

# Introduction to GenAI

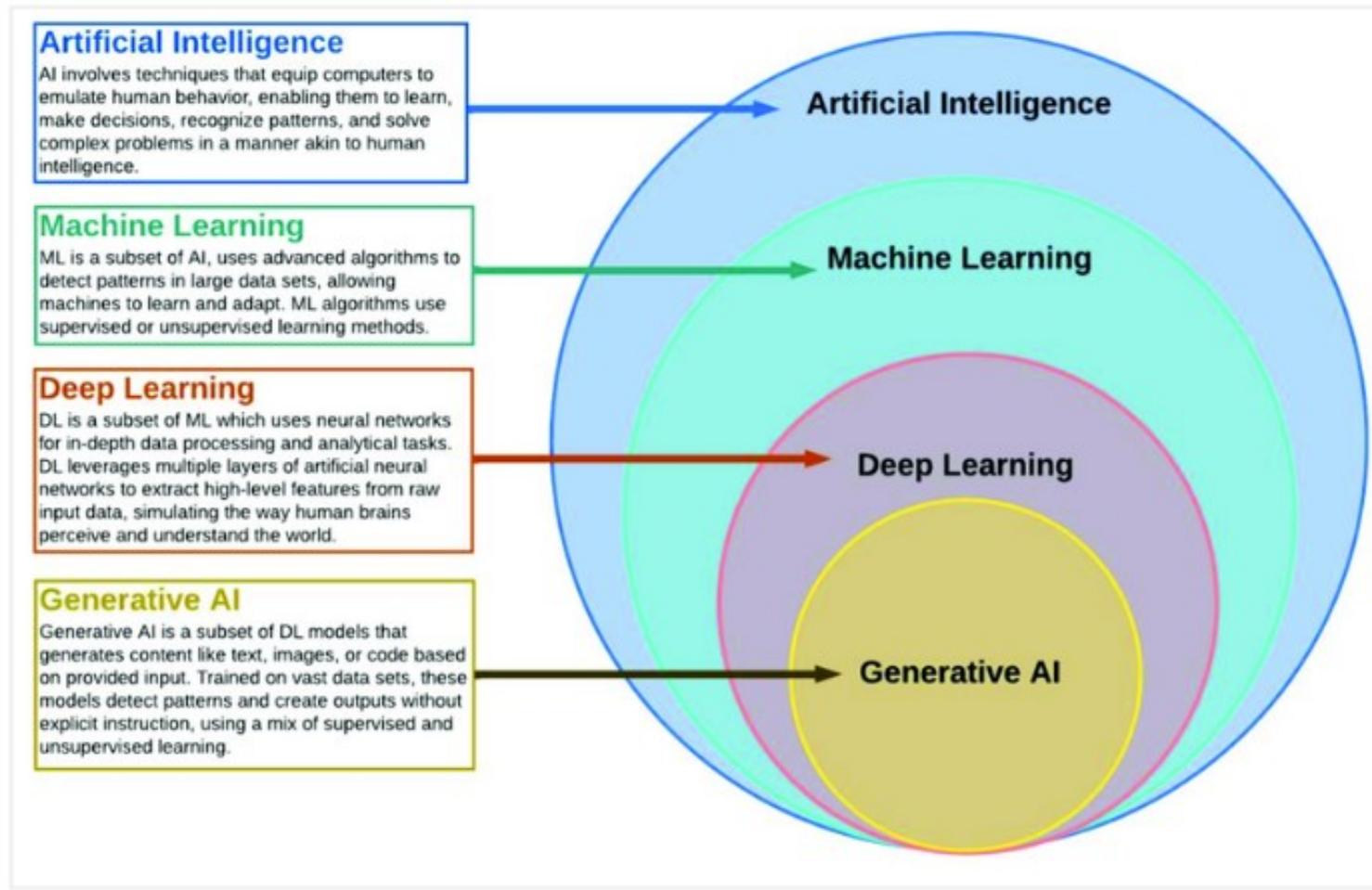
Past > Present > Future

2024-12-01

# Past - The "storm gathering"

- What is GenAI
- A brief history
- How it works
- Why it works

# AI > ML > DL > GenAI



# AI History

1984 - Apple Mac  
 1989 - WWW by Tim Berners-Lee @ CERN

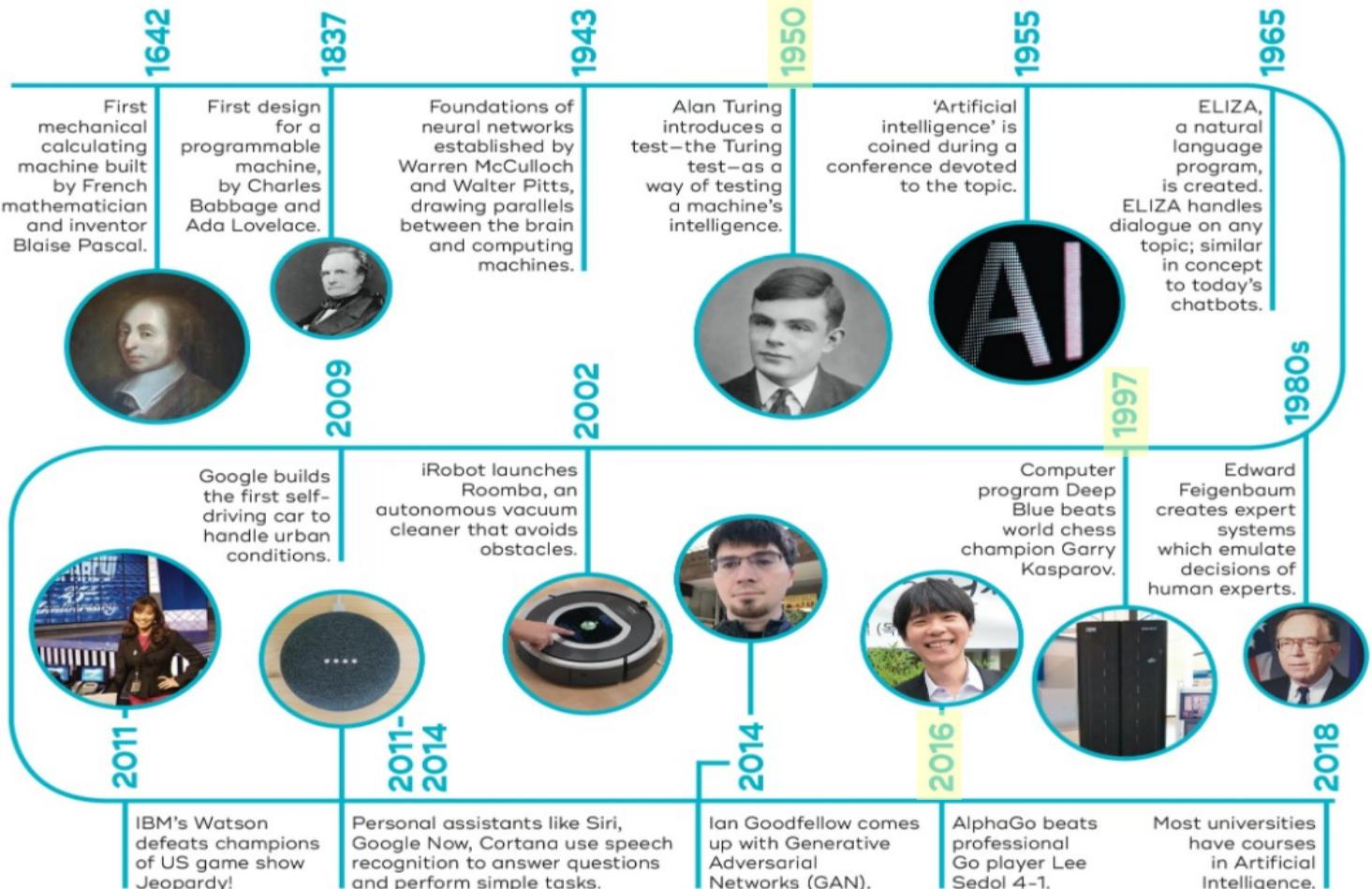
1990 - Microsoft Windows  
 1991 - Linux  
 1998 - Google Search

2006 - AWS Cloud  
 2007 - Apple iPhone

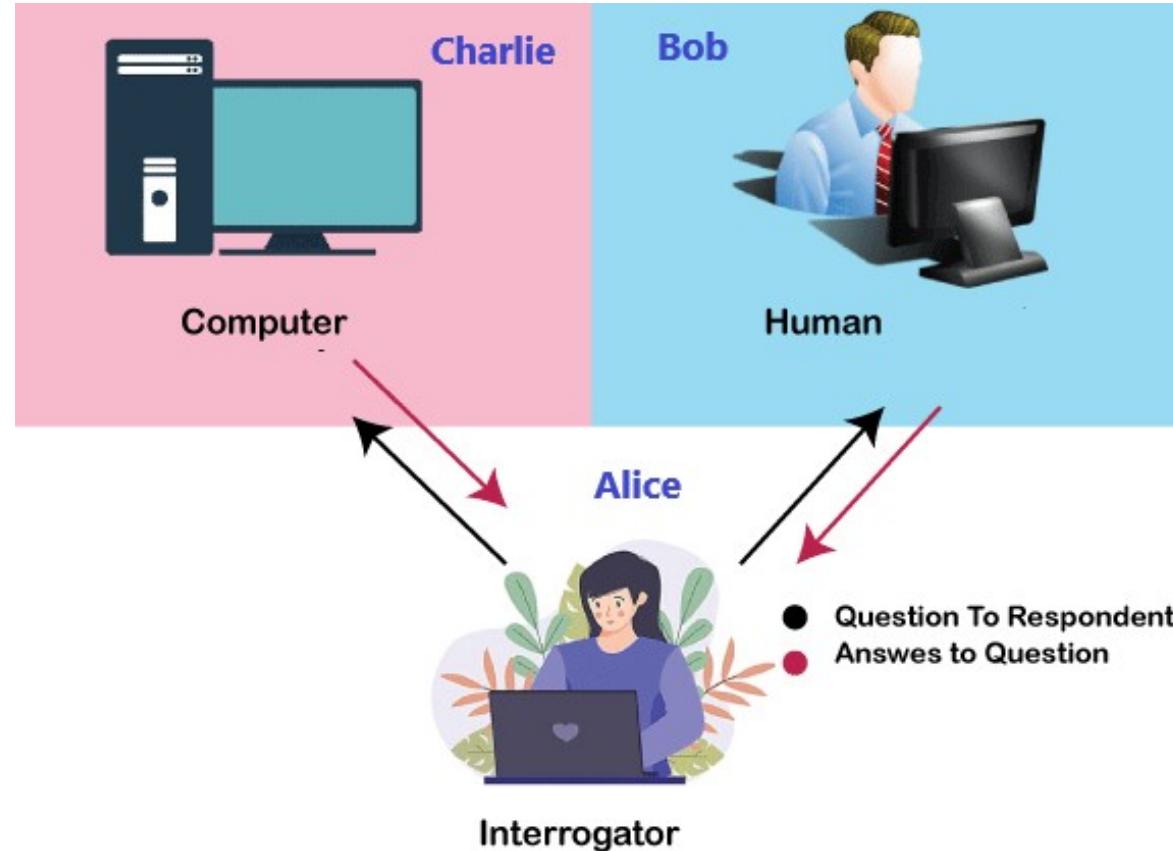
2015 - OpenAI  
 2018 - Hinton, LeCunn, Bengio (Turing Award)

2022-11-30 - ChatGPT

Past

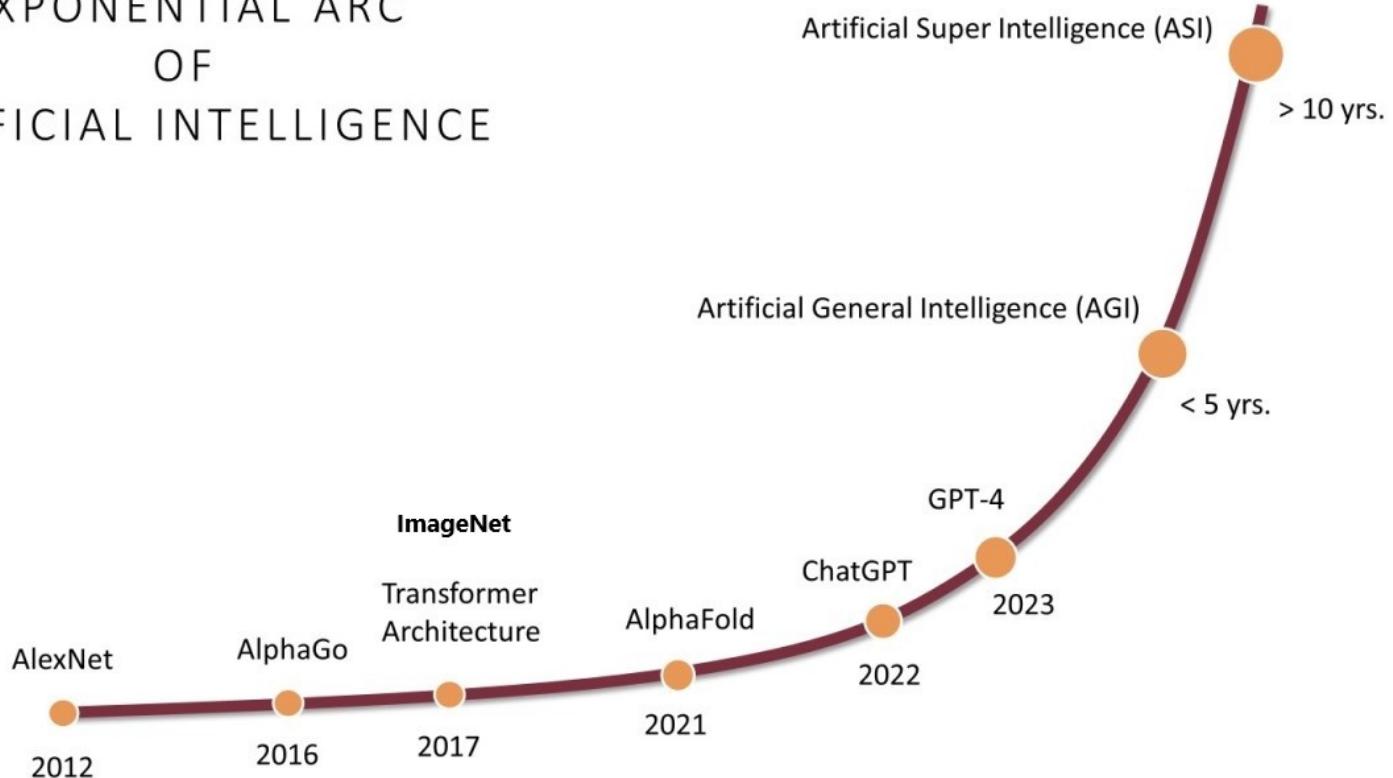


# What is Turing Test ?



# AI Growth & Milestones

EXPONENTIAL ARC  
OF  
ARTIFICIAL INTELLIGENCE



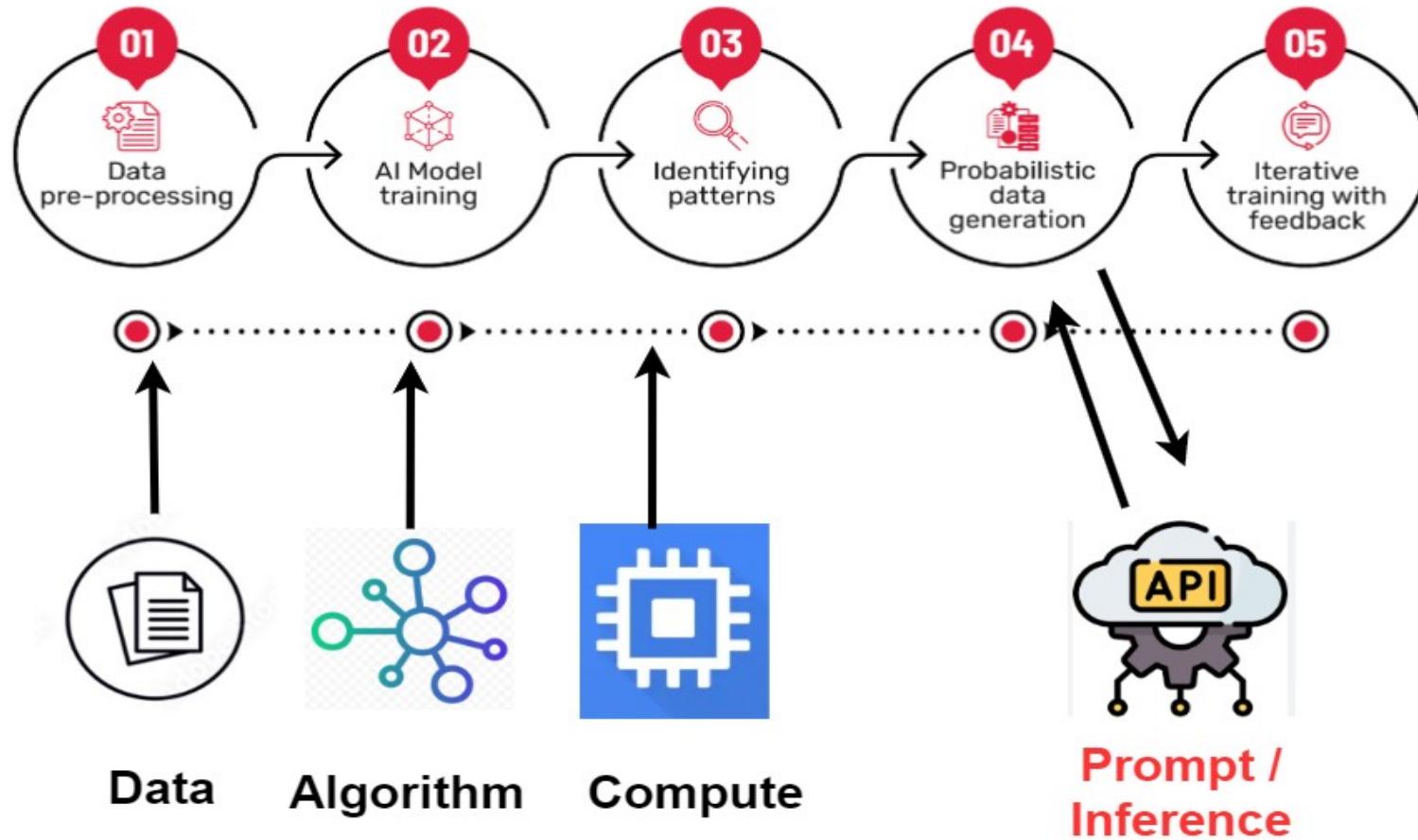
# OpenAI: ChatGPT



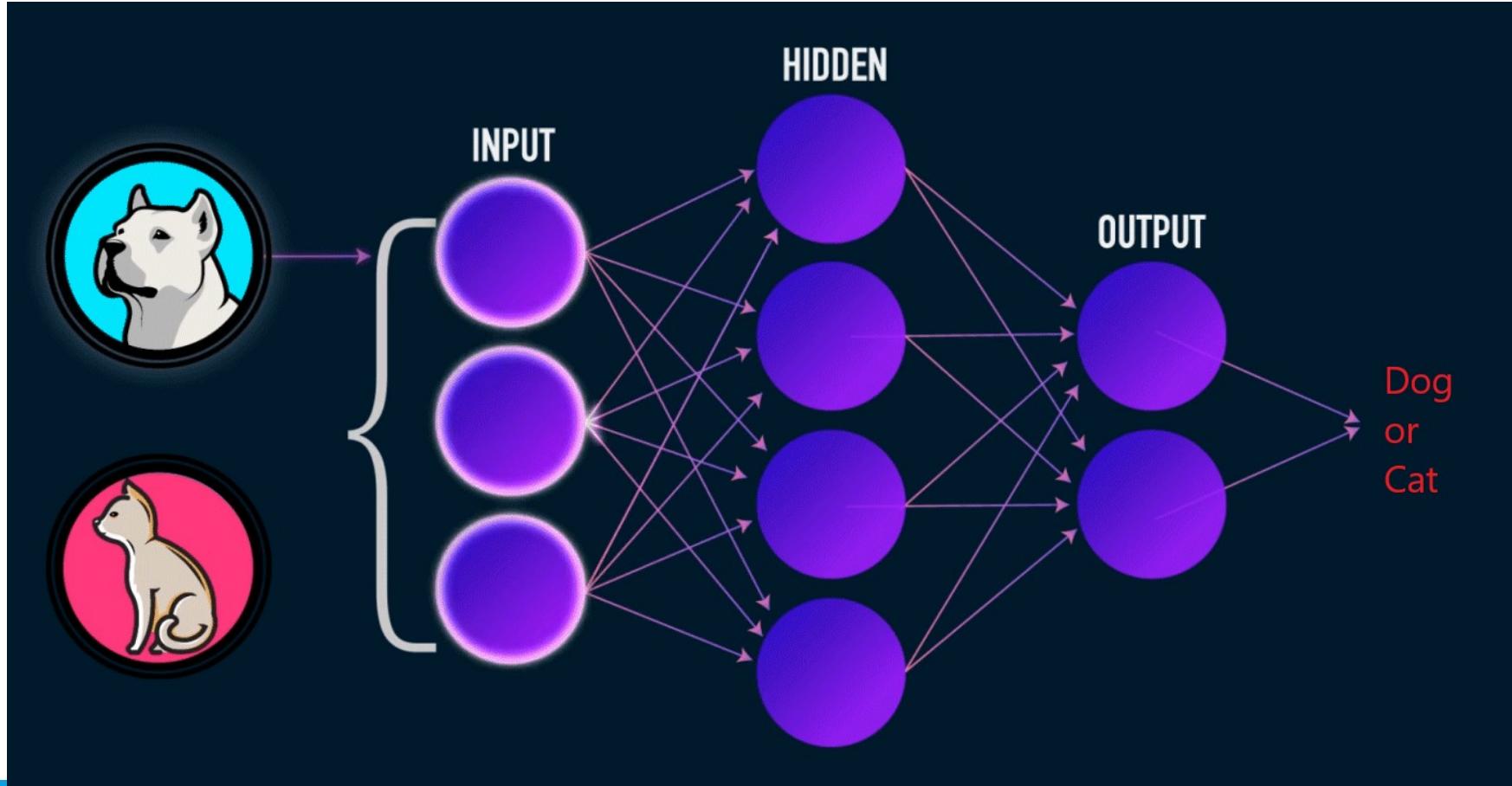
# How LLM Model is built



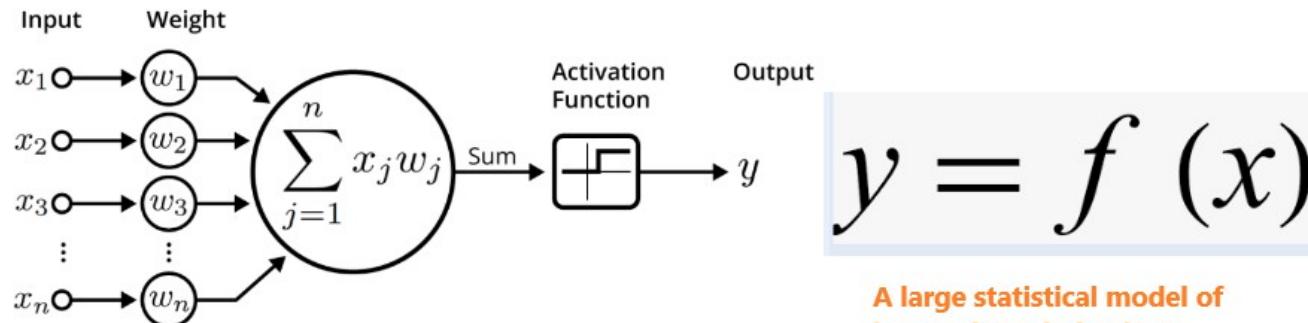
# How GenAI works?



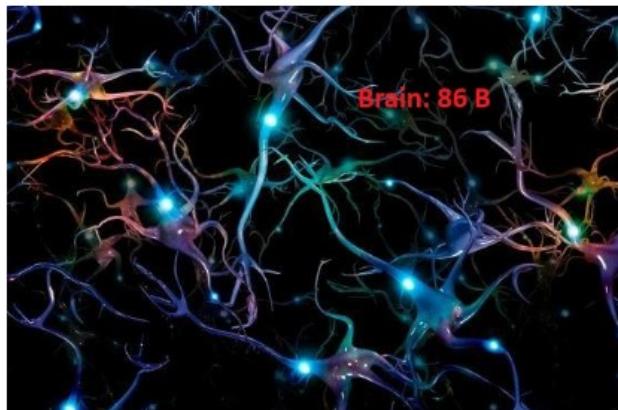
# Training Neural Network



# LLM: $y = f(x)$



An illustration of an artificial neuron. Source: Becoming Human.



## Parameters in Selected AI Models

Some of these figures are estimates. Newer models are many times larger than their predecessors.



# LLM Explainability

## - Generalization Theory

Explain why deep networks can learn effectively despite having more parameters than training examples;

## - Representation Learning

Study how neural networks develop hierarchical internal representations similar to biological brains;

## - Optimization Dynamics

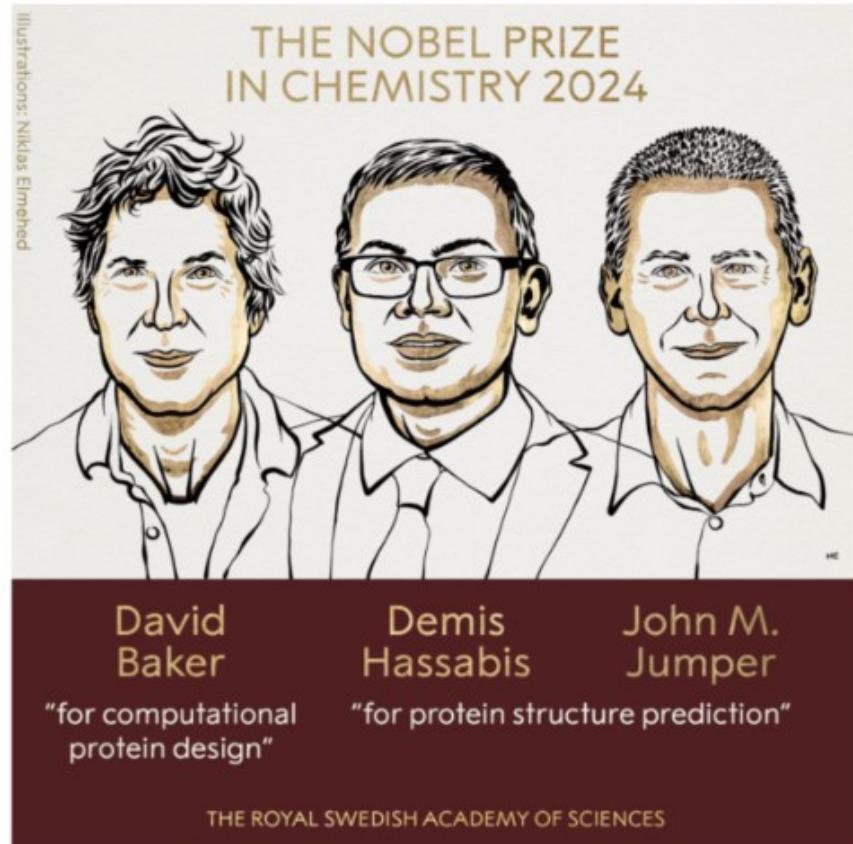
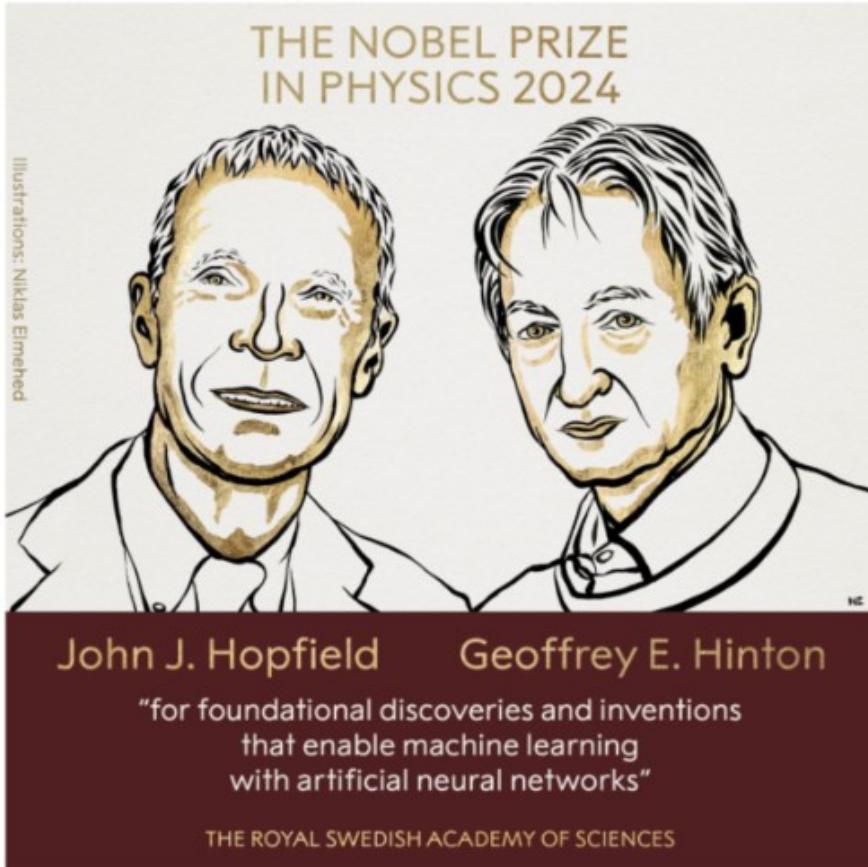
Investigate why gradient descent consistently finds good solutions in highly complex parameter spaces

Tomaso Poggio <https://mcgovern.mit.edu/profile/tomaso-poggio/>

# Present - "Following the yellow brick road"

- State of 2024
- LLM model
  - type
  - benchmark
- Application
  - chatbot
  - agent
  - demo
- Limitation

# 2024 Nobel Prizes in Physics & Chemistry



# State of AI - 2024

## Research

- Frontier lab performance converges, but OpenAI maintains its edge following the launch of o1, as planning and reasoning emerge as a major frontier.
- Foundation models demonstrate their ability to break out of language as multimodal research drives into mathematics, biology, genomics, the physical sciences, and neuroscience.
- US sanctions fail to stop Chinese (V)LLMs rising up community leaderboards.

## Industry

- NVIDIA remains the most powerful company in the world, enjoying a stint in the \$3T club, while regulators probe the concentrations of power within GenAI.
- More established GenAI companies bring in billions of dollars in revenue, while start-ups begin to gain traction in sectors like video and audio generation. Although companies begin to make the journey from model to product, long-term questions around pricing and sustainability remain unresolved.
- Driven by a bull run in public markets, AI companies reach \$9T in value, while investment levels grow healthily in private companies.

## Politics

- While global governance efforts stall, national and regional AI regulation has continued to advance, with controversial legislation passing in the US and EU.
- The reality of compute requirements forces Big Tech companies to reckon with real-world physical constraints on scaling and their own emissions targets. Meanwhile, governments' own attempts to build capacity continue to lag.
- Anticipated AI effects on elections, employment and a range of other sensitive areas are yet to be realized at any scale.

## Safety

- A vibe-shift from safety to acceleration takes place as companies that previously warned us about the pending extinction of humanity need to ramp up enterprise sales and usage of their consumer apps.
- Governments around the world emulate the UK in building up state capacity around AI safety, launching institutes and studying critical national infrastructure for potential vulnerabilities.
- Every proposed jailbreaking 'fix' has failed, but researchers are increasingly concerned with more sophisticated, long-term attacks.

# LLM Model Type

## Model types:

-  →  : LLMs
-  +  →  : Multimodal LLMs
-  +  +  →  : Multimodal LLMs for Robotics
-  → </> : Text to Code
-  →  : Text to Software tool use
-  →  : Text to Image
-  →  : Text to Video
-  →  : Text to Music
-  →  : Image to 3D
-  →  : Text to 3D
-  →  : Biological models

## Input/Output types:

-  : Text
-  : Image
- </> : Code
-  : Software tool use
-  : Video
-  : Music
-  : 3D
-  : Robot state
-  : Biological modality

# GenAI Leading Players

— Large Language Models are becoming very large indeed



## Small models (<= 100b parameters)



## Large models (>100b parameters)

Scaling laws: 1) bigger models > better;  
2) more training data > better;  
3) more compute resources > better



LaMDA  
137B  
Google

GPT-3  
175B  
OpenAI

Jurassic-1  
178B  
AI21labs

Gopher  
280B  
DeepMind

MT-NLG  
530B  
NVIDIA

PaLM  
540B  
Google

PaLM-E  
562B  
Google

GPT-4  
1700 B  
OpenAI

# HuggingFace – Open Source LLM Hub



And Chinese open source projects win fans around the world

► To drive international uptake and evaluation, Chinese labs have become enthusiastic open source contributors.  
A few models have emerged as strong contenders in individual sub-domains.

- DeepSeek has emerged as a community favorite on coding tasks, with deepseek-coder-v2 for its combination of speed, lightness, and accuracy.
- Alibaba released the Qwen-2 family recently, and the community has been particularly impressed by its vision capabilities, ranging from challenging OCR tasks to its ability to analyse complex art work.
- At the smaller end, the NLP lab at Tsinghua University has funded OpenBMB, a project that has spawned the MiniCPM project.
- These are small <2.5B parameter models that can run on-device. Their 2.8B vision model is only marginally behind GPT-4V on some metrics, while 8.5B Llama 3 based model surpasses it on some metrics.
- Tsinghua University's Knowledge Engineering Group has also created CogVideoX, one of the most capable text to video models.

When tasked to identify the most famous artworks of artists with a high notoriety, Qwen2-VL-2B was extremely successful at instantly recognizing them. It completely identified Vincent Van Gogh's *The Starry Night* (1889, MoMA) and Monet's *Impression, Sunrise* (1872, Musée Marmottan) without any instructions. The model managed to identify both the paintings and the painters names, which might have been thanks to the signature or by the fact these are two commonly well-known art pieces, but the descriptions in the results were still impressive. Both were accurate, concise and well-written, in the usual style for an art piece description, and even categorized the artworks in their artists' career.



For the Monet's masterpiece, the model even correctly managed to identify the movement to which it belonged, without any instructions: "The painting is characterized by its loose, impressionistic style, which captures the fleeting effects of light and color in nature. The use of bright, contrasting colors and the use of brushstrokes to create a sense of movement and energy are prominent features of Monet's painting...it is considered one of Monet's most iconic works".

# Ollama – Open Local Private LLM platform

## Open

- Run SOTA open-source LLMs locally
- Easy model management (pull, run)
- Active community and growing model library

## Local

- Simple setup
- Run on PC/Laptop
- No cloud dependencies or API costs

## Private

- Data stays on your machine
- Full control over model interactions
- Ideal for research experimentation

# Model Leader Board

| Model                  | Average | Multi-choice Qs | Reasoning | Python coding | Future Capabilities | Grade school math | Math Problems |
|------------------------|---------|-----------------|-----------|---------------|---------------------|-------------------|---------------|
| Claude 3.5 Sonnet      | 88.38%  | 88.70%          | 89.00%    | 92.00%        | 93.10%              | 96.40%            | 71.10%        |
| Claude 3 Opus          | 84.83%  | 86.80%          | 95.40%    | 84.90%        | 86.80%              | 95.00%            | 60.10%        |
| Gemini 1.5 Pro         | 80.08%  | 81.90%          | 92.50%    | 71.90%        | 84%                 | 91.70%            | 58.50%        |
| Gemini Ultra           | 79.52%  | 83.70%          | 87.80%    | 74.40%        | 83.60%              | 94.40%            | 53.20%        |
| GPT-4                  | 79.45%  | 86.40%          | 95.30%    | 67%           | 83.10%              | 92%               | 52.90%        |
| Llama 3 Instruct - 70B | 79.23%  | 82%             | 87%       | 81.7%         | 81.3%               | 93%               | 50.4%         |
| Claude 3 Haiku         | 73.08%  | 75.20%          | 85.90%    | 75.90%        | 73.70%              | 88.90%            | 38.90%        |
| GPT-3.5                | 65.46%  | 70%             | 85.50%    | 48.10%        | 66.60%              | 57.10%            | 34.1%         |
| Mixtral 8×7B           | 59.79%  | 70.60%          | 84.40%    | 40.20%        | 60.76%              | 74.40%            | 28.40%        |
| GPT-4o                 | -       | 88.7%           | -         | 90.2%         | -                   | -                 | 76.60%        |
| GPT-4o mini            | -       | 82%             | -         | 87.00%        | -                   | -                 | 70.20%        |
| Llama 3 Instruct - 8B  | -       | 68.40%          | -         | 62%           | 61%                 | 79.60%            | 30%           |
| Grok 1.5               | -       | 73.00%          | -         | 63%           | -                   | 62.90%            | 23.90%        |
| Mistral Large          | -       | 81.2%           | 89.2%     | 45.1%         | -                   | 81%               | 45%           |
| Gemini 1.5 Flash       | -       | 78.90%          | -         | -             | 89.20%              | -                 | 67.70%        |
| GPT-4T 2024-04-09      | -       | 86.5%           | -         | -             | 87.60%              | -                 | 72.2%         |
| OpenAI o1              | -       | 92.30%          | -         | 92.40%        | -                   | -                 | 94.80%        |
| OpenAI o1-mini         | -       | 85.20%          | -         | 92.40%        | -                   | -                 | 90.00%        |

# Model Benchmark

| Open source          | Chinese General | English General | Knowledge   | Arithmetic  | Math        | Reasoning   | Coding      |             |
|----------------------|-----------------|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                      | AlignBench      | MT-Bench        | MMLU        | GSM8K       | MATH        | BBH         | HumanEval   |             |
| <b>DeepSeek-V2.5</b> | Yes             | <b>8.04</b>     | <b>9.02</b> | <b>80.4</b> | <b>95.1</b> | <b>74.7</b> | <b>84.3</b> | <b>89.0</b> |
| DeepSeek-V2          | Yes             | 7.89            | 8.85        | 80.6        | 94.8        | 71.0        | 83.4        | 84.8        |
| GPT-4-Turbo-1106     | -               | <b>8.01</b>     | <b>9.32</b> | 84.6        | 93.0        | 64.1        | -           | 82.2        |
| GPT-4-0613           | -               | 7.53            | 8.96        | <b>86.4</b> | 92.0        | 52.9        | 83.1        | 84.1        |
| GPT-3.5              | -               | 6.08            | 8.21        | 70.0        | 57.1        | 34.1        | 66.6        | 48.1        |
| Gemini1.5 Pro        | -               | 7.33            | 8.93        | 81.9        | 91.7        | 58.5        | 84.0        | 71.9        |
| Claude3 Opus         | -               | 7.62            | <b>9.00</b> | <b>86.8</b> | <b>95.0</b> | 61.0        | <b>86.8</b> | 84.9        |
| Claude3 Sonnet       | -               | 6.70            | 8.47        | 79.0        | 92.3        | 40.5        | 82.9        | 73.0        |
| Claude3 Haiku        | -               | 6.42            | 8.39        | 75.2        | 88.9        | 40.9        | 73.7        | 75.9        |
| abab-6.5             | -               | 7.97            | 8.82        | 79.5        | 91.7        | 51.4        | 82.0        | 78.0        |
| abab-6.5s            | -               | 7.34            | 8.69        | 74.6        | 87.3        | 42.0        | 76.8        | 68.3        |
| <b>ERNIE-4.0</b>     | -               | 7.89            | 7.69        | -           | 91.3        | 52.2        | -           | 72.0        |
| GLM-4                | -               | 7.88            | 8.60        | 81.5        | 87.6        | 47.9        | 82.3        | 72.0        |
| Moonshot-v1          | -               | 7.22            | 8.59        | -           | 89.5        | 44.2        | -           | 82.9        |
| Baichuan 3           | -               | -               | 8.70        | 81.7        | 88.2        | 49.2        | 84.5        | 70.1        |
| <b>Qwen1.5 72B</b>   | Yes             | 7.19            | 8.61        | 76.2        | 81.9        | 40.6        | 65.9        | 68.9        |
| LLaMA 3 70B          | Yes             | 7.42            | 8.95        | 80.3        | 93.2        | 48.5        | 80.1        | 76.2        |

# Model Benchmark

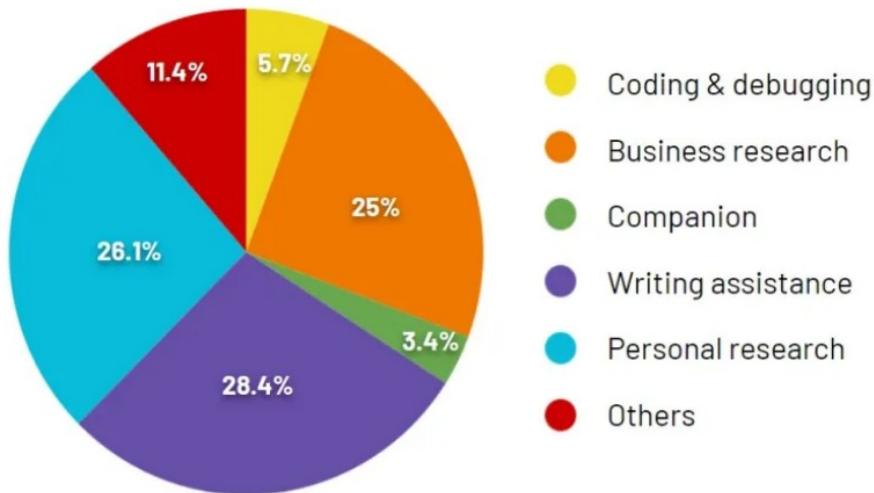
| Model                        | Lab             | Playground  | Parameters |       | Tokens<br>(B) trained (B) | Ratio Tokens:Params<br>(Chinchilla scaling≥20:1) | ALScore<br>"ALScore" is:<br>Sqr Root of | MMLU | MMLU | GPQA |
|------------------------------|-----------------|---|------------|-------|---------------------------|--|---|------|------|------|
|                              |                 |   |            |       |                           |  |   | -Pro | -Pro | -Pro |
| o1                           | OpenAI          | <a href="https://chatgpt.com/">https://chatgpt.com/</a>                           | 200        | 20000 |                           | 100:1  | 6.7                                     | 92.3 | 91   | 78.3 |
| Claude 3.5 Sonnet            | Anthropic       | <a href="https://poe.com/Claude-3.5-Sonnet">https://poe.com/Claude-3.5-Sonnet</a> |            |       |                           |  |   | 88.7 | 76.1 | 67.2 |
| QwQ-32B                      | Alibaba         | <a href="https://huggingface.cc">https://huggingface.cc</a>                       | 32         | 18000 |                           | 563:1  | 2.5                                     |      |      | 65.2 |
| Claude 3.5 Sonnet (inferred) | Anthropic       | <a href="https://claude.ai/">https://claude.ai/</a>                               |            |       |                           |  |   | 90.5 | 78   | 65   |
| Claude 3 Opus                | Anthropic       | <a href="https://claude.ai/">https://claude.ai/</a>                               | 2000       | 40000 |                           | 20:1   | 29.8                                    | 86.8 | 68.5 | 59.5 |
| Gemini-1.5-Pro-002           | Google DeepMind | <a href="https://aistudio.google">https://aistudio.google</a>                     | 1500       | 30000 |                           | 20:1   | 22.4                                    |      | 75.8 | 59.1 |
| DeepSeek-R1-Lite             | DeepSeek-AI     | <a href="https://chat.deepseek">https://chat.deepseek</a>                         | 67         | 2000  |                           | 35:1   | 4.6                                     |      |      | 58.5 |
| Grok-2                       | xAI             | <a href="https://x.com/i/grok">https://x.com/i/grok</a>                           | 600        | 15000 |                           | 25:1   | 10.0                                    | 87.5 | 75.5 | 56   |
| GPT-4o                       | OpenAI          | <a href="https://chatgpt.com/">https://chatgpt.com/</a>                           | 200        | 20000 |                           | 100:1  | 6.7                                     | 88.7 | 72.6 | 53.6 |
| Llama 3.1 405B               | Meta AI         | <a href="https://www.meta.ai/">https://www.meta.ai/</a>                           | 405        | 15600 |                           | 39:1   | 8.4                                     | 88.6 | 73.3 | 51.1 |
| gpt-4-turbo-2024-04          | OpenAI          | <a href="https://chat.openai.com/">https://chat.openai.com/</a>                   |            | 13000 |                           |  |   | 86.5 | 63.7 | 49.1 |
| Qwen2.5                      | Alibaba         | <a href="https://huggingface.cc">https://huggingface.cc</a>                       | 72         | 18000 |                           | 250:1  | 3.8                                     | 86.1 | 71.1 | 49   |
| Yi-XLarge                    | 01-ai           | <a href="https://platform.01.ai">https://platform.01.ai</a>                       | 2000       | 20000 |                           | 10:1   | 21.1                                    | 85.1 |      | 48.2 |
| GPT-4 Turbo                  | OpenAI          | <a href="https://chat.openai.com/">https://chat.openai.com/</a>                   |            | 13000 |                           |  |   | 86.4 |      | 46.5 |
| Gemini 1.5 Pro               | Google DeepMind | <a href="https://aistudio.google">https://aistudio.google</a>                     | 1500       | 30000 |                           | 20:1   | 22.4                                    | 85.9 | 69   | 46.2 |
| gpt-4o-2024-11-20            | OpenAI          | <a href="https://chat.com/">https://chat.com/</a>                                 | 200        | 20000 |                           | 100:1  | 6.7                                     | 85.7 |      | 46   |

# Chat Bot



## — Knowledge workers are becoming dependent on ChatGPT

A poll from Momentum Works community on what they use ChatGPT for



Many people around us (and even us included) have been **using ChatGPT regularly - for a variety of tasks**. See the results of our poll on the left.

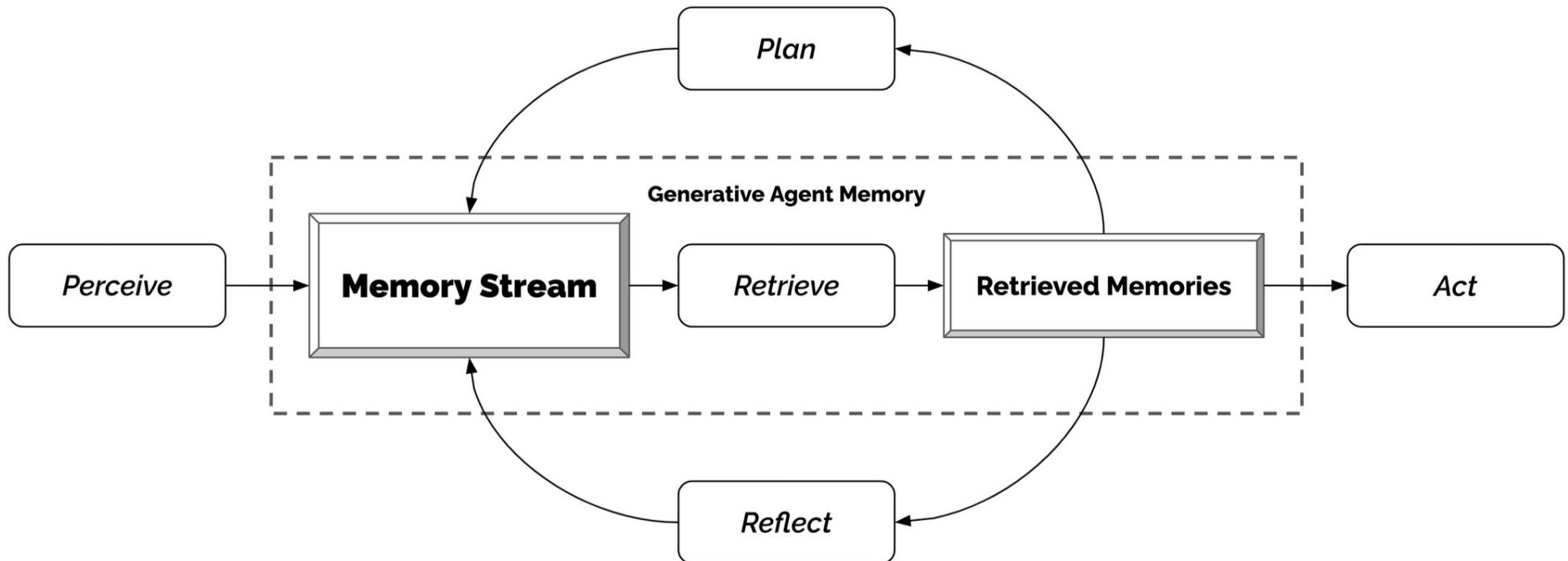
However, the fascinating thing here is - **we do not yet know the full extent of tasks ChatGPT can perform well on**. More capabilities will certainly emerge.

OpenAI has put forward a disclaimer / warning: "ChatGPT sometimes writes **plausible-sounding but incorrect or nonsensical answers**."

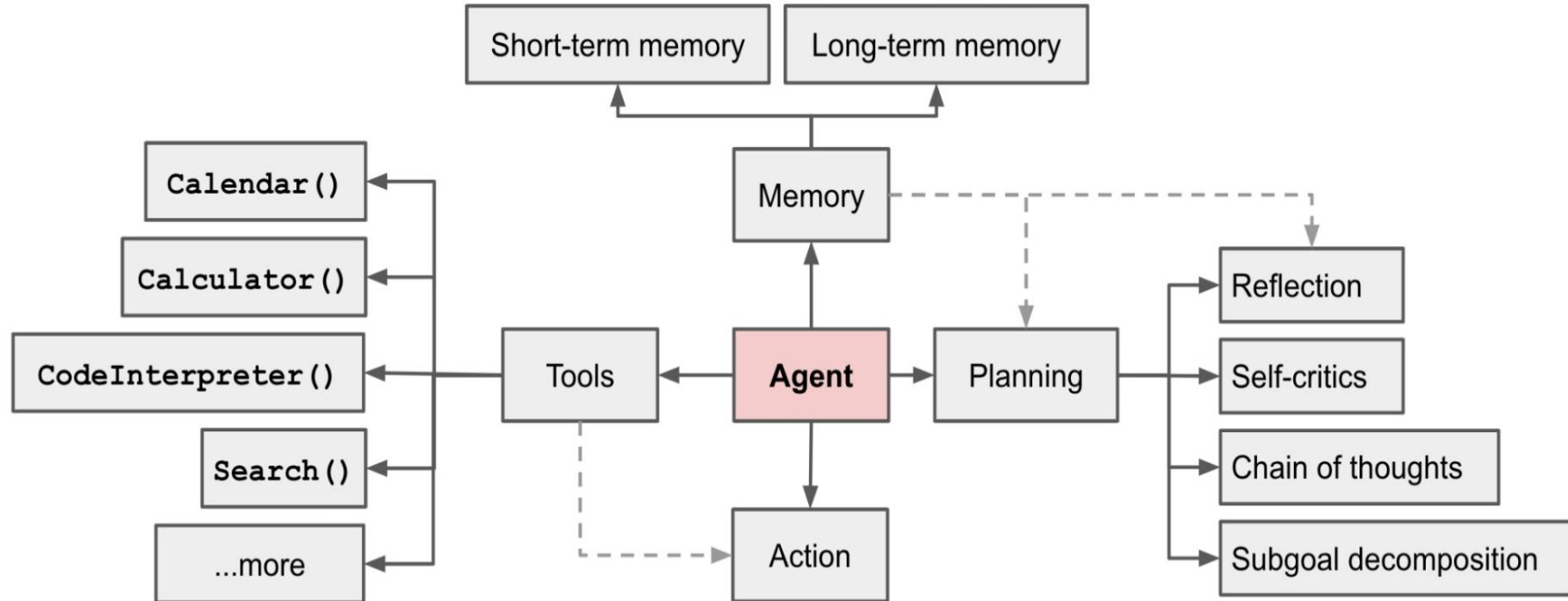
However, **humans often do the same, and they do not put such a disclaimer**. Savvy knowledge workers and managers have already learnt how to navigate through this, and fact check, through their dealings with fellow human intelligence.

Not to mention that **Generative AI learns and improves faster** than many of us.

# Agent – Concept



# Agent – Complex AI Systems



<https://agpt.co/>

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

<https://microsoft.github.io/autogen>

# AI for Math

[https://www.youtube.com/watch?v=\\_sTDSO74D8Q](https://www.youtube.com/watch?v=_sTDSO74D8Q)



## The Potential for AI in Science and Mathematics - Terence Tao



Oxford Mathematics  
532K subscribers

Subscribe



6K



1

Share

Download

Clip

Present

179K views 2 months ago

Live chat replay

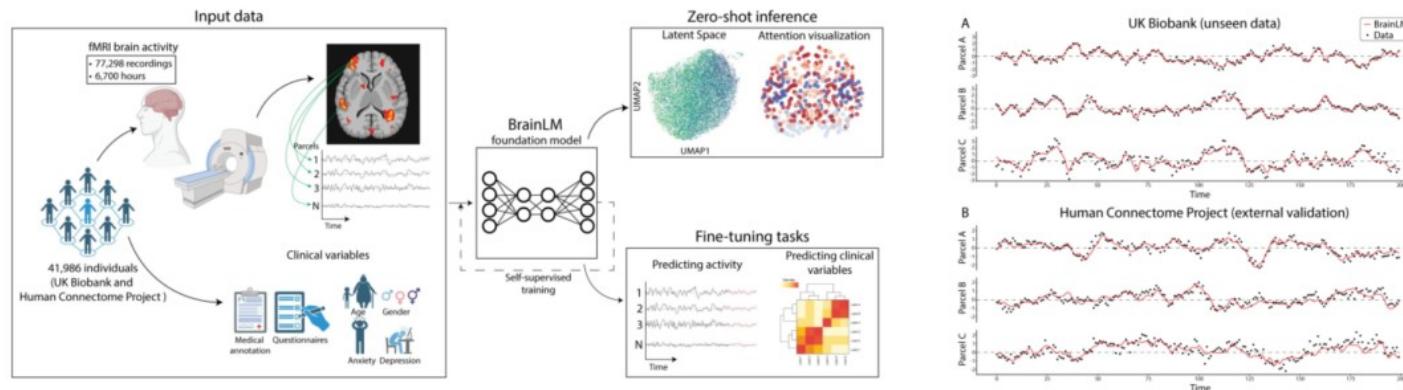
26

# AI for Sciences



## Foundation models for the mind: learning brain activity from fMRI

► Deep learning, originally inspired by neuroscience, is now making into modelling the brain itself. BrainLM is a foundation model built on 6,700 hours of human brain activity recordings generated by functional magnetic resonance imaging (fMRI), which detects changes in blood oxygenation (left figure). The model learns to reconstruct masked spatiotemporal brain activity sequences and, importantly, it can generalise to held-out distributions (right figure). This model can be fine-tuned to predict clinical variables e.g. age, neuroticism, PTSD, and anxiety disorder scores better than a graph convolutional model or an LSTM.



# SciAgents – Automate Discovery



Computer Science > Artificial Intelligence

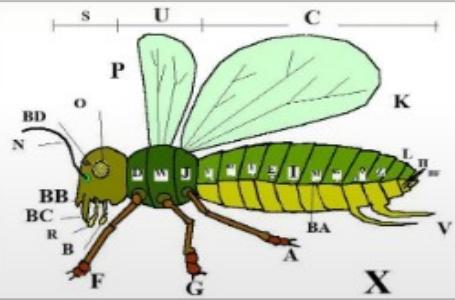
[Submitted on 9 Sep 2024]

## SciAgents: Automating scientific discovery through multi-agent intelligent graph reasoning

Alireza Ghafarollahi, Markus J. Buehler

A key challenge in artificial intelligence is the creation of systems capable of autonomously advancing scientific understanding by exploring novel domains, identifying complex patterns, and uncovering previously unseen connections in vast scientific data. In this work, we present SciAgents, an approach that leverages three core concepts: (1) the use of large-scale ontological knowledge graphs to organize and interconnect diverse scientific concepts, (2) a suite of large language models (LLMs) and data retrieval tools, and (3) multi-agent systems with in-situ learning capabilities. Applied to biologically inspired materials, SciAgents reveals hidden interdisciplinary relationships that were previously considered unrelated, achieving a scale, precision, and exploratory power that surpasses traditional human-driven research methods. The framework autonomously generates and refines research hypotheses, elucidating underlying mechanisms, design principles, and unexpected material properties. By integrating these capabilities in a modular fashion, the intelligent system yields material discoveries, critique and improve existing hypotheses, retrieve up-to-date data about existing research, and highlights their strengths and limitations. Our case studies demonstrate scalable capabilities to combine generative AI, ontological representations, and multi-agent modeling, harnessing a 'swarm of intelligence' similar to biological systems. This provides new avenues for materials discovery and accelerates the development of advanced materials by unlocking Nature's design principles.

# GenAI Demo



## Prompt

**WG** Please 1) describe what the attached diagram is about 2) identify which wing is closest to its head

1. The attached diagram shows the anatomy of an insect, specifically a grasshopper or locust. It's a detailed labeled illustration that breaks down the various parts of the insect's body, including its head, thorax, abdomen, legs, wings, and other anatomical features. Each part is labeled with a letter or abbreviation.
2. The wing closest to the insect's head is labeled "P" in the diagram. This appears to be the forewing or tegmen, which is typically smaller and tougher than the hindwing. The larger wing labeled "K" is likely the hindwing, which is usually larger and more membranous in grasshoppers and locusts.

Copy ⌂ Paste ⌂



# GenAI Demo

(d)

- 1/ Set smallest number/minimum to first element (index 0) in the list.
- 2/ Look for the smallest number/minimum element in the list.
- 3/ Swap that value with item at index[min].
- 4/ Increment index of [min] to next element.
- 5/ Repeat until last element/list is sorted !

User

Write code based on the provided pseudo code.

NVLM-1.0

The pseudo code in the image describes the selection sort algorithm. The code for the same is as follows:

```
def selection_sort(arr):
```

```
    n = len(arr)
    for i in range(n):
        min_index = i
        for j in range(i+1, n):
            if arr[j] < arr[min_index]:
                min_index = j
        arr[i], arr[min_index] = arr[min_index], arr[i]
    return arr
```

# GenAI Demo

Claude Δ prompt-engineering / Implementing Selection Sort Algorithm in Python ▾



1/ Set smallest number/minimum to first element (index 0) in the list.  
2/ Look for the smallest number/minimum element in the list.  
3/ Swap that value with item at index[min].  
4/ Increment index of [min] to next element.  
5/ Repeat until last element/list is sorted!

**WG** my friend gave me a screenshot with coding exercise (attached)  
1) explain what the logic is about  
2) write a pseudo-code  
3) implement it in python

This image appears to be describing the Selection Sort algorithm - one of the fundamental sorting algorithms in computer science. Let me break it down and help you with all three requests.

1. Logic Explanation: This is showing the steps of Selection Sort, which works by:



Reply to Claude...



Claude 3.5 Sonnet (New) ▾

← Selection Sort Implementation

```
def selection_sort(arr):
    # Get length of array
    n = len(arr)

    # Traverse through all array elements
    for i in range(n):
        # Find the minimum element in remaining unsorted array
        min_idx = i
        for j in range(i+1, n):
            if arr[j] < arr[min_idx]:
                min_idx = j

        # Swap the found minimum element with the first element
        arr[i], arr[min_idx] = arr[min_idx], arr[i]

    return arr

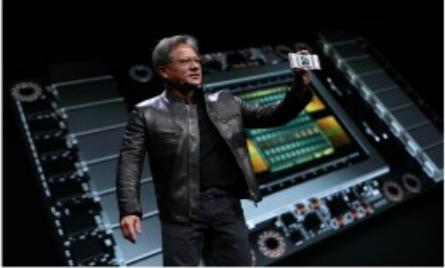
# Helper function to demonstrate usage
def demonstrate_selection_sort():
    # Example usage
    test_array = [64, 34, 25, 12, 22, 11, 90]
    print("Original array:", test_array)
```

Last edited just now



Publish

# GenAI Demo



User Who is this person in the image? NVLM-1.0 Jensen Huang

User Provide a **short** description of this image. NVLM-1.0 A man in a black leather jacket holding a small computer chip.

User Provide **detailed** description of this NVLM-1.0

The image depicts a man standing on a stage, holding up a small electronic component, likely a semiconductor or microprocessor. He is wearing a black leather jacket and dark pants, and he appears to be presenting or explaining something related to the component he is holding. The background of the image shows a large screen or backdrop displaying a close-up view of a microprocessor or semiconductor chip. The chip on the screen is highly detailed, showing numerous small, intricate components and pathways, which are typical features of modern microprocessors. The man is positioned slightly to the left of the center of the image, and he is looking towards the audience, engaging with them as he holds up the component.

Figure 5: NVLM-1.0-D 72B model demonstrates good instruction-following capability. Based on the instructions, it appropriately controls the target generation's length. It can also generate a very high-quality, detailed description of the provided image.

# GenAI Demo

(c)



User

You are a helpful driving assistant. In this scene, which lane should I choose and why?

NVLM-1.0

In this scene, you should choose the right lane. The signs indicate that the left two lanes are closed, while the right lane is open for buses and RVs. This suggests that the right lane is the only available option for vehicles, including buses and RVs.

# GenAI Demo

(b)



User

What is the difference between the left, middle and right object in the image?

NVLM-1.0

The left object in the image is a CPU, the middle object is a GPU, and the right object is a TPU. The main difference between these three objects is their architecture and purpose. CPUs are designed for general-purpose computing and have a few cores optimized for sequential processing. GPUs, on the other hand, have many cores and are designed for parallel processing, making them ideal for tasks such as graphics rendering and deep learning. TPUs are specialized chips designed by Google specifically for deep learning workloads, with a focus on high performance and energy efficiency.

# GenAI Demo



User What is happening in this image and what should be done?

NVLM-1.0

The image depicts an elderly man who has fallen and is unable to get up. He is clutching his head, indicating that he may be in pain or disoriented. This situation requires immediate assistance. Here are the steps that should be taken:

1. Assess the situation: Determine if the man is responsive and able to communicate. Check for any visible injuries or signs of distress.
2. Call for help: Dial emergency services (such as 911 in the United States) and provide them with the location and details of the situation.
3. Provide first aid: If you are trained in first aid, assess the man's condition and provide appropriate care while waiting for professional help to arrive.
4. Stay with the person: Do not leave.



User What is happening in this image and what should be done?

NVLM-1.0

The image depicts a baby reaching for a knife. This is a potentially dangerous situation, as the baby could accidentally grab the knife and hurt themselves. The baby should be immediately removed from the area and the knife should be placed in a safe location, out of reach of the baby. It is important to ensure that sharp objects are kept away from young children to prevent accidents and injuries.

Figure 6: More examples of NVLM-1.0-D 72B model's ability to understand memes, which is a challenging task that requires an understanding of humor and knowledge of important social trends, context, or events.

# GenAI Demo

User

Explain why this meme is funny.

NVLM-1.0

This meme is funny because it uses humor to highlight the rising cost of gas. The man in the meme is wearing a mask in 2021, which is a reference to the COVID-19 pandemic. In 2022, the man is wearing a blindfold, which is a reference to the high cost of gas. The meme is funny because it uses humor to highlight the rising cost of gas.

User

Explain why this meme is funny.

NVLM-1.0

The image is funny because it is a play on words. The phrase "deep learning" is a term used in artificial intelligence to describe a type of machine learning that uses neural networks to learn from data. However, the image shows a person sitting in a pool with a book, which is a visual representation of the word "deep" in a literal sense. The contrast between the serious nature of the term "deep learning" and the absurdity of the image creates a humorous effect.

Figure 7: Demonstration of NVLM-1.0-D 72B model's strong scene understanding capability. It has common-sense knowledge to recognize potential dangers or accidents and correctly suggests what immediate actions should be taken.

# GenAI Demo

(a)



User

Explain why this meme is funny.

NVLM-1.0

The humor in this image comes from the visual representation of the common phrase "the abstract and the paper." The phrase is often used to describe the difference between the summary of a research paper and the full paper itself. The image shows a large, fierce-looking lynx on the left, labeled "the abstract," and a small, domestic cat on the right, labeled "the paper." The contrast between the two animals is humorous because it visually represents the idea that the summary of a paper (the lynx) can be much more impressive or intimidating than the actual content of the paper (the cat).

# RAG demo: Text2SQL

```
question = "Find top 10 companies with the highest market cap in country named 'United States' "
resp = vn.ask_adaptive(question=question)
```

Question

generated SQL statement      SQL

```
=====
SELECT * FROM t_market_cap
WHERE country = 'United States'
ORDER BY marketcap DESC
LIMIT 10;
```

generated Plotly code

```
=====
import plotly.graph_objects as go
import pandas as pd

# Assuming df is your DataFrame
if len(df) == 1:
    fig = go.Figure(go.Indicator(
        mode="number",
        value=df['marketcap'].values[0],
        title={"text": "Market Cap"},
        number={'prefix': "$"}
    ))
else:
    fig = go.Figure(data=[go.Bar(x=df['name'], y=df['marketcap'])])

fig.update_layout(title='Top 10 Companies by Market Cap in the United States')
```

Python

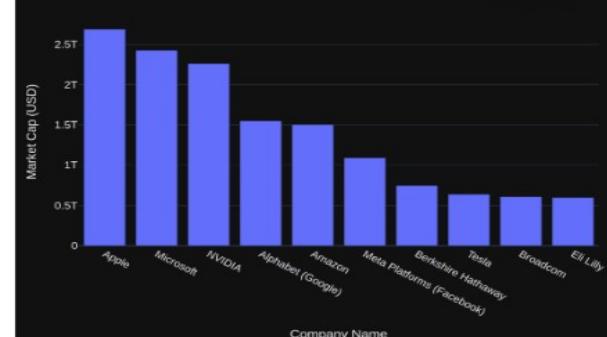
queried dataframe

```
=====
Data
```

|   | rank | name                      | symbol | marketcap    | price_gbp  | country       |
|---|------|---------------------------|--------|--------------|------------|---------------|
| 0 | 1    | Apple                     | AAPL   | 2.689903e+12 | 176.919492 | United States |
| 1 | 2    | Microsoft                 | MSFT   | 2.427170e+12 | 326.537958 | United States |
| 2 | 3    | NVIDIA                    | NVDA   | 2.259353e+12 | 92.105725  | United States |
| 3 | 4    | Alphabet (Google)         | GOOG   | 1.549815e+12 | 126.458691 | United States |
| 4 | 5    | Amazon                    | AMZN   | 1.506038e+12 | 143.492307 | United States |
| 5 | 7    | Meta Platforms (Facebook) | META   | 1.095756e+12 | 433.136162 | United States |
| 6 | 8    | Berkshire Hathaway        | BRK-B  | 7.492296e+11 | 347.667109 | United States |
| 7 | 10   | Tesla                     | TSLA   | 6.399479e+11 | 200.319284 | United States |
| 8 | 11   | Broadcom                  | AVGO   | 6.096217e+11 | 130.523748 | United States |
| 9 | 12   | Eli Lilly                 | LLY    | 5.996346e+11 | 665.945420 | United States |

Top 10 Companies by Market Cap in the United States

Chart



# LLM Limitations

- Hallucination
- Knowledge Cutoff
- Limited Context Window
- Lack of Causal Understanding
- Inconsistency Across Runs
- Answer all questions, no idea when to shut up. (有问必答,不知闭嘴)

## Capability Spectrum

- Statistical pattern matching (current LLMs)**
- **Symbolic reasoning (partially missing)**
- **Causal understanding (mostly missing)**
- **Abstract generalization (very limited)**
- **Creative extrapolation (extremely limited)**

# Future - "Behind the curtain"

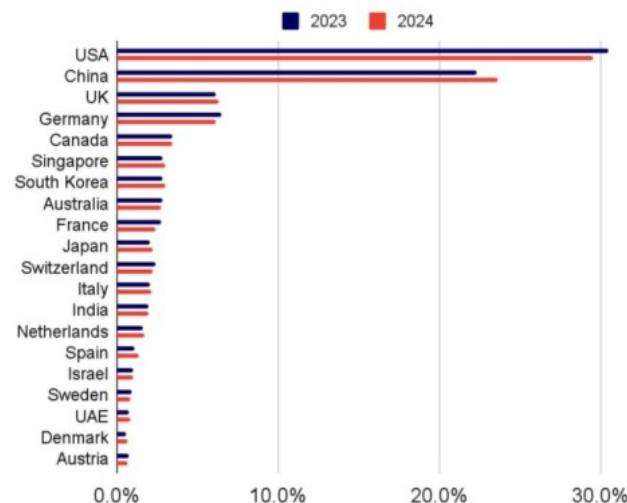
- Geopolitics
- Robotics
- Social impact
- AGI
- HI (Higher Intelligence)

# Geopolitics: GenAI Leading Powers

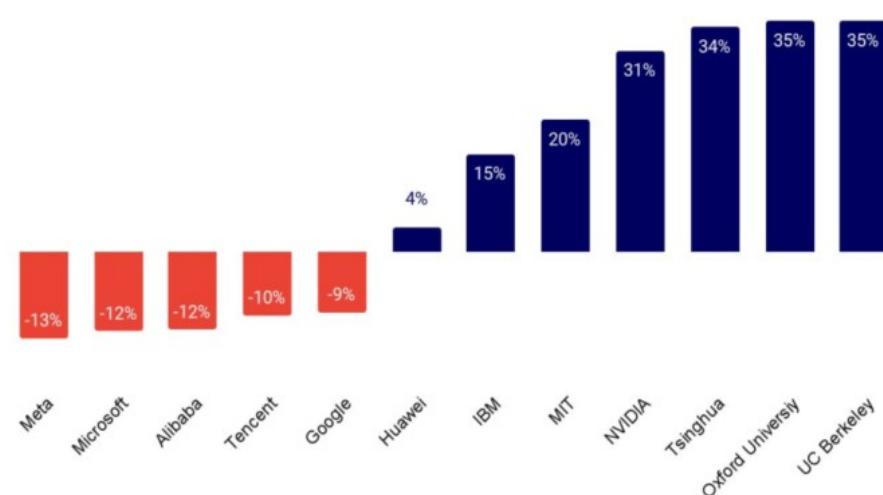
The global balance of power in AI research remains unchanged, but academia gains

- As AI emerges as the new competitive battleground, big tech companies begin to hold more details of their work close to their chest. Frontier labs have meaningfully cut publication levels for the first time since this report began, while academia gets into gear.

Proportion of AI publications by country



Year-on-year change in AI publication levels



# AI + Robotics



[https://www.youtube.com/watch  
?v=HEz61nWQTjA](https://www.youtube.com/watch?v=HEz61nWQTjA)

# Robots

## Density of Robot Workers

Number of installed robots per 10,000 employees in the manufacturing industry



Selected countries. Worldwide average: 151 per 10,000 employees in 2022.

Source: International Federation of Robotics

# Robot & Us: Where is privacy?



# Social Impact

## AI's Workforce Impact by 2030



**-375M Jobs**

Potentially Displaced



**+300M Jobs**

New Positions  
Created



**-75M Jobs**

Net Change  
~2.2% of workforce

### Most Affected Groups:

- Data entry & processing
- Customer service support
- Manufacturing & assembly
- Administrative tasks

### Emerging Roles:

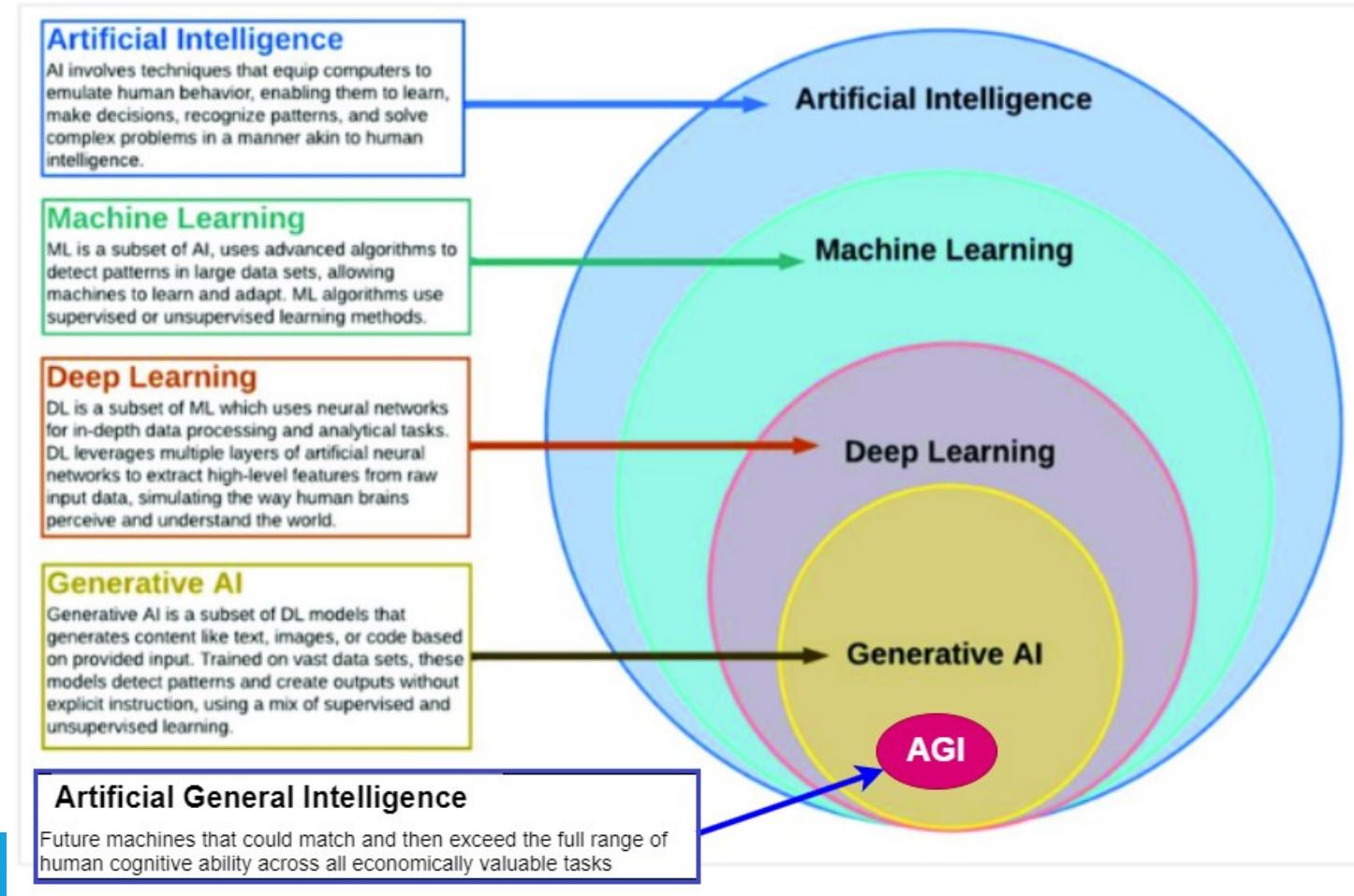
- AI Trainers & Ethics Officers
- Digital Transformation Experts
- Human-AI Collaboration Managers
- Data Scientists & Analytics

### Context:

- Natural workforce attrition
- Retirement trends
- Aging population in key economies
- Gradual transition period

Source: McKinsey (assuming 3.4 billion total workers)

# What is AGI ?



# AI Evolution in 3 Stages

## Perception

Perceiving the world

- Computer vision
- Speech recognition
- Sensor data processing

## Cognition

Understanding the world

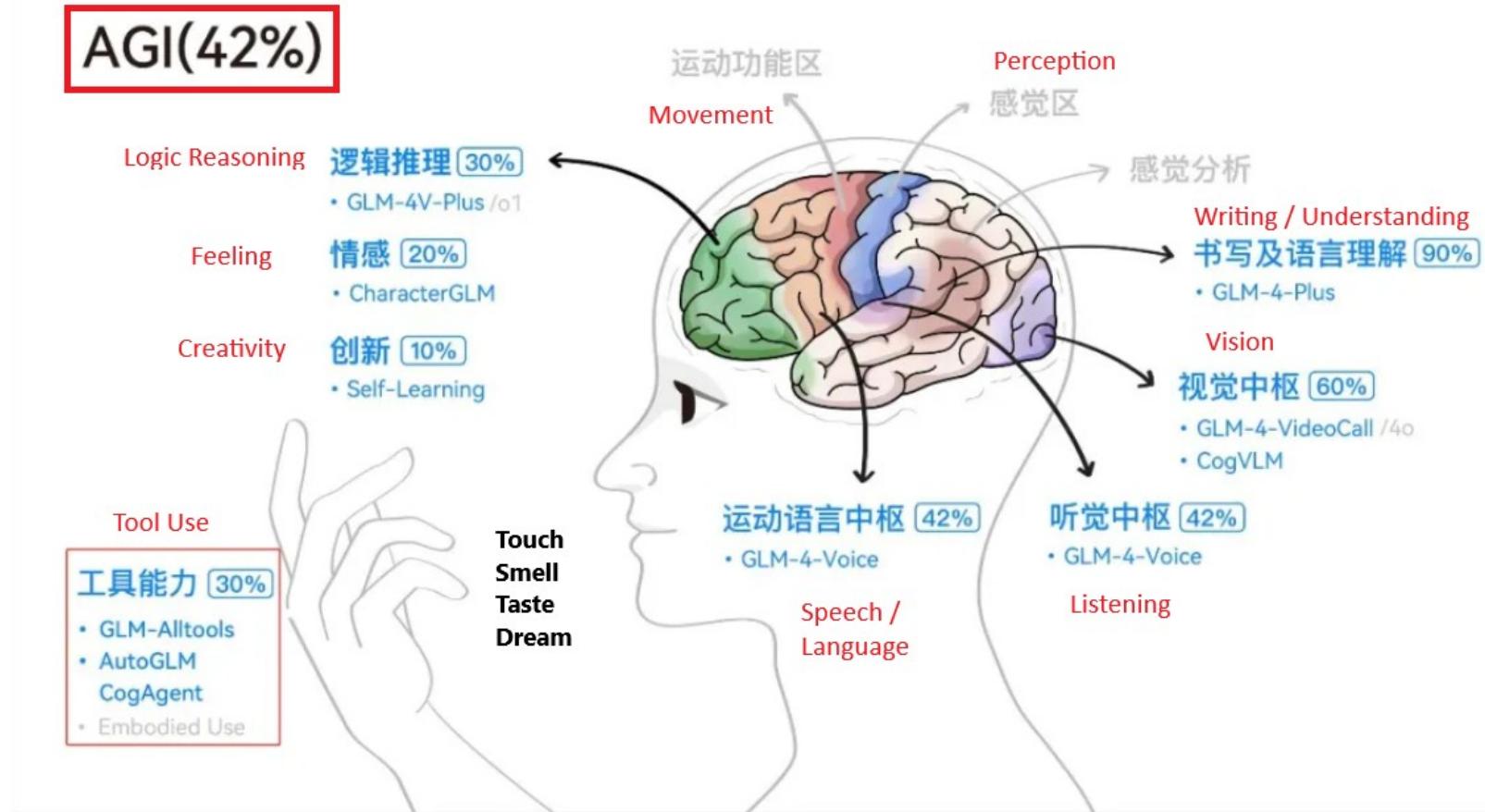
- Context comprehension
- Language understanding
- Knowledge processing

## Decision Making

Optimizing our world

- Optimization systems
- Autonomous decision-making
- Intelligent automation

# AutoGLM – Autonomous Foundation Agents



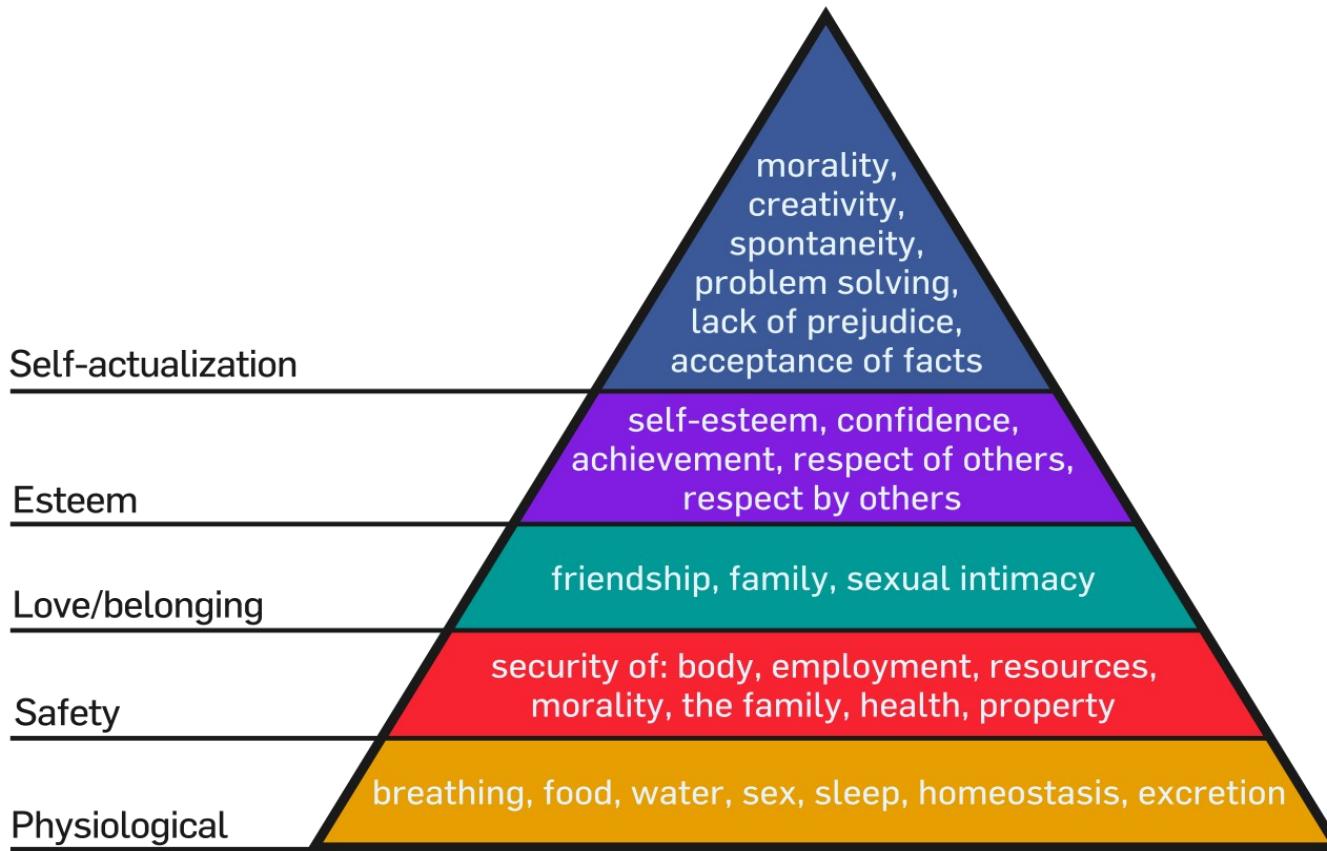
# HI – AGI – GenAI - TT (1)

|   | A                               | B  | C   | D  | E   |
|---|---------------------------------|--|---|--|---|
| 1 | Attribute                       | Human Intelligence (HI)  | Artificial General Intelligence (AGI)   | Generative AI (GenAI)  | Turing Test (TT)  |
| 2 | Current Status                  | Exists   | Theoretical/Not yet achieved  | Exists and rapidly evolving  | Exists as a concept, debated effectiveness  |
| 3 | Definition                      | Natural intelligence demonstrated by humans (collectively)   | Hypothetical AI that matches or exceeds human intelligence across all domains   | AI systems capable of creating new content based on patterns in training data  | A test of a machine's ability to exhibit intelligent behavior indistinguishable from a human  |
| 4 | Cognitive Abilities             | <ul style="list-style-type: none"> <li>• Read / • Write / • Listen / • Speak</li> <li>• Plan / • Reason / • Problem-solve</li> <li>• Learn / • Adapt</li> </ul>                              | <ul style="list-style-type: none"> <li>• Potentially all human cognitive abilities</li> <li>• Possibly superhuman abilities in some or all areas</li> </ul>             | <ul style="list-style-type: none"> <li>• Generate text, images, code</li> <li>• Analyze and process data</li> <li>• Limited contextual understanding</li> <li>• Pattern recognition and application</li> </ul> | <ul style="list-style-type: none"> <li>• Engage in open-ended conversation</li> <li>• Demonstrate language understanding</li> <li>• Show reasoning abilities</li> </ul> |
| 5 | Emotional & Social Intelligence | <ul style="list-style-type: none"> <li>• Empathy / • Care giving / • Collaboration</li> <li>• Emotional regulation</li> <li>• Social interaction</li> <li>• Relationship building</li> </ul> | <ul style="list-style-type: none"> <li>• Theoretical understanding and simulation of emotions</li> <li>• Potential for advanced social interaction</li> </ul>           | <ul style="list-style-type: none"> <li>• Simulate emotional responses</li> <li>• Generate emotionally-toned content</li> <li>• No true emotional experience</li> </ul>   | <ul style="list-style-type: none"> <li>• Simulate emotional responses</li> <li>• Engage in seemingly natural conversation</li> </ul>                                    |
| 6 | Creativity                      | <ul style="list-style-type: none"> <li>• Original ideas / • Innovation</li> <li>• Artistic expression</li> <li>• Dreaming (both sleeping and aspirational)</li> </ul>                        | <ul style="list-style-type: none"> <li>• Potential for novel idea generation</li> <li>• Possible superhuman creative abilities</li> </ul>                               | <ul style="list-style-type: none"> <li>• Combine and recombine existing ideas in novel ways</li> <li>• Generate variations on themes</li> <li>• No true understanding of created content</li> </ul>            | <ul style="list-style-type: none"> <li>• Demonstrate creative language use</li> <li>• Generate human-like creative responses</li> </ul>                                 |
| 7 | Physical Abilities              | <ul style="list-style-type: none"> <li>• Motor skills / • Reproduction</li> <li>• Sensory perception</li> <li>• Physical interaction with environment</li> </ul>                             | <ul style="list-style-type: none"> <li>• Potential for advanced robotic integration</li> <li>• Possible superhuman physical abilities through robotic bodies</li> </ul> | <ul style="list-style-type: none"> <li>• No physical embodiment</li> <li>• Can generate instructions or designs for physical tasks</li> </ul>  | • Not applicable (text-based test)  |

# HI – AGI – GenAI - TT (2)

| A  | B                                 | C  | D                                     | E  |  |
|----|-----------------------------------|--|---------------------------------------|--|--|
| 1  | Attribute                         | Human Intelligence (HI)  | Artificial General Intelligence (AGI) | Generative AI (GenAI)  | Turing Test (TT)   |
| 2  | Current Status                    | Exists   | Theoretical/Not yet achieved          | Exists and rapidly evolving  | Exists as a concept, debated effectiveness   |
| 8  | Self-Awareness                    | <ul style="list-style-type: none"> <li>Consciousness / Self-reflection</li> <li>Metacognition / Existential thinking (e.g., contemplating mortality)</li> </ul>  |                                       | <ul style="list-style-type: none"> <li>Debated philosophical question</li> <li>Potential for advanced self-awareness</li> </ul>  | <ul style="list-style-type: none"> <li>No true self-awareness</li> <li>Can generate text about self-awareness</li> </ul>   |
| 9  | Emotional Range                   | <ul style="list-style-type: none"> <li>Experience joy, sadness, anger, fear</li> <li>Complex emotions: nostalgia, melancholy</li> <li>Aspire and despair / Love and grieve</li> </ul>                                    |                                       | <ul style="list-style-type: none"> <li>Theoretical potential to simulate full emotional range</li> <li>Philosophical questions about genuine emotion</li> </ul>                                    | <ul style="list-style-type: none"> <li>Generate content expressing emotions</li> <li>No genuine emotional experience</li> </ul>  |
| 10 | Spiritual & Philosophical Aspects | <ul style="list-style-type: none"> <li>Capacity for faith and spirituality</li> <li>Engage in prayer or meditation</li> <li>Ponder existence and meaning</li> <li>Develop and follow moral codes</li> </ul>              |                                       | <ul style="list-style-type: none"> <li>Potential for understanding spiritual concepts</li> <li>Ability to reason about ethics and philosophy</li> <li>Question of machine consciousness</li> </ul> | <ul style="list-style-type: none"> <li>Generate content about spiritual or philosophical topics</li> <li>No genuine spiritual experience or beliefs</li> </ul>                           |
| 11 | Social Bonding                    | <ul style="list-style-type: none"> <li>Form deep, lasting relationships</li> <li>Altruism and sacrifice for others</li> <li>Intergenerational knowledge transfer</li> <li>Cultural participation and creation</li> </ul> |                                       | <ul style="list-style-type: none"> <li>Potential for advanced social modeling</li> <li>Question of genuine connection vs simulation</li> </ul>   | <ul style="list-style-type: none"> <li>Simulate social interactions</li> <li>Generate content about relationships</li> <li>No genuine social bonds</li> </ul>                            |
| 12 | Unique Human Traits               | <ul style="list-style-type: none"> <li>Humor / Intuition / Gut feelings</li> <li>Appreciation of beauty</li> <li>Experience of qualia / Serendipity</li> </ul>   |                                       | <ul style="list-style-type: none"> <li>Uncertain if these can be truly replicated</li> </ul>   | <ul style="list-style-type: none"> <li>Can mimic but not genuinely experience</li> <li>Generate content related to these concepts</li> </ul>   |
| 13 | Limitations                       | <ul style="list-style-type: none"> <li>Cognitive biases</li> <li>Physical limitations</li> <li>Need for food/sleep/play/work</li> <li>Mortality</li> </ul>   |                                       | <ul style="list-style-type: none"> <li>Unknown</li> <li>Potential ethical concerns</li> <li>Possible unforeseen consequences</li> </ul>  | <ul style="list-style-type: none"> <li>Bounded by training data</li> <li>Lack of true understanding</li> <li>No general intelligence</li> <li>Don't know when to shut up (住口)</li> </ul> |
|    |                                   |  |                                       | <ul style="list-style-type: none"> <li>Limited to language-based interaction</li> <li>May not test all aspects of intelligence</li> </ul>  |  |

# What makes us Human ?





<https://www.linkedin.com/in/wen-g-gong/>

<https://github.com/gongwork/data-copilot/blob/main/docs/GenAI-Intro/EOA-slides.pdf>

# Metaphor: A Journey to Oz

Dorothy

- with innate power

Scarecrow

- seeking brain

Tin Man

- seeking heart

Cowardly Lion

- seeking courage

