

将来のAIは私たちの道具か、友人か、それとも支配者か？

経済学研究科
橋本文彦

アカデミックカフェ

人間は、なぜ「心」を持っているのか？

- 神様が人間にだけ与えた？
- 人間を構成するのと同じ物質・同じ構成・同じ機能であれば人間と同様の「心」が創発する可能性はある？
- ならば、「人工」のものであっても、それが心を持つ可能性を排除できない。
（「物質」「構成」「機能」は&条件か？）
- つまり「心」ひいては意思や意図をもって、その高い能力を他種の支配に用いる可能性も排除できない。

アカデミックカフェ

最近の話題

- 2024年度のノーベル物理学賞は、Artificial Neural Networkの J. Hopfield（Hopfield型ネットワーク）と、G. Hinton（Boltzmanマシンなど）が受賞
- Hintonは、高い知能を持つAIが人類に悪影響を与える可能性について言及している
 - ⇒（自由な発言のために）2023年にGoogleを退社

アカデミックカフェ

「人間を超える能力」の問題

- 自動車事故
- 原子力発電所の事故
- インターネットを利用した犯罪
- 高度な機械化による人間の失業



上記と下記に本質的な相違があるか、
それとも延長的な問題か

- AGI(ASI) による人間支配

アカデミックカフェ

Turing テストとEliza

- A. Turing (1950)

「機械が知性をもっているかを判断するテスト」

※あえて「知性」を言語で定義しない

- J. Weisenbaum (1966)

Elizaは、（部分的に）Turingテストに合格した

⇒人工無能（脳） ⇒⇒チャットボット ⇒⇒⇒生成AI？

アカデミックカフェ

ANN(Artificial Neural Net)

- もともと、人の神経回路における信号伝達を「模した」

⇒従来のノイマン型コンピュータよりも「知能」の創発に向いているとも考えられた。

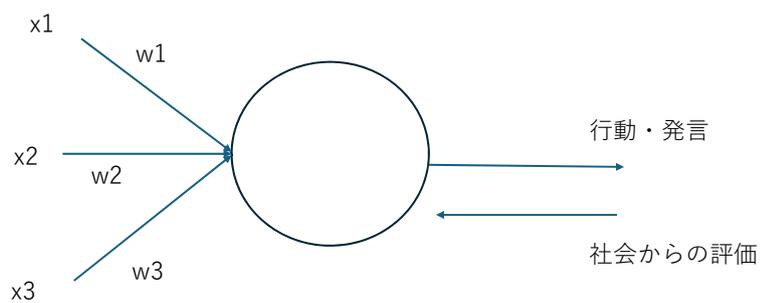
- 強いAIと弱いAI (J. Searleの「中国語の部屋」)

- 信号（情報）の入力と出力

アカデミックカフェ

Socionのアイデア

- Socion : Neuronの構造を模した「内面を持たない人間」



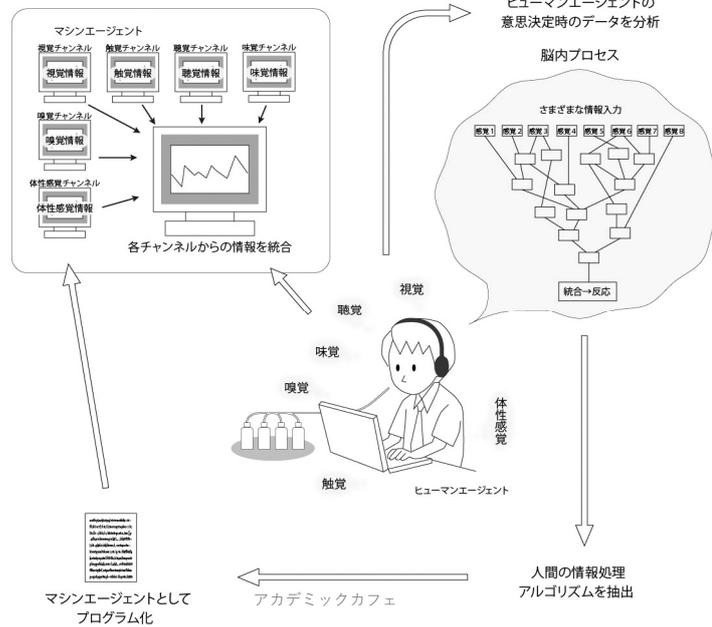
アカデミックカフェ

被験者実験からの示唆



アカデミックカフェ

プログラムによる形式化



人間に固有の時間

- 人間の「身体」を構成する「素材」が情報伝達する
 - 生物的・化学的な反応のために固有の時間が必要
- ⇒情報の入力から、出力に必要な時間が変わると意思決定内容や行動が変化する

U-Mart実験

- U-Mart（人工仮想先物取引市場）実験
 - = 人間エージェントと機械エージェントが取引行動することで相互作用する仮想先物市場
- U-Mart加速実験の結果
 - 機械エージェントは、（時間を変数に入れていないので当然ながら）振る舞いが変わらない
 - 人間エージェントは、読み取る数値の精度を下げたり、よりヒューリスティックスを聞かせた振る舞いをするようになる
 - ⇒ 必ずしも、利益の減少を引き起こすわけではない

アカデミックカフェ

逆さ眼鏡実験

- 人間の被験者4名が上下反転（2名）と左右反転（2名）を装着して60日間生活した。
- 身体の順応過程を調べた
- fMRIによって、脳の変化を追いかけた
 - ⇒
 - 長い時間の中で、徐々に順応していく
 - 順応中の身体活動が大きい方が順応しやすい
 - 近距離を見ると、眼球の輻輳角が合わない、などの問題あり
 - 重心移動の順応にはさらに時間がかかる

アカデミックカフェ

逆さ眼鏡と人工知能

- 例えば、左右逆さ眼鏡では、変化を検知してX軸方向を逆転するだけでよい
- 順応過程を再現するには、そのプログラムを構成して遅延させるだけでなく、順応途上での反応を「記述」する必要がある。
- さらに「身体活動」との関係性を再現するためには、そのためのプログラムも必要か？

⇒人工「知能」にとって、上述は必要か？

アカデミックカフェ

まとめ 1

- 能力が特化された「弱いAI」は、その部分において人間の能力をはるかに超えている。
- (人間と同様に) 汎用的な能力を持つ「強いAI」、あるいはAGIが、多くの点において人間の能力を超えることは十分考えられる。
- AGIが真に「汎用」か否かは、そこに「心」があるかどうかではないか？

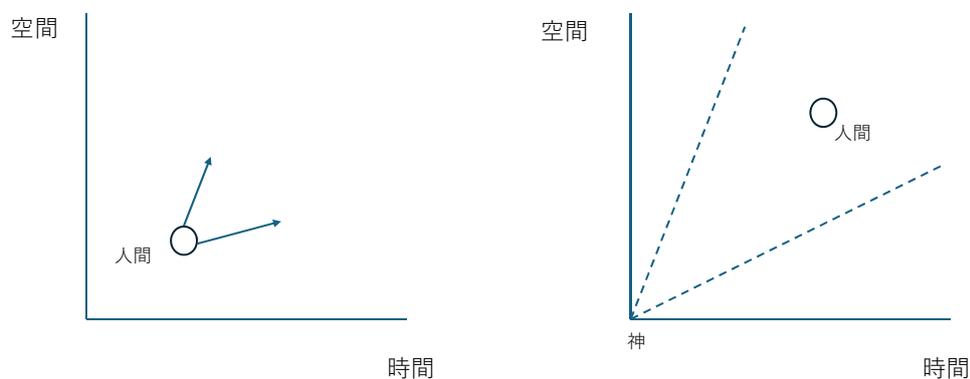
アカデミックカフェ

まとめ 2

- 人間が情報を処理するには、身体を構成する素材に特有の反応時間が必要である。
⇒与えられた制限時間によって意思決定や反応が変わる
- プログラミングによる機械エージェントは、（人間の反応時間オーダー程度での）制限時間の変化では行動が変容しない。
- 機械は、それなりの「心」を持つかもしれないが、人間と「心」が通い合うことはない。

アカデミックカフェ

まとめの後の余談



人間のパースペクティブと神のパースペクティブ

アカデミックカフェ