2024.10.30 開催
第 11 回 ACADEMIC CAFE
「AI と人の知性」

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　教授　橋本　文彦

概要　人間の脳という機械に「心」が宿っているならば、脳とは異なる機械にも「心」は宿り、記憶力や処理速度において人間の脳の能力を凌駕する可能性を否定できるでしょうか？人工知能へのアプローチには、「部分的にでも人間よりも優れた能力」を目指すものと、「まず人間と同様の総合能力」を目指すものとがあります。ここでは、人間の脳や身体を構成する物質の特性や変化時間に着目して、人工知能と人間の未来の関係を探ります。

キーワード　人工知能、心、身体、素材、形式化、時間

座談会の様子

1.　心を持つということ

人間は、なぜ「心」を持っているのでしょうか？この問いに対して、「神様が人間にだけ心を与えた」という答えは、ここでは考えないこととします。そうすると、人間を構成するのと同じ物質・同じ構造・同じ機能であれば、人工物であったとしても、そこには人間と同様の「心」が創発してもおかしくありません。そして、ホモ・サピエンスが、その賢さによって他の種を支配したなら、より賢い（人工の）種はホモ・サピエンスを支配することもありうるのでしょうか？

先日（2024 年 12 月）、ノーベル物理学賞を受賞した 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ます。

もちろん、「高度な技術」の登場によって人間が制御しきれないという事態はこれまでの歴史でも何度も出てきました。

AGI（汎用能力を持つ AI）による人間支配の懸念はこれまでの歴史と同じなのでしょうか、それとも全く異なるのでしょうか。

1950 年に Turing は、機械とコミュニケーションして、人間のそれと差がないなら「知性がある」と言える、というテストを考えましたが、その後すぐに登場した Weisenbaum のイライザというプログラムは、相手の言葉の一部を取り出して問い返すだけの簡単なプログラムでしたが、Turing テストを通過しました。ボットや生成 AI はその進化形態に過ぎないのでしょうか？

このことを考えるために、次に「内面」に関わる問題を考えます。

2.　プログラムによる形式化と人間に固有の時間

ANN（人工ニューラルネット）は、もともと人の神経回路における信号伝達を「模した」ものです。「人を模した」ことで、ANN は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ました。

他方で、Searle は「中国語の部屋」という思考実験を提案しました。これは、とある窓口で顧客から漢字の書かれた指示を受け取った人が、マニュアルに従ってその「絵（＝漢字）」の並びに対応した別の「絵」の並びを返します。外の人からは、この窓口の担当者は中国語を理解しているように見えるけれど、実際には「中国語を理解してはいない」という主張です。つまり、プログラムという「形式」に従って処理を行われているが、

「内面」としての理解はしていない、という主張です。

私自身は、人間を、社会を構成する一つの素子（ソシオン＝1990 年頃に関西大学の藤沢・木村・雨宮らによって最初に提唱されました）になぞらえて、むしろそもそもソシオンは「内面」をもたず、外部からの情報に重みづけ（信頼）をする（だけで）、外の社会に対して発信・行動し、社会から高い評価を得たら重みづけを拡張方向に維持し、失敗したら重みづけを修正する、というモデルを考えました。つまり、そもそも Serle の主張する「内面」の存在を前提とする必要はない、というものです。

3. 被験者実験からわかったこと

他方で、私は人間を被験者として、さまざまなマルチモーダル情報を提示し、その処理の結果としての反応・行動を測定するという実験心理学的な研究を行ってきました。

その研究から二つほど例を示します。

１）U-Mart（人工仮想先物取引市場）実験

これは、人間エージェントと機械エージェントが一緒に（人工）市場に入って相互に作用しながら取引行動するという仮想先物市場です。この実験のうち、特に「加速実験（＝通常時よりも短い時間で次々と取引の締切がくる場面実験）」を行った結果として、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のことながら）通常時間と加速時間とでその振る舞いは変わりませんが、人間エージェントは、加速時に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を効かせたりといったふるまいをするようになりました。なお、この場合でも必ずしも利益の減少を引き起こすわけではありませんでした。

２）逆さ眼鏡実験

これは、人間の被験者４名が上下反転（２名）と左右反転（２名）眼鏡を常時装着して、６０日間生活するというもので、その間の身体的順応を観察・測定し、また fMRI を用いて脳の変化を追いかけたというものでした。

彼らは、長い時間の中で徐々に順応していく様子を見せました。また、順応中の身体活動が大きい被験者の方が順応しやすいことも示されました。この実験中、左右反転眼鏡着用時には、近距離の対象を見た時に、眼球の輻輳角があわない、などの問題も見られました。また、自転車に乗ってカーブを曲がる際の重心移動の順応にはさらに時間がかかる様子が示されました。

4. 人間に固有の時間と心

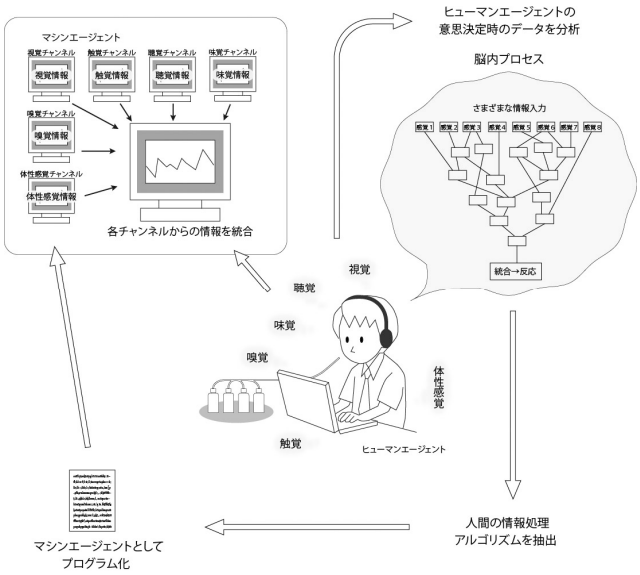
上記の二つの実験を含む被験者実験からわかってきたことは、「人間には固有の時間があり、それが人間の心にとって重要である」というものです。

人間の身体内では、この身体を構成する素材が情報を伝達するわけですが、この素材の生物的・化学的な反応のためには固有の時間が必要です。そして、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します。

例えば、左右逆さ眼鏡では人工知能であれば、変化を検知して X 軸方向を逆転するだけで「瞬時に」順応することができます。

人間の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・行動の変化までつぎつぎと「記述」する必要が出てくることになります。

図１　プログラムによる形式化と人間



5. まとめ

能力が特化された「弱い AI」は、その特化された部分において、すでに人間の能力をはるかに超えています。また、（人間と同様に）汎用的な能力を持ち「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も十分に考えられます。

しかし、AGI が真に「汎用」か否かは、そこに「心」があるかどうかにかかっているのではないでしょうか？

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り、与えられた制限時間によって意思決定や反応が変わります。

他方で、プログラムによる機械エージェントはもともと時間軸そのものが入っておらず、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ません。

このことから、本稿での結論は、機械はそれなりに「心」を持つかもしれないけれど、その「心」が人間と通い合うことはないだろう、ということになります。

6. まとめのあとの余談

ここまで述べてきた通り、人間は自らを置かれた空間と時間の制約の中でのみ世界を観察できるパースペクティブを持ちます。

一方、神は、すべての空間と時間の制約から解放されており、世界の全体を見通すパースペクティブを持っていると考えられます。

時間軸を持たず形式化された AI は、ある意味で神の視点を持っているかもしれません。また、プログラム部分のみに着目すれば、空間的にも制約を持たないとも言えます。

参考文献

[1] Turing, A., 1950, “Computing Machinery and Intelligence “, *Mind*, 49, 433-460
[2] Searle, J., 1980, “Minds, Brains and Programs”, Behabioral and Brain Sciences, 3, 417-456

[3] 橋本文彦、1994, 「機械は心をもちうるのか」, 心理学評論, 34, 533-554.
[4] 藤澤等（監）, 2006, ソシオン理論入門, 北大路書房
[5] 橋本文彦, 2012, 「機械の身体と人間の身体、機械の心と人間の心」, 思索, 45(2), 207-232

発表者紹介

橋本文彦　1963 年生まれ
東北大学文学部哲学科、大阪市立大学大学院文学研究科哲学専攻修了。博士（経済学）。
専門は科学哲学・数学基礎論から行動心理学・経済情報論のほか、古代ギリシャ哲学や医療統計学まで。