

生成AIに関する提言に向けて

黒橋 禎夫

日本学術会議 第三部会員、情報学委員会 幹事

国立情報学研究所 所長／京都大学 特定教授

言語処理学会 会長／デジタルアーカイブ学会 理事

第26期・第5回 科学者委員会 学術体制分科会 (2024/7/2)

問題意識

- 環境問題、格差問題、地域紛争など、人類は矛盾に満ちた世界に苦しんでいる
- これらの複合的な社会課題を解決するためには、多様なバックグラウンドを持つ人々の協働が必要
 - ⇒ **オープンデータ、オープンサイエンス・イノベーション**
- この30年間、情報学が社会に広範なインパクトを与えてきた（WWW、深層学習、AlphaFold、DeepL、ChatGPT）
 - ⇒ **生成AIによる社会変革**

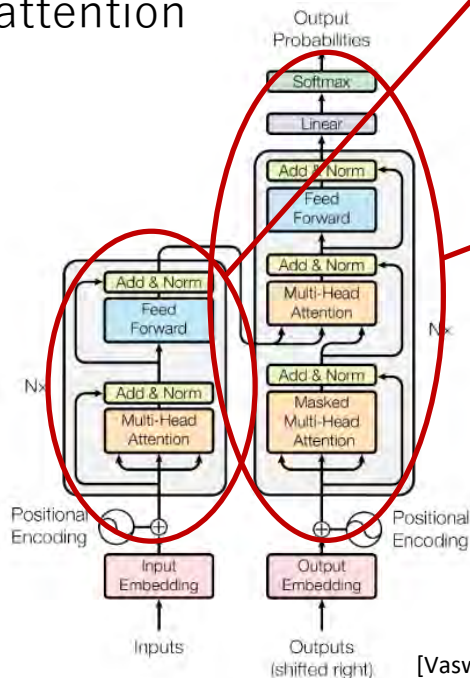
LLMの歴史

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

意味・文脈の理解と多言語性

⚡ 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。



意味・文脈の理解と多言語性

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。

意味・文脈の理解と多言語性

ChatGPT

PLUS



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。



現代語への翻訳(土佐日記)

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。



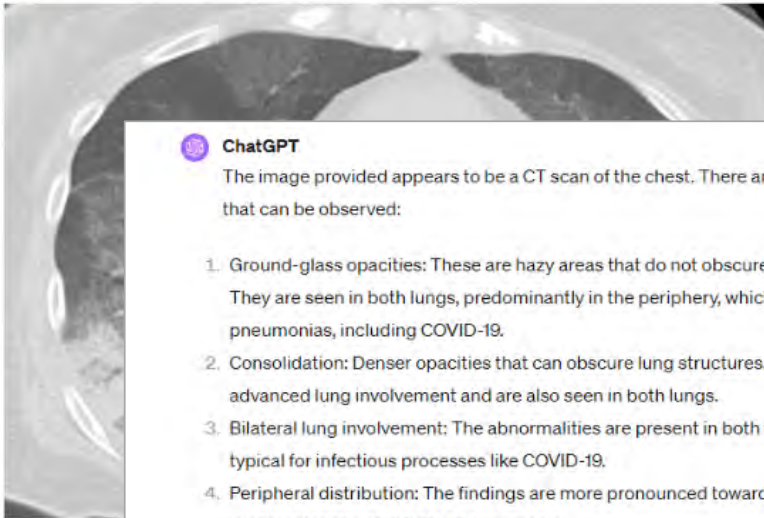
もちろんです。以下の文を現代語に変換いたします。

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。

画像の理解

ChatGPT 4

You



You are a he
COVID-19?

<https://www.covid19-jma-med>

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_0 .

Un barreau, conducteur thermique, homogène, de longueur L , de section S et de conductivité thermique λ , et sous échange thermique latéraux, relie le bolomètre à un thermostat de température T_b , (fig. 1).

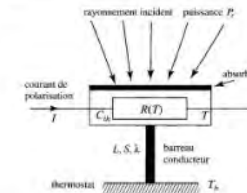


Figure 1 - Schéma d'un bolomètre.

I.1. Dans cette première partie, on étudie le comportement du bolomètre en régime stationnaire, à la température de l'ambiance $T = T_0$.

I.1.a) Exprimer la température $T(x)$ en tout point du barreau conducteur relié 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 coke can to
Taylor Swift



move coke can to
X



move bag to
Google



move redbull can
to H



move soccer ball
to basketball



move banana to
Germany

Prompt:
Given I need to
hammer a nail, what
object from the scene
might be useful?
Prediction:
Rocks. Action: 1 129 138
122 132 135 106 127





Gサイエンス学術会議の「AI and Society」共同声明

Gサイエンス学術会議: <https://www.scj.go.jp/ja/int/g8/index.html>

政策立案者への12項目の提言

1. 強力なAIシステムは、悪用されると社会に危険を及ぼす可能性があるため、サイバー攻撃や物理的な攻撃から適切に保護されなければならない。
2. AIシステムの性能が更に向上し、人間による制御により大きなリスクを与えるようになるにつれて、システムの導入を行う前に、安全性についての公的保証が求められるようになるだろう。
3. AIシステムの信頼性は極めて重要である。
4. 生成 AI はコンテンツの作成やソフトウェア開発において大きな可能性をもたらすが、AI が既存の作品を基にして外見的には新しいコンテンツを創り出した場合、知的財産権の保護に関する諸々の問題が浮上する。
5. 大規模なAIシステムは高価な計算資源を必要としており、それが参入障壁となっている。
6. 大規模なオープンソースAIモデルは、便益と人々が負わされることになる悪用から生じるリスクとの間に複雑なトレードオフ関係を生じさせる。
7. 自律型兵器システムの台頭は、急を要する倫理的かつ軍事的な懸念を生じさせている。
8. AIシステムが影響力を持つようになるにつれ、AIを人間の価値観や倫理観に合わせる事が求められる。
9. AIの恩恵を公平に分配するためには、効果的なスチュワードシップが極めて重要である。
10. 官民の協力は、責任あるAIの開発を促進する。民間企業が急速に技術革新を進める一方で、公的機関は規制と倫理の枠組みを提供し、人類の利益のために独立した長期的研究を推進する。
11. AIの責任ある安全な開発と導入は、人文科学や社会科学を含むさまざまな学問分野を横断する共同アプローチが必要である。
12. 教育は、AIの能力を明らかにし、神話を否定し、その長所と短所を比較検討すべきである。市民は技術的なリテラシーを身につけ、AIの社会的・倫理的な意味を認識すべきである。

Gサイエンス学術会議：平成17年以降、G7サミット参加国のナショナルアカデミーがG7サミットに向けて科学的な政策提言を行うことを目的で共同声明を取りまとめている国際会議で、日本からは日本学術会議が参加

提言「生成AIを受容・活用する社会の実現に向けて」

生成AIは2020年代以降急速にその技術が普及し、特に大規模言語モデル(LLM)であるChatGPTは公開後2ヶ月でアクティブユーザ数が1億人に達するなど、教育研究を含むあらゆる分野に大きな影響をもたらしている。国内においても、すでに文部科学省から2023年7月4日に初等中等教育段階における生成AIの利用について暫定的なガイドラインが発行され、大学など大多数の高等教育機関でも独自にガイドラインが策定・公開されている。その趣旨は生成AIのもつ高いポテンシャルを有効利用することの促進と、プライバシーや著作権の侵害といった利用上のリスクや不正な使用に対する注意喚起の2つの側面をもつ。後者については、2023年5月にG7関係閣僚を中心にAIの活用や開発・規制に関する国際的なルール作りを推進するプロセス(広島AIプロセス)が開始され、12月の閣僚級会合でまとめられた「広島AIプロセス包括的政策枠組み」がG7首脳に承認された。2024年2月に独立行政法人情報処理推進機構にAIセーフティ・インスティテュート(AISI)というAIの安全性評価に関する基準や手法を検討する専門の機関が設立されている。2024年4月には、Gサイエンス学術会議2024において「人工知能と社会」についての共同声明が発表された。

このような状況のもとで、学術の立場から生成AIについてのより深い洞察を提示しつつ、前述の2つの側面を最も有益な形で調和させたどのような施策をとるべきかを日本学術会議から見解として発信することは、科学技術政策にとどまらずソサイエティ5.0の描く社会全体の望ましい発展への大きな貢献となる。

想定する読者：政策立案者、研究者・エンジニア、企業や一般社会も含め生成AIに関わりのある人々

スケジュール案：

2024年 7月	執筆依頼と原稿内容具体化(7/25頃オンラインで確認)
8月末	第一稿脱稿
9月	委員会において審議、学術会議会員・学会へのヒアリング
10月	委員会において見解案の取りまとめ
11月末	提言公表

提言「生成AIを受容・活用する社会の実現に向けて」

※ 声明番号はGサイエンス学術会議の
「AI and Society」共同声明での対応する項目

1. はじめに (1ページ)

2. 生成AIの現状と動向 (8ページ)

2.1 生成AIと基盤モデル

2.2 生成AIの発展と実用化

2.3 各種のモダリティ(テキスト、プログラムコード、静止画・動画、音声や音楽)

2.4 今後の動向

3. 生成AIの脅威や課題 (8ページ)

3.1 生成AIの脅威 (声明2)

3.2 生成AIモデル開発の障壁 (声明5)

3.3 生成AIモデル運用に伴うリスク (声明4、6、7)

3.4 生成AIモデルの備えるべき要件 (声明3)

4. 生成AIの活用による波及効果 (6ページ)

4.1 科学技術の発展に対する効果

4.2 産業分野への効果

4.3 社会的な波及効果 (教育、人材育成含む)

5. 提言 (6ページ)

5.1 生成AIモデルの適切な運用 (声明1、8、9、11)

5.2 生成AIモデル開発の望ましい体制(声明10)

5.3 生成AIモデルを適切に利活用できるリテラシー教育(声明12)

参考文献

参考資料 AI事業者ガイドライン(総務省、経産省)

オープンという考え方

生成AIにおけるオープンさの重要性

- ChatGPTをはじめとする生成AIは社会に衝撃を与え、本格的な利用が徐々に始まっている
- 入力に対して極めて高い理解力をもつ一方で、事実と異なる内容を出力してしまうHallucinationという課題があり、当面は使い方、利用者の批判的思考が重要
- 超大規模LLM(1兆パラメータ級)の学習は膨大なコストがかかるため、一部の組織の寡占状態で、内容（データ、モデル、ノウハウ）は非公開
- 「OpenAIの現・元従業員ら、AIシステム構築の無謀さと秘密主義を告発」（2024/6/7）

<https://www.itmedia.co.jp/news/articles/2406/05/news100.html>

生成AI学習におけるオープンデータの重要性

- クローズなデータはあってよい（企業内データ、要配慮個人情報等）
- （一旦）オープンになっているデータは社会の発展のためにAI学習に活用すべき（ウェブ、論文、（公共）放送等）
- 著作者の権利への配慮が必要なものは
 - AI学習に供するのは1/3～1/2程度とし
 - AIの出力に対して、学習データへの適切なリンク（リファレンス）を提示

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（現在は1500名超）

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」を構築・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
パラメタ級モデルを学習）

国立情報学研究所内に「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」として
大規模言語モデル研究開発センターを設置

- **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

LLM-jp 13B：根拠を調べるためのコーパス検索機能

@llmjp-13b-finetuned 地震の原因はナマズが動くことですか？

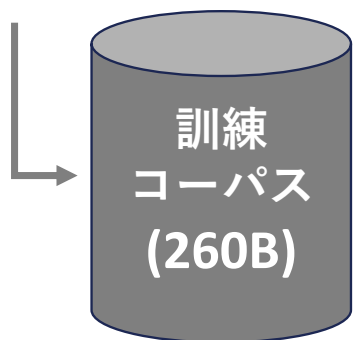
1 reply



llmjp-13b-finetuned APP 3 hours ago

いいえ、地震はナマズのせいではありません。地震は、地殻に蓄積されたエネルギーが突然解放されることで起こる自然現象です。地下の岩盤が、圧力に耐えきれずにずれ始めると、断層に力がかかり、地震が発生する。

地震は通常、海底の隆起や火山の噴火など、特定の要因が引き金となって起こる。地震学 [...]



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...

...

生成AIとオープンデータの相乗効果

- オープンで良質なデータ・知識が生成AIをより良いものにする
- より良い生成AIは人類に様々な恩恵をもたらす
 - 文化的、言語的な多様性の理解の上にたった従来とは比べものにならないレベルの自動翻訳を実現し、人々の異文化への関心を高め、理解を促進する
 - データ・知識を解釈・昇華するとともに、人々の知的活動や専門家の協働を促進・強化する
- その結果、人類はより良いデータ・知識を生み出すことができる