

Image and Video Generation

How diffusion works, then how to control it: conditioning, inpainting, and a LoRA trained on your own subject. You finish with a generation pipeline that runs at a cost you can name.

Intermediate

5 phases

26 sessions

39 contact hours

Syllabus

5 phases, 26 sessions. Each phase ends in something you have built.

1

The Denoise Loop

Know what the sampler is actually computing, well enough to debug it instead of rerolling seeds.

5 sessions

\$72 alone

1.1 Noise, Score, Velocity

Forward noising, the score function, and why the network predicts a direction and not an image.

[FREE PREVIEW](#)

1.2 Flow Matching Won

Rectified flow and velocity prediction, and why SD3.5, FLUX and Qwen-Image stopped predicting epsilon.

1.3 Latents and the VAE

Why you denoise in a compressed space, what the VAE costs you, and the banding a wrong decode gives.

1.4 Samplers, Schedulers, CFG

Euler against DPM++ 2M, timestep shift, and the distilled model that falls apart at CFG 7.

1.5 Fifty Lines of Sampler

Build the loop from Diffusers parts, then find the four things besides the seed that move your output.

By the end: **A sampling loop you wrote by hand that matches a Diffusers pipeline output on a fixed seed.**

2

Making It Obey

Get the specific image the brief asked for, using conditioning instead of prompt roulette.

6 sessions

\$72 alone

2.1 T5 Against CLIP

Why a long literal prompt works on FLUX and confuses SDXL, and what the token limit quietly cuts.

2.2 ControlNet Still Wins

Depth, canny and pose when composition must be locked, and ControlNet-Union inside one graph.

2.3 Reference, Not Structure

IP-Adapter, InstantID and Redux for identity and style, and the pose change where identity drifts.

2.4 Editing by Instruction

FLUX.2 multi-reference and Qwen-Image-Edit replacing a pile of img2img, masks and loose control.

2.5 Repair Passes

Differential diffusion, mask blur, tiled upscaling, and the seam where lighting stops matching.

2.6 Graphs, Not Scripts

ComfyUI 0.30 subgraphs, API-format export, and pinning custom nodes so the graph still runs in a month.

By the end: **A ComfyUI graph that hits a fixed composition brief from one control image and one reference.**

3

One Subject, Fitted

Fit a LoRA to a subject you own, and prove it learned the subject and not the folder.

5 sessions

\$72 alone

3.1 Twenty Images, Curated

Shot variety, backgrounds that leak into the concept, and the near-duplicate that becomes the LoRA.

3.2 Captions Decide Everything

Trigger tokens, tags against natural language, and captioning the attributes you do not want learned.

3.3 Rank, LR, Steps

Three settings that are really one, and a SimpleTuner config that converges on a 24GB card.

3.4 Catching Overfit Early

Sampling every few hundred steps, prompt adherence collapsing, and keeping step 1800 over step 3000.

3.5 Score It, Don't Squint

HPSv3, ImageReward and GenEval on a held-out prompt set, and why FID answers the wrong question.

By the end: **A subject LoRA with a scored eval report and a written reason for the checkpoint you rejected.**

4

Motion

4 sessions

Get a usable shot out of a video model, and know which failures are yours and which are the model's. \$72 alone

4.1 The 2026 Video Board

Veo 3.1, Kling 3.0 and Runway Gen-4.5 against open Wan 2.2, LTX-2.3 and HunyuanVideo-1.5.

4.2 First Frame, Last Frame

Design frame zero in an image model, then ask the video model for motion and nothing else.

4.3 Video LoRA on Wan

Musubi Tuner on Wan 2.2, and the trap where high noise and low noise experts need separate adapters.

4.4 Where Shots Break

Identity through a head turn, hands, extension drift, and cutting before the failure arrives.

By the end: **A 30 second cut assembled from generated shots, with a log of every retry and what broke it.**

5

Shipping It

6 sessions

Put a generation behind an API at a cost you can quote and a provenance trail you can defend.

\$72 alone

5.1 Cog and Replicate

Package a ComfyUI workflow with cog-comfyui, push it, and get a queued endpoint you can call.

5.2 What a Generation Costs

\$0.012 per megapixel against H100 seconds at \$0.001525, cold starts, and the idle worker eating margin.

5.3 Making It Cheaper

torch.compile on fixed shapes, FP8 and GGUF for memory, and TeaCache traded against motion quality.

5.4 Signing Every Output

C2PA 2.4 manifests, a durable watermark as the backup, and the CDN transcode that strips both.

5.5 The Licence Blocks You

FLUX.1 dev non-commercial, the Stability \$1M revenue line, and the checkpoint you cannot deploy.

5.6 Consent and Likeness

EU AI Act Article 50 live since August 2026, the ELVIS Act, and the consent record you have to store.

By the end: **A live Replicate endpoint with a measured cost per image and a signed manifest on every output.**

Tools you will use

Replicate

Cog

ComfyUI 0.30

Diffusers 0.39

FLUX.2 [dev]

Qwen-Image-Edit-2511

SimpleTuner 4.1

Musubi Tuner

Wan 2.2

c2patool (C2PA 2.4)

What you will build

1 The Sampler You Wrote

Fifty lines that match the pipeline on one fixed seed

2 One Subject, One LoRA

Twenty photos to a scored adapter, and a rejected checkpoint

3 Endpoint With a Receipt

A cost per image, a signed manifest, a licence you can name

Registration

Phase 1: The Denoise Loop	\$72
Phase 2: Making It Obey	\$72
Phase 3: One Subject, Fitted	\$72
Phase 4: Motion	\$72
Phase 5: Shipping It	\$72
Whole track, all 5 phases	\$300

Buying every phase separately costs \$360, so the whole track saves \$60.

No payment is taken online. Registering creates an order and payment instructions are sent with it. Access opens once payment is confirmed.