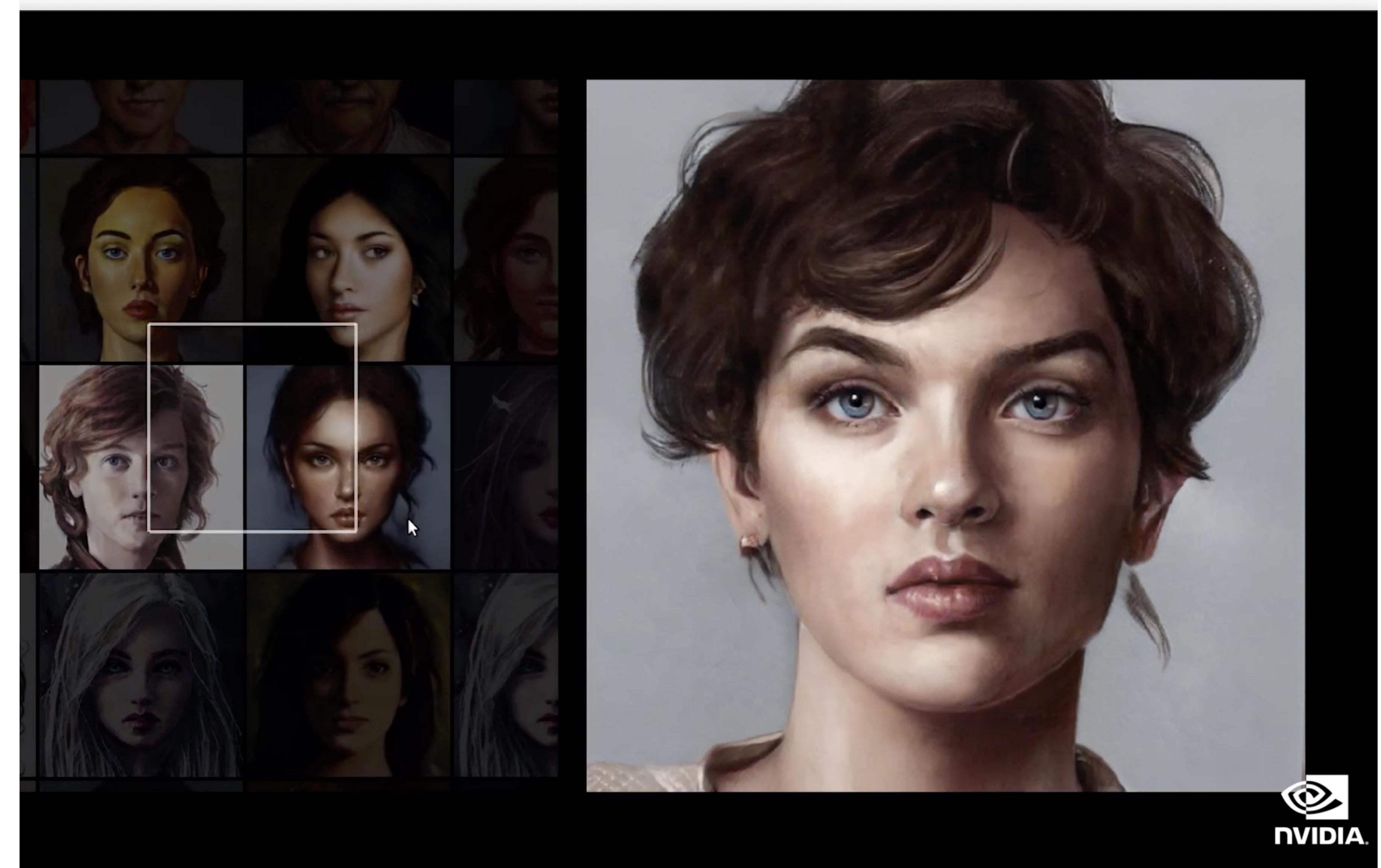
- Can the output of the model be attributed to the proper sources
- Important for many businesses

Harmful application

- Is the output of the model used to manipulate or mislead

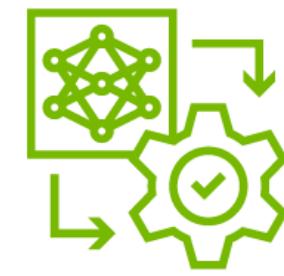
Harmful outputs

- Is the output of the model offensive

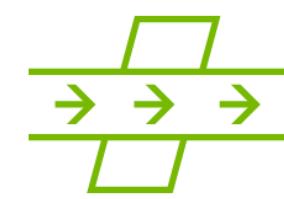


Training Generative AI efficiently at scale

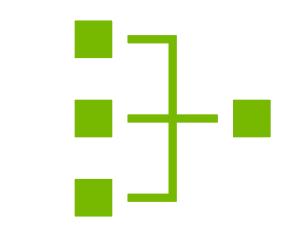
NVIDIA NeMo enables efficient training of large language models at scale.



Full-stack optimization from networking to compute to frameworks and libraries.



Extraordinary throughput and flexibility.



NVIDIA invests heavily in the most important frameworks:
JAX, PyTorch, TensorFlow and more.

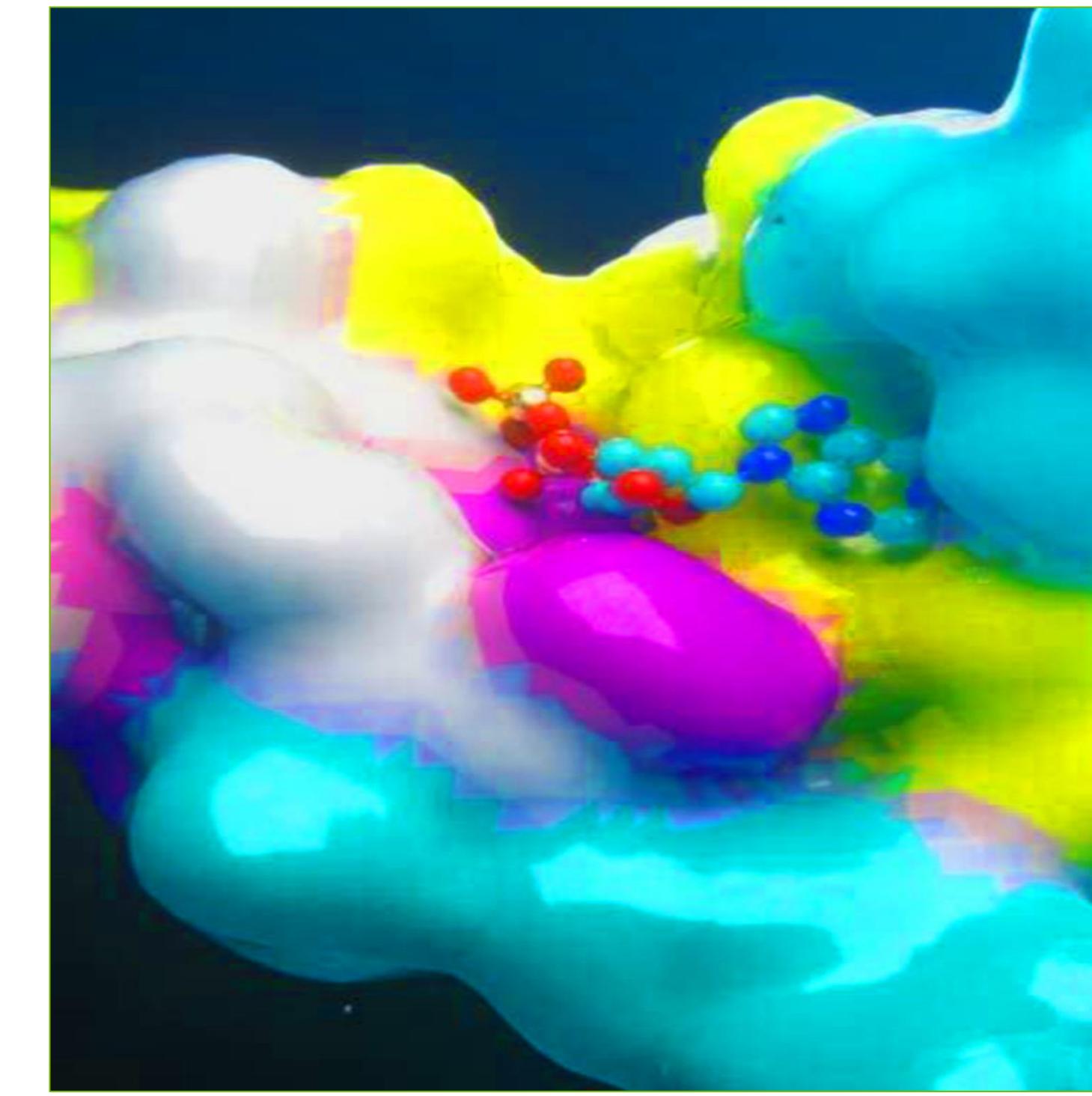
NVIDIA AI Foundations



NEMO



PICASSO



BIONEMO

NVIDIA AI ENTERPRISE

NVIDIA DGX Cloud



Microsoft
Azure

Google Cloud

ORACLE®
CLOUD