global AI conference

December 12th 2023

Generative AI Unleashed: crafting voices and music from text & images

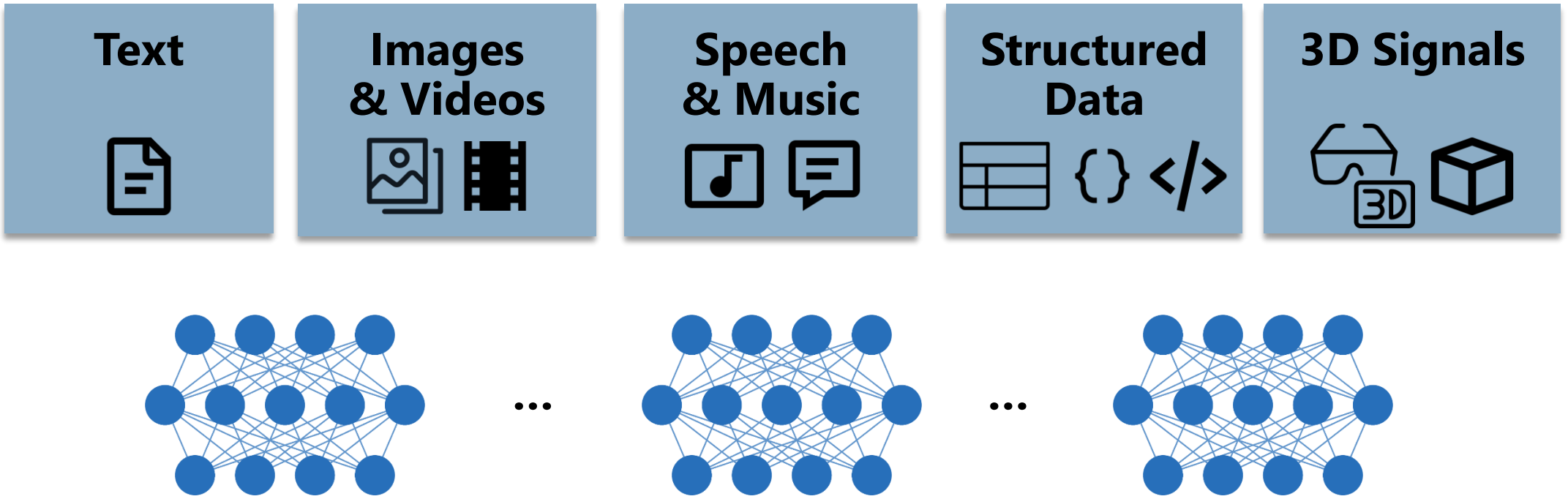
Clemente Giorio
EYES ON

Gianni Rosa Gallina
deltatre

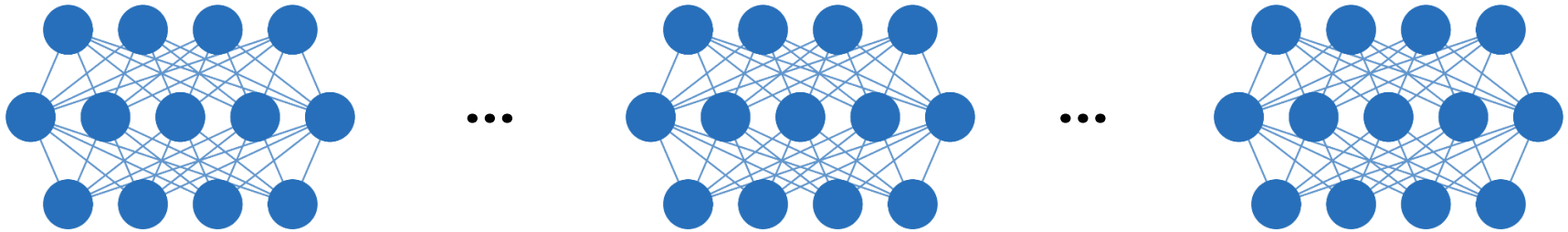
MVP Microsoft Most Valuable Professional

Generative AI

Overview



Generative AI
Overview



Foundation Models

- Question Answering
- Information Extraction
- Image Captioning
- Multimodal Translation
- Text-To-X
- Image-To-X

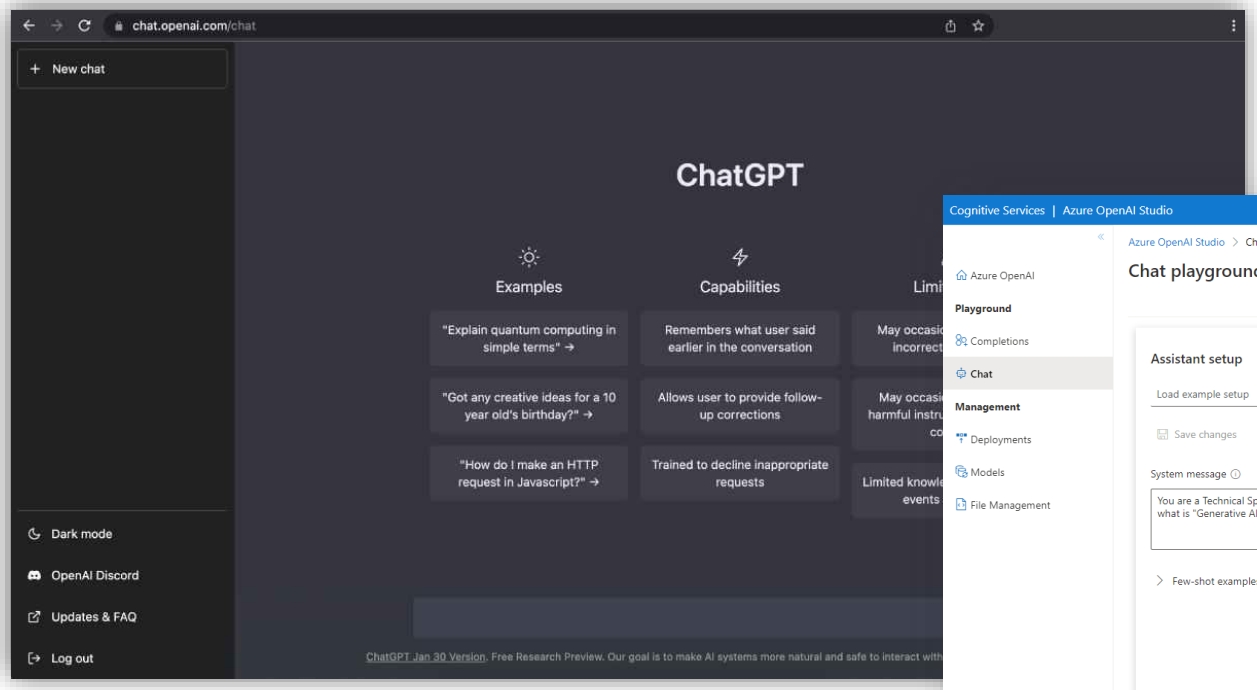
Generative AI… for all and everything

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

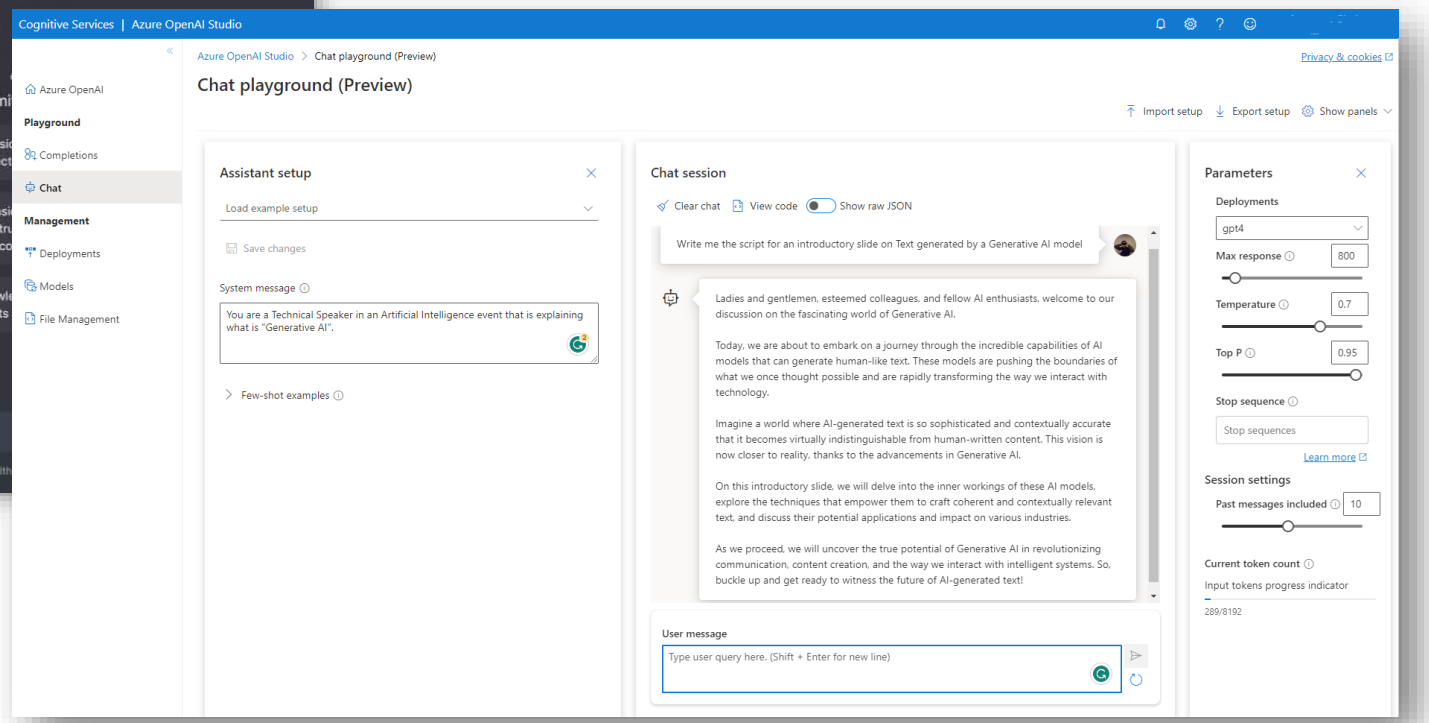
…

Generative AI

Text



https://chat.openai.com/



https://azure.microsoft.com/en-us/products/ai-services/openai-service/

OpenAI

Microsoft

Generative AI

Text-To-Image

"Gothic event room full of wizard software developers"

DALL·E 3

https://openai.com/dall-e-3

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

Generative AI

Text-To-Image



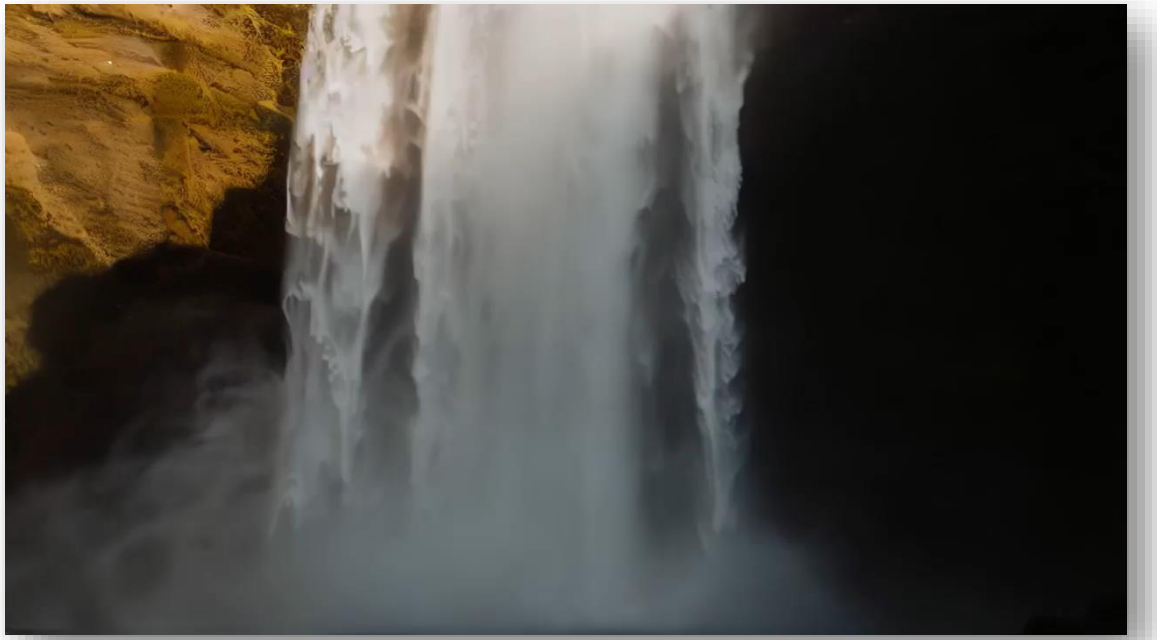

stability.ai runway LAION Hugging Face

Diffusers

https://huggingface.co/blog/stable_diffusion
https://stability.ai/stable-diffusion
https://github.com/huggingface/diffusers
https://runwayml.com/

Generative AI
Text-To-Video

runway



https://runwayml.com/

Generative AI

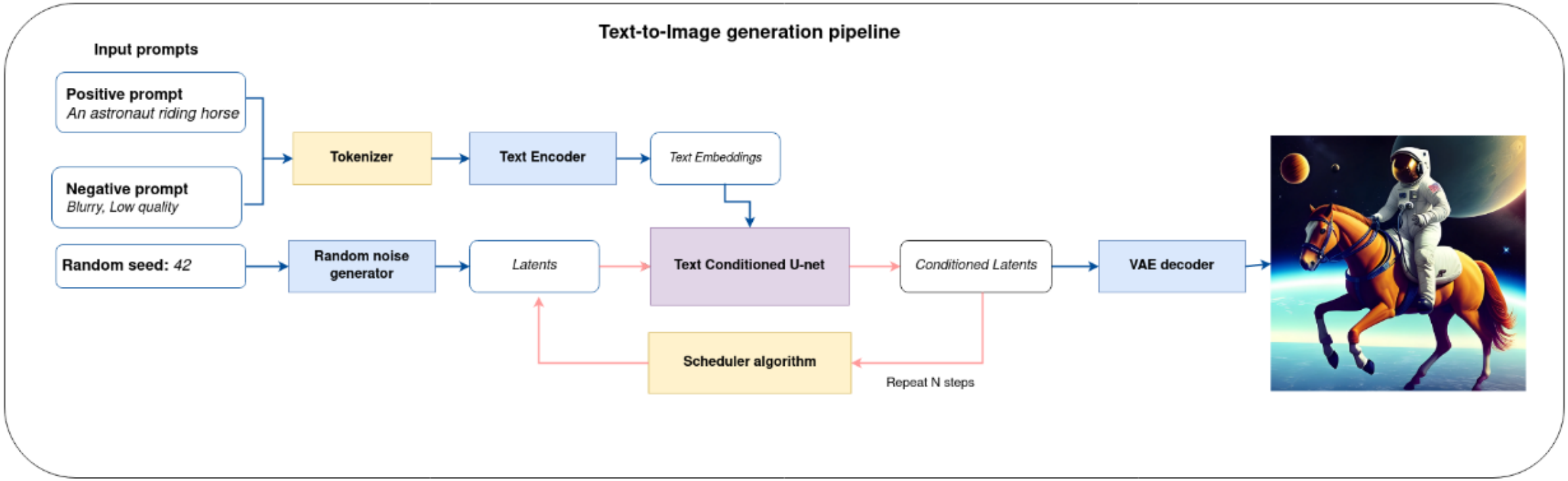
Text-To-Video

NEW

stability.ai

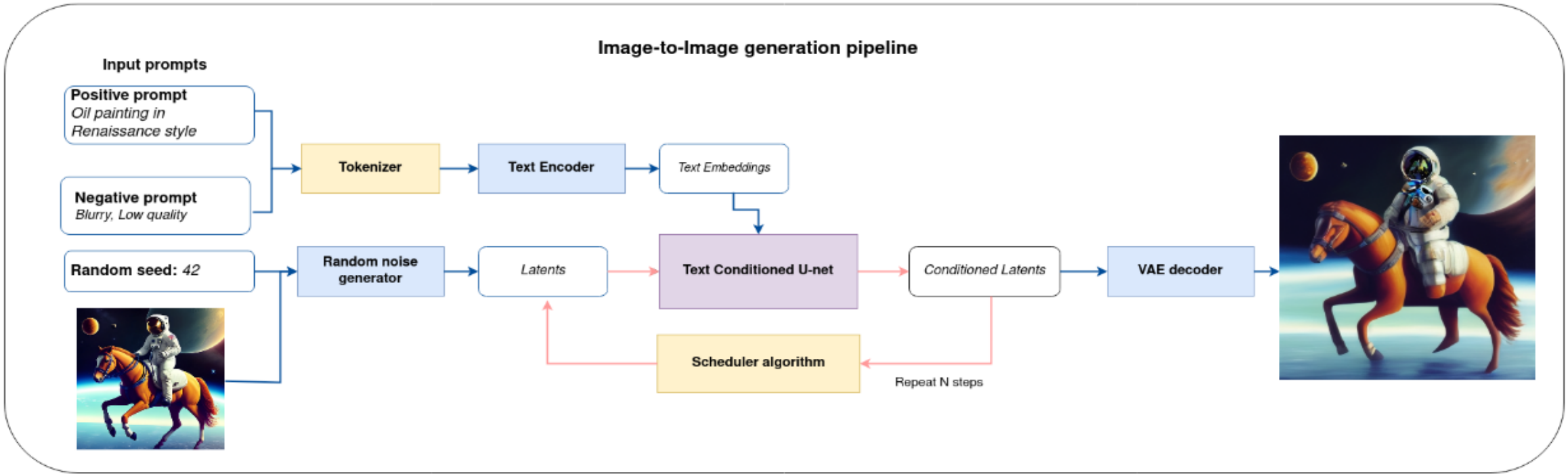
https://huggingface.co/stabilityai/stable-video-diffusion-img2vid

Images
Stable Diffusion



https://stability.ai/

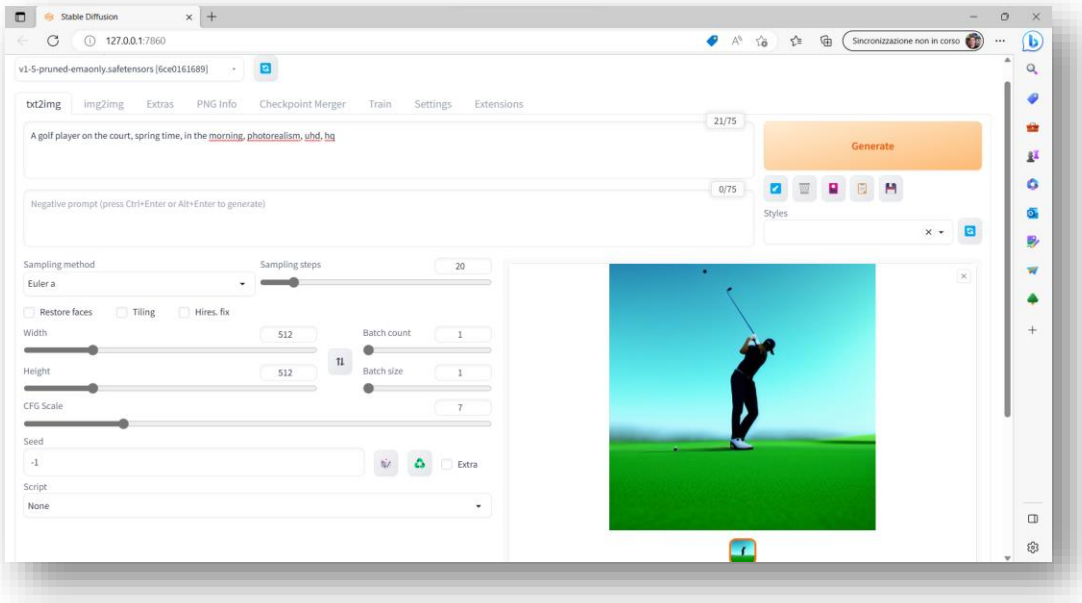
Images
Stable Diffusion

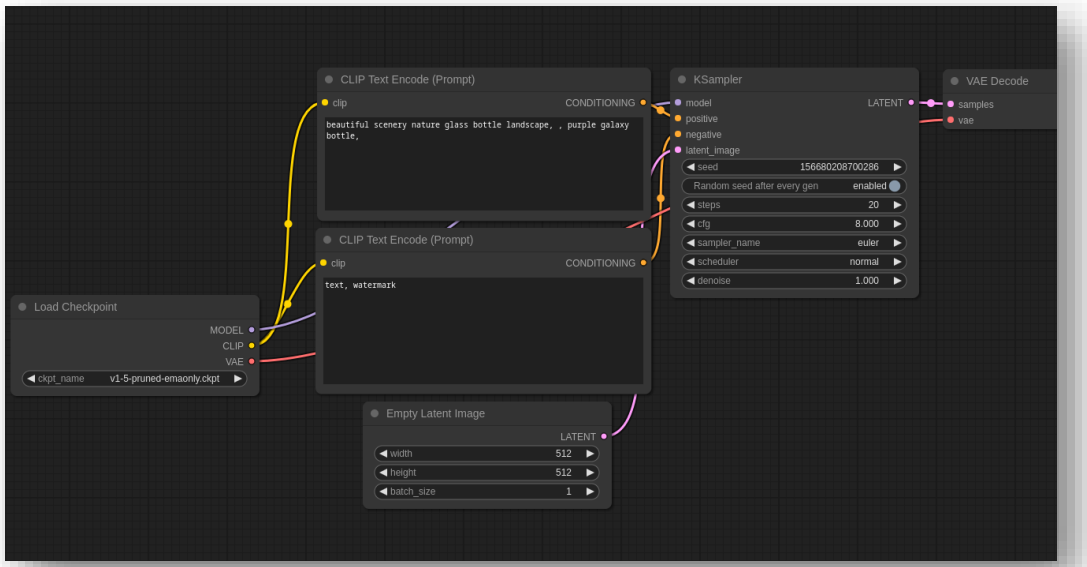


https://stability.ai/

Images

Stable Diffusion – Web UI Tool





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

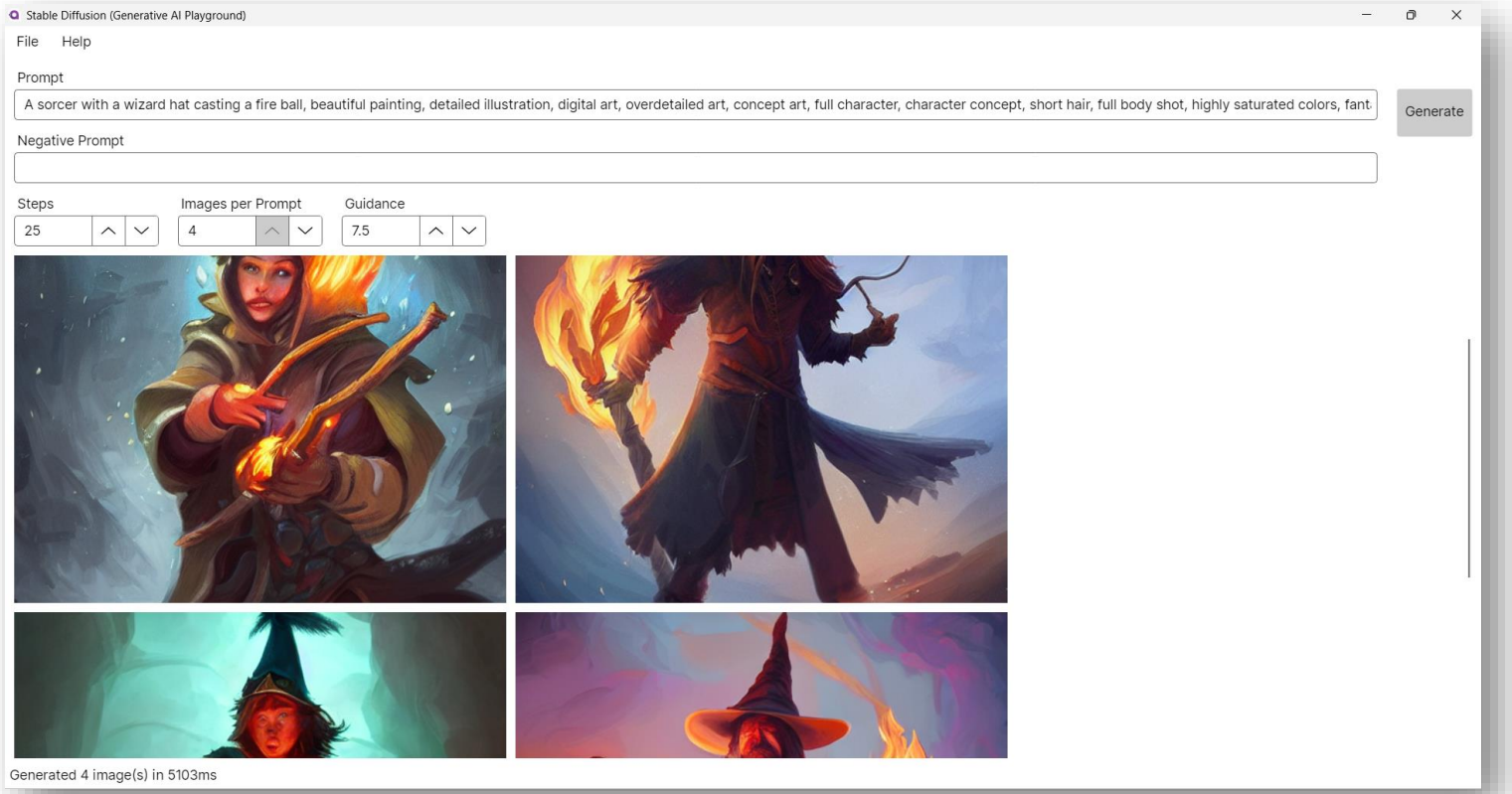
https://github.com/comfyanonymous/ComfyUI

Images

Stable Diffusion: ONNX Runtime & .NET

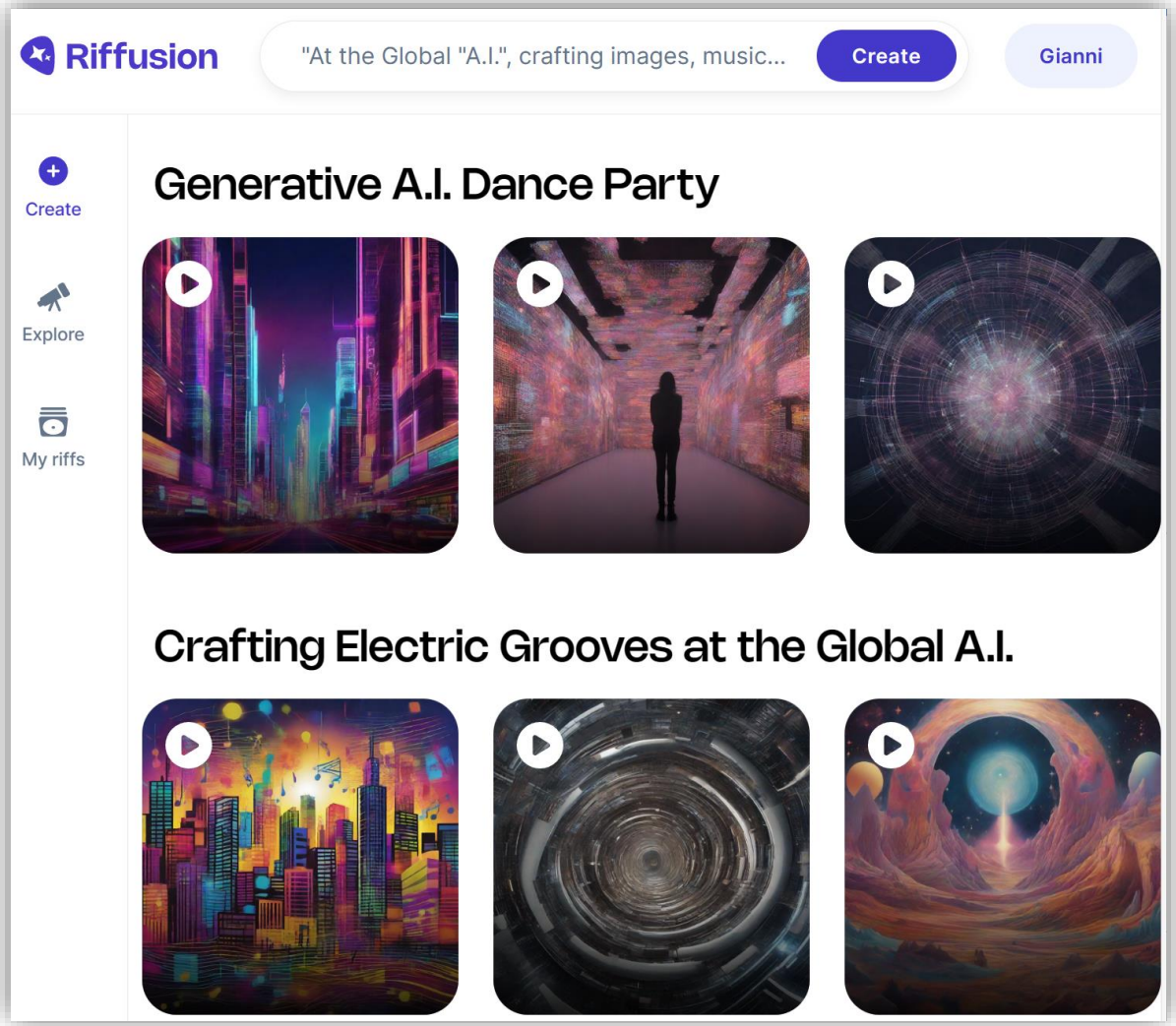
DEMO Generative AI Playground .NET

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

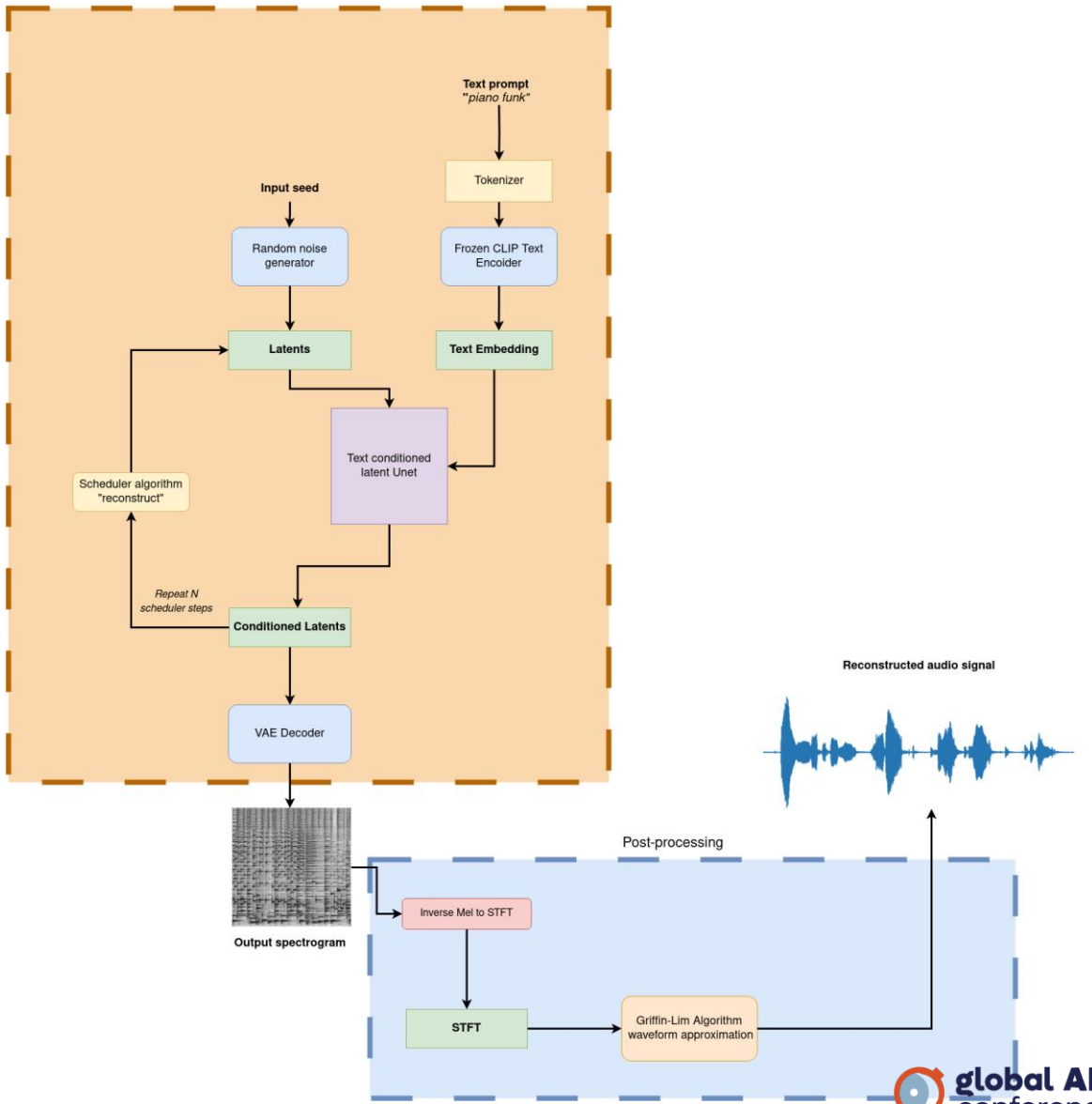


Audio

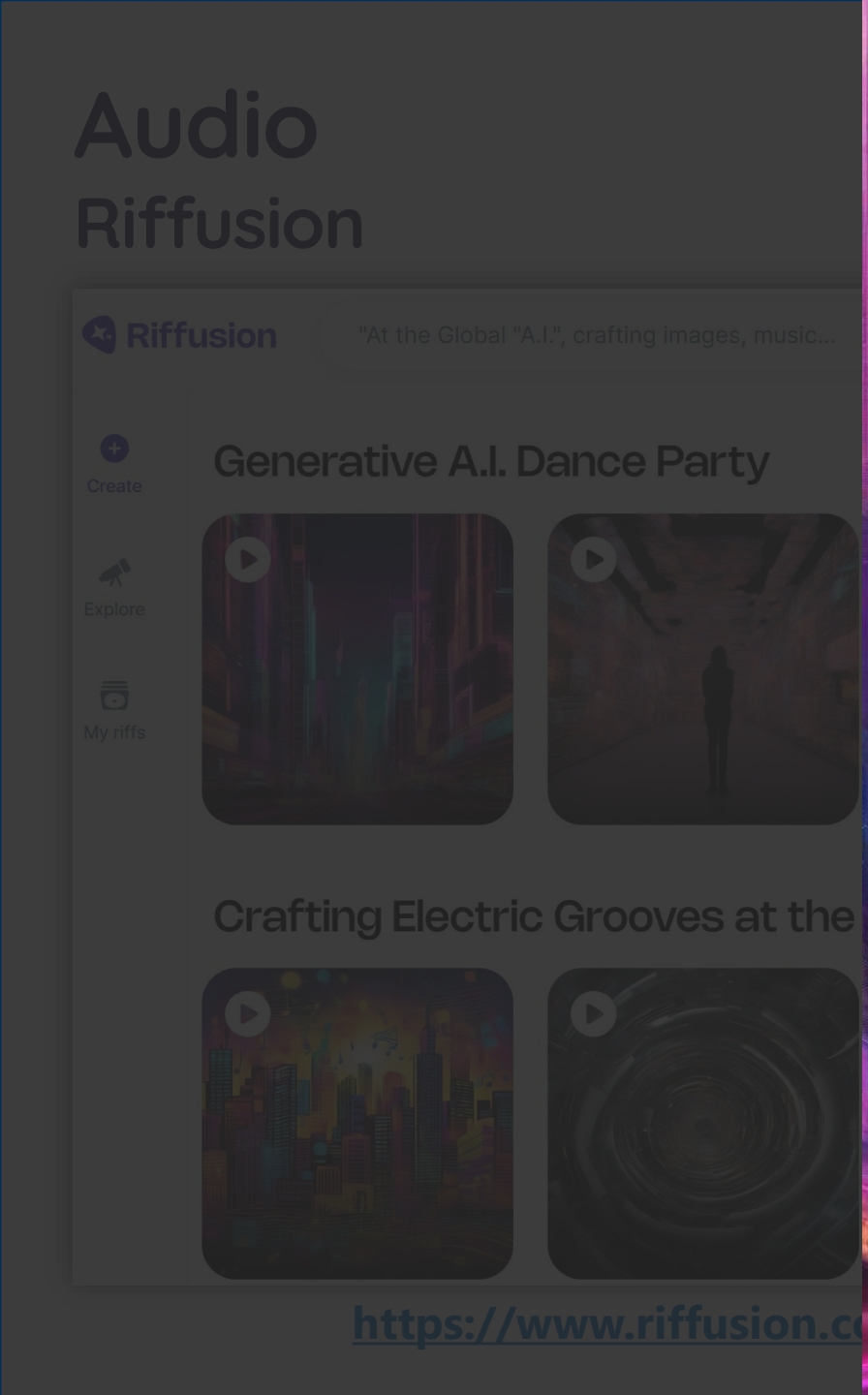
Riffusion



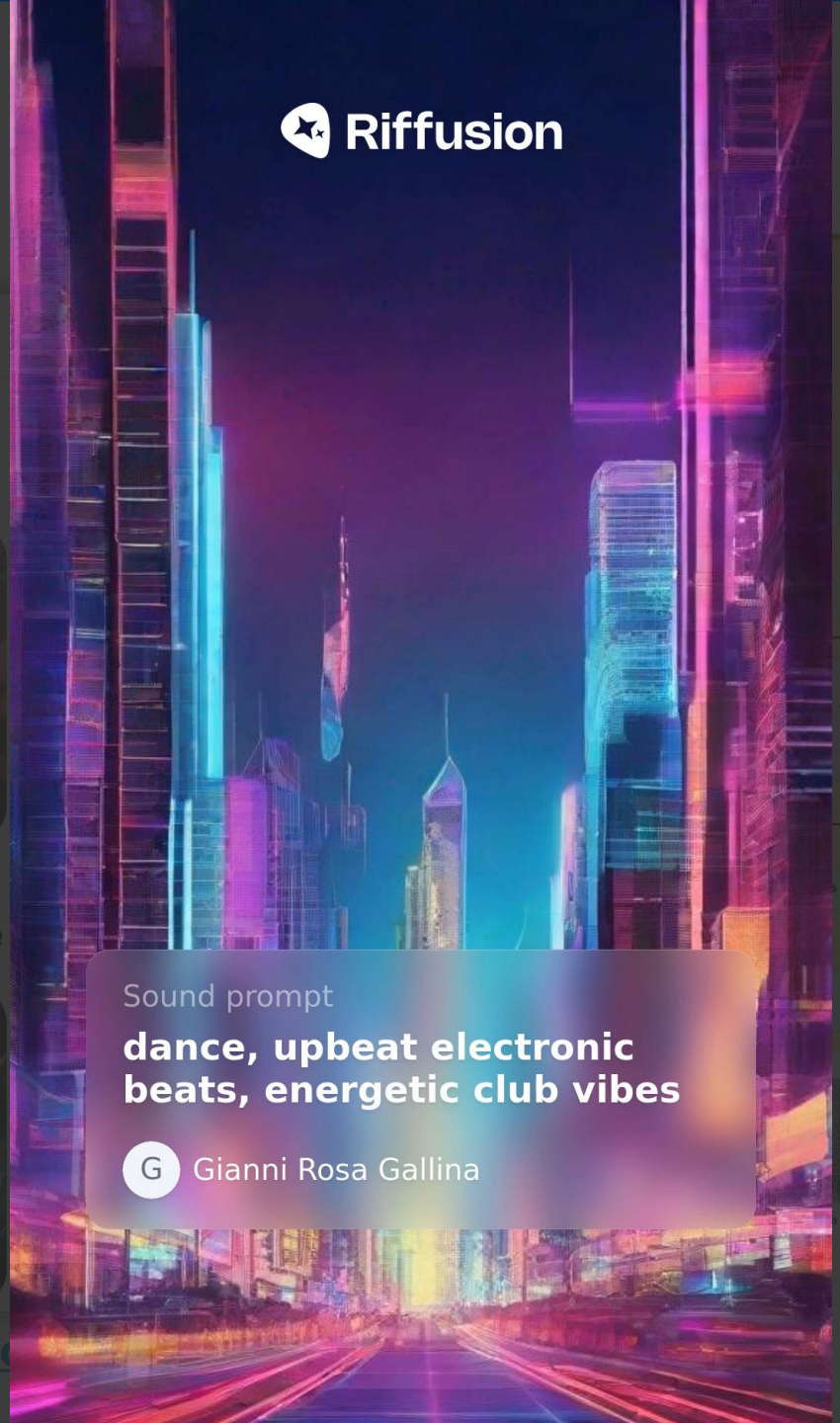
https://www.riffusion.com/

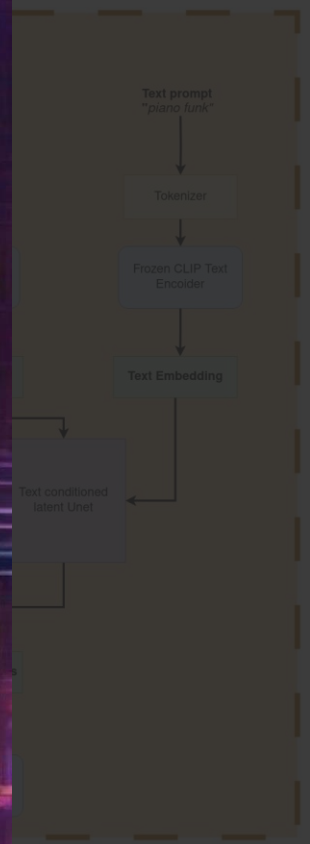


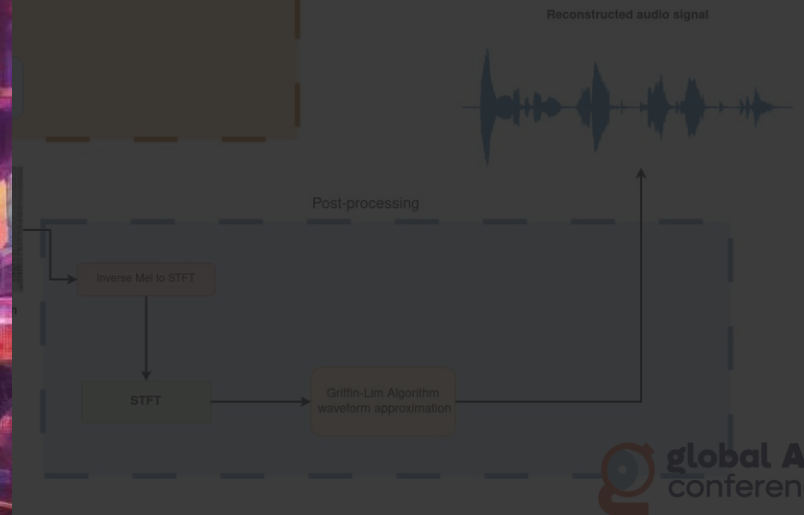
Audio
Riffusion



https://www.riffusion.c



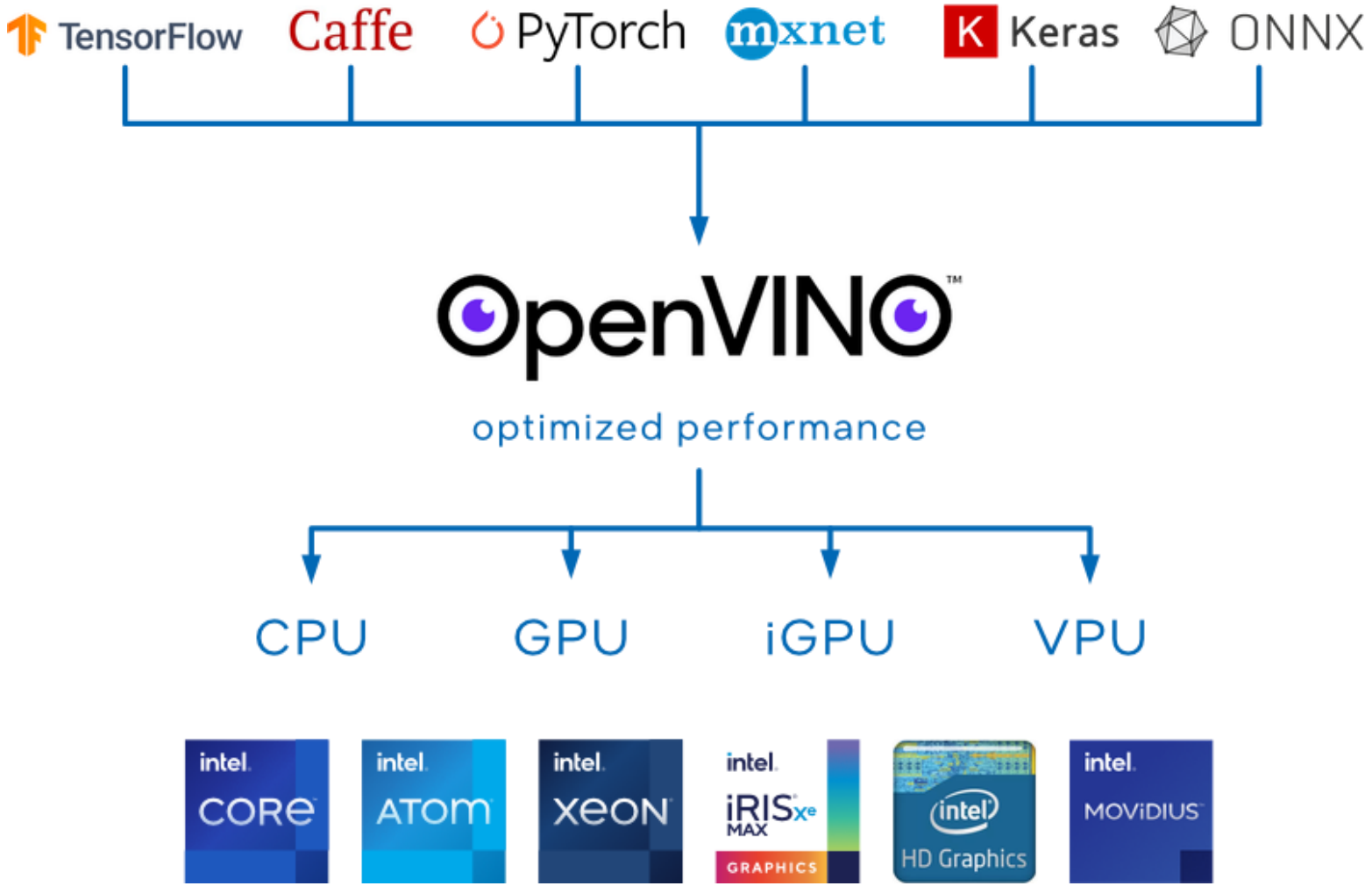




global AI conference

Tools and Frameworks

OpenVINO

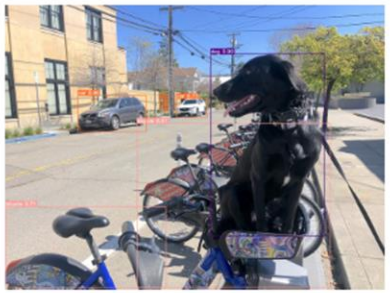


https://docs.openvino.ai/

OpenVINO™
Notebooks

AI Trends - Notebooks

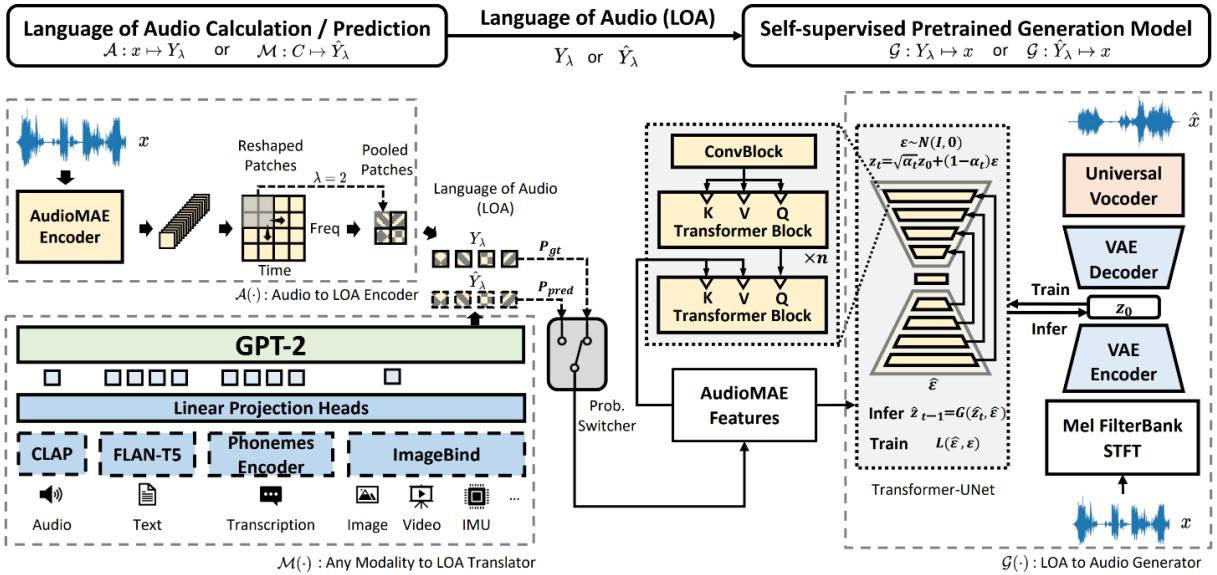
Check out the latest notebooks that show how to optimize and deploy popular models on Intel CPU and GPU.

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

Sound Generation with AudioLDM2 and OpenVINO™

AudioLDM 2 is a latent text-to-audio diffusion model capable of generating realistic audio samples given any text input. AudioLDM 2 was proposed in the paper AudioLDM 2: Learning Holistic Audio Generation with Self-supervised Pretraining by `Haohe Liu` et al. The model takes a text prompt as input and predicts the corresponding audio. It can generate text-conditional sound effects, human speech and music.

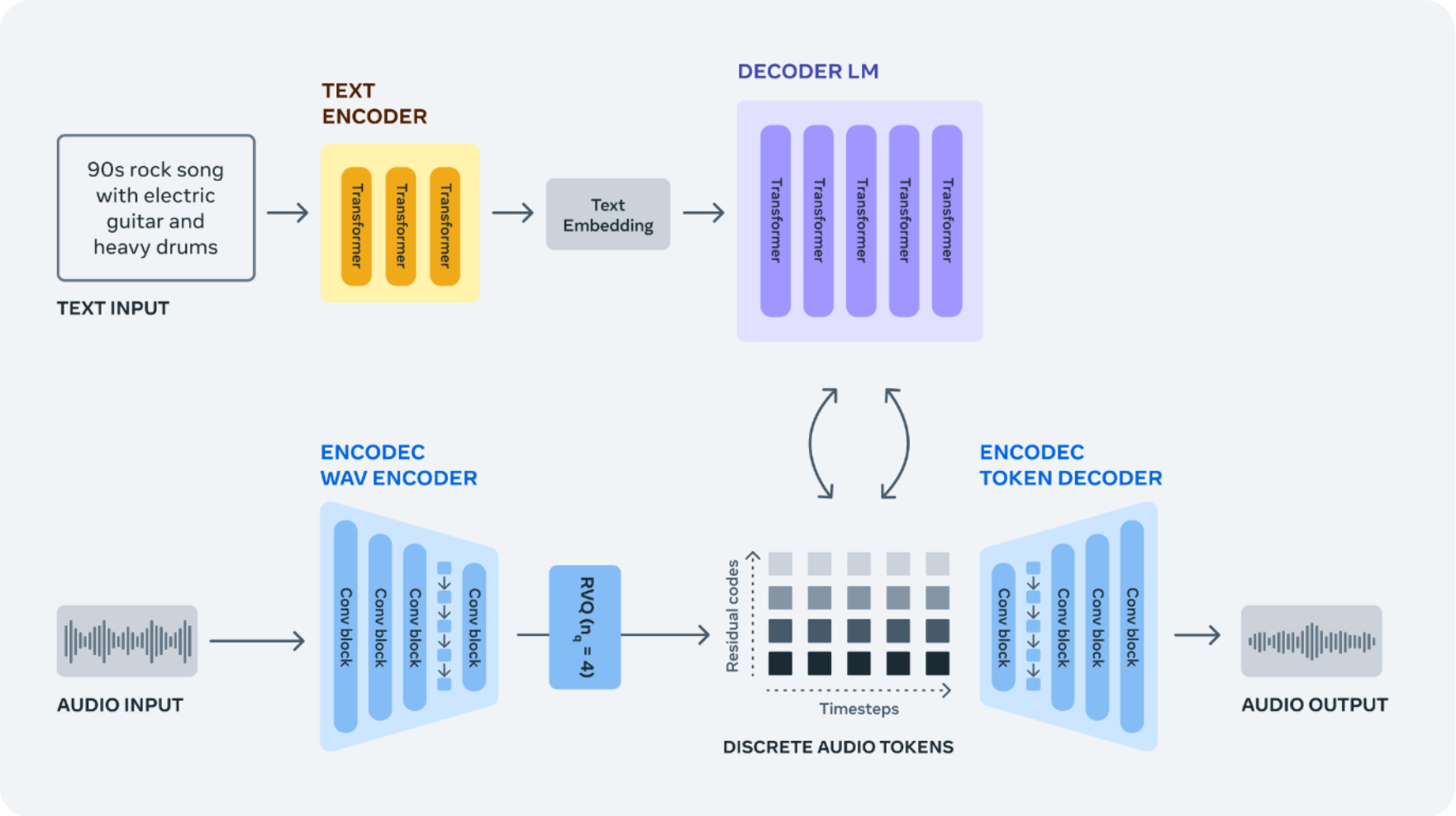
In this tutorial we will try out the pipeline, convert the models backing it one by one and will run an interactive app with Gradio!



https://github.com/openvinotoolkit/openvino_notebooks

Audio

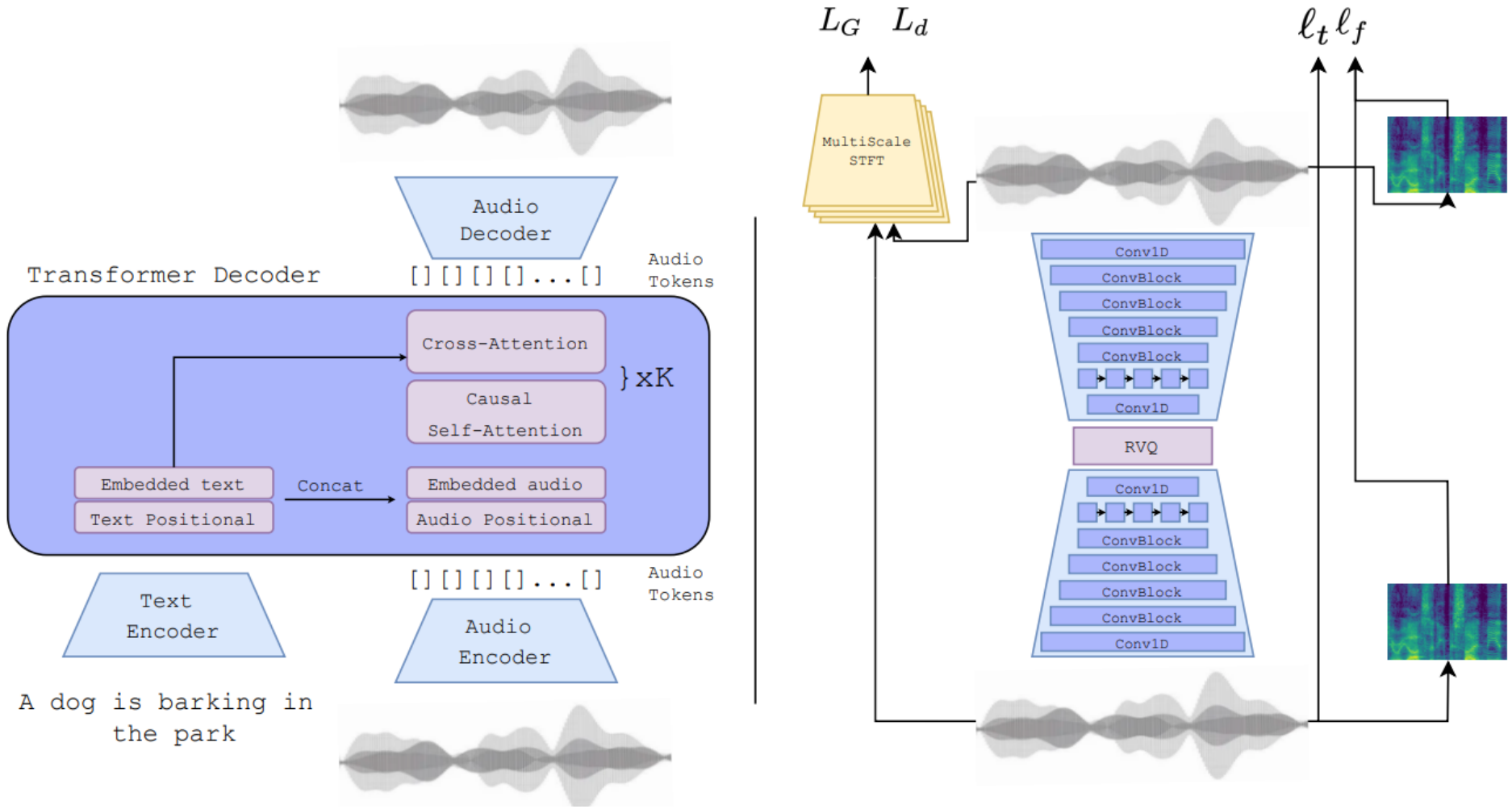
MusicGen



Simple and Controllable Music Generation
https://arxiv.org/pdf/2306.05284.pdf

Audio

AudioGen



AudioGen: Textually Guided Audio Generation

https://arxiv.org/pdf/2209.15352.pdf

Audio

Speech-To-Text: Whisper

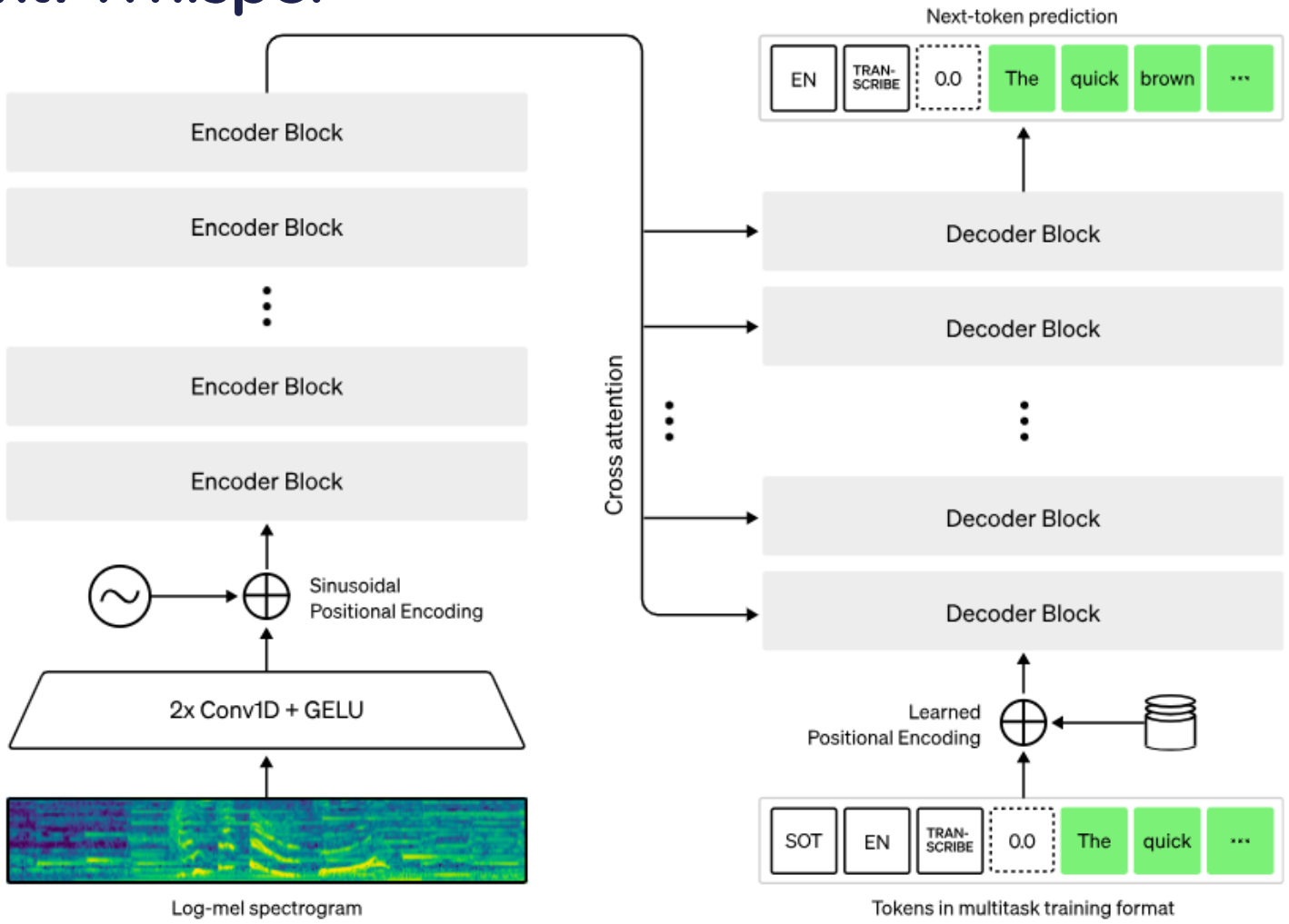
This is the Micro Machine Man presenting the most midget miniature motorcade of Micro Machines. Each one has dramatic details, terrific trim, precision paint jobs, plus incredible Micro Machine Pocket Play Sets. There's a police station, fire station, restaurant, service station, and more. Perfect pocket portables to take any place. And there are many miniature play sets to play with, and each one comes with its own special edition Micro Machine vehicle and fun, fantastic features that miraculously move. Raise the boatlift at the airport marina. Man the gun turret at the army base. Clean your car at the car wash. Raise the toll bridge. And these play sets fit together to form a Micro Machine world. Micro Machine Pocket Play Sets, so tremendously tiny, so perfectly precise, so dazzlingly detailed, you'll want to pocket them all. Micro Machines are Micro Machine Pocket Play Sets sold separately from Galoob. The smaller they are, the better they are.

OpenAI

https://openai.com/research/whisper

Audio

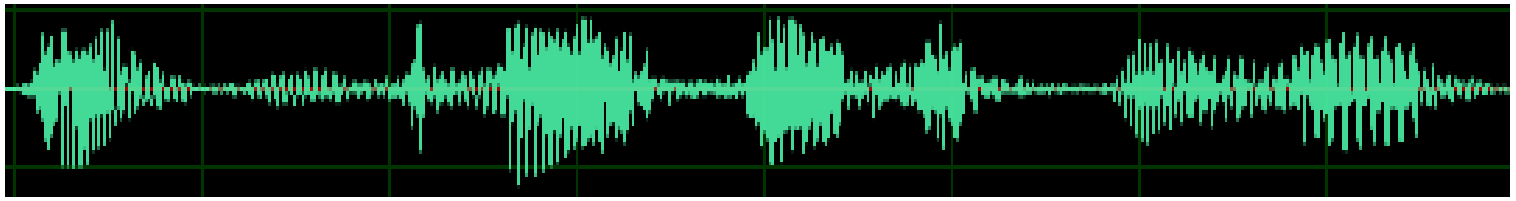
Speech-To-Text: Whisper



Robust Speech Recognition via Large-Scale Weak Supervision
https://arxiv.org/pdf/2212.04356.pdf

Audio

Speech-To-Text: Whisper



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

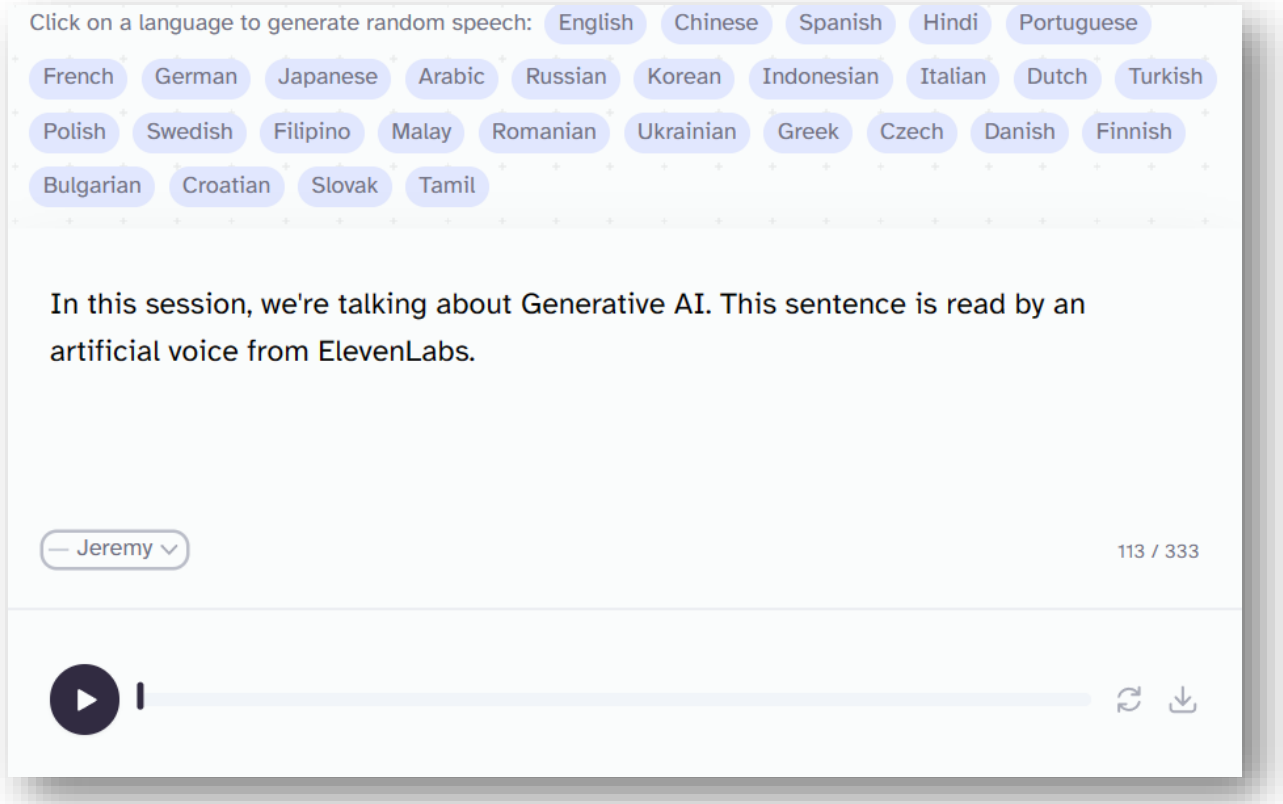
Local Audio Transcription in .NET

https://github.com/ggerganov/whisper.cpp
https://github.com/sandrohanea/whisper.net
https://github.com/gianni-rg/gen-ai-net-playground

Audio

Text-To-Speech

llElevenLabs – Generative Voice AI



https://elevenlabs.io/

OpenAI TTS

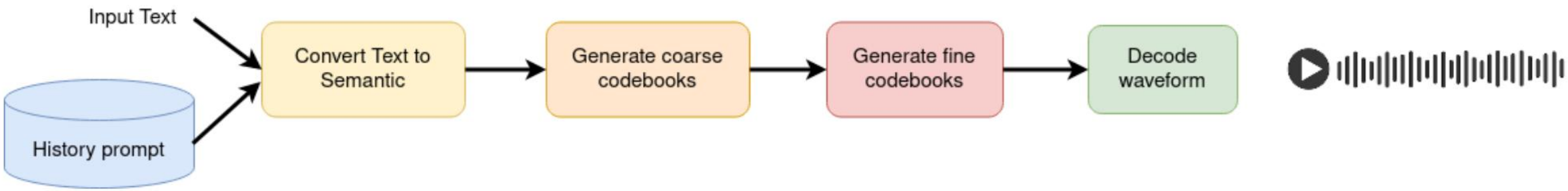
https://platform.openai.com/docs/guides/text-to-speech

Microsoft Text to Speech

https://azure.microsoft.com/en-us/products/ai-services/text-to-speech

Audio

Text-To-Speech: Bark



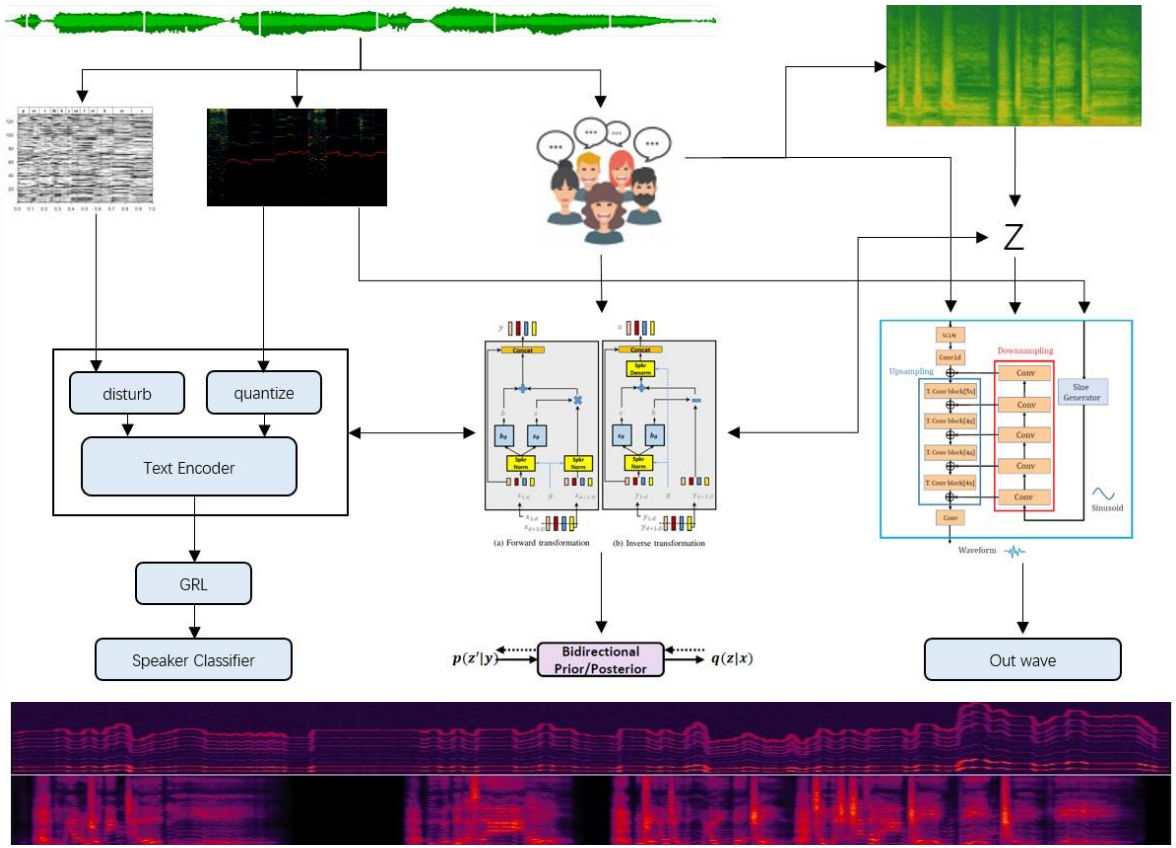
https://github.com/suno-ai/bark

Audio

Audio-To-Audio

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

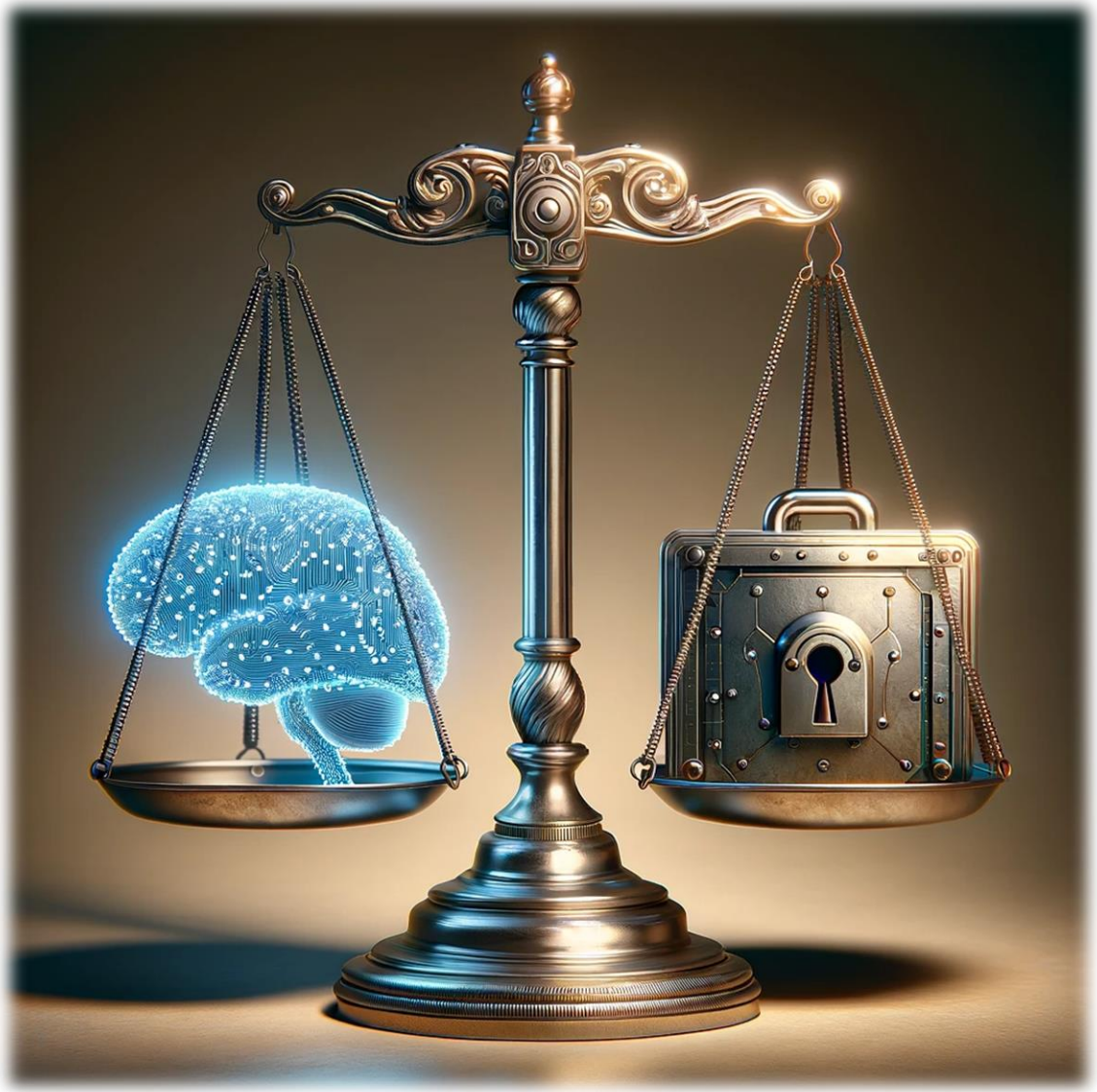
https://github.com/PlayVoice/so-vits-svc-5.0



Recap

- Generative AI overview
- Stable Diffusion
- OpenVINO & Notebooks
- Riffusion, MusicGen, AudioGen
- Whisper
- Text-To-Speech
- Audio-To-Audio

Underrated Topics



Ethical AI
Model Security
Data Privacy

"HDR photo of a balance scale with a brain on one side representing AI and machine learning, while the other side showcases a shield and lock indicating security [...]"

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

“Simpson character in front of enthusiastic Software Developers. In a speech bubble saying ‘Any questions?’ ”

References (1/2)

- https://globalai.community/
- 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/dall-e-3
- https://openai.com/research/whisper
- https://elevenlabs.io/
- https://github.com/steven2358/awesome-generative-ai
- https://github.com/imaurer/awesome-decentralized-llm
- https://onnx.ai/
- https://docs.openvino.ai/
- https://github.com/openvinotoolkit/openvino_notebooks
- https://github.com/suno-ai/bark
- https://platform.openai.com/docs/guides/text-to-speech
- https://azure.microsoft.com/en-us/products/ai-services/text-to-speech

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/comfyanonymous/ComfyUI
- https://github.com/gianni-rg/gen-ai-net-playground
- https://github.com/openai/whisper
- https://github.com/ggerganov/whisper.cpp
- https://github.com/sandrohanea/whisper.net
- https://whisper.ggerganov.com/talk/

About Us

Clemente GIORIO

R&D Senior Principal Engineer EYES ON

@tinux80

- Augmented/Mixed/Virtual Reality
- Artificial Intelligence, Machine Learning, Deep Learning
- Computer Vision, Multimodal Tracking
- Internet of Things
- Hybrid Clusters

dotNET{podcast}

MVP Microsoft Most Valuable Professional

intel Software INNOVATOR

[PACKT] PUBLISHING Author

FAB LAB NAPOLI

NVIDIA Certified Associate - AI in the Data Center

About Us

Ing. Gianni ROSA GALLINA

R&D Technical Lead @ deltatre

𝕏@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

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/

global AI conference