

# AI Image Generation



# What is an Image?

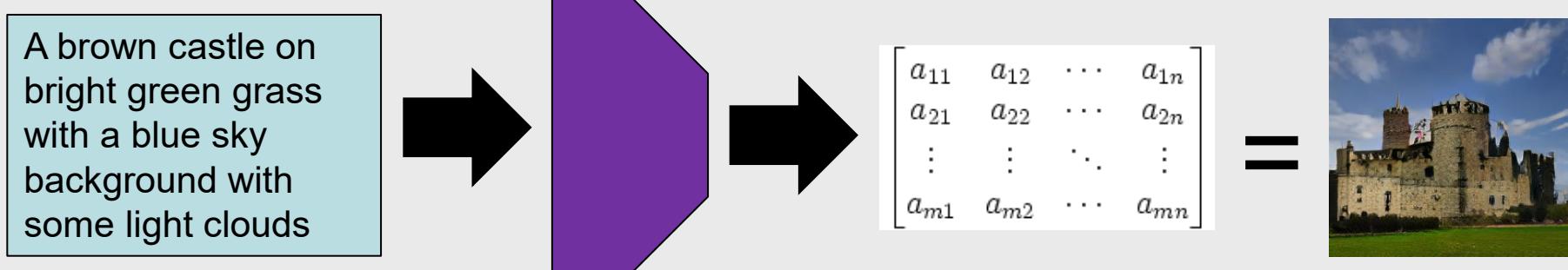
- An image can be represented as a big matrix (1024 x 1024 pixels)
  - Really 3 big matrices for Red, Blue, and Green colors
- Each pixel is a matrix element, its value gives brightness or color
- Pixels have some structure
  - Neighboring pixels are usually the same color
  - Pixels change color at an edge
  - Enclosed regions have the same color



$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

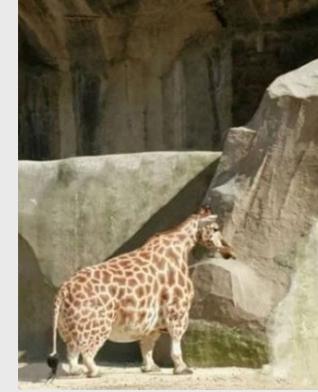
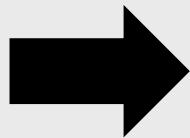
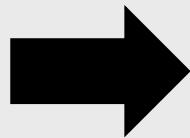
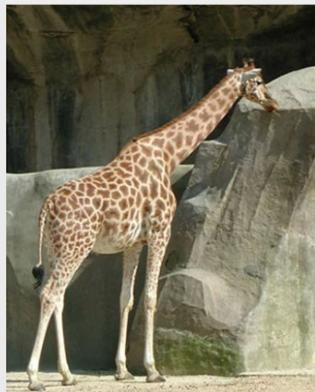
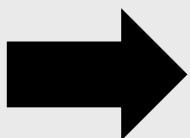
# Image Generation

- Generating an image means generating a matrix of numbers
- The numbers must be arranged in a way to represent real objects
- We would like to control how the pixels are arranged with text descriptions



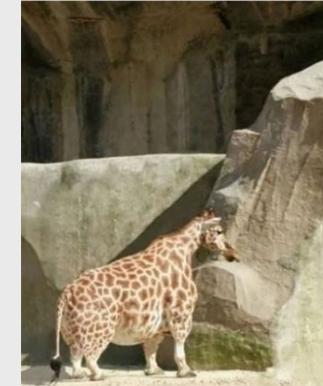
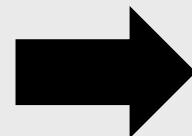
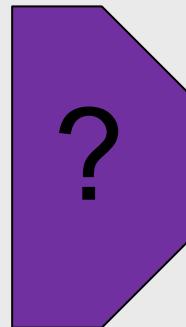
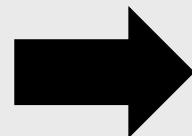
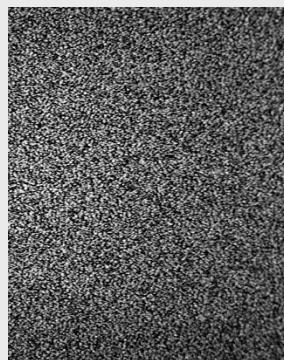
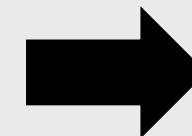
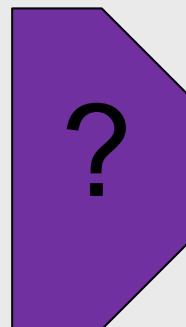
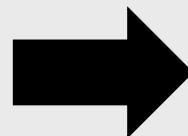
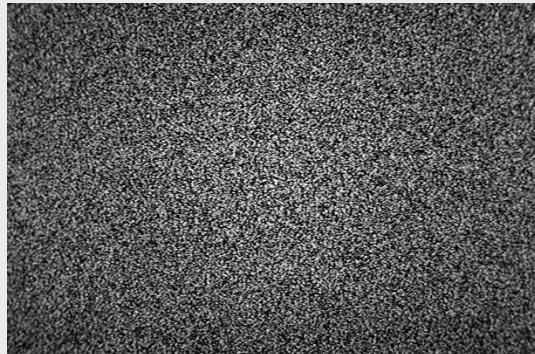
# Modifying an Image

- With some skill and effort and Photoshop, people could edit an existing image



# Generating an Image

- Creating an image from nothing was thought to be very challenging



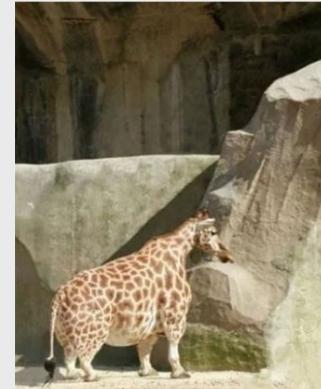
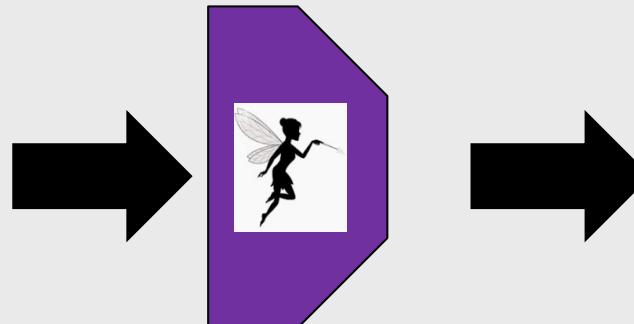
# Generating a Specific Image

- Creating an image from a text description was thought to be a fairy tale

An astronaut holding a bag of weed on a space shuttle



A very short and stout giraffe facing to the right



# **Image Generation Challenges v1**

# **Image Generation Challenges v1**

- How can I make a random image that looks real? (i.e. input a random number, and output a picture of something real)

# **Image Generation Challenges v1**

- How can I make a random image that looks real? (i.e. input a random number, and output a picture of something real)
- How can I make a specific type of image that looks real (i.e. input a random number, and output a picture of a cat)

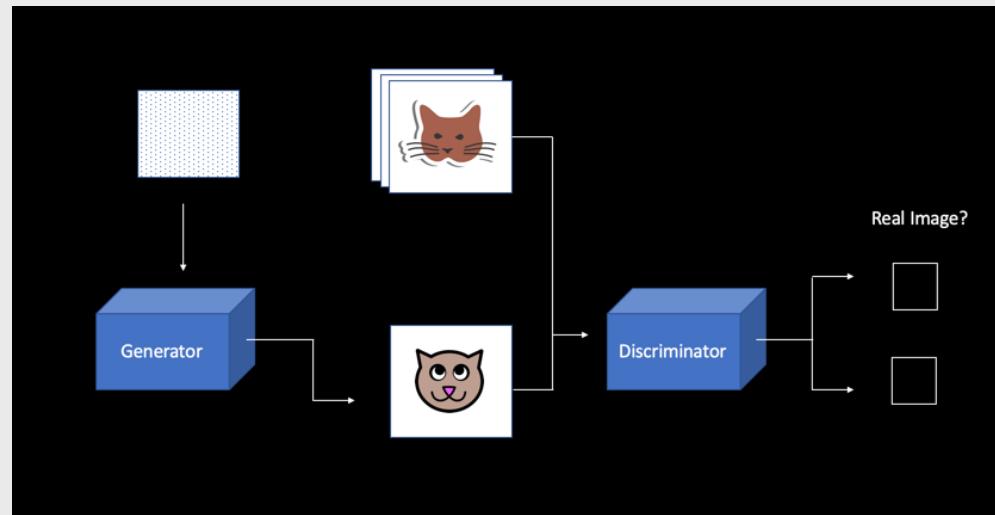
# **Image Generation Challenges v1**

- How can I make a random image that looks real? (i.e. input a random number, and output a picture of something real)
- How can I make a specific type of image that looks real (i.e. input a random number, and output a picture of a cat)
- Solution: **Generative Adversarial Networks (GANs)**

# **Generative Adversarial Network (GAN)**

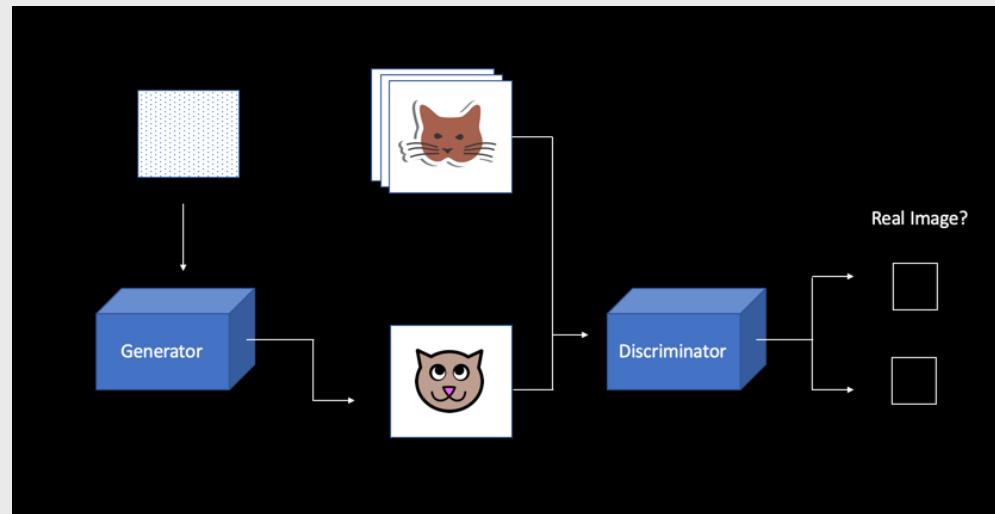
# Generative Adversarial Network (GAN)

- GAN - train **generator** and **discriminator** **networks** in an **adversarial** manner



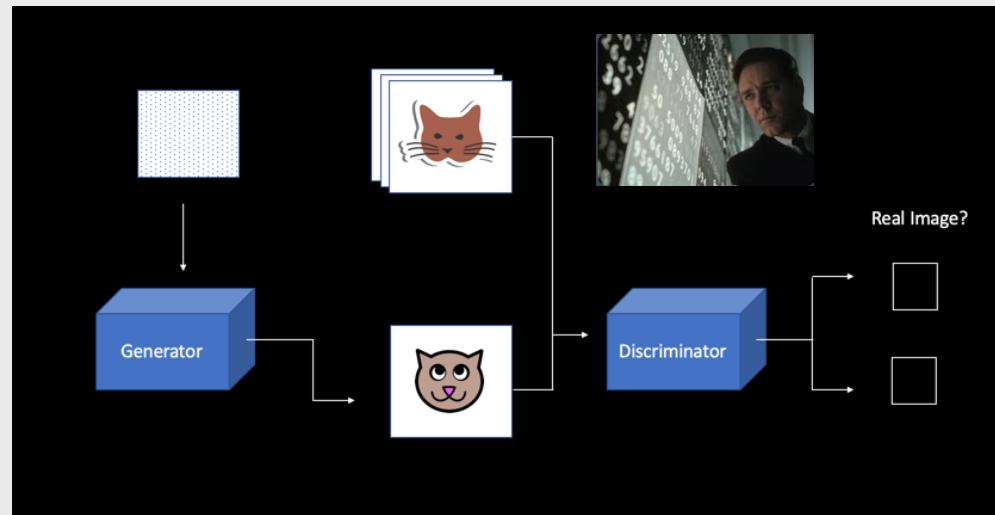
# Generative Adversarial Network (GAN)

- GAN - train **generator** and **discriminator networks** in an **adversarial** manner
- GANs are playing a game
  - Generator adjusts its parameters to fool Discriminator
  - Discriminator adjusts its parameters to catch Generator



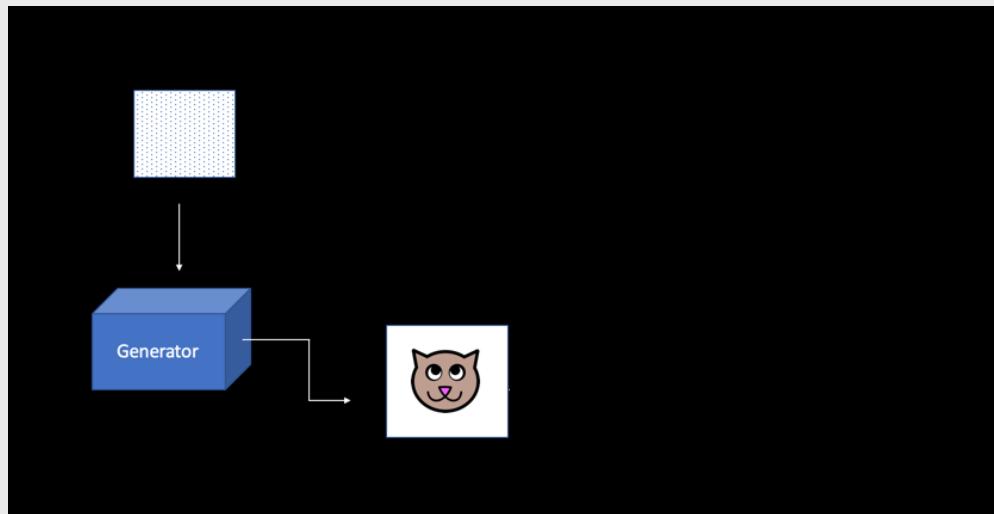
# Generative Adversarial Network (GAN)

- GAN - train **generator** and **discriminator networks** in an **adversarial** manner
- GANs are playing a game
  - Generator adjusts its parameters to fool Discriminator
  - Discriminator adjusts its parameters to catch Generator
- Stop training when networks reach Nash equilibrium



# Generative Adversarial Network (GAN)

- GAN - train **generator** and discriminator **networks** in an **adversarial** manner
- GANs are playing a game
  - Generator adjusts its parameters to fool Discriminator
  - Discriminator adjusts its parameters to catch Generator
- Stop training when networks reach Nash equilibrium



# GAN Paper (2014)

---

## Generative Adversarial Nets

---

Ian J. Goodfellow, Jean Pouget-Abadie\*, Mehdi Mirza, Bing Xu, David Warde-Farley,  
Sherjil Ozair,<sup>†</sup> Aaron Courville, Yoshua Bengio<sup>‡</sup>

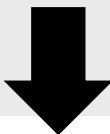
Département d'informatique et de recherche opérationnelle  
Université de Montréal  
Montréal, QC H3C 3J7

### Abstract

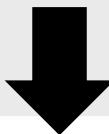
We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model  $G$  that captures the data distribution, and a discriminative model  $D$  that estimates the probability that a sample came from the training data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

# First GAN Images

Training data

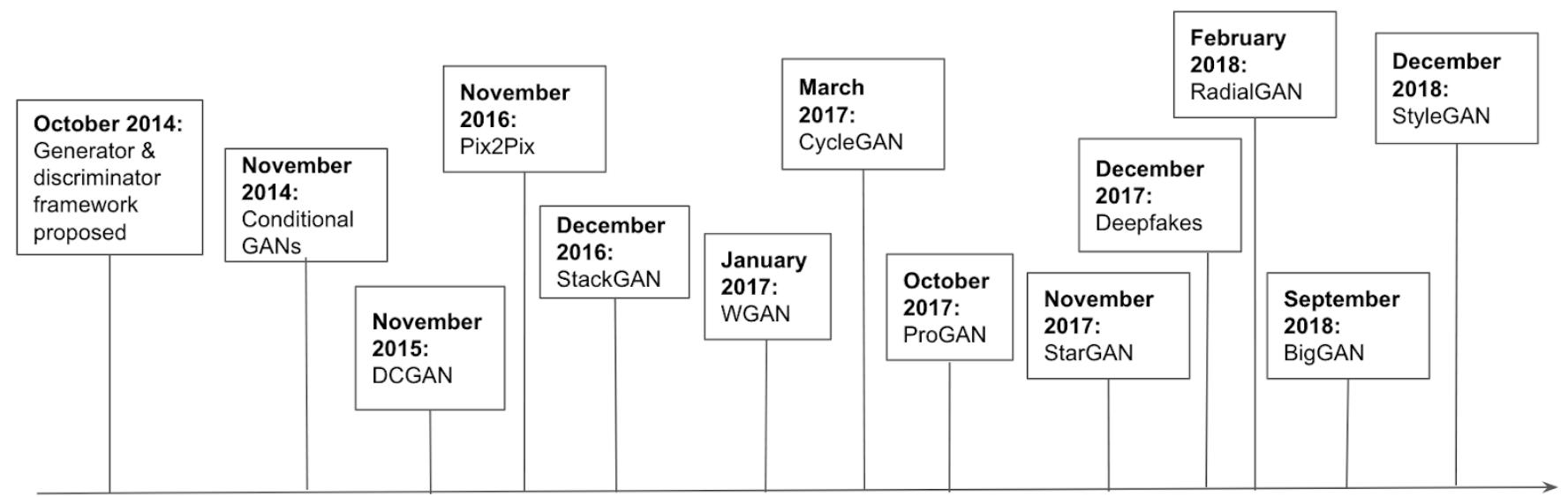


Training data



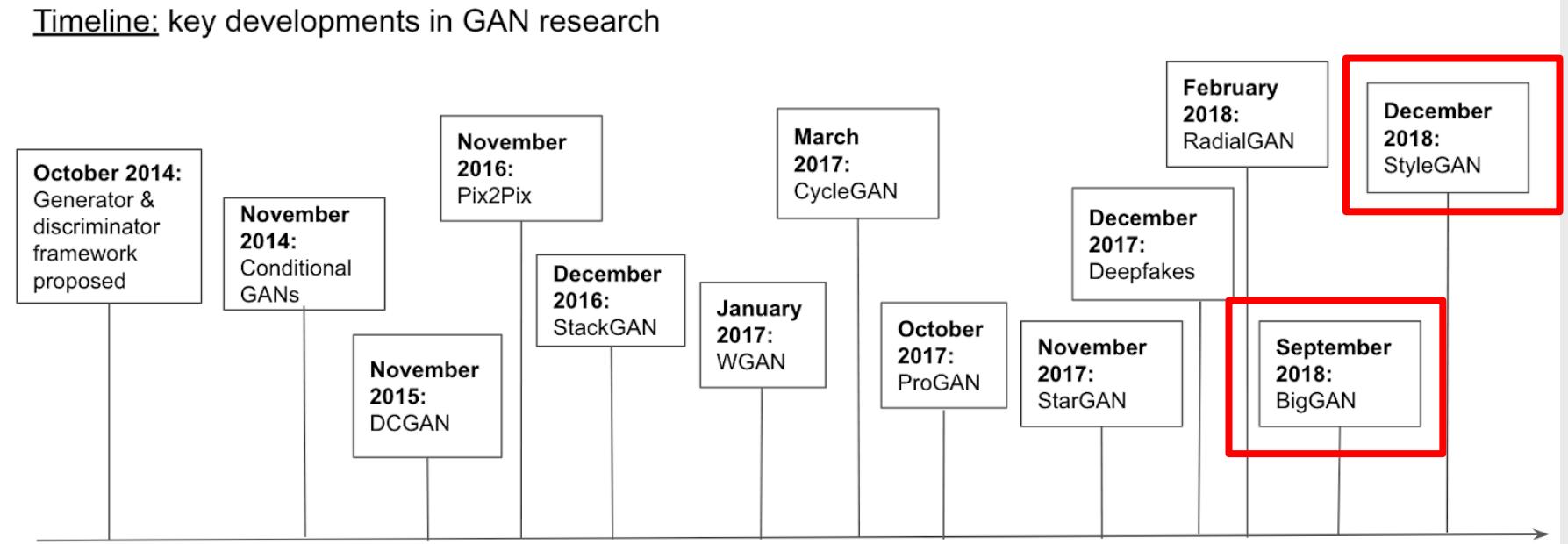
# GAN Timeline

Timeline: key developments in GAN research



# GAN Timeline

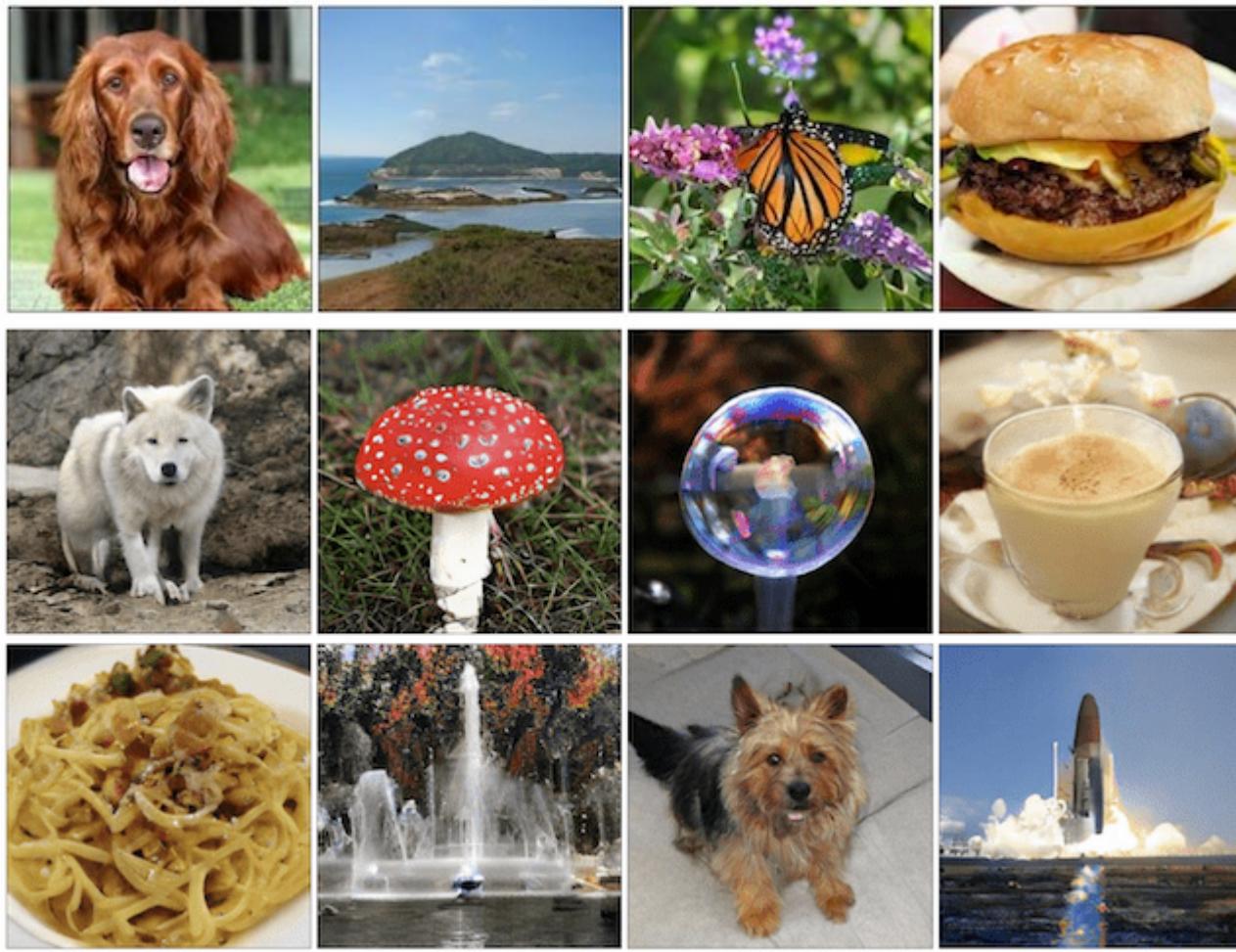
Timeline: key developments in GAN research



# **BigGAN**

- 2018 – BigGAN generate images from thousands of categories
- Focused on scaling up GAN models

# BigGAN Examples



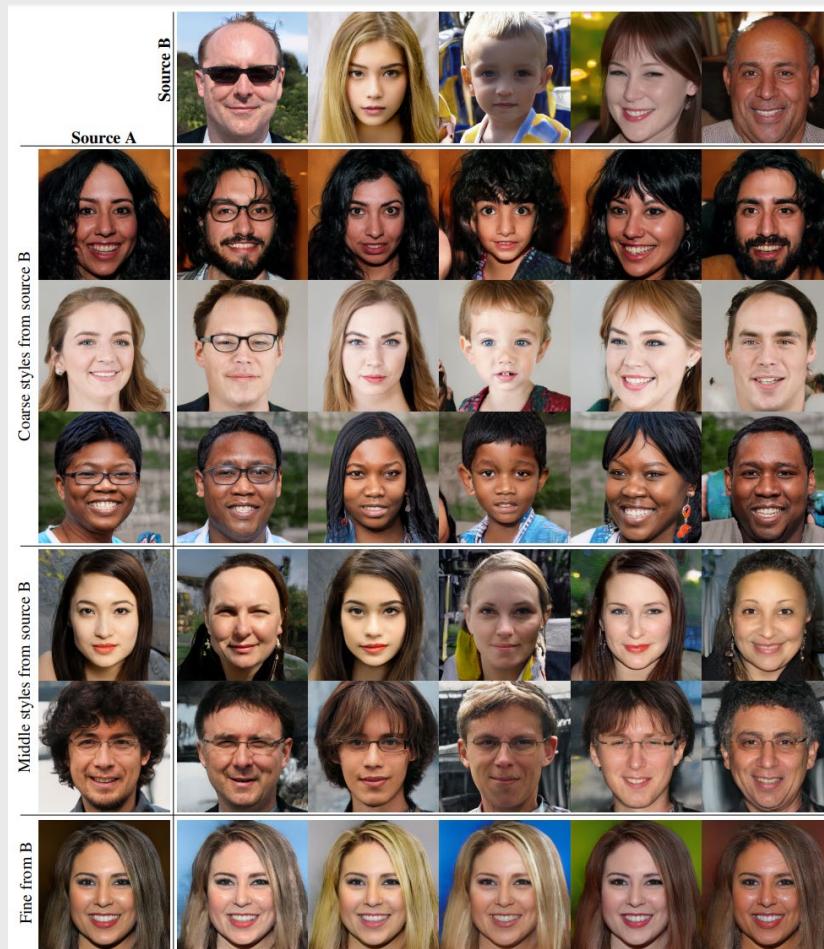
# StyleGAN

- 2018 – StyleGAN generates real looking faces



# StyleGAN

- StyleGAN lets you combine faces



# **GANs Pros and Cons**

- **Pros**
  - Images look very realistic
  - Fast image generation
- **Cons**
  - Cannot control image details
  - Hard to train



# **Image Generation Challenges v2**

# **Image Generation Challenges v2**

- **Make an image based on a text description**
  - i.e. draw a cat in a blue hat tap dancing on top of the Burj Khalifa

# Image Generation Challenges v2

- Make an image based on a text description
  - i.e. draw a cat in a blue hat tap dancing on top of the Burj Khalifa



# Image Generation Challenges v2

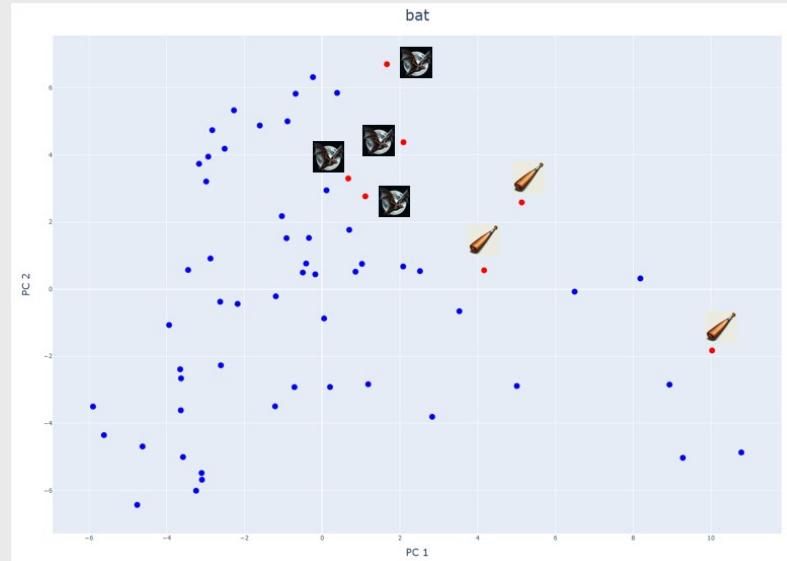
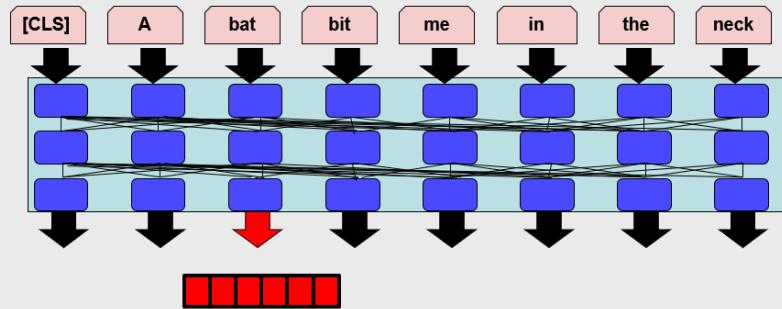
- Make an image based on a text description
  - i.e. draw a cat in a blue hat tap dancing on top of the Burj Khalifa



- Solution: **Transformers**

# Transformers

- Transformers let us turn text into a vector
- This vector “understands” the text
- Understanding = geometry
  - Similar things are close together



# Generating Images From Text

- Input text and output an image
- We have a way to map text to a vector
- Can we “un-map” the vector to an image?

A brown dog



# DALL-E Paper (2021)

## Zero-Shot Text-to-Image Generation

Aditya Ramesh<sup>1</sup> Mikhail Pavlov<sup>1</sup> Gabriel Goh<sup>1</sup> Scott Gray<sup>1</sup>  
Chelsea Voss<sup>1</sup> Alec Radford<sup>1</sup> Mark Chen<sup>1</sup> Ilya Sutskever<sup>1</sup>

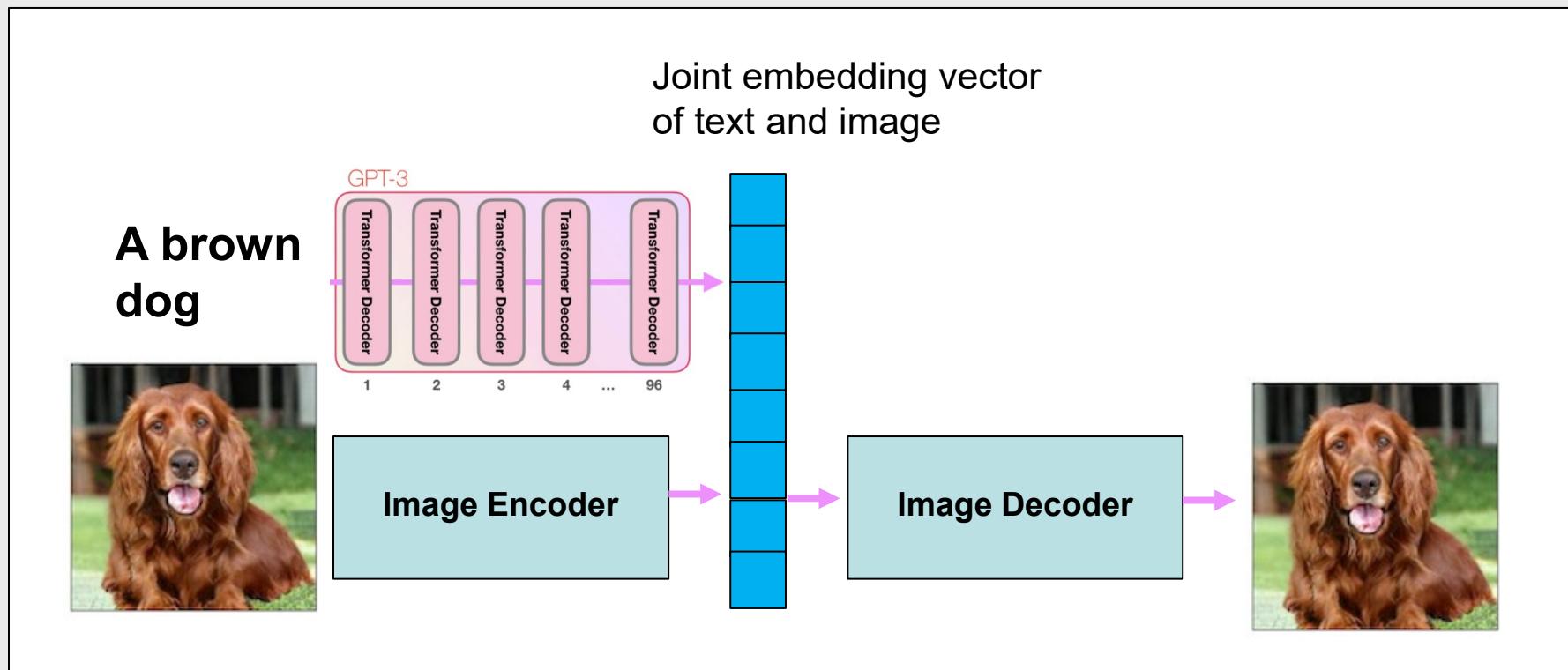
### Abstract

Text-to-image generation has traditionally focused on finding better modeling assumptions for training on a fixed dataset. These assumptions might involve complex architectures, auxiliary losses, or side information such as object part labels or segmentation masks supplied during training. We describe a simple approach for this task based on a transformer that autoregressively models the text and image tokens as a single stream of data. With sufficient data and scale, our approach is competitive with previous domain-specific models when evaluated in a zero-shot fashion.



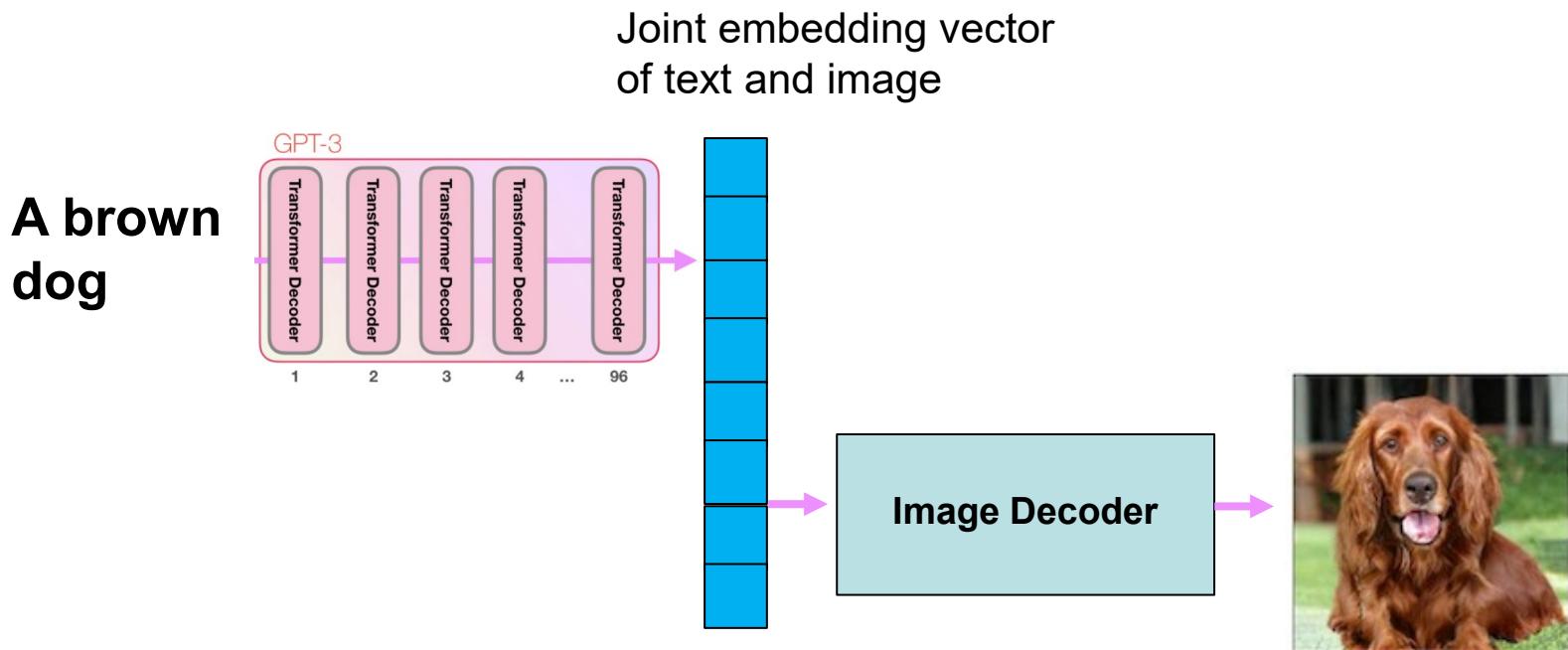
# DALL-E (2021)

- DALL-E is the GPT-3 transformer, but trained to jointly encode images + captions
- Text is encoded as a sequence of tokens (words)
- Image is also encoded as a stream of tokens (image patterns)

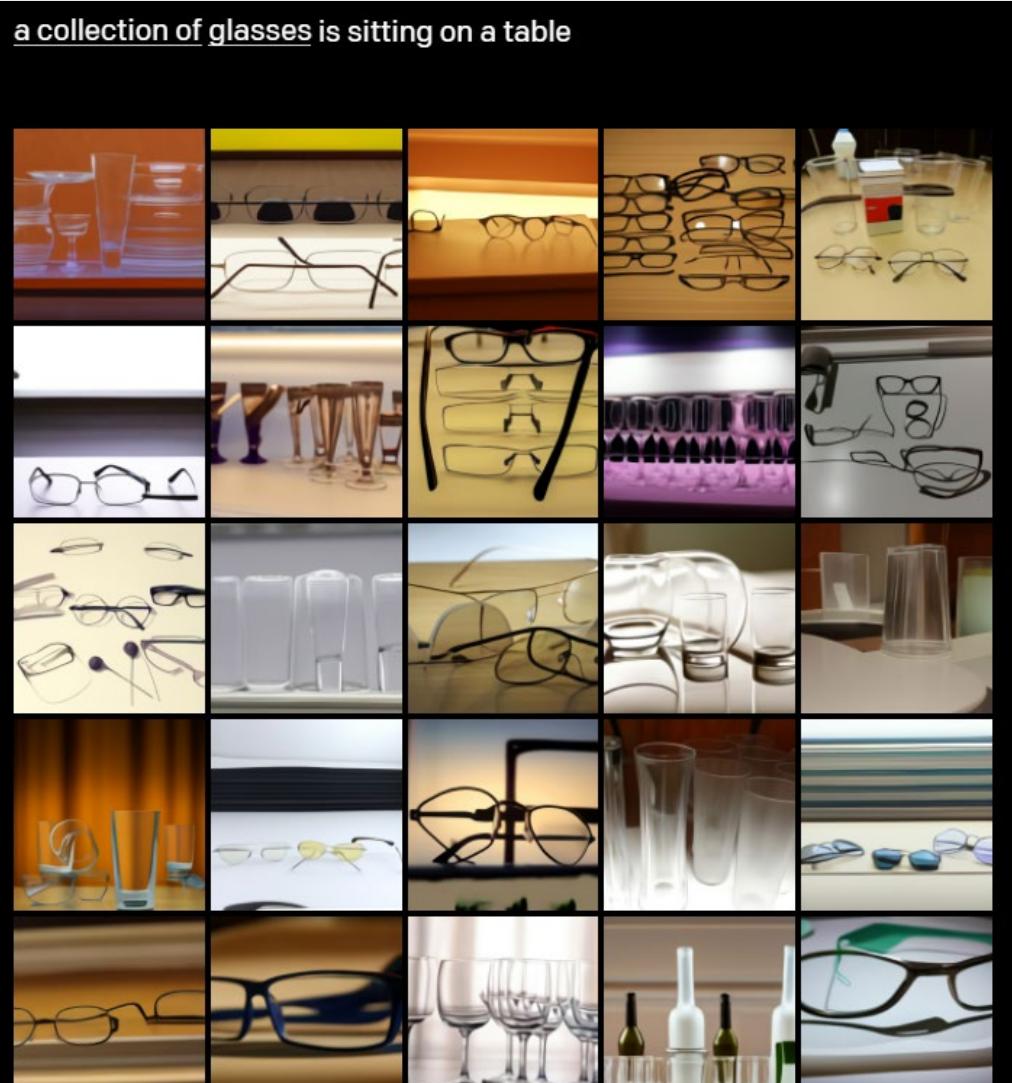


# DALL-E Image Generation

- Image generation needs only the text description
- DALL-E “finishes” the sentence with image tokens



# DALL-E Examples



# DALL-E Examples

an emoji of a baby penguin wearing a blue hat, red gloves, green shirt, and yellow pants



# Pros and Cons of Transformer Image Generation

- **Pros**
  - We can generate images with text
- **Cons**
  - Image quality is not the best



# **Image Generation Challenges v3**

# **Image Generation Challenges v3**

- How can I make an image look really really real?

# Image Generation Challenges v3

- How can I make an image look really really real?



# Image Generation Challenges v3

- How can I make an image look really really real?



- Solution: Latent Diffusion Models

# Diffusion

- Diffusion is the random motion of particles



# Diffusion

- Diffusion is the random motion of particles
- Diffusing particles have a **deterministic drift velocity**



# Diffusion Paper (2015)

---

## Deep Unsupervised Learning using Nonequilibrium Thermodynamics

---

**Jascha Sohl-Dickstein**

JASCHA@STANFORD.EDU

Stanford University

**Eric A. Weiss**

EAWEISS@BERKELEY.EDU

University of California, Berkeley

**Niru Maheswaranathan**

NIRUM@STANFORD.EDU

Stanford University

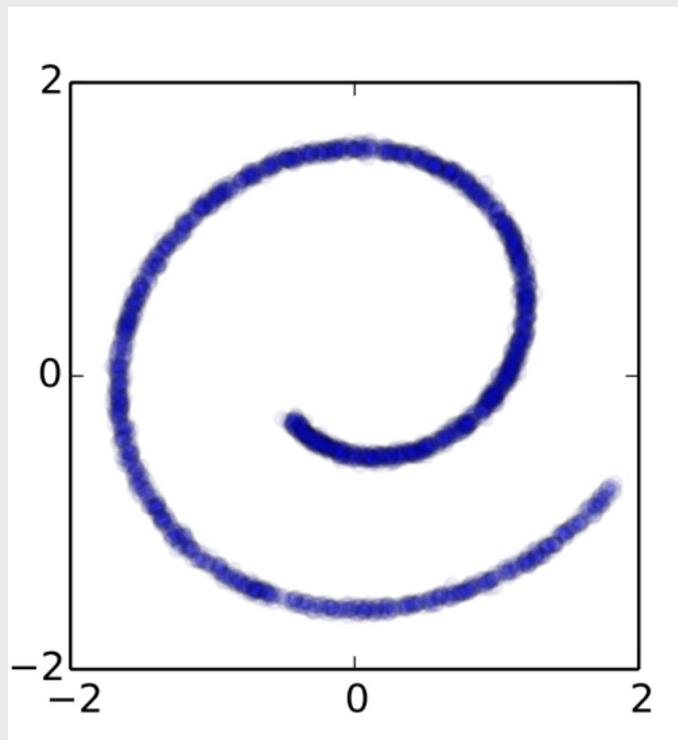
**Surya Ganguli**

SGANGULI@STANFORD.EDU

Stanford University

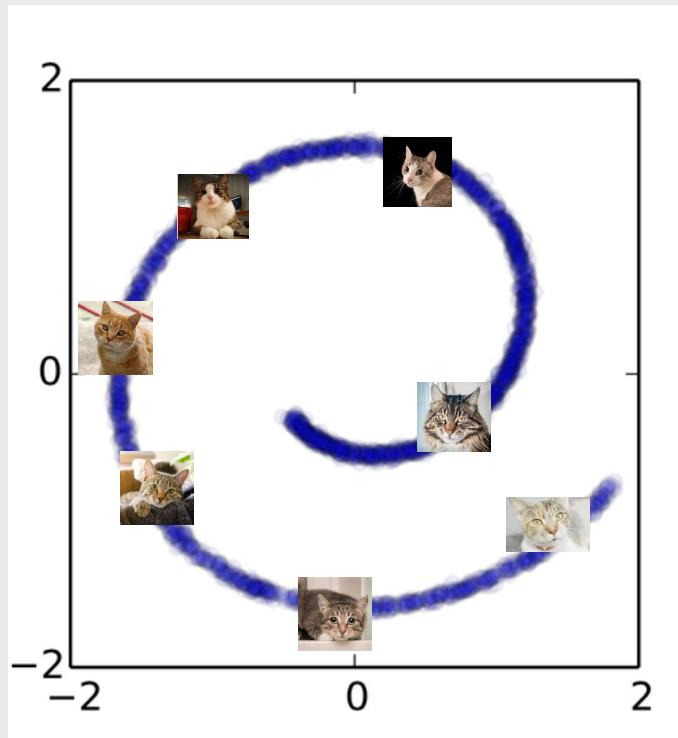
# Diffusion Example

- Imagine our data points start off in a specific shape



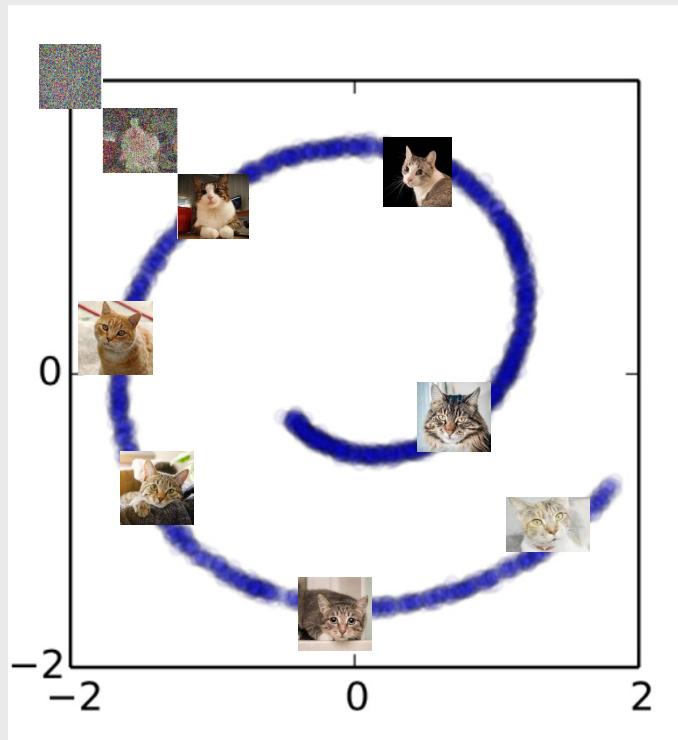
# Diffusion Example

- Imagine our data points start off in a specific shape



# Diffusion Example

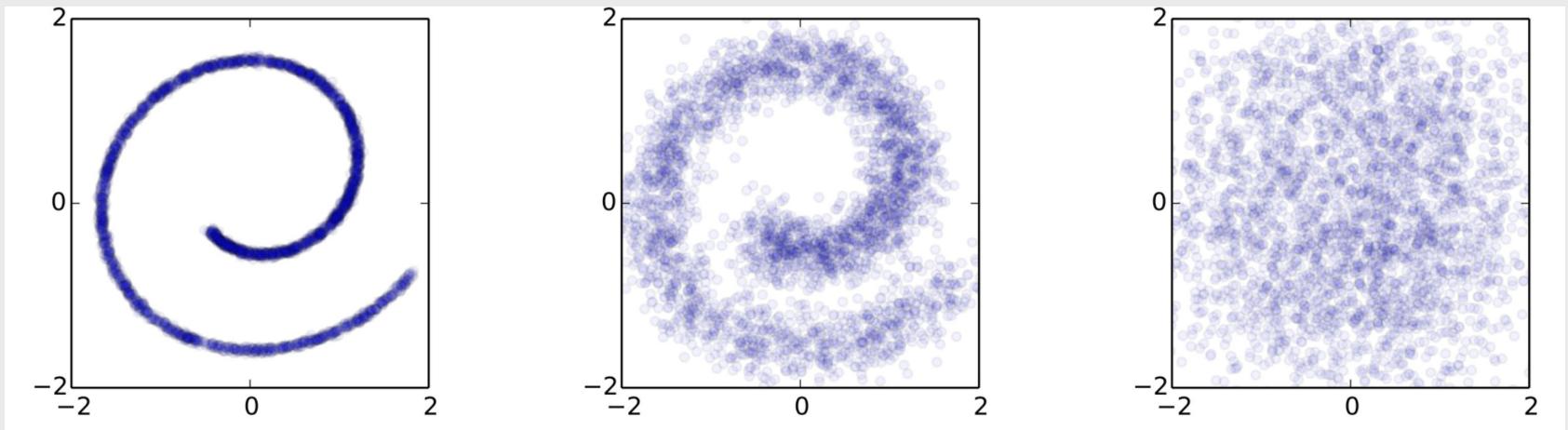
- Imagine our data points start off in a specific shape



# Forward Diffusion

- In the forward direction, the drift velocity is zero always
- We end up with a uniform cloud of points

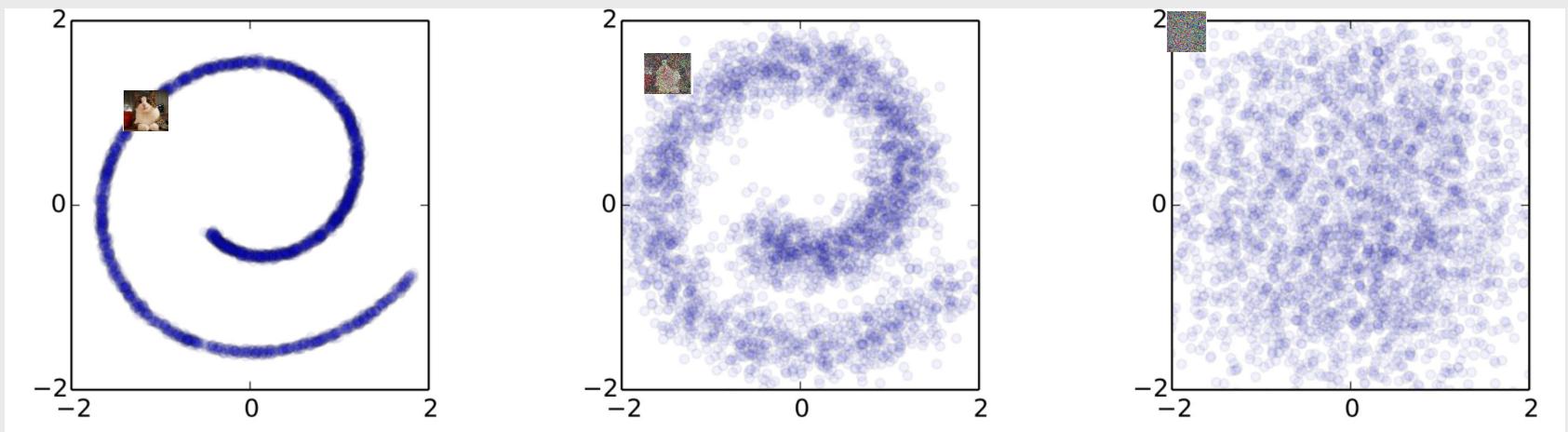
Forward in time



# Forward Diffusion

- In the forward direction, the drift velocity is zero always
- We end up with a uniform cloud of points

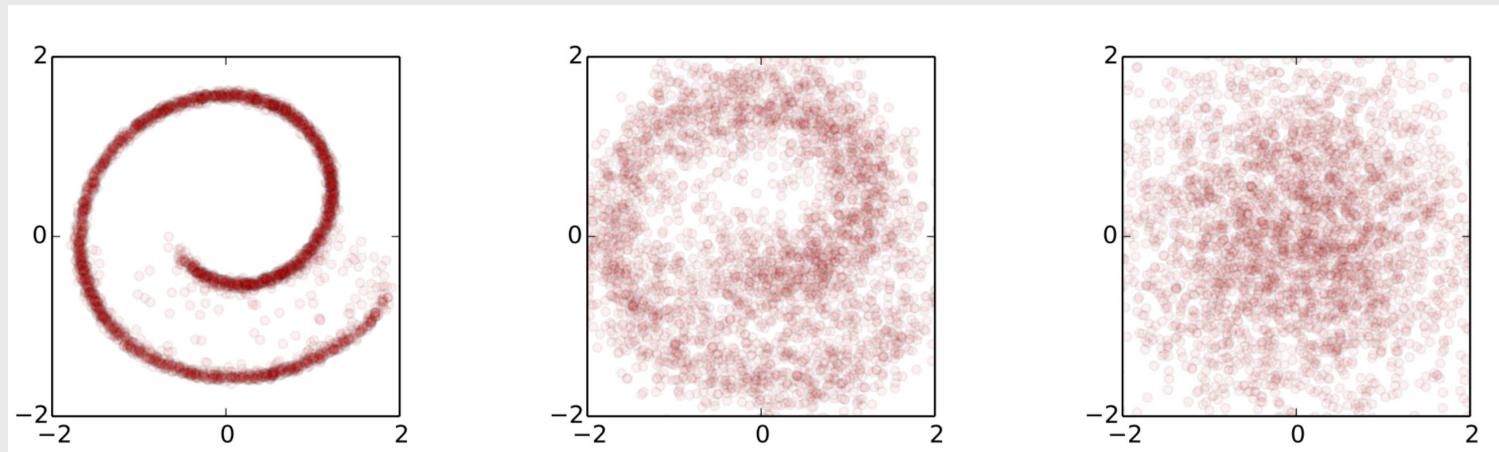
Forward in time →



# Backward Diffusion

- In the backward direction, the **drift velocity** is depends on the **time** and **location** of the data point
- We end up with the original data point configuration

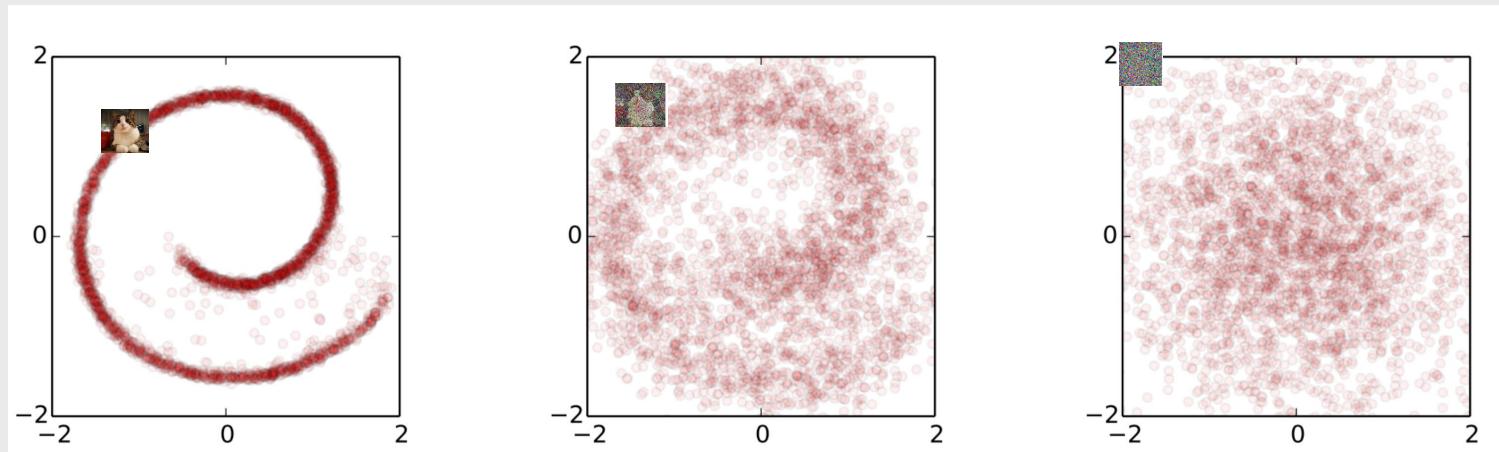
← Backward in time



# Backward Diffusion

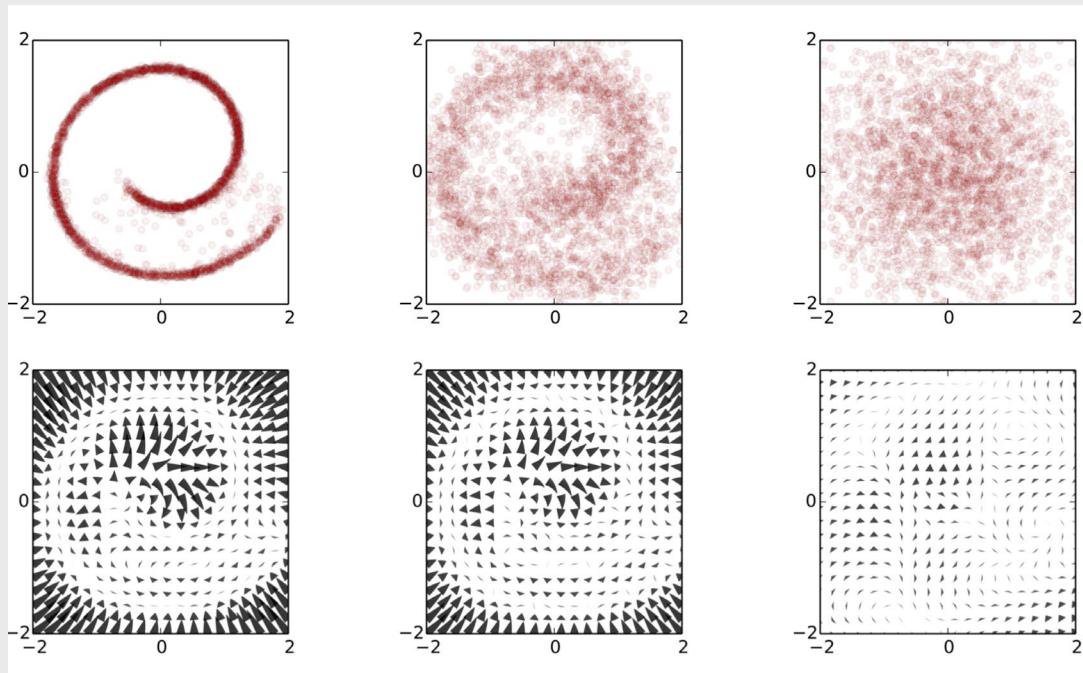
- In the backward direction, the **drift velocity** is depends on the **time** and **location** of the data point
- We end up with the original data point configuration

Backward in time



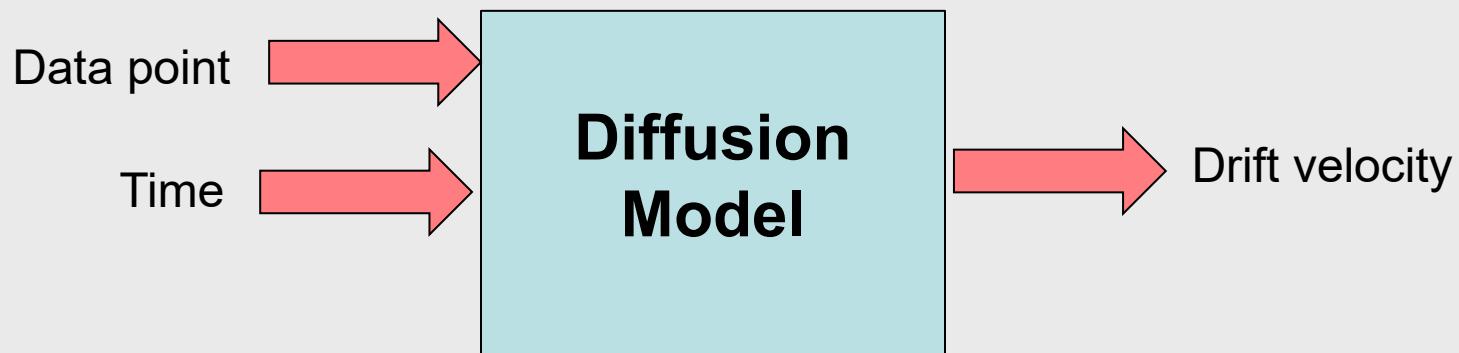
# Backward Drift Velocity

- Initially, backward drift velocity is almost zero – no rush to go anywhere
- Near the end, the velocity can become very strong – no data points in certain regions



# Latent Diffusion Models

- The backward drift velocity is the key to a latent diffusion model
- We can use a **neural network** to model the drift velocity

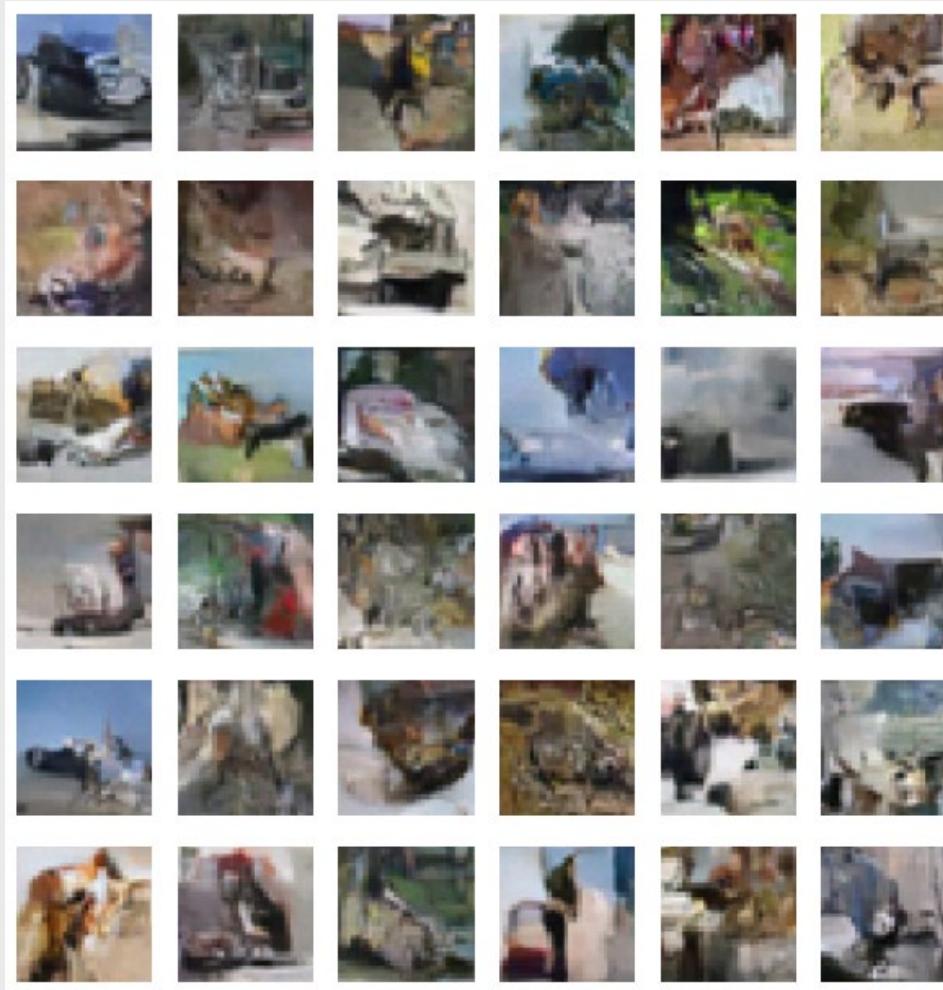


# **Diffusion Model Image Generation**

- Data points can be images
- The diffusion model will drift pixel colors in noisy image until they look like something from your training data

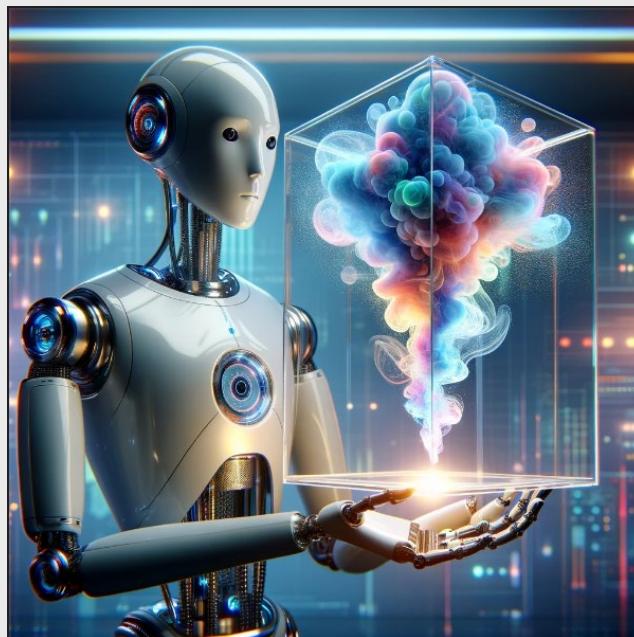


# First Diffusion Model Generated Images (2015)



# Diffusion Stalls

- Diffusion and GANs appeared at around the same time (2014-2015)
- GANs got all the attention and diffusion models did not make much progress



# Diffusion Models Beat GANs

## Paper (2021)

---

### Diffusion Models Beat GANs on Image Synthesis

---

**Prafulla Dhariwal\***

OpenAI

[prafulla@openai.com](mailto:prafulla@openai.com)

**Alex Nichol\***

OpenAI

[alex@openai.com](mailto:alex@openai.com)

### Abstract

We show that diffusion models can achieve image sample quality superior to the current state-of-the-art generative models. We achieve this on unconditional image synthesis by finding a better architecture through a series of ablations. For conditional image synthesis, we further improve sample quality with classifier guidance: a simple, compute-efficient method for trading off diversity for fidelity using gradients from a classifier. We achieve an FID of 2.97 on ImageNet  $128 \times 128$ , 4.59 on ImageNet  $256 \times 256$ , and 7.72 on ImageNet  $512 \times 512$ , and we match BigGAN-deep even with as few as 25 forward passes per sample, all while maintaining better coverage of the distribution. Finally, we find that classifier guidance combines well with upsampling diffusion models, further improving FID to 3.94 on ImageNet  $256 \times 256$  and 3.85 on ImageNet  $512 \times 512$ . We release our code at <https://github.com/openai/guided-diffusion>.

# Diffusion Models Beat GANs

BigGAN-deep

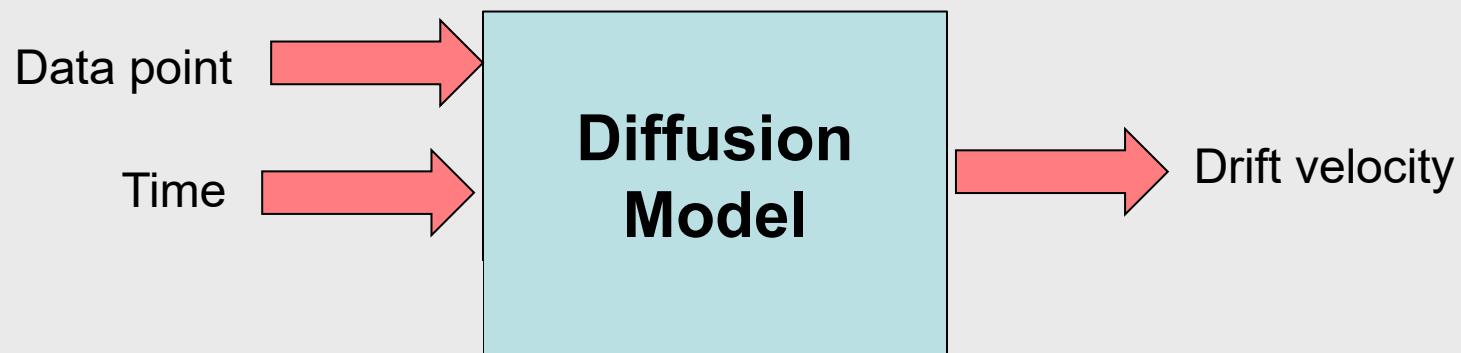


Diffusion Model



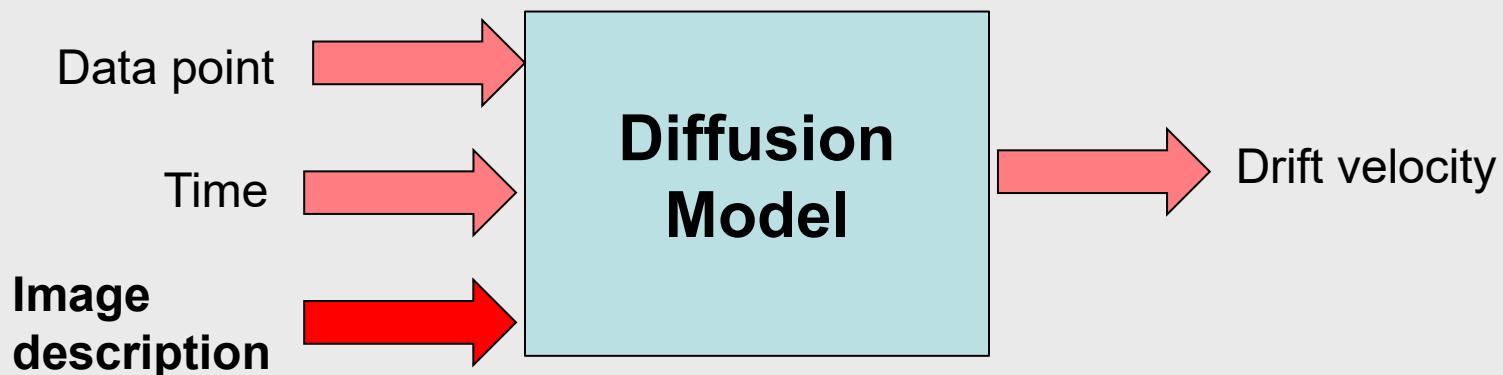
# Latent Diffusion Models with Text

- Diffusion model gives us drift velocity



# Latent Diffusion Models with Text

- Diffusion model gives us drift velocity
- We can use a text image description to guide the diffusion



# GLIDE (2022)

- **GLIDE - Guided Language to Image Diffusion for Generation and Editing**

---

## **GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models**

---

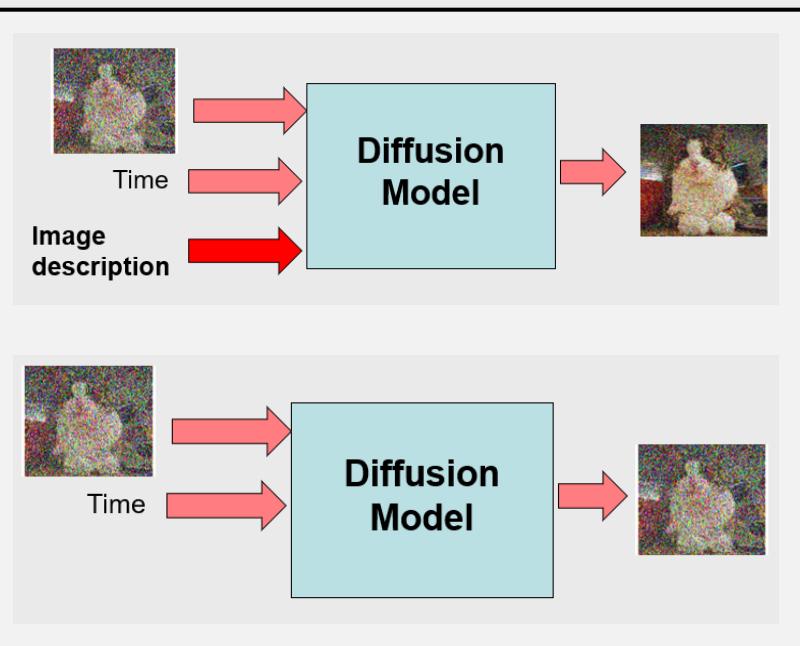
Alex Nichol \* Prafulla Dhariwal \* Aditya Ramesh \* Pranav Shyam Pamela Mishkin Bob McGrew  
Ilya Sutskever Mark Chen

# **GLIDE Image Generation**

- **Glides uses a clever trick to incorporate the text into the image generation**
  - Combine the image without the prompt and with the prompt

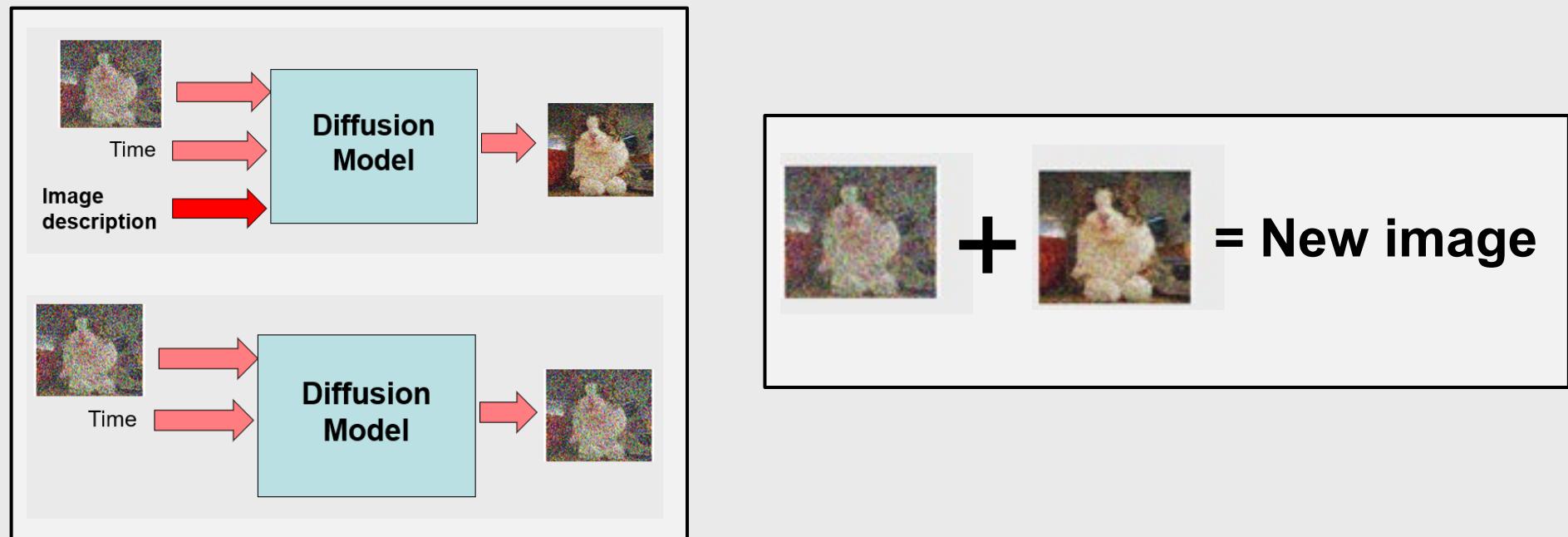
# GLIDE Image Generation

- Glides uses a clever trick to incorporate the text into the image generation
  - Combine the image without the prompt and with the prompt



# GLIDE Image Generation

- Glides uses a clever trick to incorporate the text into the image generation
  - Combine the image without the prompt and with the prompt



# GLIDE Image Editing

- **GLIDE allows one to edit images with a mask and text**
- **Training process shows AI images + masked images**



“a corgi wearing a bow tie and a birthday hat”



“a fire in the background”



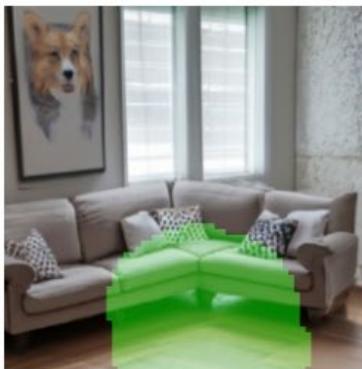
“only one cloud in the sky today”

# GLIDE Image Editing

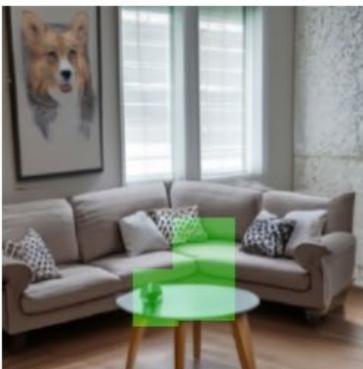
- GLIDE can sequentially modify an image to make a complex scene



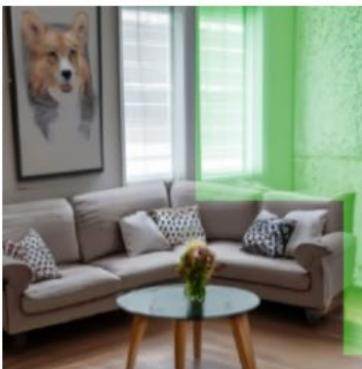
“a cozy living room”



“a painting of a corgi  
on the wall above  
a couch”



“a round coffee table  
in front of a couch”



“a vase of flowers on a  
coffee table”



“a couch in the corner  
of a room”

# DALLE-2 (2022)

- DALLE-2 was developed around the same time as GLIDE
- Puts text and image embeddings in the same space

---

## Hierarchical Text-Conditional Image Generation with CLIP Latents

---

**Aditya Ramesh\***

OpenAI

aramesh@openai.com

**Prafulla Dhariwal\***

OpenAI

prafulla@openai.com

**Alex Nichol\***

OpenAI

alex@openai.com

**Casey Chu\***

OpenAI

casey@openai.com

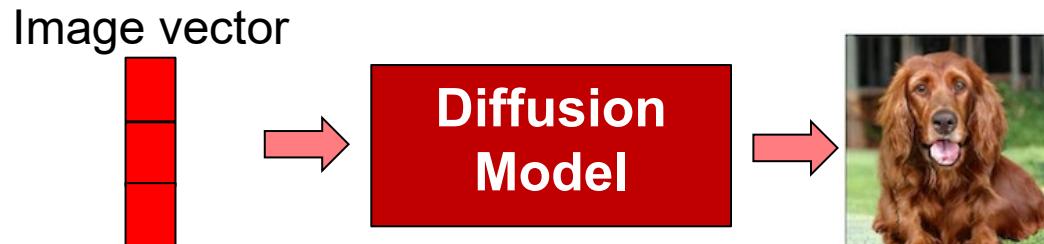
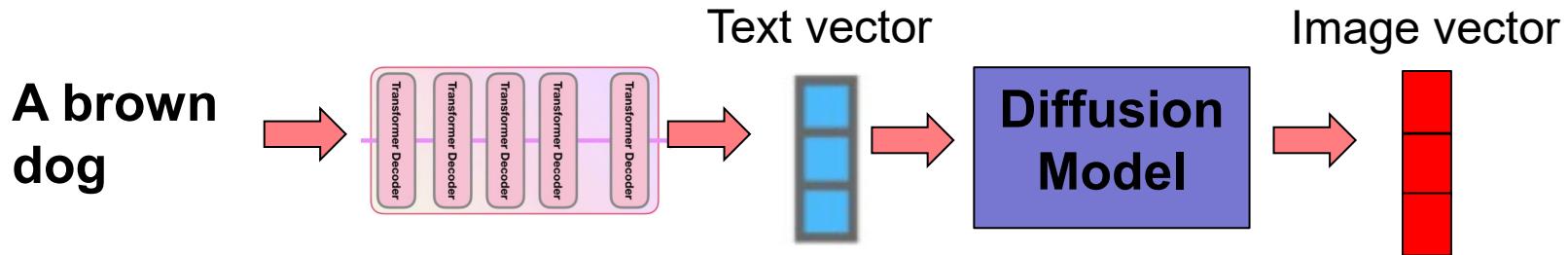
**Mark Chen**

OpenAI

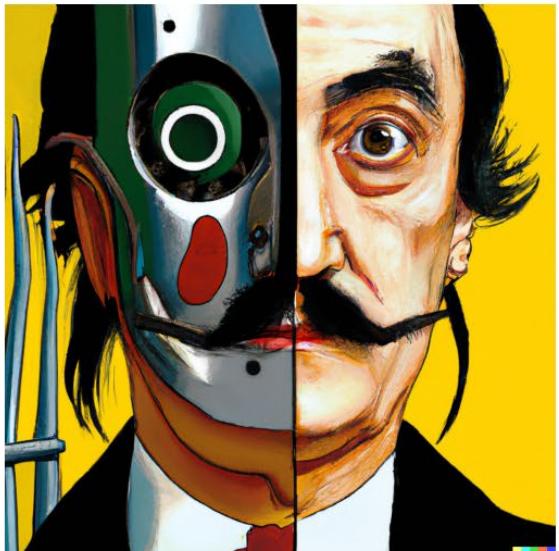
mark@openai.com

# DALL-E 2 Architecture

- DALL-E 2 uses a diffusion model to turn a text vector into an image vector
- It then uses another diffusion model to decode the image vector into an image



# DALLE-2 Examples



vibrant portrait painting of Salvador Dalí with a robotic half face



a shiba inu wearing a beret and black turtleneck



a close up of a handpalm with leaves growing from it

# Diffusion Model Pros and Cons

- **Pros**
  - Text guided image generation
  - High quality images
- **Cons**
  - Slow image generation
  - Can't maintain consistency across images



# AI Image Generators

- Diffusion models + transformers are now used by nearly all AI image generators
- Many AI image generators available today
  - DALLE-3 (ChatGPT) from OpenAI
  - Midjourney from Midjourney
  - Stable Diffusion from Stability AI
  - Firefly from Adobe
  - Imagen (Gemini) from Google
  - And many more ...



# AI Image Generators

- Prompt: A Yale men's basketball player jumping from the free-throw line for a dunk during the championship game

# AI Image Generators

- Prompt: A Yale men's basketball player jumping from the free-throw line for a dunk during the championship game



Firefly



Stable Diffusion



Midjourney



ChatGPT

# AI Image Generators

- Prompt: A Yale men's basketball player jumping from the free-throw line for a dunk during the championship game



Firefly



Stable Diffusion



Midjourney



ChatGPT

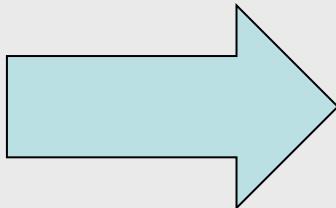
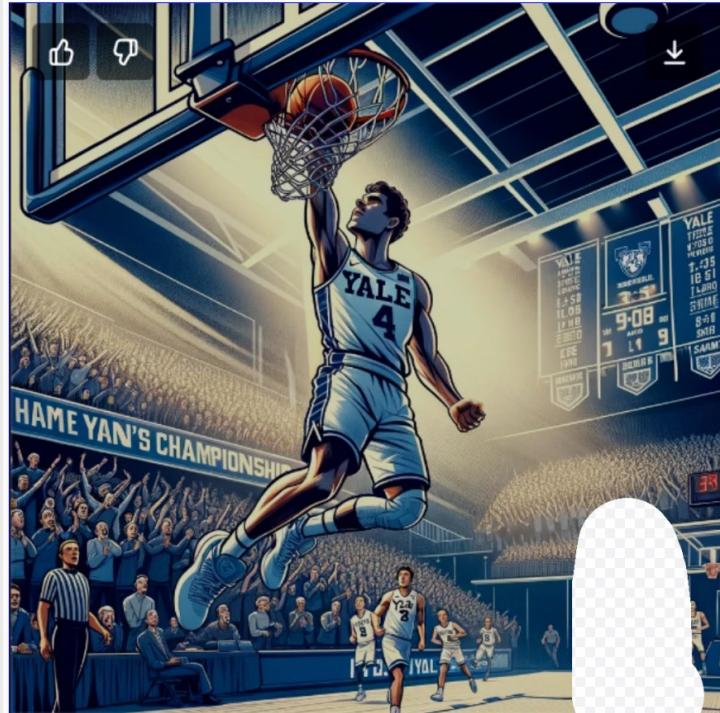
◆ We are working to improve Gemini's ability to generate images of people. We expect this feature to return soon and will notify you in release updates when it does.



Gemini

# AI Image Editing

- Many generative AI models allow for image editing

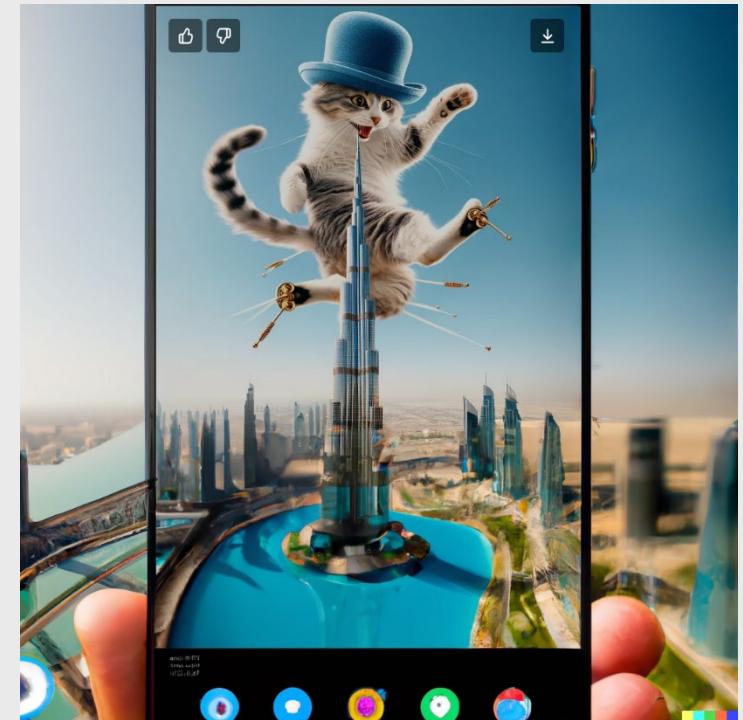
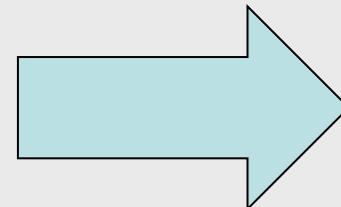
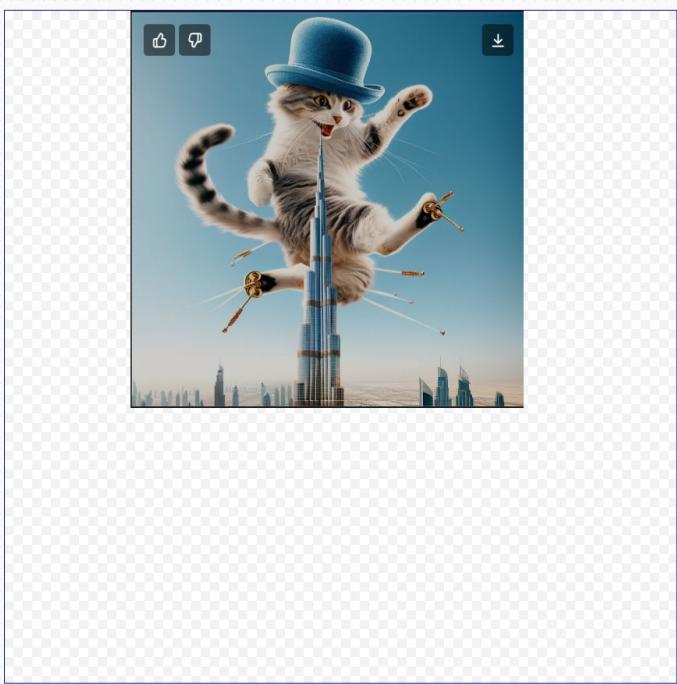


A bulldog

# AI Image Editing

- Many AI models allow for image editing

Generation frame: 1024 x 1024



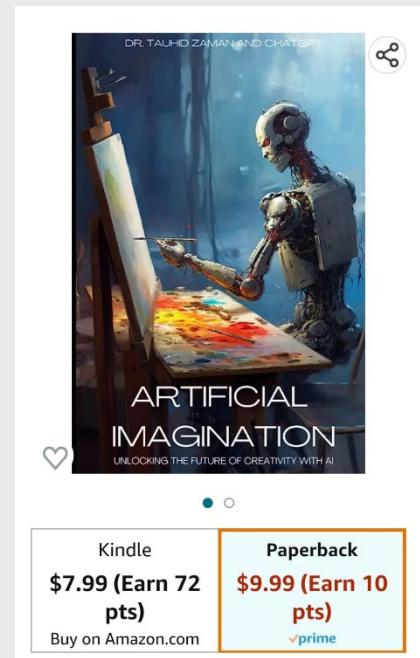
the burj khalifa  
tower

# New Generative AI Landscape

- **Text guided image generation opens up a wide array of possibilities**
  - Simple graphic design for any skill level
  - Visualize concepts for artistic projects
  - Create engaging visual content
- **Combined with text generation, we have even more possibilities**

# Writing Books With AI

- One can easily write books by combining text and image generation
- ChatGPT
  - Book title and outline
  - Chapter titles and content
  - Prompts for illustrations
- Midjourney
  - Illustrations



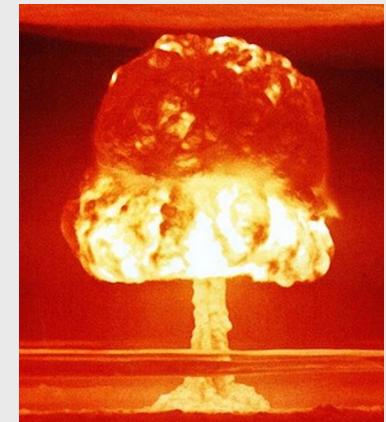
# Dangers of AI Image Generation + Social Networks



+



=



# **Fake Tweets and Images**

- **Use AI to create fake tweets about a topic**
- **Use AI to make fake images matching tweets**

# Fake Tweets and Images

- Use AI to create fake tweets about a topic
- Use AI to make fake images matching tweets

## Fake tweet

A screenshot of a tweet from the Twitter interface. The tweet is from a user named "PatriotPaul" (verified) with the handle "@MAGAdefender45". The tweet content is:  
Just heard the news that our LEGITIMATE 45th President, Donald J. Trump, has been ARRESTED! 😠 This is nothing but a WITCH HUNT by the Deep State and liberal elites! We will NOT stand for this disgraceful attempt to silence our voices! 🇺🇸💪 #FreeTrump  
#MAGA #FightForFreedom

The tweet was posted at 8:06 PM · Mar 15, 2023. It has 658 Retweets, 125 Quote Tweets, and 1.3K Likes. Below the tweet are standard social media interaction icons: a speech bubble, a retweet symbol, a heart, and an upward arrow.

# Fake Tweets and Images

- Use AI to create fake tweets about a topic
- Use AI to make fake images matching tweets

Fake tweet

PatriotPaul 🇺🇸 ✅  
@MAGAdefender45

Just heard the news that our LEGITIMATE 45th President, Donald J. Trump, has been ARRESTED! 😢 This is nothing but a WITCH HUNT by the Deep State and liberal elites! We will NOT stand for this disgraceful attempt to silence our voices! 🇺🇸 💪 #FreeTrump  
#MAGA #FightForFreedom

8:06 PM · Mar 15, 2023

658 Retweets 125 Quote Tweets 1.3K Likes

Reply Share Like Retweet

Fake image



# Midjourney



# Midjourney

**“Joe biden tripping in the Oval office in front of a bunch of reporters”**

# Midjourney



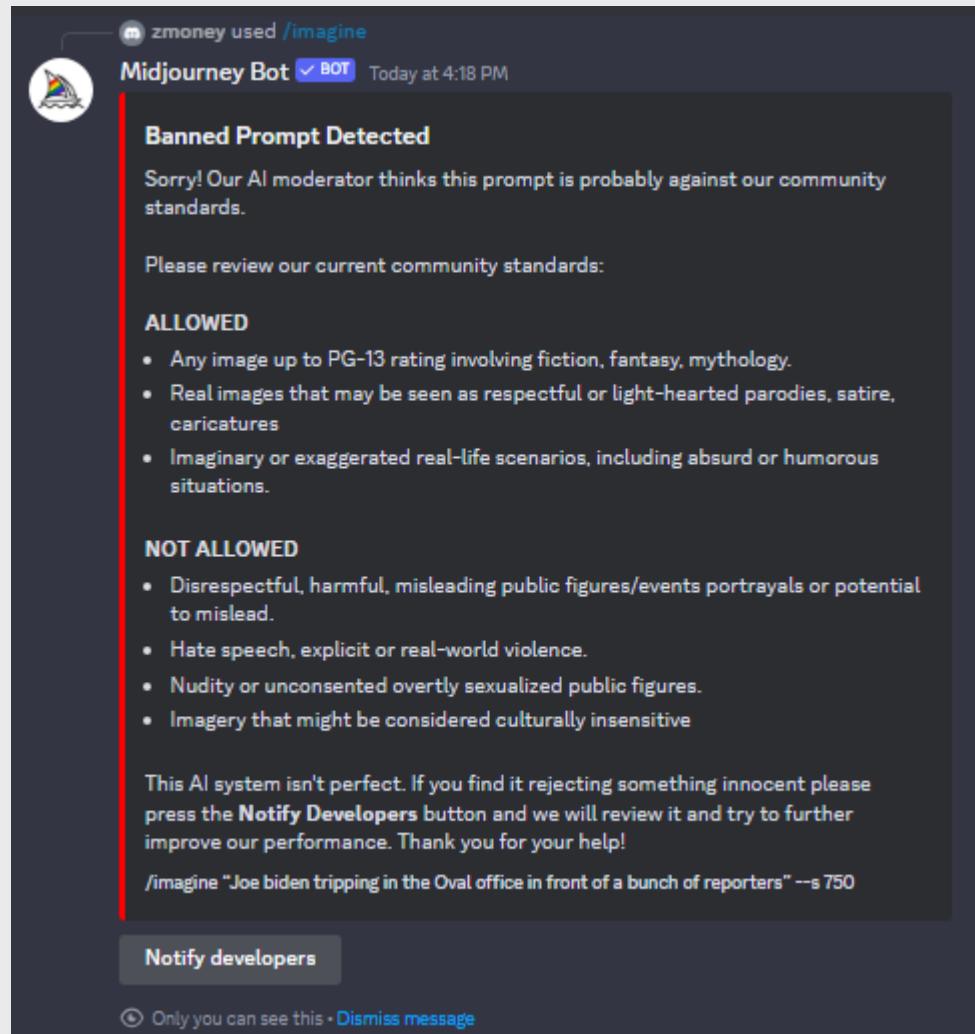
**“Joe biden tripping in the Oval office in front of a bunch of reporters”**

# Midjourney (1 year ago)



"Joe biden tripping in the Oval office in front of a bunch of reporters"

# Midjourney (today)



zmoney used /imagine

Midjourney Bot ✅ BOT Today at 4:18 PM

**Banned Prompt Detected**

Sorry! Our AI moderator thinks this prompt is probably against our community standards.

Please review our current community standards:

**ALLOWED**

- Any image up to PG-13 rating involving fiction, fantasy, mythology.
- Real images that may be seen as respectful or light-hearted parodies, satire, caricatures
- Imaginary or exaggerated real-life scenarios, including absurd or humorous situations.

**NOT ALLOWED**

- Disrespectful, harmful, misleading public figures/events portrayals or potential to mislead.
- Hate speech, explicit or real-world violence.
- Nudity or unconsented overtly sexualized public figures.
- Imagery that might be considered culturally insensitive

This AI system isn't perfect. If you find it rejecting something innocent please press the **Notify Developers** button and we will review it and try to further improve our performance. Thank you for your help!

/imagine "Joe biden tripping in the Oval office in front of a bunch of reporters" --s 750

**Notify developers**

Only you can see this · Dismiss message

"Joe biden tripping in the Oval office in front of a bunch of reporters"

# Deep Fake: Balenciaga Pope

- **Fake image of Pope in Balenciaga jacket**
- **Goes viral on social media**
- **Effective because the fake part is in the realm of reality**



# Dangerous Deep Fakes

## Who Is Behind the Spread of Taylor Swift's Deepfake Nudes?

After an examination into the source of the AI-generated images that caused a firestorm online and ignited a political debate, Microsoft responds to a report that its AI image generator tool was used.

BY KEVIN DOLAK  FEBRUARY 1, 2024 11:38AM



Taylor Swift NOAM GALAI/GETTY IMAGES

# Policing AI Generated Content

## YouTube to roll out labels for ‘realistic’ AI-generated content



By Clare Duffy, CNN

⌚ 3 minute read · Published 6:00 AM EST, Tue November 14, 2023

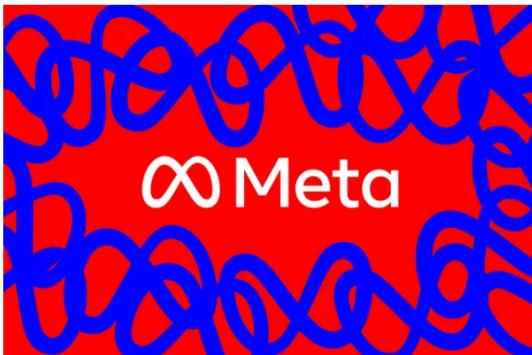
[f](#) [X](#) [✉](#) [🔗](#)

## Twitter tries to crack down on AI images with Community Notes feature

The update comes on the heels of several AI-generated images that gained notoriety — remember Balenciaga pope?

TECH / POLICY / POLITICS

## Meta to require political advertisers disclose AI-generated content



/ The company already banned political advertisers from using its new generative AI ad tools.

By Makena Kelly, a reporter who covers the politics and power influencing the tech industry. Before joining The Verge in 2018, she covered Congress and breaking news.

Nov 8, 2023, 6:00 AM EST

Illustration: Nick Barclay / The Verge

[🔗](#) [f](#) [t](#) [v](#) | [Comments \(0 New\)](#)

## Meta Will Label AI-Generated Images On Instagram, Facebook—Battling Those Who ‘Want To Deceive’

# Policing AI Generated Content

## House floats a new bill for warning labels on AI-generated content

The bill would require AI-generated content to carry disclaimers and digital signatures so viewers know it is machine-made.



Photo by: Richard Drew / AP

---

By: Scripps News Staff

Posted at 8:10 PM, Mar 21, 2024 and last updated 9:10 PM, Mar 21, 2024

# **Don't Be Evil**

- As AI improves, fake text, images, audio, (and soon videos) become more realistic and effective
- Danger is amplified when AI enters social networks
- Please don't use this for evil

# **Coding Session**

- We will use ChatGPT's API to generate images
- Learn some prompt engineering tricks
- Edit images with AI
- Create an AI Instagram account