



## WORKSHOP

# Generative AI Hands-On: from popular services to DIY solutions



Gianni ROSA GALLINA  
R&D Technical Lead  
**deltatre**



GENERATIVE AI



“Photo of a Software Engineer, seated at an office desk, staring into the distance with a thoughtful expression. A lightbulb icon, symbol of an idea.”

# ...for all and everything

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

...



# Magic?

“Any sufficiently advanced technology is indistinguishable from magic”

*Arthur C. Clarke*



# Overview

Text



Images & Videos



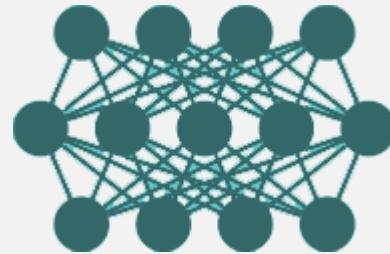
Speech & Music



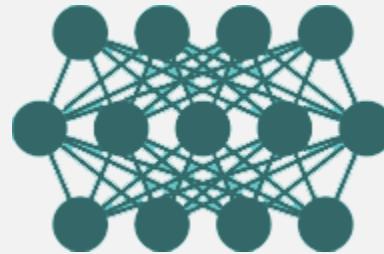
Structured Data



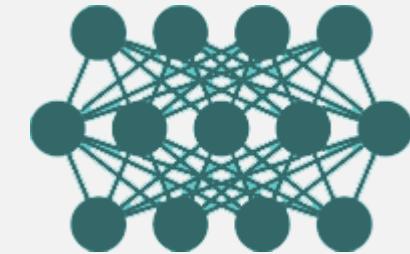
3D Signals



...

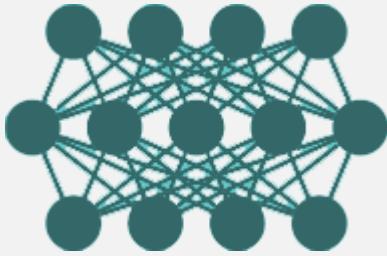


...

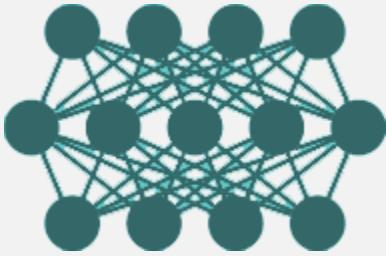


Foundation Models

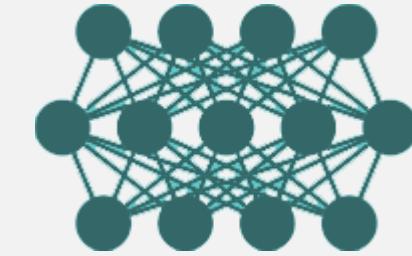
# Overview



...



...



Foundation Models

Question  
Answering

Information  
Extraction

Image  
Captioning

Multimodal  
Translation

Text-To-X

Image-To-X

# Text



**My GPTs**

**Create a GPT** Data  
Customize a version of ChatGPT for a specific purpose

**Recently Used**

- DALL-E**  
Let me turn your imagination into imagery  
By ChatGPT
- Hot Mods**  
Let's modify your image into something really wild. Upload an image and let's go!  
By ChatGPT
- Data Analysis**  
Drop in any files and I can help analyze and visualize your data  
By ChatGPT

**Made by OpenAI**

- DALL-E**  
Let me turn your imagination into imagery  
By ChatGPT
- Data Analysis**  
Drop in any files and I can help analyze and visualize your data  
By ChatGPT
- ChatGPT Classic**  
The latest version of GPT-4 with no additional features  
By ChatGPT

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

**Azure AI | Azure OpenAI Studio**

**Azure AI Studio** PUBLIC PREVIEW  
Presenting the new Azure AI Studio (Preview)  
Build, evaluate, and deploy your AI solutions from end to end.

[Explore Azure AI Studio](#)

**Welcome to Azure OpenAI service**

Explore the generative AI models, craft unique prompts for your use cases, and fine-tune select models.

**Get started**

- Chat playground**  
Design a customized AI assistant using ChatGPT. Experiment with GPT-3.5-Turbo and GPT-4 models.  
[Try it now](#)
- Completions playground**  
Experiment with completions models for use cases such as summarization, content generation, and classification.  
[Try it now](#)
- DALL-E playground** PREVIEW  
Generate unique images by writing descriptions in natural language.  
[Try it now](#)

[https://azure.microsoft.com/  
en-us/products/ai-services/openai-service/](https://azure.microsoft.com/en-us/products/ai-services/openai-service/)

# OpenAI Updates



	GPT-4 Turbo
Cut-Off date	April 2023
Context Window	128K Tokens
Multimodal Abilities	yes



<https://openai.com/blog/new-models-and-developer-products-announced-at-devday>

[Microsoft Ignite 2023: Azure OpenAI Service Announces New Models and Multimodal Advancements](#)

# Developer Assistants



**GitHub**  
Copilot



**OpenAI ChatGPT**



**Amazon CodeWhisperer**

The screenshot displays a multi-layered interface illustrating the integration of AI developer assistants. At the top, a Microsoft Visual Studio extension window titled 'Manage Extensions' shows the 'GitHub Copilot Chat' extension by Microsoft, which has been installed. Below this, the main Visual Studio IDE shows a C# code editor with the following code:

```
// write a function to check OS version
public static bool IsWindows10OrGreater()
{
    var os = Environment.OSVersion;
    return os.Platform == PlatformID.Win32NT && os.Version >= new Version(10, 0);
}
```

On the right side of the IDE, the 'Output' window shows logs from the 'GitHub Copilot' extension. In the bottom right corner of the IDE, there's a small AI icon indicating the presence of AI features. To the right of the IDE, a web browser window is open to the ChatGPT interface. A user message asks for a Python implementation to solve a Sudoku puzzle. ChatGPT responds with a Python code snippet:

```
def solve_sudoku(board):
    ...
    Solves a Sudoku puzzle using a backtracking algorithm.

    :param board: 2D list representing the Sudoku puzzle (9x9 grid)
    :return: True if the puzzle is solvable, False otherwise
    ...

    # Find an empty cell (cell with 0) to start solving
    empty_cell = find_empty_cell(board)
```

At the bottom of the browser window, there's a note: "ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT September 26 Version".

# Text-To-Image

“Style painting of  
a futuristic Torino,  
with the Mole,  
mountain background”

DALL·E 3

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

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



# Text-To-Image

i s loyi aaron o i ytu neettt mptdi uneoimatisisannrti mn  
it ie ainegcteys iittrest rvtidoltueoi/a enen aeiahuiuvnb  
roe oslolemnslaril ttomtiosdita nr ornh ctaneden e ha m nh  
g i /psui ncncrie n rilegsil tt lhicrtrunrarhneca ncn poalr  
gifncao pwnnhevtsnv igeocies tesepisodoleite lpeaoausme  
b1 eeeilaoga afn ws ne irstnm epyoohu rpdgeyd hn ae lc/ril  
in wnasnnft fgrfetiiam r drdaad psa ,a uxlia ni09 nuymen  
ggn ineo thetee8aigcg snsbasey ol tys ol,ts or doeg i /so  
nr gmr ev inowhinwt rlfwr si hai lrso,tad n o ogt bc v  
l mdc nmo ttd/eo tli tpaitrrtttnsl swi ktsinna hegpr /d  
ls yc n otgopyttpot ,apw a isrela y e/Ap  
n itvn MREXVEXLXEXLXEXLXEXLXEXLXEXLXEXLXEXLXEXLX  
nn u /vndrc i lbe v  
ba yltt e r rrlrts rIpw gnfotebi oct d bil/i  
re /teainde adieo cf sanili Ip atedwybdneitreaiaAigsoefra  
a simbaigirdseco awunaudif scopele dst esoddedimmtta oa i  
pi nnagmita a hm otni tealoren c s n ie5Ptwnitl an aasnlo  
ffs/naha eutesortdogi ttea la ur nsfh o r nd i glao im  
lyi rrhat roglrtith scaemrtbddw nwapt akmtfbwqwgigemr/  
ys g lnouginfit,dawoea iso re oioittsidott relokaagi/en  
vr nmni ea rzsdvor mufehsueannhnoitadr.aheoi e igniagendoi  
miv,odrae nea at dee erves ceg ewpp udio yci r arm iun  
a wmainaeuf roiidemesrnfolan n,Oy rwddrh tal ri moief

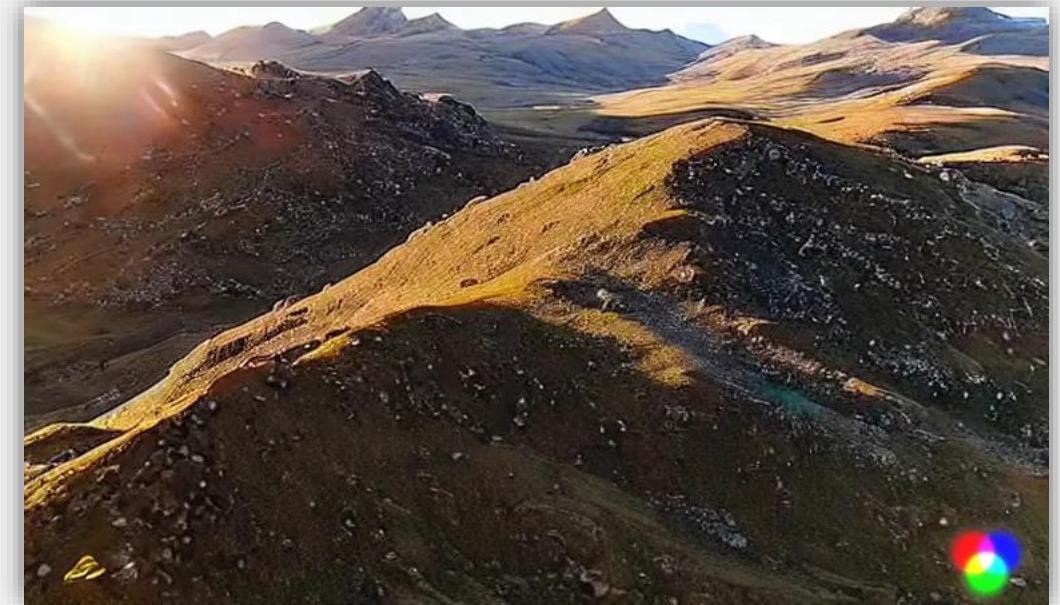


“[Inter/Juventus/Milan] club as woman, She wears the [Inter/Juventus/Milan] jersey, ultrarealistic, ultrahd, 4K”

# Text-To-Video



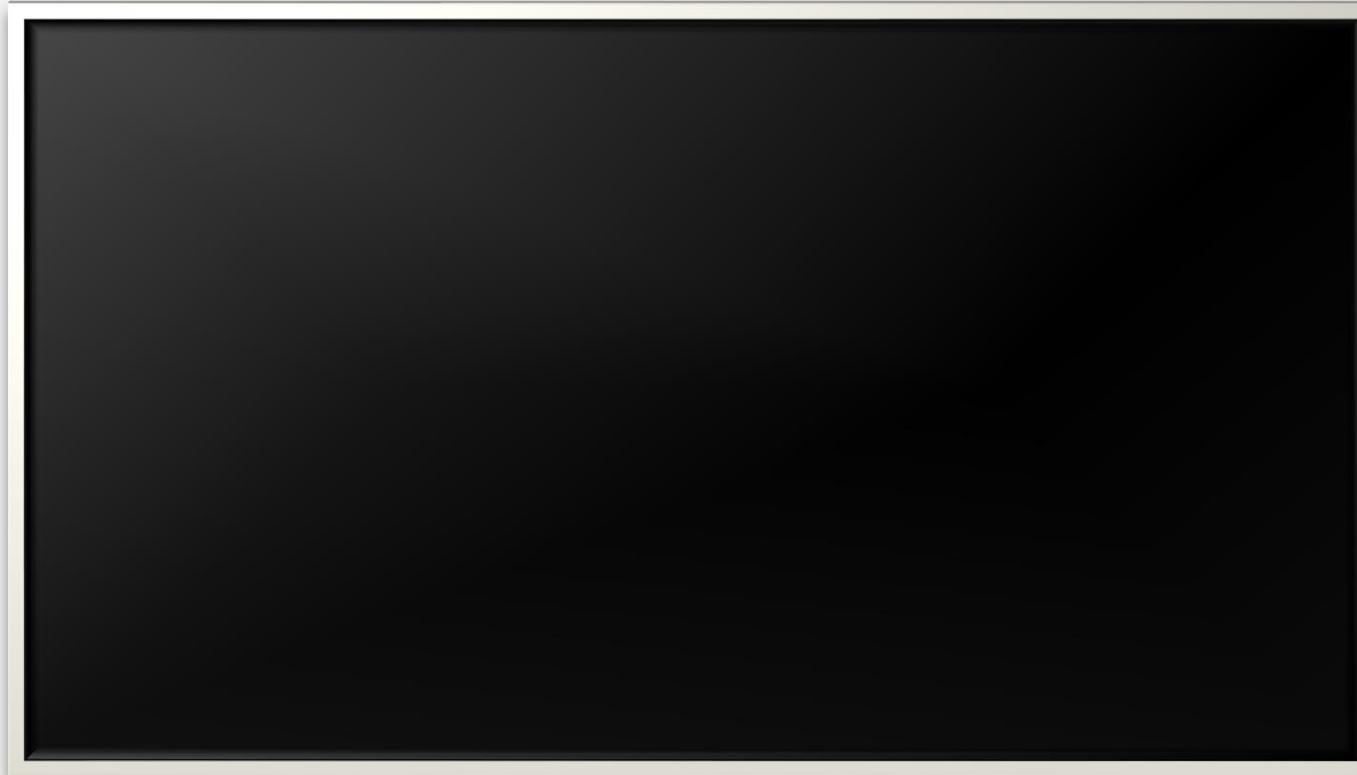
*“The late afternoon sun peeking through the window of a New York City loft.”*



*“Aerial drone footage of a mountain range”*

# Text-To-Speech

IIElevenLabs – Voice Dubbing demo



[https://www.youtube.com/watch?v=17\\_xLsqny9E](https://www.youtube.com/watch?v=17_xLsqny9E)

# Text-To-Speech

IIIElevenLabs – Generative Voice AI

Click on a language to generate random speech:

English Chinese Spanish Hindi Portuguese French

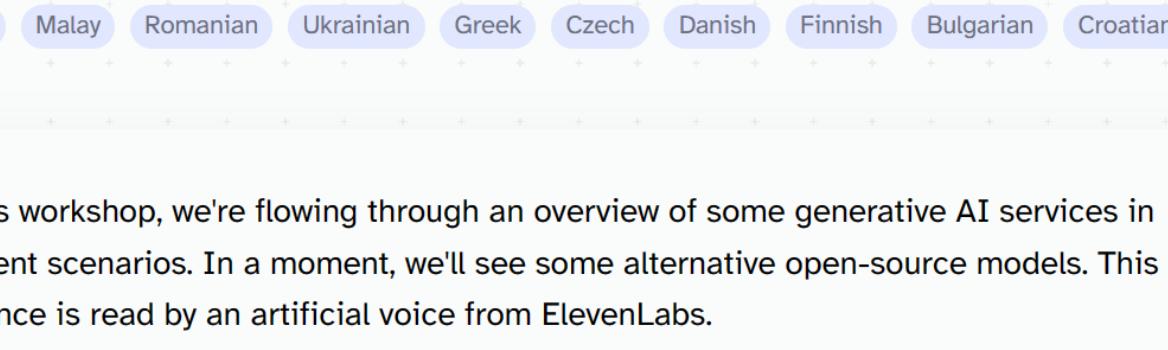
German Japanese Arabic Russian Korean Indonesian Italian Dutch Turkish Polish Swedish

Filipino Malay Romanian Ukrainian Greek Czech Danish Finnish Bulgarian Croatian Slovak

Tamil

In this workshop, we're flowing through an overview of some generative AI services in different scenarios. In a moment, we'll see some alternative open-source models. This sentence is read by an artificial voice from ElevenLabs.

— Thomas ▾ 228 / 333



<https://elevenlabs.io/>

# Text-to-3D



“A Lego bunny sitting  
on top of a stack of books.”



“A baby bunny sitting on top  
of a stack of pancakes.”

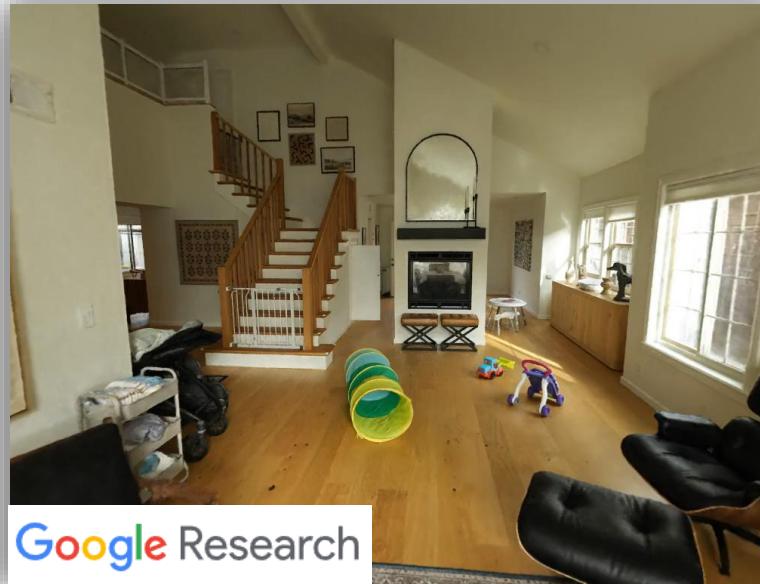


“A metal bunny sitting on top of a  
stack of chocolate cookies.”

# 2D-to-3D



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



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



<https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting>

# Cinema: Animations



# Cinema: De-Aging & Face Swapping



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



<https://metaphysic.ai/>

# Real-Time: High-Quality Effects



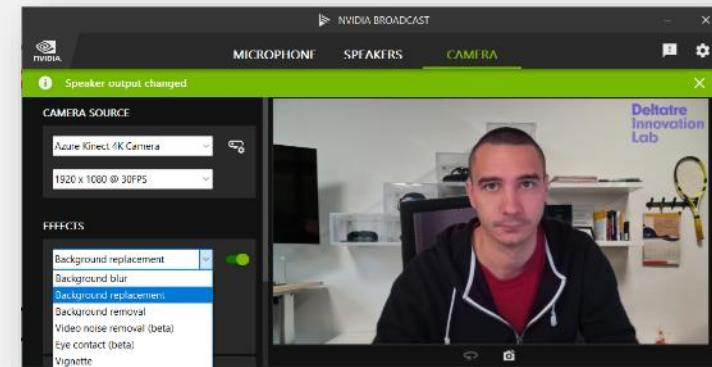
Speaker Focus  
Noise removal  
Audio Super-resolution  
Acoustic Echo Cancellation



Virtual Background  
Super Resolution  
Artifact Reduction  
Video Noise Removal



Eye Contact  
Face Tracking  
Face Landmark Tracking  
Face Mesh



# Alternative Models

## GitHub - awesome-generative-ai

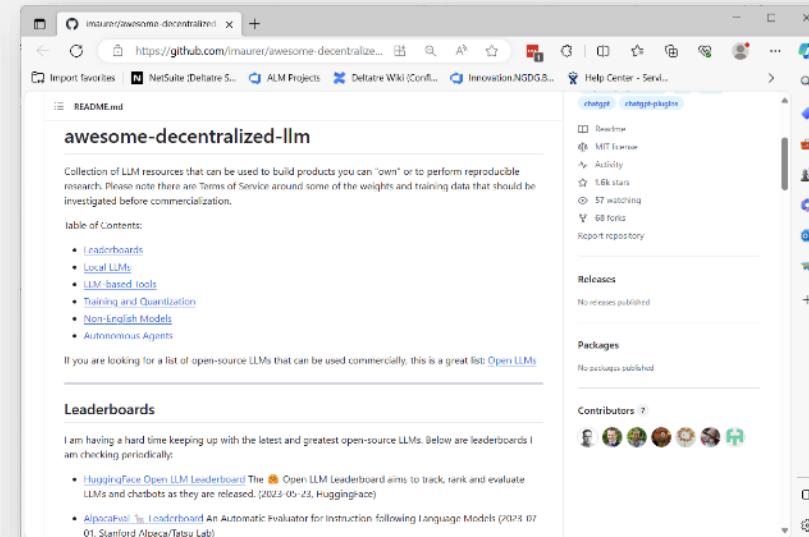
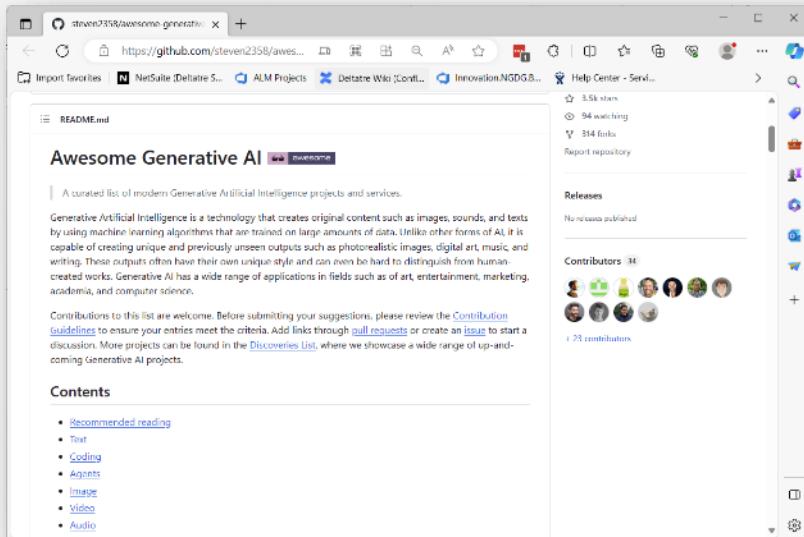
*A curated list of modern Generative Artificial Intelligence projects and services*

<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>



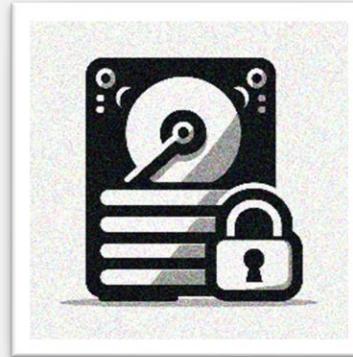
# Open-Source Models



Customization &  
Flexibility



Embedded/Mobile  
Devices



Data Policies  
& Ownership

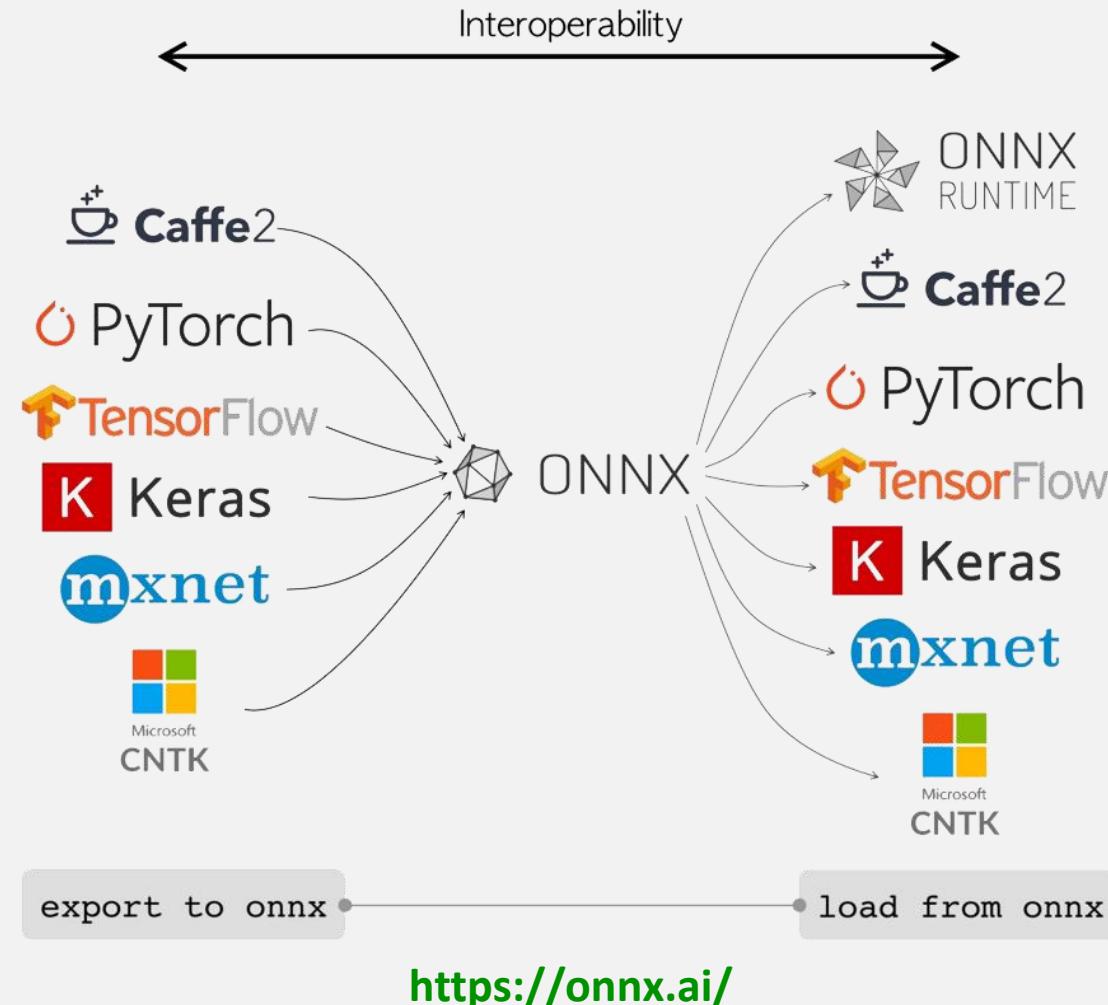


No Connectivity

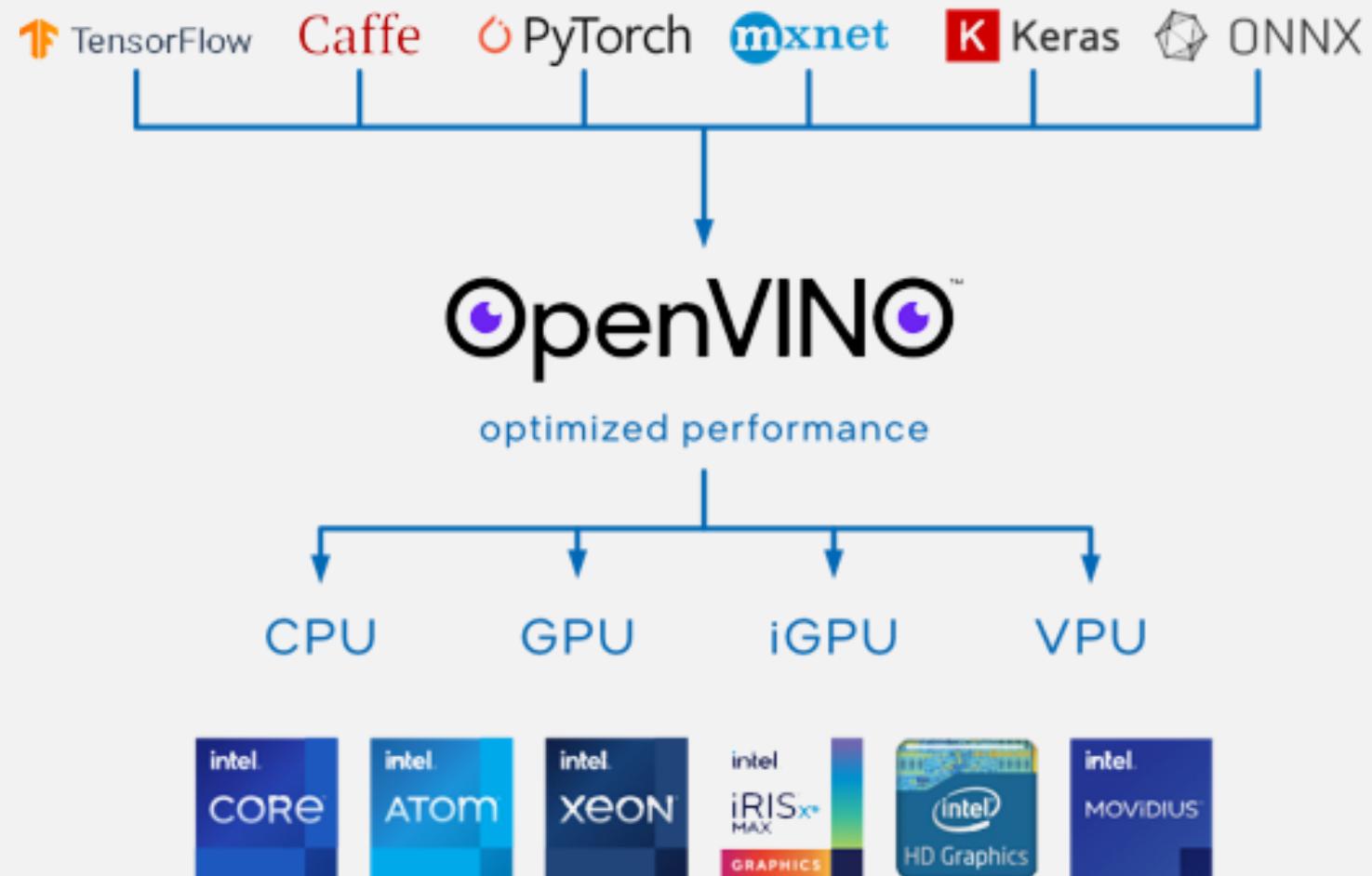


Savings &  
Optimization

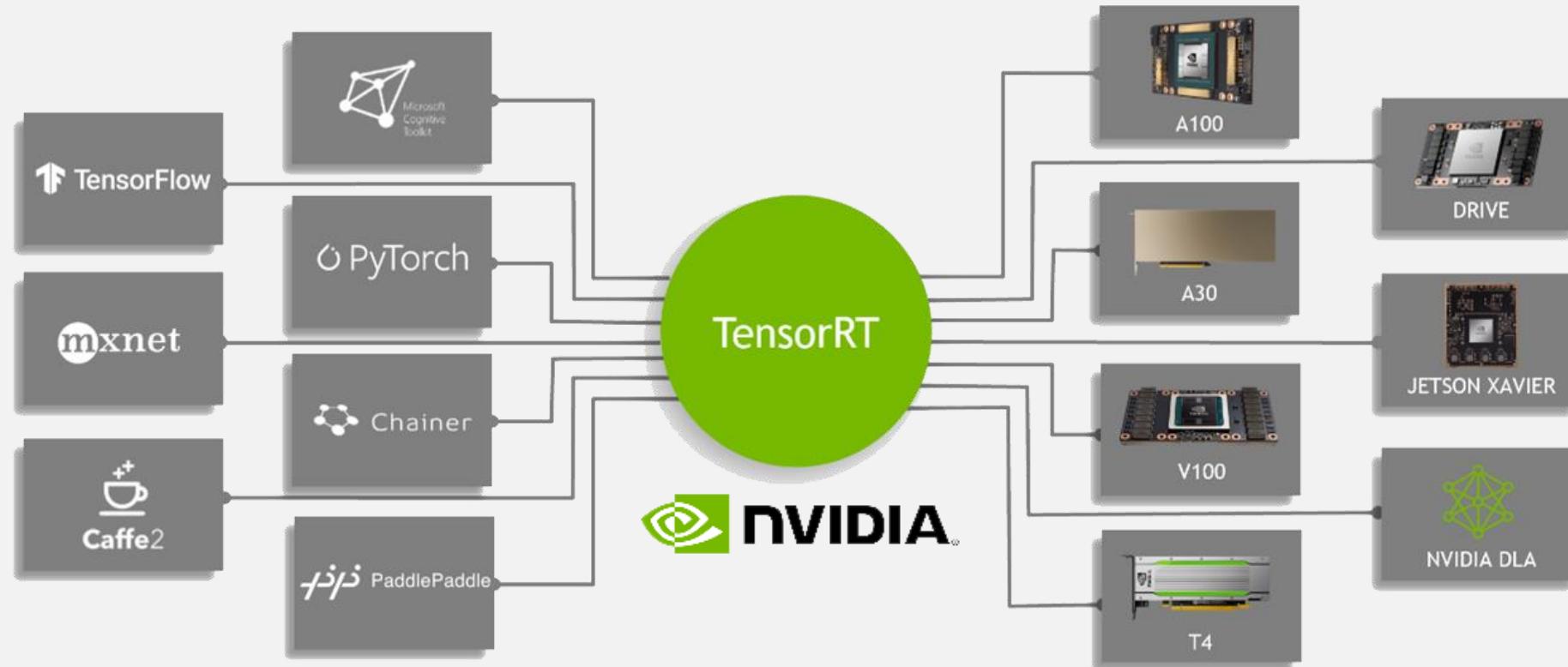
# Tools and Framework



# Intel OpenVINO™



# NVIDIA TensorRT™



IMAGES

# Stable Diffusion



**stability.ai**  **runway**

**LAION** 



**Hugging Face**

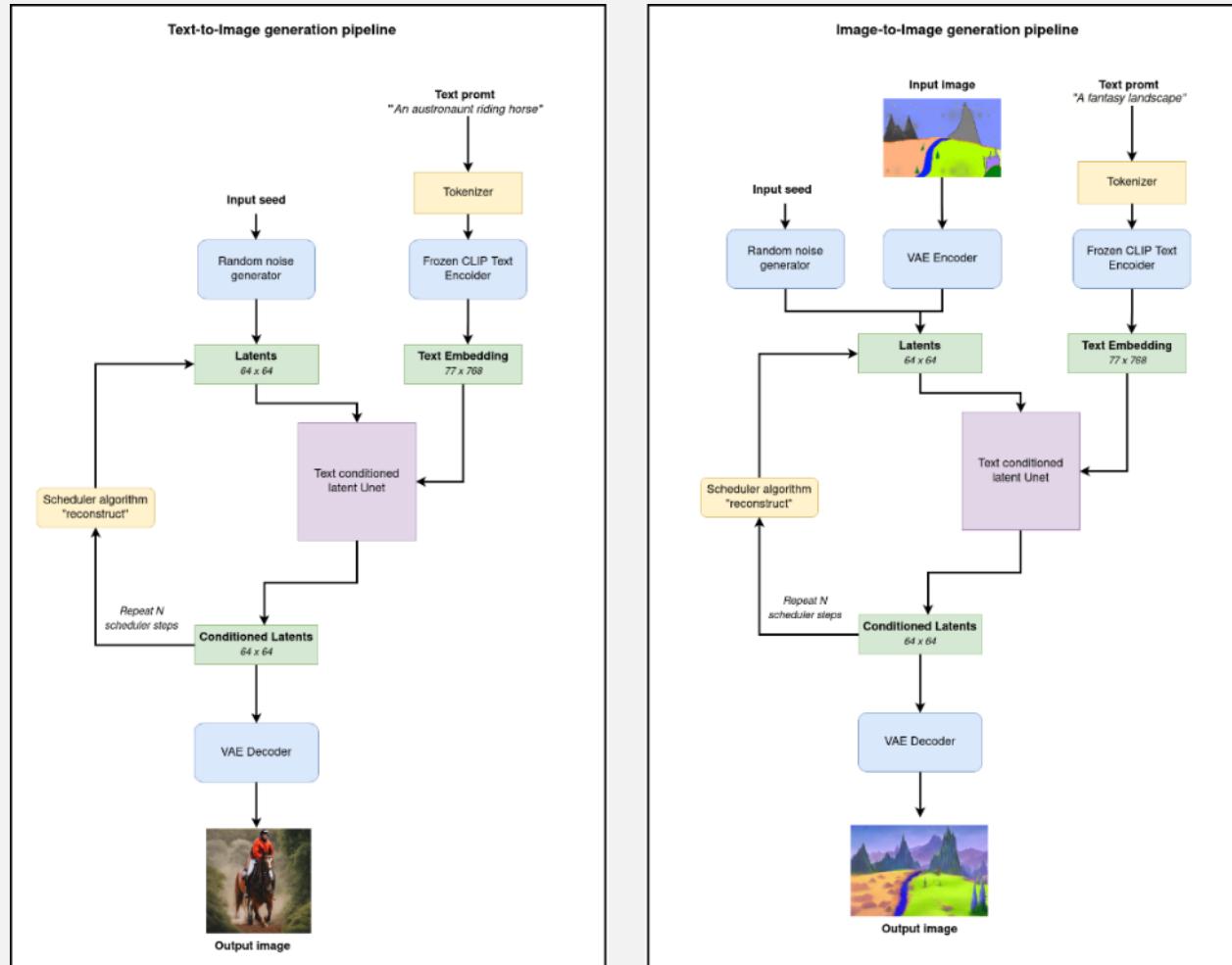
[https://huggingface.co/blog/stable\\_diffusion](https://huggingface.co/blog/stable_diffusion)

<https://stability.ai/stable-diffusion>

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

<https://runwayml.com/>

# Stable Diffusion

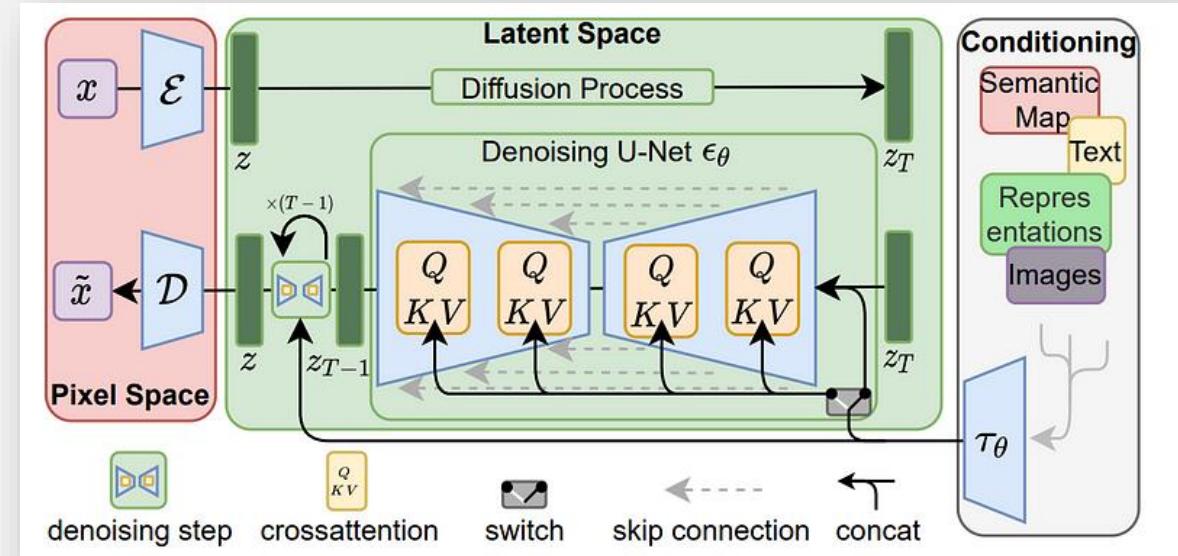


IMAGES

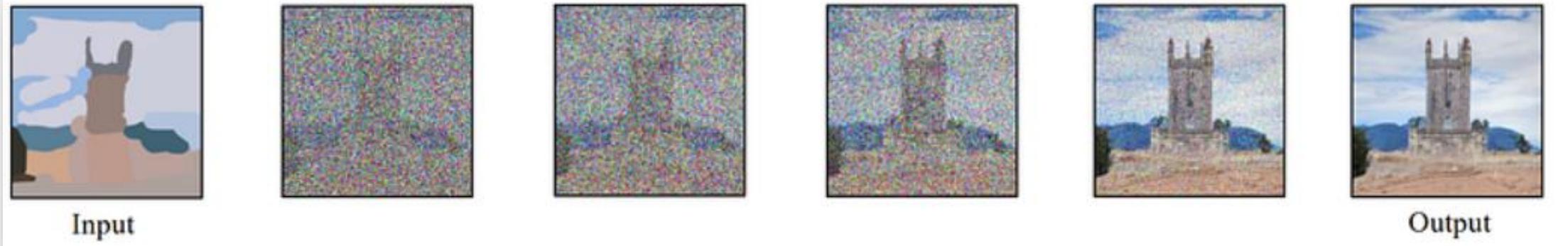
# Stable Diffusion

High-Resolution Image Synthesis with Latent Diffusion Models

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



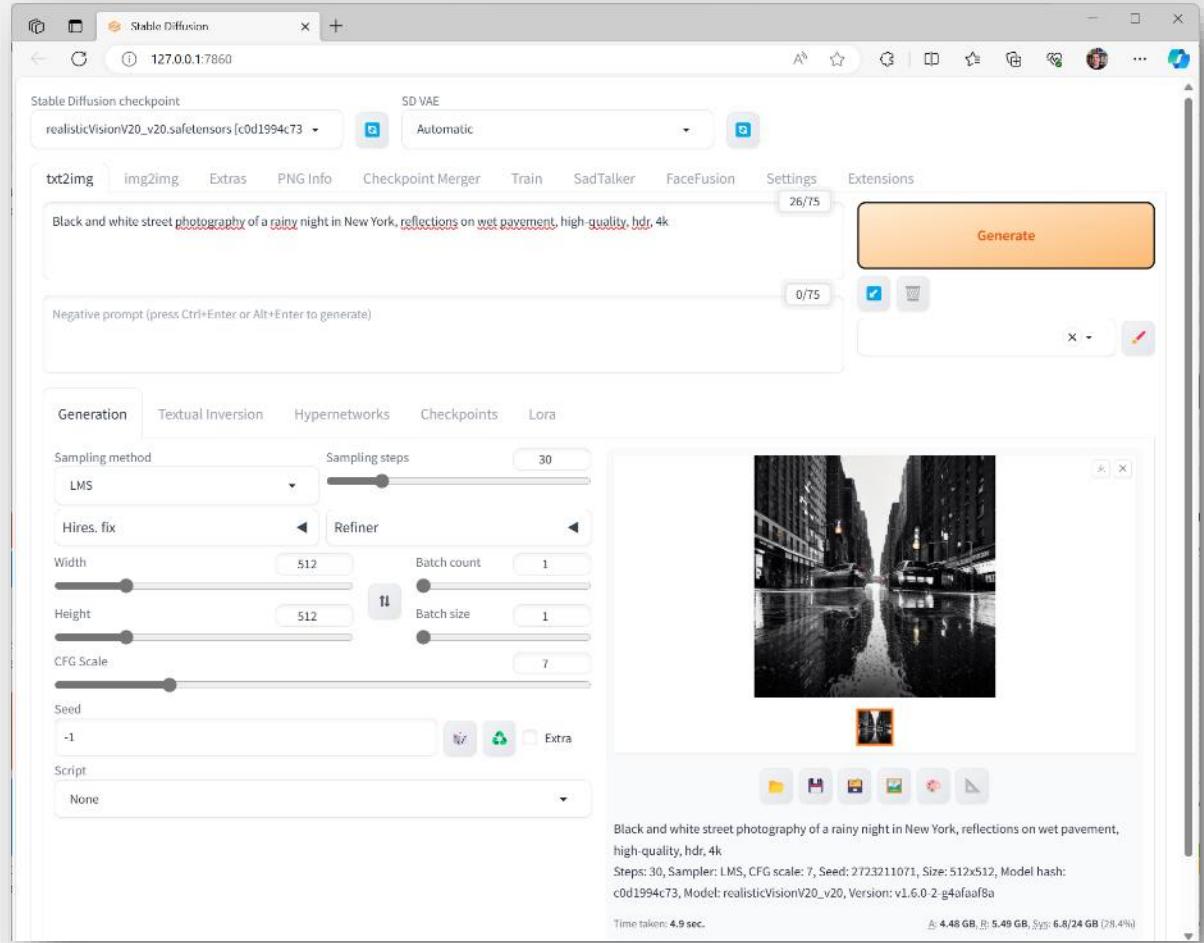
steps



IMAGES

# Stable Diffusion: Web UI Tool

DEMO



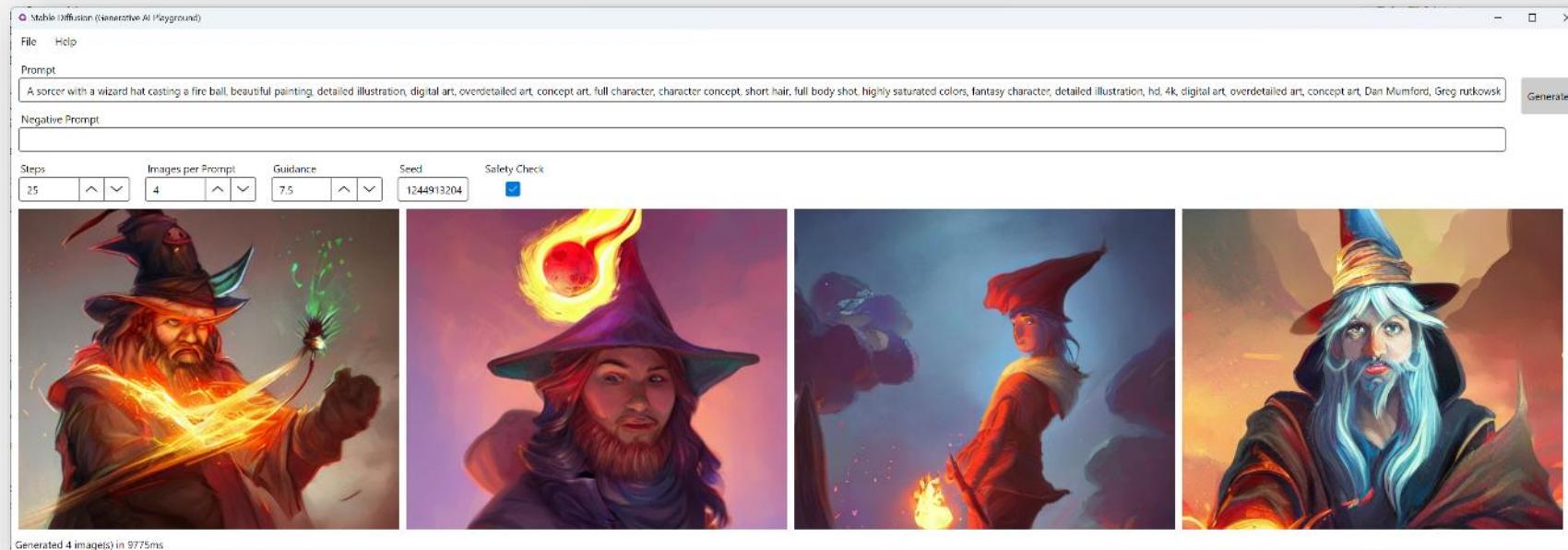
IMAGES

# Stable Diffusion: ONNX Runtime & .NET

## DEMO

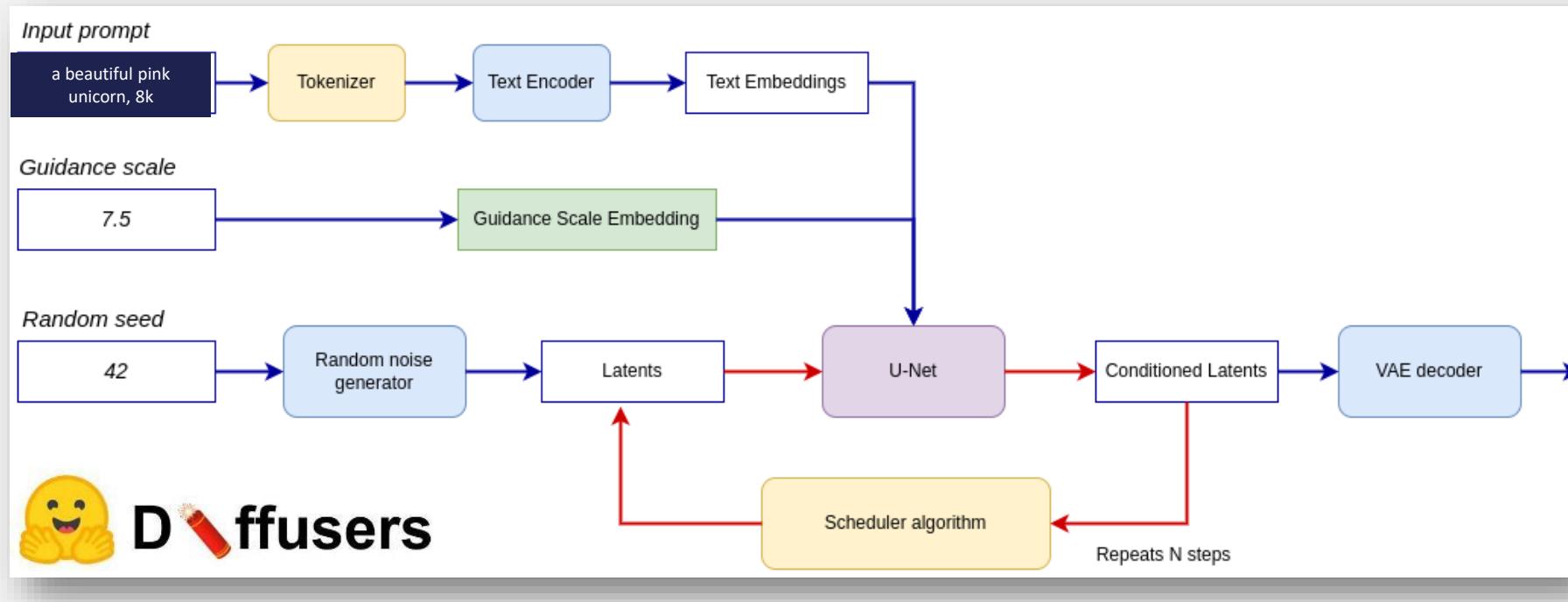
<https://github.com/gianni-rg/gen-ai-net-playground>

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



# LCM - Latent Consistency Models

**Latent Consistency Models: Synthesizing High-Resolution Images with Few-Step Inference**  
<https://arxiv.org/abs/2310.04378>



# Notebooks

English | 简体中文

## OpenVINO™ Notebooks

license Apache 2.0 treon passing docker\_treon passing

A collection of ready-to-run Jupyter notebooks for learning and experimenting with the OpenVINO™ Toolkit. The notebooks provide an introduction to OpenVINO basics and teach developers how to leverage our API for optimized deep learning inference.

### AI Trends - Notebooks

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

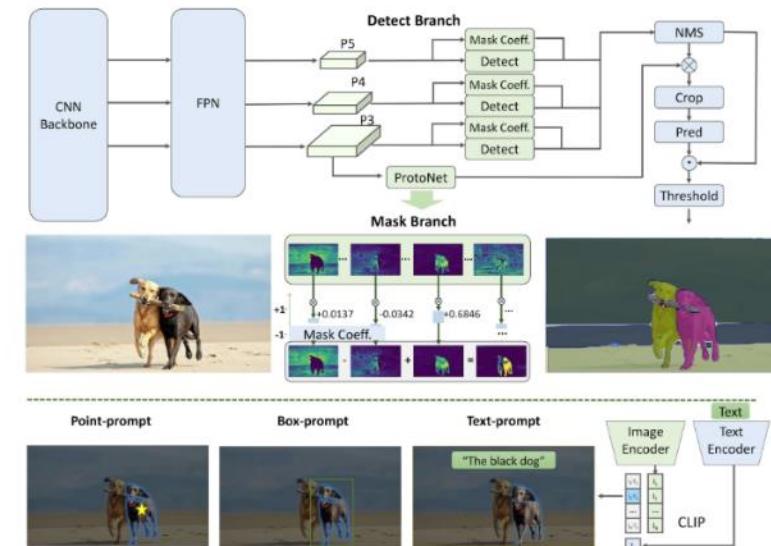
Notebook	Description	Preview	Complementary Materials
<a href="#">SAM - Segment Anything Model</a>	Prompt based object segmentation mask generation using Segment Anything and OpenVINO™		<a href="#">Blog - SAM: Segment Anything Model — Versatile by itself and Faster by OpenVINO</a>
<a href="#">ControlNet - Stable-Diffusion</a>	A Text-to-Image Generation with ControlNet Conditioning and OpenVINO™		<a href="#">Blog - Control your Stable Diffusion Model with ControlNet and OpenVINO</a>

[https://github.com/openvinotoolkit/openvino\\_notebooks](https://github.com/openvinotoolkit/openvino_notebooks)

### Object segmentations with FastSAM and OpenVINO™

[launch binder](#) [Open in Colab](#)

The Fast Segment Anything Model (FastSAM) is a real-time CNN-based model that can segment any object within an image based on various user prompts. Segment Anything task is designed to make vision tasks easier by providing an efficient way to identify objects in an image. FastSAM significantly reduces computational demands while maintaining competitive performance, making it a practical choice for a variety of vision tasks.



Notebook Contents

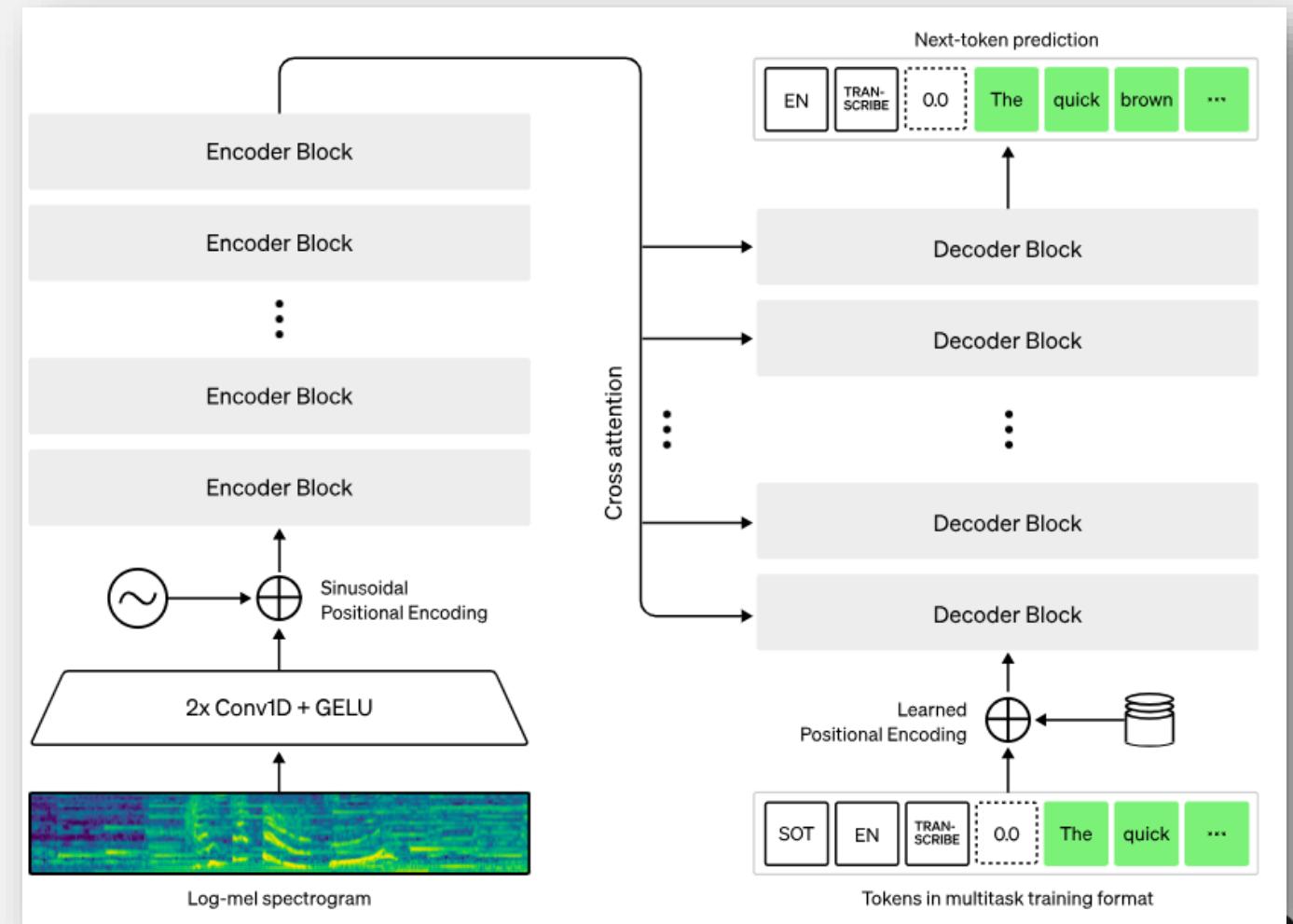
AUDIO

# Speech-To-Text: Whisper

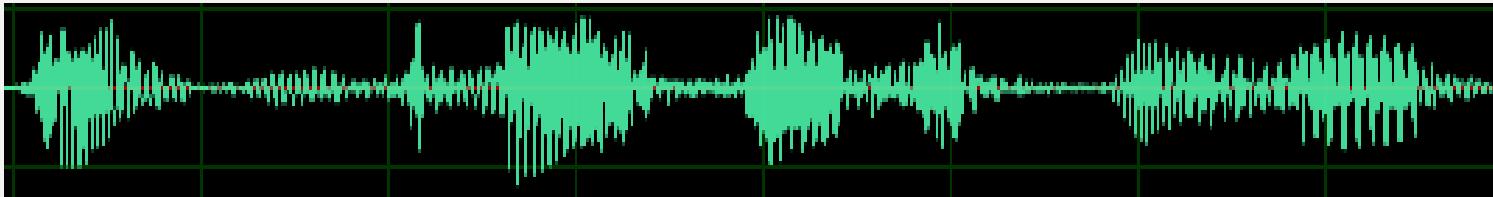


# Speech-To-Text: Whisper

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



# Speech-To-Text: Whisper v3



Robust Speech Recognition via Large-Scale Weak Supervision

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

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

<https://github.com/openai/whisper/discussions/1762>

## 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>

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



# LLaMA 2

Meta and Microsoft introduce the Next Generation of Llama  
<https://about.fb.com/news/2023/07/llama-2/>  
<https://github.com/facebookresearch/llama>

Llama 2: Open Foundation and Fine-Tuned Chat Models  
<https://arxiv.org/pdf/2307.09288.pdf>

## License

Model and weights are licensed for both **research AND commercial use**, upholding the principles of openness.

<https://ai.meta.com/llama/license/>  
<https://github.com/facebookresearch/llama/blob/main/LICENSE>



*“The central character is an anthropomorphic llama, standing on two legs, shouting out loud. As background enthusiastic software developers.”*

LARGE LANGUAGE MODELS (LLMS)

# LLaMA 2 with 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

Model	Measure	F16	Q4_0	Q4_1	Q5_0	Q5_1	Q8_0
7B	perplexity	5.9066	6.1565	6.0912	5.9862	5.9481	5.9070
7B	file size	13.0G	3.5G	3.9G	4.3G	4.7G	6.7G
7B	ms/tok @ 4th	127	55	54	76	83	72
7B	ms/tok @ 8th	122	43	45	52	56	67
7B	bits/weight	16.0	4.5	5.0	5.5	6.0	8.5

## LLaMa Sharp

<https://github.com/SciSharp/LLamaSharp>

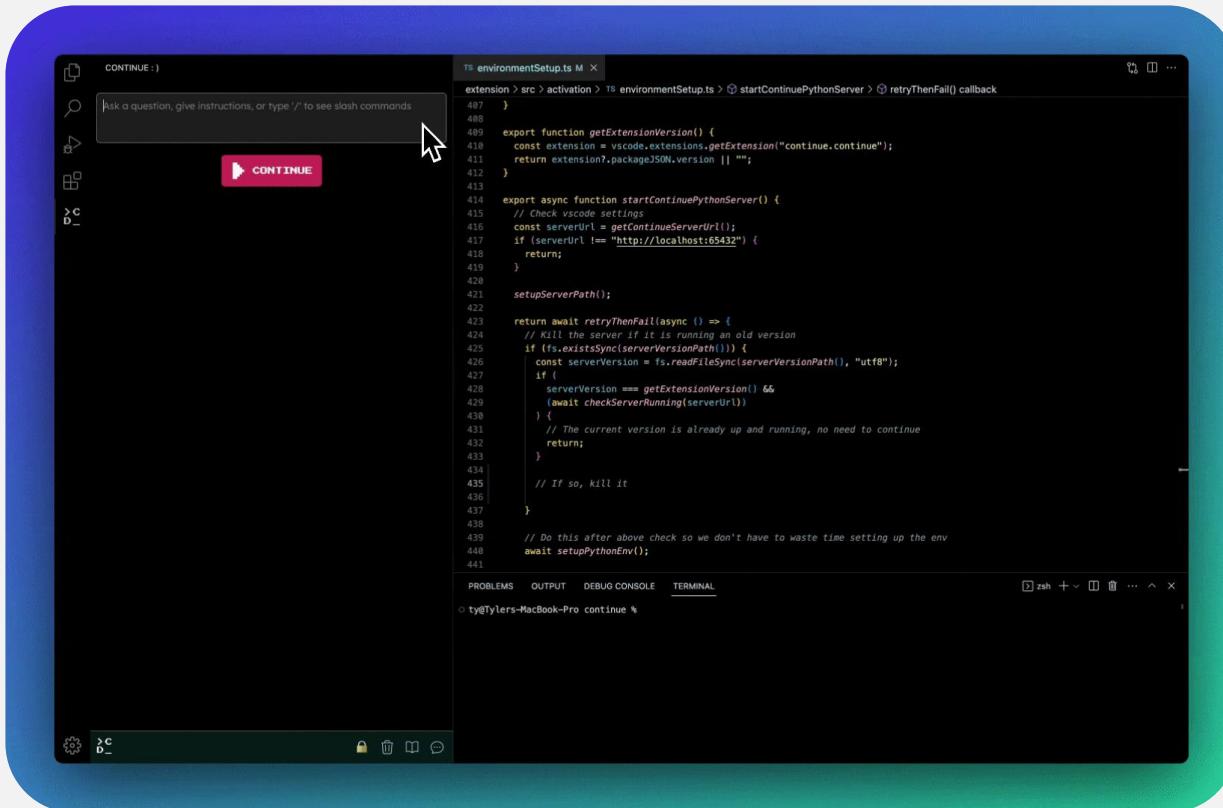
<https://github.com/gianni-rg/gen-ai-net-playground>



### Supported models:

- LLaMA 🦙
- LLaMA 2 🦙🦙
- Falcon
- [Alpaca](#)
- [GPT4All](#)
- [Chinese LLaMA / Alpaca](#) and [Chinese LLaMA-2 / Alpaca-2](#)
- [Vigogne \(French\)](#)
- [Vicuna](#)
- [Koala](#)
- [OpenBuddy 🐶 \(Multilingual\)](#)
- [Pygmalion/Metharme](#)
- [WizardLM](#)
- [Baichuan 1 & 2 + derivations](#)
- [Aquila 1 & 2](#)
- [Starcoder models](#)
- [Mistral AI v0.1](#)
- [Refact](#)
- [Persimmon 8B](#)
- [MPT](#)
- [Bloom](#)

# Developer Assistants



**Deltatre  
Innovation  
Lab**



<https://continue.dev/>

# The open-source autopilot for software development

# A VS Code extension that brings the power of ChatGPT (and other LLMs) to your IDE

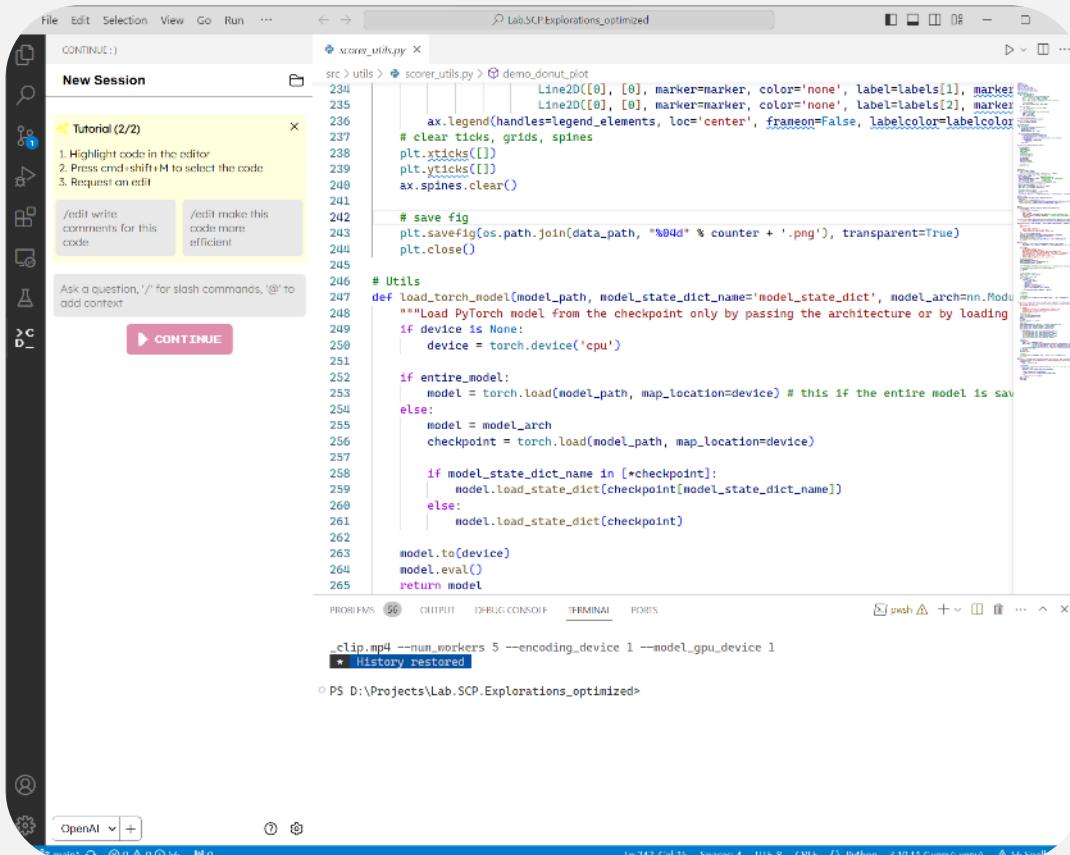
# LLaMa Sharp

<https://github.com/SciSharp/LLamaSharp>

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



# Developer Assistants



## Answer coding questions

Highlight sections of code and ask Continue for another perspective

# Edit in natural language

Highlight a section of code and instruct Continue to refactor it

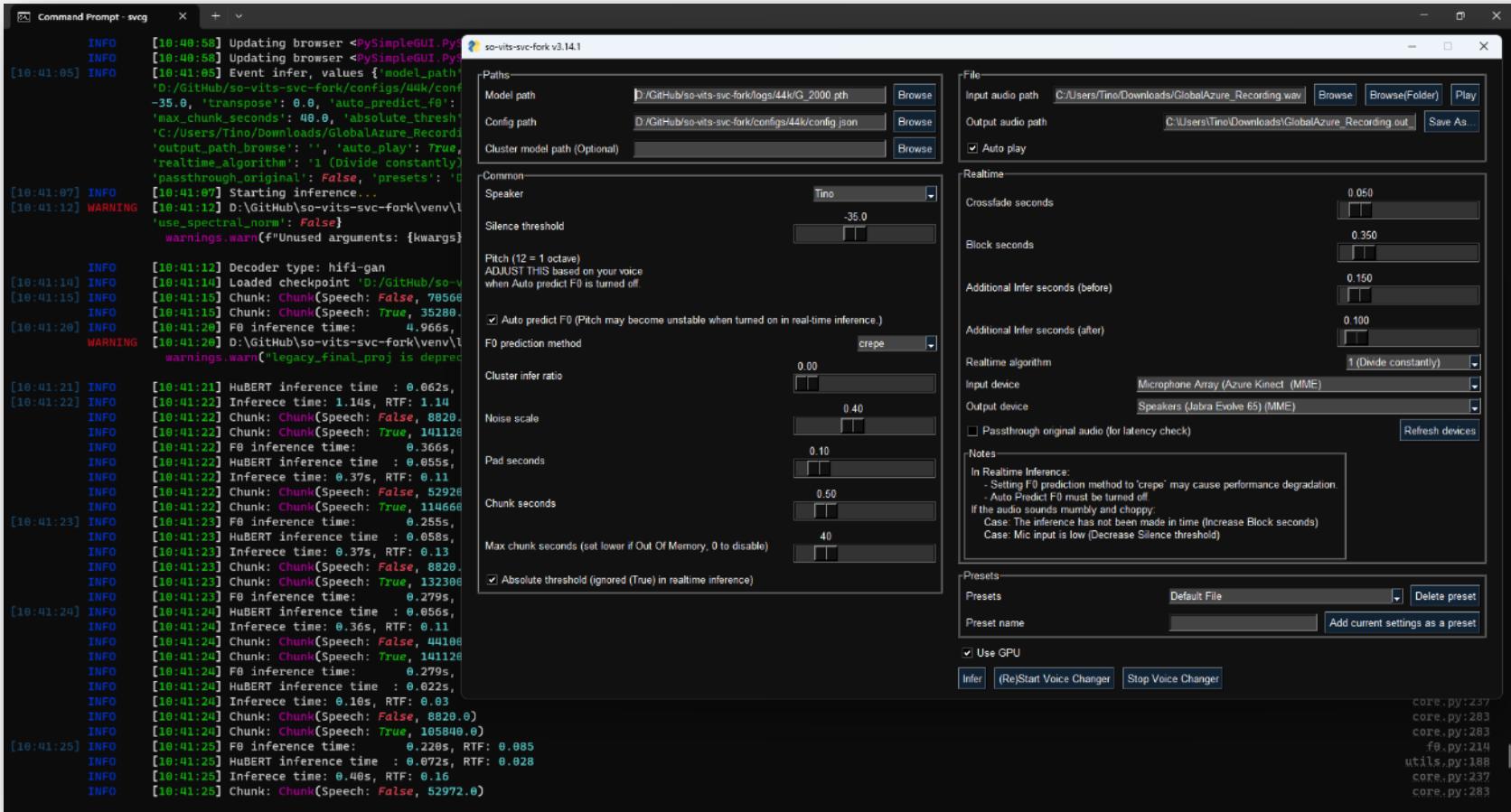
# Generate files from scratch

Open a blank file and let Continue start new Python scripts, React components, C# classes, C++ methods, etc.

# Understand errors and exceptions

In case of error or exception, you can send the stack trace into Continue and ask for it to explain the issue to you.

# Audio-To-Audio



# Foundation Models

The screenshot shows the Azure AI | Machine Learning Studio interface. The left sidebar includes sections for All workspaces, Home, Model catalog (selected), Authoring (Notebooks, Automated ML, Designer, Prompt flow), Assets (Data, Jobs, Components, Pipelines, Environments), Models, Endpoints, Manage (Compute, Monitoring, Data Labeling, Linked Services), and a preview section for Compute. The main content area displays the 'Model catalog' (PREVIEW) for the workspace 'Deltatre S.p.A. > aml-aiday-test'. It features sections for 'Announcements' (Azure OpenAI language models, Introducing Llama 2, NVIDIA AI foundation models) and 'Models' (a grid of 20 model cards). The right sidebar contains filters for 'Collections' (Curated by Azure AI, Meta, Hugging Face, NVIDIA, Microsoft Research) and 'Inference tasks' (Text classification, Token classification, Table question answering, Question answering, Zero-shot classification, Translation, Summarization, Conversational, Text generation, Fill mask, Speech recognition, Chat completion, Embeddings, Image classification, Image segmentation, Object detection, Text to image, Zero-shot image classification).

# 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, laws, policies."*



# Recap

- Generative AI Overview & Tools
- Stable Diffusion (Web & .NET)
- Latent Consistency Models
- OpenVINO & Notebooks
- Whisper
- LLMs: LLaMA 2, llama.cpp & .NET
- Developer Assistants
- Audio-To-Audio
- Azure Machine Learning Catalog
- Ethical AI & Security



# Thank You!

ευχαριστώ

Salamat Po

متشكرم

شكراً

Grazie

благодаря

ありがとうございます

Kiitos

Teşekkürler

謝謝

ឧបម្ពុណ្ឌរំបៀប

Obrigado

شكريه

Terima Kasih

Dziękuje

Hvala

Köszönöm

Tak

Dank u wel

ДЯКУЮ

Tack

Mulțumesc

спасибо

Danke

Cám ơn

Gracias

多謝晒

Ďakujem

הודות

ശ്രദ്ധി

Děkuji

감사합니다

# About Me



Ing. Gianni ROSA GALLINA

R&D Technical Lead @ **deltatre**

X @giannirg



**Microsoft** Specialist  
**Microsoft** CERTIFIED

Programming in C#  
Programming in HTML5  
with JavaScript & CSS3



Deltatre  
Innovation  
Lab



PLURALSIGHT  
Author

- 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
- Blog: <https://gianni.rosagallina.com/en/>
- LinkedIn: <https://www.linkedin.com/in/gianni-rosa-gallina-b206a821/>



# References (1/2)

- <https://www.bing.com/images/create/>
- <https://midjourney.com/>
- <https://azure.microsoft.com/en-us/products/ai-services/openai-service/>
- <https://runwayml.com/>
- <https://research.runwayml.com/gen2>
- <https://openai.com/dall-e-3>
- <https://openai.com/research/whisper>
- [https://www.youtube.com/watch?v=17\\_xLsqny9E](https://www.youtube.com/watch?v=17_xLsqny9E)
- <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/openvinotoolkit/openvino_notebooks)
- [https://huggingface.co/blog/stable\\_diffusion](https://huggingface.co/blog/stable_diffusion)

# References (2/2)

- <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/SciSharp/LLamaSharp>
- <https://github.com/gianni-rg/LlamaSharpApiServer/>
- <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://whisper.ggerganov.com/talk/>
- <https://about.fb.com/news/2023/07/llama-2/>
- <https://arxiv.org/pdf/2307.09288.pdf>