# LLMs for me

## Multimodal LLMs & Frameworks

llmsfor.me

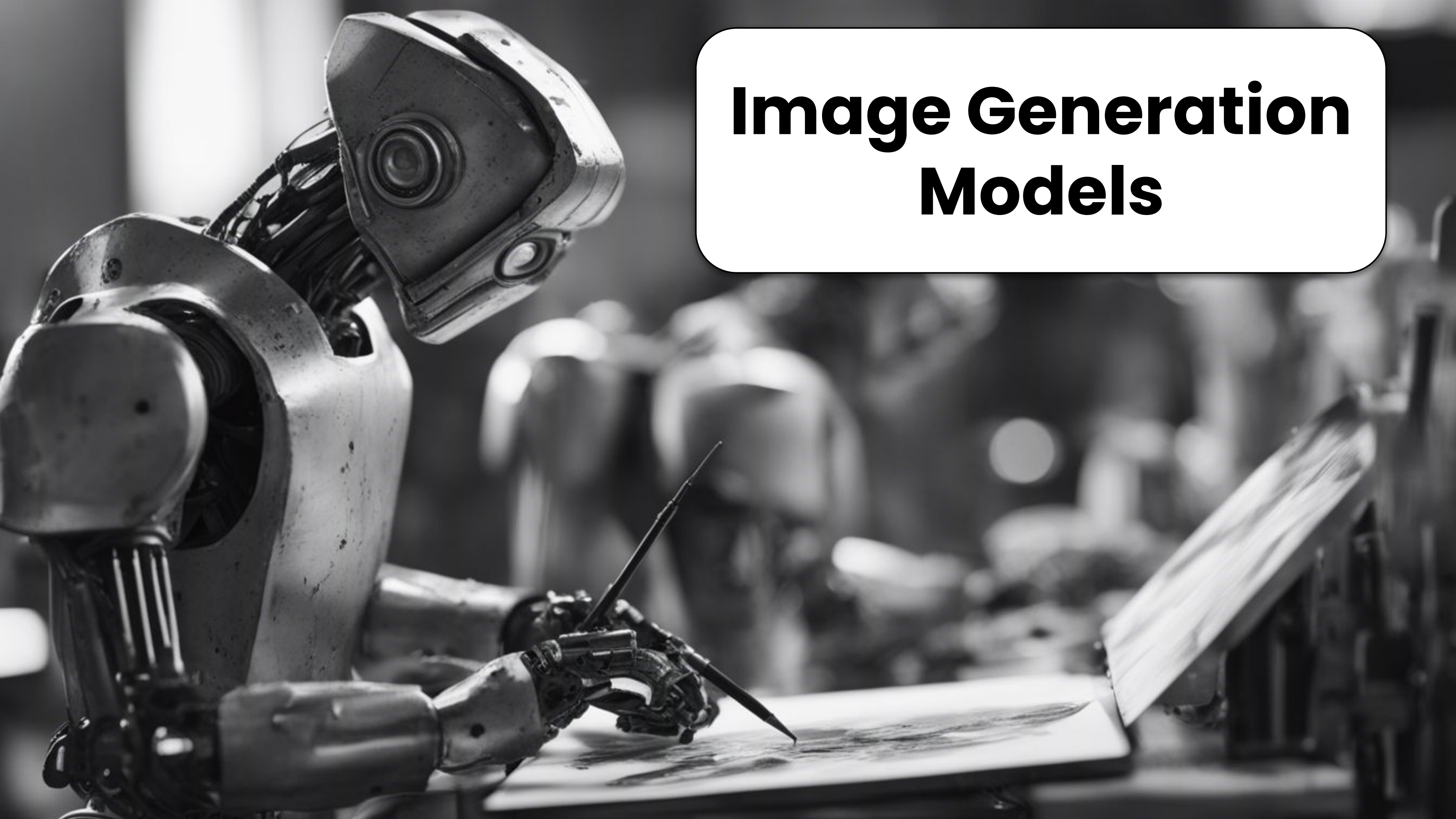

**Myles Harrison,**
AI Consultant & Trainer


**February 3rd, 2025**

NLP from scratch

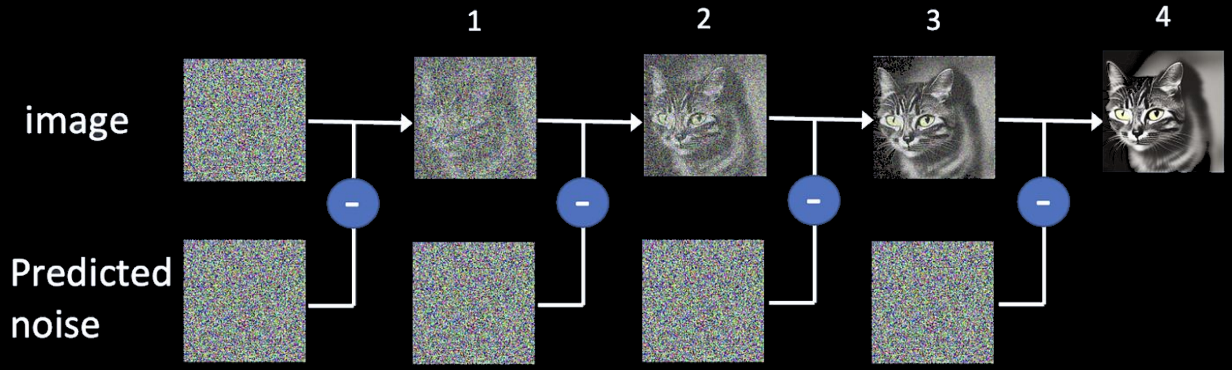
# Agenda

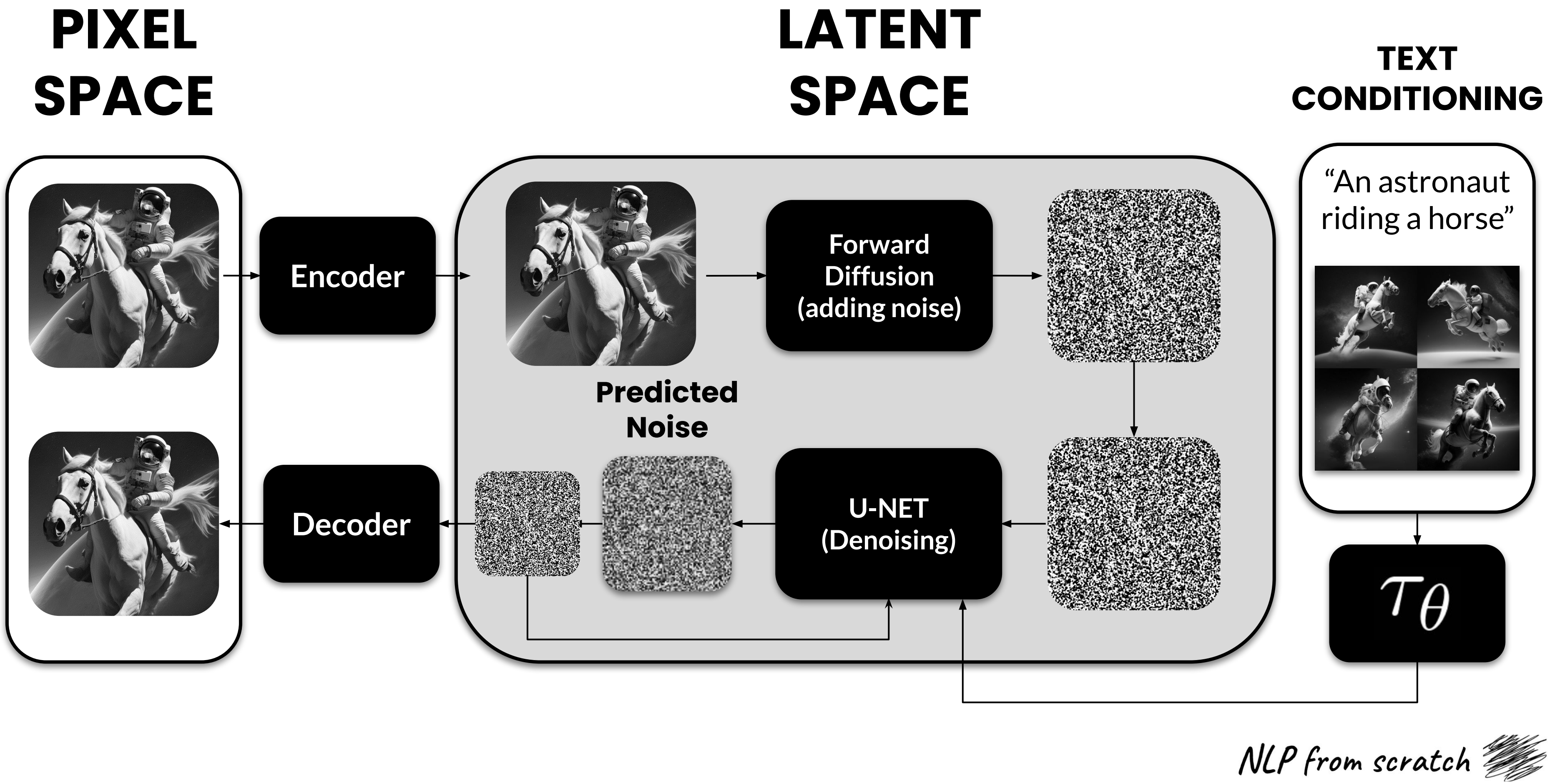
# Image Generation Models



# Stable Diffusion

- Latent diffusion models (Rombach et. al, 2021) are a type of generative AI model that can create images by iteratively refining random noise, guided by a learned representation of patterns in data (a “latent space”)
- These models start with random noise and use a neural network to "denoise" step by step, transforming it into a detailed image by following patterns learned from a large dataset of images and captions.
- Model learns to predict the added noise during training, then denoises during prediction
- Can be conditioned with text via the important CLIP model learning representations between text and images (OpenAI, 2021)




PIXEL SPACE
LATENT SPACE
TEXT CONDITIONING
"An astronaut riding a horse"
Encoder
Forward Diffusion (adding noise)
Predicted Noise
U-NET (Denoising)
Decoder
$\tau_\theta$
NLP from scratch

# Stable Diffusion Timeline

**December 2021**: Initial Latent Diffusion (LDM) paper and code are published.

2022

**August 2022**: Stability AI releases the initial model versions publicly (***v1.1 - 1.4***, 863M parameters)

**October 2022**: RunwayML releases Stable Diffusion ***1.5***.

**Nov-Dec 2022**: Versions ***2.0 & 2.1*** are released, retrained from scratch on new datasets.

2023

**July 2023**: ***SDXL***, a new larger model with separate “base” and “refiner” models is released (3.5 B parameters)

**November 2023**: ***SDXL Turbo*** is released.

2024

**June 2024**: ***SD3*** open release (8B parameters)

**October 2024**: ***SD3.5*** medium open model is released for general use.

2025

# Stable Diffusion XL (SDXL)

- Released July 2023 by researchers at Stability AI, the successor to Stable Diffusion 2.1
- 3x in size to (core of) original model
- Additional refiner model (image-to-image) for denoising used in a supplementary fashion after base model for high fidelity outputs
- Available through Clipdrop (paid) and on Hugging Face spaces (free, various)
- Now near real-time image generation “as you type” with SDXL Turbo

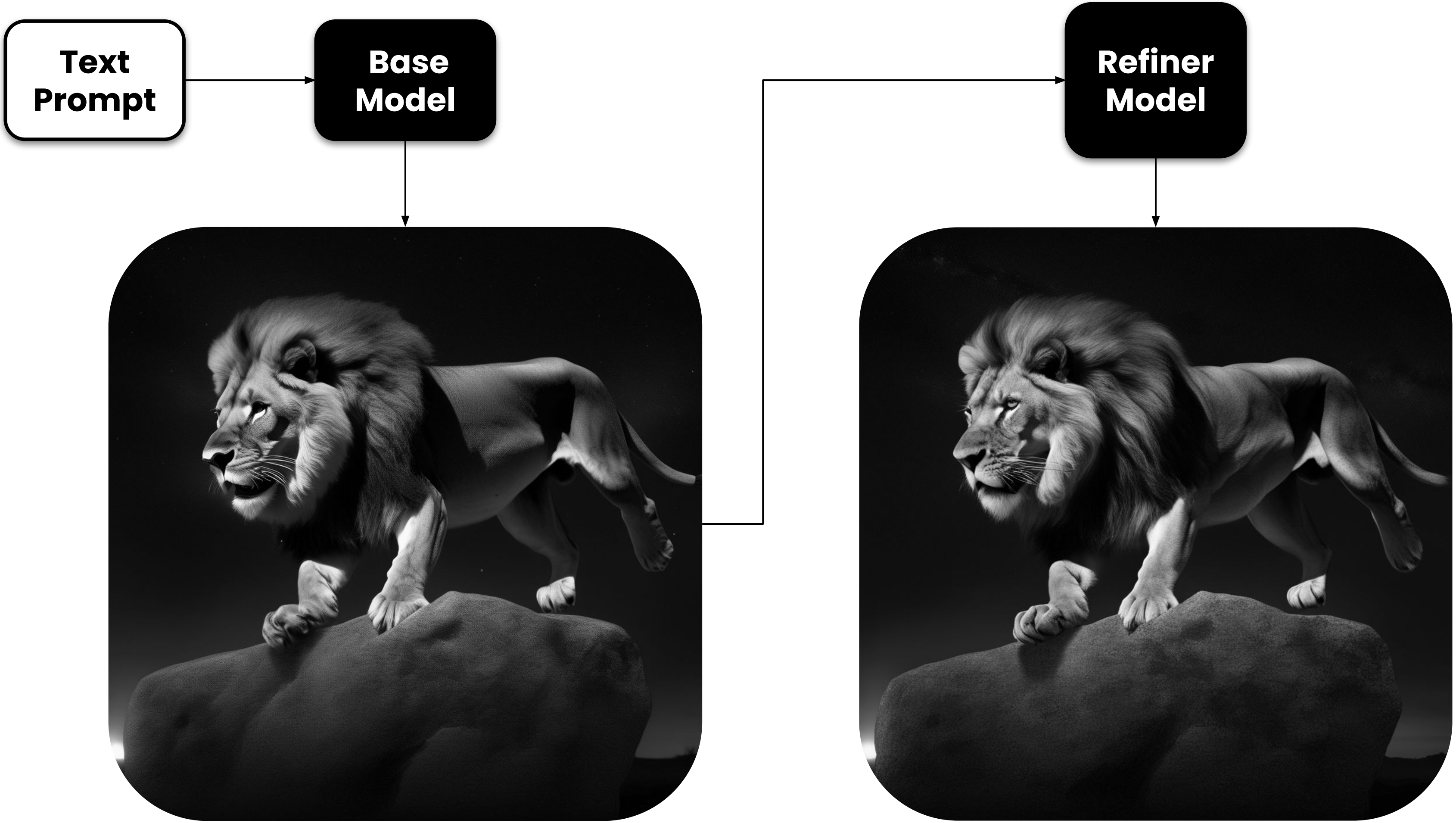
# Hugging Face 🤗: SDXL in 5 lines of code

```
from diffusers import AutoPipelineForText2Image
import torch

pipeline =
AutoPipelineForText2Image.from_pretrained(
    "stabilityai/stable-diffusion-xl-base-1.0",
     torch_dtype=torch.float16, variant="fp16",
     use_safetensors=True
).to("cuda")

image = pipeline(prompt="A cute dog in a fuzzy
sweater").images[0]

image.save("dog.png")
```



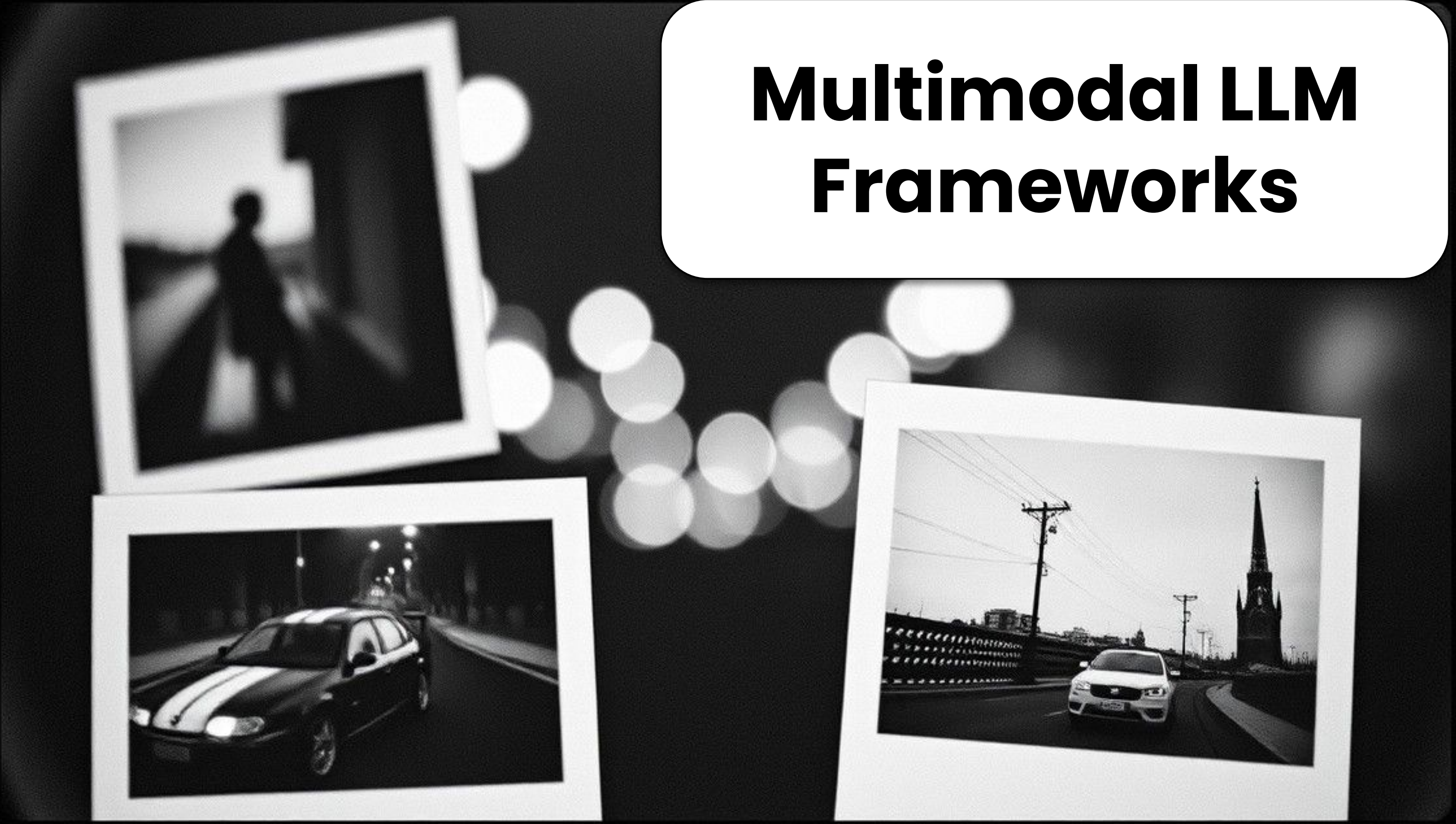

Images from huggingface.co/docs/diffusers/en/using-diffusers/sdxl

# Flux

- Announced 2024/08/01
- Team of original creators of Stable Diffusion created startup Black Forest Labs
- $231M in seed from a16z
- 12B transformer/diffusion flow-based model in 3 versions: Pro, Dev, and Schnell (Apache 2.0 licensed)

[blackforestlabs.ai/announcing-black-forest-labs/]

# Multimodal LLM Frameworks



# Image Generation with Services (no coding required)

- Most chatbot services have image generation built in for a free (or inexpensive) AI image generation
- For example, Copilot and Meta AI both allow image generation through their chatbots if the user is signed in at no cost
- Pay services include DALL-E for ChatGPT subscribers and every-popular Midjourney via Discord bot
- Other options out there, *e.g.* Hugging Face spaces (https://huggingface.co/spaces/google/sdxl)

Copilot

Meta

DALL·E

Midjourney

# AUTOMATIC1111 & ComfyUI

- Stable Diffusion web UI (or more familiarly, AUTOMATIC1111) was original open-source, web-based UI for running Stable Diffusion models, based upon the Gradio framework
- Features extensive customization, including inpainting, outpainting, and upscaling, face rendering models and offered an intuitive interface with real-time previews and batch processing
- ComfyUI offshoot from the original creator of SD web UI, gaining popularity and traction
- Node-based, modular UI for Stable Diffusion, allowing for deep customization of workflows, enables users to create complex AI image pipelines with drag-and-drop functionality but with a steep learning curve

# DEMO

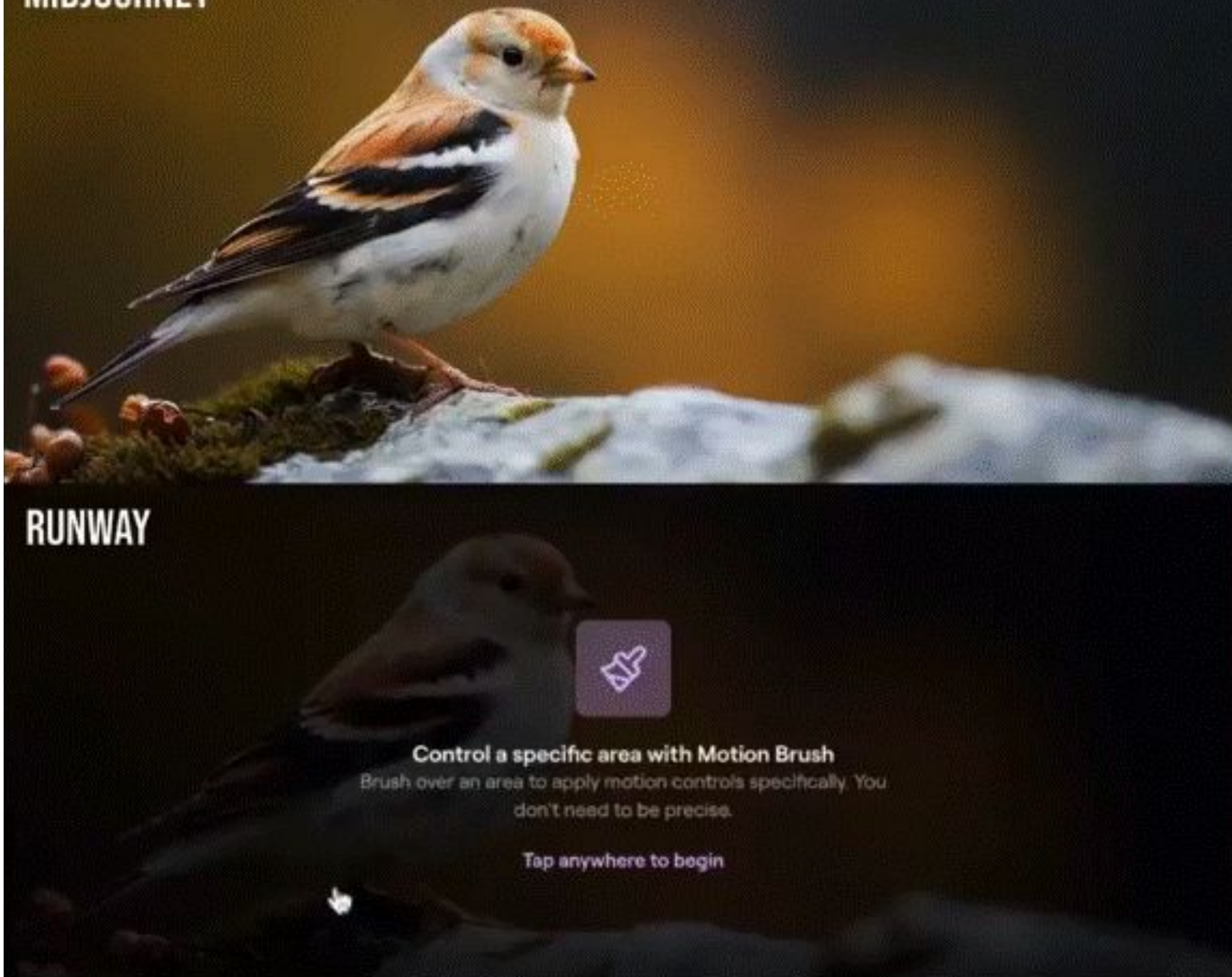
ComfyUI

# Beyond Images

# Beyond images

runway

runwayml.com



# Google’s Veo 2

Veo 2

deepmind.google/technologies/veo/veo-2

# SPAR3D - Text Prompt to 3D Asset Creation

stability.ai

spar3d.github.io

# End of Part 5

LLMsfor.me
PWYC Microcourse in LLMs and Generative AI
January 2025

**Part 5 – Multimodal LLMs and Frameworks**

**Monday, February 3rd, 2025**

llmsfor.me

NLP from scratch