



# Generative AI Model Families

A comprehensive guide to understanding the core architectures powering  
Gen AI

Created By : Kumar Siddharth Nirmal

# What We'll Explore



## Transformers

The language masters behind ChatGPT and modern AI



## GANs

Competing networks that create stunning visuals



## VAEs

Encoding and reconstructing data in creative ways



## Diffusion Models

From noise to art in iterative steps

# Transformers: The Language Powerhouses

## Core Principle

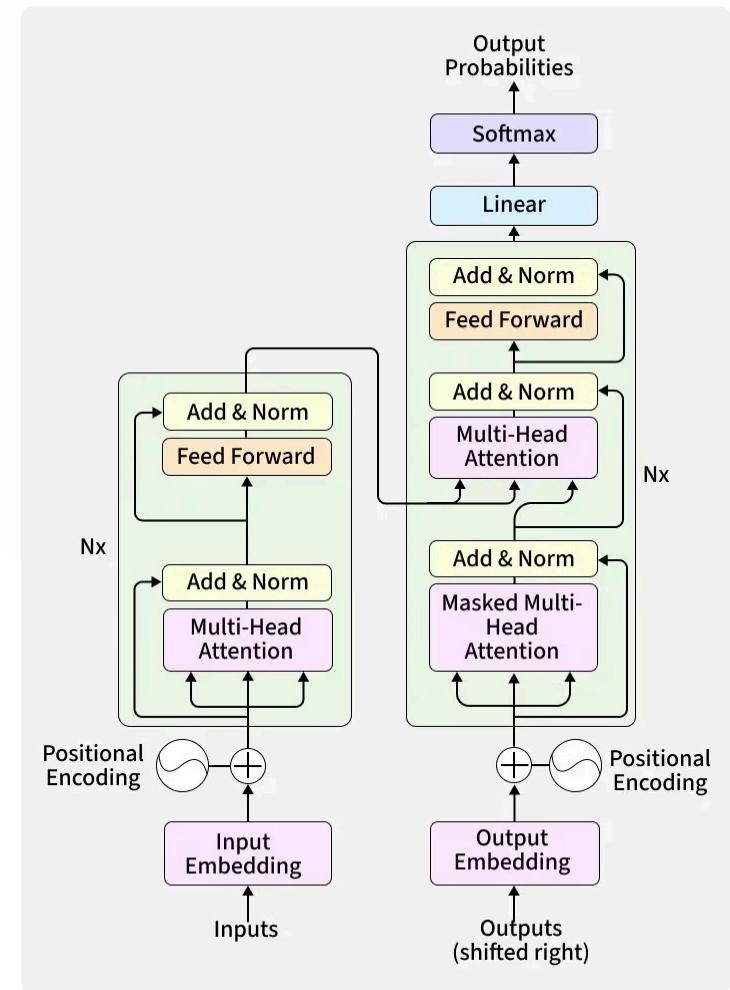
Transformers predict the next token based on context using attention mechanisms. They excel at understanding relationships between words, no matter how far apart they are in a sentence.

## How It Generates

Sequential token prediction works by analyzing all previous tokens simultaneously, choosing the most likely next word based on learned patterns from massive datasets.

## Famous Examples

- GPT-4 (OpenAI)
- Gemini (Google)
- Claude (Anthropic)
- LLaMA (Meta)
- Mistral AI



## Typical Use Cases

- Conversational chatbots
- Text summarisation
- Code completion
- Language translation
- Content creation

# GANs: Creative Adversaries

## Two Networks, One Goal

GANs feature a unique competitive architecture where two neural networks engage in a continuous game of creation and critique.

### The Generator

Creates synthetic data trying to mimic real examples. Starts with random noise and learns to produce increasingly realistic outputs.

### The Discriminator

Acts as a critic, distinguishing between real and fake data. Provides feedback that helps the generator improve its creations.

**The Result:** Through this adversarial training, GANs produce remarkably realistic images, videos, and synthetic media.

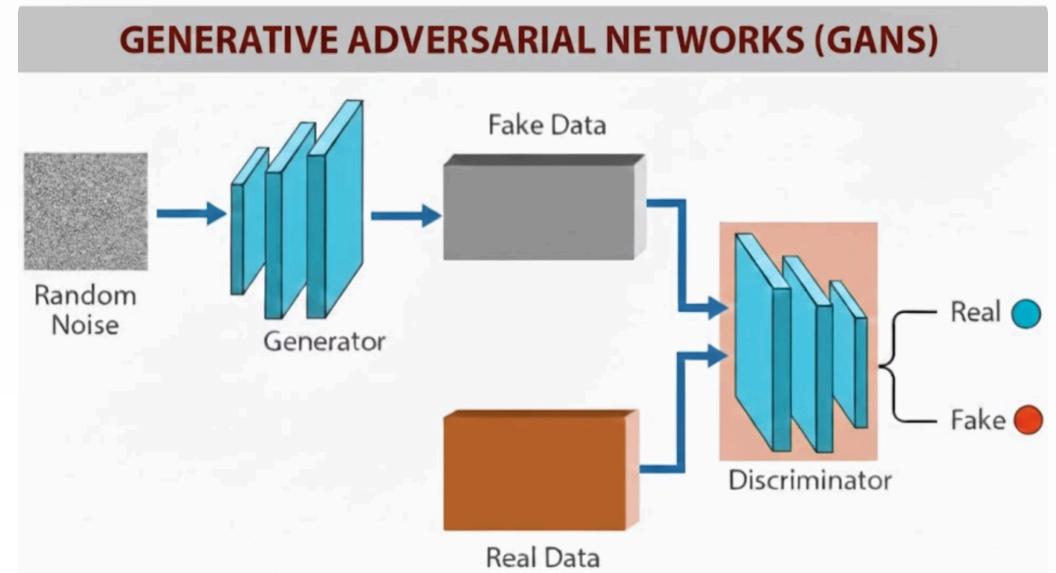
# GANs in Action

## Famous Examples

- **StyleGAN:** Creates photorealistic faces and artistic portraits
- **BigGAN:** Generates high-resolution, diverse images
- **CycleGAN:** Transforms images between domains (photo to painting)

## Typical Use Cases

- Realistic image creation
- Artistic design and style transfer
- Video synthesis
- Deepfake technology
- Data augmentation



# VAEs: The Compression Artists



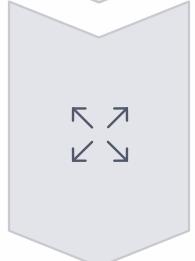
## Encoding

Input data is compressed into a compact latent space representation, capturing essential features while reducing dimensionality.



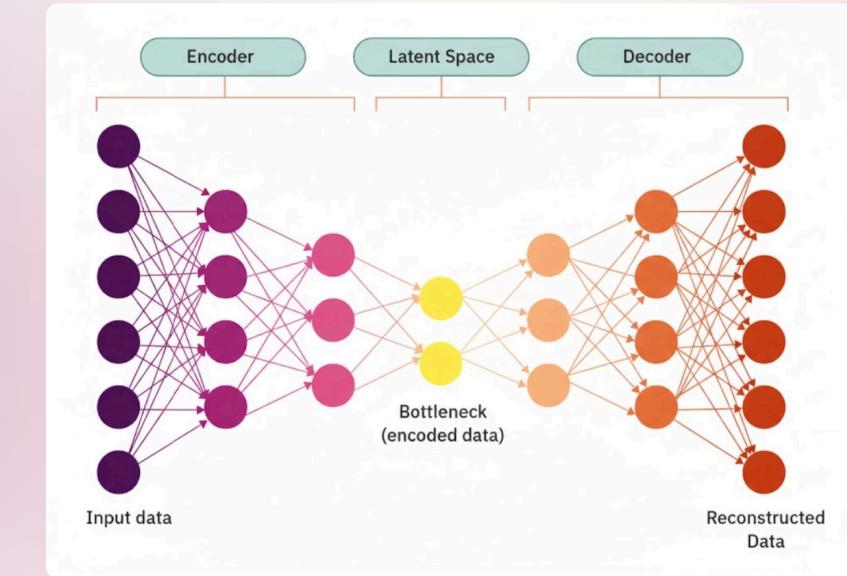
## Sampling

Random samples are drawn from the learned latent distribution, enabling generation of new variations.



## Decoding

The decoder reconstructs full data from latent vectors, creating novel outputs that resemble the training data.



# Understanding VAEs

## Core Principle

VAEs encode input into a latent space, then decode to generate new data, enabling smooth variations.

## Typical Use Cases

- Image reconstruction
- Anomaly detection
- Representation learning
- Data compression
- Feature extraction

## Everyday Examples

- **Photo Compression:** Compress and recreate photos with variations.
- **Face Generation:** Create new, realistic faces.
- **Music Variation:** Generate similar new melodies from a song's style.
- **Handwriting Styles:** Learn and generate new words in a specific handwriting style.

# Diffusion Models

Diffusion models have transformed generative AI by mastering a unique approach: starting with pure random noise and gradually refining it into coherent, high-quality outputs.

## 1 Random Noise

Process begins with completely random pixel values

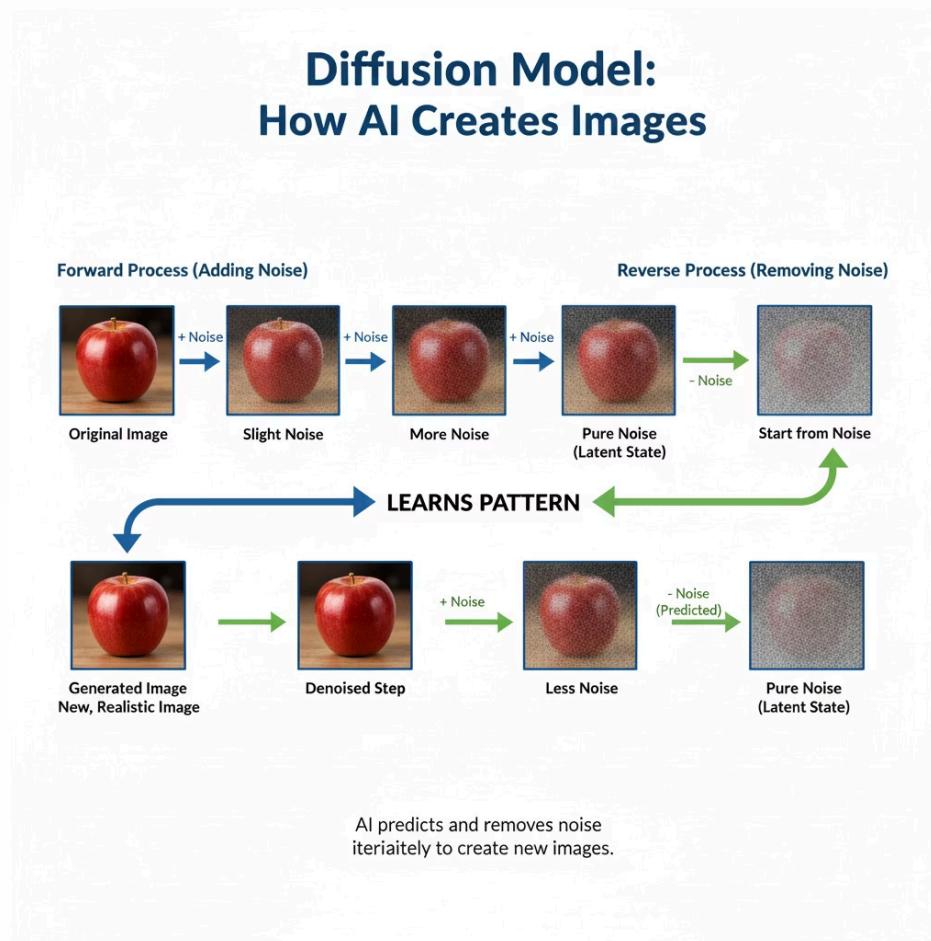
## 2 Iterative Denoising

Model progressively removes noise over many steps

## 3 Refined Output

Final result is a coherent, high-quality image

# Diffusion Models: Leaders in Visual Generation



## Stable Diffusion

Open-source powerhouse enabling customizable image generation at scale

## DALL·E 2

OpenAI's breakthrough in text-to-image generation with photorealistic results

## Midjourney

Artistic excellence with stunning visual quality and creative interpretation

**Current Dominance:** Diffusion models now lead the field in image, audio and video generation for their exceptional quality and controllability.

# Quick Reference Guide



## Transformers

Best for text, reasoning, and multimodal understanding



## GANs

Excel at realism and creativity in visual media



## VAEs

Great for structured latent learning and reconstruction



## Diffusion Models

Dominant current generative imagery for quality and control

Each model family has unique strengths. Transformers revolutionized language AI, GANs pioneered realistic synthesis, VAEs enable efficient representation learning, and diffusion models set new standards for visual generation quality.

# Thank You

Connect with me on [LinkedIn](#).

Created By: Kumar Siddharth Nirmal