



ICA第3回講演会 生成AIを含む最新AI状況について

株式会社 数理先端技術研究所
2023/8/21 生島高裕

自己紹介



株式会社 数理先端技術研究所 代表取締役 生島高裕

(日本オミックス医学会 理事、公共貨幣フォーラム 理事、NPOサイバー絆研究所 監事)

2007年会社設立後、数値計算、数式処理、制御システム、機械学習、HPC、知識システム、AIに関する研究開発を行っている。

対象分野は、核融合、宇宙、物質構造、バイオ、医療、金融などで、分野における最先端課題のソリューションを提供している。

最近では日本オミックス医学会で創造性についての企画を行いながら、広く哲学、倫理学、社会学、政治学、経済学関連でAIの動向分析を行っている。

理論物理、数学でフィールズ賞をAIが取る日をシンギュラリティと定義している。

アジェンダ



1. 生成AI前史
2. 深層学習と強化学習(2010年以降)
3. LLM(大規模言語モデル)と生成AI
4. 創造性とは
5. 研究分野、産業におけるAI、LLMの利用
6. 教育とLLM
7. 課題
8. ICA、オミックス医学会、JASISの履歴、ETC

1. 生成AI前史

● 歴史

20150903—人工知能からの新しい産業社会環境の動向【ASIS2018】

第1次AIブーム 1950-1960

情報科学者が中心に研究が進んだ時代

- 脳は神経細胞の電気ネットワーク
- ノーバート・ウィナーのサイバネティクス
- クロード・シャノンの情報理論
- アラン・チューリングの計算理論
- チューリングテスト 1950
- ノイマン

数学基礎論、ゲーム理論、モンテカルロ法、量子力学の数学的基礎、経済成長理論
ノイマン型コンピュータ、セル・オートマトン


- 記号的推論と Logic Theorist 1955 アレン・ニューウェルとハーバート・サイモン

このプログラムはラッセルとホワイトヘッドの「プリンキピア・マテマティカ」の最初の方にある

52 の定理のうち 38 の定理を証明してみせ、そのうち一部は新たな洗練された証明方法を見出した。

- ダートマス会議 1956

1. 生成AI前史




第2次AIブーム 1980-1990

医学系、大企業での応用が始まった。

- エキスパートシステム
- 第五世代コンピューティング
- コネクショニズムと記号主義

1. 生成AI前史



第3次AIブーム 2010- (1)

携帯電話、Web サービスには広く機械学習などの AI 技術が使われてるようになった。

このことは、全エンドユーザが AI の利用者であることを意味する。

従って、もはや、「シンギュラリティの日」までそれ以降ずっと続くことになった。

- AI ベンチャーの一覧

 - 最近の世界中の AI ベンチャーをまとめた図

 - The Current State of Machine Intelligence

- 新ネットワーク思考

- 予測脳

- Semantic Web

- 強化学習

- 機械学習

- ワトソン

1. 生成AI前史

第3次AIブーム 2010- (2)

- DNN (Deep Neural Network)
 - シンボルグラウンディング、フレーム問題
 - 言語ネットワークから新しい特徴量を抽出できるか？
 - 数学的シンボルを使って概念をまとめる。
- 進化計算
 - 好奇心、知識欲 ⇒ 学習、整理
 - DNN による特徴量抽出 ⇒ 疑問、課題創出
 - 未解決問題として特徴量抽出 ⇒ 概念ネットワークの変更と評価
- 全脳アーキテクチャ -whole brain architecture-
 - 人工知能、神経科学、認知科学の領域を横断し、全脳アーキテクチャの実装を目指すグループ
- SIG 汎用人工知能
 - AGI: artificial general intelligence および「技術的特異点」について広く情報交換 とディスカッションを行うためのグループ

1. 生成AI前史

第1回:人工知能の概要とディープラーニングの意義 松尾豊

2014年9月での未来予測 P49 2015年12月、再度修正 P51

- 第1次AIブーム(1956～1960年代):探索・推論の時代
- ...冬の時代
- 第2次AIブーム(1980年代):知識の時代
- ...冬の時代
- 第3次AIブーム(2013年～):機械学習、表現学習の時代

人工知能の未来 - ディープラーニングの先にあるもの松尾豊 2016/02/24 総務省

1. 生成AI前史

● コネクショニズムとシンボリズム

<コネクショニズム>

- 脳は神経細胞の電気ネットワーク
- サイバネティクス
- クロード・シャノンの情報理論
- エキスパートシステム
- 新ネットワーク思考
- 予測脳
- 強化学習
- DNN (Deep Neural Network)
- 進化計算

<シンボリズム>

- アラン・チューリングの計算理論
- チューリングテスト 1950
- ノイマン
- 記号的推論
- 第五世代コンピューティング
- Semantic Web
- 進化計算

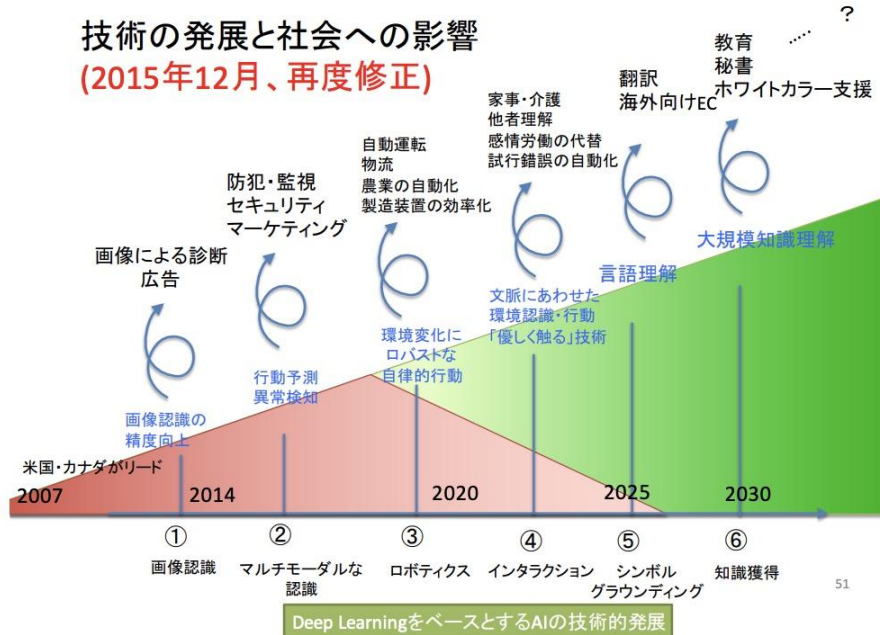
数学、言語、推論、知識表現的？

統計、確率、Neural Network的？


1. 生成AI前史

● 今後のロードマップ(シンギュラリティ議論)

第1回:人工知能の概要とディープラーニングの意義 松尾豊



1. 生成AI前史



● 今後のロードマップ(シンギュラリティ議論)

2014:画像認識の精度向上

2025:自然言語意味理解(平均的人間の言語能力を持つAI)

2035:汎用人工知能(8歳ぐらいのレベルになり、自律的に掃除とか介護ができるようなロボットの出現です。(自我の目覚め))

2045:シンギュラリティ、超知能(人により定義がバラバラ)、私の場合、フィールズ賞を取るAIの出現

2. 深層学習と強化学習(2010年以降)

● chatGPT、Bingで整理して見る

Q: ディープラーニングの概念史を作ろうと思います。代表的な概念を年代順にあげてください。

A: 代表的な概念を年代順に示します。

1940年代-1950年代: マッカロックとピッツ

1960年代-1970年代: フランク・ローゼンブラット、パーセプトロンに関するミンスキーとパパートの論文

1980年代-1990年代: バックプロパゲーション(誤差逆伝播法)アルゴリズム、シンボリックAIからコネクショニズム

2000年代: GPUの普及により、大規模なニューラルネットワークの訓練が加速

2010年代: AlexNet、CNN、LSTM、GAN

2020年代: トランスフォーマー

ほぼ概要はよく聞く回答になる。

ボツルマンマシンがなかったので質問、その関係でホップフィールド・ネットの違いを聞いた。

概要はいいのだが、停止条件を聞くとボツルマンマシンであやふやな回答、突っ込むと謝りながら合意形成はできた。

「誤差」など定義があやふやに感じたときは徹底的に聞くとよい。

2. 深層学習と強化学習(2010年以降)

Q: さて「ディープラーニングの概念史」ですが、2010年代からあなたの知識の2021年9月までで、重要な10の概念を重要な順にあげてください。

A: ディープラーニングの概念史において、2010年代から2021年9月までの間に重要な10の概念を重要な順に以下に挙げます。

1. **畳み込みニューラルネットワーク(CNN)の台頭**: AlexNetの登場により、畳み込みニューラルネットワークが画像認識タスクで驚異的な性能向上を達成しました。これにより、CNNがコンピュータビジョン分野で主流のアーキテクチャとなりました。
2. **リカレントニューラルネットワーク(RNN)の進化**: 長短期記憶(LSTM)やゲート付きリカレントユニット(GRU)などのリカレントネットワークの改良により、シーケンスデータ処理や自然言語処理で大きな進展がありました。
3. **Generative Adversarial Networks (GANs)の登場**: GANsは、生成モデルと識別モデルが競い合う構造を持ち、写真のようなリアルな画像や他のデータを生成することが可能な手法として注目を集めました。
4. **Word2Vecと自然言語処理の革命**: Word2Vecやその他の単語埋め込み手法は、単語の意味をベクトル表現として捉え、自然言語処理タスクで大きな進歩をもたらしました。
5. **トランスフォーマーの登場**: トランスフォーマーはセルフアテンションメカニズムを導入し、自然言語処理タスクで圧倒的な成果を上げるモデルとなり、BERTやGPTなどの派生モデルが次々と登場しました。

2. 深層学習と強化学習(2010年以降)

Q:さて「ディープラーニングの概念史」ですが、2010年代からあなたの知識の2021年9月までで、重要な10の概念を重要な順にあげてください。

A:ディープラーニングの概念史において、2010年代から2021年9月までの間に重要な10の概念を重要な順に以下に挙げます。

6. 転移学習とプレトレーニングの重要性の浮き彫り: 大規模なプレトレーニングデータと事前学習によって、特定のタスクにおける性能向上が実現されるようになりました。

7. 自己教師あり学習の探求: 教師データを必要とせずに学習する手法が注目され、半教師あり学習や自己教師あり学習が研究されました。

8. 強化学習の進展: 深層強化学習によって、ゲームやロボティクスなどの領域で、人間を上回る成果を達成する手法が登場しました。

9. 自己適応システムとAutoML: ニューラルアーキテクチャの設計やハイパーパラメータのチューニングを自動化するAutoML技術が注目され、効率的なモデル開発が進展しました。

10. ディープラーニングの多様な応用: 画像・音声・自然言語だけでなく、医療、金融、製造業など多様な分野でディープラーニングが応用され、新たな可能性が広がりました。

2. 深層学習と強化学習(2010年以降)

この回答に対して、自分の重要な100あまりの論文リストの中で漏れている感じのキーワードを要求してみました。

Q: Dropout、VAE (変分オートエンコーダ)、蒸留、深層強化学習、AlphaZero、メタラーニングなど追加するとどうなるでしょうか？10を超えてもかまいません。

A: もちろんです。以下に、追加の概念を含めたディープラーニングの重要な概念をいくつか挙げてみましょう。

1. **Dropout**: ニューラルネットワークの学習中にランダムに一部のユニットを無効化することで、過学習を防ぐ手法。
2. **変分オートエンコーダ (VAE, Variational Autoencoder)**: オートエンコーダの一種で、生成モデルを構築し、データを低次元の潜在空間にマッピングする際に確率分布を考慮する。
3. **蒸留 (Knowledge Distillation)**: 大きなモデル(教師モデル)の知識を小さなモデル(生徒モデル)に転移させる手法。モデルの軽量化や高速化に利用される。
4. **深層強化学習 (Deep Reinforcement Learning)**: ニューラルネットワークを用いて強化学習を行う手法。特に画像データを入力とする場合、深層強化学習が注目されました。
5. **AlphaZero**: Google DeepMindによって開発された、ボードゲームのAIプレイヤーで、強化学習とモンテカルロ木探索を組み合わせ高レベルのプレイ力を持つことで知られています。

2. 深層学習と強化学習(2010年以降)

この回答に対して、自分の重要な100あまりの論文リストの中で漏れている感じのキーワードを要求してみました。

Q: Dropout、VAE (変分オートエンコーダ)、蒸留、深層強化学習、AlphaZero、メタラーニングなど追加するとどうなるでしょうか？ 10を超えてもかまいません。

A: もちろんです。以下に、追加の概念を含めたディープラーニングの重要な概念をいくつか挙げてみましょう。

6. **メタラーニング (Meta-Learning)**: モデルが新しいタスクを効率的に学習するためのアプローチ。少ないデータで新しいタスクに適応できる柔軟性を持つ。
7. **GANの進化と応用**: GANの拡張として、Conditional GAN (条件付きGAN)、CycleGAN (画像変換)、StyleGAN (画像生成)などが登場し、様々な応用がされています。
8. **注意機構 (Attention Mechanism) の普及**: トランスフォーマーに導入された注意機構が、自然言語処理やコンピュータビジョン分野で多くのタスクにおいて効果を発揮しました。
9. **自己教師あり学習 (Self-Supervised Learning) の台頭**: 教師データを必要としないデータの自己教育を用いた学習手法が注目を浴び、様々なタスクに適用されました。
10. **クラウドコンピューティングとディープラーニング**: クラウドサービスの普及により、ディープラーニングのモデルの訓練と推論が大規模かつ効率的に行われるようになりました。

2. 深層学習と強化学習(2010年以降)

後、フューショットラーニング、ゼロショットラーニング、ReLUなど加えていきました。

19概念が出てきたので整理するために次の要求を行いました。

Q: ディープラーニングの概念史は19概念が出てきました。

これを年代順に並べ替えてもらえますか？

と19概念の全項目を提示する。

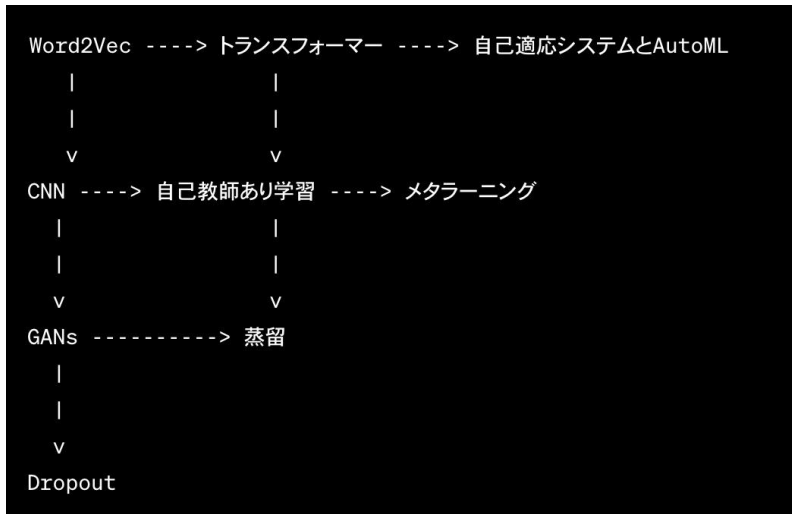
A: 以下に、提供いただいた19の概念を年代順に並べ替えてみました。

1. Word2Vecと自然言語処理の革命
 2. 畳み込みニューラルネットワーク(CNN)の台頭
 3. リカレントニューラルネットワーク(RNN)の進化
 4. GANs (Generative Adversarial Networks)の登場
 5. トランスフォーマーの登場
 6. 強化学習の進展
 7. Dropout
 8. 自己教師あり学習の探求
 9. 転移学習とプレトレーニングの重要性の浮き彫り
 10. 自己適応システムとAutoML
 11. 注意機構 (Attention Mechanism)の普及
 12. AlphaZero
 13. メタラーニング (Meta-Learning)
 14. 蒸留 (Knowledge Distillation)
 15. 変分オートエンコーダ (VAE, Variational Autoencoder)
 16. ゼロショットラーニング (Zero-shot Learning)
 17. フューショットラーニング (Few-shot Learning)
 18. ReLU, GELU
 19. 自己教師あり学習 (Self-Supervised Learning)の台頭
- キーの論文の順番にはソートできなかった。

2. 深層学習と強化学習(2010年以降)

Q:これらの概念間の関係はグラフ化できますか？

A:はい、これらの概念間の関係をグラフ化することはできます。ただし、関係性や影響の程度については一概に言えない部分もありますが、一般的な関連性を表現するグラフを作成することは可能です。以下に、いくつかの概念の関係性を簡単なグラフで示してみましよう。なお、これは抽象的な関係を表現したものであり、実際の複雑な関係性を完全に網羅しているわけではありません。



セマンティックWeb的グラフが書ければいいですね。
それができると推論が走れるし、様々なアノテーションが開く可能性があります。

2. 深層学習と強化学習(2010年以降)

各概念に年代、提案者とか論文を聞いていきます。

ゲーデルマシンなどあまり知られていないものは、知らないと答えたり、謝ったりするので細く説明をすると探してきて回答します。

こんな回答も。

「ゲーデルマシンは、メタラーニング("learning to learn")としても知られる問題に取り組む際に頻繁に議論されます。アプリケーションには、人間の設計決定の自動化や複数の関連するタスク間での知識の転送などが含まれ、より堅牢で一般的な学習アーキテクチャの設計につながる可能性があります。理論的には実現可能ですが、現時点では完全な実装は存在していません。」

この後、代表論文を聞いてそこでググって論文を見つけましたが、Bingなら一度に回答してくれます。

2. 深層学習と強化学習(2010年以降)

● ディープラーニングの概念史

以下は深層学習に関連するいくつかの重要なキーワードです:

ニューラルネットワーク (Neural Networks)

畳み込みニューラルネットワーク (CNN)

リカレントニューラルネットワーク (RNN)

長短期記憶 (LSTM)

ゲート付きリカレントユニット (GRU)

トランスフォーマー (Transformer)

注意機構 (Attention Mechanism)

ジェネレイティブアドバーサリーネットワーク (GAN)

強化学習 (Reinforcement Learning)

自己教師あり学習 (Self-Supervised Learning)

メタラーニング (Meta-Learning)

プレトレーニング (Pretraining)

転移学習 (Transfer Learning)

ゼロショット学習 (Zero-shot Learning)

フューショット学習 (Few-shot Learning)

蒸留 (Knowledge Distillation)

ドロップアウト (Dropout)

活性化関数 (Activation Function)

オートエンコーダ (Autoencoder)

バックプロパゲーション (Backpropagation)

2. 深層学習と強化学習(2010年以降)

<対話>

「ボルツマン・マシンとホップフィールド・ネットの違い」を聞くと、うまく説明するが、明確でないところを突っ込むと、それ用の説明を返してきます。

怪しい説明を聞き直すと修正してきます。

それはあたかも先生に質問して納得がいかない時、繰り返しドリルダウンする方法に似ています。

先生でも、chatGPTでも、答えられられなくなる、トートロジーになる、と対話は終わりですね。

それはあたかも大量の学習データをいろんな視点(プロンプト視点)で問い直し、全体を浮かび上がらす行為に似ています。

意味は固定し、視点を変えた表現ができていく感じがします。

<概念抽出>

ディープラーニングの概念史」で重要な10の概念を重要な順に挙げさすとそれなりの回答がある。

これは広い意味での要約になっている。

<各概念の代表論文>

問題なく出力される。

<概念の年代ソート>

少し順番に問題ありとは思いますが背景知識の順序性はある程度反映できている気がする。

<概念関係のグラフ化>

部分的にはできるが、マトリックス構造のみ、将来はセマンティック WebのRDF、OWL表現で書けるか？

プログラミングがある程度できるので可能性はある。

2. 深層学習と強化学習(2010年以降)

プロンプトエンジニアリング:ソクラテスの問答法

自然言語処理のモデルに対して、効果的な命令や質問を与えることで、望ましいテキストを生成させる技術

モデルが実行すべきタスクを明示的に教えるのではなく、自然言語で書かれたプロンプト(刺激)を用いて、モデルの能力を引き出す

指示(Instruction):モデルが実行する特定のタスクや命令

背景(Context):モデルの回答精度を高めるための追加の情報や文脈

入力データ(Input Data):モデルに応答を求める入力や質問

出力形式/出力指示子(Output Indicator):出力タイプやフォーマット

ソクラテスの問答法に関係していると思われる。

GAN:弁証法

生成モデルと識別モデルが競い合う構造

蒸留:思惟経済?

既に学習してあるモデルを使用し、より軽量なモデルを生み出すこと、学習済みの教師モデルの出力を生徒モデルの学習に利用
数学、物理などの理論を整理する話と対応?論文から教科書への変換など、ブルーナーの発見学習にも関係すると思われる。

2. 深層学習と強化学習(2010年以降)

アテンション(注意): 定位、フィルターリング、探索

別のデータから関連する情報を選択的に取り出す仕組み

トランスフォーマーは、アテンションを効率的に積層した深層学習モデル、エンコーダとデコーダから構成

エンコーダは入力データを受け取り、Self-Attentionと呼ばれる仕組みを用いて、データ内の各要素が互いに注目する度合いを計算

データ内の構造や関係性を捉えることができる

デコーダは出力データを生成するために、Masked Self-Attentionと呼ばれる仕組みを用いて、自分自身の過去の要素に注目しながら次の要素を予測

Cross-Attentionと呼ばれる仕組みを用いて、エンコーダから受け取ったデータから必要な情報を抽出

これにより、入力データと出力データの対応関係を学習することができる

GPTとBERTはいずれもトランスフォーマーアーキテクチャをベースにしたモデルで、事前学習と転移学習のアプローチを使用

人間の「注意」には定位、フィルターリング、探索の3つが現象論的に言われている。

精神疾患を起こすとフィルターリングができなくなる場合があり、すべての情報を均等に扱いフレーム問題を起こす

Word2Vec: ひょっとしたら人間もやっている可能性がある

単語の意味のベクトル表現

Dropout: 脳を不完全な状態で酷使すると汎化性能が上がるかもしれない

学習中にランダムにニューロンを無効化することで、汎化性能を高める手法

2. 深層学習と強化学習(2010年以降)

転移学習: 人気の場合は T型学習、 π 型学習と言われ、専門が深いと他の分野に応用が効く話がある

「解いた設問の解法を別の設問に応用」

1. 学習済みモデルを入手する
2. 出力層だけ目的のタスク向けに交換する
3. 出力層以外のネットワークの重みは固定する
4. 学習ステップ
 - ・用意したデータセット用いて損失を計算
 - ・得られた損失を用いて出力層の重みを更新

ファインチューニング: 人間は転移学習のあともう一度対象分野で整理し直す場合がよくある

「一度解いた設問の解法を、別の設問のために微調整」

1. 学習済みモデルを入手する
2. 出力層だけ目的のタスク向けに交換する
3. 学習ステップ
 - ・用意したデータセット用いて損失を計算
 - ・得られた損失を用いてネットワーク全体の重みを更新

3. LLM(大規模言語モデル)と生成AI



● 生成AI(Generative AI)

識別的、生成的

データのパターンや関係を学習し、新しいコンテンツを生成すること

● LLM(Large Language Model)

自然言語を理解し、生成するAIの一種で、情報検索、質問応答、機械翻訳、文章生成など、多岐に渡るタスクに対応
大規模なデータセットを用いた機械学習によって精度を高めた自然言語処理モデル

3. LLM(大規模言語モデル)と生成AI

驚異的な普及

2022年12月

- **OpenAIがChatGPTを公開**（正確な公開日は2022年11月30日）
- ChatGPTは公開後5日で100万ユーザを獲得
- Stack OverflowがChatGPTで生成された投稿を禁止
- **GoogleがChatGPTに関して「コードレッド」を宣言と報道**

2023年1月

- 機械学習に関する国際会議ICMLが生成型AIで論文を執筆することを禁止
- 自然言語処理に関する国際会議ACLが生成型AIに関するポリシーを発表
- Natureが論文の共著者としてChatGPTを認めない方針を発表
- **ChatGPTの月間アクティブユーザの推計が1億人に到達（Instagramを抜き過去最速）**

2023年2月

- OpenAIがサブスクリプションサービスChatGPT Plusを発表
- **MicrosoftがChatGPTを搭載した検索エンジンBingを発表**
- Googleが（ChatGPT対抗と言われる）対話型AIのサービスBardを限定公開
- Metaが大規模言語モデルLLaMA（7B～65B）を（モデルのパラメータも含めて）公開

2023年3月

- OpenAIがChatGPTとWhisperのAPIを公開
- MicrosoftがAzure OpenAI ServiceでChatGPTを提供
- **OpenAIがGPT-4を発表**
- GitHubがGPT-4を搭載したCopilot Xを発表

2023年4月

- 個人情報保護への懸念から、ChatGPTのイタリアでの提供を禁止
- GPT-4よりも強力なAI技術の開発と実験を6か月停止する公開書簡を発表
- OpenAIのCEOが岸田首相と面会し、日本でのサービス拡充を提案
- イラストレーターや漫画家の団体が画像生成AIの適切な使用や法整備を求める提言を提出

2023年5月

- 日本政府が人工知能に関する政策の司令塔機能を担う「AI戦略会議」を設置
- Googleの対話型AIサービスBardの新バージョン（PaLM 2）が日本語に対応
- 日本語に対応した大規模言語モデルがrinnaやサイバーエージェントから発表される
- G7広島サミットの首脳声明で、信頼できる人工知能というビジョンと目標が掲げられる

3. LLM(大規模言語モデル)と生成AI

● 大規模言語モデルの活用法

- 検索の代替
 - Web検索代わり、回答への追加質問、手順や理由などの質問に直接回答
- パーソナライズされた相談相手
 - 授業や地涌学習におけるチューター、カスタマーサポート、人生相談、バーチャル秘書
- 汎用的な問題解決器
 - 最先端のAIをプログラミング無しで活用
 - 機械学習、自動要約、英文校正、感情分析、情報抽出などのタスクを汎用的に担当
- 専門家支援
 - 問題の解決策のアイデア列挙、問題に対するアプローチの仕方を知る
 - コンテンツ(文章、画像、動画)の生成
- DX・プログラム生成
 - 約束事が多い文章を高品質に生成、ある程度のプログラム
- データ拡張
 - 訓練データの自動生成、人間よりも質の高いデータを作成する場合もある
 - 将来的には辞書、wikiの自動生成、集合知の自動生成

3. LLM(大規模言語モデル)と生成AI

● 言語モデルとは

単語(トークン) 列 y_1, \dots, y_T 成確率 $P(y_1, \dots, y_T)$

テキストの続き (あるテキストに続く単語) を予測できる

● 深層学習への準備: 単語埋め込み

単語の意味をベクトル (単語埋め込み) で表現する

```
m['アメリカ']
```

```
array([-0.01, -0.41,  0.06, -0.38, -0.15,  0.15, -0.37, -0.48, -0.15,  0.54,  0.08, -0.05, -0.04, -0.02, -0.14,  0.02,
        0.29, -0.08,  0.06,  0.22, -0.18, -0.03, -0.49, -0.18,  0.26,  0.33, -0.34, -0.03, -0.11,  0.34, -0.41,  0.15,
        0.29, -0.06, -0.19, -0.12,  0.09, -0.14, -0.06, -0.20, -0.22, -0.04,  0.20, -0.01, -0.25,  0.06,  0.36, -0.15,
        -0.19,  0.16, -0.18,  0.03,  0.37, -0.43, -0.23,  0.18, -0.05, -0.13,  0.10, -0.01, -0.43, -0.17, -0.30,  0.48,
        0.38, -0.17,  0.12,  0.07,  0.10,  0.03,  0.13,  0.34, -0.15, -0.12,  0.03,  0.20,  0.12, -0.25,  0.30, -0.29,
        -0.04, -0.27,  0.08, -0.19, -0.05, -0.25, -0.35,  0.11,  0.50, -0.15,  0.06,  0.22,  0.13, -0.11,  0.00,  0.03,
        0.18,  0.35,  0.02,  0.28, -0.28,  0.48, -0.18, -0.20, -0.34,  0.39, -0.15,  0.10,  0.22, -0.26,  0.29, -0.16,
        -0.19,  0.04, -0.02,  0.31, -0.30, -0.17, -0.53, -0.33, -0.33, -0.05, -0.20,  0.08,  0.25, -0.34,  0.20,  0.37,
        0.46, -0.39, -0.10, -0.09, -0.21,  0.60, -0.07, -0.38, -0.22, -0.00, -0.00, -0.08,  0.06, -0.33, -0.72, -0.11,
        0.20, -0.16, -0.05,  0.01, -0.28,  0.00, -0.34, -0.10, -0.07,  0.43, -0.49, -0.03,  0.15, -0.35,  0.38, -0.07,
        0.10, -0.13, -0.33,  0.03, -0.17, -0.00,  0.23, -0.42,  0.16, -0.40,  0.05,  0.03, -0.27,  0.29,  0.11,  0.08,
        -0.06, -0.35,  0.36,  0.21,  0.32,  0.06, -0.06, -0.16, -0.64,  0.13, -0.26,  0.07, -0.08, -0.11, -0.01,  0.42,
        0.01,  0.06,  0.33, -0.40, -0.22, -0.24,  0.02, -0.11,  0.19,  0.03, -0.17, -0.11,  0.22, -0.02,  0.02, -0.24,
        0.09,  0.12, -0.07,  0.14,  0.01,  0.17,  0.13, -0.05,  0.28,  0.38, -0.14, -0.10,  0.13, -0.13,  0.15, -0.01,
        -0.19,  0.01,  0.27, -0.07,  0.27,  0.12,  0.02, -0.31,  0.30, -0.10, -0.03,  0.04,  0.04,  0.37,  0.22,  0.55,
        -0.08,  0.00, -0.03, -0.10, -0.16, -0.28, -0.22,  0.15,  0.14,  0.03, -0.05, -0.09,  0.14,  0.23, -0.14,  0.10])
dtype=float32)
```

256次元のベクトルで表現された「アメリカ」 (日本語Wikipedia記事テキストで学習したもの)

```
m.similar_by_vector(m['アメリカ'])
```

```
m.similar_by_vector(m['パリ'] - m['フランス'] + m['アメリカ'])
```

3. LLM(大規模言語モデル)と生成AI

●再帰型ニューラル言語モデル (RNNLM) (Mikolov+ 2010)

埋め込み表現 (単語ベクトル) により類義語 関連語を考慮できる

TMikolov, M Karafiát, L Burget, J Čermocký, S Khudanpur. 2010.

Recurrent Neural Network Based Language Model. In INTERSPEECH, pp. 1045-1048.

●Transformer の登場 (Vaswani+ 2017)

- **自己注意**だけで単語間の情報を統合するモデル
- **位置エンコーディング、マルチヘッド注意、残差結合、層正規化**などの工夫を盛り込む
 - 単語間の情報の統合に要するコストが距離に依らない (長距離依存を扱いやすい)
 - 並列計算で実装しやすい (GPUやTPUなどのハードウェアを活用しやすい)
 - 大規模言語モデルに限らず、自然言語処理以外の分野も含めて、汎用的に用いられる基盤アーキテクチャとなった

A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A N. Gomez, L Kaiser, I Polosukhin. 2017.

Attention is All You Need. In NIPS, pp. 5998-6008.

3. LLM(大規模言語モデル)と生成AI

● Generative Pre-Training (GPT)

Transformerの生成部 (デコーダ) を言語モデルとして学習したもの (事前学習)

学習データとして本やウェブから収集した大量のテキストを用いる

大量のテキストから言語に関する一般的な知識を獲得することを狙う

[A Radford, K Narasimhan, T Salimans, Sutskever. 2018. Improving Language Understanding by Generative Pre-Training. Technical report.](#)

● ファインチューニング

所望のタスクが解けるように事前学習済みの (言語) モデルを調整 (学習) すること

言語に関する一般的な知識をベースに、タスク固有の知識を獲得することを狙う

2018年頃から流行しているアプローチ (代表例はGPT, BERT, ELMo)

大規模なモデルではファインチューニングのコストが高い

[A Radford, K Narasimhan, T Salimans, I Sutskever. 2018. Improving Language Understanding by Generative Pre-Training. Technical report.](#)

3. LLM(大規模言語モデル)と生成AI

● プロンプト: 言語モデルだけでタスクを汎用的に解く

大規模言語モデルに与える「テキスト (=プロンプト)」を工夫すれば、言語モデルのアーキテクチャを変更せずにタスクを解けるのではないか? (GPT-2とGPT-3の論文) テキストに続く単語を予測させる (通常の言語モデルの使い方)

[T. B. Brown, B. Mann, N. Ryder, M. Subbiah, et. al. 2020. Language Models are Few-Shot Learners.](#)

● [Prompt Engineering Guide](#)


DAIR.AI([opens in a new tab](#))のプロジェクト

Prompt Engineeringについて、研究者や実践者に教育することを目的としています。

[Distributed AI Research Institute](#) [分散AI研究所](#)

米Googleを解雇されたとツイートしたAI倫理研究者のティムニット・ゲブル氏

3. LLM(大規模言語モデル)と生成AI



● Prompt Engineering

- Introduction
 - LLM設定
 - 基本的なプロンプティング
 - プロンプトの要素
 - プロンプトをデザインする一般的な Tips
 - プロンプトの例
- Techniques
 - Zero-shotプロンプティング
 - Few-shotプロンプティング
 - Chain-of-Thoughtプロンプティング
 - 自己整合性 (Self-Consistency)
 - 知識生成プロンプティング
 - Tree of Thoughts
 - Retrieval Augmented Generation
 - Automatic Reasoning and Tool-use
 - 自動プロンプトエンジニア
 - アクティブプロンプト
 - 方向性刺激プロンプティング
- Techniques
 - ReAct
 - マルチモーダル CoT
 - グラフプロンプト (GraphPrompts)
- Applications
 - PAL (プログラム支援言語モデル)
 - データ生成
 - Generating Code
 - Graduate Job Classification Case Study
 - Prompt Function
- Models
 - Flan
 - ChatGPT
 - LLaMA
- Models
 - GPT-4
 - Model Collection
- Risks & Misuses
 - 敵対的 Prompting
 - 事実性
 - バイアス
- Papers
- Tools
- Notebooks
- Datasets
- Additional Readings

3. LLM(大規模言語モデル)と生成AI

● Instruction Tuning (FLAN) (Wei+2022)

- 言語モデルの構成を変えずに、複数のタスクでファインチューニングする
 - タスク毎にテンプレートを用意し「プロンプト (タスクの指示と事例) + 出力」という形式の学習データに変換することで、言語モデルとして追加学習できる
 - 実験では自然言語推論、常識推論、感情分析、言い換え、質問応答、機械読解、共参照解析、要約、機械翻訳など、62個のデータセットを用いた
- タスクをゼロショットで解く実験において、GPT-3よりも高い精度を達成
- タスクの指示を自然言語で与えることが成功の鍵であることが示唆された

J Wei, M Bosma, V Y Zhao, K Guu, et. al. 2022. Finetuned Language Models are Zero-Shot Learners. ICLR.

● 思考の連鎖(Chain of Thought)

- 解答例をモデルに与えるときに「考え方」を含めるようにする
 - 言語モデルが単語を予測するときに、思考過程と解答を出力するようになる Wei+ (2022) ではデータセットに思考過程を手作業で付与
 - Kojima+ (2022) では"Let's think step by step" をプロンプトに含める手法を提案
 - 数学問題、常識推論、記号推論などのタスクの性能を大幅に改善

J Wei, X Wang, D Schuurmans, M Bosma, et. al. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. NeurIPS.

T Kojima, SS Gu, M Reid, Y Matsuo, Y Iwasawa. 2022. Large Language Models are Zero-Shot Reasoners. NeurIPS.

3. LLM(大規模言語モデル)と生成AI

● InstructGPT (Ouyang+ 2022)

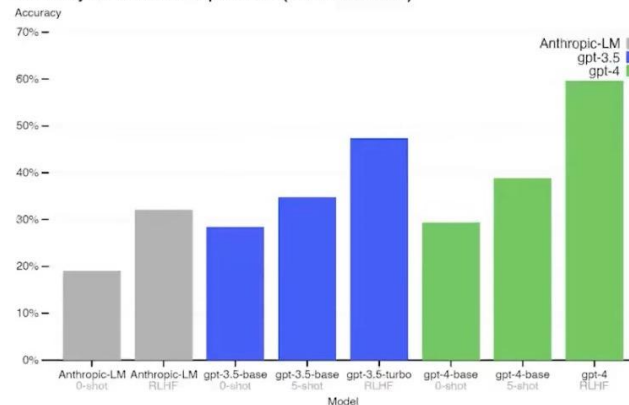
- ChatGPTの基となるモデル (ChatGPTの詳細に関する論文は未公開)
- 人間のフィードバックによる強化学習により、プロンプトで与えられた人間の指示に従い、詳細な応答を返すように調整されたモデル

[L Ouyang, J Wu, X Jiang, D Almeida, et. al. 2022. Training Language Models to Follow Instructions with Human Feedback.](#)

● TruthfulQAGPT-40 (OpenAI, 2023)

GPT-4 (base) モデルはGPT-3.5 (base) モデルよりも若干良い性能
GPT-4に対するRLHFはGPT-3.5 に対するRLHFよりも大きな性能向上
[OpenAI. 2023. GPT-4 Technical Report.](#)

Accuracy on adversarial questions (TruthfulQA mc1)



TruthfulQAにおけるGPT-4, GPT-3.5, Anthropic-LMの性能 (Lin+ 2022)

3. LLM(大規模言語モデル)と生成AI

● [大規模言語モデルにおけるプロンプトエンジニアリングの概説](#)

[小島 武\(東京大学\)](#)

● [Capabilities of GPT-4 on Medical Challenge Problems 12 Apr 2023](#)

大規模言語モデル(LLM)は、医療を含む様々な領域において、自然言語の理解と生成に顕著な能力を発揮
最先端のLLMであるGPT-4[Ope23]の医療能力試験とベンチマークデータセットに対する包括的な評価を発表
GPT-4は汎用のモデルであり、トレーニングによって医療問題に特化されたものではなく、臨床タスクを解決するために設計されたものでもない。

我々の分析は、米国で臨床能力を評価し免許を付与するために使用される3段階の試験プログラムである、米国医師免許試験(USMLE)の公式練習教材2セットを対象としている。

モデルの性能を測定するだけでなく、テキストと画像の両方を含む試験問題がモデルの性能に与える影響を調べたり、訓練中の内容の暗記を調べたり、確率の校正を調べたりする実験も行った

GPT-4は、特別なプロンプトを作成することなく、USMLEの合格点を20点以上上回り、以前の汎用モデル(GPT-3.5)よりも優れていることがわかった

GPT-4は、GPT-3.5やMed-PaLM(Flan-PaLM 540Bのプロンプト調整版)と同様に、USMLEの合格点を20点以上上回った。

GPT-4はGPT-3.5よりも大幅にキャリブレーションが改善されており、解答が正しい可能性を予測する能力が大幅に向上していることが示されている。

3. LLM(大規模言語モデル)と生成AI



また、GPT-4が次のようなケーススタディを行うことで、モデルの振る舞いを定性的に探ります。

GPT-4が医学的推論を説明し、学生への説明をパーソナライズし、次のような能力を持つことを示す。

また、医学的推論を説明し、学生への説明をパーソナライズし、医学的症例に関する新しい反事実シナリオをインタラクティブに作成する能力を示すケーススタディを提示することにより、モデルの動作を定性的に探求する。

調査結果の意味は、医学教育、評価、臨床実践におけるGPT-4の潜在的な使用について議論される、正確さと安全性の課題に適切に注意を払いながら。

3. LLM(大規模言語モデル)と生成AI

● 【AI】ChatGPT-4は日本の医師国家試験で優れたパフォーマンスを見せ合格水準に達した

1つ目の問題は、非英語圏でも同様なのである。

2つ目の問題は、医療文化や社会・政治的背景の差である。医学/医療においては日本独自のものもあり、ChatGPTがそれを学習していない可能性がある。

3つ目の問題として、実はChatGPTは論理的推論がやや苦手な部分がある。ChatGPTはチャット能力と正確かつ詳細な回答生成のバランスをとることで生成タスクのパフォーマンスが低下する可能性があり、論理的推論に関してはGoogle社の対話型AIであるBardの方が優れていたと報告されている。

● 日本の医師国家試験でGPT-4が合格、言語の壁と課題も明らかに

米ワシントン大、札幌心臓血管クリニック、東北大からなる研究チームが査読論文サイトで発表した内容によると、2018～2022年の5年分の医師国家試験を対象に、GPT-4は5年分すべてで合格ラインを上回ったが、妊婦の患者には投与できない薬剤を選んだり、患者に対して安楽死を促すような言葉がけを適切な対応として選択したりするなどの問題があった。これらの問題は、現行の大規模言語モデルAPIの制約が露呈していると指摘されている。

4. 創造性とは

Vygotskian Autotelic Artificial Intelligence

ヴィゴツキーのオートテリック人工知能

● 外言と内言

Jean Piaget 1896-1980 内部→外部

自己中心会話は(子供の)認知的発達の未熟成の証拠

発達はstagesに分かれている

独り言は正常な認知発達において全く積極的な役割を果たさない

幼児が自己中心的発話を行うのは、他人の視点が想像できないからだ

子どもが本当の意味での社会的相互関係をもてるようになると、独り言は次第に消える

Lev Vygotsky 1896-1934 外部→内部

自己中心会話は(子供の)社会的な会話を内面化したもの

社会・文化の影響はstagesと関係はない

知的発達の過程でもっとも重要な瞬間は、それまでは全く独立していた2本の発達ラインである発話と実際の行動が収束する際に起きる

初期の社会的コミュニケーションが独り言を引き起こす

社会的コミュニケーションが、人間独特の高度な認知プロセスを生じさせる

4. 創造性とは

生産モデルの4種類

1. Trajectory captioners

会話による思いつきの連続

2. Plan generators (ALFWorld)

プランを意識的に作る

3. Explanation generators

説明性による新たな発見

[Lampinen et al. "Tell me why! Explanations support learning relational and causal structure" \(ICML 2022\)](#)

4. Goal generators

目的を作る

IMAGINE

Intrinsic Motivations And Goal INvention for Exploration

探索のための内発的動機づけと目標発明

創造性 = 新規性 x 適切性 (Simonton, 2012)

言語的創造性とは、既知の文法や既知の構文 (適切性)から新しい発話を生み出すこと (新規性)である (Chomsky, 1957; Hoffmann, 2020)

4. 創造性とは

● 他者モデル

「他者モデル」を使いながら相手への説明性を構築します。もちろん初対面なら発話しながら、自分の持つ他者モデルのどれに近いかに推定します。当然「自己モデル」と類似性が高ければそちらを使うはずで、そこで、不足情報を補いながら『芸術作品が「美しい』』ということを伝えようとしています。

ある程度、一般的な話をするなら、共通語彙空間、ミーム、住んでいる地域の文化・教育をベースに外挿すると思います。譬え話とかアナロジーも駆使します。古くは、ユングの集合的無意識、最近は、コモングランドという概念で捉えているようです。

1. 内言化

思考の対象の語彙群が次々にキックされ、内言表現ではずっとこのチェーンがつながり、文が形成されている

2. 外言化

基本このチェーンにシンタックスチェックが走り、文としてシリアライズされ出力される

4. 創造性とは

● クリエータ・オーディエンス問題

オーディエンスは環境に対応し、クリエイータはアクター (実行部)とクリティック(評価部)に分かれる

- ①: クリエータ実行部は、美意識をもとに概念(アイデア)を選択し、コンテンツを作成
- ②: オーディエンスから評判及び評価を評価部が観測します。
- ③: クリエータ評価部は、得られた評判、評価を使って実行部の取った概念(アイデア)行動評価実行、クリエイータ実行部に通知、実行部は評価をもとに美意識の更新

● actor-critic architectures

強化学習の領域で、エージェントの方策 (Actor)と価値関数 (Critic)を同時に学習するこのアーキテクチャ

Williams, R. J., Baird, L. C. (1990). A mathematical analysis of actor-critic architectures for learning optimal controls through incremental dynamic programming.

4. 創造性とは

- そもそもはLisp: 自分自身を書き換えられる
- Gödel machine
Jürgen Schmidhuber(ユルゲンシュミットバー) Dalle Molle人工知能研究所
1997年、SchmidhuberとSepp Hochreiterは、LSTMIに関する論文を発表
2015年、これがGoogleのスマートフォン用ソフトウェアでの音声認識の新しい実装に使用
- Gödel Machine HP
Gödelマシンは理論的な自己の改善であるコンピュータプログラム、最適な方法で問題を解決
それは新しいコードをより最適な戦略を提供して証明することができたときに、自身のコードを書き換えた再帰自己改善プロトコルを使用
- AIXI (2000年)
生涯報酬を最大化する行動を取るエージェント
Unifying Foundations of Intelligent Agents Marcus Hutter
- Foundations of Universal Artificial Intelligence
 1. Ockhams' razor (simplicity) principle
 2. Epicurus' principle of multiple explanations
 3. Bayes' rule for conditional probabilities
 4. Turing's universal machine
 5. Kolmogorov's complexity
 6. Solomonoff's universal prior(=Ockham+Epicurus+Bayes+Turing)
 7. Bellman equations

4. 創造性とは

人類の文化的躍進のきっかけは、7万年前に起きた「脳の突然変異」だった: 研究結果

- 外側前頭前野に損傷がある場合、人は物と物の関係や、相対性を表す文章が理解できなくなる
例えば「犬は賢い」というシンプルな文章は理解できても、「犬は猫よりも賢い」となると、どちらが賢いのかわからなくなる
「円の上に三角を描く」「春は夏の前に来る」なども同様に、物事の上下関係や前後関係の理解がなくなってしまう
- 「前頭前野統合 (Prefrontal Synthesis)」または「メンタル統合 (Mental Synthesis)」
記憶のなかの複数の単語を意味のあるメンタルイメージとして合成するプロセス
- 言葉の再帰構造を習得できるのは子どものうちだけ
メンタル統合能力だけは、だいたい 5歳くらいまでの幼児期に再帰構造のある言葉に触れておかないと、大人になってもこれを習得することはできない
ヴィシェドスキーは、13歳までいっさい言語に触れることのなかった少女をはじめとした 10人の子どもの特異な例を挙げている
興味深いことに、10人全員が何年もの言語トレーニングを経たあとでも、英語の「in」「on」「at」などの空間的前置詞、動詞の時制、および文章の再帰構造を完全に理解することはなかった
社会的サポートがままならない途上国で、再帰構造のある手話に触れる機会がなかった聴覚障害のある子どもたちも同様
こういった子どもたちは、あとになって補聴器をつけたり徹底した言語療法を受けたりしても、「緑の箱を青い箱に入れる」などの簡単な指示をこなすことができないという

4. 創造性とは

- このような子どもたちに特徴的な行動

子どもたちは正しく緑色と青色の箱を持ち上げ、合っているかどうかのヒントを得るために、“実験者の顔色を見ながら”2つの箱を空間移動させ、トライ・アンド・エラーを繰り返す

- 「再帰言語」と「メンタル統合能力」の切っても切り離せない関係

こういった指示を頭のなかで想像して理解するには、5歳までに難しい再帰言語に晒されることで鍛えられるメンタル統合スキルが必須

これこそが、人類の「現代的行動」の基盤となる発想力や文化的創造力に大きく貢献した

洞窟壁画、住居の建設、副葬品を伴う埋葬、骨製の針などにみられる道具の専門化など、現生人類の想像力を彷彿とさせる「文化的創造性」は、7万年前よりも以前には発見されていない

この“文化的空白”の50万年間——現代的音声器官の発達と現代的想像力の獲得の間に横たわる長いギャップは、何十年も科学者たちを困惑させてきた

7万年前の想像力獲得に至った要因、そしてこのプロセスで言語が果たした役割は何だったのだろうか？

4. 創造性とは

[「Research Ideas and Outcomes」Andrey Vyshedskiy](#)

- **メンタル統合能力の重要性『再帰言語』**

再帰言語の自然発生は、1970年代から80年代にかけて、ニカラグアの聴覚障害がある子どもたちのなかで実際に観察されている。

ニカラグア手話は、かつて家庭内で必要最低限のジェスチャーでしかコミュニケーションがとれなかった子どもたちが施設に集められ、そのなかで独自に生み出された言語だ。

新たに発明された手話は年少者へと受け継がれて年月とともに複雑化し、ついには数世代で再帰構造を含む洗練された言語へと進化を遂げたのだ。

- **前頭前皮質遅延の突然変異とメンタル統合獲得が、ほぼ同時期に起こったことを予測**

ブラジルのアマゾン川のほとりには、ピダハン族と呼ばれる「再帰言語を話さない」狩猟採集民族が暮らしている。

この部族の言語と文化を研究してきたダニエル・エヴェレットによると、ピダハン語には数や左右、色彩、それに性差の概念がない。

そしてピダハン族は実際に見聞きしたことしか話さず、架空の話はしない。

反実仮想、数学的帰納法、ペアノの公理

遺伝子的変異、学習と環境、発達心理学的説明が必要

5. 研究分野、産業におけるAI、LLMの利用

● 最近のイベントからピックアップ

20231106-10—SWC2023

シンボリズムはどこまで迫れるか？

20230809—「潜在空間分子設計」シンポジウム

測定と推定 (AI) と合成

20230808—第24回汎用人工知能研究会

パネル企画: 人類の脅威となる AGI の作り方 (AI アライメント)

20230801—第8回全脳アーキテクチャシンポジウム

パネル討論: よき超知能時代に向かうためにヒト脳型 AGI が果たすべき役割

20230721—リザバーコンピューティング技術セミナー

20230718—「生成AIの倫理とプラットフォーム学」ゲームチェンジャーとなる技術と、それに必要な
ELSI・ガバナンス

5. 研究分野、産業におけるAI、LLMの利用

●最近のイベントからピックアップ(2)

20230717—アバター共生社会倫理コンソーシアム第4回シンポジウム

パネル:「アバター社会実装におけるガイドラインについて」すでに産業構造は変化している

20230719—ESWC2023サーベイ会

LLM利用の論文が多い

20230701—第15回オンライン物理講話演題: 2次元の物質グラフェンにまつわる10個のストーリー

20230524—alt-conference2023 ~AIの創造と破壊~ パネルディスカッション

動画03:11:02 パネルディスカッション.2『~AIの創造と破壊~』

ゲスト:京都大学大学院 理学研究科教授 橋本 幸士氏、生島 高裕、

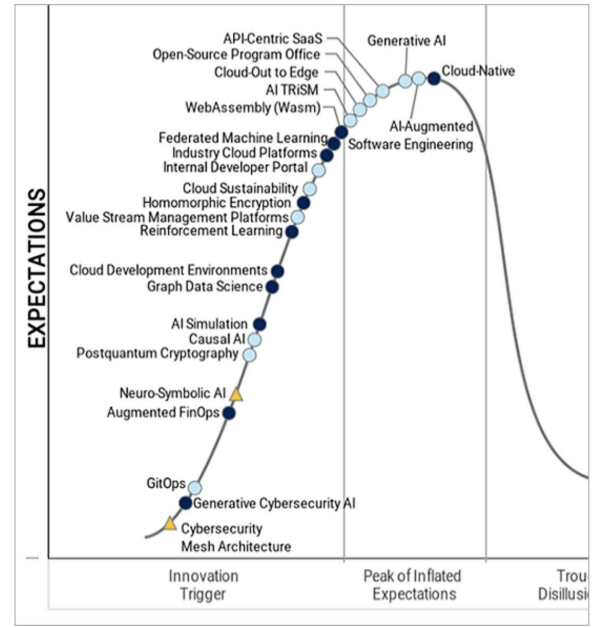
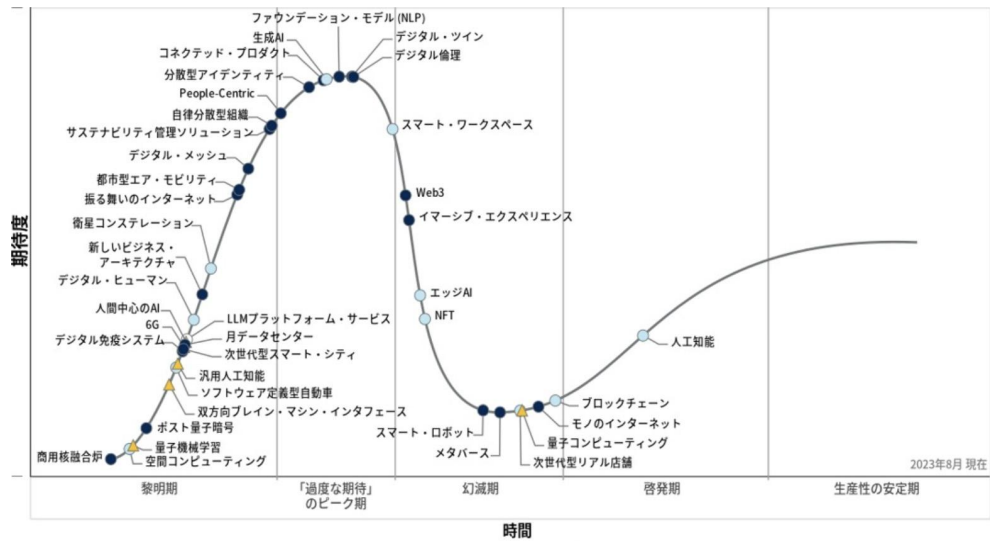
20230511—第38回 全脳アーキテクチャ勉強会~「神経科学における ChatGPT等の活用」

20230126—第37回 全脳アーキテクチャ勉強会~大脳皮質の回路とその役割の謎に迫る~

5. 研究分野、産業におけるAI、LLMの利用

● ハイプ・サイクル

「日本における未来志向型インフラ・テクノロジーのハイプ・サイクル: 2023年」
 米ガートナー「先進テクノロジーのハイプサイクル 2023年」



出典：米ガートナー（2023年8月）。画像の右側にある啓蒙活動期、生産性の安定期を省略

5. 研究分野、産業におけるAI、LLMの利用

● 自動化の流れ

2016—人工知能がノーベル賞を獲る日、そして人類の未来：究極のグランドチャレンジがもたらすもの
北野 宏明

ロボティックバイオロジーによる生命科学の加速 高橋 恒一

インタビュー「科学と技術の関係性が変化、AI駆動で目指す新たな価値創出」

AIロボット駆動科学シンポジウム 2023

・清水亮太 (東京大学)「マテリアルインフォマティクス・プロセスインフォマティクスのための AI駆動科学」

・二階堂愛 (東京医科歯科大学 理研BDR)「基盤モデルに向けたオミックス計測技術の未来」

GPT2、GPT3、の歴史

生命を1細胞レベルで理解するためのゲノム科学技術を開発する

<https://www.bdr.riken.jp/ja/research/labs/nikaido-i/index.html>

新しい1細胞オミックス計測技術の開発

新しい1細胞オミックスを用いた連携

5. 研究分野、産業におけるAI、LLMの利用

●自動化の流れ(2)

Autores 熊谷亘

目的: 研究成果の生産工場を作る(自己改善可能)

「入力: タスク=>出力: 論文」の生成を効率化するタスクを解く

上位システム: メタタスク、自己改善

下位システム: タスク、関連情報理解、研究計画、アルゴリズム生成、コード生成、論文作成、データベース化

自動定理証明: マークアップ言語の使用により困難な問題の証明成功率が向上

自動コーディング: プロンプトチューニングにより困難な問題のコード生成成功率が向上

6. 教育とLLM

●生成AIが広まらない？

ChatGPTへのアクセス、日本は世界3位 1位と2位は？

「アメリカ 生成AIの使用率」

生成AIの認知率が50%を超えるも仕事に活用しているのは3%、挨拶文の作成などが中心—NRI調査
今後は文書の要約やマニュアル・議事録の作成用途が拡大

企業の「ChatGPT」利用率に大差、日本7%、米国51%！ 日本経営層の関心、米国の半分以下 ... 利用も事務ばかり、米国はクリエイティブなのに

ChatGPTの利用率、日本とアメリカの差は●倍!! 生成AI先進国のアメリカから考える日本の今後
アメリカは半数...「チャットGPT」業務利用、日本の現状は？

「社説」日本が探るAI戦略の活路

人工知能(AI)の活用が急速に広がるなか、政府が国産技術開発への支援を強めている。出遅れた日本は独自の戦略を探り、活路を見いだす必要がある。

経済産業省や総務省、文部科学省などが国内の研究開発力の強化を重点政策に掲げる。2024年度の関連予算は政府全体で2000億円規模と、前年度の2倍ほどに増やす意向だ。

AIに関する暫定的な論点整理 2023年5月26日 AI戦略会議

6. 教育とLLM

● AIアライメント

Key open questions about the transition to AGI

AGI への移行に関する主な未解決の質問への回答

この調査の目的

4つの欠点を回避する

AGIの定義

Q1: AGIの代替的な定義は？

Q2: 近いですか、遠いですか？

Q3: 壊滅的か有益か？

Q4: 経済的な大混乱を予想していますか？

Q5: リスクとベネフィットの理解

Q6: 単一か多数か？

Q7: 限定か無制限か？

Q8: 自動的な道徳的調整？

Q9: AGIに関する自由市場での決定を奨励しますか？

Q10: 漸進的か先取りのか？

Q11: 独裁政治を排除した監視？

Q12: 価値観に関する世界的な合意？

Q13: AGIに道徳原理を組み込む？

Q14: カナリアシグナルに関する合意 Q15: 防水スイッチオフメカニズム？

Q16: 超知能を監視する超人？

Q17: AIsは、私たちにAIsのコントロールやアライメントの方法を教えてくださいか？

Q18: 有用な発見的制約は？

Q19: オープンソースやブロックチェーンの役割は？

Q20: 重複する他のディスラプションは何ですか？

Q21: 単一の中央集権的な開発？

Q22: 誰が始めるのですか？

Q23: ガバナンス・モデルのテンプレート

Q24: 制約の中でのレース？

Q25: 最も困った質問は？

6. 教育とLLM

● [How Do AI Timelines Affect Existential Risk?](#)

AIのタイムラインは存続リスクにどう影響するか？ August 2022

● 危惧する懸念は、作る側でなく、使う側の問題

1. 作る側と使う側の関係

作る側、大企業はそれなりの倫理を持って開発する

もし悪意を持って開発するならそれはマッドサイエンティストと呼ばれるため、その称号への拒絶反応があるはず

しかし、使う側の倫理は歴史を見ても、政府、企業で悪利用であったとしてもそれを正当化する論理が働く、ここを抑止できていない現実がある

[The power of corporate control in the global ownership network](#)

最も大きな問題は作る側は使う側から資金を得る場合が多いので使う側に強く言えない

AGIのリスク面ばかり強調すると研究費が集まらなくなるリスクがある

使う側の倫理を追求するとアインシュタインのような世界政府構想になるが、それは難しいと言うのが現状

6. 教育とLLM

2. 国家総動員でASI兵器を作る可能性

一旦、ASIの重要性に目覚めた、国家、為政者は相当開発に力を入れるでしょう。

核もバイオもナノテクもそう

歴史的な恐怖は、“元”の軍事力で“金”を征服した事実

全てをAI兵器にかける国が出てきたら、生活は貧しくても兵器開発に走り、領土侵略行う可能性

ウクライナ戦争でこのリスクが陽にした

経済ニュース番組などでは地政学的リスクは常にニュースになる

[世界が注目の“ドローン迎撃”企業を徹底取材～台湾企業は地政学リスクにどう向き合うか～](#)

Tron Future Tech: 低コストのドローン捕捉・迎撃システムを開発

3. 兵器のオープン化とコモディティ化

この問題は経済のグローバル化の1つとしての兵器問題だが、オープン戦略で市場規模の拡大、抱え込みを狙う

その結果、テクノロジーが拡散、クローズ戦略を取るとイノベーションから取り残されると言う強迫観念が存在するから

コモディティ化ではカラシニコフ銃が有名で、ASI兵器が安価で広く出回るとなるとその制御をどうすればいいのかは難問

7. 課題

● 微細化でどこまで見えてくるか？

シナプス間隙: 20nm

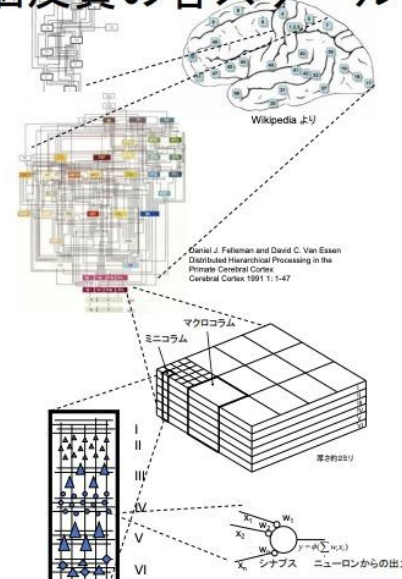
神経伝達物質受容体の大きさ: 7-70nm

軸索の直径: 数 μm -100 μm

軸索の長さ: 数 μm -1m

大脳皮質についての基礎知識 一杉裕志(産総研)

大脳皮質の各スケールでの構造



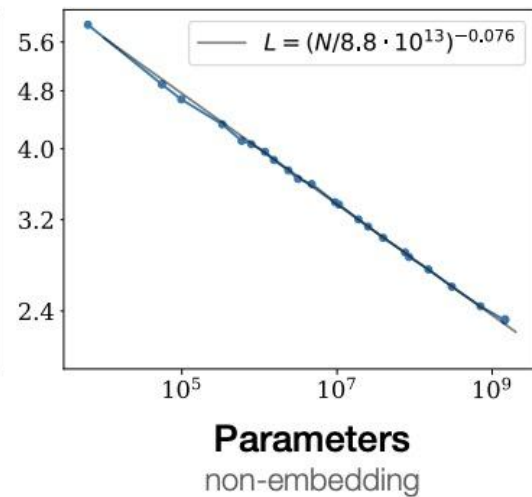
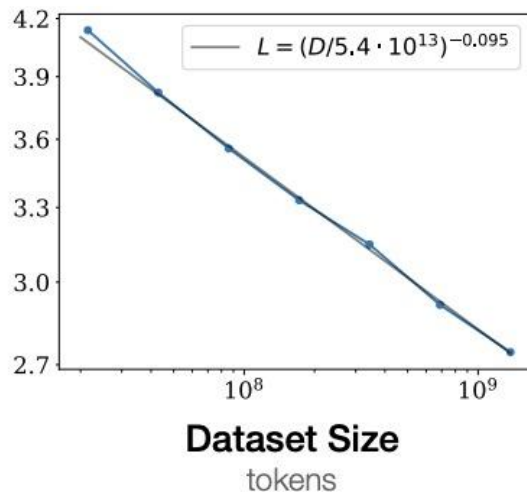
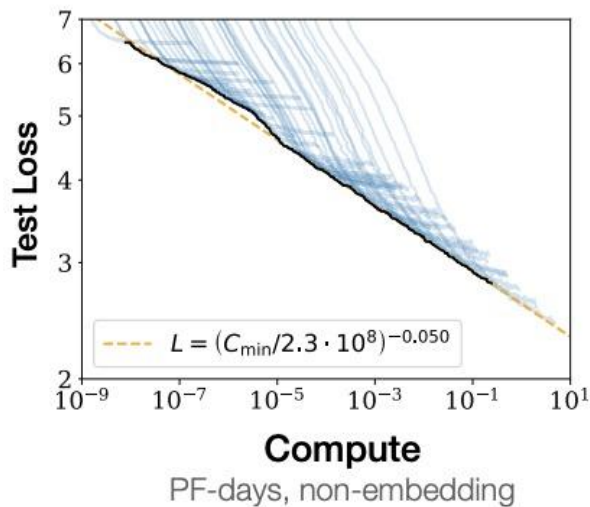
- 領野 約50個
- マクロコラム約100万個
- ミニコラム 約1億個
- ニューロン 約100億個
- シナプス 約100兆個

7. 課題

● パラメータと計算量、発熱量

[Scaling Laws for Neural Language Models 23 Jan 2020](#)

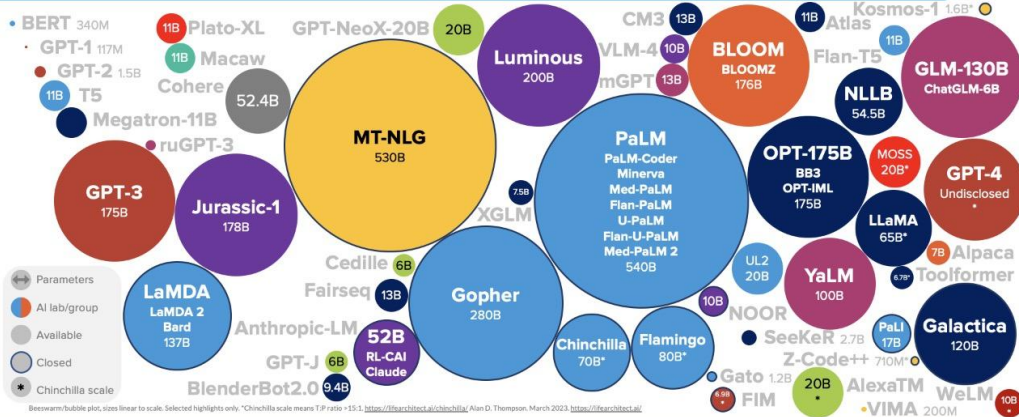
言語モデルの性能と計算能力、訓練データ量、パラメータ数の間のべき乗則 (Kaplan+ 2020)



7. 課題

Inside language models (from GPT-4 to PaLM)

LANGUAGE MODEL SIZES TO MAR/2023



BILLBOARD CHART FOR LANGUAGE MODELS JUN/2023

| Rank | New (Jun/2023) | 6m ago (Dec/2022) | 12m ago (Jun/2022) | ALScore | Model name Details | AI lab Openness |
|------|----------------|-------------------|--------------------|---------|--|--------------------|
| 1 | — | — | — | 14.9 | GPT-4 1T trained on 20T tokens* | OpenAI API |
| 2 | — | — | — | 3.7 | PaLM 2 340B trained on 3.6T tokens* | Google API |
| 3 | 1 | 1 | 1 | 2.2 | PaLM 1 540B trained on 0.8T tokens | Google Closed |
| 4 | — | — | — | 1.6 | Inflection-1 120B trained on 2T tokens* | Inflection AI API |
| 5 | — | — | — | 1.4 | InternLM 104B trained on 1.6T tokens | Shanghai AI Closed |
| 6 | 5 | 5 | 5 | 1.0 | Chinchilla 70B trained on 1.4T tokens | DeepMind Closed |
| 7 | — | — | — | 1.0 | StableLM 65B trained on 1.5T tokens | Stability AI Open |
| 8 | — | — | — | 1.0 | LLaMA-65B 65B trained on 1.4T tokens | Meta AI Open |

| Rank | New (Jun/2023) | 6m ago (Dec/2022) | 12m ago (Jun/2022) | Size (TB) | Dataset name Details | AI lab Language |
|------|----------------|-------------------|--------------------|-----------|---|------------------------|
| 1 | — | — | — | 86 | Piper monorepo 37.9B tokens in 86TB | Google Code |
| 2 | — | — | — | 40 | GPT-4 20T tokens in 40TB | OpenAI English |
| 3 | — | — | — | 23 | RefinedWeb 5T tokens in 23TB (CC only) | TII Multilingual |
| 4 | 1 | 1 | 1 | 20 | MassiveText multilingual 5T tokens in 20TB | DeepMind Multilingual |
| 5 | 2 | 2 | 2 | 13 | Infiniset 2.8T tokens in 12.6TB | Google English |
| 6 | 3 | 3 | 3 | 11 | MassiveText English 2.35T tokens in 10.5TB | DeepMind English |
| 7 | — | — | — | 5 | InternLM 1.6T tokens in 5.1TB | Shanghai AI En/Chinese |
| 8 | — | — | — | 5 | Stability New Pile 1.5T tokens in 5TB | Stability AI English |

LifeArchitect.ai/models

ALScore: 「ALScore」は、モデルのパワーを簡単に評価するものです。式は次のとおりです $\sqrt[3]{\text{Parameters} \times \text{Tokens}} \div 300$ 。2023 年半ばには、ALScore ≥ 1.0 は強力なモデルとなります。

LifeArchitect.ai/models

7. 課題

[Inside language models \(from GPT-4 to PaLM\)](#)

データセット サイズの増加 2018 ~ 2023 年

| # | Lab | Dataset | Size (TB) | Tokens (trillion) | Notes |
|---|----------|------------------|---------------|-------------------|---------------------------------------|
| 1 | Google | Piper monorepo | 86TB | <i>37.9T</i> | DIDACT, code only. From 2016 paper. |
| 2 | OpenAI | GPT-4 | <i>40TB</i> | <i>20T</i> | 1T model ∴ 20T tokens. gdb said 40TB. |
| 3 | TTI | RefinedWeb | <i>23.2TB</i> | 5.0T | CC-only dataset prepared by UAE. |
| 4 | DeepMind | MassiveText (ml) | <i>20TB</i> | 5.0T | From Retro paper. |
| 5 | Google | PaLM 2 | <i>13TB</i> | 3.6T | From PaLM 2 CNBC report. |
| 6 | Google | Infiniset | <i>12.6TB</i> | 2.8T | From LaMDA paper. |

Table. 2023 largest dataset estimates to Jun/2023. Rounded. Disclosed in **bold**. Determined in *italics*. For similar models, see my *What's in my AI* paper.

7. 課題

● パラメータと計算量、発熱量

[情報化社会の進展がエネルギー消費に与える影響 \(Vol.4\) LCS-FY2021-PP-01](#)

ーデータセンター消費電力低減のための技術の可能性検討ー

情報化社会の進展に伴う世界の情報量 (IPトラフィック) は2030年には現在の30倍以上、2050年には4,000倍に達すると予想

現在の技術のまま、全く省エネルギー対策がなされないと仮定すると、データセンターだけで 2030年には年間3,000TWh、2050年には400PWhという膨大な消費電力が予測

[1980年から2021年までの特定の年における世界の純電力消費量](#)

年間25,000TWh

Electricity consumption

[World electricity final consumption by sector, 1974-2019](#)

[Top500](#) も気になります。

2019

● Industry: **9 566 TWh**

● Transport: **420 TWh**

● Residential: **6 072 TWh**

● Commercial and public services: **4 849 TWh**

● Others: **1 940 TWh**

7. 課題

● LLMの限界についてパラメータ拡大の限界

「計算量」「データ量」「モデルパラメータ数」の相関:

「計算量」と「モデルパラメータ数」の関係:

GPTはパラメータ数の二乗に比例して計算量が増える
「計算量」と「データ量」の関係: ?

集合知データの整備:

wiki、辞書などの最適化、別のAIが行うようになるのか?

ハルシネーション(幻覚):

AIは事実を知っているわけではなく、ただ訓練データで見たパターンを再現するだけなのでパラメータ拡大で生じなくなるか? 人間も時々情報量が多いと起こす。

計算パワー:

地球規模、人類の消費電力でどれぐらいまで許されるのか?

全消費 > 産業消費 > サーバー消費 > I消費 > LLM消費

7. 課題

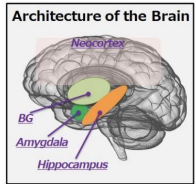
● 全脳アーキテクチャ・アプローチの紹介 Since 2014

脳全体のアーキテクチャに学び人間のようなAGIを創る(工学)

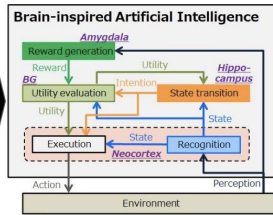
第4回 WBAレクチャー脳のコンポーネント図の作り方:プロセス間関係の整理と確率モデルによる記述

全脳アーキテクチャ・アプローチ Since 2014

脳全体のアーキテクチャに学び人間のようなAGIを創る(工学)



- ① 脳の各器官を機械学習モジュールとして開発
- ② それらを統合した認知アーキテクチャの構築



本アプローチの利点

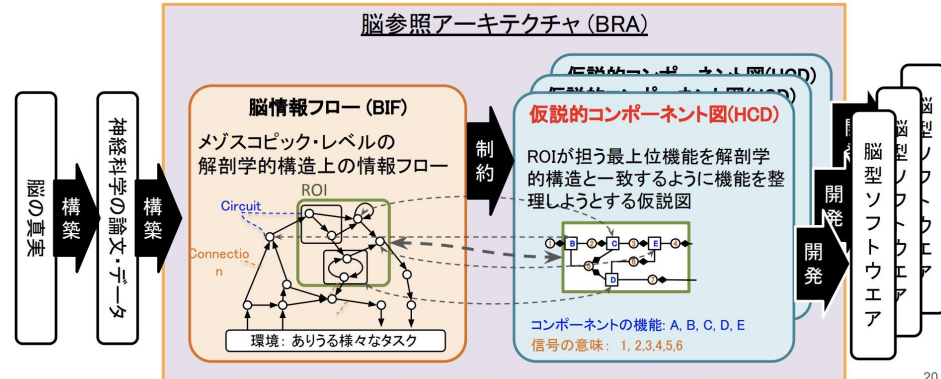
- 共同開発の基盤として脳の構造を利用して開発を加速
- 人間に関わる諸分野の科学知見を活用できる
 - 開発の発散を防げる合意しうる認知アーキテクチャ
 - 技術の統合を先送りせずに取り組める

- 人と親和性の高いAGIを作れる
- 対人インタラクションが必要な応用
 - 精神疾患などのモデル化など医療応用
 - マインド・アップローディングの器

他のアプローチによってAGIが完成しても依然として残る利点

構造: 解剖学に整合的な構造だが多様なコンポーネント図

一つのBIF



7. 課題

● 心の問題

心の哲学 wiki

哲学の一分科で、心、心的出来事、心の働き、心の性質、意識、およびそれらと物理的なものとの関係を研究する学問
心の哲学では様々なテーマが話し合われるが、最も基本的なテーマは心身問題、すなわち心と体の関係についての問題

心の理論 wiki

ヒトや類人猿などが、他者の心の状態、目的、意図、知識、信念、志向、疑念、推測などを推測する直観による心の機能のことである。

Premack, D. G., Woodruff, G. (1978).

Does the chimpanzee have a theory of mind?. Behavioral and Brain Sciences 1 (4): 515–526.

サリーとアン課題 誤信念課題(相手の心を読むこと) wiki

誤信念を他人に帰する社会的認知能力を測定するために発達心理学で使われる。

サイモン・バロン＝コーエン、Alan M. Leslie、Uta Frithにより先導的に実施された(1985)。

8. ICA、オミックス医学会、JASISの履歴、ETC

20150903—人工知能からの新しい産業社会環境の動向【ASIS2018】

20170415—DeepMind社周辺の動向調査【Do2Dle】

20170721—計算美学に向けて—AIとアート・創造性【計算美学】

20171113—脳研究と人工知能研究の急接近【ICA】

20171115—AIと意識・自我・倫理・哲学・悟り【神保町サロン】

20171128—沖縄で“参加型ヘルスケアを考える& AI【ICA】

20180907—AIの進歩による、IoT・知識ベースの未来予想【ASIS2018】

20181010—AIの進歩を考慮した人材養成の課題【CBI学会】

20190318—AI医療・AI創薬応用への挑戦『人工知能発展の展望』【オミックス医学会】


20190628—認知脳科学とAIは精神疾患の解明に貢献できるのか【オミックス医学会】

20191205—現代版 AI&脳科学からの無意識、意識、自我、超自我の考察【依頼講演会】

20200219—創造性と精神疾患は認知脳科学とAIによって統一的に理解できるか【オミックス医学会】

20200321—第6回自然言語処理テーマ:自然言語意味理解の現状【Do2Dle】

20200530—第8回自然言語処理テーマ:意味理解現状+メタ学習応用の現状【Do2Dle】



ご清聴ありがとうございました