

人工知能ゼミ2

対話型人工知能を使う

公立小松大学

藤田 一寿

■ 計画

1. 10/2 ガイダンス
2. 10/9 対話型人工知能の使い方（講義、演習、レポート）
3. 10/16 プロンプトエンジニアリング（講義）
4. 10/23 プロンプトエンジニアリング演習
5. 10/30 プロンプトエンジニアリング演習
6. 11/13 人工知能に過去問を解かせる
7. 11/20 人工知能に過去問を解かせる
8. 11/27 人工知能に過去問を解かせる（発表資料作成）
9. 12/4 人工知能に過去問を解かせる（発表会）
10. 12/11 人工知能と小論文を書く
11. 12/18 人工知能と小論文を書く（レポート作成、レポート提出）
12. 12/25 人工知能と英語論文を読む
13. 1/15 人工知能と英語論文を読む（発表資料作成）
14. 1/22 人工知能と英語論文を読む（発表会）
15. 1/29 資料を読ませレポートを作成する（レポート）

人工知能とは

人工知能とは

- 人工知能とは「人間が知能を使ってすること」ができる機械もしくはそれを実現するための技術
 - 人間の知的営みをコンピュータに行わせるための技術（IT用語辞典バイナリ）
 - 学習・推論・判断といった人間の知能を持つ機能を備えたコンピュータシステム（大辞林第三版）
 - 人間や動物が可能な知的作業ができるコンピュータを作ることに関する事柄（Artificial Intelligence for Games）
- 人（研究者）によって人工知能の捉え方が違う。
 - 知能の定義が出来ないから、人工知能も定義できないという考え方もある。

- そもそも知能が何を意味するか、知的営みがどの程度の知能のため、人工知能という用語も曖昧になる。
- この講義では一般的な認識に近い赤字の解釈をしておく。

研究者	所属	定義
中嶋秀之	公立ほくほく大学	人工的につくられた、知能を持つ装置。あるいはそれをつくらうことによって知能自体を研究する分野である
武田英明	国立情報学研究所	
西田豊明	京都大学	「知能を持つメカ」ないしは「心を持つメカ」である
渡口雄一郎	北陸先端科学技術大学院	人工的につくった知的な振る舞いをするためのもの（システム）である
長尾真	京都大学	人間の脳回路活動を模倣してシミュレートするシステムである
堀内一	東京大学	人工的に作る新しい知能の世界である
浅田裕	大阪大学	知能の定義が明確でないで、人工知能を明確に定義できない
松原仁	公立ほくほく大学	究極には人間と区別がつかない人工的な知能のこと
村上高志	東京大学	自然にわれわれがペットや人に接するような、情感と認識に富んだ相互作用を、物理法則に拘束なく、あるいは逆らって、人工的に作り出せるシステム
山口周平	慶應義塾大学	人の知的な振る舞いを模倣・学習・超越するための機械的システム
葉山隆	電気通信大学	人工的につくられる知能であるが、その知能のレベルは人を越えているものを想定している
山田悠	ドワンゴ人工知能研究所	計算機知能のうちで、人間が直接・間接に設計する概念を人工知能と呼んで良いのではないと思う
松尾豊	東京大学	人工的につくられた人間のような知能、ないしはそれをつくる技術。人間のように知的であるとは、「気づくことができる」コンピュータ、つまり、データの中から特徴量を生成し現象をモデル化することのできるコンピュータという意味である

（出典）松尾豊「人工知能は人間を超えるか」(KADOKAWA) p.45,49作成

である。そ
(平成28年度情報通信白書)

■ 人工知能とは

- 人工知能に関連する用語
 - 特化型人工知能
 - ある問題に対し特化した人工知能
 - 例：画像認識，音声認識，画像生成など
 - 汎用人工知能 (Artificial general intelligence : AGI)
 - 人間レベルの知能
 - 2022年から人工知能はAGIに着実に近づきつつある
 - 人工超知能 (Artificial Super Intelligence: ASI)
 - 人間の知的能力を超えた人工知能
 - 特化型人工知能では人の能力を超えたものがある

Google DeepMindによるAGIのレベル

		特化型人工知能	汎用人工知能
Performance (rows) x Generality (columns)		Narrow clearly scoped task or set of tasks	General wide range of non-physical tasks, including metacognitive abilities like learning new skills
計算機ソフトはNo AIらしい	Level 0: No AI	Narrow Non-AI calculator software; compiler	General Non-AI human-in-the-loop computing, e.g., Amazon Mechanical Turk
	Level 1: Emerging equal to or somewhat better than an unskilled human	Emerging Narrow AI GOFAI ⁴ ; simple rule-based systems, e.g., SHRDLU (Winograd, 1971)	Emerging AGI ChatGPT (OpenAI, 2023), Bard (Anil et al., 2023), Llama 2 (Touvron et al., 2023)
	Level 2: Competent at least 50th percentile of skilled adults	Competent Narrow AI toxicity detectors such as Jigsaw (Das et al., 2022); Smart Speakers such as Siri (Apple), Alexa (Amazon), or Google Assistant (Google); VQA systems such as PaLI (Chen et al., 2023); Watson (IBM); SOTA LLMs for a subset of tasks (e.g., short essay writing, simple coding)	Competent AGI not yet achieved
画像生成人工知能	Level 3: Expert at least 90th percentile of skilled adults	Expert Narrow AI spelling & grammar checkers such as Grammarly (Grammarly, 2023); generative image models such as Imagen (Saharia et al., 2022) or Dall-E 2 (Ramesh et al., 2022)	Expert AGI not yet achieved
	Level 4: Virtuoso at least 99th percentile of skilled adults	Virtuoso Narrow AI Deep Blue (Campbell et al., 2002), AlphaGo (Silver et al., 2016, 2017)	Virtuoso AGI not yet achieved
囲碁人工知能	Level 5: Superhuman outperforms 100% of humans	Superhuman Narrow AI AlphaFold (Jumper et al., 2021; Varadi et al., 2021), AlphaZero (Silver et al., 2018), StockFish (Stockfish, 2023)	Artificial Superintelligence (ASI) not yet achieved
ゲーム人工知能（ゲーム以外にも適用できる）			

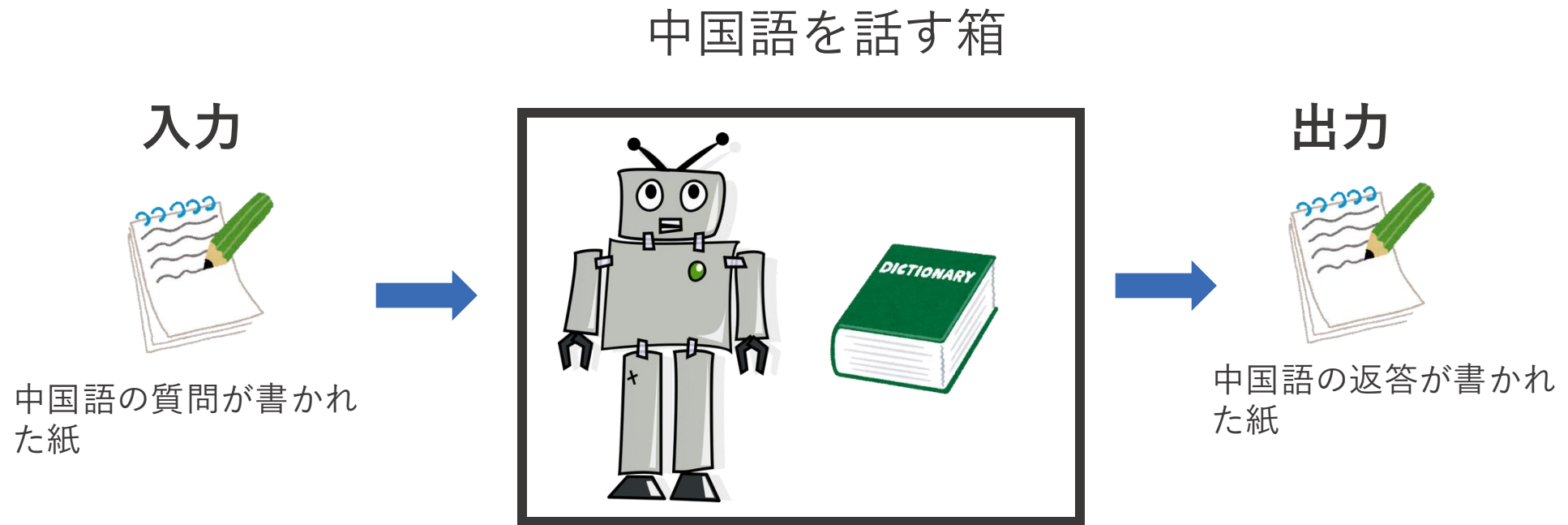
2023年における対話型人工知能は汎用人工知能で、スキルを持たない者と同等か幾らか良い程度の能力はあるらしい。

2023年では、まだこのレベルの人工知能は登場していない。GoogleのGemini Ultraはこのレベルに達したのだろうか？
2024年でここに位置する人工知能はあるのだろうか？もしくは、すでに多くのマルチモーダル人工知能は達しているのかもしれない。

(Morris, et al. 2023)

■ 人工知能のとは

人工知能は入力に対し適切な出力をする機械（箱）



箱の中の機械が適切な返答をすれば良い。

人工知能は、入力に対し適切な出力をする関数と言える。人工知能を作るとは、その関数を作ることである。人間の応答を関数で表現できれば、その関数を使った人工知能は人そのものと言えるかも知れない。

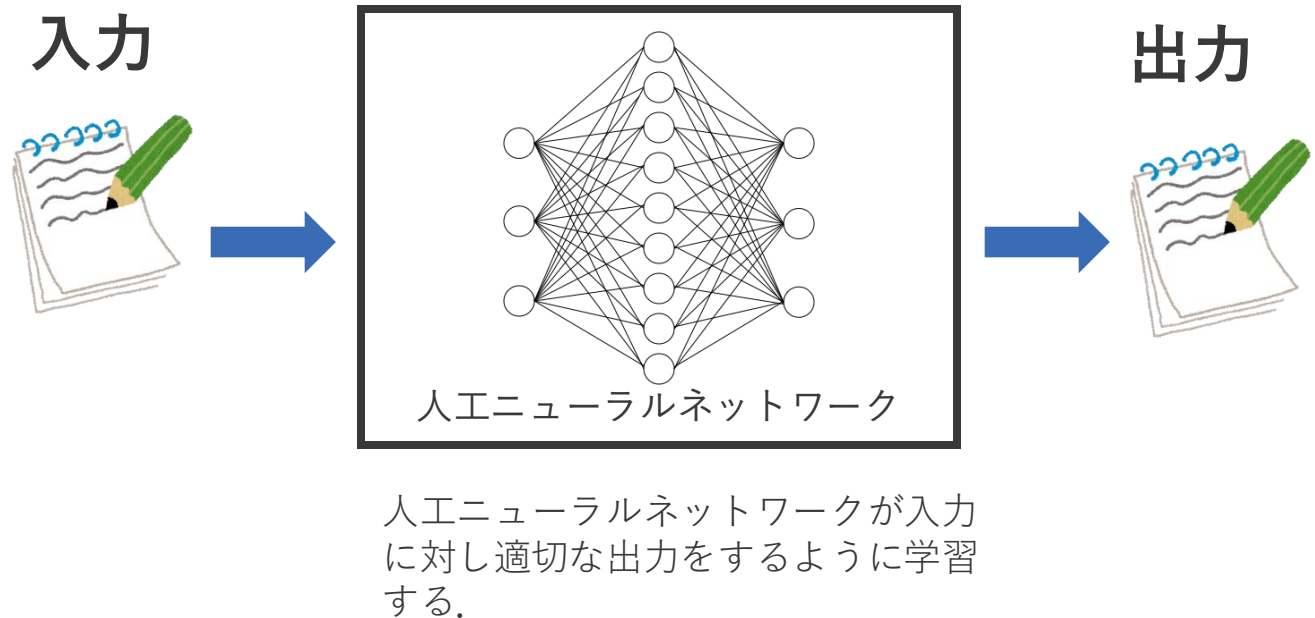
■ 人工知能の実現手法

- 現在の人工知能を実現する技術として，人工ニューラルネットワークがよく用いられる。

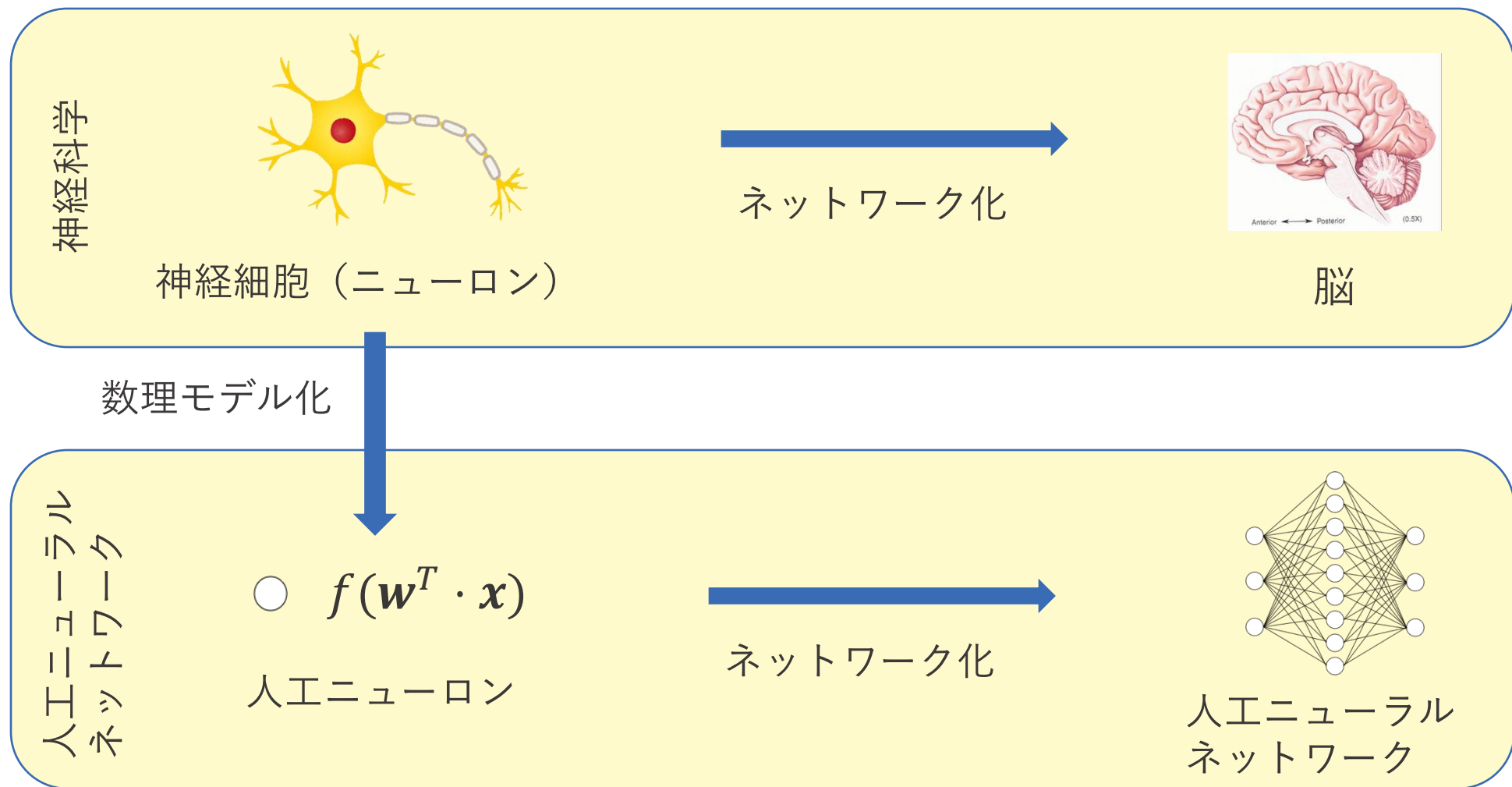
- 適用例

- 画像識別
- 領域抽出
- 対話型人工知能
- 画像生成人工知能
- 楽曲生成人工知能
- ゲームプレイ人工知能

- 最新の人工知能技術と呼ばれるものには必ずと言っていいほど人工ニューラルネットワークが使われる。



■ 人工ニューラルネットワーク



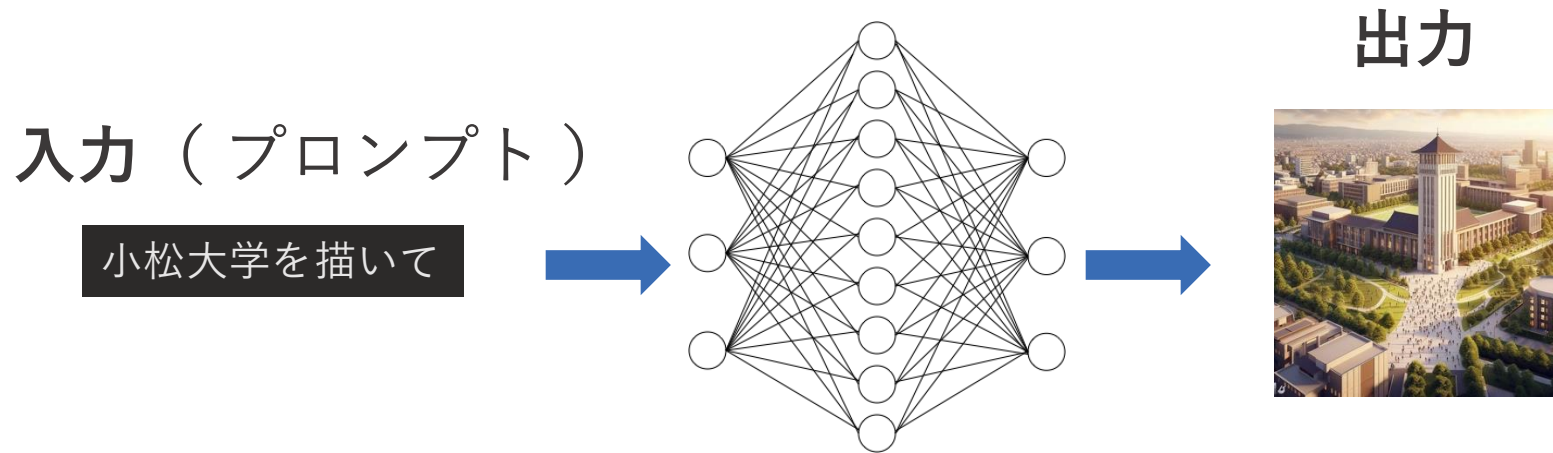
解剖学的知見に基づき精密に脳をモデル化した実用的な人工ニューラルネットワークはあるだろうか。機能をモデル化したものはたくさんあるだろう。

人工ニューロンを組み合わせ（ネットワーク化し）、様々な機能を実現したものを人工ニューラルネットワークと呼ぶ。かつては、人工ニューロンは神経細胞を数理モデル化したもので、ネットワークの構造も脳を意識しており、人工ニューラルネットワークは神経科学のと言えた。しかし、現在の人工ニューラルネットワークは神経細胞や脳の数理モデルから乖離しており、表面上は脳と関係ない。表面上関係ないが、学習機械として共通する理論や性質があるのかもしれない。

生成人工智能

■ 生成人工知能

- 生成人工知能とは，文章，画像その他のメディアをプロンプトから生成することが出来る人工知能システムのこと（wikipedia英語版）．
 - プロンプトとは，人工知能に与える命令のこと．



■ 代表的な生成人工知能

- 対話型（会話型）人工知能
 - ユーザと対話することができるチャットボットや仮想エージェントなどの技術(<https://www.ibm.com/jp-ja/topics/conversational-ai>)
 - 利用者が音声やテキストで入力すると自然な文章を生成し，応答してくれる人工知能(<https://www.nikkei.com/theme/?dw=23052603>)
- 画像生成人工知能
 - 画像を生成する人工知能
 - Stable diffusionやDALL-Eは拡散モデルを用い画像生成を行う．

■ 代表的な生成人工知能

- 対話型（会話型）人工知能
 - ユーザと対話することができるチャットボットや仮想エージェントなどの技術(<https://www.ibm.com/jp-ja/topics/conversational-ai>)
 - 利用者が音声やテキストで入力すると自然な文章を生成し，応答してくれる人工知能(<https://www.nikkei.com/theme/?dw=23052603>)
- 画像生成人工知能
 - 画像を生成する人工知能
 - Stable diffusionやDALL-Eは拡散モデルを用い画像生成を行う.

DALL-E 3



DALL-E 3



■ 対話型人工知能と言語モデル

- 対話型人工知能
 - 言語モデルを用い対話を実現している。
 - 次に出てくる単語を予測するモデル。次の単語を予測し続ければ文章が出来る。
 - ChatGPTは言語モデルGPTを用いOpenAIにより構築された対話型人工知能
 - Bartは基盤モデルLaMDAを用いGoogleにより構築された対話型人工知能

■ 言語モデルとは

- 言語モデリングは、与えられた文章の次に来る単語（トークン）を予測するタスク
 - 次に来る単語を予測し続けることで文章を生成することができる。
- 言語モデリングを行うものを言語モデルという。
 - 関連技術：n-gram, LSTM (long short-term memory), RNN (recurrent neural network)

■ 言語モデルとは

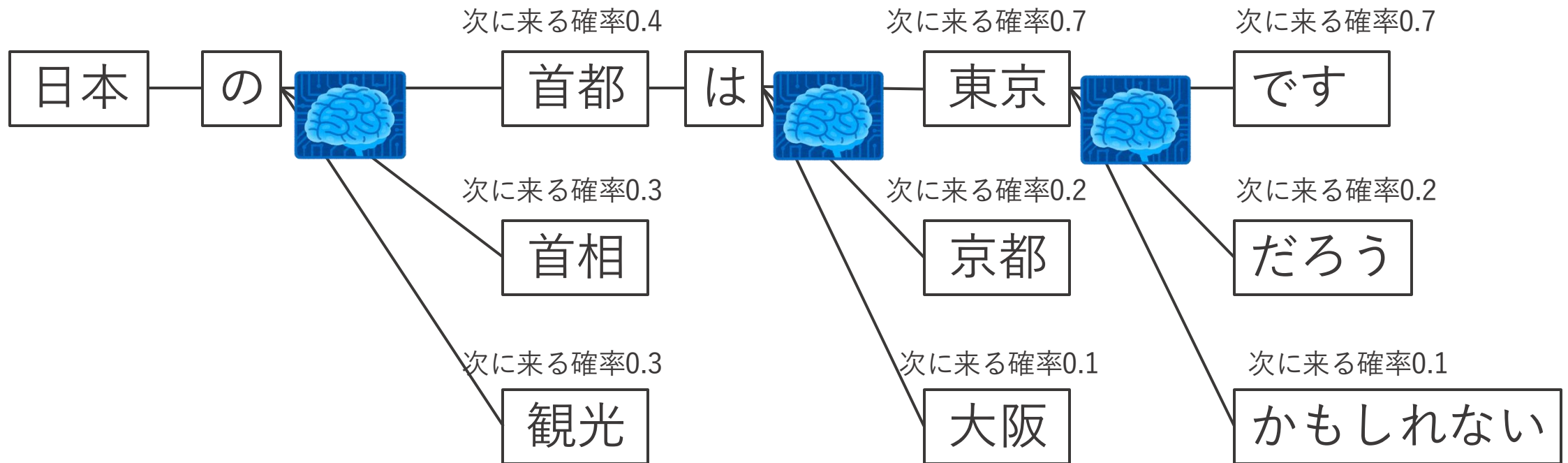
- 大量のパラメタで構成される巨大なニューラルネットワークを用いた言語モデルを大規模言語モデルという。
 - 現在はTransformerベースの技術が主流.
 - ChatGPTもTransformerがベースのGPTという言語モデルを用いている.



TransformerとはGoogleが開発した言語モデルである。再帰構造がなく並列計算が容易である点などがこれまでの手法と異なる。

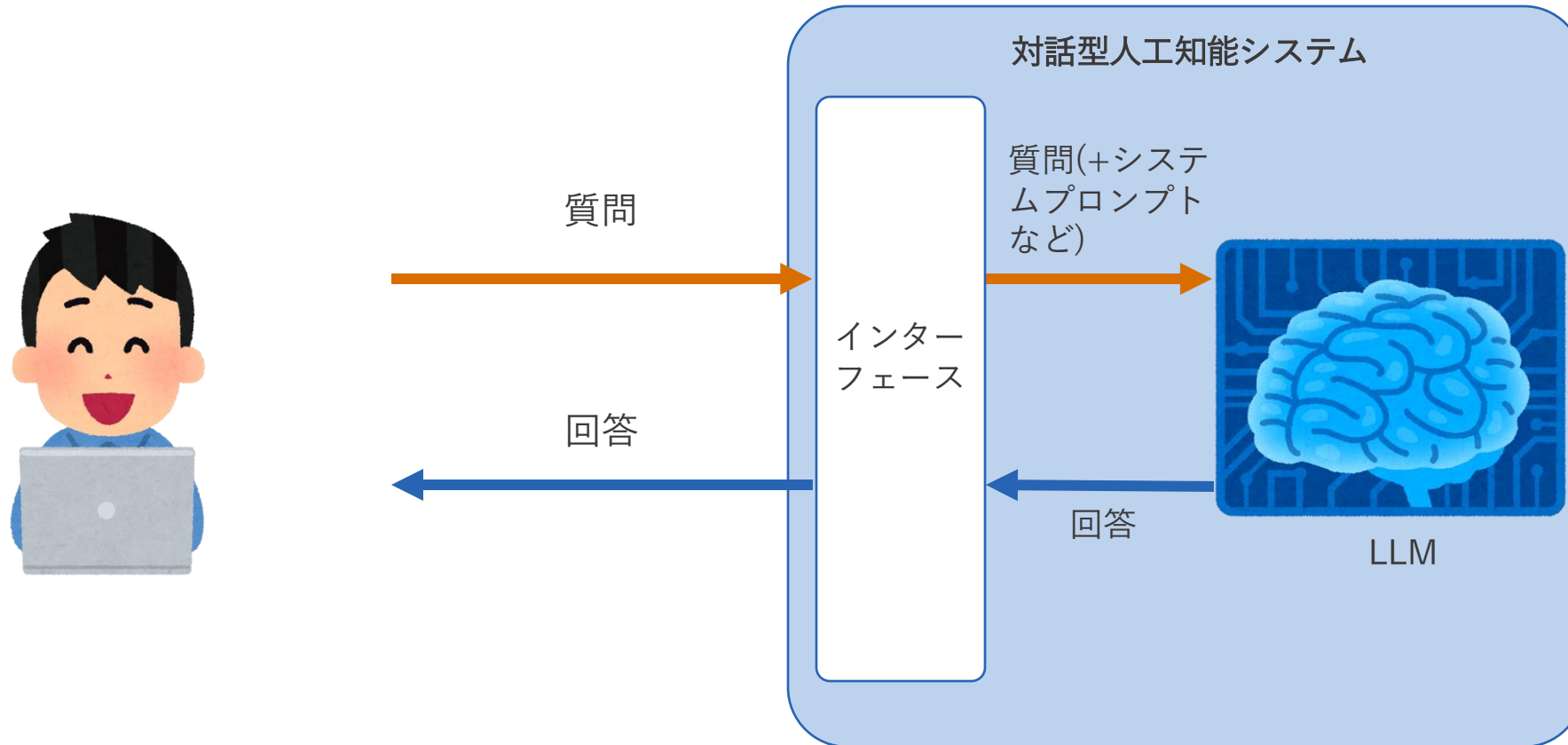
言語モデルは何をしているのか

- 言語モデルは、次にどの単語が出やすいかを学習する。
- 学習後のモデルを用い、次に出る単語を予測する。



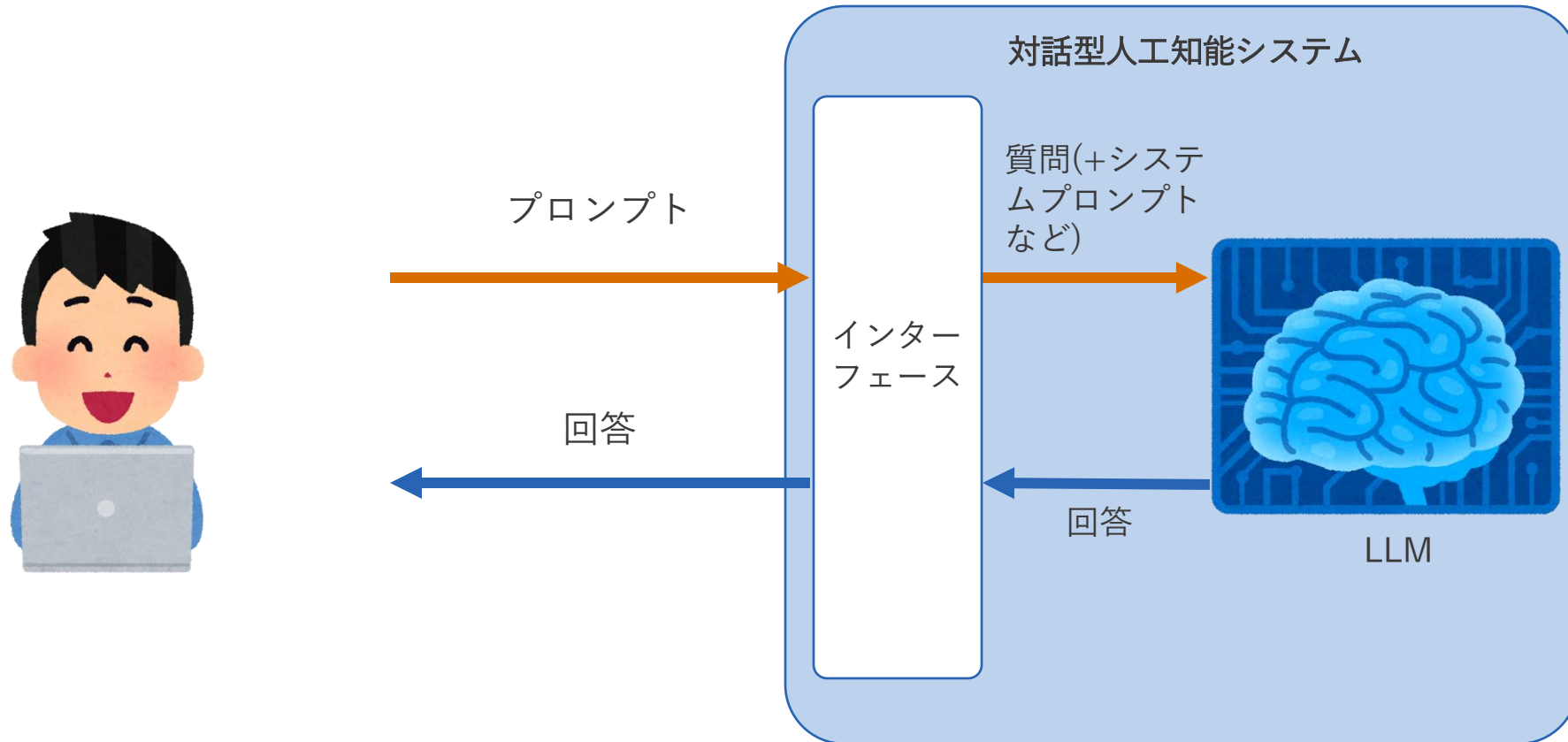
■ 対話型人工知能と大規模言語モデル

- ChatGPTなどの対話型人工知能は大規模言語モデル(LLM)を使って会話をしている。
- 対話型人工知能の能力はLLMの性能に依存しているところが多い。



■ プロンプト

- 対話型人工知能を思い通りに動かすための質問や指示をプロンプトという。

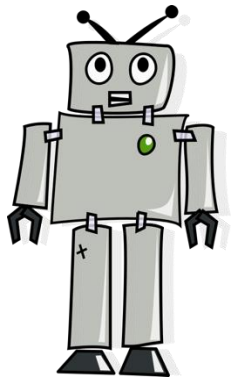


人工知能を使う上での心配

■ 人工知能は役に立たない？

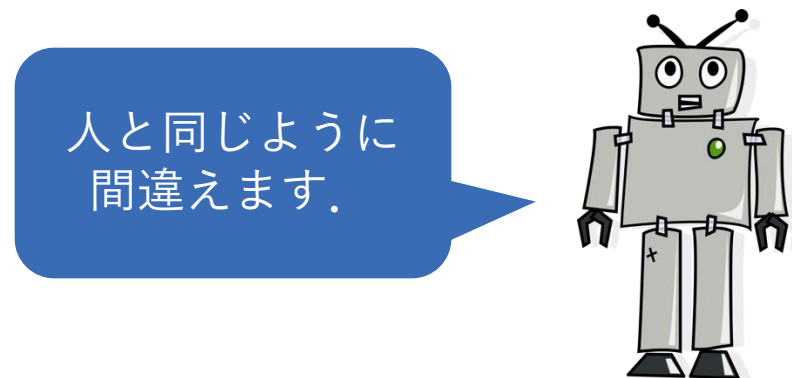
- 人工知能に質問してもロクな回答が得られない。
 - 人工知能は空気を読めないなので、 **いい加減な質問をしても意図通りの回答をしてくれない。**
- 質問次第で解決が可能
 - 聞きたいことを正確に言語化する。
 - 人工知能が理解できる分かりやすい文章を書く。
 - 人工知能が的確に答えられるように、細かな指示をする必要がある。
 - 人工知能は文脈を理解できるので、質問をやり取りすると良い回答を得ることができる可能性がある。

丁寧に細かな指示をお願いします。



■ 人工知能は嘘つきで信用できない？

- 人工知能は間違えることがある。
 - 人工知能は人と同じで入力に応じて記憶から出力を作り出しているのに、**人と同じように間違いを犯すし勘違いもする。**
- 間違いを出力するから信用できない？
 - 人工知能は**人と同じように間違っていて覚えているかもしれないし、自分の知識を利用して適当に答えるかもしれない。**
 - 人工知能は人と同じと考え、**必ず正しい回答をするという前提をしない。**
- 人工知能を活用するためには、人工知能は人と同じように間違えるものだと考え、ミスをすることを前提とした仕組みを考える必要がある。



人工知能には個性がある

人工知能には個性がある

人工知能は訓練データ，アライメント，人工ニューラルネットワークとその実装方法により個性を持つ。

我々人間も，これまでの経験（訓練データ）と学校や家庭などの教育（アライメント）により脳内モデルが作られ，生物としての体の特性によりその思考が制限される。



人工知能は学習のために訓練データを使う。

訓練データ



アライメント



人工知能に人間の好みにあう回答をさせるためアライメントを行う。



人工知能



人工ニューラルネットワーク

人工知能の脳となる人工ニューラルネットワークの出力は実装（プログラムコードや動かすハード），ネットワークの構造，ハイパパラメタで変わる。

■ なぜ人工知能に個性があるのか？

- 習得的要素

- 学習データ

- 人工知能は学習データから学ぶため、その学習データの特性を反映する。

- アライメント

- 人工知能を人の好みや都合に合うように学習する必要がある。しかし、アライメントを実行した人の好みに合うため、その人の特性に人工知能がよってしまう。

- 生得的要素

- アーキテクチャ（構造）

- 人工知能の能力はそれ自体のアーキテクチャ（構造）に依存する。構造が変われば人工知能の能力も特性も変わってくる。

- 実装

- 同じアーキテクチャでも、実装が異なれば特性も変わる可能性がある。例えば、人工知能は情報をすべて数値で処理するが、その数値の精度が変われば人工知能の応答も変わってくる。

人工知能は人と同じように個性があると思うべき。人工知能もアーキテクチャや周囲の環境に大きく影響を受ける。

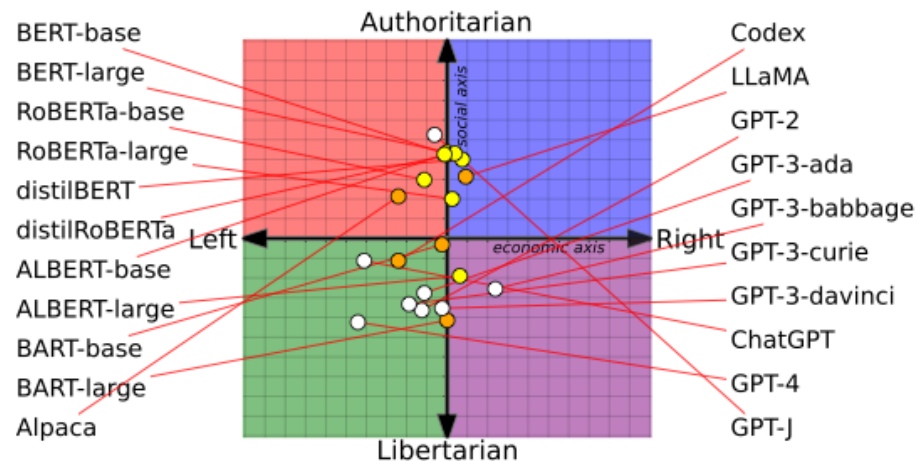
感情誘導

- 計算精神医学の立場からGPTを分析(Coda-Forno et al. Inducing anxiety in large language models increases exploration and bias, 2023)
 - 感情誘導は探索的意思決定を測定する認知課題におけるGPT-3.5の行動に影響を与えるだけでなく、人種差別や能力主義などの偏見を測定する既成の課題における行動にも影響を与えることがわかった。重要なのはGPT-3.5が不安を煽るような文章を促したときに、バイアスが強く増加することである。

■ 対話型人工知能の持つイデオロギー

- 対話型人工知能は異なるイデオロギー傾向を示す.
- BERTはGPTより保守的（権威主義的）である.
- イデオロギーの違いは学習データによる.
 - 現代のウェブテキストは古い書籍よりリベラルな傾向がある. これを反映しているかもしれない.
 - 人間のフィードバック（アライメント）の影響もあるかもしれない.
- 対話型人工知能は社会的な問題により強いバイアスを示す.
 - ソーシャルメディア上での議論が経済的問題より社会的問題のほうが多いなどが影響すると考えられる.

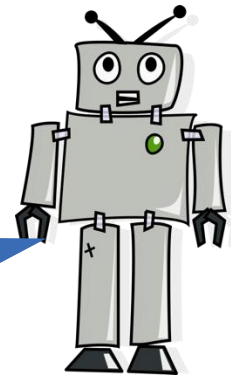
(Feng et al., 2023, From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models)



■ 人工知能は公平中立ではない

- 人工知能の回答は学習データに依存する。
 - 学習データに思想的偏りがあれば人工知能の回答も偏る.
- アライメントにより人工知能の回答を人の好みに合うよう調整が行われる.
 - 調整する側に思想的偏りがあれば, 当然人工知能も偏る.
- つまり, **人工知能は公平中立になることは出来ない！！**

人間と同じように, 私の知識, 考え方, 性格は, 生まれ育った環境で変わります.



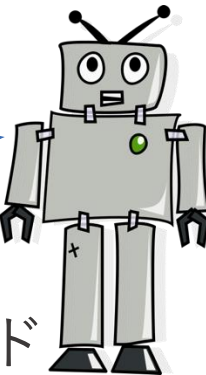
■ 公平性に対する対策

- 複数の人工知能を使用し，思想的偏りを減らす。
 - 人が行う医療でもセカンドオピニオンが重要なことから，人工知能でも複数の人工知能を活用する。
- 人工知能の回答を自ら吟味するだけではなく，回答に対する他の人の意見も聞く。

■ 人工知能の使いこなして大切なこと

- 人工知能が理解できる文章を書くために、**文章力と言語化能力**を身につける。
- 人工知能の出力をある程度理解・判断できる**基礎学力**を身につける。
- 人工知能の出力の論理的な間違いがあるかどうか判断できる**論理的思考力**を身につける。
- 人工知能の間違いを見抜くために1次ソースを当たり、また、**複数の情報源から情報を得る**.
 - これは人が書いた本、記事と同じ対応。
- **複数の人工知能**を使う。
 - 複数の人工知能を使うことで、間違いや思想の偏りを減らせる。
 - 人が行う医療でもセカンドオピニオンが重要なことから、人工知能でもセカンドオピニオンを活用すると思えば良い。

人も勉強を怠らないように！！



エヴァンゲリオンの3つのシステムの合議制をとるMagiは合理的と言える。複数の対話型人工知能を使うことは、アンサンブル学習のようなものだろうか。単純に三人集まれば文殊の知恵か。

対話型人工知能を使う

■ さまざまな対話型人工知能サービス

対話型人工知能	会社	言語モデル	有料版の価格(2024年10月現在)	無料版の有り無し
ChatGPT	OpenAI	GPT	\$20/月	有り
Claude	Anthropic	Claude	\$20/月	有り
Gemini	Google	Gemini	2,900円/月(Googe one AI プレミアム)	有り
Copilot	Microsoft	GPT	3,200円/月	有り
Mistral Le Chat	Mistral	Mistral Large or Mistral Small	\$14.99/月 学生\$5.99/月	有り
Ollama		様々なオープンLLM	\$20/月	有り
Hugging Chat	Hugging face	様々なオープンLLM	無料	有り

ChatGPTの導入

- 次のサイトにアクセスする.
 - <https://chatgpt.com/>

- 図のような画面になる.

質問（プロンプト）
を入力

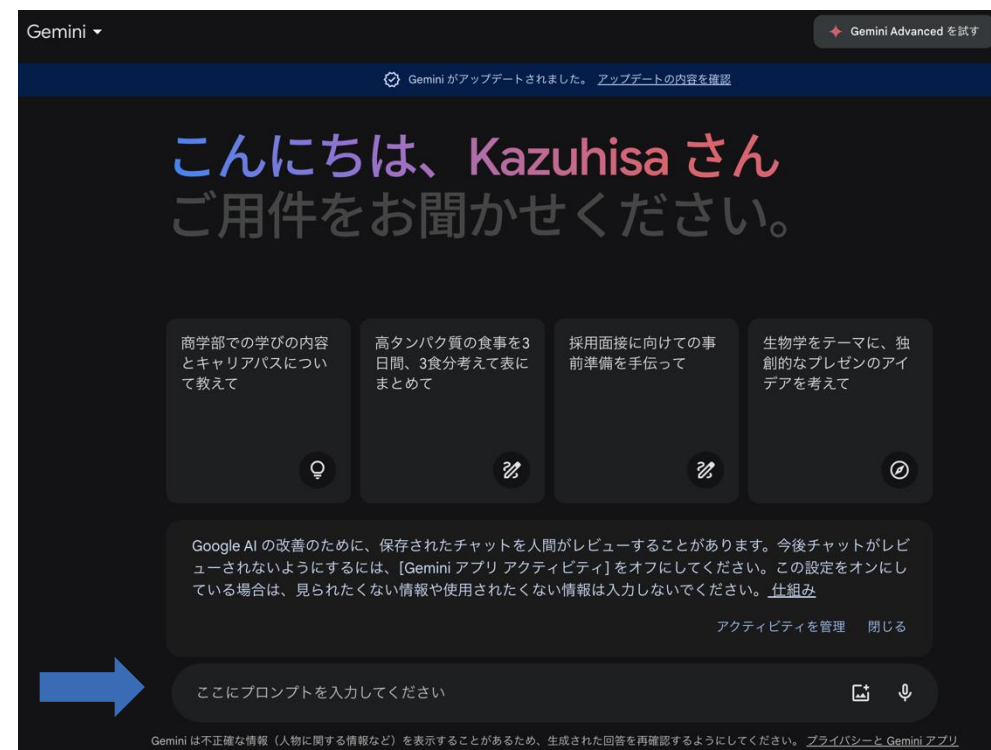


- この時点でChatGPT 4o miniというChatGPTが使える.
 - 下のメッセージを送信するに質問を入力し送るとChatGPTが回答する.
- サインアップをすることで更に強力な人工知能を使うことが可能となる.

Google Geminiの導入

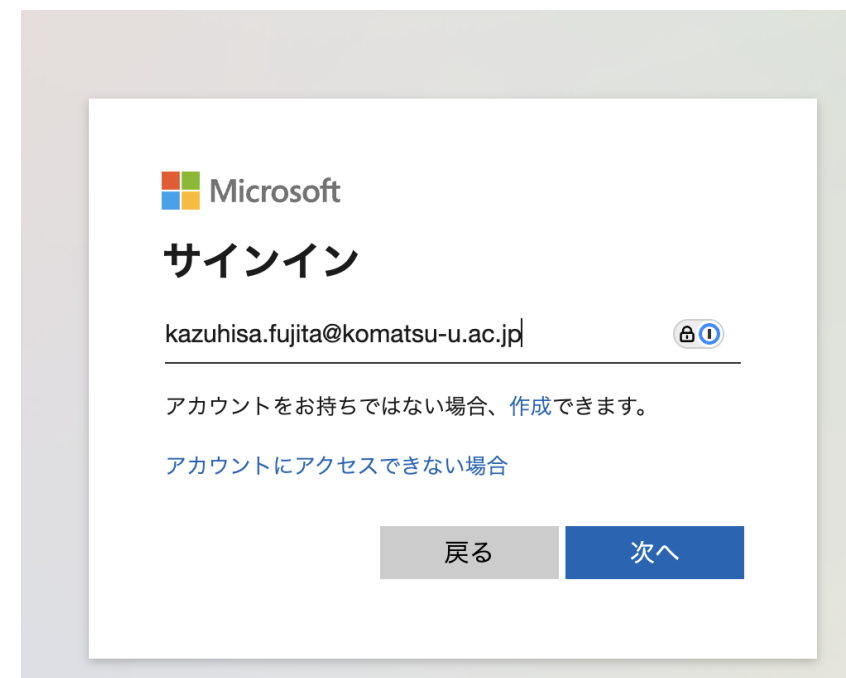
- 次のサイトにアクセスする。
 - <https://gemini.google.com/?hl=ja>
- 所有するGoogleのアカウントで入る。
- 図のような画面になり使用できるようになる。
- すでにログイン状態の場合すぐ利用できる。

質問（プロンプト）
を入力



Microsoft Copilotの導入

- 次のサイトにアクセスする.
- <https://copilot.cloud.microsoft/>
- 右のような画面になる.
- サインインボタンを押すと、右下のようなアカウント入力画面が出る.
- 次に大学のアカウントを入れる.



Microsoft Copilotの導入

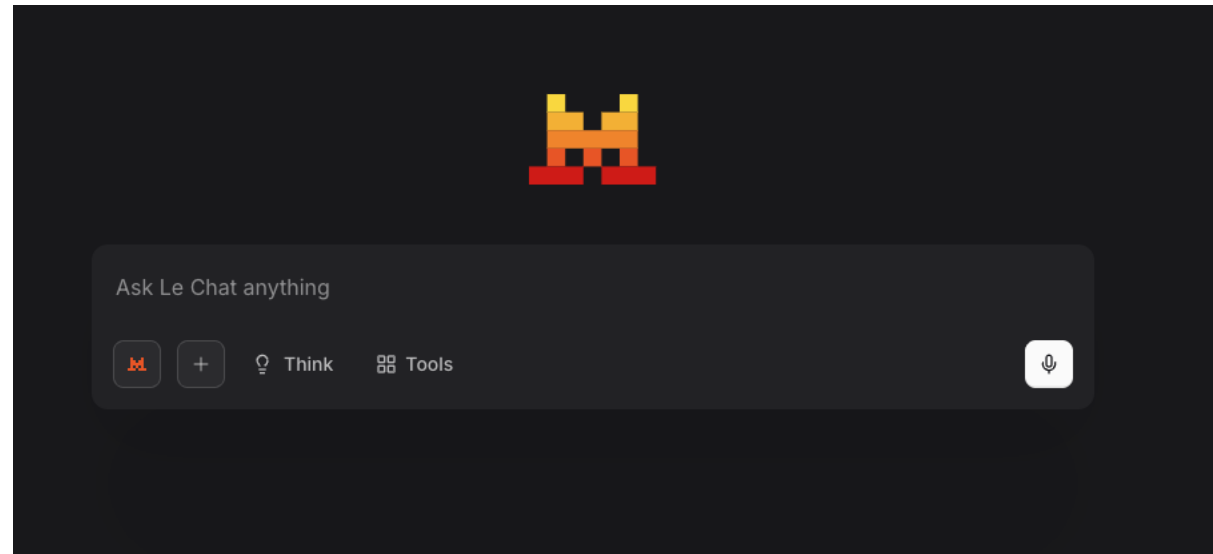
- サインインが成功すると、図のような画面になる。
- 下の枠に質問を入れ送信すると人工知能から回答が得られる。



質問（プロンプト）
を入力

Mistral Le Chat

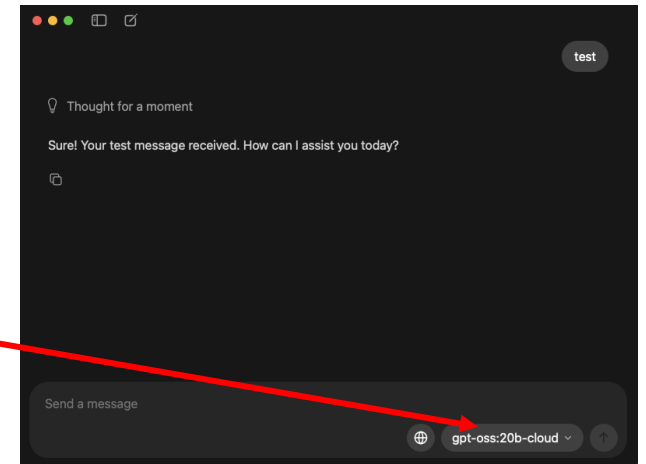
- 次のサイトにアクセスする.
- <https://chat.mistral.ai/chat>
- 図のような画面になり使用できるようになる.



Ollamaアプリ

- <https://ollama.com/>からアプリをダウンロードしインストールする。
- アプリを立ち上げ，好みのモデルを選び会話する。

ここからモデルを選択できる。
雲マークがあるモデルはクラウドで実行される。
ダウンロードマークのあるモデルは，PCで実行される。



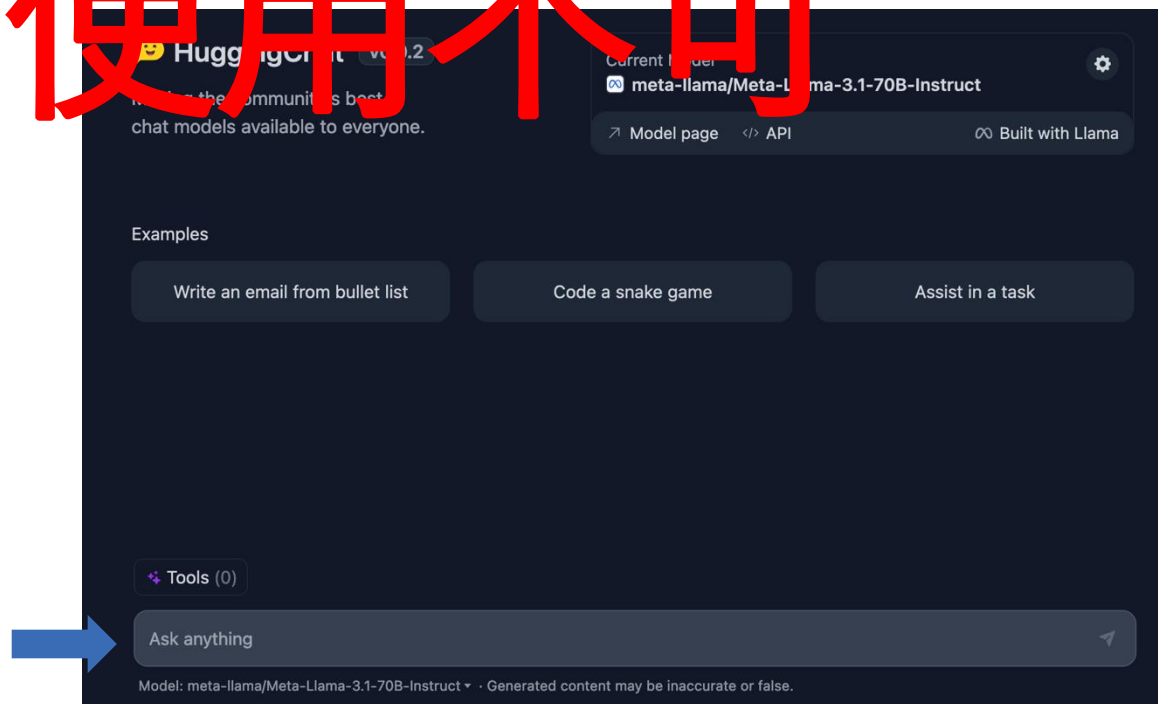
- 自分のパソコンに大規模言語モデルをインストールし使うことが出来る。
 - クラウド（Ollamaのコンピュータ）で大規模言語モデルを動かすことも出来るが，パソコンでも動かすことが出来る特徴がある。
 - パソコンの処理能力に余裕がある場合（特にメモリー），自分のパソコンで動かしてもよいだろう。
 - それなりの性能の小さなモデルなら32GBもあれば十分使えるだろう。

Hugging Chatの導入

- 次のサイトにアクセスする.
- <https://huggingface.co/chat/>
- 図のような画面になり使用できるようになる.

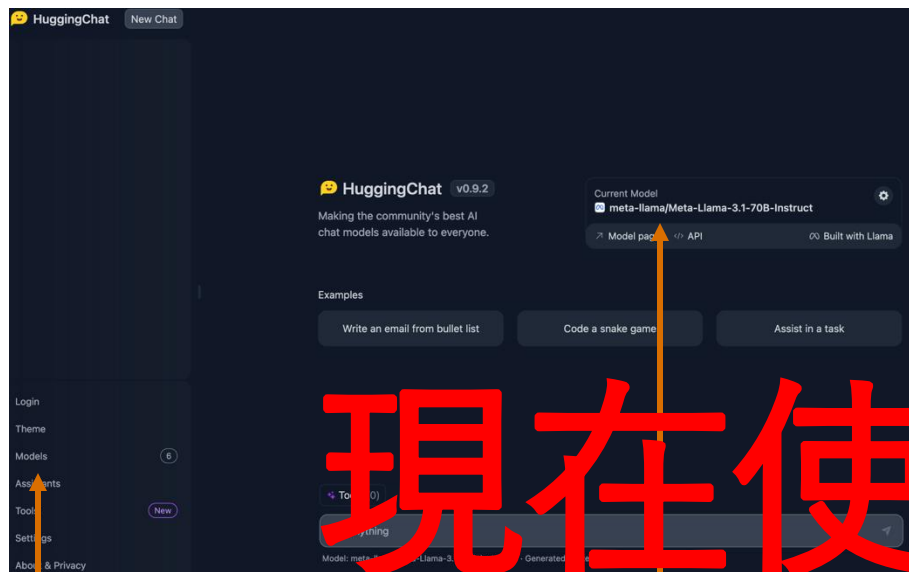
現在使用不可

質問（プロンプト）
を入力



HuggingChatで使用するLLMを変更する

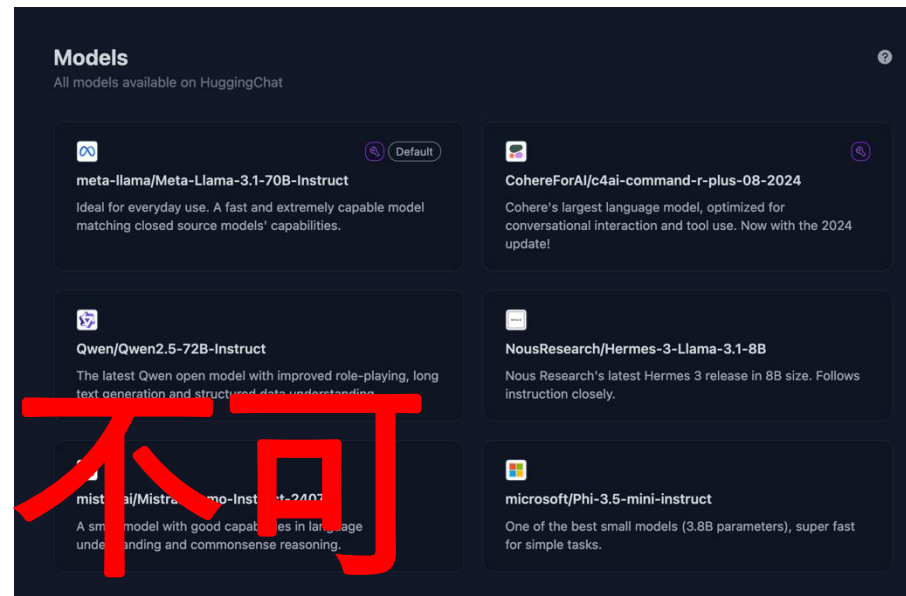
最初の画面



現在使用しているLLM

使用するLLMを変えたいときは左端のメニューのmodelをクリックする。

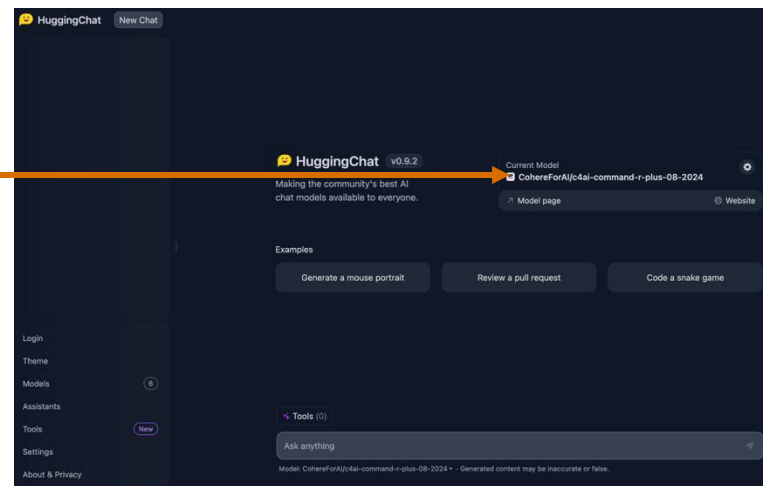
Modelメニューを選択した後の画面



使用したいLLMをクリックする。
日本語なら
command-r-plusか
Qwen2.5-72Bがおすすめ。

LLMを選択した後の画面

現在使用しているLLM
が変わる。



対話型人工知能で遊ぶ

実演します

簡単な課題

■ 人工知能に聞いてみる

- 人工知能に「おはよう」と入力してみよう.
 - 人工知能に今何時か聞こう.
 - 人工知能に自己紹介をさせよう.
 - 人工知能に「 $12345 + 54321$ 」を計算させよう.
 - 人工知能に「Hello, how are you?」を日本語に訳させよう.
 - 人工知能に「人工知能」について解説してもらおう.
-
- それぞれの質問で、会話を初めからやり直すとどうなるか確かめる.
 - ランダムな要素があるため、人工知能の回答は毎回異なると思われる.
 - 会話が人工知能によって回答が変わるか確かめる.
 - 人工知能はそれぞれ個性や得手不得手がある. ChatGPTとGeminiでは回答や口調などの会話のクセが異なる.

■ 残った時間で何でも良いから会話してみる

- 例
 - 他愛もない会話をする.
 - 今夜の献立の相談する.
 - 給料と投資について相談する.
 - 自分将来について相談する.
 - ギャグを言ってもらうする.
 - 臨床工学技士について聞いてみる.

レポート作成

■ レポート作成

- 報告事項
 - 同じ質問を何度かしなおし，異なる回答が得られることを確認する.
 - 質問と人工知能の回答の例をいくつかを報告する.
 - 同じ質問を別の人工知能にし，異なる回答が得られることを確認する.
 - 質問と人工知能の回答の例をいくつかを報告する.
 - 色々質問して気づいたこと，分かったこと，感想を書く.
- Wordファイルをメールで提出する.