# Week 0e: Generative AI

## The Creation Challenge

Machine Learning for Smarter Innovation

BSc-Level Course

October 6, 2025

# Outline

# 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

# Mathematical Foundation

## Two Approaches to Learning

**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:**

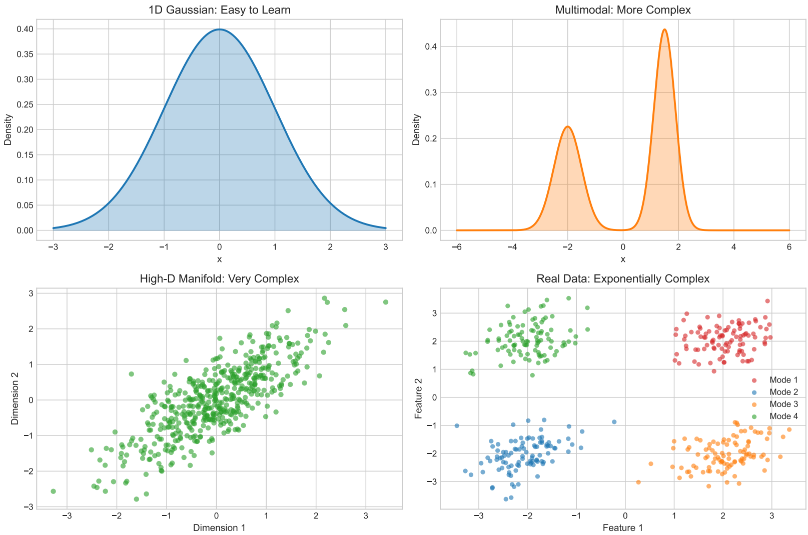
- VAEs, GANs
- Diffusion models

**Goal:** Data generation

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

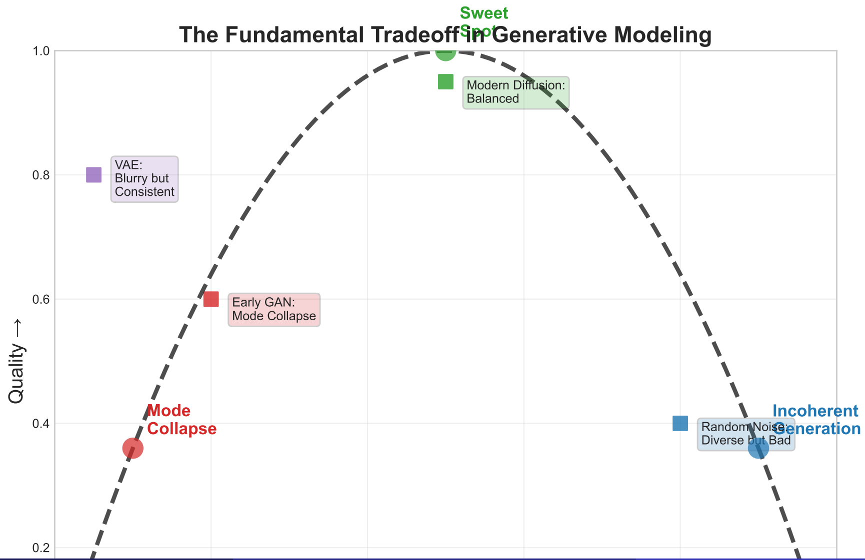
# The Hard Problem

## Why Generation is Fundamentally Difficult



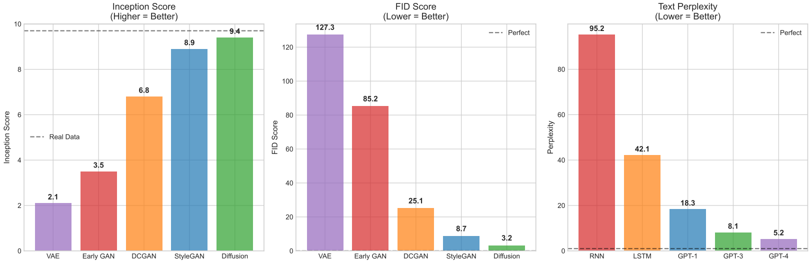
# The Fundamental Tradeoff

## Quality vs Diversity Dilemma



# Measuring Generation Quality

Metrics for Evaluating Generative Models




**Inception Score**

- Higher = better
- Quality & diversity

**FID Score**

- Lower = better
- Feature distance

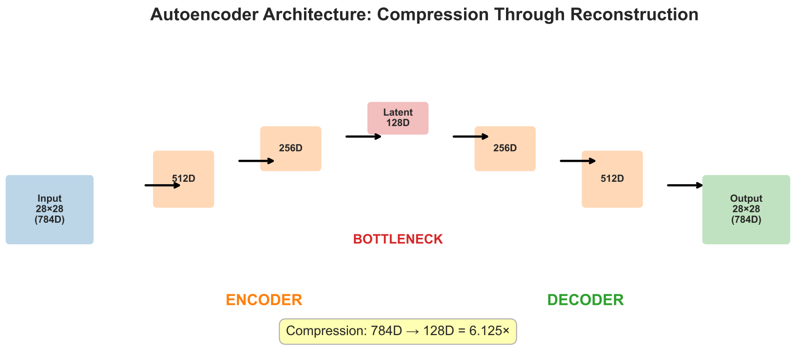
**Perplexity (Text)**

- Lower = better
- Language fluency

IS¿300 excellent, FID¡10 photorealistic, PPL¡20 human-like

# Autoencoders: The Foundation

Learning Compressed Representations




**Encoder**

- 784D -¿ 128D
- $z = f_{enc}(x)$

**Latent Space**

- Bottleneck: 128D
- Key features

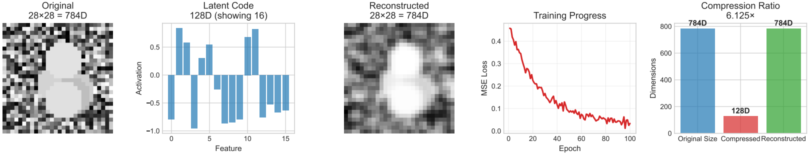
**Decoder**

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

Force information through bottleneck, learn to reconstruct

# Worked Example: MNIST Compression

From 784 Pixels to 128 Features




**Architecture:**

- Input: 784 pixels
- Encoder: 784 -¿ 128
- Decoder: 128 -¿ 784

**Training:**

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

MSE drops 0.45 -¿ 0.03 over 100 epochs

# Autoencoder Successes

## What Works Well

Autoencoder Successes

Visualization Placeholder

(Chart 12)

**[+] SUCCESSES:**

- Dimensionality reduction: 784D -¿ 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:**

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

Vae Framework

Visualization Placeholder

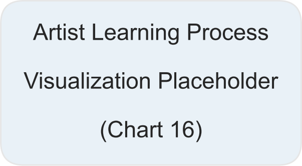
(Chart 15)

**Key Innovation:**

- Encode to distribution, not point

**VAE Loss:**

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




**Traditional Art Education:**

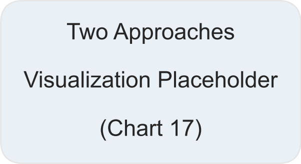
- Student creates artwork

**Key Insights:**

- Adversarial feedback drives improvement

# Two Revolutionary Approaches

## Beyond VAEs to Better Generation




**Approach 1: Adversarial**

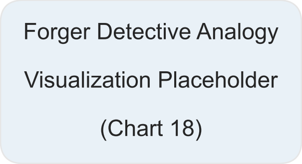
- Two networks compete

**Approach 2: Diffusion**

- Iterative denoising

# GANs: The Forger vs Detective Game

Adversarial Training in Plain English




**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 -¿ pure noise

**Reverse (Generation):**

- Pure noise -¿ clean image

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

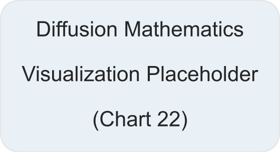
(Chart 21)

**Epoch 1:**

- D_loss: 1.386

**Epoch 100:**

- D_loss: 0.695




**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

# Experimental Validation

## Quality Metrics Throughout Training

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


**VAEs:** Probabilistic, smooth latent, blurry
**GANs:** Adversarial, sharp outputs, unstable

**Diffusion:** Iterative denoising, high quality, slow
**Transformers:** Sequential, excellent text, scalable

Each approach has unique strengths - modern systems combine techniques

# 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

# State-of-the-Art Applications

Production Generative AI Systems

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