Week 0e: Generative AI

The Creation Challenge

Machine Learning for Smarter Innovation

BSc-Level Course

October 7, 2025

Outline

1. Act 1: The Challenge

2. Act 2: Variational Autoencoders

3. Act 3: Adversarial & Diffusion

4. Act 4: Synthesis

The Creation Challenge

Moving Beyond Classification

Traditional ML: "What is this?"

- Email spam detector: Classify existing emails
- Medical diagnosis: Analyze X-ray images
- Sentiment analysis: Judge customer reviews

Limitation: Only analyzes, never creates

Generative AI: "Create something new"

- Generate phishing emails for security training
- Synthesize medical images for rare diseases
- Write product descriptions automatically
- Compose music for video backgrounds

Power: Creation enables innovation

Fundamental shift: from pattern recognition to content generation

Mathematical Foundation

Two Approaches to Learning

Discriminative Models
Learn: $P(y|x)$ - Conditional probability
What it does:

- Given x, predict label y
- Learns decision boundaries
- Divides input space

Examples: Logistic, RF, SVM
Can sample new x? NO - only classifies existing data

Generative Models
Learn: $P(x)$ - Joint or marginal distribution
What it does:

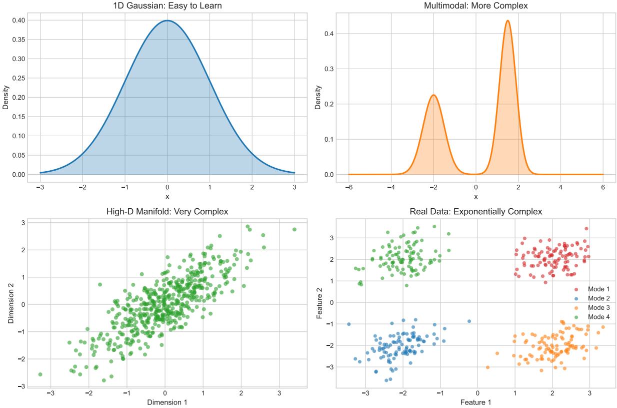
- Models entire data distribution
- Samples new $x \sim P(x)$
- Creates novel instances

Examples: VAEs, GANs, Diffusion
Can sample new x? YES - generates from distribution

Key distinction: Discriminative draws boundaries, Generative learns distributions enabling sampling

The Hard Problem

Why Generation is Fundamentally Difficult



Challenges:

- High-dimensional spaces
- Multimodal distributions

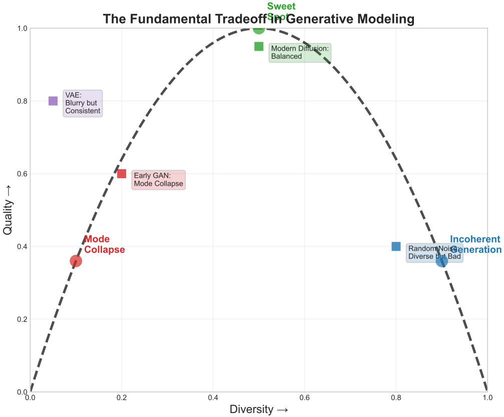
Requirements:

- Capture all patterns
- Maintain realism

Real data lives on complex manifolds - learning full distribution is exponentially hard

The Fundamental Tradeoff

Quality vs Diversity Dilemma

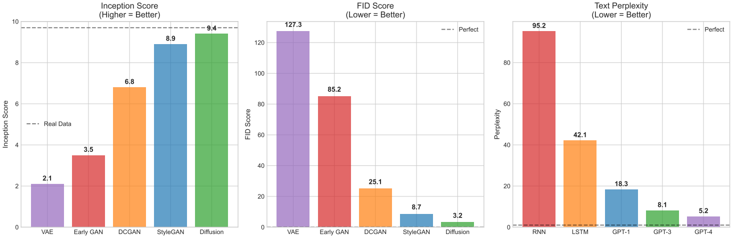


High Quality: Mode collapse, repetitive
Balanced: Realistic variety
High Diversity: Unrealistic

Realistic AND diverse remains the central challenge

Measuring Generation Quality

Metrics for Evaluating Generative Models



Inception Score (IS)

- Range: 1-1000
- Higher = better
- Quality & diversity

Interpretation:

- $>$300: Excellent
- 100-300: Good
- $<$100: Poor

FID Score

- Range: 0-500
- Lower = better
- Feature distance

Interpretation:

- $<$10: Photorealistic
- 10-50: Good quality
- $>$50: Noticeable artifacts

Perplexity (Text)

- Range: 1-10,000
- Lower = better
- Predictability

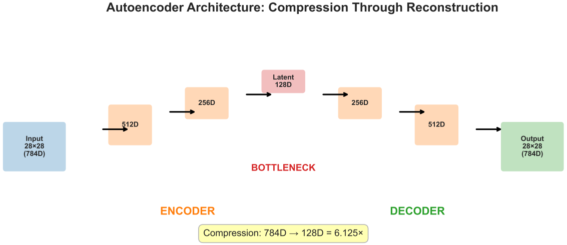
Interpretation:

- $<$20: Human-like
- 20-100: Coherent
- $>$100: Gibberish

Quantitative metrics enable objective quality assessment and model comparison

Autoencoders: The Foundation

Learning Compressed Representations


Autoencoder Architecture: Compression Through Reconstruction
Input 28×28 (784D)
512D
256D
Latent 128D
256D
512D
Output 28×28 (784D)
BOTTLENECK
ENCODER
DECODER
Compression: 784D → 128D = 6.125×

Encoder

- 784D -¿ 128D
- Forces selective encoding
- Filters noise

Latent

- 128D bottleneck
- Key features only
- 6.1x compressed

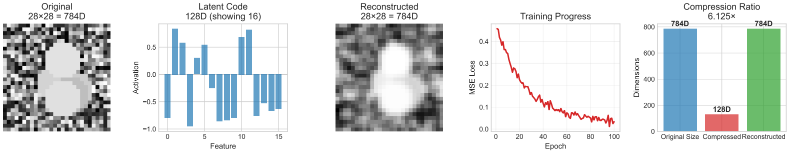
Decoder

- 128D -¿ 784D
- Lossy reconstruction
- Preserves essentials

Bottleneck forces meaningful compression

Worked Example: MNIST Compression

From 784 Pixels to 128 Features



Architecture:

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

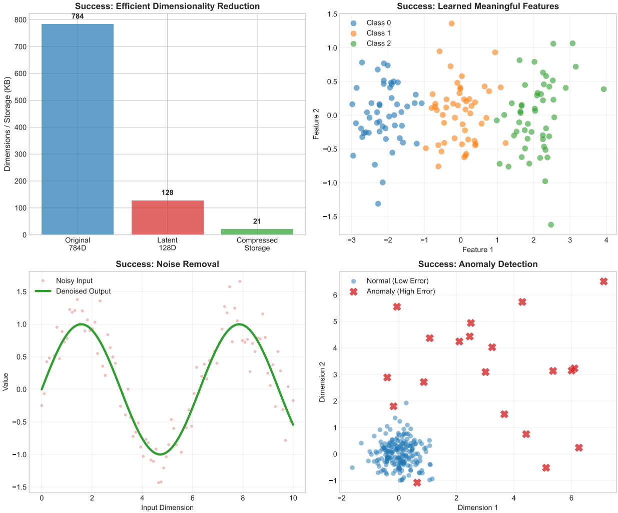
Training:

- 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

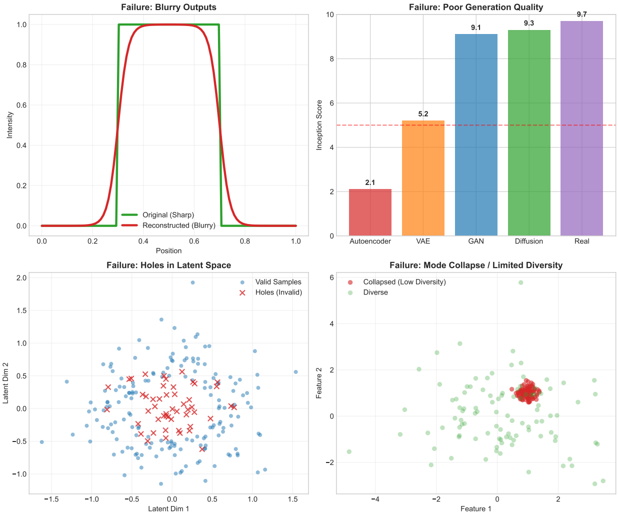


[+] SUCCESSES:

Results:

Autoencoder Limitations

The Generation Problem



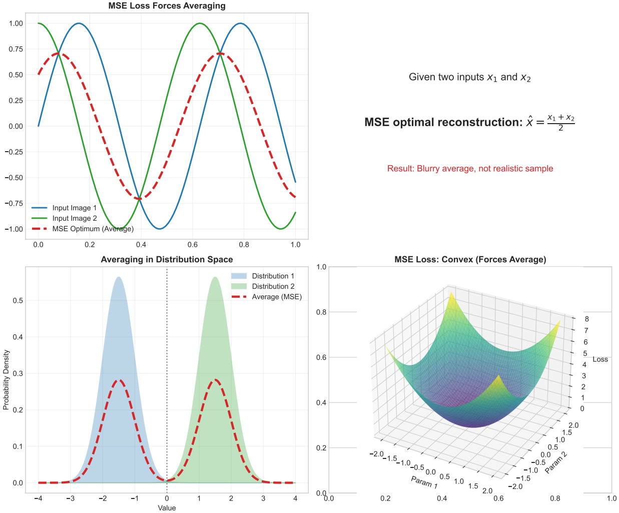
[-] FAILURES:

Metrics:

IS	2.1

Root Cause Analysis

Why Autoencoders Generate Poorly



Given two inputs x_1 and x_2

MSE optimal reconstruction: $\hat{x} = \frac{x_1 + x_2}{2}$

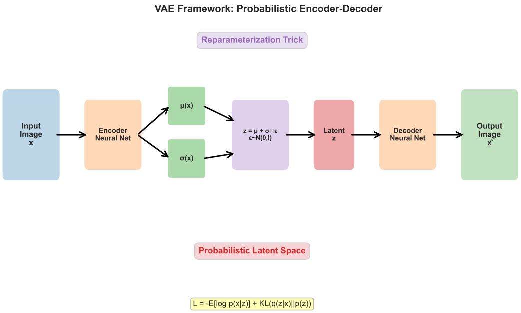
Result: Blurry average, not realistic sample

Problem:

Math:

Variational Autoencoders (VAEs)

The Probabilistic Solution



VAE Framework: Probabilistic Encoder-Decoder

Key Innovation:

- Encode to distribution: $q_\phi(z|x) = \mathcal{N}(\mu, \sigma^2)$
- Sample: $z = \mu + \sigma \odot \epsilon$

Reparameterization:

- Make z deterministic
- Gradient flows

Reparameterization enables gradient optimization

VAE Loss:

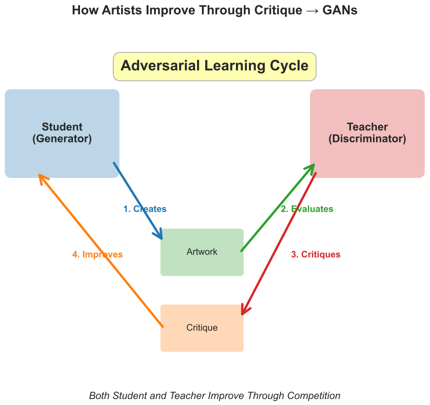
$$\mathcal{L} = -E[\log p(x|z)] + KL(q||p)$$

Two terms:

- Reconstruction
- KL regularization
- β-VAE balances

Human Learning Analogy

How Artists Develop Mastery


How Artists Improve Through Critique → GANs
Adversarial Learning Cycle
Student (Generator)
Teacher (Discriminator)
1. Creates
2. Evaluates
3. Critiques
4. Improves
Artwork
Critique
Both Student and Teacher Improve Through Competition

Art Education:

- Student creates
- Teacher critiques
- Student improves

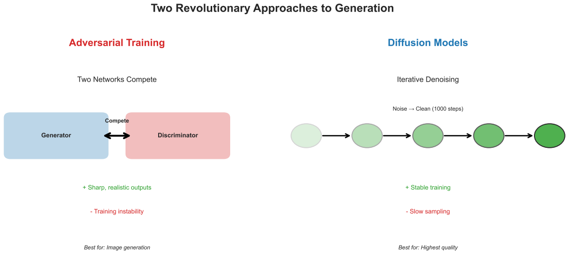
Insights:

- Adversarial feedback drives improvement
- Both improve together

Adversarial learning inspired GANs

Two Revolutionary Approaches

Beyond VAEs to Better Generation



Adversarial

- Two networks compete
- Sharp, realistic

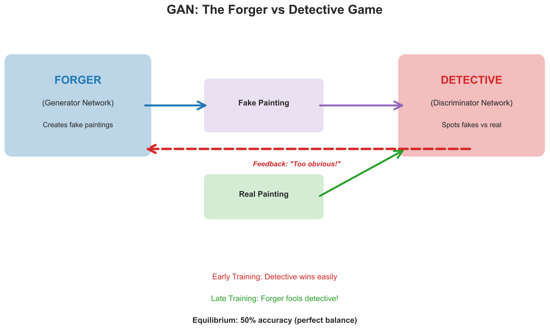
Diffusion

- Iterative denoising
- Stable, controllable

Both address VAE limitations

GANs: The Forger vs Detective Game

Adversarial Training in Plain English


GAN: The Forger vs Detective Game
FORGER
(Generator Network)
Creates fake paintings
Fake Painting
DETECTIVE
(Discriminator Network)
Spots fakes vs real
Feedback: "Too obvious!"
Real Painting
Early Training: Detective wins easily
Late Training: Forger fools detective!
Equilibrium: 50% accuracy (perfect balance)

Forger:

- Creates fakes
- Fools detective

Detective:

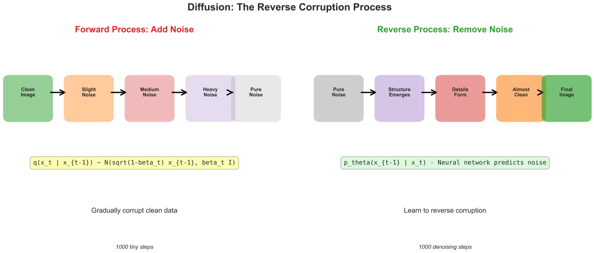
- Examines: real/fake?
- Gets better at detection

Result: Detective can't tell fake from real!

Competition drives both to excellence

Diffusion: The Reverse Corruption Process

Denoising in Plain English


Diffusion: The Reverse Corruption Process
Forward Process: Add Noise
Reverse Process: Remove Noise
Clean Image
Slight Noise
Medium Noise
Heavy Noise
Pure Noise
Pure Noise
Structure Emerges
Details Form
Almost Clean
Final Image
q(x_t | x_{t-1}) ~ N(sqrt(1-beta_t) x_{t-1}, beta_t I)
p_theta(x_{t-1} | x_t) - Neural network predicts noise
Gradually corrupt clean data
Learn to reverse corruption
1000 tiny steps
1000 denoising steps

Forward:

- Clean -¿ noise
- 1000 steps

Key: Learn to undo corruption

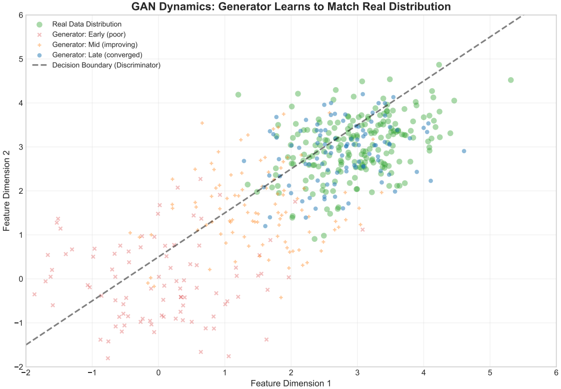
Like sculptor revealing statue

Reverse:

- Noise -¿ clean
- 1000 steps

GAN Dynamics: Geometric View

Understanding the Adversarial Process



Generator:

- Maps z to x
- Loss: $-\log D(G(z))$

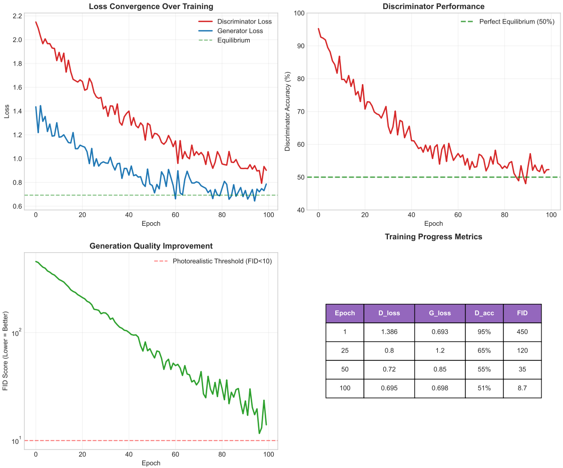
Discriminator:

- Separates real/fake
- Loss: $-\log D(x) - \log(1 - D(G))$

Equilibrium: Generator = Real, D accuracy = 50%

GAN Training: Step-by-Step Example

Real Loss Values from MNIST Training



Training Progress Metrics

Epoch	D_loss	G_loss	D_acc	FID
1	1.386	0.693	95%	450
25	0.8	1.2	65%	120
50	0.72	0.85	55%	35
100	0.695	0.698	51%	8.7

Epoch 1:

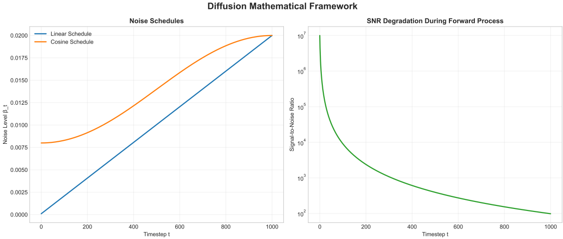
- D: 1.386, G: 0.693
- Images: noise

Epoch 100:

- D: 0.695, G: 0.698
- Images: realistic

Diffusion Mathematical Framework

Forward and Reverse Processes



Forward:

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

Noise Schedule:

- Linear: 0.0001 -¿ 0.02
- Cosine: Variable rate
- Matters: Smooth degradation

Reverse:

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

Training:

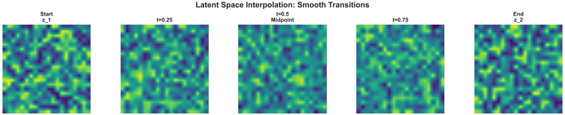
$$L = E[||\epsilon - \epsilon_\theta(x_t, t)||^2]$$

Intuition: Predict noise, subtract it

Linear noise schedule works for most cases

Latent Space Interpolation

Smooth Transitions in Generated Content



Method:

- Sample z_1, z_2
- Interpolate: $z_t = (1-t)z_1 + tz_2$
- Generate: $x_t = G(z_t)$

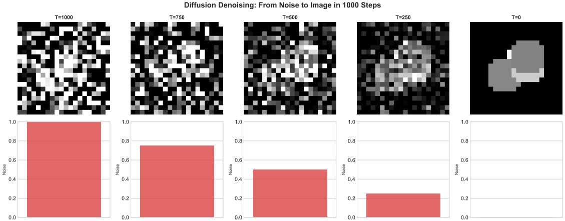
Applications:

- Style transfer
- Face morphing
- Drug discovery

Meaningful latent spaces enable smooth interpolation

Diffusion Denoising Visualization

From Noise to Image in 1000 Steps



Steps:

- T=1000: Noise
- T=500: Structure
- T=0: High quality

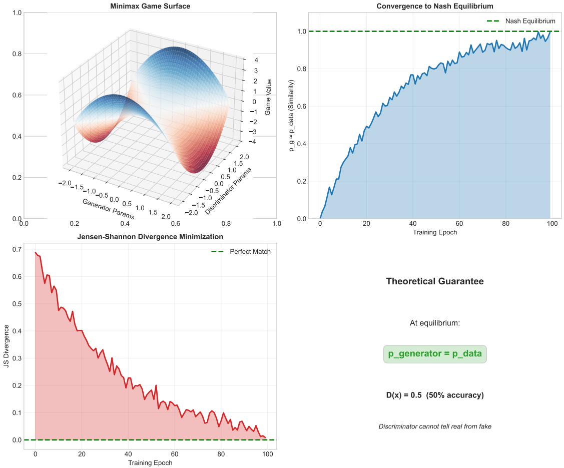
Control:

- Guidance scale
- Step count

Gradual refinement

Why Adversarial Training Works

The Mathematical Guarantee



Minimax Game Surface

Convergence to Nash Equilibrium

Jensen-Shannon Divergence Minimization

Theoretical Guarantee

At equilibrium:

p_generator = p_data

D(x) = 0.5 (50% accuracy)

Discriminator cannot tell real from fake

Theory:

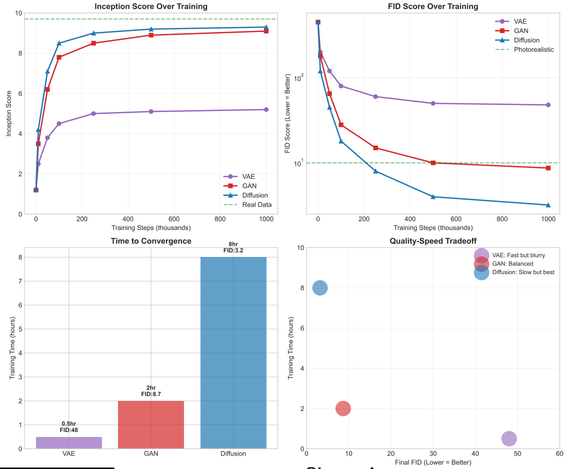
- Minimax convergence
- Equilibrium: $p_g = p_{data}$

Benefits:

- Sharp, realistic
- Fine details

Experimental Validation

Quality Metrics vs Training Progress



Results (MNIST):

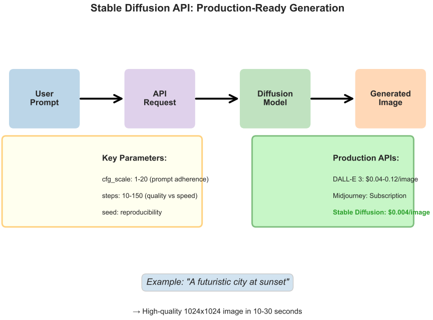
Method	IS	FID	Time
Random	1.0	500	-
VAE	5.2	48	30min
GAN	9.1	9	2hr
Diffusion	9.3	3	8hr

Observations:

- Diffusion: Best
- GAN: 4x faster
- VAE: Fast, blurry

Implementation: Stable Diffusion API

Production-Ready Generative AI



Stable Diffusion API: Production-Ready Generation

User Prompt

API Request

Diffusion Model

Generated Image

Key Parameters:

cfg_scale: 1-20 (prompt adherence)

steps: 10-150 (quality vs speed)

seed: reproducibility

Production APIs:

DALL-E 3: $0.04-0.12/image

Midjourney: Subscription

Stable Diffusion: $0.004/image

Example: "A futuristic city at sunset"

→ High-quality 1024x1024 image in 10-30 seconds

Usage:

```
response = requests.post(
    api_url,
    headers={"Auth": key},
    json={
        "text_prompts": [{"text": "city"}],
        "cfg_scale": 7,
        "steps": 30
    })
```

APIs: DALL-E 3, Midjourney, Stable Diffusion

Parameters:

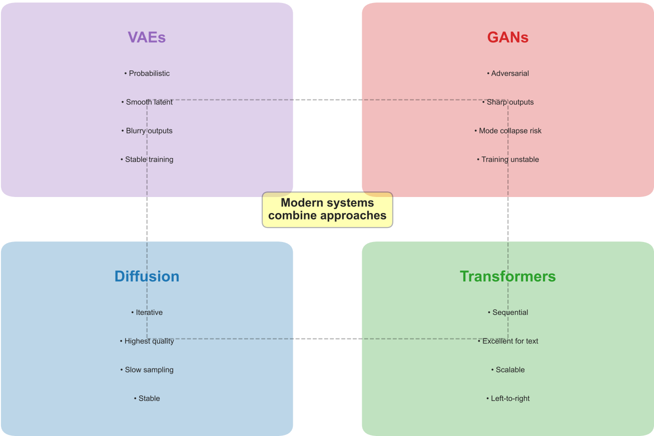
- `cfg_scale`: 1-20
- `steps`: 10-150

Cost: $0.004/image

The Generative AI Landscape

Four Fundamental Approaches

The Generative AI Landscape



VAEs
- Probabilistic
- Smooth latent
- Blurry outputs
- Stable training

GANs
- Adversarial
- Sharp outputs
- Mode collapse risk
- Training unstable

Diffusion
- Iterative
- Highest quality
- Slow sampling
- Stable

Transformers
- Sequential
- Excellent for text
- Scalable
- Left-to-right

Modern systems combine approaches

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

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 ¡ 5)
- 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**
   - KL -¿ 0
   - Fix: β-VAE, warm-up
2. **Blurry**
   - MSE averages
   - Fix: Perceptual loss

GAN Pitfalls

1. **Mode Collapse**
   - Limited variety
   - Fix: Minibatch disc
2. **Unstable**
   - Oscillates
   - Fix: Wasserstein, spectral norm

Diffusion Pitfalls

1. **Slow (1000 steps)**
   - Latency issue
   - Fix: DDIM (50 steps)
2. **Memory**
   - High-res costly
   - Fix: Latent diffusion

Each approach has characteristic failure modes with specific solutions

Generative AI Best Practices

From Research to Production

Training:

1. **Start Simple**
   - Low res first (64x64 before 1024x1024)
   - Validate on toy datasets
2. **Monitor Obsessively**
   - Log every 100 steps
   - Visual sample inspection
   - Track FID/IS
3. **Use Pretrained**
   - Transfer learning saves weeks
   - Fine-tune Stable Diffusion
4. **Ablation Studies**
   - Test components independently

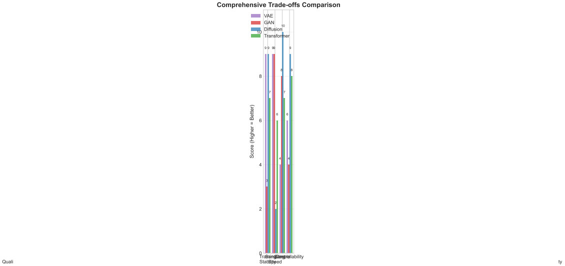
Deployment:

1. **Quality Control**
   - Human-in-the-loop review
   - Content filtering
   - Watermarking
2. **Performance**
   - Quantization (FP16, INT8)
   - Distillation for speed
   - Caching
3. **Safety**
   - Rate limiting
   - Content moderation
   - Prompt injection defenses
4. **Continuous Improvement**
   - User feedback
   - A/B testing

Production requires systematic validation and continuous monitoring

Comprehensive Trade-offs

No Free Lunch in Generative Modeling



Stability:

- VAEs, Diffusion: Stable
- GANs: Unstable

Speed:

- VAEs, GANs: Fast
- Diffusion: Slow

Quality:

- Diffusion, GANs: Excellent
- VAEs: Blurry

Control:

- Diffusion, Transformers: High
- GANs: Limited

Choose based on requirements

State-of-the-Art Applications

Production Generative AI Systems

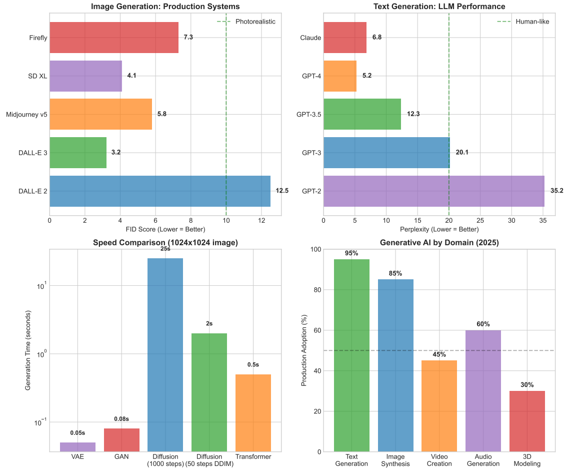


Image:

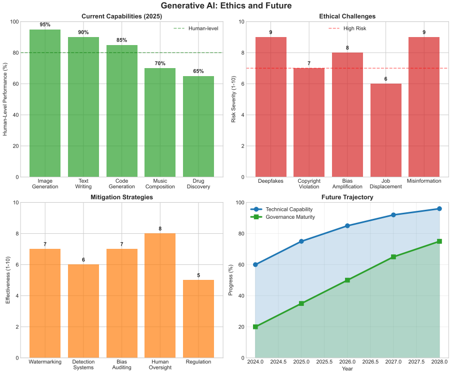
- DALL-E 3, Midjourney
- Stable Diffusion, Firefly

Text:

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

Summary & Future of Generative AI

What We Learned and What's Next



Learned:

- VAEs: Probabilistic, blurry
- GANs: Adversarial, realistic
- Diffusion: Best quality
- Decision framework, pitfalls

Future:

Ethics:

- Deepfakes, copyright
- Bias, displacement

Solutions:

- Watermarking, auditing
- Governance