

UGIDOTNET SMALL TALKS



GENERATIVE AI: A LOOK INTO OPEN-SOURCE MODELS AND TOOLKITS

GIANNI ROSA GALLINA

R&D TECHNICAL LEAD @ DELTATRE
MICROSOFT MVP

GIOVEDÌ, 4 MAGGIO 2023, 17:00 CET

YOUTUBE - TWITCH - LINKEDIN

Generative AI... for all and everything



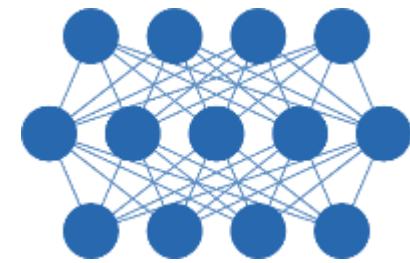
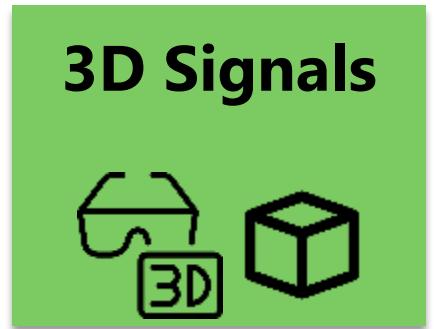
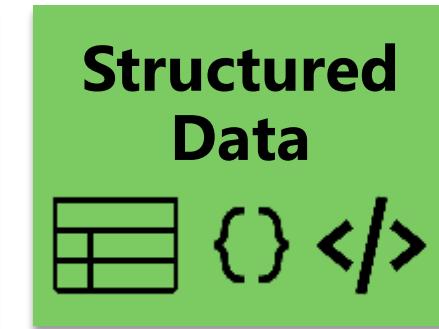
- Arts & Photography
- Design
- Fashion
- Writing
- Sounds & Music
- Gaming
- Architecture
- Marketing
- Customer Support
- Advertising
- Programming
- Scientific Research
- Cinema

...

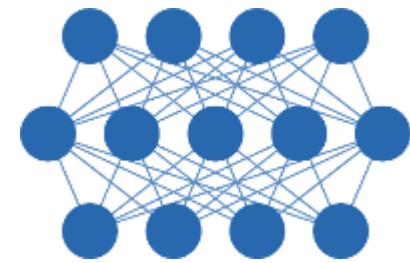


Generative AI

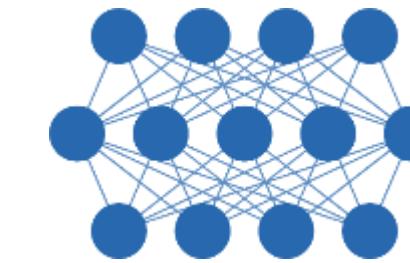
Overview



...



...

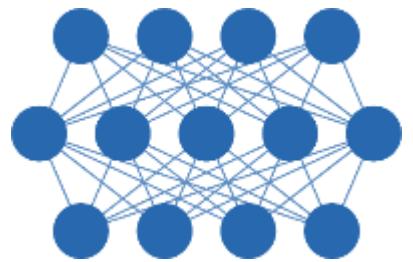


Foundation Models

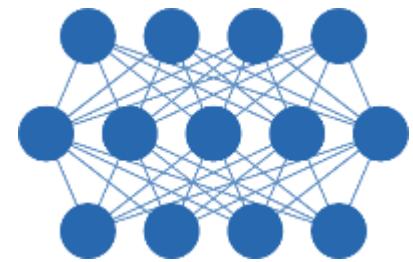


Generative AI

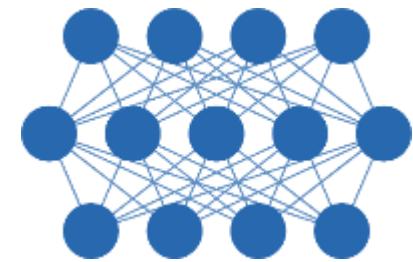
Overview



...



...



Foundation Models



Question
Answering

Information
Extraction

Image
Captioning

Multimodal
Translation

Text-To-X

Img-To-X

Generative AI

A little bit of history



AI Origins	Neural Networks	Deep Learning boom	Transformers everywhere	Breakthroughs	Generative AI for all & everything
Foundational Research (Logic, Math, Stats, IT)		Audio, Vision, 3D RNNs CNNs GANs RL	Audio Text Images Video 3D	GPT-1, GPT-2, GPT-3 Dall-E / Dall-E 2 Copilot Imagen DreamFusion ChatGPT MidJourney NeRF ...	Images Text Video Speech Music 3D Programming ...

Generative AI Images

"Milan City skyline landscape in van gogh style, 16K resolution, DeviantArt, Flickr, rendered in Enscape, Miyazaki, Nausicaa Ghibli, detailed post processing, atmospheric, hyper realistic, 8k, epic composition, cinematic, artstation"



<https://www.bing.com/images/create/>



Image Creator
powered by DALL-E

PREVIEW

Generative AI

Images



"star wars animals"

<https://midjourney.com/>

Generative AI

Text

The screenshot shows the ChatGPT interface. On the left, there's a sidebar with 'Dark mode', 'OpenAI Discord', 'Updates & FAQ', and 'Log out'. The main area has three tabs: 'Examples' (with rows for quantum computing, creative ideas, and HTTP requests), 'Capabilities' (with rows for remembering past messages, providing follow-up corrections, and declining inappropriate requests), and 'Limitations' (with rows for occasional errors in facts, generating harmful instructions, and limited knowledge after 2022). At the bottom, it says 'DIVERSE AI, 2023. Works. Free Research Preview. Our goal is to make AI systems more natural and easy to interact with. Your feedback is important to us.'

<https://chat.openai.com/>

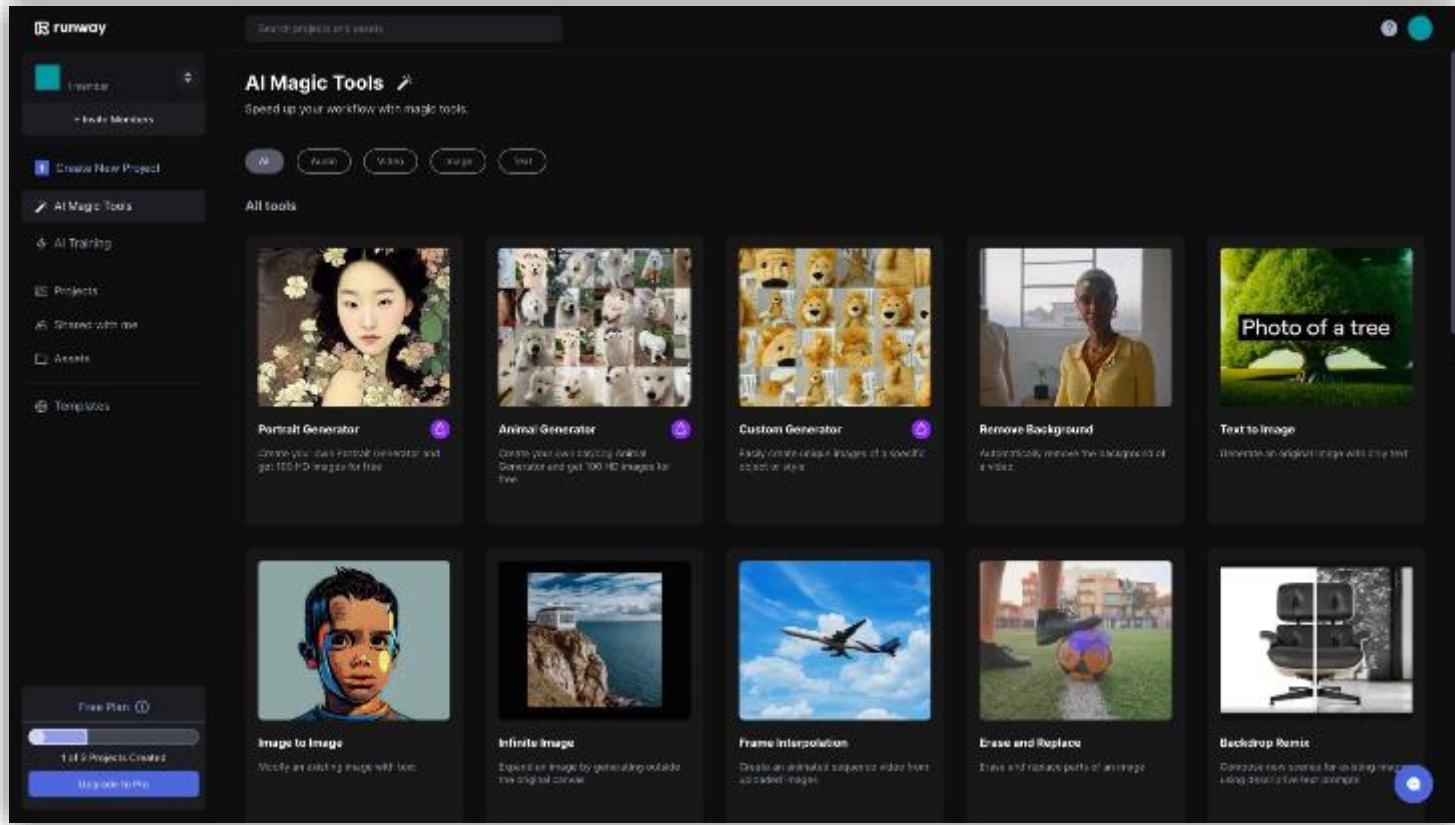


The screenshot shows the Azure OpenAI Studio 'Chat playground (Preview)' interface. It includes sections for 'Assistant setup' (with a dropdown menu for 'Test example 1'), 'Chat session' (with a message from 'TuringBot' about Generative AI), and 'Parameters' (with sliders for 'Max response', 'Temperature', 'Top P', 'Stop sequence', 'Stop sequence', and 'Session settings'). The 'Session settings' section includes a 'Past messages included' slider and a 'Current token count' indicator (200/512) with a 'Reset token progress indicator' button.

<https://azure.microsoft.com/en-us/products/cognitive-services/openai-service>

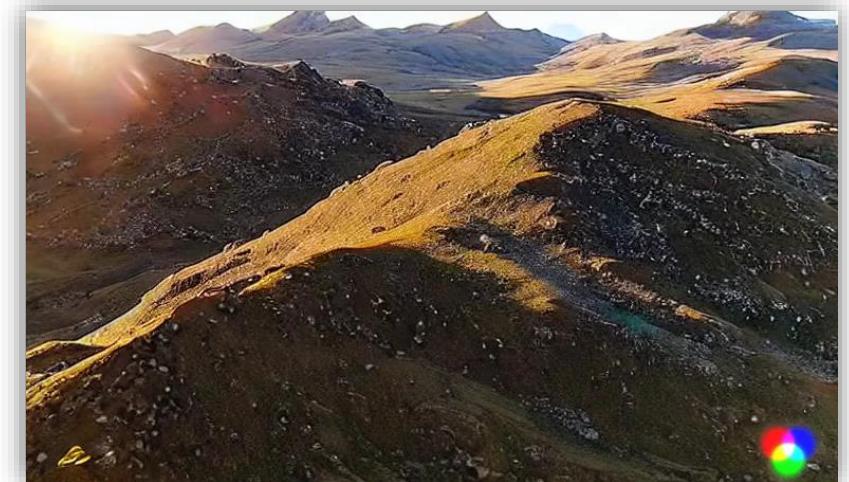
Generative AI

Video



<https://runwayml.com/>

"Aerial drone
footage of a
mountain range."



<https://research.runwayml.com/gen2>

Generative AI

Audio (Speech-To-Text)



Generative AI

Audio (Text-To-Speech)

ElevenLabs – Voice Dubbing demo



https://www.youtube.com/watch?v=17_xLsqny9E

Generative AI

Audio

ElevenLabs – Prime Voice AI

The screenshot shows the ElevenLabs AI voice generator interface. At the top, there's a text input field with placeholder text "Try entering any text or give me an idea" and a button labeled "give me an idea". Below the input field, it says "works in" followed by language options: English, German, Polish, Spanish, Italian, French, Portuguese, and Hindi. The "Portuguese" and "Hindi" buttons are highlighted in blue. In the center, there's a text area containing the Italian sentence: "In questa sessione, stiamo parlando di IA Generativa. Questa frase è letta da un modello di voce artificiale di ElevenLabs!". Below this text, there's a dropdown menu showing "premade/Bella" and a page number "124 / 333". At the bottom, there's a large play button icon, a progress bar, and download/cancel icons.

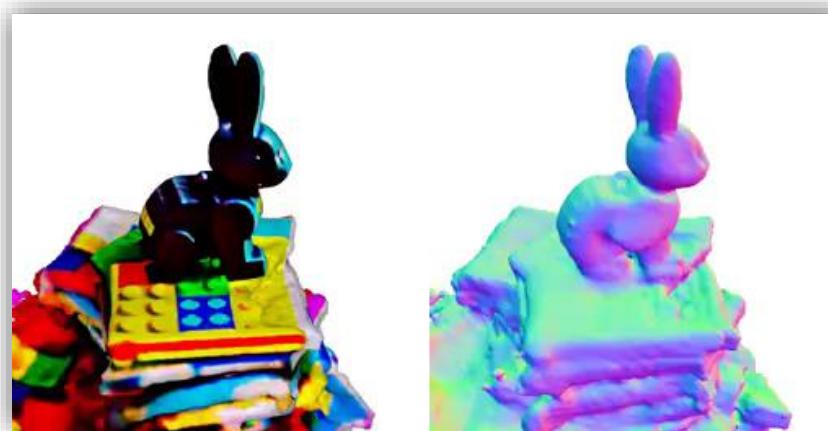
<https://beta.elevenlabs.io/>

Generative AI

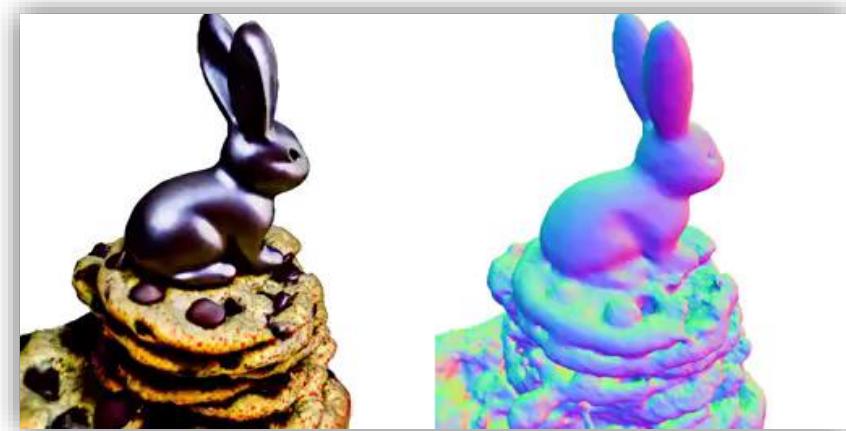
Text to 3D



"A baby bunny sitting on top of a stack of pancakes."



"A Lego bunny sitting on top of a stack of books."



"A metal bunny sitting on top of a stack of chocolate cookies."

Generative AI

2D to 3D generation



<https://developer.nvidia.com/blog/getting-started-with-nvidia-instant-nerfs/>

<https://jonbarron.info/zipnerf/>

Alternative Generative AI Models

GitHub - awesome-generative-ai

A curated list of modern Generative Artificial Intelligence projects and services

<https://github.com/steven2358/awesome-generative-ai>

GitHub - awesome-decentralized-lm

Collection of LLM resources that can be used to build products you can "own" or to perform reproducible research

<https://github.com/imaurer/awesome-decentralized-lm>



Hugging Face

stability.ai

LAION The LAION logo consists of the word "LAION" in white capital letters on a dark blue background, followed by a white silhouette of a paw print.

A screenshot of a GitHub repository page titled "awesome-decentralized-lm". The README.md file contains the following text:

awesome-decentralized-lm

Collection of LLM resources that can be used to build products you can "own" or to perform reproducible research.

Awesome Generative AI

A curated list of modern Generative Artificial Intelligence projects and services.

Generative Artificial Intelligence is a technology that creates original content such as images, sounds, and texts by using machine learning algorithms that are trained on large amounts of data. Unlike other forms of AI, it is capable of creating unique and unexpected outputs such as poems, images, digital art, music, and more. These outputs often have their own unique style and can even be hard to distinguish from human-created ones. Generative AI has a wide range of applications in fields such as art, entertainment, marketing, academia, and computer science.

Contributions to this list are welcome. Add links through pull requests or create an issue to start a discussion.

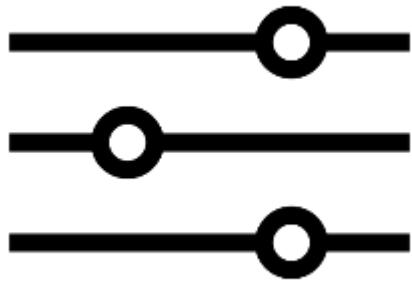
Contents

- Recommender reading
- Text
- Code
- Image
- Video
- Audio
- Other
- More info
- Autonomous agents

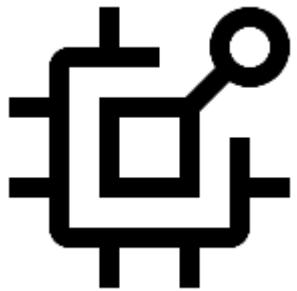
Recommended reading

The repository page also shows a sidebar with navigation links like "Issues", "Pull requests", "Commits", and "Contributors", along with a list of recent activity.

Why alternative & open-source models?



Customization
& Flexibility



Embedded/Mobile
Devices



Data Policies &
Ownership



No/Limited
Connectivity



Savings &
Optimization

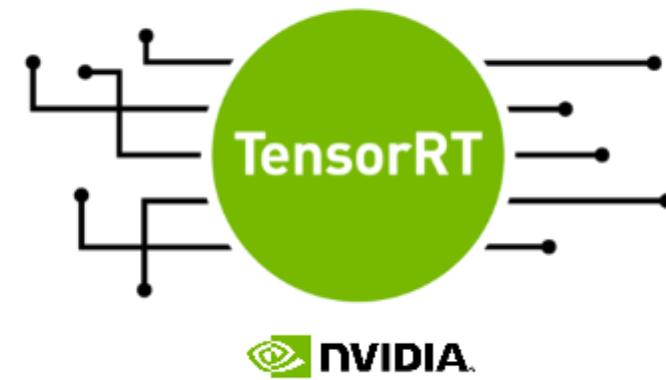
Tools and Frameworks



Python



VS Code

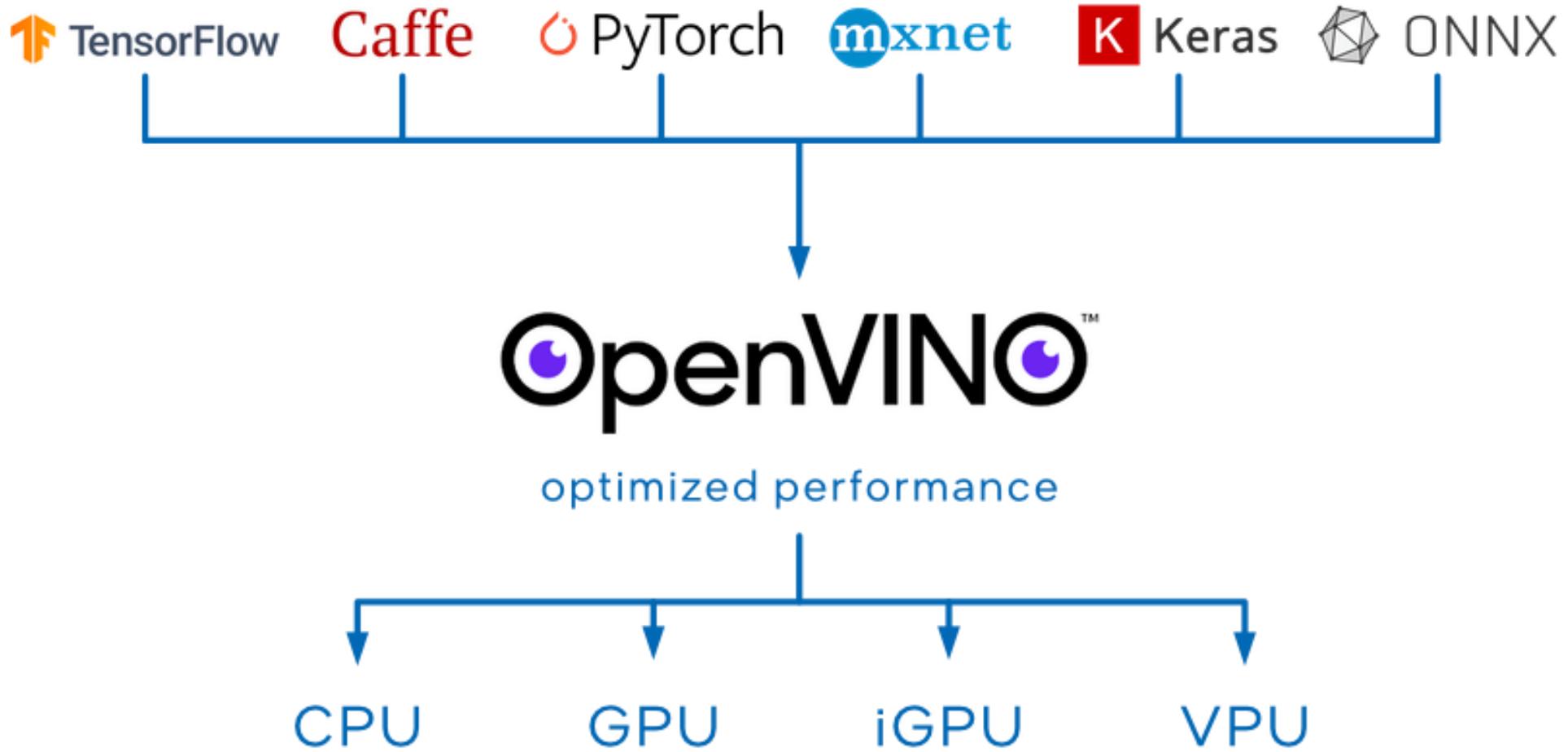


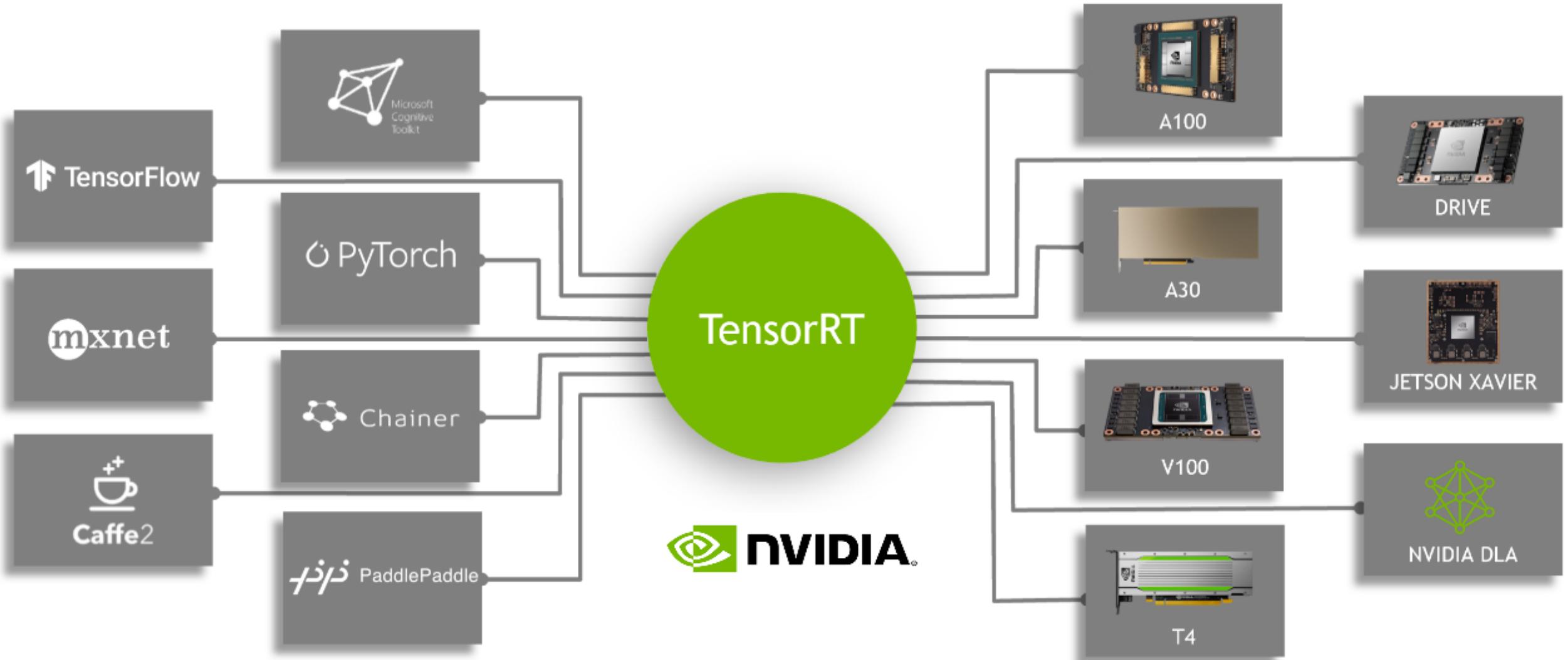
PyTorch

scikit
learn

TensorFlow



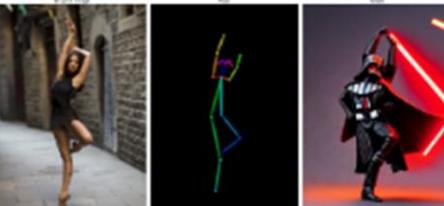




DEMO



OpenVINO™ Notebooks

Description	Preview	Complementary Materials
Optimize YOLOv8 using NNCF PTQ API		Blog - How to get YOLOv8 Over 1000 fps with Intel GPUs?
Prompt based object segmentation mask generation using Segment Anything and OpenVINO™		Blog - SAM: Segment Anything Model — Versatile by itself and Faster by OpenVINO
A Text-to-Image Generation with ControlNet Conditioning and OpenVINO™		Blog - Control your Stable Diffusion Model with ControlNet and OpenVINO

Text-to-Image Generation with Stable Diffusion v2 and OpenVINO™

Stable Diffusion v2 is the next generation of Stable Diffusion model a Text-to-Image latent diffusion model created by the researchers and engineers from Stability AI and LAION.

General diffusion models are machine learning systems that are trained to denoise random gaussian noise step by step, to get to a sample of interest, such as an image. Diffusion models have shown to achieve state-of-the-art results for generating image data. But one downside of diffusion models is that the reverse denoising process is slow. In addition, these models consume a lot of memory because they operate in pixel space, which becomes unreasonably expensive when generating high-resolution images. Therefore, it is challenging to train these models and also use them for inference. OpenVINO brings capabilities to run model inference on Intel hardware and opens the door to the fantastic world of diffusion models for everyone!

In previous notebooks, we already discussed how to run Text-to-Image generation and Image-to-Image generation using Stable Diffusion v1 and controlling its generation process using ControlNet. Now is turn of Stable Diffusion v2.

Stable Diffusion v2: What's new?

The new stable diffusion model offers a bunch of new features inspired by the other models that have emerged since the introduction of the first iteration. Some of the features that can be found in the new model are:

- The model comes with a new robust encoder, OpenCLIP, created by LAION and aided by Stability AI; this version v2 significantly enhances the produced photos over the V1 versions.
- The model can now generate images in a 768x768 resolution, offering more information to be shown in the generated images.

Stable Diffusion

Images

stability.ai

 **runway**

LAION 



Hugging Face



Dffusers



https://huggingface.co/blog/stable_diffusion

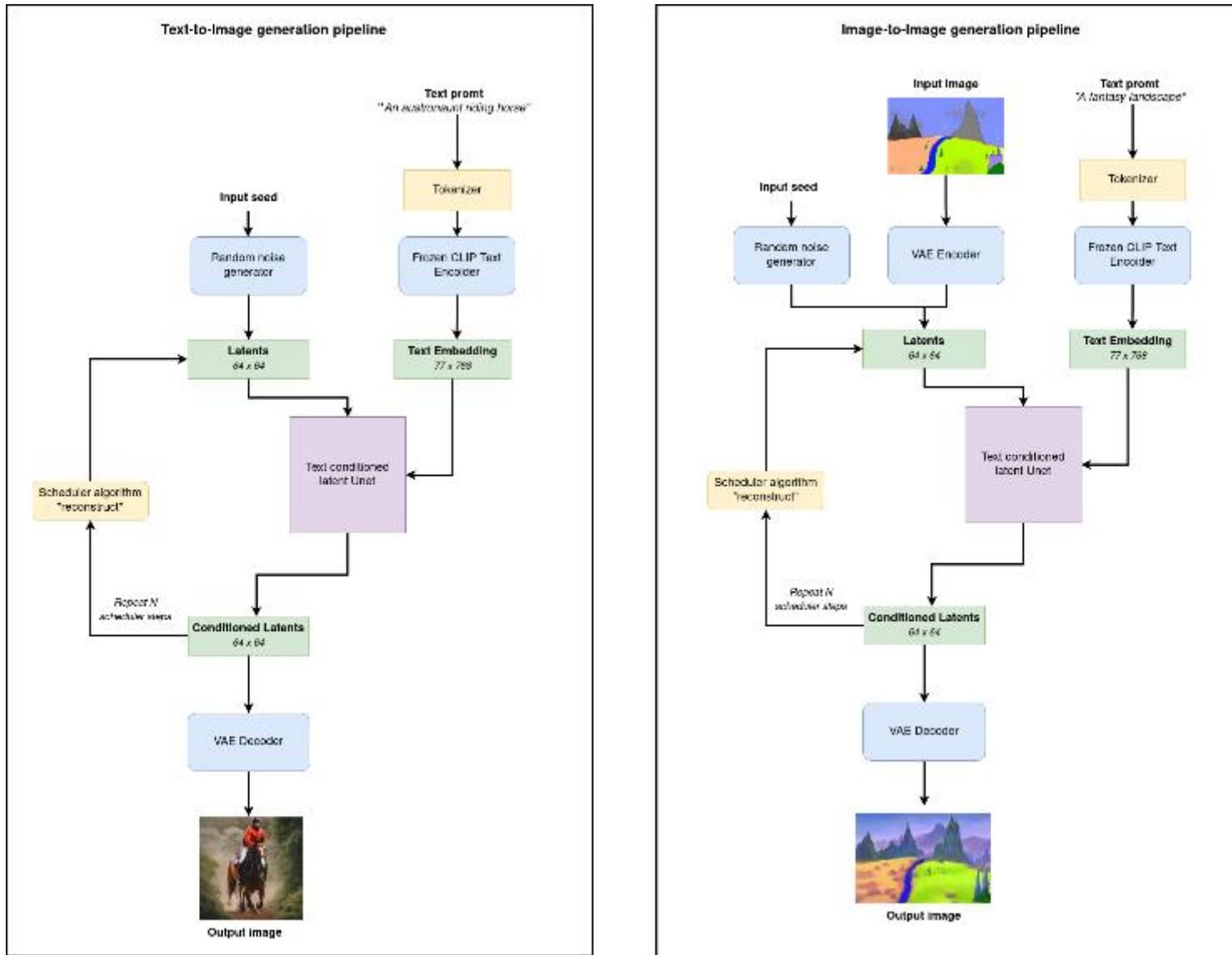
<https://huggingface.co/blog/annotated-diffusion>

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

<https://github.com/runwayml/stable-diffusion>

Stable Diffusion

Images

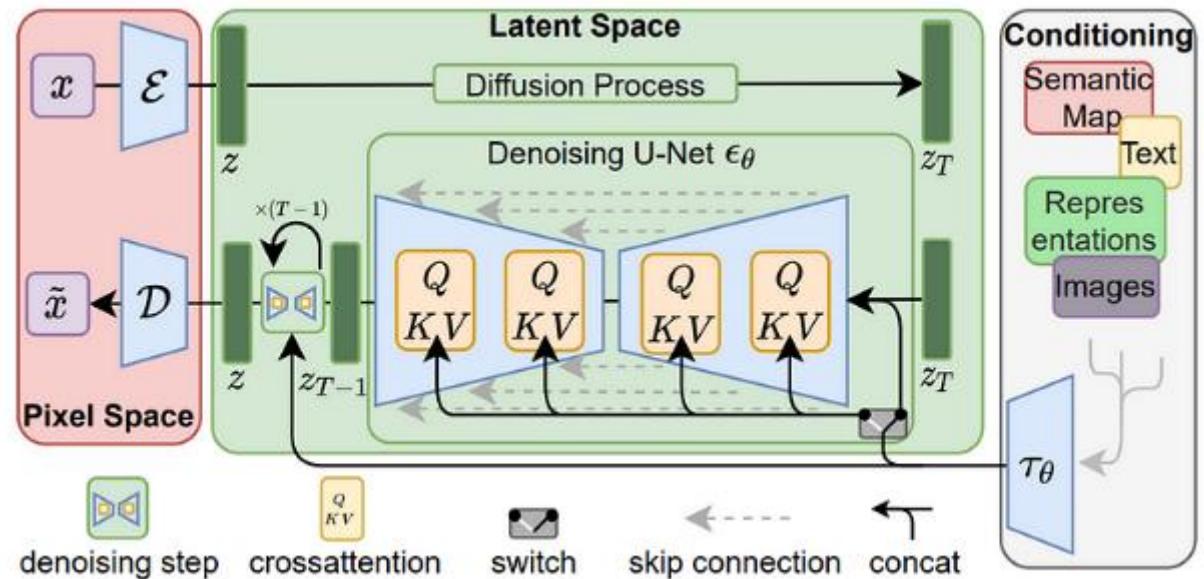


Source: https://github.com/openvinotoolkit/openvino_notebooks (225_stable-diffusion)

Stable Diffusion Images

High-Resolution Image Synthesis with
Latent Diffusion Models

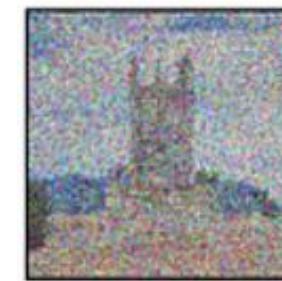
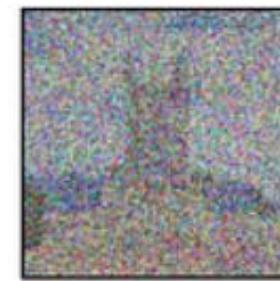
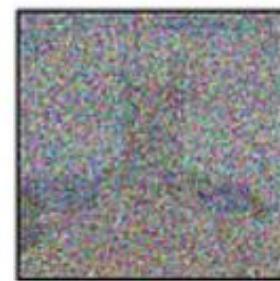
<https://arxiv.org/abs/2112.10752>



steps



Input

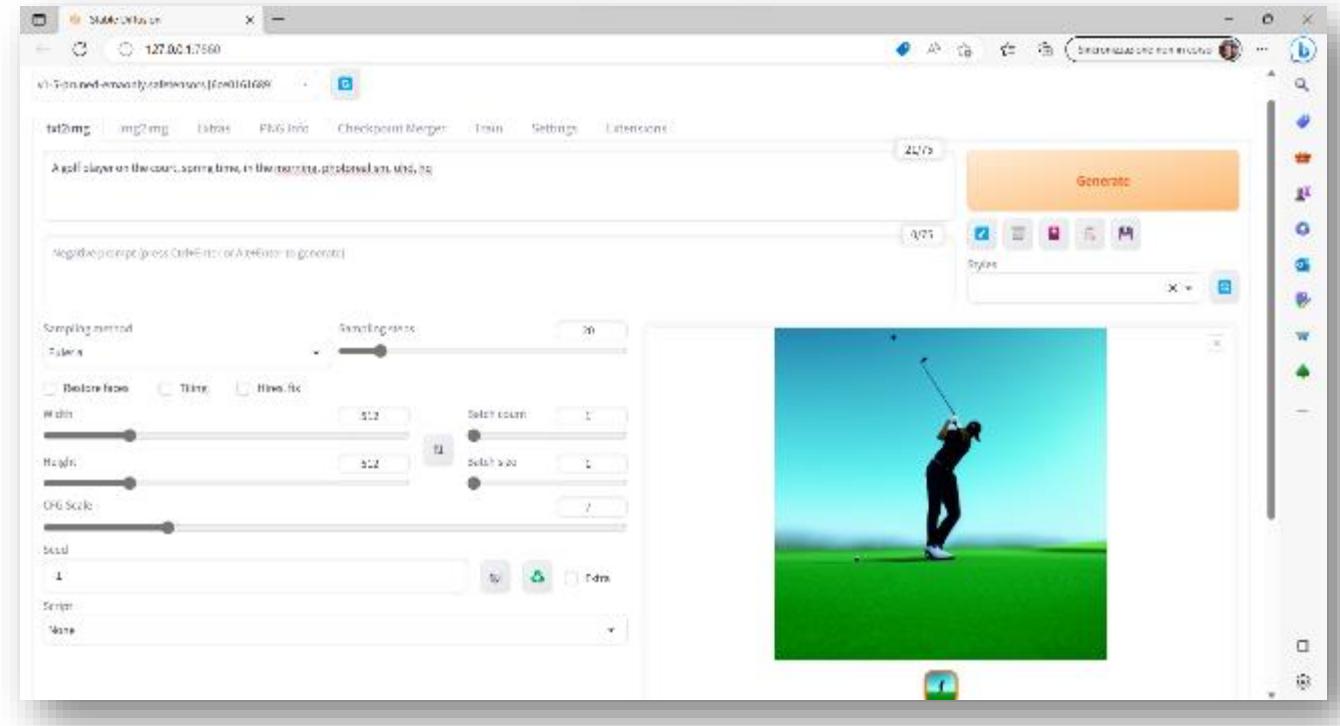


Output

Stable Diffusion Images

DEMO Web UI Tool

<https://github.com/AUTOMATIC1111/stable-diffusion-webui>



DEMO Generative AI Playground .NET

<https://github.com/gianni-rg/gen-ai-net-playground>



LLaMA

Text

Introducing LLaMA: A foundational, 65-billion-parameter large language model (LLM)

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

<https://github.com/facebookresearch/llama>

LLaMA: Open and Efficient Foundation Language Models

<https://arxiv.org/abs/2302.13971>

License:

source code → GPLv3

pre-trained models → **Non-Commercial**



LLaMA.cpp

Text

LLaMA^{©+}

DEMO

LOCAL chat experience in the Terminal

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

Model	Original size	Quantized size (4-bit)
7B	13 GB	3.9 GB
13B	24 GB	7.8 GB
30B	60 GB	19.5 GB
65B	120 GB	38.5 GB

https://github.com/tatsu-lab/stanford_alpaca

<https://github.com/nomic-ai/gpt4all>

<https://github.com/oobabooga/text-generation-webui>

Supported models:

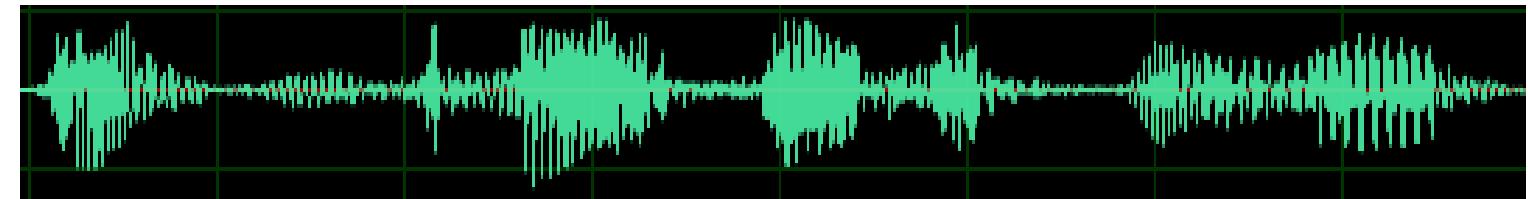
- LLaMA 
- Alpaca
- GPT4All
- Chinese LLaMA / Alpaca
- Vigogne (French)
- Vicuna
- Koala

Bindings:

- Python: [abetlen/llama-cpp-python](#)
- Go: [go-skynet/go-llama.cpp](#)
- Node.js: [hlhr202/llama-node](#)
- Ruby: [yoshoku/llama_cpp.rb](#)

Whisper

Speech-To-Text



Robust Speech Recognition via Large-Scale Weak Supervision

<https://arxiv.org/abs/2212.04356>

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

Size	Parameters	English-only model	Multilingual model	Required VRAM	Relative speed
tiny	39 M	tiny.en	tiny	~1 GB	~32x
base	74 M	base.en	base	~1 GB	~16x
small	244 M	small.en	small	~2 GB	~6x
medium	769 M	medium.en	medium	~5 GB	~2x
large	1550 M	N/A	large	~10 GB	1x

DEMO

LOCAL Audio Transcription (IT/EN) in .NET

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

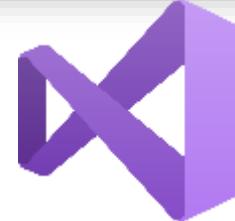
<https://github.com/sandrohanea/whisper.net>

<https://github.com/gianni-rg/gen-ai-net-playground>

DEMO

Browser (chat) STT → GPT2 → TTS

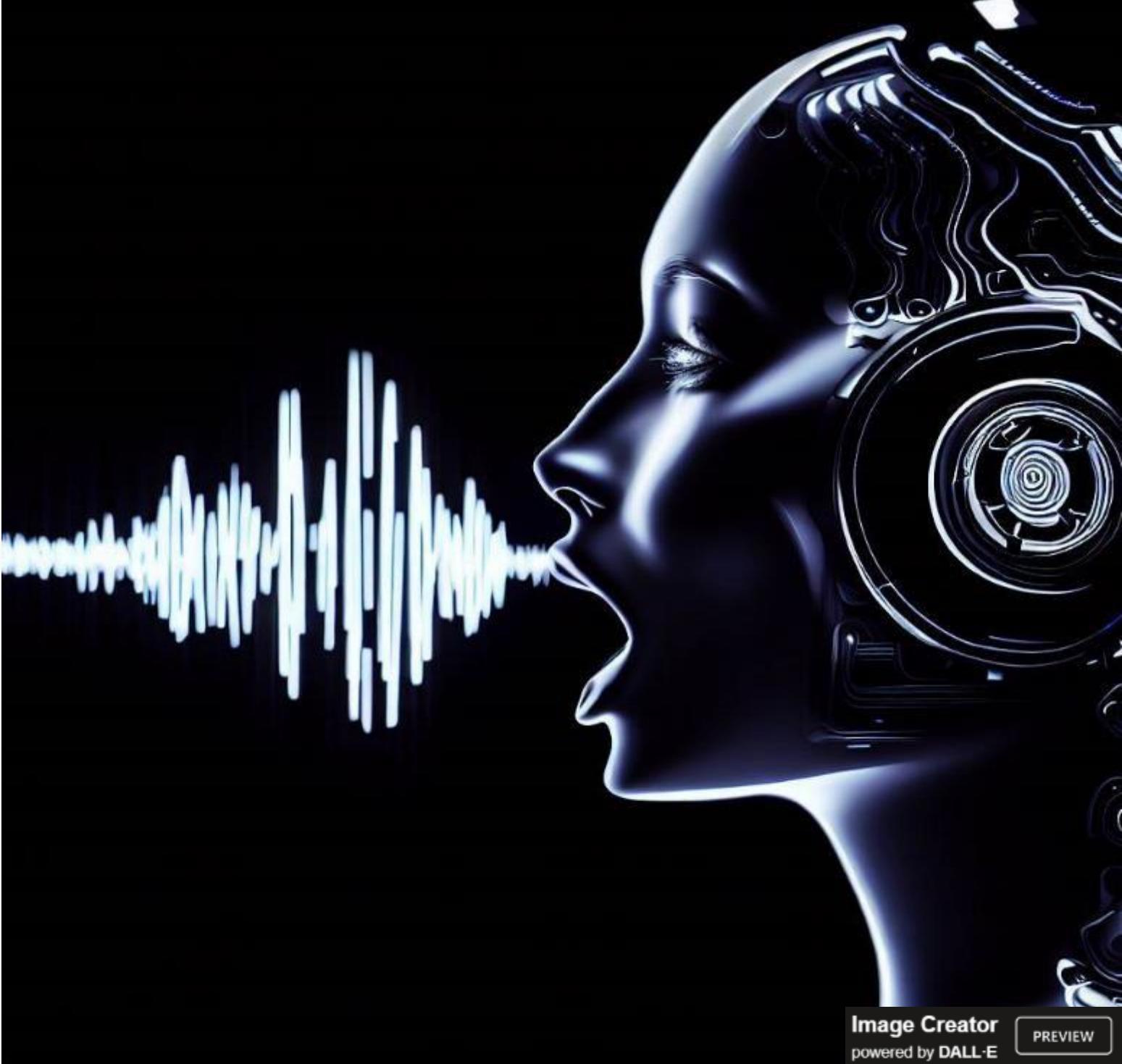
<https://whisper.ggerganov.com/talk/>



Voice Cloning

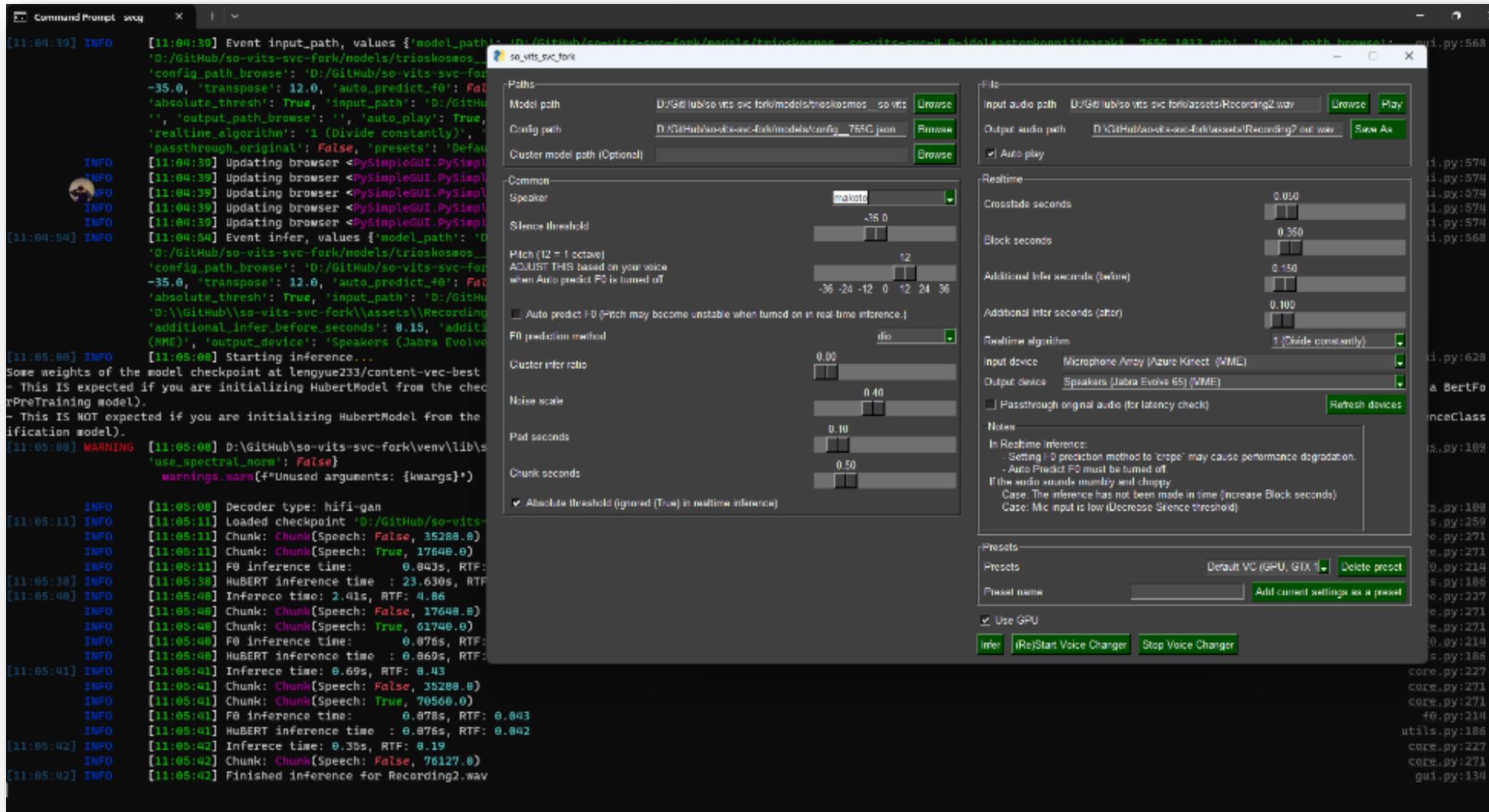
Audio to Audio

"Futuristic Female Head Shot
with waveform speech from
her mouth, high quality"



Voice Cloning

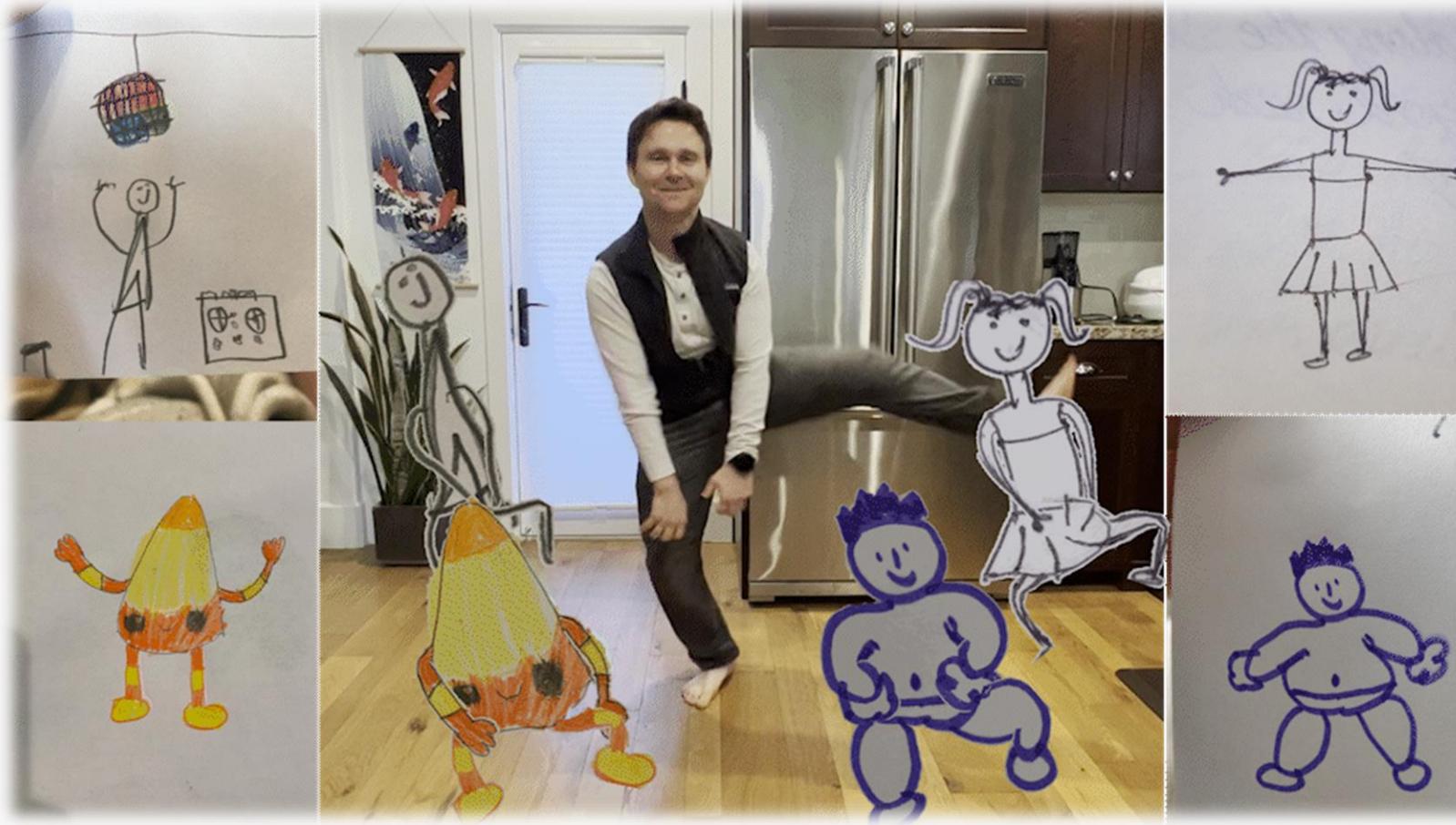
Audio to Audio



<https://github.com/voicepaw/so-vits-svc-fork>

Animated Drawings

Image to Video



Cinema DE-AGING



<https://www.youtube.com/watch?v=Pal1Vv9MpYY>

Cinema

Animate



<https://www.youtube.com/watch?v=Y1HGgICqZ3c>
<https://ebsynth.com/>

Real Time

High quality effects



- Speaker Focus
- Noise removal
- Room echo removal
- Audio Super-resolution
- Acoustic echo cancellation



- Virtual Background
- Super Resolution/ Upscaler
- Artifact Reduction
- Video Noise Removal



<https://developer.nvidia.com/maxine>

AR⁺

- Face Expression Estimation
- Eye Contact
- Face Tracking
- Face Landmark Tracking
- Face Mesh
- Body Pose Estimation

Thank You!

ευχαριστώ

Salamat Po

متشكر م

شَكْرًا

Grazie

благодаря

ありがとうございます

Kiitos

Teşekkürler

謝謝

ឧបម្ពុណ្ឌរំបៀប

Obrigado

شُكْرِيَّة

Terima Kasih

Dziękuję

Hvala

Köszönöm

Tak

Dank u wel

дякую

Tack

Mulțumesc

спасибо

Danke

Cám ơn

Gracias

多謝晒

Ďakujem

הַדְוָת

ശ്രദ്ധി

Děkuji

감사합니다

Code & Slides



<https://github.com/gianni-rg>

<https://globalazuretorino.welol.it/>



References (1/2)

- <https://www.bing.com/images/create/>
- <https://midjourney.com/>
- <https://azure.microsoft.com/en-us/products/cognitive-services/openai-service>
- <https://runwayml.com/>
- <https://research.runwayml.com/gen2>
- <https://openai.com/research/whisper>
- https://www.youtube.com/watch?v=17_xLsqny9E
- <https://beta.elevenlabs.io/>
- <https://research.nvidia.com/labs/dir/magic3d/>
- <https://developer.nvidia.com/blog/getting-started-with-nvidia-instant-nerfs/>
- <https://jonbarron.info/zipnerf/>
- <https://github.com/steven2358/awesome-generative-ai>
- <https://github.com/imaurer/awesome-decentralized-llm>
- <https://onnx.ai/>
- <https://docs.openvino.ai/>
- <https://www.nvidia.com>
- https://github.com/openvinotoolkit/openvino_notebooks
- <https://www.youtube.com/watch?v=Pal1Vv9MpYY>
- <https://developer.nvidia.com/maxine>

References (2/2)

- https://huggingface.co/blog/stable_diffusion
- <https://huggingface.co/blog/annotated-diffusion>
- <https://github.com/huggingface/diffusers>
- <https://github.com/runwayml/stable-diffusion>
- <https://github.com/AUTOMATIC1111/stable-diffusion-webui/>
- <https://github.com/gianni-rg/gen-ai-net-playground>
- <https://github.com/facebookresearch/llama>
- <https://arxiv.org/abs/2302.13971>
- <https://github.com/ggerganov/llama.cpp>
- https://github.com/tatsu-lab/stanford_alpaca
- <https://github.com/nomic-ai/gpt4all>
- <https://github.com/oobabooga/text-generation-webui>
- <https://github.com/openai/whisper>
- <https://github.com/ggerganov/whisper.cpp>
- <https://github.com/sandrohanea/whisper.net>
- <https://github.com/gianni-rg/gen-ai-net-playground>
- <https://whisper.ggerganov.com/talk/>
- <https://github.com/voicepaw/so-vits-svc-fork>
- <https://github.com/facebookresearch/AnimatedDrawings>

About me



Microsoft

Specialist

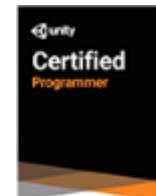
Programming in C#
Programming in HTML5
with JavaScript & CSS3



Microsoft
CERTIFIED

Solutions Developer

Windows Store Apps Using C#
Web Applications



PLURALSIGHT
Author

Ing. Gianni ROSA GALLINA
R&D Technical Lead @ **deltatre**



@giannirg

- AI, Machine Learning, Deep Learning on multimedia content
- Virtual/Augmented/Mixed Reality
- Immersive video streaming & 3D graphics for sport events
- Cloud solutions, web backends, serverless, video workflows
- Mobile apps dev (Windows / Android / .NET MAUI / Avalonia)
- End-to-end solutions with Microsoft Azure

<https://gianni.rosagallina.com/en/>

