Lecture 11: Generative AI

https://github.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2025s1

Teerapong Panboonyuen
https://kaopanboonyuen.github.io

Reference

- https://huggingface.co/models
- https://www.geeksforgeeks.org/deep-learning/transformers-library/
- https://arxiv.org/pdf/1706.03762
- https://research.google/blog/transformer-a-novel-neural-network-architecture-for-language-understanding/
- https://openai.com/
- https://gemini.google.com/

🚀 What is Generative AI?

- AI that **creates** new content instead of just predicting.
- Examples: **Images, Text, Music, 3D, Chatbots** 🎨 📝 🎶 🏗️
- Powered by **neural networks** like GANs, Diffusion, Transformers.

https://arxiv.org/pdf/1706.03762

[cs.CL] 2 Aug 2023

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

https://arxiv.org/pdf/1706.03762

Attention Visualizations

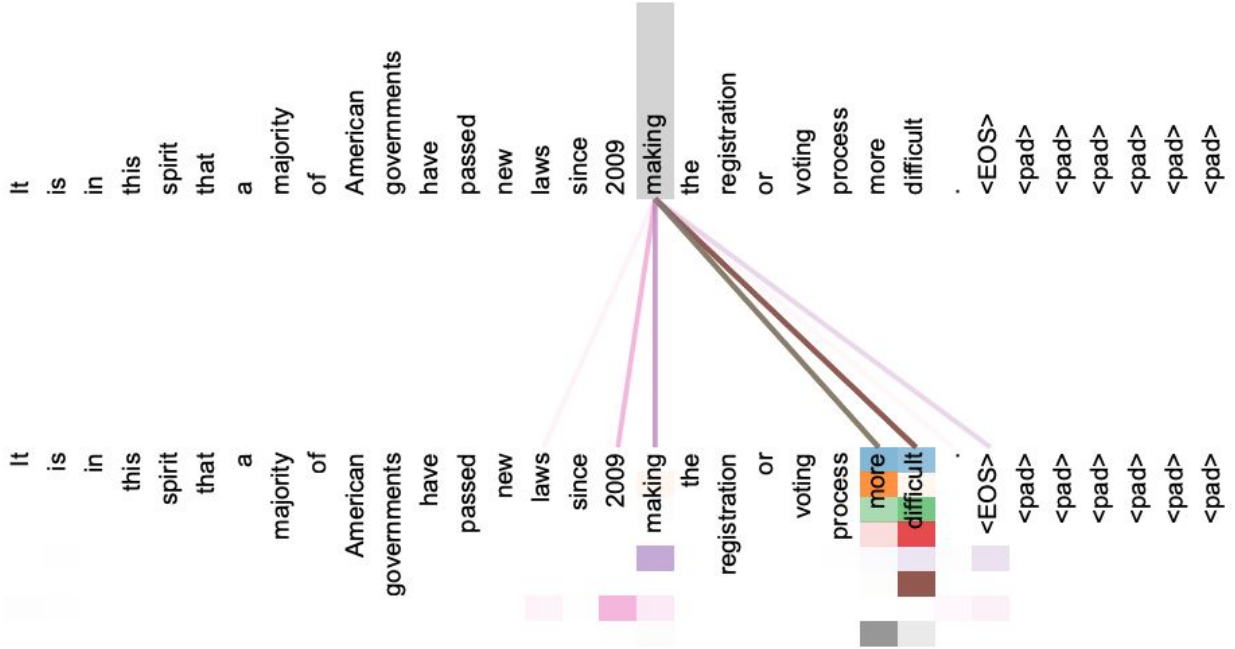


Figure 3: An example of the attention mechanism following long-distance dependencies in the encoder self-attention in layer 5 of 6. Many of the attention heads attend to a distant dependency of the verb 'making', completing the phrase 'making...more difficult'. Attentions here shown only for the word 'making'. Different colors represent different heads. Best viewed in color.

👨‍🎨 GANs — Generative Adversarial Networks

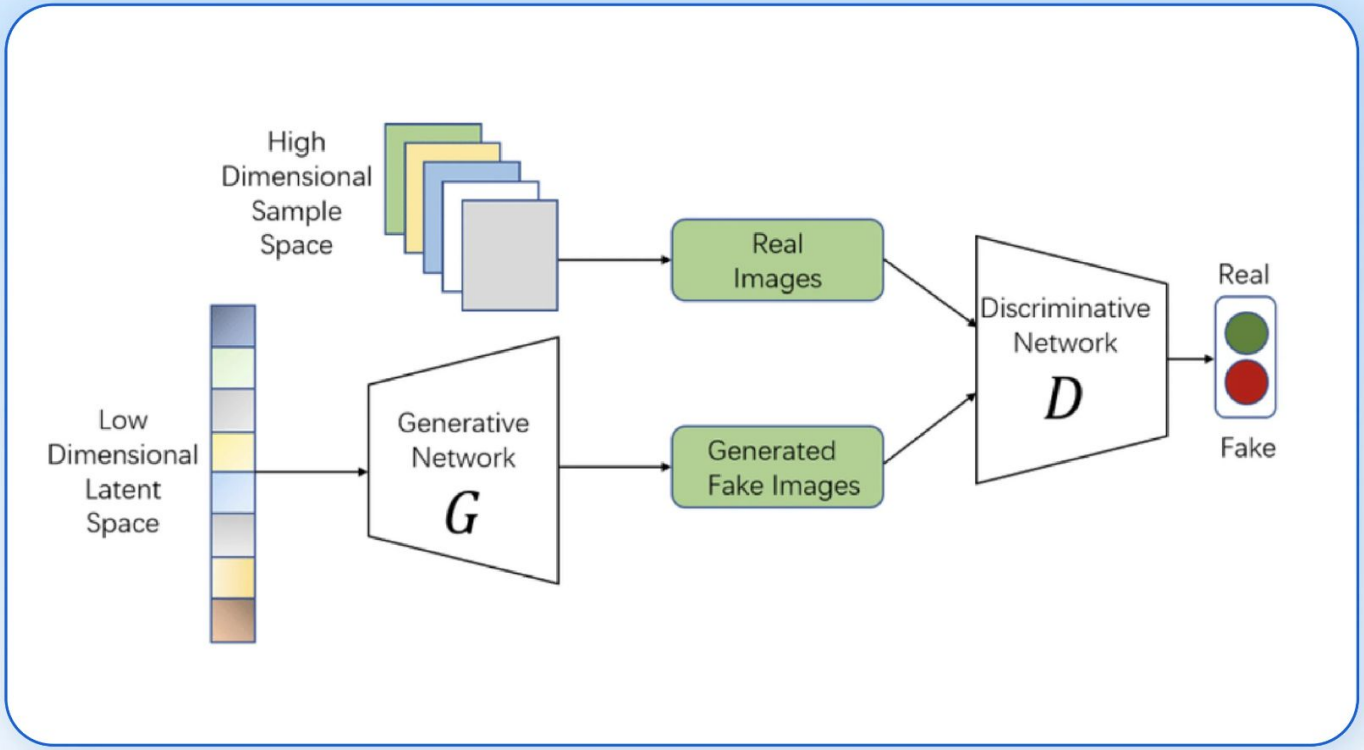
- **Two networks** compete:
  - **Generator** 🎨 : Creates fake data.
  - **Discriminator** 🔎 : Detects real vs fake.
- Like **art student vs art teacher** 🎭

👉 Example: Generate handwritten digits (MNIST)

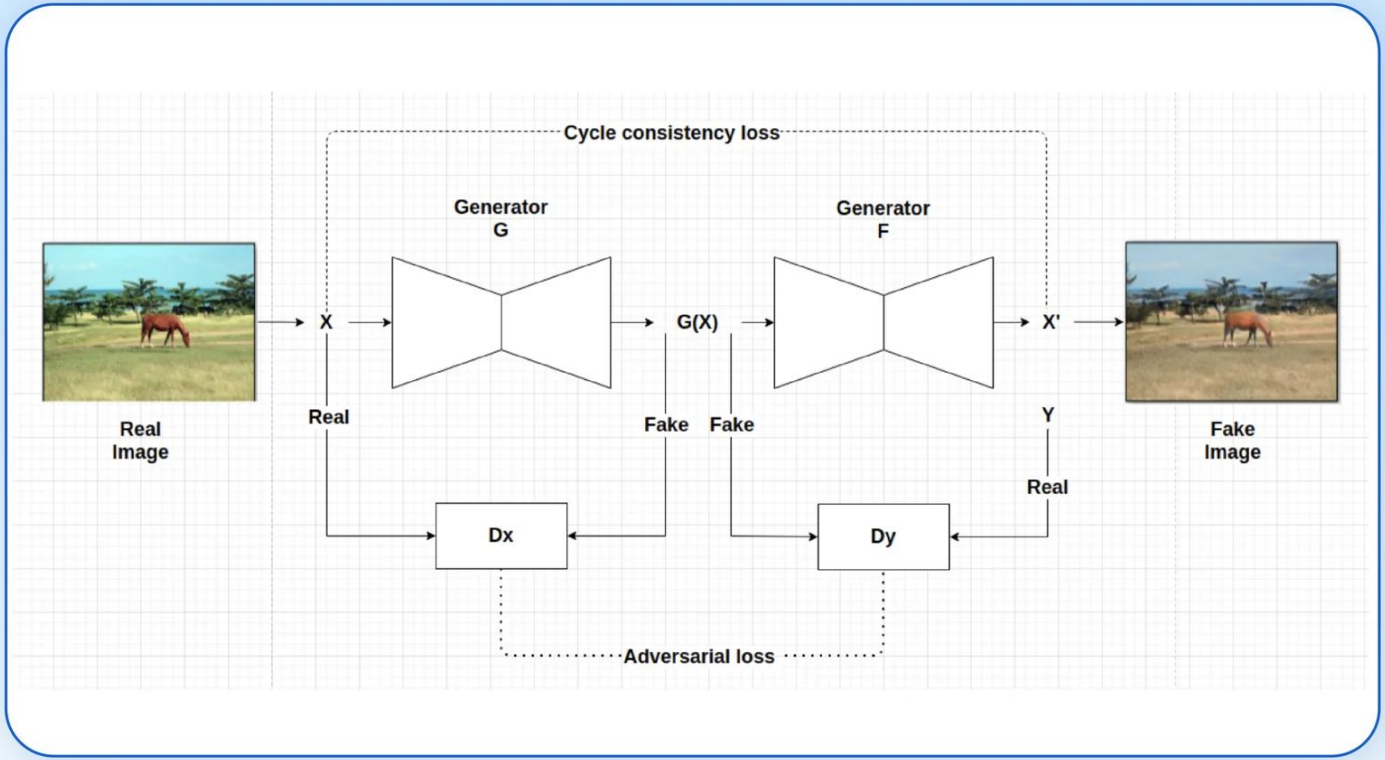
```
plt.imshow(fake_imgs[0].cpu().view(28,28), cmap="gray")
plt.title("🖼️ GAN-generated digit")
```

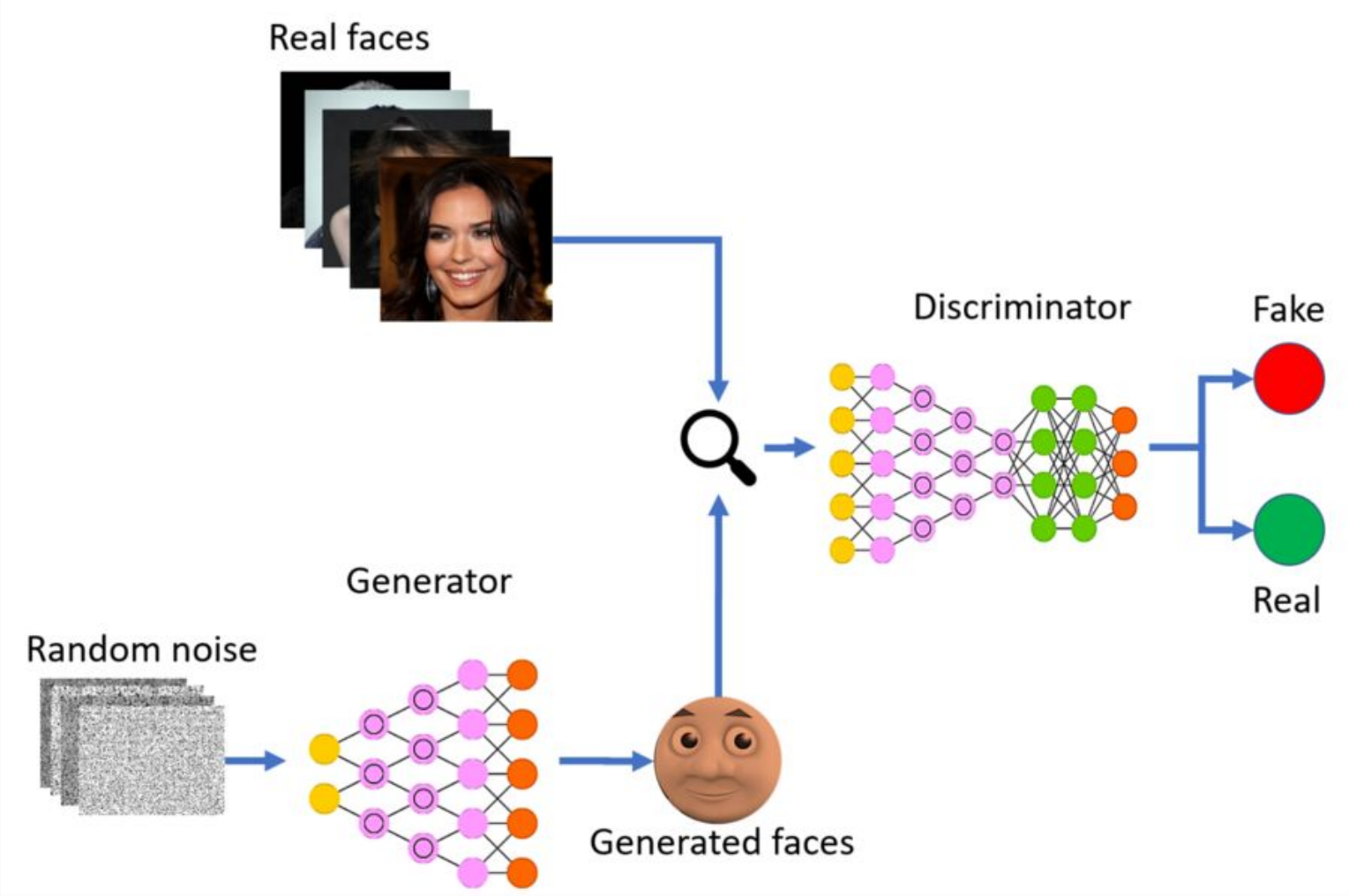
https://www.xenonstack.com/blog/gans-for-image-synthesis

XENONSTACK



https://www.xenonstack.com/blog/gans-for-image-synthesis




Real faces
Discriminator
Fake
Real
Generator
Random noise
Generated faces

All faces apart from one were imagined by GAN (generative adversarial network) StyleGAN2 (Karras et al, arXiv:1912.04958) and retrieved from https://thispersondoesnotexist.com/

1️⃣ GANs (Generative Adversarial Networks)

Goal: Generator G creates fake data that fools Discriminator D.

Math (simplified):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))]$$

- x = real data
- z = random noise
- $D(x)$ = probability real
- $G(z)$ = fake data

Intuition:

- D = "judge" → try to detect real/fake
- G = "artist" → try to fool judge
- They play a **minimax game** until G generates realistic samples 🎨

🎭 Modern GAN — StyleGAN2

- Can generate **photorealistic faces** 👩 👨
- Trained on **FFHQ dataset** (Flickr-Faces-HQ)

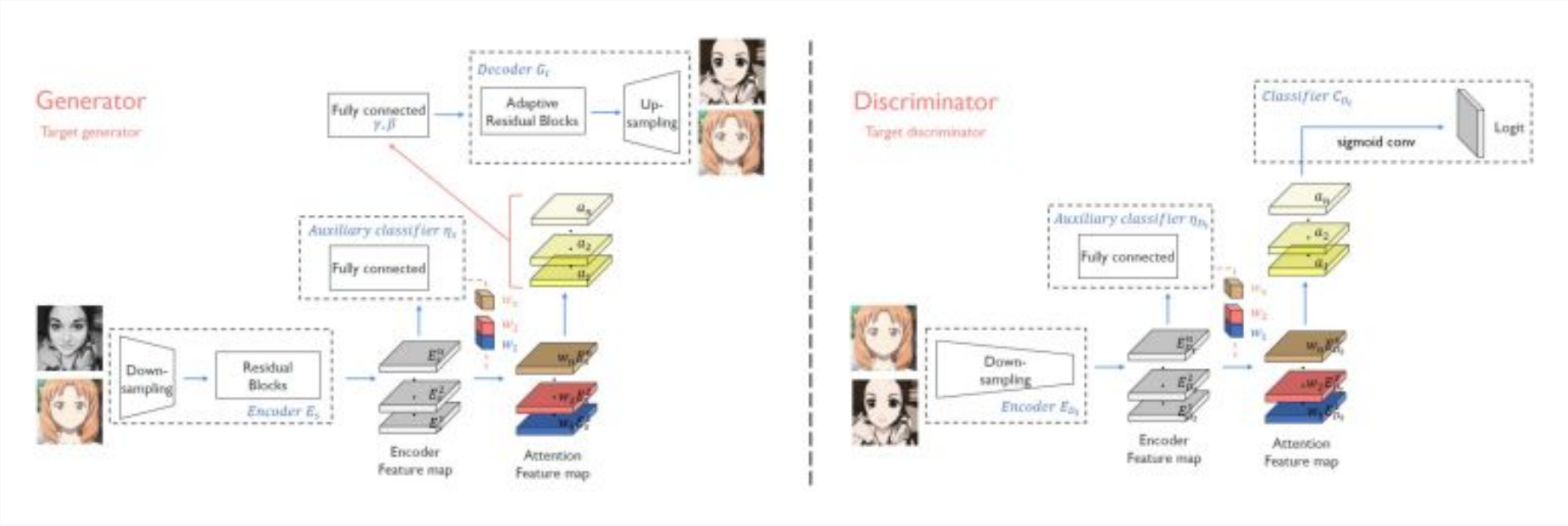
👉 Example: Generate a human-like face

```
sample = model.generate_images(num=1)
Image.open(sample[0]).resize((256,256))
```

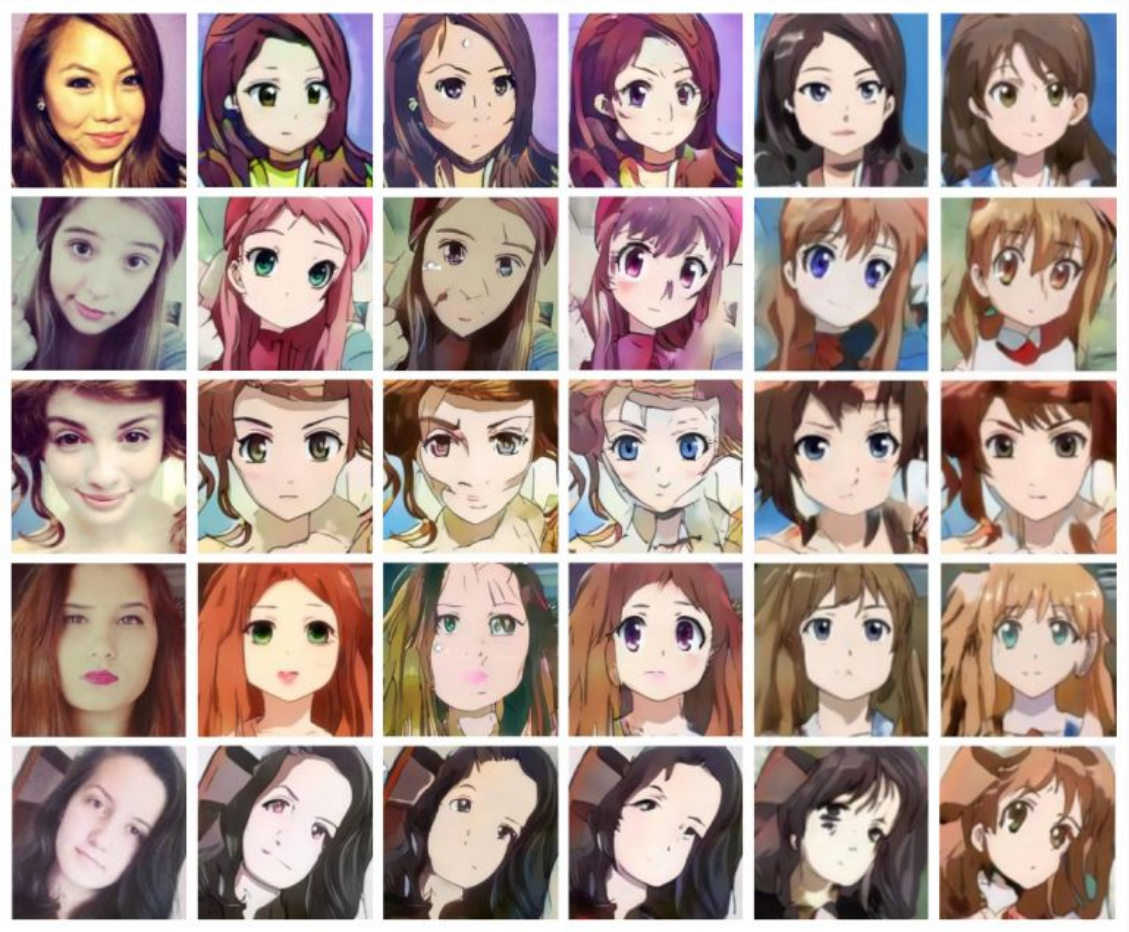
🎨 Other GAN Model Ideas

Task	Model	Output
Cartoon	`cartoonGAN`	Anime-like
Face Aging	`photo2cartoon`	Add age to faces
Style Transfer	`stylegan`	Picasso, Van Gogh, etc.
Pix2Pix	Custom	Maps → Satellite, Sketch → Photo

https://syncedreview.com/2019/08/08/south-korean-game-developers-ai-turns-your-selfie-into-an-anime-face/



https://syncedreview.com/2019/08/08/south-korean-game-developers-ai-turns-your-selfie-into-an-anime-face/



🌌 Diffusion Models (Stable Diffusion)

- Add **noise** to an image, then **denoise step by step** 🔄
- Great for **text-to-image** 🎨

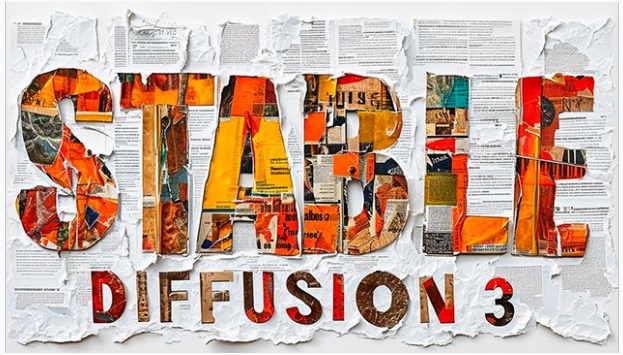
👉 Example:

Prompt: "A magical castle in the sky, fantasy art 🌈🏰 "

```
image = pipe(prompt).images[0]
image.show()
```

https://stability.ai/news/stable-diffusion-3


GO BIG
OR go HOME


STABLE
DIFFUSION 3


Go
DREAM ON


STABLE
DIFFUSION.

💥 What is Stable Diffusion?

- **Stable Diffusion** is a **deep generative model** based on **diffusion processes**.
- It generates high-quality images from text prompts or modifies existing images.

How it works (simplified):

1. Start with **pure noise** (random pixels).
2. A neural network (U-Net + Transformer) **gradually denoises** the image step by step.
3. Conditioning on text (via CLIP / text encoder) guides the denoising → image matches the description.

Key Features:

- 🎨 Text → Image: "A cat wearing sunglasses"
- 🔄 Image → Image: apply style transfer, inpainting, or super-resolution
- ⚡ Efficient: Runs on consumer GPUs (optimized with latent diffusion)
- 🌍 Open-source: Hugely popular, community-driven, with many fine-tuned variants

huggingface.co/stabilityai/stable-diffusion-3-medium

Hugging Face | Search models, datasets, users... | Models | Datasets | Spaces | Community | Docs | Enterprise | Pricing | Log In | Sign Up

stabilityai / **stable-diffusion-3-medium** | like 4.83k | Follow Stability AI 30k

Text-to-Image | Diffusion Single File | English | stable-diffusion | arxiv:2403.03206 | License: stabilityai-ai-community

Model card | Files and versions xet | Community 217 | Deploy | Use this model

You need to agree to share your contact information to access this model

This repository is publicly accessible, but **you have to accept the conditions to access its files and content.**

By clicking "Agree", you agree to the License Agreement and acknowledge Stability AI's Privacy Policy.

Log in or Sign Up to review the conditions and access this model content.

Stable Diffusion 3 Medium

Downloads last month
11,888

Inference Providers NEW | fal

Text-to-Image

Your sentence here... | Generate

View Code Snippets | Maximize

Model tree for stabilityai/stable-diffusion-3-medium

Adapters	67 models
Finetunes	7 models
Quantizations	4 models

Spaces using stabilityai/stable-diffusion-3-medium 100

stabilityai/stable-diffusion-3-medium | jasperai/flash-sd3 | alfredplpl/sd3-with-LLM | madebyollin/sd3-with-taesd3-previews | jbilcke-hf/OmniAvatar | fffiloni/Stand-In | aipicasso/emi-3 | rphrp1985/stable-diffusion-3-medium | jbilcke-hf/ReCamMaster | Kunbyte/Lumen | kasper-boy/Transform_Ordinary_Photos_into_Extraordinary_Art | hf-demo-linux/sili

+ 88 Spaces

2️⃣ Diffusion Models

Goal: Learn to **denoise** data step by step.

Forward process (add noise):

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

Reverse process (denoise using neural network):

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

Intuition:

- Start from noise → denoise step by step → get realistic image 🌌
- Network predicts mean μ_θ of clean image

3️⃣ Transformers (Self-Attention)

Attention formula:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

- Q, K, V = queries, keys, values (vectors)
- Compute **weighted sum** of values → network "focuses" on important words/tokens

Intuition:

- Words attend to other words in the sequence
- Enables GPT to generate coherent text ✍️

https://arxiv.org/pdf/1706.03762

[cs.CL] 2 Aug 2023

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

🌟 Self-Attention Explained with Example

Sentence:

> "The cat sat on the mat"

Step 1: Represent Words as Vectors

- Each word → vector (embedding)
- Q (query), K (key), V (value) derived from embedding

Step 2: Compute Attention Scores

- Attention between words = similarity of Q(word) & K(other words)
- Example attention weights (simplified):

Word →	The	cat	sat	on	the	mat
cat	0.1	0.6	0.2	0.05	0.02	0.03

- "cat" attends **most to itself (0.6)** and "sat" (0.2)

Step 3: Weighted Sum

- Multiply weights × V values → new vector for "cat"
- This vector now encodes context from important words

Step 4: Intuition

- Attention = **focus mechanism**
- Each word "looks" at other words to understand context
- Enables GPT to generate **coherent and context-aware text** ✍️

Visual:

arduino Copy code

```
cat → [The:0.1, cat:0.6, sat:0.2, on:0.05, the:0.02, mat:0.03] → new vector
```

Key Takeaway:

- Self-Attention allows the model to **decide which words matter most** for each word in the sentence.
- Makes LLMs capable of **long-range dependencies** (e.g., "The cat...mat")

4️⃣ LLM / Text Generation (Autoregressive)

Goal: Predict next token x_t given previous tokens $x_{<t}$

$$P(x_1, x_2, ..., x_T) = \prod_{t=1}^{T} P(x_t | x_1, x_2, ..., x_{t-1})$$

- Model trained to **maximize likelihood** of dataset
- Decoder generates one token at a time → fluent text

Intuition:

- Treat text like a **probabilistic chain**
- Each word depends on previous words 🤖

5️⃣ Optional: LoRA (Low-Rank Adaptation)

- Represent **weight update** as low-rank matrices:

$$W_{\text{new}} = W + AB$$

- W = pretrained weight, A, B = small trainable matrices
- Efficient fine-tuning → train small fraction of parameters ⚡

Tip

- GAN = competition (Generator vs Discriminator)
- Diffusion = “noise → clean” iterative process
- Transformers = “focus on important words” via attention
- LLM = chain rule for text
- LoRA = “lightweight fine-tuning”

🌌 Text-to-Image

- Models: **Stable Diffusion, DALL·E, MidJourney**
- Input: Text prompt → Output: Image 🎨
- How it works: Latent diffusion + cross-attention between text & image
- Applications: Digital art, game assets, concept design
- Sample Prompt: `"A futuristic Bangkok skyline at sunset 🌇✨"`

python Copy code

```python
from diffusers import StableDiffusionPipeline
pipe = StableDiffusionPipeline.from_pretrained("runwayml/stable-diffusion-v1-5").to("cuda"
image = pipe("A futuristic Bangkok skyline at sunset 🌇✨").images[0]
image.show()
```

🏗️ Text-to-3D Image

- Models: **Point-E (OpenAI), DreamFusion (Google)**
- Input: Text → Output: 3D model (mesh, voxel, point cloud)
- How it works: Diffusion in 3D latent space + Neural Radiance Fields (NeRF)
- Applications: Game assets, AR/VR, 3D printing
- Concept Example: `"A cute robot cat on a chair"`

🎬 Text-to-Video

- Models: **Make-A-Video, Imagen Video, Runway Gen-2**
- Input: Text prompt → Output: Short video clip
- How it works: Temporal diffusion + frame consistency + motion modeling
- Applications: Movie concept scenes, education, marketing videos
- Example Prompt: `"A dragon flying over futuristic city at night"`

🎶 Text-to-Music / Composition

- Models: **MusicGen (Meta), Jukebox (OpenAI), Riffusion**
- Input: Text / Lyrics → Output: Music / Sound
- How it works: Transformer / diffusion for audio spectrogram generation
- Applications: Original music creation, game soundtracks, AI-assisted composition
- Example Prompt: `"Upbeat electronic music with Thai instruments"`

🤖 Basic Chatbot (GPT-2)

- Transformer-based **text generator** 📚
- Learns from huge datasets

👉 Example:

```
👩 (Student): Hello AI, can you tell me about the future of technology?
🤖 (AI): In the future, technology will...
```

Text-to-Text Generation

- **Core Models**: GPT-5, Gemini, Claude, LLaMA 3
- **Applications**: Chatbots, summarization, translation, code generation
- **Demo**: ChatGPT (OpenAI), Gemini (Google)

🎓 How to Fine-Tune GPT-2 to Generate Funny AI Texts

Teach your AI to tell jokes, speak like you, or ace the class!

🤖 What Is GPT-2?

- A powerful language model from OpenAI
- Trained on internet text
- Can generate human-like sentences
- Used for writing, chatting, coding, and more!

Fun Fact: It doesn't know the future—but it's great at pretending!

🔧 What's Fine-Tuning?

- Giving GPT-2 your own data
- Teaching it new patterns (jokes, answers, poems, etc.)
- Like training a parrot 🦜 — but way smarter!

Why?

To make GPT-2 sound like *you*, your classroom, or your sense of humor!

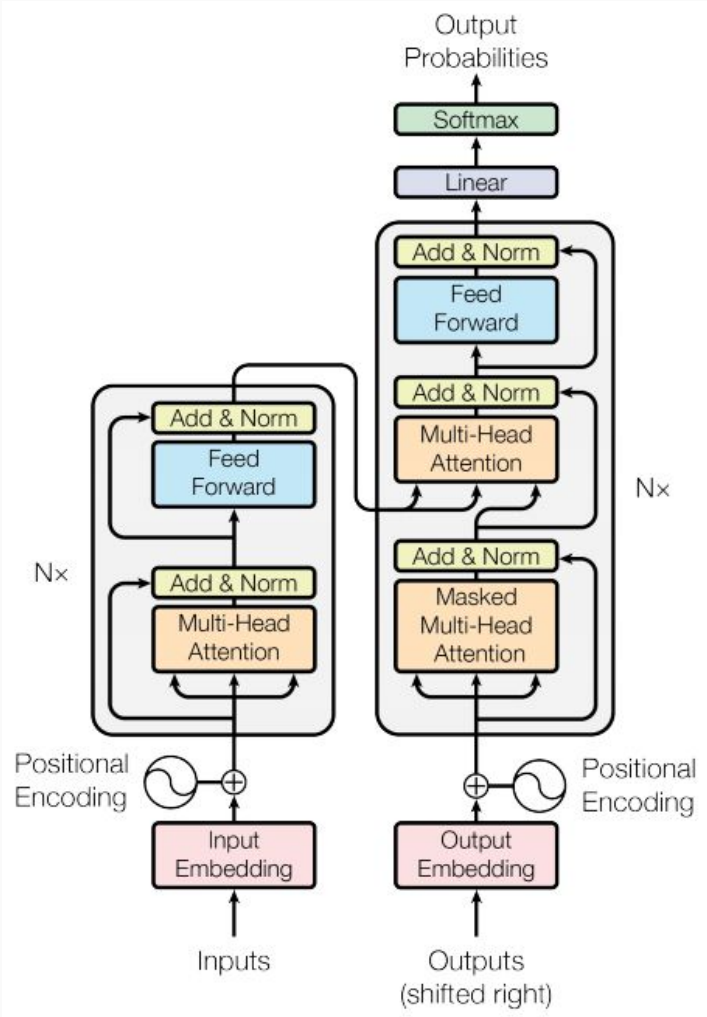
📄 Step 1: Create a Funny Text File

Example: `funny_quotes.txt`

txt

```
In the future, AI will teach math by rapping equations.
My robot teacher gave me candy instead of homework.
AI grading: “You got an A for effort, and a B for bandwidth.”
```

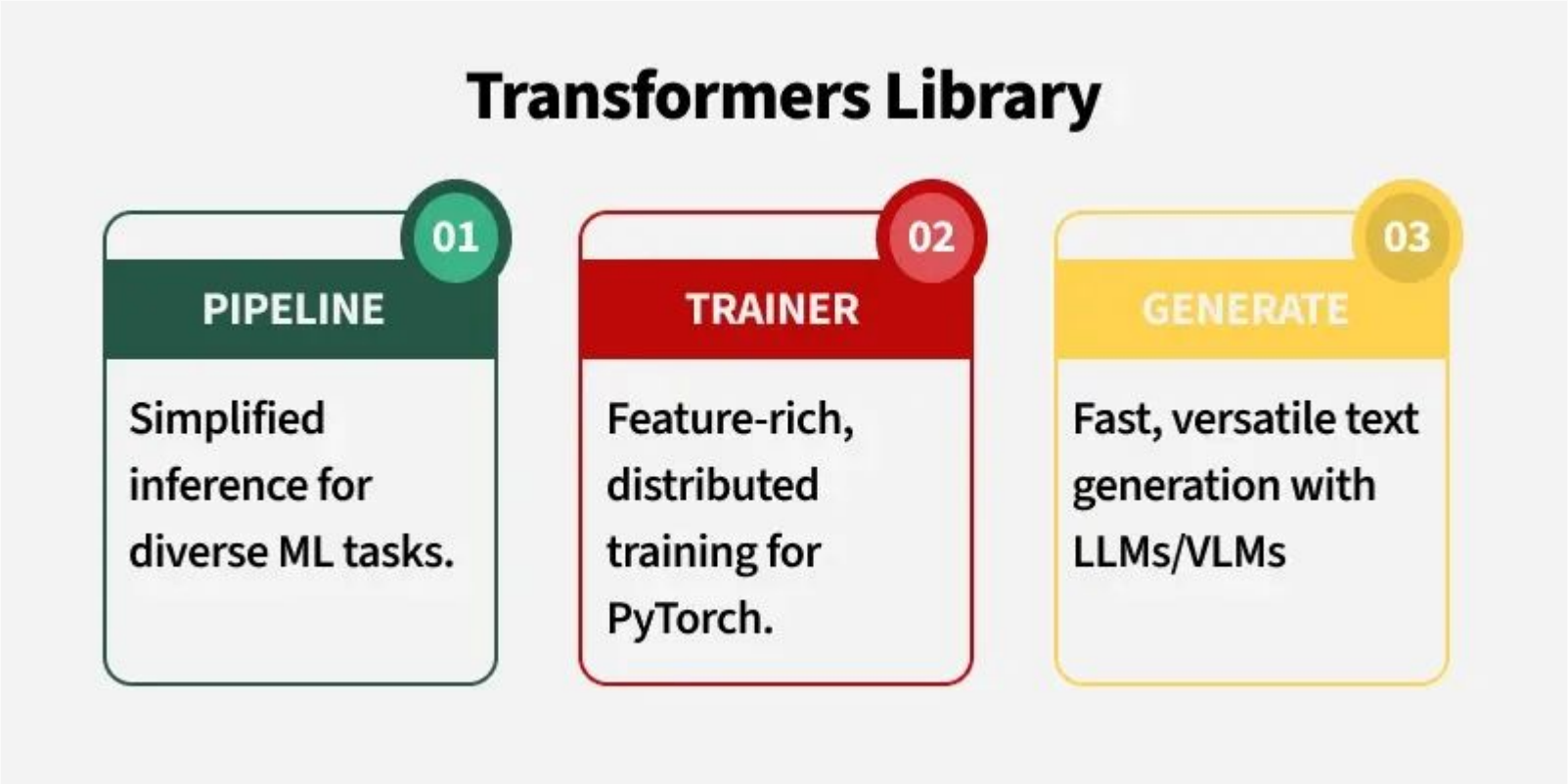
➡️ Think of your dataset as the AI’s bedtime story!



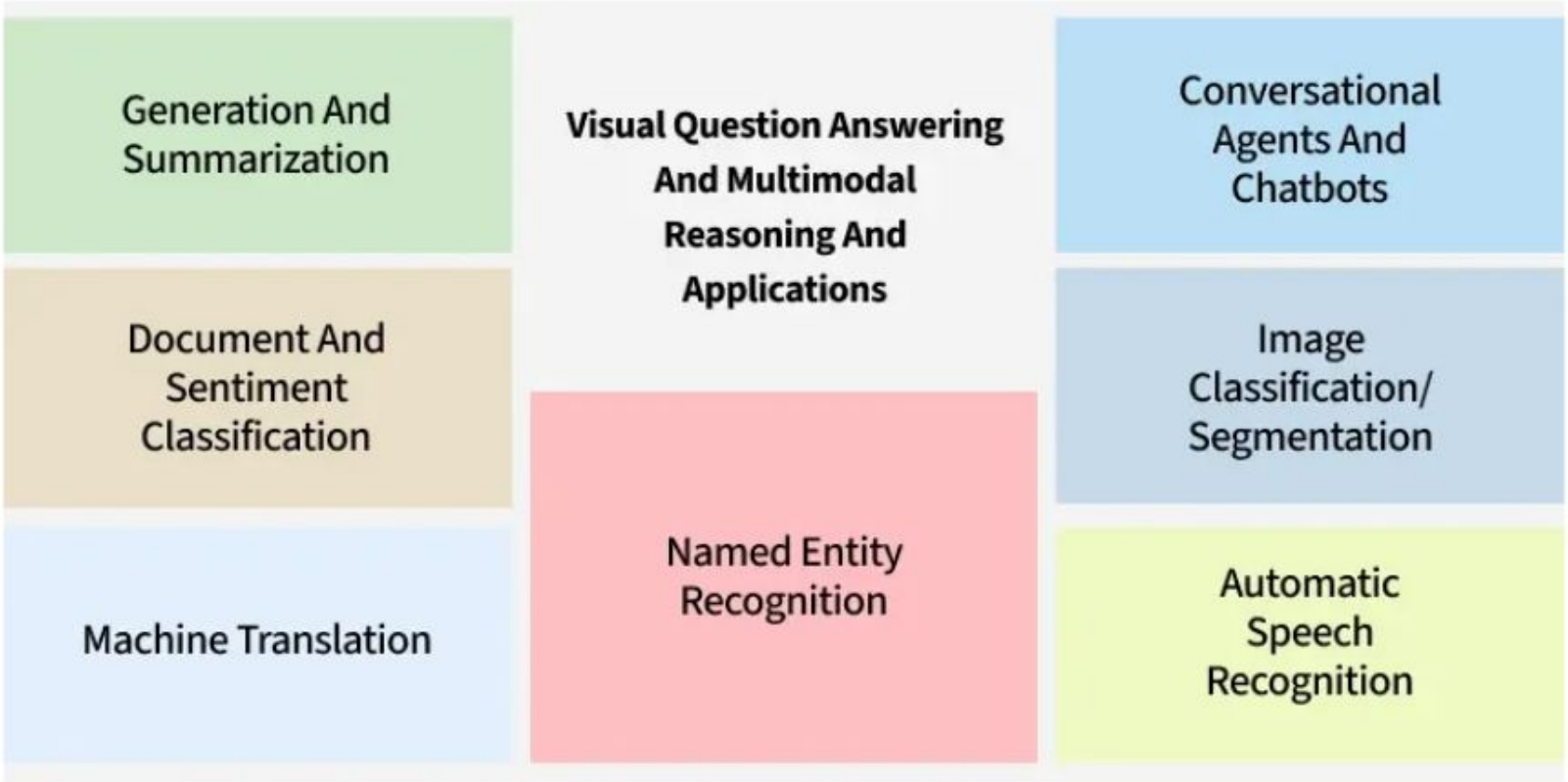
Transformers

build passing | license Apache-2.0 | website online | release v2.0.0

https://www.geeksforgeeks.org/deep-learning/transformers-library/



Use Cases of Transformers library



⚙️ Step 2: Install Tools

```bash
pip install transformers datasets accelerate
```

This gives you the AI "toolbox" from Hugging Face.

🧠 Step 3: Fine-Tune GPT-2

python Copy code

```python
from transformers import GPT2Tokenizer, GPT2LMHeadModel, Trainer, TrainingArguments

tokenizer = GPT2Tokenizer.from_pretrained("gpt2")
model = GPT2LMHeadModel.from_pretrained("gpt2")

# Load your text dataset
train_dataset = TextDataset(
    tokenizer=tokenizer,
    file_path="funny_quotes.txt",
    block_size=128
)
```

We're training it on our jokes!

🏋️ Step 4: Train and Save

python Copy code

```python
trainer = Trainer(
    model=model,
    args=TrainingArguments(
        output_dir="./gpt2-funny",
        per_device_train_batch_size=2,
        num_train_epochs=5,
        fp16=True,
    ),
    train_dataset=train_dataset,
)
trainer.train()
```

💾 Save the model:

python Copy code

```python
model.save_pretrained("./gpt2-funny")
```

🤖 Generate New Funny Text!

python

Copy code

```python
from transformers import pipeline

generator = pipeline("text-generation", model="./gpt2-funny")

prompt = "In the future, AI will"
print(generator(prompt, max_new_tokens=50)[0]["generated_text"])
```

✨ Now GPT-2 makes jokes your way!

🚀 Upgrade Ideas

- Use `gpt2-medium` or `gpt2-xl` for better results
- Train on song lyrics, speeches, or memes
- Use Google Colab to train in the cloud
- Upload your model to Hugging Face Hub and share with the world!

🖼️➡️📝 Image Captioning (BLIP)

- Input: Image 🖼️
- Output: Text caption 📝

👉 Example: Obama face → “A man in a suit smiling”

```
caption = processor.decode(out[0], skip_special_tokens=True)
print("🖼️➡️📝:", caption)
```

Real-World Tools You Should Know

Task	Model	Tool/Platform	Notes
Text-to-Text	GPT-4	ChatGPT	Code, summarization, chat
Text-to-Text	Gemini Pro	Google Gemini	Search-integrated reasoning
Text-to-Image	Midjourney	Discord Bot	Artistic image generation
Text-to-Image	DALL·E 3	ChatGPT	OpenAI native integration
Image-to-Text	GPT-4V	ChatGPT	Multimodal input
Text-to-3D	Magic3D	NVIDIA/Luma	3D mesh and NeRF

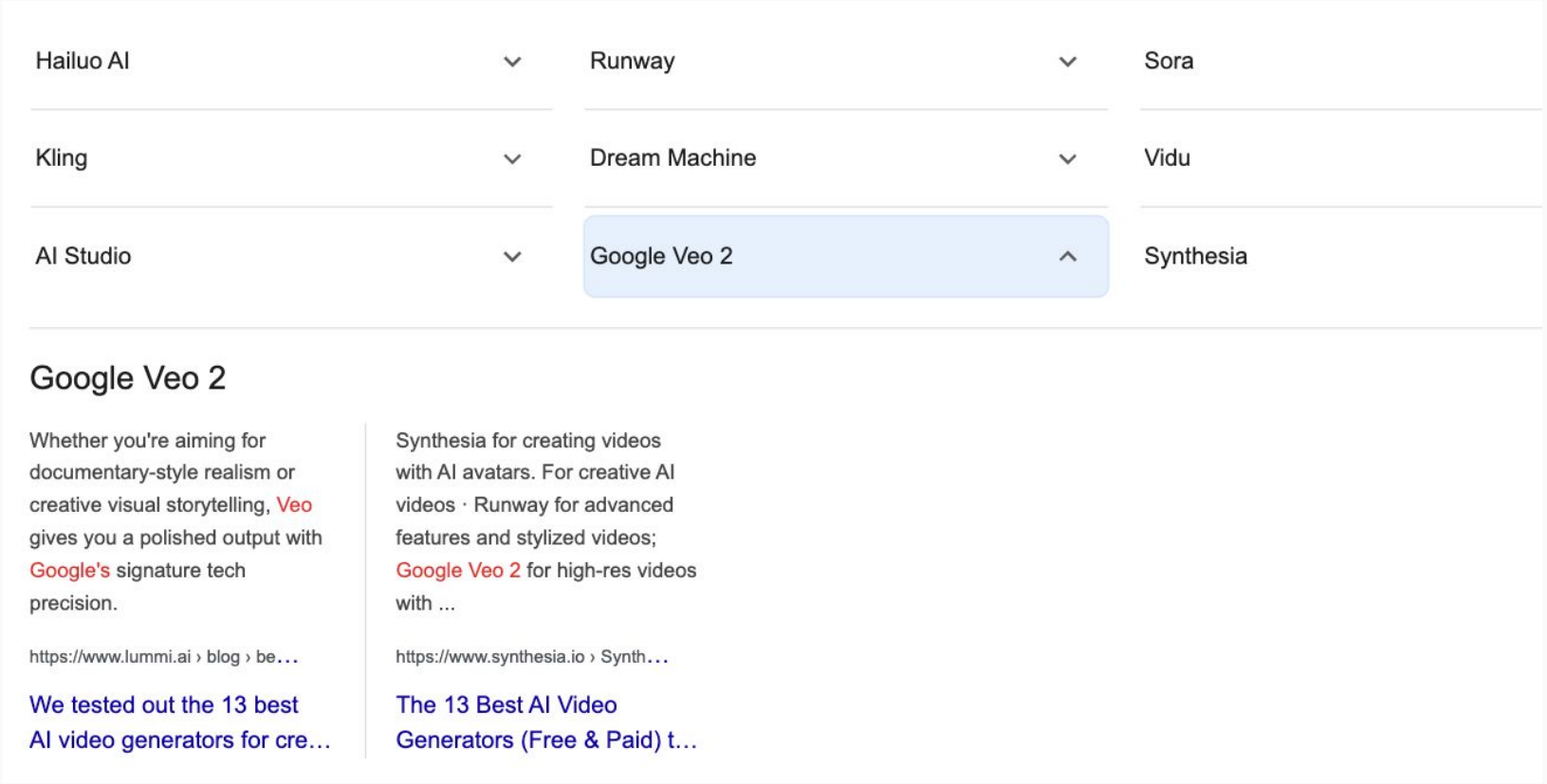
Real-World Tools You Should Know

Midjourney

ChatGPT

DALL·E

Gemini


Hailuo AI
Runway
Sora
Kling
Dream Machine
Vidu
AI Studio
Google Veo 2
Synthesia
Google Veo 2
Whether you're aiming for documentary-style realism or creative visual storytelling, Veo gives you a polished output with Google's signature tech precision.
https://www.lummi.ai › blog › be…
We tested out the 13 best AI video generators for cre…
Synthesia for creating videos with AI avatars. For creative AI videos · Runway for advanced features and stylized videos; Google Veo 2 for high-res videos with ...
https://www.synthesia.io › Synth…
The 13 Best AI Video Generators (Free & Paid) t…

Backend Technologies for GenAI Apps

- APIs: OpenAI, Stability AI, Google AI
- Hugging Face Transformers
- LangChain / LlamaIndex (for building apps)
- Vector DBs: Pinecone, Weaviate, FAISS
- Deployment: Streamlit, Gradio, Flask

Text-to-Image

- **Core Models: Midjourney, DALL·E 3, Stable Diffusion, Imagen**
- **How it works: Diffusion models + guidance**
- **Demo: Prompt to image (Midjourney, DALL·E)**

Image-to-Text (Vision-Language Models)

- **Core Models: GPT-5V, Gemini, Flamingo (DeepMind)**
- **Use Cases: Image captioning, accessibility, multimodal reasoning**
- **Demo: Upload image to ChatGPT/Gemini → generate caption**

Text-to-3D

- **Models: DreamFusion, Magic3D, Luma AI**
- **Techniques: NeRFs, Score Distillation Sampling (SDS)**
- **Use Cases: AR/VR, games, architecture**

🎨 GANs vs Diffusion

Feature	GANs (e.g., StyleGAN2)	Diffusion Models (e.g., SDXL)
Core Mechanism	Generator vs Discriminator	Iterative noise removal process
Training Stability	Often unstable, mode collapse common	More stable, though slower to train
Output Quality	High-quality, but prone to artifacts	Very high-quality, realistic and coherent
Control over Output	Limited, needs tricks (e.g., conditional GANs)	Easily guided by text or image prompts
Training Data Needs	Requires lots of data	Also data-hungry, but handles noise better
Inference Speed	Very fast once trained	Slower due to multi-step generation
Popular Use Cases	Face generation, deepfakes, image enhancement	Text-to-image, inpainting, 3D model generation
Popular Models	StyleGAN2, BigGAN	DALL·E 3, Stable Diffusion, Imagen, SDXL
Current Trend	Less common in modern GenAI	Dominant in SOTA generative models

🎨 Image ➡️ Image with Generative AI

- 🔄 **What is it?**

  Transform one image into another while preserving structure but changing **style, texture, or details**.

- 🛠️ **Applications**
  - Style transfer (photo → painting)
  - Super-resolution (low → high quality)
  - Inpainting (fill missing parts)
  - Face editing / enhancement
- 🤗 **Hugging Face Models**
  - `runwayml/stable-diffusion-inpainting` → remove/replace objects
  - `CompVis/stable-diffusion-img2img` → style transfer & modifications
  - `stabilityai/stable-diffusion-x4-upscaler` → image upscaling

- 💻 **Code Example (Stable Diffusion Img2Img)**

python Copy code

```python
from diffusers import StableDiffusionImg2ImgPipeline
import torch, PIL.Image

pipe = StableDiffusionImg2ImgPipeline.from_pretrained(
    "CompVis/stable-diffusion-img2img", torch_dtype=torch.float16
).to("cuda")

init_image = PIL.Image.open("Barack_Obama_Sample_Face.png").convert("RGB")
prompt = "Barack Obama in Van Gogh painting style"

image = pipe(prompt=prompt, image=init_image, strength=0.75, guidance_scale=7.5).images[0]
image.save("styled_obama.png")
```

- 🌟 **Key Idea:** Preserve input image → add creativity via prompts ✍️ + diffusion 🎨

kaze Toolkit Pricing API Blog Log in Try Kaze free

Home > AI Headshot Generator

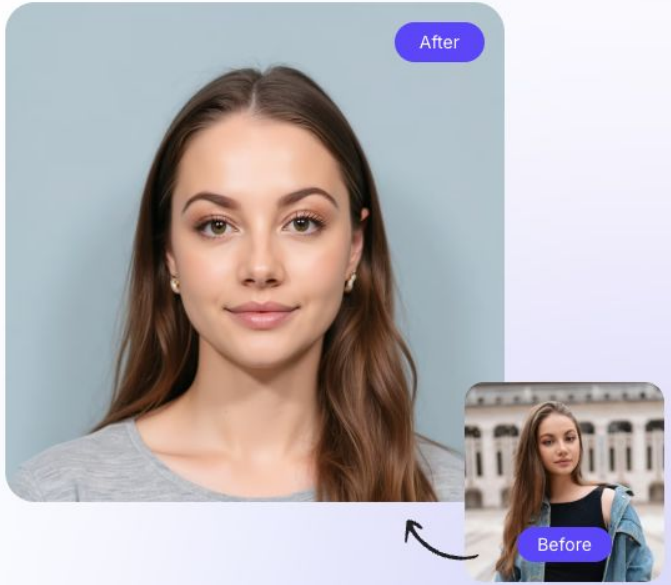
AI headshot generator

Create stunning AI-generated headshots in minutes. Look professional, feel confident, and leave a lasting impression!

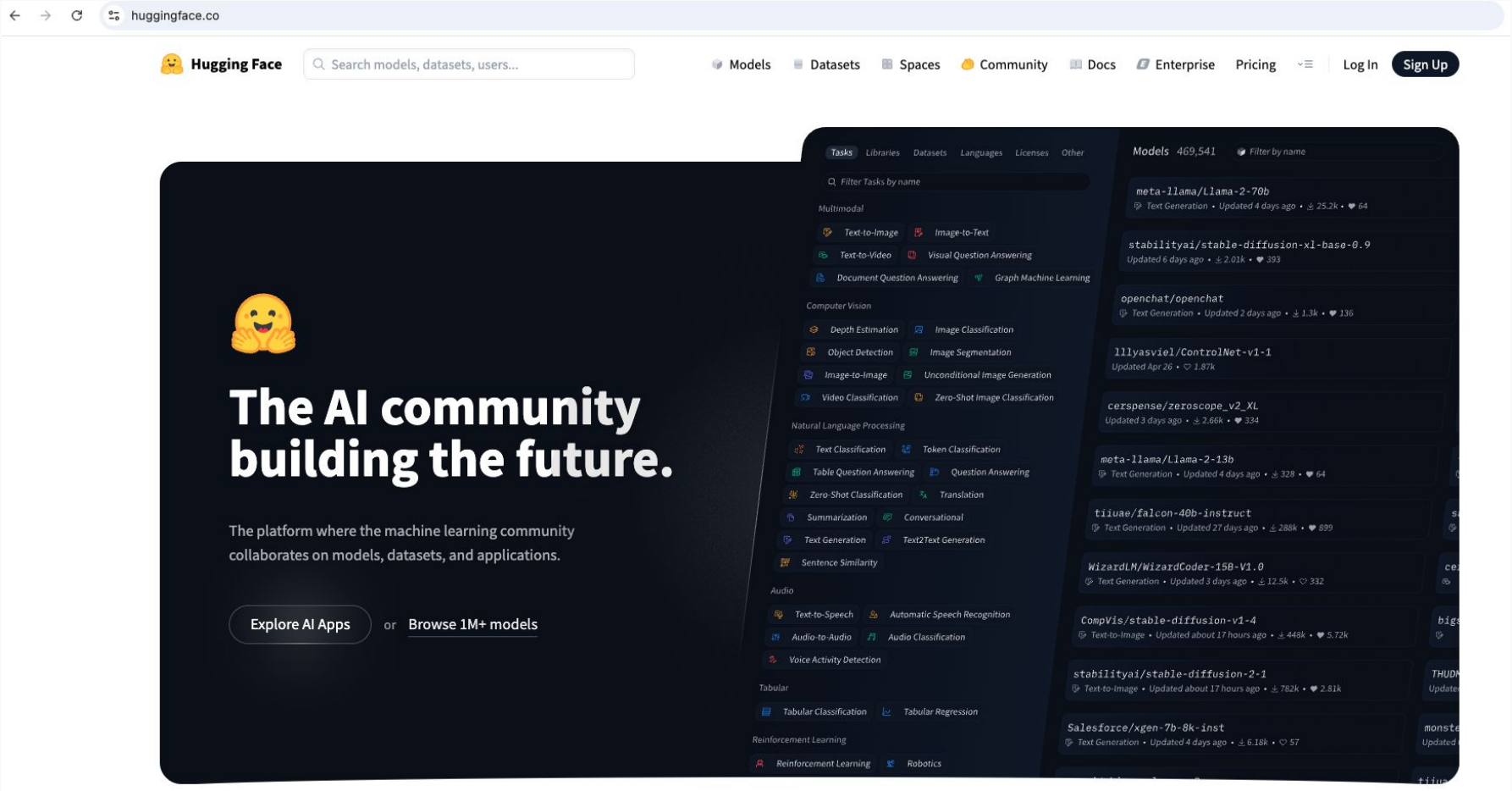
Create your headshot now

10K+ stunning professionals headshots generated

4.8 ⭐ average rating from google reviews



Hugging Face


huggingface.co
Hugging Face
Search models, datasets, users...
Models
Datasets
Spaces
Community
Docs
Enterprise
Pricing
Log In
Sign Up
The AI community building the future.
The platform where the machine learning community collaborates on models, datasets, and applications.
Explore AI Apps
or
Browse 1M+ models
Tasks
Libraries
Datasets
Languages
Licenses
Other
Filter Tasks by name
Multimodal
Text-to-Image
Image-to-Text
Text-to-Video
Visual Question Answering
Document Question Answering
Graph Machine Learning
Computer Vision
Depth Estimation
Image Classification
Object Detection
Image Segmentation
Image-to-Image
Unconditional Image Generation
Video Classification
Zero-Shot Image Classification
Natural Language Processing
Text Classification
Token Classification
Table Question Answering
Question Answering
Zero-Shot Classification
Translation
Summarization
Conversational
Text Generation
Text2Text Generation
Sentence Similarity
Audio
Text-to-Speech
Automatic Speech Recognition
Audio-to-Audio
Audio Classification
Voice Activity Detection
Tabular
Tabular Classification
Tabular Regression
Reinforcement Learning
Reinforcement Learning
Robotics
Models 469,541
Filter by name
meta-llama/Llama-2-70b
Text Generation • Updated 4 days ago • 25.2k • 64
stabilityai/stable-diffusion-xl-base-0.9
Updated 6 days ago • 2.01k • 393
openchat/openchat
Text Generation • Updated 2 days ago • 1.3k • 136
lllyasviel/ControlNet-v1-1
Updated Apr 26 • 1.87k
cerspense/zeroscope_v2_XL
Updated 3 days ago • 2.66k • 334
meta-llama/Llama-2-13b
Text Generation • Updated 4 days ago • 328 • 64
tiiuae/falcon-40b-instruct
Text Generation • Updated 27 days ago • 288k • 899
WizardLM/WizardCoder-15B-V1.0
Text Generation • Updated 3 days ago • 12.5k • 332
CompVis/stable-diffusion-v1-4
Text-to-Image • Updated about 17 hours ago • 448k • 5.72k
stabilityai/stable-diffusion-2-1
Text-to-Image • Updated about 17 hours ago • 782k • 2.81k
Salesforce/xgen-7b-8k-inst
Text Generation • Updated 4 days ago • 6.18k • 57

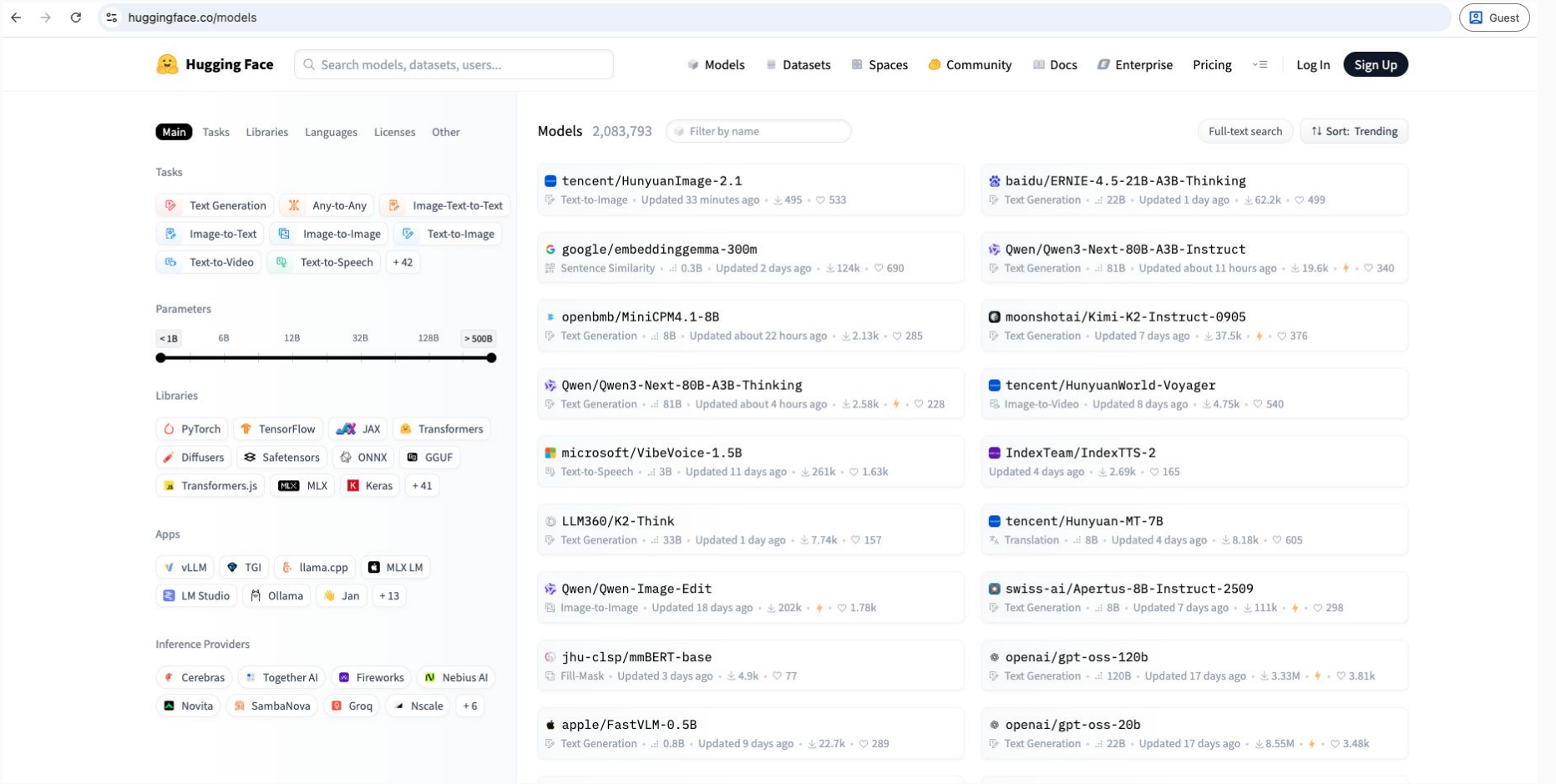
Hugging Face

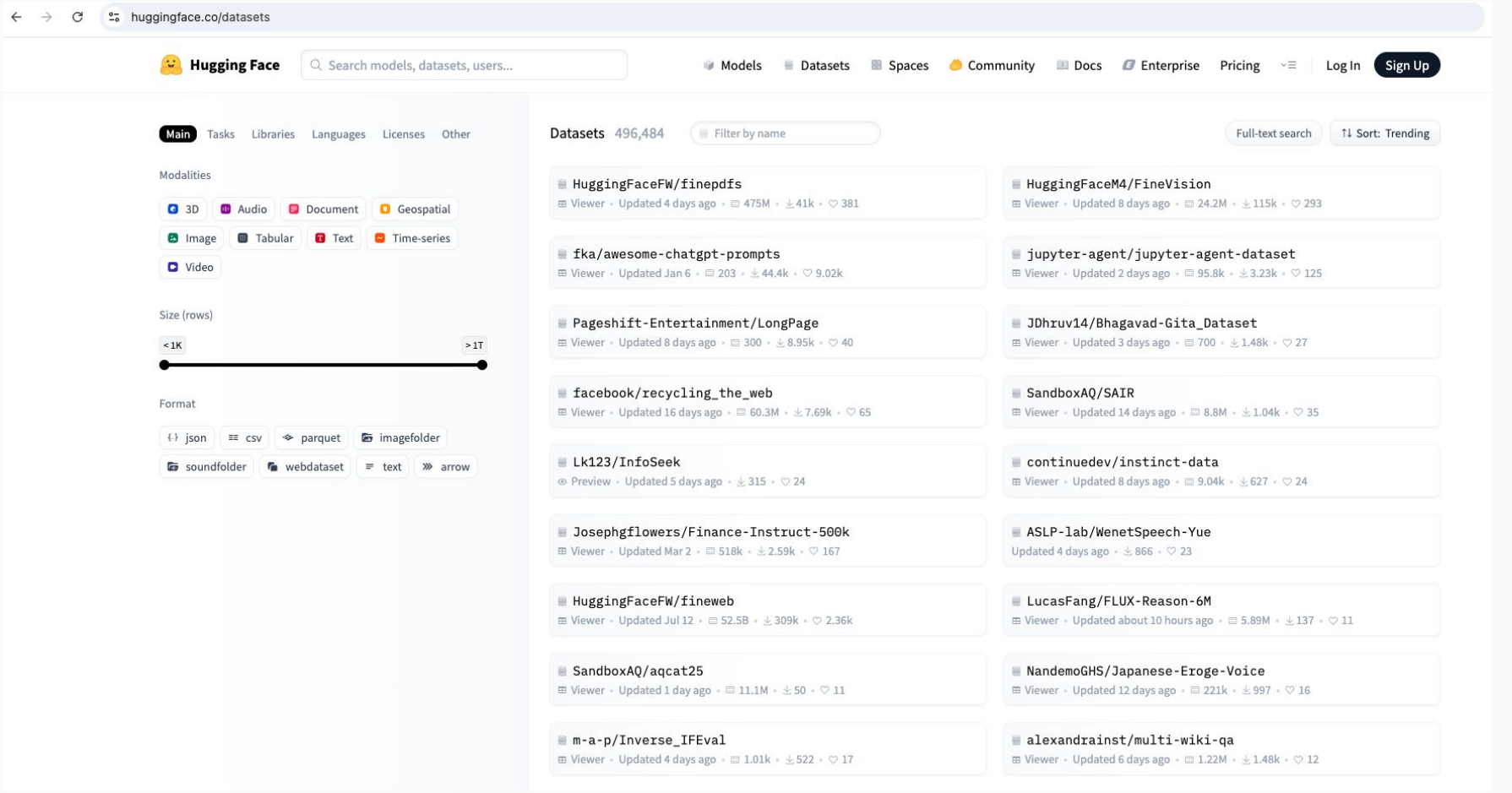
Search models, datasets, users...

Main | Tasks | Libraries | Languages | Licenses | Other

Tasks

- Text Generation
- Any-to-Any
- Image-Text-to-Text
- Image-to-Text
- Image-to-Image
- Text-to-Image
- Text-to-Video
- Text-to-Speech
- + 42


huggingface.co/models
Guest
Hugging Face
Search models, datasets, users...
Models
Datasets
Spaces
Community
Docs
Enterprise
Pricing
Log In
Sign Up
Main
Tasks
Libraries
Languages
Licenses
Other
Tasks
Text Generation
Any-to-Any
Image-Text-to-Text
Image-to-Text
Image-to-Image
Text-to-Image
Text-to-Video
Text-to-Speech
+ 42
Parameters
< 1B
6B
12B
32B
128B
> 500B
Libraries
PyTorch
TensorFlow
JAX
Transformers
Diffusers
Safetensors
ONNX
GGUF
Transformers.js
MLX
Keras
+ 41
Apps
vLLM
TGI
llama.cpp
MLX LM
LM Studio
Ollama
Jan
+ 13
Inference Providers
Cerebras
Together AI
Fireworks
Nebius AI
Novita
SambaNova
Groq
Nscale
+ 6
Models 2,083,793
Filter by name
Full-text search
Sort: Trending
tencent/HunyuanImage-2.1
Text-to-Image • Updated 33 minutes ago • 495 • 533
baidu/ERNIE-4.5-21B-A3B-Thinking
Text Generation • 22B • Updated 1 day ago • 62.2k • 499
google/embeddinggemma-300m
Sentence Similarity • 0.3B • Updated 2 days ago • 124k • 690
Qwen/Qwen3-Next-80B-A3B-Instruct
Text Generation • 81B • Updated about 11 hours ago • 19.6k • 340
openbmb/MiniCPM4.1-8B
Text Generation • 8B • Updated about 22 hours ago • 2.13k • 285
moonshotai/Kimi-K2-Instruct-0905
Text Generation • Updated 7 days ago • 37.5k • 376
Qwen/Qwen3-Next-80B-A3B-Thinking
Text Generation • 81B • Updated about 4 hours ago • 2.58k • 228
tencent/HunyuanWorld-Voyager
Image-to-Video • Updated 8 days ago • 4.75k • 540
microsoft/VibeVoice-1.5B
Text-to-Speech • 3B • Updated 11 days ago • 261k • 1.63k
IndexTeam/IndexTTS-2
Updated 4 days ago • 2.69k • 165
LLM360/K2-Think
Text Generation • 33B • Updated 1 day ago • 7.74k • 157
tencent/Hunyuan-MT-7B
Translation • 8B • Updated 4 days ago • 8.18k • 605
Qwen/Qwen-Image-Edit
Image-to-Image • Updated 18 days ago • 202k • 1.78k
swiss-ai/Apertus-8B-Instruct-2509
Text Generation • 8B • Updated 7 days ago • 111k • 298
jhu-clsp/mmBERT-base
Fill-Mask • Updated 3 days ago • 4.9k • 77
openai/gpt-oss-120b
Text Generation • 120B • Updated 17 days ago • 3.33M • 3.81k
apple/FastVLM-0.5B
Text Generation • 0.8B • Updated 9 days ago • 22.7k • 289
openai/gpt-oss-20b
Text Generation • 22B • Updated 17 days ago • 8.55M • 3.48k


huggingface.co/datasets
Hugging Face
Search models, datasets, users...
Models
Datasets
Spaces
Community
Docs
Enterprise
Pricing
Log In
Sign Up
Main
Tasks
Libraries
Languages
Licenses
Other
Modalities
3D
Audio
Document
Geospatial
Image
Tabular
Text
Time-series
Video
Size (rows)
< 1K
> 1T
Format
json
csv
parquet
imagefolder
soundfolder
webdataset
text
arrow
Datasets 496,484
Filter by name
Full-text search
Sort: Trending
HuggingFaceFW/finepdfs
Viewer • Updated 4 days ago • 475M • 41k • 381
HuggingFaceM4/FineVision
Viewer • Updated 8 days ago • 24.2M • 115k • 293
fka/awesome-chatgpt-prompts
Viewer • Updated Jan 6 • 203 • 44.4k • 9.02k
jupyter-agent/jupyter-agent-dataset
Viewer • Updated 2 days ago • 95.8k • 3.23k • 125
Pageshift-Entertainment/LongPage
Viewer • Updated 8 days ago • 300 • 8.95k • 40
JDhruv14/Bhagavad-Gita_Dataset
Viewer • Updated 3 days ago • 700 • 1.48k • 27
facebook/recycling_the_web
Viewer • Updated 16 days ago • 60.3M • 7.69k • 65
SandboxAQ/SAIR
Viewer • Updated 14 days ago • 8.8M • 1.04k • 35
Lk123/InfoSeek
Preview • Updated 5 days ago • 315 • 24
continuedev/instinct-data
Viewer • Updated 8 days ago • 9.04k • 627 • 24
Josephgflowers/Finance-Instruct-500k
Viewer • Updated Mar 2 • 518k • 2.59k • 167
ASLP-lab/WenetSpeech-Yue
Updated 4 days ago • 866 • 23
HuggingFaceFW/fineweb
Viewer • Updated Jul 12 • 52.5B • 309k • 2.36k
LucasFang/FLUX-Reason-6M
Viewer • Updated about 10 hours ago • 5.89M • 137 • 11
SandboxAQ/aqcat25
Viewer • Updated 1 day ago • 11.1M • 50 • 11
NandemoGHS/Japanese-Eroge-Voice
Viewer • Updated 12 days ago • 221k • 997 • 16
m-a-p/Inverse_IFEval
Viewer • Updated 4 days ago • 1.01k • 522 • 17
alexandrainst/multi-wiki-qa
Viewer • Updated 6 days ago • 1.22M • 1.48k • 12

🌟 Hugging Face Introduction

- Hugging Face = “GitHub for AI Models”
- Ecosystem: 🤗 Transformers, Datasets, Tokenizers, Diffusers
- 10,000+ pretrained models → NLP, Vision, Audio, Multimodal
- Community-driven + Open Source

🔍 Finding Models on Hugging Face Hub

- Website: huggingface.co/models ↗
- Filter by:
  - ✅ Task (NLP, CV, Speech, Multimodal)
  - ✅ Library (PyTorch, TensorFlow, JAX)
  - ✅ Dataset
- Example models: `bert-base-uncased`, `runwayml/stable-diffusion-v1-5`

🚀 Using Pretrained Models (Text-to-Text)

python

Copy code

```python
from transformers import pipeline

classifier = pipeline("sentiment-analysis")
print(classifier("Generative AI is inspiring! 🚀"))
```

- **Pipeline API** = model + tokenizer + inference in one line
- Perfect for class demos & prototyping

🎨 Using Diffusion (Text-to-Image)

python

Copy code

```python
from diffusers import StableDiffusionPipeline
import torch

pipe = StableDiffusionPipeline.from_pretrained(
    "runwayml/stable-diffusion-v1-5",
    torch_dtype=torch.float16
).to("cuda")

image = pipe("A futuristic city at sunset 🌇").images[0]
image.show()
```

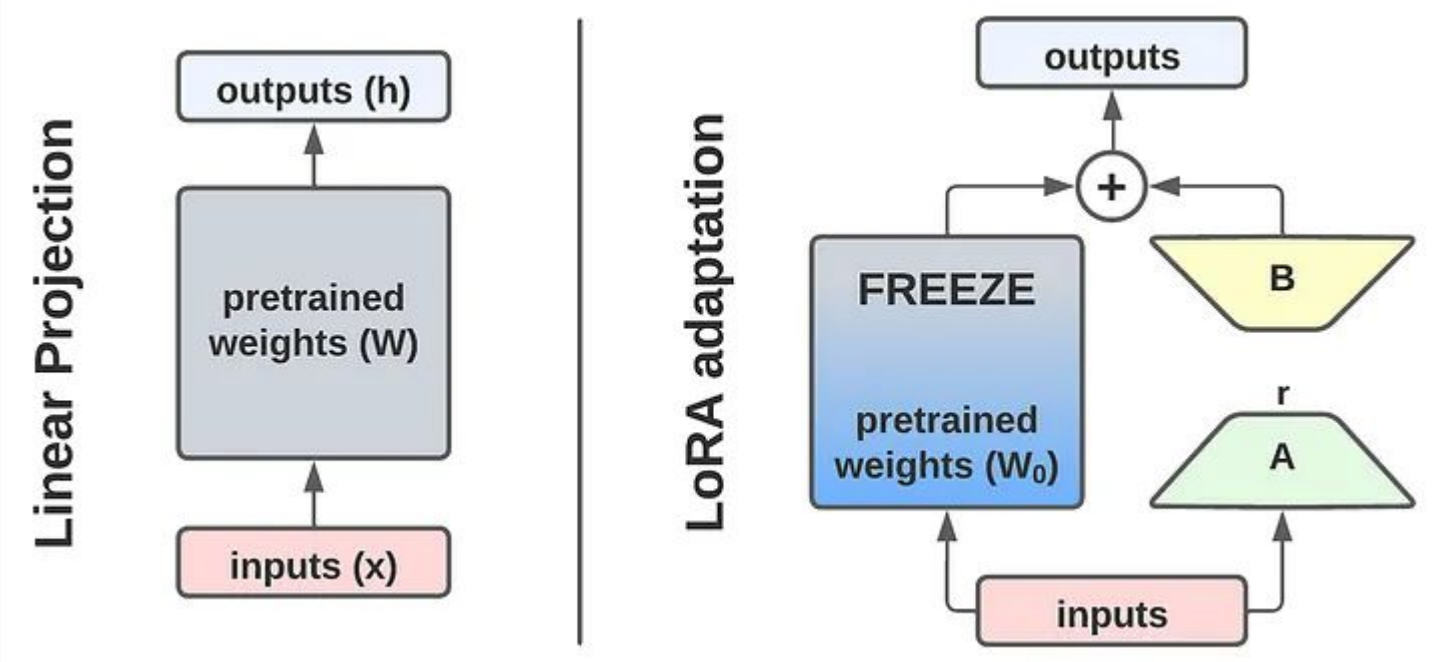
- Stable Diffusion = state-of-the-art text-to-image
- Runs on free Colab GPU

🛠️ Fine-Tuning Models (Concept)

- Fine-tuning = start from pretrained → adapt to your dataset
- Approaches:
  - Full fine-tuning (all weights updated)
  - LoRA / PEFT (train small adapters → faster, lightweight)
- Example: fine-tune BERT for Thai text classification

https://medium.com/@arnabdutta8986/fhow-to-implement-lora-low-rank-adaptation-from-scratch-adaa36e2e275

https://pub.towardsai.net/low-rank-adaptation-lora-fedf37b92026



✨ LoRA: Low-Rank Adaptation for Fine-Tuning

- 🔷 **Problem**: Full fine-tuning = heavy (billions of parameters)
- 🔷 **Solution**: Insert small trainable adapters → freeze original weights
- 🔷 **Benefit**:
  - Train only **0.1–1%** of parameters
  - Faster, cheaper, less GPU memory
  - Reusable adapters for multiple tasks
- 🔷 **Code Example** (with 🤗 PEFT):

python Copy code

```python
from peft import LoraConfig, get_peft_model
from transformers import AutoModelForCausalLM

model = AutoModelForCausalLM.from_pretrained("gpt2")
config = LoraConfig(r=8, lora_alpha=32, target_modules=["q_proj","v_proj"])
model = get_peft_model(model, config)
```

- 🎯 **Use case**: Fine-tune GPT / BERT / Stable Diffusion in class with limited GPU

⚡ Tips & Best Practices

- Use `pipeline` for quick experiments
- Use `from_pretrained()` for direct model access
- Prefer lightweight fine-tuning (LoRA, adapters) in class demos
- Cache models to avoid re-downloads on Colab
- Explore Spaces on Hugging Face for interactive demos

⚡ Training vs Fine-Tuning vs Pretraining vs Inference

Concept	Definition	Purpose	Example	Cost 💰
Training	Learn model weights from scratch on a dataset	Build a model without prior knowledge	Train CNN on cats vs dogs	🔴 Very High
Pretraining	Train on a large, general dataset first	Provide strong initial weights for many tasks	BERT trained on Wikipedia	🟠 High
Fine-Tuning	Adapt a pretrained model to a specific dataset/task	Customize for your problem	Fine-tune BERT for sentiment analysis	🟡 Medium
Inference	Use the trained/fine-tuned model to make predictions	Apply the model in real-world tasks	GPT generates a response	🟢 Low

https://www.brilworks.com/blog/limitations-of-generative-ai/

Limitations of Generative AI

Limited by Training Data

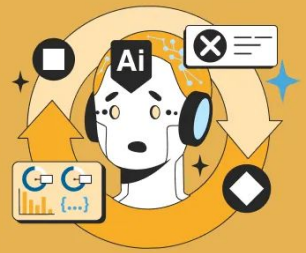
Generative AI models are only as good as the data they're trained on, leading to potential biases and inaccuracies.

Struggling with True Creativity

While impressive, generative AI can't truly conceive entirely new ideas, but rather assemble existing ones.

Lacking Nuanced Understanding

While impressive, generative AI can't truly conceive entirely new ideas, but rather assemble existing ones.



Difficulty with Adaptability

Adapting to new situations or tasks beyond their training proves challenging for generative AI.

Data Privacy and Security

Ensuring the privacy and security of vast amounts of data used to train generative AI remains a concern.

Q&A in Generative AI?