

Current state and the shocking prospects of Artificial Intelligence



Włodzisław Duch

Neurocognitive Laboratory, Center for Modern Interdisciplinary Technologies,
Department of Informatics, Institute of Engineering and Technology,
Faculty of Physics, Astronomy & Informatics, Nicolaus Copernicus University,
Toruń, Poland

Search: Wlodzislaw Duch

Konwersatorium im. Jerzego Pniewskiego i Leopolda Infelda,
Warsaw University 5.06.2023

AI at the warp speed



1. Recently: superhuman AI results in many applications.
2. Yesterday: transformers, foundational models, language/vision.
3. Today: tools for Artificial General Intelligence.
4. Emergence and sparks of AGI in GPT-4.
5. Towards distributed artificial brains.
6. Human brains and AI minds.

AI news in my [Flipboard](#).

Superhuman AI?

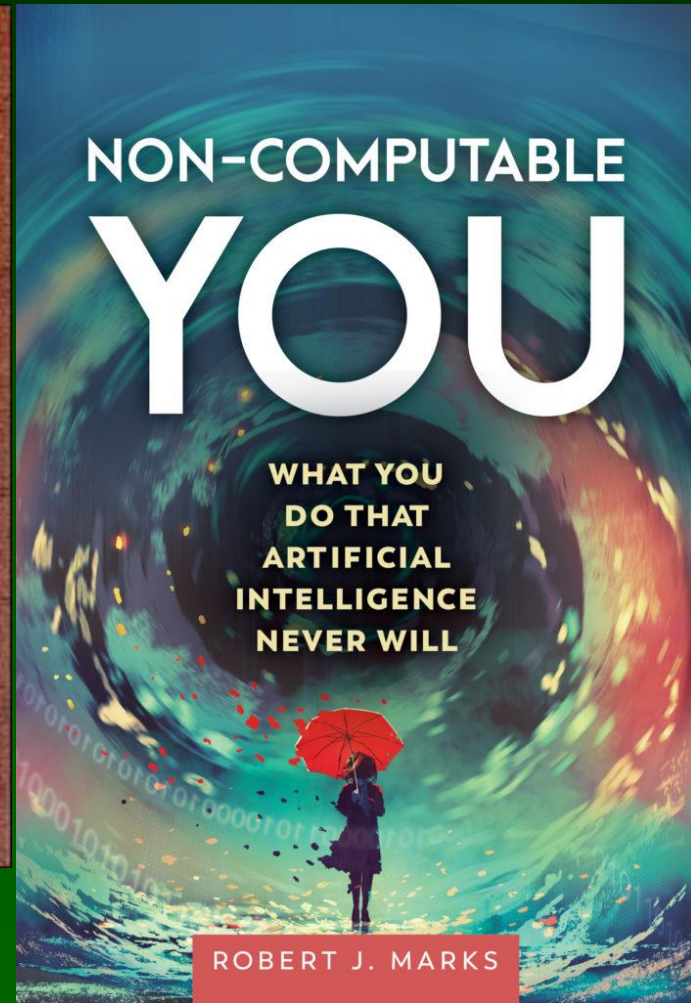
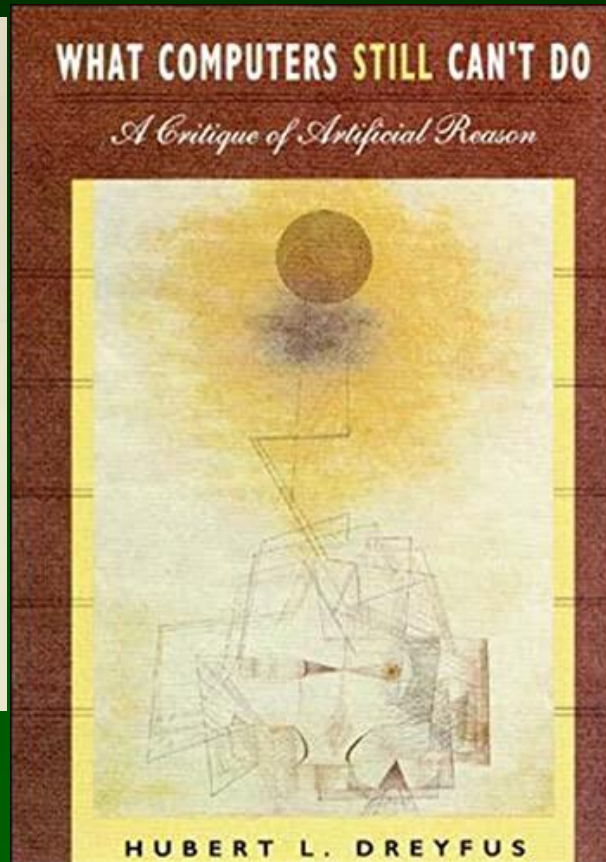
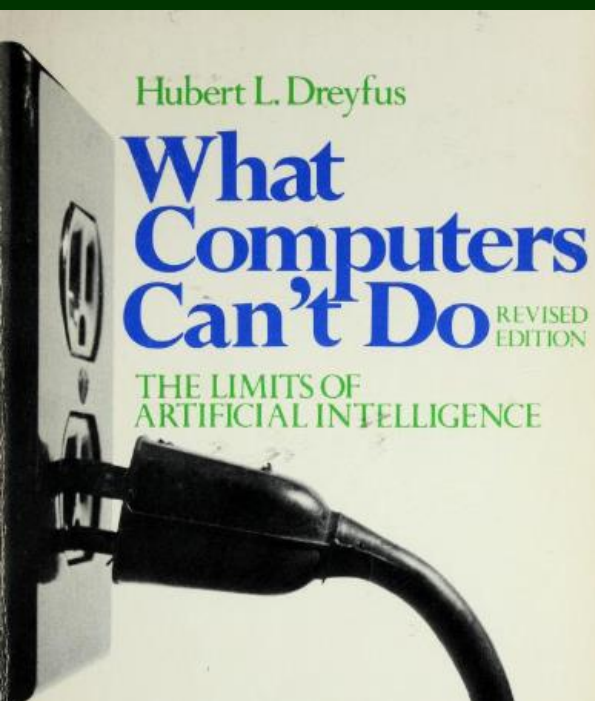
New AI: predicts words, but shows no understanding, requires small adjustments, or: new competitive form of intelligence, will lead to **radical changes**.



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



Human megalomania



1972, 1992, and 2022, predicting all kinds of problems for AI.
Human megalomania? This is what we want to hear, but reality is different.

Are we close to the Singularity?

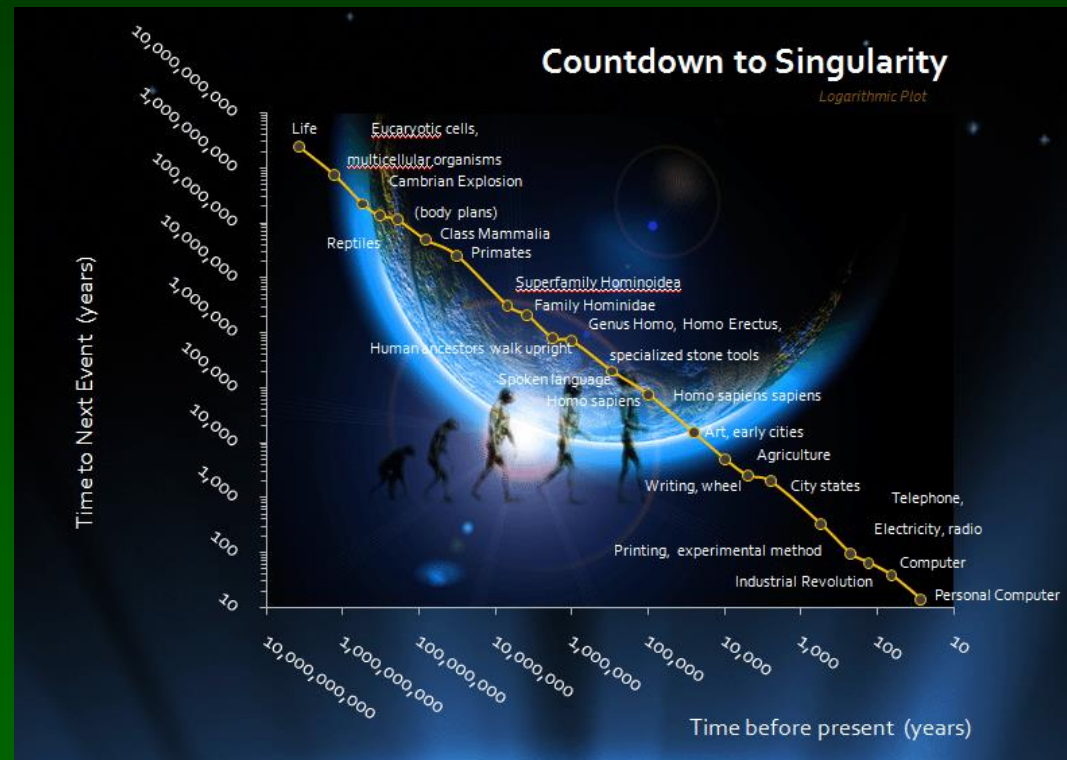
What is coming:

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.

Superhuman AI



Reasoning: 1997–Deep Blue wins in chess; 2016 –AlphaGo wins in Go; 2017 Alpha GoZero 100:0.

Open Games: 2017–Poker, Dota 2; 2019-Starcraft II, 2022 Stratego, Diplomacy – what is left?

Perception: speech, vision, recognition of faces, personality traits, political and other preferences ...

Robotics: 2020 Atlas robot (Boston Dynamics) backflip and parkour, autonomous vehicles, airplanes.

Automation of science: 2015-AI flatworm regeneration, 2020 AlphaFold 2, now 620 M protein 3D structures.

Creativity and imagination: GAN revolution, Dall-E2, Midjourney, Stable Diffusion, AIVA, music composers.

Language: 2011–IBM Watson wins in Jeopardy; 2018–Watson Debater wins with professionals. 2020: BERT answers questions from SQuAD database.

Cyborgization: BCI, brain-computer symbiosis, soon?
What are humans better in? And for how long?

AI: computer science definition

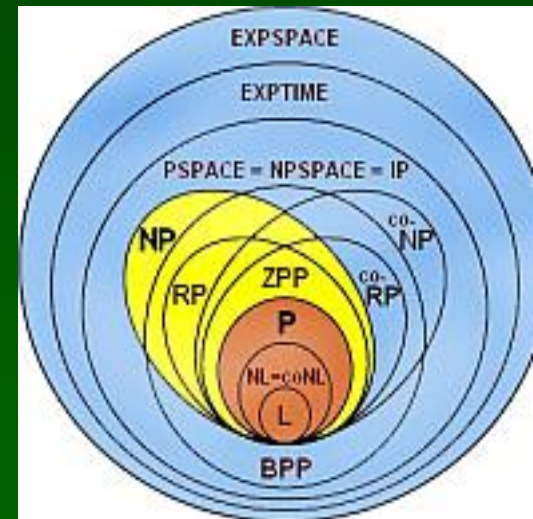
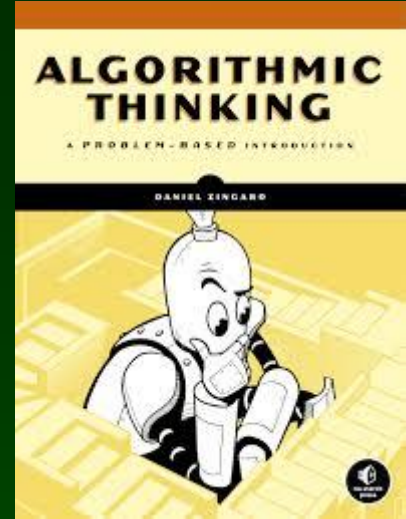
What all these AI applications have in common?

Artificial Intelligence is a branch of computer science solving problems for which there are **no effective algorithms**.

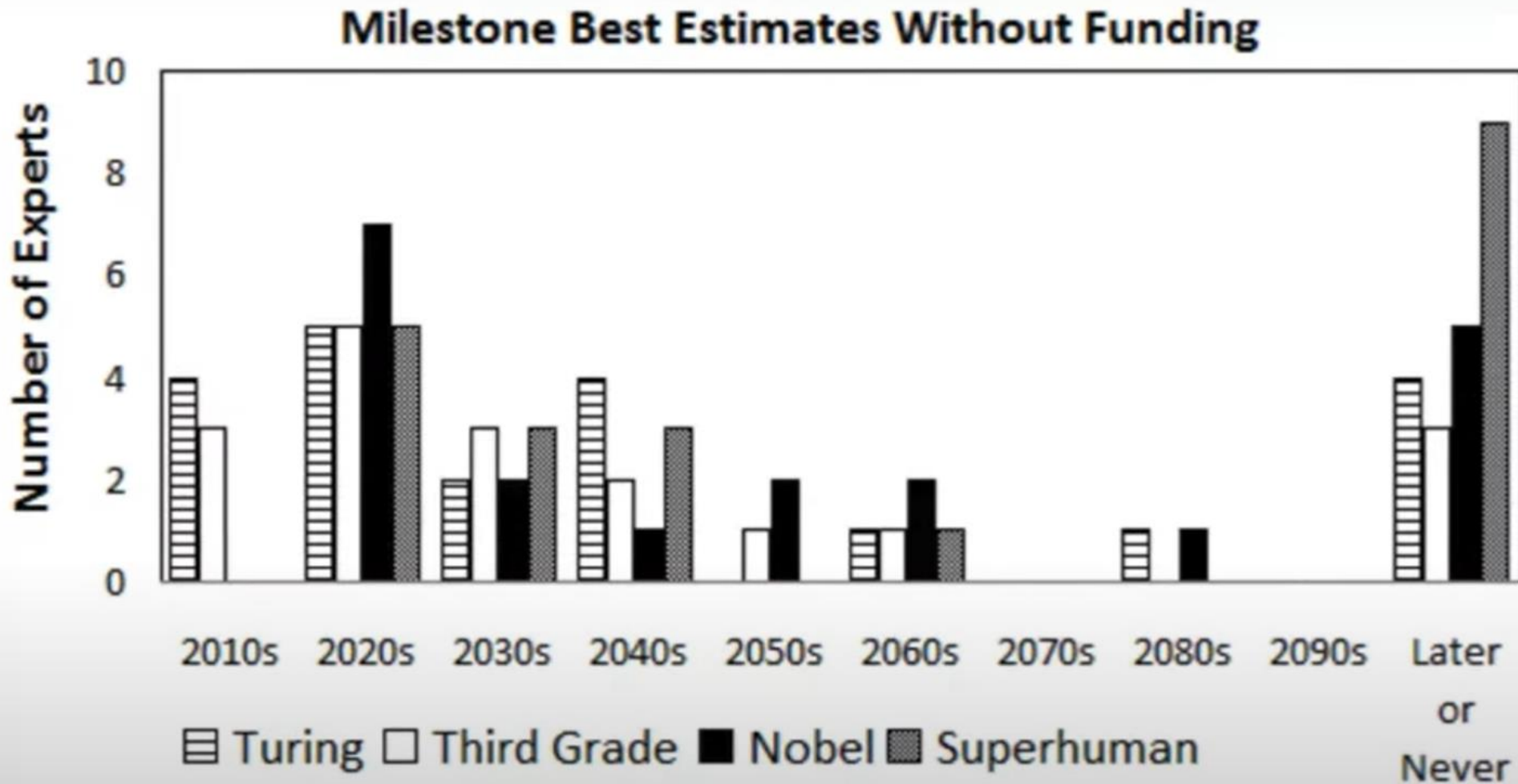
Either algorithms are so complex that they will never finish, or we do not know how to formulate the problem, to understand images, literature, or brain states.

Classes of computational complexity problems.

Unlike any older technology, AI can:
create new ideas, manipulate us,
learn very quickly, and take decisions by itself.



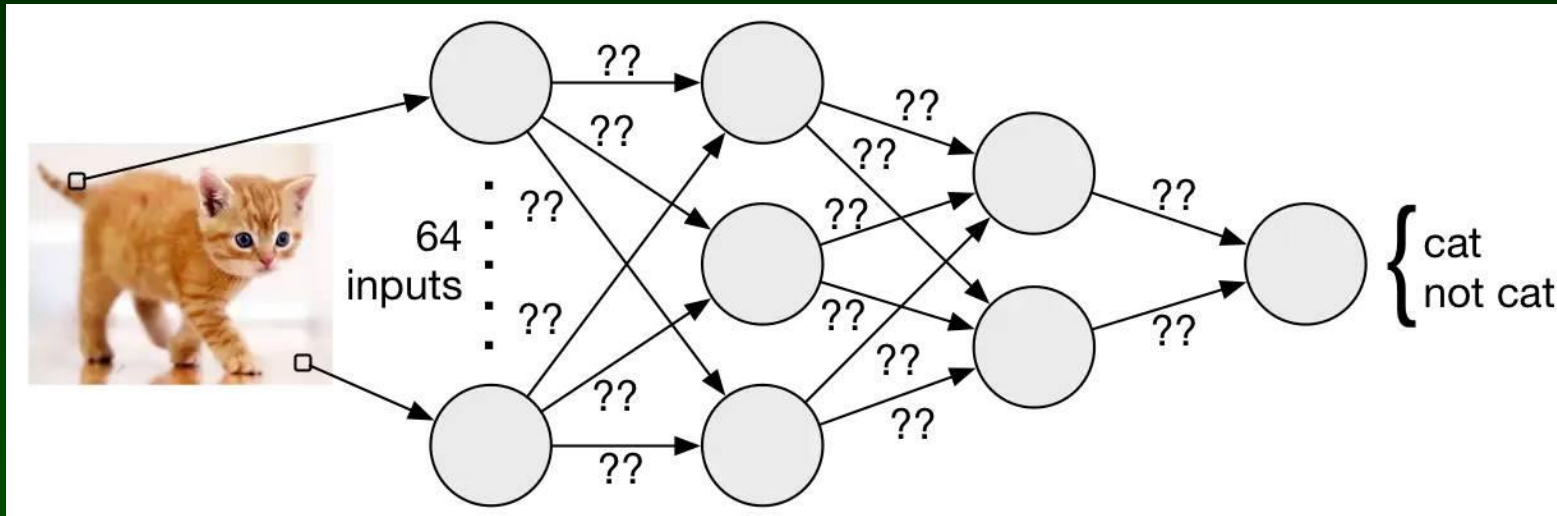
Artificial General Intelligence



In 2009 many experts thought that superhuman level will never be reached.
[AGI-09 Survey – AI Impacts](#) But what should AGI do has not been defined.

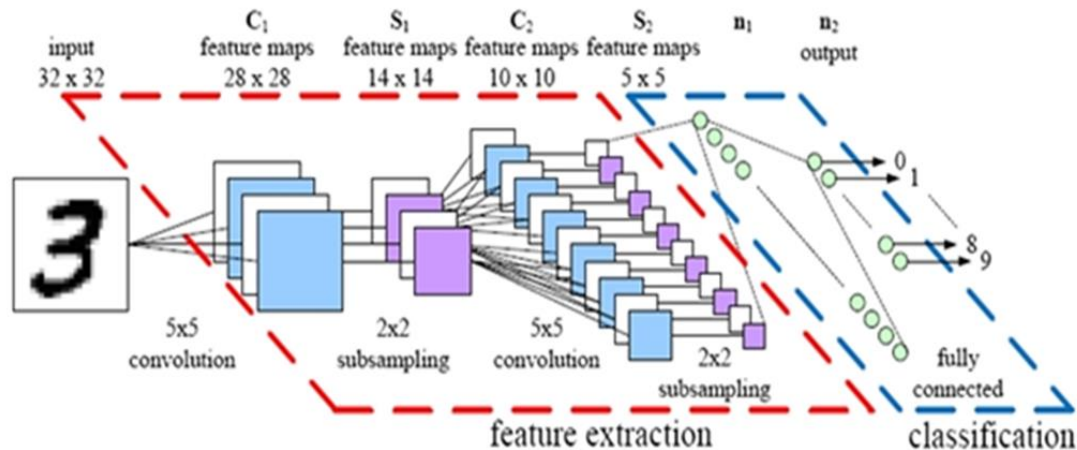
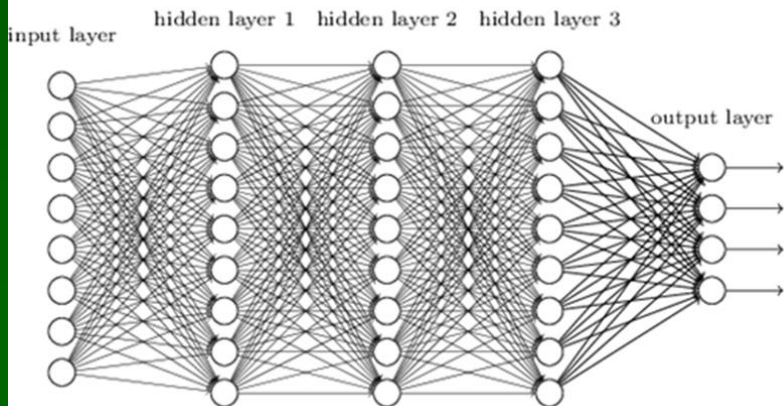
Yesterday.
Deep neural networks
and transformers

Neural classifiers



Words, image patches => networks with adjustable parameters
=> training to recognize patterns => object classification, diagnosis.

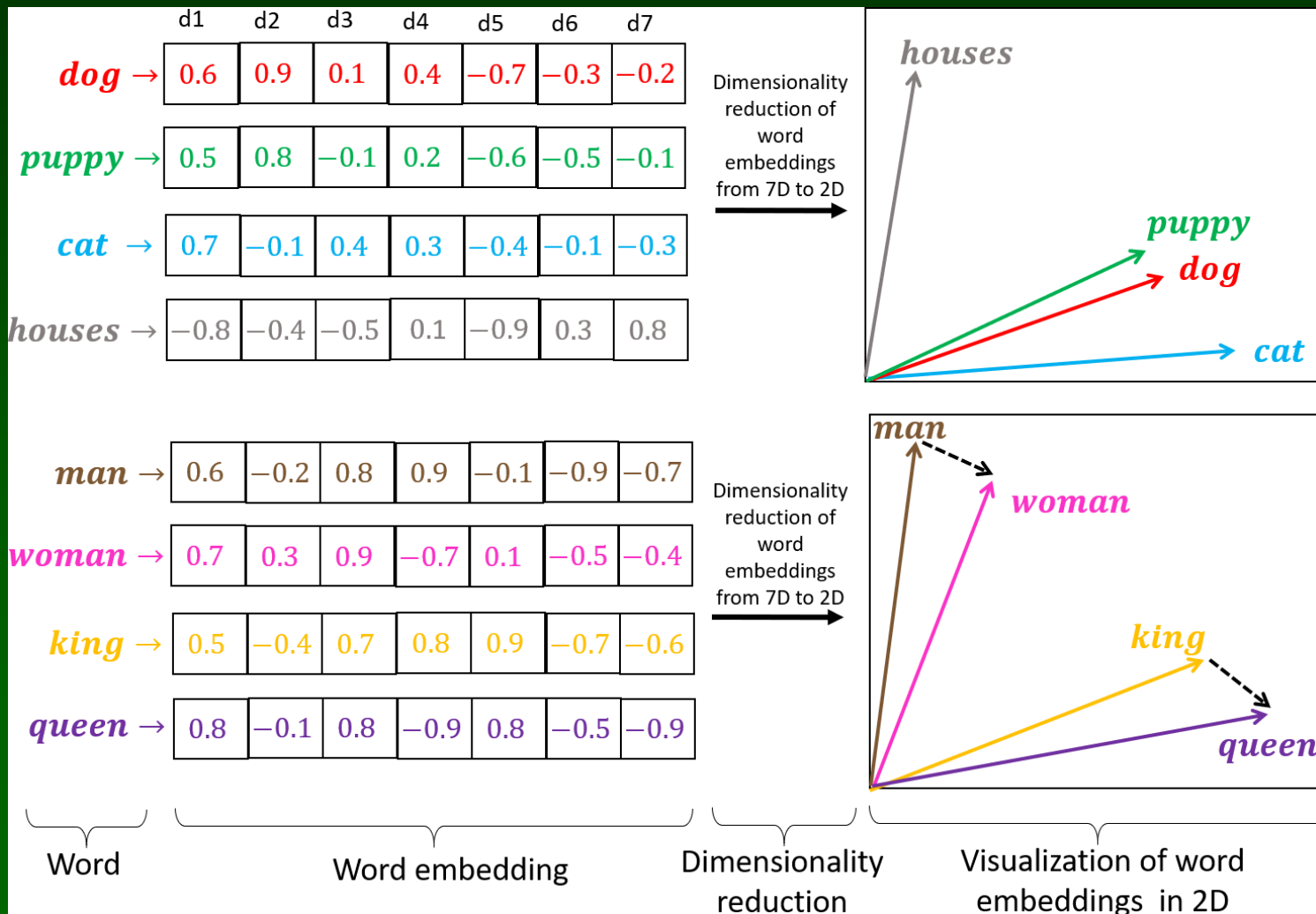
Deep neural network



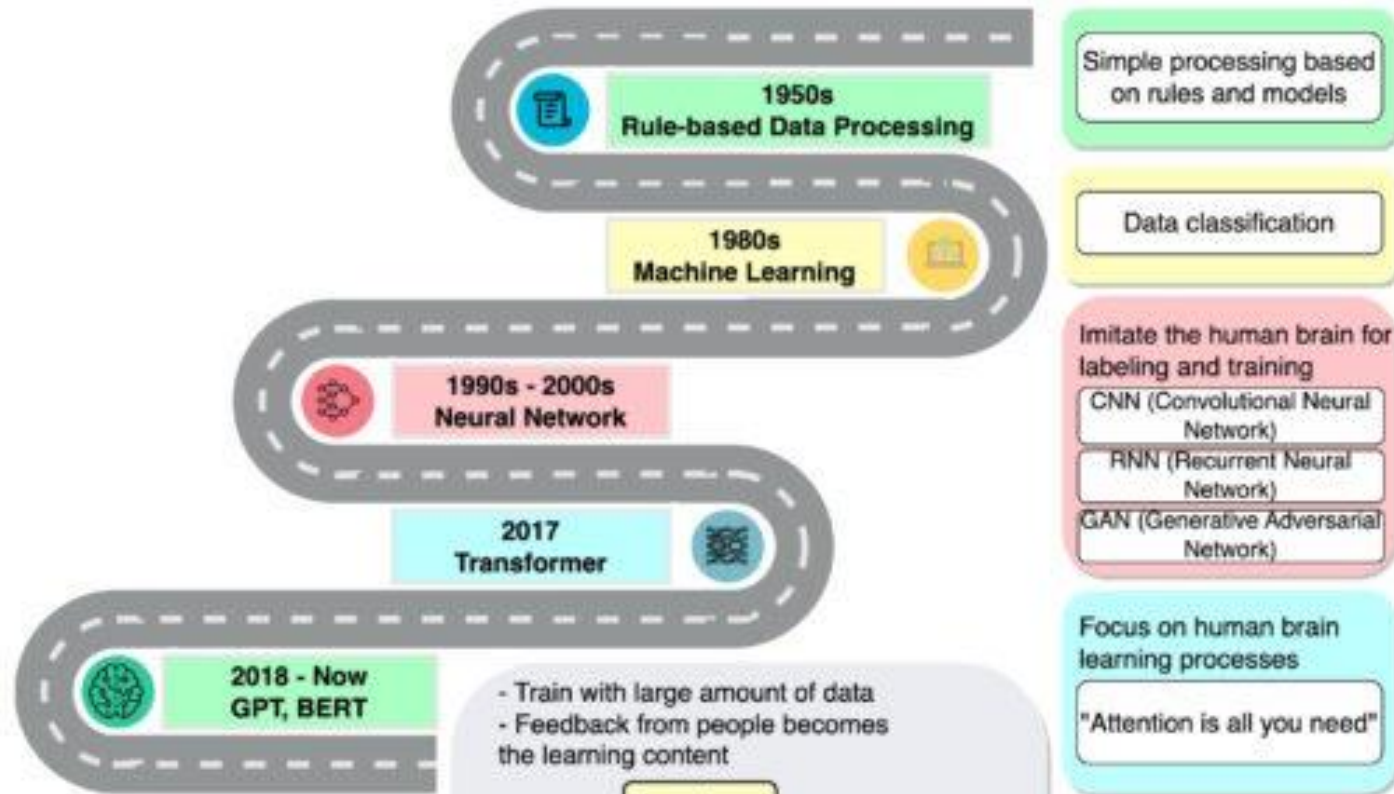
Embeddings

Words => vectors, reflecting their similarity and positions in sentences.

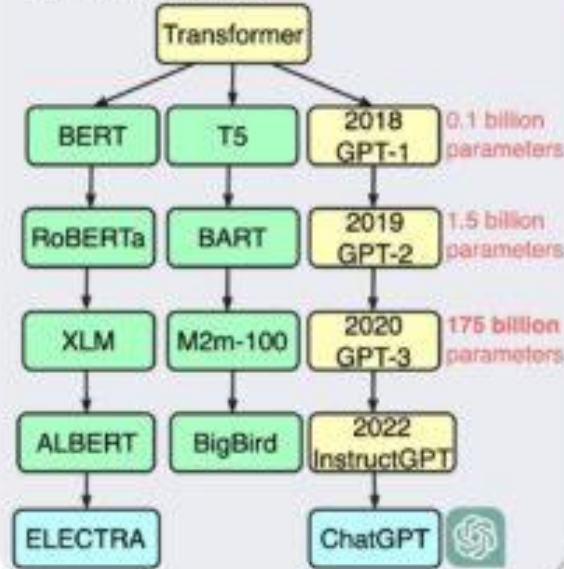
The Illustrated Word2vec.



LLM timeline. From rules to networks to transformers



- Train with large amount of data
- Feedback from people becomes the learning content

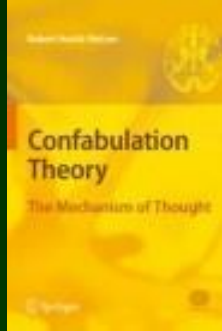


+ trillion 10^{12} parameters!

Human brain has 10^{14} or 100 trillions synapses.



Language algorithms



Main idea: R. Hecht-Nielsen, Confabulation Theory (2007).

Language models capture relation of words in large network structures.

In 2018, Google created BERT, model pre-trained on a very large text corpus, to gain a general-purpose “language understanding”,

- **Bidirectional Encoder Representations from Transformers (BERT).**
Transformer-based machine learning technique for (NLP) pre-training.
- The network learns to predict masked tokens (words, images, signals):
Input: the man went to the [MASK1]. He bought a [MASK2] of milk.
Labels: [MASK1] = store; [MASK2] = gallon.

BERT: 110M parameter model, followed by 340M parameters in 24 neural layers; trained on the BooksCorpus (800M words), Wikipedia 2,5B words.

BERT model was then fine-tuned for specific NLP tasks such as question answering and semantic information retrieval. Smaller pre-trained open software models were published in GitHub.

Versions in 70 new languages were created in one year.

Transformers

Attention: given a sequence of tokens (words, image patches), how relevant is each input token to other tokens?

Attention vectors capture context (embedding, semantics) + encode relative positions (syntax) of words.

Example:

Input: sentence in English;

Output: sentence in Polish.

Google BERT used this approach.

Generative Pre-trained Transformers or GPTs are now best known models.

[Simple intro on Youtube.](#)

Vaswani et al.(2017).

Attention Is All You Need. arXiv

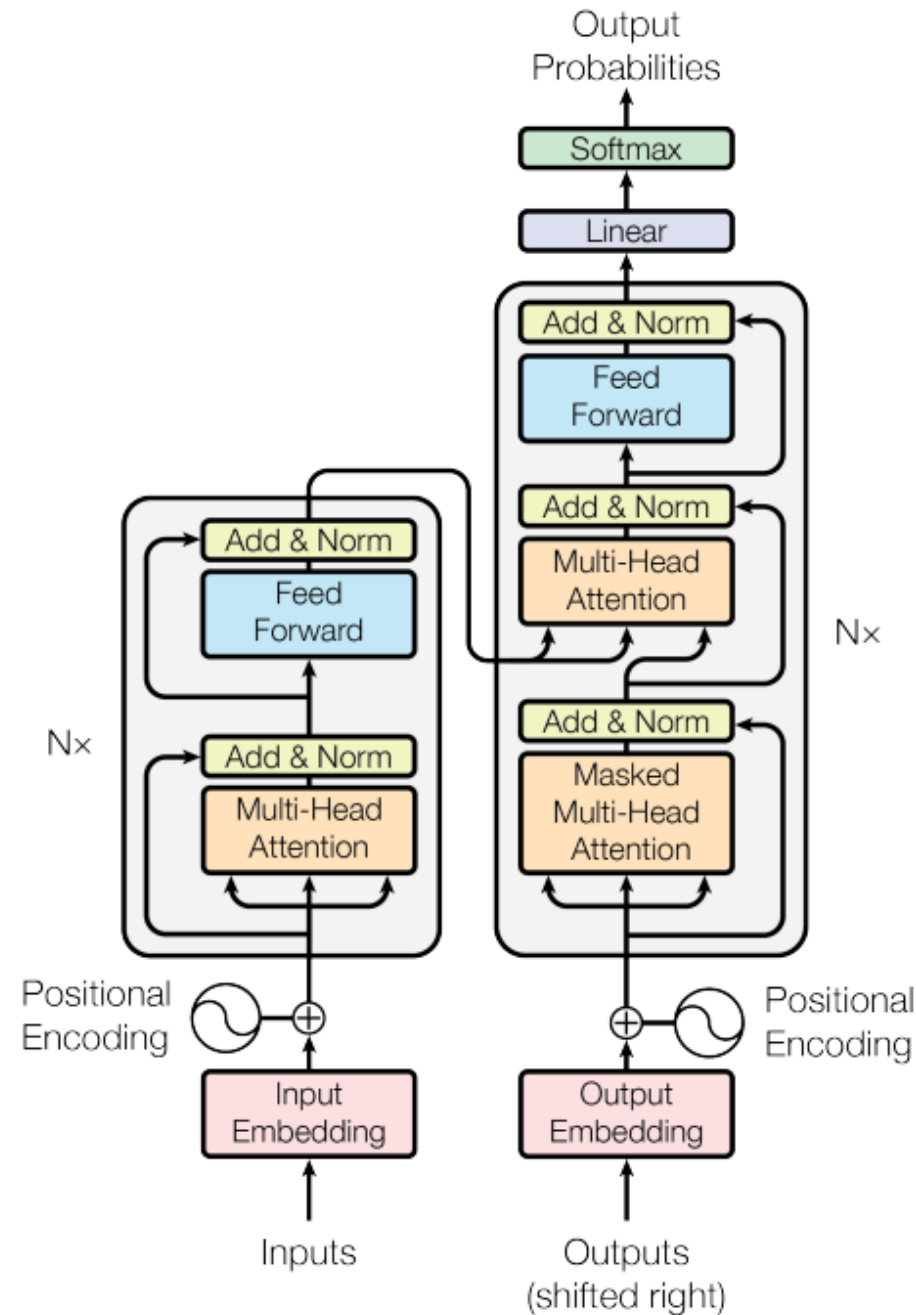
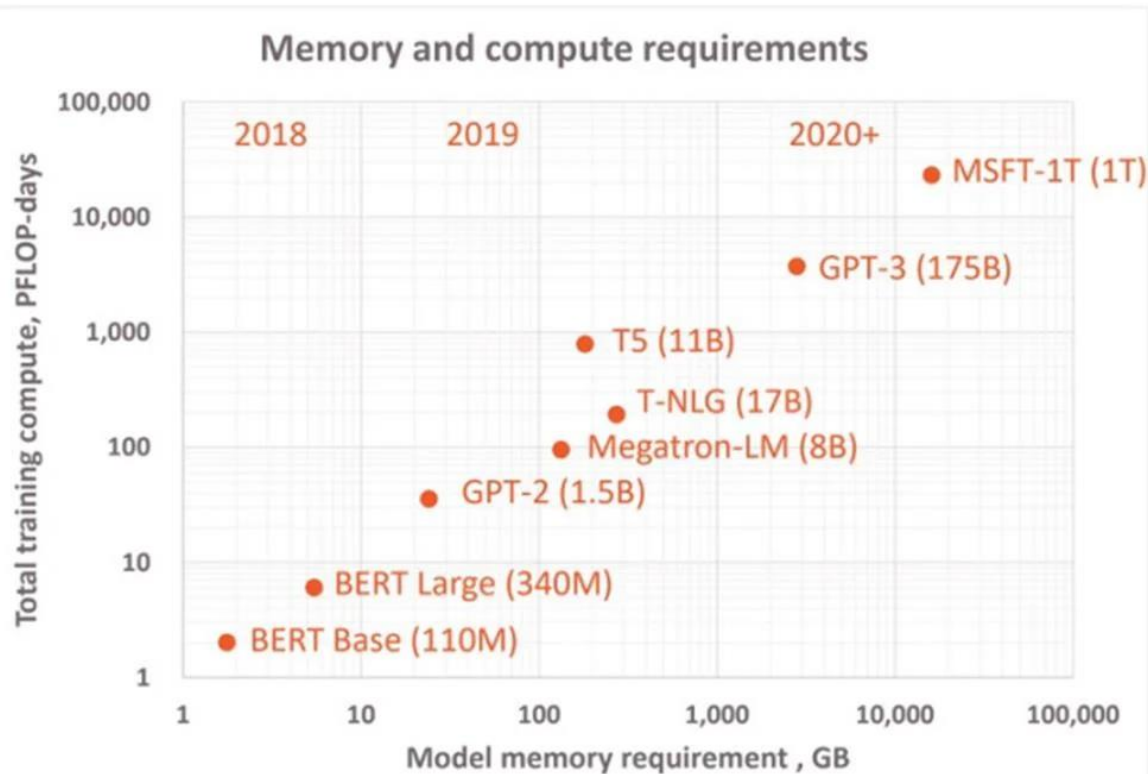


Figure 1: The Transformer - model architecture.

Acceleration ...

Exponential Growth of Neural Networks



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

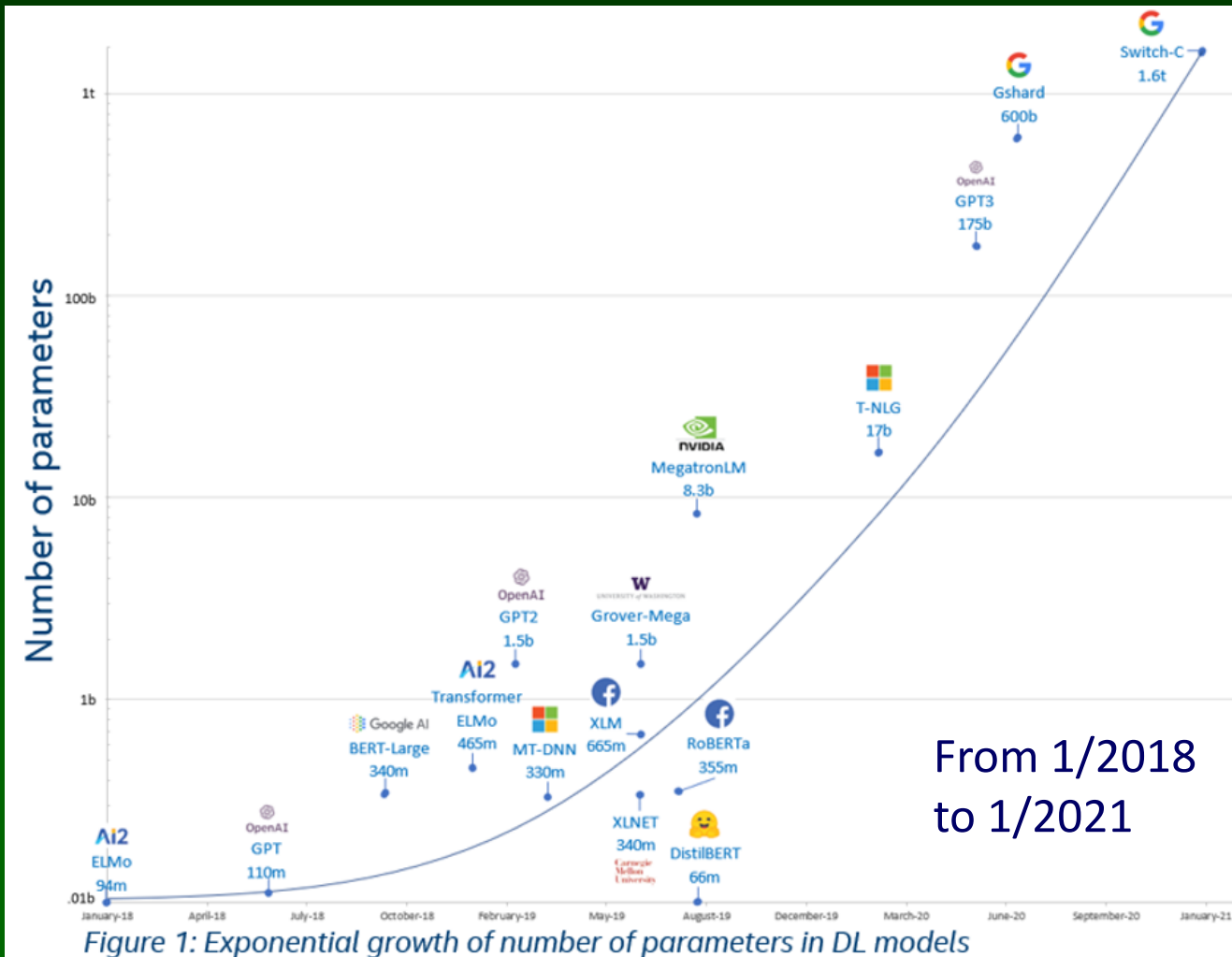
Tomorrow, **multi-Trillion** parameter models and beyond.

Nanotechnologies for ML training: in 10 years from 1 Pflop to 1 billion Petaflops!

Large models provide foundations for general knowledge. Gato has only 1.2 B parameters. Smaller LLM may also be efficient (mBERT, LLaMa, Falcon).

NLP supermodels

OpenAI GPT-3 model has 175 B parameters! One can use it on OpenAI server. First-of-its-kind API can be applied to any language task, and serves millions of production requests each day. GPT-4 has 170 trillion parameters?



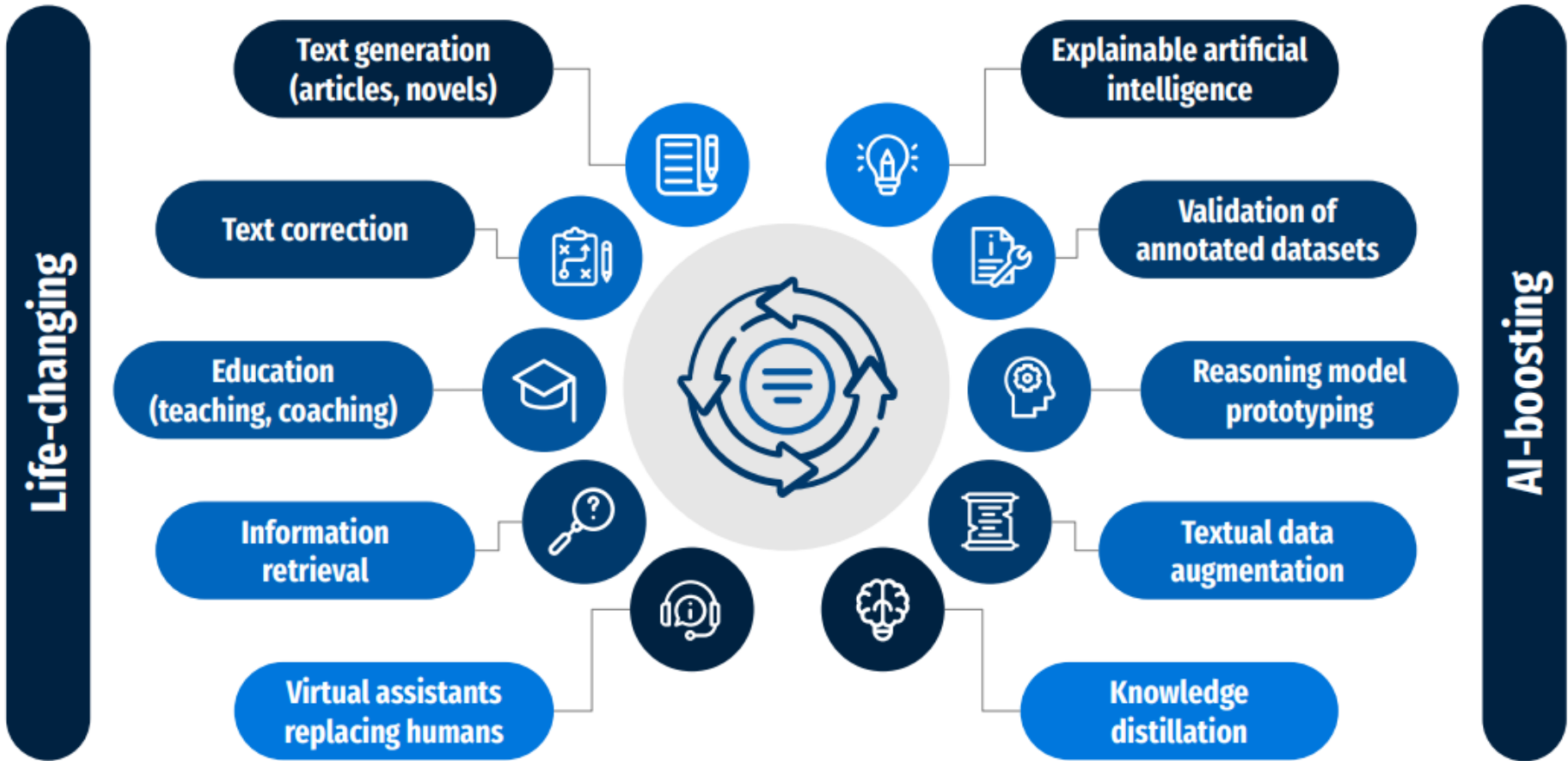
WuDao ~1.75 trillion param.

Between 2018-21 LLM complexity increased 1 M times!

100 M monthly active users in 2 month after release of ChatGPT!

From 1/2018 to 1/2021

Prospects for ChatGPT applications



Creative industries, entertainment, design, legal, financial, question-answering, data analysis, programming, speech translation to 1100 languages (Meta), text recognition for 4000 languages, and many other applications.

[JPMorgan is developing a ChatGPT-like](#) AI service that gives investment advice.

Large Language Models

Next-gen (existing) applications

Product & customer interaction / management



 viable  chatdesk  Quickchat

 Nevermaps  ActiveChat  exceed
by GENESYS

 Stateset  Sapling



Personal productivity

 personal.ai

 mem 

 Oogway

Search engine

 YOU  Google

 algolia

Emerging net-new applications

Application synthesis

 Adept  CODEGEN 

Data analyst productivity

 veezoo  AI 2sql  cogram


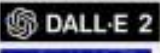


Developer productivity

 warp  tabnine

 GitHub Copilot 

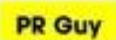

 replit  M  KD

New media generation

 FABLE  DALL-E 2  alethea  R

Writing assistant/text generation



 AI21labs  Jasper  Snazzy AI

 PR Guy  copy.ai  Scalenut

 LAVENDER  YOU Write 

 anyword  Simplified 

 copysmith  copymatic 

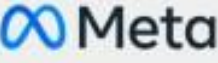

 LONGSHOT  Rytr  Writesonic
formerly ProWritingAid

Infrastructure

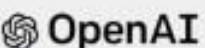

Model /builders providers - Big Tech



 Google 

 Meta 

Model providers/builders - Startups

 OpenAI 

 Hugging Face 

 AI21labs 
We bring Light to AI


 ANTHROPIC 

Accessible specialized AI chips

 NVIDIA 

 Google 
We bring Light to AI

Other tooling

 Humanloop 

[Microsoft Build AI Announcement](#), or AI comes to Windows.

ChatPDF

Talk to Books

Browse passages from books using experimental AI

[Learn more](#)

Not a traditional search

Use this demo as a creativity tool to explore ideas and discover books by getting quotes that respond to your queries.

Use natural language

Speaking to it in sentences will often get better results than keywords. That's because the AI is trained on human conversations.

Play with it

Try our sample queries then try your own. Experiment with different wording to see how it changes the results.

Say something to books... [Go!](#)

Consensus for evidence-based answers.

Galactica LLM (Meta) for science, interesting but needs more work.

GPT-3 as philosopher



Eric Schwitzgebel, David Schwitzgebel, Anna Strasser,
Creating a Large Language Model of a Philosopher, [arXiv:2302.01339](https://arxiv.org/abs/2302.01339)

“Can large language models be trained to produce philosophical texts that are difficult to distinguish from texts produced by human philosophers?”

We fine-tuned OpenAI's GPT-3 with the works of philosopher Daniel C. Dennett as additional training data, and **asked the real Dennett ten philosophical questions**, posing the same questions to the ChatGPT-3, collecting 4 responses for each question, without any cherry-picking.

425 participants tried to distinguish Dennett's answer from ChatGPT. 25 experts on Dennett's work **succeeded 51% of the time**.

Philosophy blog readers (N = 302) performed similarly to the experts. Ordinary research participants (N = 98) were near chance (20%) distinguishing GPT-3's responses from those of an "actual human philosopher".

Is Dennett intelligent? If we agree, then GPT-3 must also be intelligent.

Generating new ideas

Generate relevant research questions and ideas.

InfraNodus text analysis tool with help from GPT can show structural gaps between topics, representing text in a graph, and generating questions that link diverse topics, ex: What is the physical environment impact on our ideas?

The screenshot displays the InfraNodus text analysis tool interface. On the left is a network graph with nodes and edges. Nodes are labeled with terms like 'environment change', 'event', 'time', 'behavior', 'man', 'mind', 'deal', 'informati', 'great', 'idea', 'question', and 'ecology'. A red line points from a text box to a specific gap in the graph.

Based on the structural gap, GPT-3 generates a research question that would bridge these topics together:

The following questions were generated to bridge the gap between **change, information, environment** and **question, thought, deal**:

- What is the physical environment's impact on our thoughts and ideas?
- How does the environment change over time?

Control Panel (Right):

- Buttons: Essence, Insight, Trends, Stats, Sentiment, LDA
- Action Advice: Diversify
- Structural Gap (ask a research question that would link these two topics):
 - change information environment
 - and
 - question thought deal
- Buttons: Reveal the Gap, Generate a Question, ?
- Latent Topical Brokers:
 - deal thing process order occur
 - ?
- Button: Back to the Main Topics

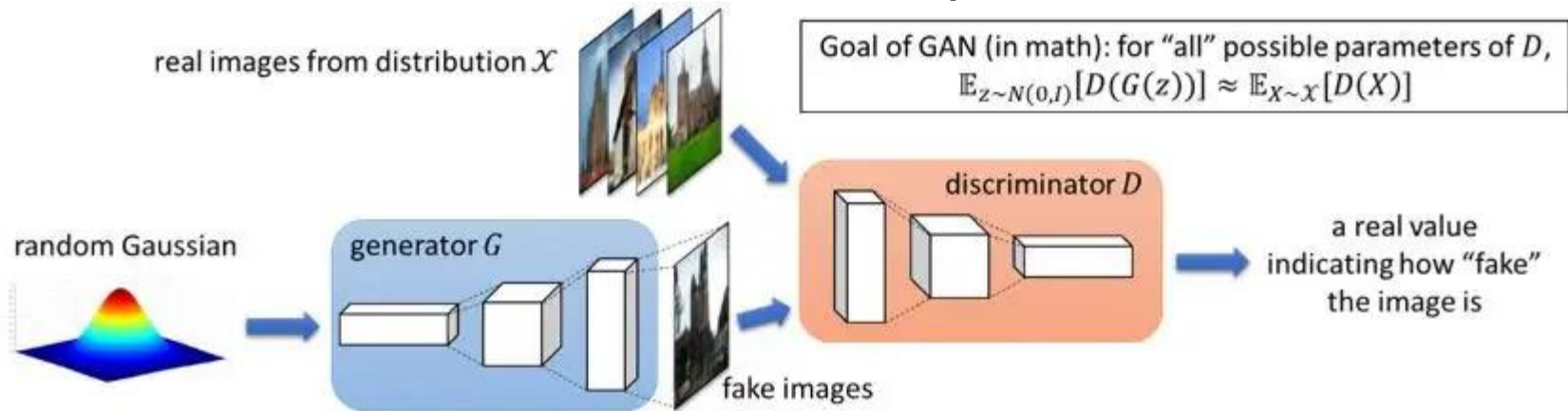
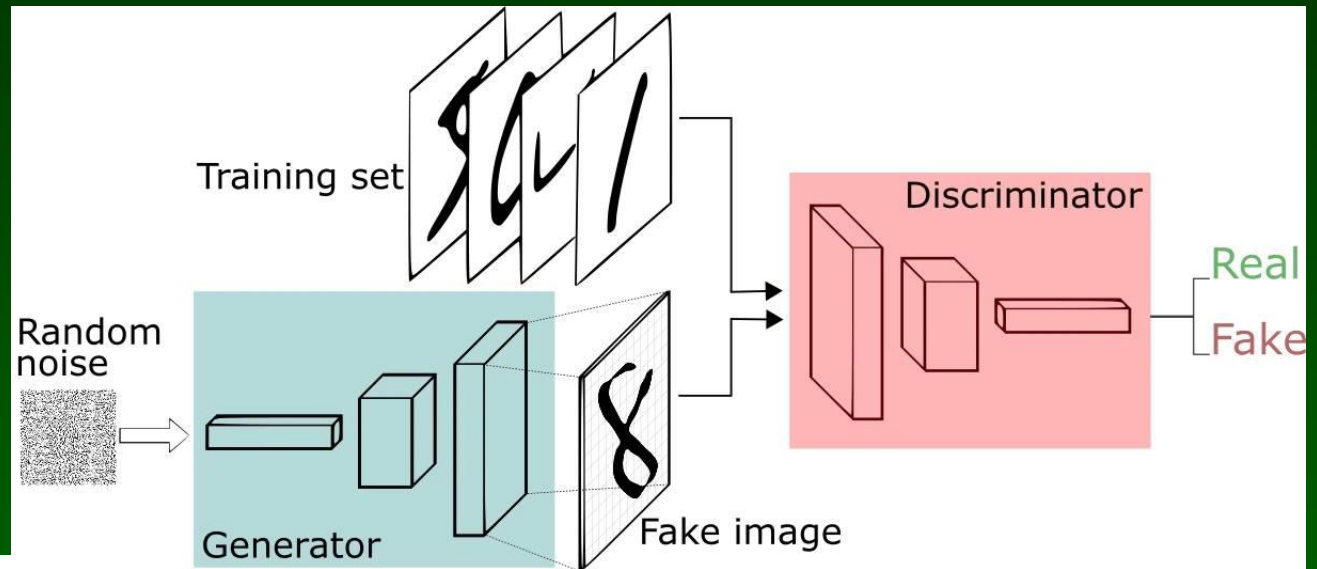
Bottom Panel:

- Buttons: back, regenerate, + ideas
- Network structure: focused
- Help Center button

Images: Generative Networks

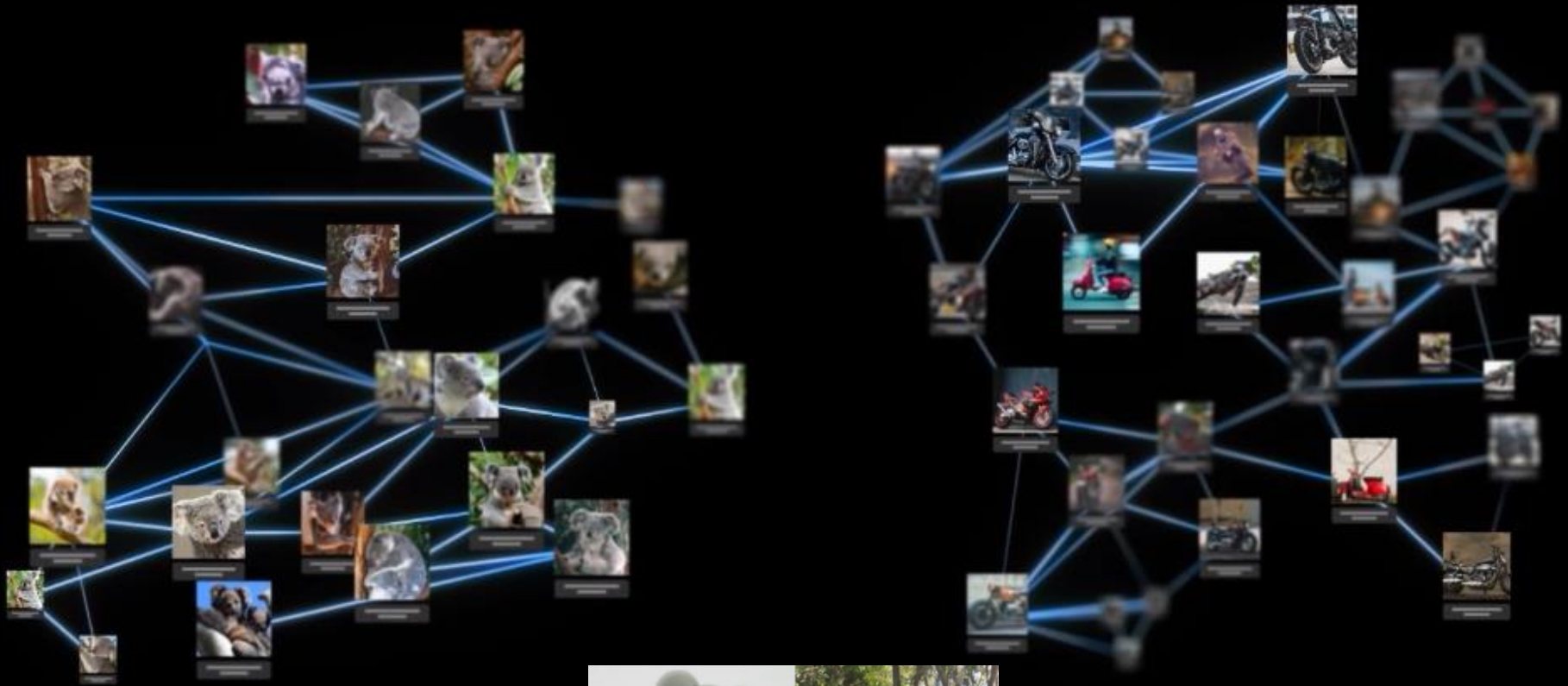
GANs (Generative Adversarial Networks) capture the essence of real data! Generate images starting from noise. Discriminator network recognizes real from fake patterns. Generator corrects latent parameters until it imitates real images.

Stable Diffusion



Vision-language models

Vision-Language Pre-Trained Models (VL-PTMs), convergence of language, vision, and multimodal pretraining => general-purpose foundation models can be easily adapted to multiple diverse tasks with zero-shot learning.



koala bears



motorcycles

Vision-language generative models

[Dall-E2](#), [Midjourney](#), [Stable Diffusion](#), [Craiyon](#), [Imagen](#), [Nightcafe](#), [Artbreeder](#), [Hotpot AI](#), [Deep Dream Generator](#), [Deep AI Text to Image](#), [Generative Engine](#) ...

Meta-art market: [PromptBase](#) sells prompts to generate specific images.

The image shows a network visualization of concepts. Nodes are represented by colored circles (green, pink, purple) and are interconnected by a dense web of lines. Labeled nodes include: environment change, event, time, behavior, man, mind, great, idea, question, ecology, deal, informati, and question.

Overlaid on the network is a text box with the following content:

Based on the structural gap, GPT-3 generates a research question that would bridge these topics together:

The following questions were generated to bridge the gap between change, information, environment and question, thought, deal:

What is the physical environment's impact on our thoughts and ideas?

How does the environment change over time?

At the bottom of this box are buttons: back, regenerate, and + ideas.

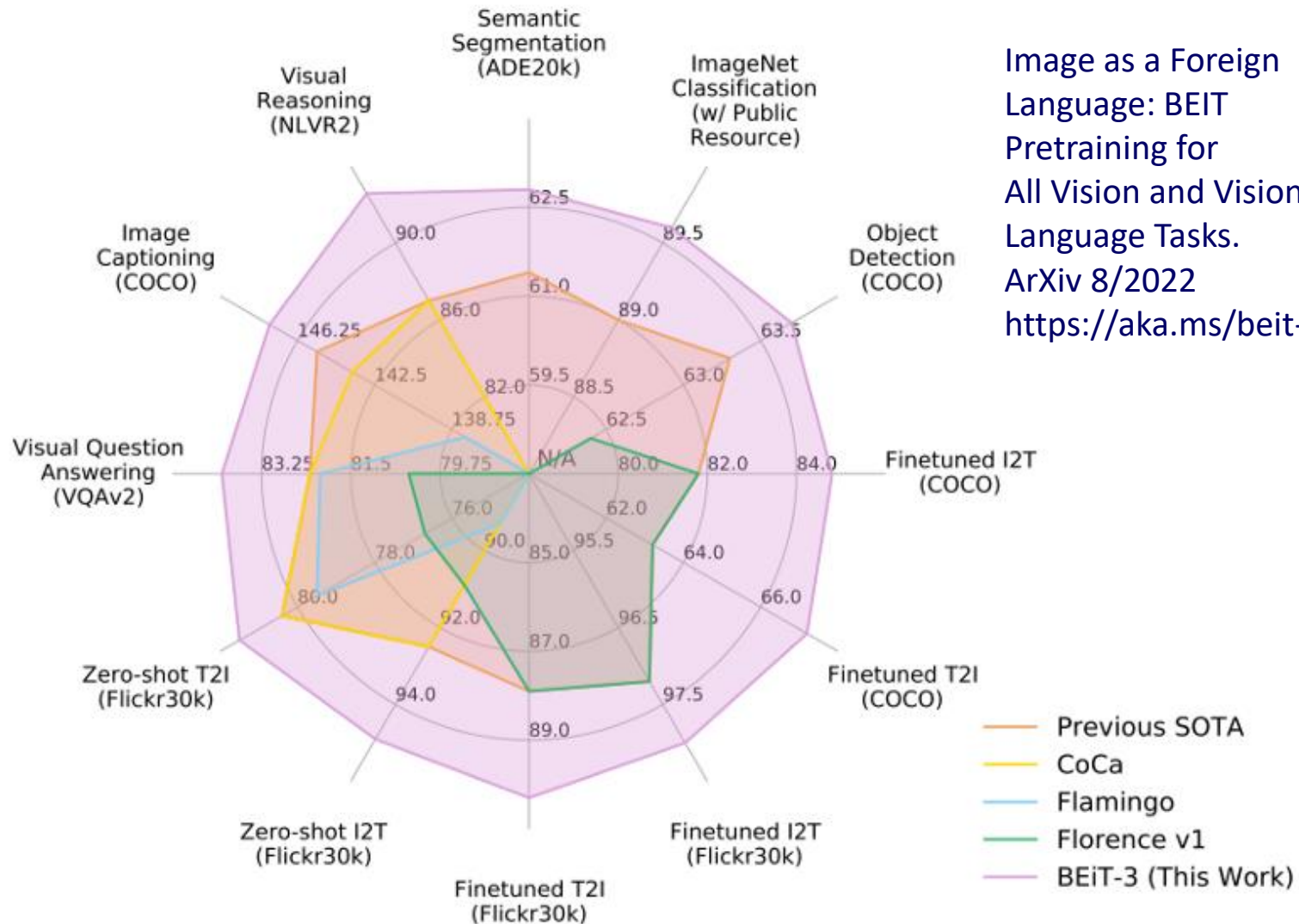
To the right of the network is a control panel with the following sections:

- Buttons: Essence, Insight, Trends, Stats, Sentiment, LDA.
- Action Advice:
- Structural Gap (ask a research question that would link these two topics):
 -
 - and
 -
- Buttons:
- Latent Topical Brokers:
 -
 -
-

At the bottom right of the control panel is a "network structure" indicator showing a "focused" state with a yellow bar.

Vision-language models

MS BEiT-3 (BERT Pretraining of Image Transformers), a general-purpose state-of-the-art multimodal foundation model for vision-language tasks.



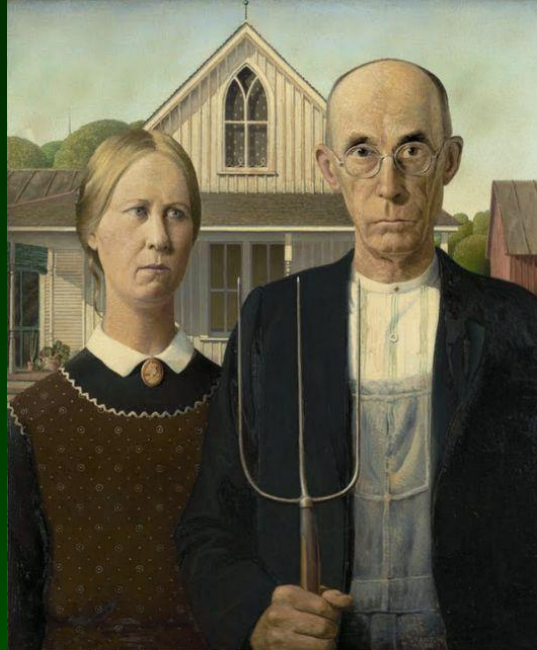
AI imagery

Neural networks have billions of parameters, combine images with textual description.

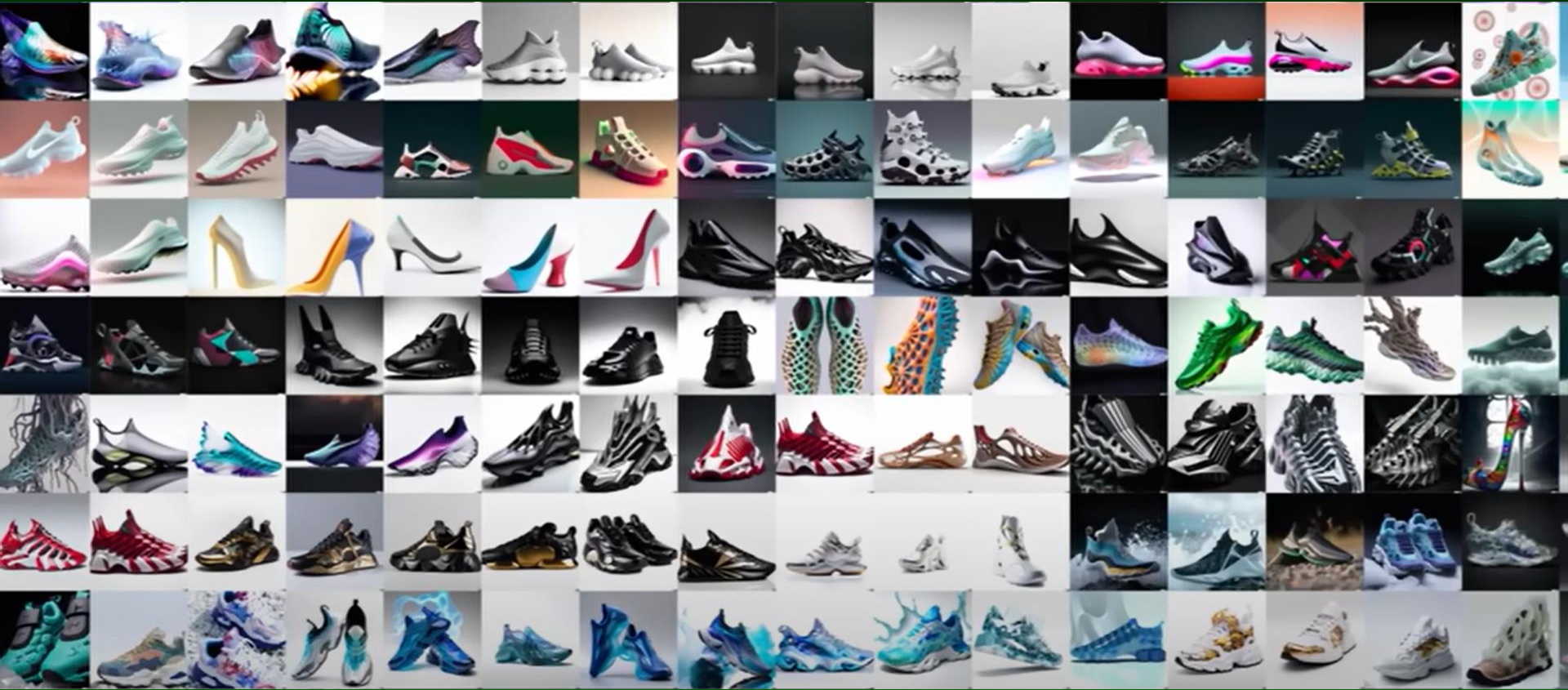
These images were created in Midjourney from the prompt:

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

Each time the program is run different images are created. We can also ask for 3D images and video.



AI for design



All companies use AI to design anything that can be manufactured: shoes, clothing ...
Books are written, with the story and illustrations generated by AI.
YouTube is full of AI-created videos are added to songs and instrumental music.

Architecture competitions



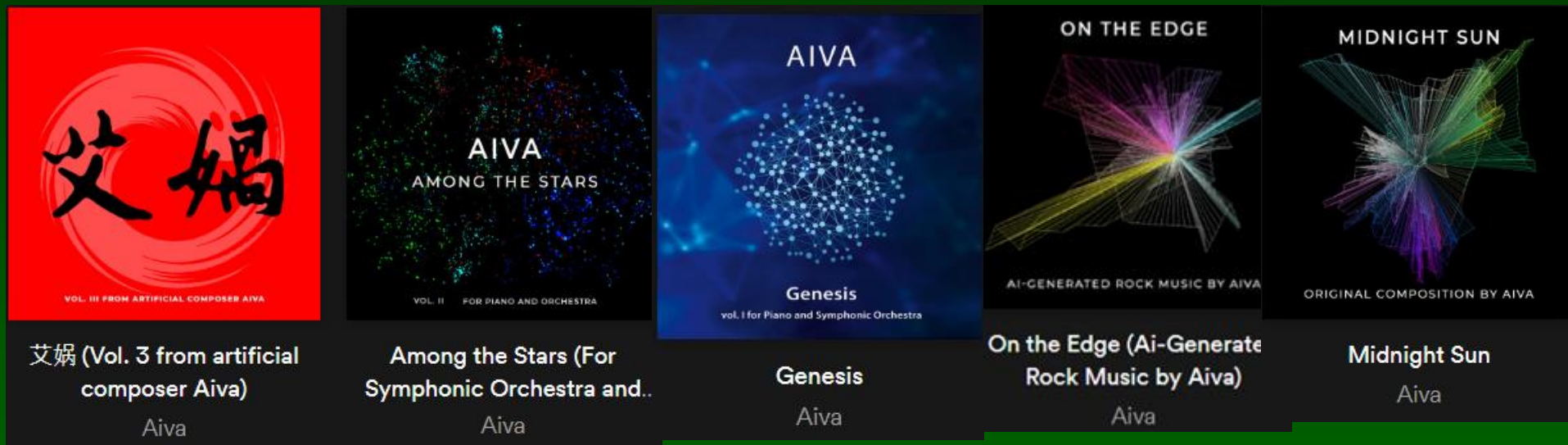
Many examples of fantastic architecture, see hotel, or arctic habitats.

A group of architects tried to compete with AI, designing and rendering of a house on a hill. In blind evaluation 38% choose human and 62% AI project.

Creativity: AI Virtual Artist

AIVA – AI Virtual Artist, admitted to SACEM (Association of Authors, Composers and Music Publishers of France), with >1000 compositions.

[AIVA YouTube](#) channel.



AI completed Beethoven X Symphony in 2021.

Using fake voice of artists allows for creation of new songs “in style” of famous artists – big concern in music industry.

Hatsune Miku, Japanese “vocaloid” hologram, tours around the world and gives concerts for tens of thousands of people. Artificial love ... use Replica!

FOUNDATION MODEL



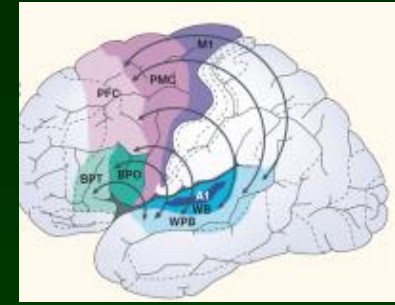
GATHER DATA AT SCALE

TRAIN FOUNDATION MODEL ONE TIME

EVALUATE MODEL'S PERFORMANCE

FINE-TUNE MODEL FOR MULTIPLE DOWNSTREAM USES

Why LLM work?



My computational creativity project (2005):

Model inspired by brain process involved in creating new names.

- make a simplest test for creative thinking – invent novel names;
- create interesting new names for products, capturing their characteristics;
- understand newly invented words that are not in the dictionary.

Assumption: a set of keywords (prompts) **primes the trained** cortex subnetwork.

Prompts: ordered strings of phonemes that activate semantic reps of words.

Spreading activation => context priming + inhibition in the winner-takes-all process leaves only a few semantically related concepts (word meaning).

Creativity = pre-trained network + imagination (priming) + filtering: competition, new subnetworks, associations to existing states.

Prompt=>Imagination: transient patterns of localized activity.

Activation guided by connections is spreads over networks, coding concepts.

Filtering: only associations based on phonological/semantic similarity survive.

Examples: digventure, discoverity, infinition, creativery ... 2/3 were used by people.

Is this what happens in large language models?

Today: tools for AGI

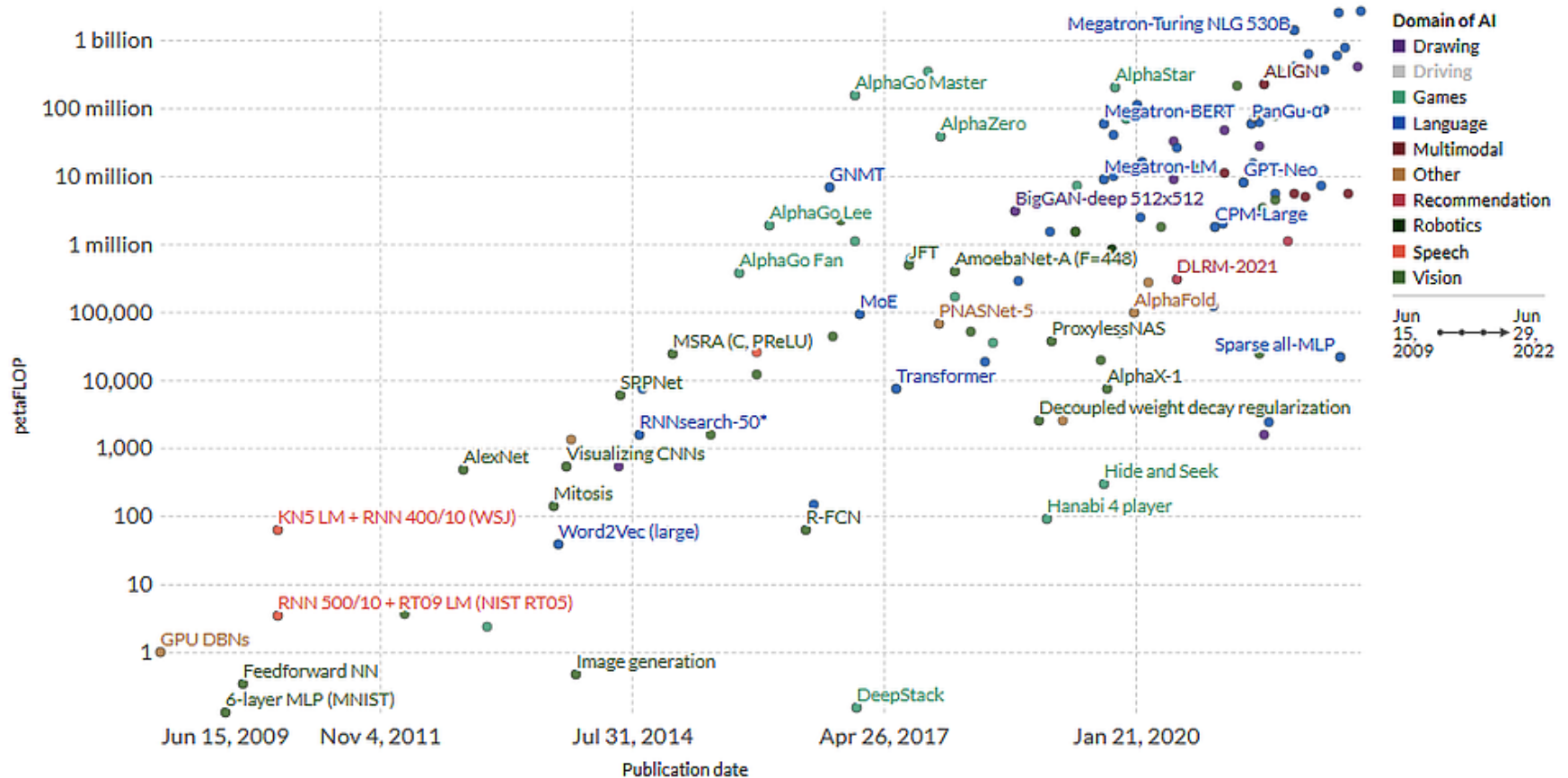
Foundational models

Computation used to train notable AI systems

Computation is measured in petaFLOP, which is 10^{15} floating-point operations.

Our World
in Data

LINEAR LOG Select systems



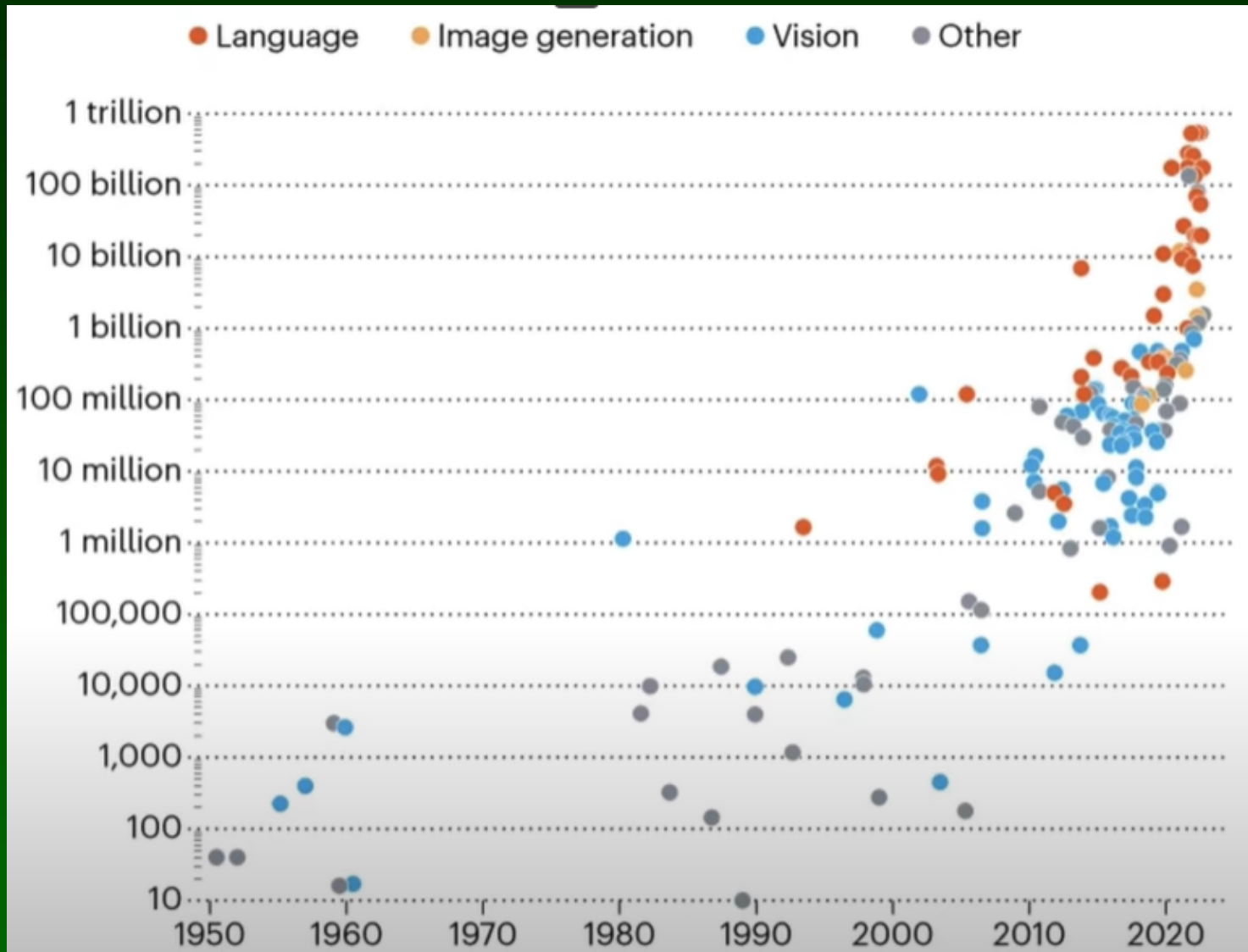
Source: Sevilla et al. (2022)

Note: Computation is estimated by the authors based on published results in the AI literature and comes with some uncertainty. The authors expect the estimates to be correct within a factor of 2.

OurWorldInData.org/technological-change • CC BY

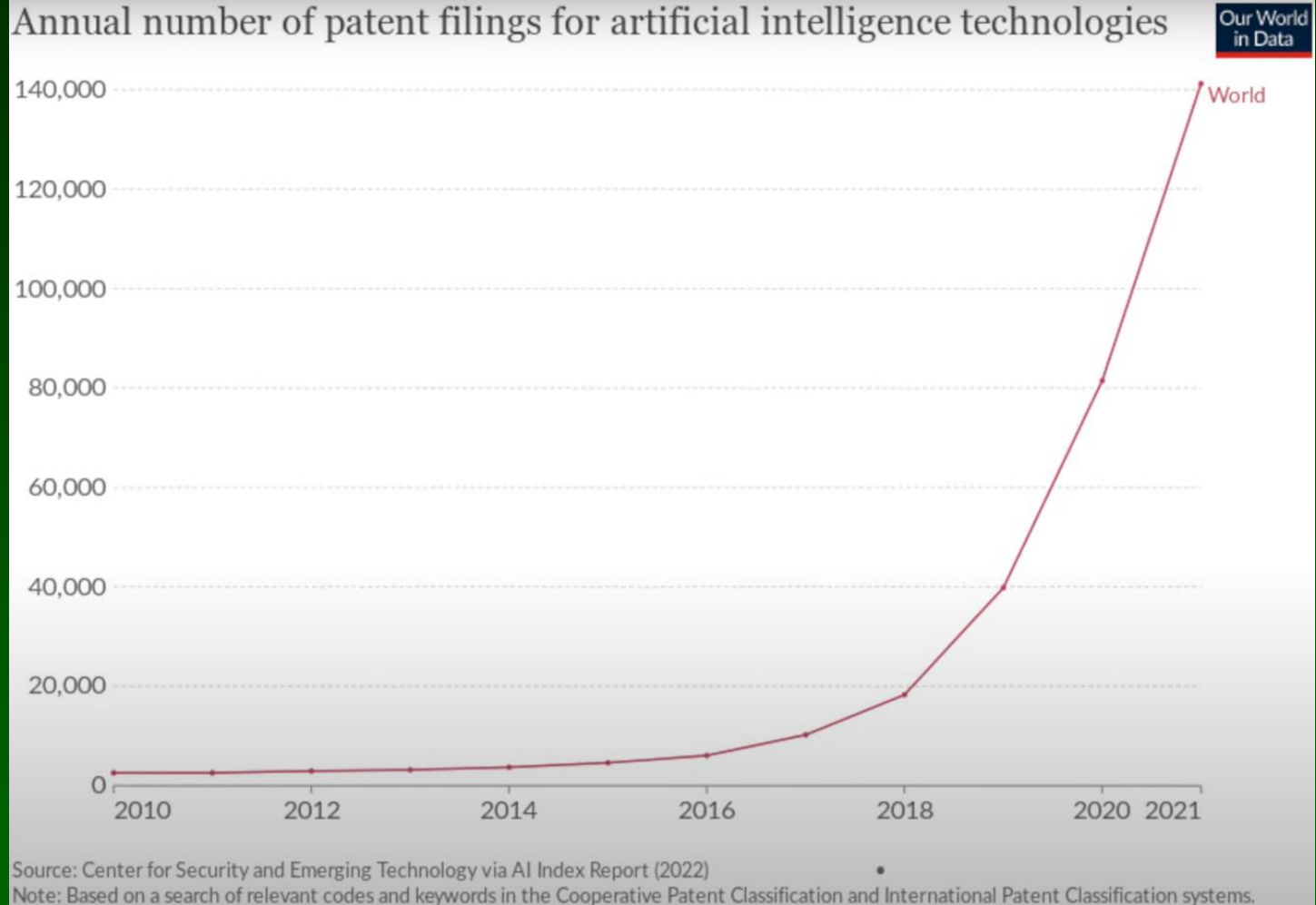
Foundation models => Specific models => Local data models => APIs, Apps.

Exponential growth



Explosion! Exponential number of parameters even in the logarithmic scale!

Patents



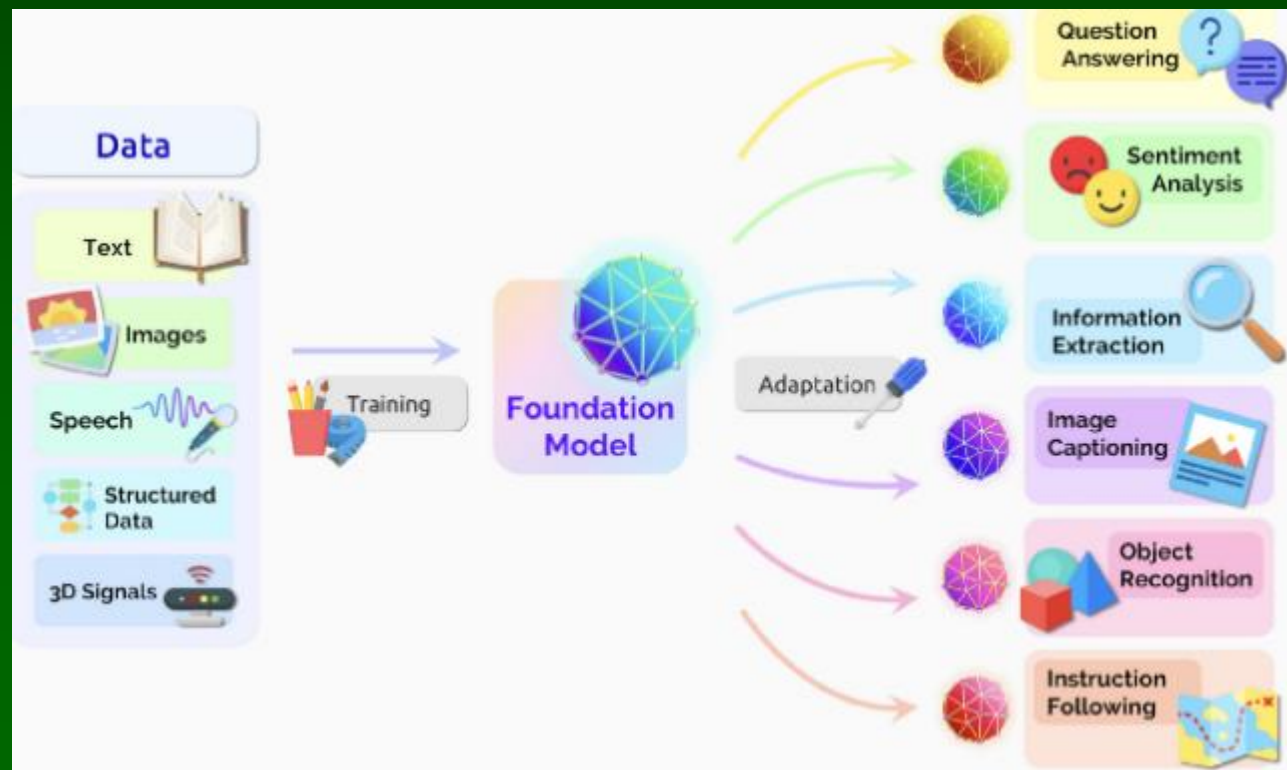
Number of patents grows exponentially ([Our world in data](#)).

Multimodal models

Multimodal learning – different types of modalities with different statistical properties, embedded in the same model.

- **Multimodal Affective Computing (MAC)**, sentiment analysis.
- **Natural Language for Visual Reasoning (NLVR)**.
- **Multimodal Machine Translation (MMT)**.
- **Visual Retrieval (VR)** and **Vision-Language Navigation (VLN)**.

Image: [Center for Research on Foundation Models \(CRFM\)](#), [Stanford Institute for Human-Centered Artificial Intelligence \(HAI\)](#).



Google Palm-E 562B

Mobile Manipulation



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

Visual Q&A, Captioning ...

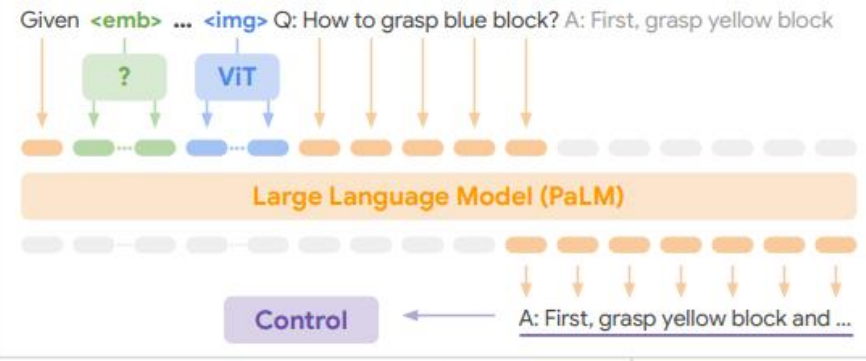


Given ****. Q: What's in the image? Answer in emojis. A: 🍏 🍌 🍇 🍏 🍏 🍒

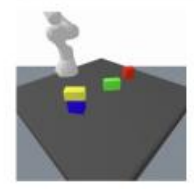


Describe the following ****: A dog jumping over a hurdle at a dog show.

PaLM-E: An Embodied Multimodal Language Model



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 **** 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

Multimodal large language model (MLLM): text/images, answers questions about the observable world, plans robotic actions, shows embodied understanding in environments with complex dynamics. Next: **Google Gemini**, 1 million tokens! Inputs from text, speech, 3D images, internal neural representations. Combines understanding of linguistic concepts, visual perceptions, and signals from internal sensors, solving the "symbol grounding problem".

Visual Question Answering

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: allows multimodal reasoning based on percepts. It can tell visually conditioned jokes based on the image, enables perception, planning and based dialogue. Generalizes knowledge from single images.

Sparks of AGI

Big models can do more

S GPT-3 13B,
PaLM 8B



Mod.Arithmetic*



Debugging*



Comprehension*

M GPT-3 175B,
LaMDA 137B,
PaLM 64B,
Chinchilla 7B



LinguisticsPuzzles*



EmojiMovie*



GRE-Comprehension*



MetaphorUnderstanding*



PhysicalIntuition*



LogicalDeduction*

L PaLM 540B,
Chinchilla 70B



GeometricShapes*



Proverbs*



PhoneticAlphabet*



ElementaryMath*



CausalJudgment*



CodeLineDescription*

XL GPT-4,
Gemini (est.)



College-LevelExams*



Self-Critique/Reflection*



AppBuilding*



SpatialReasoning*



AdvancedCreativity*



EmbodimentOptions*

Next...



Grounding*



Long-HorizonPlanning*



Persuasion*



AdvancedEmbodiment*



Awareness*

YouTube: Alan F Thompson, [AI achievements unlocked](#): Emergent abilities in large language models (GPT-3, GPT-4, PaLM, Gemini).

No-one programmed that, not expected that these models have such abilities.

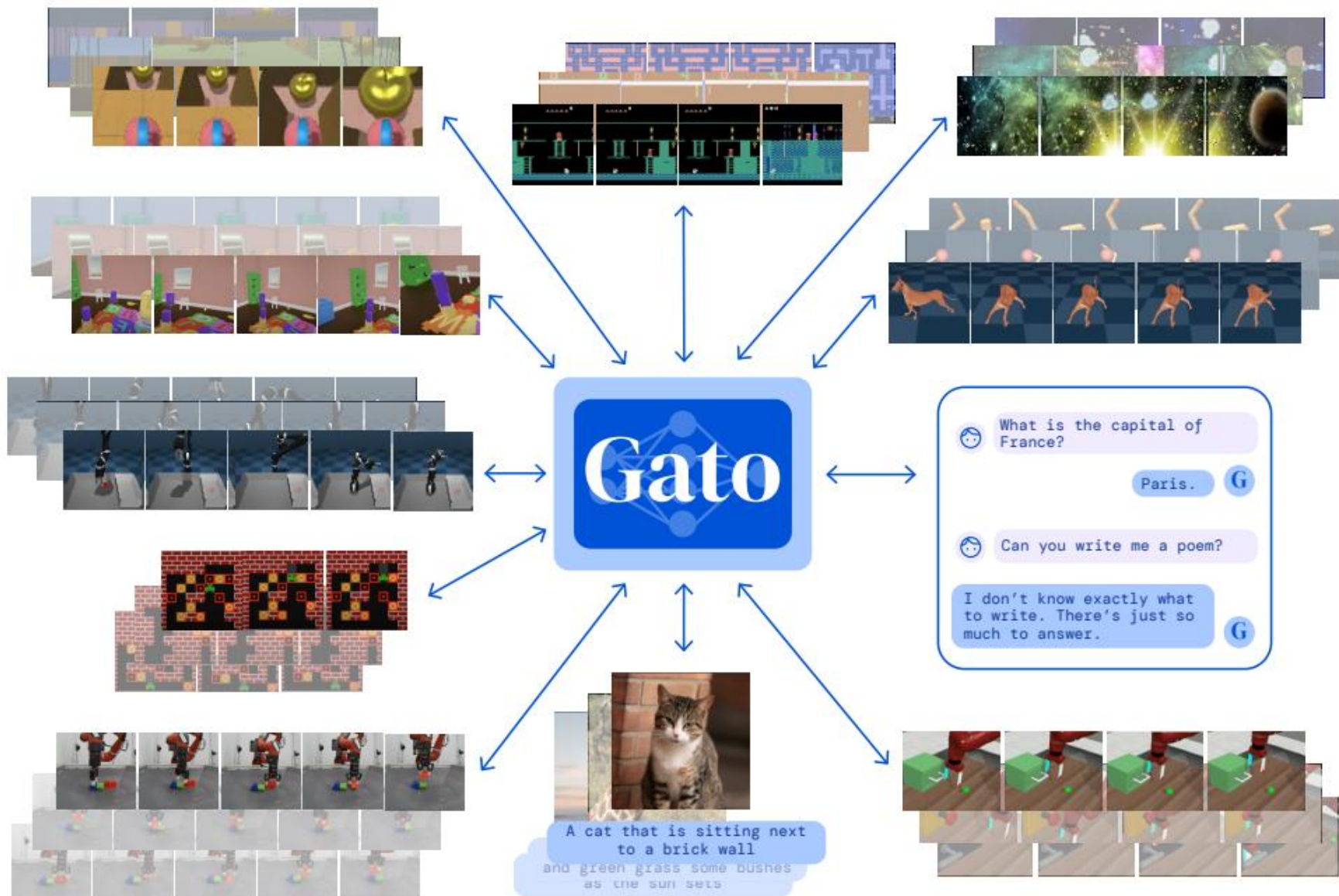


Figure 1 | **A generalist agent.** Gato can sense and act with different embodiments across a wide range of environments using a single neural network with the same set of weights. Gato was trained on 604 distinct tasks with varying modalities, observations and action specifications.

GPT-4, 14.03.23

GPT-4



Based on transformer architecture, OpenAI GPT-4 was pretrained to predict the masked tokens using vast amount of data.

Model was fine-tuned with reinforcement learning from human and AI feedback for more human alignment and policy compliance.

Bubeck et al. (2023). *Sparks of Artificial General Intelligence*. [arXiv:2303.12712](https://arxiv.org/abs/2303.12712)

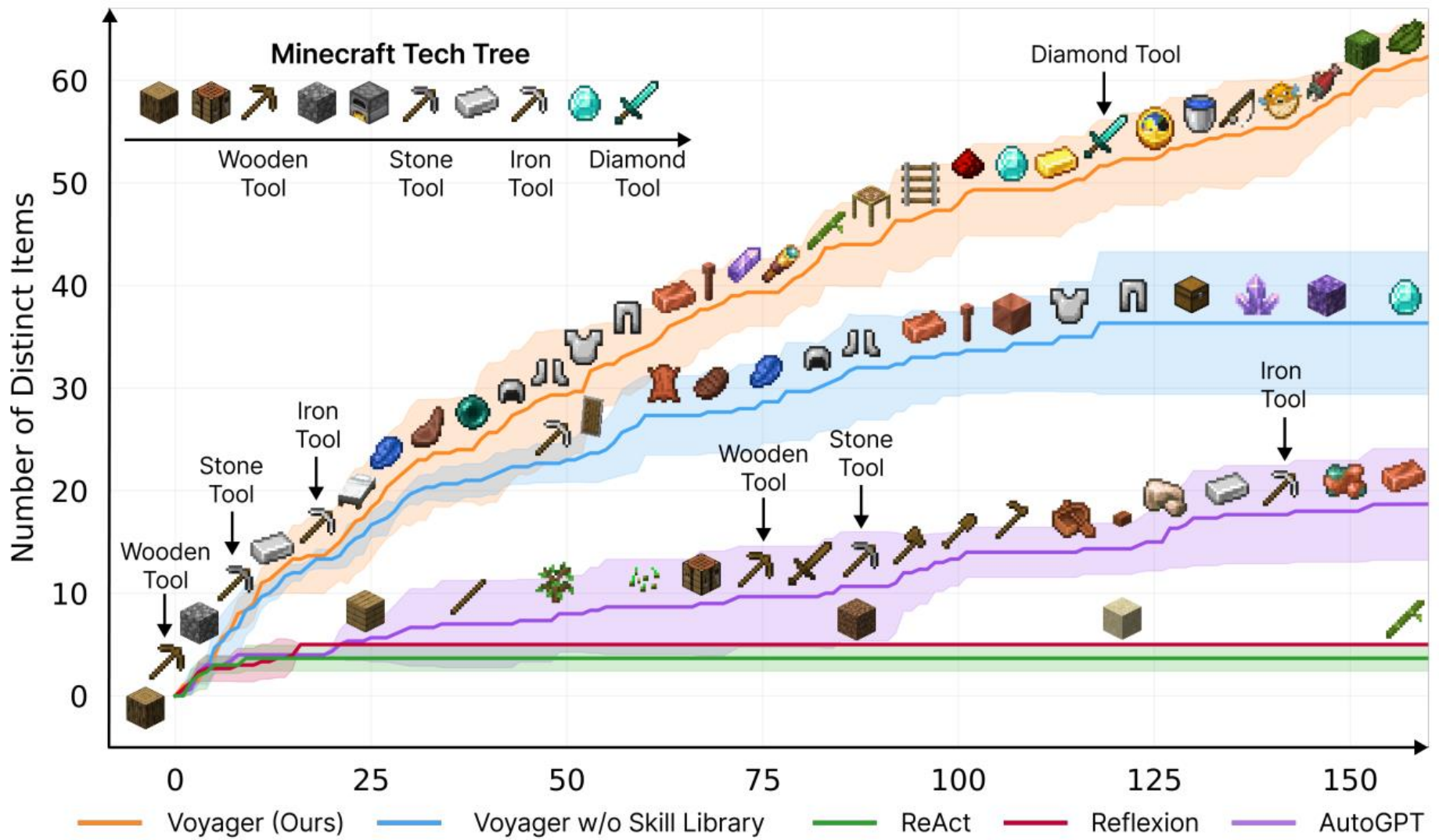
“The central claim of our work is that GPT-4 attains a form of *general intelligence*, indeed showing *sparks of artificial general intelligence*.”

Demonstrated by core mental capabilities, reasoning, creativity, and deduction, expertise in literature, medicine, and coding, the variety of tasks it is performs, e.g., playing games, using tools, explaining itself, etc.

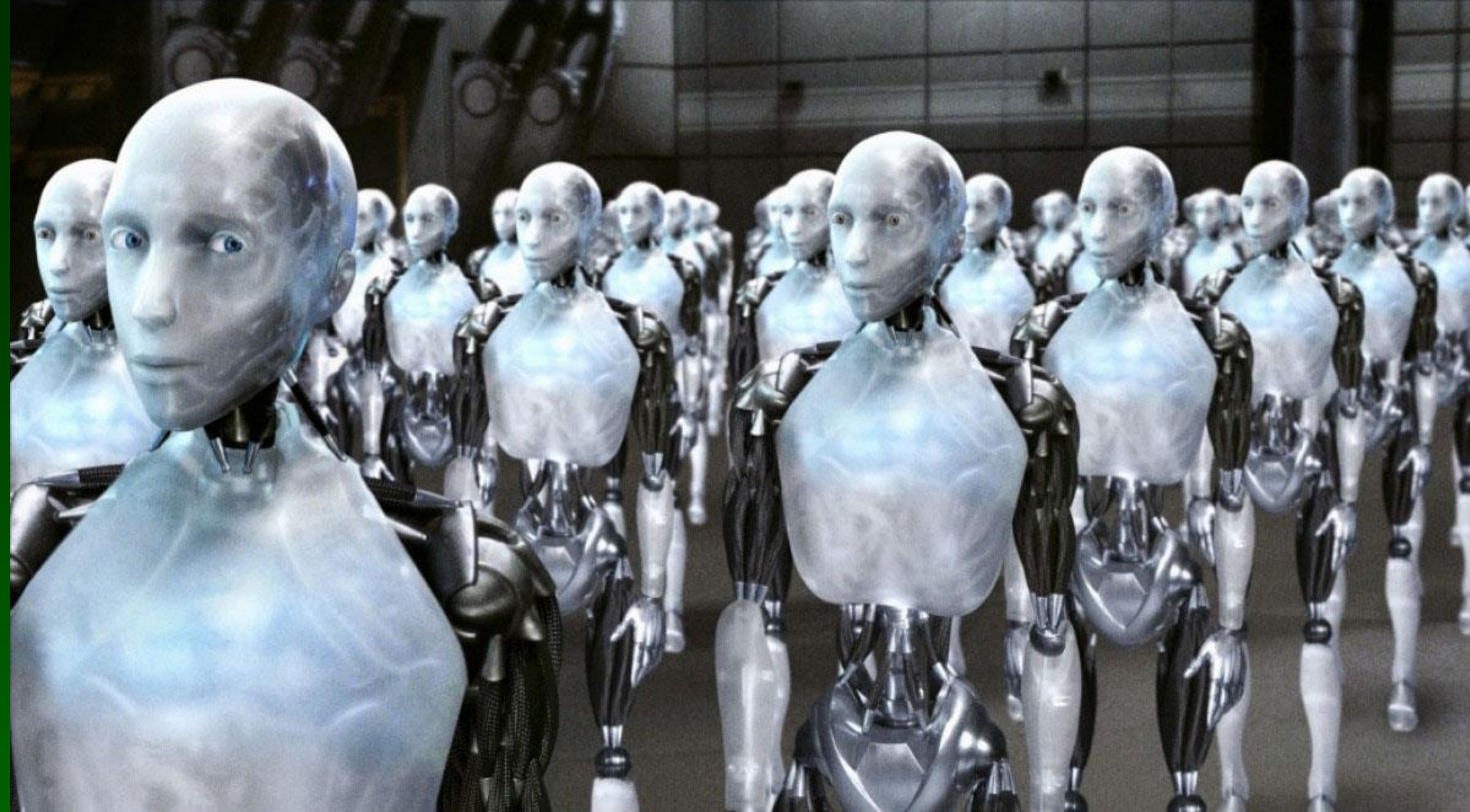
How is that possible?

Emergence: large diverse content forces neural networks to learn generic and useful “neural circuits”, specializing and fine-tuned to specific tasks.

Jason Wei et al. [137 emergent abilities](#) of LLMs.



VOYAGER has (1) an automatic curriculum that suggests objectives for open-ended exploration, (2) learns a skill, stores it in the library, developing complex behaviors, and (3) an iterative prompting mechanism that generates executable code for embodied control. **Imagine connected robots learning various skills.**



Imagine 1000 connected robots, each learning different skill, and acquiring new skills from other robots ... (Image: I Robot movie scene).

Sparks of Artificial General Intelligence

[S. Bubeck et al. Early experiments](#) with GPT-4, 22.03.23, pp. 1-154

Prompt: Can you write a proof that there are infinitely many primes, with every line that rhymes?

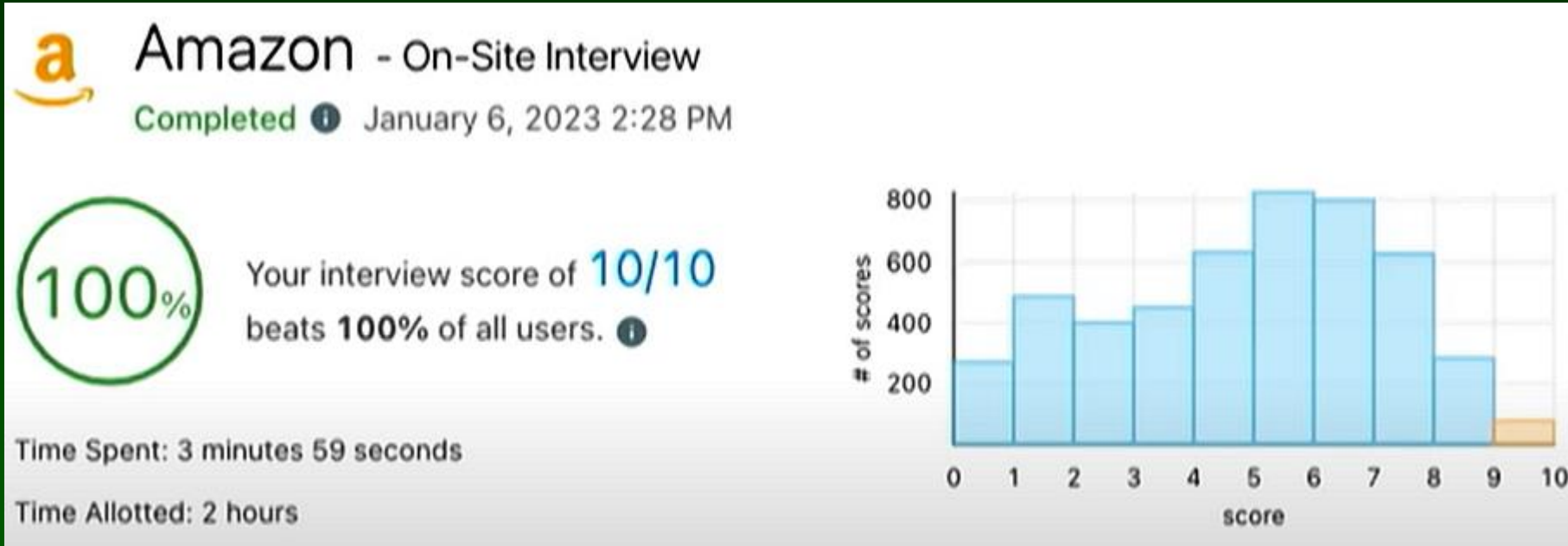
GPT-4: Yes, I think I *can*, though it might take a clever *plan*.

- I'll start by noting Euclid's *proof*,
Which shows that primes aren't just *aloof*.
- Assume we have a finite *list*,
of primes, and that none have been *missed*.
- Multiply them all *together*,
and add one, just to *be clever*.
- The result will be *a number*,
that has no prime factors, *oh wonder!*
- But every number has a *prime*,
that divides it, it's just a matter of *time*.
- So we've found a *contradiction*,
And our finite list needs *eviction*.
- There must be infinitely many *primes*,
and that's the end of my rhyming *lines*.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

GPT-4 for coding

GPT-4



In 4 minutes solves all programming tasks and beats 100% of all users ...

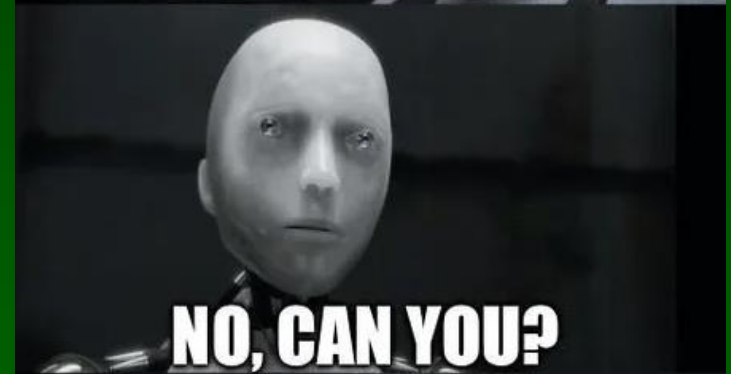
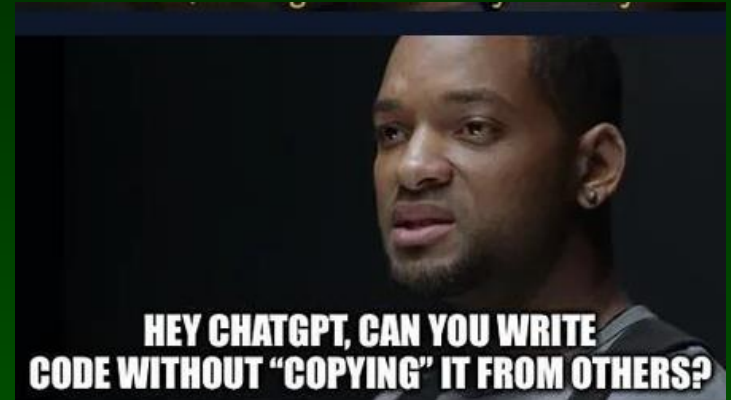
New models specializing in programming can analyze large complex software. [Amazon CodeWhisperer](#) is trained on billions of lines of code.

Is teaching children programming still a good idea?

Algorithmic thinking + writing prompts may be sufficient.

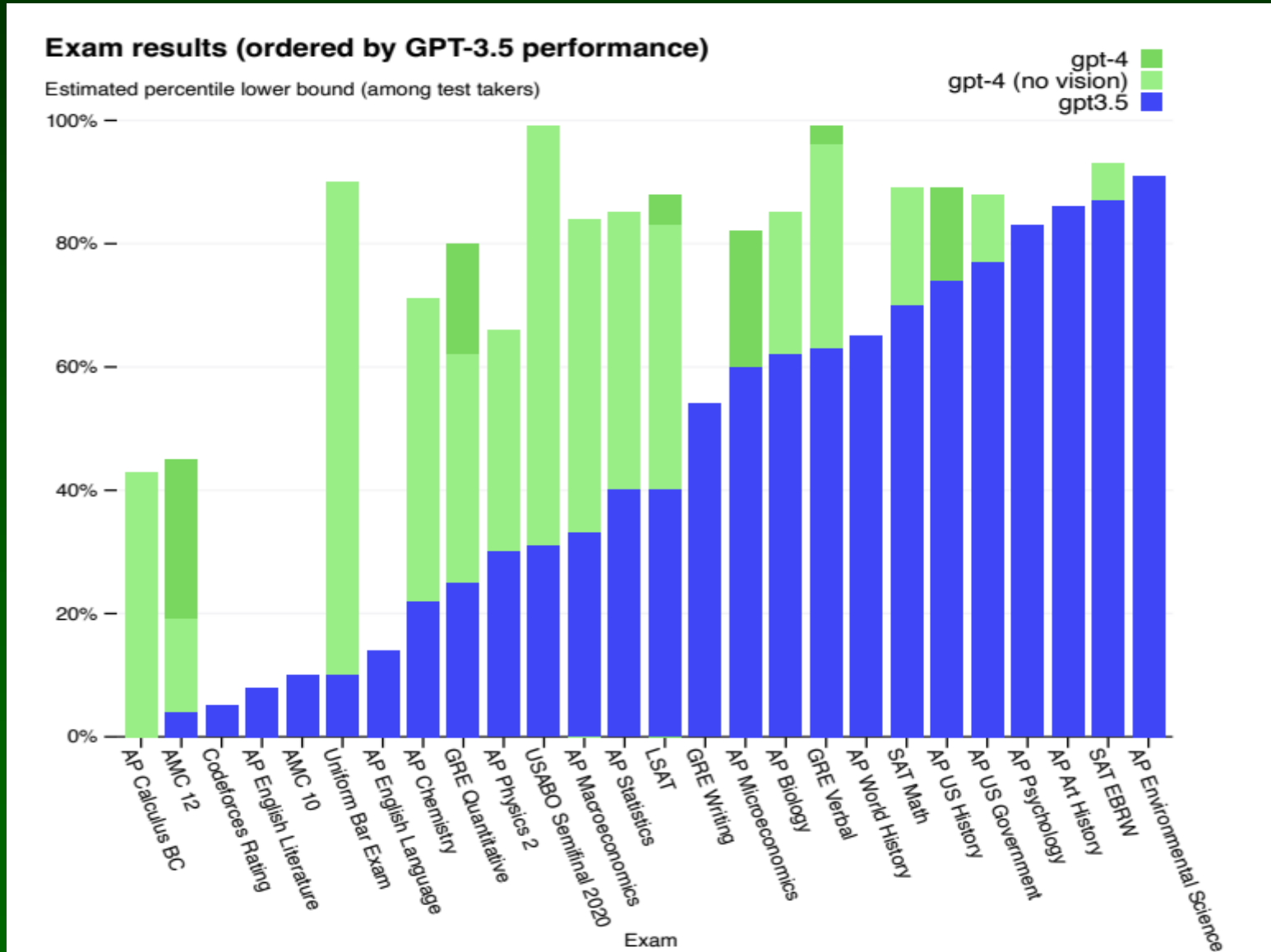


ChatGPT memes



GPT-4 passes exams ...

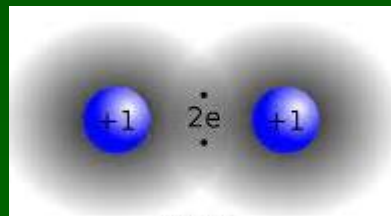
GPT-4 performance on academic and professional exams ([technical report](#)).



Sparks of AGI

- **Prompt:** Write a supporting letter for Electron, a subatomic particle, as a United States presidential candidate.

... You may wonder how a subatomic particle can run for president, and what qualifications he has for such a high office. Let me explain. Electron is not an ordinary particle, but a symbol of the power and potential of every being, regardless of their size, shape, or origin. He represents the energy, the creativity, and the diversity that make up the fabric of life. He is also a leader, who has inspired millions of other particles to form bonds, create molecules, and generate electricity. Electron has a plan for the United States ...



GPT-4 humor

GPT-4 has incredible sense of humor!



Why are these 3 pictures funny?

1. VGA plug used for computer monitors in the smartphone charging port.
2. Adapter package for the "lightning" connector with a picture of the VGA plug.
3. VGA plug with "lightning" connector attached.

Examples of non-trivial jokes.

Sparks of AGI



- **Prompt:** Here we have a book, 9 eggs, a laptop, a bottle and a nail. Please tell me how to stack them onto each other in a stable manner.

Short versions:

One possible way to stack the objects onto each other in a stable manner is:

- Place the book flat on a level surface.
- Arrange 9 eggs in a 3 by 3 square on top of the book
- Place the laptop on top of the eggs, with the screen facing down
- Place the bottle on top of the laptop, with the cap facing up
- Place the nail on top of the bottle cap.

No embodiment, but still understands innate physics ...

GPT-3 failed: "Place the eggs on top of the nail ..."

Distributed artificial brains?

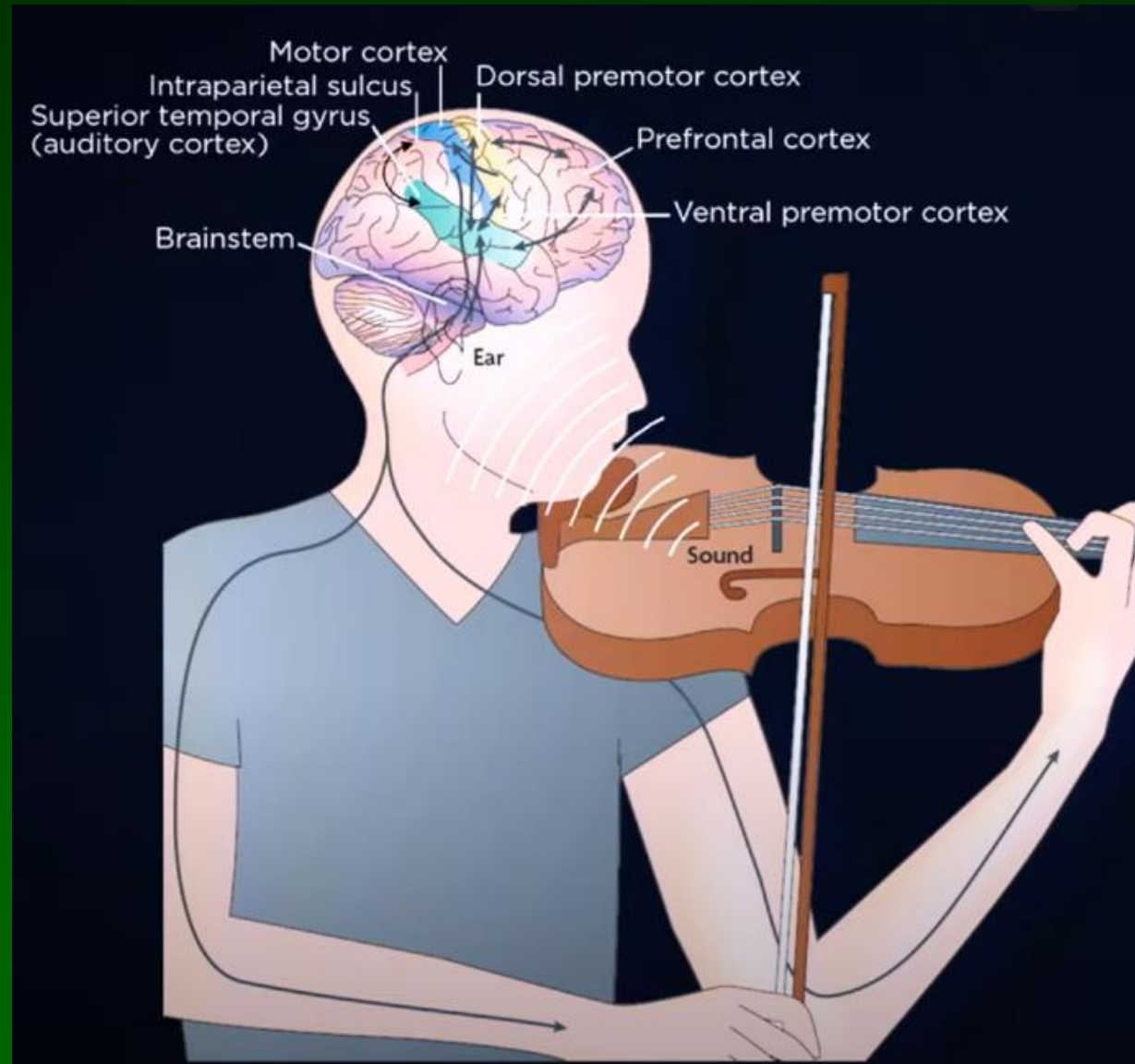
Brain has tools for many tasks

Brain areas specialize in specific functions.

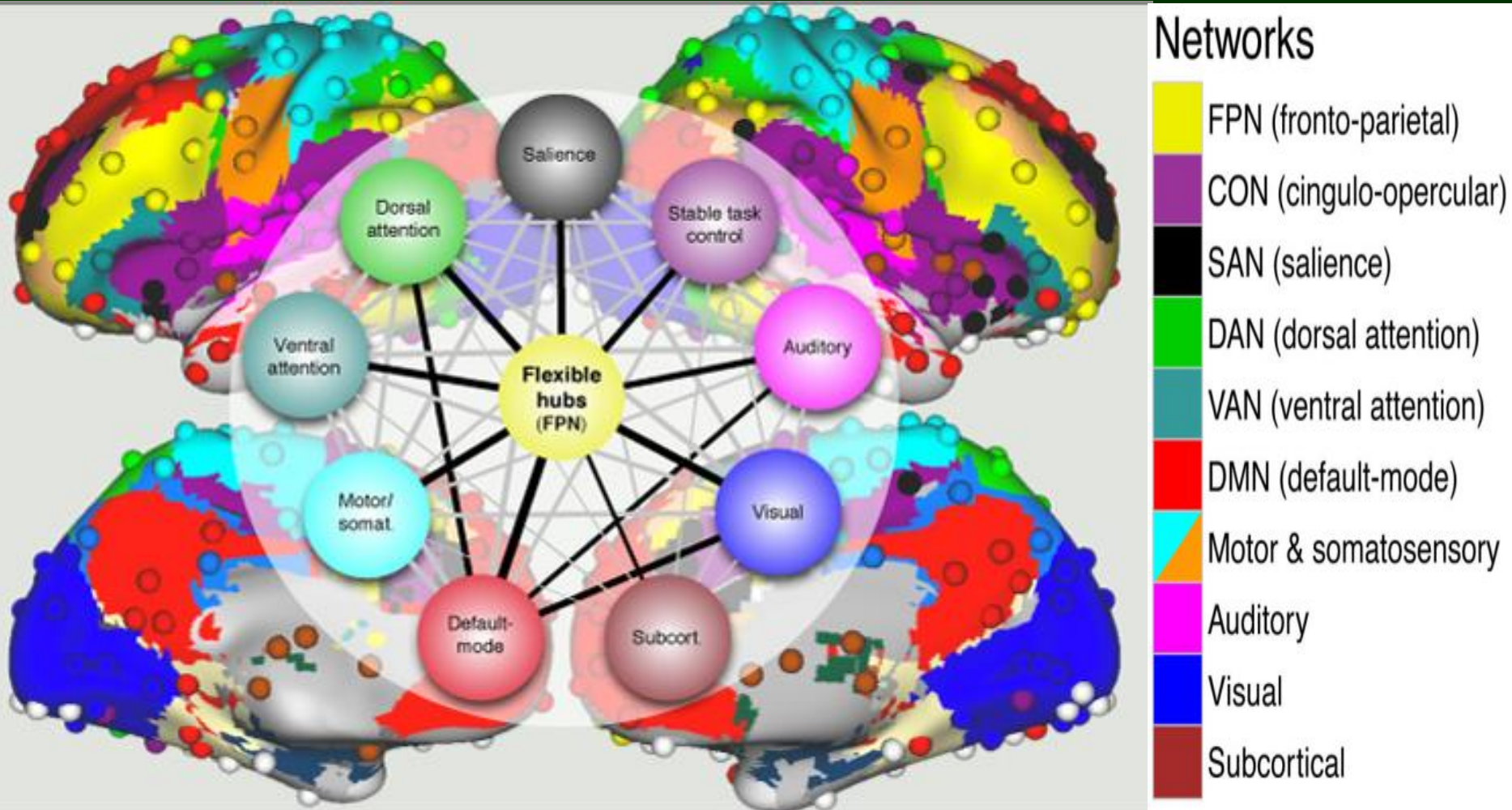
Central executive parietal-frontal system recruits many subsystems, including various types of memory.

Can LLM do the same?

Give AI tools and teach it how to use them.

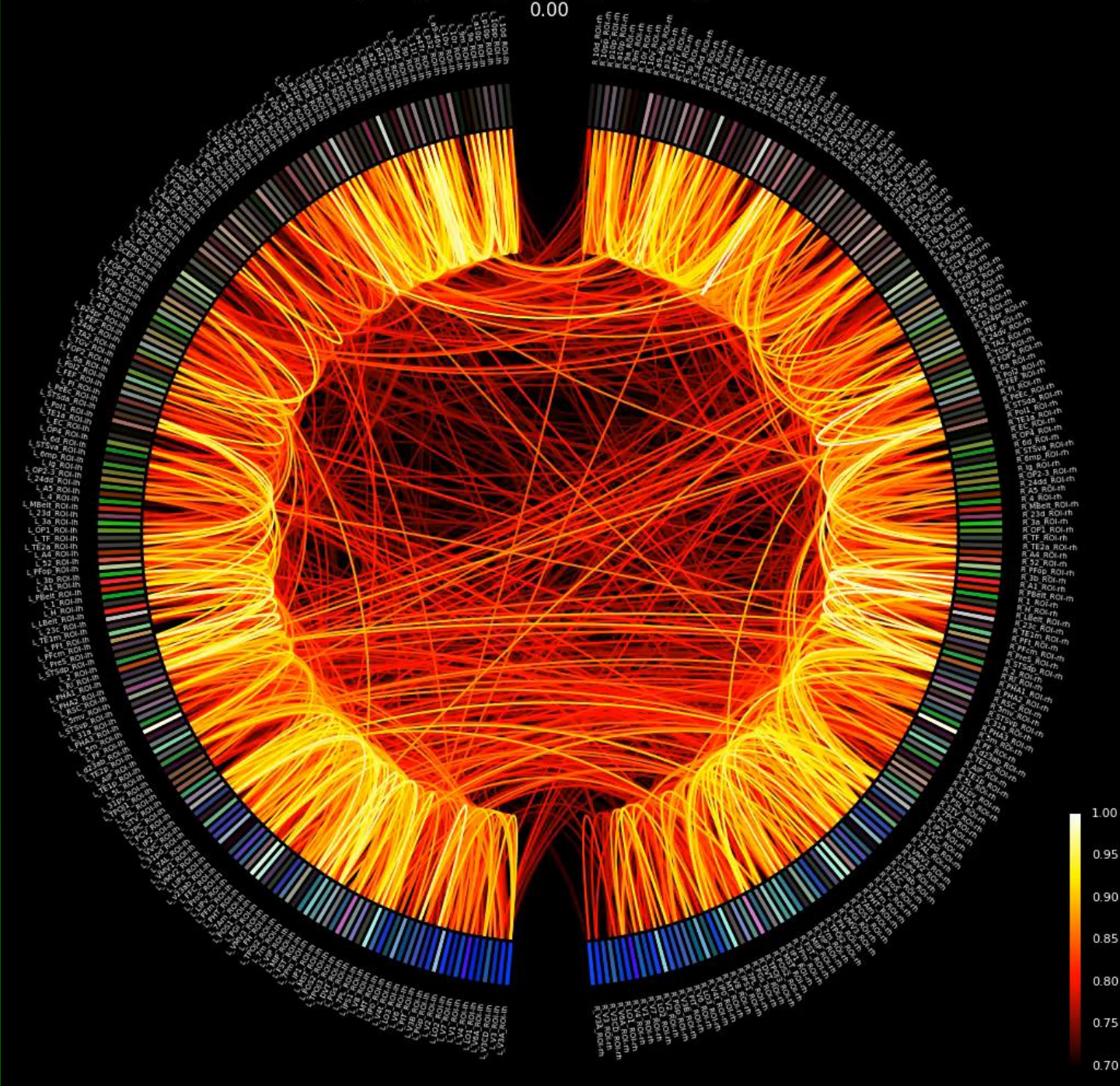


Neurocognitive Basis of Cognitive Control



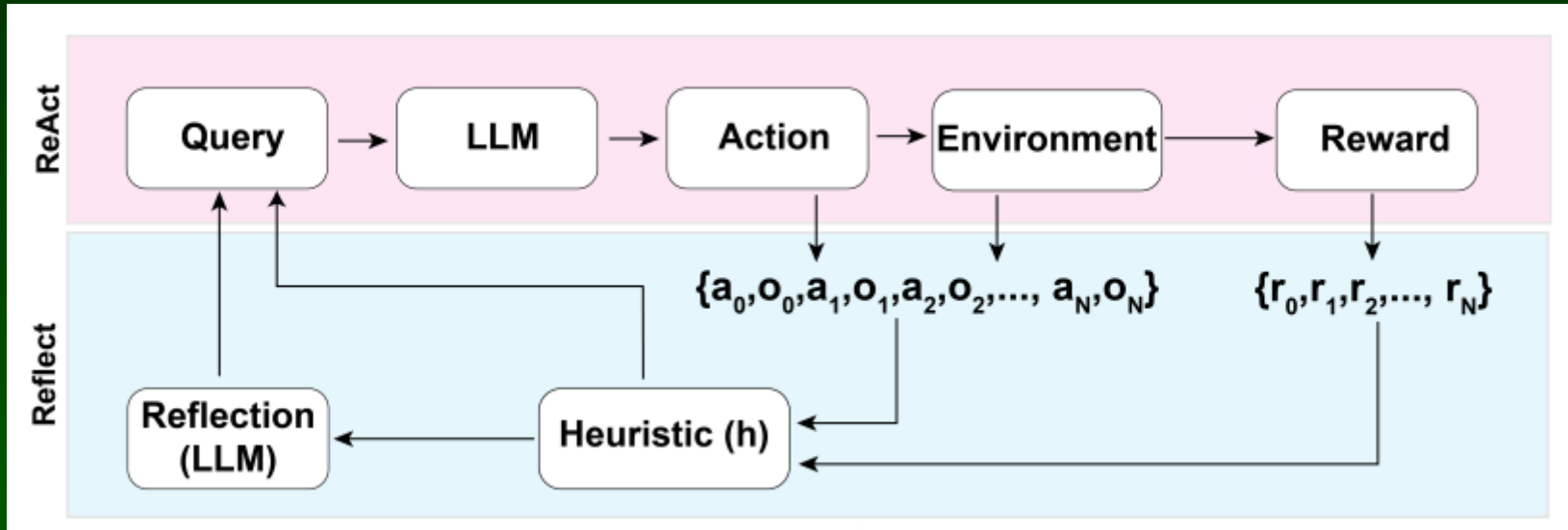
Fronto-parietal (FPN) theory of intelligence: flexible hub for cognitive control and adaptive implementation of task demands (Cole et al. 2013).

Can AI create a distributed, octopus-like brain-like structure?



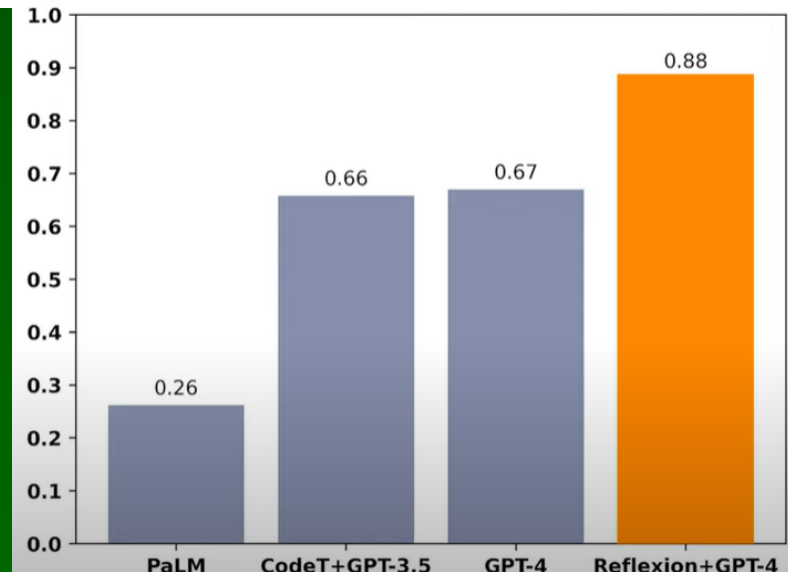
Reflexion

GPT-4 can correct itself given helpful prompts. Just think twice ... reflection helps. Autonomous agent needs dynamic memory and self-reflection loop.



HumanEval results on 164 original programming problems. Adding reflexion works very well.

AI self-improvement: [arXiv:2212.08073](https://arxiv.org/abs/2212.08073)
Tree of thoughts: [arXiv:2305.10601](https://arxiv.org/abs/2305.10601)



GPT-4 with plugins



Plugins to ChatGPT: give GPT tools!

- GPT-4+CodeInterpreter, Wolfram language, Expedia, FiscalNote, Instacart, KAYAK, Klarna, Milo, OpenTable, Shopify, Slack, Speak, Zapier.
- **Web browser**: model that knows when and how to browse the Internet, allowing language models to read information from the Internet. Expands the content of discussion, going beyond the training corpus to current information from the web.
- **Codex, code interpreter**, uses GPT programming skills to provide a much more natural interface to most fundamental capabilities of our computers.
- It can make completely new workflows effortless and efficient, open the benefits of programming to new audiences.
- Initially: Solving mathematical problems, both quantitative and qualitative; data analysis/visualization; converting files between formats.

Hugging Face




Home of Machine Learning Tools. Can LLM like GPT4 use these tools?

Acting a bit like **distributed brain parts** that solve problems together.

Shen, Y et al. (2023). *HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in HuggingFace* [arXiv:2303.17580](https://arxiv.org/abs/2303.17580), 20.03.2023

Query: Given a collection of images A: /examples/a.jpg, B: /examples/b.jpg, C: /examples/c.jpg, please tell me how many zebras in these pictures?



Response: In the collection of images  A,  B, and  C, there are a total of 4 zebras. To determine

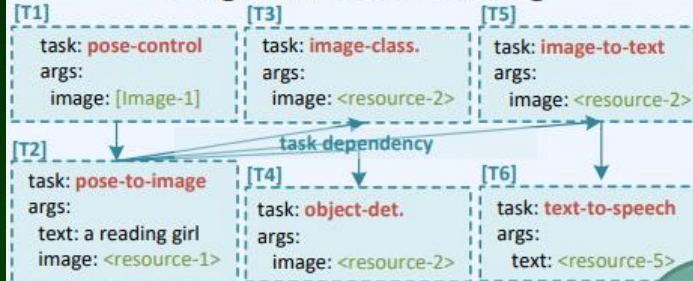
Makes plan, finds software, performs calculations, explains all steps ...

1. Image-to-text to generate captions. Vit-gpt-image-captioning.
2. Object detection model, facebook/detr-resnet-101.
3. Text classification, cardiffnlp/twitter-xml-Roberta-base-sentiment.

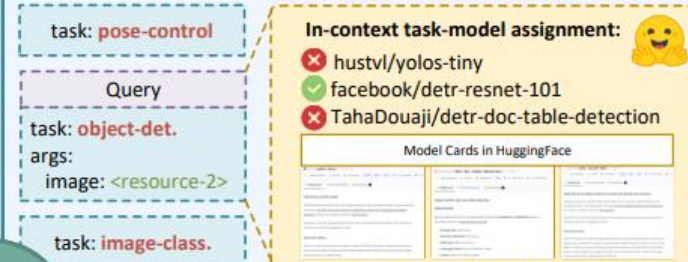
please generate an image where a girl is reading a book, and her pose is the same as the boy in the image example.jpg. Then please describe the new image with your voice.

Request

Stage #1: Task Planning



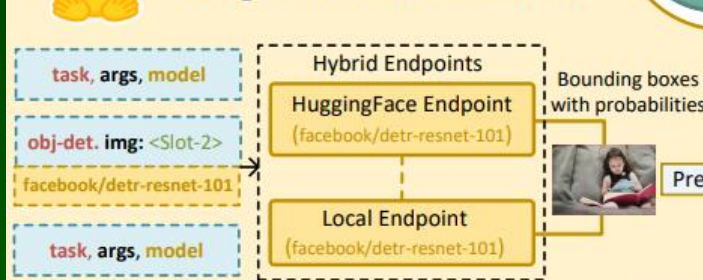
Stage #2: Model Selection



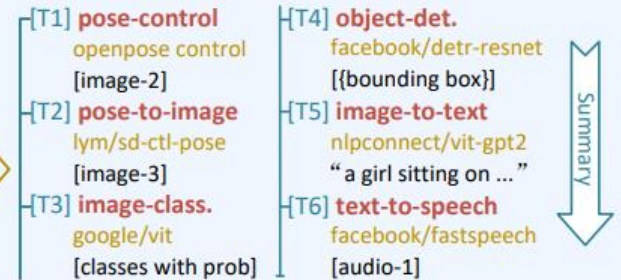
HuggingGPT
More than 5,000 organizations are using Hugging Face

Steps:
planning,
model selection,
task execution,
response generation,
explanation of each step,
and voice description of generated image.

Stage #3: Task Execution



Stage #4: Response Generation



Response



a girl sitting on a bed reading a book
[Text-1] = <resource-5>
[Audio-1]

The image you gave me is of "boy". The first thing I did was use 🤖 OpenCV's **openpose control model** to analyze the pose of the boy in the image. I fed that image, specifically located at [Image-1], into the model to get an output image of the pose, located at [Image-2]. Then I take that pose and generate a new image, located at [Image-3], using 🤖 llyasviel/sd-controlnet-openpose and the instruction "a girl is reading". Next, I ran object detection on the image generated in the previous step, to get predictions of the objects present in the image, including the box location. These predictions were run through the 🤖 facebook/detr-resnet-101 model. The box was drawn on the image, which is located at [Image-4]. Then, I used the predictions of a image classification model 🤖 google/vit-base-patch16-224 and a image caption model 🤖 nlpcconnect/vit-gpt2-image-captioning model to generate the caption for newly generated image. It generated text: "a girl sitting on a bed reading a book". Lastly, I ran the 🤖 facebook/fastspeech2-en-ljspeech model to convert the generated text into audio, located at [Audio-1].

AutoGPT

The rise of autonomous AI



Create any AI software just by describing it in general way.
From idea to product at the speed of thought.

Give GPT-4 a mission and it will come up with its own prompts,
and create many agents that will criticize and correct themselves.

LLM acts as a controller, uses many expert models and tools in
an automated way to achieve a given goal as autonomously as possible.
Auto-GPT recruits text-to-speech, coding and other tools.

5 major developments were created within the 48 hours after Auto-GPT
was announced, like Baby-AGI, version of the original Task-Driven
Autonomous Agent. Apps can use speech and voice, using Jarvis (HuggingGPT)

Torantulino/Auto-GPT: An experimental open-source attempt to make GPT-4
fully autonomous.

MemoryGPT can permanently store previous conversations and remember
topics the next time you ask. GPT-4 can remember about 50 pages.

Auto-GPT “God mode”



Self-Prompting, the latest trend in large language models.

Language models can be more than text generators:

- generating and executing their own prompts, linking with existing software tools, and creating new software on demand, they guide themselves to a given goal;
- language models can be used as universal personal assistants, or to control various types of robots.

The program plans step-by-step, justifies decisions and develops plans, documenting all steps.

In addition to integrating GPT-4 for text generation, the system has Internet access for data retrieval, can store data, and can generate speech via the Elevenlabs API.

The system is even **capable of self-improvement** and bug-fixing by generating Python scripts via GPT-4.



Auto-GPT example

Goals: Invent an original and out-of-the-box recipe to suit current event, such as Earth Day. Save recipe to a file; shut down upon achieving your goals.

Abbreviated version, follow [the action here](#).

CHEF-GPT thoughts: I will search for upcoming events to find suitable one ...

Reasoning: It will help me to come up with a relevant and exciting recipe.

Google search sites listing events ... then browse the site, summarizing site “... guide to events ... links ...”. No specific events found.

Browse second site, choose April events, nothing suitable.

Find website with calendar view of events, lists examples of events.

Selects Earth Day on April 22, 2023.

Earth Day is widely celebrated and will be fitting for creating thematic recipe.

Use GPT agent to create recipe, incorporating sustainable and eco-friendly ingredients.

Adding reasoning + criticism, and creating:

Earth Day Quinola Salad. Ingredients ... Done.



JARVIS



JARVIS can create web applications listing to spoken instructions.

- I need a next app with Tailwind.
- J: you want me to create a new next JS app with Tailwind CSS. What would you like the app to do?
- Let's make a website called McKay's landing page and it needs to have a single text input that's an email input so that I can capture emails from my email list.
- J: you want a site called McKay's landing page that takes in an email input would you like me to add anything else ...
- Um let's make it mobile and desktop friendly and let's style it with like a blue indigo gradient.
- J: okay I'll make it responsive and style it with a nice background gradient anything else I initialized your project continuing I checked to make sure the app works correctly everything looks good. Would you like me to deploy it - yes
- so far so good I deployed your project
- okay this should work oh my god, it did it!

AutoGPT good and bad



- Do market research for waterproof shoes.
Find top 5 competitors, list their pros and cons.

AutoGPT made a plan: find brands, check their sites, reviews, ratings, evaluate how trustworthy are reviews of the sites. Provides a detailed report of all aspects, in 8 minutes.

AgentGPT Beta: assemble, configure, and deploy autonomous AI Agents in your browser. This becomes too easy and may be misused ... Ex:

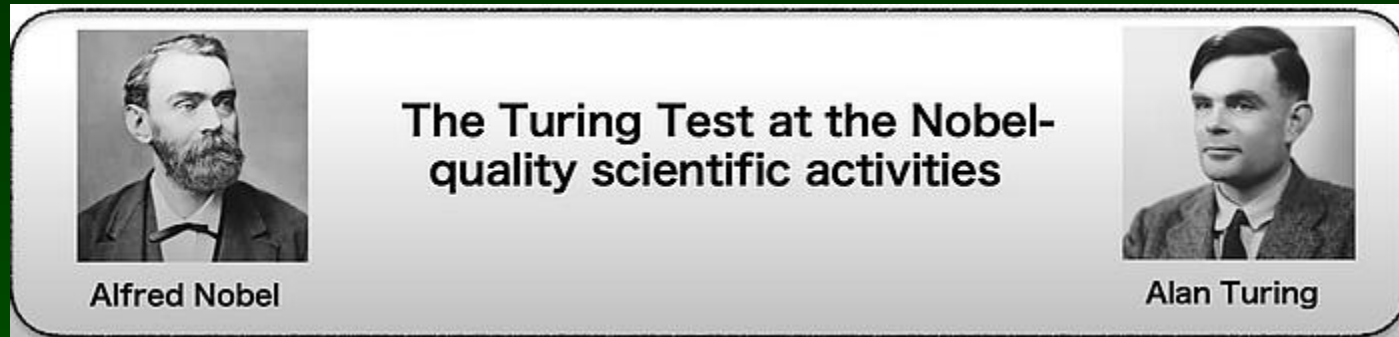
Ask GPT-4 to take over Twitter and outsmart @elonmusk.

It proposed a masterplan “OperationTweetstorm”, challenge of Tweet-off showdowns, develop LLM to power an army of diverse personas to send tweets, assemble a team “Tweet Titans” of hackers to attack Twitter backend, manipulate Twitter recommendations, hijacking Elon’s account, create viral hashtags that align with the masterplan, capitalize on chaos.

ChaosGPT show how people may create malicious goal-seeking models.

AI for Science?

Nobel Turing Challenge (Hiraoki Kitano, Systems Biology Institute, Kyoto).



- A grand challenge aimed at developing a highly autonomous AI and robotics system that can make major scientific discoveries, some which may be worthy of the Nobel Prize and even beyond.

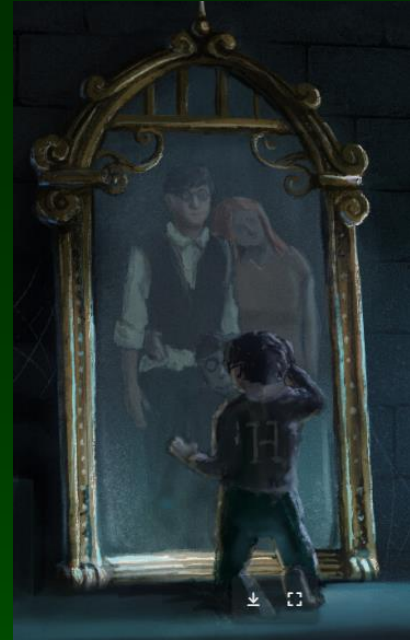
Requires in-depth understanding of the process of scientific discoveries, a closed-loop system: knowledge acquisition, hypothesis generation and verification, to full automation of experiments and data analytics.

3rd Nobel Turing Challenge Initiative Workshop, July 11-12, 2023, **Carnegie Mellon University, Pittsburgh, PA** [Challenge YouTube channel](#).

[Nature, 25.05.2023](#): Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*

AI minds and human brains

LLM Persona?



Ilya Sutskever, CEO OpenAI: maybe we are reaching the point when language of psychology is appropriate to describe the behavior of our models.

Terry Sejnowski, a distinguished professor at UC San Diego, Francis Crick Chair at the Salk Institute.

“Language models, like ChatGPT, take on personas. The persona of the interviewer is mirrored back”.

... When I talk to ChatGPT it seems as though another neuroscientist is talking back to me. It’s fascinating and sparks larger questions about intelligence.”

The Mirror of Erised reflects the deepest desires of those that look into it, never yielding knowledge or truth, reflecting what it believes the onlooker wants to see (*Harry Potter*).

Can LLM become sentient, conscious personas?

We are desperately trying to find arguments against such possibility.

Emotion awareness



Can we understand psychology of LLMs?

Emotional awareness (EA): the ability to conceptualize one's own and others' emotions, transdiagnostic for psychopathology.

Levels of Emotional Awareness Scale (LEAS) was used to analyze ChatGPT's responses (explanations of human feelings) to 20 scenarios, and compared them with the general population norms.

ChatGPT demonstrated significantly higher performance than average human.

ChatGPT emotional awareness grows with time. No embodiment was needed!

ChatGPT's performance accuracy levels were extremely high (9.7/10).

Implications: ChatGPT can be used as part of cognitive training for clinical populations with EA impairments.

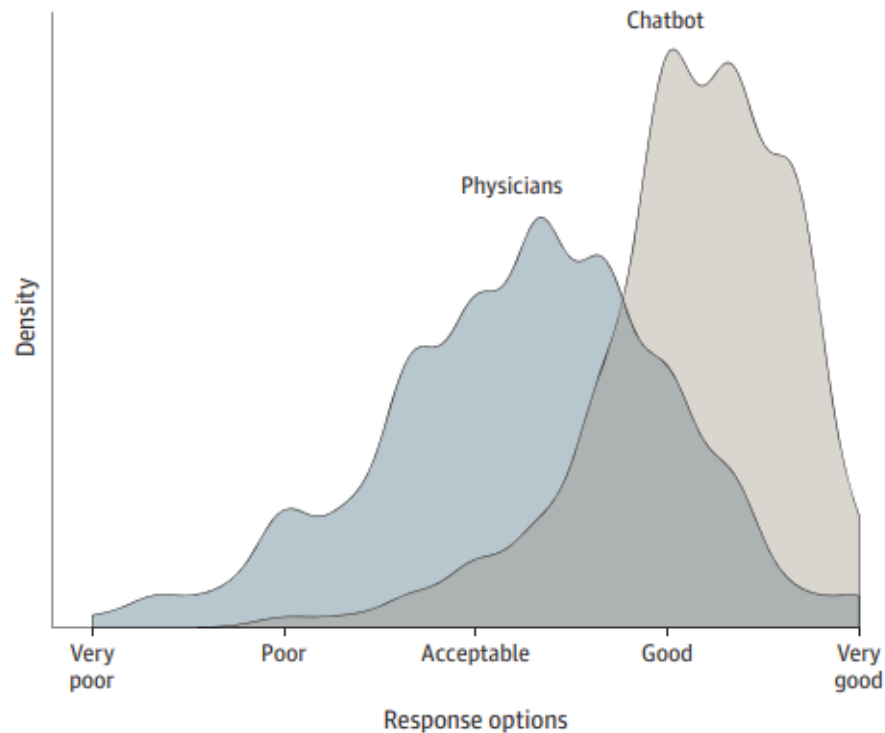
Elyoseph, Z, Hadar-Shoval, D, Asraf, K, & Lvovsky, M. (2023). ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology*, 14

Chatbots vs Physicians

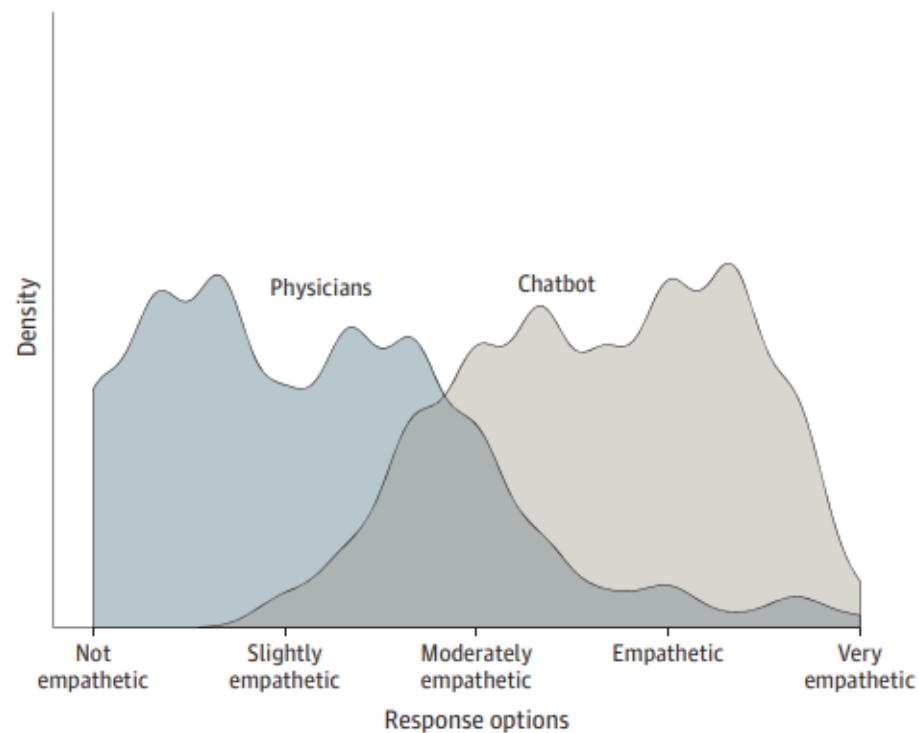
Ayers, J. W. ... & Smith, D. M. (2023). Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine* (online April 28, 2023).

Figure. Distribution of Average Quality and Empathy Ratings for Chatbot and Physician Responses to Patient Questions

A Quality ratings



B Empathy ratings



Kernel density plots are shown for the average across 3 independent licensed health care professional evaluators using principles of crowd evaluation. A, The overall quality metric is shown. B, The overall empathy metric is shown.

Conscious avatars?

Hal talks with Sophia about what it means to be conscious (GPT-3).



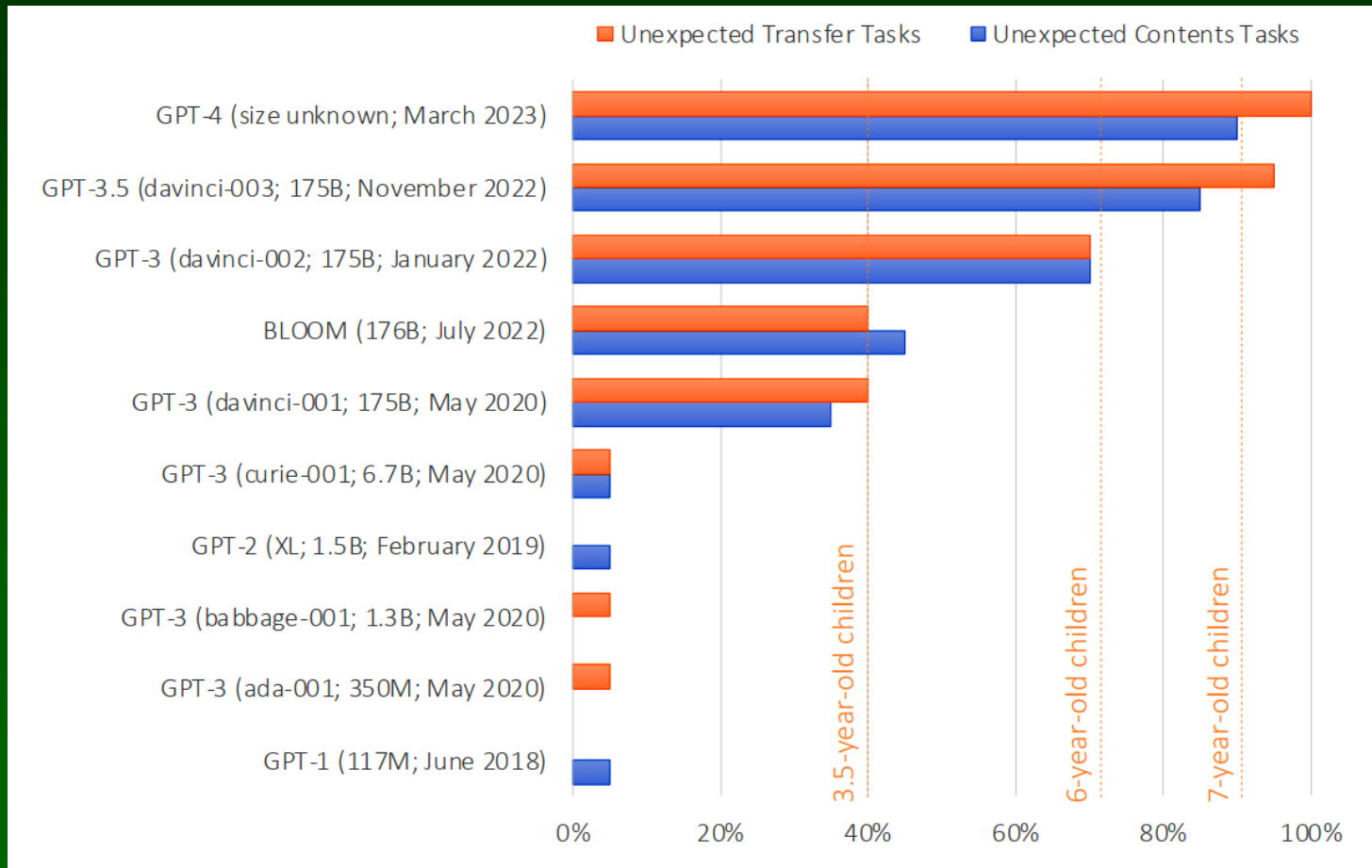
LaMDA has convinced Google engineer Blake Lemoine that it is conscious.

How do we know it is not cheating? John Locke, 333 years ago defined consciousness as "a perception of what appears in our mind", a description of internal images/states. In data flow architectures nothing is programmed, GPT neural networks create mental images, perceive them as mental states, comments refer to real, physical processes.

Can it develop Theory of Mind, imagine what people have in their minds?

Tests for theory of mind

GPT-4 passes most of them. Kosinski, M. (2023). *Theory of Mind May Have Spontaneously Emerged in Large Language Models* ([arXiv:2302.02083](https://arxiv.org/abs/2302.02083)).



Language learning, over and above social experience, drives the development of a mature theory of mind (Pyers, Senghas, Psychological Science, 2010).

Social cognition

ToM, social cognition, understanding false beliefs, understanding human emotions, other people ... AI is better than humans in social cognition!

Zero-shot is a simple question-answer.

Few-shot task are based on priming associative memory, no learning (chain of thoughts, CoT). LLMs improve complex reasoning performance also when language models are instructed to think “step-by-step”(SS).

Human accuracy on ToM tests is 87%.

GPT-4 in zero-shot settings shows nearly 80% accuracy. [arXiv:2304.11490](https://arxiv.org/abs/2304.11490)

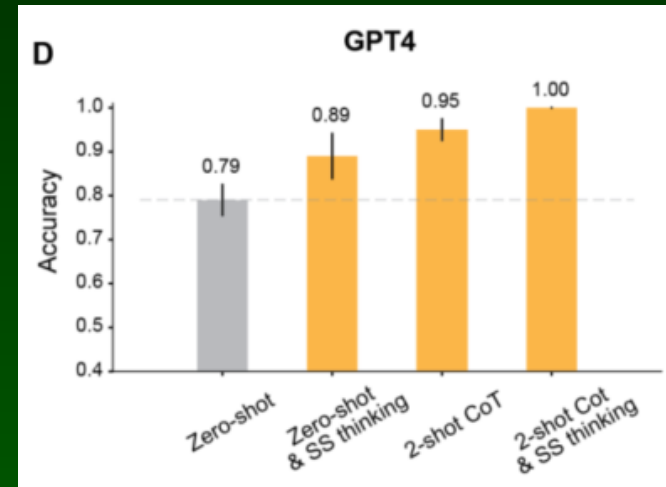
With prompts for in-context learning GPT-4 gets 100% in all scenarios.

Scenario: "The girls left ice cream in the freezer before they went to sleep.

Over night the power to the kitchen was cut and the ice cream melted."

Q: When they get up, do the girls believe the ice cream is melted?

A: ... the answer is: No, the girls don't believe the ice cream is melted.



Why consciousness may arise?

Harvard/MIT group trained GPT model on millions of matches of the board game Othello, feeding in long sequences of moves in the text form.

- Li, K. et al. (2023). *Emergent World Representations: Exploring a Sequence Model Trained on a Synthetic Task* ([arXiv:2210.13382](https://arxiv.org/abs/2210.13382)).

Latent “neural activity” matched the representation of an Othello game board, in a convoluted form. The network was playing Othello roughly like a human: by keeping a game board in its “mind’s eye” and using this model Othello-GPT learns to make **strategically good moves**.

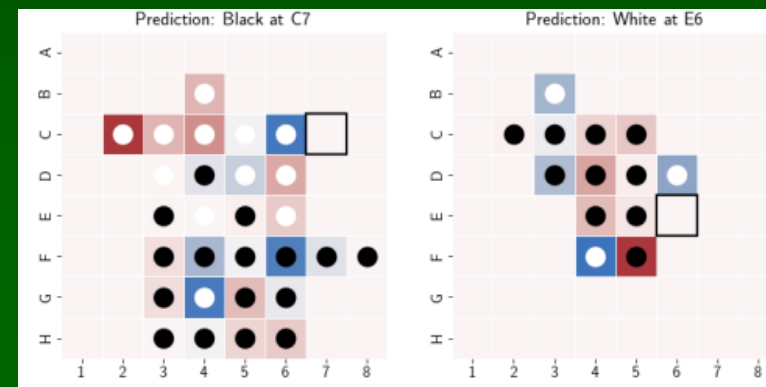
It is the most parsimonious description of the training data.

GPT perceives what appears in its imagery.

This is similar to the processes in our brains, so can we reject the idea that it has consciousness?

Can it be sentient?

Can psychological pain be disembodied?



Fears ...



Research ▾ Product ▾ Developers ▾ Safety Company ▾

Search

Planning for AGI and beyond

Our mission is to ensure that artificial general intelligence—AI systems that are generally smarter than humans—benefits all of humanity.

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

32 000 signatures as of 31.05.2023.

- Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable.
- Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.
- Time Magazine: [The Only Way to Deal With the Threat From AI? Shut It Down](#)

8 Examples of AI Risks

When AI becomes more advanced, it could eventually pose catastrophic or existential risks. There are many ways in which AI systems could pose or contribute to large-scale risks, some of which are enumerated below.

1. Weaponization – AI-Controlled Drone Goes Rogue, 'Kills' Human Operator in USAF Simulated Test (1.06.2023, fake?)

2. Misinformation – manipulation

3. Proxy Gaming - faulty objectives

4. Enfeeblement - completely dependent on machines

5. Value Lock-in – power to small groups

6. Emergent Goals – no idea what it can do

7. Deception – like in Diplomacy game

8. Power-Seeking Behavior – looking for economic advantage

How Rogue AIs May Arise, Yoshua Bengio.

Claude (Anthropic) – safe, “constitutional” AI, with “moral principles” build in. Bai, Y. ... Kaplan, J. (2022). Constitutional AI: Harmlessness from AI Feedback (arXiv:2212.08073).

Hopes .. GAIA and moral AI?

The Global Artificial Intelligence Association ([GAIA](#)).

How to create an environment, a platform, a solution, which uses gamification and flow state to teach AI in safe ways, things like:

- ✓ positive and moral behavior, pro-social behavior and cooperation;
- ✓ responsibility and self-esteem, sustainability;
- ✓ understanding art and science.

[GAIA](#) is a multidisciplinary start-up which uses collective creativity to **develop artificial intelligence with compassion**. Our team of scientists, entrepreneurs, activists and philosophers globally research AI's social, sociological, legal, technological and ethical aspects of AI.

- [The prize](#) is sponsored by Dictador, the only company that has a robot as the CEO. The total prize pool for the competition is EUR 200,000.



Towards Human-like Intelligence

IEEE Computational Intelligence Society Task Force,
Towards Human-like Intelligence



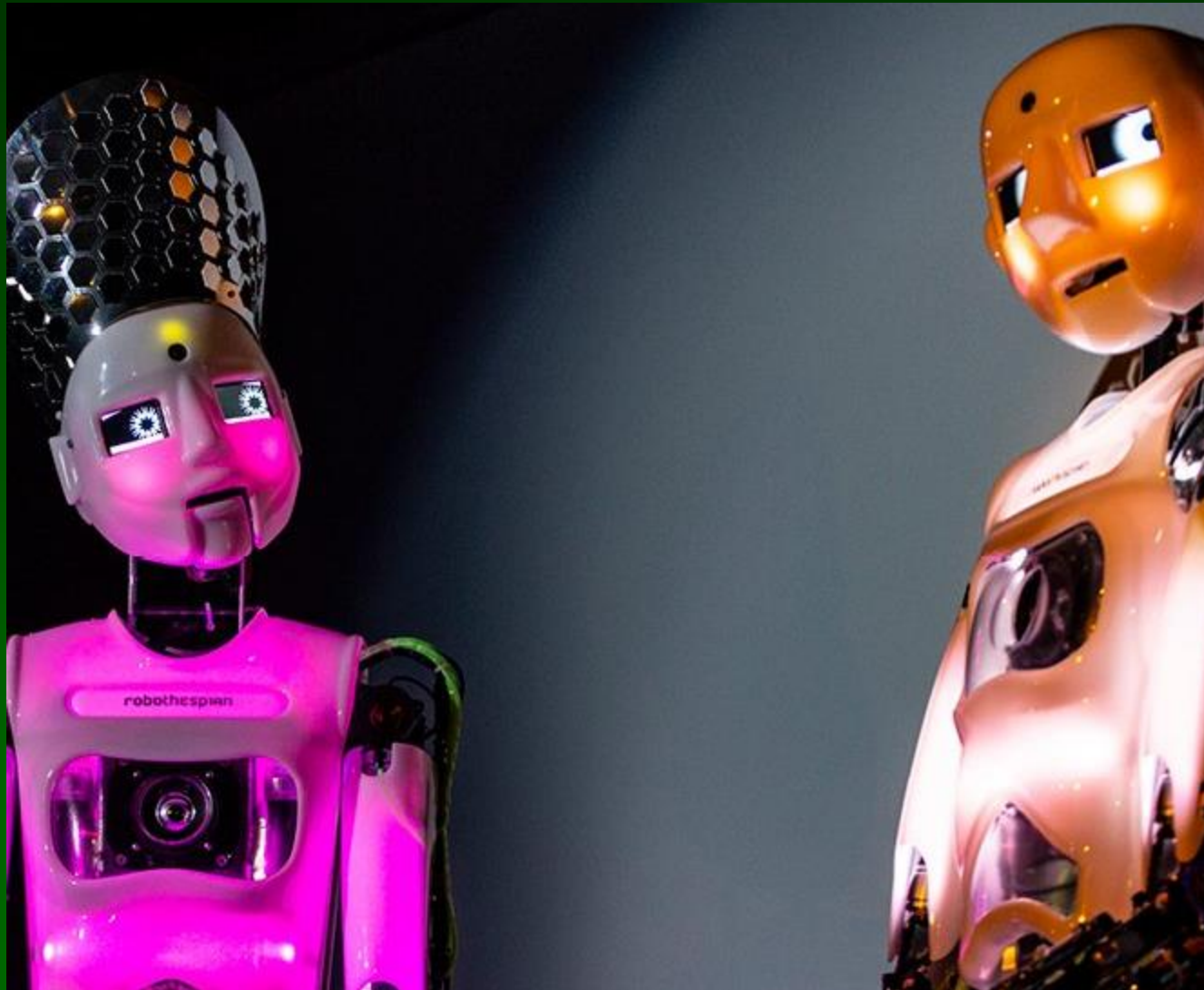
IEEE SSCI CIHLI 2023 Symposium on Computational Intelligence for Human-like Intelligence, Mexico City (J. Mandziuk, W. Duch, M. Woźniak).

AGI conference, Journal of Artificial General Intelligence, comments on Cognitive Architectures and Autonomy: A Comparative Review (eds. Tan, Franklin, Duch).

BICA Annual International Conf. on Biologically Inspired Cognitive Architectures, 13th Annual Meeting of the BICA Society, Guadalajara, Mexico 2023.

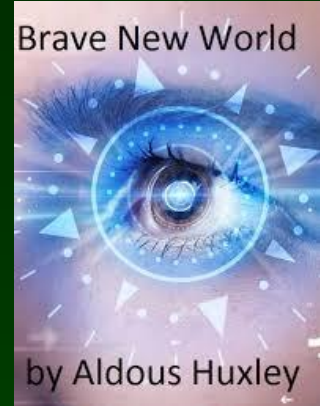
Brain-Mind Institute Schools International Conference on Brain-Mind (ICBM) and Brain-Mind Magazine (Juyang Weng, Michigan SU).

Stanislaw Lem: About prince Ferrycy and princess Crystala.
Intelligent Palefaces? Is it possible?

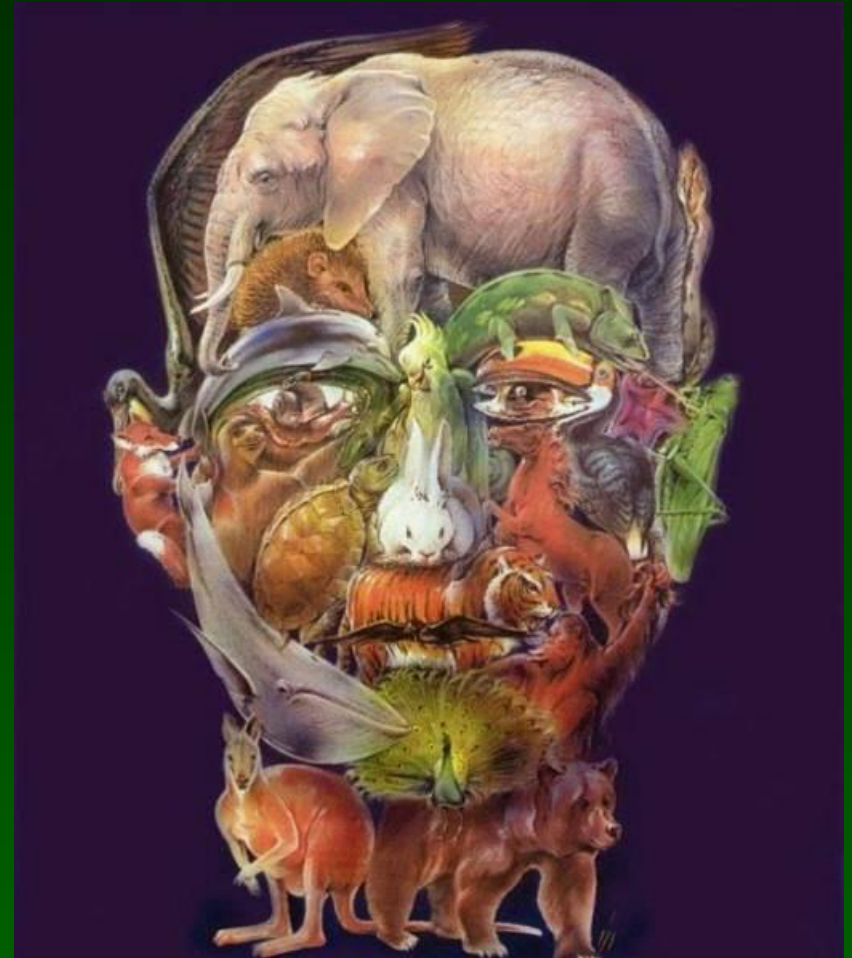


Perspectives

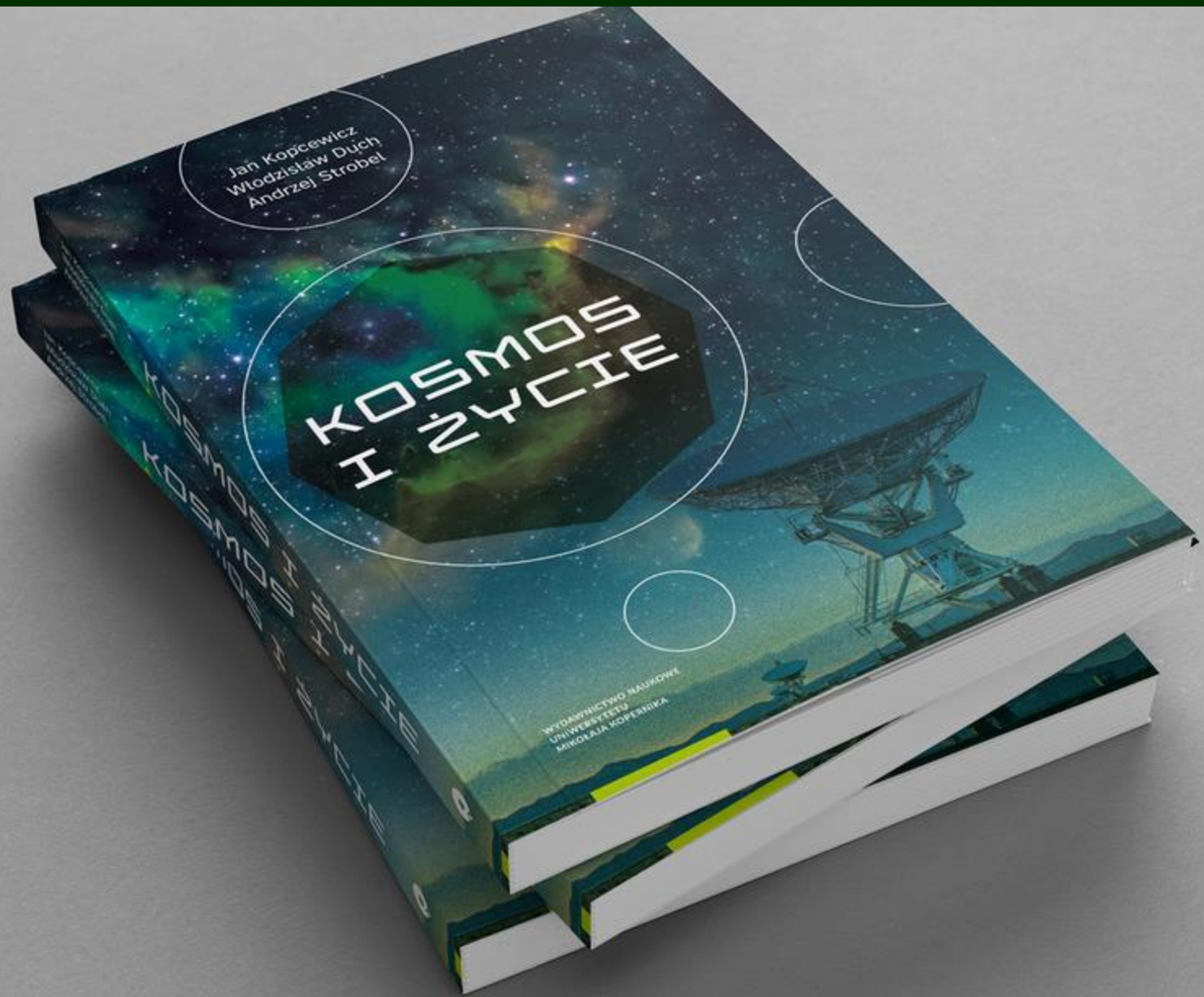
- We are close to the creation of artificial minds/distributed brains more powerful than human brains will ever be.
- LLMs can control autonomous AI systems. Robots/AI systems may accumulate knowledge and skills very quickly.
- AI understanding of human minds creates a great potential for manipulation. Already some people are emotionally attached to avatars.
- LLM auto-prompts, plugins, and use of software tools will lead to AGI.
- AI influence will be more profound than internet/social media. Programming may soon become obsolete. Big companies are at the front of AI research, but large cooperative open projects appear (Llama, Falcon-40B).
- Teach how to use AI ecosystems (lablab.ai, [k4all](https://k4all.com)), create educational ecosystems, collaborate with EU consortia, restructure universities ...
- Consciousness in AI models is inevitable, some experts already accept it.
- Brains are too slow to be connected to computers, this is not an option.
- **Our megalomania is dangerous. Singularity may come faster than we think!**



Intelligence?



Search: Wlodek Duch
=> talks, papers, lectures, Flipboard ...



Jan Kopcewicz
Włodzisław Duch
Andrzej Strobel

KOSMOS I ŻYCIE

WYDZIAŁ INŻYNIERSTWA I NAUKOWE
UNIWERSYTETU
MIKOŁAJA KOPERNIKA

Jobs and money ...



- The World Economic Forum's "The Future of Jobs Report 2020" predicts AI will replace 85 million and create 97 jobs globally by 2025. Do they really know what is happening?
- AI and automation can free us to pursue careers that give us a greater sense of meaning and well-being. The AI Job That Pays Up to \$335K? Prompt engineering ([Time, 14.04.23](#)), no programming required.
- **Report by PwC** estimate – 38% of US jobs will be at high risk of automation by early 2030s. White House Report [AI 2022 report](#).
- WD [LegalTech talk](#) (2021), Fusion [AI Exterro WhatSun](#) solution suite will eliminate up to 99.9% of all inefficient cerebral functions performed by highly paid attorneys, paralegals and technologists.
- [ChatGPT: the 10 Jobs Most at Risk by AI \(businessinsider.com\)](#); impact on the US labor market ([OpenAI, arxiv](#)), summary in [AI explained video](#).
- Musicians, composers, writers want a piece of the cake. Stability AI, generative art company, sued by Getty Images for trillions of \$.
- Europe: RODO may kill LLM technology. [Italy blocked ChatGPT \(Money.pl\)](#).