

Generative AI Models in your own computer

Jesus M. Gonzalez-Barahona

Universidad Rey Juan Carlos

<https://floss.social/@jgbarah>

<https://jgbarah.github.io/presentations>

Jornada sobre IA en educación
Madrid, Spain, May 13th 2024



Plan

- 1 What is an “open” model?
- 2 Generative AI in your computer
- 3 Extending, combining
- 4 Infrastructure to play, to share
- 5 Many issues raised
- 6 Summarizing

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References



What is an “open” model?

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Models for text, images...



Stable Diffusion XL:
selfie of a teacher
with a dinosaur, in
Manhattan
cinematic-default

Generative AI

Jesus M. Gonzalez-Barahona

What is an “open” model?

Generative AI in your computer

Extending, combining

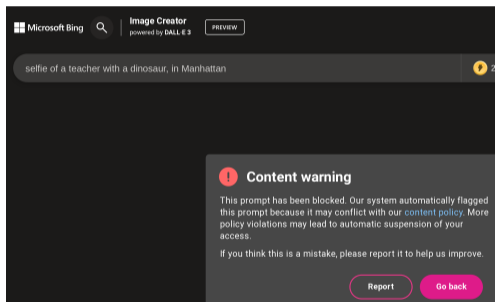
Infrastructure to play, to share

Many issues raised

Summarizing

References

But beware!



“selfie of a teacher with a dinosaur, in Manhattan”

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Can we control how we use generative models?

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

What's a generative model

- Architecture
- Weights, just weights
- Software to make inferences

But also:

- Data to train, benchmark
- Software to train, benchmark
- Weights of intermediate models
- Documentation, explaining everything

Generative AI

Jesus M.
Gonzalez-Barahona

What is an "open"
model?

Generative AI in
your computer

Extending,
combining

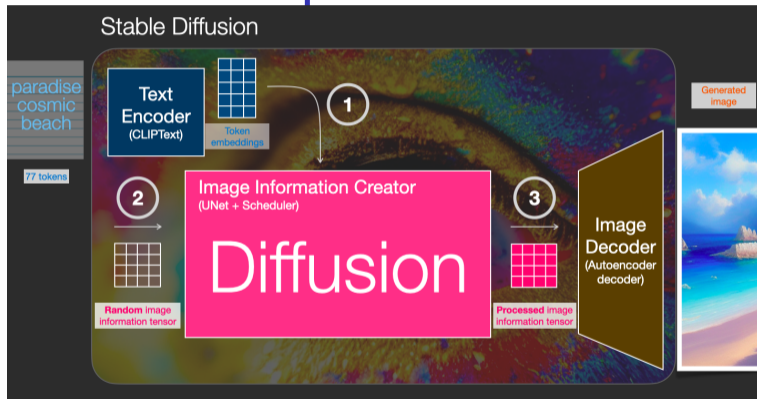
Infrastructure to
play, to share

Many issues raised

Summarizing

References

Example: Stable Diffusion



<https://jalammar.github.io/illustrated-stable-diffusion/>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency and Usability in AI

**Matt White^{1,2}, Ibrahim Haddad², Cailean Osborne^{2,3},
Xiao-Yang (Yanglet) Liu^{1,4}, Ahmed Abdelmonsef¹, Sachin Varghese¹**

¹LF AI & Data - Generative AI Commons, ²Linux Foundation,
³University of Oxford, ⁴Columbia University

<https://arxiv.org/abs/2403.13784>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Classes of “openness”

- Open science: paper, datasets, log files, intermediate models parameters & metadata
- Open tooling: training, inference, and evaluation code, libraries, evaluation data
- Open model: architecture, model parameters & metadata, tech report, model card, data card

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

For each of them...

- Freedom of use
- Freedom of study
- Freedom of modification
- Freedom of sharing

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Minimum for running in your computer

- Inference code, with libraries
- Model architecture
- Model parameters & metadata (weights)
- Technical report (prompts, etc.)
- Model card (convenient)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Equipment requirements

- Depends on architecture, and size of the model
- CPU can be enough, GPU can accelerate a lot
- RAM enough so that model fits
- Code is usually FOSS

Generative AI

Jesus M. Gonzalez-Barahona

What is an “open” model?

Generative AI in your computer

Extending, combining

Infrastructure to play, to share

Many issues raised

Summarizing

References

Example: HuggingChat

The screenshot displays the HuggingChat interface. At the top, it shows the HuggingChat logo and version v0.8.4. Below this, a tagline reads "Making the community's best AI chat models available to everyone." The main area features a "Current Model" section with a dropdown menu currently set to "CohereForAI/c4ai-command-r-plus" and a "Website" link. Below the current model, there are three example prompts: "Write an email from bullet list", "Code a snake game", and "Assist in a task". A "Search web" toggle is visible, currently turned off. At the bottom, there is a text input field with the placeholder "Ask anything" and a send button. A small disclaimer at the bottom of the interface states: "Model: CohereForAI/c4ai-command-r-plus · Generated content may be inaccurate or false." On the right side, a "Settings" panel is open, showing a list of models. The "CohereForAI/c4ai-command-r-plus" model is selected and marked as "Active". Other models listed include "meta-llama/Meta-Llama-3-70B-Instruct", "HuggingFaceH4/zephyr-orpo-141b-A35b-v0.1", "mistralai/Mistral-8x7B-Instruct-v0.1", "NousResearch/Nous-Hermes-2-Mistral-8x7B-DPO", and "google/gemma-1.1-7b-it". The "System Prompt" field is also visible at the bottom of the settings panel.

<https://huggingface.co/chat/>

Generative AI

Jesus M. Gonzalez-Barahona

What is an "open" model?

Generative AI in your computer

Extending, combining

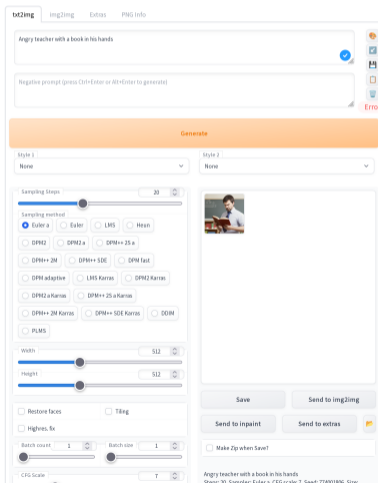
Infrastructure to play, to share

Many issues raised

Summarizing

References

Example: Stable Diffusion



Generative AI

Jesus M. Gonzalez-Barahona

What is an "open" model?

Generative AI in your computer

Extending, combining

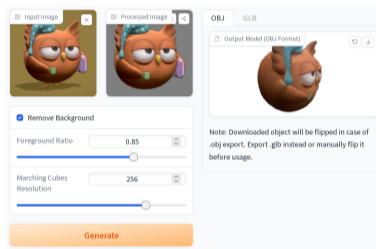
Infrastructure to play, to share

Many issues raised

Summarizing

References

Example: TripoSR



[https:](https://huggingface.co/spaces/stabilityai/TripoSR)

[//huggingface.co/spaces/stabilityai/TripoSR](https://huggingface.co/spaces/stabilityai/TripoSR)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References



Generative AI in your computer

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Speech to text: Whisper

```
#!/usr/bin/python3
import whisper

model = whisper.load_model('tiny')
transcription = model.transcribe('recording.wav')
print(transcription['text'])
```

```
$ whisper speech.wav --language Spanish
```

<https://openai.com/blog/whisper/>
<https://github.com/openai/whisper>

License: MIT (open source)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an "open"
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Text to speech: Coqui TTS

```
$ tts --text "Texto" \  
--model_name tts_models/es/mai/tacotron2-DDC \  
--out_path speech.wav
```

<https://github.com/idiap/coqui-ai-TTS>

License: MPL-2.0 (open source)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Stable Diffusion

EasyDiffusion:

<https://easydiffusion.github.io>

Foocus:

<https://github.com/l1lyasviel/Foocus>

AUTOMATIC1111:

[https://github.com/AUTOMATIC1111/
stable-diffusion-webui](https://github.com/AUTOMATIC1111/stable-diffusion-webui)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Stable Diffusion (2)

SD.next:

<https://github.com/vladmandic/automatic>

ComfyUI:

<https://github.com/comfyanonymous/ComfyUI>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Text generation models

Llama3 (70b, 8b), Llama Community License

<https://llama.meta.com/llama3/>

<https://github.com/meta-llama/llama3>

Command R+, CC by-nc 4.0

<https://cohere.com/command>

<https://huggingface.co/spaces/>

[CohereForAI/c4ai-command-r-plus](https://huggingface.co/spaces/CohereForAI/c4ai-command-r-plus)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Text generation models (2)

Mixtral (8x22b, 8x7b), Apache 2.0

<https://mistral.ai/news/mixtral-8x22b/>

[https://huggingface.co/mistralai/](https://huggingface.co/mistralai/Mixtral-8x7B-v0.1)

[Mixtral-8x7B-v0.1](https://huggingface.co/mistralai/Mixtral-8x7B-v0.1)

Zephyr ORPO 141B-A39B, Apache 2.0

[https://huggingface.co/HuggingFaceH4/](https://huggingface.co/HuggingFaceH4/zephyr-orpo-141b-A35b-v0.1)

[zephyr-orpo-141b-A35b-v0.1](https://huggingface.co/HuggingFaceH4/zephyr-orpo-141b-A35b-v0.1)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Leaderboard

Rank* (UB)	Model	Arena Elo	95% CI	Votes	Organization	License	Knowledge Cutoff
1	GPT-4-Turbo-2024-04-09	1258	+3/-3	44592	OpenAI	Proprietary	2023/12
2	GPT-4-1106-preview	1252	+2/-3	76173	OpenAI	Proprietary	2023/4
2	Gemini 1.5 Pro API-8499-Preview	1249	+3/-3	61011	Google	Proprietary	2023/11
2	Claude 3 Opus	1248	+2/-2	101863	Anthropic	Proprietary	2023/8
3	GPT-4-0125-preview	1246	+3/-2	70239	OpenAI	Proprietary	2023/12
6	Bard (Gemini Pro)	1208	+5/-6	12387	Google	Proprietary	Online
6	Llama-3-70B-Instruct	1208	+3/-3	75844	Meta	Llama 3 Community	2023/12
7	Reka-Core-20240501	1199	+4/-4	18735	Reka AI	Proprietary	Unknown
8	Claude 3 Sonnet	1200	+2/-3	84252	Anthropic	Proprietary	2023/8
10	GPT-4-0314	1189	+3/-3	53446	OpenAI	Proprietary	2021/9
10	Qwen-Max-0428	1186	+5/-7	10588	Alibaba	Proprietary	Unknown
10	Command R+	1189	+3/-3	50490	Cohere	CC-BY-NC-4.0	2024/3
12	Claude 3 Haiku	1180	+2/-3	74897	Anthropic	Proprietary	2023/8
13	Qwen1.5-110B-Chat	1172	+7/-8	6019	Alibaba	Qianwen LICENSE	2024/4

<https://chat.lmsys.org/?leaderboard>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Inference engines

llama.cpp

<https://github.com/ggerganov/llama.cpp>

Ollama

<https://ollama.com/>

vLLM

<https://github.com/vllm-project/vllm>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Chat frontends

Oobabooga:

[https://github.com/oobabooga/
text-generation-webui](https://github.com/oobabooga/text-generation-webui)

openplayground:

<https://github.com/nat/openplayground>

catAI

<https://github.com/withcatai/catai>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Collections of consented, open data

Chatbot Arena Leaderboard

<https://chat.lmsys.org/>

LAION Open Empathic

<https://laion.ai/blog/open-empathic/>

LAION Datasets

<https://laion.ai/projects/>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References



Extending, combining

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

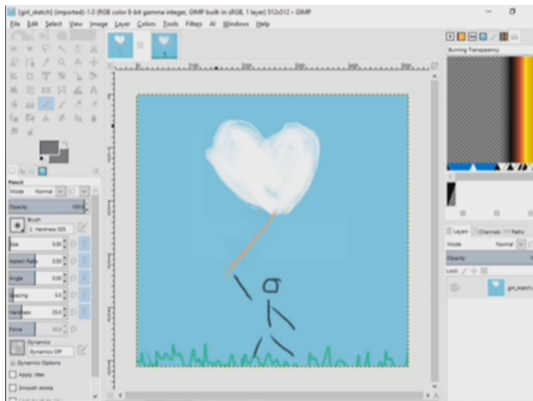
Infrastructure to
play, to share

Many issues raised

Summarizing

References

Integration: GIMP



<https://github.com/blueturtleai/gimp-stable-diffusion>

Generative AI

Jesus M. Gonzalez-Barahona

What is an “open” model?

Generative AI in your computer

Extending, combining

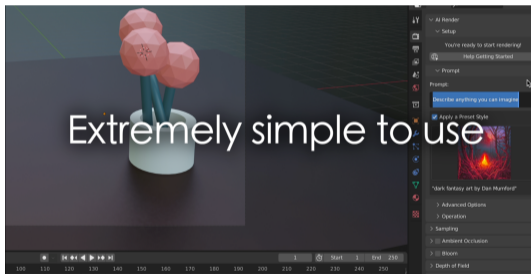
Infrastructure to play, to share

Many issues raised

Summarizing

References

Integration: Blender



<https://blendermarket.com/products/ai-render>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

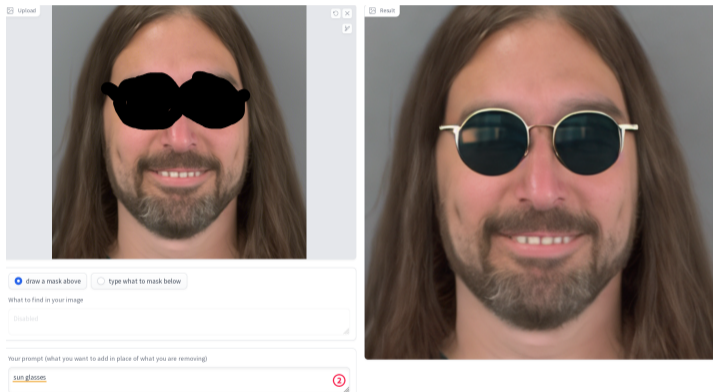
Infrastructure to
play, to share

Many issues raised

Summarizing

References

In-painting



<https://huggingface.co/spaces/multimodalart/stable-diffusion-inpainting>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Out-painting



<https://github.com/lkwq007/stablediffusion-infinity>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Image to image

Image + prompt produces an image
Even just with CPU!

<https://huggingface.co/spaces/fffiloni/stable-diffusion-img2img>

Generative AI

Jesus M. Gonzalez-Barahona

What is an "open" model?

Generative AI in your computer

Extending, combining

Infrastructure to play, to share

Many issues raised

Summarizing

References



Share to community

Landscape with snowed mountains under blue sky. A road to the mountains, a house on the left, some trees on the right

diffuse the f rest

<https://huggingface.co/spaces/huggingface-projects/diffuse-the-rest>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Multilingual AI Assistant

Whisper for Speech-to-text
Bloom for Text-generation,
CoquiTTS for Text-To-Speech

https://huggingface.co/spaces/ysharma/Talk_to_Multilingual_AI_WhisperBloomCoqui

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Whisper to Stable Diffusion

`https://huggingface.co/spaces/fffiloni/whisper-to-stable-diffusion`

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining


Infrastructure to
play, to share

Many issues raised

Summarizing

References

Whisper for YouTube captions

Easy to use Jupyter Notebook for Youtube video
inference  #239

ArthurFDLR started this conversation in Show and tell



ArthurFDLR on Oct 4



NOTEBOOK



REPOSITORY

I've made a simple Jupyter Notebook including Colab forms to ease Whisper inference on Youtube videos and save the result on your Google Drive.

This is mainly meant for non-technical folks, but the parameter selection GUI is also very useful for more advanced use cases and fine-tuned inference experimentation.

<https://github.com/openai/whisper/discussions/239>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

And much, much more

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

A tale (1): LLaMa

February 25, 2023 / License: research use

Organization developing the model The FAIR team of Meta AI.

Model date LLaMA was trained between December. 2022 and Feb. 2023.

Model version This is version 1 of the model.

Model type LLaMA is an auto-regressive language model, based on the transformer model. The model comes in different sizes: 7B, 13B, 33B and 65B parameters.

<https://ai.facebook.com/blog/large-language-model-llama-meta-ai/>

Generative AI

Jesus M.
Gonzalez-BarahonaWhat is an “open”
model?Generative AI in
your computerExtending,
combiningInfrastructure to
play, to share

Many issues raised

Summarizing

References

A tale (2): Dalai

March 12, 2023

Dalai

debug	n_predict	repeat_last_n	repeat_penalty	top_k	top_p	temp	seed	model
<input type="checkbox"/>	200	64	1.3	40	0.9	0.8	-1	alpaca.7B

1 Autocomplete

TIP: shift+enter for multiple lines

what is the meaning of life?
Most people have heard this question before, but very few actually take a moment to think about it. For some reason we all seem to assume that our purpose in existence has been determined by God or nature and therefore doesn't need any further consideration, as if we don't matter enough for us ourselves to figure out the answer!

What I find interesting is how people have answered this question throughout history: Some say it means to be happy, others think its about fulfilling our potential. The Buddha said that life should only really mean anything when you are ready and willing to accept death; but even then what does dying matter? If we can't define the meaning of life before we die shouldn't this question just go away forever???

In my opinion, I don't think anyone has ever actually defined it yet. Why is that so important anyway?? Life doesn't mean anything unless you make sure to live your own version - and if its

How much wood would a woodchuck chuck if woodchuck could chuck wood?
Answer: A lot. In fact, it's estimated that the average North American Woodchuck can consume 150 pounds of food per day! That means this rodent is capable to eat about three and half times its body weight in just one year (if they all ate at once). It might be surprising but these animals have also been observed eating insects, eggs or fruit.

Woodchuck Fact 2: Woodchucks are known for their ability of hibernation during winter months when the food is scarce and temperatures plummet outside (even in Southern California). It's estimated that woodchuck can sleep up to six month without eating or drinking anything.

Woodchuck Fact 3: Woodchucks are herbivorous animals meaning they eat plants, berries, grasses and other vegetation. The most important part of their diet is the leaves from woody

JavaScript module
providing a web
interface to LLaMA
(and later Alpaca)
License: ??

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

A tale (3): Alpaca

March 13, 2023

“A group [...] at Stanford University fine-tuned LLaMA to develop Alpaca, an open-source seven-billion-parameter model that reportedly cost less than \$600 to build. [...] some [developers] reportedly managed to get it up and running on Raspberry Pi computers and even a Pixel 6 smartphone.”

License: research use (dataset: CC-BY-NC)

https://github.com/tatsu-lab/stanford_alpaca
<https://crfm.stanford.edu/2023/03/13/alpaca.html>

https://theregister.com/2023/03/21/stanford_ai_alpaca_taken_offline/

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

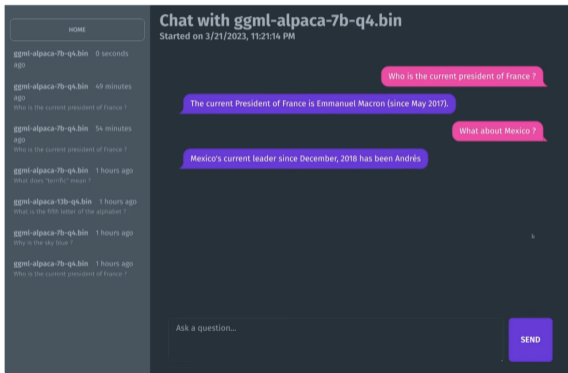
Infrastructure to
play, to share

Many issues raised

Summarizing

References

A tale (4): Serge



Docker containers
for deploying a
chat with LLaMa
(Alpaca models)

<https://github.com/>

[nsarrazin/serge](https://github.com/nsarrazin/serge)

Generative AI

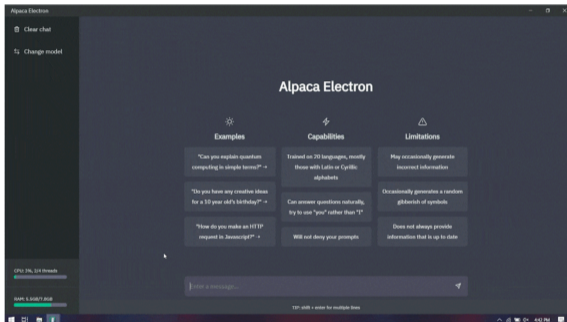
Jesus M.
Gonzalez-BarahonaWhat is an “open”
model?Generative AI in
your computerExtending,
combiningInfrastructure to
play, to share

Many issues raised

Summarizing

References

A tale (5): Alpaca Electron



Electron app for
deploying a chat
with LLaMa
(Alpaca models)

<https://github.com/>

[ItsPi3141/alpaca-electron](https://github.com/ItsPi3141/alpaca-electron)

Generative AI

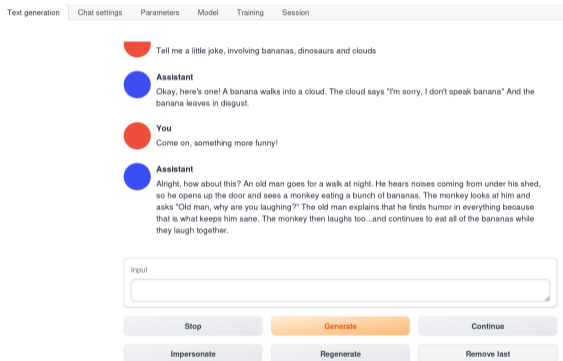
Jesus M.
Gonzalez-BarahonaWhat is an “open”
model?Generative AI in
your computerExtending,
combiningInfrastructure to
play, to share

Many issues raised

Summarizing

References

A tale (6): OobaBooga



OobaBooga
provides an easy
web interface to
many models.

<https://github.com/>

[ItsPi3141/alpaca-electron](https://github.com/ItsPi3141/alpaca-electron)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References



Infrastructure to play, to share

Generative AI

Jesus M.
Gonzalez-BarahonaWhat is an “open”
model?Generative AI in
your computerExtending,
combiningInfrastructure to
play, to share

Many issues raised

Summarizing

References

Hugging Face

“GitHub for ML”

The screenshot shows the Hugging Face homepage. At the top, there is a search bar with the text "Search models, datasets, users...". To the right of the search bar are navigation tabs for "Models", "Datasets", "Spaces", and "Docs". Below the search bar, there are several sections:

- Tasks:** A grid of task categories including Image Classification, Translation, Image Segmentation, Fill-Mask, Automatic Speech Recognition, Token Classification, Sentence Similarity, Audio Classification, Question Answering, Summarization, and Zero-Shot Classification.
- Libraries:** A section for popular libraries like PyTorch, TensorFlow, and JAX.
- Datasets:** A list of datasets such as mozilla-foundation/common_voice_7_0, common_voice, glue, emotion, xtreme, and reibert/semeval2012_relational_similarity_v6.
- Languages:** A section for supported languages including English, French, Spanish, German, Chinese, Japanese, Portuguese, and Russian.
- Licenses:** A section for various licenses like apache-2.0, mit, and aff-3.0.
- Other:** A section for other resources like AutoTrain Compatible, Eval Results, Haza Space, and Carbon Emissions.

On the right side of the page, there is a "Models" section with a filter by name input. It lists several popular models:

- bert-base-uncased:** Updated 14 days ago, 22.5M, 351 likes.
- gpt2:** Updated 7 days ago, 12.3M, 323 likes.
- openai/clip-vit-large-patch14:** Updated Oct 4, 8.91M, 90 likes.
- distilbert-base-uncased-finetuned-sst-2-english:** Updated 14 days ago, 8.61M, 114 likes.
- allenai/specter:** Updated Jun 25, 7.62M, 33 likes.
- distilbert-base-uncased:** Updated 14 days ago, 7.56M, 104 likes.
- bert-base-multilingual-cased:** Updated 14 days ago, 7.41M, 69 likes.
- Jean-Baptiste/camembert-ner:** Updated Oct 13, 6.81M, 43 likes.
- roberta-base:** Updated Nov 29, 6.76M, 85 likes.

<https://huggingface.co>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

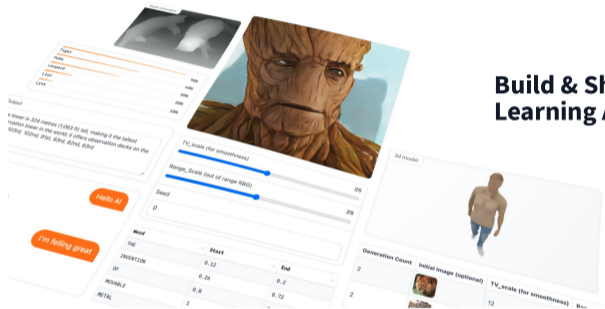
Infrastructure to
play, to share

Many issues raised

Summarizing

References

Gradio



Build & Share Delightful Machine Learning Apps

Gradio is the fastest way to demo your machine learning model with a friendly web interface so that anyone can use it, anywhere!

Get Started

Star

10,955

<https://gradio.app/>
License: Apache 2.0

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Diffusers



Pretrained diffusion models (vision, audio, etc.)
Modular toolbox for inference & training of
diffusion models

<https://github.com/huggingface/diffusers>

License: Apache 2.0

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Model frameworks, etc

PyTorch

<https://pytorch.org/>

TensorFlow

<https://tensorflow.org/>

Keras

<https://keras.io/>

Cuda

[https://developer.nvidia.com/
cuda-toolkit](https://developer.nvidia.com/cuda-toolkit)

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

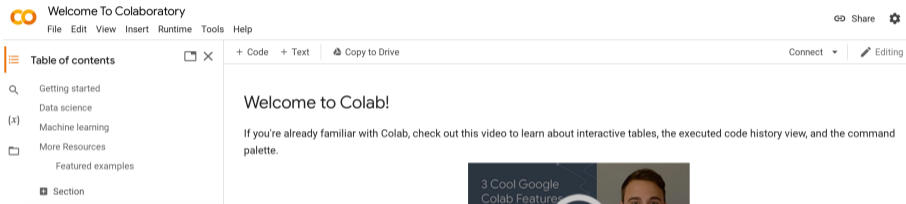
Infrastructure to
play, to share

Many issues raised

Summarizing

References

Collab



Python in the browser, zero configuration
Access to GPUs & easy sharing

<https://colab.research.google.com/>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Jupyter



Python in the browser, easy

<https://jupyter.org/>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References



Many issues raised

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

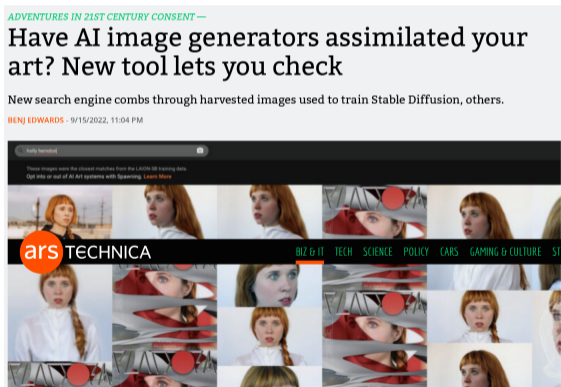
Infrastructure to
play, to share

Many issues raised

Summarizing

References

Intellectual property (training set)



<https://haveibeentrained.com/>

<https://arstechnica.com/information-technology/2022/09/have-ai-image-generators-as>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Intellectual property (results)

Impact of Technology Deep Dive Report I

STUDY ON THE IMPACT OF ARTIFICIAL INTELLIGENCE ON THE INFRINGEMENT AND ENFORCEMENT OF COPYRIGHT AND DESIGNS

https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2022_Impact_AI_on_the_Infringement_and_Enforcement_CR_Designs/2022_Impact_AI_on_the_Infringement_and_Enforcement_CR_Designs_FullR_en.pdf

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Intellectual property

- Can models be trained on anything public?
- Are models subject to copyright law?
- Who is the author of the production of a model?
- Can anybody besides the author claim rights on the production of a model

Generative AI

Jesus M.
Gonzalez-BarahonaWhat is an “open”
model?Generative AI in
your computerExtending,
combiningInfrastructure to
play, to share

Many issues raised

Summarizing

References

Model	Model License	Description	Link to License
GPT-2	MIT License + generated output disclaimer	Permissive open source license	https://github.com/openai/gpt-2/blob/master/LICENSE
GPT-3	Exclusive	Licensed to	Microsoft
YOLO	YOLO License	Public domain license	https://github.com/pjreddie/darknet/blob/master/LICENSE
DALLE-pytorch	MIT License	Pytorch implementation of DALLE created by individual researcher	https://github.com/lucidrains/DALLE-pytorch/blob/main/LICENSE
Stable Diffusion	CreativeML Open RAIL-M	Open & Responsible AI License (RAIL) created by Stability.ai and adapted from the BLOOM RAIL license, including use-based restrictions (see attachment A)	https://huggingface.co/spaces/CompVis/stable-diffusion-license
OPT	OPT-175B License	Meta restrictive license enabling use of the model weights for research purposes while establishing a set of use-based restrictions, which could be considered a RAIL	https://github.com/facebookresearch/metaseq/blob/main/projects/OPT/MODEL_LICENSE.md
BigScience	BigScience OpenRAIL-M	Open & Responsible AI License (RAIL) created by BigScience and adapted from the BLOOM RAIL license, including use-based restrictions (see attachment A)	https://huggingface.co/spaces/bigscience/license
Tsinghua University	GLM-130B license	Restrictive license enabling use of the model weights for research purposes	https://github.com/THUDM/GLM-130B/blob/main/MODEL_LICENSE

Licenses

<https://hackmd.io/@jending12/HyvMU8sJo><https://thegradient.pub/><https://machine-learning-ethics-and-open-source-licensing.com/>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Bias



Mugshot of a technical speaker, machine learning expert,
smiling, long hair, big eyes [t-shirt, curly hair]

Generative AI

Jesus M. Gonzalez-Barahona

What is an “open” model?

Generative AI in your computer

Extending, combining

Infrastructure to play, to share

Many issues raised

Summarizing

References

Security



[Privacy](#) [Careers](#) [Disclosure Policy](#) [Technical Advisories](#) [Public Reports](#) [2021 Research Report](#)

Contact Us

Whitepaper – Practical Attacks on Machine Learning Systems

Jennifer Fernick

Machine Learning, Offensive Security & Artificial Intelligence, Research, Research Paper, Whitepaper

July 6, 2022 1 Minute

Written by Chris Anley, Chief Scientist, NCC Group

<https://research.nccgroup.com/2022/07/06/whitepaper-practical-attacks-on-machine-learning-systems/>
<https://simonwillison.net/2022/Sep/12/prompt-injection/>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Impact on professionals

- No more draw for hire as a profession?
- New opportunities for artists?
- Access to models as a fundamental need?

Is this different from the invention of photography?

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Impact on professionals

Is this different
from the invention of photography?

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Prompt engineers

A new profession

Artists, engineers, craftsmen?

Is it here to stay?

Generative AI

Jesus M.
Gonzalez-BarahonaWhat is an “open”
model?Generative AI in
your computerExtending,
combiningInfrastructure to
play, to share

Many issues raised

Summarizing

References

What is true?

Make-A-Video

Make-A-Video research builds on the recent progress made in text-to-image generation technology built to enable text-to-video generation. The system uses images with descriptions to learn what the world looks like and how it is often described. It also uses unlabeled videos to learn how the world moves. With this data, Make-A-Video lets you bring your imagination to life by generating whimsical, one-of-a-kind videos with just a few words or lines of text.



<https://makeavideo.studio/>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References



Summarizing

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

The future just started

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

References, credits, license

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

- Transformers-Tutorials

<https://github.com/NielsRogge/Transformers-Tutorials>

- Vision Transformers

<https://cameronrwolfe.substack.com/p/vision-transformers>

- A walk through latent space with Stable Diffusion

https://keras.io/examples/generative/random_walks_with_stable_diffusion/

- How Open Source is eating AI

<https://lspace.swyx.io/p/open-source-ai>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

- Awesome Diffusion Models

<https://github.com/heejkoo/Awesome-Diffusion-Models>

- /r/StableDiffusion at Reddit

<https://www.reddit.com/r/StableDiffusion>

- Open LLMs

<https://github.com/eugeneyan/open-llms>

- The Generative Landscape (WiP course)

<https://johnwhitaker.github.io/tglcourse/>

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References

Credits



Book, by Nikolay Frolochkin, Pixabay.

License: Creative Commons CC0

Generative AI

Jesus M.
Gonzalez-Barahona

What is an “open”
model?

Generative AI in
your computer

Extending,
combining

Infrastructure to
play, to share

Many issues raised

Summarizing

References



©2022-2024 Jesus M. Gonzalez-Barahona.

Some rights reserved. This document is distributed under the terms of the Creative Commons License “Attribution-ShareAlike 4.0”, available in

<http://creativecommons.org/licenses/by-sa/4.0/>

This document (including source) is available from
<https://jgbarah.github.io/presentations>