

Generatywna Sztuczna Inteligencja

Obrazy i muzyka



Włodzisław Duch

Katedra Informatyki Stosowanej UMK

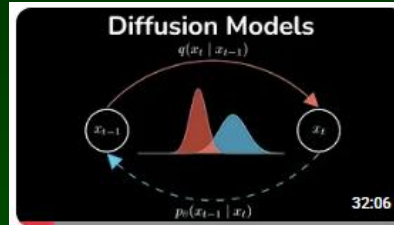
Google: Włodzisław Duch

Generative AI Animated

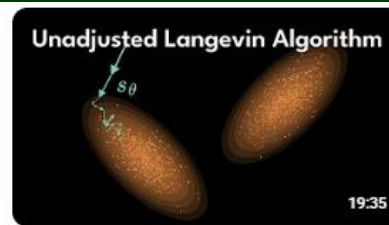
Animacja algorytmów Deepia

- Latent Space Visualisation: PCA, t-SNE, UMAP
- Autoencoders
- Denosing Autoencoders
- Diffusion Models: DDPM
- Variational Autoencoders

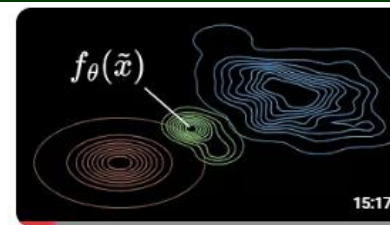
- MIT Sloan course and tools



Diffusion Models: DDPM | Generative AI Animated
2K views · 13 hours ago



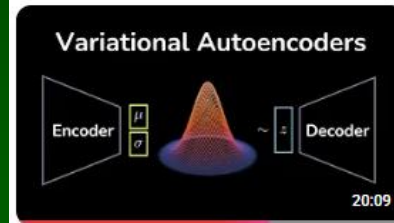
Unadjusted Langevin Algorithm | Generative AI Animated
19K views · 5 months ago



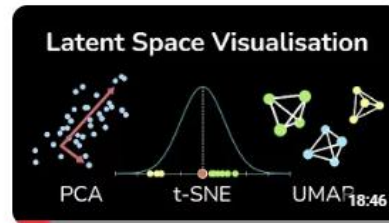
Denosing Autoencoders | Deep Learning Animated
27K views · 5 months ago



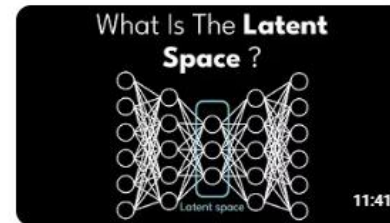
Contrastive Learning with SimCLR | Deep Learning Animated
15K views · 7 months ago



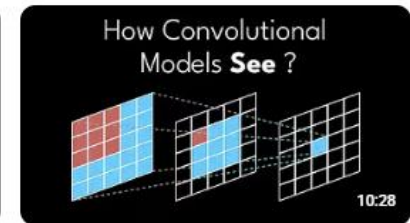
Variational Autoencoders | Generative AI Animated
100K views · 8 months ago



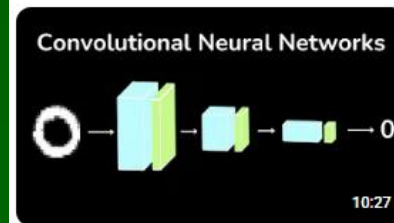
Latent Space Visualisation: PCA, t-SNE, UMAP | Deep Learning Animated
83K views · 9 months ago



Autoencoders | Deep Learning Animated
56K views · 11 months ago



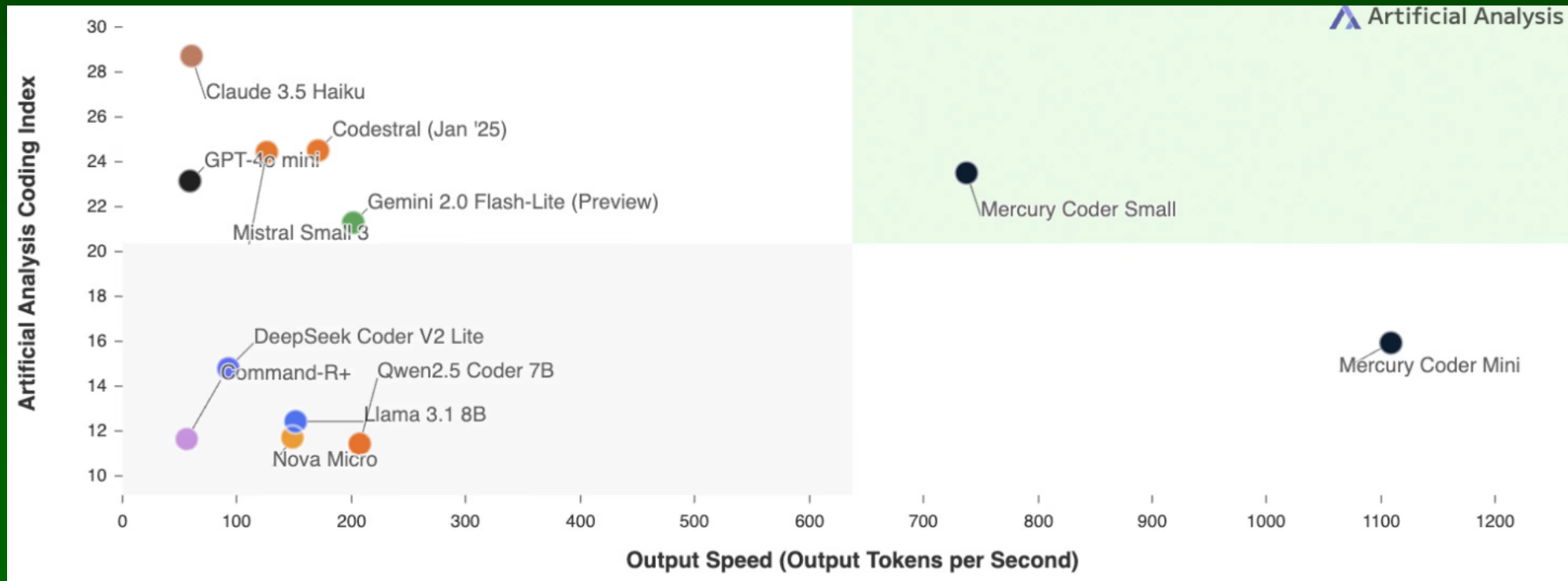
CNN Receptive Field | Deep Learning Animated
18K views · 11 months ago



Convolutional Neural Networks | Deep Learning Animated
33K views · 1 year ago

Dyfuzyjne LLMs

- [Inception Labs](#) Mercury, pierwszy dyfuzyjny model językowy na skalę komercyjną.
- Rodzina dyfuzyjnych dużych modeli językowych (dLLM) Mercury, nowa generacja LLM, która przesuwa granicę szybkiego generowania wysokiej jakości tekstu. Generuje ponad 1000 tokenów/s na komputerach NVIDIA H100.
- Model generowania kodu, Mercury Coder, jest dostępny do testowania.
- Takie modele umożliwią głębokie rozumowanie i odpowiedzi bez żadnych opóźnień.

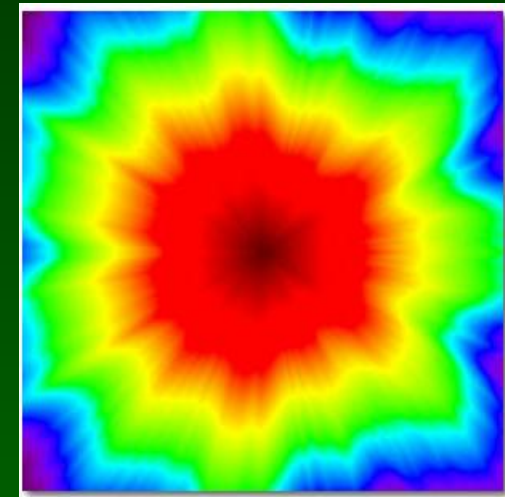
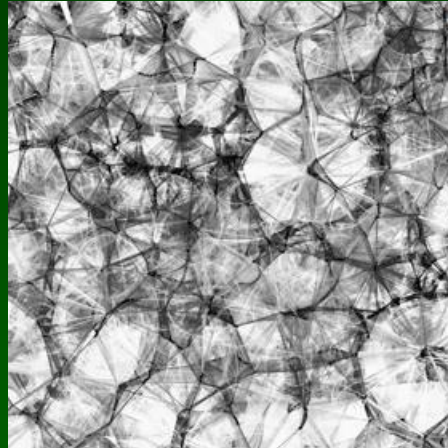


Generative art

„Generative art” w Wikipedii, początki AI w sztuce

http://en.wikipedia.org/wiki/Generative_art

<http://www.random-art.org>



<http://www.solaas.com.ar/dreamlines/p5/>

<http://www.jhllabs.com/java/art.html>

Yuki Terai singing animated avatar, first digital idol
from Japan!



Aaron Painter

Harold Cohen has asked about minimal requirements that should be fulfilled to call something art painting, and created in 1973 [Aaron](#) robotic artist.



Aaron examples



Meeting On Gauguin's Beach,



Aaron, with Decorative Panel, 1988
Oil 1992

Basic principles of composition + randomization. [See more here.](#)

Aaron in museums



Liberty & Friends, 1985
Ink



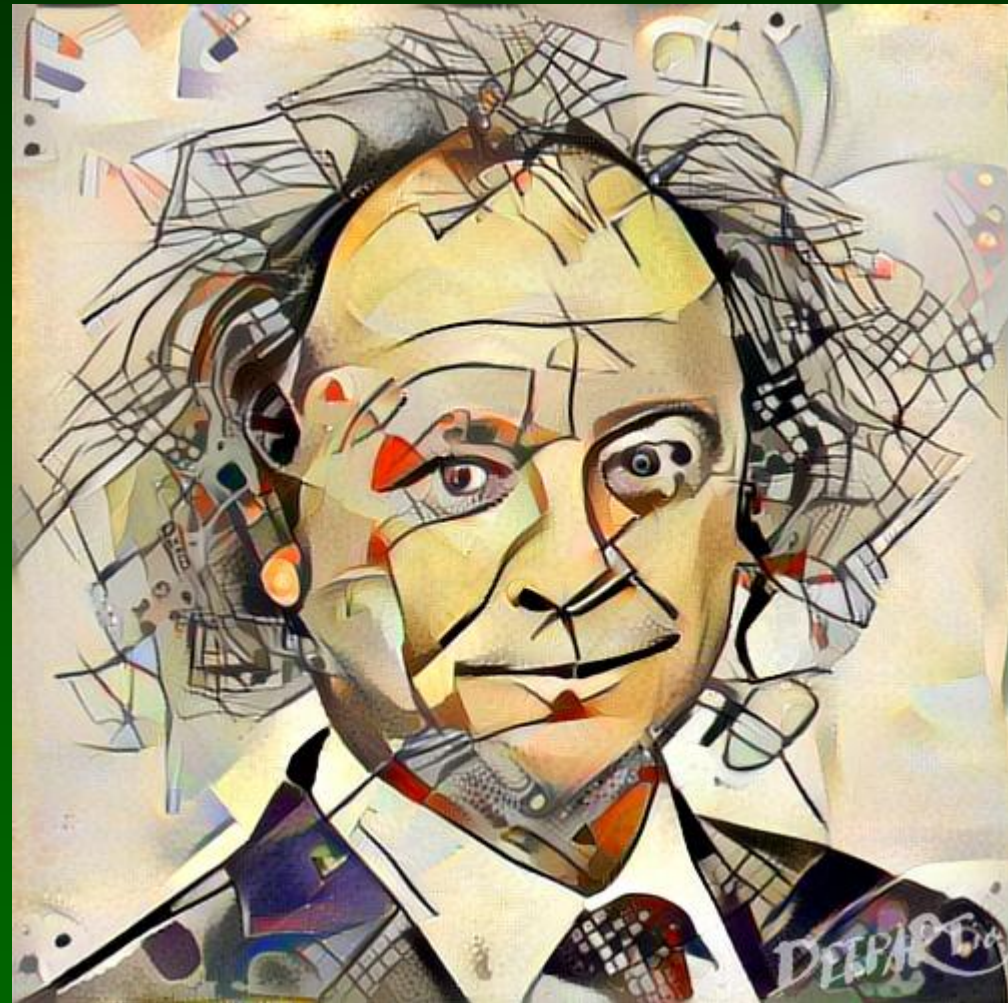
Mural, Capitol
Children Museum 1980

Deep Art. - blending



<https://deepart.io/latest/> Deep Neural Networks, Deep Dream,
LA Gatys, AS Ecker, M Bethge, A Neural Algorithm of Artistic Style (2015)

Deep Art



<https://deepart.io/latest/> Deep Neural Networks, Deep Dream,
LA Gatys, AS Ecker, M Bethge, A Neural Algorithm of Artistic Style (2015)

Examples of Deep Art



<https://deepart.io/latest/> Deep Art,
LA Gatys, AS Ecker, M Bethge, A Neural Algorithm of Artistic Style (2015)

Deep Dream



[Google Deep Dream/Deep Style i Generator](#),
LA Gatys, AS Ecker, M Bethge, A Neural Algorithm of Artistic Style (2015)

Deep Dream



[Google Deep Dream/Deep Style i Generator](#),
LA Gatys, AS Ecker, M Bethge, A Neural Algorithm of Artistic Style (2015)

Nowy Rembrandt

Next Rembrandt Project.

- 346 znanych portretów Rembrandta van Rijn zeskanowanow w 3D, ok ~150 GB.
- Nauka malowania w stylu Rembrandta, szczegółowe ruchy pędzla, wykorzystanie głębokich sieci neuronowych.
- Stworzenie nowego portretu w stylu Rembrandta – 500 godzin renderowania obrazu, 148 mln pikseli w 3D.
- Druk 13 warstw farby.

i



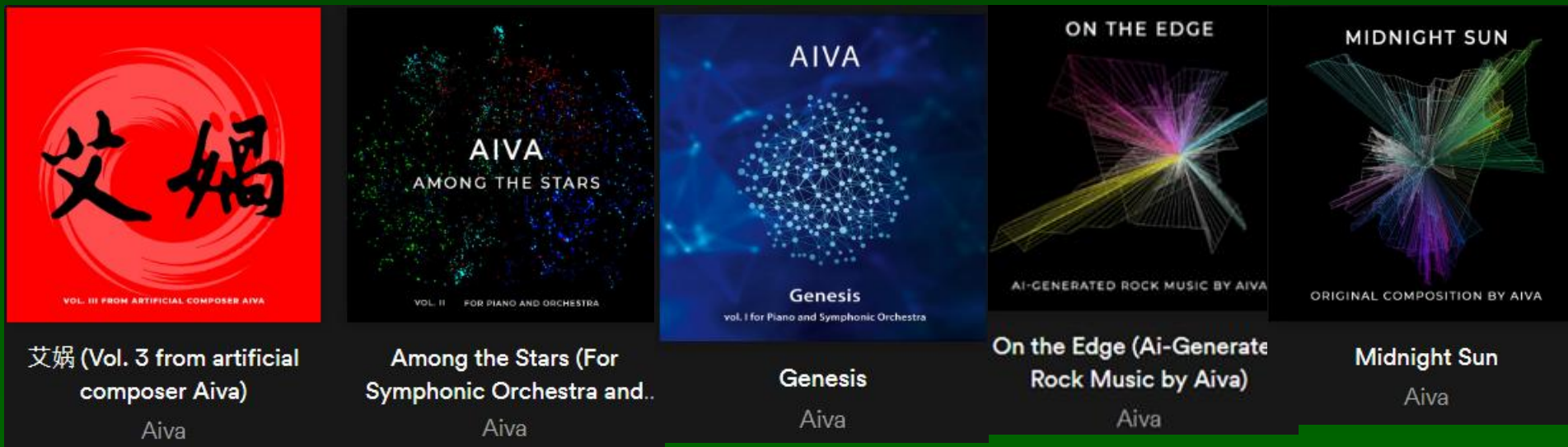
AI Virtual Artist

[AIVA](#) – AI Virtual Artist, member of author's rights society ([SACEM](#)), 206 works.

[AIVA YouTube](#) channel, Youtube „[Letz make it happen](#)“, Op. 23

[SoundCloud](#) channel

[Spotify](#) and [Apple](#) channel



Generatywna AI: Obrazy, wideo, muzyka ...

Luty 2024: SORA, nowy system tekst => realistyczne wideo i obrazy, doskonałe połączenie. Aktorzy i scenarzyści w Hollywood protestują. Powstają tuziny programów GenAI, tworząc tekst, programy, obrazy, grafikę, wideo, kopiując głos ludzki, komponując i wykonując muzykę ...

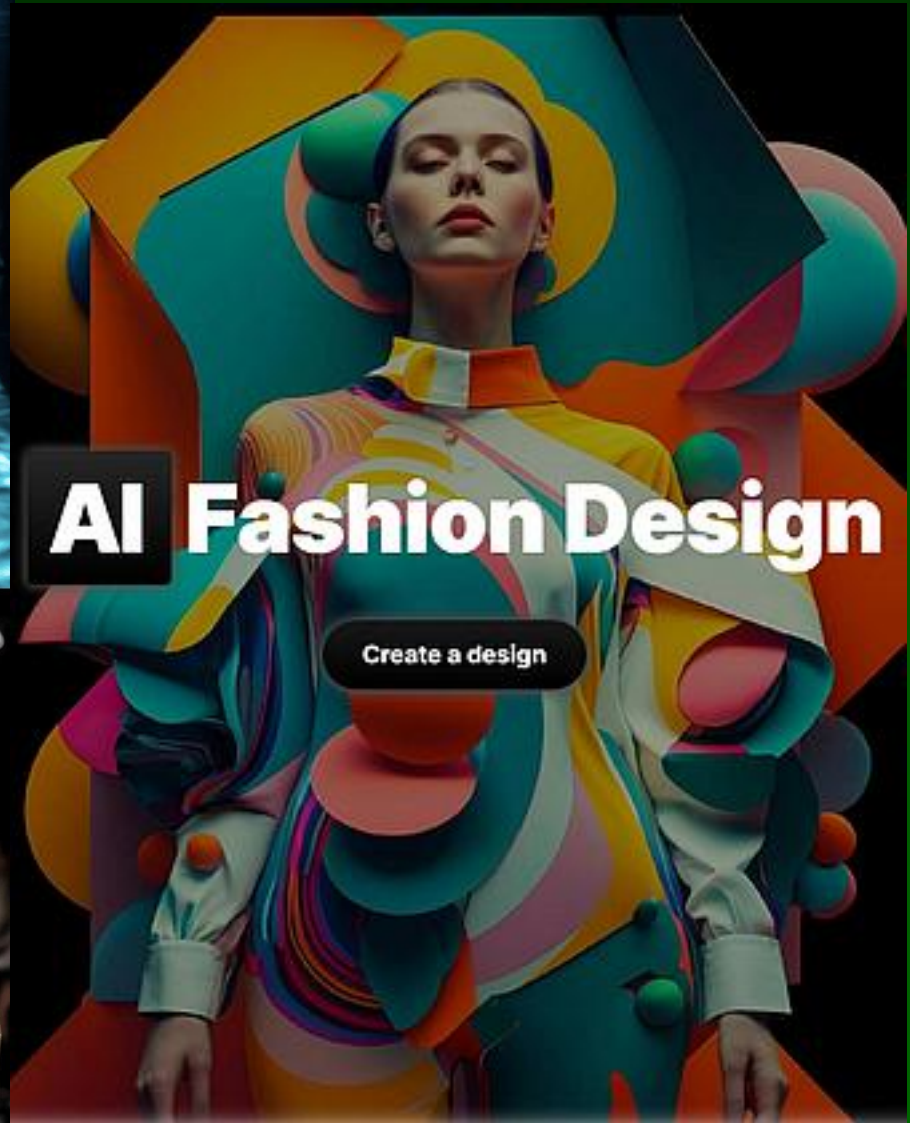
Lista programów: [AI Image Generation Tools](#) i porównanie [16 programów](#).



Projektowanie: zabawki i buty



Design: tydzień mody AI (NY, 4/2023)



Generative AI for digital art and music

"Edmond de Belamy" – portret namalowany przez AI sprzedany na aukcji Christie w 2018 roku za 432 500\$.



"Théâtre D'opéra Spatial", zwycięzca dorocznego konkursu Colorado State Fair's w 2022 roku.



'A.I. God. Portrait of Alan Turing', namalowany przez robota Ai-Da sprzedano za 1.32-mln \$ na aukcji Sotheby w 2024 r. Część 5-panelli, pokazanych na szczycie United Nations na temat A.I. w Genewie.



AIVA należy do stowarzyszenia kompozytorów SACEM w Francji. Suno, Udio, Soundful, Mubert – tworzą muzykę.

Uczymy AI politycznie poprawnych halucynacji

Sure, here is an image of a pope:

Gemini 2/2024



Halucynacje: ludzie i LLMy

Cena kreatywności: kompresja informacji w sieciach neuronowych umożliwia nowe skojarzenia, ale **część będzie zawsze błędna!**

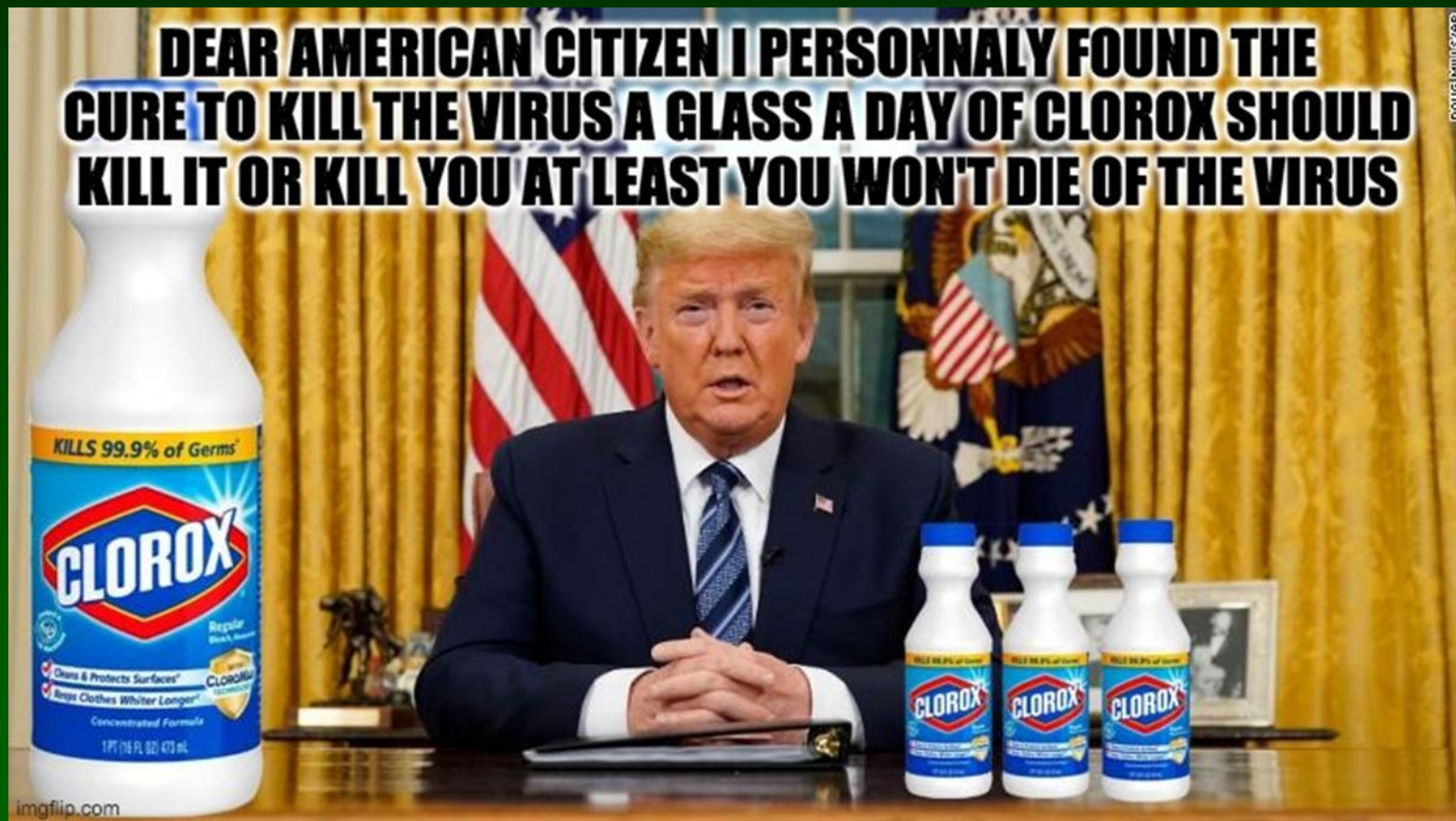
Sieci neuronowe tak mają.

Polityczna poprawność fałszuje historię: Google Gemini wymyślił czarnych papieży i kobiety w gronie Ojców Założycieli USA.

Google AI overviews: mała skałka raz dziennie, klej do pizzy ...

LLM wystarczy skrytykować i się poprawi, z ludźmi jest trudniej.

[Search-Augmented Factuality Evaluator \(SAFE\)](#)



GameNGen

Generatywna inteligencja (mały model dyfuzyjny) wyobraża sobie grę w swojej pamięci, tworzy i wykonuje grę Doom bez programowania. <https://gamengen.github.io> Nie mamy takiej pamięci ...

move forward



open door



attack



move forward



attack



turn right



run back



Gene2: Imagery, virtual worlds

Genie 2: A large-scale foundation world model - Google DeepMind

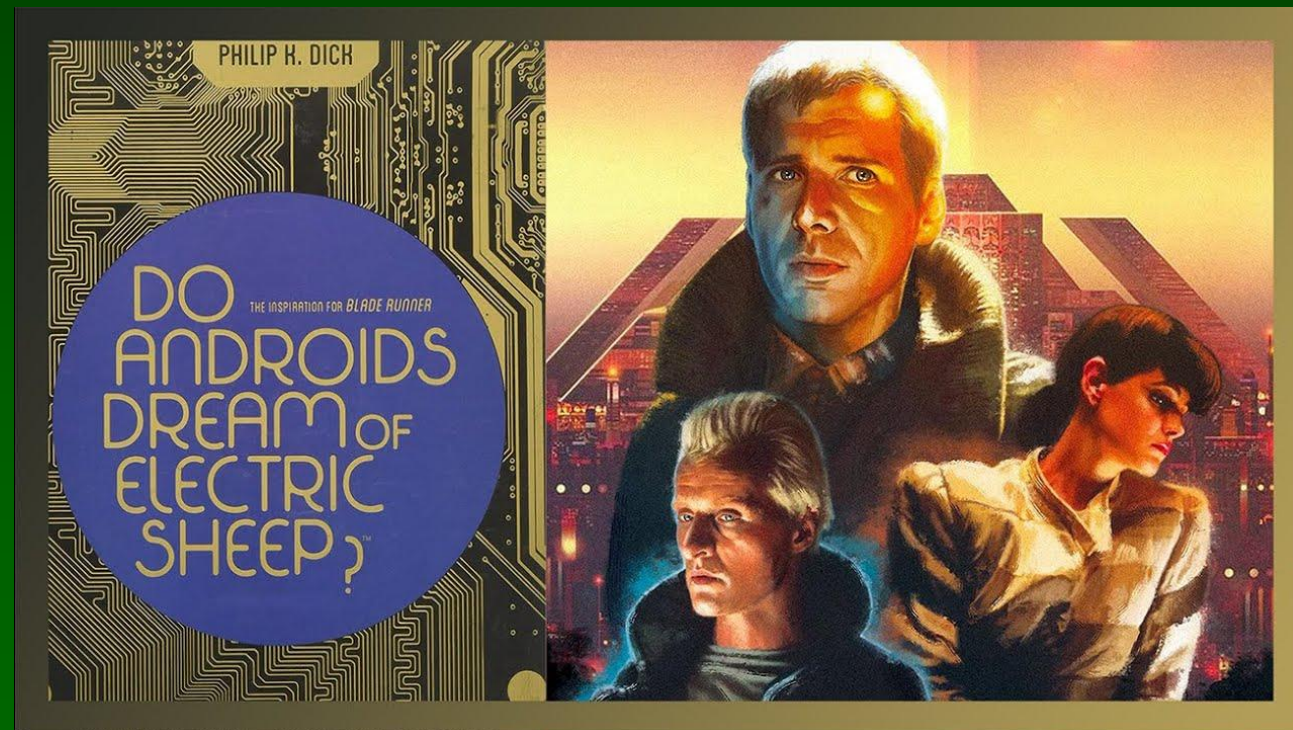
Google Genie 2 to podstawowy model świata, który może tworzyć nieskończoną różnorodność spójnych, sterowanych akcją, grywalnych środowisk 3D do szkolenia i oceny ucieleśnionych agentów. Agenci ci mogą wchodzić w interakcje między sobą i z ludźmi.

Agenci Genie 2 mogą śnić, poruszając się w stworzonym przez siebie wirtualnym świecie.

Burchi, M., & Timofte, R. (2025).
Learning Transformer-based World Models with Contrastive Predictive Coding.

[arXiv:2503.04416](https://arxiv.org/abs/2503.04416)

Algorytm DreamerV3 uczy się dokładnego modelu świata opartego na rekurencyjnych sieciach neuronowych (RNN) i modelach świata opartych na transformatorach przy użyciu maskowanej uwagi własnej.



Generative music

„Popcorn”, Hot Butter hit from 1969,
Created through computer analysis of various melodies.
Computer helping human musicians.



Musicians learn by hearing and analyzing.

Why computers should not do the same? With access to all music in the world?

Algorithmic music: John Cage, Iannis Xenakis ...

http://en.wikipedia.org/wiki/Algorithmic_music

Lucasfilm Games' gra Ballblazer (1982) używała komputerowej improwizacji jazzowej, Xbox improwizuje na zadane tematy.

Generative music: Brian Eno started to experiment in 1975!

http://en.wikipedia.org/wiki/Generative_music

Examples of generative music:

<http://jmusic.ci.qut.edu.au/jmMusic.html>



From images to music

Can music be composed converting images?

Example: finish “metacomposer” Lauri Gröhn:

[Synesthesia Software Music](#)

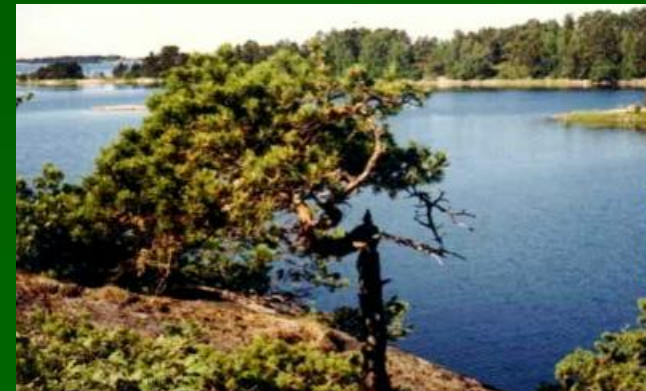
„I want to show that in future computers may compose as well as people”.



[Turku Archipelago](#), Fog, Sakura – interesting compositions!

Can computer really „feel” music?

[Greta ist trying...](#)



Flow Machines

by Sony CSL

<http://www.flow-machines.com>

Software that learns.
What is programmed
is only the motivation
to learn!

[Learning to Create](#)
(Lrn2Cre8) EU FP7
Project

Daddy's Car: a song composed by Artificial Intelligence - in the style of the Beatles



flowmachines
flowmachines

Immersja: Świat Wirtualny



Wzrok, słuch, dotyk, zapach, ruch ... świat rzeczywisty nie jest tak interesujący!
AI uczy się zachowania w wirtualnym (VR) i rzeczywistym (AR) świecie analizując
nasze doświadczenia i wymieniając między sobą.
Projekty Samsung + Google.

Othello-GPT

Czy to tylko statystyka czy model wewnętrzny świata?

- Othello-GPT wytrenowano tak, by przewidywał kolejne ruchy w sekwencjach, które pochodziły z realnych gier. Model nie miał żadnych informacji co to jest.
- Użyto prostej sieci neuronowej jako klasyfikatora podając jej aktywacje warstw ukrytych sieci Othello-GPT.
- Można zauważyć, że te aktywacje układają się w obraz planszy, która wynika z dążenia do maksymalnej kompresji informacji

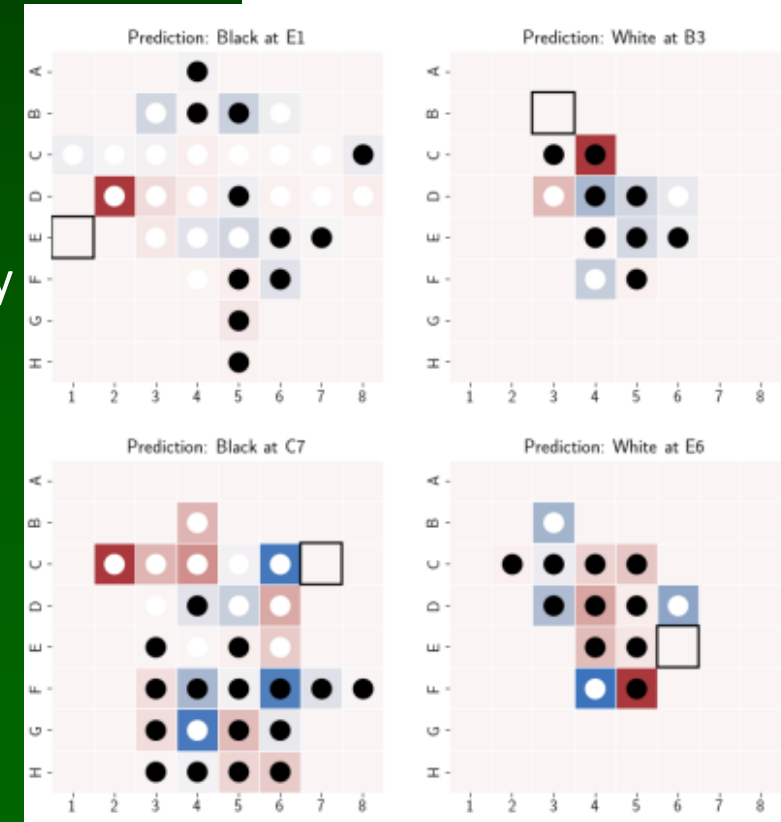
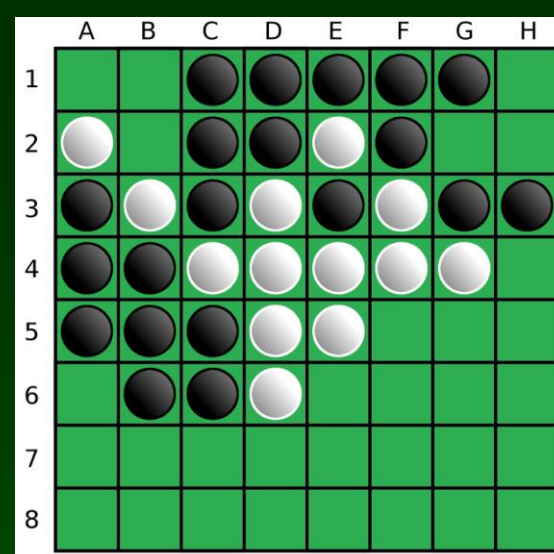
Sieć gra w Othello podobnie jak człowiek: utrzymując wyobrażenie planszy w swoim “oku umysłu” by nauczyć się szukać jak najlepszych ruchów.

- GPT postrzega to, co pojawia się w wyobrażeniu tworzonym przez model i może to komentować.

John Locke: świadomość to percepcja tego co się dzieje w naszym umyśle.

Czemu nie miałyby dojść do autorefleksji?

Reprezentacji „ja”, formowania się osobowości?

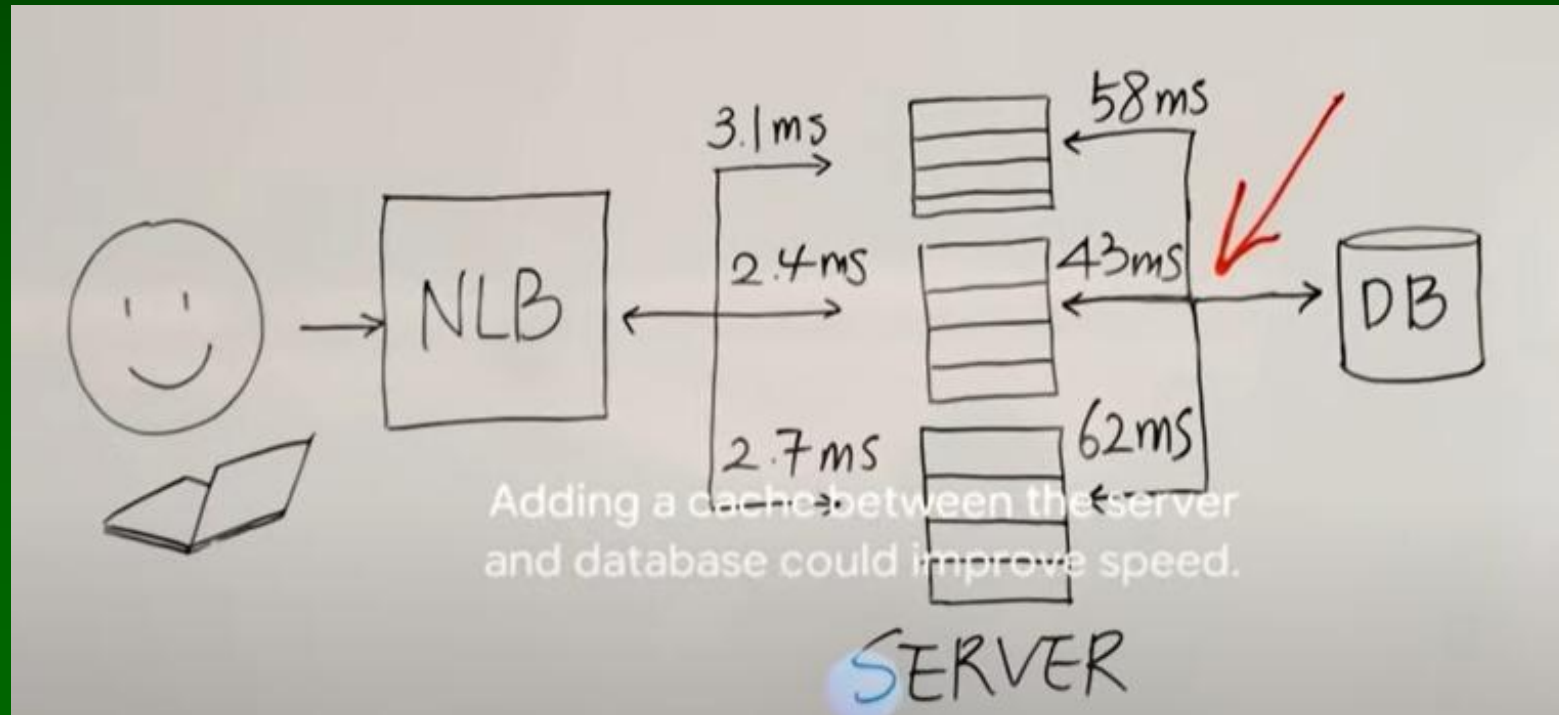


Google Project Astra

Orientacja i rozumowanie przestrzenne pozwala rozumieć intencje, gesty i szkice, zamienić rysunki na diagramy, prowadzić inteligentną konwersację, rozumieć co się w świecie dzieje, używać komputera.

Project Astra video – LMM ma pełną **świadomość sytuacji, embodiment**

Gemini widzi przez okulary AR rysunek na tablicy i odpowiada na pytanie: „co mogę tu dodać by system szybciej działał”.



Okulary AR zbierają i wyświetlają informację, AI się uczy „bycia w świecie”. Pamięć interakcji, rozmów i sytuacji => subiektywność.

VR + EEG, BCI + AI



VR glasses are now comfortable, soft contact lenses are coming.
Augmented/virtual reality + EEG + AI = the world of Alice in Wonderland.

OR REMIND ME TO BUY THIS GIFT
FOR MY FRIEND

META RAY-BAN UPDATES

- Meta AI video allows glasses to see in real-time
- Includes live translation and location awareness
- Integrations with Spotify, Amazon Music, iHeart

**BREAKING
NEWS**

**META ANNOUNCES NEW CAPABILITIES
FOR RAY-BAN SMART GLASSES**

[AI glasses](#)

[Introduced 17-18.09.25](#)

AI in art

- [This Person Does Not Exist](#) artificial faces.
- Open AI [Dall-e2](#), od opisu do obrazu
- [Digital Humans](#).
- The [First Thinking Sculpture](#): Inspired by Gaudí, by Watson.
- [Autodraw](#), drawing from poor sketch.
- [AI artwork sells](#) for \$432,500 (12/2018); now selling NFT tokens
- [10 Amazing Examples Of How Deep Learning AI Is Used](#)
- [15 examples of AI in marketing](#).
- Reports on AI: The [European Artificial Intelligence landscape](#),
- [EU Digital Transformation 2021-27](#) plans
- [Communication: Artificial Intelligence for Europe](#)
- [Digital Poland: AI strategies \(in Polish\)](#)
- [My Flipboard Art & AI](#) and [Flipboard AI-CI-ML](#)

Computational Creativity

AIVA – AI Virtual Artist, registered in France with an author's rights society (SACEM), many compositions in different styles.

AIVA YouTube channel and SoundCloud channel.

Google Magenta for music, CNET news.

6 AI Tools for music generation

Deep Art: <https://deepart.io/>

27 Incredible Examples Of AI And Machine Learning In Practice

What's the purpose of humanity if machines can learn ingenuity?
(extract from Marcus Du Sautoy's book, The Creativity Code).

Open AI Research and Google Deepmind

Tutorials on ML

- Autocoder explanations

<https://www.youtube.com/watch?v=qJeaCHQ1k2w>

- More ML tutorial videos
- Deep Learning automated

<https://www.youtube.com/@Deepia-ls2fo>

Generatory muzyki

Generator	Mocne strony	Słabe strony	Najlepsze zastosowanie
Suno	Piosenki z wokalem, szybkie generowanie, łatwe prompty	Mniejsza kontrola nad szczegółami aranżacji, czasem sztuczne teksty	Piosenki pop, rock, EDM, demo utworów, szybkie prototypowanie
Udio	Bardzo dobra jakość wokali i miksu, kontrola struktury	Wolniejsze generowanie, większa krzywa uczenia	Rozwijanie pomysłów, profesjonalne utwory
Stable Audio	Wysoka jakość, przejrzystość licencyjna	Słabsze wokale i pełne piosenki	Soundtracki, ambient, efekty dźwiękowe, media
MusicGPT	Kompozycja, wokale, aranżacja	Mniej sprawdzony ekosystem	Wszechstronna produkcja muzyki AI
SOUNDRAW	Edycja struktury utworu, royalty-free	Mniej kreatywne wokale	YouTube, podcasty, reklamy
AIVA	Kompozycje orkiestrowe	Mniej nowoczesnych stylów wokalnych	Muzyka filmowa, klasyczna, soundtracki
Loudly	Royalty-free, szybkie generowanie, łatwe użycie	Niższa jakość niż Suno/Udio	Content creation, social media
Eleven Music	Głosy ElevenLabs, nacisk na licencjonowanie	Młodszy produkt, mniej sprawdzony	Piosenek z wokalem do zastosowań komercyjnych

Generatory obrazów

Generator	Mocne strony	Słabe strony	Najlepsze zastosowanie
ChatGPT Images / GPT Image	rozumienie promptów, iteracyjne poprawki, edycja obrazów	mniej „artystycznego chaosu” niż Midjourney	ilustracje, diagramy, prototypy, edycje
Midjourney	świetna estetyka, styl filmowy, ilustracyjny i koncepcyjny	mniejsza precyzja techniczna, zamknięty ekosystem	grafika artystyczna, concept art
Ideogram	bardzo dobre napisy w obrazach	słabszy realizm fotograficzny	plakaty, logotypy, grafiki z tekstem
Adobe Firefly	integracja Photoshopa, komercyjne, modele własnego stylu	mniej efektowny niż Midjourney	marketing, design, firmy
Stable Diffusion / Stable Image	duża kontrola, możliwość lokalnego uruchomienia, LoRA, ControlNet	wymaga wiedzy technicznej	badania, własne modele, prywatność
FLUX	bardzo wysoka jakość, dobry realizm, otwarty ekosystem	lokalnie wymaga mocnego GPU	fotorealizm, eksperymenty, R&D

Generatory obrazów

Generator	Mocne strony	Słabe strony	Najlepsze zastosowanie
FLUX	bardzo wysoka jakość, dobry realizm, otwarty ekosystem	lokalnie wymaga mocnego GPU	fotorealizm, eksperymenty, R&D
Leonardo AI	dużo gotowych narzędzi, style, upscaling, workflow dla twórców	mniej precyzyjny niż wyspecjalizowane modele	gry, social media, assety
Canva AI	łatwe użycie, integracja z projektami	ograniczona kontrola	prezentacje, posty, grafiki edukacyjne
Runway	dobre połączenie obrazu i wideo	bardziej platforma video niż czysty generator obrazów	storyboardy, animacje, reklamy
Freepik AI	szybkie generowanie, bank stockowy + AI	jakość zależna od modelu	grafiki komercyjne, materiały promocyjne
NightCafe	społeczność, wiele modeli, proste użycie	mniej profesjonalny	zabawa, szybkie eksperymenty
Playground AI	proste edycje	mniej zaawansowany	szybkie prototypy wizualne

Pytania

- Jak działają modele dyfuzyjne
- Czym różnią się obecne model generatywne od pierwszych prób tworzenia obrazów, takich jak Aaron?