# **Introduction to Generative AI)**

### Khoat Than

## Hanoi University of Science and Technology

2024

Challenge: understand complex data?



Computer Vision



Natural Language Processing



### Computational Speech



Robotics

(Some slides were borrowed from Prof. Stefano Ermon, Stanford Univ.)

## Introduction





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

Generative model: "What I understand, I can create"

How to enable a machine to generate an image?



Many generative models have: generation + inference

## **Generative model**

## Generative models are trained from data



Prior knowledge (about Physics, Chemistry, ...)

### Data (e.g.: pictures about rooms)

## Prior knowledge is often useful



Statistical generative model is a probability distribution P(x)

- Data: examples/samples (e.g., pictures about rooms)
- Prior knowledge: distribution shape (e.g., Gaussian?), loss function (e.g., likelihood?), training method, ...

x 
$$\longrightarrow$$
 Distribution  
P(x)  $\longrightarrow$  probability p(x)

## Generation: sample from P(x)















### **Discriminative**: Dinning room > < Bedroom



Given an image X. **Target**: Decision boundary

E.g.: SVM, CNN, ...



Generative: generate X



Y=D , X=





Y=B , X=

Y=B , X=





## **Target**: Joint distribution of X & Y



Bayes' rule:

$$P(Y = Bedroom \mid X = V = F(Y = Bedroom, X = V) = P(Y = Bedroom, X = V)$$

Discriminative model: Since Y is simple and X is given, we do not consider P(X)

Those models may not work with cases of missing data

P(image | caption)

Caption: lychee-inspired spherical chair





Caption: a small hedgehog holding a piece of watermelon



We can generate samples towards a given preference or desire

P(X= | Y = Bedroom)

Given information in Y, we can generate **samples with some desired properties** 



A discriminative model is a simple example of conditional generative model

## Models for image: GANs



Ian Goodfellow, 2019

## Models for image: Diffusion Models



Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *In NeurIPS*.

# Text2Image Diffusion Models

Text:

An astronaut riding a horse



# Text2Image Diffusion Models

Text:

A perfect Italian meal



## DALLE-3

## Text:

 A minimap diorama of a cafe adorned with indoor plants. Wooden beams crisscross above, and a cold brew station stands out with tiny bottles and glasses



### https://openai.com/index/dall-e-3/

# Some applications: Art





Generative model of realistic images



Stroke paintings to realistic images [Meng, He, Song, et al., ICLR 2022]



Generative model of paintings



### Language-guided artwork creation

https://chainbreakers.kath.io @RiversHaveWings

## Some applications: Outlier detection



[Song et al., ICLR 2018]



SPEED

LIMIT







# $P(\text{high resolution} \mid \text{low resolution})$ $P(\text{high resolution)$

Menon et al, 2020

### P(full image | mask)



Liu al, 2018

## P(color image | greyscale)



 Invert a generative process of a sample

Input

sketch



output



Input

Output

"A photo of an open box"



"A goat jumping over a cat"



"A photo of a sitting dog"

"A children's drawing of a waterfall"

Kawar et al., 2023

## Some applications: WaveNet

A model for speech generation



### Text to Speech





van den Oord et al, arXiv:1609.03499, 2016

Diffusion model for speech



Betker, Better speech synthesis through scaling, arXiv:2305.07243, 2023

## Some applications: Audio Super Resolution

Conditional generative model P(high-res signal | low-res audio signal)







High res audio signal

Kuleshov et al., 2017

### Completion

Custom prompt

To get an A+ in deep generative models, students have to

To get an A+ in deep generative models, students have to be willing to work with problems that are a whole lot more interesting than, say, the ones that most students work on in class. If you're a great student, the question above can be avoided and you'll be able to do great work, but if you're not, you will need to go beyond the basics before getting good.

Now to be clear, this advice is not just for the deep-learning crowd; it is good advice for any student who is taking his or her first course in machine learning.

The key point is that if you have a deep, deep brain of a computer scientist, that's just as important to you.

Radford et al., 2019 Demo from talktotransformer.com

### P(next word | previous words)

I will spend two days to explore Venice. What should be a great plan?

Exploring Venice in two days requires a well-organized plan to make the most of your time. Here's a suggested itinerary that covers the major highlights and allows you to experience the unique charm of this beautiful city.

### Day 1: Exploring the Heart of Venice

#### Morning

- 1. Piazza San Marco (St. Mark's Square)
  - Start your day at the iconic St. Mark's Square.
  - Visit St. Mark's Basilica and admire its stunning mosaics.
  - Climb the Campanile (Bell Tower) for a panoramic view of Venice.
- 2. Doge's Palace (Palazzo Ducale)
  - Explore the Doge's Palace, a masterpiece of Gothic architecture.

bạn lên kế hoạch du lịch đồng bằng sông cửu long trong 2 ngày nhe?

Kế hoạch du lịch Đồng bằng sông Cửu Long trong 2 ngày có thể bao gồm nhiều hoạt động thú vị và địa điểm nổi bật. Dưới đây là một gợi ý cho lịch trình:

### Ngày 1: Mỹ Tho - Bến Tre

### Buổi sáng

\$

- 1. Khởi hành từ Sài Gòn
  - Xuất phát lúc 6:00 sáng để tránh kẹt xe và tận hưởng ngày dài ở Mỹ Tho.
  - Thời gian di chuyển: khoảng 1,5 2 giờ.
- 2. Tham quan chùa Vĩnh Tràng
  - Địa chỉ: Ấp Mỹ An, xã Mỹ Phong, TP. Mỹ Tho.
  - Thời gian tham quan: khoảng 1 giờ.

<pre> • parse_expenses.py •• write_sql.go  sentiments dddresses.rb  1 import datetime 2 3 def parse_expenses(expenses_string): 4 """Parse the list of expenses and return the list of triples (date, va) 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</pre>	
<pre>import datetime def parse_expenses(expenses_string):     """Parse the list of expenses and return the list of triples (date, va)     """Parse the list of expenses and return the list of triples (date, va)     """"     """     """"     """"     """"     """"     """"     """"     """"     """"     """""     """"     """"     """"     """"     """"     """"     """""     """""     """"     """""     """""     """""     """""     """""     """"""</pre>	🝦 parse_expenses.py 👓 write_sql.go 🖪 sentiment.ts 🛃 addresses.rb
<pre>2 3 def parse_expenses(expenses_string): 4 """Parse the list of expenses and return the list of triples (date, va 5 6 7 8 8 9 9 10 11 12 13 14 15 16 17 18 19 20 </pre>	1 import datetime
<pre>3 def parse_expenses(expenses_string): 4  """Parse the list of expenses and return the list of triples (date, va 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 </pre>	2
<pre>"""Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """Parse the list of expenses and return the list of triples (date, va """"Parse the list of expenses and return the list of triples (date, va """" """" """"Parse the list of expenses and return the list of triples (date, va """"" """" """" """" """"" """"" """"" """"</pre>	3 def parse_expenses(expenses_string):
5       6       7         7       7       7         8       9       10         10       11       12         11       12       13         12       13       14         15       16       17         18       19       19         20       10       10	4 """Parse the list of expenses and return the list of triples (date, va
6       7         7       8         9       9         10       11         11       12         13       14         15       16         16       17         18       19         20	5
7       8       9         9       9       9         10       11       12         12       13       14         13       14       15         16       17       18         19       20       9       9	6
8	7
9       10         11       11         12       13         13       14         15       16         16       17         18       19         20	8
10         11         12         13         14         15         16         17         18         19         20	9
11         12         13         14         15         16         17         18         19         20	10
12 13 14 15 16 17 18 19 20	
13         14         15         16         17         18         19         20	12
14       15       16       17       18       19       20	13
15       16       17       18       19       20	14
16     17       18     19       20	15
17 18 19 20	16
18 19 20	17
19 20	18
20	19
	20



Create video by providing a short description

a couple sledding down a snowy hill on a tire roman chariot style



# Some applications: Video



## Some applications: Imitation

Conditional model: P(actions | past observations)





Janner et al., 2022

Li et al., 2017

## Some applications: Materials



## Some applications: ...



\$

### Many more

- Text to song
- Text to room design
- Text to house design
- Image to text
- Image to translated text, ...

Đây là hình ảnh của một hộp sản phẩm "Bioscalin NOVA Genina" với chức năng là "Shampoo Fortificante Rivitalizzante," có nghĩa là dầu gội tăng cường và phục hồi. Sản phẩm này giúp tăng cường sức khỏe của tóc và da đầu, đặc biệt là dành cho tóc yếu. Sản phẩm này được sản xuất bởi công ty Giuliani.

0

Là cái gì?