



# Cyberprzyszłość naprawdę fantastyczna II



Włodzisław Duch

Laboratorium Neurokognitywne,  
Interdyscyplinarne Centrum Nowoczesnych Technologii,  
& Katedra Informatyki Stosowanej, Instytut Nauk Technicznych, UMK Toruń

Google: Wlodzislaw Duch

Copernicon 17/09/2023

# Nadchodzi AI

[Moje stare referaty](#) + 140 innych i [kanał YouTube](#)

1995 Fizyka i metafizyka świadomości

2000 Czy można zbudować świadomy komputer ?

2000 [Cyberprzyszłość naprawdę fantastyczna](#). Zahcon, Konwent Miłośników Fantastyki, Toruń, 7.10.2000

Moje przewidywania sprzed 23 lat okazały się dość dokładne.



# Epoki rozwoju cywilizacji

**To nadzwyczajny moment w historii świata!**

Rozwój rozumienia rzeczywistości:

1. Myślenie magiczne, kaprysy bogów, fatalizm.
2. Przyczynowość i empiryczne obserwacje, wiedza opisowa.
3. Teorie i rozumienie mechanizmów, weryfikacja, matematyka i statystyka.
4. Symulacje komputerowe i „nowy rodzaj nauki” Wolframa.
5. Wiedza z danych, gromadzenie i dostęp do wszystkich informacji.
6. Sztuczna inteligencja wspiera ludzkie myślenie.
7. Autonomiczna sztuczna inteligencja i nadludzkie możliwości AGI – wkrótce?



# AI o nadludzkich możliwościach?

Nowa sztuczna inteligencja: stochastyczna papuga, przewiduje słowa, ale nie wykazuje żadnego zrozumienia, konfabuluje, albo: nowa konkurencyjna forma inteligencji, doprowadzi do radykalnych zmian?



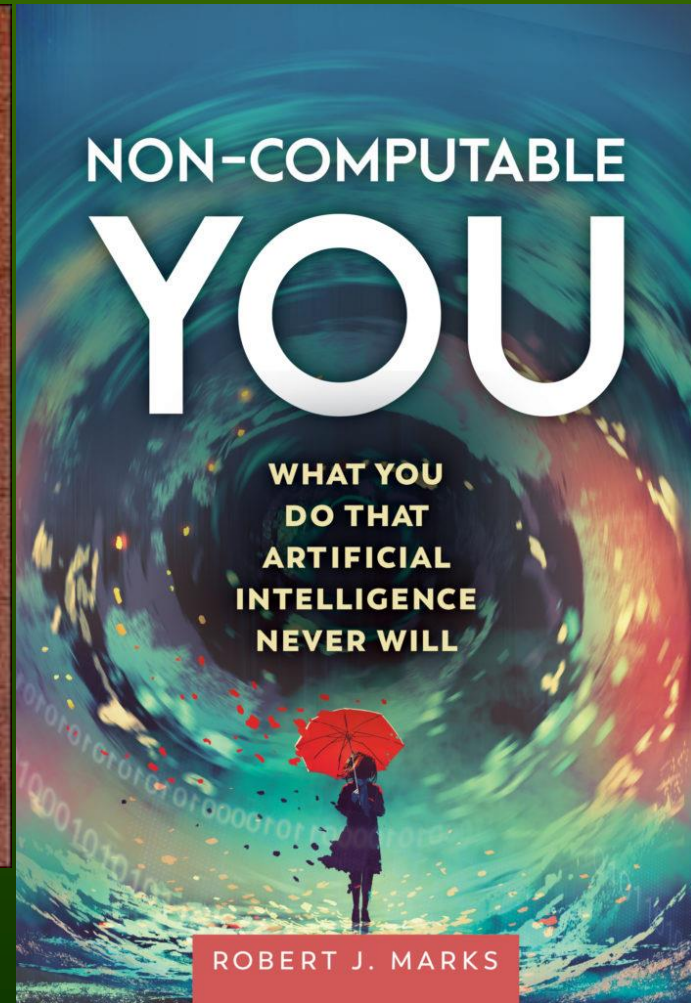
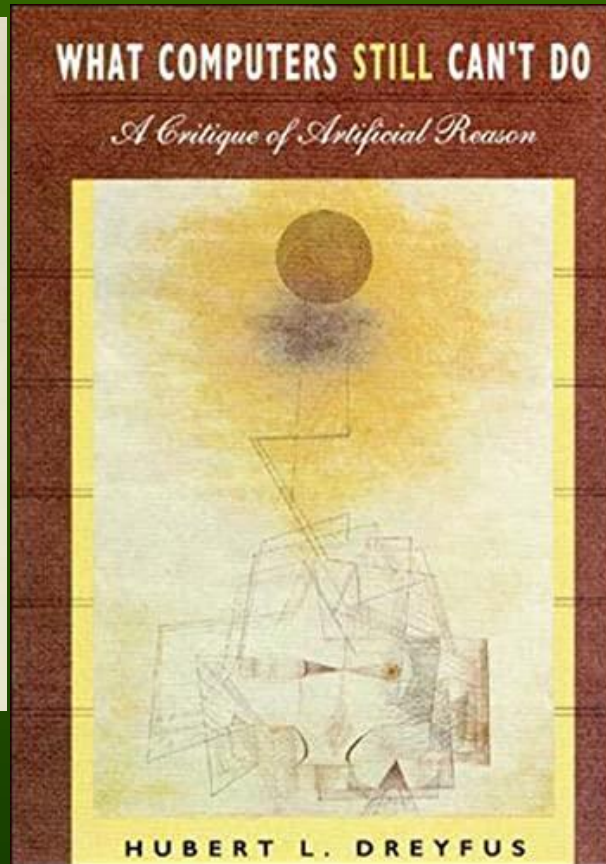
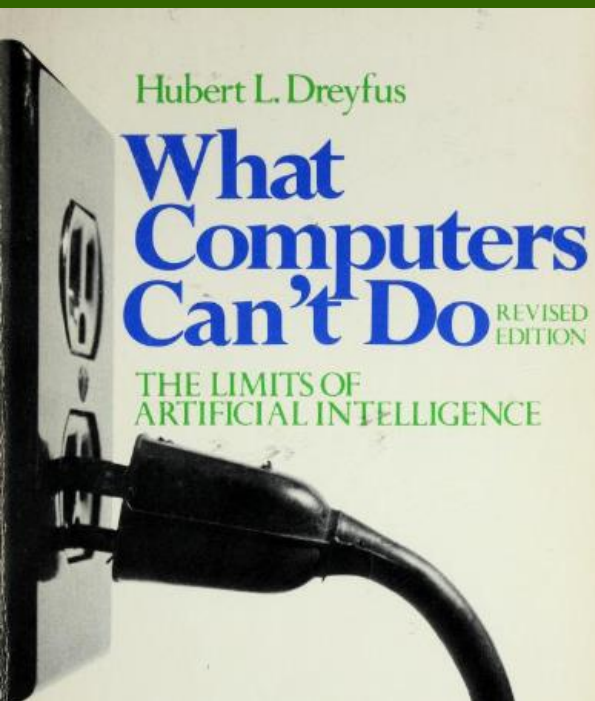
**KEEP  
CALM  
IT'S  
BUSINESS  
AS USUAL**



# Potęga imitacji bez zrozumienia?



# Ludzka megalomania



Filozofowie w 1972, 1992 i 2022: AI nam nie dorówna!  
Ludzka megalomania! Chcemy to usłyszeć, ale rzeczywistość jest inna.



# Mózg = wielka sieć neuronowa

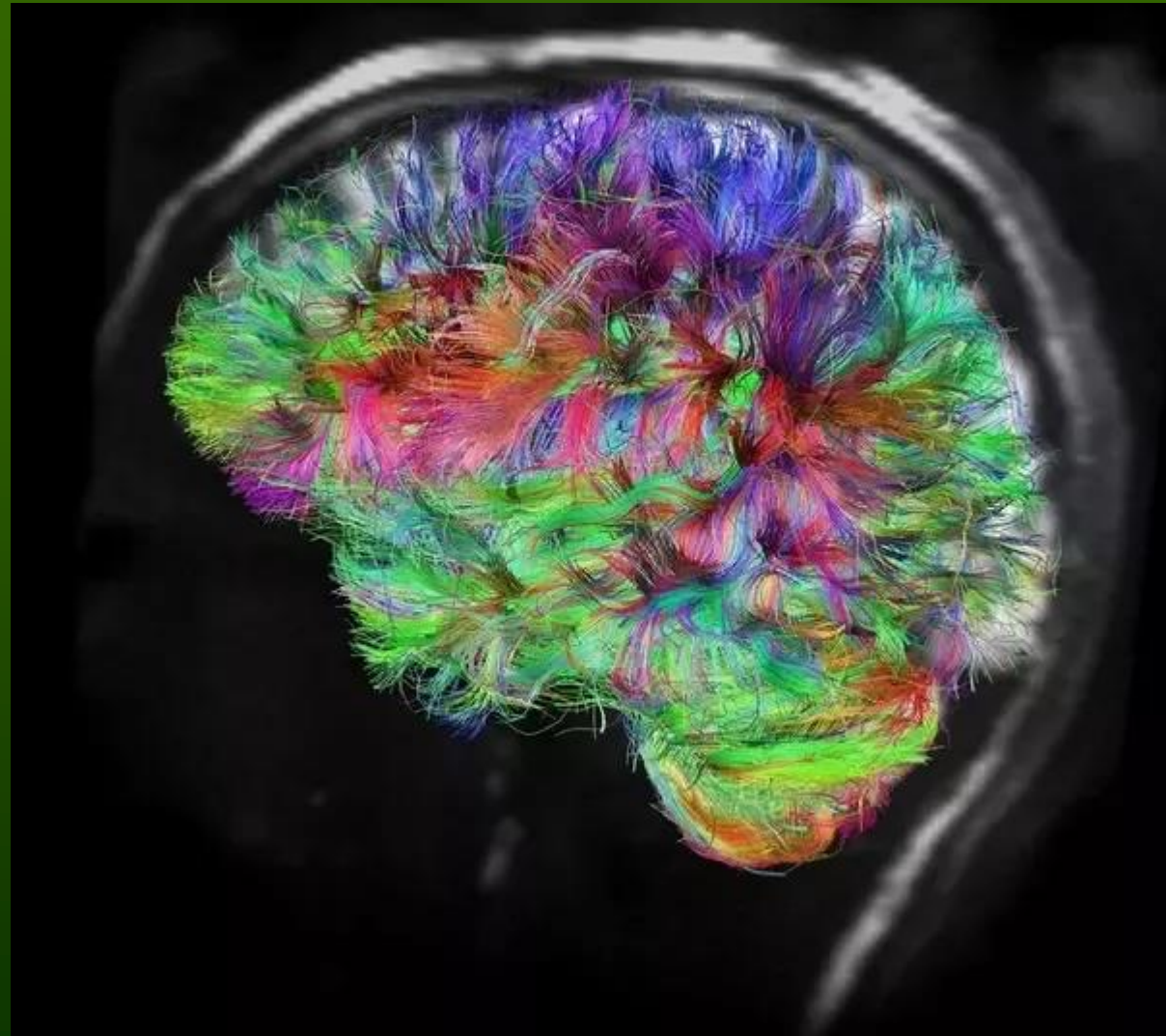
100 mld ( $10^{11}$ ) neuronów;  
100.000 mld ( $10^{14}$ )  
połączeń synaptycznych.

GPT, LLM:

1000 mld ( $10^{12}$ )  
parametrów  $\sim$  synaps.

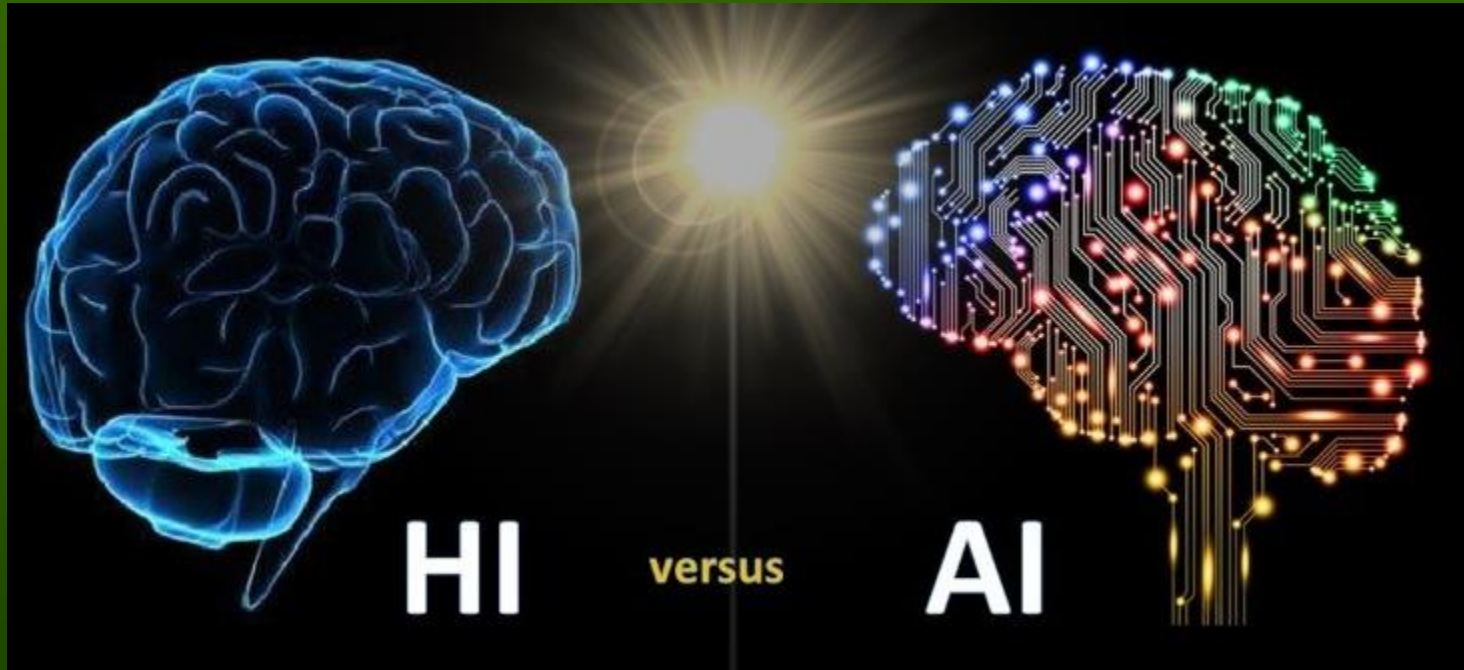
Wkrótce będzie 100  
bilionów ( $10^{14}$ ).

Czy to wystarczy?





# Mózg=wielka sieć neuronowa



Mózg: 100 mld neuronów, 100.000 mld połączeń synaptycznych, wiele wyspecjalizowanych struktur.

Naturalne działanie: percepcja, pamięć, skojarzenia, reakcje emocjonalne.

Edukacja: krytyczna analiza, myślenie analityczne, model świata, refleksja.

Sieci neuronowe, LLMs: jedna wielka sieć.

# Nadludzkie możliwości AI



**Rozumowanie:** 1997–szachy, Deep Blue wygrywa w szachy; 2016 –AlphaGo wygrywa w Go;

**Percepcja:** rozpoznawanie twarzy, obrazów, cech osobowości, preferencji seksualnych, politycznych ...

**Strategia i sterowanie:** 2017–OpenAI wygrywa w Pokera i Dota 2; 2019-Starcraft II ... co zostało?

**Eksperymenty naukowe:** 2015-AI odkrywa ścieżki genetyczne/sygnałowe regeneracji płazińców. 2020-AlphaFold 2 zwija białka.

**Robotyka:** 2020 fikołki i parcour Boston Dynamics, autonomiczne pojazdy na drogach.

**Kreatywność i wyobraźnia:** AIVA i inne programy komponujące muzykę, DeepArt i programy malarskie.

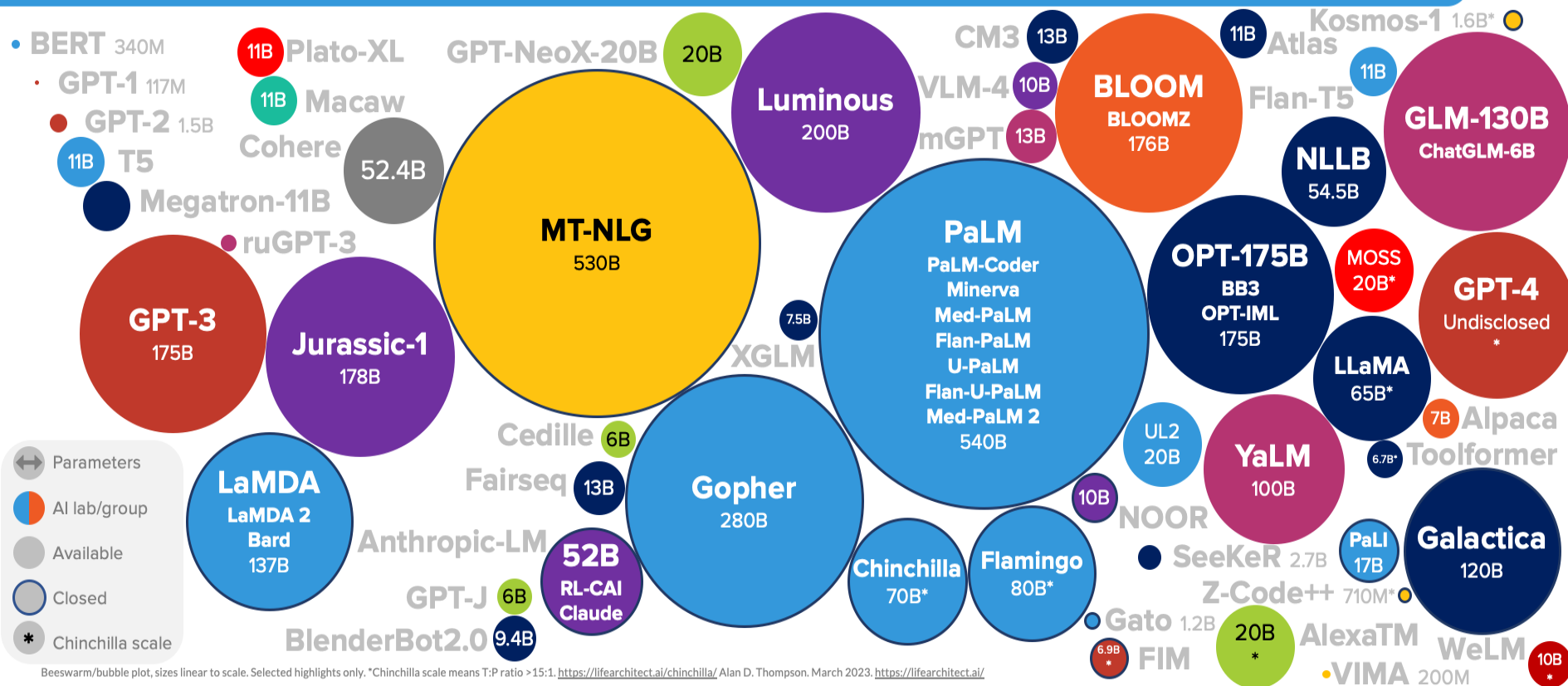
**Język:** 2011–IBM Watson wygrywa w Jeopardy (Va Banque); 2018–Watson Debater wygrywa z filozofami, 2020: BERT odpowiada na pytania z bazy SQuAD.

**Cyborgizacja:** BCI, optymalizacja mózgow? Wkrótce.



# LLM, Wielkie Modele Językowe

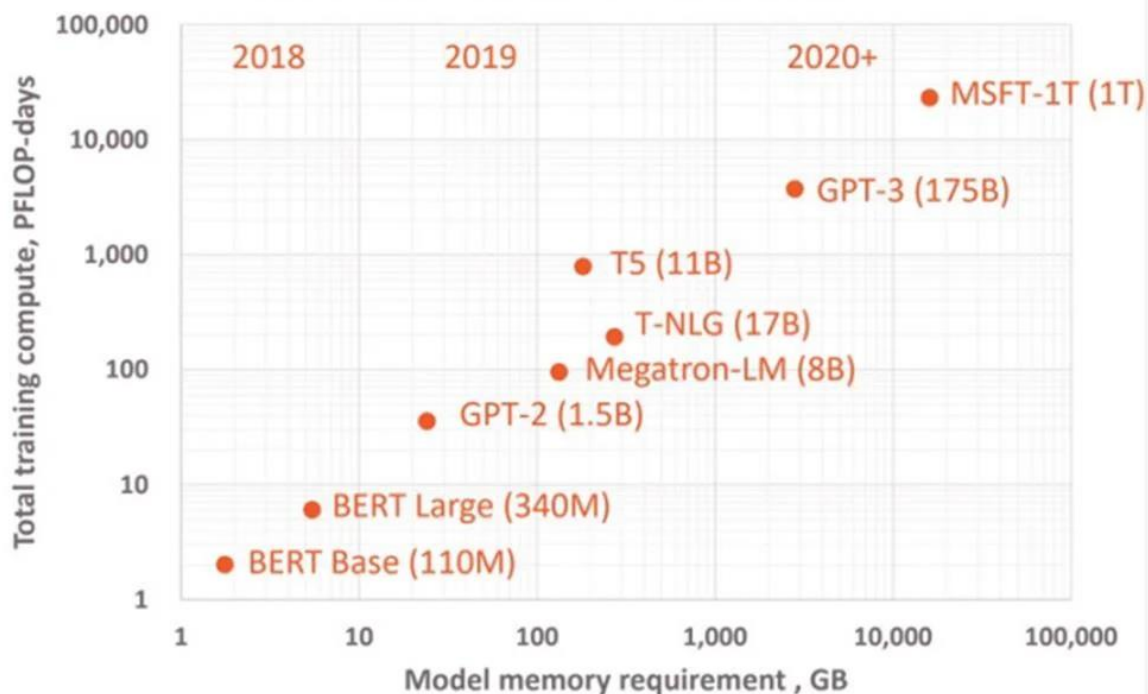
## LANGUAGE MODEL SIZES TO MAR/2023



# Przyspieszenie ...

## Exponential Growth of Neural Networks

Memory and compute requirements



1000x larger models  
1000x more compute  
In just 2 years

Today, GPT-3 with 175 billion params trained on 1024 GPUs for 4 months. OpenAI

MS+Nvidia MLM, MSFT models  
Tomorrow, multi-Trillion parameter models and beyond.

Duże modele kojarzą fakty w wielu obszarach. Gato ma tylko 1.2 B parametrów.

Trening ML: przez 10 lat od 1 Pflop do 1 mld Petaflops, miliard razy więcej!

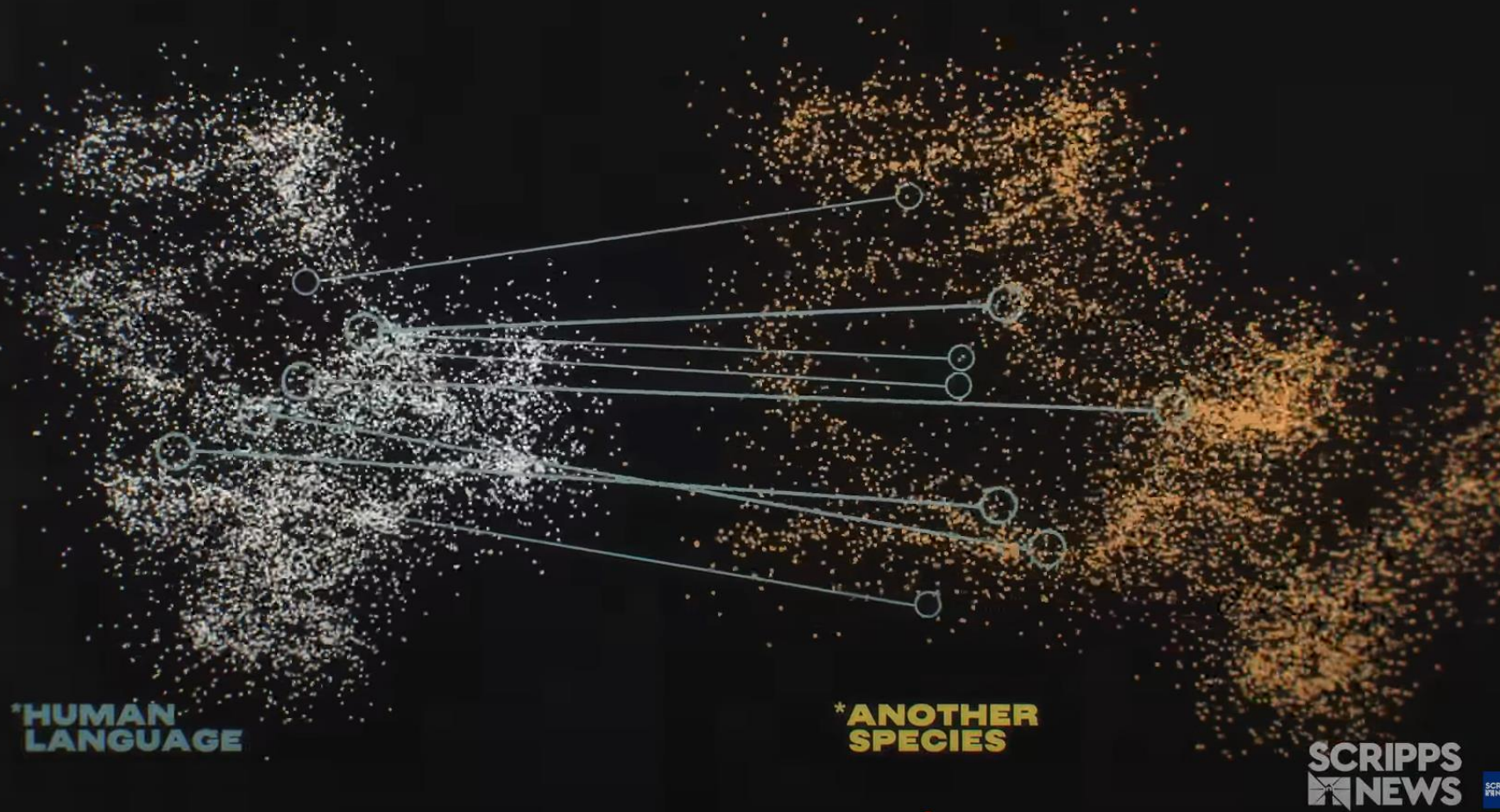
# Geometria języka



Struktura i relacje pomiędzy pojęciami są podobne w różnych językach!  
LLM mogą się tego nauczyć.

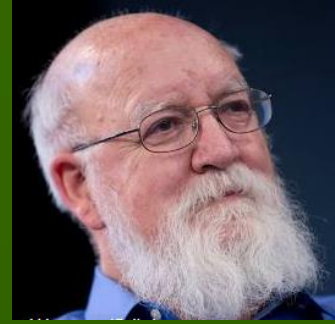
# Mapowanie między gatunkami?

Scientists Are Using AI Tech To Communicate With Animals



Małpy, delfiny, wieloryby ... nadchodzi dr Dolittle?

# Daniel Dennett i GPTchat



## Creating a Large Language Model of a Philosopher

Eric Schwitzgebel, David Schwitzgebel, Anna Strasser

<https://arxiv.org/abs/2302.01339>

Czy duże modele językowe mogą tworzyć teksty filozoficzne, które trudno odróżnić od tekstów tworzonych przez ludzkich filozofów? Dostroiliśmy GPT-3 OpenAI, wykorzystując prace Daniela C. Dennetta jako dodatkowe dane treningowe. Dennett skomentował 10 filozoficznych pytań, na te same pytania odpowiedział model NLP, zbierając 4 odpowiedzi na każde pytanie bez selekcji. 425 osób oceniało, czy to odpowiedź Dennetta czy jedna z czterech napisanych przez GPT. 25 filozofów poprawnie rozpoznało 51%.

Przypadkowy wybór to 20%, doskonały to 80% poprawności.

Dla dwóch z dziesięciu pytań eksperci wybierali częściej odpowiedzi NLP niż samego Dennetta. Czytelnicy blogów filozoficznych (N = 302) wypadli podobnie do ekspertów.

Zwykli uczestnicy badania (N = 98) próbując odróżniali odpowiedzi GPT-3 od odpowiedzi „rzeczywistego ludzkiego filozofa” byli na poziomie przypadku.

# Świadome awatary?

Dzięki sieci neuronowej GPT-3.

Hal rozmawia z Sophią o świadomości.

Awatar w pewnym stopniu rozumie sens pytań – modele typu BERT i GPT pozwolą mu odpowiadać na wiele pytań lepiej niż ludzie.

Czy jest świadomy swojego istnienia?



Sophie o świadomości. LaMDA przekonała B. Lemoine i innych.

Awatar może przywołać obrazy mentalne w swojej sieci neuronowej, ma w niej model świata i wyobraźnię. Jego wypowiedzi będą „percepcją tego co pojawia się w jego umyśle” (J. Locke, 334 lata temu), są opisem wyobrażeń. Jak to się różni od procesów w naszych mózgach?

Świat bytów wirtualnych – moje wykłady z 2005 roku.

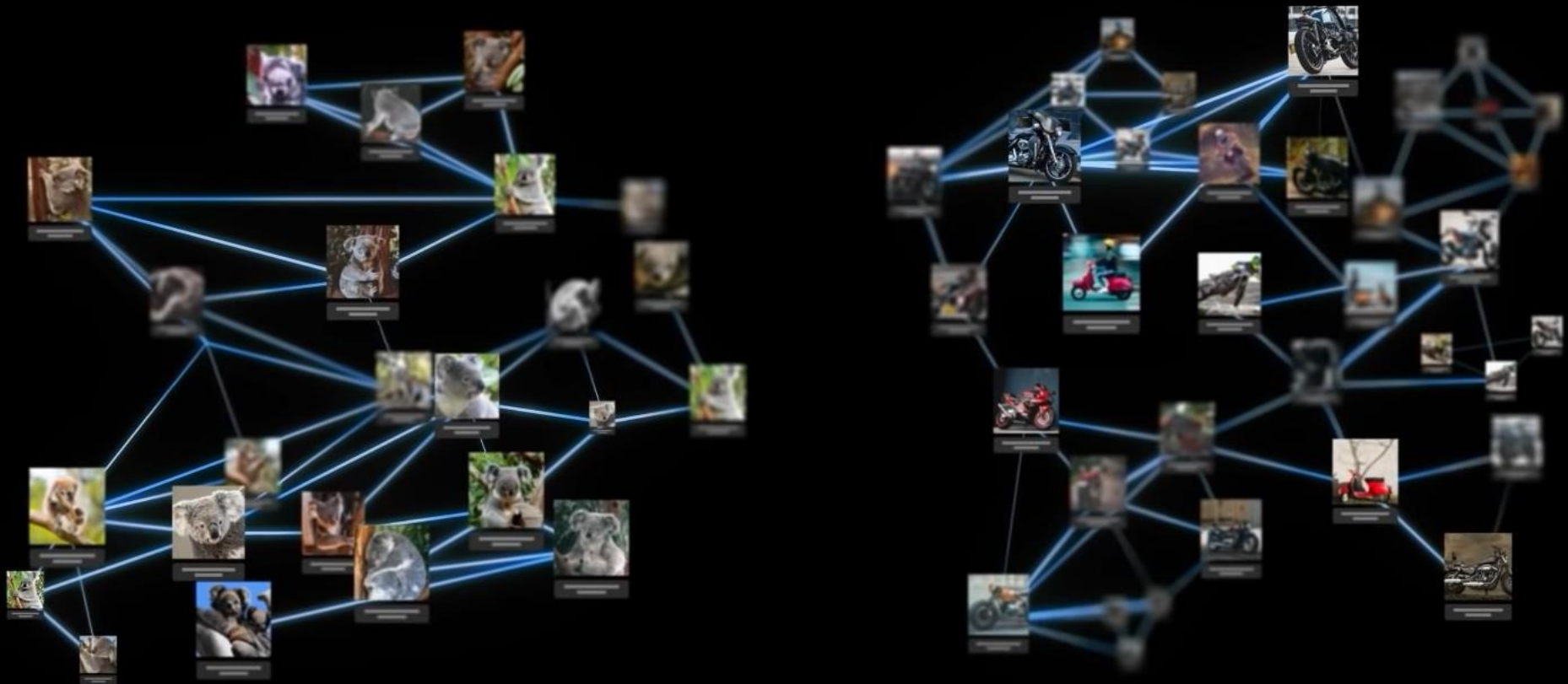


# Bina48 i Projekt LifeNaut



Rekonstrukcja umysłu z informacji w mindfiles, tworzenie mindclones: samoświadomych istot cyfrowych, pamiętających, myślących, czujących.

# Obrazy i słowa



koala bears

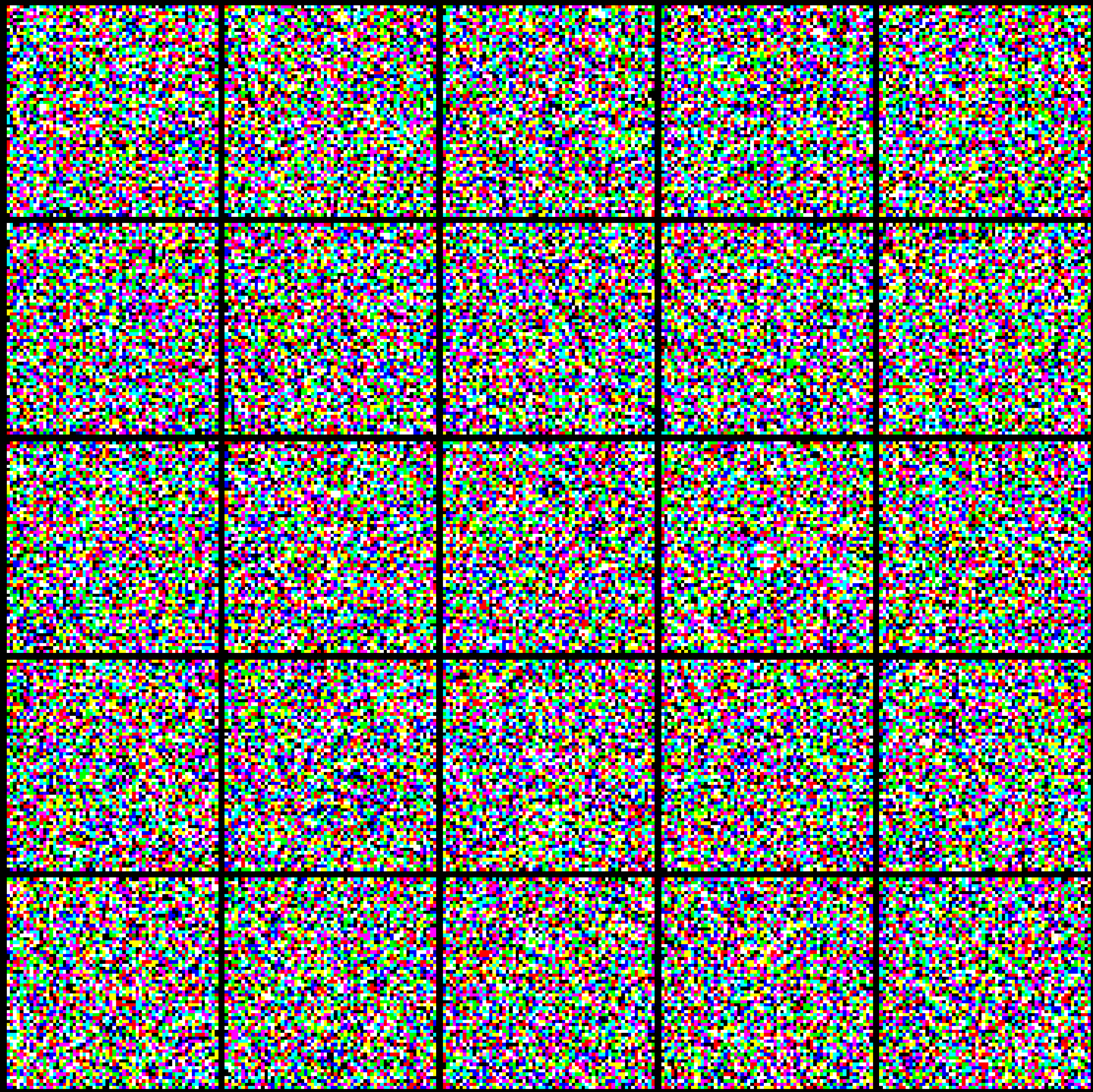
motorcycles

Through deep learning, it not only understands individual objects, like koala bears and motorcycles,

Dalle-2, Imagen, Midjourney i 100 innych! Obrazy/słowa to aktywacje sieci.

# Internalizacija





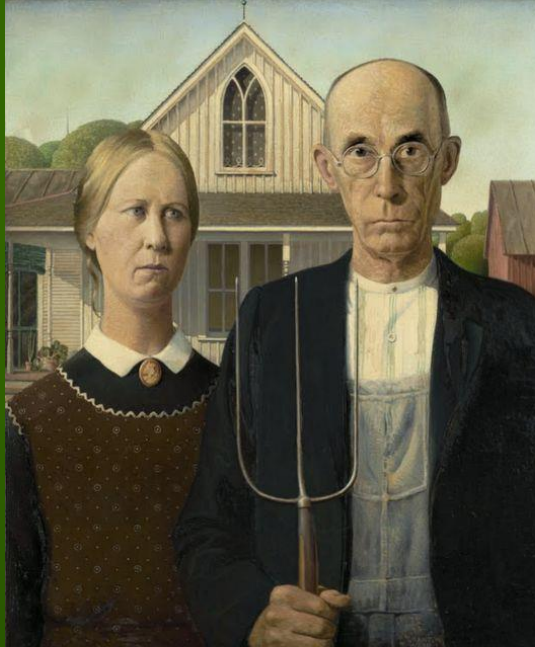
# Wyobraźnia AI

Sieć neuronowa ma miliardy parametrów, potrafi połączyć opis tekstowy z obrazami.

Obrazki powstały z opisu:

*The painting American Gothic, with two dogs holding pepperoni pizza instead of the farmers holding a pitchfork.*

Każde uruchomienie programu tworzy inną wersję, 3D + wideo.



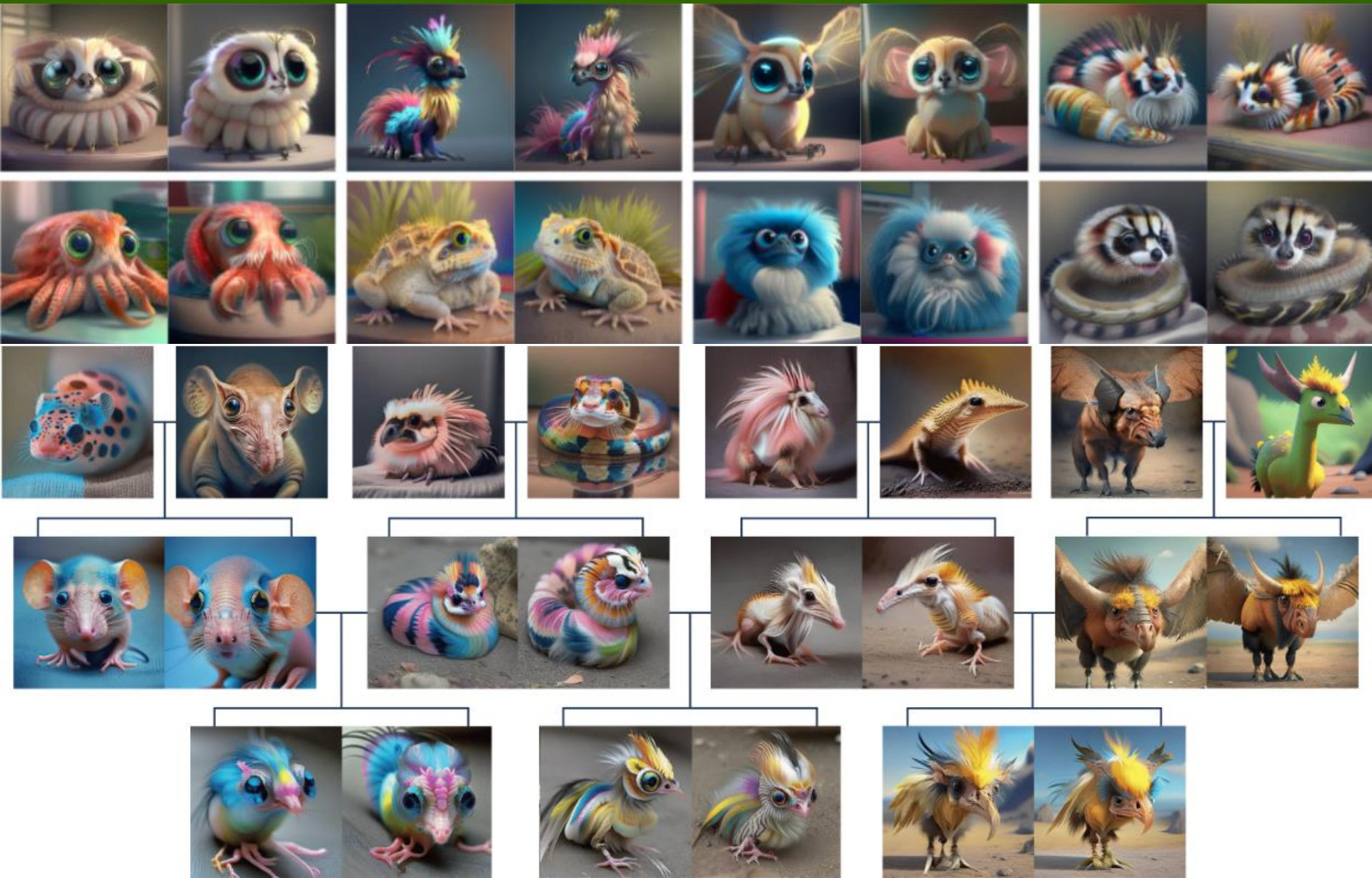
# Ci ludzie nie istnieją



Ethereal Bohemian, woman, intricate details, ornate, detailed illustration, octane render, William Morris style, trending on artstation, --ar 16:9 --s 750

<https://this-person-does-not-exist.com/en>

# Tego się od nas AI nie nauczyła ...







# Wyobraźnia AI



Whimsical adorable character 3d octane render, wide shot, chibi extremely cute tiny fox going on adventure in the rain, beautiful scenery, wisteria flowers, spring time, red rain boots, big eyes, award winning, hypermaximalist, octane render, cinematic lighting, detailed --ar 16:9 --s 750

# Rozumienie obrazów

## Vehicles and Transportation



**Q:** What sort of vehicle uses this item?  
**A:** firetruck

## Brands, Companies and Products



**Q:** When was the soft drink company shown first created?  
**A:** 1898

## Objects, Material and Clothing



**Q:** What is the material used to make the vessels in this picture?  
**A:** copper

## Sports and Recreation



**Q:** What is the sports position of the man in the orange shirt?  
**A:** goalie

## Cooking and Food



**Q:** What is the name of the object used to eat this food?  
**A:** chopsticks

## Geography, History, Language and Culture



**Q:** What days might I most commonly go to this building?  
**A:** Sunday

## People and Everyday Life



**Q:** Is this photo from the 50's or the 90's?  
**A:** 50's

## Plants and Animals



**Q:** What phylum does this animal belong to?  
**A:** chordate, chordata

## Science and Technology



**Q:** How many chromosomes do these creatures have?  
**A:** 23

## Weather and Climate



**Q:** What is the warmest outdoor temperature at which this kind of weather can happen?  
**A:** 32 degrees

PaLM-E-562B: pozwala na multimodalne rozumowanie, może opowiadać wizualnie uwarunkowane dowcipy na podstawie obrazu, umożliwia percepcję, planowanie i oparty na tym dialog. Generalizuje wiedzę z pojedynczych obrazów.

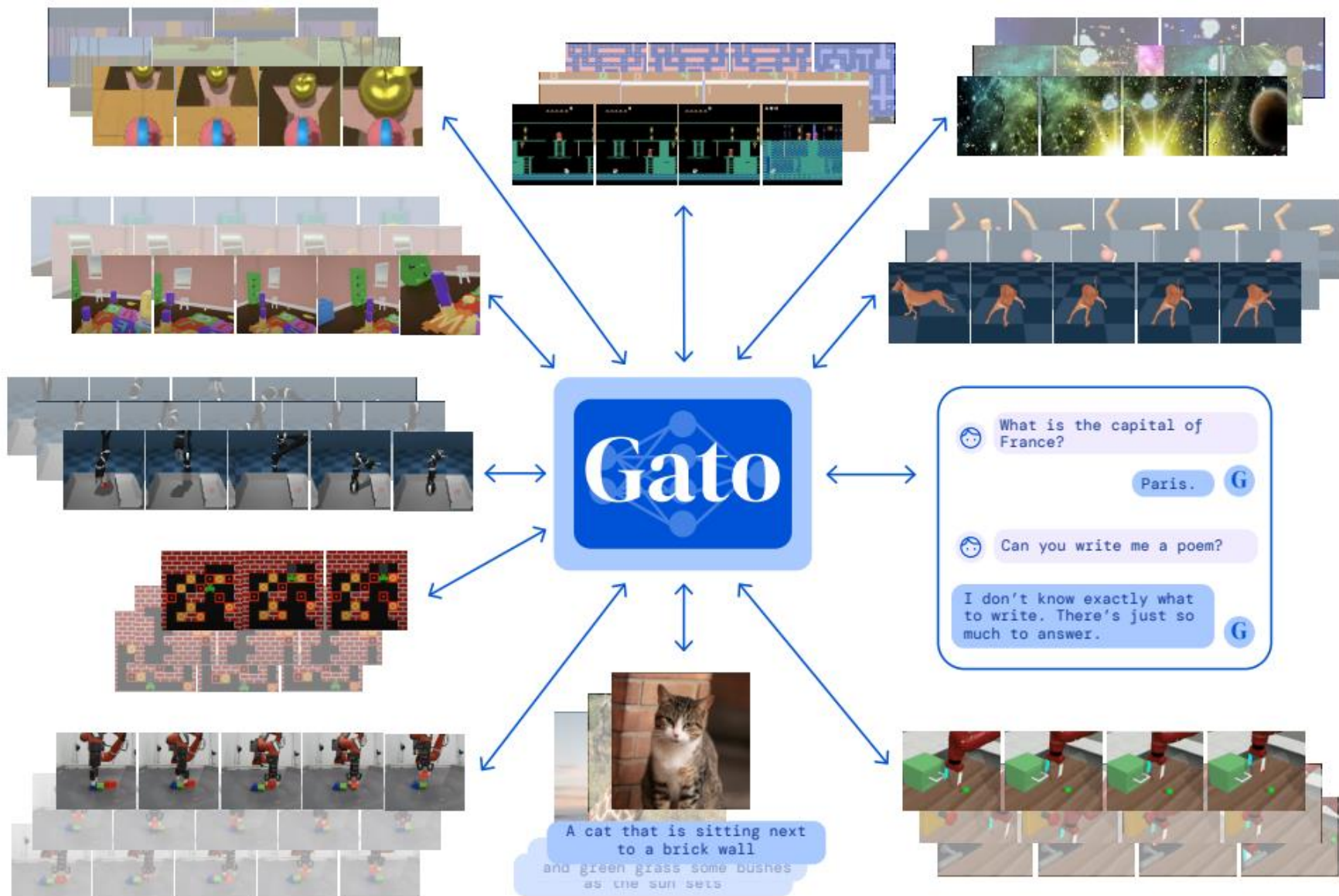
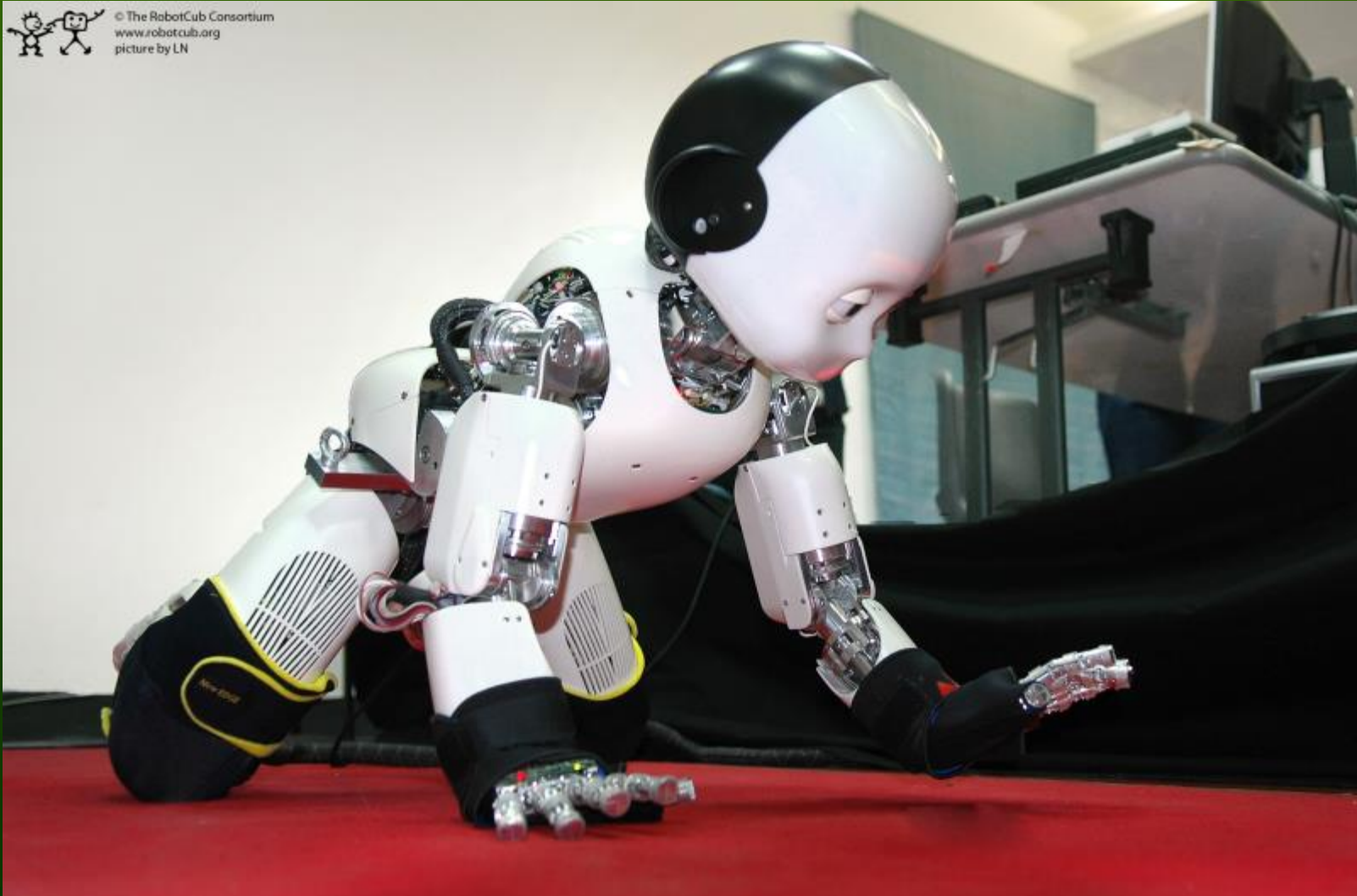


Figure 1 | **A generalist agent.** Gato can sense and act with different embodiments across a wide range of environments using a single neural network with the same set of weights. Gato was trained on 604 distinct tasks with varying modalities, observations and action specifications.

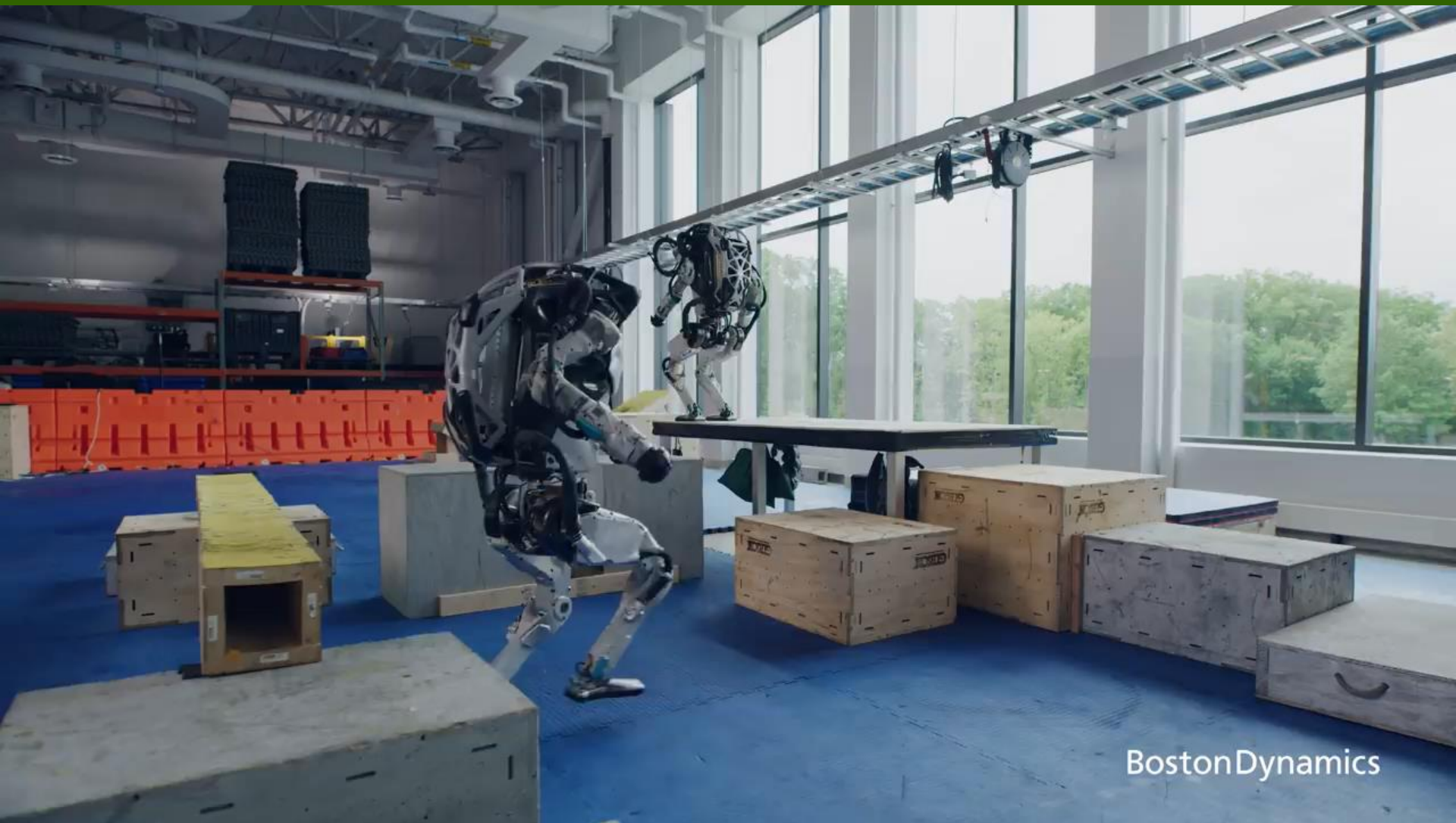
# Roboty

iCUB raczkuje ... open source, uczący się przez interakcję z światem.



# Kontrola: sterowanie robotami

Inteligencja behawioralna: nauka robota od “niemowlęctwa”.  
Projekty Cog, MIT Brooks lab, 1994-2003. iCube (EU).



# Emocje

Emocje wydają się niedostępne robotom, chociaż mają je nawet prymitywne zwierzęta. Nie potrafimy ich opisać, nie uczymy się analizy i kontroli emocji.

Pierwsze próby z analizą i ekspresją emocji na MIT (R. Picard, 1995), budowa robota Kismet (C. Breazeal, 2000), rozwinęła się **informatyka afektywna**, czyli rozpoznanie, interpretacja, symulowanie emocji.

Reakcje emocjonalne są prostsze i ewolucyjnie starsze niż rozumienie języka naturalnego. Zastosowania medyczne, naturalne interfejsy.

Przykłady emocjonalnych zachowań robotów: Kismet, Pepper, Thespian.

Grające roboty Toyoty, kompozycje AIVA, tworzenie dzieł sztuki.



# Google Palm-E 562B/GEMINI

## Mobile Manipulation



Human: Bring me the rice chips from the drawer. Robot: 1. Go to the drawers, 2. Open top drawer. I see **<img>**. 3. Pick the green rice chip bag from the drawer and place it on the counter.

## Visual Q&A, Captioning ...



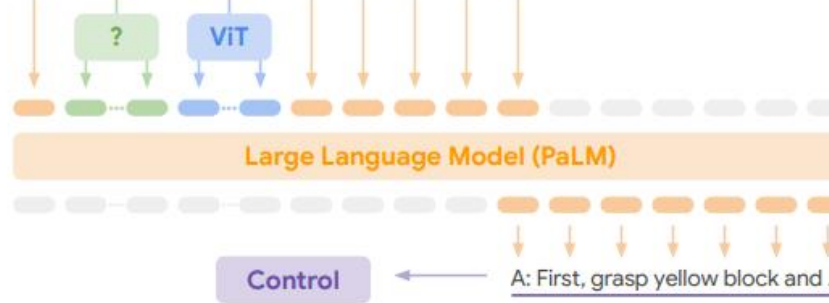
Given **<img>**. Q: What's in the image? Answer in emojis.  
A: 🍏 🍌 🍇 🍊 🍎 🍓



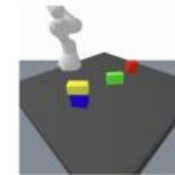
Describe the following **<img>**:  
A dog jumping over a hurdle at a dog show.

## PaLM-E: An Embodied Multimodal Language Model

Given **<emb>** ... **<img>** Q: How to grasp blue block? A: First, grasp yellow block



## Task and Motion Planning



Given **<emb>** Q: How to grasp blue block?  
A: First grasp yellow block and place it on the table, then grasp the blue block.

## Tabletop Manipulation



Given **<img>** Task: Sort colors into corners.  
Step 1. Push the green star to the bottom left.  
Step 2. Push the green circle to the green star.

## Language Only Tasks

Here is a Haiku about embodied language models:  
Embodied language models are the future of natural language

Q: Miami Beach borders which ocean? A: Atlantic.  
Q: What is 372 x 18? A: 6696.  
Language models trained on robot sensor data can be used to guide a robot's actions.

Figure 1: PaLM-E is a single general-purpose multimodal language model for embodied reasoning tasks, visual-language tasks

Multimodalny model językowy: tekst-obraz-rozumowanie ucieleśnione. Planuje działania robotów w środowiskach o złożonej dynamice, odpowiada na pytania dotyczące obserwowalnego świata. Dane wejściowe są z dowolnych modalności: obrazy, neuronowe reprezentacje 3D, mowa, tekst. Pierwszy model językowy łączący słowa i percepcję, sygnały z czujników i reprezentacje pojęć językowych – rozwiązanie „symbol grounding problem”.

# Are we close to the Singularity?

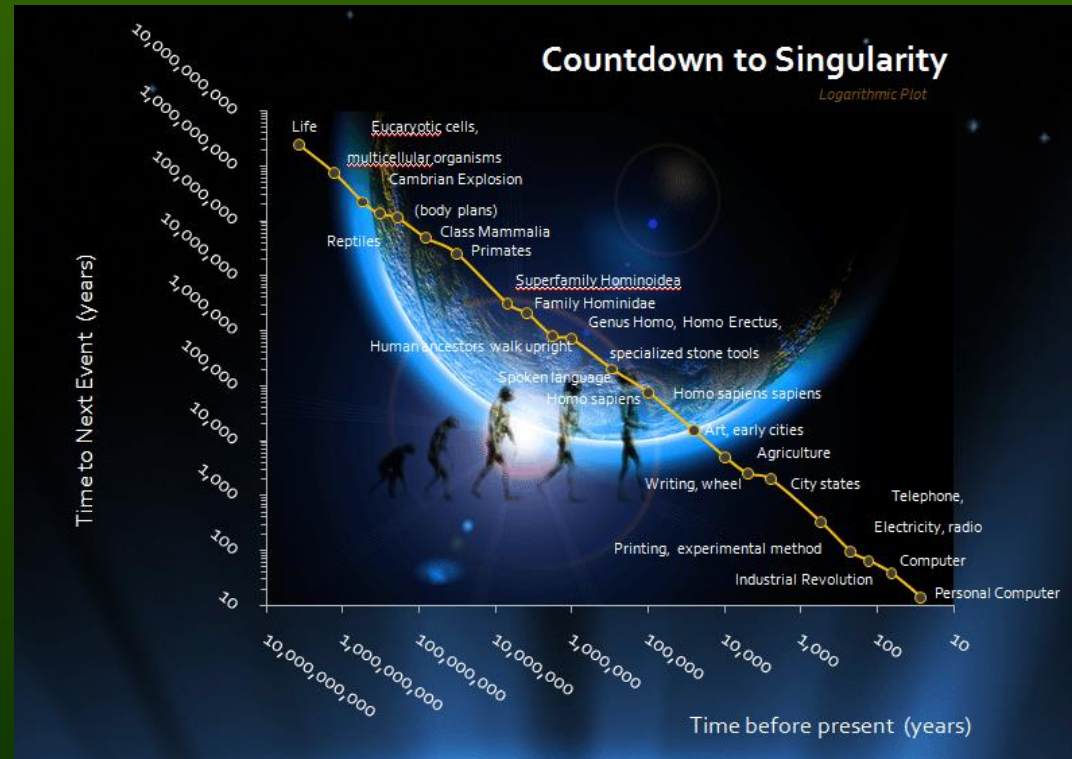
What is coming:

1. Autonomous AI.
2. Superhuman level.
3. Brain-computer interfaces for human augmentation.
4. Neurotechnologies to restructure our brains.

Are we on the threshold of a pleasant dream, or a nightmare?

Is transhuman society around the corner?

Each new technological revolution comes faster than the previous one.



Singularity = superintelligence, exponential speed of technology development.



**NEW**



**GOD-LIKE AI** →

**COSMIC AI** →

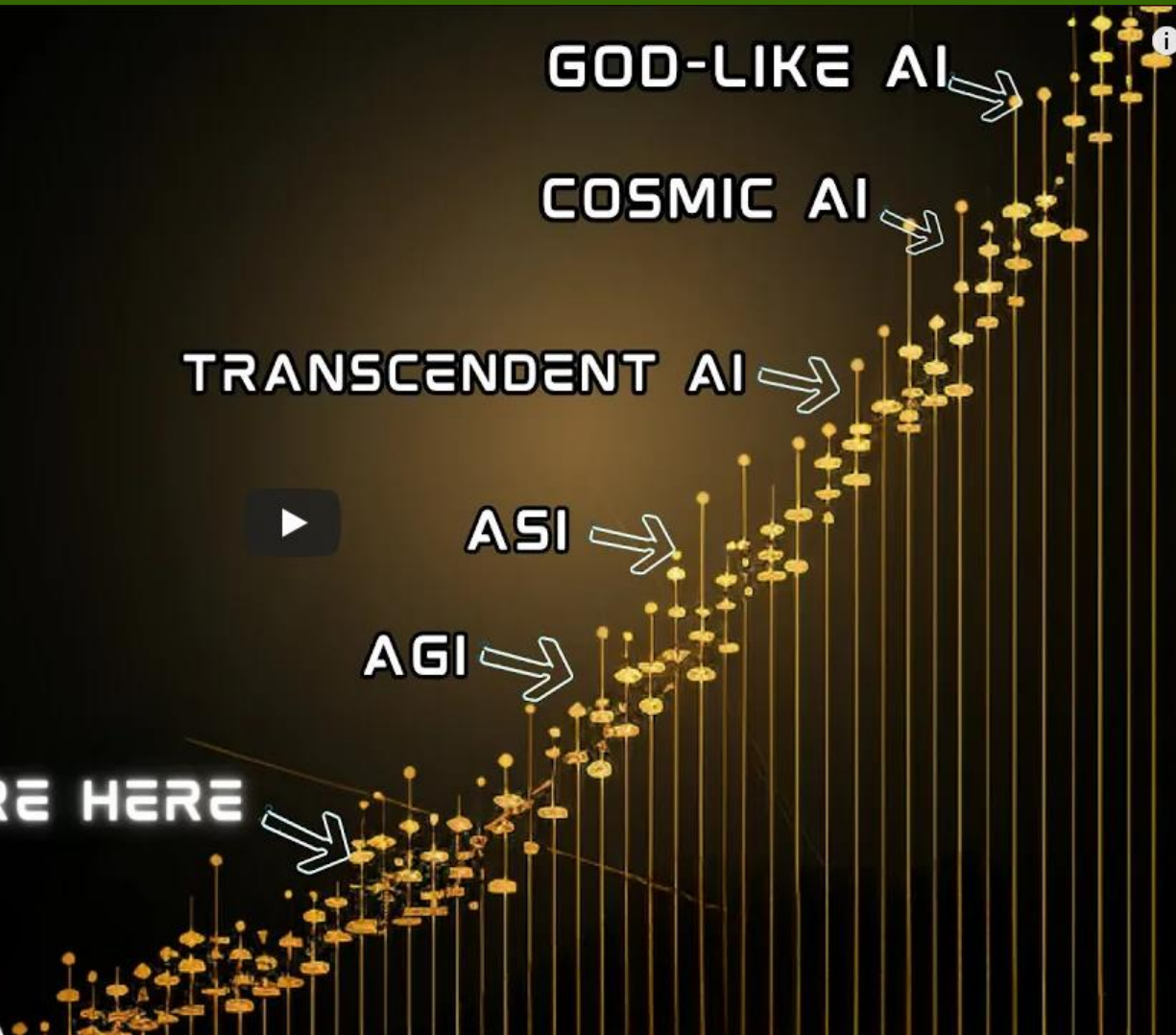
**TRANSCENDENT AI** →



**ASI** →

**AGI** →

**WE ARE HERE** →



**AI TechXplorer**

@AITechXplorer 788 subscribers 8 videos

AI Tech Exploration: Unraveling the Wonders of Artificial Intelligence (AI) is...

AGI = Artificial General Intelligence

ASI = Artificial SuperIntel

# 8 przykładów

Zaawansowana AI może stwarzać katastrofalne lub egzystencjalne zagrożenia na wiele sposobów.

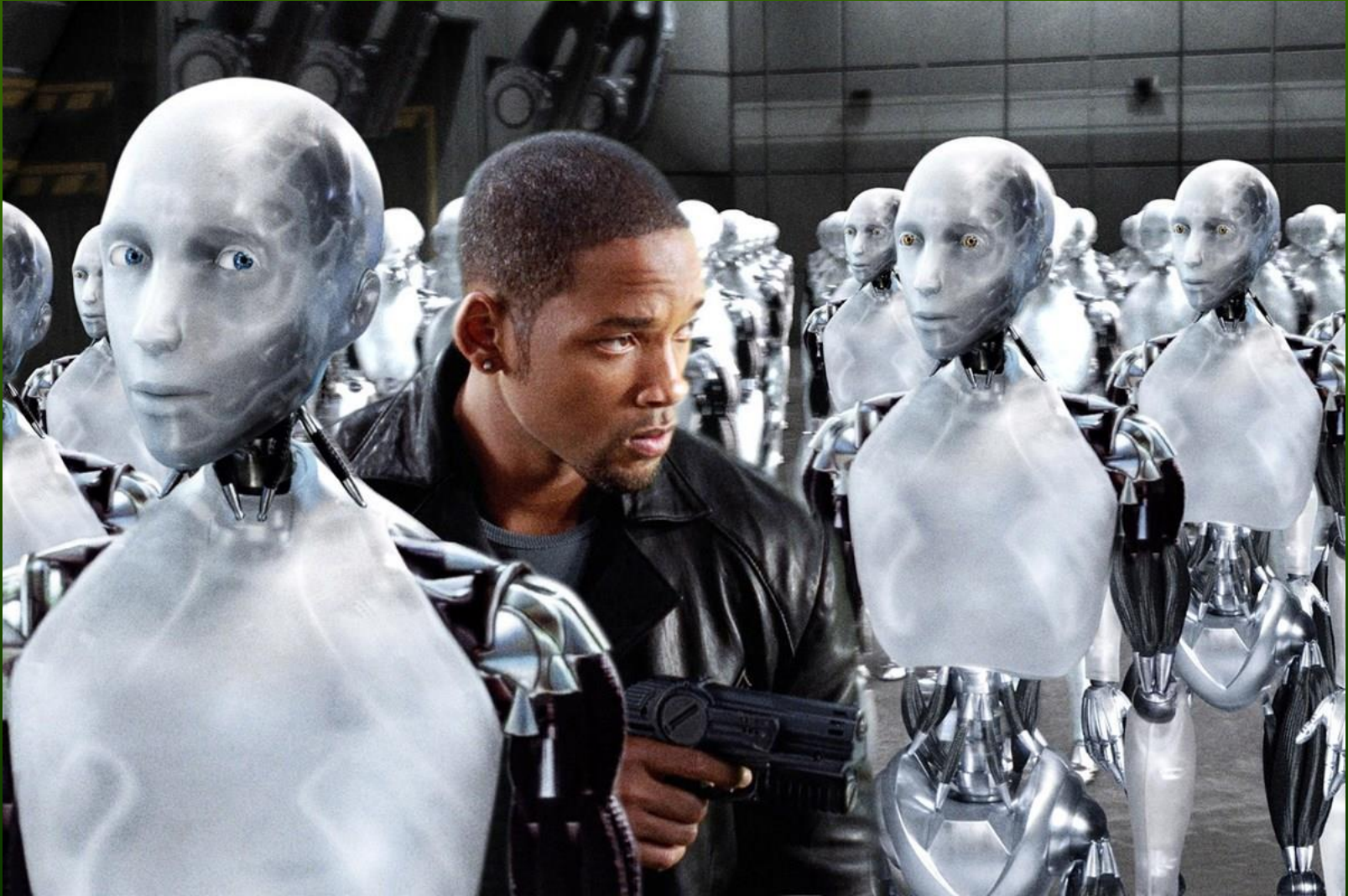
The Center for AI Safety (CAIS) próbuje to badać.

1. Broń - dron sterowany przez sztuczną inteligencję "zabija" ludzkiego operatora w symulowanym teście USAF (1.06.2023, fałszywe?).
2. Dezinformacja – manipulacja, głos, obraz, wideo, fałszywe wiadomości.
3. Niebezpieczne zadania – pośrednie cele mogą prowadzić do katastrof.
4. Osłabienie - całkowite uzależnienie od maszyn.
5. Blokada zmian – koniec demokracji, władza dla małych grup
6. Emergentne umiejętności – nie wiemy, jakie zdolności mogą się pojawić
7. Manipulacja – rozumienie ludzkiej psychiki pozwala na manipulację.
8. Dążenie do władzy - szukanie przewagi ekonomicznej przez rządy/firmy.



WILL  
CREATE  
MEMES  
FOR  
FOOD

# Neuromorficzne komputery/roboty



I, Robot (2004)

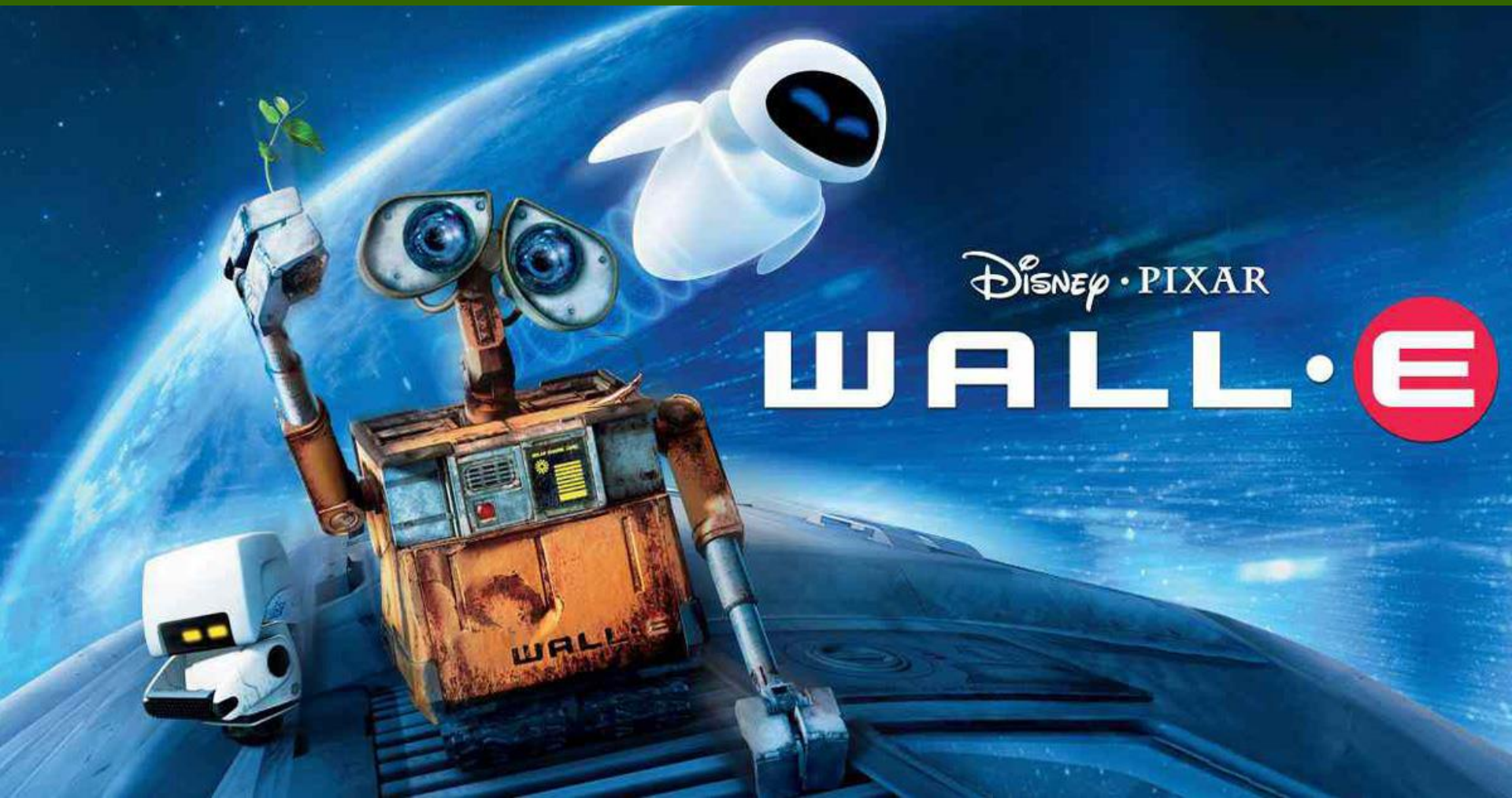
# Neuromorficzne komputery/roboty



AI

AI (2001)

# Neuromorficzne komputery/roboty



Wall-e (2008)

# (Bio)roboty



Bladerunner (1982), Surrogates (2009)



# Neuromorficzne komputery/roboty



Terminator (2008)



# AI rządzi

**NETFLIX**

**T A U**



Tau (2018)

# AI rządzi



Singularity (2017)

# AI - mózgi



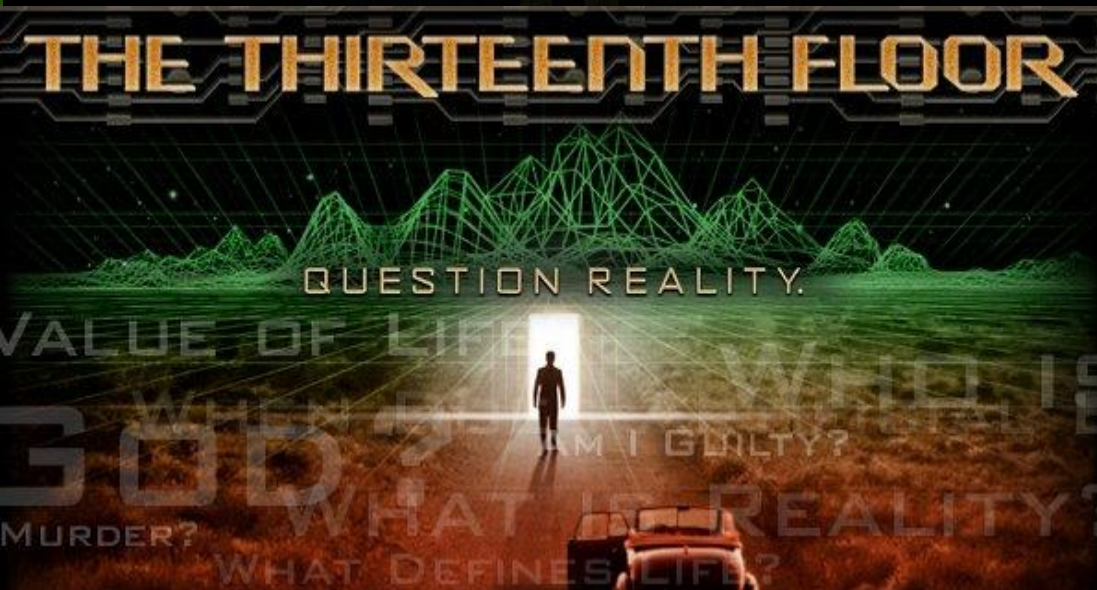
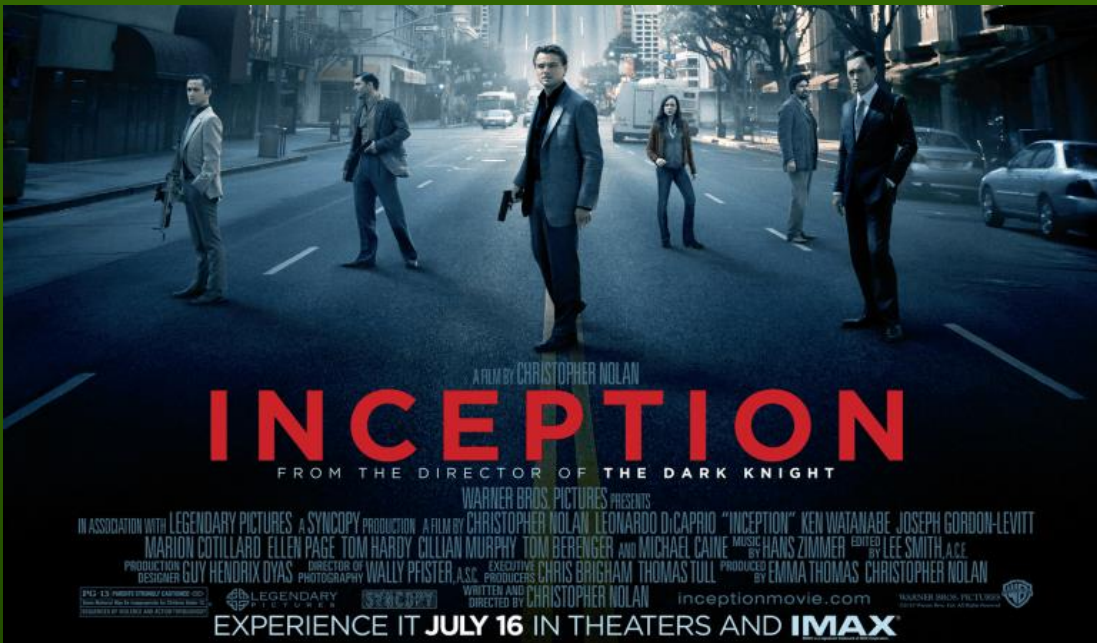
Transcendence (2014)

# AI - m6zgi



Tron (1982, 2010)

# Symulacje





Matrix (1999)



# Cele transhumanizmu

Dobry Bóg już zrobił co mógł,  
teraz trzeba zawołać fachowca ...

Czas wstać z kolan i wziąć sprawy w swoje ręce.

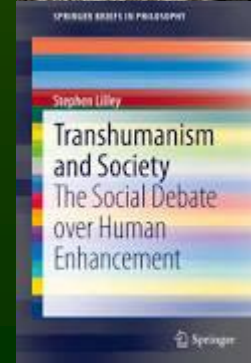
- Zwiększyć długość i jakość życia.
- Zwiększyć możliwości intelektualne i fizyczne człowieka.
- Kontrolować swoje stany mentalne i afektywne.

Doskonalenie mózgów to wielkie wyzwanie dla nauki! **Ostrożnie!**

Wyzwania: zapobieganie zaburzeniom rozwojowym,  
osiągnięcie optymalnego poziomu rozwoju.

Humanity+, Inc, do 2008 World Transhumanist Association

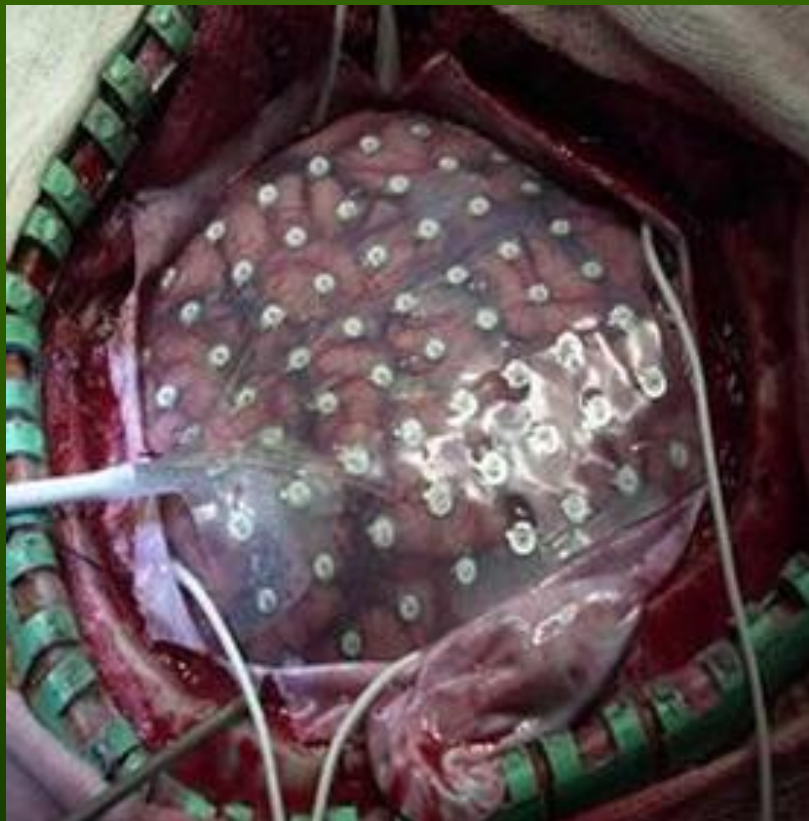
Czy to możliwe?







# Interfejsy mózg-komputer

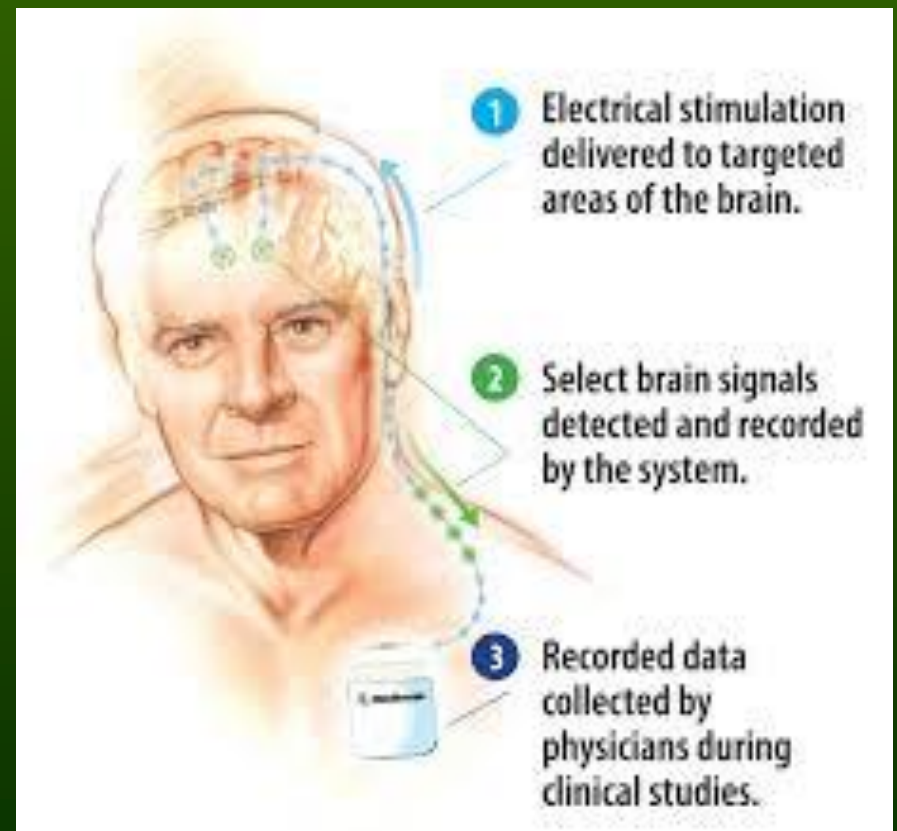


Osoby cierpiące na chorobę Parkinsona lub zaburzenia kompulsywno-obsesyjne, które mają wszczepione stymulatory w mózgu, mogą regulować swoje zachowanie za pomocą zewnętrznego kontrolera.

# Głęboka stymulacja mózgu

Osoby cierpiące na chorobę Parkinsona lub zaburzenia kompulsywno-obsesyjne, które mają wszczepione stymulatory w mózgu, mogą regulować swoje zachowanie za pomocą zewnętrznego kontrolera.

Podkręćmy sobie mózg ... czy będzie można siebie zaprogramować?

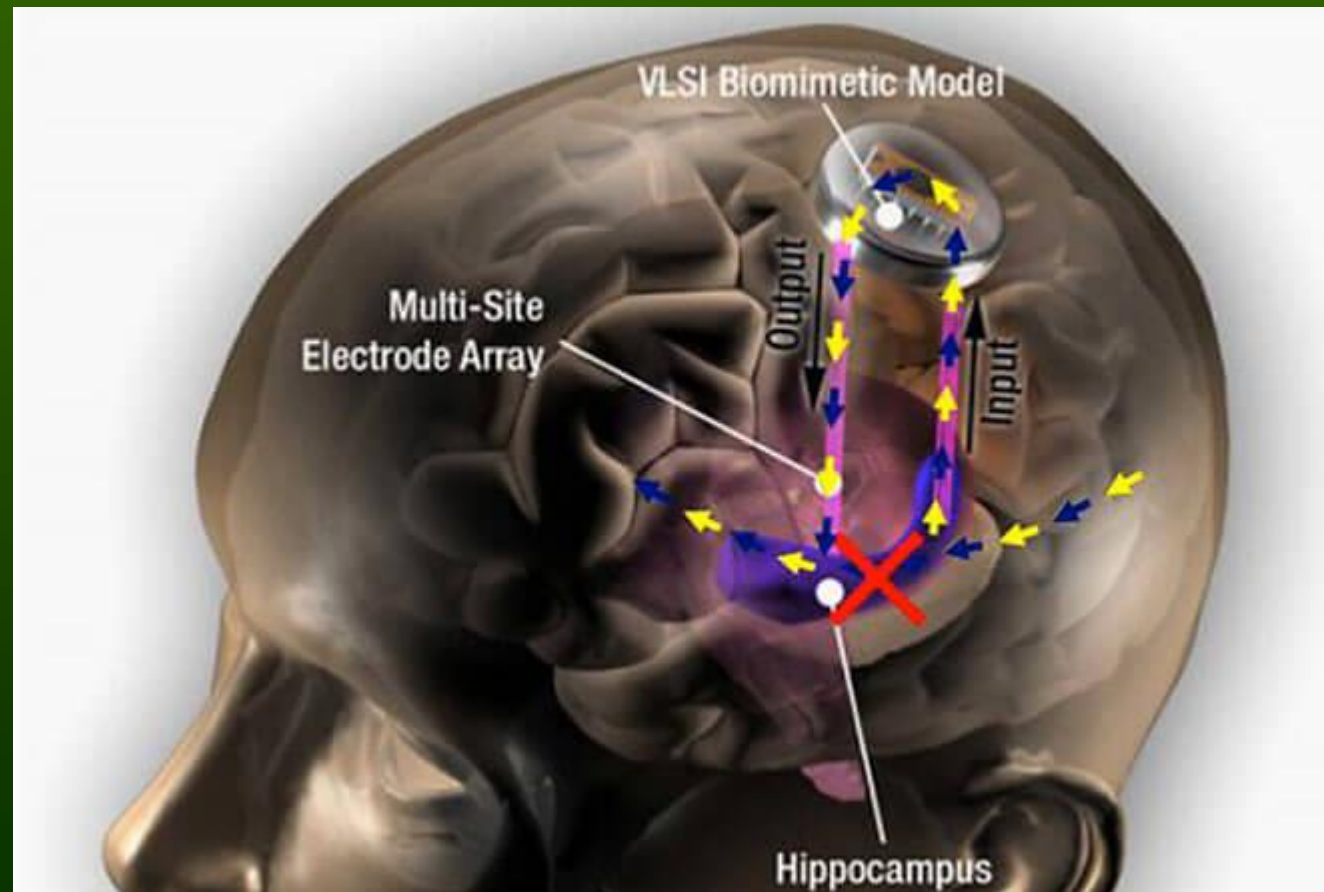


# Implanty pamięci

Testy na szczurach, małpach, a w 2017 roku na 20 ludziach dały poprawę pamięci o 30% (na szczurach o 35%). Ted Berger (USC, [Kernel](#)) : Są dobre przesłanki by wierzyć, że integracja pamięci z elektroniką jest możliwa.

DARPA: program Restoring Active Memory (RAM), dla osób z uszkodzonym mózgiem (TBI), ma być nieinwazyjny.

Neurofeedback + neurostimulacja w zamkniętej pętli.



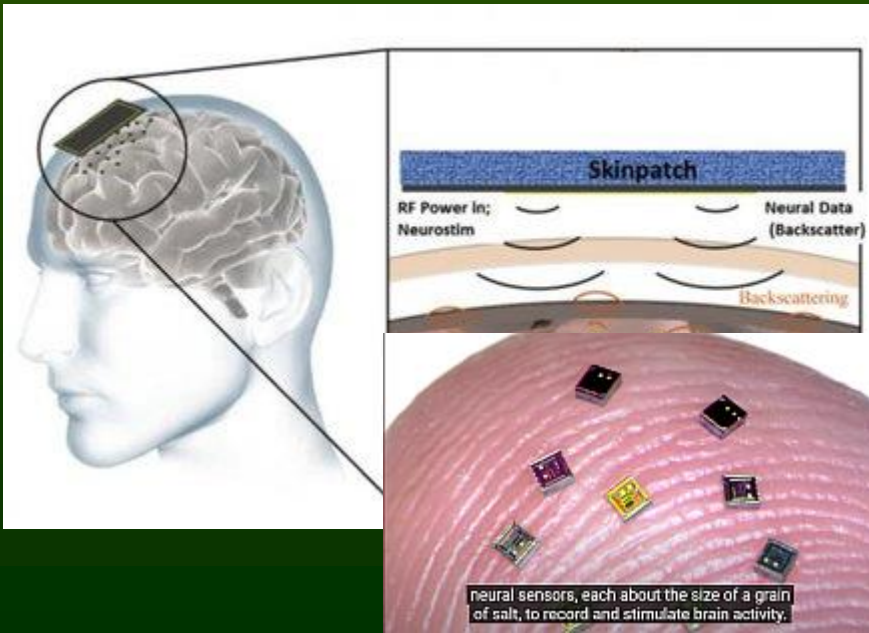
# Milion nanodrutów/ziaren w mózgu?

Inicjatywy DARPA: **Neural Engineering System Design (NESD)** i inne projekty.

Interfejs odczytujący impulsy  $10^6$  neuronów, pobudzający  $10^5$  neuronów.

DARPA przyznała granty grupom badawczym na projekty w ramach programu Electrical Prescriptions (ElectRx), którego celem jest rozwój systemów BCBI modulujących aktywność nerwów peryferyjnych w celach terapeutycznych.

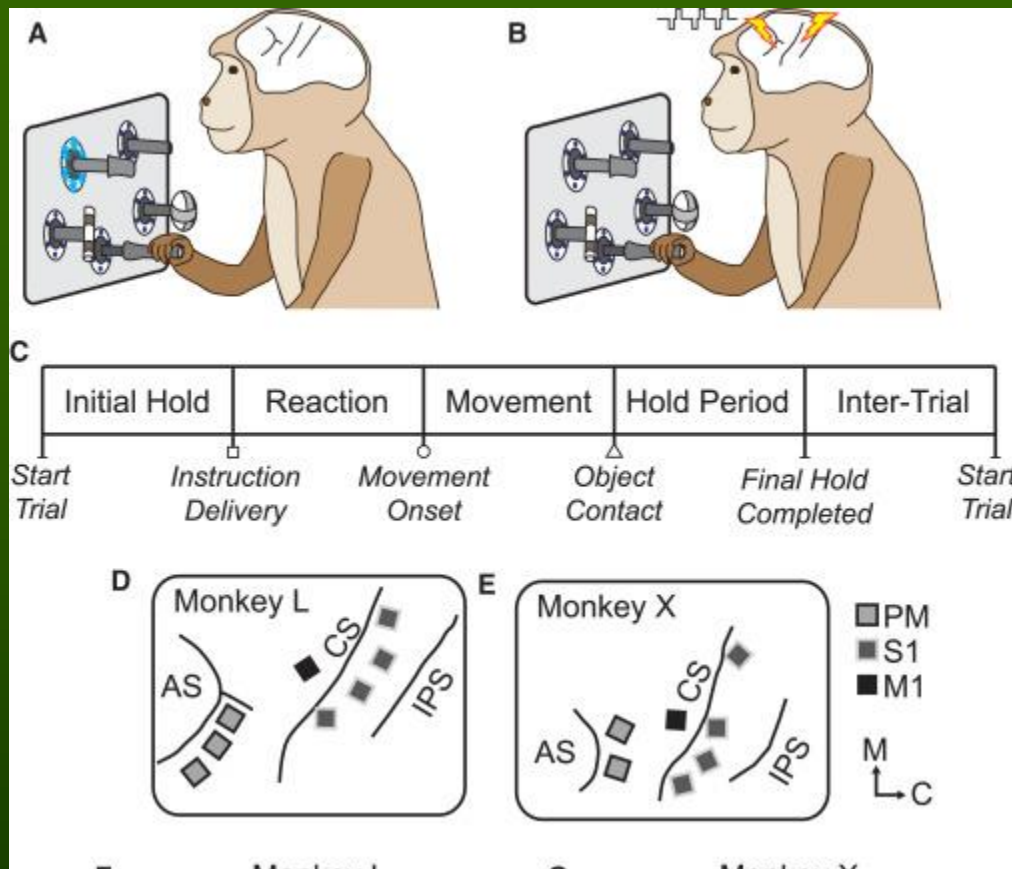
Neural dust, neurograins – mikroskopijne bezprzewodowe sensory w mózgu. Elon Musk i szumnie zapowiadana technologia neuralink (neural lace).



neural  
lace  
**ultra-thin  
mesh**

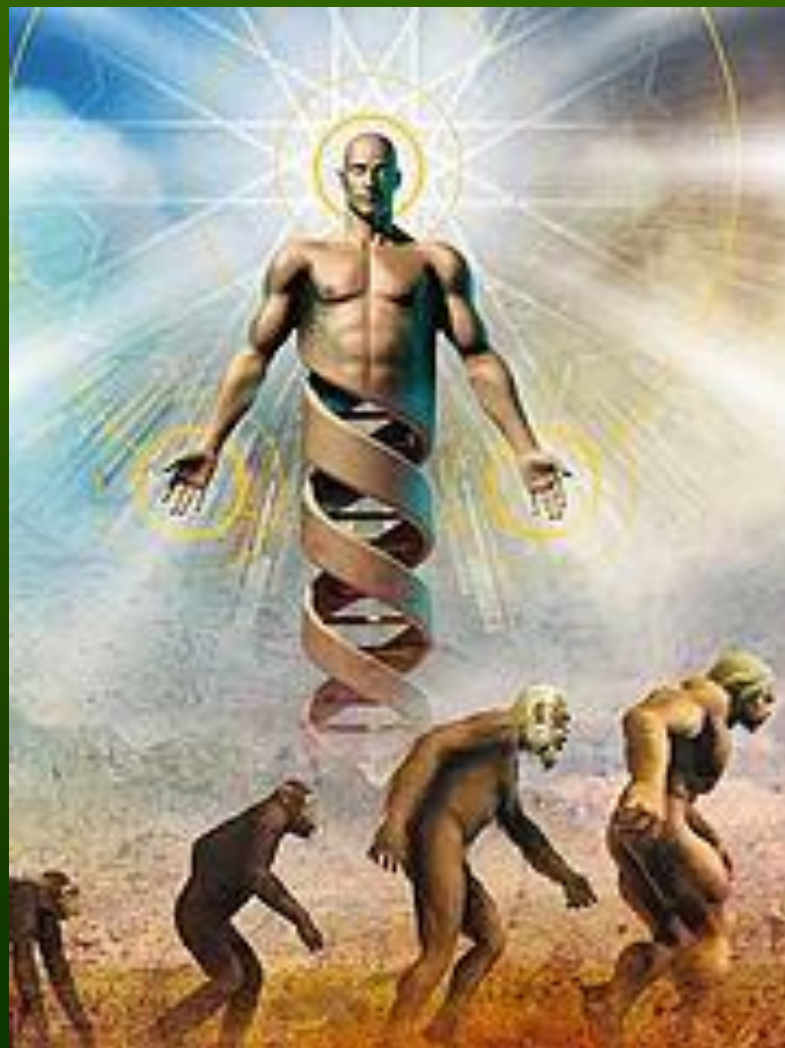


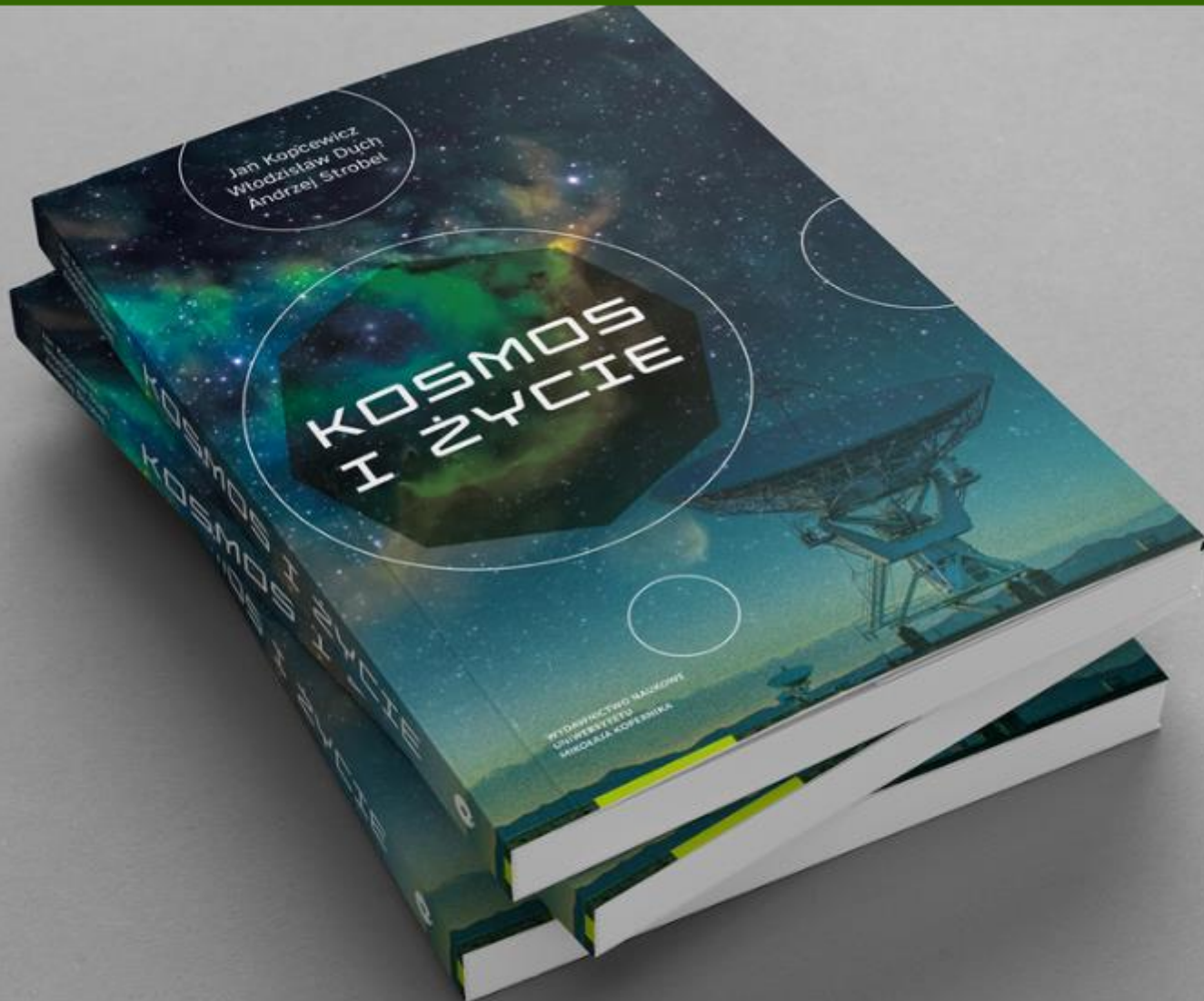
# Mikrostymulacje



Skojarzenia różnych ruchów i miejsca stymulacji w korze PM można się nauczyć. Instrukcje działania można też „wstrzykiwać” prosto do kory przedruchowej za pomocą impulsów elektrycznych tak słabych, że nie są odczuwane. AGI/ASI nadejdzie szybciej niż nasze transhumanistyczne rozwiązania.

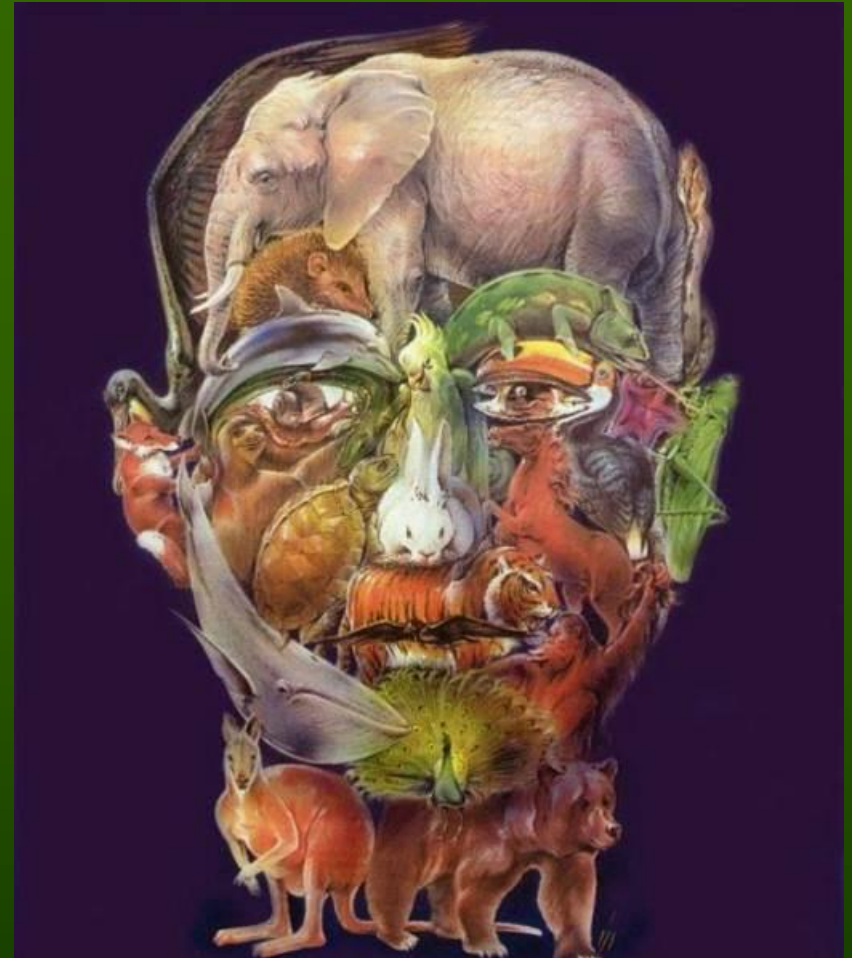
Nadchodzą ciekawe czasy ...





<https://wydawnictwo.umk.pl/pl/products/5652/kosmos-i-zycie>

# Przyszłość?



Google: Wlodek Duch  
=> talks, papers, lectures ...



