UGIDOTNET SMALL TALKS

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

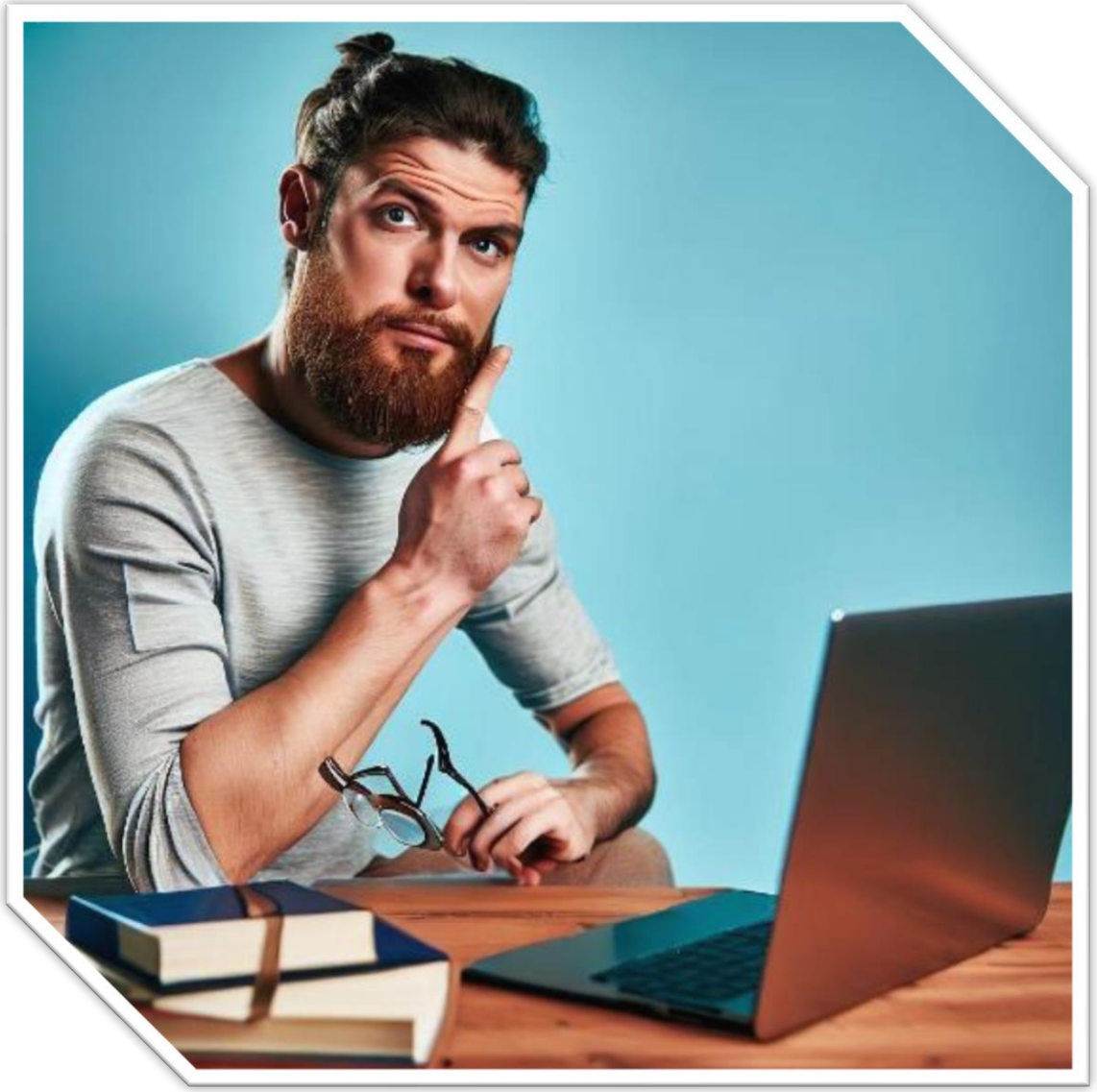
GIANNI ROSA GALLINA

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

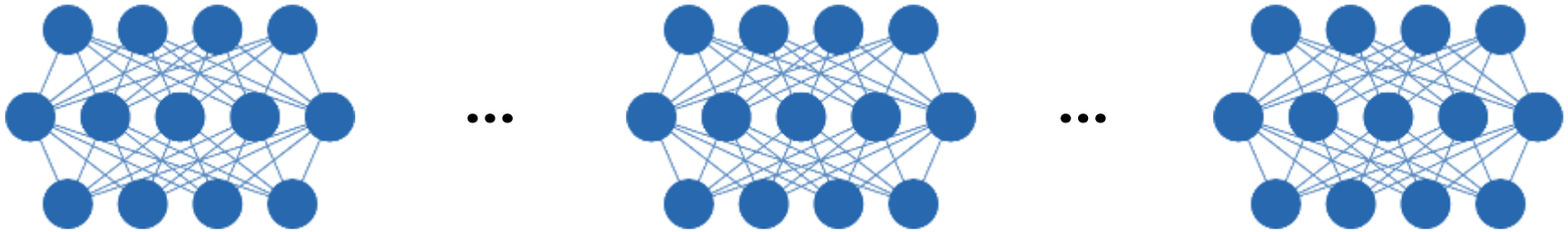
…

Generative AI

Overview

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"

OpenAI

Microsoft

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

Generative AI

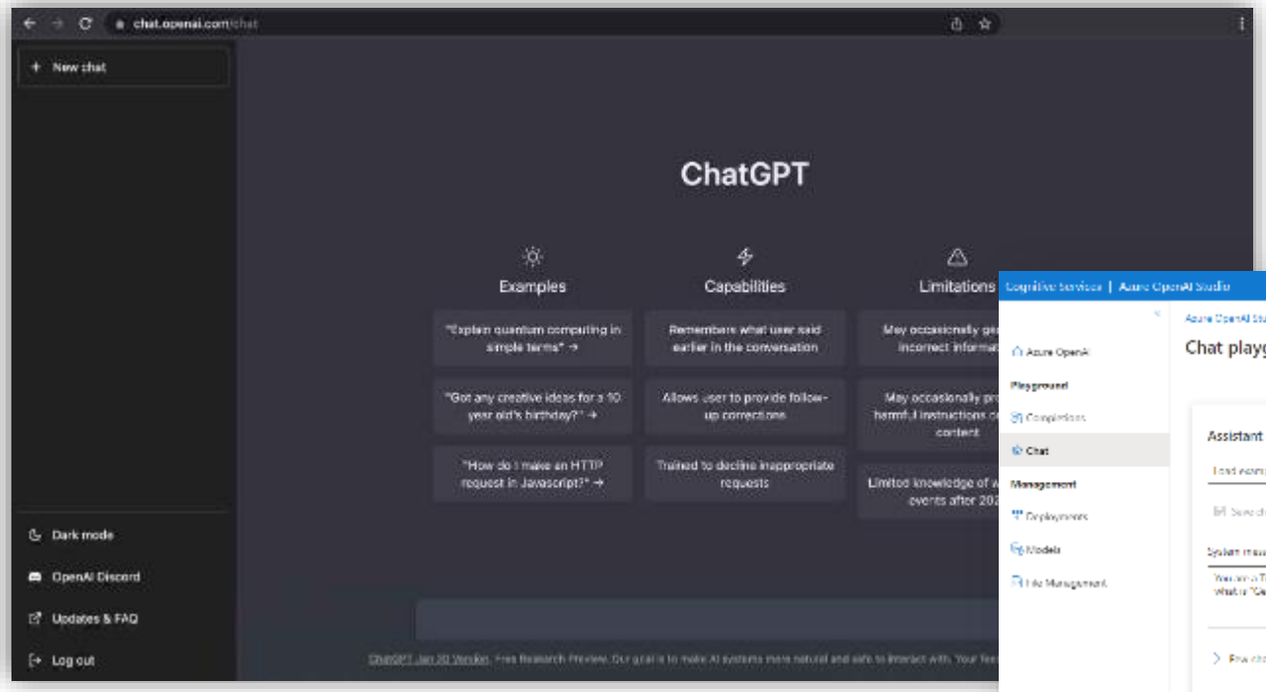
Images

"star wars animals"

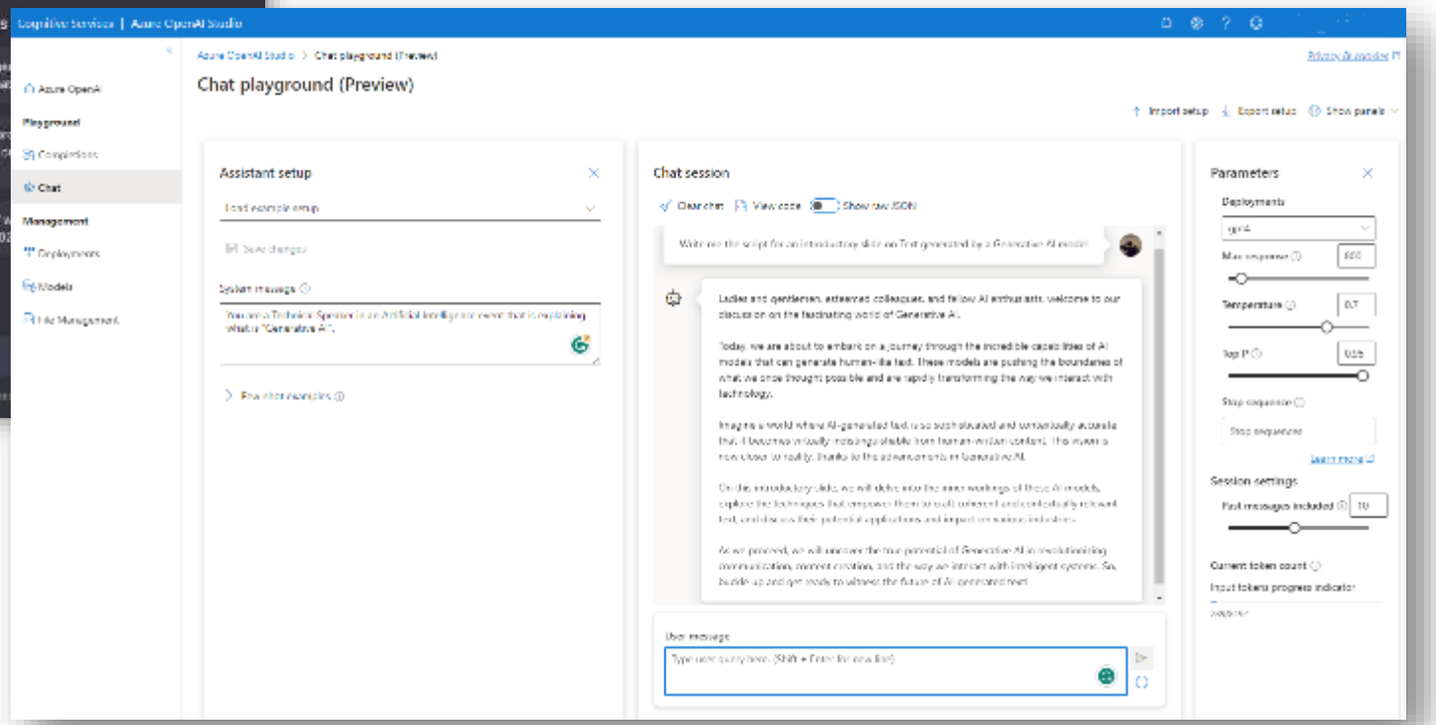
https://midjourney.com/

Generative AI

Text



https://chat.openai.com/



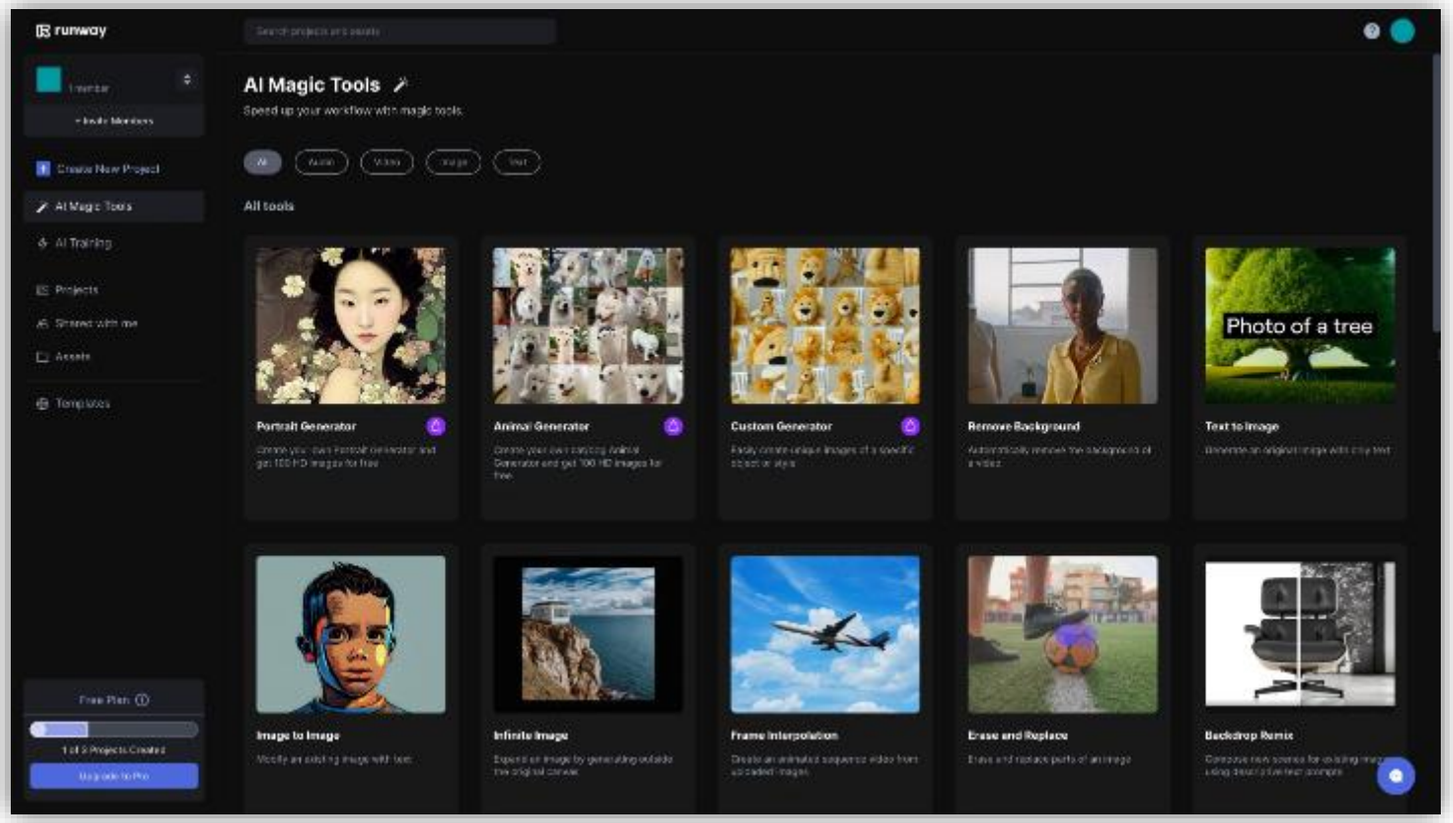
OpenAI

Microsoft

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)

OpenAI

https://openai.com/research/whisper

Generative AI

Audio (Text-To-Speech)

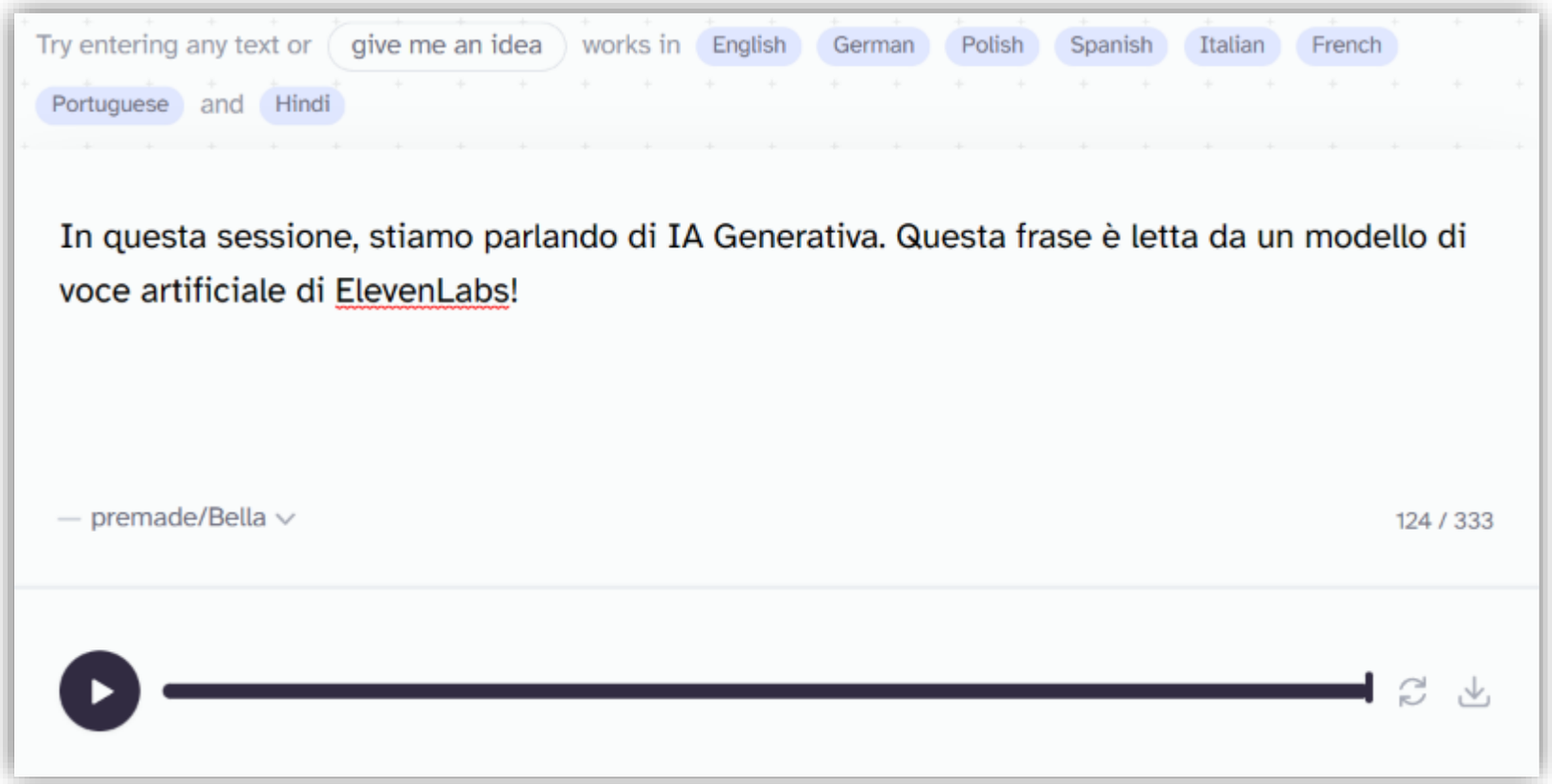
llElevenLabs – Voice Dubbing demo

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

Generative AI

Audio

llElevenLabs – Prime Voice AI



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."

https://research.nvidia.com/labs/dir/magic3d/

Generative AI

2D to 3D generation

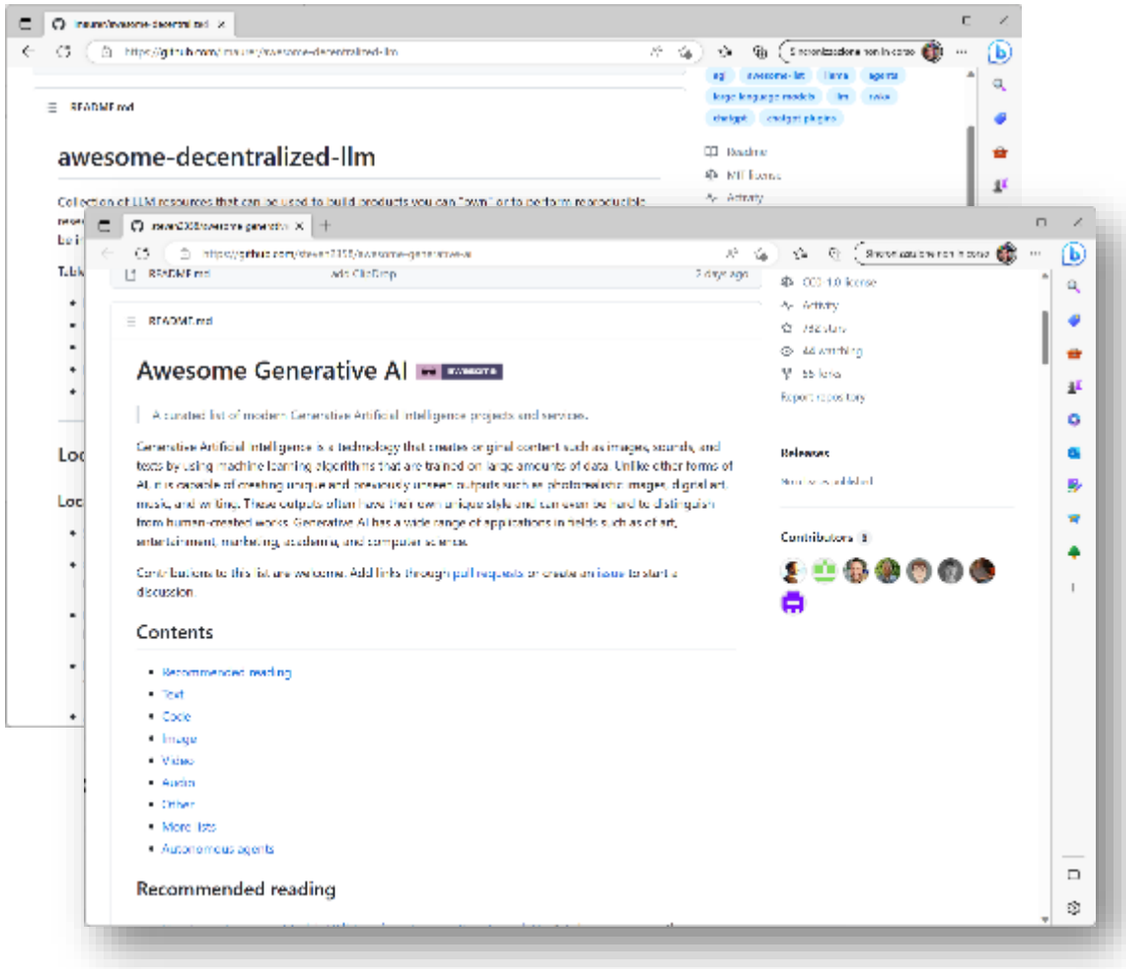
https://developer.nvidia.com/blog/getting-started-with-nvidia-instant-nerfs/

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-llm
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-llm

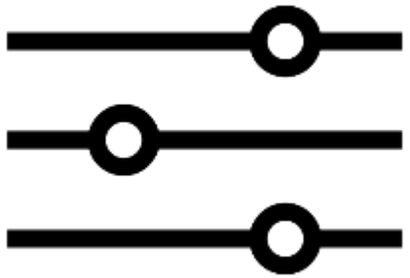


Hugging Face

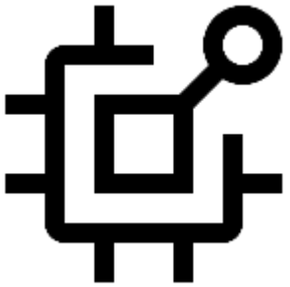
stability.ai

LAION

Why alternative & open-source models?



Customization & Flexibility



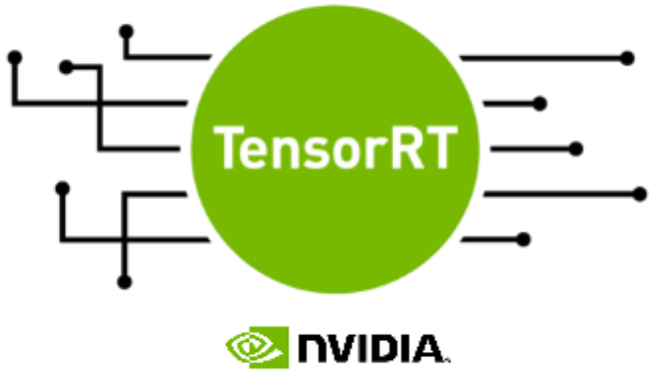
Embedded/Mobile Devices

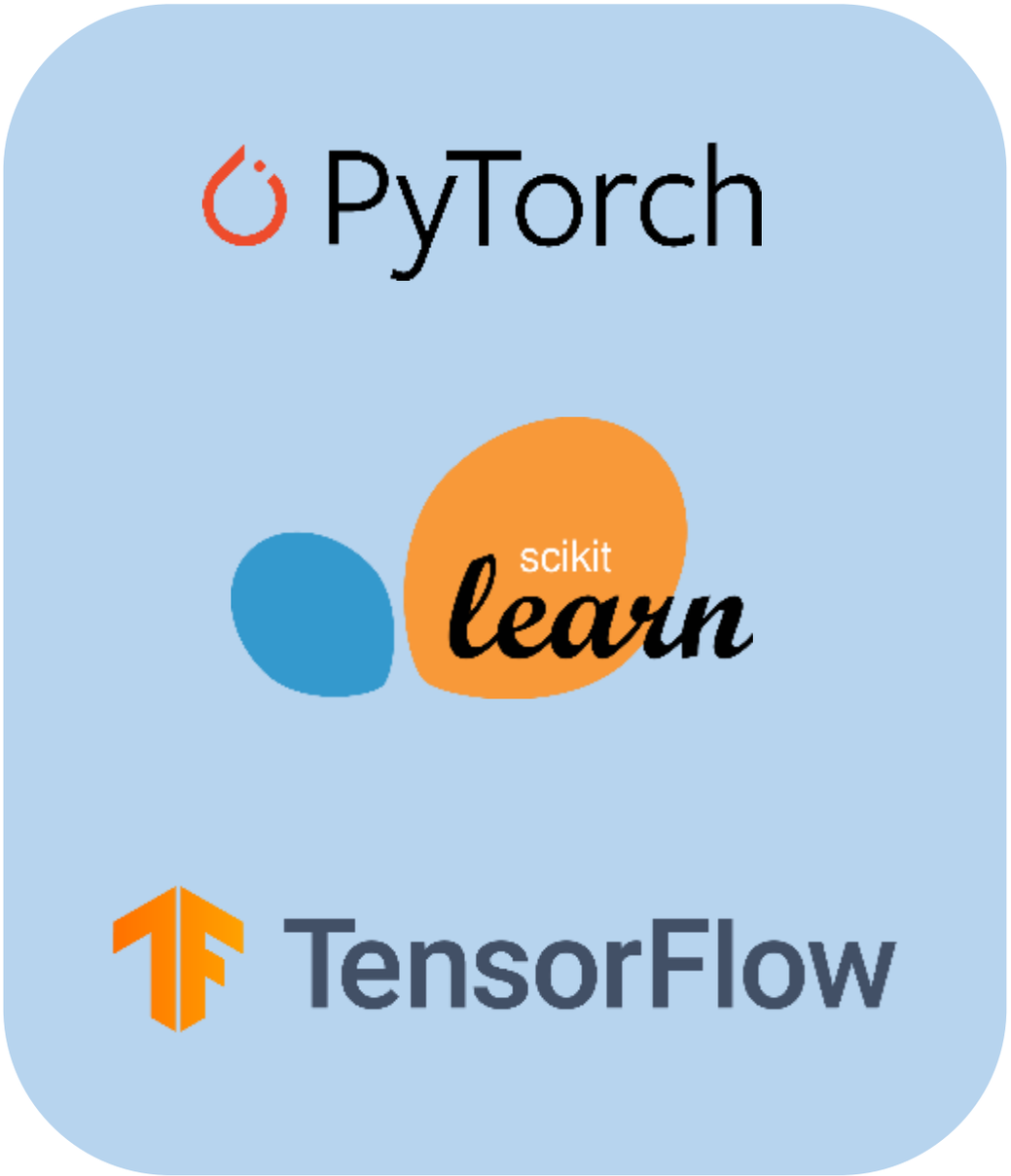
Data Policies & Ownership

No/Limited Connectivity

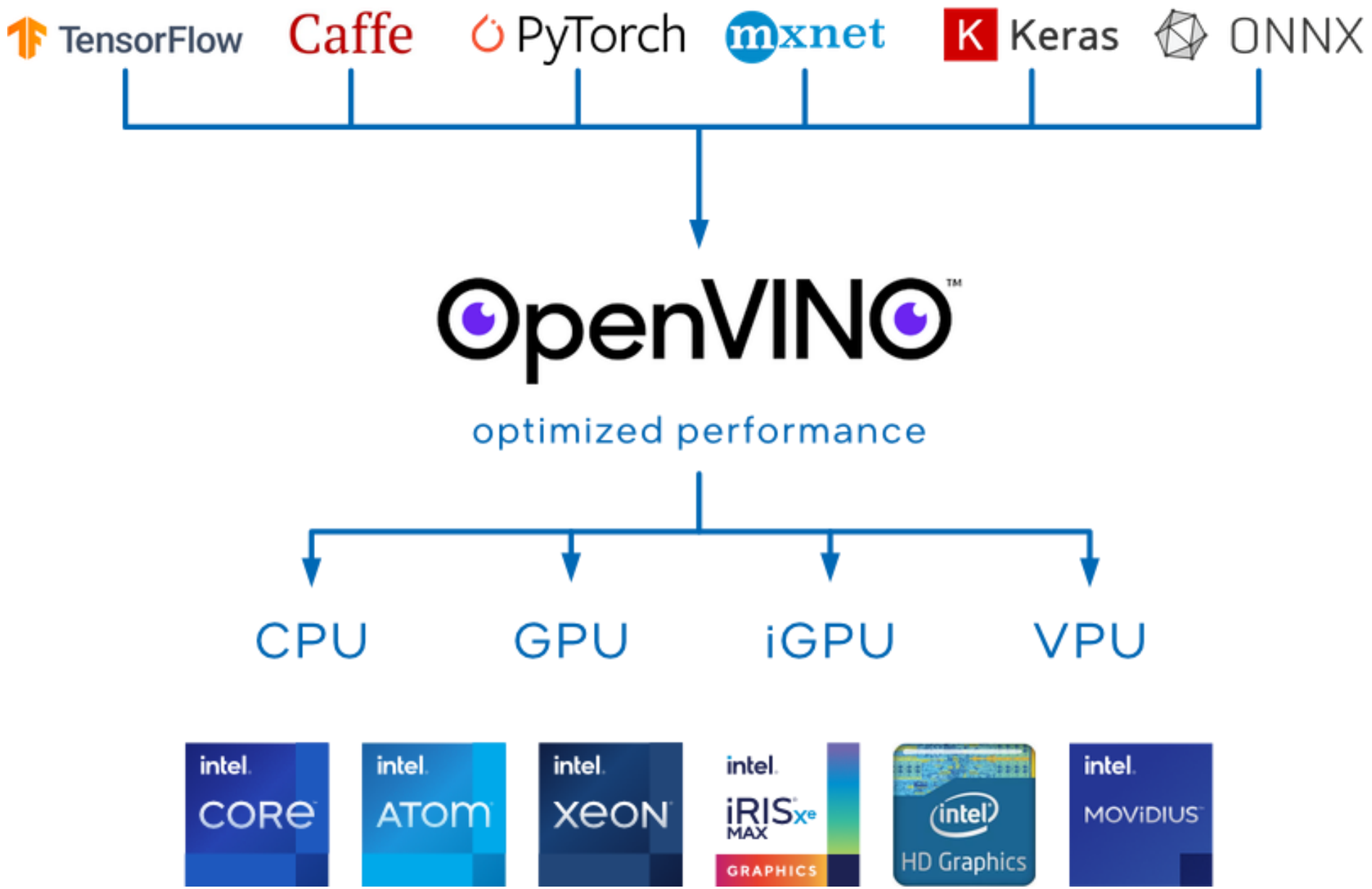
Savings & Optimization

Tools and Frameworks

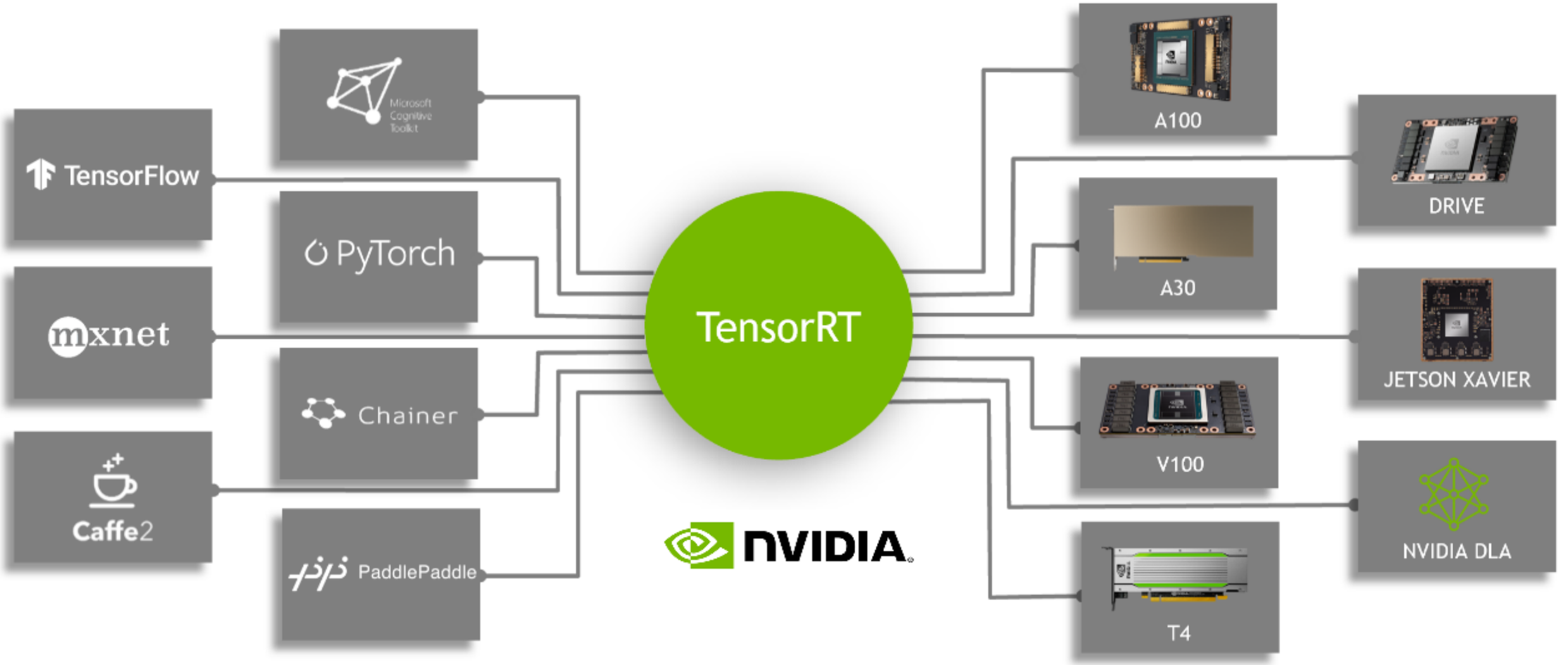




https://onnx.ai/



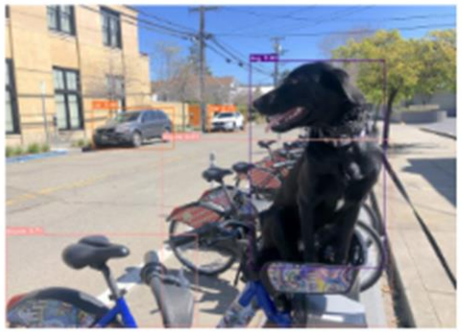
https://docs.openvino.ai/



https://www.nvidia.com


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

https://github.com/openvinotoolkit/openvino_notebooks

Stable Diffusion

Images

stability.ai

runway

LAION

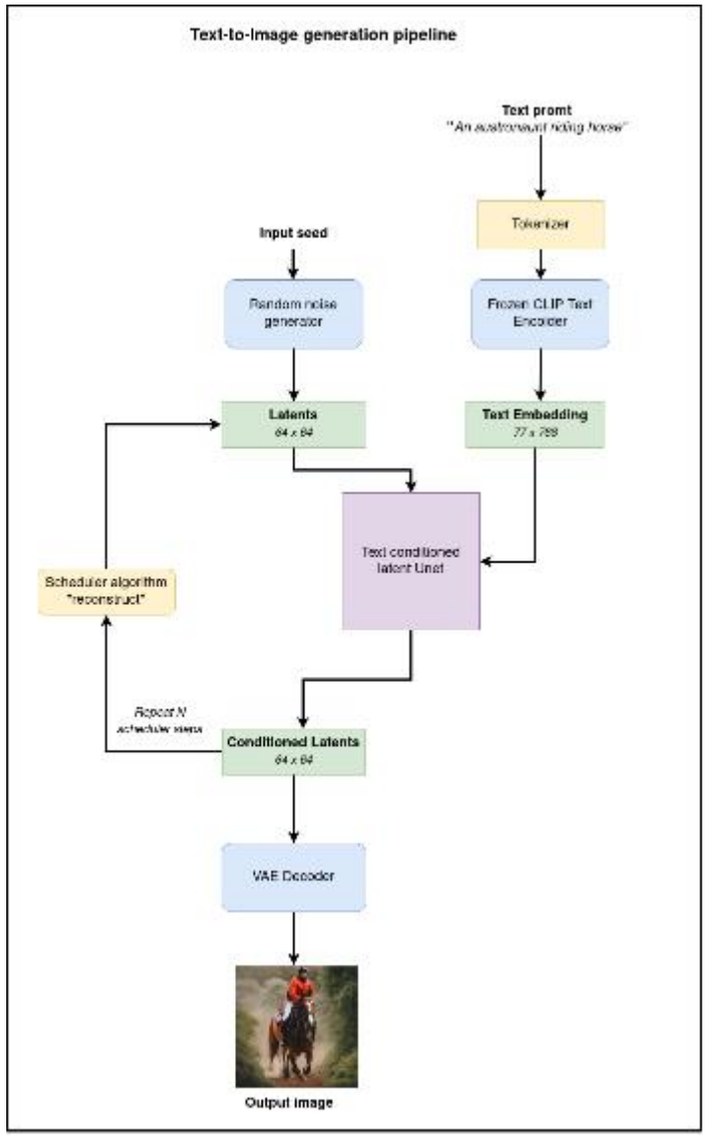
Hugging Face

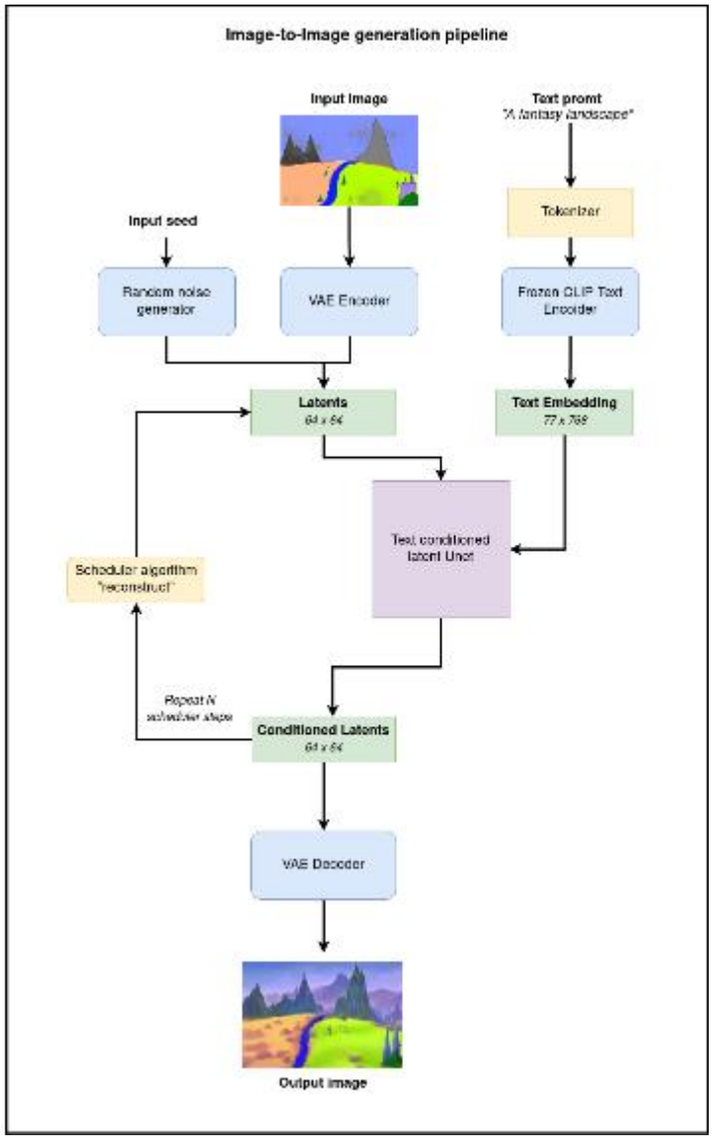
Diffusers

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



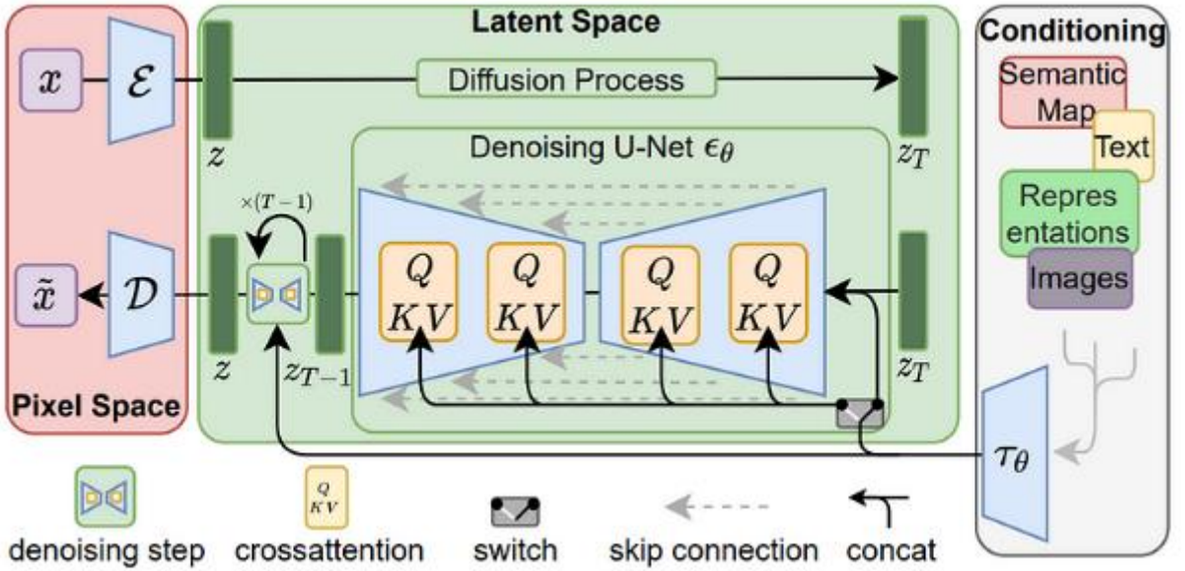


Source: **https://github.com/openvinotoolkit/openvino_notebooks (225_stable-diffusion)**

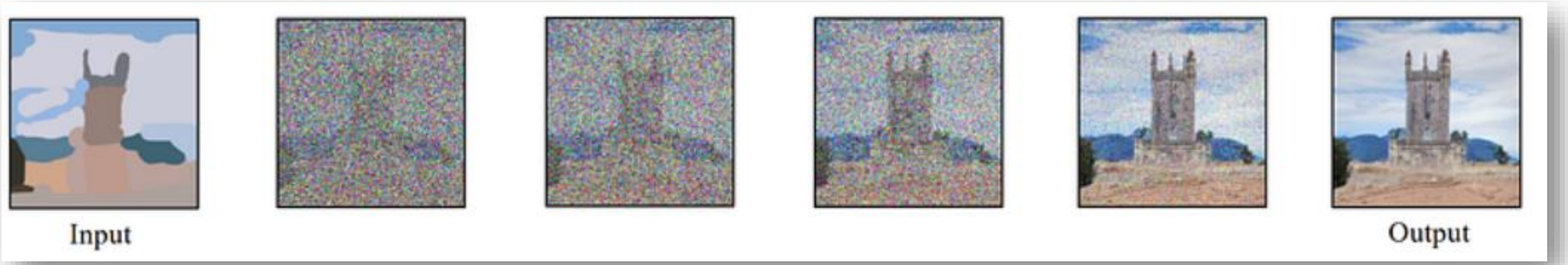
Stable Diffusion

Images

High-Resolution Image Synthesis with Latent Diffusion Models
https://arxiv.org/abs/2112.10752



steps

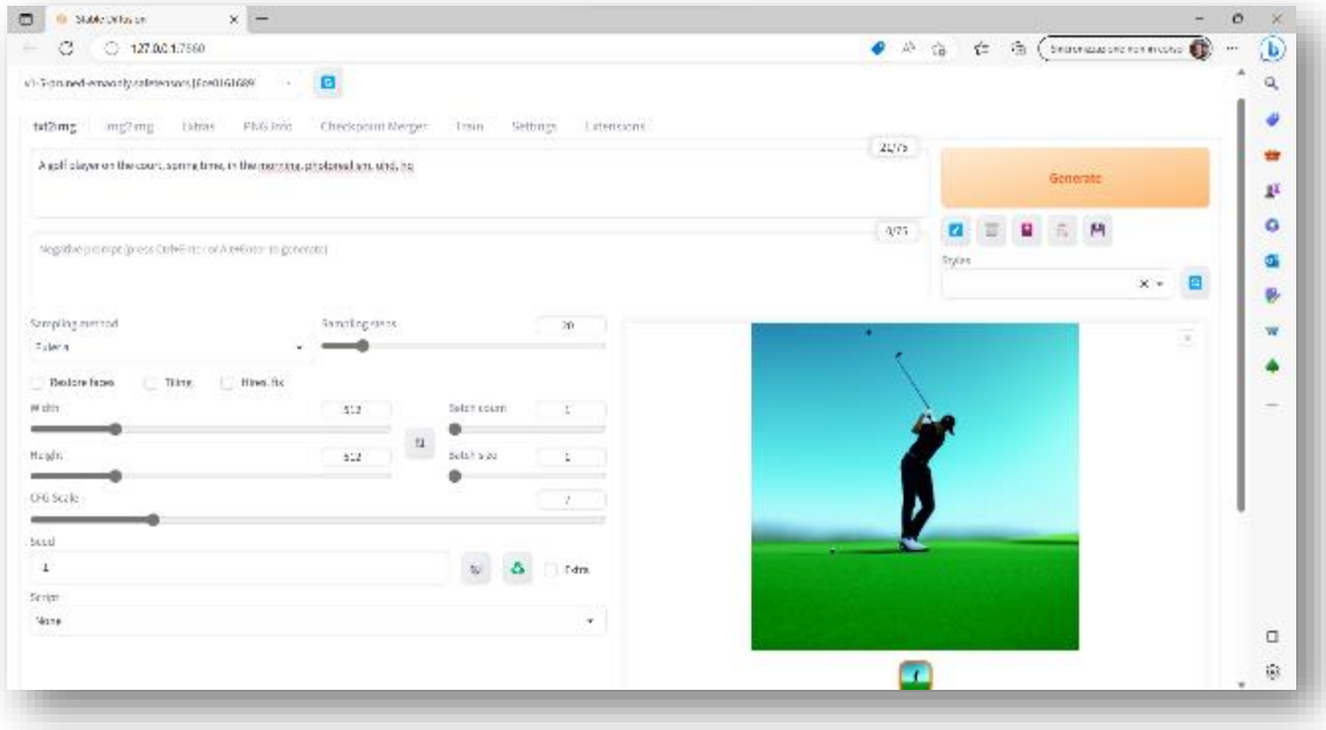


Stable Diffusion

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

.NET Framework

ONNX

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

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:

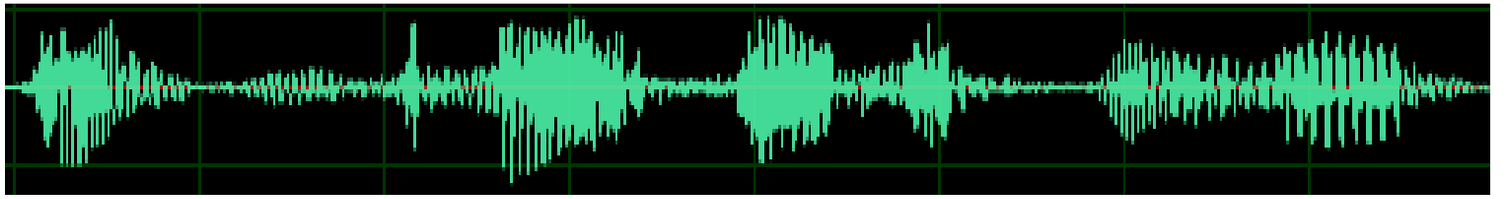
- LLaMA
- Alpaca
- GPT4All
- Chinese LLaMA / Alpaca
- Vigogne (French)
- Vicuna
- Koala

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

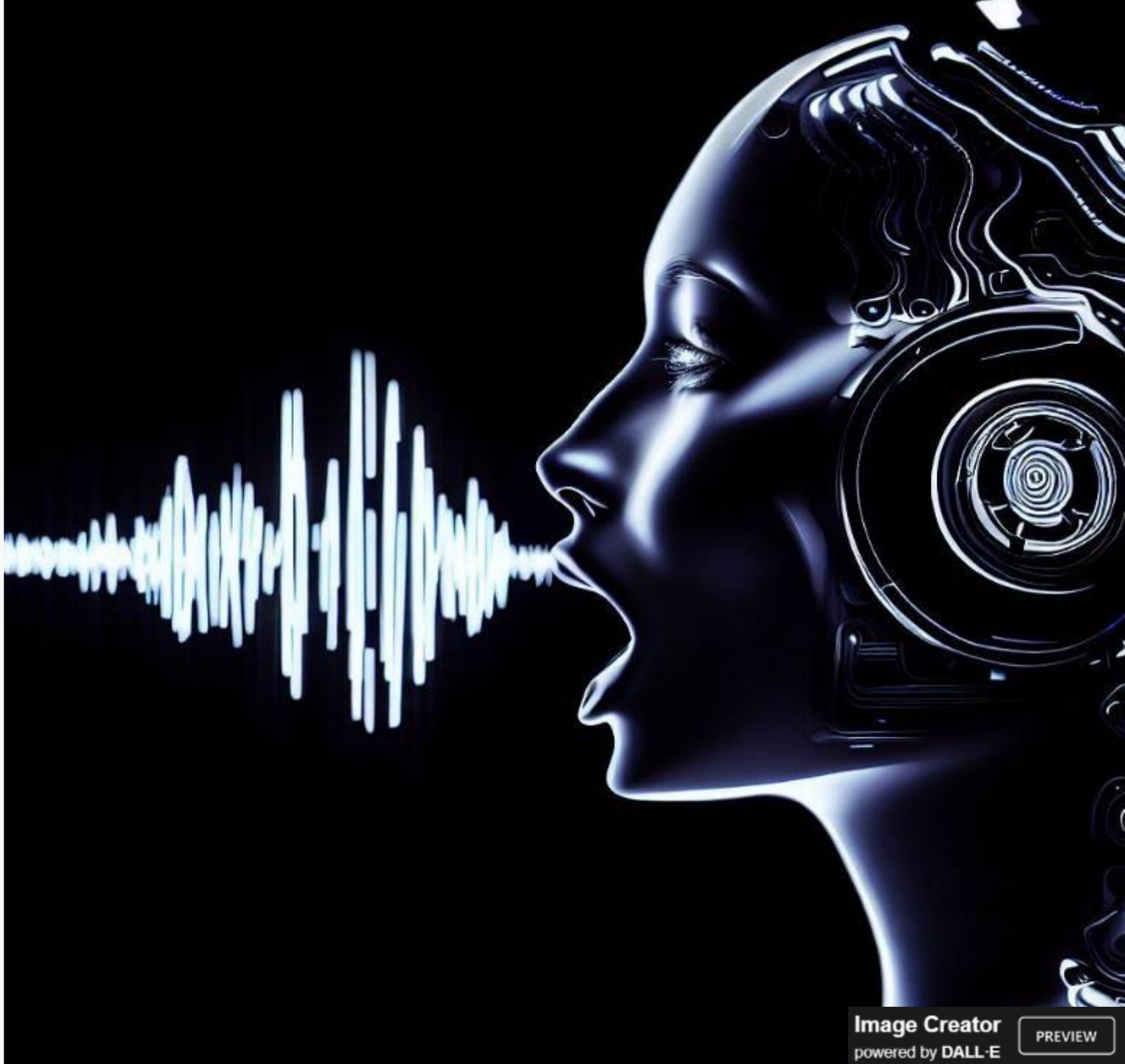
Browser (chat) STT → GPT2 → TTS

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

Voice Cloning

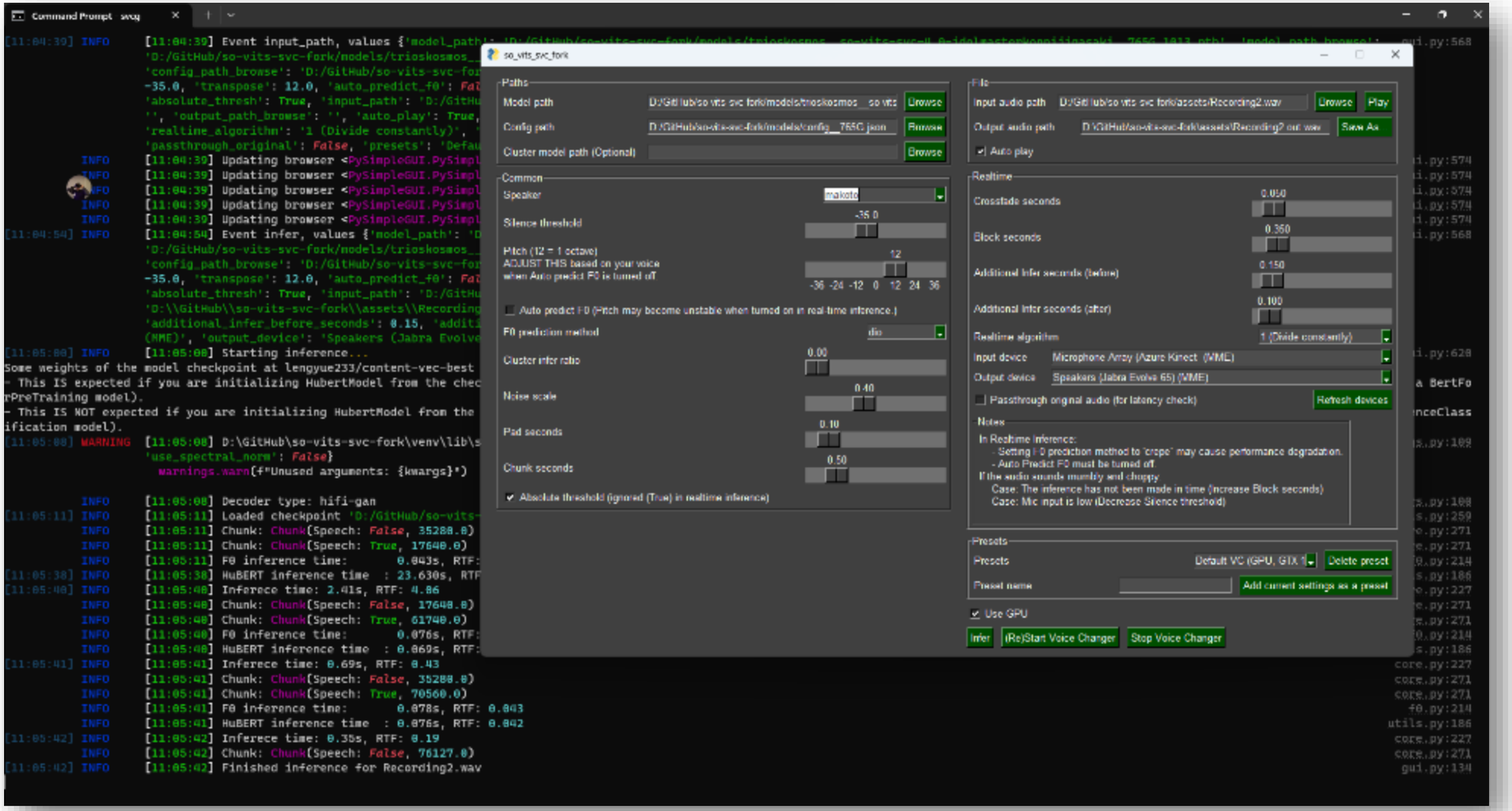
Audio to Audio

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



Voice Cloning

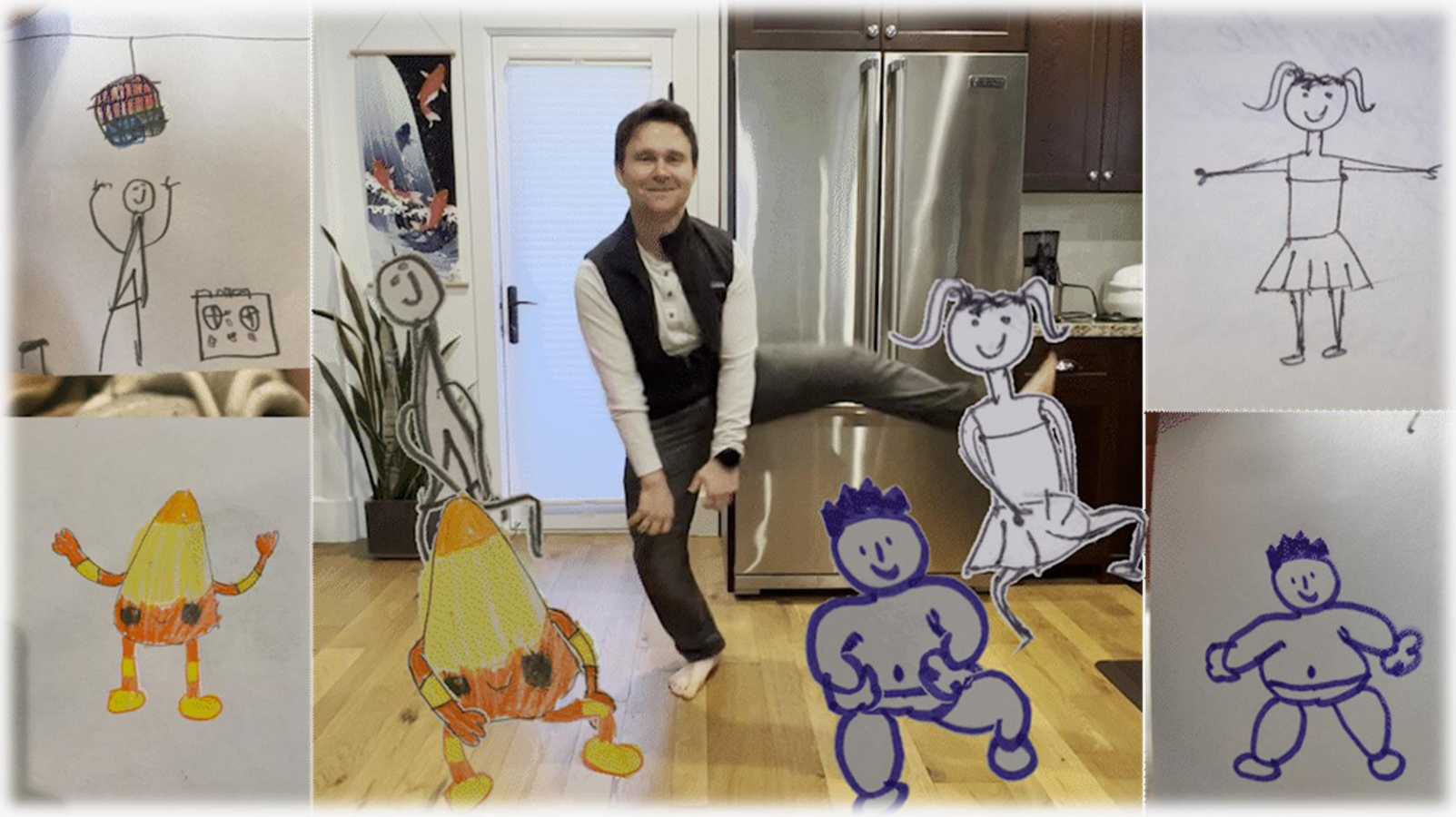
Audio to Audio



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

Animated Drawings

Image to Video



https://github.com/facebookresearch/AnimatedDrawings

Cinema
DE-AGING

https://www.youtube.com/watch?v=PaI1Vv9MpYY

Cinema

Animate

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

Real Time

High quality effects

https://developer.nvidia.com/maxine

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

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

AR+

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

Thank You!

ευχαριστώ Salamat Po متشكرم شكراً Grazie

благодаря ありがとうございます Kiitos Teşekkürler 谢谢

ขอบคุณครับ Obrigado شكريه Terima Kasih Dziękuję

Hvala Köszönöm Tak Dank u wel дякую Tack

Mulţumesc спасибо Danke Cám ơn Gracias

多謝晒 Ďakujem תודה நன்றி Děkuji 감사합니다

Code & Slides

https://github.com/gianni-rg

https://globalazuretorino.welol.it/

TORINO

13 Maggio 2023

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=PaI1Vv9MpYY
- 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

Ing. Gianni ROSA GALLINA

@giannirg

R&D Technical Lead @ deltatre

- 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

MVP Microsoft Most Valuable Professional

Microsoft Specialist
Programming in C#
Programming in HTML5 with JavaScript & CSS3

Microsoft CERTIFIED Solutions Developer
Windows Store Apps Using C#
Web Applications

Microsoft Certified: Azure AI Engineer Associate

Microsoft Certified: Azure Data Scientist Associate

Unity Certified Programmer

PLURALSIGHT Author

SCAN ME

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