Day 1 – SESSION 2 Introduction to Diffusion Models (in Generative AI)

Diffusion models are a class of **generative models** used to **create high-quality data**, such as **images, audio, or video**, by **reversing a gradual noise process**. They're at the heart of modern tools like **DALL·E 3**, **Stable Diffusion**, and **MidJourney**.

What Is a Diffusion Model?

At a high level:

- A **diffusion model** starts with **pure noise**.
- It then **gradually removes noise**, step by step, to reveal a new sample (e.g., an image).
- This process is **learned by training** the model on real data.

Think of it like watching a **photo develop from static**.

How It Works (Simplified)

1. Forward Process (Adding Noise)

- The model takes a real image (or data sample) and **adds noise to it in steps**.
- After many steps, the image becomes **pure noise**.
- This teaches the model how data gets "destroyed."

2. Reverse Process (Denoising)

- The model is then trained to **reverse this process**:
  - It learns to **predict and remove noise** at each step.
  - By doing so iteratively, it **generates new data** from pure noise.

□ Key Terms

Term	Meaning
Latent Space	A compressed representation of the data
Noise Schedule	The way noise is added across timesteps
Timestep	A step in the diffusion (or denoising) process
DDPM (Denoising Diffusion Probabilistic Model)	The foundational architecture for diffusion models
UNet	A common neural network used for the denoising steps

□ Why Are Diffusion Models Popular?

- **Very high-quality outputs** — realistic and detailed images.
- **Stable training** — unlike GANs, which can be unstable.
- **Flexible** — can be conditioned on prompts, images, audio, etc.

□ Examples of Diffusion-Based Tools

Tool	What it Does
DALL·E 3	Text-to-image generation
Stable Diffusion	Open-source image generation
MidJourney	Artistic image generation
Imagen (Google)	High-quality image synthesis

□ Variants of Diffusion Models

- **DDIM (Denoising Diffusion Implicit Models)** — faster generation with fewer steps.
- **Latent Diffusion Models (LDM)** — work in compressed space for speed (used in Stable Diffusion).
- **ControlNet** — guides generation with extra inputs (like sketches or poses).

□ Applications

- **Text-to-image generation**
- **Inpainting** (filling in missing parts of images)
- **Image super-resolution**
- **Music and speech generation**
- **Video synthesis (early stage)**

□ Pros vs. Cons

Pros	Cons
Very high-quality output	Slow generation (many steps)
Stable training process	High compute requirements
Good at diversity	Complex model architecture

□ Want to Dive Deeper?

You can explore:

- **Math foundations**: Score matching, variational inference
- **Implementations**: Hugging Face `diffusers` library, PyTorch/Colab tutorials
- **Code walkthroughs**: Build your own diffusion model (starting from MNIST or CIFAR-10)