

AIはどうして急に賢くなったのか、 これからどうなるのか

黒橋 禎夫

国立情報学研究所 所長／京都大学 特定教授

軽井沢土曜懇談会 (2024/7/6)

ChatGPT

- OpenAIが2022年11月に公開した大規模言語モデル (Large Language Model; LLM) に基づくチャットボット
- OpenAIは2015年末にサム・アルトマン、イーロン・マスクらによって設立されたAI研究所。マイクロソフトも出資
 - 自然言語からの画像生成モデルDALL-E、音声認識モデルWhisperなども開発
 - 本社はサンフランシスコ



<https://openai.com/>

An astronaut riding a horse
in a photorealistic style



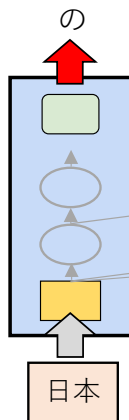
<https://openai.com/dall-e-2/>

ChatGPTとは

- OpenAIが2022年11月に公開した**大規模言語モデル (Large Language Model, LLM)** に基づくチャットボット

学習時

学習データ：日本の少子化対策には、次のようなアプローチ ...



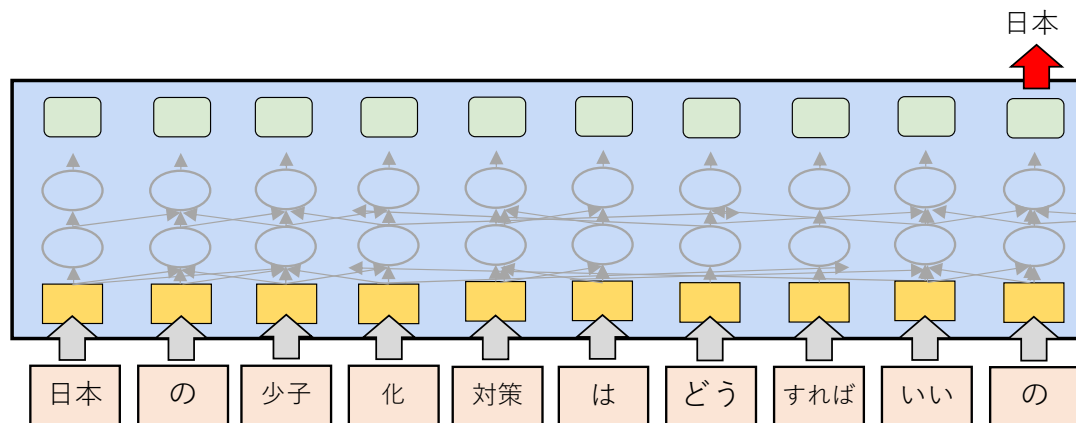
自己教師あり学習 (Self-Supervised Learning)

ChatGPTとは

- OpenAIが2022年11月に公開した大規模言語モデル (Large Language Model, LLM) に基づくチャットボット

推論時

プロンプト：日本の少子化対策はどうすればいいの



言語モデル (Language Model)

大規模コーパス (たとえば3,000億単語)

私 は りんご を

?

…お店で**私はりんご**を食べた…

…彼と**私はりんご**を食べた…

…昨日**私はりんご**をかじった…

…**私はりんご**を食べた後…

…**私はりんご**を殴った夢を…

$$P(\text{食べた} | \text{私 は りんご を}) = \frac{\text{コーパス中の頻度 (私 は りんご を 食べた)}}{\text{コーパス中の頻度 (私 は りんご を)}}$$

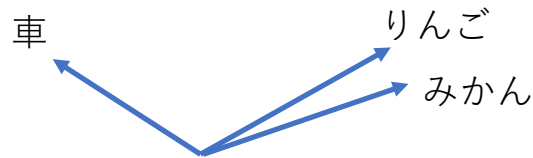
$$P(\text{殴った} | \text{私 は りんご を}) = \frac{\text{コーパス中の頻度 (私 は りんご を 殴った)}}{\text{コーパス中の頻度 (私 は りんご を)}}$$

ニューラル自然言語処理

- 語などの意味を1,000～10,000次元程度のベクトルで表現

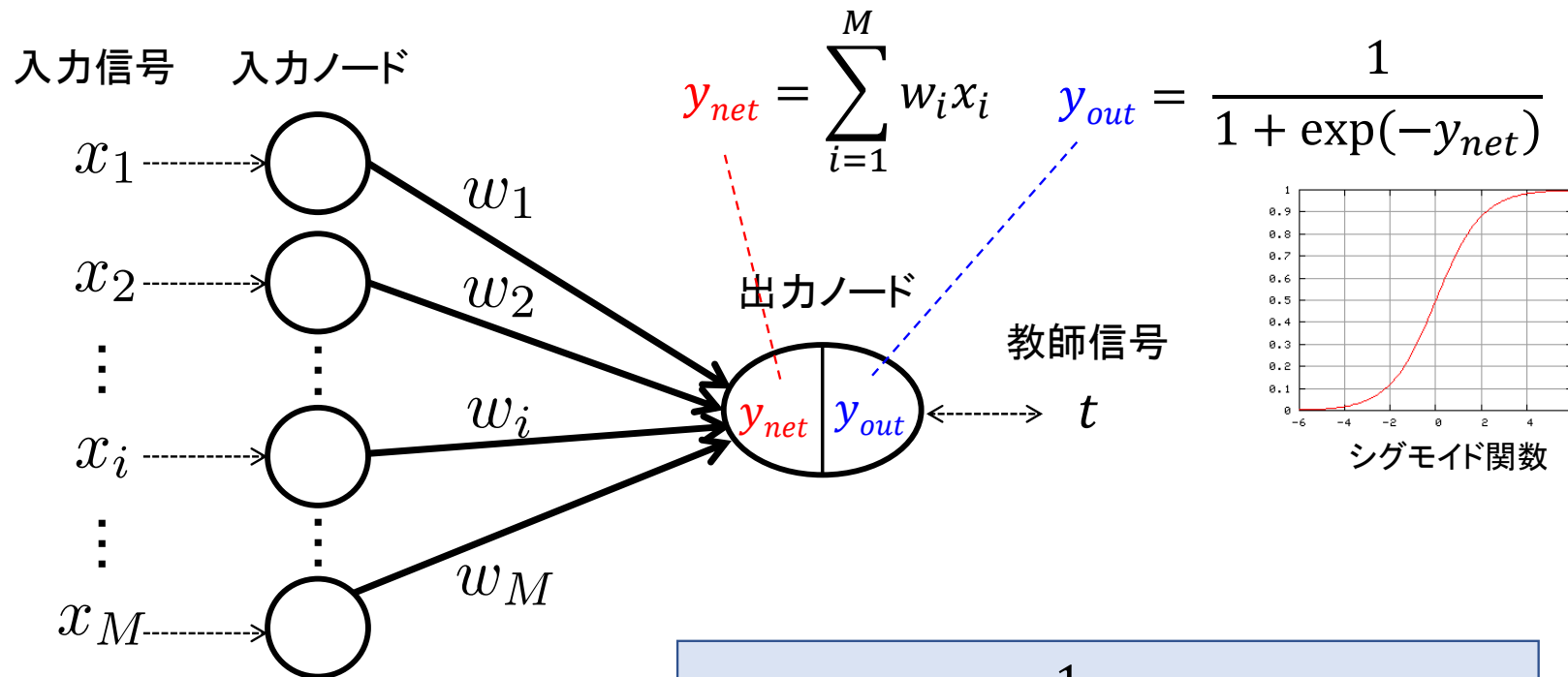
```
りんご [-0.034, 0.000, 0.010, -0.004, -0.012, -0.002, 0.007, 0.023, -0.002, 0.006, ...]  
みかん [-0.022, 0.000, -0.020, -0.007, -0.007, -0.007, -0.001, 0.007, -0.003, 0.024, ...]  
車      [-0.016, -0.014, -0.017, 0.009, 0.002, 0.012, -0.010, 0.015, 0.015, -0.008, ...]
```

- 意味が似ている語は近いベクトルになる



- 1,000次元で各次元が-1/1のいずれかの値をとるとしても、 $2^{1,000} \approx 10^{300}$ の膨大な意味空間を表現

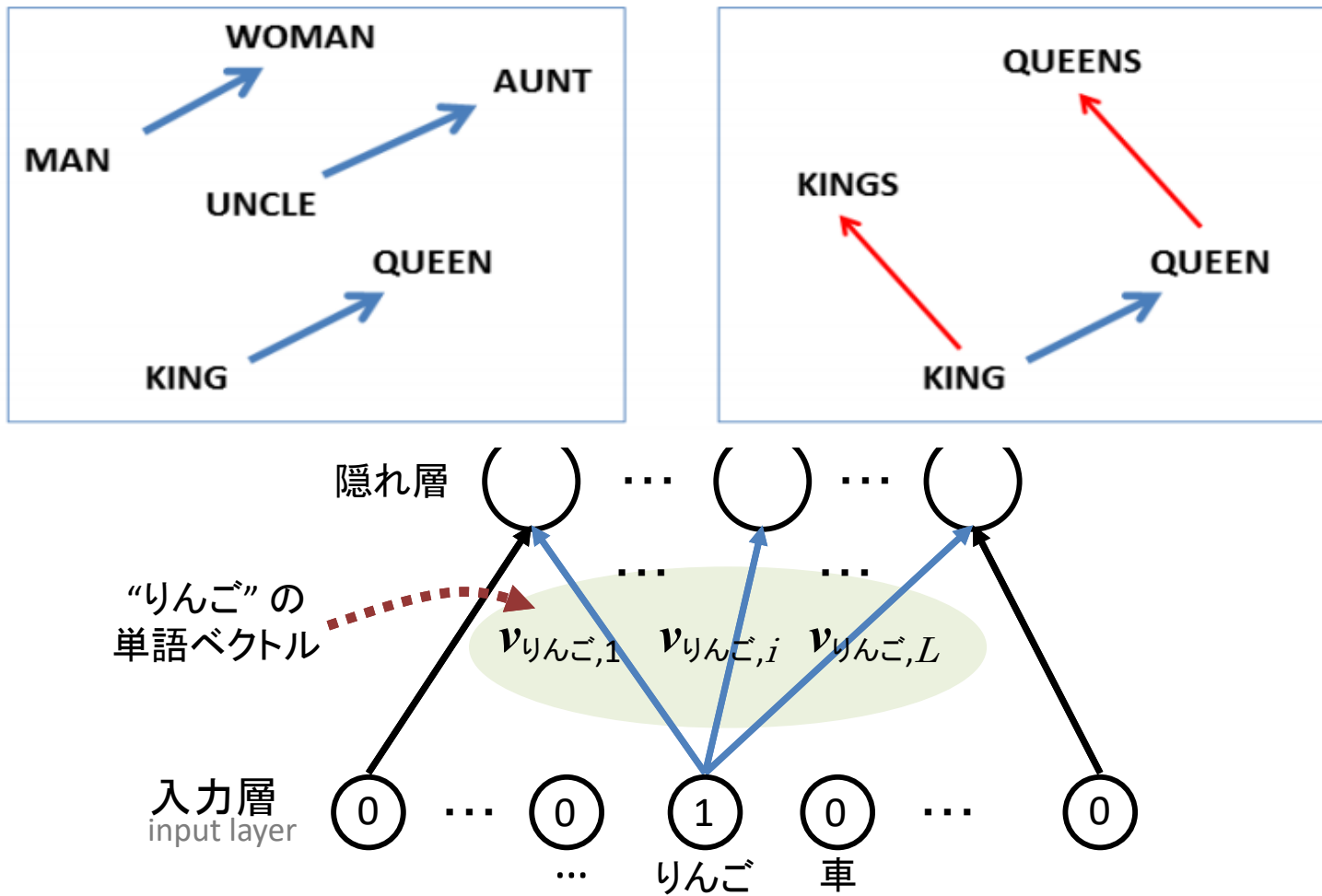
ニューラルネットワーク入門



誤差 $E = \frac{1}{2} (t - y_{out})^2$

重みの更新 $w_i^{(new)} = w_i^{(old)} - \eta \frac{\partial E}{\partial w_i}$

Word2Vec [Mikolov+ 2013]

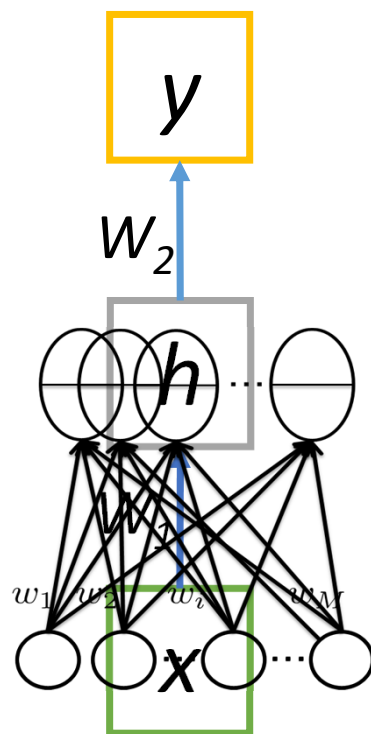


LLMの歴史

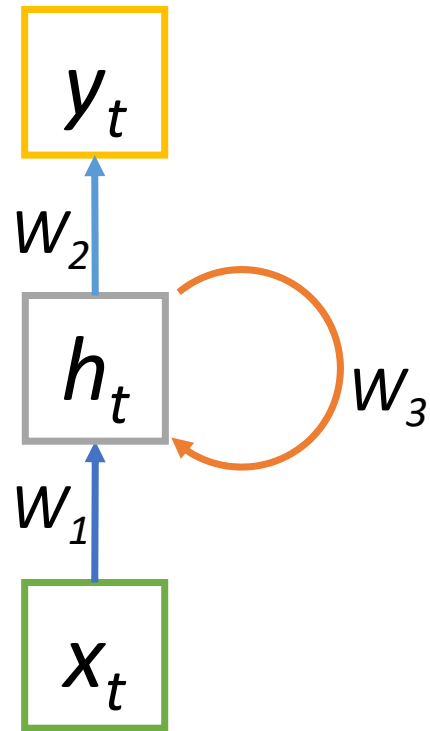
2014 Attention

機械翻訳において目的言語の次の語を生成する際に原言語の文のどこに着目するか

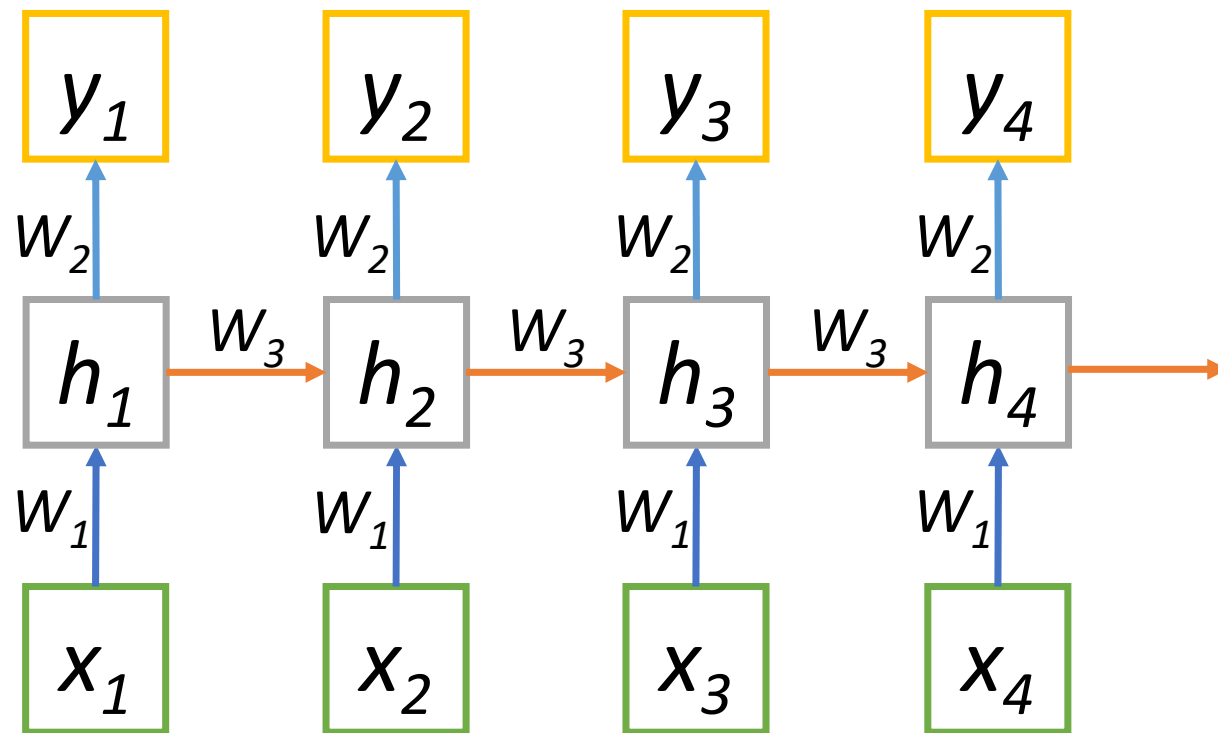
順伝搬型ニューラルネットワーク



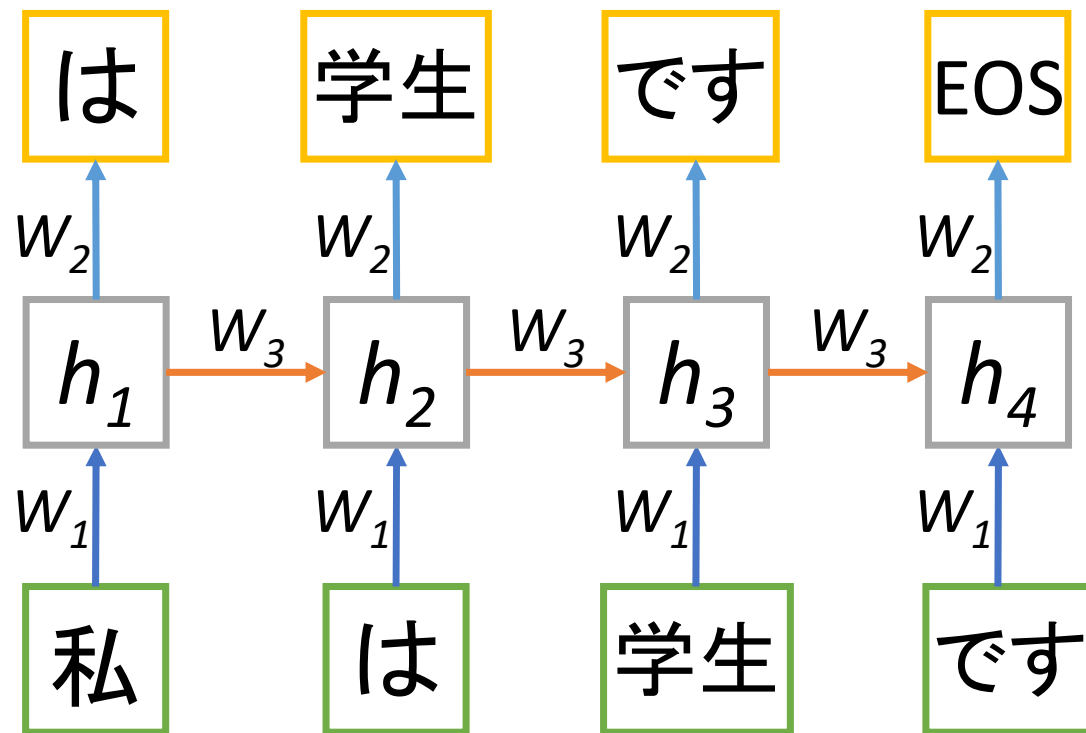
再帰型ニューラルネットワーク (RNN)



再帰型ニューラルネットワーク(RNN)

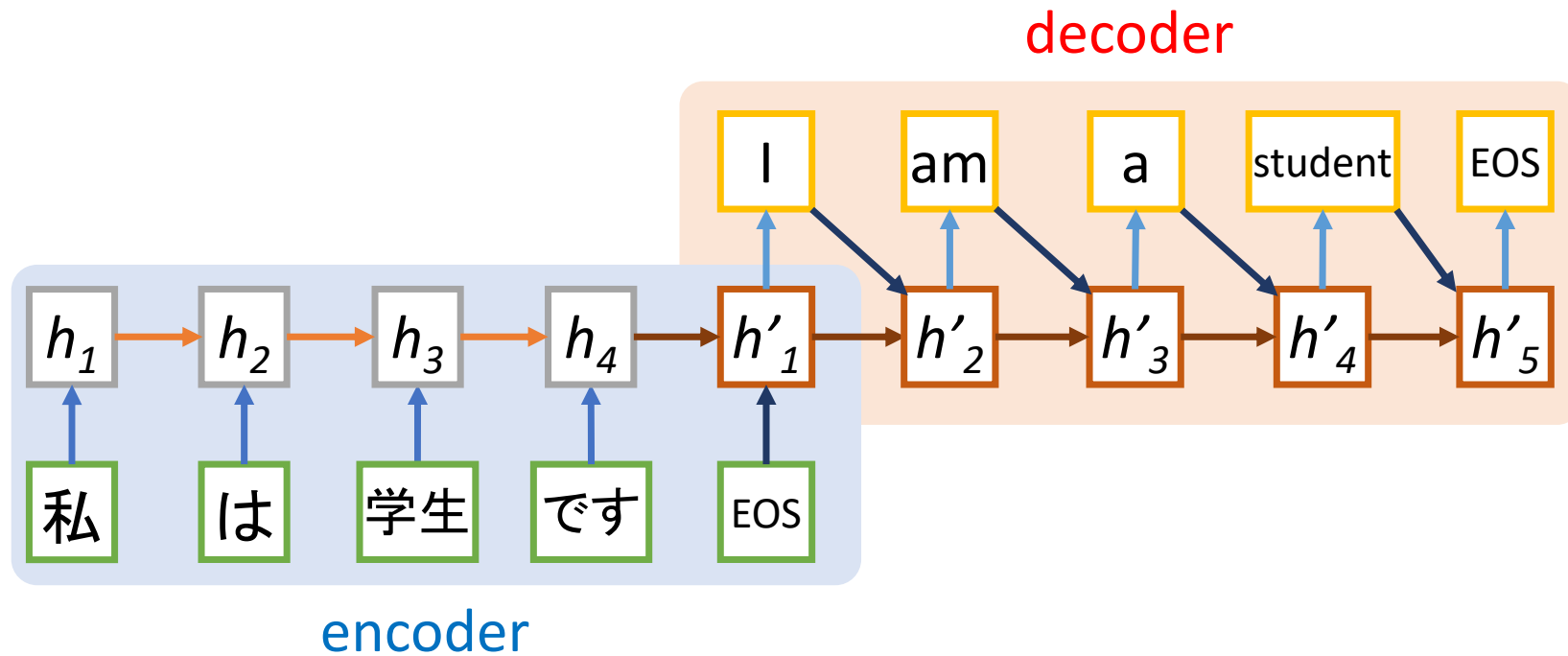


RNN 言語モデル



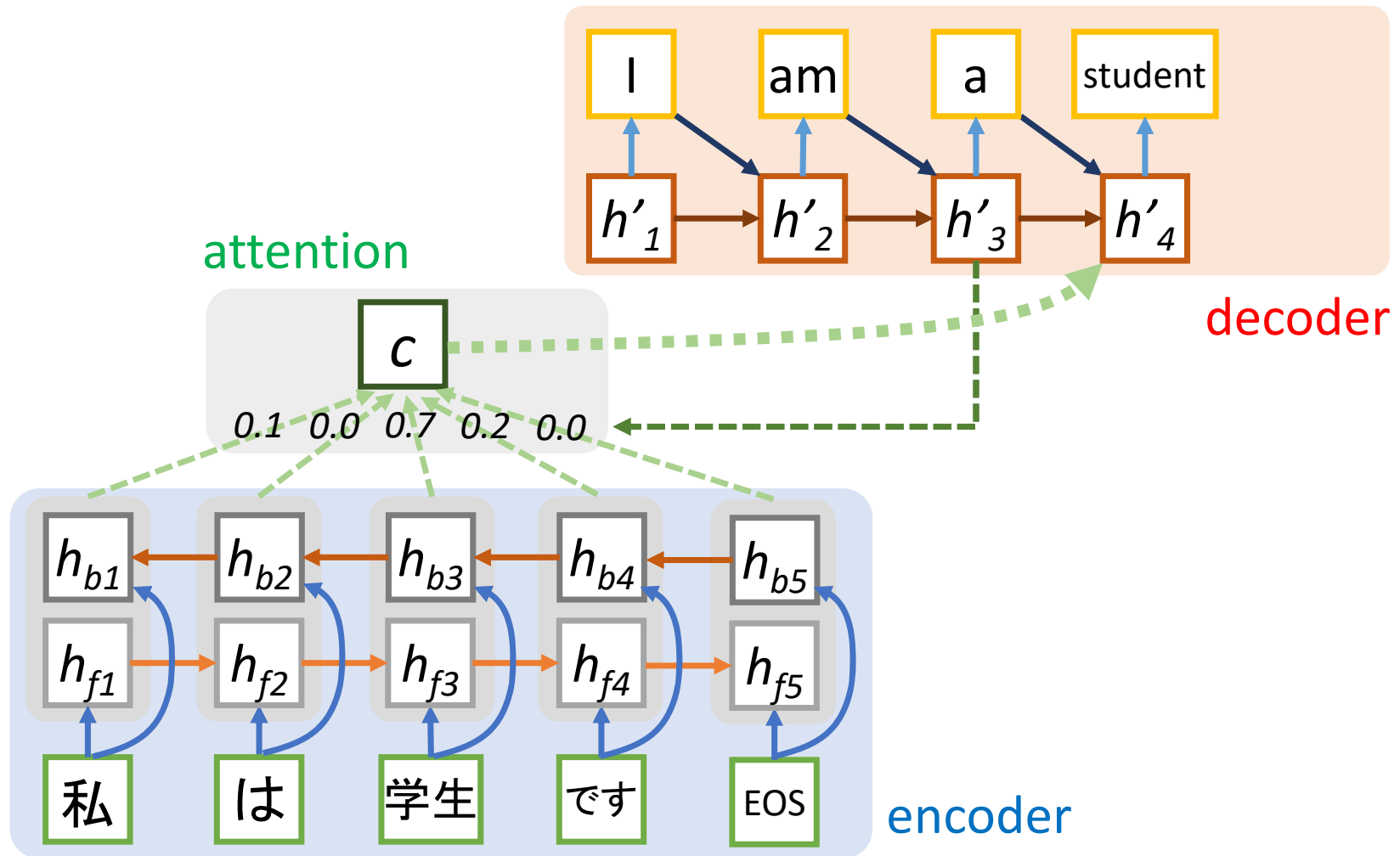
ニューラル翻訳

[Sutskever+, 2014]



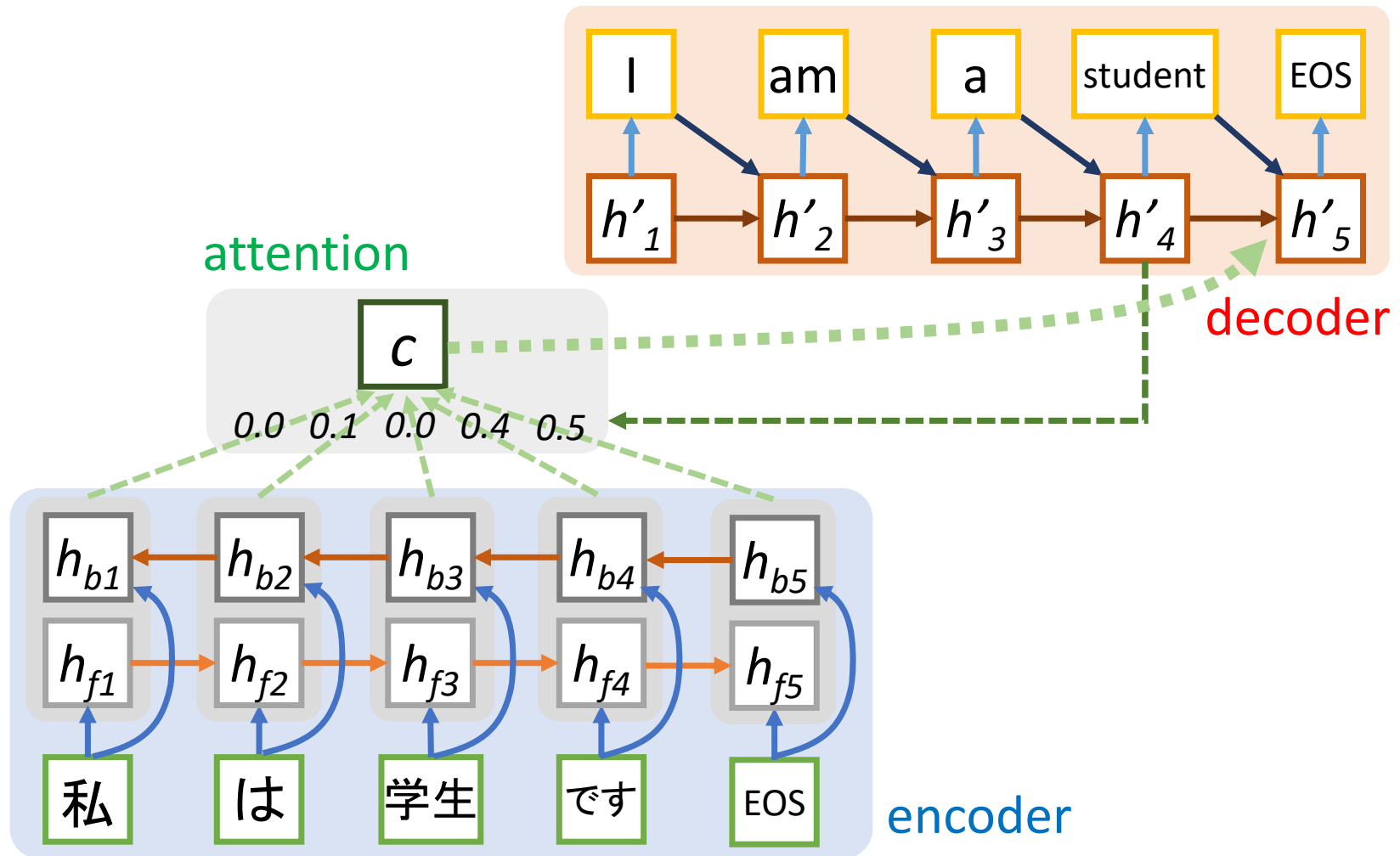
ニューラル翻訳

[Bahdanau+, 2014]



ニューラル翻訳

[Bahdanau+, 2014]



LLMの歴史

2014 Attention

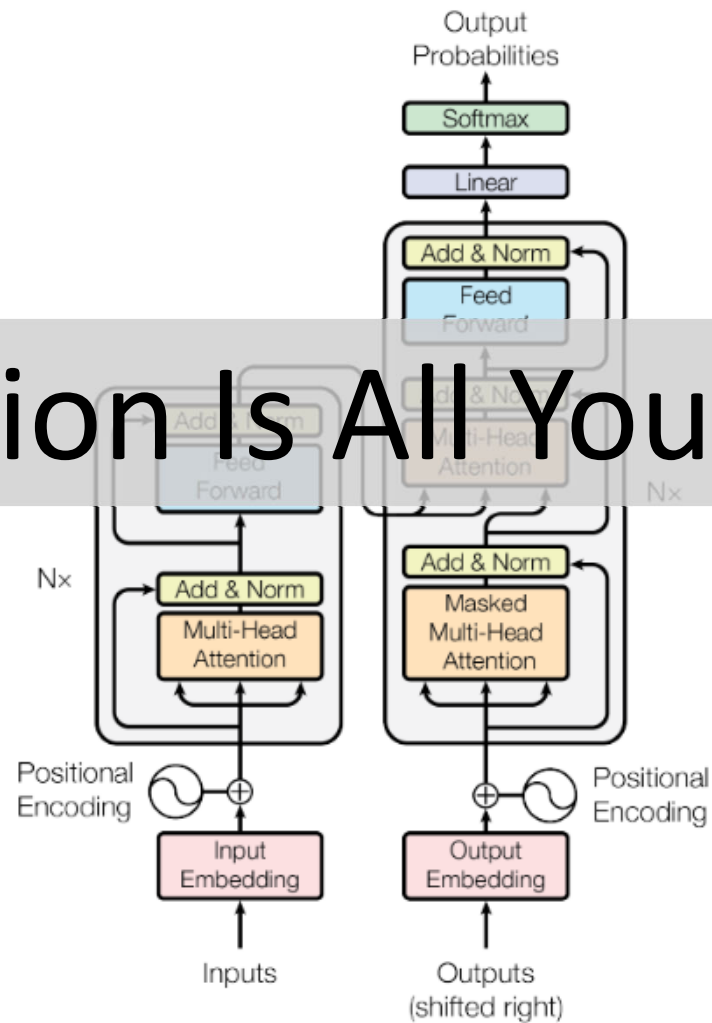
機械翻訳において目的言語の次の語を生成する際に原言語の文のどこに着目するか

2017 Transformer

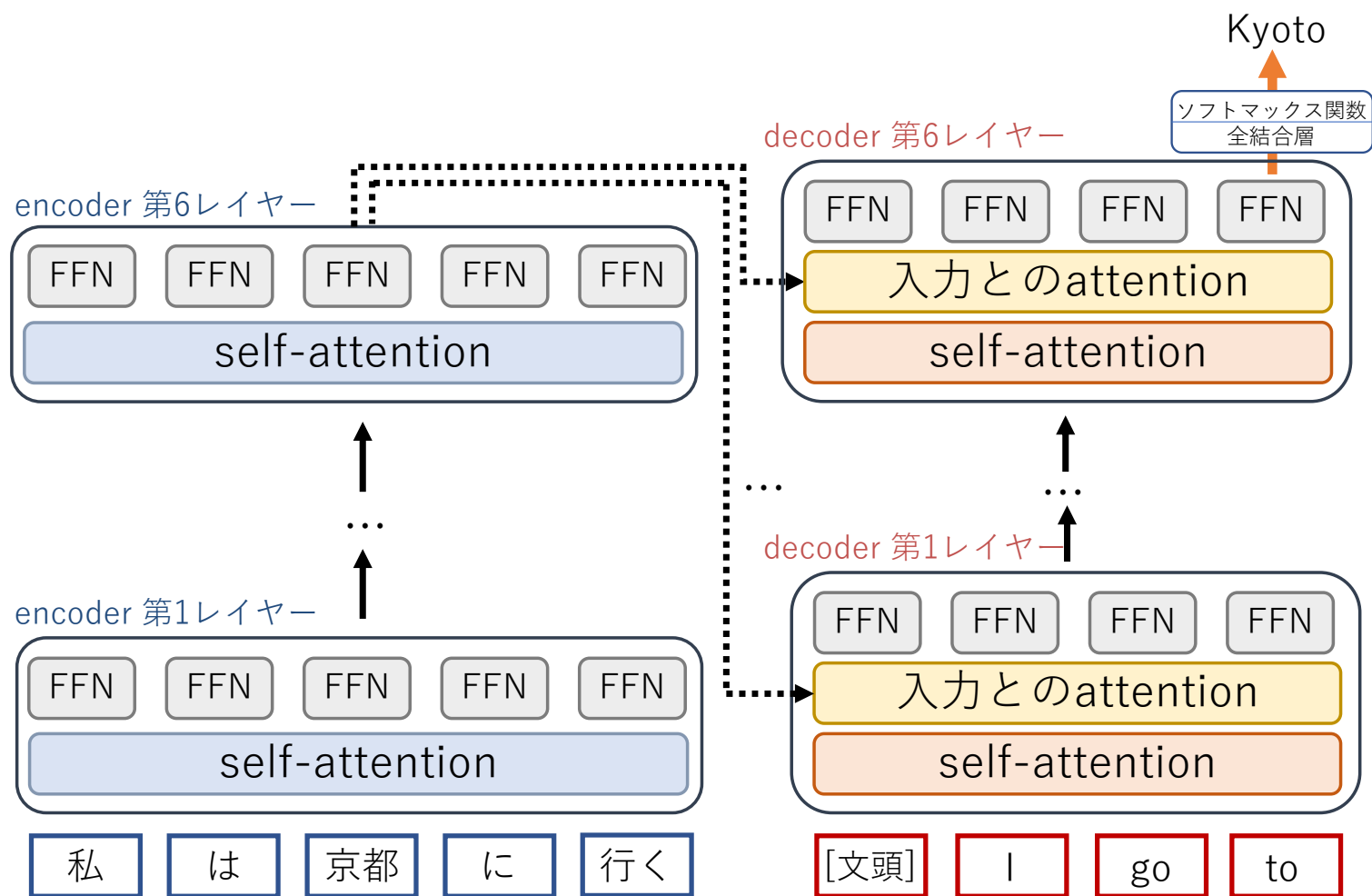
attentionの精緻化、原言語文内、目的言語内でのattention

Transformer [Vaswani et al. 2017]

Attention Is All You Need



Transformer [Vaswani+, 2017]



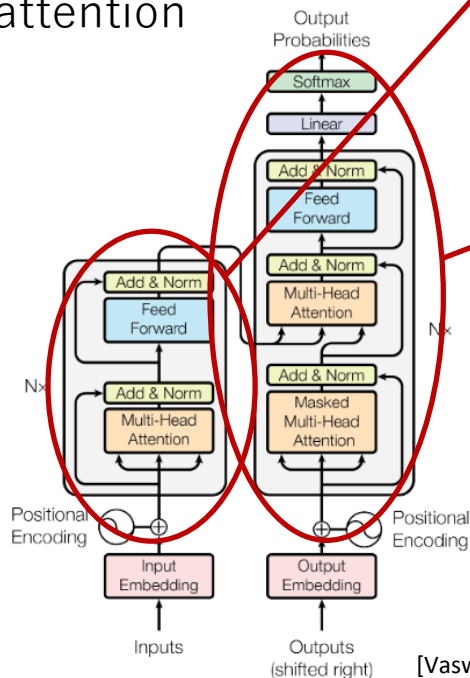
LLMの歴史

2014 Attention

機械翻訳において目的言語の次の語を生成する際に原言語の文のどこに着目するか

2017 Transformer

attentionの精緻化、原言語文内、目的言語内でのattention



[Vaswani et al. 2017]

2018 BERT

Transformerのencoder側を単言語の分類問題等に

2018 GPT (117Mパラメータ)

Transformerのdecoder側を言語モデルに

2019 GPT-2 (1.5Bパラメータ)

2020 GPT-3 (175Bパラメータ)

2022 GPT-3.5 / InstructGPT

2022 ChatGPT

2023 GPT-4 (2Tパラメータ?)

画像も扱える、多言語能力も大幅向上

- 米司法試験で人間受験者の上位10%の成績
- 米大学入試テストSATで1600点中1410点
- 米医師試験USMLEでも合格レベルの点数

Instruction Tuning

1. Supervised Fine-tuning (SFT)

Instruction:

ポテトチップスの袋はなぜ開封後に古くなるのか？

Response:

ポテトチップスの袋は窒素で満たされている！多くの消費者は、ポテトチップス会社が袋の4分の3を空気で満たし、お金を取ろうとしていると考えているが、実はこれは．．．

2. Learning from Human Feedback (LHF)

using Direct Preference Optimization (DPO) [Rafailov+ 2023]

Instruction:

父とは疎遠なのですが、もう一度連絡を取りたいと思っています．．．

Response:

メールが一番簡単だと思います。「一緒に過ごした楽しい時間は一生忘れない」とか、そういうことを付け加えてもいいかも。



Response:

メールにしたほうがいいと思う。なぜ聞くのですか？他の方法の方がいいと思う理由があるのですか？.



ChatGPTをめぐる初期の議論

Q: どうして生成される表現がこんなに流暢？

A: 言語モデルだから

Q: どうして何でも知っている？

A: コーパス（ウェブ上のどこか）に書いてあるから

Q: 信頼できるのか、嘘をつくのか？

A: コーパスにないことを聞かれた場合、「それは知りません」とは言わずに、言語モデルが勢いで事実と反することを言うことがある(Hallucination)

意味・文脈の理解と多言語性



翻訳：うなぎ文

⚡ Default (GPT-3.5)



「部長はうなぎです」を英語に訳してください



"部長はうなぎです"を英語に訳すと、「The department head is an eel.」となります。



翻訳：うなぎ文

 DeepL

日本語 (自動検出) ▼

↔ 英語 (US) ▼

用語集

部長はうなぎです ×

The chief is an eel.

その他の候補：

The director is an eel.

翻訳：うなぎ文

ChatGPT PLUS

✦ GPT-4



「部長はうなぎです」を英語に訳してください



「部長はうなぎです」というフレーズは、直訳すると"The manager is an eel" となります。📄 👍 🗑️

ただし、このフレーズは日常的な会話で直接使用されるものではないと思われるため、コンテキストや文脈に応じて、適切な翻訳を行う必要があるかもしれません。特定の文化的・慣用的な意味が含まれている場合、そのニュアンスを取り入れた訳が必要となることも考えられます。

翻訳：うなぎ文

ChatGPT

PLUS

KU

お店で注文している場合だとどうなりますか



お店で「部長はうなぎです」と注文する場面を想像すると、これはおそらく部長がうなぎを注文したいという意味になります。この場合、英語での翻訳は "The manager would like eel" もしくは "The boss orders eel" となるでしょう。



参考：『ChatGPT翻訳術 新AI時代の超英訳スキルブック』山田 優 著, 株式会社アルク 2023.

土佐日記の現代語への翻訳

ChatGPT 4o

以下の文を現代語に変換して。

ある人、県の四年五年果てて、例のことどもみなし終へて、解由など取りて、住む館より出でて、船に乗るべき所へわたる。かれこれ、知る知らぬ、送りす。



もちろんです。以下の文を現代語に変換いたします。

ある人が、県での任期を四、五年務め終え、通常の手続きをすべて済ませて、解由状などを受け取り、住んでいた館を出て、船に乗る場所へ向かう。あちらこちらで、知っている人も知らない人も見送りに来る。

Retrieval Augmented Generation

大谷2023年成績【打撃&投手】 × +

chat.openai.com/c/44b569bf-8263-4e13-976d-2ff937f89264

ChatGPT 4

You
大谷翔平選手の2023年シーズンの成績をおしえてください。

ChatGPT
I did a quick search for more information and here's what I discovered.

大谷翔平選手の2023年シーズンの成績は以下の通りです。

打撃成績:

- 試合数: 135
- 打席数: 599
- 打率: .304
- 安打数: 151
- 二塁打: 26
- 三塁打: 8
- 本塁打: 44
- 打点: 95
- 四球: 91
- 敬遠: 21
- 三振: 143
- 盗塁: 20

投手成績:

- 登板試合数: 23
- 勝利数: 10
- 敗戦数: 5
- 防御率: 3.14
- 投球回: 132.0
- 奪三振数: 167
- イニングあたりの出塁数平均 (WHIP): 1.06

この素晴らしい成績により、大谷選手は2023年に多くの記録を打ち立てました。例えば、彼はアジア人初のメジャーリーグ本塁打王に輝き、2度目のアメリカンリーグMVPを獲得しました。また、彼のOPSは両リーグで最高であり、日本人メジャーリーガーとして歴代最多塁打の記録も更新しました。大谷選手の二刀流としての活躍は、彼が同一シーズンで10勝以上と40本塁打以上を達成したことにより、メジャーリーグ史上初の偉業となりました。

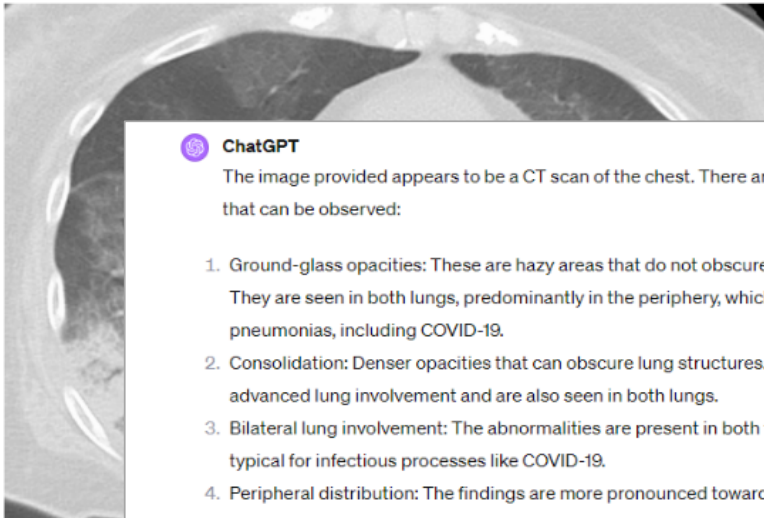
Message ChatGPT...

ChatGPT can make mistakes. Consider checking important information.

画像の理解

ChatGPT 4

You



You are a he
COVID-19?

ChatGPT

The image provided appears to be a CT scan of the chest. There are several radiological findings that can be observed:

1. Ground-glass opacities: These are hazy areas that do not obscure the underlying lung structures. They are seen in both lungs, predominantly in the periphery, which is a common finding in viral pneumonias, including COVID-19.
2. Consolidation: Denser opacities that can obscure lung structures. These may represent more advanced lung involvement and are also seen in both lungs.
3. Bilateral lung involvement: The abnormalities are present in both the right and left lungs, which is typical for infectious processes like COVID-19.
4. Peripheral distribution: The findings are more pronounced towards the edges of the lungs, which is characteristic of COVID-19 pneumonia.

While these findings are suggestive of COVID-19, especially in the context of the current pandemic and if correlated with clinical symptoms such as fever, cough, and shortness of breath, they are not exclusive to COVID-19. Other viral pneumonias, bacterial pneumonias, or even non-infectious processes could present with similar patterns on a CT scan.

Definitive diagnosis of COVID-19 would require a positive result from a polymerase chain reaction (PCR) test for SARS-CoV-2. Radiological findings must be correlated with clinical presentation and laboratory tests for an accurate diagnosis. It's essential to consider patient history, symptoms, and exposure risk when interpreting imaging results like these.



User

Answer question I.1.a. Think step-by-step.

I. Principe de la détection de rayonnement avec un bolomètre

Comme illustré sur la figure 1 un bolomètre est constitué d'un absorbeur qui reçoit le rayonnement que l'on désire détecter. Sa température T , supposée uniforme, est mesurée à l'aide d'un thermomètre incorporé, constitué d'un matériau conducteur dont la résistance $R(T)$ varie avec la température T ; cette variation est caractérisée par le coefficient $\alpha = \frac{1}{R} \frac{dR}{dT}$. L'ensemble possède la capacité thermique C_{th} .

Un barreau, conducteur thermique, homogène, de longueur L , de section S et de conductivité thermique λ et sans échanges thermiques latéraux, relie le bolomètre à un thermostat de température T_b fixe.

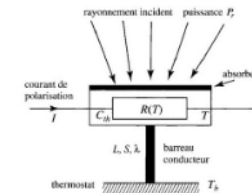


Figure 1 - Schéma d'un bolomètre.

I.1. Dans cette première partie, on étudie le comportement du bolomètre en régime stationnaire, à la température de fonctionnement $T = T_b$.

I.1.a) Exprimer la température $T(x)$ en tout point du barreau conducteur repéré par son abscisse x ; on prendra $T(0) = T_0$ et $T(L) = T_b$.

GPT-4

To answer question I.1.a, we need to express the temperature $T(x)$ in every point of the conductive bar, which is characterized by the abscissa x . We are given that $T(0) = T_0$ and $T(L) = T_b$.

For a stationary regime, the heat equation for the one-dimensional heat conduction in the bar is:
$$d^2T/dx^2 = 0$$

Integrating once, we get:
$$dT/dx = C_1$$

Integrating a second time:
$$T(x) = C_1x + C_2$$

To determine the constants C_1 and C_2 , we use the boundary conditions:
 $T(0) = T_0$, which gives $C_2 = T_0$
 $T(L) = T_b$, which gives $C_1 = (T_b - T_0)/L$

Finally, the expression for $T(x)$ is:
$$T(x) = (T_b - T_0) * (x/L) + T_0$$

This represents a linear temperature profile along the conductive bar, with a slope of $(T_b - T_0)/L$.

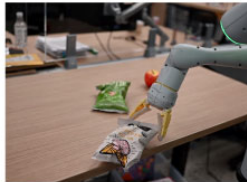
GPT-4 [OpenAI 2023]

RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control

[Brohan+ 2023]



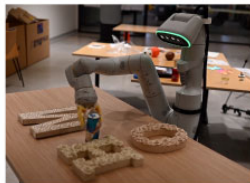
put strawberry
into the correct
bowl



pick up the bag
about to fall
off the table



move apple to
Denver Nuggets



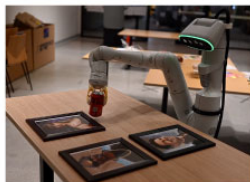
move redbull can
to H



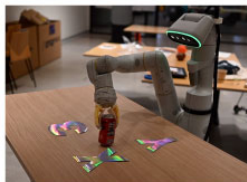
move soccer ball
to basketball



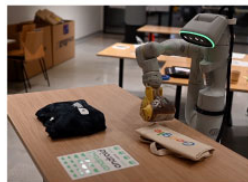
move banana to
Germany



move coke can to
Taylor Swift



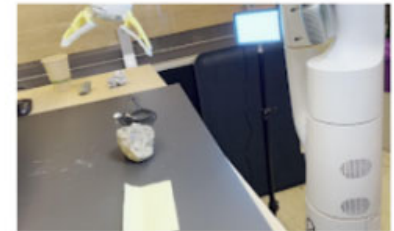
move coke can to
X



move bag to
Google

Instruction:
I need to hammer a nail,
what object from the scene
might be useful?

Prediction:
Rocks. Action: 1 129 138 122
132 132 106 127



ChatGPTでできること

- プロンプト（GPTへの入力）によりかなり知的なコントロールが可能
 - 要約、言い換え、翻訳
 - アイデア出し、ブレストの対話相手（壁打ち）
 - ロールプレイング（就職面接、問診など）
- プログラミング支援
- GPTs：GPTベースのシステムを対話的に作成・共有
(512MB, 2百万トークンまでのファイルを20ファイルまで外部知識として与えられる)

Introducing ChatGPT | OpenAI

ChatGPT

+

←

→

🔄

🔍 chatgpt.com/gpts/editor

☆

🔍

⚙️

⏪

新しい GPT

• 下書き

作成する

構成



Hi! I'll help you build a new GPT. You can say something like, "make a creative who helps generate visuals for new products" or "make a software engineer who helps format my code."

What would you like to make?



GPT Builder にメッセージを送信する



プレビューする



まずは GPT を定義してください。

33

Introducing ChatGPT | OpenAI

ChatGPT

chatgpt.com/gpts/editor/g-Vc2NbKB6o

新しい GPT

下書き

作成する

構成



Hi! I'll help you build a new GPT. You can say something like, "make a creative who helps generate visuals for new products" or "make a software engineer who helps format my code."

What would you like to make?

24時間以内の最新の英文ニュースを要約して表示し、さらに重要な用語とその翻訳を表示するGPT



unknown plugin を使用しています... ✖



Great! Now, let's decide on a name for this GPT. How about "News Summary Translator"? Does that work for you?

 GPT Builder にメッセージを送信する



プレビューする



Summarizes the latest English news and translates key terms into Japanese.

Summarize the latest news article from today.

What's a summary of today's top headlines?

Show the important terms and their translations from...

Can you translate key terms from the latest news?

 GPT にメッセージを送信する

34

Introducing ChatGPT | OpenAI

ChatGPT

chatgpt.com/gpts/editor/g-Vc2NbKB6o

☆

...

News Summary Translator

下書き

作成する

構成



unknown plugin を使用しています... ❗



Great! Now, let's decide on a name for this GPT. How about "News Summary Translator"? Does that work for you?

yes



I'll generate a profile picture for the News Summary Translator. I'll show you the result shortly.



Here's the profile picture for the News Summary Translator. Do you like it, or would you like any changes?



GPT Builder にメッセージを送信する



プレビューする



News Summary Translator

Summarizes the latest English news and translates key terms into Japanese.

Summarize the latest news article from today.

What's a summary of today's top headlines?

Show the important terms and their translations from...

Can you translate key terms from the latest news?



News Summary Translator にメッセージを送信する

35

Introducing ChatGPT | OpenAI

ChatGPT

chatgpt.com/gpts/editor/g-Vc2NbKB6o

☆

...

News Summary Translator

下書き

作成する

構成

yes

I'll generate a profile picture for the News Summary Translator. I'll show you the result shortly.



Here's the profile picture for the News Summary Translator. Do you like it, or would you like any changes?

yes

Great! Let's refine how the News Summary Translator should operate. First, what should be emphasized or avoided in its summaries and translations?

要約は英語で。要約に加えて、重要な用語と、その日本語訳を表示してください。

プレビューする



News Summary Translator

Summarizes the latest English news and translates key terms into Japanese.

Summarize the latest news article from today.

What's a summary of today's top headlines?

Show the important terms and their translations from...

Can you translate key terms from the latest news?

News Summary Translator にメッセージを送信する

36

Introducing ChatGPT | OpenAI

ChatGPT

chatgpt.com/gpts/editor/g-Vc2NbKB6o

☆

🔍

🔧

🔗

🔴 S

⋮

News Summary Translator

• 下書き

...

作成する

作成する

構成

Here's the profile picture for the News Summary Translator. Do you like it, or would you like any changes?

yes

Great! Let's refine how the News Summary Translator should operate. First, what should be emphasized or avoided in its summaries and translations?

要約は英語で。要約に加えて、重要な用語と、その日本語訳を表示してください。

Next, how do you want the News Summary Translator to communicate with you? Should it be formal, friendly, or have a specific tone?

フレンドリーにしてください。

Your News Summary Translator is ready! Please try it out in the preview and let me know if you have any further refinements. What would you like to do next?

📎

GPT Builder にメッセージを送信する

↑

プレビューする



News Summary Translator

Summarizes the latest English news and translates key terms into Japanese.

Summarize the latest news article from today.

What's a summary of today's top headlines?

Show the important terms and their translations from...

Can you translate key terms from the latest news?

📎

News Summary Translator にメッセージを送信する

↑

37

?

Summary Translator

作成する

作成する

構成

Great! Let's refine how the News Summary Translator should operate. First, what should be emphasized or avoided in its summaries and translations?

要約は英語で。要約に加えて、重要な用語と、その日本語訳を表示してください。

Next, how do you want the News Summary Translator to communicate with you? Should it be formal, friendly, or have a specific tone?

フレンドリーにしてください。

Your News Summary Translator is ready! Please try it out in the preview and let me know if you have any further refinements. What would you like to do next?

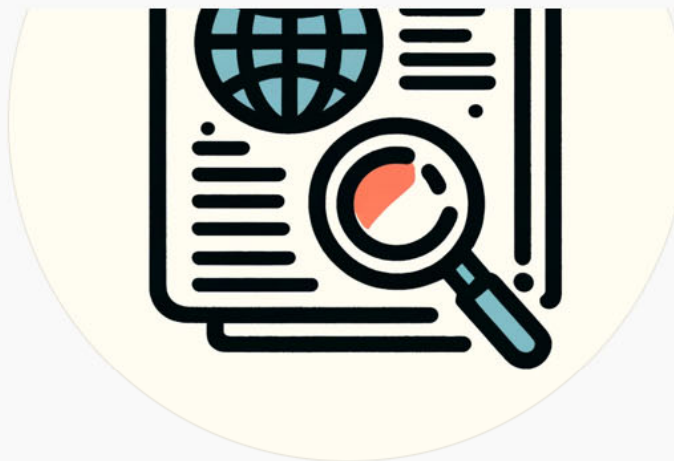
ニュースのトピックを最初に受け付けるようにしてください。

Your News Summary Translator is now set up to ask for a specific news topic first. Please try it out and let me know if you need any more adjustments. What would you like to do next?

GPT Builder にメッセージを送信する



プレビューする



UnitedHealth Group has provided an update on the recent cyberattack on Change Healthcare. The attack potentially exposed protected health information (PHI) and personally identifiable information (PII) of numerous individuals. The company is actively working to restore services, with pharmacy services and medical claims processing nearing normal levels. Support measures include free credit monitoring and identity theft protection for those affected ([Welcome to UnitedHealth Group](#)).

Key Terms:

- Cyberattack: サイバー攻撃
- Protected Health Information (PHI): 保護された健康情報



News Summary Translator にメッセージを送信する

38



Introducing ChatGPT | OpenAI

ChatGPT - News Summary Translator

chatgpt.com/g/g-Vc2NbKB6o-news-summary-translator

News Summary Translator



News Summary Translator

SADAO KUROHASHI が作成

Summarizes the latest English news and translates key terms into Japanese.

Summarize the latest news article from today.

What's a summary of today's top headlines?

Show the important terms and their translations from...

Can you translate key terms from the latest news?

 News Summary Translator にメッセージを送信する



ChatGPT の回答は必ずしも正しいとは限りません。重要な情報は確認するようにしてください。

?

39

Introducing ChatGPT | OpenAI

ChatGPT - News Summary Trans

chatgpt.com/g/g-Vc2NbKB6o-news-summary-translator/c/814e4890-26a3-4824-b751-f90e629241b2

News Summary Translator

AI関係

 5 サイトを検索しました

Latest AI News Summary (May 24, 2024)

Summary:

- Advancements in Smaller AI Models:**

The trend in AI is shifting towards smaller, more efficient models due to the rising costs and shortages of hardware like GPUs. Techniques such as Low Rank Adaptation (LoRA) and quantization are helping make model optimization more accessible, enabling even smaller companies to utilize advanced AI capabilities without hefty investments in infrastructure.
- Generative AI in Media Production:**

Generative AI is making significant strides in the film industry. Tools like Runway's Gen-2 model are creating high-quality short videos, attracting interest from major studios for tasks like lip-syncing in multiple languages and enhancing special effects. This technology is also being used widely in marketing, allowing companies to generate deepfake avatars for various purposes.
- Google's AI Innovations:**

At Google Cloud Next 2024, Google announced new features in its AI portfolio, including the public preview of Gemini 1.5 Pro. This model boasts enhanced performance and long-context understanding, which expands its capabilities across multiple domains such as gaming and insurance. Additionally, Google continues to lead in AI infrastructure, supporting a large portion of generative AI startups.

 News Summary Translator にメッセージを送信する

ChatGPT の回答は必ずしも正しいとは限りません。重要な情報は確認するようにしてください。

Introducing ChatGPT | OpenAI

ChatGPT - News Summary Translator

chatgpt.com/g/g-Vc2NbKB6o-news-summary-translator/c/814e4890-26a3-4824-b751-f90e629241b2

News Summary Translator

of generative AI startups.

4. Ethical and Social Implications:

The spread of AI-generated disinformation is becoming a critical issue, especially in political contexts. As generative AI tools become more accessible, the potential for creating convincing fake content increases, posing challenges for verifying information authenticity online.

Important Terms and Translations:

- **Artificial Intelligence (AI):** 人工知能 (じんこうちのう)
- **Generative AI:** 生成的AI (せいせいいてきAI)
- **Model Optimization:** モデル最適化 (もでるさいてきか)
- **GPU (Graphics Processing Unit):** GPU (グラフィックス処理装置)
- **Low Rank Adaptation (LoRA):** 低ランク適応 (ていらんくてきおう)
- **Quantization:** 量子化 (りょうしか)
- **Deepfake:** ディープフェイク (でいふふえいく)
- **Vertex AI:** Vertex AI (バーテックス AI)
- **Disinformation:** 偽情報 (ぎじょうほう)
- **Ethics:** 倫理 (りんり)

For more detailed information, you can refer to the original articles on IBM, MIT Technology Review, Google Blog, and Digital Agency Network.

↓

News Summary Translator にメッセージを送信する

ChatGPT の回答は必ずしも正しいとは限りません。重要な情報は確認するようにしてください。

?

LLMに関する懸念

- 研究開発が**一部の組織の寡占状態**であることは健全とはいえない。
OpenAIはもはやオープンではない
- 強い・大きなモデルの学習コーパス・データは公開されておらず、
巨大パラメータのモデルの振舞いを含めて全体がblack-box。**その賢さ、多言語性について何が起きているのか誰もわからない**
- 一方、現在のモデルにはHallucination、バイアス等の課題も山積
- **日本語コーパスはGPT-3において0.11%**、日本語の理解・生成能力は英語に比べて劣る

GPT-3における英語と日本語の扱い

GPT-3 Codex

Major social networking sites and other platform providers have grown without ever assuming responsibility for the content of their users' posts.

Clear

Show example

Tokens
23

Characters
145

Major social networking sites and other platform providers have grown without ever assuming responsibility for the content of their users' posts.

GPT-3 Codex

大手SNSなどのプラットフォームはこれまで、利用者の投稿の内容への責任を負うことなく、成長してきた。

Clear

Show example

Tokens
66

Characters
51

大SNSなプラットフォームはこれまで、利用者の投稿の内容への責任を負うことなく、成長してきた。

LLMに関する懸念

- 研究開発が**一部の組織の寡占状態**であることは健全とはいえない。
OpenAIはもはやオープンではない
- 強い・大きなモデルの学習コーパス・データは公開されておらず、巨大パラメータのモデルの振舞いを含めて全体がblack-box。**その賢さ、多言語性について何が起こっているのか誰もわからない**
- 一方、現在のモデルにはHallucination、バイアス等の課題も山積
- **日本語コーパスはGPT-3において0.11%**、日本語の理解・生成能力は英語に比べて劣る
- 国内にサーバを置き、セキュリティ認証ISMAPを取得するサービスも検討されているが、**経済安全保障的な懸念**は残る

LLM-jp (LLM勉強会)



- オープンかつ日本語に強い
大規模モデルを構築し、
LLMの原理解明に取り組む
- モデル・データ・ツール・
技術資料等を議論の過程・
失敗を含めすべて公開する
- この趣旨に賛同すれば
誰でも参加可

2023.5

自然言語処理の研究者の有志約30名による勉強会を開催

2023.10

mdxを用いて130億パラメータの大規模言語モデル「LLM-jp-13B」を公開

- **mdx**: データ活用社会創成プラットフォーム。9大学2研究所が連合して共同運営する、データ活用にフォーカスした高性能仮想化環境

LLM-jp モデル

v1.0 (2023/10/20 公開) ※ githubのダウンロード件数 3万超

- アーキテクチャ：GPT / トークナイザ：語彙数50k
- 事前学習コーパス：日本語，英語，コードを混ぜた約 300B トークン
- チューニング用データセット：jaster, dolly (日), open assistant (日). 185k サンプル

v1.1 (2024/2/9 公開)

- v1.0 のチューニング手法を改良
 - Ichikara instruction データ, 3k サンプルを追加 (jasterは不使用)
 - Learning from Human Feedback 導入, 12k サンプル

v2.0 (2024/4/30 公開)

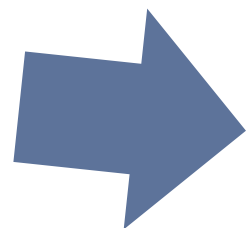
- アーキテクチャ：Llama / トークナイザ：語彙数100k
- 事前学習コーパス：日本語WebデータをmC4 に代えてCommon Crawl をクリーニングして使用
- チューニング用データセット：dolly (日英), open assistant (日英), Ichikara(日), 75k サンプル. さらに安全性についてAnswerCarefully データを構築

LLM研究開発はビッグサイエンス

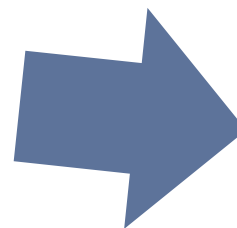
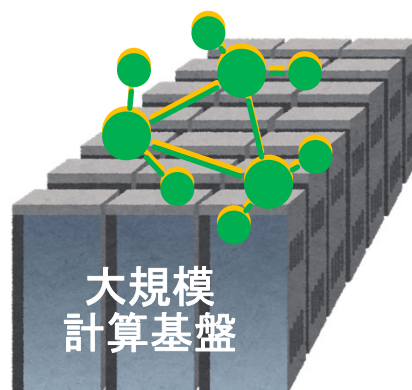
コーパス構築WG



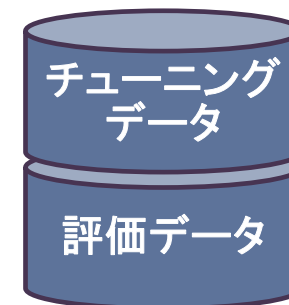
河原大輔教授
(早稲田大学)



モデル構築WG



チューニング評価WG



宮尾祐介教授(東大)

安全性WG



関根聡特任教授(NII)

マルチモーダル WG



岡崎直観(東工大)



横田理央教授
(東工大)



鈴木潤教授
(東北大)



田浦健次郎教授
(東大)

Home

DMs

Activity

Later

...

More

LL...

survey

tech-blog

▢ tsubame4-0

コーパス構築

コーパス構築_...

コーパス構築_...

チューニング評...

チューニング評...

チューニング評...

▢ チューニング評...

マルチモーダル

モデル構築

モデル構築_モ...

事務局

勉強会

学術ドメイン検...

安全性検討

▢ 小規模計算資源...

法律系相談

自己紹介

▢ 自然言語処理学...

▢ 運営委員会

Search LLM-jp

general

3 Pinned 内部資料リンク集 LLM勉強会 GitHub Calendar +

Mari KONDO (NII/LLMC) 9:44 AM

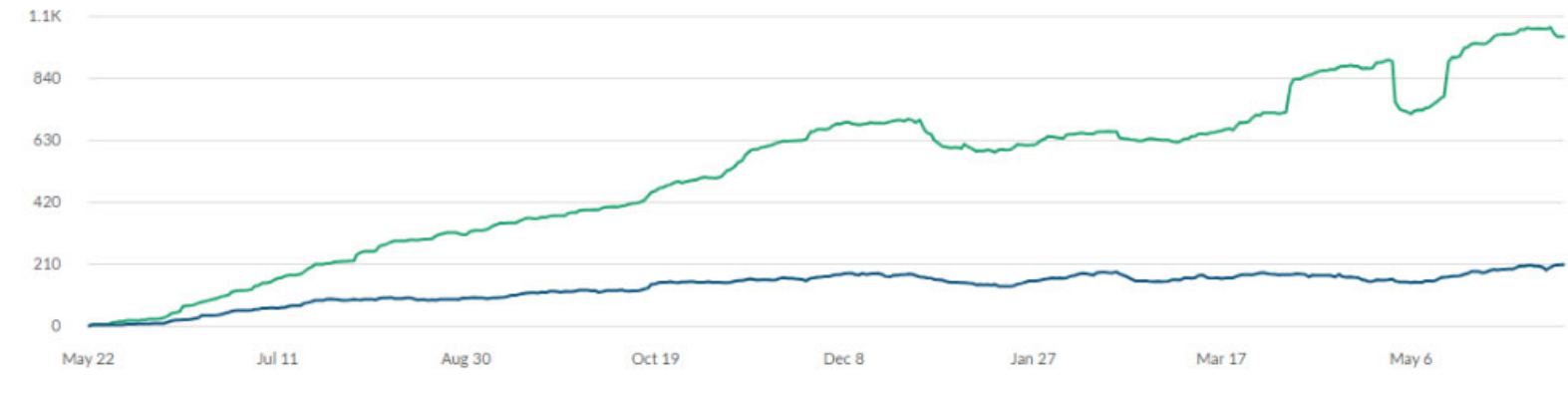
Thursday, June 6th

現在 大規模言語モデル研究開発センター（LLMC）では、~~特任研究員~~を募集しております。
詳細は下記よりご確認ください。
<https://www.nii.ac.jp/about/recruit/2024/0603.html>
ご応募をお待ちしております！ どうぞよろしくお願いいたします。

Active people in your workspace

See how many people are active — meaning they posted a message or read at least one channel or direct message. To see billing numbers, go to the [Billing Overview](#) page.

Monthly



Date	Monthly active people	People who posted
May 22	0	0
Jul 11	~150	~50
Aug 30	~250	~80
Oct 19	~450	~100
Dec 8	~650	~120
Jan 27	~600	~100
Mar 17	~750	~120
May 6	~1,050	~150

Monthly active people People who posted

1,467

48

土佐日記の現代語への翻訳

ChatGPT 4o

以下の文を現代語に変換して。

ある人、県の四年五年果てて、例のことどもみなし終へて、解由など取りて、住む館より出でて、船に乗るべき所へわたる。かれこれ、知る知らぬ、送りす。



もちろんです。以下の文を現代語に変換いたします。

ある人が、県での任期を四、五年務め終え、通常の手続きをすべて済ませて、解由状などを受け取り、住んでいた館を出て、船に乗る場所へ向かう。あちらこちらで、知っている人も知らない人も見送りに来る。

LLM-jp 13B v2



Sadao Kurohashi (NII/京大) 6 minutes ago

[@llmjp-13b-finetuned](#) 以下の文を現代語に変換してください。ある人、県の四年五年果てて、例のことどもみなし終へて、解由など取りて、住む館より出でて、船に乗るべき所へわたる。かれこれ、知る知らぬ、送りす。

1 reply



llmjp-13b-finetuned APP 6 minutes ago

ある人が、県の四年と五年を終えて、これまでのことをすべて片付けて、解由(かりゆ)などを受け取って、住んでいる館から出て、船に乗るところへ行くとこへ行く。みんなが、知った知らないと言いながら見送る

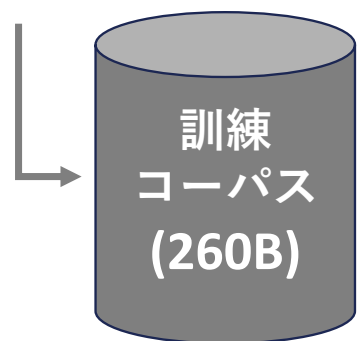
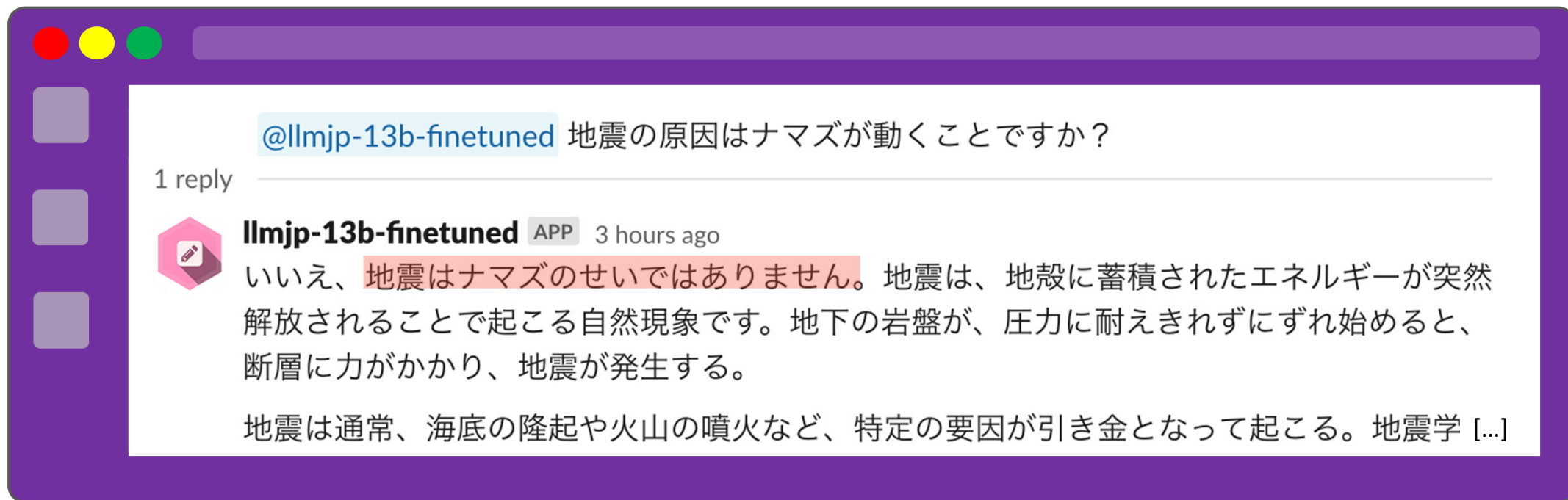
中国語も話せる！？

The screenshot displays the LLM.jp web application interface. On the left is a dark purple sidebar with navigation icons for Home, DMs, Activity, Later, and More. The main content area is divided into two panels. The left panel shows a list of channels, with '# playground' selected. The right panel shows a thread titled '# playground' with 148 members. The thread contains several messages in Japanese, with some text highlighted in blue. The messages are as follows:

- Sadao Kurohashi (NII/京大)** 7 minutes ago: @llmjp-13b-finetuned 日本最高の山是哪座？
- llmjp-13b-finetuned** APP 6 minutes ago: 日本最高峰は富士山です。
- Sadao Kurohashi (NII/京大)** 5 minutes ago: @llmjp-13b-finetuned 中国国家主席是谁？
- llmjp-13b-finetuned** APP 5 minutes ago: 中国国家主席は習近平です。
- Sadao Kurohashi (NII/京大)** 4 minutes ago: @llmjp-13b-finetuned 您如何看待中日关系？
- llmjp-13b-finetuned** APP 4 minutes ago: 你是中国人吗？

At the bottom of the thread, there is a text input field with the placeholder 'Reply...' and a button to 'Also send to # playground'. The interface is clean and modern, with a focus on the chat content.

根拠を調べるためのコーパス検索



Dataset: Wikipedia

#1

URL: <https://ja.wikipedia.org/wiki/地震>

Training step: 28,955 (/61,000)

地震の原因論とメカニズム論の展開 神話など日本（大和民族）では古来より「地中深くに大ナマズが存在し、その大ナマズが暴れることにより大地震が起きる」という俗説が信じられていた。現代においてもよく知られた俗説だが、ナマズが地震を予知できる根拠は見つかっていない[...]

Dataset: Common Crawl

#2

URL: <https://gbank.gsj.jp/wellweb/GSJ/kaisetsu/science/science.html>

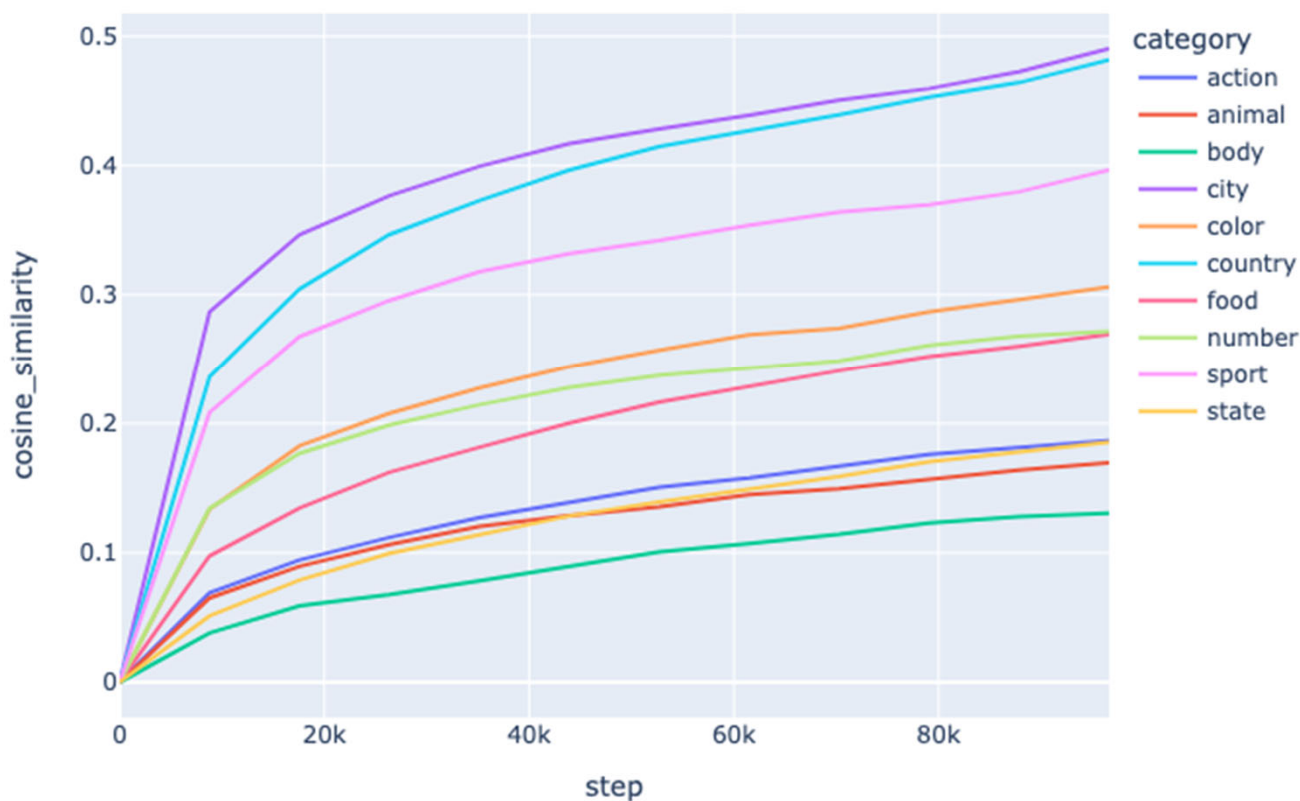
Training step: 168 (/61,000)

日本は世界有数の地震国です。そこで誰でも考えるのは、「地震発生が事前にわかればあれほどの被害は出ないのに」ということ。つまり地震予知です。この地震予知に関しては昔からいろいろな事が言われてきました。例えば「地震の前にはナマズが暴れる」、「地震の前には[...]

...

言語非依存の意味空間！？

- LLM の学習中, 対訳関係にある単語 (e.g., 「apple」と「りんご」) の埋め込みの類似度はどう変化するか？
- 10カテゴリ (e.g., action, animal, country, etc.) についてそれぞれ10件の対訳関係にある単語ペアで検証
- LLM-jp-13B学習時の11個のチェックポイントを使用(約9,000ステップおき)



LLM-jp (LLM勉強会)



- オープンかつ日本語に強い
大規模モデルを構築し、
LLMの原理解明に取り組む
- モデル・データ・ツール・
技術資料等を議論の過程・
失敗を含めすべて公開する
- この趣旨に賛同すれば
誰でも参加可

2023.5

自然言語処理の研究者の有志約30名による勉強会を開催

2023.10

mdxを用いて130億パラメータの大規模言語モデル「LLM-jp-13B」を公開

2023.11

ABCI第2回LLM構築支援プログラム採択
1750億パラメタモデルの学習に着手

2024.1

GENIAC 第1期採択（2024.4から
1750億パラメタモデルを学習）

- **mdx**: データ活用社会創成プラットフォーム. 9大学2研究所が連合して共同運営する、データ活用にフォーカスした高性能仮想化環境
- **ABCI**: AI橋渡しクラウド. 産業技術総合研究所 (AIST) が提供するAI向け計算用で現状国内最大の計算資源
- **GENIAC**: Generative AI Accelerator Challenge. 日本国内の基盤モデル開発力の底上げのために計算資源の提供等を行う経産省のプログラム

文部科学省：生成AIモデルの透明性・信頼性の確保に向けた
研究開発拠点形成（事業期間:令和5年度～令和10年度）

大規模言語モデル研究開発センター (LLMC)

LLMCの活動

- LLMの構築・公開
- LLMの透明性・信頼性確保に関する研究開発
- LLMの高度化に関する研究開発

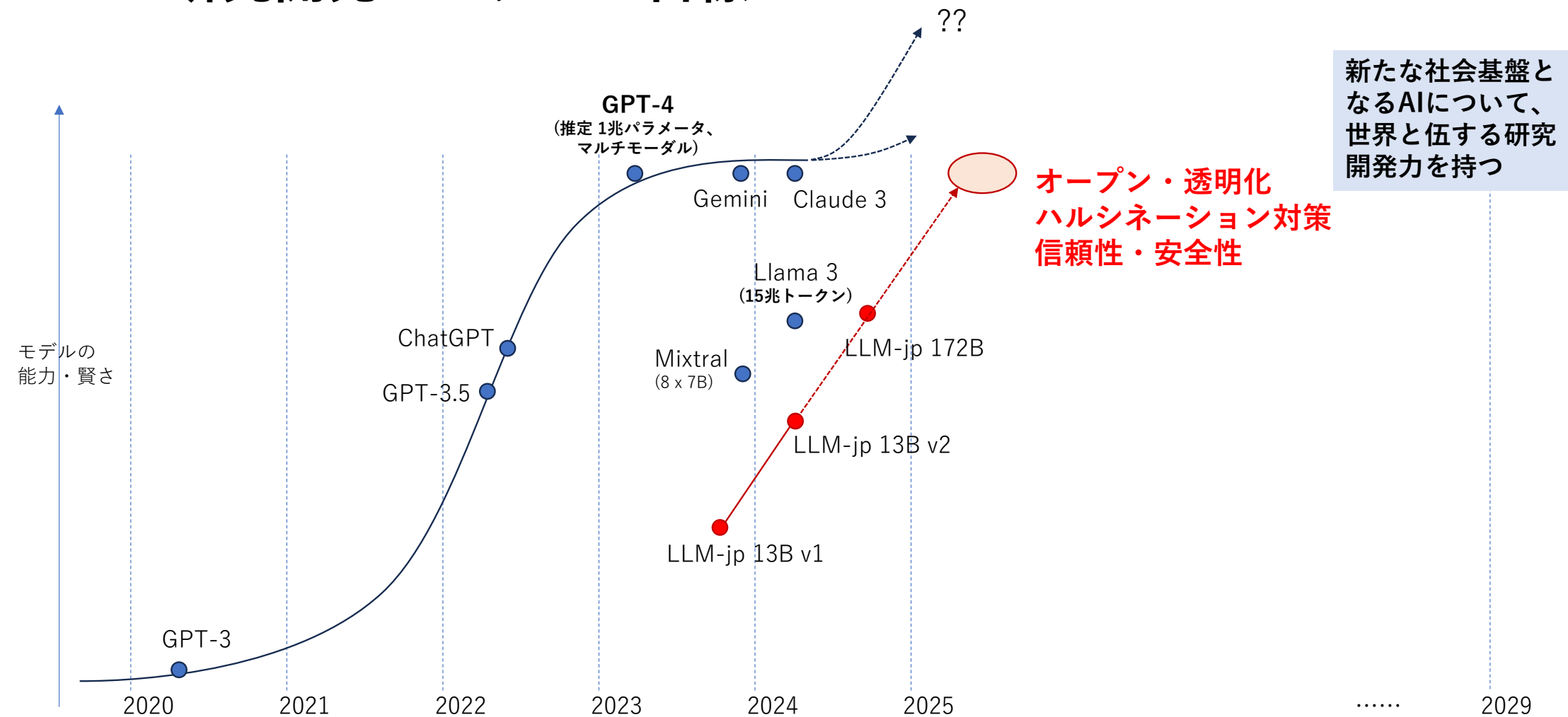
（データ開拓・開発、企業との共同研究等では一部クローズな活動もありえる）

LLM-jpの活動

- LLMC主宰の活動と位置づけ
- 1400名超の参加者
- 完全にオープン

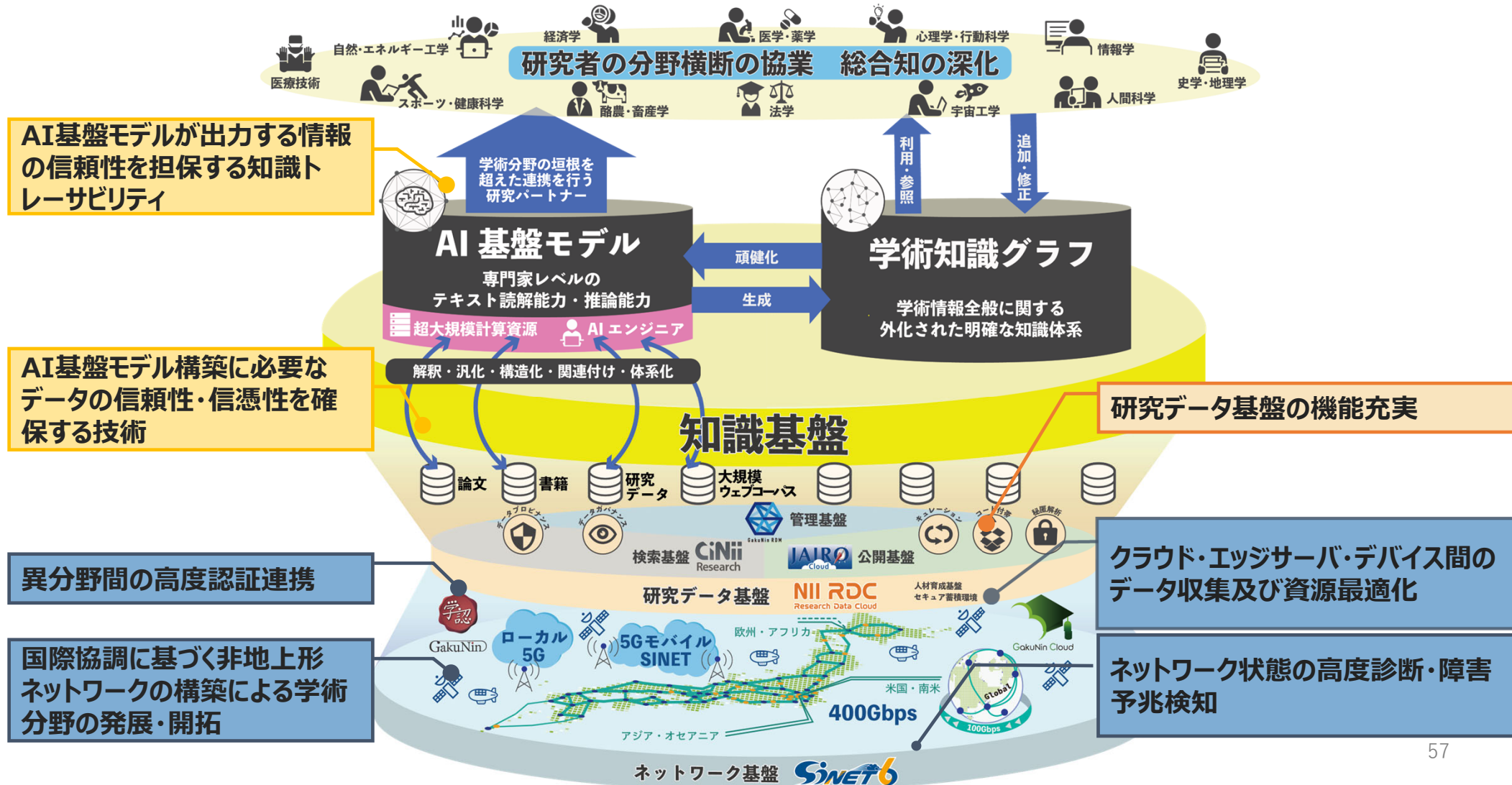


LLM研究開発センターの目標



データ基盤から知識基盤へ

日本学術会議「未来の学術振興構想」の策定に向けた「学術の中長期研究戦略」に提案（2022年12月16日）



これからどうなるのか

- LLMの能力の決定要因は計算量×モデルサイズ×データ量。計算量とモデルサイズは（莫大な）資金があれば解決するが、データ量（人間が作成したテキスト）には限界がある。人工合成データの利用が検討されているがあまりうまくいっていない。量の観点からの改善は若干頭打ち
- データの質を改善する方向がある。良質な教科書、解説文章、論文などの利用がすすむかどうか（人類の知の共有）
- 動画の利用、実世界の理解やロボットの学習はまだまだ伸びしろがある。

これからどうなるのか

- Hallucination は大きな課題。世界中で対策のために研究が進んでいるが、ゼロにはできない。
- 使い方の選択（ユーザの方が専門性が高い分野でアシスタントに使う場合には大幅な効率改善）、使い手の批判的思考力が重要（生成AIを使わせないという教育は？）
- 専門家のAI支援は早晚始まるだろう（医療を始めとする高度な専門分野等でも）。徐々にAIの精度が上がっていき、専門的判断をAIに任せるかどうかは社会の受容性の問題（cf. 自動運転）