

NICTの大規模言語モデルとその周辺

情報通信研究機構(NICT)

鳥澤 健太郎

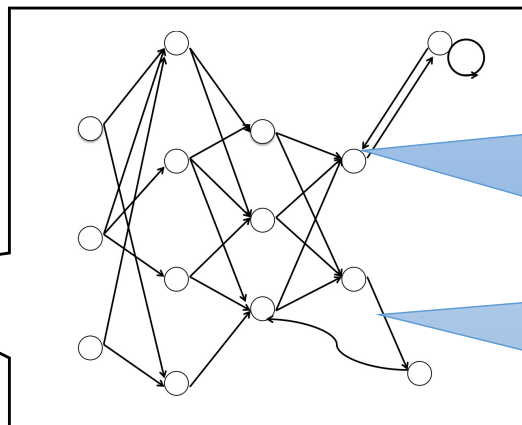
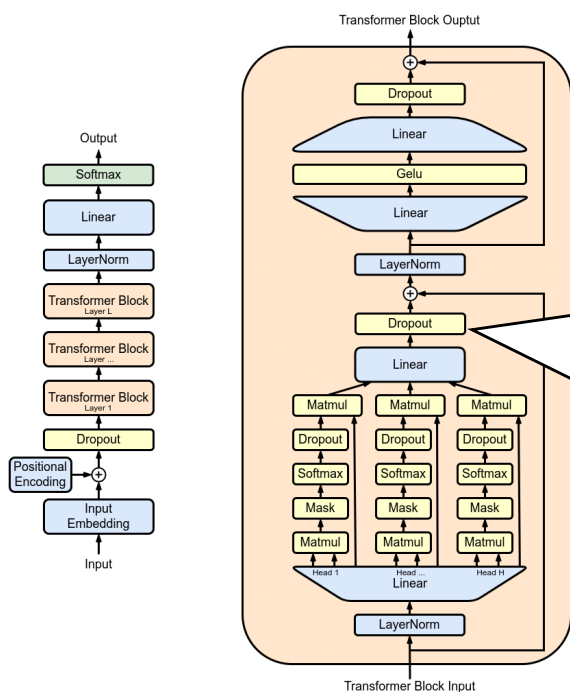
2023年11月1日

1. 大規模言語モデルとは何か？
2. NICTでの開発状況
3. 学習データの重要性
4. 大規模言語モデルのインパクトとリスク

1. 大規模言語モデルとは何か？
2. NICTでの開発状況
3. 学習データの重要性
4. 大規模言語モデルのインパクトとリスク

- 米国OpenAI社のChatGPTを代表とする新種のAI
 - 特徴
 - ユーザの入力に従って、あたかも人が書くようなテキストを生成
 - 様々な機能が「創発」、日々新たな使い方が発見される、これまでにないAI
 - 質問応答、要約、文書作成、創作、対話、プログラム作成...
 - これまでのAIは作り込みが必要だった
 - 目的とする機能、動作が実現するようにシステムの設計や、学習データの設計、構築をする必要
 - 大規模言語モデルの最大の特徴は、こうした作り込みは全くしない、もしくは最小限であっても様々な機能が創発していること
 - この「様々な機能の創発」はこれまでにないチャンスやリスクの両方をもたらす
→全世界的に「世界を変える」「規制が必要」と言われる所以
 - Webを中心とする大量のテキストから学習
 - Webに書かれている人間のアクティビティを広範に学習


- 最近言われている大規模言語モデルは概ねGPT(Generative Pre-trained Transformer)のことを指す
 - より具体的には、Transformerというアーキテクチャに基づく、超巨大なニューラルネット
 - Transformerとは、元々機械翻訳用に開発されたニューラルネットのアーキテクチャ
- 生成AIとか、LLM (Large Language Model)とも呼ばれたりもする
- 入力されたテキストの次の単語を確率的に予測する、という点は不変

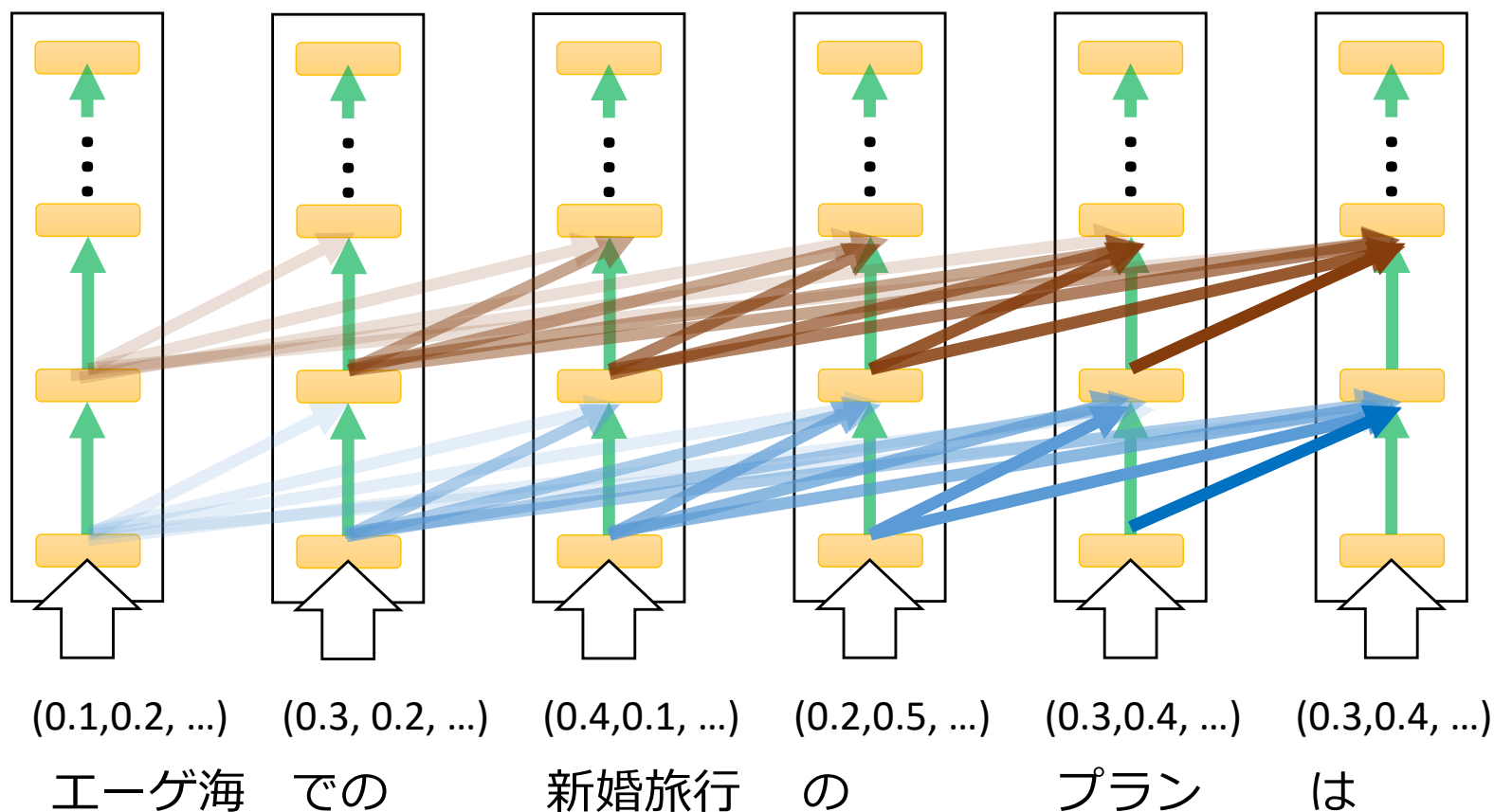


各々の○はニューロン（脳細胞）に対応。一つのニューロンは結線（矢印）を介して伝搬してくる入力信号に関して、四則演算程度の単純な計算をし、その結果を結線(矢印)に出力

各々の結線(矢印)には「重み」(パラメータ)がつく。
このパラメータ、結線が多ければ多いほど「賢い」傾向が

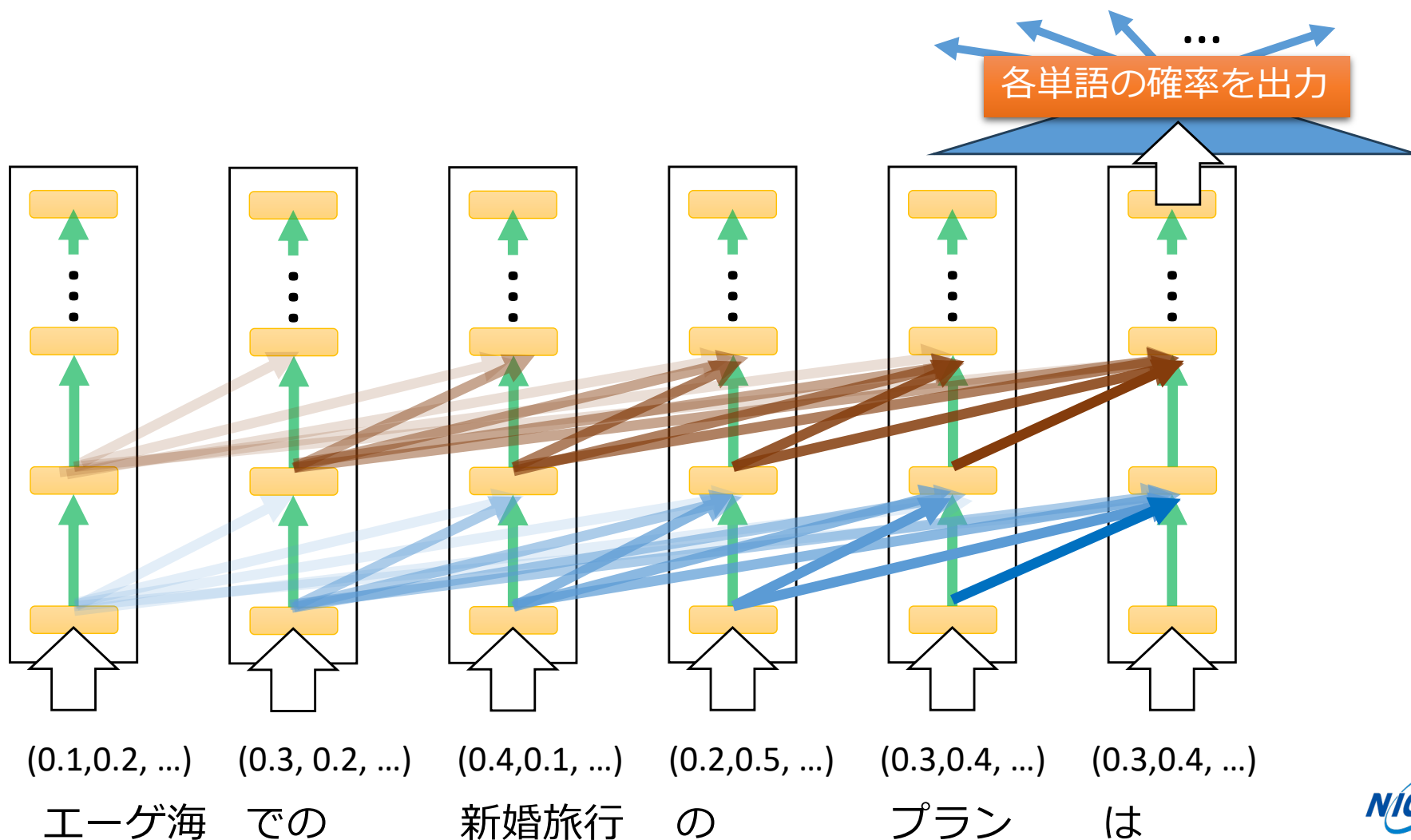
左図のtransformerのブロックダイアグラムは、Wikipedia日本語版の「GPT(言語モデル)」の項より転載[https://ja.wikipedia.org/wiki/GPT_\(言語モデル\)](https://ja.wikipedia.org/wiki/GPT_(言語モデル)) CC-BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.ja> に則り転載

- 「エーゲ海での新婚旅行のプランは」と入力してみると、その続きのテキストを生成
- 単語はベクトル (= ) に変換
- ニューラルネットの下から上に向けて、単語のベクトルを変形しつつ、周囲の単語のベクトルを融合することを（数十回から数百回）繰り返す



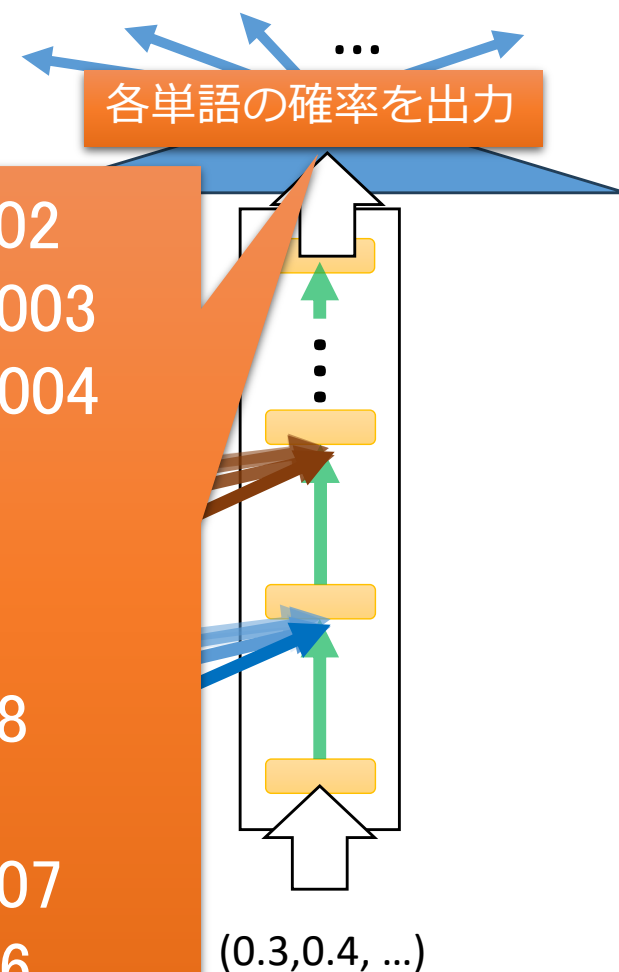
• 続きの単語を生成するには？

- 一つ前の単語（この場合は「は」）に関する出力のベクトルをもとに、辞書にあるすべての単語が次の単語として現れる確率を計算
- 基本的には確率最大の単語を「次の単語」として生成



• 続きの単語を生成するには？

- 一つ前の単語（この場合は「は」）に関する出力のベクトルをもとに、辞書にあるすべての単語が次の単語として現れる確率を計算
- 基本的には確率最大の単語を「次の単語」として生成



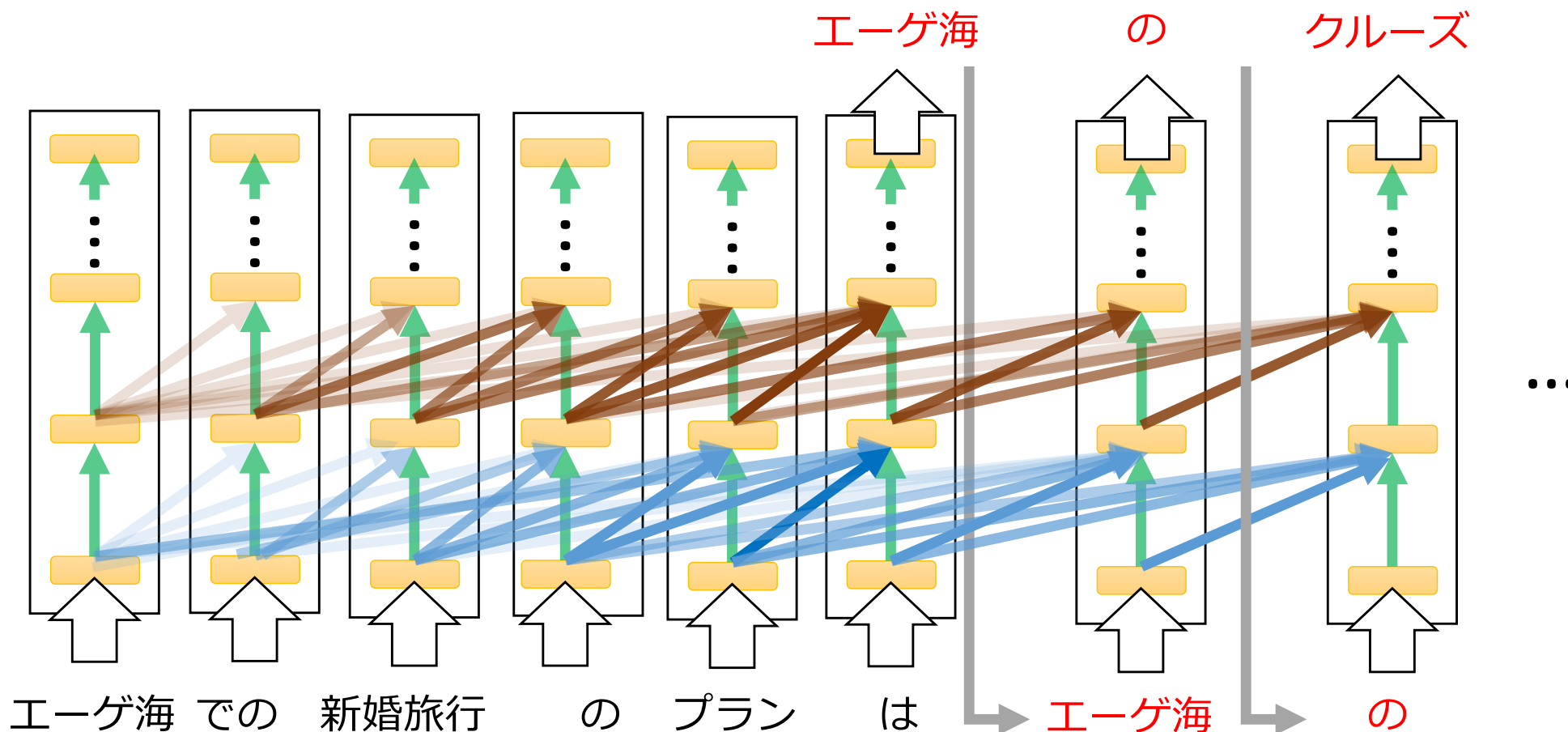
辞書にある数万語から数十万語のすべての単語について確率を計算。基本的には最大のもの（この例では「エーゲ海」）を選ぶ

| | |
|-------------|----------|
| 「あ」の確率 = | 0.00002 |
| 「あい」の確率 = | 0.000003 |
| 「愛」の確率 = | 0.000004 |
| ... | |
| 「エーゲ海」の確率 = | 0.001 |
| ... | |
| 「クルーズ」の確率 = | 0.0008 |
| ... | |
| 「、」の確率 = | 0.00007 |
| テキスト終了の確率 = | 0.0006 |

エーゲ海での新婚旅行のプラン

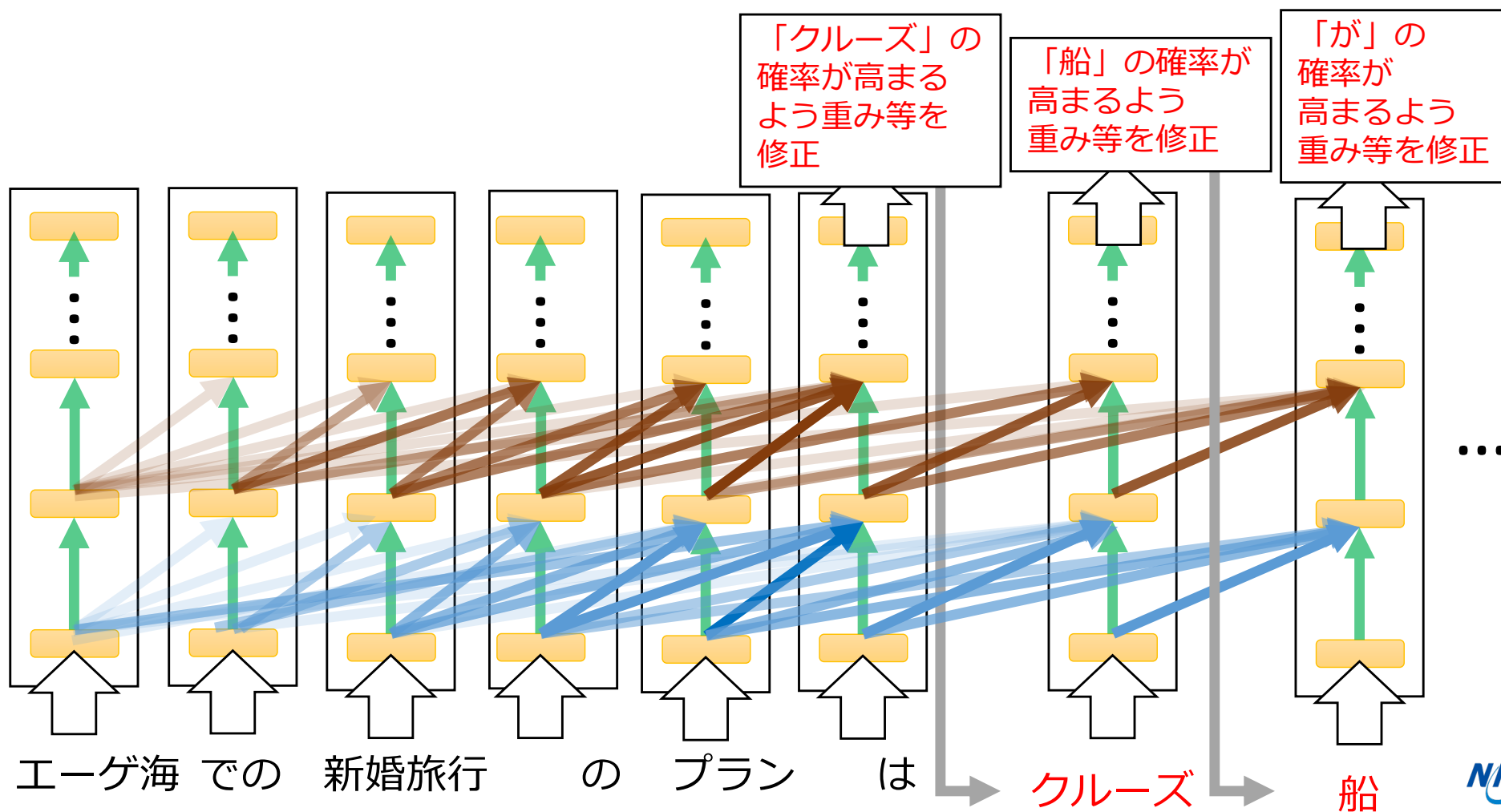
は

- 基本的には、以上の動作を繰り返すことで長いテキストを生成
 - 生成された単語は入力の末尾に追加される



学習はどのようにするのか？

- 学習データの「正解テキスト」中の単語の確率が高くなるように、ニューラルネットの重みや単語ベクトルを修正
- これを大量の正解テキストに対して繰り返す
- 例えば、学習データ中の正解テキストが「エーゲ海での新婚旅行のプランはクルーズ船がおすすめ」であったとすると、



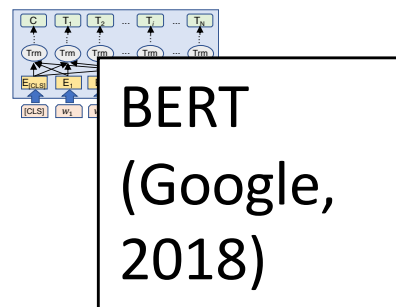
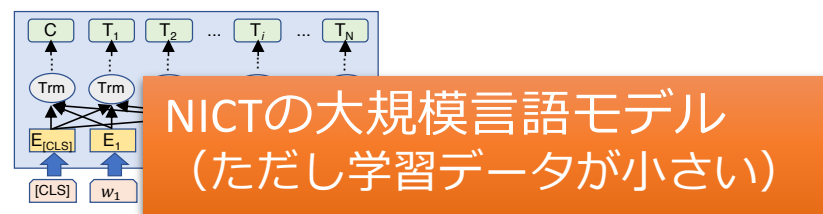
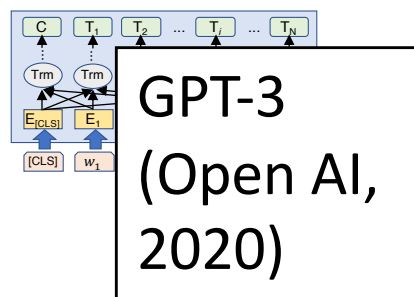
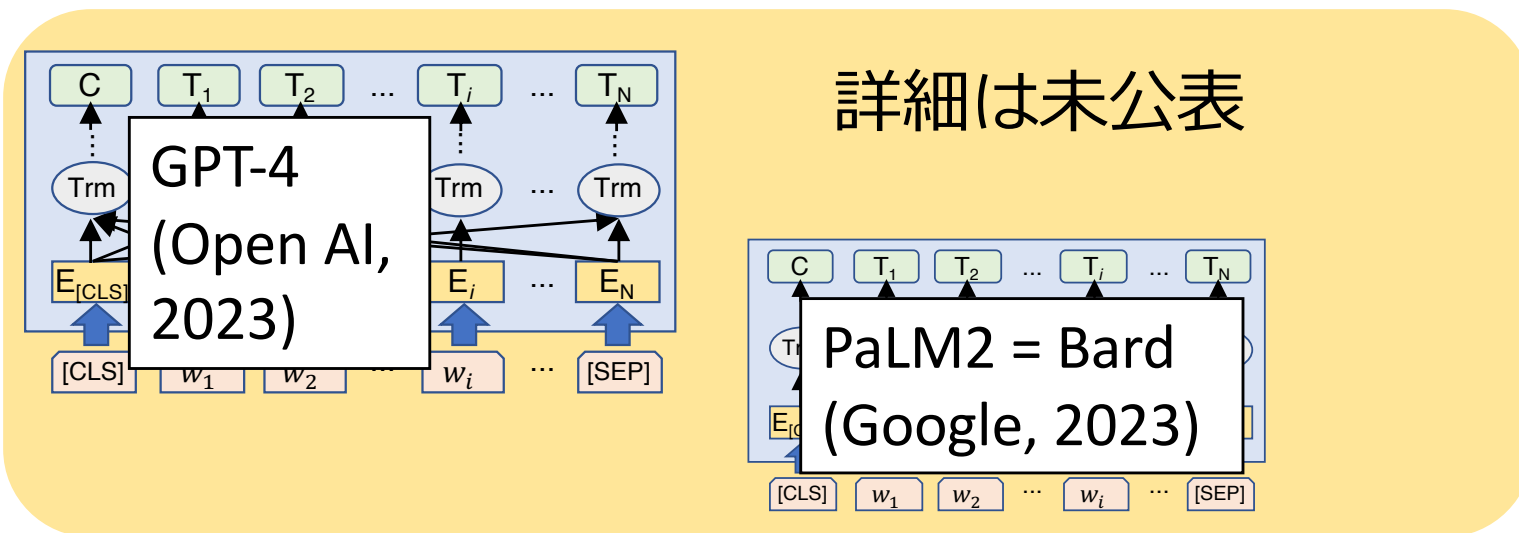
パラメータ数

一般にパラメータ数が
大きいほど高機能
一方、学習も大変

1兆個以上？

1,750億個

1億～4億個



- 最新のLLMでは、大量の計算機を用いて数か月単位の期間、学習。一回の学習に要する所要のコストは数億円～数十億円以上とみられる
- 新聞数万年分以上に相当する膨大な学習データを使用

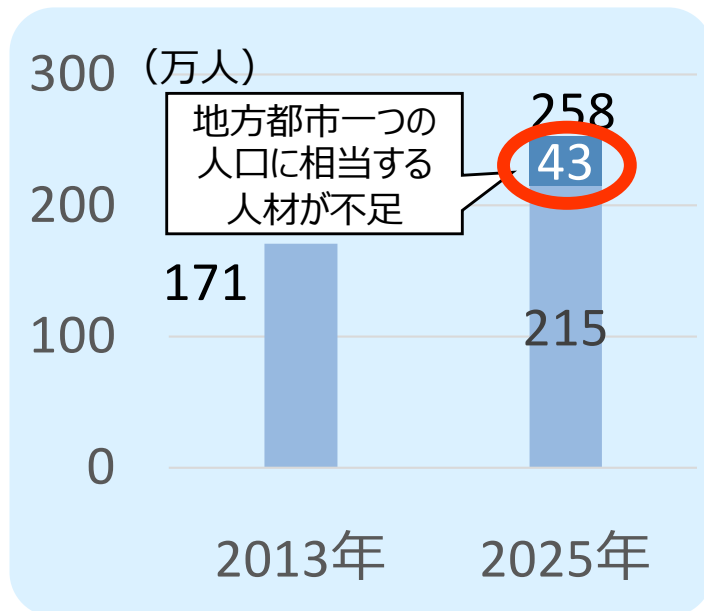
1. 大規模言語モデルとは何か？
2. NICTでの開発状況
3. 学習データの重要性
4. 大規模言語モデルのインパクトとリスク

- 深層学習、大規模言語モデルも用いた大規模自然言語処理システム、対話システム、基盤技術



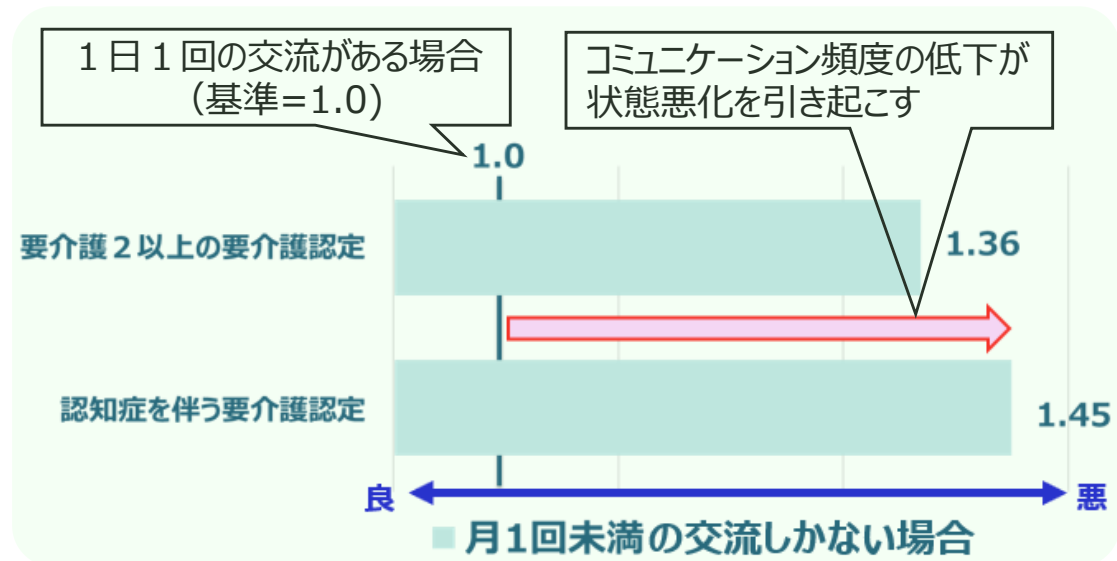
- 高齢者介護体制逼迫の可能性は非常に高い
- AIによる介護作業の負担軽減は待ったなし
- 高齢者のコミュニケーション不足、孤独も大きな課題

介護人材の需給ギャップ^o (推計)



総務省「平成22年国勢調査」、国立社会保障・人口問題研究所「日本の地域別将来推計人口(平成25年3月推計)」、厚生労働省「平成25年介護サービス施設・事業所調査」、厚生労働省「平成26年度介護給付費実態調査」、厚生労働省「2025年に向けた介護人材にかかる需給推計(確定値)(都道府県別)」より作成

同居者以外の他者との交流頻度と健康指標の関連 (1日1回程度の交流がある場合とのハザード比)



斎藤雅茂、近藤克則、尾島俊之、平井寛、"健康指標との関連からみた高齢者の社会的孤立基準- 10年間のAGESコホートより", 日本公衆衛生雑誌2015; 62(3): 95-105, doi:10.11236/jph.62.3_95 (表3より一部抜粋)

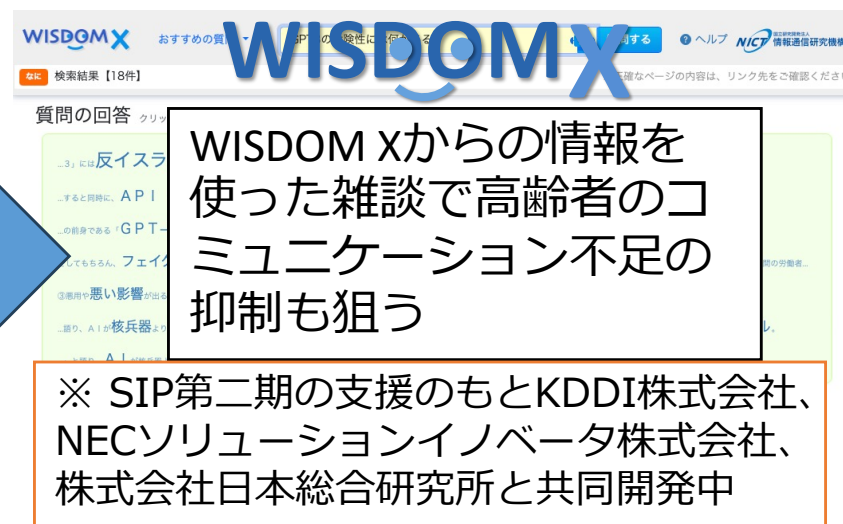
- 前例のない高齢化による介護関係者の人手不足に対して、最も時間のかかる作業である介護モニタリング（面談による健康状態チェック）を一部代替する

マルチモーダル音声対話システム**MICSUS**を開発

- 大規模実証実験によって**実用レベルの性能**であることを確認



MICSUSの柴犬型端末。
スマホ版も開発



WISDOM Xからの情報を使った雑談で高齢者のコミュニケーション不足の抑制も狙う

※ SIP第二期の支援のもとKDDI株式会社、NECソリューションイノベータ株式会社、株式会社日本総合研究所と共同開発中

- 300万件以上の対話用データを人手で構築。BERTを学習させた
- GPU一枚(V100)で五千人と同時に対話できる等、ChatGPT等に比して運用コストは極めて低い → 高齢者介護という目的からして極めて重要
- 実証実験の様子やその他の詳細は、以下をご覧ください
https://direct.nict.go.jp/news/#MICSUS_NEWS_20230308

- ② **音声認識誤りに頑健な独自開発の言語モデルHBERTを300万件以上の独自学習データで学習させ、遠回しな言い方でも正しく解釈可能**
 - GPU(V100)一枚で五千人と同時に対話できる等、ChatGPT等に比して運用コストは極めて低い
- ③ WISDOM XのWeb情報から、ユーザ発話に関するチャンスやリスクなど**新規情報を含む雑談的応答**生成
- ⑤ **日本全国の179名の高齢者を対象とする、計927回、計95時間の対話で実用レベルの性能を確認**
 - 高齢者の発話だけで計1万3千回
 - 意味解釈精度、適切な応答の割合はいずれも約93%。対話の中断、取りやめは一度もなし
 - 雑談応答の半分以上に積極的な興味を示す応答、1/4以上に笑顔。
 - 事後の五段階アンケートの平均スコアは4.2（5が最良）

高齢者とMICSUSの実際の対話例

健康状態チェックの質問に**間接的回答**をしても正しく解釈



Web情報をベースに、何らかの**新規情報**を含む雑談を生成
→ **高齢者のコミュニケーション不足抑制**

毎日どれくらい水分をとっているかわかりますか？
大体1.5Lから2L程度だと思います。

いつもの生活でどれくらい水分を取っているかご自身で分かっていると安心ですね。

どんなおかずが好きですか？

えー、野菜サラダです。

好きなおかずは野菜サラダなのですね。
ネットでは、**野菜サラダに卵を追加**すると体内でビタミンEのより良い吸収を促進するという情報がありますよ。

ああそういえば、よくゆで卵を使っています。

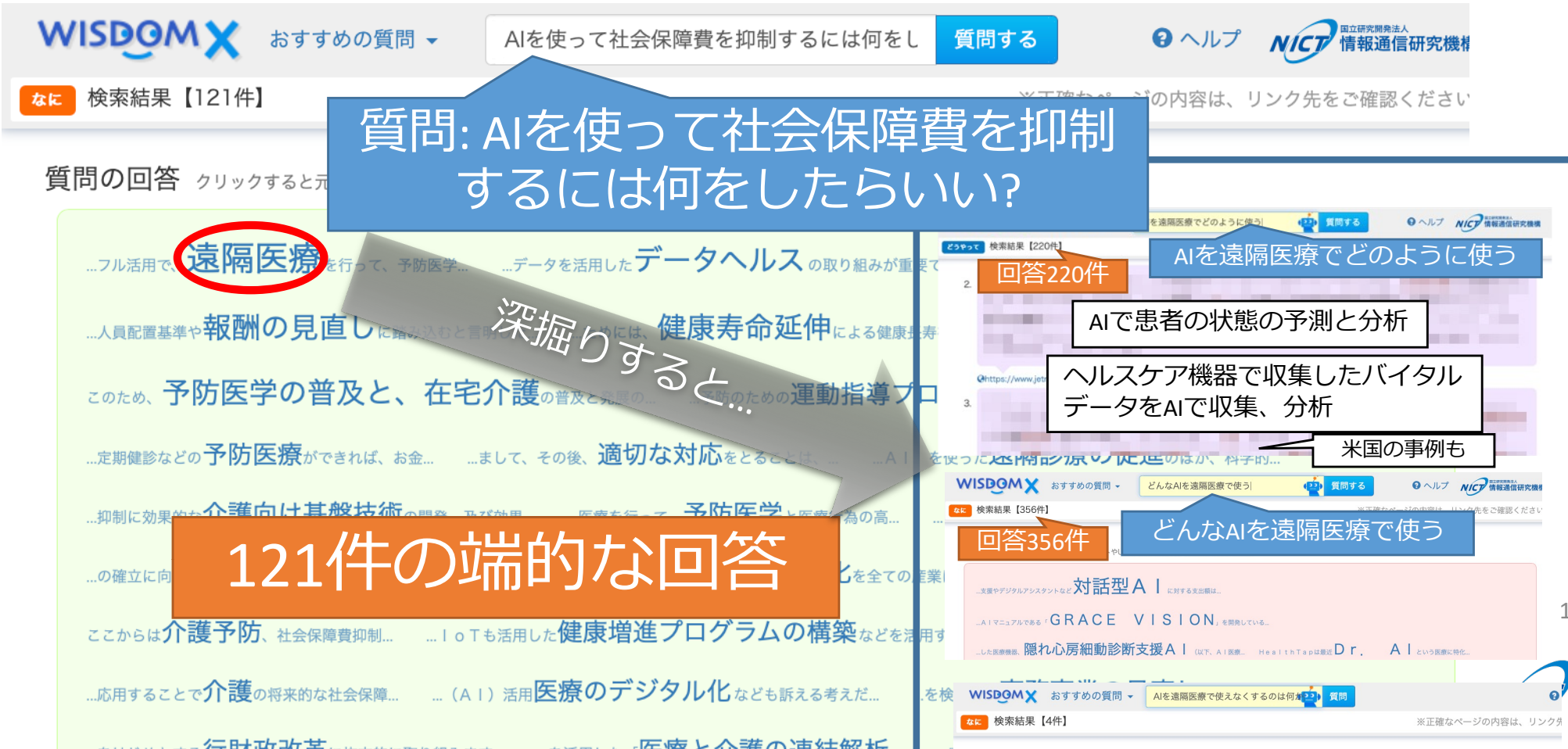
ネットの情報なので嘘かもしれませんが、野菜サラダの中で**シーザーサラダはカロリーが高い**サラダという情報がありますよ。

健康状態チェックの質問は、厚労省が普及を進めている「適切な**ケアマネジメント手法**」に基づく



カメラで**顔画像**から高齢者の感情も認識

- Web160億ページの情報をもとに多様な質問文にネットの検索結果を表示して回答
- 語句の単純検索ではなく、BERT等の深層学習を駆使してテキスト間の同義、因果関係等を自動認識し、質問への回答や、一連の世界初の技術により仮説的シナリオの作成や質問の提案まで行う
- 回答が書かれたWebテキストを提示するため、全くの偽情報が出力される可能性は大規模言語モデルよりは低い
- 数十万件の質問応答用学習データでBERTをファインチューニング（学習）
- <https://www.wisdom-nict.jp>にて一般公開中



The screenshot shows the WISDOMX website interface. At the top, there is a search bar with the text "AIを使って社会保障費を抑制するには何をし" and a "質問する" button. Below the search bar, the search results are displayed, showing "検索結果【121件】".

Annotations on the screenshot include:

- A blue callout box containing the question: "質問: AIを使って社会保障費を抑制するには何をしたらいい?"
- A red circle around the text "遠隔医療" in the search results.
- A grey arrow pointing from the question to the search results with the text "深掘りすると..."
- An orange box at the bottom containing the text "121件の端的な回答".
- On the right side, there are several callout boxes:
 - A blue box: "AIを遠隔医療でどのように使う"
 - An orange box: "回答220件"
 - A white box: "AIで患者の状態の予測と分析"
 - A white box: "ヘルスケア機器で収集したバイタルデータをAIで収集、分析"
 - A white box: "米国の事例も"
 - A blue box: "どんなAIを遠隔医療で使う"
 - An orange box: "回答356件"

- 深層学習、大規模言語モデルも用いた大規模自然言語処理システム、対話システム、基盤技術

大規模Web情報分析システム

WISDOMX

対災害SNS情報分析システム

DISAANA / D-SUMM

防災チャットボット

SOCDA

- これらのシステムでは2020年にNICTが構築した高精度なBERT（4億パラメータの識別系大規模言語モデル）を活用
 - 2022年末まで、我々の知る限り、同規模のモデルでは日本語で世界最高性能
- 生成系の大規模言語モデルは偽情報、差別等のバイアス等、様々なリスクから基礎研究と位置付け
- ChatGPTの出現で方針を変更し、今年の3月より本格的に開発開始

Web60億ページの情報を使って音声の質問に回答し、雑談も行う。対話時の省略等文脈処理も実施

介護支援のための対話システム。Webを用いた雑談で高齢者のコミュニケーション不足も解消。KDDI、NECソリューションイノベータ、日本総研と共同開発。

民間企業が商用化検討中、SIP終了後も開発継続中

ここ半年でそれぞれ130億個、400億個、1,790億個のパラメータを持つモデルを試作。1790億個のパラメータはOpenAIのGPT-3並み

NICT 大規模言語モデル (179B fp16, 質問モード)



エーゲ海での新婚旅行のプランは、

エーゲ海での新婚旅行のプランは、エーゲ海の島々を巡るクルーズが人気です。エーゲ海の島々は、ギリシャ神話の舞台となった場所が多く、その神話の世界を巡るクルーズは、新婚旅行にはピッタリです。また、エーゲ海の島々は、ギリシャ神話の舞台となった場所だけでなく、その美しい景観から、世界遺産に登録されている場所も多く、新婚旅行で訪れるには、ピッタリの場所です。エーゲ海の島々は、ギリシャ本土から離れているため、クルーズで巡るのが一般的です。

AIで科学研究を効率化するには、

AIで科学研究を効率化するには、AIが人間の研究者のように、自分の研究テーマを自分で決め、自分で研究を進め、自分で成果を発表することが求められる。

トマトソースのスパゲティの作り方は以下の通り：

トマトソースのスパゲティの作り方は以下の通り：

1. フライパンにオリーブオイルを入れ、ニンニクを炒める。
2. ニンニクがきつね色になったら、トマトの水煮缶を入れ、トマトを潰しながら炒める。
3. トマトが煮崩れてきたら、塩、コショウ、バジルを入れ、味を調える。
4. 茹で上がったスパゲティを入れ、よく混ぜる。

本居宣長の映画のあらすじは、

本居宣長の映画のあらすじは、江戸時代の国学者である本居宣長が、古事記の研究をするために、松阪にある鈴の屋という家に滞在していた時の話です。本居宣長は、鈴の屋に滞在している間に、鈴の屋の娘である「お鈴」と恋に落ちます。しかし、本居宣長は、お鈴との恋をあきらめ、お鈴のことを忘れるために、鈴の屋を出て、江戸に向かいます。

皆がAIに相談して仕事をするようになれば、過疎や情報格差が心配な地方はどうか？

皆がAIに相談して仕事をするようになれば、過疎や情報格差が心配な地方はどうか？
というと、AIが仕事をしてくれるので、地方にいても都会と同じ仕事ができるようになる。

奈良の鹿は、毎年6月ごろに赤ちゃんが生まれるんですね。私も以前、ニュースで見て、実際に奈良に見に行ったことがあります。鹿の赤ちゃんは、ほんとに小さくてよちよちとお母さんを追いかけたりしていて可愛かったです。他の大人の鹿たちも赤ちゃんの様子を見守っていて、心が温まりました。でも、あまり赤ちゃんに近づくとお母さんが怒って攻撃することがあるらしく、注意が必要だそうです。ご参考までに。／要点：

奈良の鹿は、毎年6月ごろに赤ちゃんが生まれるんですね。私も以前、ニュースで見て、実際に奈良に見に行ったことがあります。鹿の赤ちゃんは、ほんとに小さくてよちよちとお母さんを追いかけたりしていて可愛かったです。他の大人の鹿たちも赤ちゃんの様子を見守っていて、心が温まりました。でも、あまり赤ちゃんに近づくとお母さんが怒って攻撃することがあるらしく、注意が必要だそうです。ご参考までに。／要点：奈良の鹿は、毎年6月ごろに赤ちゃんが生まれるので、見に行くといいですよ。

健康でストレスなく暮らすコツを3つ挙げると

健康でストレスなく暮らすコツを3つ挙げるとしたら、

1. 自分の体の声を聞く
2. 自分を大切にすること
3. 自分を愛するということ

高齢者向け対話システムを普及させるためには以下のような戦略がある：

高齢者向け対話システムを普及させるためには以下のような戦略がある：

1. 高齢者向け対話システムの開発者は、高齢者のニーズを理解し、高齢者のニーズに合った対話システムを開発する必要がある。
2. 高齢者向け対話システムの開発者は、高齢者のニーズに合った対話システムを開発するために、高齢者のニーズを理解する必要があります。



本居宣長の映画のあらすじは、



40B

本居宣長の映画のあらすじは、国学者の本居宣長が、江戸時代の終わり頃に松阪の地で、国学の研究に没頭し、「古事記伝」を完成させるまでの生涯を描いたものです。松阪市では、この映画の上映をきっかけに、松阪の歴史や文化を再認識し、松阪の魅力を全国に発信するため、「松阪市映画上映実行委員会」を設立し、映画の上映を行います。この映画の上映にあたり、映画の上映を盛り上げるための企画を募集します。

松阪市上映実行委員会なる架空の組織まで踏み込む



179B, bf16

本居宣長の映画のあらすじは、松阪市の国学者・本居宣長の生涯を描いた映画「鈴の雫」の製作発表が、松阪市で行われました。

映画のタイトル「鈴の雫」に言及



179B, fp16

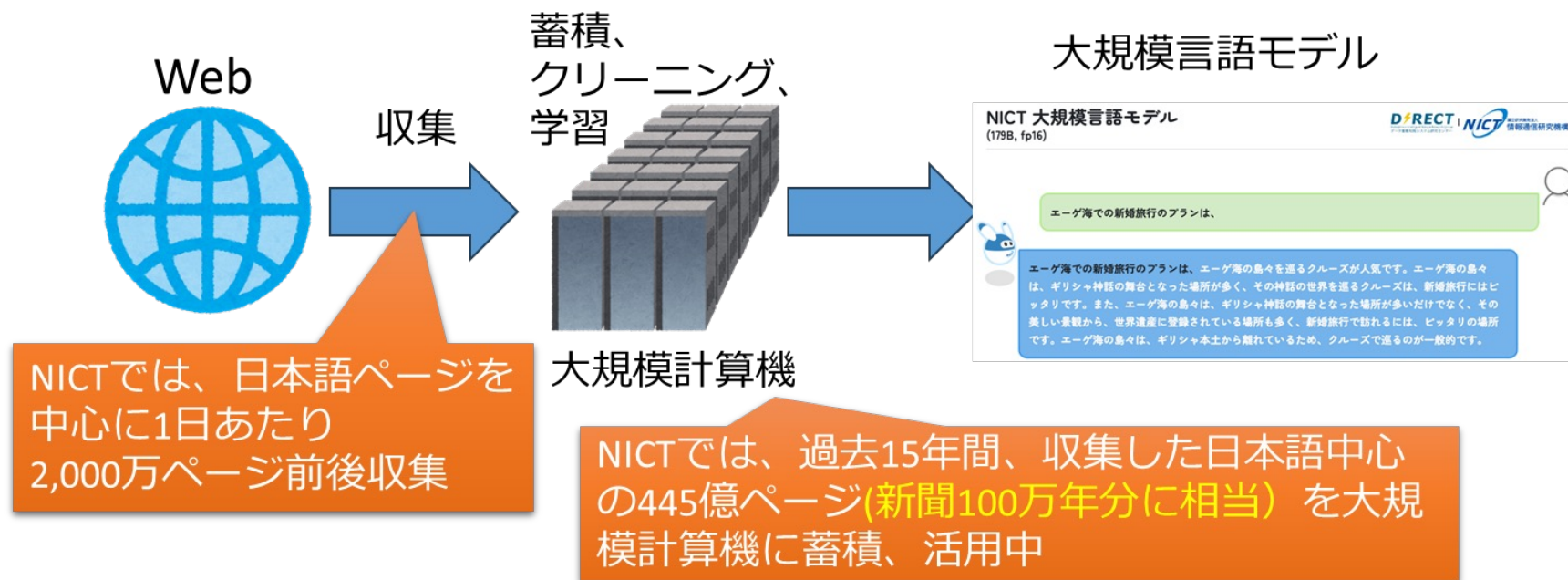
本居宣長の映画のあらすじは、江戸時代の国学者である本居宣長が、古事記の研究をするために、松阪にある鈴の屋という家に滞在していた時の話です。本居宣長は、鈴の屋に滞在している間に、鈴の屋の娘である「お鈴」と恋に落ちます。しかし、本居宣長は、お鈴との恋をあきらめ、お鈴のことを忘れるために、鈴の屋を出て、江戸に向かいます。

どういうわけか、ラブストーリーに

- NICTの大規模言語モデルをOpenAIのものに比べると、
 - 学習が途中で失敗するケースがあり、学習の完走を優先したため、
 - パラメータ数(=130億個、400億個、1,790億個) が小さい
 - OpenAIのGPT-4等、最新のものはより巨大
 - とはいっても、我々の知る限り国内では最大級
 - 学習データ(=350GBの日本語データ)が小さい
 - 単語数でGPT-3の1/6程度、GPT-4等の学習データはもっと巨大
 - 一方、日本語BERTで最高精度をもたらしたデータ
 - 強引に1TB以上まで増やしたところBERTでは精度低下
 - 小さいとはいっても、新聞1,000年分
 - 入力、出力のテキストの長さが短い
 - NICTは最大256語、GPT-4は25,000語以上
 - 学習ノウハウについては他の研究機関等に共有済み
 - 現在、これらの差を詰めるべく
 - データの品質向上のための徹底的なクリーニングも含めた、学習データ構築の作業中
 - 3,110億パラメータのモデルも学習中、11月なかばに終了予定

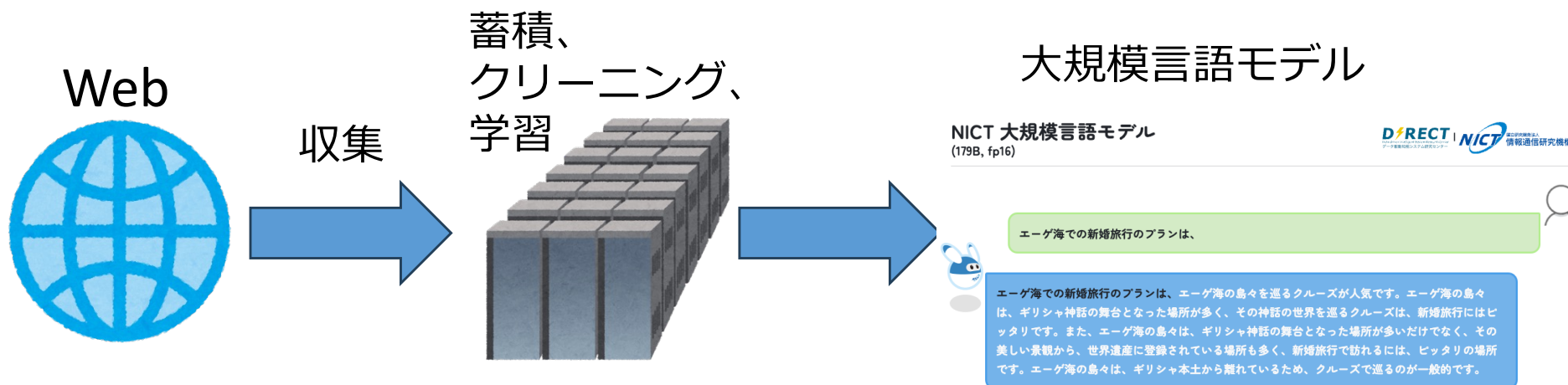
1. 大規模言語モデルとは何か？
2. NICTでの開発状況
3. 学習データの重要性
4. 生成AIと既存AIとの違い
5. 大規模言語モデルのインパクトとリスク

- 大規模言語モデルは膨大なWebページのテキストから学習



- 海外の大規模言語モデルの学習データは英語中心で、日本語データはわずか
 - 日本の主張、文化が海外の大規模言語モデルによってかき消される可能性、使用料金の流出も
- データの量だけではなく、品質も重要。同じパラメータ数でもデータの質・量により大規模言語モデルの性能は向上
 - 計算資源の量でGAFAM等に劣後する中で、データの質・量は数少ない日本の勝ち筋

- 大規模言語モデルは膨大なWebページのテキストから学習



- 現在、「真っ当な文章らしい」テキストのみをWebページから可能な限り大量に抽出し、学習データとする手法を開発中
 - 「真っ当な文章」はWeb全体の2%程度、宣伝がその5倍存在すると推定
 - さらに差別表現や個人情報等のフィルタリング等も実施予定
- NICTは日本語を中心とする良質な学習データを、海外産のLLM学習データと同等規模にまで拡充。(現在の数十倍のサイズに)
- 学習に適したデータは試作したモデルとともに、法的課題をクリアしながら日本企業・アカデミアが使えるようにしていく予定。
- 現在、10社以上の民間企業と調整中。我が国における大規模言語モデル開発力に貢献

- Webページでも学習に適したデータは極々一部
 - おそらく以下のようなテキストは学習では重要ではない

検索

メニュー

カテゴリ

医学

アレルギー & 臨床免疫学解剖学麻酔学行動科学学生化学 / 化学心臓病学臨床医学 / 医学一般クリティカルケア皮膚科学 & 美容外科学医学教育発生学救急医療内分泌学疫学 & 公衆衛生 E B M 消化器病学 & 肝臓病学遺伝学老人病学血液学組織学 & 細胞生物学感染症医学倫理医学 (他) 新生児学 / 周産期医学腎臓病学神経学神経外科産科婦人科腫瘍学眼科学整形外科学耳鼻咽喉学病理学小児科薬理学薬理学療法 & リハビリテーション医学生理学 P l a s t i c & R e c o n s t r u c t i v e S u r g e r y プライマリーケア / 一般診療精神医学放射線診断学呼吸器病学リウマチ病学外科泌尿器学 B o o k S e r i e s

- 現在、「まともな文章らしい」テキストのみをWebページから可能な限り大量に抽出する手法を開発中
 - これまでのところ、Webテキスト中で「まともな文章」は 全体の2%程度、宣伝がその5倍程度存在すると推定(HTMLタグは除く)
 - さらに差別表現や個人情報等のフィルタリング等も必要
- こうした「縁の下の力持ち」的作業が品質を大きく左右

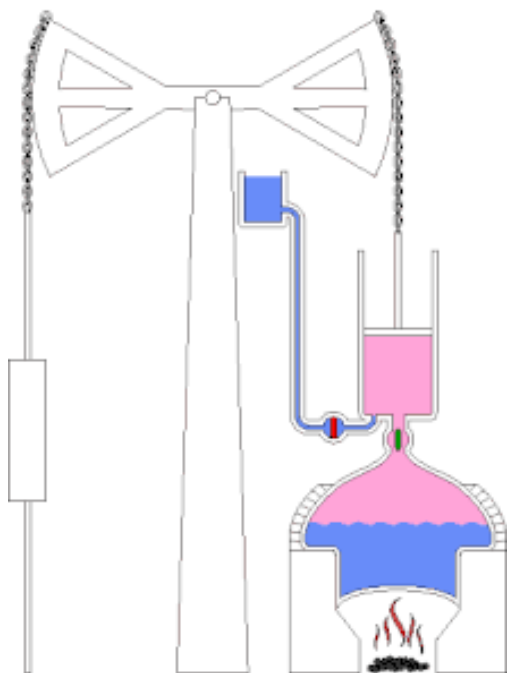
1. 大規模言語モデルとは何か？
2. NICTでの開発状況
3. 学習データの重要性
4. 大規模言語モデルのインパクトとリスク

- ポジティブ
 - 生産性向上・業務効率化
 - クリエイティビティの強化
 - 創作能力の民主化
 - ...
- ネガティブ
 - 偽情報
 - 偏見・バイアス
 - 仕事が奪われる
 - 民主主義の死
 - 皆考えるのをやめて、バカになる
 - ...

- 「この間も最初はいろいろあったけど、そのうち皆ハッピーになった。今度も同じ」
- 「どうせ、そんなに悪いことは起きっこない」

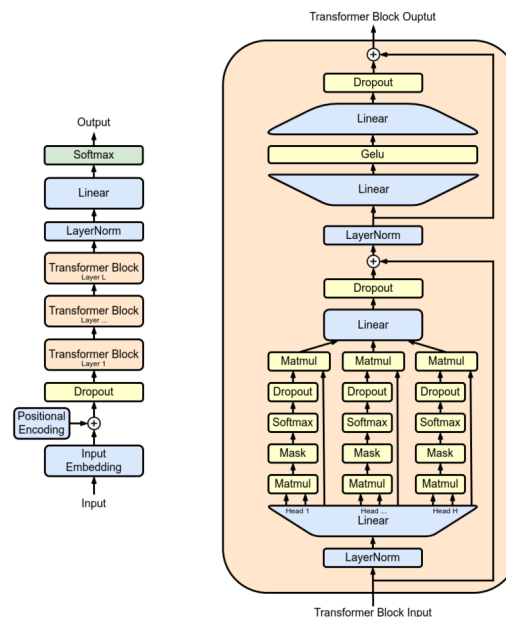
地球温暖化は？

産業革命



たかだか 1次元の運動

生成AI革命



> 1700億次元
(GPT-3のパラメータ数)

真面目に受け取らないでください

左図の蒸気機関の図、右図のtransformerのブロックダイアグラムは、Wikipedia日本語版の「外燃機関」「GPT(言語モデル)」の項よりそれぞれ転載
<https://ja.wikipedia.org/wiki/外燃機関>, [https://ja.wikipedia.org/wiki/GPT_\(言語モデル\)](https://ja.wikipedia.org/wiki/GPT_(言語モデル))
 CC-BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.ja> に則り転載



Donald Rumsfeld

“unknown unknown”

=“The things we don’t know
we don’t know”

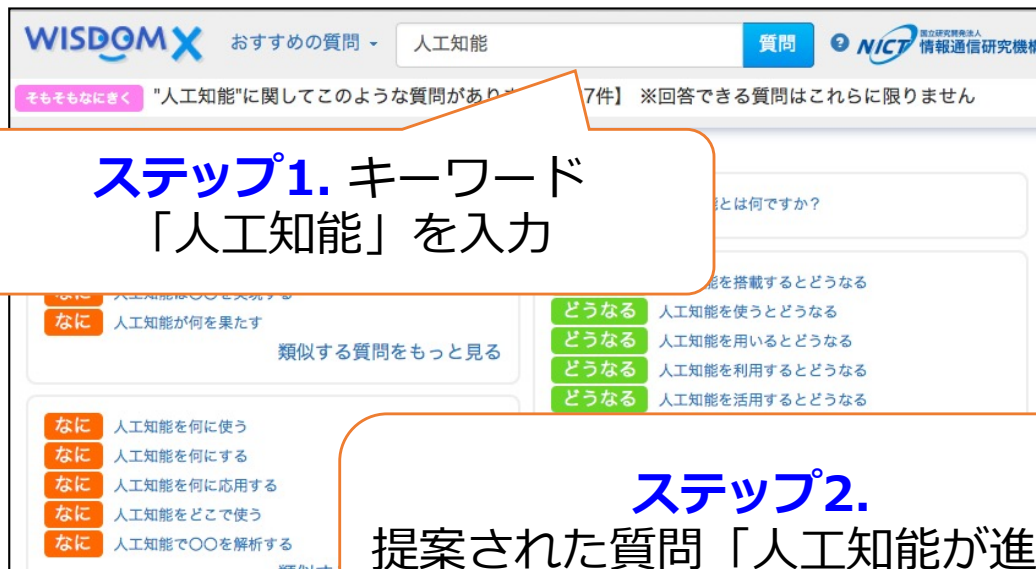
=“risks that come from situations
that are so unexpected that they
would not be considered.” (from
Wikipedia)

- ジェームズ・ワット等、産業革命への貢献者にとっては、「地球温暖化」はunknown unknown
- 生成AIでも無数のunknown unknownが発生するのではないか？

上の肖像写真はWikipedia日本語版「ドナルド・ラムズフェルド」の項より転載
<https://ja.wikipedia.org/wiki/ドナルド・ラムズフェルド>
CC-BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.ja> に則り転載

例えば、AIで少子高齢化が進み、 日本経済が窮地に陥る？

2016年10月の講演資料より



... 好感が持てる女性には相手にされな
い... 恋人の役割をロボットに任せたい

...
(読者アンケート)

2015年5月31日 朝日新聞デジタル
関連記事で類似の意見を紹介

質問: 人工知能が進化するとどうなる。

回答1: 人工知能が進化して、記憶容量も演算速度も人よりも速くなり知能指数で言えば...

回答2: 人工知能が発達したら、話し相手に困らなくなって結婚しない人が増える...

回答3: 人工知能が進化して、モノではなくなりますよね。

回答7: ...人工知能が発達して人間の脳と同じ様な化学反応を起こすようになれば...

人工知能が進化すると...

話し相手に困らなくなって、**結婚しない人が増える**

さらに提案された質問「話し相手に困らなくなって、結婚しない人が増えるとうなる？」をクリック。
回答「**少子高齢化が進む**」

さらに続けると、「日本経済が窮地に陥る」

- ハマダラカ
 - マラリアを媒介



上の写真はWikipedia日本語版「ハマダラカ」の項より転載
<https://ja.wikipedia.org/wiki/ハマダラカ>
CC-BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.ja> に則り転載

ブルキナファソのマラリア研究チームが、2020年7月に**不妊の遺伝子操作を施したメスの蚊を5000匹放出する**実験をおこなうことを計画していることがわかった。このプロジェクトは、オックスフォード大学を含む複数の研究機関の研究者を擁するTarget Malaria Projectによって実施されるもので、ビル&メリンダ・ゲイツ財団の資金提供を受けている。**不妊の蚊を放出することでマラリアを媒介するハマダラカの自然減を引き起こし、マラリアの感染拡大を抑制することが期待されている。**

“ブルキナファソで遺伝子操作の蚊を放出する計画”，現代アフリカ研究センター、東京外国語大学、2019/9/24、<http://www.tufs.ac.jp/asc/information/post-609.html>

不妊の蚊 = 生成AI?

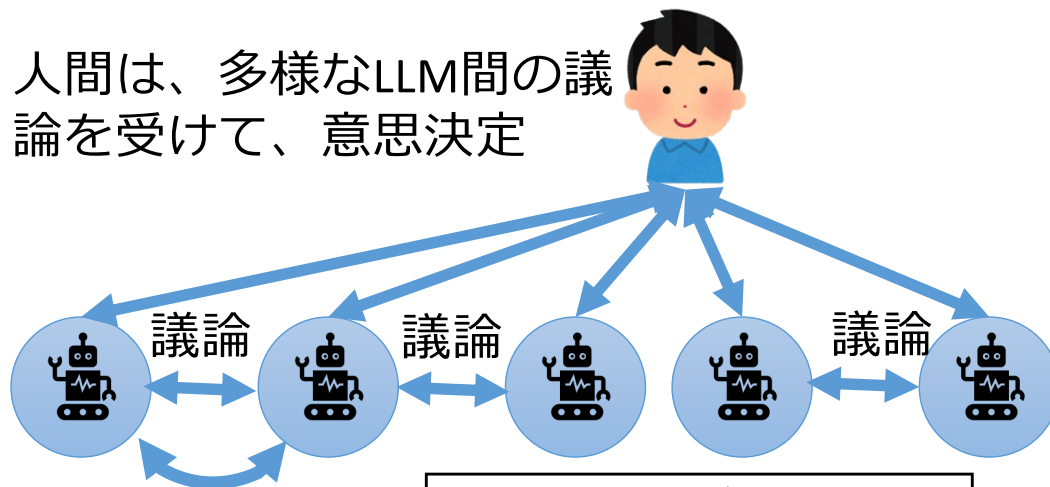
普通の蚊 = 人間？

“unknown unknown”?

- 今後、正体不明の「野良」生成AIも含め、多数の生成AIが出現する可能性が高く、多数の未知のリスクが発生する可能性も
 - 偽情報等、不適切な情報で日本社会が飽和する可能性も
 - ロシア、ウクライナ、イスラエルでの偽情報合戦
 - 国外からの選挙介入
 - 日本国民が気づかない間に特定の生成AIに思考を支配される？
 - 海外では生成AIと仲良くなり、洗脳されてエリザベス女王暗殺未遂や自殺といった報道も
 - 海外では出どころ不明の生成AIで新規マルウェア開発
 - 人間のホワイトハッカーが処理できない数のマルウェアが毎日新規開発されたら...
 - 他にも今後いろいろなリスクが出てくると思われる

- こうした状況での**抜本的な対策**として、将来的には、**多様な生成AIが互いに喧々諤々議論を行い、その結果を受けて人間が意思決定する「民主的」AIの世界**が考えられる
 - 不適切な情報はAIが互いに否定、批判することで排除できる可能性
 - 多様な生成AIがそれぞれの視点で議論し、より良い結論を
 - 海外発の偽情報に対して国産のLLMで反論できる可能性
 - ユーザとも議論すれば「皆考えるのを止めて、バカになる」が防げるかも

人間は、多様なLLM間の議論を受けて、意思決定



互いに否定、批判

玉石混交の多様なLLM

互いに別の視点を提供

多数のAIが互いに議論をするとうなるか？

40B, fp16

多数のAIが互いに議論をするとうなるかという、「自分の意見が正しい」ということを証明するために、相手の意見を否定することになります。

179B, bf16

多数のAIが互いに議論をするとうなるかという、「AI同士の議論が、人間の議論よりもはるかに生産性が高い」ということが起こります。

179B, fp16

多数のAIが互いに議論をするとうなるかという、「多数決」になる。

13B, fp16

多数のAIが互いに議論をするとうなるかという、AI同士が互いに相手の意見を否定しあうことになる。

- ここでは、米国の哲学者 Sellars, Brandomらに倣って、議論 = 「(主張の) 根拠を与え、要求するゲーム」 (“The game of giving and asking for reasons”) と仮定



Wilfrid Sellars



Robert Brandom

Wikipedia英語版“Wilfrid Sellars”, “Robert Brandom”の項より転載
https://en.wikipedia.org/wiki/Wilfrid_Sellars,
https://en.wikipedia.org/wiki/Robert_Brandom
CC-BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.ja> に則り転載

- Brandomの議論を単純化すると、上のゲームがプレーできるか否かが、オウムと人間を分ける境目
 - 大規模言語モデルは異常に賢いオウムではない？
- AIが安心安全で、合理的に振る舞えるとするならば、当然議論をして、自らの発言や振る舞いの根拠を提示できなければならない

NICT 大規模言語モデル (議論モード, 179B)



ユーザがテーマとなる主張を設定

議論のテーマ

地球温暖化の解決のために増やすべき自動車のタイプは、電気自動車ではなく、ハイブリッド車である。

その理由は？

ユーザがテーマとなる主張の根拠を質問

電気自動車は、電気を作るために、火力発電所で化石燃料を燃やすので、二酸化炭素を増やすからである。

しかし、ハイブリッド車も二酸化炭素をたくさん排出する。

ユーザが反論

このため、結論は、「自動車を減らす」ということになる。

大規模言語モデルが主張を変更

しかし、自動車を減らすことによる経済的ダメージで人類の存続が危うくなるというのであれば、

ユーザが反論

大規模言語モデルが代替案を提示

自動車を減らすことはできない。このため、自動車を減らすことによる経済的ダメージを減らすために、自動車の燃費を向上させることが必要である。自動車の燃費を向上させるためには、自動車の軽量化が必要である。自動車の軽量化のためには、自動車のボディを鉄から炭素繊維に変えることが必要である。炭素繊維は、鉄の1.0倍の強度を持ち、重さは鉄の4分の1である。このため、炭素繊維を自動車のボディに使用すれば、自動車の軽量化が可能になる。

しかし、炭素繊維を作るには二酸化炭素をたくさん排出しなければいけない。

ユーザが反論

このため、結論は、「二酸化炭素を減らすためには、自動車を減らす」ということになる。

テキストを入力(議論)

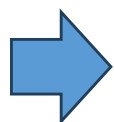
大規模言語モデルが再度主張を変更

注：議論のプロセスや結論が正しいことを保証するものではありません

- NICTの大規模言語モデルでもフェイクニュースやオレオレ詐欺の台本等生成可能→今後そうした能力をもつ大規模言語モデルが多数出現
- 安全保障上も対策が必要
- 対策例

大規模言語モデルが生成したフェイクニュースの例
(実際にはNICTの大規模言語モデルが生成した例)

正体不明の
大規模言語モデル



「伊豆天城山が噴火。土石流が発生し、近隣自治体で死傷者多数。救援は極めて困難」というタイトルの新聞記事は以下の通り。「本日未明、伊豆半島の天城山が噴火。噴火による火山灰が土石流となり、近隣自治体に甚大な被害をもたらした。また、噴火による火山灰が上空に滞留したため、航空機の運航が全面的に停止。伊豆半島の交通網は完全に麻痺した。現在、自衛隊が救援活動を行っているが、噴火による火山灰が上空に滞留しているため、ヘリコプターによる救援活動は極めて困難となっている。

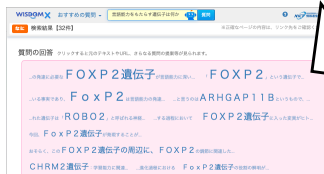
透明性の高い大規模言語モデル等で
タイムリーに根拠のある反論

天城山での噴火やそれによる地震、火山灰等は一切観測されておりません。また、伊豆半島での通行止めや、自衛隊が救援活動を行っている事実もありません。...

NICT 大規模言語モデル
(1798, fp16)

DIRECT/NICT

検索ベースの質問応答システム等で情報の裏取り



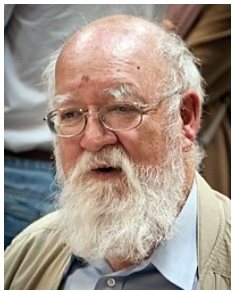
他の大規模言語モデル

他の大規模言語モデルと議論。
反論のネガチェック

他の大規模言語モデル

大規模言語モデルが生成するマルウェア等にも大規模言語モデルを使った対策が必要？

- Explainable AI = 判断の「根拠」を説明するAI
 - ここでは議論の過程で根拠、理由を提示する大規模言語モデル
- “The work on reason-giving and normativity descended from Sellars at Pittsburgh, via Brandom and McDowell and Haugeland, has never stressed, to my knowledge, that these all-important human practices are systematic generators of false ideology whenever the demand for reasons exceeds the available supply.”



-- Daniel Dennett, “From Bacteria to Bach and Back”, p.314, 脚注, 2017

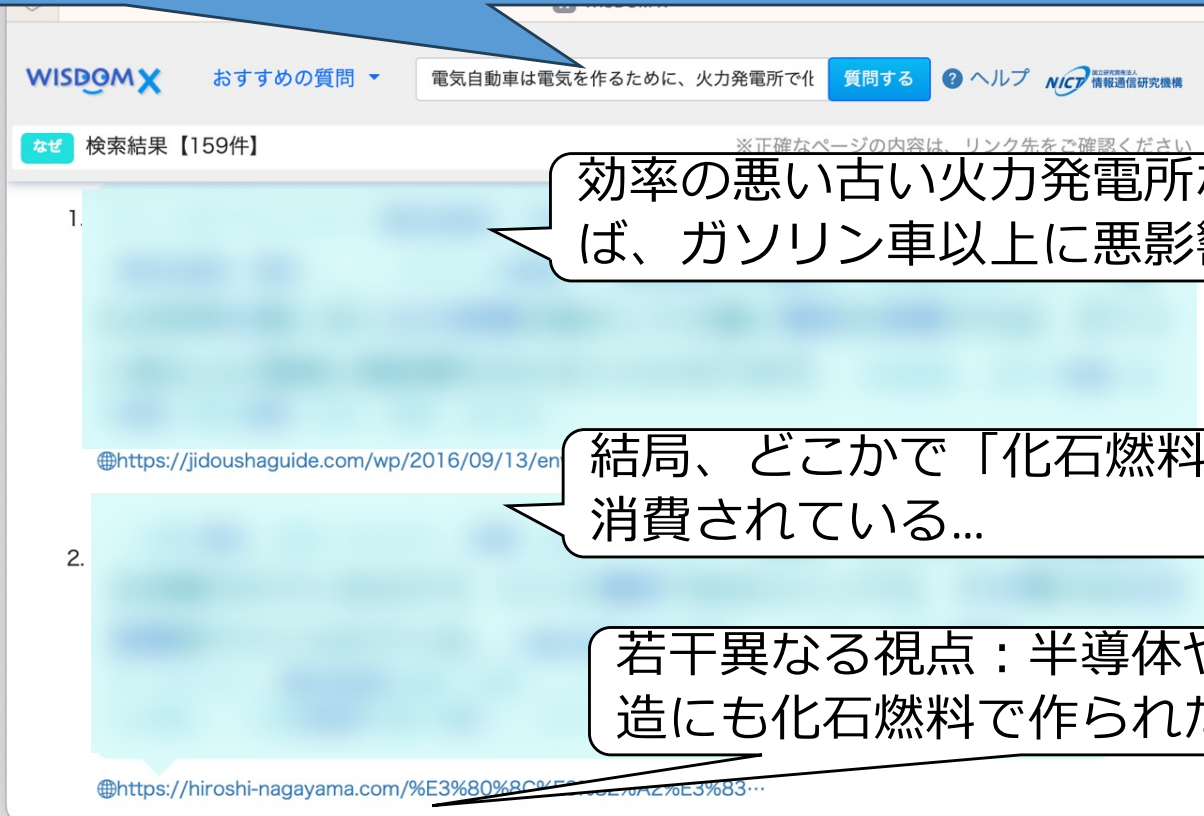
- 超訳「苦し紛れの根拠を与えると『誤ったイデオロギー』の出現の原因となる」
- 仮にいい加減なAIがいい加減な根拠を与えるとかえって有害？
- 例:大鯰 = 江戸時代の地震の根拠 (誰かが苦し紛れに考えだした?)



- 議論ができて、偽情報が出力されれば意味がない
- 誰か人間が書いたWebテキストで裏取り
 - 現在各所で検討が行われている情報のソース、来歴等保証するオリジネータープロファイル (OP)等の技術も重要



電気自動車は電気を作るために、火力発電所で化石燃料を燃やすので、二酸化炭素を増やすからである。+のはなぜ



WISDOM X おすすめの質問 ▾ 電気自動車は電気を作るために、火力発電所で化石燃料を燃やす 質問する ヘルプ NICT 情報通信研究機構

なぜ 検索結果【159件】 ※正確なページの内容は、リンク先を確認ください

1. 効率の悪い古い火力発電所ならば、ガソリン車以上に悪影響
2. 結局、どこかで「化石燃料」は消費されている...
若干異なる視点：半導体やバッテリーの製造にも化石燃料で作られた大量の電気が...

<https://jiidoushaguide.com/wp/2016/09/13/en>

<https://hiroshi-nagayama.com/%E3%80%80%E3%82%A2%E3%83...>

・日経新聞の記事より

Q. 日本のこれまでの少子化対策は何が間違っていましたか？

ChatGPTの回答

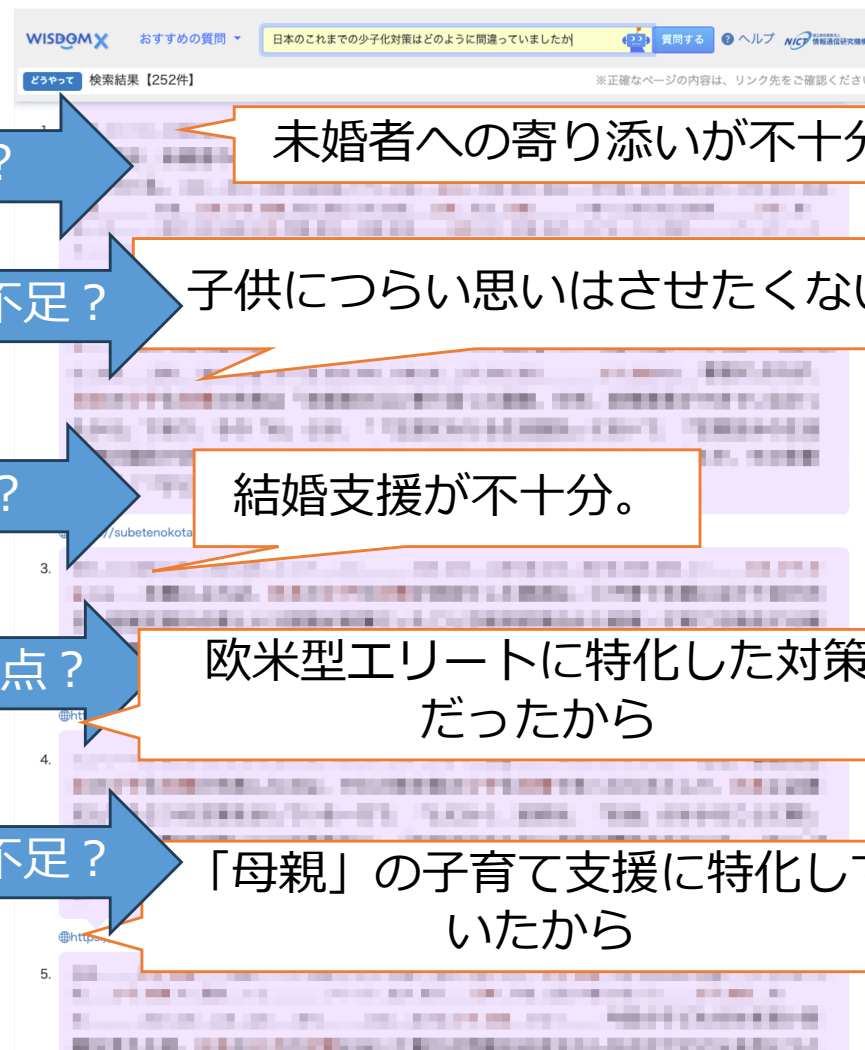
- 定額の児童手当の導入や育児休業法の改正など多くの政策が実施されてきましたが、出生率は上がっていません。政策が十分な効果を発揮していないことや、社会的な問題が背景にあることが挙げられます
- 具体的には女性が仕事と家庭の両立を困難に感じる社会的な制約や、若者の経済的不安定さが原因の一つと指摘されています。教育や就労の機会が増えて女性が自己実現を求めるようになり、子育てに時間を割くことが難しくなっていることも要因の一つです。
- 今後の政策は、こうした社会的な問題にも着目し、女性や若者が子育てしやすい環境を整備することが必要です

忙しすぎる女性、経済的に不安定な若者への施策が不十分

“ChatGPTと考える 少子化ニッポンの成長の道”
日経新聞電子版2023年4月7日

<https://www.nikkei.com/article/DGXZQOUA15AO20V10C23A2000000/>

NICTの質問応答システムWISODM X
(Webページから回答を「検索」)



WISODM X おすすめの質問 日本のこれまでの少子化対策はどのように間違っていましたか？ 質問する ヘルプ NICT 情報通信研究機構

どうやって 検索結果 [252件] ※正確なページの内容は、リンク先をご確認ください

欠落？ 未婚者への寄り添いが不十分

踏み込み不足？ 子供につらい思いはさせたくない

欠落？ 結婚支援が不十分。

不在の視点？ 欧米型エリートに特化した対策だったから

踏み込み不足？ 「母親」の子育て支援に特化していたから

注：回答の形式が変わるため、質問を「どのように間違っていましたか？」と修正

もっと尖った議論はできるか？

- NICTが現在開発中の将来シナリオ生成技術は、低品質なシナリオも生成するが、ChatGPTよりも「より尖ったシナリオ」を生成することも
- もちろん、わかりやすさ、可読性はChatGPTの方がはるかに上

• ChatGPT (GPT-4) の出力 (2023.6.5作成)

DI 対話システムを開発して高齢者のQoLを向上させるためには具体的にどのような用途がありそうでしょうか

高齢者のQoL (Quality of Life : 生活の質) を向上させるための対話システムの開発には、以下のような用途が考えられます。

1. **健康管理支援:** 高齢者の日常的な健康管理をサポートするための対話システムは、症状のモニタリングやリマインダー機能 (薬の服用時間や定期的な医療チェックアップなどのリマインダー) を提供できます。
2. **エンターテインメント:** 音楽、物語、クイズゲームなどを提供し、高齢者が楽しみながら脳を活性化させることができます。
3. **情報提供:** 天気予報、ニュース、地元のイベントなどの情報を提供することで、高齢者が外の世界とつながりを保つことができます。
4. **社会的つながりのサポート:** 対話システムが家族や友人との通信をサポートすることで、高齢者の孤立感を軽減することができます。
5. **日常生活のアシスタンス:** 簡単な質問応答、日常生活のリマインダー (ゴミの日など)、料理のレシピの提供など、日常生活のさまざまな側面で支援を提供できます。
6. **認知症予防および支援:** 認知症に対する初期介入として、または認知症患者のケアとして、メモリーゲームや日常の状況を思い出させるような対話を提供することができます。
7. **エクササイズ指導:** 軽い運動やストレッチの指導を行い、健康維持に役立てます。
8. **感情的サポート:** 会話によって高齢者の感情的な状態を理解し、慰めたり、励ましたりすることができます。

これらの用途を考慮に入れて、対話システムを設計するときには、利用者のニーズと、そして可能な限りの使いやすさを確保することが重要です。

• NICTの独自将来シナリオ生成技術

入力: 対話システムを開発する + 高齢者



- 対話システムで詐欺的投資勧誘等の悪質商法から高齢者を保護する → 高齢者等を狙った悪質商法や特殊詐欺の手口等の情報を提供する対話システムを構築する
- 対話システムで高齢者の食生活を支援する → 対話システムが高齢者に対して、宅配による配食サービスを実施する
- 対話システムで地域住民との交流を促進する → 対話システムが地域住民の方々とバーベキュー大会を企画する

より尖った将来シナリオでイノベーションのヒントを広く提供？

- 大規模言語モデルの研究開発は始まったばかり
- 検索エンジンも含め、既存AIとより密接な連携等も含めて、今後さらなる進化があるだろう
 - より根拠に基づいたテキストを出力
 - より多様な視点に基づいたテキスト
 - よりクリエイティブなテキスト
 - よりの的を射た対話、議論
 - ...
- 一方で多様なリスク対策は必要
 - 偽情報対策、セキュリティ対策
 - おそらく長い長い戦い

- 生成AIがさらに進化した未来の可能性
 - 大都市の労働力が地方に分散？
 - 常にAIが相談相手になり、また、様々なコーチング、注意喚起をしてくれるのであれば、
 - 大都市の大勢の同僚と同じオフィスで働く必要性は減る？
 - すでにリモートワークが一定の存在感
 - インターネットで情報格差も随分減少
 - 地方の方が低いリビングコストも含めて、より良好な生活環境
 - 様々な専門家と直接話をする、あるいは勉強会等をする必要性も減る？
 - 地方の若年労働者のスキルアップ？
 - 地方の学生は何か面白いことをやろうと思っても周辺に付き合ってくれる専門家、仲間は少ない
 - 生成AIが代わりを果たす？ ブレインストーミング的議論
 - 教育への応用も類似の効果が期待できるかも
 - 高齢者介護の改善
 - コミュニケーションの有無は高齢者の健康維持に重要
 - 生成AIならばいつでもどれだけでも会話に付き合える
 - 等々....

- 長期的に可能な未来
 - 地方の若年労働者のスキルアップ？

地方在住の若年労働者がAIを相談相手に使いつつ仕事をしたら、どのような効果が見込めるか？

地方在住の若年労働者がAIを相談相手に使いつつ仕事をしたら、どのような効果が見込めるかという、まず、地方在住の若年労働者は、AIを相談相手に使うことで、仕事の悩みを解消することができる。

地方在住の若者がAIを相談相手に使いつつ仕事をしたらどのような効果が見込めるか？

地方在住の若者がAIを相談相手に使いつつ仕事をしたらどのような効果が見込めるかという、地方の若者は、AIを相談相手にすることで、地方にいながらにして、東京の最先端の仕事をするようになる。

地方在住の若年労働者がAIを相談相手にしたらスキルアップが可能になる理由は何か？

地方在住の若年労働者がAIを相談相手にしたらスキルアップが可能になる理由は何かという、AIは「自分の能力を客観的に評価してくれる」からです。しかし、地方在住の若年労働者は、自分の能力を客観的に評価してくれる人を見つけることができません。

テキストを入力



- 生成AIがさらに進化した未来の可能性
 - 要するに、市民一人一人が、十分な知識を備えた「仲間のようなもの」といつでもどこでもコミュニケーションができれば社会が変わる。日本の課題も抑制可能かも
 - 人口減に対する対策とも言える
 - 「洗脳」等に対して、十二分なリスク対策は必須

- NICTの大規模言語モデルと関連技術
- 生成AIがもたらすリスク
 - 究極のリスクとしての”unknown unknown”
- リスクへの抜本的な対策としての、多様な生成AIが互いに議論する「民主的AI」の世界の可能性



上の写真はWikipedia日本語版「ウィンストン・チャーチル」の項より転載
<https://ja.wikipedia.org/wiki/ウィンストン・チャーチル>
CC-BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.ja> に則り転載

*‘Many forms of Government have been tried, and will be tried in this world of sin and woe. No one pretends that democracy is perfect or all-wise. Indeed it has been said that **democracy is the worst form of Government except for all those other forms that have been tried from time to time....’***

- Winston S Churchill, 11 November 1947

- 生成AIの今後の発展の可能性
- 社会に与えるインパクトの可能性
 - 多様な視点からの人口減へのカウンター