Introduction to the Course

Generative Models: Fundamentals and Applications

Shuigeng Zhou, Yuxi Mi College of Computer Science and Artificial Intelligence 2025/9/8

Outline

- What are generative models?
- Why generative models?
- Discriminative models vs. Generative models
- Applications of generative models
- Arrangements for the course

Generative Models Now Everywhere...

Challenge

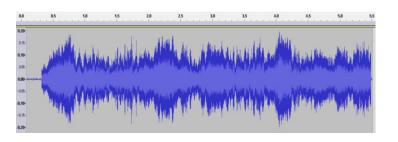
- Understand *complex*, *unstructured* inputs
- Generate *unseen* samples



Computer Vision image/video generation, restoration, style transfer, scene understanding



Natural Language Processing text generation, dialogue systems, machine translation, code generation



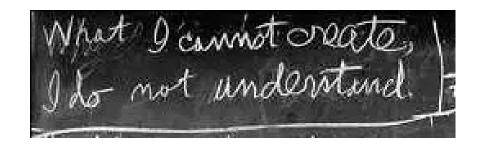
Speech and Audio text-to-speech, voice conversion, music generation, speech enhancement

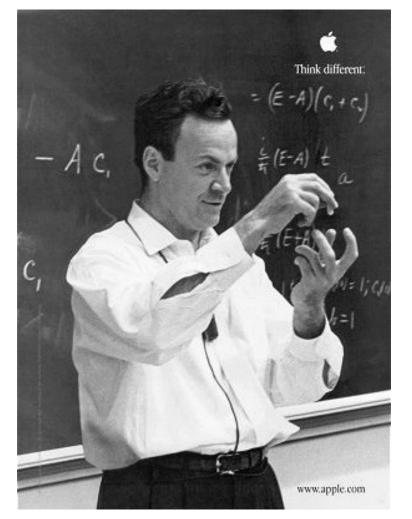


Robotics dynamics modeling, trajectory prediction, motion planning, environment simulation

What are Generative Models?

- "What I cannot create, I do not understand"
 - -- Richard Feynman (Physicist, 1965 Nobel Prize winner)
- Generative modeling
 - What I understand, I can create



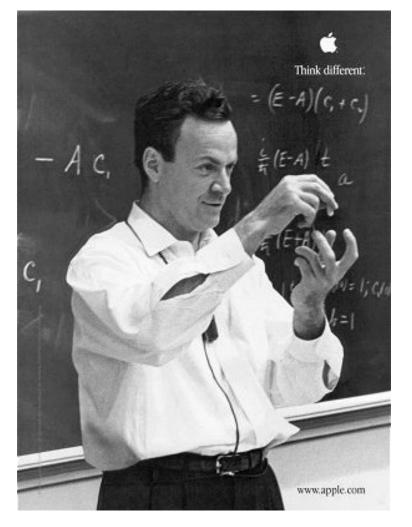


What are Generative Models?

- "What I cannot create, I do not understand"
 - -- Richard Feynman (Physicist, 1965 Nobel Prize winner)
- Generative modeling
 - What I understand, I can create

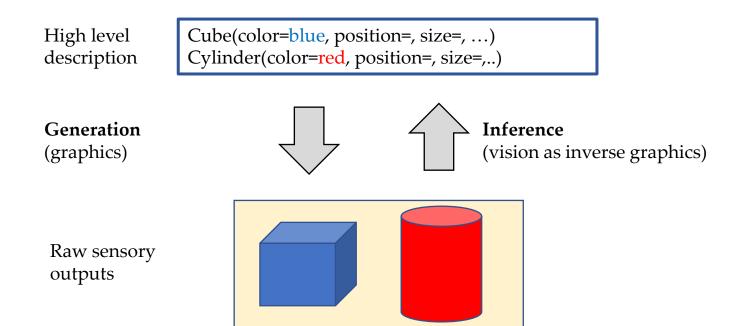
*How?*Through a probabilistic / statistical view

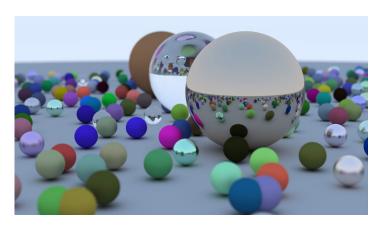




Generative Modeling

• An example case: How to generate natural images with a computer?





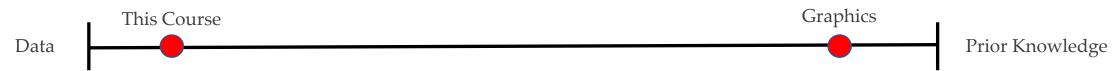
Generated balls via ray tracing Shirley et al., 2016

• Many generative models will have similar structure (generation + inference)

• Statistical generative models are **learned from data**

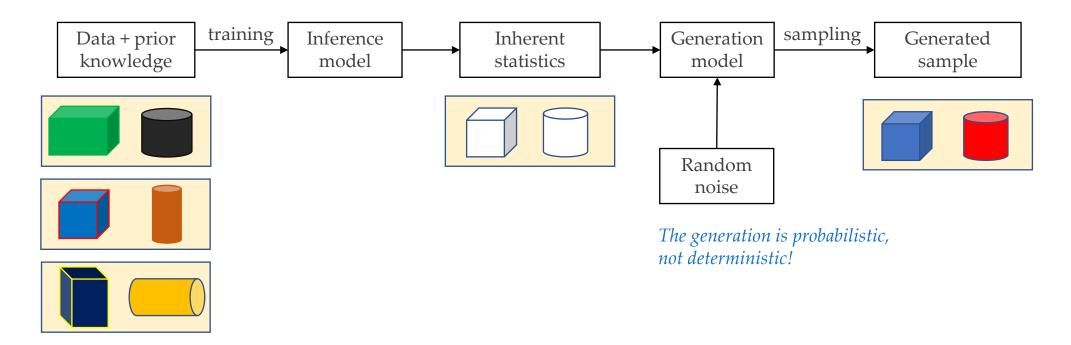


• Priors are always necessary *, but there is a spectrum

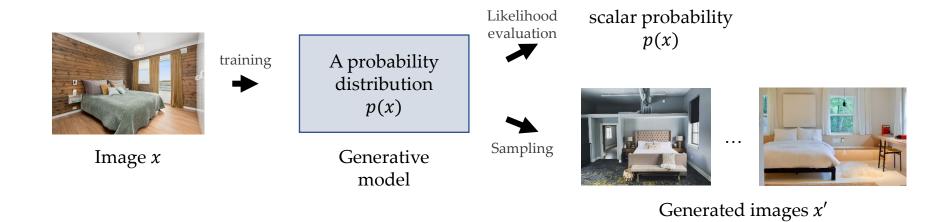


^{*} Wolpert, D. H., & Macready, W. G, 1997. No Free Lunch Theorems for Optimization.

- How to generate natural images with a statistical generative model?
 - A general (*not decisive*) pipeline



- A statistical generative model is a probability distribution p(x)
 - Data: samples (e.g., images of bedrooms)
 - **Prior knowledge:** parametric form (e.g., Gaussian?), loss function (e.g., maximum likelihood?), optimization algorithm, etc.

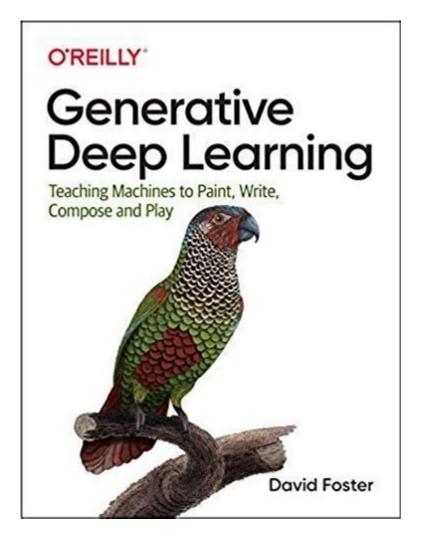


• It is generative because sampling from p(x) generates new images

A broad definition

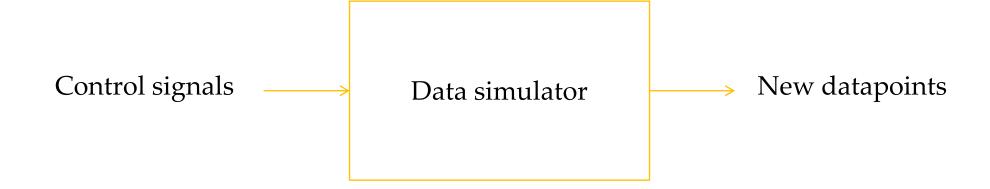
 A generative model describes how a dataset is generated, in terms of a probabilistic model.
 By sampling from this model, we are able to generate new data.

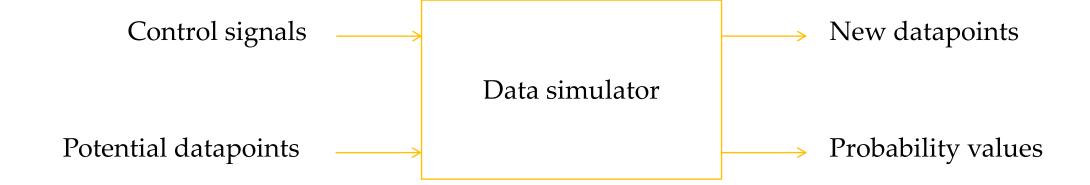
----David Foster (ed.), Generative deep learning, 2019

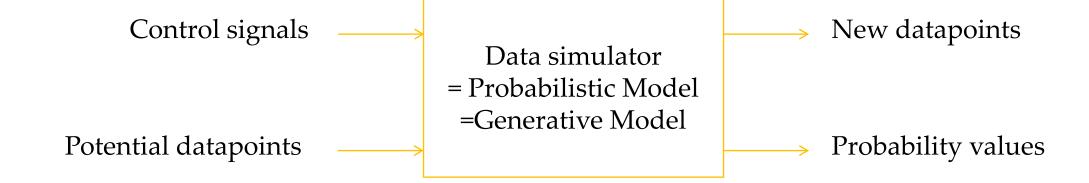


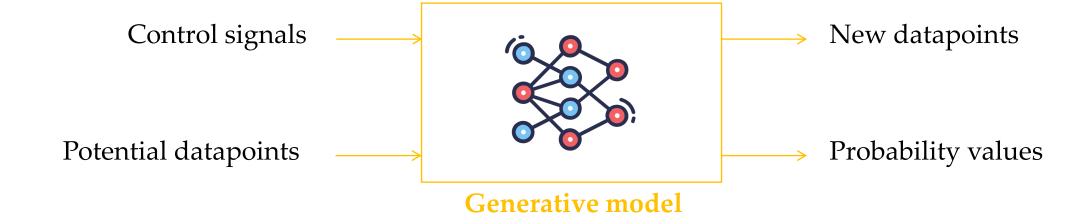
• Rethink from an application view...



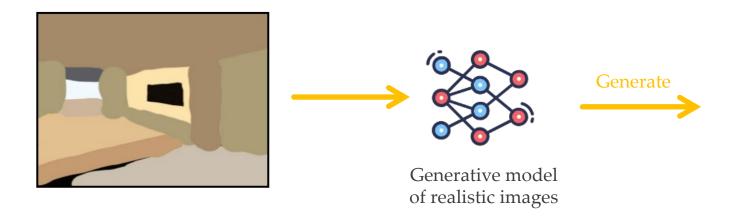








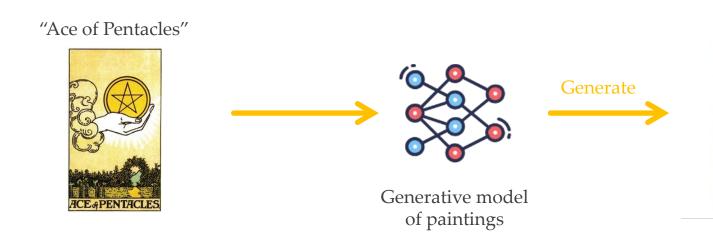
Apply to Real-World Data Generation







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

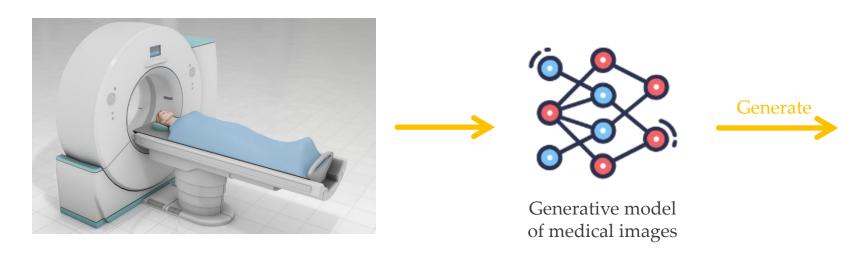


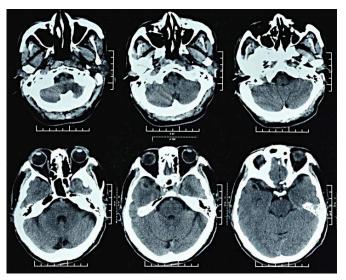




Language-guided artwork creation RiversHaveWings, https://chainbreakers.kath.io

Apply to Solve Inverse Problems





Medical image reconstruction Song et al., ICLR 2022

Apply for Outlier Detection







Generative model of traffic signs

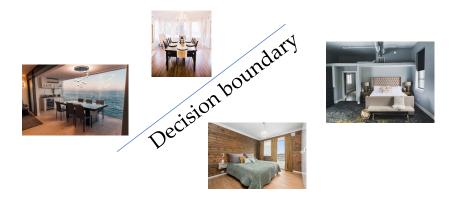






Outlier detection Song et al., ICLR 2018

• Discriminative: classify bedroom vs. dining room



- The input image *X* is given
- Goal: a good decision boundary, via conditional distribution over label Y

$$P(Y = \text{Bedroom} \mid X = 0.0003)$$

• E.g., SVM, logistic regression, CNN, etc.

• **Generative**: generate *X*

$$Y = B$$
, $X = \begin{bmatrix} Y = D \\ X = \end{bmatrix}$
 $Y = D$, $X = \begin{bmatrix} Y = D \\ X = \end{bmatrix}$
 $X = \begin{bmatrix} Y = D \\ X = \end{bmatrix}$

- The input *X* is not given
- Requires a model of the **joint distribution**

$$P(Y = \text{Bedroom}, X = 0.0002)$$

E.g., naïve Bayes, hidden Markov, etc.

• Joint and conditional are both related via **Bayes Rule**:

Discriminative

• *Y* is simple; *X* is always given, so not need to model P(X = X)

• However, it cannot handle missing data P(Y = Bedroom | X = X)

Generative Models

• Class conditional is possible:

Class conditional is possible:
$$P(X = \{Y = Bedroom\}) = P(Y = Bedroom)$$

$$P(X = \{Y = Bedroom\}) = P(Y = Bedroom)$$

In the real world, it's often useful to condition on rich side information *Y*

$$P(X = | Y = \text{``A black table with 6 chairs''})$$

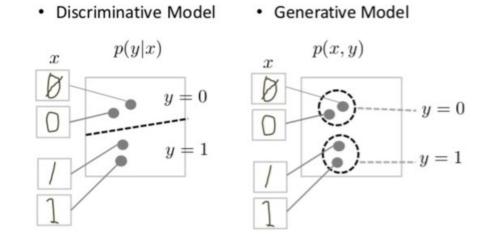
A discriminative model is also a very simple conditional generative model of *Y*:

$$P(Y = \text{Bedroom} | X =)$$

Recall

The probabilistic distribution

- Formally, given a set of data instances *X* and a set of labels *Y*:
 - Generative models capture the joint probability p(X,Y), or just p(X) if there are no labels
 - Modeling the distribution of individual classes
 - Discriminative models capture the conditional probability p(Y|X)
 - Modeling the boundaries between classes



Discriminative and generative models of handwritten digits

Discriminative

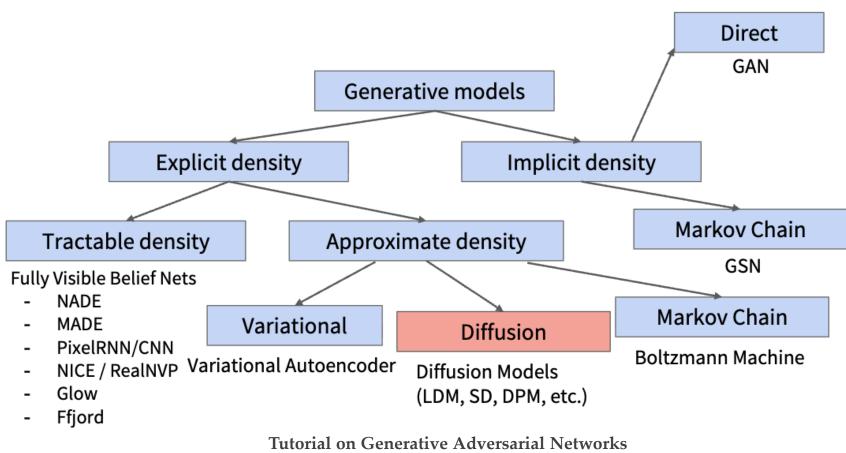
- Pros:
 - Simple and direct
 - Require fewer samples, lower training cost
 - High performance on classification tasks
- Cons:
 - Capture limited information about data distribution
 - Cannot directly perform data generation or outlier detection

Generative

- Pros:
 - Capture richer information in data
 - More versatile: can handle generation, unsupervised learning, anomaly detection
 - Can also be used for discriminative tasks
- Cons:
 - Require more samples and higher computational cost
 - Sensitive to noise and outliers
 - Training is often more complex

Taxonomy of Generative Models

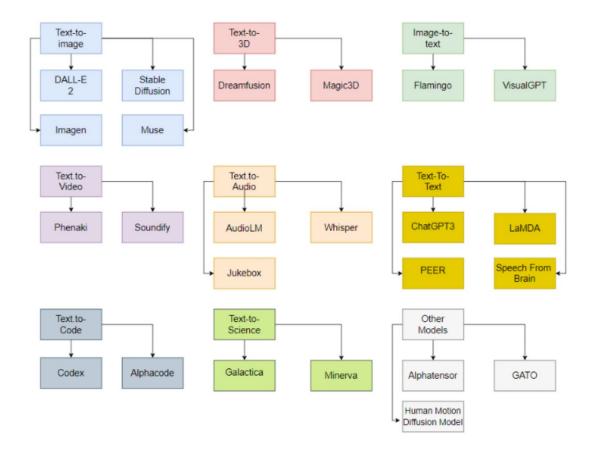
• From a probabilistic modeling perspective



Ian Goodfellow, 2017

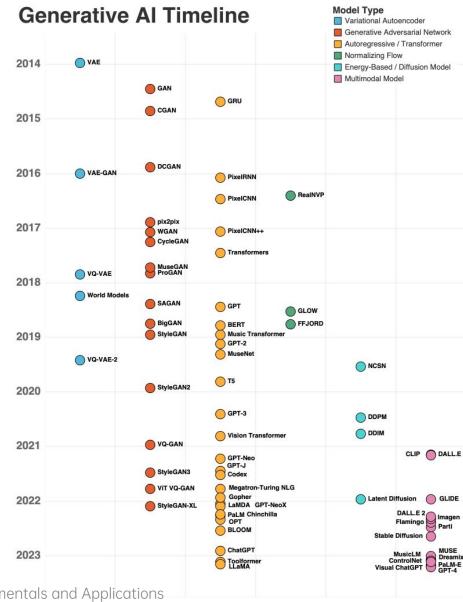
Taxonomy of Generative Models

• From an input/generated data modality perspective



Generative Models: A Timeline

• Recent models by date of release



Properties of Generative Models

- Common criteria
 - Stability
 - How easily can the model be trained?
 - Sample quality & diversity
 - Are the generated samples realistic and varied?
 - Efficiency
 - Is the model computationally efficient in training and inference?
 - Likelihood estimation
 - Can the model assign probabilities to samples?
 - Capacity
 - Can the model represent and generate complex data?
 - Flexibility
 - How constrained is the model by its structure or assumptions?
 - Evaluability
 - How easy is it to evaluate and compare this model?

| | Stability | Quality | Diversity | Efficiency | Likelihood |
|------|-----------|----------|-----------|------------|------------|
| VAE | © | 8 | © | © | © |
| GAN | 8 | © | 8 | © | 8 |
| DM | ☺ | © | © | 8 | |
| Flow | <u>=</u> | © | © | ☺ | © |

There is no one-for-all solution

Recent Progress: Image Generation

• Face generation via VAE and GANs



Ian Goodfellow, 2019

Recent Progress: Image Generation

• Face generation via diffusion models



Score-Based Generative Modeling through Stochastic Differential Equations Song et al., 2021



Flux. 1 2024

Recent Progress: Image Generation

• Face dataset generation



MorphFace Mi et al., 2025

- T2I User input:
 - "An astronaut riding a horse"



Dall-E 2 2022

- T2I User input:
 - "A perfect Italian meal"



Stable Diffusion 3

- T2I User input:
 - "泰迪熊穿着戏服,站在太和殿前唱京剧"



文心一言 2023

• T2I User input:

• "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"



Dall-E 3 (Built natively on ChatGPT)

- T2I User input:
 - "A story about LeCun, Hinton, and Benjio winning the Novel Prize"



Recent Progress: Inverse Problems

• Inpainting / Super-Resolution / Restoration

P(high resolution | low resolution)



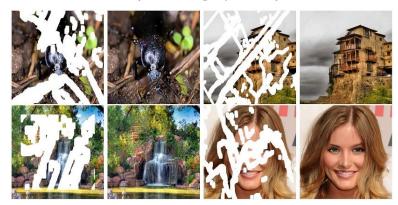
Menon et al, 2020

P(color image | greyscale)



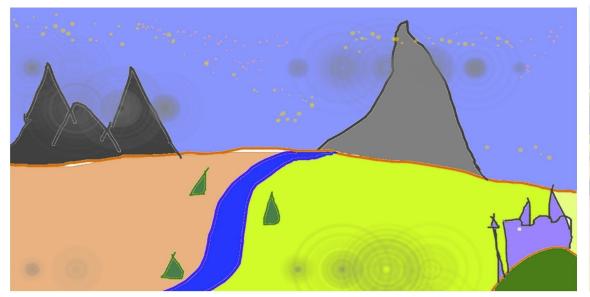
Antic et al., 2020

P(full image | mask)



Liu al, 2018

• Sketch-to-image





• Sketch-to-image

Stroke Painting to Image

Stroke-based Editing

Input

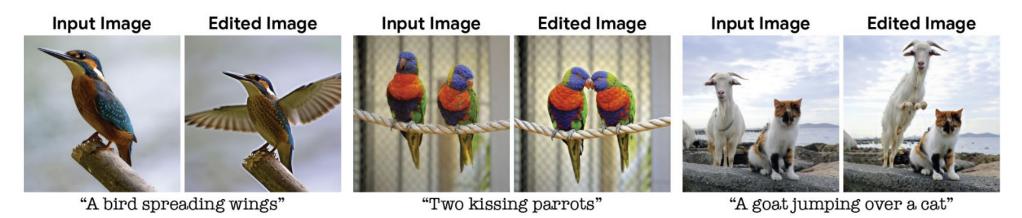
Output

Stroke-based Editing

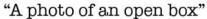
Output

SDEdit *Meng et al.,* 2021

Text-based image editing









"A photo of a sitting dog"



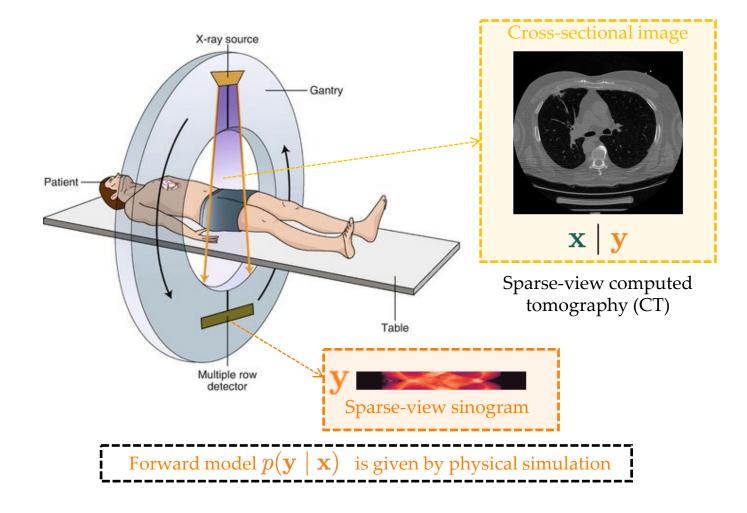




"A children's drawing of a waterfall"

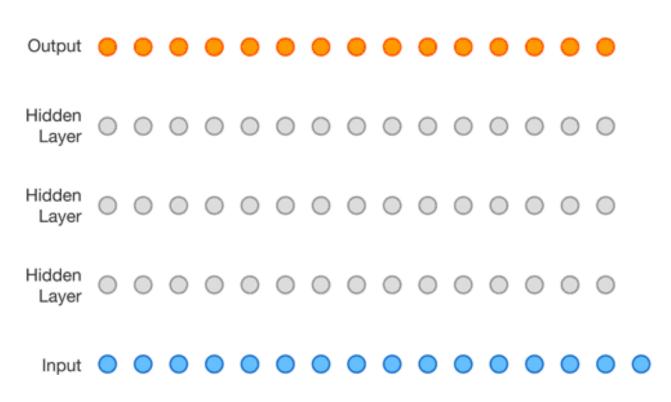
Kawar et al., 2023

• Medical image reconstruction



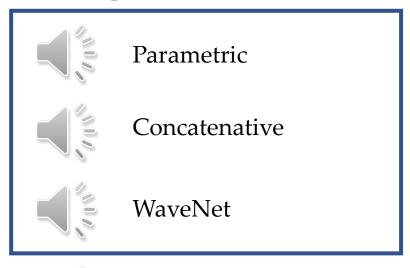
Recent Progress: Audio & Speech

• Generative model of speech signals



WaveNetvan den Oord et al., 2016

Text to Speech





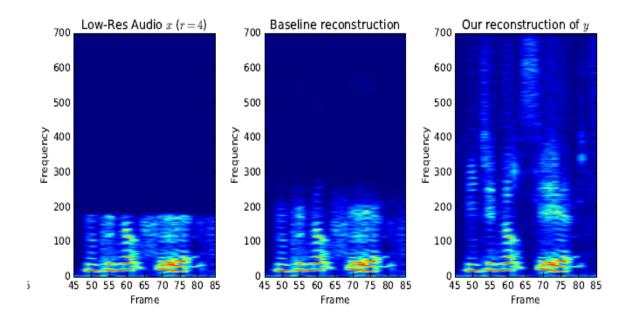
Unconditional



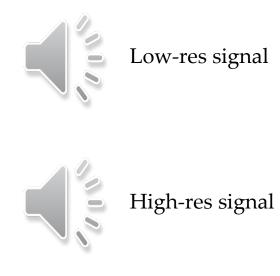
Music

Recent Progress: Audio & Speech

- Audio super-resolution
 - Conditional generative model *P*(highres signal | lowres signal)







Recent Progress: Language Generation

- Text auto completion
 - Next-word prediction: *P*(next word | previous words)

Custom prompt

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

Completion

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

Recent Progress: Language Generation

GPT

How to learn generative models well?

Learning generative models well is a big journey, but you can approach it step by step depending on your background and goals. Here's a structured roadmap I'd suggest:

1. Build Core Foundations

Before diving into generative models, strengthen the basics:

- Probability & Statistics: distributions, Bayes' rule, KL divergence, entropy.
- Linear Algebra & Optimization: matrix operations, eigen-decomposition, gradient descent, convexity.
- Deep Learning Fundamentals: backpropagation, CNNs, RNNs, transformers.
- recommended: Deep Learning by Goodfellow, Bengio, Courville (especially Chapters 3–5, 17).

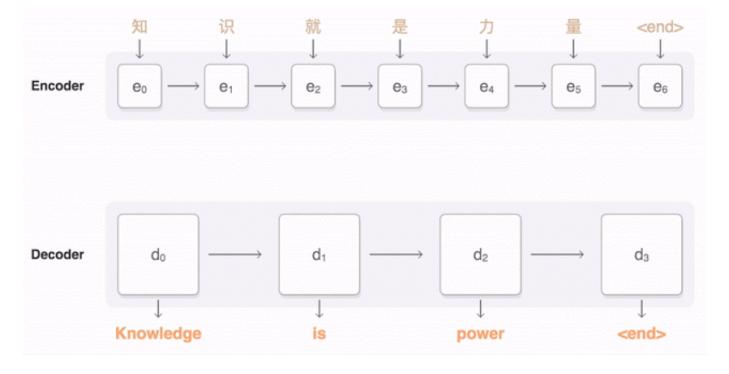
2. Understand the Landscape of Generative Models

Generative models aim to learn p(x) or p(x|y). The major families are:

- Explicit likelihood models
 - Gaussian Mixture Models (classical start)
 - Variational Autoencoders (VAE)
 - Normalizing Flows (e.g., RealNVP, Glow)
- Implicit models
 - Generative Adversarial Networks (GANs, StyleGAN, BigGAN)
 - Diffusion Models (DDPM, Stable Diffusion, Imagen)

Recent Progress: Language Generation

- Machine translation
 - Conditional generative model *P*(English text| Chinese text)



Google AI research

- OpenAI Codex
 - Aka. Copilot

```
Ts sentiment.ts
                                                  🔏 addresses.rb
parse_expenses.py
 1 import datetime
 3 def parse_expenses(expenses_string):
        """Parse the list of expenses and return the list of triples (date, va
12
```

- T2V User Input:
 - "Suddenly, the walls of the embankment broke and there was a huge flood."



Pika Pika Labs

- T2V User Input:
 - "A couple sledding down a snowy hill on a tire roman chariot style."



Pika Pika Labs

• Video and audio generation



• Talking head



Zhong et al., 2025

• Talking head



Zhong et al., 2025

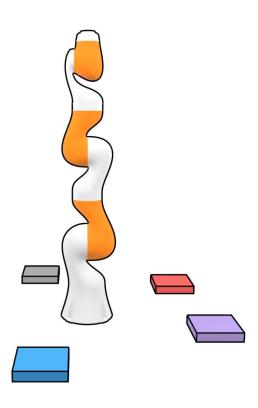
Recent Progress: Robotics

• Imitation Learning

• Conditional generative model *P*(actions | past observations)



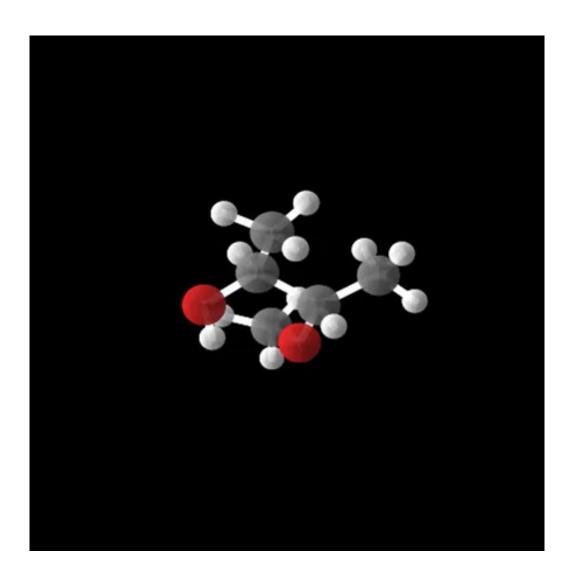
Li et al., 2017



Janner et al., 2022

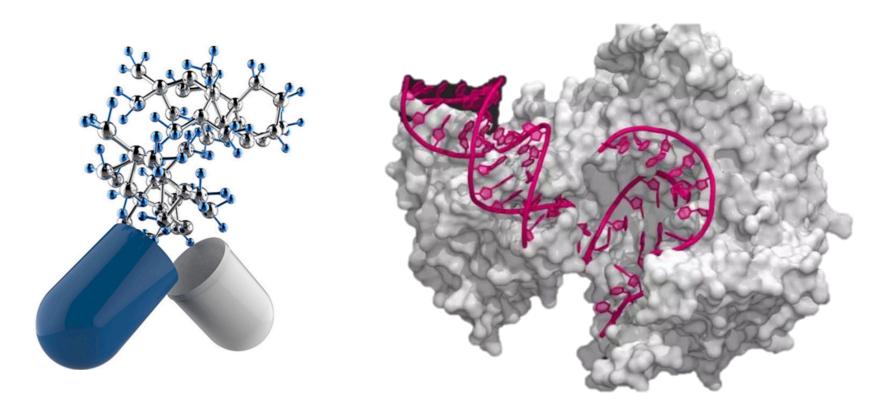
Recent Progress: AI4S

• Molecule generation



Recent Progress: AI4S

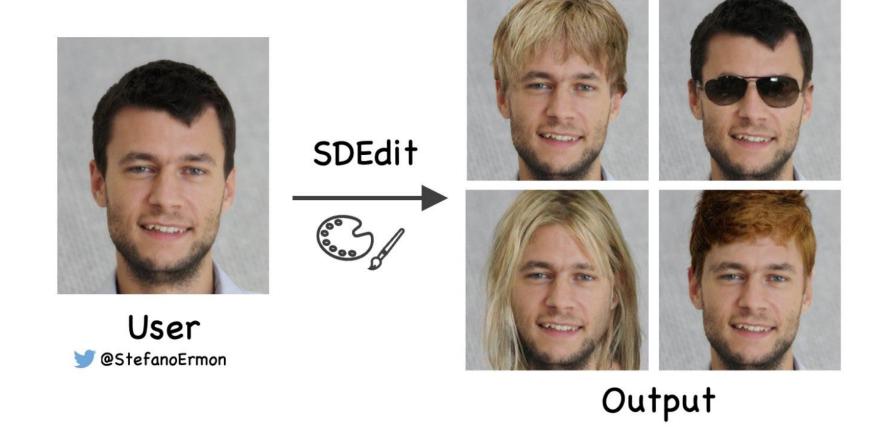
• Bioinformatics



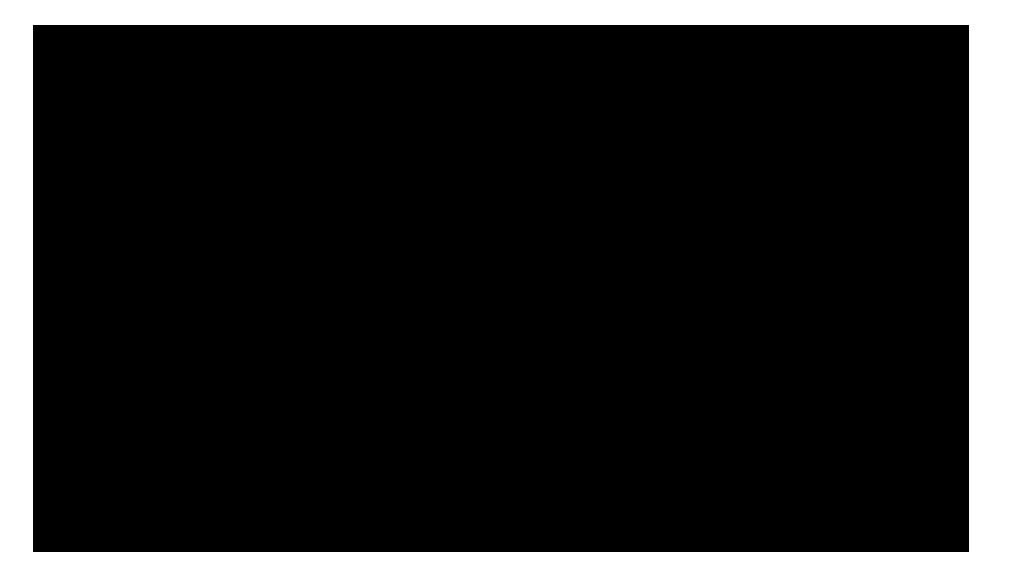
ProteinGAN Repecka et al., Nature Machine Intelligence, 2021

Recent Progress: DeepFake

• Which image is real?



Recent Progress: DeepFake



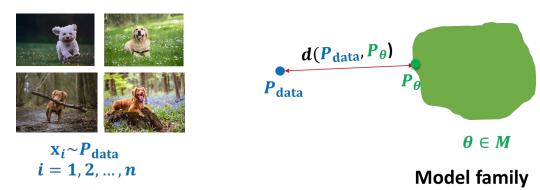
Roadmap and Key Challenges

• Representation

- How do we model the joint distribution of many random variables?
- Need compact representation

• Learning

What is the right way to compare probability distributions?



Inference

- How do we invert the generation process (e.g., vision as inverse graphics)?
- Unsupervised learning: recover high-level descriptions (features) from raw data

Syllabus

- Basic concepts of generative models
- Generative models in statistical learning
 - Naive Bayesian models, Hidden Markov models, Gaussian mixture models
 - Probabilistic latent semantic analysis models (PLSA), latent Dirichlet allocation

Generative models in deep learning

- Autoregressive models
- Variational auto-encoders (VAEs)
- Generative adversarial networks (GANs)
- Flow-based models
- Energy-based models (EBMs)
- Score-based / diffusion generative models (DMs)
- Large-language models (LLMs)

Applications in different areas

- Image processing: image generation
- NLP: text generation
- Speech recognition: audio generation
- Video processing: video generation
- Drug design & discovery: molecule generation

Syllabus

- Basic concepts of generative models
- Generative models in statistical learning
 - Naive Bayesian models, Hidden Markov models, Gaussian mixture models
 - Probabilistic latent semantic analysis models (PLSA), latent Dirichlet allocation

Generative models in deep learning

| • | Autoregressive models | ~2011 | |
|---|---|-------|-------------------------|
| • | Variational auto-encoders (VAEs) | ~2013 | |
| • | Generative adversarial networks (GANs) | ~2014 | time rose to prominence |
| • | Flow-based models | ~2015 | |
| • | Energy-based models (EBMs) | ~2016 | |
| • | Score-based / diffusion generative models (DMs) | ~2019 | |
| • | Large-language models (LLMs) | ~2020 | • |

• Applications in different areas

- Image processing: image generation
- NLP: text generation
- Speech recognition: audio generation
- Video processing: video generation
- Drug design & discovery: molecule generation

Syllabus

- Basic concepts of generative models
- Generative models in statistical learning
 - · Naive Bayesian models, Hidden Markov models, Gaussian mixture models
 - Probabilistic latent semantic analysis models (PLSA), latent Dirichlet allocation

Generative models in deep learning

- Autoregressive models
- Flow-based models
- Variational auto-encoders (VAEs)
- Generative adversarial networks (GANs)
- Energy-based models (EBMs)
- Score-based / diffusion generative models (DMs)
- Large-language models (LLMs)
- Applications in different areas
 - Image processing: image generation
 - NLP: text generation
 - · Speech recognition: audio generation
 - Video processing: video generation
 - Drug design & discovery: molecule generation

Prerequisites

Mathematics

- Basic knowledge of probability and calculus
 - Random variables, independence, conditional independence
 - Gradients, gradient-descent optimization
 - Bayes rule, chain rule, change of variables formulas

• Computer science

- basic data structures/algorithms
- basic knowledge of machine learning/deep learning
- Programming language, preferably Python

Course Goal

Understand

- Grasp the fundamentals of generative models
- Read the latest research papers without obstacles

• Familiarize

• Learn major generative models, especially deep ones, with their pros & cons

Recognize

• Identify the latest achievements and development trends in typical application areas

Explore

- If possible, try to conduct some research
- Say, improve existing models or apply them to solve real problems

Grading Policies

- Attendance: 20%
- Paper presentation: 30%
 - Introduce at least one paper from top venues
- Course article: 40%
 - A survey or a research-style article related to generative models
 - Preferably use LaTeX
- Course project: 10%
 - Reproduce a generative model; if possible, explore extensions
 - Or, use generative models to solve a tiny real problem
 - Several project options will be provided

Logistics

- Instructors
 - Prof. Shuigeng Zhou (周水庚)
 - D4021, No. 2 Interdisciplinary Building
 - Email: <u>sgzhou@fudan.edu.cn</u>
 - Homepage: http://admis.fudan.edu.cn/sgzhou
 - Dr. Yuxi Mi (糜予曦)
 - D4023, No. 2 Interdisciplinary Building
 - Email: <u>yxmi@fudan.edu.cn</u>
- *The best way to reach us is email!*

Logistics

• Time

- 1:30 PM 4:10 PM
 Mondays from September 8 to December 22, 2025
- There will be no class on October 6 (national holidays)
- *So, we have totally 15 weeks of class*

Venue

• JA 103

Recommended Texts

- S. Bond-Taylor, A. Leach, Y. Long, et al. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. arXiv, 2021.
- E. Nalisnick, A. Matsukawa, Y. W. Teh, et al. Do Deep Generative Models Know What They Don't Know? ICLR, 2019.
- R. Salakhutdinov. Learning deep generative models. Annual Review of Statistics and Its Application, 2015.
- Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. The MIT Press, 2012.
- David Foster (ed.). Generative Deep Learning. O'Reilly Media, 2019.
- I. Goodfellow, Y. Bengio, A. Courville (eds.), Deep Learning. The MIT Press, 2016.
- 李航. 统计学习方法. 清华大学出版社, 2012.

Recommended Top Venues

Journals

- Nature, Science, Nature Machine Intelligence
- TPAMI, IJCV, AI, JMLR
- TIP

Conferences

- NIPS, ICLR, ICML
- ICCV, CVPR, ECCV
- AAAI, IJCAI, MM, COLT, ACL, COLM

Thank you for listening!

Questions?

