

WEEK 1 · DAY 1 · AFTERNOON

A bit of AI history and current AI Tech Stacks

My background is in Computer Vision.

A few years ago I submitted with some colleagues a “Trustworthy AI Systems” NSF/NRT grant

In my research I do AI/ML/Vision systems, including running a project where we deploy to NCMEC a tool to detect the hotel rooms in the background of sex trafficking images, which generates my interest in security of high stakes image ML tools.

<https://pless.github.io/trustworthyAI/>

1. Introduction to Trust & AI by Robert Pless
2. AI and High Tech Product Liability
3. AI, Publishing, and Ethics/Research Integrity
4. Mechanisms of Human Oversight in Training AI Systems
5. AI and Federal Food and Nutrition Programs
6. SOTA: AI Agents
7. Anthropological understanding of language; What is missing from Large Language Models (LLMs)
8. AI and Art Production

[Project 1: Characterize AI Model Errors](#)

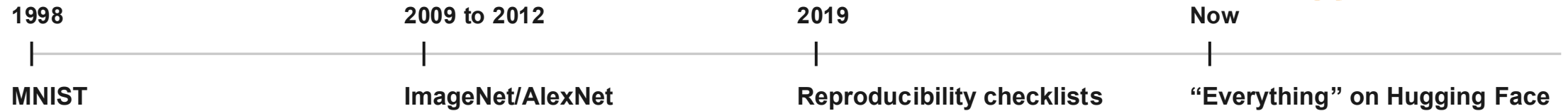
[Project 2: Games as a model of LLM alignment](#)

[Project 3: Selective Truthiness, Do LLMs tell different things to different people](#)

[Project 4: Personalizing Trust, Use AI for something that matters to you](#)

[Final Project: Write a paper submittable to AAAI AI Ethics and Society Conference](#)

AI/ML was early on the "replicability" train

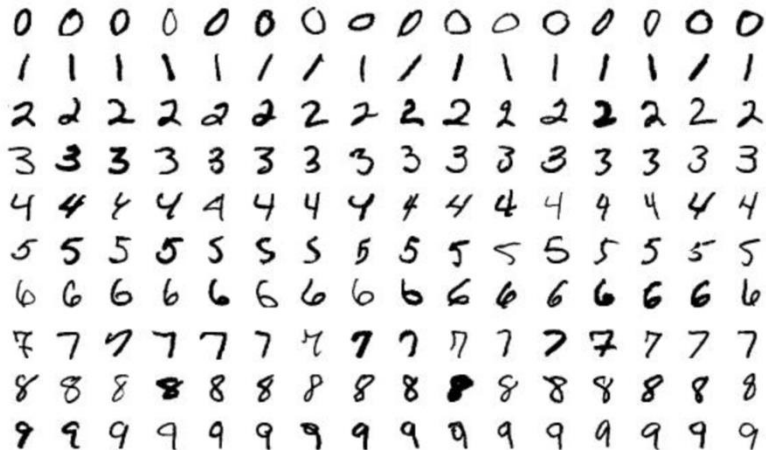


LeCun posts a handwriting dataset on his homepage.

An open dataset and a public contest. "AlexNet" in 2012 dominates the competition and shares complete specifications of their architecture.

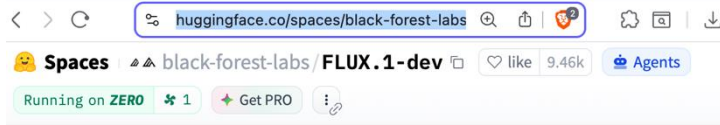
NeurIPS now requires a data checklist and code (& now an ethics statement). Papers sharing materials jump from about half to over three quarters.

Weights, datasets, and demos sit in one place. Sharing is the default, not the exception.



You can access/modify SOTA academic ML tools

Image generation



FLUX.2 [dev] is here! [Try it out here](#)

FLUX.1 [dev]

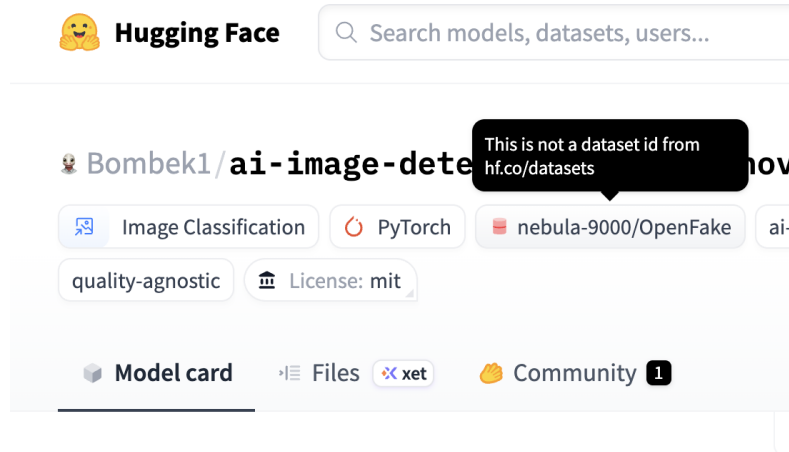
12B param rectified flow transformer guidance-distilled from [FLUX.1 \[pro\]](#)
[\[non-commercial license\]](#) [\[blog\]](#) [\[model\]](#)

an army of robots chanting "YOLO"

Run



AI Image detection



AI Image Detector (SigLIP2 + DINOv2 Ensemble)

A high-accuracy, **quality-agnostic** model for detecting AI-generated images, achieving **0.9997 AUC** on validation and strong cross-dataset generalization.



Afternoon goal:

Share some of the landscape of existing AI models and tools.

Give some examples of ways that you (or your students) can interact with these models.

Explore the "model zoo" in huggingface and share your favorite discoveries

PART B · WHAT IS AVAILABLE

Image generation

huggingface.co/spaces/black-forest-labs

Spaces black-forest-labs / FLUX.1-dev like 9.46k Agents

Running on ZERO 1 Get PRO

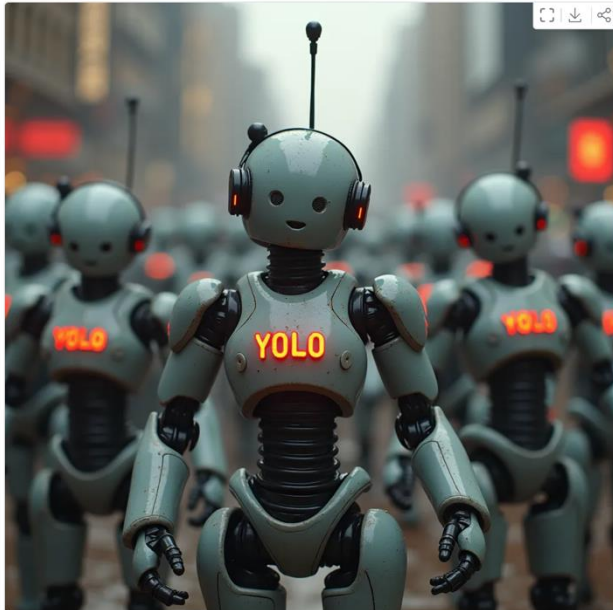
FLUX.2 [dev] is here! [Try it out here](#)

FLUX.1 [dev]

12B param rectified flow transformer guidance-distilled from [FLUX.1 \[pro\]](#)
[\[non-commercial license\]](#) [\[blog\]](#) [\[model\]](#)

an army of robots chanting "YOLO"

Run



<https://huggingface.co/black-forest-labs/FLUX.1-dev>

huggingface.co/black-forest-labs/FLUX.1-dev

black-forest-labs / FLUX.1-dev like 13k Follow Black Forest Labs 37.7k

Text-to-Image Diffusers Safetensors English FluxPipeline image-generation flux

License: flux-1-dev-non-commercial-license

Model card Files xet Community 593 Deploy Copy to bucket NEW Use this model

Gated model You have been granted access to this model



FLUX.1 [dev] is a 12 billion parameter rectified flow transformer capable of generating images from text descriptions. For more information, please read our [blog post](#).

Key Features

1. Cutting-edge output quality, second only to our state-of-the-art model [FLUX.1 \[pro\]](#).
2. Competitive prompt following, matching the performance of closed source alternatives.
3. Trained using guidance distillation, making FLUX.1 [dev] more efficient.

Downloads last month
701,789



Inference Providers NEW fal

Text-to-Image

Your prompt here...

Generate

View Code

Maximize

Model tree for black-forest-labs/FLUX.1-dev

| | |
|---------------|--------------|
| Adapters | 41937 models |
| Finetunes | 573 models |
| Merges | 10 models |
| Quantizations | 69 models |

Spaces using black-forest-labs/FLUX.1-dev 100

black-forest-labs/FLUX.1-dev

black-forest-labs/FLUX.1-Krea-dev

black-forest-labs/FLUX.1-Fill-dev

black-forest-labs/FLUX.1-Redux-dev

black-forest-labs/FLUX.1-Depth-dev

black-forest-labs/FLUX.1-canny-dev

TencentARC/Pixal3D yanze/PuLID-FLUX

BoobyBoobs/Flux_Lustly_AI_Uncensored_NSFW_V1

PART B · WHAT IS AVAILABLE

Image generation

huggingface.co/spaces/black-forest-labs

Spaces black-forest-labs / FLUX.1-dev like 9.46k Agents

Running on ZERO 1 Get PRO

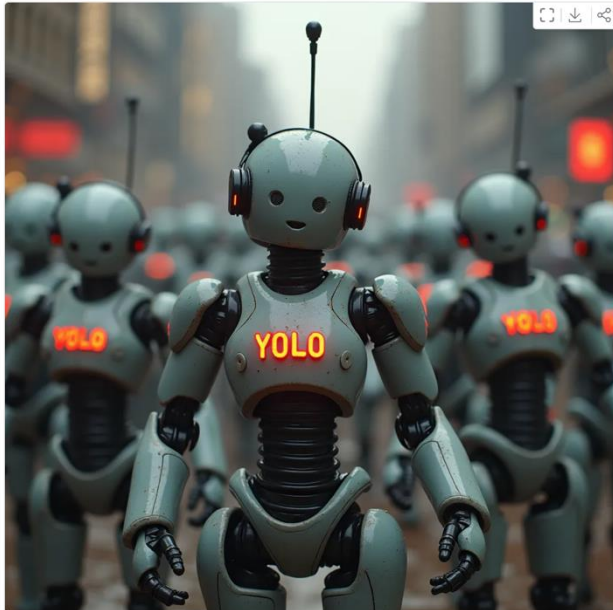
FLUX.2 [dev] is here! [Try it out here](#)

FLUX.1 [dev]

12B param rectified flow transformer guidance-distilled from [FLUX.1 \[pro\]](#)
[\[non-commercial license\]](#) [\[blog\]](#) [\[model\]](#)

an army of robots chanting "YOLO"

Run



<https://huggingface.co/black-forest-labs/FLUX.1-dev>

huggingface.co/black-forest-labs/FLUX.1-dev

black-forest-labs / FLUX.1-dev like 13k Follow Black Forest Labs 37.7k

Text-to-Image Diffusers Safetensors English FluxPipeline image-generation flux

License: flux-1-dev-non-commercial-license

Model card Files xet Community 593 Deploy Copy to bucket Use this model

Gated model You have been granted access to this model

Downloads last month 701,789

Inference Providers Text-to-Image Your prompt here... Generate View Code Maximize

Model tree for black-forest-labs/FLUX.1-dev

- Adapters 41937 models
- Finetunes 573 models
- Merges 10 models
- Quantizations 69 models

Spaces using black-forest-labs/FLUX.1-dev 100

- black-forest-labs/FLUX.1-dev
- black-forest-labs/FLUX.1-Krea-dev
- black-forest-labs/FLUX.1-Fill-dev
- black-forest-labs/FLUX.1-Redux-dev
- black-forest-labs/FLUX.1-Depth-dev
- black-forest-labs/FLUX.1-canny-dev
- TencentARC/Pixel3D
- yanze/PuLID-FLUX
- BoobyBoobs/Flux_Lustly_AI_Uncensored_NSFW_V1

FLUX.1 [dev] is a 12 billion parameter rectified flow transformer capable of generating images from text descriptions. For more information, please read our [blog post](#).

Key Features

1. Cutting-edge output quality, second only to our state-of-the-art model FLUX.1 [pro].
2. Competitive prompt following, matching the performance of closed source alternatives .
3. Trained using guidance distillation, making FLUX.1 [dev] more efficient.

PART B · WHAT IS AVAILABLE

<https://huggingface.co/black-forest-labs/FLUX.1-dev>

Image generation

huggingface.co/spaces/black-forest-labs

Spaces black-forest-labs / FLUX.1-dev like 9.46k Agents

Running on ZERO 1 Get PRO

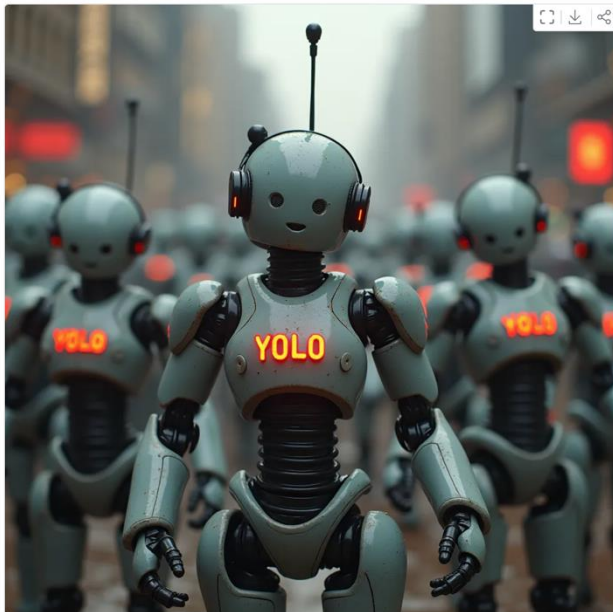
FLUX.2 [dev] is here! [Try it out here](#)

FLUX.1 [dev]

12B param rectified flow transformer guidance-distilled from [FLUX.1 \[pro\]](#)
[\[non-commercial license\]](#) [\[blog\]](#) [\[model\]](#)

an army of robots chanting "YOLO"

Run



black-forest-labs / FLUX.1-dev like 13k Follow Black Forest Labs 37.7k

Text-to-Image Diffusers Safetensors English FluxPipeline image-generation flux

License: flux-1-dev-non-commercial-license

Model card Files xet Community 593

Deploy Copy to bucket NEW Use this model

main FLUX.1-dev 57.9 GB Go to file 7 contributors History: 11 commits + Contribute

| | | | | |
|-----------------------|------------------------------------|----------|------------------------------------|------------------|
| jaymueller | Update LICENSE.md | 3de623f | VERIFIED | 11 months ago |
| scheduler | Add diffusers format weights (#... | | | almost 2 year... |
| text_encoder | Add diffusers format weights (#... | | | almost 2 year... |
| text_encoder_2 | Add diffusers format weights (#... | | | almost 2 year... |
| tokenizer | Add diffusers format weights (#... | | | almost 2 year... |
| tokenizer_2 | Add diffusers format weights (#... | | | almost 2 year... |
| transformer | Add diffusers format weights (#... | | | almost 2 year... |
| vae | Add diffusers format weights (#... | | | almost 2 year... |
| .gitattributes | 1.66 kB | Safe | Initial commit | almost 2 year... |
| LICENSE.md | 18.6 kB | Safe | Update LICENSE.md | 11 months ago |
| README.md | 4.39 kB | Safe | Update README.md | almost 2 year... |
| ae.safetensors | 335 MB | Safe xet | Rename ae.sft to ae.safetensors | almost 2 year... |
| dev_grid.jpg | 1.3 MB | Safe xet | Initial commit | almost 2 year... |
| flux1-dev.safetensors | 23.8 GB | Safe xet | Rename flux1-dev.sft to flux1-d... | almost 2 year... |
| model_index.json | 536 Bytes | Safe | Add diffusers format weights (#... | almost 2 year... |

PART B · WHAT IS AVAILABLE

Image generation requires ~24 GB for inference (e.g. in order to produce an output), making it a laptop/server based inference model.

<https://huggingface.co/black-forest-labs/FLUX.1-dev>

black-forest-labs / **FLUX.1-dev** like 13k Follow Black Forest Labs 37.7k

Text-to-Image Diffusers Safetensors English FluxPipeline image-generation flux

License: flux-1-dev-non-commercial-license

Model card Files xet Community 593 Deploy Copy to bucket Use this model

main FLUX.1-dev 57.9 GB Go to file 7 contributors History: 11 commits + Contribute

| | | | | |
|-----------------------|------------------------------------|-------------------|------------------------------------|------------------|
| jaymueller | Update LICENSE.md | 3de623f | VERIFIED | 11 months ago |
| scheduler | Add diffusers format weights (#... | almost 2 year... | | |
| text_encoder | Add diffusers format weights (#... | almost 2 year... | | |
| text_encoder_2 | Add diffusers format weights (#... | almost 2 year... | | |
| tokenizer | Add diffusers format weights (#... | almost 2 year... | | |
| tokenizer_2 | Add diffusers format weights (#... | almost 2 year... | | |
| transformer | Add diffusers format weights (#... | almost 2 year... | | |
| vae | Add diffusers format weights (#... | almost 2 year... | | |
| .gitattributes | 1.66 kB | Initial commit | almost 2 year... | |
| LICENSE.md | 18.6 kB | Update LICENSE.md | 11 months ago | |
| README.md | 4.39 kB | Update README.md | almost 2 year... | |
| ae.safetensors | 335 MB | xet | Rename ae.sft to ae.safetensors | almost 2 year... |
| dev_grid.jpg | 1.3 MB | xet | Initial commit | almost 2 year... |
| flux1-dev.safetensors | 23.8 GB | xet | Rename flux1-dev.sft to flux1-d... | almost 2 year... |
| model_index.json | 536 Bytes | | Add diffusers format weights (#... | almost 2 year... |

Inference Model Size

Even if you can download weights, how hard is it to run the inference?

Laptop

CPU, Apple Silicon, or a small GPU (about 8 GB)

BERT

text / NLP

YOLO

object detection

Whisper

speech to text

SAM

image segmentation

Stable Diffusion 1.5

text to image

GPU server

one 16 to 80 GB GPU (free Colab T4 at the entry)

FLUX.1

~24 GB native, ~8 to 13 GB quantized

Stable Diffusion XL / 3.5

text to image

Llama 8B to 70B

text / LLM

Data center

many GPUs in a cluster

Llama 405B

text / LLM

DeepSeek-scale (~671B)

text / LLM

The largest open models

frontier scale

“Open weights”

”Open”

| Type | What it lets you do | Example models |
|-----------------------------|---|------------------------------------|
| Permissive | Use, modify, and ship almost anything, including commercial work. (MIT, Apache-2.0) | Whisper, SAM, Qwen, Mistral |
| Copyleft | Free to use, but derivatives and often network use must be shared back. (AGPL-3.0) | Some YOLO versions |
| Restricted community | Free with conditions: usage caps, banned uses, or no-training-other-models clauses. | Llama, FLUX.1-dev (non-commercial) |

The first question is not "can I download it." It is "what does the license actually allow."

Closed models (for comparison)

Do not share their weights. Offered via API or as part of a device.

On-device

runs locally on your phone or laptop, but closed

Apple Intelligence

~3B on iPhone, iPad, Mac

Gemini Nano

Android AI Core, Chrome

Phi-Silica

Copilot+ PCs

Mid tier (API)

the cheap, fast members of each lineup

Claude Haiku

fast, low cost

GPT-5 mini / nano

high volume

Gemini Flash / Flash-Lite

cheap, quick

Frontier (API)

served from large clusters, API only

Claude Opus 4.8

Anthropic

GPT-5.5

OpenAI

Gemini 3.1 Pro

Google

How to reach each tier

Laptop

Install Ollama for local language models.

Use Python with scikit-learn or PyTorch.

GPU server

Use free Colab, rent a GPU from the Cloud or use university HPC.

Data center scale

Rent inference through a hosted API. There is no hardware to manage. University systems/Hospitals sometime run their own model for data privacy reasons

Anatomy of an API call

A hosted model is just an HTTP endpoint. You send a request describing what you want, and you get a structured reply back.

```
# talk to a model over the network
from openai import OpenAI
client = OpenAI(api_key="sk-...")

resp = client.chat.completions.create(
    model="gpt-4o",    # the one line you'll swap next
    messages=[
        {"role": "system", "content": "You are a tutor."},
        {"role": "user", "content": "What is a contrail?"},
    ],
)

print(resp.choices[0].message.content)
```

You send a request.

A **model name** plus a list of **messages**, each tagged with a role: system, user, or assistant.

You get structured JSON back.

The text lives at `choices[0].message.content`, alongside token counts and metadata.

OpenRouter: one key, many models

OpenRouter is the practical way to rent inference. One key reaches many hosted models, both open and closed.

```
# you change one string to swap models
model = "anthropic/claude-sonnet-4.6"
model = "meta-llama/llama-3.3-70b"
client.chat.completions.create(model, ...)
```

Routes your request across providers of a model finding "best" (cheap, quick) response. Open models (Llama, DeepSeek, Qwen) have many companies that run data centers and serve the model to you.

Closed models (Claude, GPT) = there is only 1 provider. OpenRouter charges 5.5% fee.

Model price hugely variable across variables

| Open model (hosted) | Input \$/M | Output \$/M |
|--------------------------|------------|-------------|
| gpt-oss 20B | 0.05 | 0.20 |
| gpt-oss 120B | 0.15 | 0.60 |
| Llama-class 8B | 0.18 | 0.18 |
| Llama-class 70B | 0.88 | 0.88 |
| DeepSeek V4 Flash | 0.14 | 0.28 |
| Qwen 3.6 Plus | 0.50 | 3.00 |
| DeepSeek V4 Pro | 2.10 | 4.40 |

Small open models undercut frontier APIs by 10–50× per token for routine tasks.

Output is the costly half

Generated tokens cost several times more than input tokens. Terse output saves money.

Planning for a class: student keys and costs

- **Never paste keys in code.** Keys are credentials. Use Colab Secrets, and never commit them to git.
- **Set hard spend caps.** Put a budget limit on every account before the first call.
- **Start on free tiers.** Free credits cover a first lab. No card is needed to begin.
- **No student credit cards.** Provision keys for students. If they enter payment details they have high upside risk.

Class issues:

Where student data goes. Anything sent to a hosted API leaves your perimeter and may be logged. Student work and real logs can raise privacy and FERPA issues. The fix is to use synthetic data, or run an open model on your own hardware.

Models drift under you. A hosted model updates, and your assignment's expected output changes. A jailbreak you taught last week may get patched. The fix is to pin a specific open model for anything graded.

Goal for today

Where student data goes. Anything sent to a hosted API leaves your perimeter and may be logged. Student work and real logs can raise privacy and FERPA issues. The fix is to use synthetic data, or run an open model on your own hardware.

Models drift under you. A hosted model updates, and your assignment's expected output changes. A jailbreak you taught last week may get patched. The fix is to pin a specific open model for anything graded.

Survey:

Lab exercises: https://colab.research.google.com/drive/1_5hXaFWqS1hfkqGNX9dCWISTo2tC0KG1