



# global AI conference

December 12<sup>th</sup> 2023

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

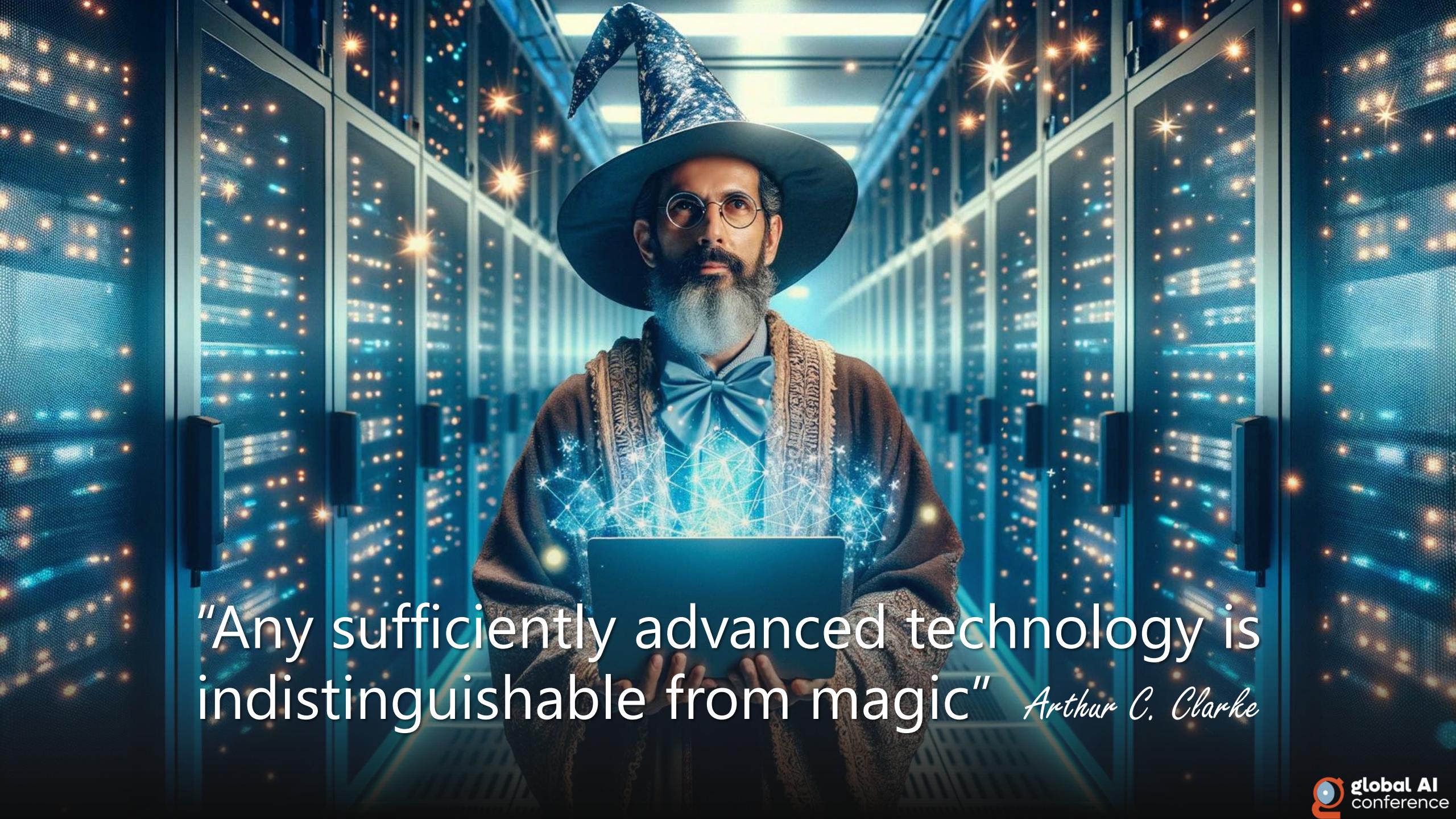


Clemente Giorio  
**EYES ON**



Gianni Rosa Gallina  
**deltatre**





“Any sufficiently advanced technology is  
indistinguishable from magic” *Arthur C. Clarke*



"Standing on the Shoulders of the Giants"

# Generative AI

## Overview

**Text**



**Images & Videos**



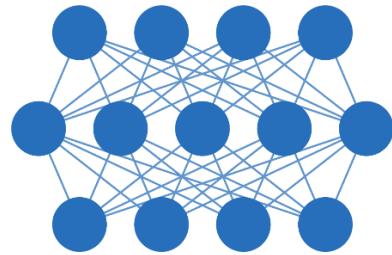
**Speech & Music**



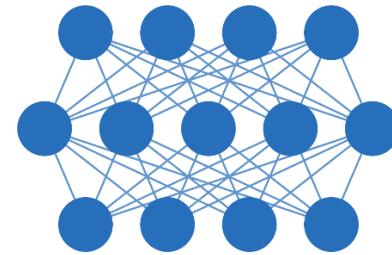
**Structured Data**



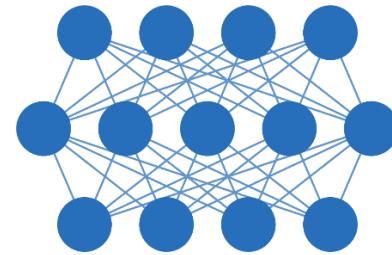
**3D Signals**



...



...

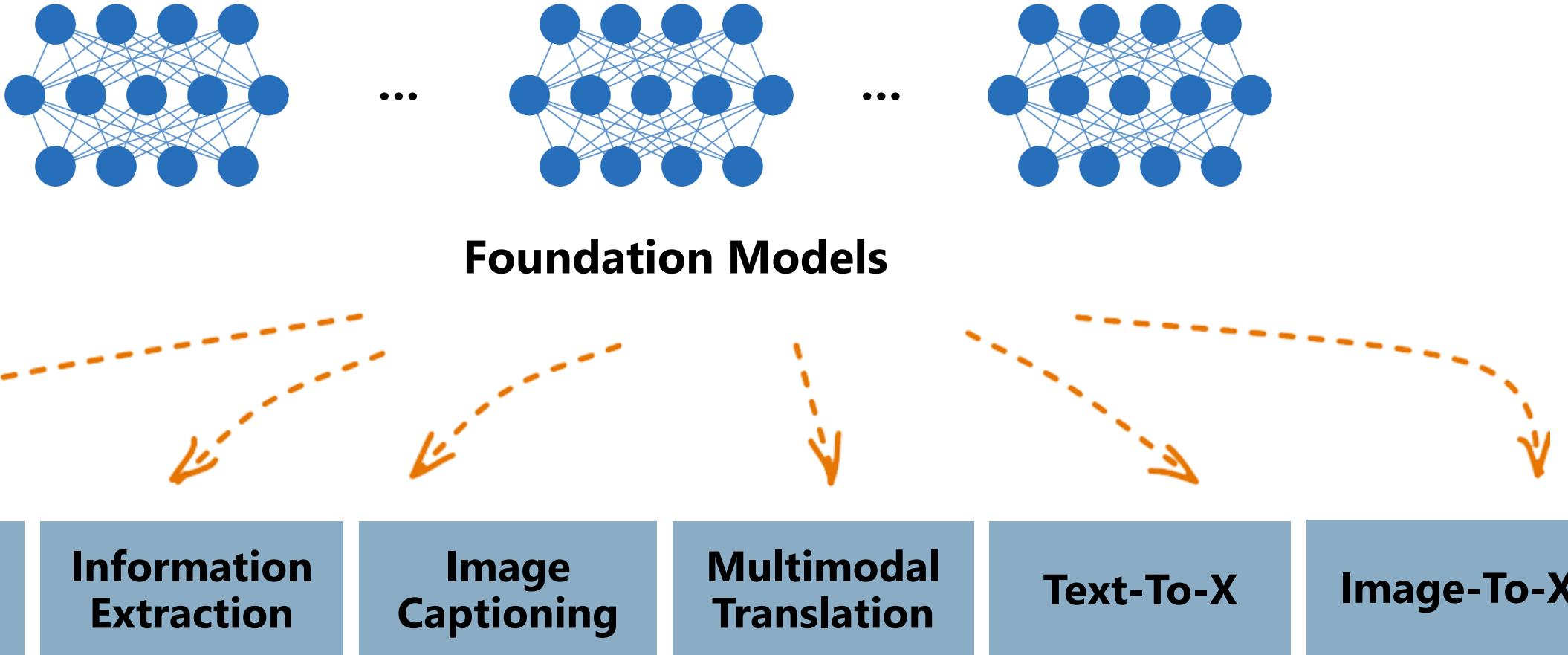


**Foundation Models**



# Generative AI

## Overview



# 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

ChatGPT

Examples

- "Explain quantum computing in simple terms" →
- "Got any creative ideas for a 10 year old's birthday?" →
- "How do I make an HTTP request in Javascript?" →

Capabilities

- Remembers what user said earlier in the conversation
- Allows user to provide follow-up corrections
- Trained to decline inappropriate requests

Limitations

- May occasionally provide incorrect information
- May occasionally generate harmful instructions
- Limited knowledge of current events

Dark mode

OpenAI Discord

Updates & FAQ

Log out

ChatGPT Jan 30 Version. Our goal is to make AI systems more natural and safe to interact with.

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



Cognitive Services | Azure OpenAI Studio

Azure OpenAI Studio > Chat playground (Preview)

Chat playground (Preview)

Assistant setup

Load example setup

Save changes

System message

You are a Technical Speaker in an Artificial Intelligence event that is explaining what is "Generative AI".

Few-shot examples

Chat session

Clear chat View code Show raw JSON

Write me the script for an introductory slide on Text generated by a Generative AI model

Ladies and gentlemen, esteemed colleagues, and fellow AI enthusiasts, welcome to our discussion on the fascinating world of Generative AI.

Today, we are about to embark on a journey through the incredible capabilities of AI models that can generate human-like text. These models are pushing the boundaries of what we once thought possible and are rapidly transforming the way we interact with technology.

Imagine a world where AI-generated text is so sophisticated and contextually accurate that it becomes virtually indistinguishable from human-written content. This vision is now closer to reality, thanks to the advancements in Generative AI.

On this introductory slide, we will delve into the inner workings of these AI models, explore the techniques that empower them to craft coherent and contextually relevant text, and discuss their potential applications and impact on various industries.

As we proceed, we will uncover the true potential of Generative AI in revolutionizing communication, content creation, and the way we interact with intelligent systems. So, buckle up and get ready to witness the future of AI-generated text!

User message

Type user query here. (Shift + Enter for new line)

Parameters

Deployments

Max response

Temperature

Top P

Stop sequence

Session settings

Past messages included

Current token count

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

# 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** 

[https://huggingface.co/blog/stable\\_diffusion](https://huggingface.co/blog/stable_diffusion)  
<https://stability.ai/stable-diffusion>  
<https://github.com/huggingface/diffusers>  
<https://runwayml.com/>



**Hugging Face**



**Diffusers**

# Generative AI

## Text-To-Video

 runway



<https://runwayml.com/>

# Generative AI

## Text-To-Video

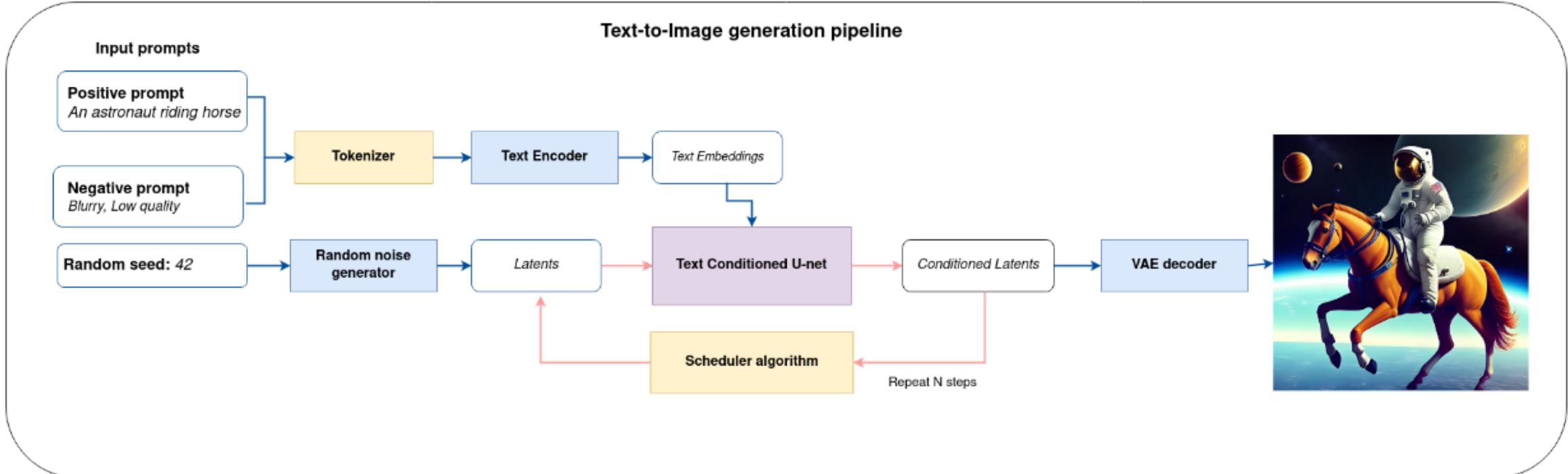
stability.ai



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

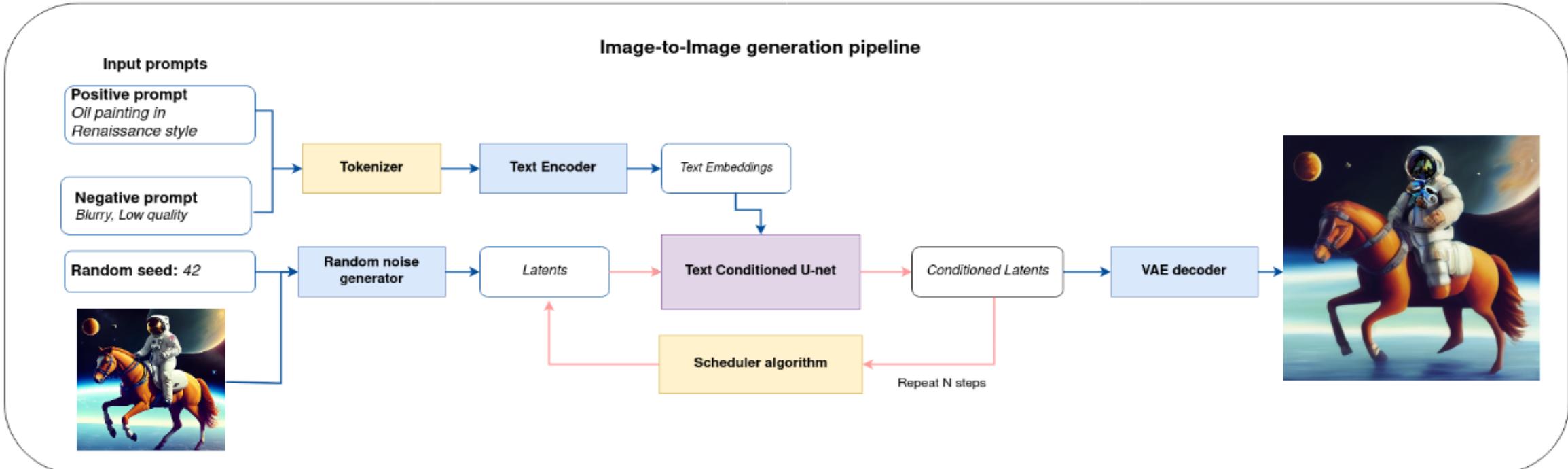
# Images

## Stable Diffusion



# Images

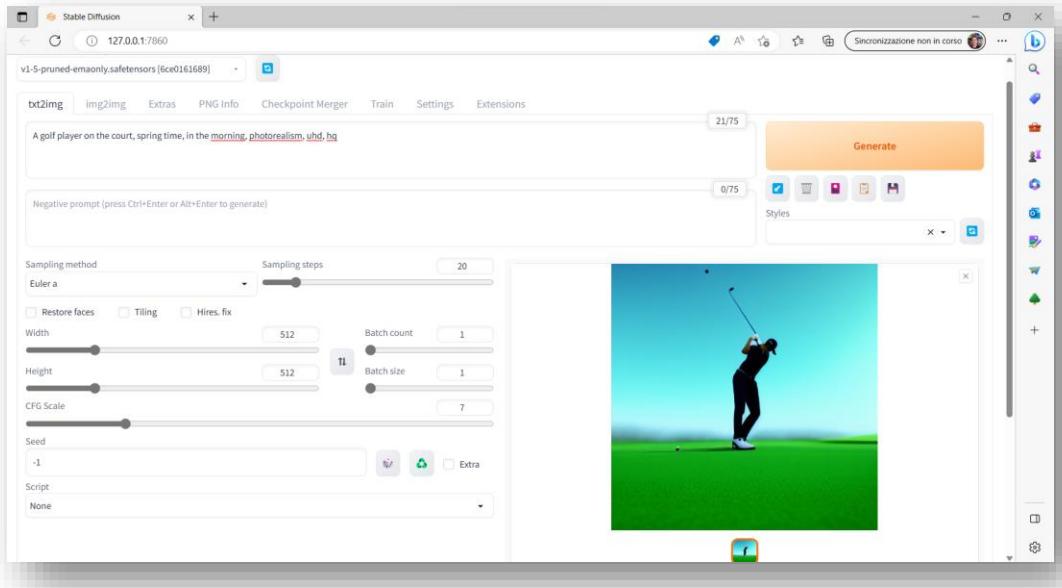
## Stable Diffusion



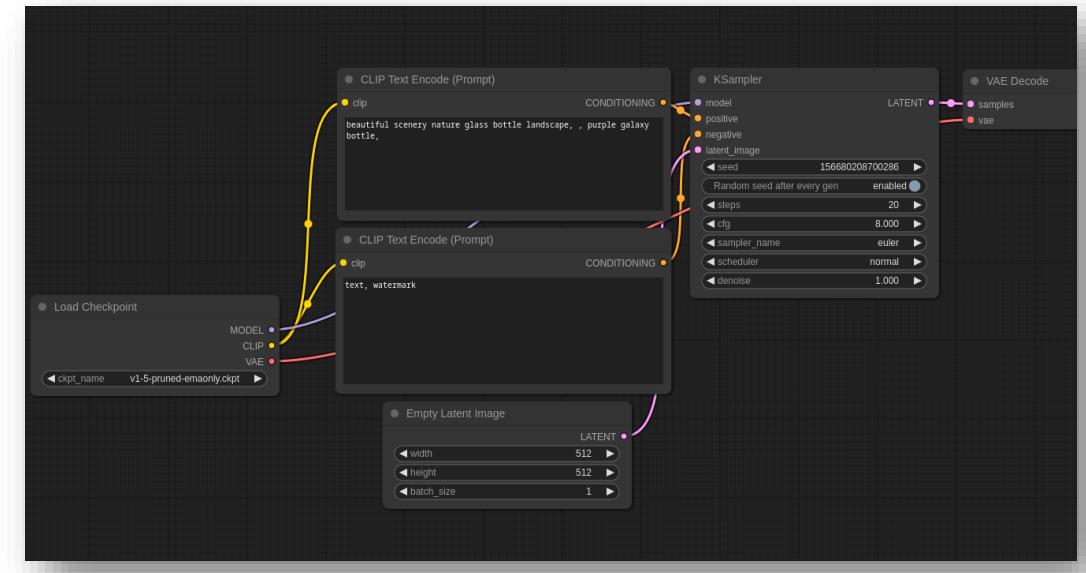
<https://stability.ai/>

# Images

## Stable Diffusion - Web UI Tool



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



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

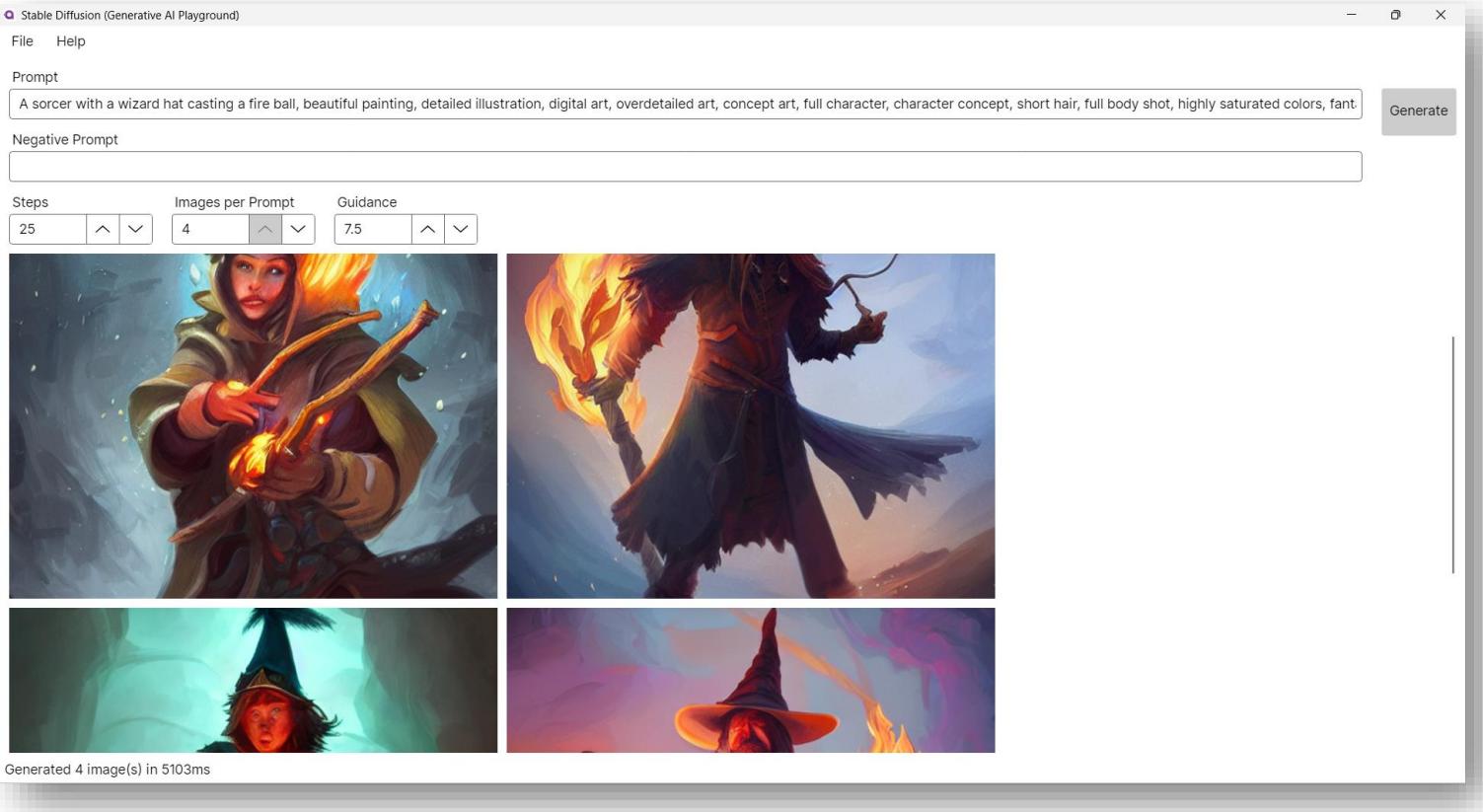
# Images

## 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>



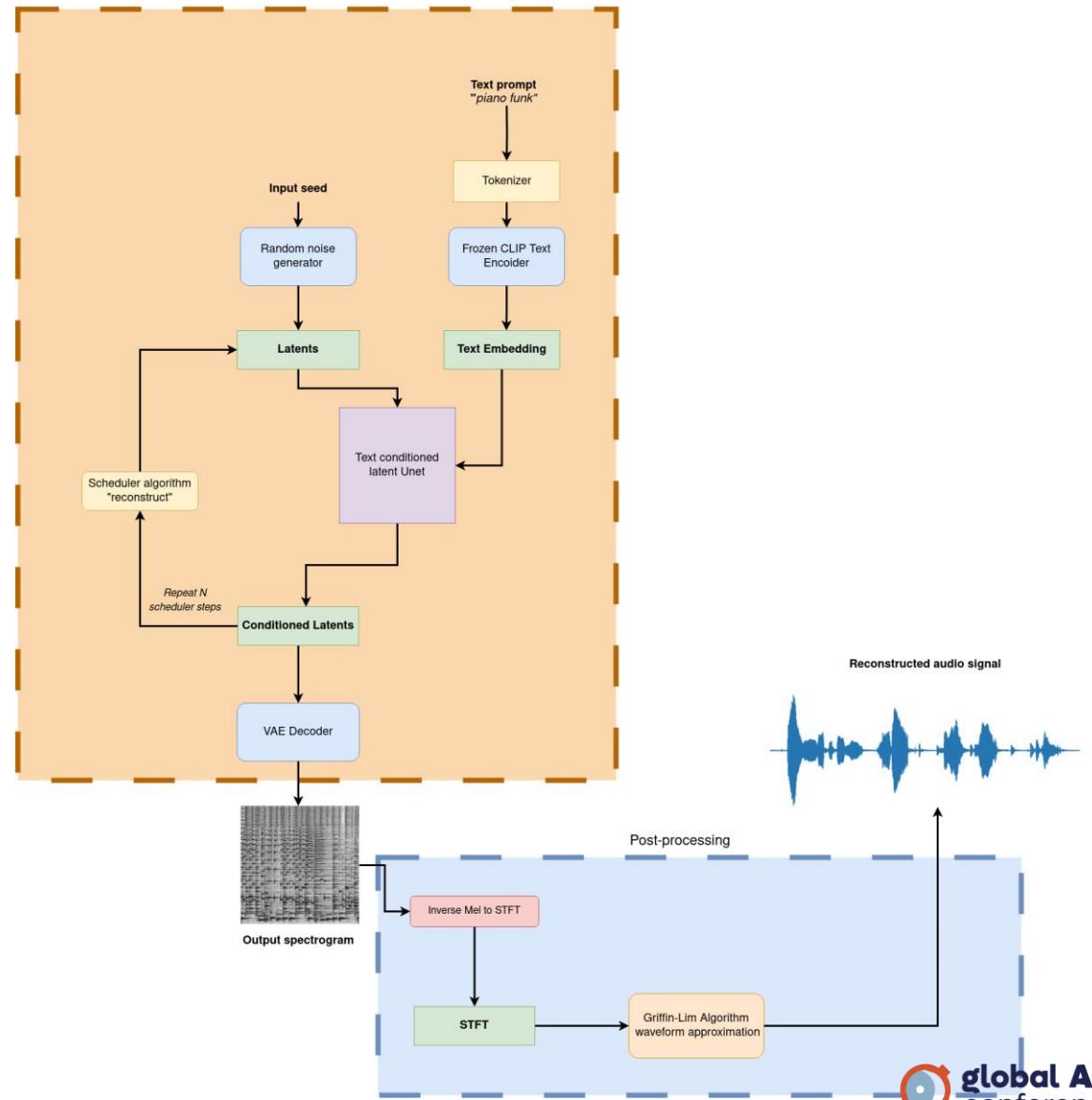
# Audio Riffusion

**Riffusion** "At the Global "A.I.", crafting images, music... Create Gianni

**Generate A.I. Dance Party**

**Crafting Electric Grooves at the Global A.I.**

<https://www.riffusion.com/>



# Audio Riffusion

**Riffusion** "At the Global "A.I.", crafting images, music...

**Create**

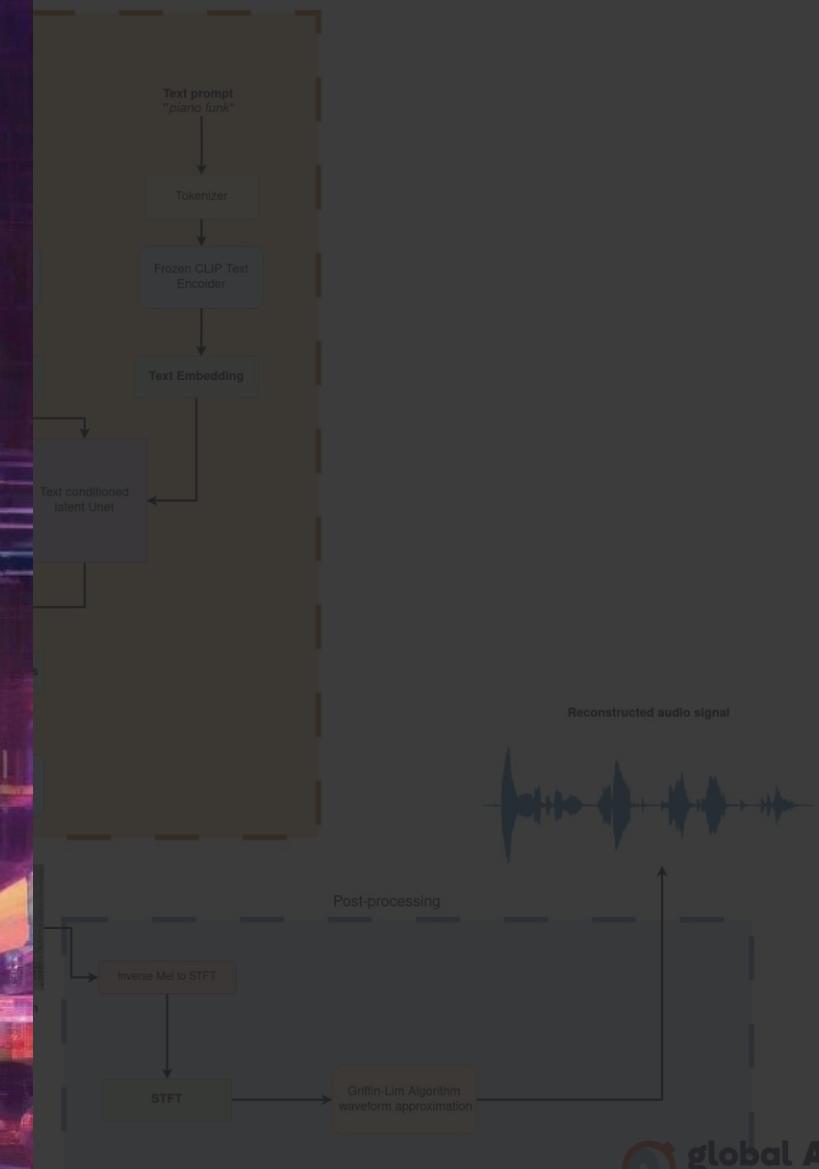
**Explore**

**My riffs**

**Generative A.I. Dance Party**

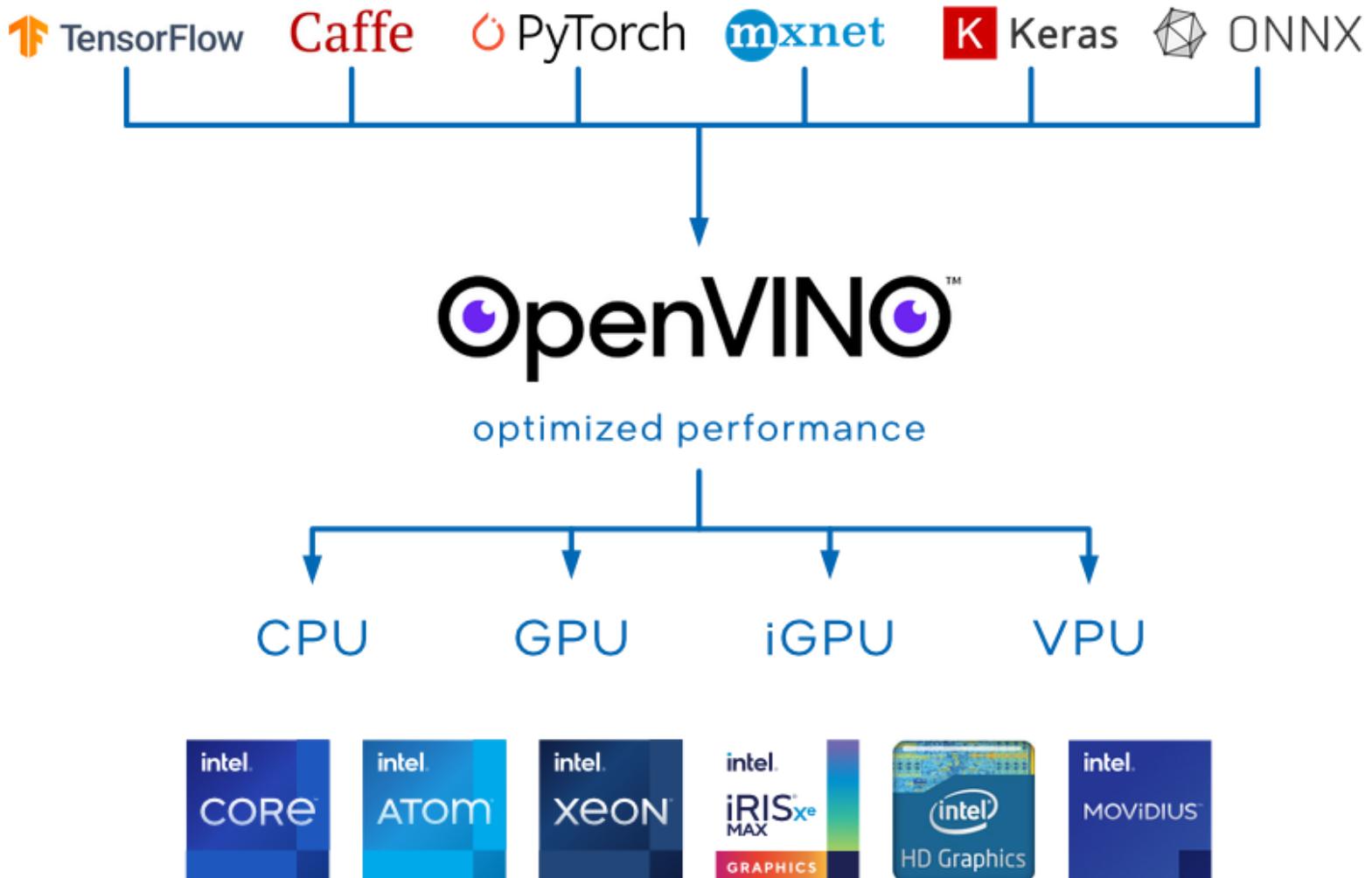
**Crafting Electric Grooves at the Global AI Conference**

<https://www.riffusion.com>



# Tools and Frameworks

## OpenVINO



<https://docs.openvino.ai/>

# OpenVINO™ Notebooks

## AI Trends - Notebooks

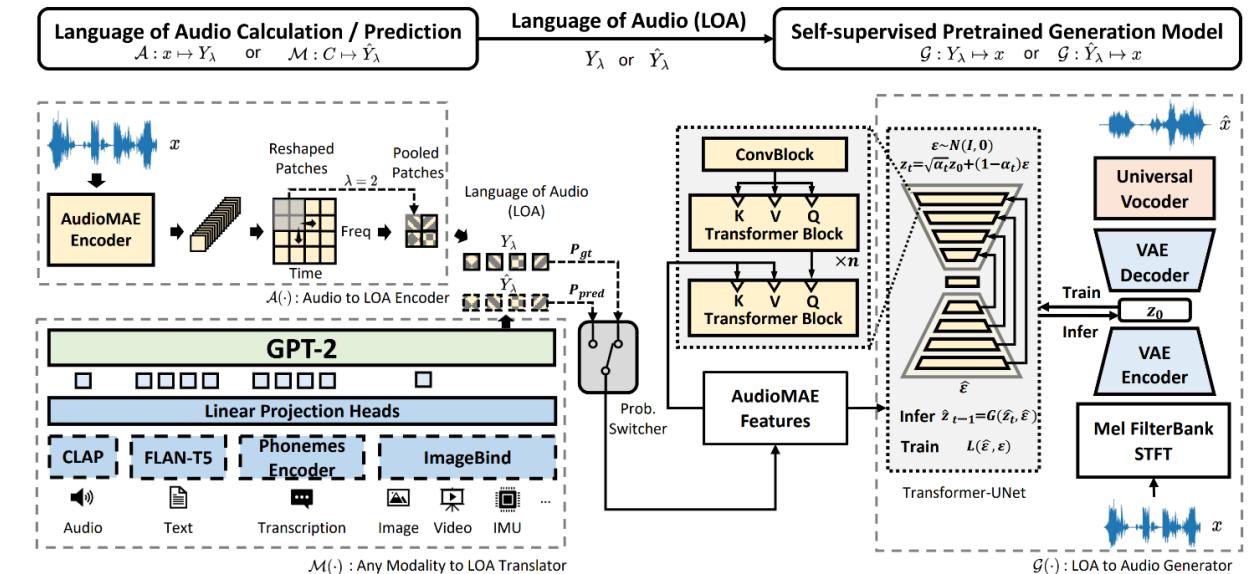
Check out the latest notebooks that show how to optimize and deploy popular models on Intel CPU and GPU.

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

## Sound Generation with AudioLDM2 and OpenVINO™

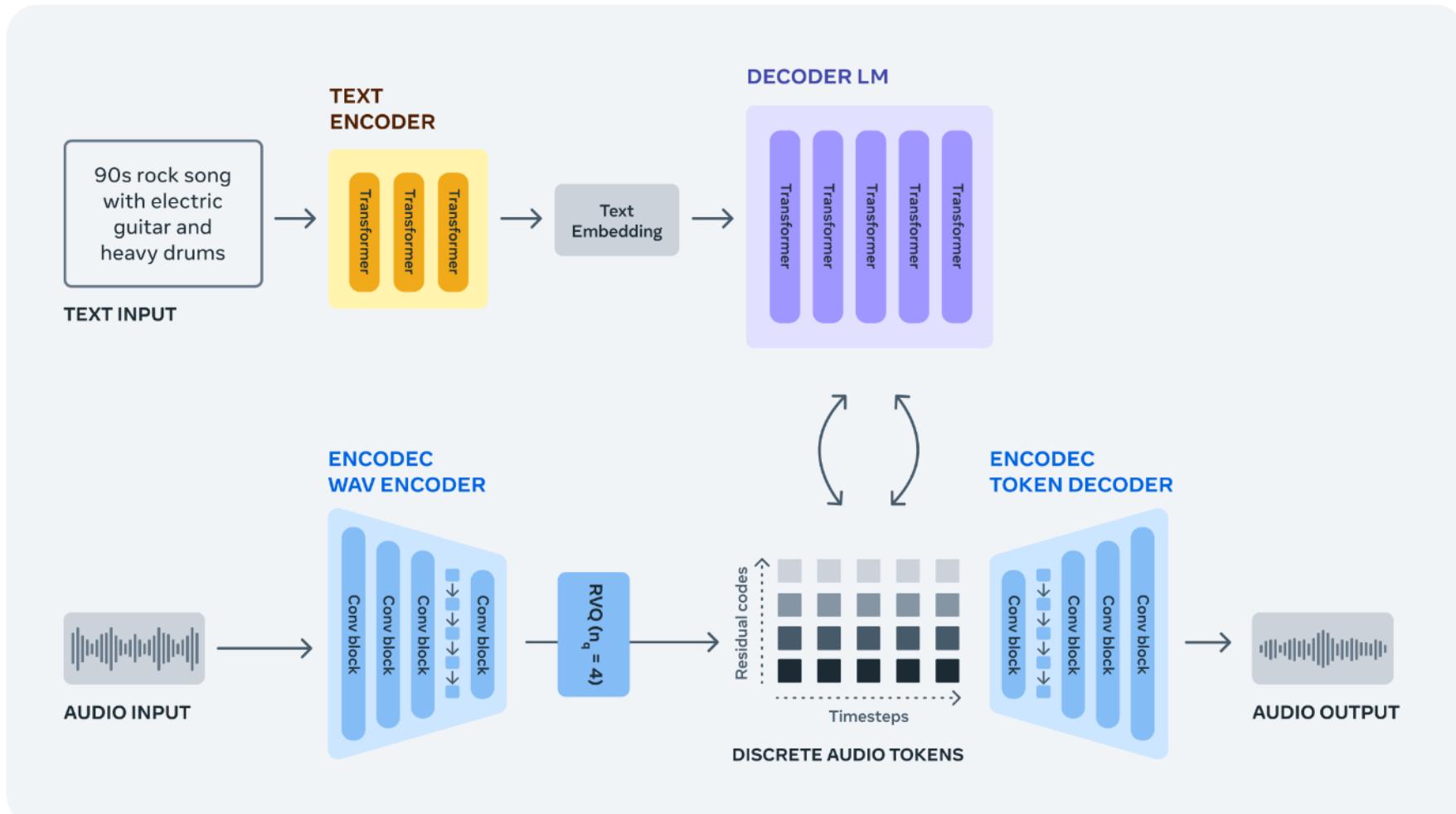
[AudioLDM 2](#) is a latent text-to-audio diffusion model capable of generating realistic audio samples given any text input. AudioLDM 2 was proposed in the paper [AudioLDM 2: Learning Holistic Audio Generation with Self-supervised Pretraining](#) by Haohe Liu et al. The model takes a text prompt as 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](https://github.com/openvinotoolkit/openvino_notebooks)

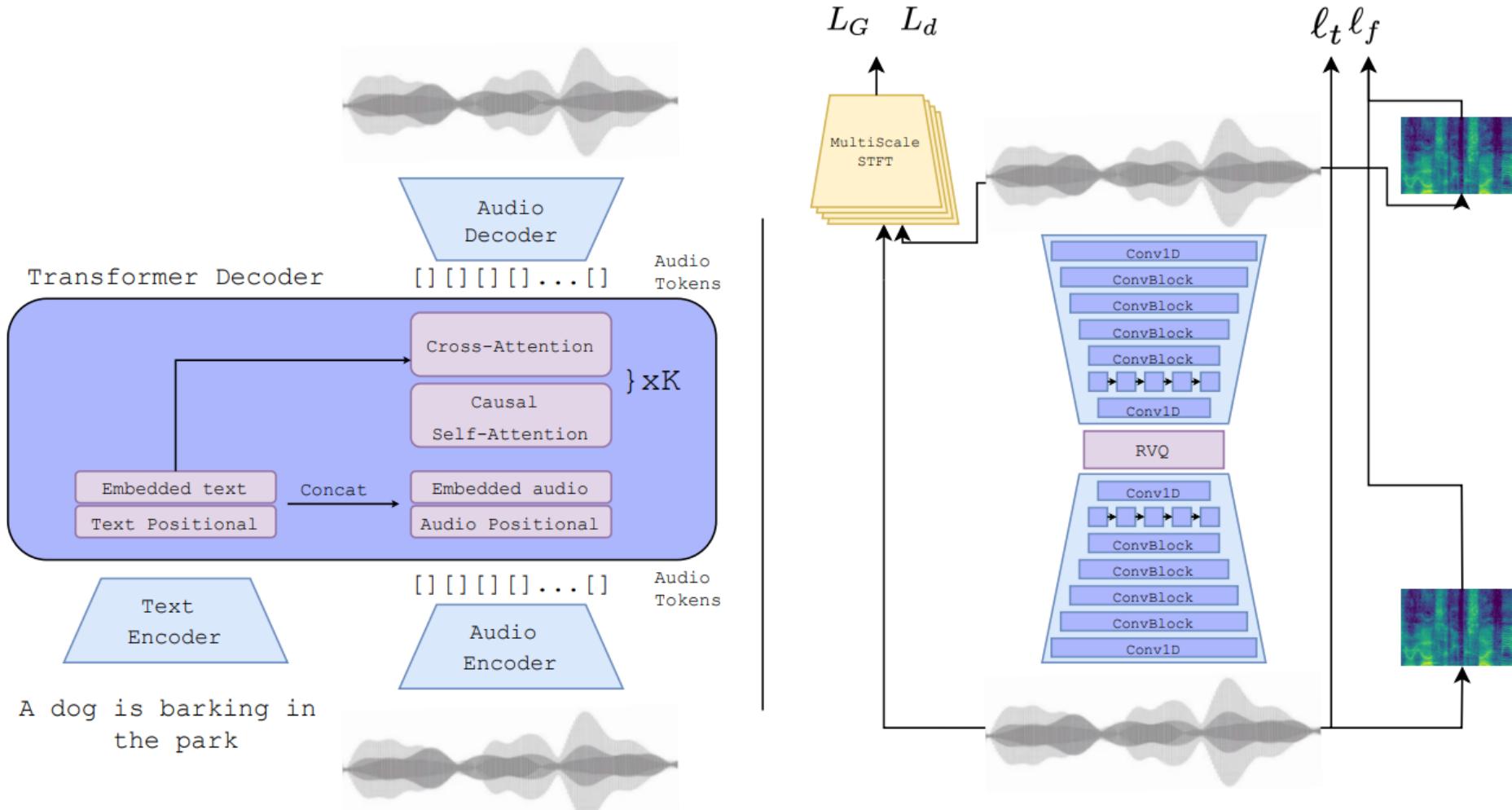
# Audio MusicGen



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

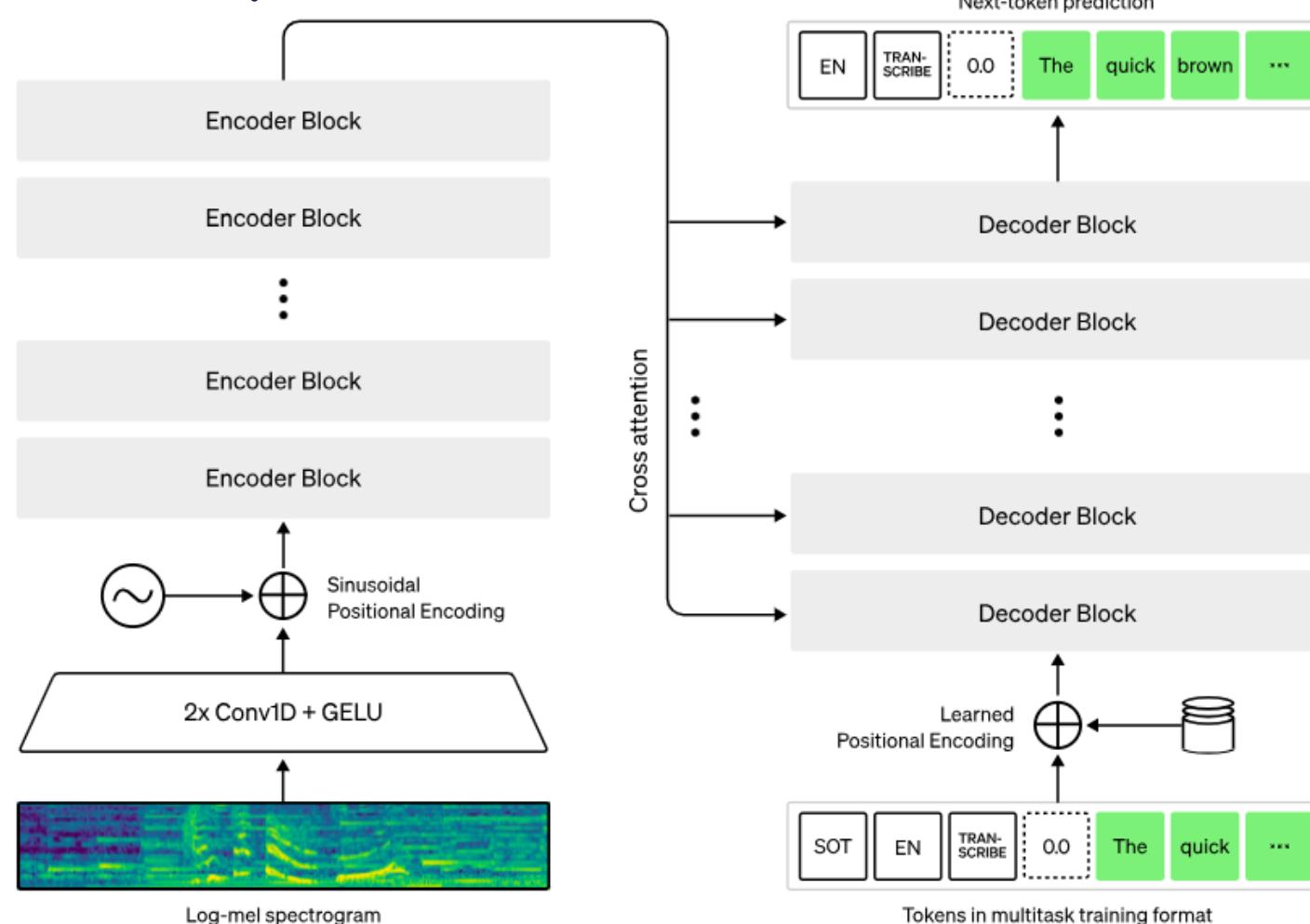
## Speech-To-Text: Whisper



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 anywhere. 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.

# 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	<code>tiny.en</code>	<code>tiny</code>	~1 GB	~32x
base	74 M	<code>base.en</code>	<code>base</code>	~1 GB	~16x
small	244 M	<code>small.en</code>	<code>small</code>	~2 GB	~6x
medium	769 M	<code>medium.en</code>	<code>medium</code>	~5 GB	~2x
large	1550 M	N/A	<code>large</code>	~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

### ElevenLabs – Generative Voice AI

Click on a language to generate random speech:

- English
- Chinese
- Spanish
- Hindi
- Portuguese
- French
- German
- Japanese
- Arabic
- Russian
- Korean
- Indonesian
- Italian
- Dutch
- Turkish
- Polish
- Swedish
- Filipino
- Malay
- Romanian
- Ukrainian
- Greek
- Czech
- Danish
- Finnish
- Bulgarian
- Croatian
- Slovak
- Tamil

In this session, we're talking about Generative AI. This sentence is read by an artificial voice from ElevenLabs.

— Jeremy ▾

113 / 333

▶ | ↻ ⏪

<https://elevenlabs.io/>

### OpenAI TTS

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

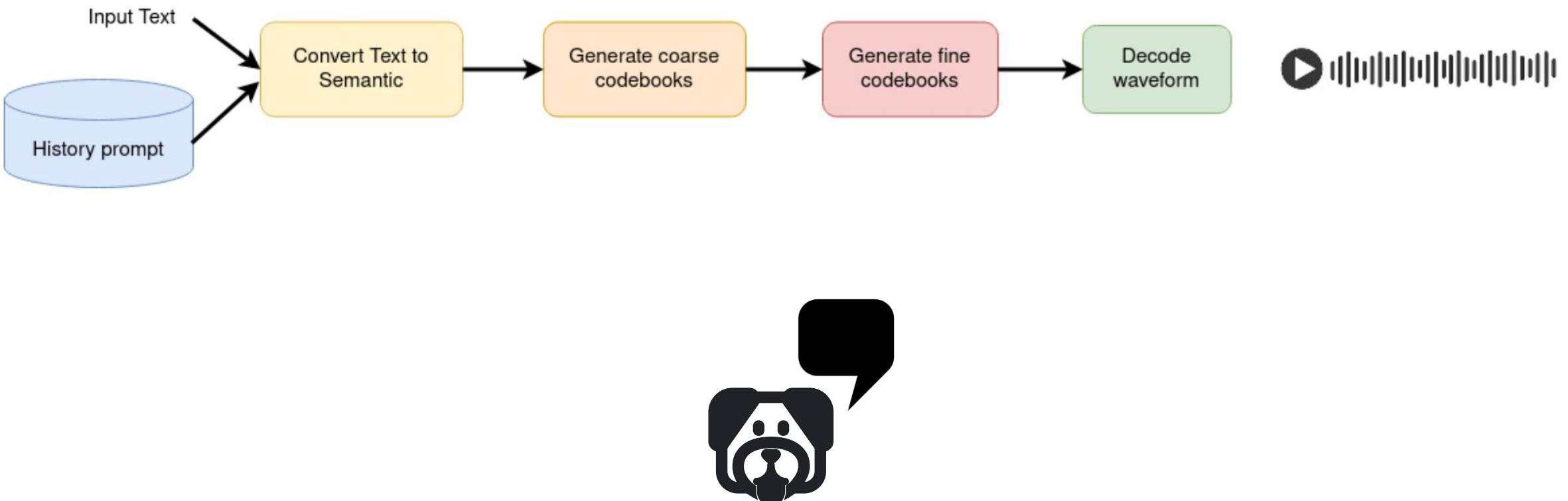
### Microsoft Text to Speech



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

# Audio

## Text-To-Speech: Bark

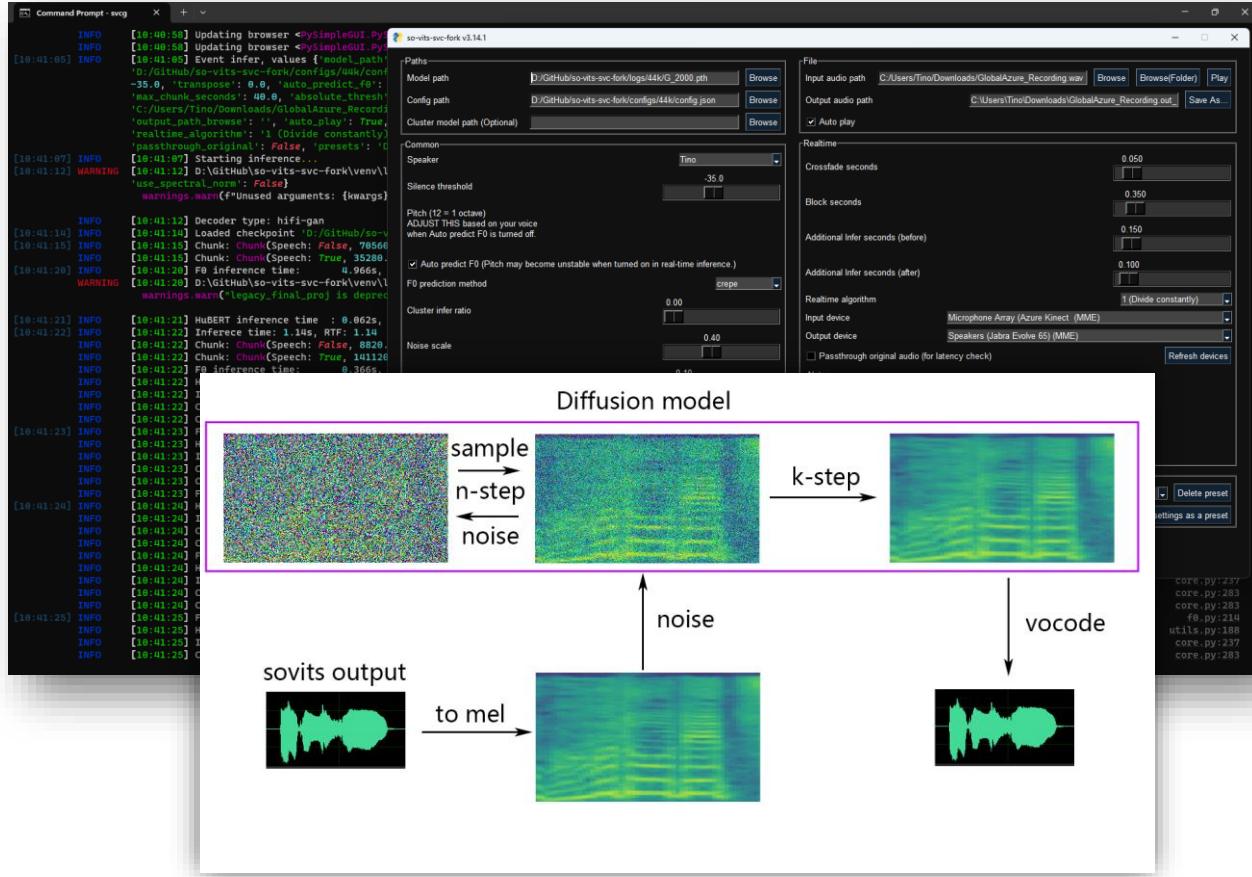


<https://github.com/suno-ai/bark>

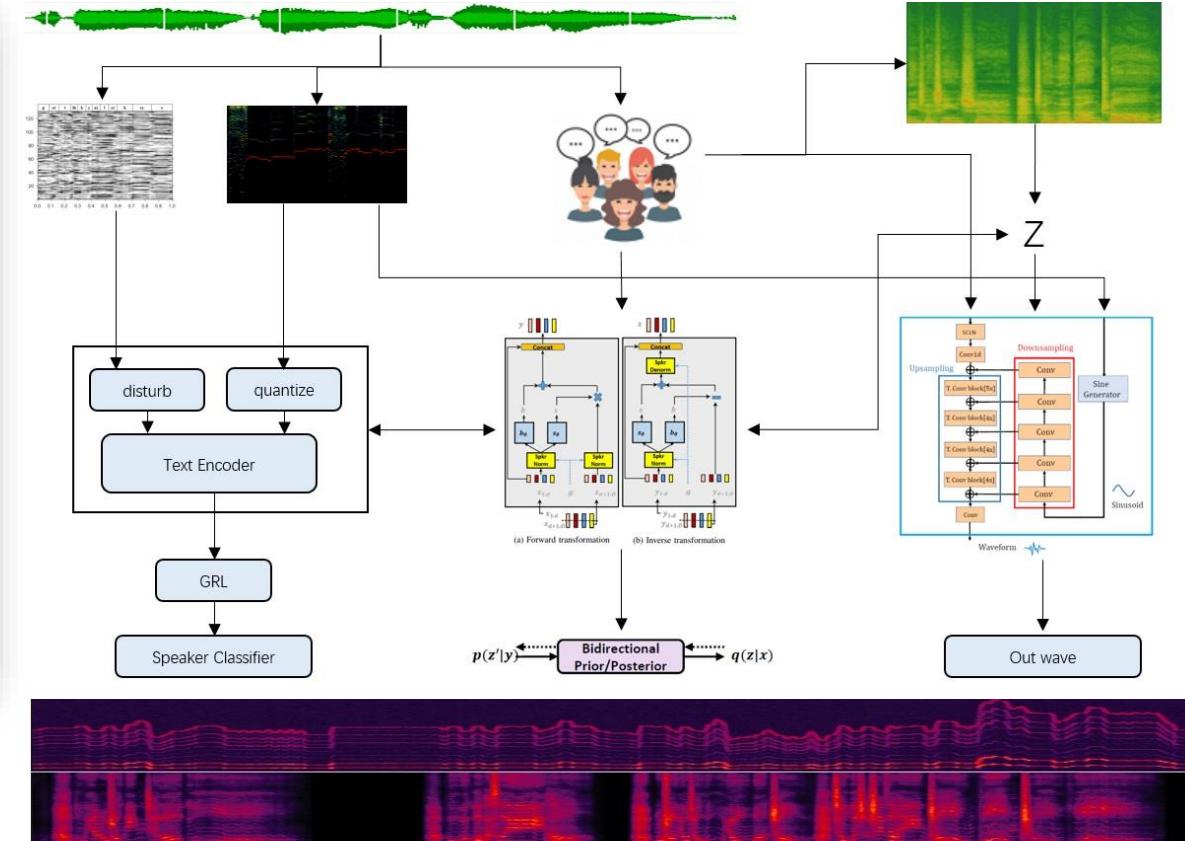
# Audio

## 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

شَكْرًا

Grazie

благодаря

ありがとうございます

Kiitos

Teşekkürler

謝謝

ឧបម្ពុណមរ៉ា

Obrigado

شُكْرِيَّه

Terima Kasih

Dziękuję

Hvala

Köszönöm

Tak

Dank u wel

дякую

Tack

Mulțumesc

спасибо

Danke

Cám ơn

Gracias

多謝晒

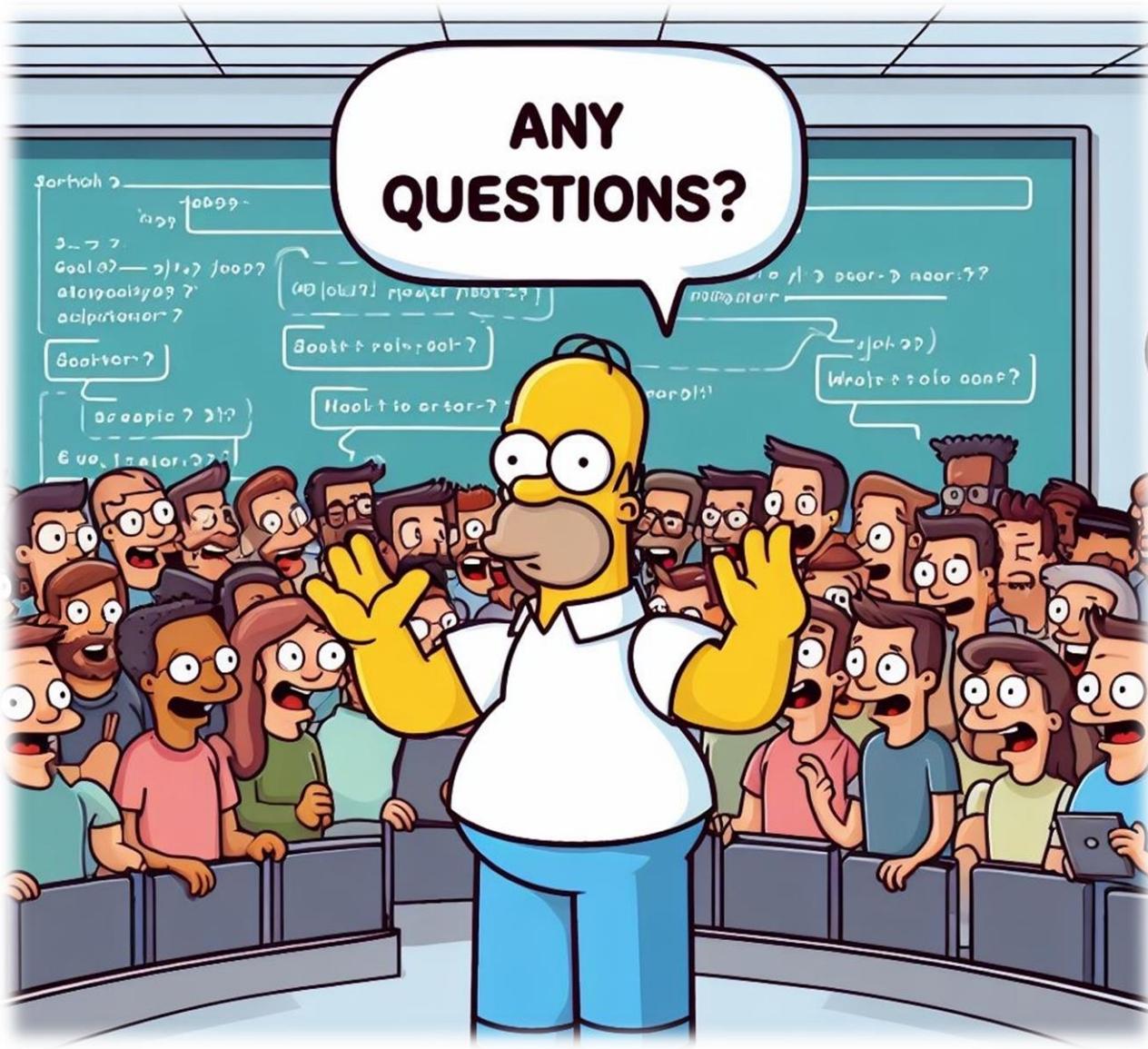
Ďakujem

הַתִּוְל

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-lm>
- <https://onnx.ai/>
- <https://docs.openvino.ai/>
- [https://github.com/openvinotoolkit/openvino\\_notebooks](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/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



INNOVATOR



NVIDIA Certified Associate - AI in the Data Center

**Clemente GIORIO**

R&D Senior Principal Engineer



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

X@tinux80



dotNET{podcast}

[PACKT]  
PUBLISHING Author



FAB  
LAB  
NAPOLI

global AI  
conference

# About Us



**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**

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



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