Generative Models

Guide to AI and Data Courses · Part II: Branches of AI

Data distribution | Latent space | VAEs and GANs | Diffusion | Autoregressive models | Neighboring courses

Sercan KÜLCÜ · Giresun University · Computer Engineering

© Sercan Külcü. All rights reserved. · Interactive HTML version of these slides: https://sercankulcu.github.io/files/ai_courses/slides/Chapter_13_Generative_Models.html

INTRODUCTION

What Question Does It Ask?

> **How does a model generate new samples that look like real data?**

Generative models learn the probability distribution of data and **generate new samples** from it. A discriminative model asks "is this image a cat?", while a generative model asks "how are cat images distributed, and how do I draw a new one?"

At a glance

APPROACH learning the data distribution	**INPUT** huge, mostly unlabeled data	**OUTPUT** new samples: images, text, audio
SUCCESS MEASURE realism and diversity (FID etc.)	**MATH LOAD** high	**DATA NEEDS** very large

A Short History

2013
Kingma and Welling propose the variational autoencoder (VAE).

2014
Goodfellow and colleagues introduce generative adversarial networks (GANs), pitting a generator against a discriminator.

2015
Sohl-Dickstein and colleagues publish the idea of diffusion models that undo noise step by step.

2018
OpenAI's first GPT: an autoregressive transformer pretrained on a large text corpus.

2020
Ho, Jain and Abbeel's DDPM brings diffusion models to the front of image generation.

2022
Stable Diffusion is released with open weights; ChatGPT brings generative models to a wide audience.

THE COURSE

Core Concepts

Likelihood
The probability the model assigns to real data; training increases it.

Latent space
Low-dimensional hidden variables that summarize samples.

Autoencoder and VAE
Encode, decode; a VAE forces the latent space into a regular distribution.

GAN
A generator makes fake samples, a discriminator tries to tell them from real ones.

Diffusion
Adds noise to data; a network learns to remove it step by step.

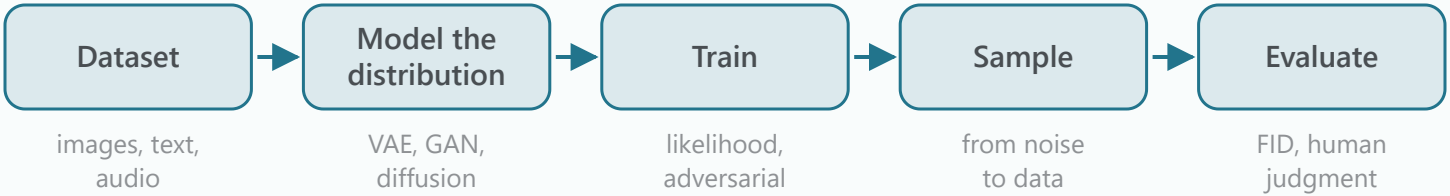
Autoregressive model
Predicts each token from the previous ones: GPT.

Conditional generation
Generation guided by text, a class or a sketch.

Evaluation
FID, perplexity and human evaluation.

THE COURSE

A Typical Workflow



Example: A diffusion model adds noise step by step to millions of images and trains a network to predict that noise. At generation time it starts from pure noise, removes it over dozens of steps guided by a text prompt, and arrives at an image never seen before.

THE COURSE

What Does the Course Cover?

1. Generative modelling; likelihood and KL divergence
2. Autoregressive models: PixelCNN, transformers
3. Variational autoencoders (VAE) and the ELBO
4. Normalizing flows
5. Generative adversarial networks and training stability
6. Energy-based models
7. Score matching and diffusion models
8. Fast sampling and latent diffusion
9. Conditional generation, classifier-free guidance
10. Large language models: pretraining and fine-tuning
11. Multimodal models: text to image
12. Evaluation: FID, perplexity, human judgement

Skills you will gain

training VAEs and GANs in PyTorch

using diffusion models with Diffusers

fine-tuning language models with Transformers

measuring sample quality with FID

tracking and comparing training runs

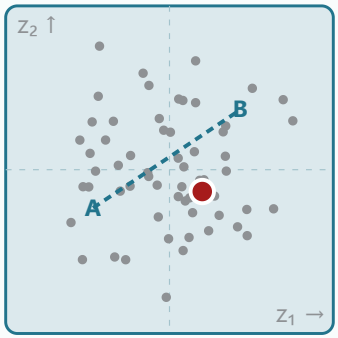
A typical semester; the order and coverage vary between universities.

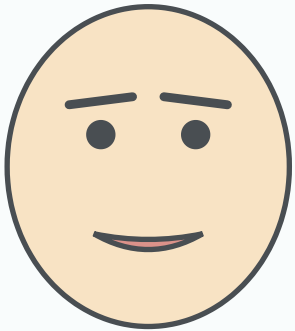
TRY IT

New Faces from Latent Space

A game studio wants new characters from 60 cartoon faces. The model learned to compress each face into two numbers (z_1, z_2); generate a new face with the sliders or by clicking the plane on the left.

Latent space (click)





z_1 = **0.6**

z_2 = **-0.4**

Random face | Interpolate A → B

Interpolation from A to B:



Two numbers → 6 face measures. Smile **0.60**, eye size **0.34**, mouth opening **0.21**.
Distance to nearest training face 0.05: very close to a training face.
Two dimensions carry 98% of the variation in the data.

The model did not memorize the faces; it learned the **latent structure** of the data. Every point on the plane gives a new but plausible face; VAEs and diffusion carry the same idea to millions of dimensions.

CONTEXT

Prerequisites

Courses worth knowing before this one, and why they are needed.

Course	Why it is needed
Deep Learning	GANs, VAEs, diffusion, large language models
Information Theory	KL divergence, likelihood
Calculus	working comfortably with integrals, derivatives and probability densities
Programming	Python and a deep learning library (PyTorch)

CONTEXT

Are You Ready for This Course?

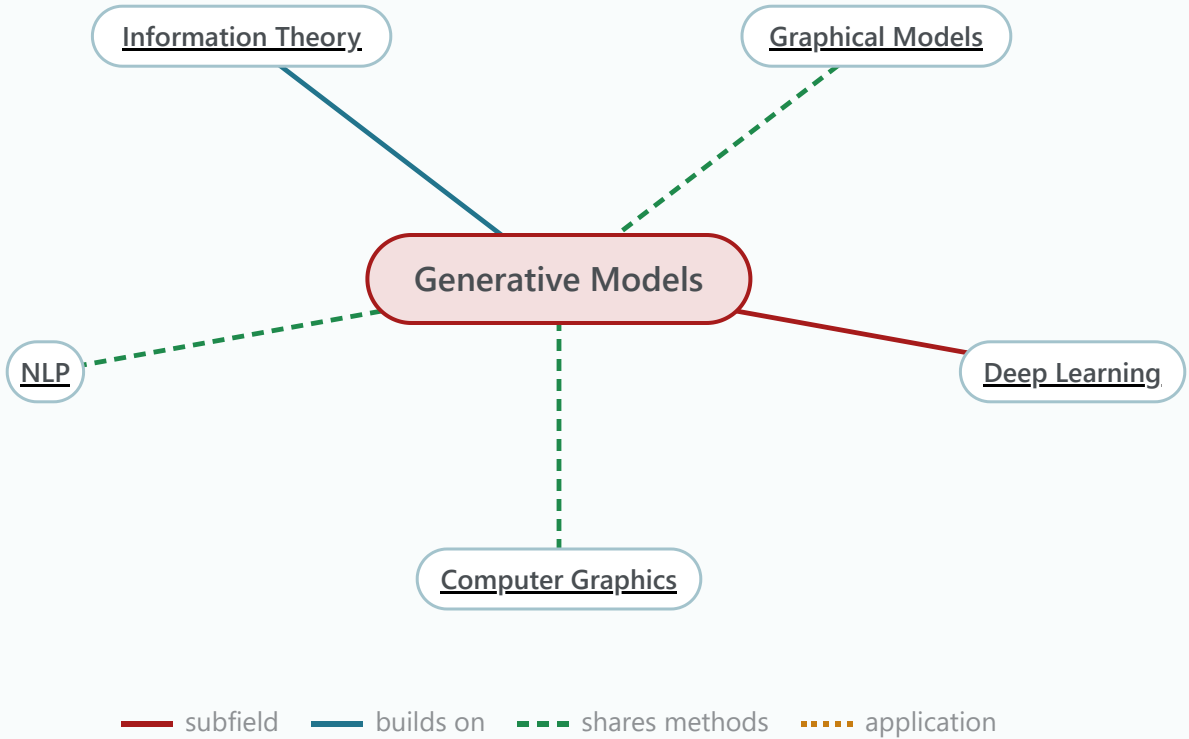
Ask yourself; for every "no", the right-hand column shows where to start.

Ask yourself	If not, start with
Can you build and train a neural network in PyTorch?	Deep Learning
Can you explain probability densities, expectation and Bayes' rule?	Probability and Statistics
Can you explain how backpropagation uses the chain rule to get gradients?	Artificial Neural Networks
Can you compute an integral and the derivative of a logarithm?	Integrals and derivatives from calculus

If you can say "yes" to most of these, you are ready; one or two gaps can be closed during the semester.

CONTEXT

Neighbouring Courses



Overlapping Courses

Relation	Course	Shared ground
subfield	This course is a subfield of Deep Learning	GANs, VAEs, diffusion, large language models
builds on	This course builds on Information Theory	KL divergence, likelihood
shares methods	Shares methods with Natural Language Processing and Text Mining	text generation
shares methods	Shares methods with Probabilistic Graphical Models	latent variable models, variational inference
shares methods	Shares methods with Computer Graphics	image and scene synthesis

Where It Differs (1/2)

Criterion	Generative Models	Deep Learning	Graphical Models
Core question	How does a model generate new samples that look like real data?	How do many-layered networks learn useful representations from raw data?	How do we model and query many interdependent, uncertain variables?
Approach	learning the data distribution	multilayer representation learning	probability and graphs combined
Input	huge, mostly unlabeled data	raw data: images, text, audio	variables, dependencies, data
Output	new samples: images, text, audio	learned representation and prediction	conditional probability, best explanation
Success measure	realism and diversity (FID etc.)	accuracy, loss, scaling	likelihood, inference accuracy
Math load	high	high: derivatives, matrices, probability	high: probability
Data needs	very large	large; needs GPUs	moderate

- **Deep Learning:** Deep learning is the general toolkit and mostly makes predictions; this course uses those tools to generate new samples from a data distribution.
- **Graphical Models:** Graphical models build the structure between variables explicitly; the deep generative models here usually learn it implicitly.

Where It Differs (2/2)

Criterion	Generative Models	NLP	Computer Graphics
Core question	How does a model generate new samples that look like real data?	How does a computer analyze, understand and generate human language?	How is a realistic image produced from the description of a scene?
Approach	learning the data distribution	linguistics, statistics and deep learning	geometry and light simulation
Input	huge, mostly unlabeled data	text	scene: models, lights, camera
Output	new samples: images, text, audio	label, translation, summary, answer	image or animation
Success measure	realism and diversity (FID etc.)	accuracy, BLEU, human evaluation	realism, frame rate
Math load	high	medium	high: linear algebra, geometry
Data needs	very large	large text corpora	none

- **NLP:** NLP also covers parsing and understanding language; this course looks at the shared principles of generating text, images and audio.
- **Computer Graphics:** Computer graphics computes images from a physical description of a scene; generative models produce images by learning from data.

PRACTICE

Applications

Text-to-image
Pictures and illustrations from a prompt.

Chat and writing
Text generation with large language models.

Audio synthesis
Natural-sounding speech and music.

Data augmentation
Synthetic samples for classes with little data.

Drug design
Proposing new candidate molecules.

Image editing
Inpainting missing regions, super-resolution.

PRACTICE

Risks and Ethics

Deepfakes

A person's face and voice are imitated without consent and used for fraud, blackmail and election disinformation.

Training data without consent

Web-scraped datasets such as LAION-5B contain works without their owners' permission; copyright lawsuits are ongoing.

Model collapse

Models retrained on generated data lose the tails of the distribution over generations, and their outputs grow uniform.

The metric illusion

Metrics such as FID do not always agree with human judgement; a model tuned to the metric can look worse to the eye.

In depth: Explainable AI and AI Ethics.

PRACTICE

Tools and Datasets

Tools and libraries

PyTorch | Hugging Face Diffusers | Hugging Face Transformers | JAX | Stable Diffusion | Weights & Biases

Well-known datasets

MNIST | CelebA | CIFAR-10 | LSUN | LAION-5B | FFHQ

PRACTICE

Common Misconceptions

Common belief	In fact
A generative model copies its training data.	The goal is to learn the distribution; memorization can still happen and must be checked.
Generative AI is just chatbots.	Models that generate images, audio, video, molecules and code are also part of this course.
Realistic-looking output is correct output.	A model can produce plausible but wrong content; correctness is checked separately.
GANs are outdated; diffusion is always better.	Each family trades off speed, quality and diversity differently; a GAN generates in one step.

NEXT STEPS

After This Course

Courses that build on or extend this one

This course is mostly an end point: it builds on the others.

Related careers

Generative AI engineer | Researcher | Machine learning engineer | Creative technologist

After generative models you can take Natural Language Processing for text generation, Computer Graphics for image and scene synthesis, or Explainable AI and AI Ethics for safe and responsible use of models. Information Theory and Probabilistic Graphical Models deepen the theoretical basis.

NEXT STEPS

Further Reading

- Foster, *Generative Deep Learning*, 2nd ed., O'Reilly, 2023.
- Tomczak, *Deep Generative Modeling*, Springer, 2022.
- Goodfellow et al., "Generative Adversarial Nets", *Advances in Neural Information Processing Systems* 27, 2014.
- Kingma and Welling, "Auto-Encoding Variational Bayes", *ICLR*, 2014.
- Ho, Jain and Abbeel, "Denoising Diffusion Probabilistic Models", *Advances in Neural Information Processing Systems* 33, 2020.

The End

Chapter 13: Generative Models

← Previous chapter · Map of the field · Next chapter →