



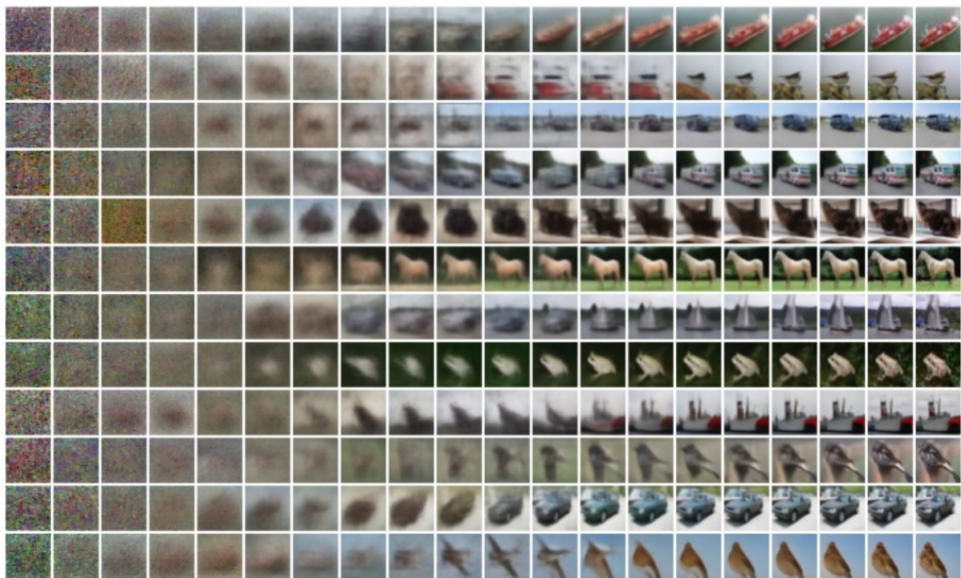
# Vision–Language Models

Jordan Boyd-Graber

University of Maryland

Diffusion Models

Slides from Saurabh Gupta, Sarah Jabbour,  
Yiming Dou, Daniel Geng



From Lana Lazenbik

# Image from Text Generation

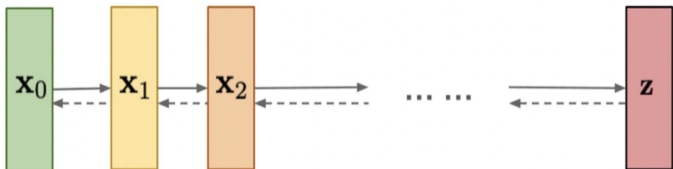
- Multiple models: DALL-E, GAN, ImageGen, Glide

# Image from Text Generation

- Multiple models: DALL-E, GAN, ImageGen, Glide
- Most accessible: Stable Diffusion

## Basic Idea of Diffusion

- Small changes easier to handle than doing everything at once
- Convert noisy inputs to target distribution iteratively

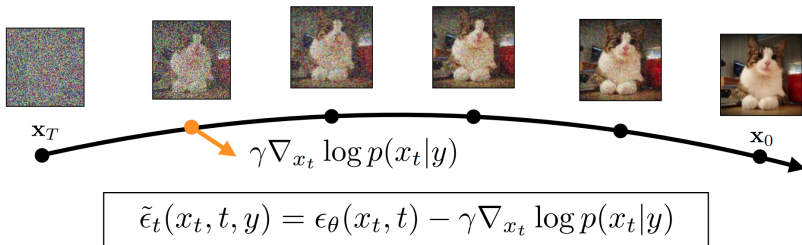


- Markov Chain: each step adds Gaussian noise, learn reverse process

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (1)$$

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}) \quad (2)$$

## Adding Guidance



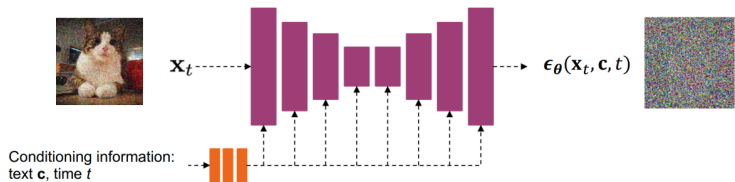
Reverse the process given a classifier that models probability of pixels given label.

# Problems with Classifier Guidance

Need to fine-tune or re-train a classifier on noisy data

- Need a pre-trained classification model
- What if we want to use any text prompt as input?
- Classifier gradients are poor. They can suffer from “shortcuts”

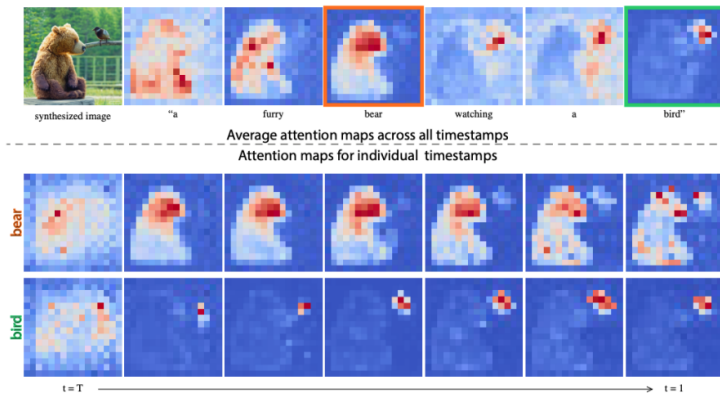
# Longer Captions



Instead of a class label,  $y$  can be an encoded text prompt, injected into the U-Net using cross-attention

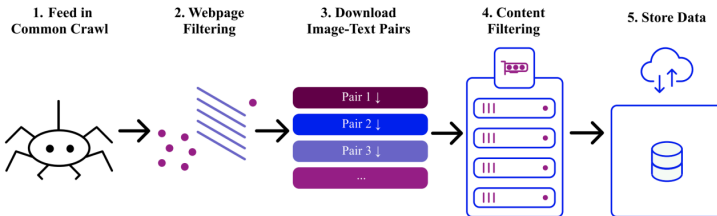


# Longer Captions



Instead of a class label,  $y$  can be an encoded text prompt, injected into the U-Net using cross-attention

# LAION



t

Backend url:

<https://knn5.laion>

Index:

laion\_SB

french cat



[Clip retrieval](#) works by converting the text query to a CLIP embedding, then using that embedding to query a knn index of clip image embeddings

Display captions ☒

Display full

captions ☐

Display similarities

☐Safe mode ☒

Hide duplicate urls

☒

Hide (near)

duplicate images ☒

Search over

image Search with  
multilingual clip☐

french cat



french cat



How to tell if your feline is french. He wears a b...

イケメン猫モデル  
「トキ・ナンタケッ  
ト」がかっこいい-  
NAVER まとめ

Hilarious pics of funny cats! funnycatsgif.com



Hipster cat

網友挑戰「加幾筆畫  
出最創意貓咪圖片」，  
笑到岔氣之後我也手cat in a suit Georgian  
sells tomatoesFrench Bread Cat Loaf  
Metal Print

ARTIFICIAL INTELLIGENCE / TECH / LAW

## Getty Images is suing the creators of AI art tool Stable Diffusion for scraping its content



An image created by Stable Diffusion showing a recreation of Getty Images' watermark. Image: The Verge / Stable Diffusion

/ Getty Images claims Stability AI 'unlawfully' scraped millions of images from its site. It's a significant escalation in the developing legal battles between generative AI firms and content creators.

By **JAMES VINCENT**

Jan 17, 2023, 4:30 AM CST | [18 Comments](#) / [18 New](#)



# Issues

- Copyright
- Stereotypes
- Fake news