

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

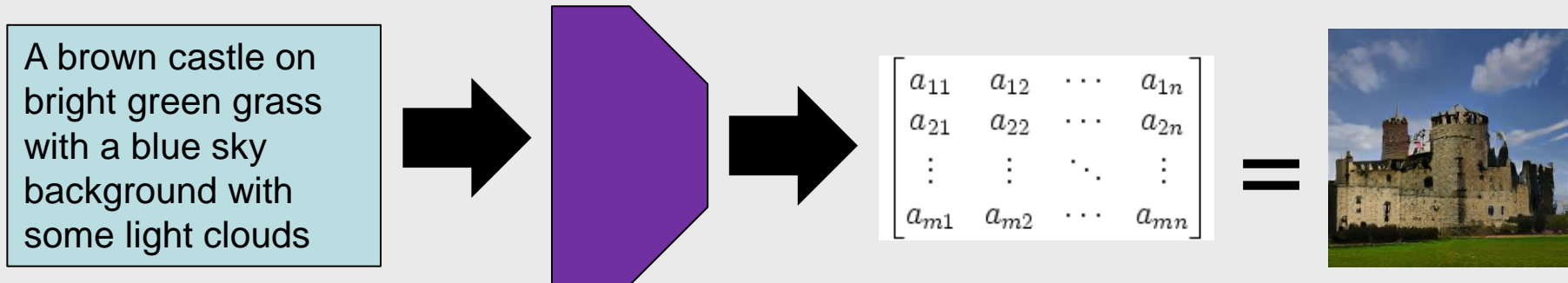


=

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

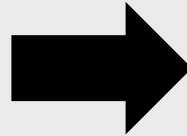
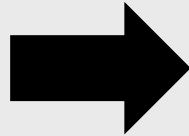
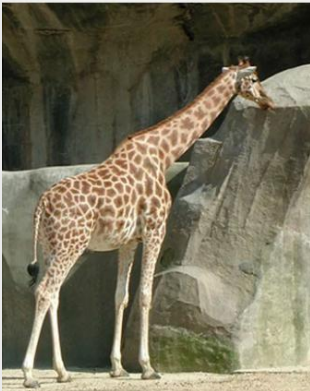
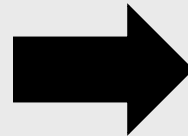
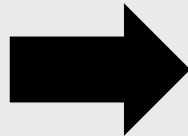
Image Generation

- **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**



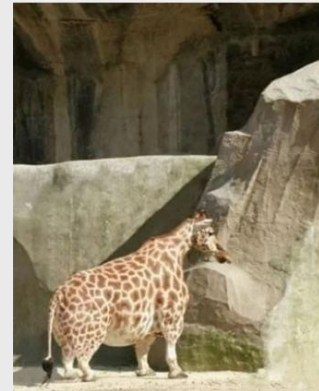
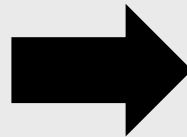
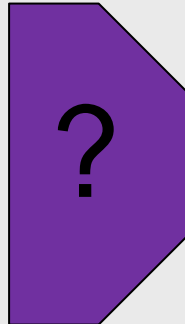
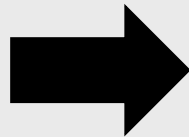
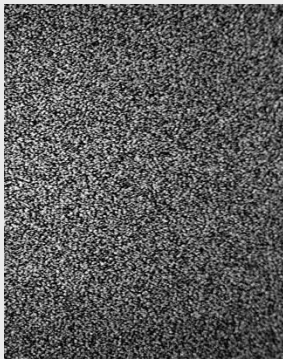
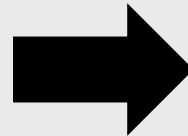
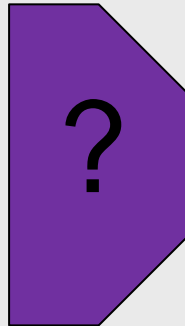
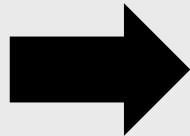
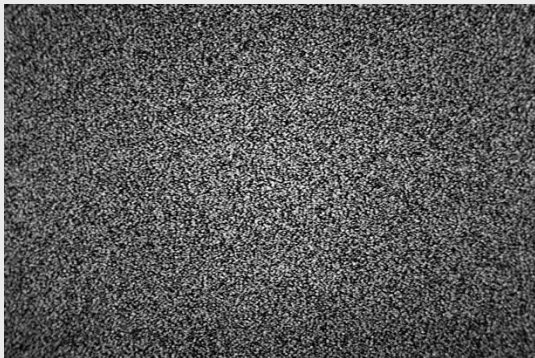
Modifying an Image

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



Generating an Image

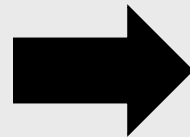
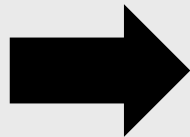
- Creating an image from nothing was thought to be very challenging



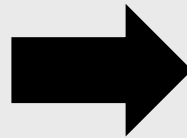
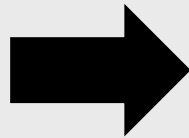
Generating a Specific Image

- 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

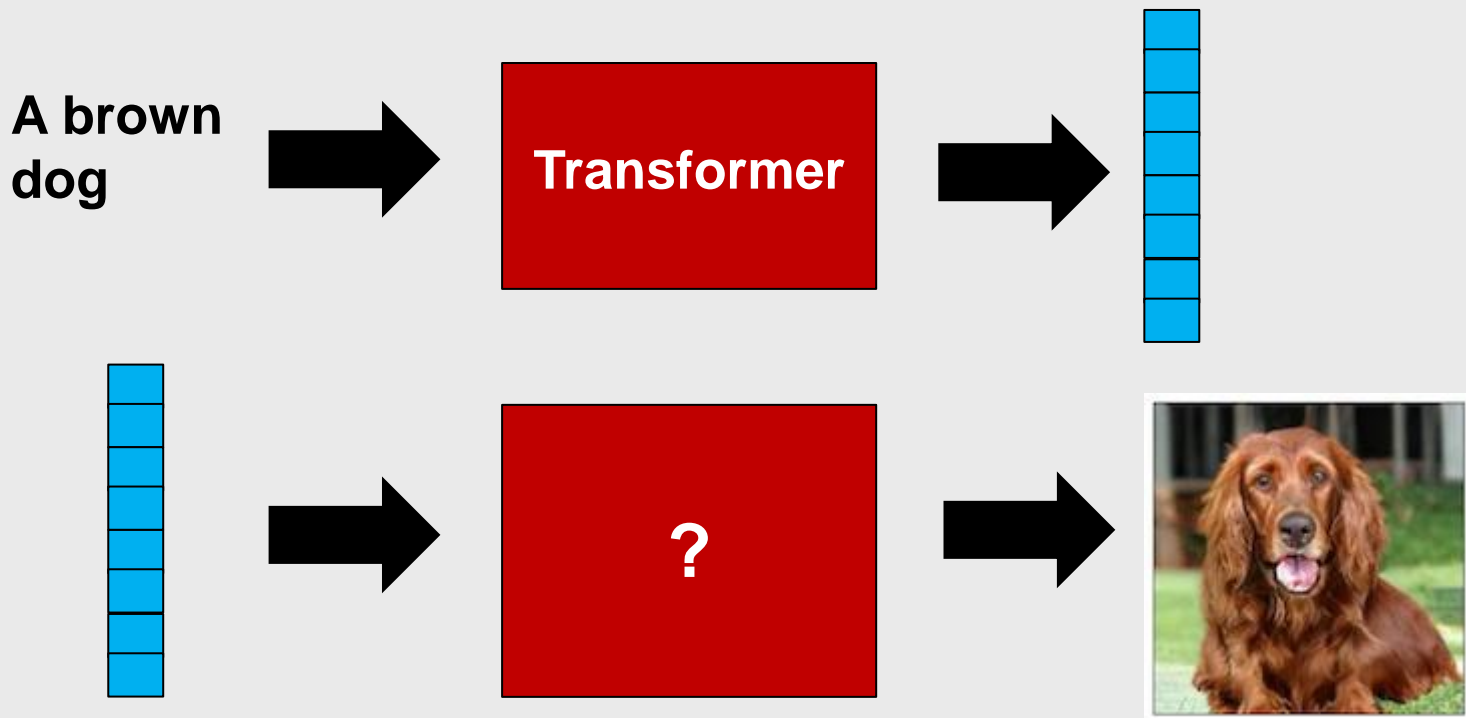


A very short and stout giraffe facing to the right



Generating Images From Text

- We have a way to map text to a vector (transformer)
- How do we map the vector to an image?



Diffusion Models

- Diffusion is the random motion of particles
- Diffusion models create images by letting vectors “diffuse” backwards in time



Diffusion Paper (2015)

Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein

Stanford University

JASCHA@STANFORD.EDU

Eric A. Weiss

University of California, Berkeley

EAWISS@BERKELEY.EDU

Niru Maheswaranathan

Stanford University

NIRUM@STANFORD.EDU

Surya Ganguli

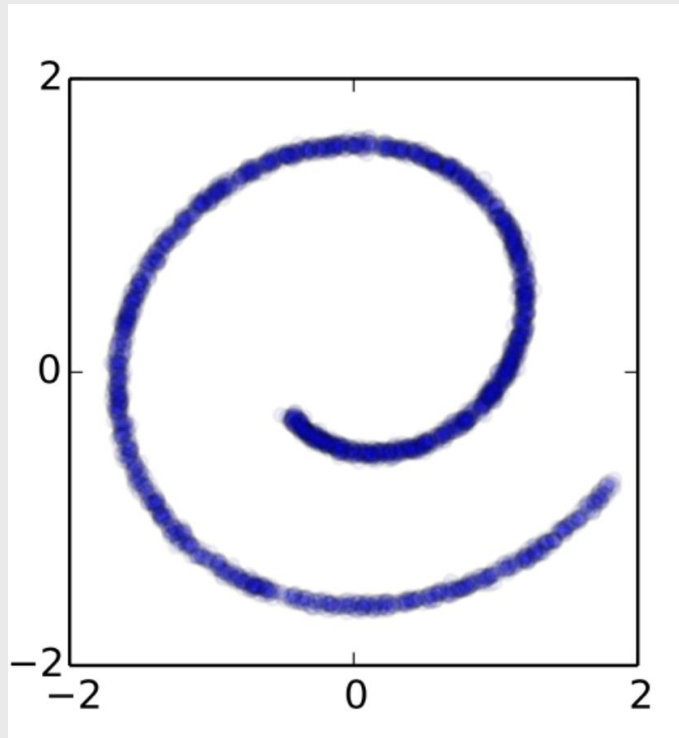
Stanford University

SGANGULI@STANFORD.EDU

18 Nov 2015

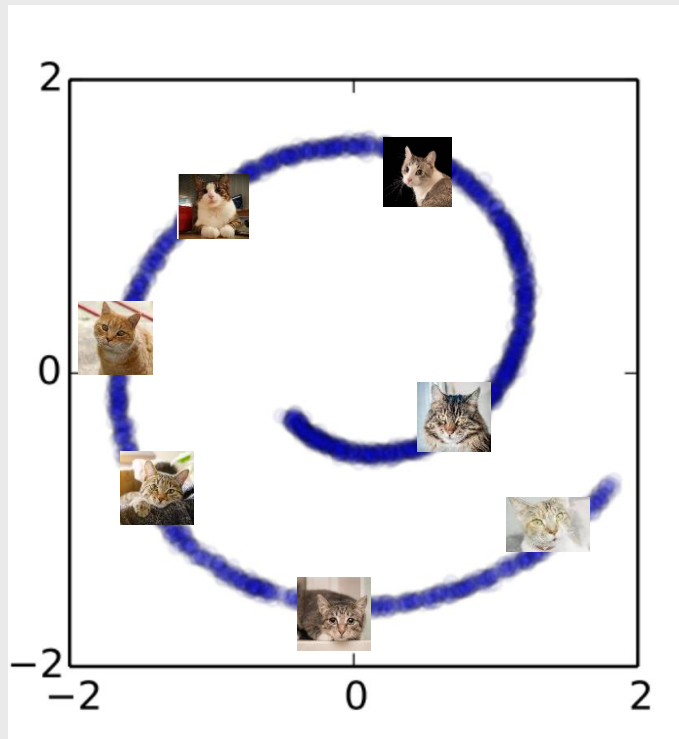
Diffusion Example

- Imagine our data points start off in a specific shape



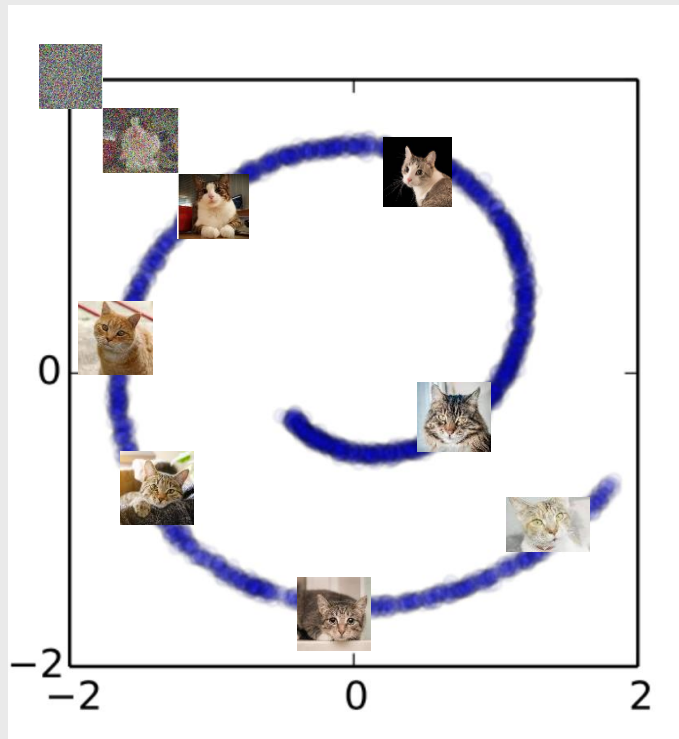
Diffusion Example

- Imagine our data points start off in a specific shape



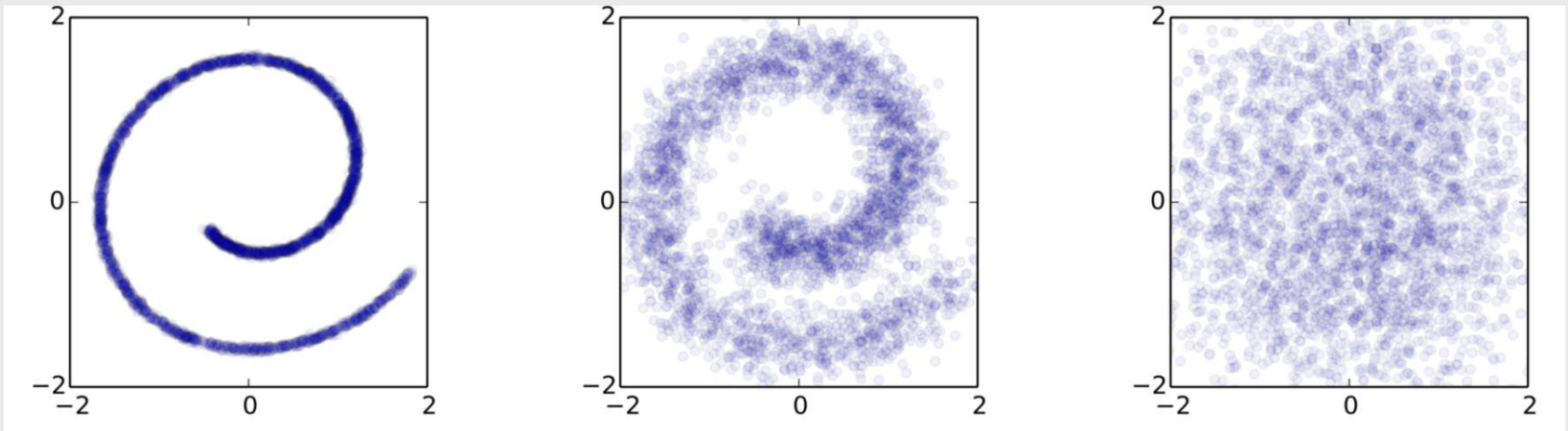
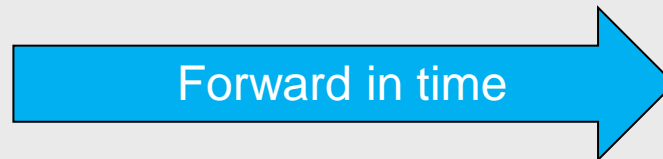
Diffusion Example

- Imagine our data points start off in a specific shape



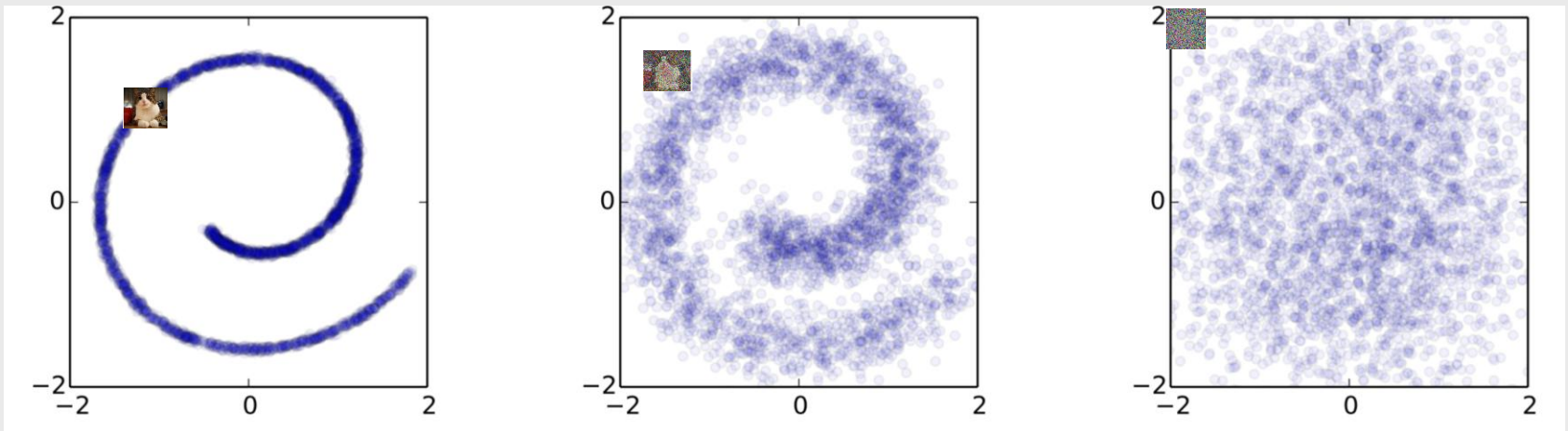
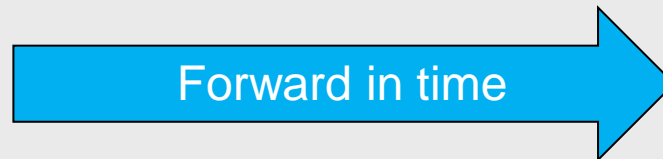
Forward Diffusion

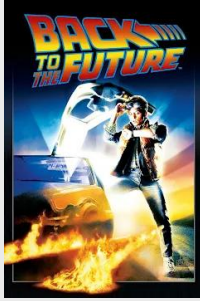
- In the forward direction, the drift velocity is zero always
- We end up with a uniform cloud of points



Forward Diffusion

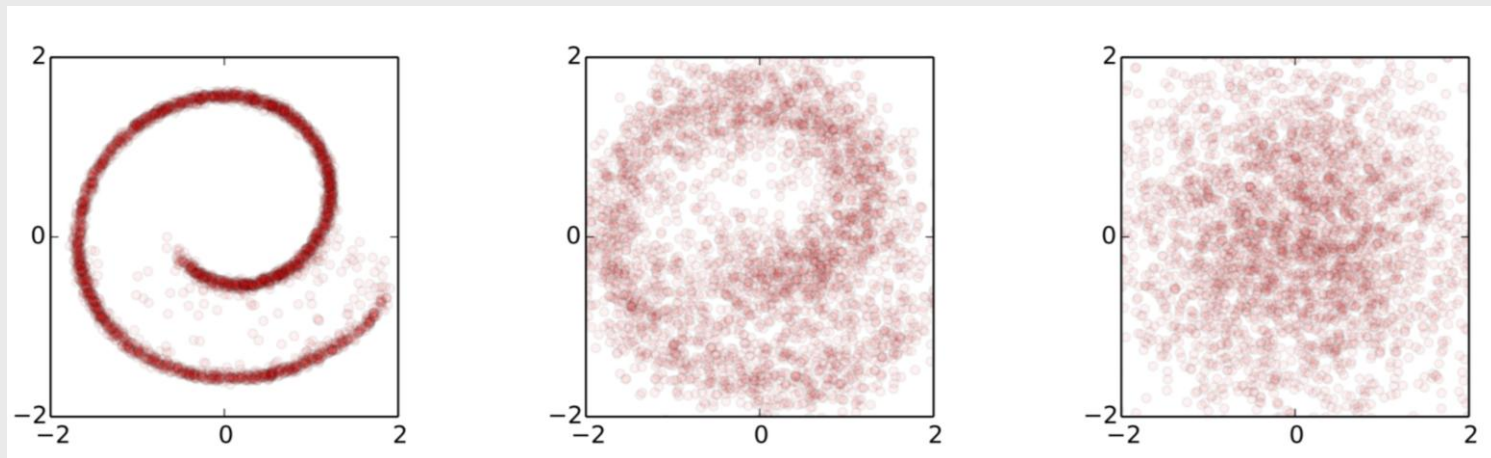
- In the forward direction, the drift velocity is zero always
- We end up with a uniform cloud of points

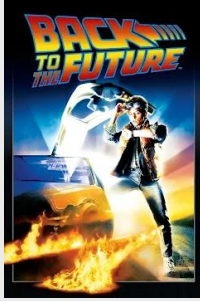




Backward Diffusion

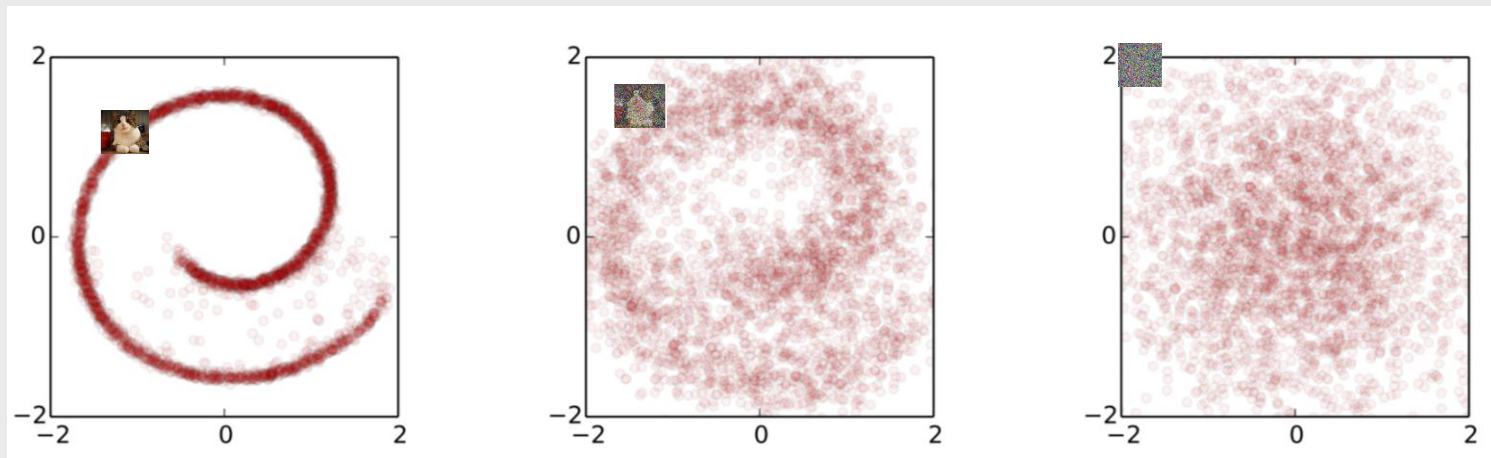
- In the backward direction, the **drift velocity** is depends on the **time** and **location** of the data point
- We end up with the original data point configuration





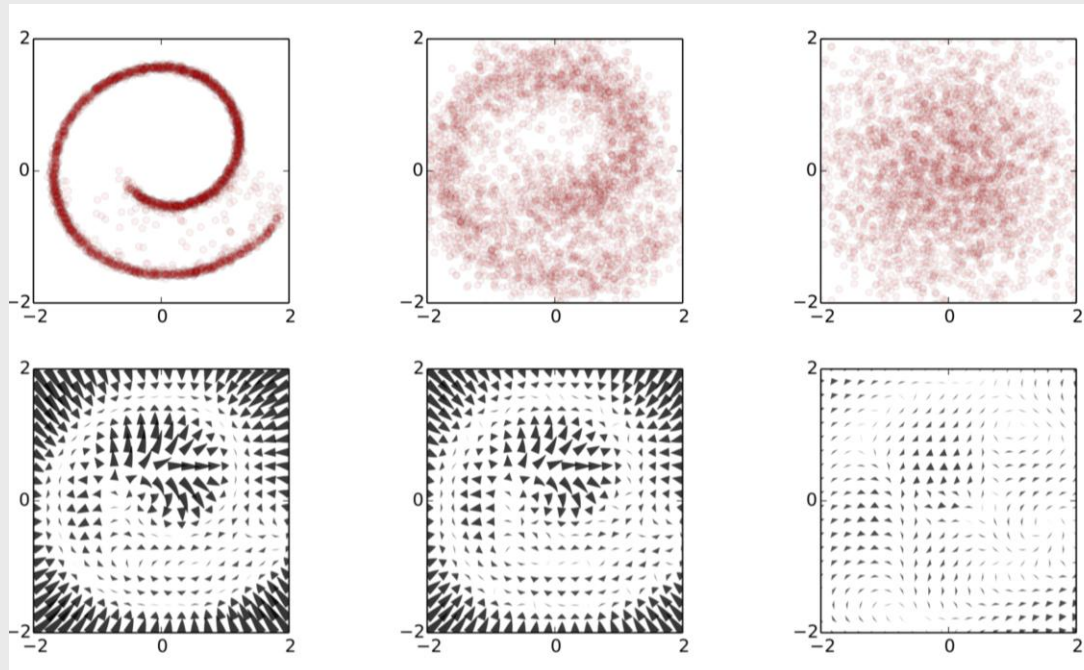
Backward Diffusion

- 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



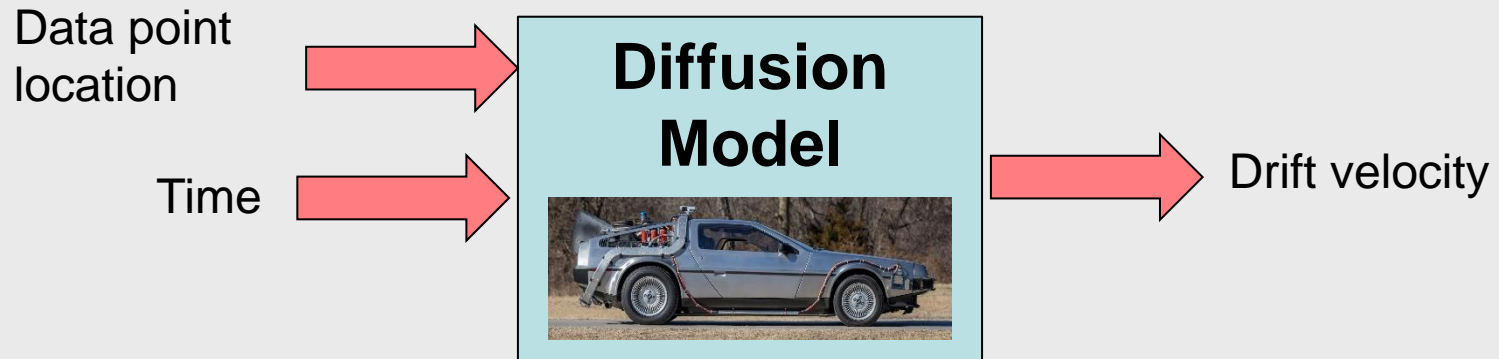
Backward Drift Velocity

- 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



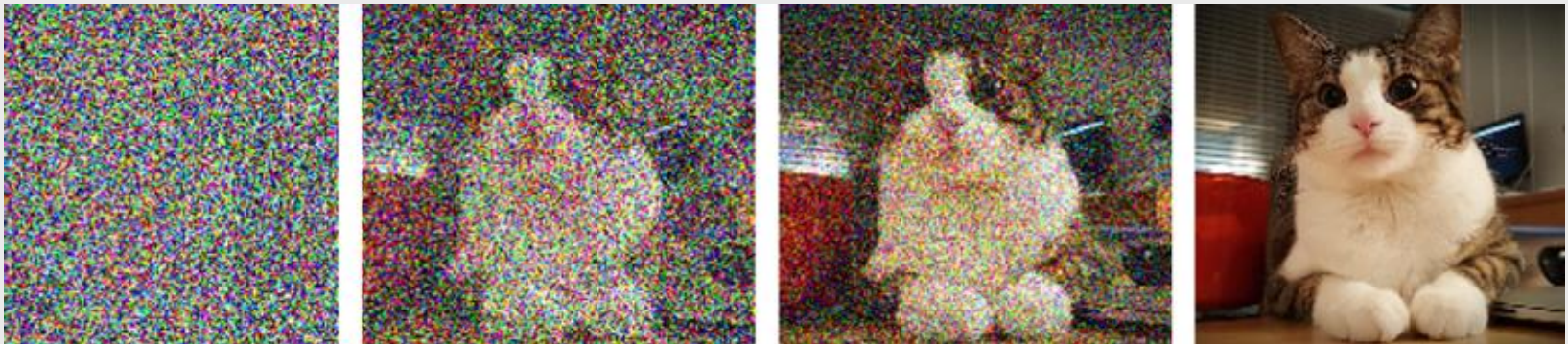
Latent Diffusion Models

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

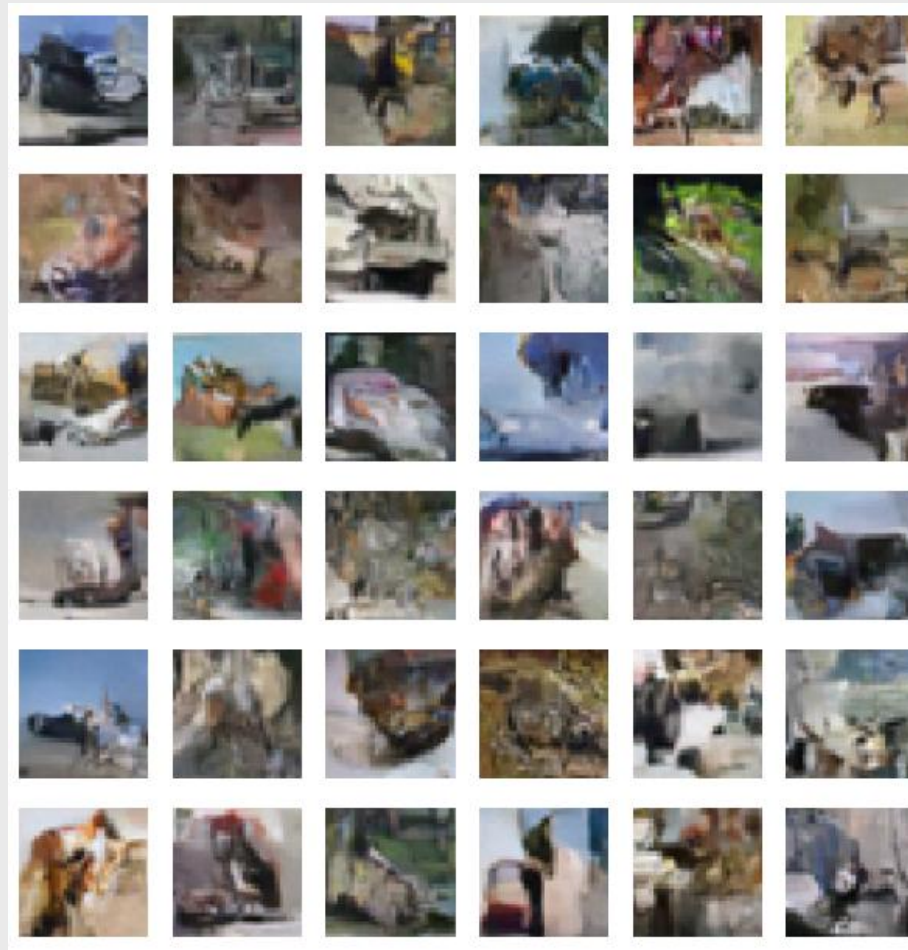


Diffusion Model Image Generation

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

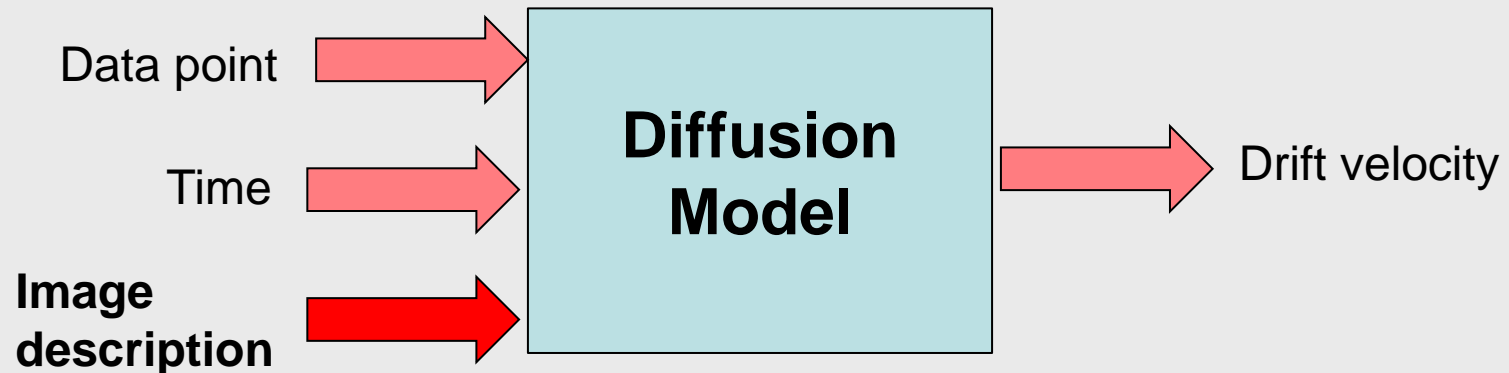


First Diffusion Model Generated Images (2015)



Latent Diffusion Models with Text

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



DALLE-2 (2022)

- DALLE-2 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

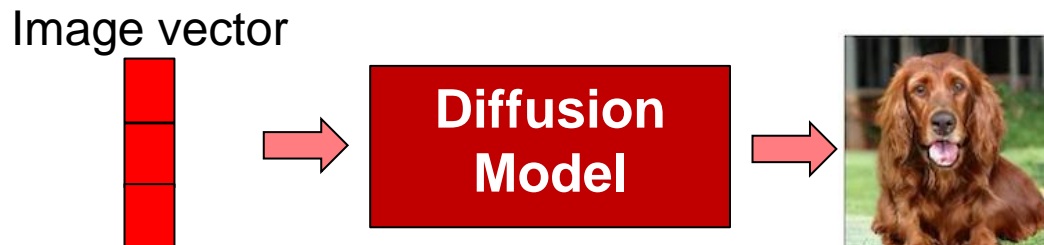
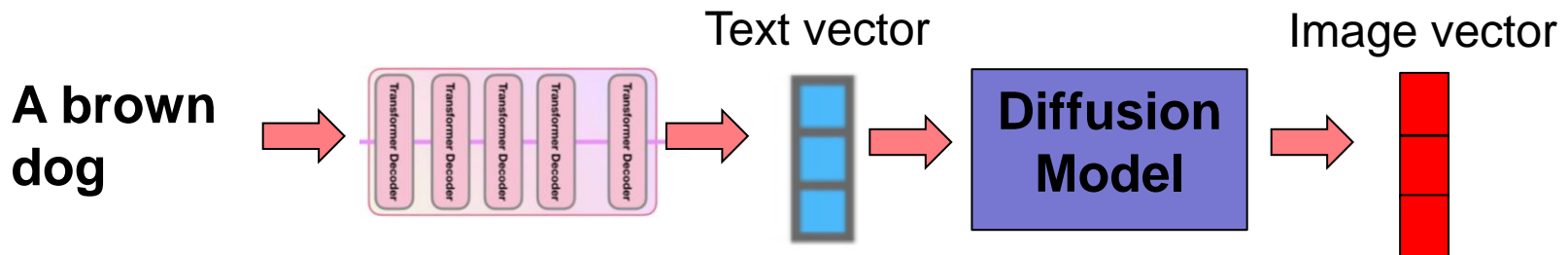
Alex Nichol*
OpenAI
alex@openai.com

Casey Chu*
OpenAI
casey@openai.com

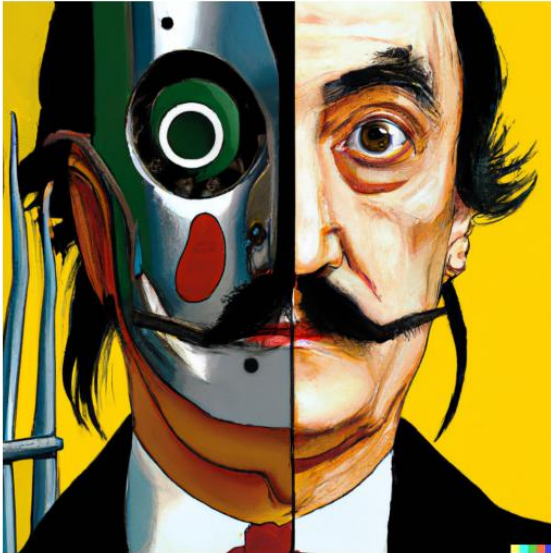
Mark Chen
OpenAI
mark@openai.com

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
 - DALL-E-3 (ChatGPT) from OpenAI
 - Midjourney from Midjourney
 - Stable Diffusion from Stability AI
 - Firefly from Adobe
 - Imagen (Gemini) from Google
 - And many more¹ ...



1. <https://zapier.com/blog/best-ai-image-generator>

AI Image Generators

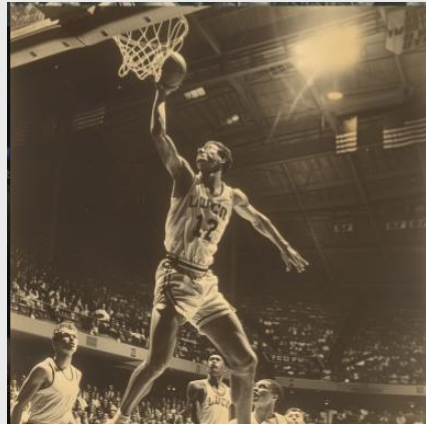
- **Prompt: A Yale men's basketball player jumping from the free-throw line for a dunk during the championship game**

AI Image Generators

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



Firefly



Midjourney



ChatGPT



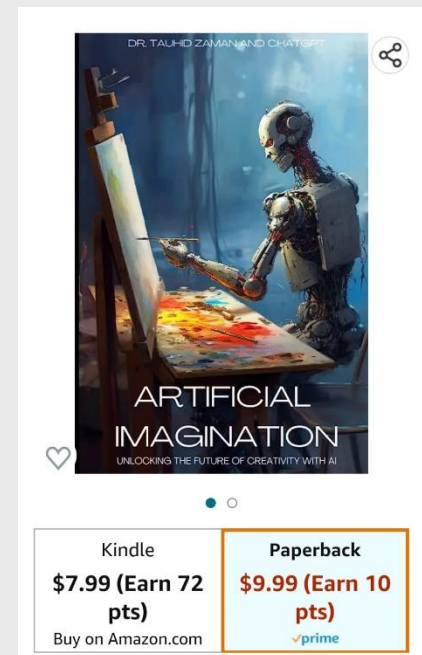
Gemini

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



AI Video Generation

- **Now we can generate an image from text**
- **A video is just a sequence of images which are very similar**
- **Many different approaches for text to video currently being developed**

Sora

- Sora is OpenAI's video generation model



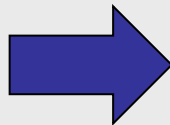
Veo 3

- **Veo 3 is Google's video generation model**



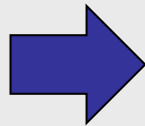
HeyGen

- HeyGen is a platform that lets you create videos from text or images

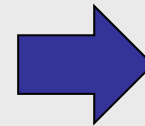


HeyGen Video Avatars

- You can also create video avatars in HeyGen and make them say anything
- You can't make image or video based avatars without permission
- You can make text based avatars

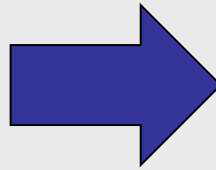


Slim woman with wavy ash-brown hair, bangs, Celine tee, black jeans, gold jewelry, and confidence.



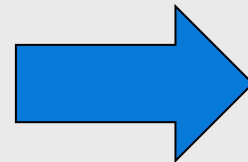
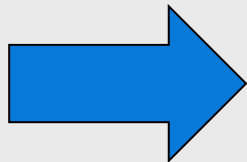
HeyGen Video Avatars

- Or showcase different items



ElevenLabs Voice Dubbing

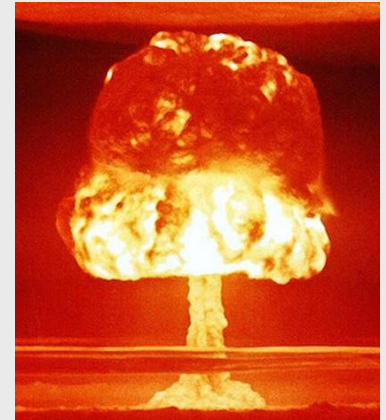
- ElevenLabs lets you dub your video in multiple languages



Dangers of AI Image/Video Generation + Social Media



██████████
██████████



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



Fake Tweets and Images

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

Fake tweet



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 (2 years 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



Inducing Emotion with AI Images

- AI images can make us feel emotion even if we know they are fake
- Prompt: a robot comforting a human after he lost his wife to cancer as a photorealistic image (MidJourney)



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



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

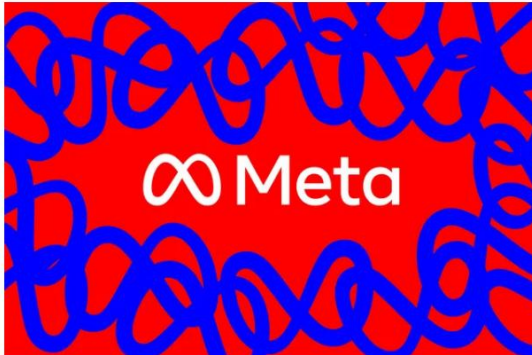


Illustration: Nick Barclay / The Verge

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

By Makensia Kelly, a reporter who covers the politics and power influencing the tech industry. Before joining The Verge in 2016, she covered Congress and breaking news.

Nov 8, 2023, 6:00 AM EST



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

Don't Be Evil

- **As AI improves, fake text, images, audio, and videos become more realistic and effective**
- **Danger is amplified when AI enters social media**
- **Please don't be evil**

Coding Session

- **Dub videos in different languages in ElevenLabs**
- **Create a video of a virtual influencer promoting a fashion item by wearing it in HeyGen**
- **Vibe code an app to generate an image and caption it using Cursor and n8n**