

Week 0e: Generative AI

The Creation Challenge

Machine Learning for Smarter Innovation

BSc-Level Course

October 6, 2025

The Creation Challenge

Moving Beyond Classification

Traditional ML: “What is this?”

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

Limitation: Analysis only

Generative AI: “Create something new”

- Generate realistic images
- Write coherent articles
- Compose music

Power: Creation & innovation

Fundamental shift: from understanding to creating

Discriminative Models

Learn: $P(y|x)$

Examples:

- Logistic regression
- Random Forest, SVM

Goal: Decision boundaries

Generative Models

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

Examples:

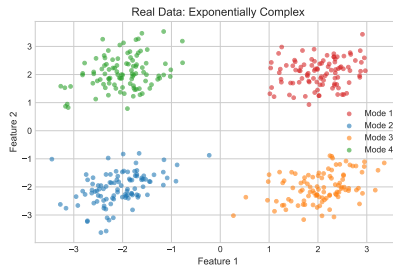
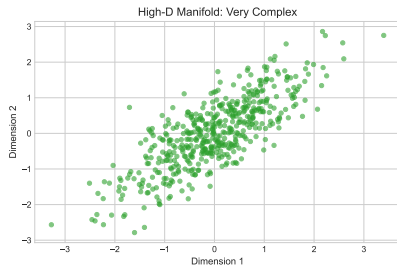
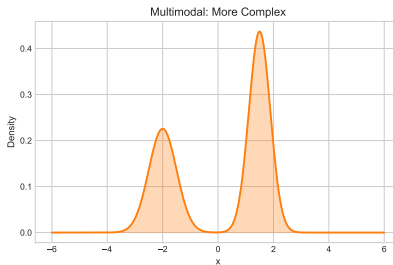
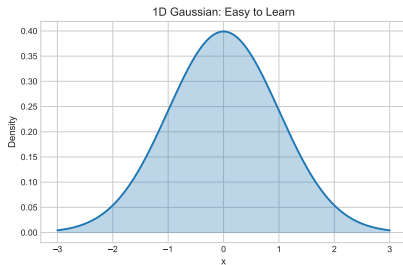
- VAEs, GANs
- Diffusion models

Goal: Data generation

Discriminative: "Is this X?" — Generative: "Create X"

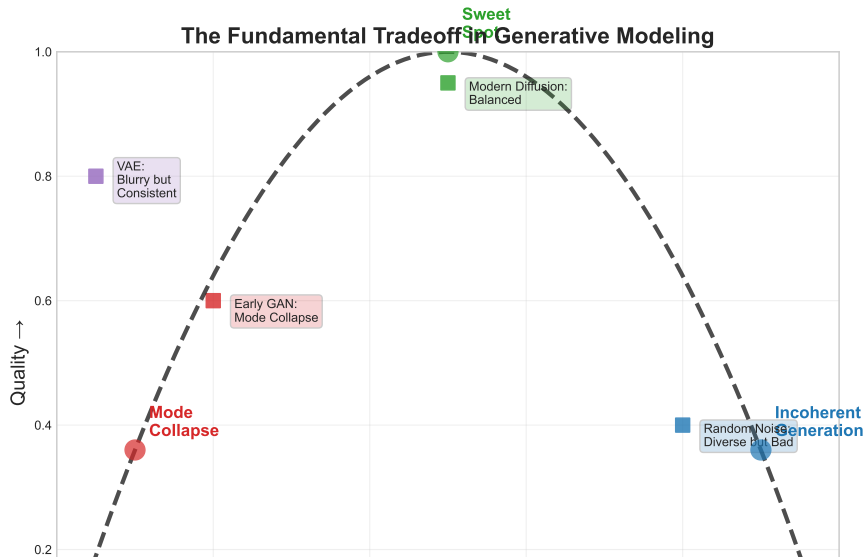
The Hard Problem

Why Generation is Fundamentally Difficult



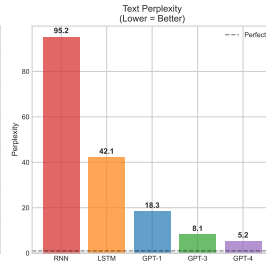
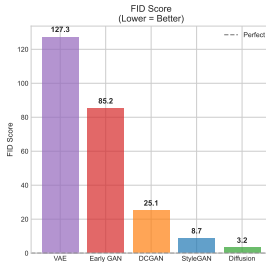
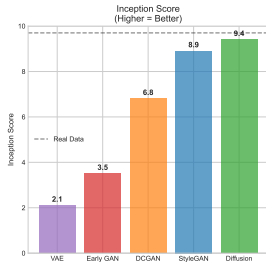
The Fundamental Tradeoff

Quality vs Diversity Dilemma



Measuring Generation Quality

Metrics for Evaluating Generative Models



Inception Score

- Higher = better
- Quality & diversity

FID Score

- Lower = better
- Feature distance

Perplexity (Text)

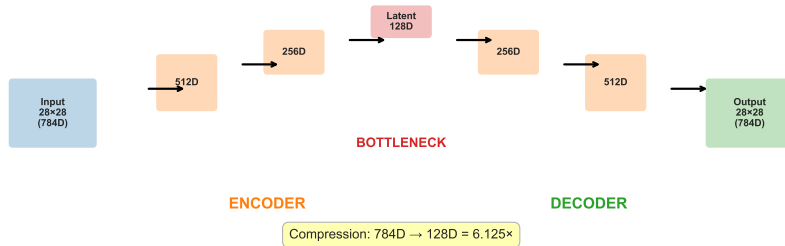
- Lower = better
- Language fluency

IS₃₀₀ excellent, FID₁₀ photorealistic, PPL₂₀ human-like

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

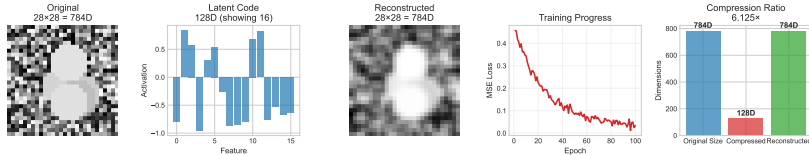
- Bottleneck: 128D
- Key features

Decoder

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

Worked Example: MNIST Compression

From 784 Pixels to 128 Features



Architecture:

- Input: 784 pixels
- Encoder: 784 \rightarrow 128
- Decoder: 128 \rightarrow 784

Training:

- Loss: $L = ||x - \hat{x}||^2$
- Optimizer: Adam
- Compression: 6.125x

MSE drops 0.45 \rightarrow 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: β_t (0.0001 : 0.02)

Reverse (Learned):

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

Network ϵ_θ 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)

Basic Usage:

```
import requests

response = requests.post(
    api_url,
    headers={"Authorization": key}).
```

Parameters:

- `cfg_scale`: Adherence (1-20)
- `steps`: Quality (10-150)
- `seed`: Reproducible

Cost: \$0.004 per image

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

Choosing Your Generative Model

Decision Framework for Practitioners

Decision Criteria:

1. What are you generating?

- Images: Diffusion or GAN
- Text: Transformer (GPT family)
- Structured data: VAE
- Multimodal: Diffusion + Transformer

2. Data size?

- < 10k samples: VAE (stable)
- 10k-100k: GAN or VAE
- > 100k: Diffusion or Transformer

3. Priority?

- Quality: Diffusion (FID ↓)
- Speed: GAN (single pass)
- Stability: VAE (always converges)
- Control: Diffusion (guidance)

Recommendation Table:

Use Case	Best	Why
Photorealistic	Diffusion	Quality
Fast prototype	GAN	Speed
Data augment	VAE	Stable
Text gen	Transformer	Sequential
Style transfer	VAE	Interpolate
Research	VAE	Interpret

When NOT to Use:

- VAE: Need sharp images
- GAN: Limited data, need stability
- Diffusion: Real-time inference required
- All: Insufficient compute resources

Model selection requires balancing quality, speed, stability against problem constraints

Common Pitfalls: What Can Go Wrong

Failure Modes and Solutions

VAE Pitfalls

1. Posterior Collapse

- Symptom: KL term $\rightarrow 0$
- Decoder ignores latent
- Fix: -VAE, KL warm-up

2. Blurry Outputs

- Inherent to MSE loss
- Reconstruction averages
- Fix: Perceptual loss, adversarial

GAN Pitfalls

1. Mode Collapse

- Generator ignores diversity
- Produces limited variety
- Fix: Minibatch discrimination, unrolled GAN

2. Training Instability

- Loss oscillates wildly
- No convergence
- Fix: Wasserstein loss, spectral norm, TTUR

Diffusion Pitfalls

1. Slow Sampling

- 1000 steps = slow
- Latency bottleneck
- Fix: DDIM, distillation, 50 steps

2. Memory Intensive

- High-res = huge memory
- Training cost
- Fix: Latent diffusion, gradient checkpointing

Each approach has characteristic failure modes requiring specific mitigation strategies

Generative AI Best Practices

From Research to Production

Training Best Practices:

1. Start Simple

- Low resolution first (64x64 before 1024x1024)
- Small models before large
- Validate on toy datasets

2. Monitor Obsessively

- Log every 100 steps
- Visual inspection of samples
- Track FID/IS throughout
- Watch for collapse/divergence

3. Use Pretrained Models

- Transfer learning saves weeks
- Fine-tune from Stable Diffusion
- Leverage foundation models

4. Ablation Studies

- Test each component independently
- Measure contribution to performance
- Simplify when possible

Deployment Best Practices:

1. Quality Control

- Human-in-the-loop review
- NSFW content filtering
- Bias detection systems
- Watermarking for traceability

2. Performance Optimization

- Model quantization (FP16, INT8)
- Distillation for speed
- Caching common generations
- Batch processing

3. Safety Measures

- Rate limiting to prevent abuse
- Content moderation
- Prompt injection defenses
- User consent and attribution

4. Continuous Improvement

- Collect user feedback
- Retrain on curated data

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