

# Week 0e: Generative AI

## The Creation Challenge

Machine Learning for Smarter Innovation

BSc-Level Course

October 6, 2025



## Traditional ML:

“What is this?”

- Email: spam or not?
- Image: cat or dog?
- Text: positive sentiment?
- Patient: high risk?

**Limitation:** Analysis only

## Generative AI:

“Create something new”

- Generate: realistic images
- Write: coherent articles
- Compose: original music
- Design: novel molecules

**Power:** Creation & innovation

The fundamental shift: from understanding existing data to creating new possibilities

## Discriminative Models

Learn:  $P(y|x)$

“Given input  $x$ , what’s the label  $y$ ?”

### Examples:

- Logistic regression
- Random Forest
- Neural networks (classification)
- SVM

**Goal:** Decision boundaries

Discriminative: “Is this a cat?” — Generative: “Draw me a cat”

## Generative Models

Learn:  $P(x)$  or  $P(x, y)$

“What does the data distribution look like?”

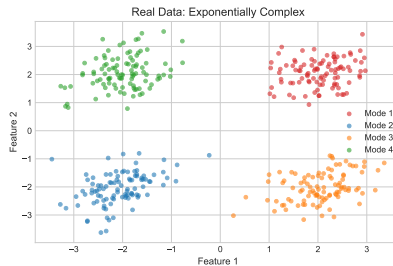
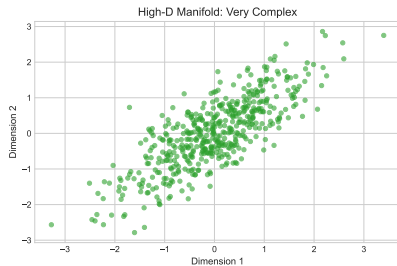
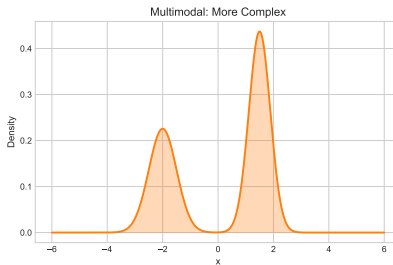
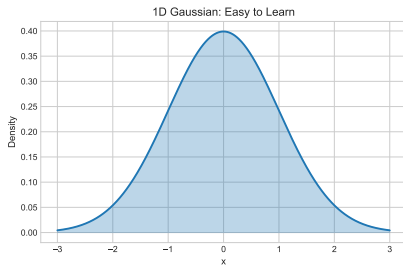
### Examples:

- Gaussian Mixture Models
- Variational Autoencoders
- GANs
- Diffusion models

**Goal:** Data generation

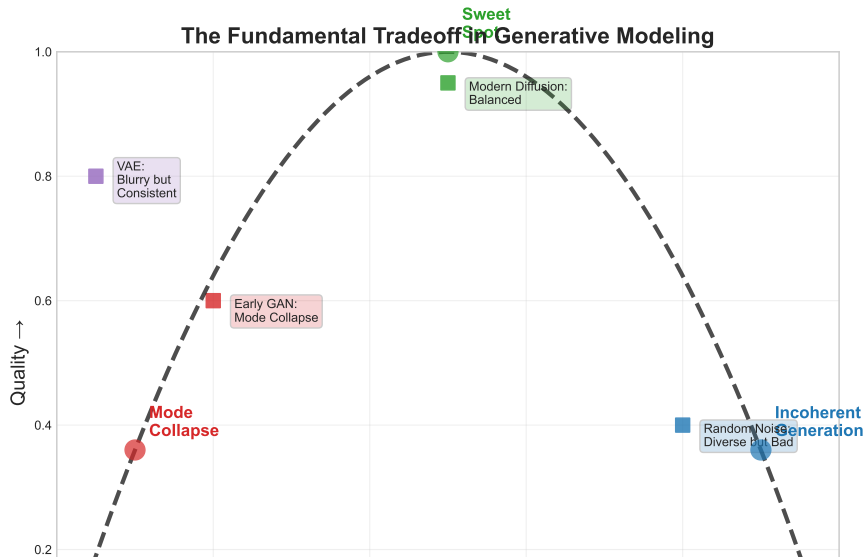
# The Hard Problem

Why Generation is Fundamentally Difficult



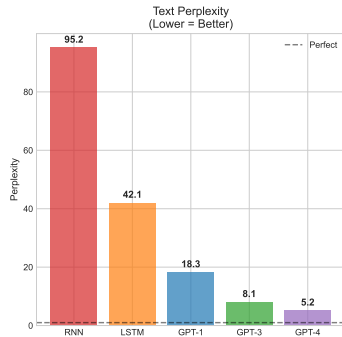
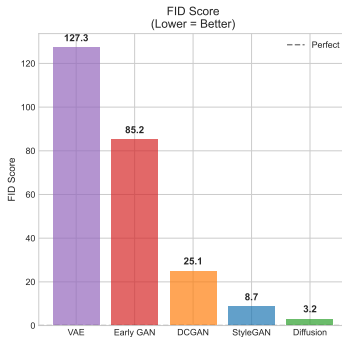
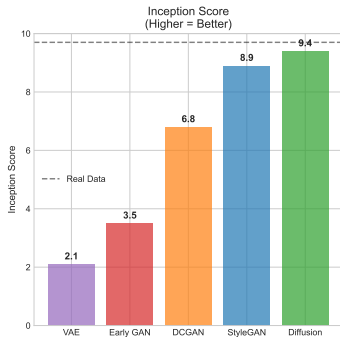
# The Fundamental Tradeoff

Quality vs Diversity Dilemma



# Measuring Generation Quality

## Metrics for Evaluating Generative Models



### Inception Score (IS)

- Range: 1-1000+
- Higher = better
- Quality & diversity
- $IS = \exp(E[KL(p(y|x)||p(y))])$

### FID Score

- Range: 0-500+
- Lower = better
- Feature distance
- Real vs generated

### Perplexity (Text)

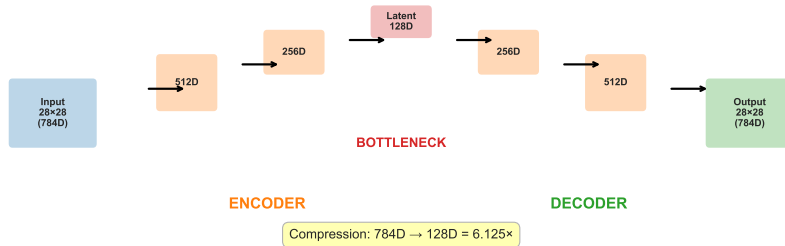
- Range: 1-10,000+
- Lower = better
- Predictability
- Language fluency

Quantitative evaluation: IS=300+ (excellent), FID<sub>i</sub>10 (photorealistic), Perplexity<sub>i</sub>20 (human-like text)

# Autoencoders: The Foundation

## Learning Compressed Representations

### Autoencoder Architecture: Compression Through Reconstruction



#### Encoder

- 784D  $\rightarrow$  128D
- $z = f_{enc}(x)$

Force information through bottleneck, learn to reconstruct

#### Latent Space

- Bottleneck: 128D
- Key features

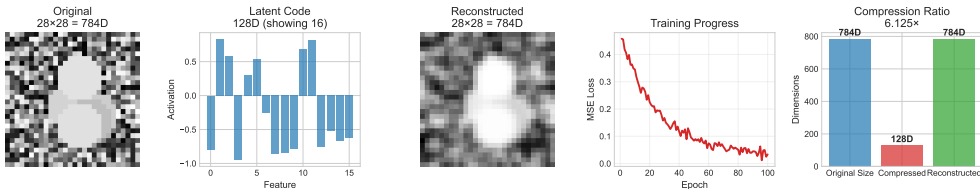
#### Decoder

- 128D  $\rightarrow$  784D
- $\hat{x} = f_{dec}(z)$



# Worked Example: MNIST Compression

From 784 Pixels to 128 Features



## Architecture Details:

- Input:  $28 \times 28 = 784$  pixels
- Encoder:  $784 \rightarrow 512 \rightarrow 256 \rightarrow 128$
- Decoder:  $128 \rightarrow 256 \rightarrow 512 \rightarrow 784$
- Activation: ReLU (hidden), Sigmoid (output)

## Training Process:

- Loss:  $L = ||x - \hat{x}||^2$
- Optimizer: Adam,  $lr=0.001$
- Epochs: 100
- Compression ratio:  $784/128 = 6.125\times$

Reconstruction loss: MSE drops from 0.45 to 0.03 over 100 epochs

# Autoencoder Successes

What Works Well

Autoencoder Successes  
Visualization Placeholder  
(Chart 12)

## [+] SUCCESSES:

- Dimensionality reduction: 784D  $\rightarrow$  128D

## Quantitative Results:

- MSE: 0.031, Compression: 6.125x

# Autoencoder Limitations

The Generation Problem

Autoencoder Failures  
Visualization Placeholder  
(Chart 13)

**[ - ] FAILURES:**

- Blurry outputs (averaging)

Generation Metrics:

Metric	Score
IS	2.1

# Root Cause Analysis

Why Autoencoders Generate Poorly

Averaging Problem  
Visualization Placeholder  
(Chart 14)

## The Averaging Problem:

- Loss:  $L = ||x - \hat{x}||^2$

## Mathematical Insight:

- $\hat{x} = \arg \min E[||x - \hat{x}||^2]$

# Variational Autoencoders (VAEs)

The Probabilistic Solution

Vae Framework  
Visualization Placeholder  
(Chart 15)

## Key Innovation:

- Encode to distribution, not point

## VAE Loss:

$$\mathcal{L} = -E[\log p_{\theta}(x|z)] + KL(q||p)$$

# Human Learning Analogy

How Artists Develop Mastery

Artist Learning Process  
Visualization Placeholder  
(Chart 16)

## Traditional Art Education:

- Student creates artwork

## Key Insights:

- Adversarial feedback drives improvement

# Two Revolutionary Approaches

Beyond VAEs to Better Generation

Two Approaches  
Visualization Placeholder  
(Chart 17)

## Approach 1: Adversarial

- Two networks compete

## Approach 2: Diffusion

- Iterative denoising

# GANs: The Forger vs Detective Game

Adversarial Training in Plain English

Forger Detective Analogy

Visualization Placeholder

(Chart 18)

## Forger (Generator):

- Creates fakes from noise

## Detective (Discriminator):

- Examines: real or fake?



# Diffusion: The Reverse Corruption Process

Denoising in Plain English

Reverse Corruption Analogy  
Visualization Placeholder  
(Chart 19)

**Forward (Corruption):**

- Clean image  $\rightarrow$  pure noise

**Reverse (Generation):**

- Pure noise  $\rightarrow$  clean image

# GAN Dynamics: Geometric View

Understanding the Adversarial Process

Gan Geometric Dynamics  
Visualization Placeholder  
(Chart 20)

## Generator:

- Maps noise  $z$  to data  $x$

## Discriminator:

- Separates real from fake

# GAN Training: Step-by-Step Example

Real Loss Values from MNIST Training

Gan Training Walkthrough  
Visualization Placeholder  
(Chart 21)

**Epoch 1:**

• D\_loss: 1.386

**Epoch 100:**

• D\_loss: 0.695

# Diffusion Mathematical Framework

Forward and Reverse Processes

Diffusion Mathematics  
Visualization Placeholder  
(Chart 22)

**Forward (Fixed):**

$$q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1 - \beta_t}x_{t-1}, \beta_t I)$$

Noise schedule:  $\beta_t$  (0.0001 : 0.02)

**Reverse (Learned):**

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta, \Sigma_\theta)$$

Network  $\epsilon_\theta$  predicts noise

# Latent Space Interpolation

Smooth Transitions in Generated Content

Latent Interpolation  
Visualization Placeholder  
(Chart 23)

## GAN Interpolation:

- Sample  $z_1, z_2 \sim \mathcal{N}(0, I)$

## Applications:

- Style transfer, face morphing

# Diffusion Denoising Visualization

From Noise to Image in 1000 Steps

Denoising Steps  
Visualization Placeholder  
(Chart 24)

## Key Time Steps:

- $T=1000$ : Pure noise

## Process Control:

- Guidance scale

# Why Adversarial Training Works

The Mathematical Guarantee

Adversarial Theory  
Visualization Placeholder  
(Chart 25)

## Theory:

- Minimax convergence

## Benefits:

- Sharp, realistic images

Quality Metrics Over Time

Visualization Placeholder

(Chart 26)

## GAN Progress:

- Start: IS=1.2, FID=450

## Diffusion Progress:

- 100k: FID=200



# Implementation: Stable Diffusion API

Production-Ready Generative AI

Stable Diffusion Api  
Visualization Placeholder  
(Chart 27)

# The Generative AI Landscape

Four Fundamental Approaches

Generative Landscape  
Visualization Placeholder  
(Chart 28)

**VAEs:** Probabilistic, smooth latent, blurry

**GANs:** Adversarial, sharp outputs, unstable

Each approach has unique strengths - modern systems combine techniques

**Diffusion:** Iterative denoising, high quality, slow

**Transformers:** Sequential, excellent text, scalable

# Fundamental Trade-offs

No Free Lunch in Generative Modeling

Generative Tradeoffs  
Visualization Placeholder  
(Chart 29)

## Training Stability:

- VAEs, Diffusion: Stable
- GANs: Unstable

## Quality:

- Diffusion, GANs: Excellent
- VAEs: Blurry

Modern Applications  
Visualization Placeholder  
(Chart 30)

## Image Generation:

- DALL-E 3, Midjourney
- Stable Diffusion, Firefly

## Text Generation:

- GPT-4, Claude, Gemini
- Llama 2 (open)

# Summary & Ethical Considerations

Power and Responsibility in Generative AI

Ethics Summary  
Visualization Placeholder  
(Chart 31)

## Capabilities:

- Realistic images from text
- Human-like writing

## Challenges:

- Deepfakes, misinformation
- Copyright issues