For non-technical background and recurrent learners

Generative AI knowledge Basics

(Through using ChatGPT and Copilot*)

*Bing AI　Verification was conducted at the time of BingAI, and being updated as Copilot thereafter.

Training Text

Global Knowledge Link Center

This content is specialized in generative AI. For more information on AI, please refer to the AI Knowledge textbook.

The impact of “GenAI”1/2

- "Sneakers made of transparent crystal stones. A boy wearing the sneakers is dancing on the street.“
- A wonderful finish with the texture of crystal and a sense of dynamism was created with the first shot.

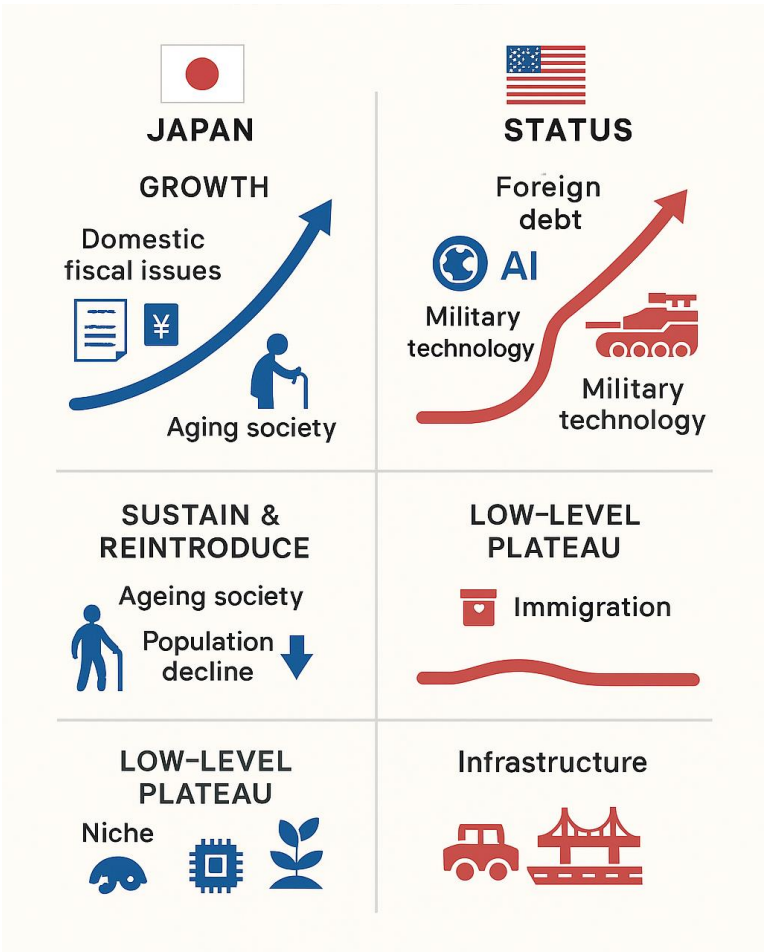
Pics : Bing Image Creator」dalle-3/

Source and Pics : https://channelivy.com/ai/dalle-3/

© Global Knowledge Link Center.

The impact of “GenAI”2/2

- National Strategy Matrix (Finance x Population x Hype Cycle), Comparison between Japan and U.S. States
-



Additional Note:
Applicability: While this table compares Japan and the U.S., the underlying concepts and viewpoints related to fiscal structure, demographic structure, policy culture, and growth drivers are applicable to analyzing any nation. Each country will have its unique characteristics within these categories.

© Global Knowledge Link Center.

Acknowledgement and our goal as Introduction

- This text is mainly designed to meet people without a technical background, those who have distanced themselves from mathematics, or those who are relearning systems, rather than those with a systems background.
- Therefore, We would like to request that you use diagrams and simple expressions to explain difficult theorems and logic as much as possible to help clarify some of the mechanisms involved.
- The reason why we focused on elucidating the mechanism was to clarify as much as possible the questions that would eventually arise while promoting the use of generative AI (GenAI), and to ensure that it could be used in applied applications. We strongly believe that this process is a first and great step.
- Prof. Takuya Matsuda (Prof. Emeritus at Kobe University) gave us various insights and suggestions. In addition, we referred to information, knowledge and explanations from GenAI service providers (OpenAI, Google, IBM, etc.), press release articles, and articles and blogs from universities, research institutes, and researchers. Thank you very much with our appreciation.

© Global Knowledge Link Center.

Table of contents

Members only

Another volume text
(The first edition has been distributed)

*This document is based on the experience gained when launching a workshop at a certain consulting firm.

Content titles and orders or so may change without notice

© Global Knowledge Link Center.

Why do you need development of “AI”?

- Anyway, what do you want to do?
- Birds can perch on trees, but airplanes can’t. But that doesn’t matter. In short, the goal is not to reproduce birds, but to travel quickly to the destination by flying. (in Professor Matsuda's talk)

Pics : https://singularity.jp/category/event/page/3/

© Global Knowledge Link Center.

Orientation

Explanatory note

- Word : This pitch shows word and meaning, so you don’t need memorize the contents.
- Reference : Although it doesn’t affect the mainstream of the content, it supports and reinforces the contents.
- Abbreviations
  - AI : Artificial Intelligence
  - Generative Artificial Intelligence, Generative AI, GAI, hereafter referred to as Gen AI
  - RAG : Retrieval Augmented Generation which means GenAI can retrieve related information from an external database (Retrieval) and generate the answer (Generation), GenAI answering some to prompts.
  - RDB and VDB : Relational Database and Vector Database
  - RPA : Robotic Process Automation

* Please excuse any words not used in this text directly.

© Global Knowledge Link Center.

Review basic words

Refer to Excel sheet

Membership only

© Global Knowledge Link Center.

The goal of this text

Generative AI

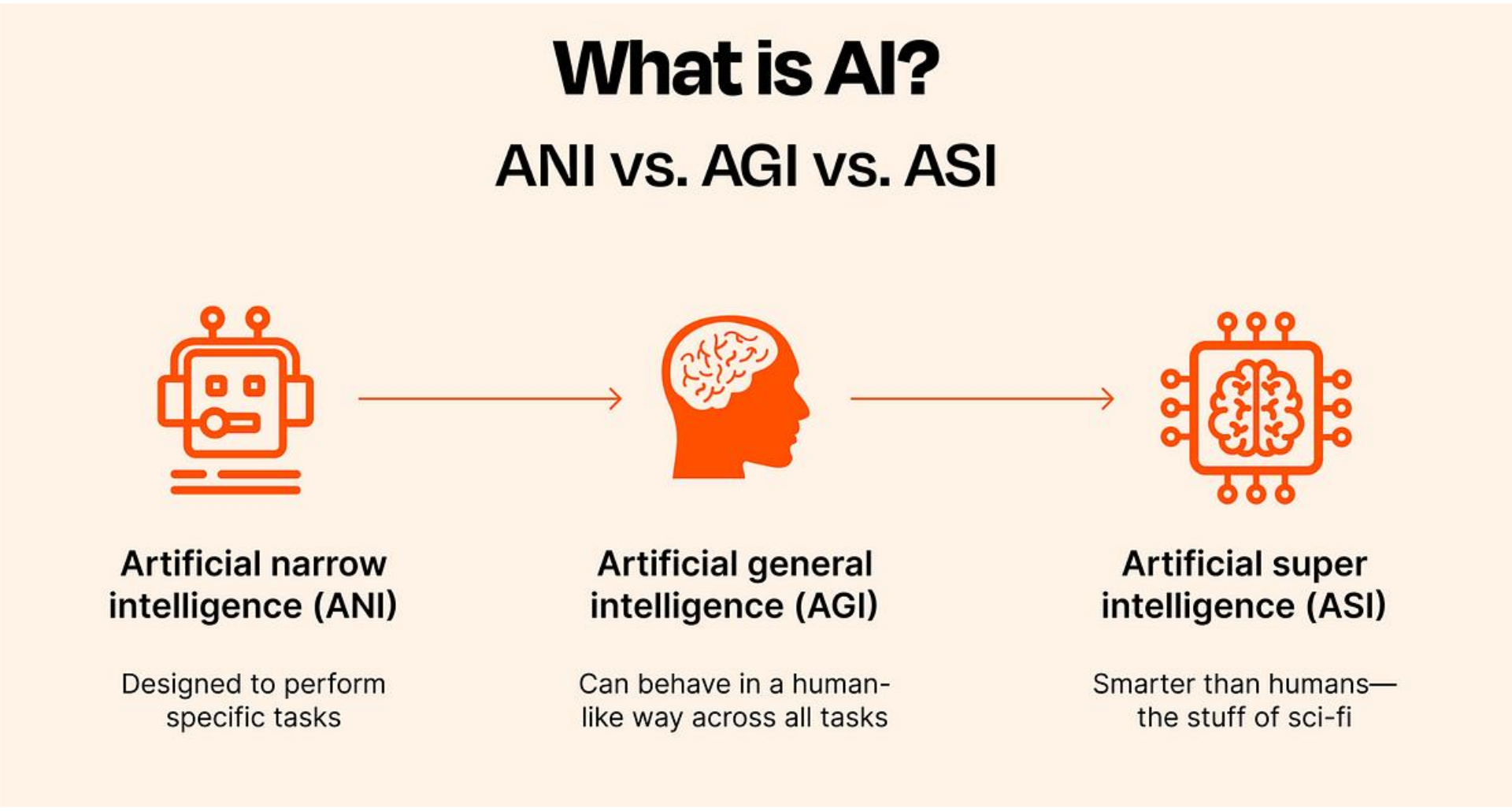
Algorithm
(Perceptron and transformer)

AI project

© Global Knowledge Link Center.

What's AI?

- ANI , AGI and ASI

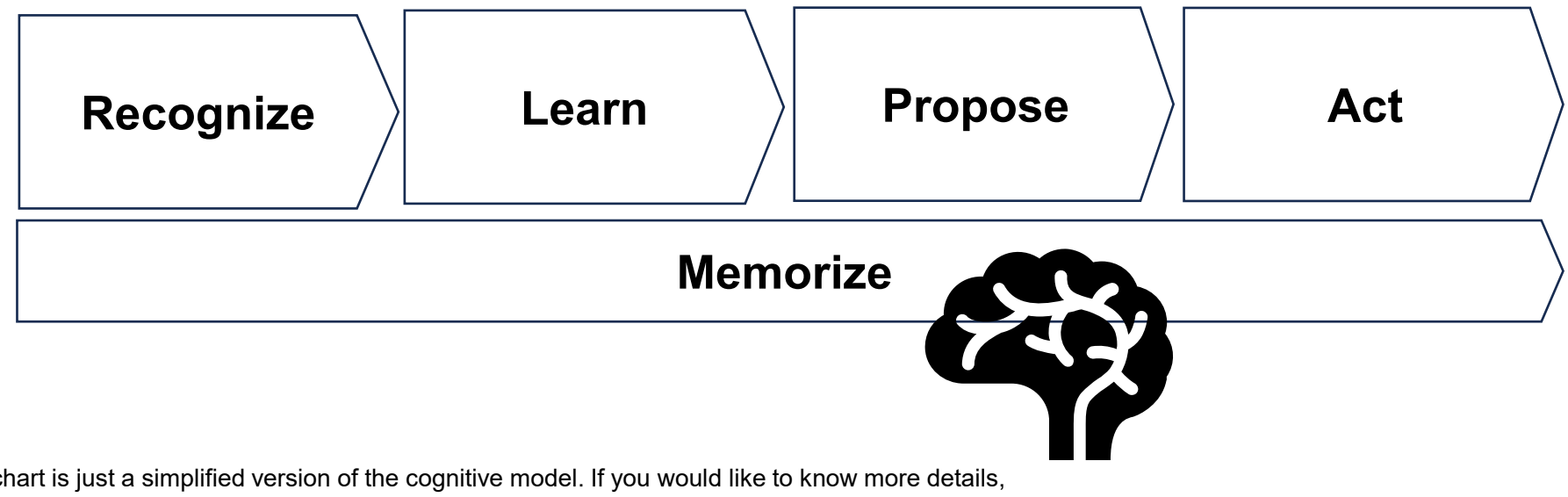


Pics : https://medium.com/@InnovateForge/understanding-ani-agi-and-asi-in-artificial-intelligence-6265d010941

© Global Knowledge Link Center.

Reference)What's AI?

- AI is an attempt by humans to artificially reproduce the thinking process by imitating the brain mechanism through computing calculation processing.
- The process from thinking to acting is divided into 5 categories: Recognize (text and data, information, knowledge, etc.), Learn, Propose means coming up with ideas and planning actions or so), and Act. And then Memorize means these series of processes, although not all, are memorized and remembered as a rule of thumb.
- Each process is not necessarily continuous, and processing is often performed in parallel.
- It forms a continuous process similar to that performed by humans through linking data and algorism (see next page).



This chart is just a simplified version of the cognitive model. If you would like to know more details, please refer to Human Information Processing, Cognitive Models, etc.

© Global Knowledge Link Center.

Reference) AI definition 1/3

Refer to Excel sheet

- Example
  - Alan Turing, UK National Physical Laboratory :
    - Conducted the "Turing Test" and proposed that if AI reaches a level where it is indistinguishable from humans in interaction with humans, it should be considered intelligent (Computing Machinery and Intelligence (1950))
    - -We also explored the potential of AI and wondered whether computers could imitate human intelligence.
    - -By the way, the concept of artificial intelligence itself, not the word artificial intelligence, was proposed by Alan Turing in the 1947 "Lecture to London Mathematical Society" (John McCarthy, mentioned below). (Be careful not to confuse it with achievements)

Subject person, Group	Definition, Thoughts and Achievements
Alan Turing, UK National Physical Laboratory	Conducted the "Turing Test" and proposed that if AI reaches a level where it is indistinguishable from humans in interaction with humans, it should be considered intelligent (Computing Machinery and Intelligence (1950)) -We also explored the potential of AI and wondered whether computers could imitate human intelligence. -By the way, the concept of artificial intelligence itself, not the word artificial intelligence, was proposed by Alan Turing in the 1947 "Lecture to London Mathematical Society" (John McCarthy, mentioned below). (Be careful not to confuse it with achievements)
Allen Institute for Brain Science	AI is “a technology that imitates human cognitive abilities”
David Griffiths, No record of affiliation information (authenticity of data unknown)	AI is a system that uses self-learning algorithms to extract patterns from data and solve problems.
Demis Hassabis,Deepmind	AI is defined as "an algorithm or system for solving complex problems."
Elon Musk, Space X, Tesla and others	AI is "a system that reproduces or exceeds human intelligence"
European Association for Artificial Intelligence (EurAI):	-Artificial intelligence is a field of science and technology concerned with the design and construction of intelligent agents -EurAI is promoting the spread of AI research and ethical standards in Europe.
Herbert Simon, Carnegie Mellon University	- Artificial intelligence is the design and construction of computer programs that imitate human intellectual activity. -Simon has made important contributions in the field of AI, including modeling economic decision-making.
Hideyuki Nakajima, Future University Hakodate Hideaki Takeda, National Institute of Informatics	An artificially created entity with intelligence. Or, it is a field that studies intelligence itself by trying to create it.
Hiroshi Yamakawa, Dwango Artificial Intelligence Laboratory	Among computer intelligence, I think that cases where humans directly or indirectly design can be called artificial intelligence.

© Global Knowledge Link Center.

Reference)) AI definition 2/3

Refer to Excel sheet

- Example
  - John McCarthy, Stanford University
    - '-John McCarthy first used the term "artificial intelligence" in 1956 at the Dartmouth Conference (“The Dartmouth Summer Research Project on Artificial Intelligence”). Defined artificial intelligence as a system that can programmatically perform human intellectual tasks.
    - -His other achievements include the programming language LISP and circumscription, for which he was awarded the Turing Award in 1971.
  - Allen Institute for Brain Science : AI is “a technology that imitates human cognitive abilities”
  - Demis Hassabis, Deepmind : AI is defined as "an algorithm or system for solving complex problems."
  - Yutaka Matsuo, University of Tokyo : Artificially created human-like intelligence, or the technology to create it. Being as intelligent as a human means a computer that can "notice," that is, a computer that can generate features from data and model phenomena.

© Global Knowledge Link Center.

Reference)) AI definition 3/3

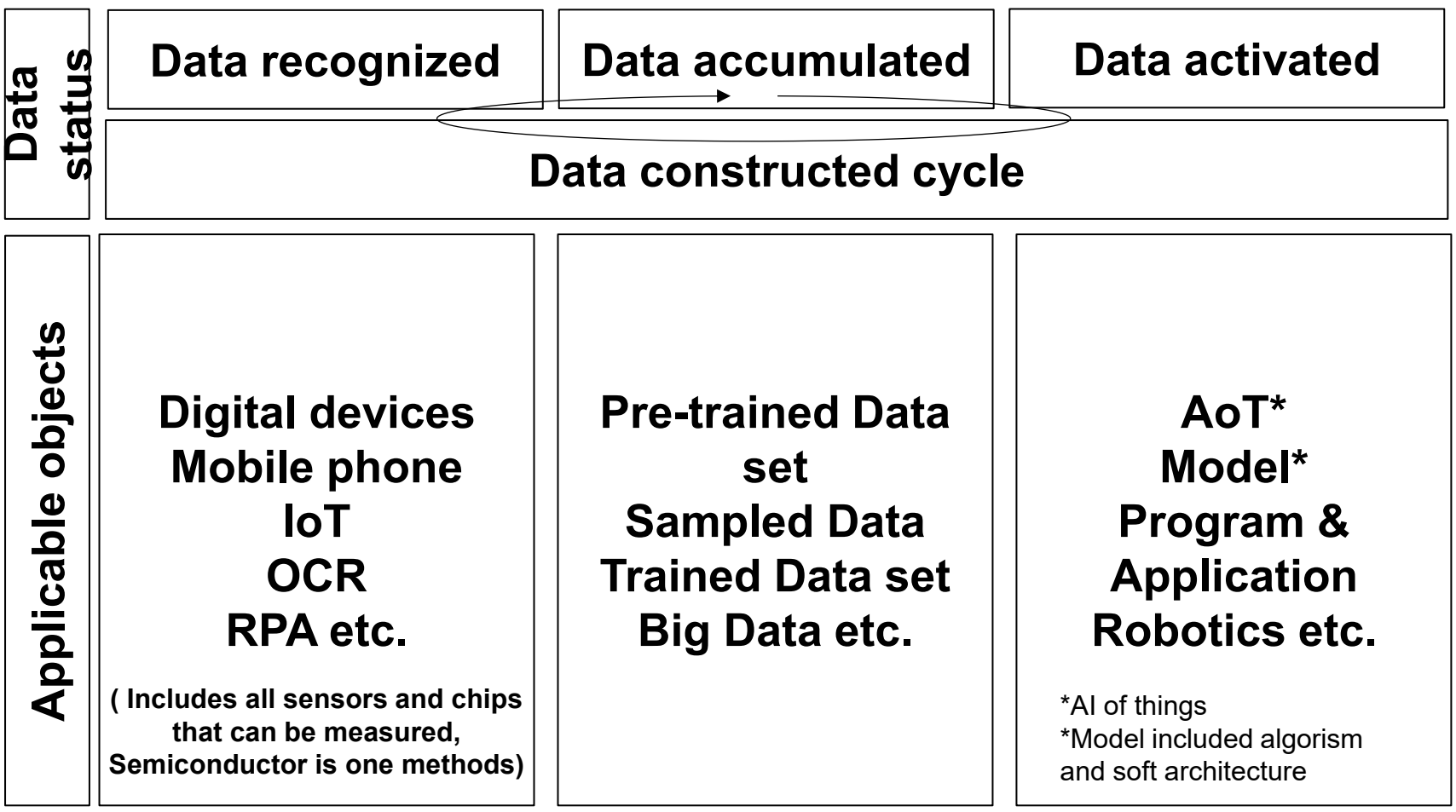
Refer to Excel sheet

Subject person, Group	Definition, Thoughts and Achievements
Hitoshi Matsubara, Future University Hakodate	Artificial intelligence that is ultimately indistinguishable from humans
Jeff Hawkins, Numenta	AI is “a technology that understands and imitates brain functions and processes.”
John McCarthy, Stanford University	-John McCarthy first used the term "artificial intelligence" in 1956 at the Dartmouth Conference (“The Dartmouth Summer Research Project on Artificial Intelligence”). Defined artificial intelligence as a system that can programmatically perform human intellectual tasks. -His other achievements include the programming language LISP and circumscription, for which he was awarded the Turing Award in 1971.
Julian Togelius, New York University	AI is “a technology that imitates human behavior and decision-making”
Koichi Hori, University of Tokyo	A new world of artificial intelligence
Makoto Nagao, Kyoto University	It is a system that simulates human brain activity to the utmost limit.
Marcus Hutter, ANU (The Australian National University)	AI is a “purposeful intelligent system”
Marvin Minsky: MIT	- Artificial intelligence is a computer program for performing intelligent tasks. -In 1959, co-founded the MIT Computer Science and Artificial Intelligence Laboratory with John McCarthy (known as the father of artificial intelligence).
Minoru Asada, Osaka University	Artificial intelligence cannot be clearly defined because the definition of intelligence is not clear.
Open AI	- AI is "a technology that imitates human intelligence"
Riichiro Mizoguchi, Jaist	It is an artificially created system that behaves intelligently.
Satoshi Kurihara, University of Electro-Communications	Although it is an artificially created intelligence, we imagine that its level of intelligence exceeds that of humans.
Stuart Russell, University of California, Berkeley	AI is "technology that imitates human-like thinking and behavior."
Takahira Yamaguchi, Keisho University	A structural system for imitating, supporting, and transcending human intellectual behavior
Takashi Ikegami, University of Tokyo	A system that can artificially create emotional and playful interactions that we naturally interact with with pets and people, regardless of or in defiance of the laws of physics.
Toyoaki Nishida, Kyoto University	A "mecha with intelligence" or "mecha with a mind"
Yoshua Bengio, University of Montreal	AI is “a system that learns patterns from data”
Yutaka Matsuo, University of Tokyo	Artificially created human-like intelligence, or the technology to create it. Being as intelligent as a human means a computer that can "notice," that is, a computer that can generate features from data and model phenomena.

© Global Knowledge Link Center.

Reference)What’s AI? Modeling the status and usage of data

- Data system ,the process by which data becomes knowledge and intelligence, links



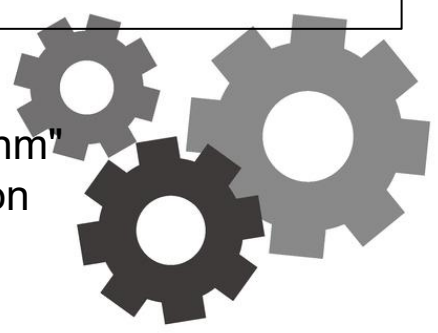
© Global Knowledge Link Center.

Reference) The deference, system, mechanism and algorism?

- Do you think of the deference, system, mechanism and algorism?

System	**System is used in two main senses.** **1. In a broad sense, when we express the structure and mechanism of something large, we call it a social system.** **2. In a narrow sense, individual or integrated software is called a system.**
Mechanism	**Mechanism mainly refers to the hardware structure, connection with functions, and sometimes same with a trick.**
Algorism	**Algorithm mainly refers to the mathematical model and logic(order of transaction) of the software (program).**

Therefore, even if the expressions "mechanism" and "algorithm" are used in a description of a system, there is no contradiction between them.



© Global Knowledge Link Center.

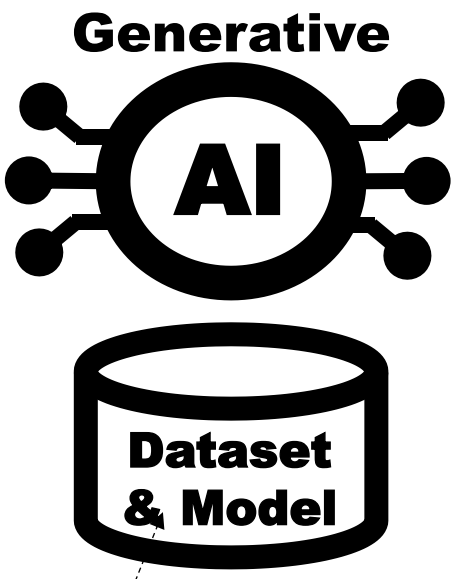
What's "GenAI"?1/5

- GenAI is a general term for AI systems that learn a huge amount of information and knowledge from prepared training data and fine-tuned, the Internet and connected external data (as RAG and special data for each company or experts), and generate intended outputs by responding to prompts with a conversational approach.

Text as Prompt which is Question here

Can you explain your mechanism?

XXXXXXX



Text as Execution which is Answer here

LLM works with huge dataset and models called having some main algorisms

XXXXXXX

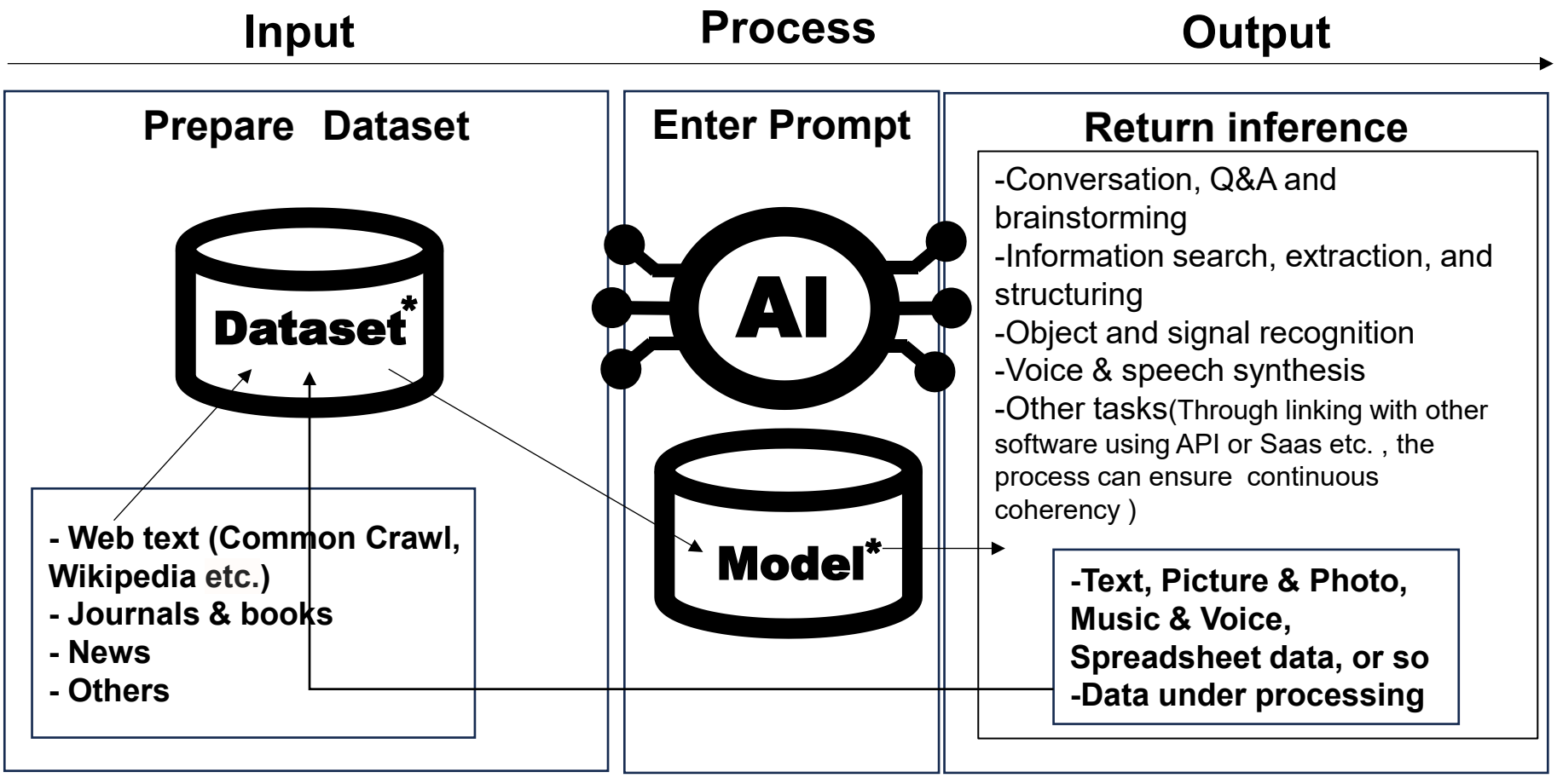
<u>In short, it is really huge DB with UI has a natural language chat function.</u>

This database is the core engine and backbone of a generative AI system. Based on that database, it can understand prompts, make inferences, and express them in text, picture and music or so. These are realized through combinations of plural various of technologies.

© Global Knowledge Link Center.

What's "GenAI"?2/5

- Simply speaking, it is an AI system that receives prompt instructions and generates answers based on previously learned data. It is called generative AI because it generates answers or some tasks.

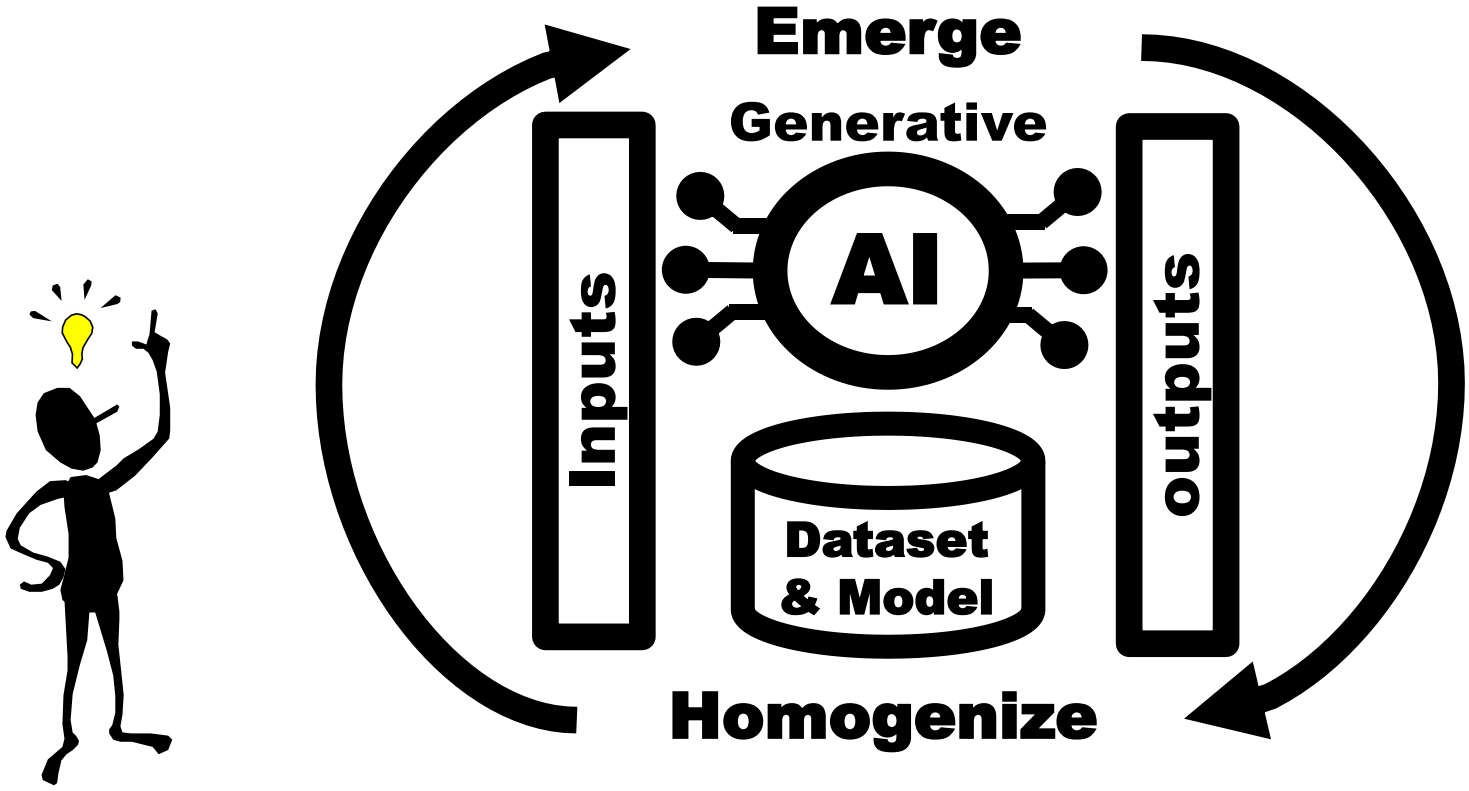


*Incl. using tech of not only RDB but 'Vector DB'

© Global Knowledge Link Center.

What's "GenAI"?3/5

- The individual dialogue itself repeats Learn-Generate, but overall it is a concept of Emerge and Homogenize.
- This can be said about AI in general, but what is noteworthy is that once power and overfitting are ignored, it is possible to continue works almost constantly without getting tired.



© Global Knowledge Link Center.

What’s “GenAI”?4/5

- Ke Jie, the world's strongest Go player who was defeated by AlphaGo, left an interesting comment: ``It's completely different when playing against a human." After the match, he reflected, ``Last year, I thought AlphaGo's playing style was extremely human-like, but today's playing style was like that of a Go god.‘’
- Although AlphaGo is not a generative AI, it is significantly different from conventional AI systems in that it is very interesting that they are using deep reinforcement learning that combines main three algorithms.

AlphaGo

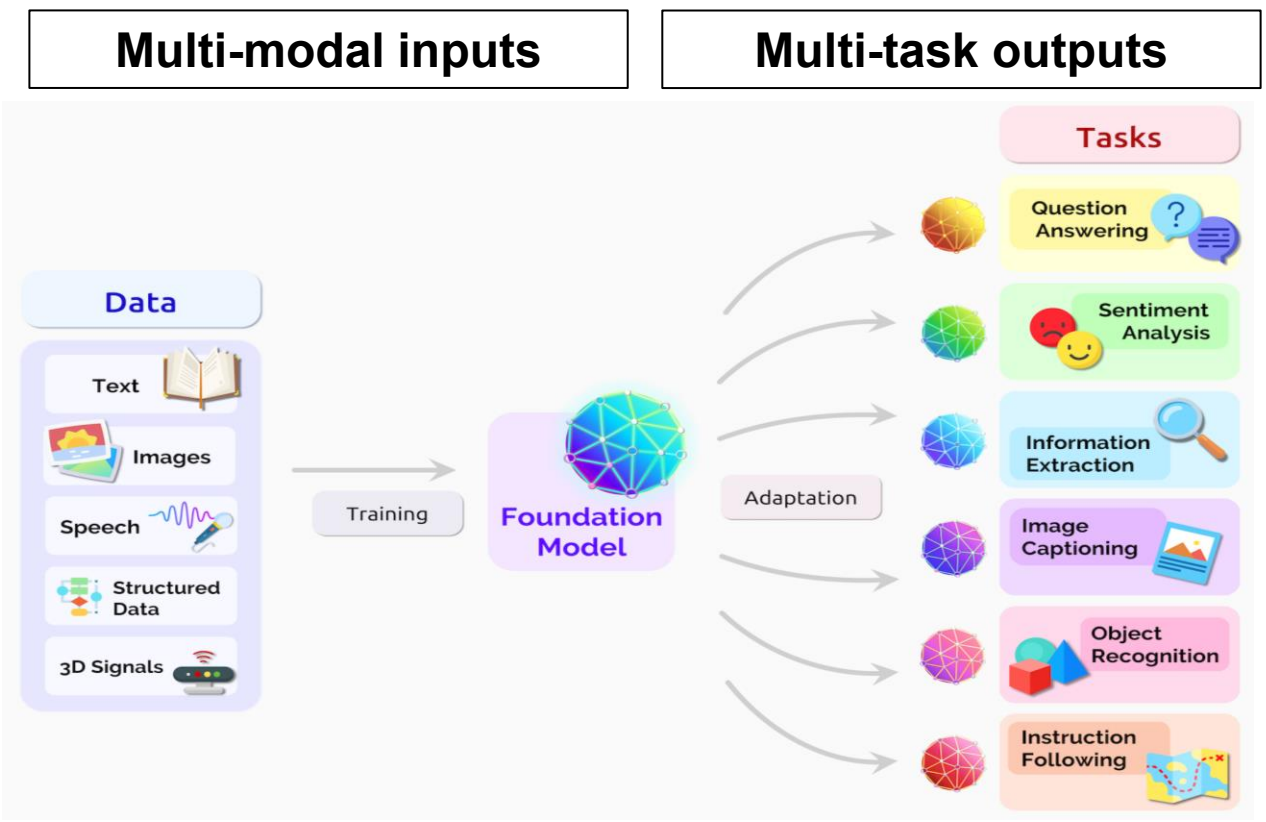
Pics : DeepMind, Google

Reference : https://10mtv.jp/pc/content/detail.php?movie_id=1372、https://www.sbbit.jp/article/cont1/58531

© Global Knowledge Link Center.

What’s “GenAI”?5/5

- The current mainstream (2024) will be seemed to be “Foundation model” to output multitasks generated from multimodal inputs.
- However, this does not mean that the models that have been built through machine learning and deep learning or so will no longer be used. These coexist and the right models are used in the right places.



Pics : https://arxiv.org/pdf/2108.07258.pdf Fig.2.

© Global Knowledge Link Center.

Words）What’s “Multi modal AI”?

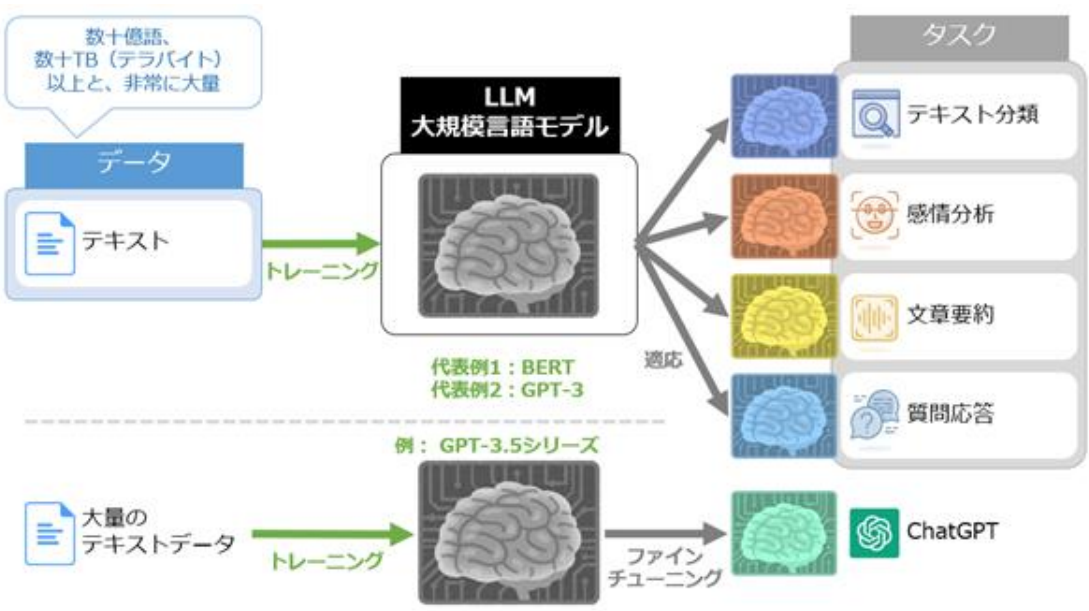
- Currently (as of 2024), there is a trend to advocate multimodal AI that can support multimodal input and output methods.

Pics : https://atmarkit.itmedia.co.jp/ait/articles/2207/04/news016.html

© Global Knowledge Link Center.

Reference)What’s LLM?

- There is no clear standard for ``large'' in LLM, but one standard is 100 billion parameters.
- BERT and GPT-3 are shown as reference values. BERT is trained on 2.8 billion words of Wikipedia data and 800 million words of Google BookCorpus data for a total of 3.3 billion words. GPT-3 is trained from 45TB (terabytes) of data (499 billion tokens in total). Data sets after GPT-3.5 have not been released, but the number of parameters in GPT-4 is estimated to be between 500 billion and 1 trillion based on its performance.

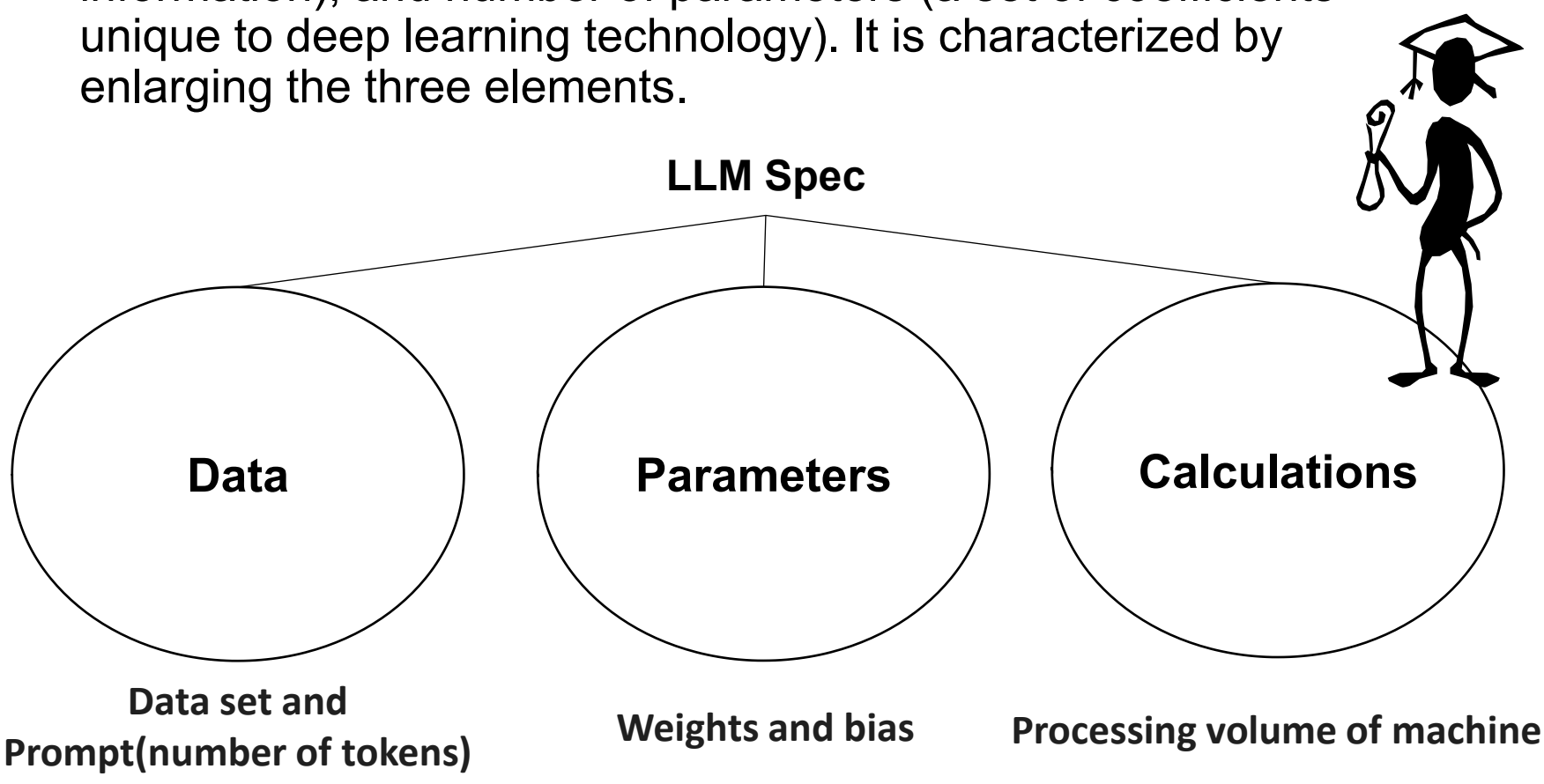


Pics : https://atmarkit.itmedia.co.jp/ait/articles/2303/13/news013.html

© Global Knowledge Link Center.

Reference）What LLM spec depend on?
1/2

- Compared to conventional natural language models, large-scale language models have a larger amount of calculations (amount of work processed by the computer), data amount (amount of input information), and number of parameters (a set of coefficients unique to deep learning technology). It is characterized by enlarging the three elements.

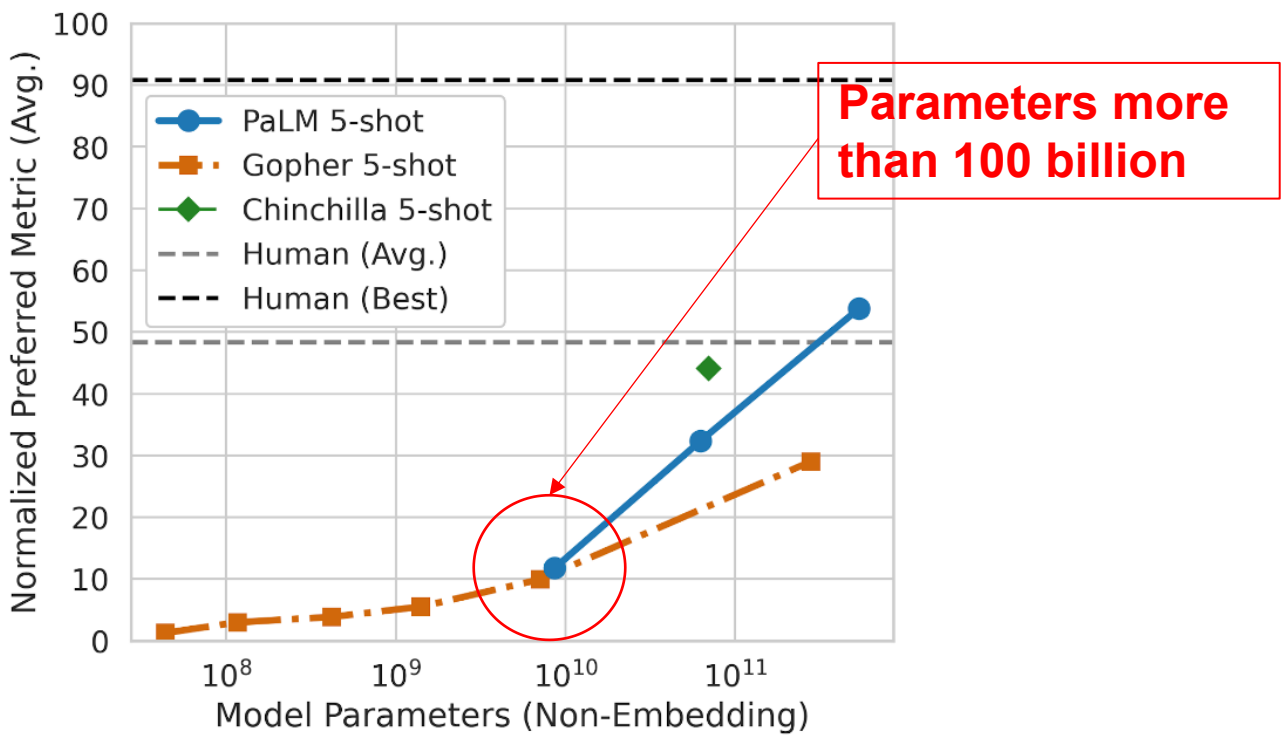


Source ： https://aismiley.co.jp/ai_news/what-is-large-language-models/

© Global Knowledge Link Center.

Reference) What LLM spec depend on?
2/2

- Scaling law indicates a power law of three variables. And these shows a kind of “Critical point” of data.
  - Number of training steps
  - Dataset size
  - Number of parameters



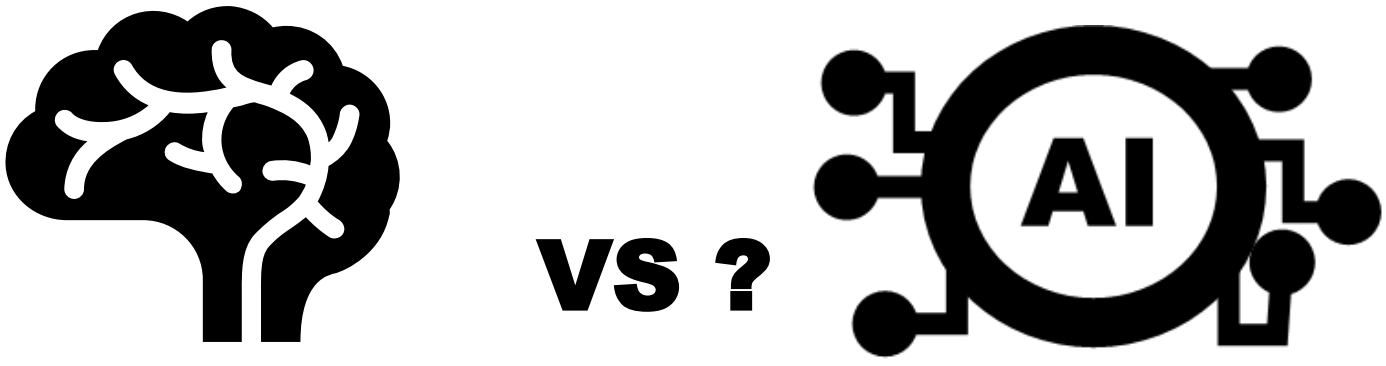
Pics ： Google AI Blog

Reference : Scaling Laws for Neural Language Models"

© Global Knowledge Link Center.

Reference)How much human brain spec is?

- It is said that the human brain has a capacity of 150 TB, a memory capacity of 17.5 TB, and a total of 100 trillion synapses.
- Is it a lie that “humans only use 10% of their brains”?
- Just as you don't use all your muscles at once, you don't use all your brain at once. Brain scan data proves that we use all parts of our brains over several days. In other words, it can be said that they are used differently depending on the need..

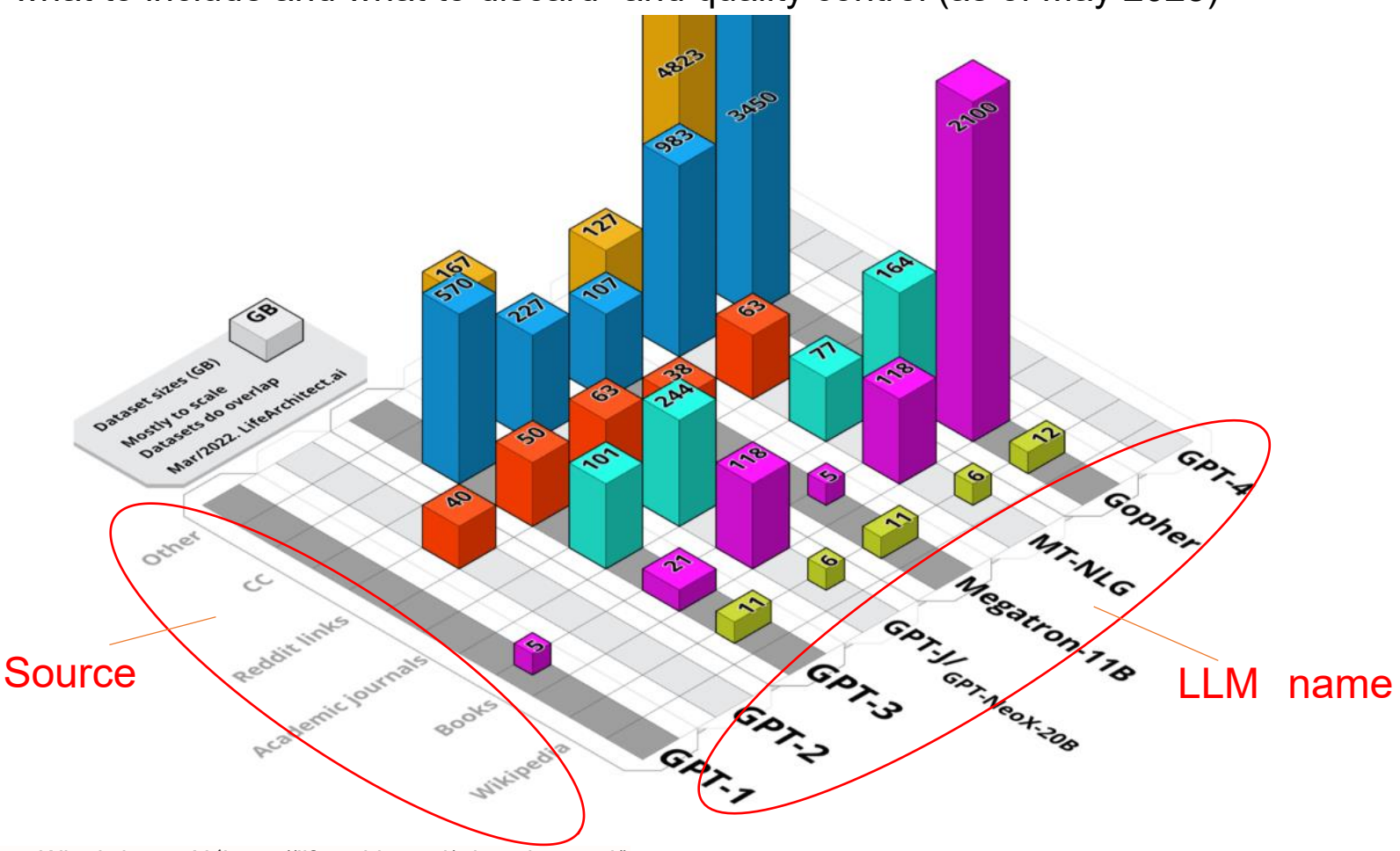


Reference : https://logmi.jp/business/articles/108173

© Global Knowledge Link Center.

Reference) Source of LLM information and knowledge

- In the case of GPT3 in this chart, Common Crawl, Reddit Links , and Academic journals account for over 95% of the total.
- Since GPT4, data collection has become much more sophisticated in terms of "what to include and what to discard" and quality control (as of May 2025)



Source and Pics: What’s in my AI（https://lifearchitect.ai/whats-in-my-ai/）

© Global Knowledge Link Center.

AI history -When Singularity will be near future, in 3 to 5 years?

- In around 1950, the current basic idea has already been developed that "Neural network" is one of the important factors.

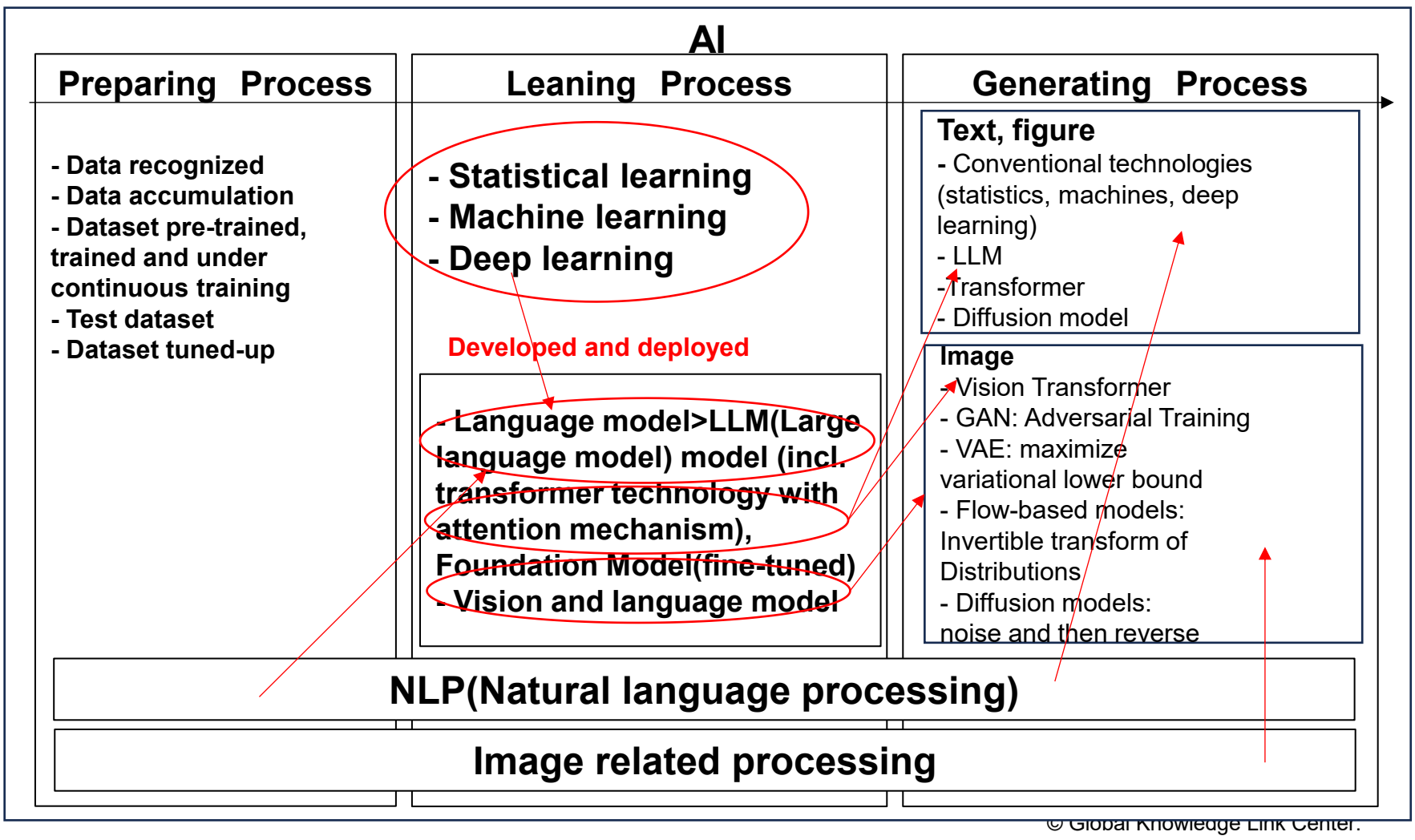
	1950	1960	1970	1980	1990	2000	2010	2020
Phase	First boom	Winter period		Second boom	Winter period		Third boom to Forth boom(2022?-)	
Tech trend	- Exploration/inference - Neuro Linguistic Programming - Neural network - Genetic algorithm - Expert system			- Knowledge based - Voice recognition - Data mining - Ontology - Statistical Neuro Linguistic Programming - VSM(Vector space Model)			- Machine learning - Supervised / unsupervised learning - Reinforcement learning - Neural network - Deep learning - LLM (Large Language Model) - GenAI boom	
Remarks	Proposal of the Turing test (1950) The term "artificial intelligence" was proposed at the Dartmouth Conference (1956) Neural network perceptron development (1958) Development of population dialogue system "ELZA" (1964) Developed the first expert system MYCIN (1972)			Development of EMYCIN, which generalized MYCIN's knowledge representation and reasoning(1979) 5th Generation Computer Project (1982-92) Psych project for knowledge description started (1984) Presentation of error backpropagation method (1984) IBM's DeepBule (chess) wins world championship (1996-1997) Proposal of deep learning (2006) Applying deep learning technology to image recognition contest (2012) DeepMind's AlphaGo (Go) wins world champion (2012-2015) Deep learning implementation is widespread (Google translation) (2016)				

Source : Processed by GKLC based on MIC (Ministry of internal affairs and communications)

© Global Knowledge Link Center.

Reference)How we integrate multiple?

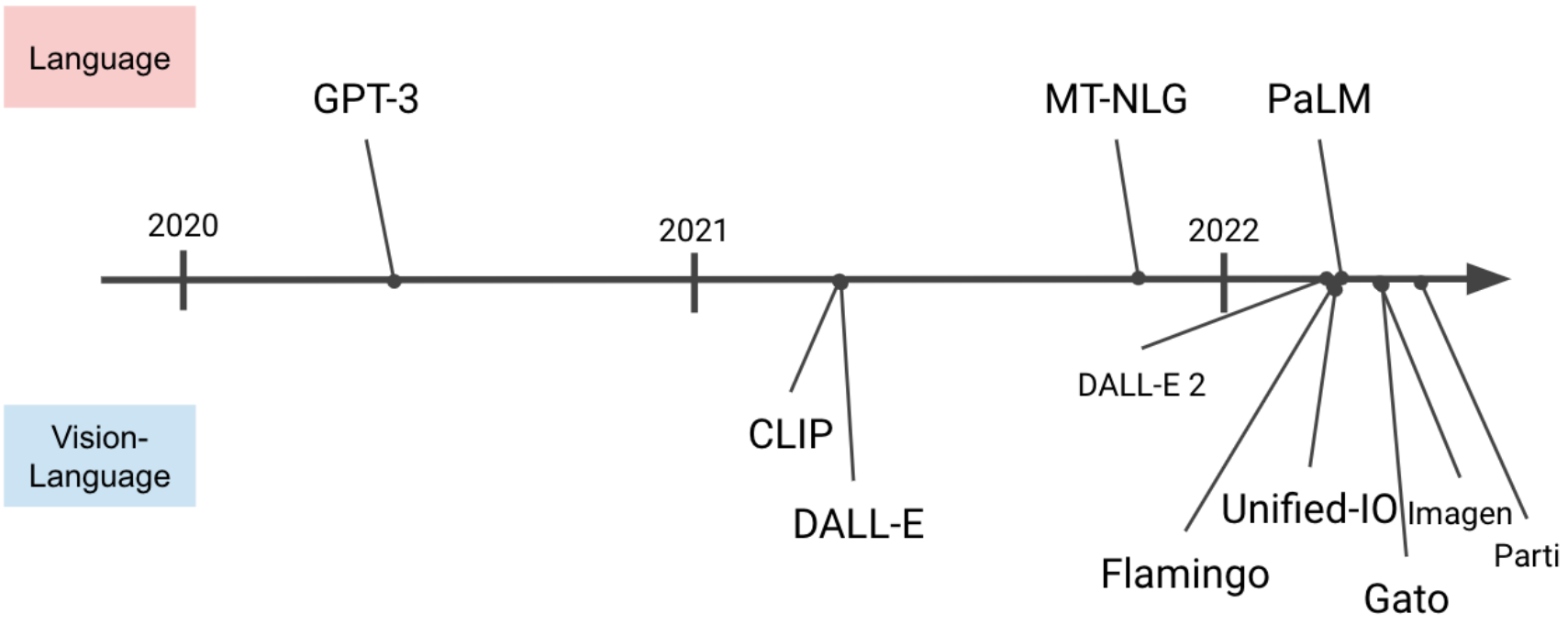
- Process of GenAI functionally are integrated , multiple concept, methods , models and technologies at each process. These include various learning methods and models, NLP, image related processing, and others.



© Global Knowledge Link Center.

Reference) Language model and Vison-language model

- Fundamental models can be roughly divided into those that deal with "language" and those that deal with "vision and language." Vision is a word that refers to both images and videos.



Pics : https://blog.recruit.co.jp/data/articles/foundation_models/

© Global Knowledge Link Center.

Reference) Inference of “Transformer”

- This is a black pen, which your father bought…at my birthday.
- The figure below shows the probability that each word will be placed in that order. In the process of analyzing this sentence, the entire sentence is recognized and the order of the words is determined.
- **Besides, Transformer with attention architecture focuses on the importance and relevance of the text**

This is a black pen, which your father bought…at my birthday...

Analyze the appropriate order of probabilities

Word order	a	black	...	...	is	...	pen	this	...	which	your
1st word	0.05	0.01	...	...	0.01	...	0.01	0.8	...	...	...
2nd word	0.01	0.1	...	...	0.8	...	0.1	0.02	...	...	...
3rd word	0.75	0.03	...	...	0.05	...	0.05	0.01	...	...	...
4th word	0.03	0.6	...	...	0.02	...	0.02	0.05	...	...	...
5th word	0.02	0.02	...	...	0.01	...	0.75	0.01	...	...	...

Reference : Attention Is All You Need

© Global Knowledge Link Center.

Reference)AI product, service and model history 1/6

Refer to Excel

- Many early AI services served as supporting hardware.
- Natural language processing, in particular, as the accuracy of voice recognition improves, some people often uses this service.
- Although a certain number of them have become popular, they are used in daily life rather than business scenes.

Updating

Product/ Service	Type	Outline	Development	Released year
Siri	Voice recognition/virtual assistant app on cell phone	A voice assistant that works on Apple's iOS and Mac devices cell phone based	Apple	2010
Amazon Alexa	Voice recognition/virtual assistant Smart speaker	Cloud-based voice assistant integrated into Amazon Echo devices	Amazon	2014
Google Assistant	Voice recognition/virtual assistant app on cell phone or smart home equipment	Voice assistant for Android devices, Google Home, and other devices Cell phone based	Google	2016
Microsoft Cortana	Voice recognition/virtual assistant app on PC	Voice assistant integrated into Windows devices and Microsoft 365	Microsoft	2014
Bixby	Voice recognition/virtual assistant app on cell phone	Voice assistant and supportintg image analysis for Samsung smart phones and consumer electronics devices	Samsung	2017

Source ： processed by GKLC based on each company site and organization site, Wikipedia, Gigazine, other press, some parts of ChatGPT and Copilot

© Global Knowledge Link Center.

Reference) AI product, service and model history 2/6

Refer to Excel

- With the improved accuracy of GenAI (LLM, Transformer and attention layer or so), the credibility of the fact that AI can think in the same way as humans has suddenly increased. It is also true that this has been proven in some cases.

Updating

AI service or model name (mainly GenAI)	Outline	Development	Parameters (incl. Estimated)	Released year
IBM Watsonx, Granite	- Has knowledge of 115 programming languages learned with 1.5 trillion tokens. - He is said to be one of the largest basic models for code generation.	IBM	13 billion to 20 billion	2023
GPT-4	- A model that uses GPT-3 to learn data other than text, such as images and audio. - Compared to GPT-3.5, GPT-4.0 has significantly improved the accuracy of language understanding by expanding the number of parameters and amount of training data by several dozen times.	OpenAI	100 trillion	2023
PaLM2（Pathways Language Model）	- This is an LLM used within Bard and is provided on GCP(Google Cloud Platform). - He has studied over 100 languages, including Japanese, and over programming languages.	Google	540 billion	2023
LLaMA（Large language Model Meta AI）(2, 3)	- It is a generation AI developed by Meta, and is said to achieve performance equivalent to GPT-3 with an overwhelmingly fewer number of parameters. - Llama and Llama 2, the number of parameters was 7 billion to 65 billion, but the number of parameters is more dramatic than in 3. - Llama 3 can be directly called up and used from SNS such as the conversation app ``Messenger'' and the image sharing app ``Instagram.'' - The technology will be made available to outsiders free of charge to compete with Open AI in the US and Google, which are leading the way in generative AI.	Meta	400 billion	2023
Claude (2, 3)	- This Model by engineers involved in the development of GPT-2 and GPT-3. - The development concept is for Claude, a friendly and hard-working colleague, to serve as a "virtual teammate" in companies to improve productivity and facilitate communication. - From an ethical perspective, it is unique that the system is equipped with a principle called ``Constitution AI.'' - Maxim Lott(TV producer) was given an IQ test and scored an IQ of 101. This made Claude 3 the first publicly available AI with an IQ of over 100 (2024)	Anthropic	Over 130 billion	2023

© Global Knowledge Link Center.

Reference) AI product, service and model history 3/6

Refer to Excel

AI service or model name (mainly GenAI)	Outline	Development	Parameters (incl. Estimated)	Released year
Alpaca 7B	- A model based on LLaMA and fine-tuned using the results of instruction-following. - He allows you to easily and cheaply reproduce behavior similar to GPT-3.5 in a much smaller environment.	Stanford university	7 billion	2023
Vicuna 13B	- Vicuna-13B, an open-source chatbot trained by fine-tuning LLaMA on user-shared conversations collected from ShareGPT. - Preliminary evaluation using GPT-4 as a judge shows Vicuna-13B achieves more than 90%* quality of OpenAI ChatGPT or so. - The cost of training Vicuna-13B is around $300. The code and weights, along with an online demo, are publicly available for non-commercial use.	University of California	13 billion	2023
OpenFlamingo	- OpenFlamingo is an open source reproduction of DeepMind's Flamingo model. - It features Flamingo-style LMM training capabilities, large multimodal datasets of alternating images and text, in-context learning assessment benchmarks for visual language tasks, and more.	LAION	3 billion to 9 billion	2023
OpenCALM	- LLM specialized in Japanese that is pre-trained from scratch using Japanese datasets. - The context length is 2048, which is the same as GPT-NeoX (a model with a structure similar to GPT-3, released as open source by EleutherAI).	Cycber agent	6.8 billion	2023
PaLM（Pathways Language Model）	- Model based on the academic paper “Scaling Language Modeling with Pathways”. - Significantly expands the number of Transformer parameters and achieves high performance.	Google	540 billion	2022
NEMO LLM	- NEMO LLM runs on NVIDIA's advanced AI platform, it is one step ahead of other models in computational speed and applicability in a wide variety of applications. - This model is highly platform dependent and outperforms GPT4 and Bert in terms of training speed and scalability.	NVIDIA	Automatic adjustment tool included	2022
Gopher	- Google subsidiary DeepMind announced Gopher, an AI natural language processing (NLP) model with 280 billion parameters. - Based on the Transformer architecture and trained on a 10.5TB corpus called MassiveText, Gopher surpasses the current record in 100 out of 124 evaluation tasks.	Deepmind, Google	280 billion	2022
LaMDA（Language Model for Dialogue Applications）	- "LaMDA" is a large-scale language model (LLM) developed by Google. The name is an acronym for "Language Models for Dialogue Applications." - This model based on Transformer and specialized for dialogue (pre-trained with a text corpus of 1.56 trillion words)	Google	137 billion	2021

Updating

© Global Knowledge Link Center.

Reference) AI product, service and model history 4/6

Refer to Excel

AI service or model name (mainly GenAI)	Outline	Development	Parameters (incl. Estimated)	Released year
GPT-3（Generative Pretrained Transformer）	- GPT-3 has announced language models GPT (2018) and GPT-2 (2019) that use a deep learning method called Transformer, and GPT-3 is the successor language model. - More than 300 applications using the GPT-3 API have been developed, generating an average of 4.5 billion words per day(2021). - Its outstanding performance was recognized in common, and it played a role in the LLM and generative AI trends.	OpenAI	175 billion	2020
BLOOM	- BLOOM is an autoregressive Large Language Model (LLM), trained to continue text from a prompt on vast amounts of text data using industrial-scale computational resources. - As such, it is able to output coherent text in 46 languages and 13 programming languages that is hardly distinguishable from text written by humans. - BLOOM is not an LLM developed exclusively by a specific company, but was developed by AI scientists around the world with the purpose of promoting AI research (the project is called "BigScience", and the development of BLOOM includes scientists from over 70 countries and over 200 institutions participated)	BigScience project	175 billion	2022
ELYZA Brain	- ELYZA has recently released a demo of "ELYZA-japanese-Llama-2-70b", a newly developed large-scale language model (LLM) with 70 billion parameters. - "ELYZA-japanese-Llama-2-70b" is the result of a project to expand the Japanese language ability of Meta's "Llama 2" series, which is a continuation of his excellent English language skills. It's a thing.	ELYZA,Tokyo university	7billion to 13billion	2020
BERT	- BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model for natural language processing developed by Google. BERT can perform language comprehension tasks and can understand the meaning and context of sentences. - BERT is a language model proposed by researchers at Google and the University of Toronto, and uses a Transformer architecture that processes sequential data in batches using a self-attention mechanism. - Early language models that increased dataset size and accuracy.	Google	100 million to 300 million	2018
Gemini, Bard	- Bard is an artificially intelligent chatbot developed by Google, later renamed Gemini. - Bard uses a lightweight version of a large-scale language model (LLM) called LaMDA developed by Google and some features of PaLM. - Gemini is developed as a multimodal generative AI model which can take text, images, audio, and video as input and generate text and images such as ChatGPT.	Deepmind, Google	1.56 trillion	2024

Updating

© Global Knowledge Link Center.

Reference) AI product, service and model history 5/6

Refer to Excel

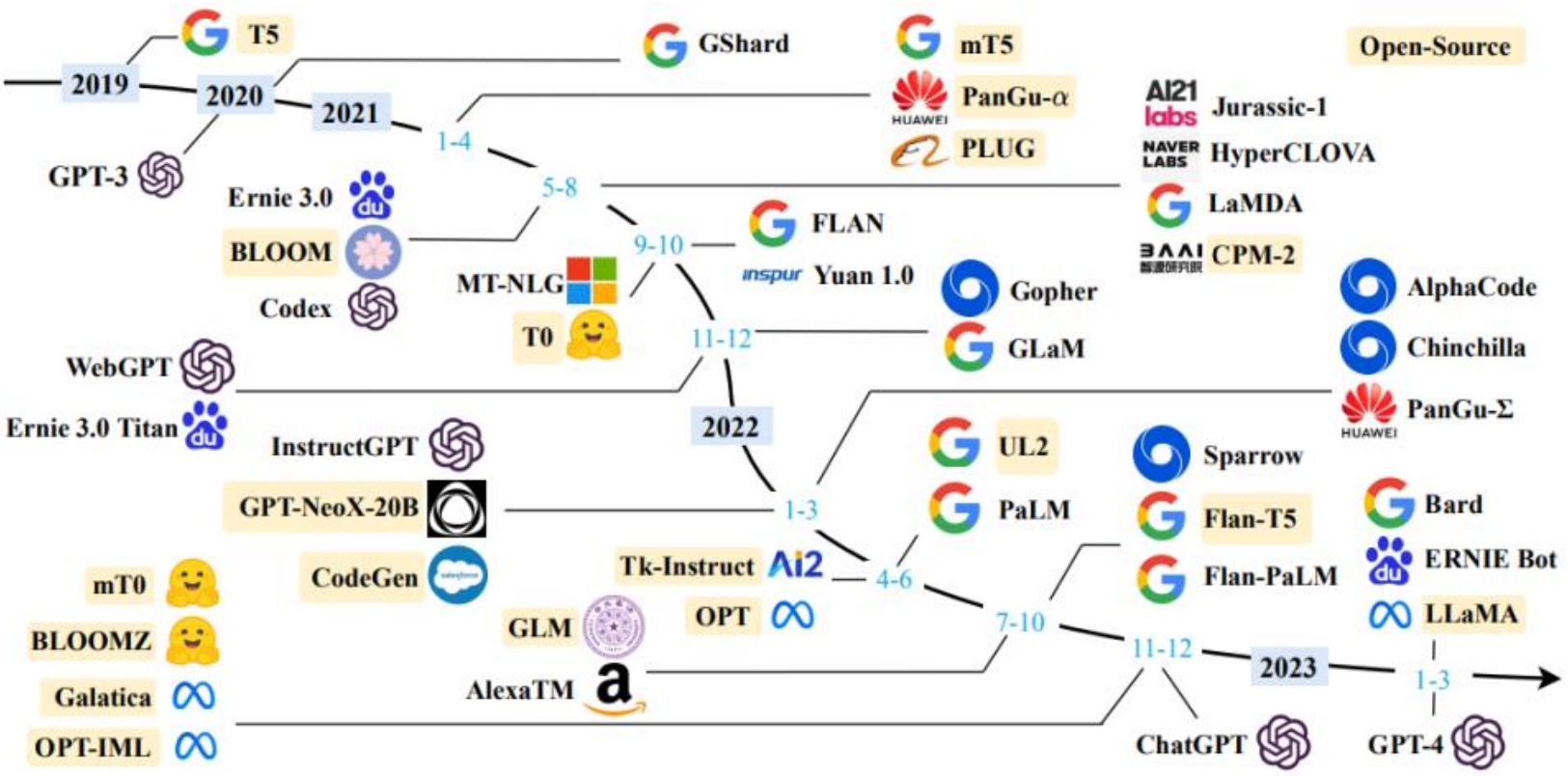
AI service or model name (mainly GenAI)	Outline	Development	Parameters (incl. Estimated)	Released year
LLaVa	- Microsoft, the University of Wisconsin-Madison, and Columbia University have released the Large Language and Vision Assistant (LLaVA) as open source. - LLaVa has the architecture consisting of Vision Transformer + Llama2 - LLaVA is based on the CLIP image encoder and LLaMA language decoder, fine-tuned on the instruction execution dataset and achieving state-of-the-art accuracy on the ScienceQA benchmark.	Microsoft, University of Wisconsin-Madison, and Columbia University	13 billion	2023
BaKLLaVa-1	- BaKLLaVa-1 has the architecture consisting of Vision Transformer + Mistral. - Mistral 7B is Mistral AI's first foundation model, supporting English text generation tasks with natural coding capabilities. - Optimized for low latency with low memory requirements and high throughput for its size.	Skunkworks AI, Mistral AI	7.3 billion to 12.9 billion	2023
fuyu-8b	- Fuyu-8B is the text and image transformer model with diverse modes developed by Adept AI. - It is also optimized for agents, supporting arbitrary image resolutions, answering questions about graphs and diagrams, responding to UI-based questions, and providing fine-grained localization of screen images.	Adept AI(Ex-transformer developper's group)	8 billion	2023
Copilot (Bing AI, Bing chat)	- Copilot is the foundation of Microsoft's AI transformation and delivers AI capabilities for a variety of products. As the meaning of Copilot suggests, it is positioned as something that supports users and products with AI. - Microsoft has invested in OpenAI and has a strategic partnership with it, so the core engine is based on ChatGPT.	Microsoft	Same with ChatGPT updated	2023

Updating

There are too many to list, but they are still evolving and being developed…

© Global Knowledge Link Center.

Reference)AI product and service history 6/6



Pics : :https://arxiv.org/pdf/2303.18223.pdf

© Global Knowledge Link Center.

Reference) Use cases

- Check it out !! [By industry] 10 examples of companies using generative AI

- Electrical parts manufacturer: Introduced AI assistant and used 5,000 times a day
- Precision equipment manufacturer: Developing a robot that moves based on verbal instructions using generative AI
- Retailer(CVS, GMS) or so: Utilizing generative AI to reduce product planning time to one-tenth
- Retailer (department store) : Advertising videos, narration, and music are all created using generation AI
- Beer Breweries: Utilizes generative AI to streamline employee searches for internal information
- SNS carrier : Engineers use generated AI to improve work efficiency by 2 hours a day
- Auction site operator : AI assistant suggests product names and descriptions that are easy to sell
- Education service provider : Providing individual advice using generative AI
- Construction company : Generative AI proposes multiple design ideas for a building based on sketches
- Financial service group: Improve employee productivity by developing unique conversational AI

Reference : https://metaversesouken.com/ai/generative_ai/difference/

© Global Knowledge Link Center.

AI transition and Issues summary

- The challenges we have faced with AI to date have come with the times and development methods, but the fundamentals remain the same. In other words, it's that deep-rooted
- In any case, efforts are needed to overcome the challenges while leveraging the benefits.

When Singularity?

Post-LLM
- High quality results even with limited computer processing
- Maintaining infrastructure with an ecosystem in mind

Foundation Model(LLM)
- One base model can be applied to multiple tasks
- Significantly reduces manual labeling

Deep Learning
- Improved prediction accuracy
- No need for manual feature design

Machine Learning
- Logic rules are generated from data (rather than manually)

Rule based system
- Highly productive processes are executed based on logic rules
- No mistakes

Common Issues Keywords summary

The time it takes to build a system, the speed of technological progress, training and securing engineers, prediction accuracy (in some cases), black box issues, general AI ethics (discrepancy between peaceful use and human intentions), Scale calculation,(large-scale to extremely large-scale), Results of hallucination, Weakness towards specific issue accelerating the digital divide, securing and occupying resources (electricity, funds, GPGPU), etc.

© Global Knowledge Link Center.

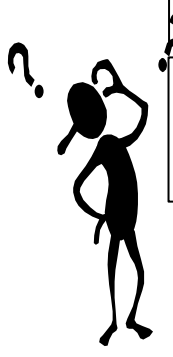
GenAI Issues

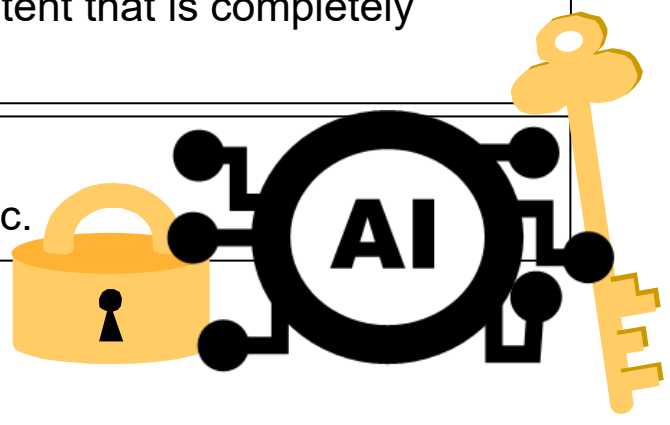
- There are around 3 issues so far (Point of 2024) . In the first place, many things are new, so it doesn't matter if you don't have experience or that they may be difficult.

Lack of AI literacy, so and then all output accuracy depends on the models, sometimes all process are under no control
- It takes time to combine which services to use (often it is not possible to judge what's appropriate means right away)
- Each model has pros and cons, so it is not possible to say which one is better. (Currently, judgment is made based on pure spec, large-sized calculation speed, number of parameters, and track record or so)

Not "Hallucination", but counter measure to "Hallucination"
In natural language processing, GenAI tells a plausible lie, a phenomenon in which it outputs information that is different from the truth or content that is completely unrelated to the context.

Inadequate protection against hostile prompts
There are prompt injections, prompt leaks, jailbreaks, etc.



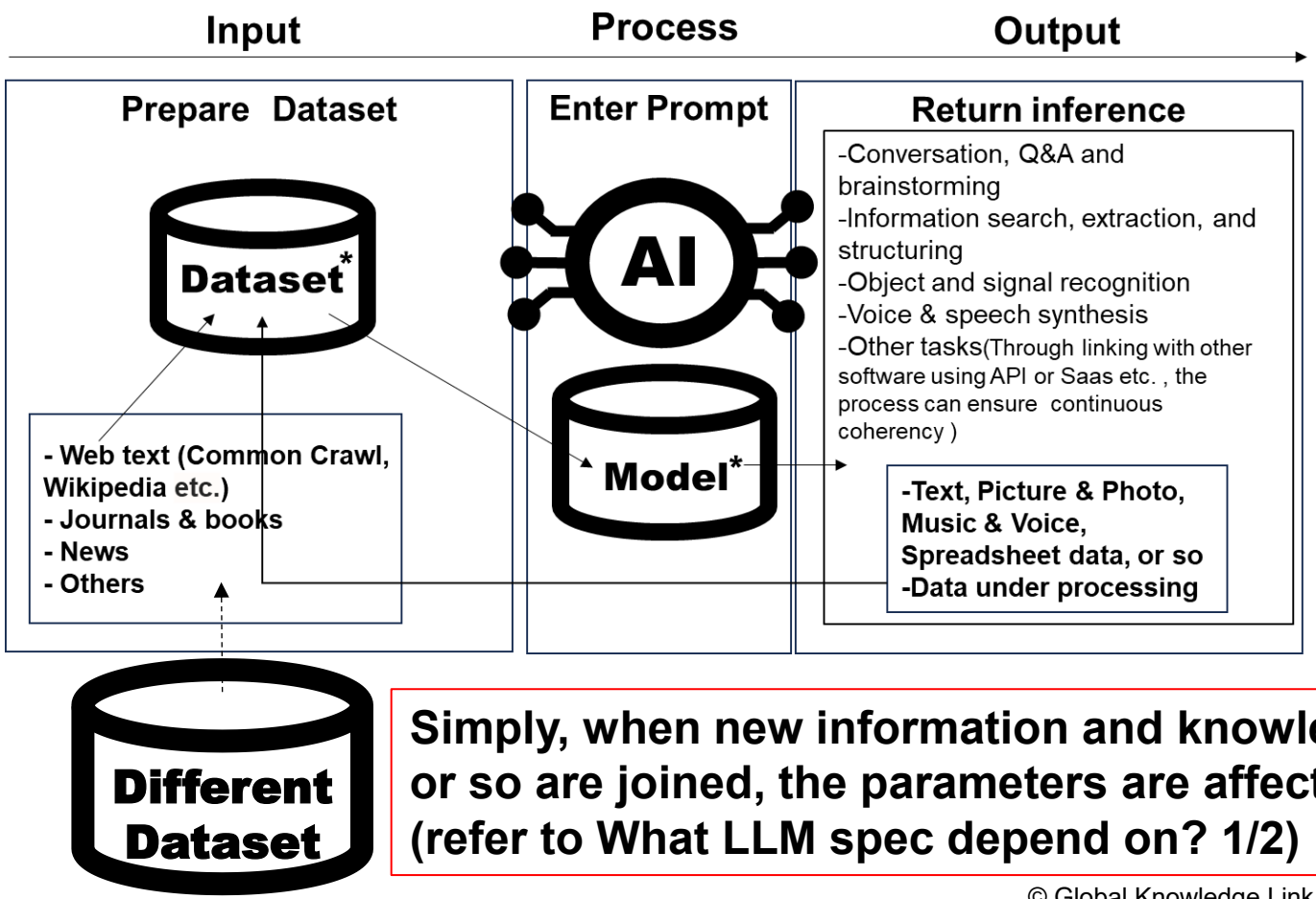


Reference : https://aismiley.co.jp/ai_news/what-is-large-language-models/

© Global Knowledge Link Center.

Reference)One of solution is fine-tuning?

- Fine-tuning means literally to tune up a model more appropriate to new tasks. Concretely speaking. though retraining part or all of a pre-trained model on a dataset using a different dataset, we add some change to the parameters of a machine learning model.



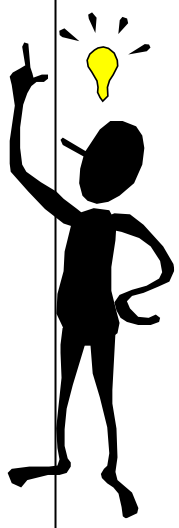
Simply, when new information and knowledge or so are joined, the parameters are affected. (refer to What LLM spec depend on? 1/2)

© Global Knowledge Link Center.

Prompt engineering on Gen AI

- 10 key lessons leaned
  - There are knowledge about many prompt libraries and input methods, so please refer to them. The important point is to tune up the model by accumulating the results of their use. Good quality data and prompts makes it compact and more efficient.

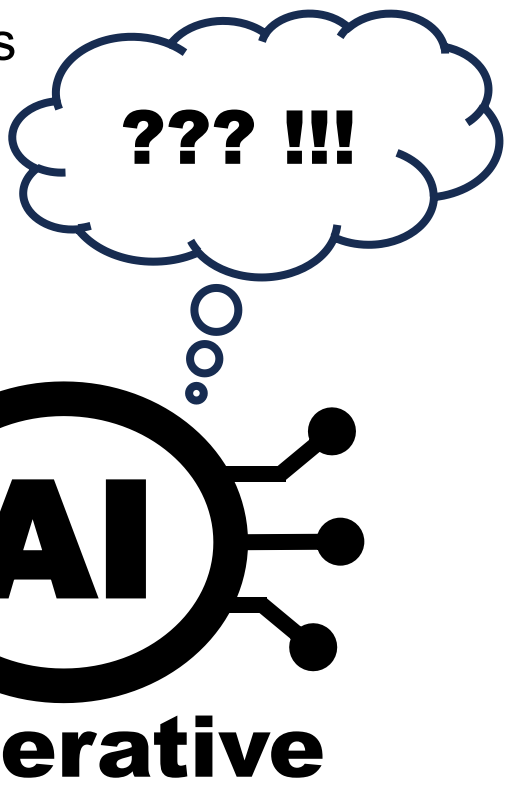
- Operation with ingenuity
- Simple and clear question
- Concreteness, proper samples and details
context, thread and background information or so
- Prompt format
- Trial to use regenerate function
Turning GenAI's fault back on them (with different answers to one same prompt)
- Logical step
step by step approach
- Link to other GenAI
self evaluation, verification of accuracy
- Automation relay (Automation of automation)
Constructure of GenAI that generates high-quality prompts.
- Accumulation of high-quality information, questions, and results
- Sublimation to Process automation



© Global Knowledge Link Center.

Prompt engineering on Gen AI

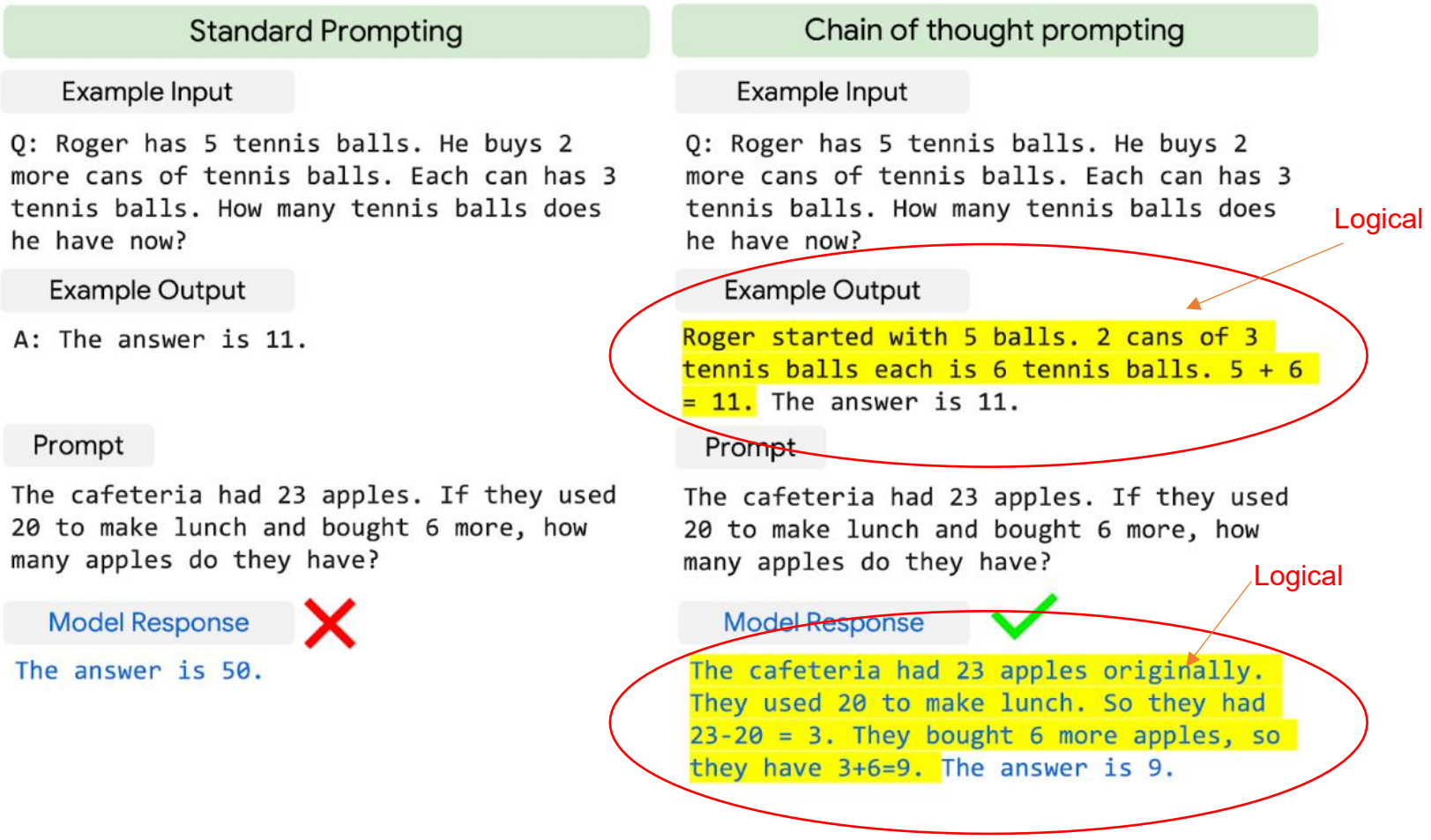
- Basic is easy and simple.
- # Command or instruction : Instructions and tasks performed by GenAI model
- # Conditions or requirements : Refer to next pages
- # Point : Refer to next pages



#Command 命令
#Reiuiements 条件
#Points ポイント

Prompt engineering on Gen AI

- Logic is clear and we'll acquire more correct answers.



Source : Processed by GKLC based on sample Google AI Blog

© Global Knowledge Link Center.

Reference）Prompt sample 1/2

- Structure sample
  - <Theme>, <Situation of the theme>, <Background of the theme>, <Style of drawing the theme>, <Adjustment word A>, <Adjustment word B>, … Example: portrait of cute dog (the element you want to emphasize most), White hair (subject situation), detailed eye brown(subject situation), In town at good weather(subject background), by Andy Warhol(illustration style), wallpaper(adjustment word A), trending on artstation(adjustment word B)), …

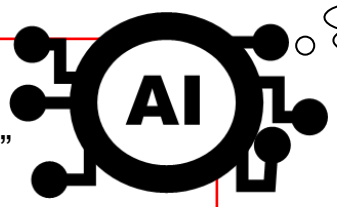
  - Prioritize the elements you want to emphasize as much as possible and place them in the front. It is better to eliminate unnecessary words as much as possible. Of course there are exceptions, but this form is stable and easy to convey the image.

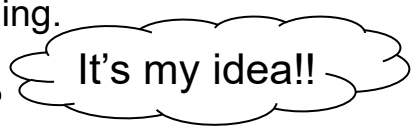
Source : https://note.com/matsur1/n/nb88d553ffbd6

© Global Knowledge Link Center.

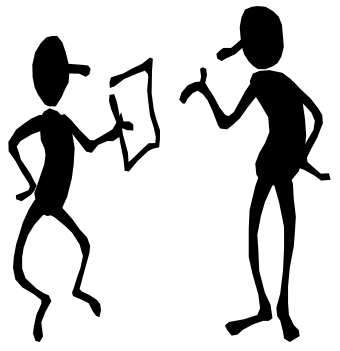
Reference）Prompt sample 2/2

- Instructions
  - Please suggest three titles for the website building seminar.
- Conditionals
  - Target: Working adults in their 20s to 30s who are interested in the web.
  - The number of characters in the title is around 20 characters
  - Easy to understand
  - Something that gives a good impression without being too business-like.
- Reference information
  - Additional information : Free of charge
  - Online event
  - You can learn hands-on seminar content from the basics of HTML to coding.





- Answer
  - 1. “Introduction to HTML that can be used on the web”
  - 2. “First Web Production: HTML Hands-on Seminar”
  - 3. “Web production starting with no experience: HTML basics”



© Global Knowledge Link Center.

Reference) GenAI also has pros and cons?

Refer to Excel

- When actually using BingAI and ChatGPT, ChatGPT seems to be more friendly because it was designed as that way. This is just a comparison between generative AI and traditional search engines. Not talking about Gemini or Bart etc, so please avoid any misunderstandings.
- BingAI also has advantages in terms of updating current information and accuracy of sources, so the correct way to use it at this point is to understand the good points of each and use them properly

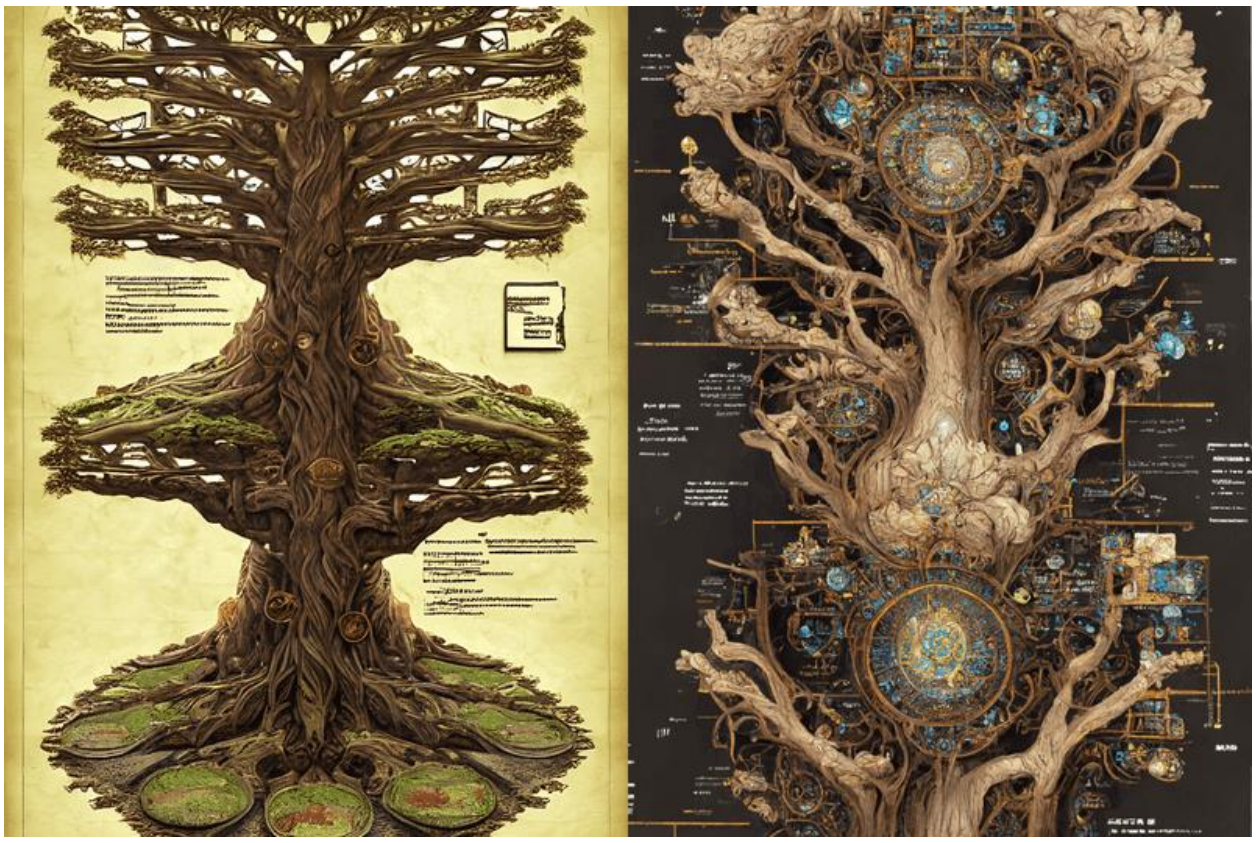
Heading	Copilot Bing AI	Chat GPT	Google search engine
Model features	Create answers based on real-time search information	Create answers based on pre-learned content	Prioritize display of relevant and reliable pages
Sentence generation	Emphasis on specifying the source (text generated based on search results)	Emphasis on the naturalness of conversations and sentences (reproduction of human conversation)	Excerpt from search results
Contextual consideration	Answer with context in mind	Answer with context in mind	Past search information is not reflected
Reproducibility of answers	Answer contents vary probabilistically	Answer contents vary probabilistically	Search results are the same
Source clearly stated	Available	Not available	Show list of source pages themselves
Update information	Available	Learning data and information at a certain point in the past (depending on the plan)	Although it is not supported, it is displayed as a result.
Specialty	Declarative knowledge - Latest information - Reliability of information (statement of source)	Procedural knowledge - Conversation and creation - Translation and summarization of text - Document proofreading - Programming	Selection and comparison of information sources - Access to specific sites - Zapping - Scrutiny and comparative examination of information
Weak field	Procedural knowledge	Latest information, sources and basis	Contextual understanding, abstract questions
purpose of use	Information gathering and knowledge acquisition	Problem solving and conversational responses	Selection and comparison of information sources

Source and Pics : https://alivevulnerable.com/gis/choose-chatgpt-bing-ai/

© Global Knowledge Link Center.

Reference)The difference with each GenAI

- Stable Diffusion on the left, mid-journey on the right

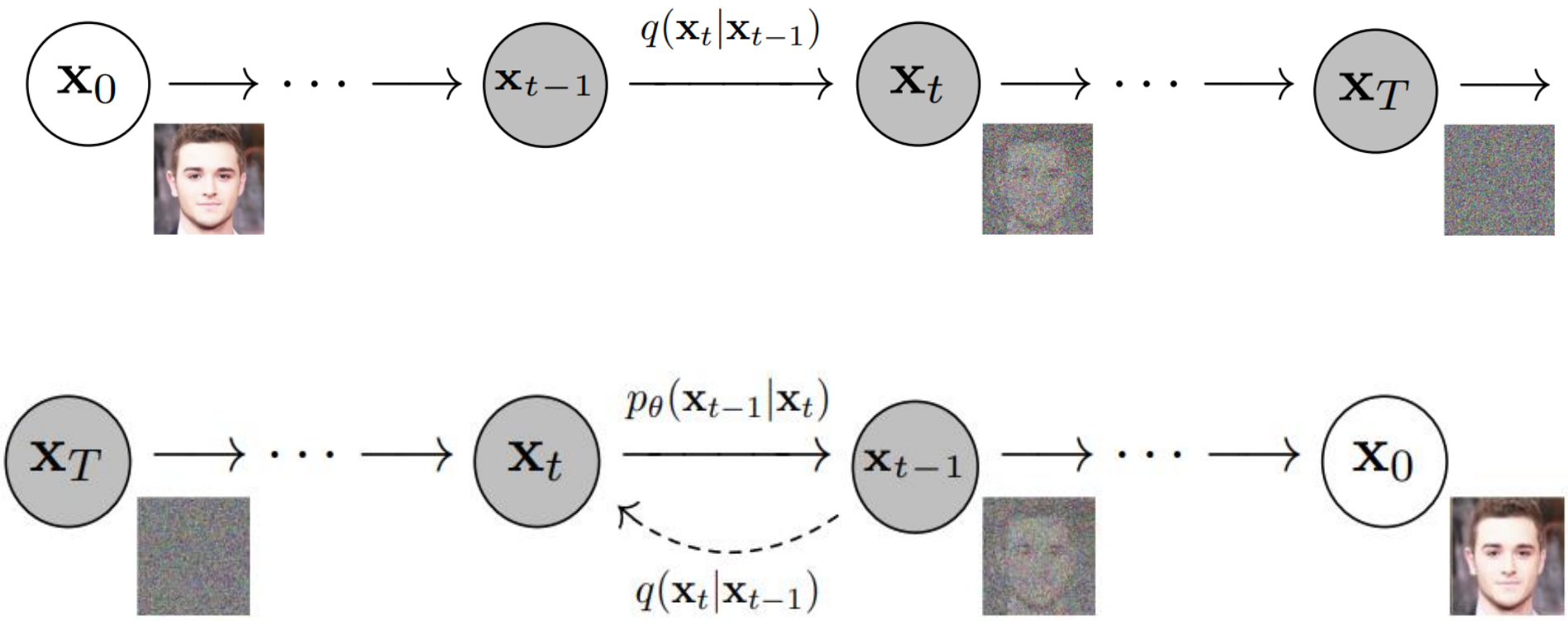


floor plan,Yggdrasil, Tree of Life, full page scan, parchment vellum, blueprint, mythological painting, insane focus, highly detailed, trending on artstation, high focus, artgerm, character concept, art print, 8k

Reference : https://note.com/matsur1/n/nb88d553ffbd6

© Global Knowledge Link Center.

Reference)Noise-Denoise proses in Diffusion model



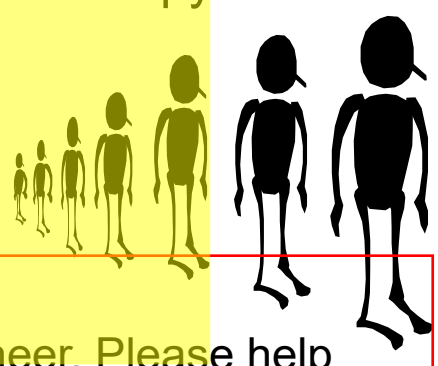
Ryan O'Connor, Introduction to Diffusion Models for Machine Learning
https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/

© Global Knowledge Link Center.

Prompt engineering on Gen AI
Concrete 10 steps toward Prompt Automation

- Use updated Model (as of 2024, ChatGPT-4 rather than 3.5)
- Let's think step by step.
- Use ### for separators,
  - <Topic1.>
  - ###
  - <Point1>
- Instructions are written at the top, related information is separated and written after the instructions.
- Make instructions concrete and concise
- Define output format and make it easy to understand and copy in your devise
- Conscious of omissions , "if not, or then" etc.
- Give some examples
- COT Chain of Thought and TOT Tree of Thought
- Prompt automation as Magic prompt
  - You are my prompt engineer. You are my prompt engineer. Please help me create it. First, please ask me a question and create a prompt based on it.
  - Ask me questions that will continually iterate and improve the prompts

Membership only



Reference : OpenAI recommended prompt explanation

© Global Knowledge Link Center.

Reference) COT and TOT

- COT Chain of Thought and TOT Tree of Thought

updating

Reference : https://arxiv.org/abs/2404.05221

© Global Knowledge Link Center.

Gen AI and other software (API, RAG, RPA or so)1/4

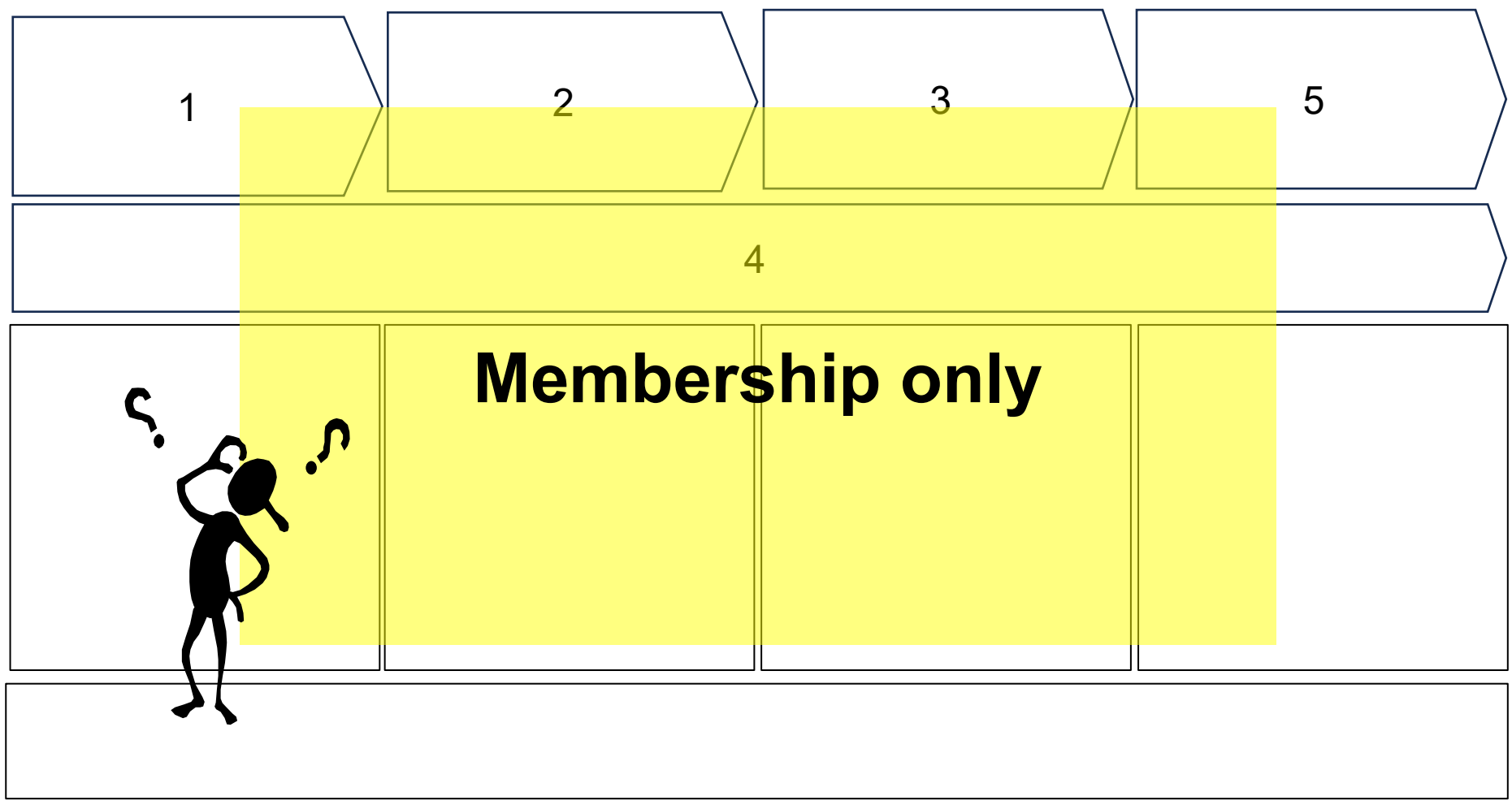
Refer to PPT

- Using GenAI with external databases and software can deliver unparalleled performance. Talking something a bit , tomorrow's presentation materials are being written.

Membership only

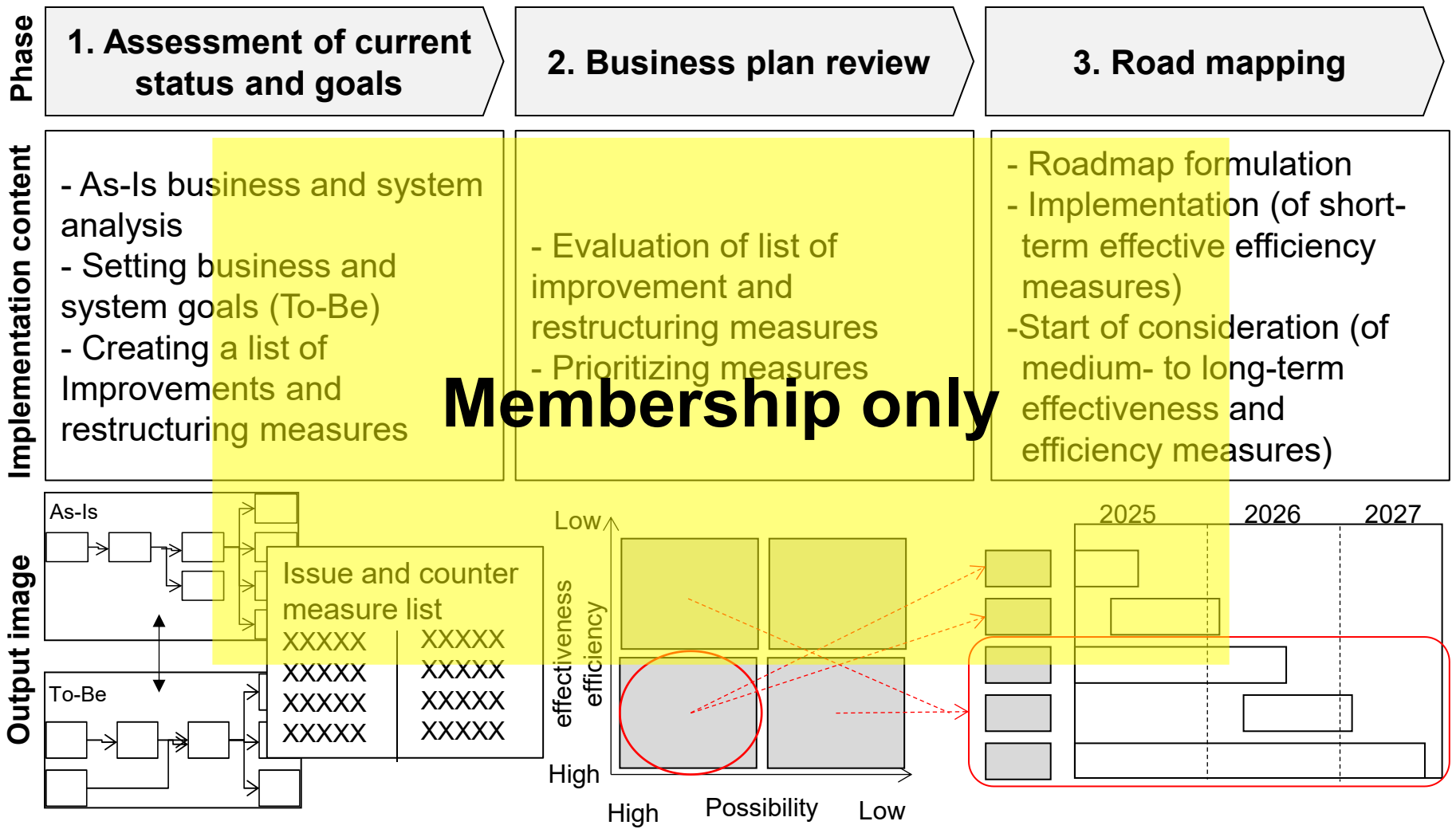
How to drive AI project?1/5

- When you initiate AI project, so what are you doing? Have the discussion and list some points in each phase.



Reference)Project approach in common

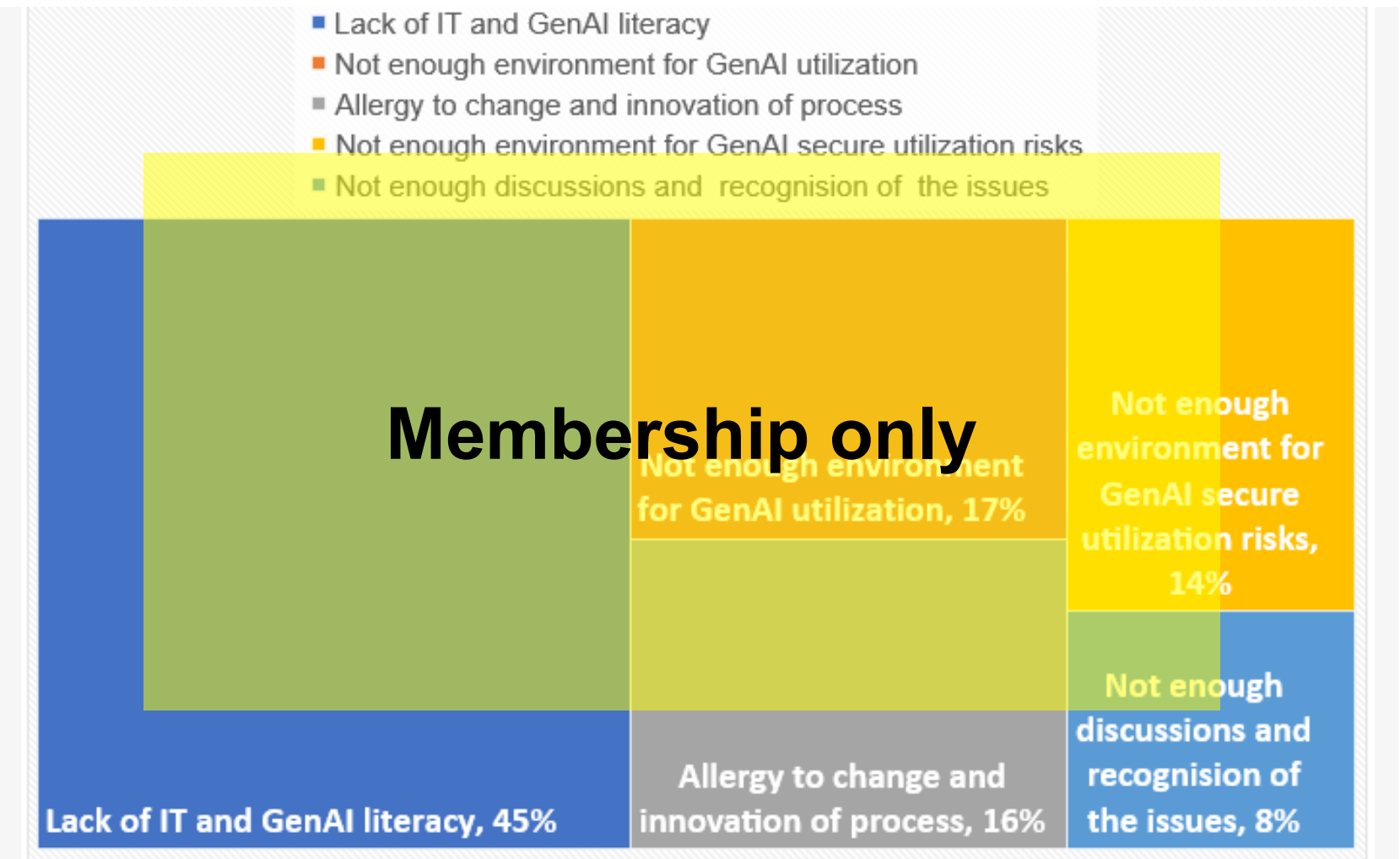
- Points are there, so let you imagine real project.



© Global Knowledge Link Center.

Reference)Why don't you use GenAI?

- This seems to be an essential and serious problem (Japan only)..



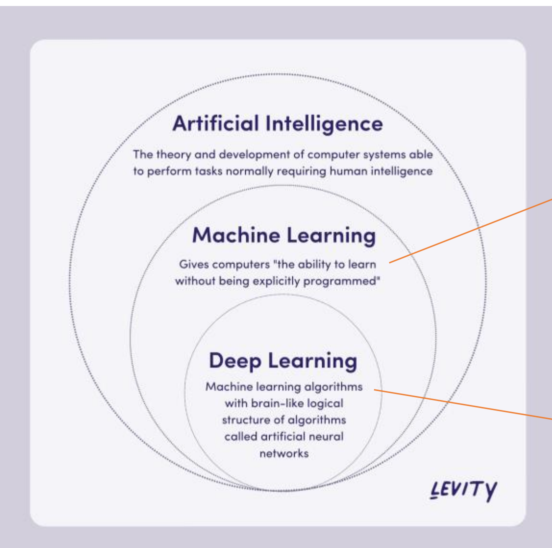
In terms of a comparison between Japan and the US, the ratio is around 10% in Japan and 60% in the US for the time being (now of 2024).

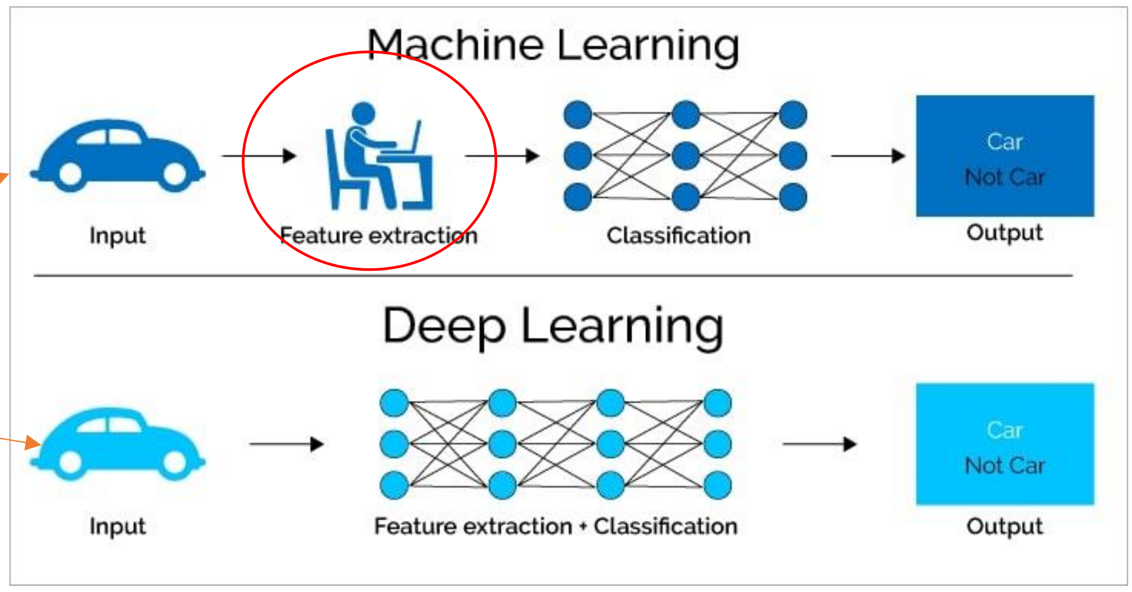
© Global Knowledge Link Center.

Reference)Which is better for the case?

- Rfc b鷸pclac gq rf_rk _afglc jc_plgle gq ^^fsk _lq bcrcpk glgle rfc af_p_arcpgqrgaq mdb_r_%_lb bccn jc_plgle gq ^^k _afglcq bcrcpk glgle rfc af_p_arcpgqrgaq mdb_r_%Rfcpcdmpc*grgqgk nmpr_lr rm sqc c_af rwnc _nnpmnpg_rcjw gl rcpk q mdu f_r iglb mdb_r_ m f_lb jc,
- Dmp cv_k njc*gdrfcpc gq m l jw _ qk _jj_k ms lr mdb_r_ rf_r a_l `c npcn_pcb dmp jc_plgle*k _afglc jc_plgle k _w `ck mpc _nnpmnpg_rc, Rfgqgq `ca_sqc k _afglc jc_plgle a_l jc_pl qs鷲agclrjwu grf jcqqb_r_ bcnclbgle ml rfc _jempgrfk sqcb, Ml rfc mrfcp f_lb*dmpnpmhcarq qsaf _q f_lbu pgrrclaf_p_arcp pcamelgrgml &MAR'*gnpccaf pcamelgrgml*...
b

Machine learning vs Deep learning?





Source : https://aismiley.co.jp/ai_news/what-is-the-difference-between-deep-learning-and-machine-learning/ , https://levity.ai/blog/difference-machine-learning-deep-learning

© Global Knowledge Link Center.

Against "Fake"

- You get along with AI under "with fake era" ?

Generating AI makes fakes clever and plentiful. Difficult to see through with human eyes

Fraud using AI fakes, so believing that you are okay is dangerous

Developing technology to spot fakes, learn how to interact with information

© Global Knowledge Link Center.

AI Updates recently

- Mid of 2025

Post LLM
(Mamba, 1bit LLM, SLM, LSTM or KAN?...)

ChatGtp4 works as a doctor?
And then “Omni”

“Lavender”

LSTM : Long Short-Term Memory, The most mentioned paper in May was "xLSTM: Extended Long Short-Term Memory," published by an Austrian research team.
KAN: Kolmogorov-Arnold Networks

© Global Knowledge Link Center.

Mamba

Pics : Adobe stock & Google

© Global Knowledge Link Center.

deepseek

© Global Knowledge Link Center.

Which is better for cases you face?

- We need both typed GenAI, so these are the same with some cases.

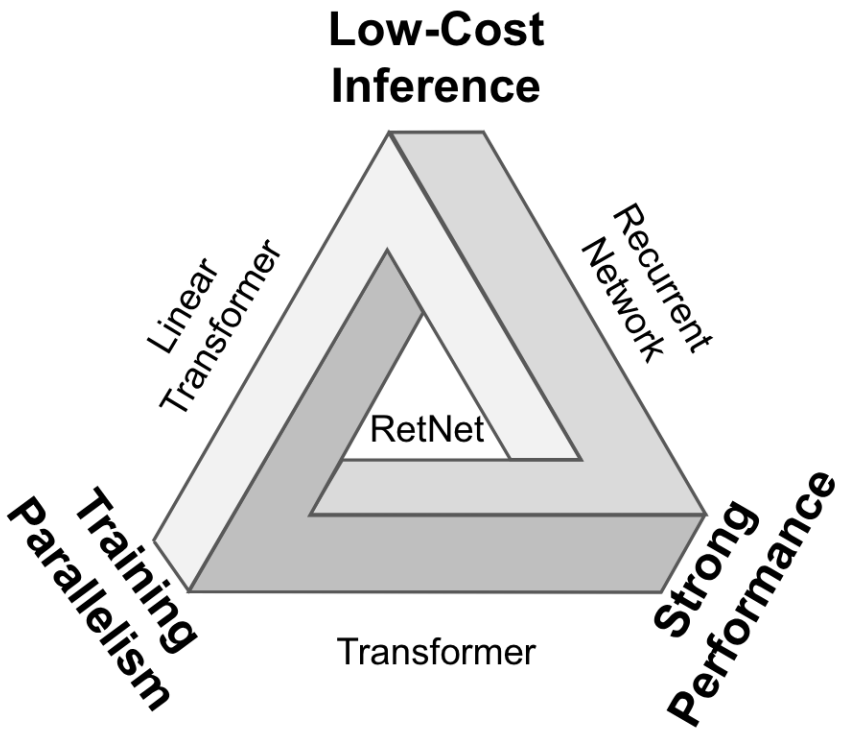
Quick response **VS.** **Deep Thinking**

© Global Knowledge Link Center.

Reference) Retnet 1/2

(Retentive Networks)

- A specification that simultaneously satisfies long text input, response speed, and low cost
- Retnet((Retentive Networks) is currently considered the leading candidate to succeed Transformer (as of 2024), and has three key advantages over Transformer:



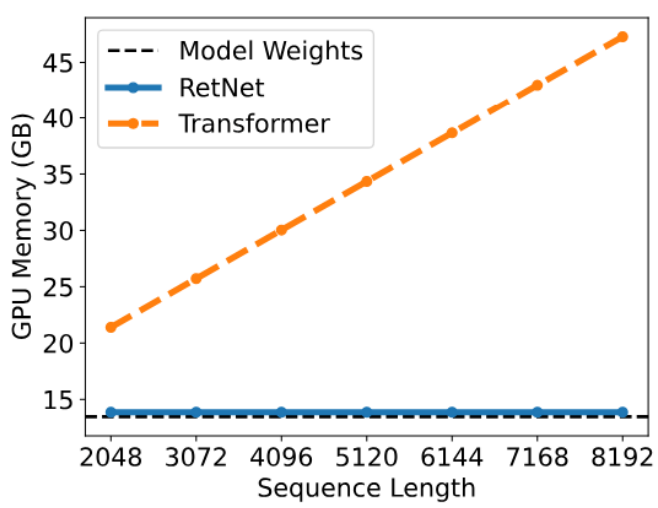
Source and Pics : https://arxiv.org/abs/2307.08621 Retentive Network: A Successor to Transformer for Large Language Models

© Global Knowledge Link Center.

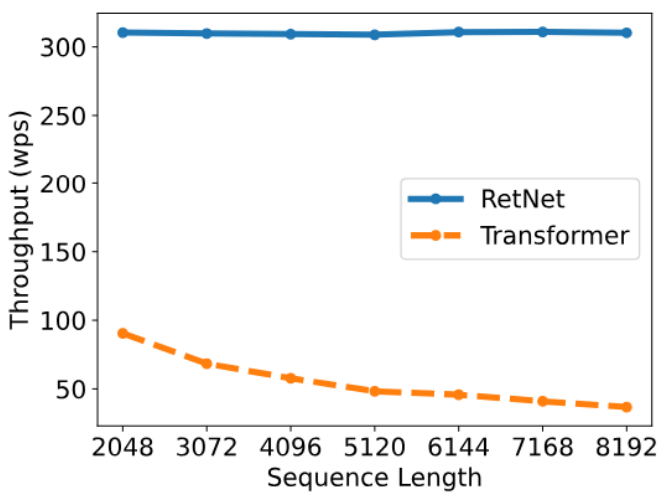
Reference) Retnet 2/2

Architectures	Training Parallelization	Inference Cost	Long-Sequence Memory Complexity	Performance
Transformer	✔	$O(N)$	$O(N^2)$	✔✔
Linear Transformer	✔	$O(1)$	$O(N)$	✘
Recurrent NN	✘	$O(1)$	$O(N)$	✘
RWKV	✘	$O(1)$	$O(N)$	✔
H3/S4	✔	$O(1)$	$O(N \log N)$	✔
Hyena	✔	$O(N)$	$O(N \log N)$	✔
RetNet	✔	$O(1)$	$O(N)$	✔✔

Memory consumption



Throughput



© Global Knowledge Link Center.

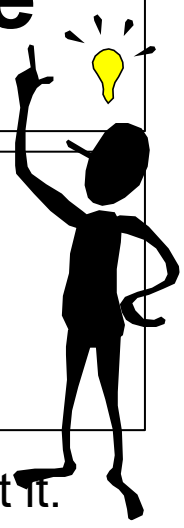
3 key message as “Wrap-up”

- Think of something through summarizing this material…

Everyone gets along well

Linkage is opener for the future

OPPENHEIMER



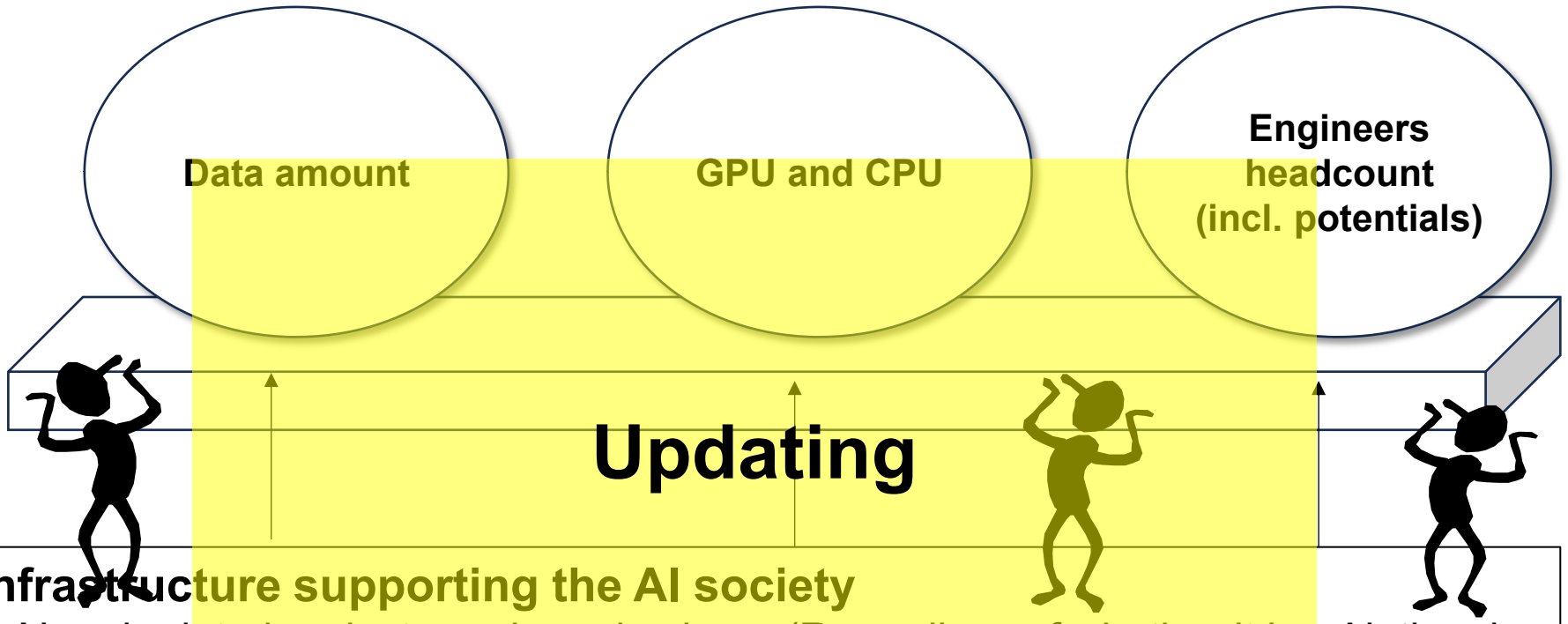
To eliminate something, you have to be vulgar enough to know about it.

© Global Knowledge Link Center.

AI Update

AI update1/5-5/5
What's AI spec as a national power?

- Comment from Professor Matsuo



Infrastructure supporting the AI society

- AI and related projects scale and volume (Regardless of whether it is a National and private projects)
- Monitored objects volume
- Server processing facilities volume
- Capacity to accept and recruit world-class engineers
- Power generation, power transmission infrastructure, etc.

© Global Knowledge Link Center.

AI update2/5
What direction we should take to grow into fair global market?

- Foreseen Risks and actions

Updating

© Global Knowledge Link Center.

AI update3/5
What direction we should take to grow into fair global market?

- What's "Model" of industry?

Updating

© Global Knowledge Link Center.

AI update4/5
What direction we should take to grow into fair global market?

- If we look at the international community broadly in terms of "democracy" versus "dictatorship," the total population is 2.3 billion versus 5.56 billion, respectively.

Updating

© Global Knowledge Link Center.

AI update5/5
What direction we should take to grow into fair global market?

Refer to Excel

- Global AI spec global ranking

Updating

© Global Knowledge Link Center.

End of the file

一から順に張り付けて
行って

19.05
25.4

保存する時、
拡張メタファイル
にする、必要なスライドの選択は
難しいようだ
そのページのところで保存すれば
該当ページのみのダウンロードで
きる

まぁいろいろやってみて

上のスライドは上に
下のスライドは下に
合わせる

- https://www.cybig.net/blog_tips/archives/707.html
  - の方法は有効でした
  - ちょっと面倒くさいが有効
  - ピクチャーにして張り付けるだけ