

次世代人工知能の構想 I

令和 1 年 9 月

公益財団法人 未来工学研究所

目 次

1. 概要	1
1.1 背景と目的	1
1.2 調査研究の概要	1
2. 第3世代の発展型の事例	5
2.1 人間中心型への接近	8
2.2 根拠説明型等の方法論的進展	15
2.3 「拡張機械学習型」	18
3. 脳科学の知見を反映した事例	21
3.1 脳科学の進展と脳機能の理解	21
3.2 脳と外部機器との接続および信号入出力 (BMI/BCI)	25
3.3 BMI/BCI による脳の活性化・強化に関する研究例	28
3.4 次世代に向けた考察	31
4. 脳科学の知見を反映した AI 模擬事例	35
4.1 AI チップ (ハードウェア) の現状	35
4.2 神経伝達、脳の機能と情動の脳内メカニズム	37
4.3 次世代型 AI チップ (ハードウェア) へのパラダイム変換 (「人間活動システム」の模擬) へのアプローチ	41
5. 脳科学の興味深い研究事例	47
5.1 「昆虫の脳」を AI に活用する (MBRAIN)	47
5.2 「3 人」の脳内ネットワーク共有 (BRAINNET)	47
5.3 「脳が量子コンピュータとして働いている」という仮説検証 (BIOCHEMICAL QUBITS)	48
6. 「パーソナル AI」を指向した事例	49
6.1 主導哲学転換型の事例	49
6.2 クリエイターに芽生えた先見的事例	65
6.3 映画『空気人形』: 人工知能と「心」・「意識」について	83
7. 今後の展望	105
7.1 現時点でのまとめ	105
7.2 次期計画「次世代人工知能の構想Ⅱ」の構想	106

1. 概要

1.1 背景と目的

1.1.1 背景

人工知能（AI）が万能であるかのような言説が一時期メディアを覆っていた。特に、ディープラーニング（以下 DL）の手法が驚異的な成果を挙げ始めた 2010 年代の中頃以降、物体認識の精度向上に伴い、「機械が眼を獲得した」という比喩的表現を真に受けて、まさに人工知能ブームが再燃した。このような AI の社会的認識状況に関し、我々は本調査研究に先立ち、認識論を基盤とした知識論の枠組みを用いて、2017 年の時点で我が国の実態を分析した。対象としては日本経済新聞本紙掲載記事、人工知能学会誌論説、Preferred Networks 社開示の実施事例、それに政府の取り組み事例である¹。

本調査研究は、この実態的調査結果を踏まえ、DL を中心とする機械学習によるアプローチでは扱いがたい領域を想定し、その領域に踏み込むための次世代型人工知能とはどのようなものであるかについて検討を加えることにした。

1.1.2 目的

本調査研究では、次世代の人工知能、つまり現在の第 3 次人工知能ブームの後に再び人工知能ブームの惹起が期待される次世代（第 4 次）型人工知能のパラダイムを構想し、その実態を具体的に探ることを目的とする。

1.2 調査研究の概要

1.2.1 調査実施者と概要

(1) 調査実施者（カッコ内は執筆箇所を示す）

非常勤上席研究員： 平澤 洽(1、7)

常勤主席研究員： 林 隆臣(6.2)、中崎孝一(6.1)、依田達郎(6.3)

常勤研究員： 山本智史

研究参与： 宮下 永、宮林正恭、太田与洋(2)、橋本 健(3)、藤江健介、大川晋司

非常勤主席研究員： 関 直康(5)、依田 孝(4)

(2) 概要

人工知能研究開発の歴史を概観し、次期パラダイムが何であるかについて認識論を踏まえた知識論の立場から構想する。DL を先鋒とする機械学習の限界が見えてきたことを踏まえ、「拡張機械学習型」の他に次期人工知能のパラダイムの可能性を第 3 世代型手法では近寄り難いヒトの「内界」に置き「人間活動システム」を模擬できる新たな AI の可能性を追究する。

¹ 公益財団法人未来工学研究所、ディープラーニング（深層学習）を中心とする第 3 世代型人工知能の限界を探る調査研究、一般社団法人新技術振興渡辺記念会「科学技術調査研究助成」（2018）

1.2.2 本調査研究の背景的枠組み

(1) システム論に基づく対象システムの区分

まず AI を適用しようとする対象知識領域をシステム論の立場から区分する²。個別システム領域の誕生を宇宙の歴史に沿って考えてみよう。

宇宙の起源と生物の進化過程を経てヒトが登場するまで世界は「自然システム」のみで構成されていた。ヒトは自己意識を抱き個別の価値観や情動に促され自身の存在に基づき「人間活動システム」を構成し、また「自然システム」に手を加え「人工的物理システム」を創出した。そしてこれら3システムに対する思考世界「人工的抽象システム」を想起する。さらにこれら知識に根ざしたシステムには依拠しない「超越的システム」を創造することもあった。

(2) 対象システムの知識論的特徴

各システムの知識はそれぞれ次表にある特徴を有している。

知識カテゴリー	構成システム	事例的説明	特徴
ハード1	自然システム、人工的物理システム	自然や機械	対象に普遍的に内在する 法則 がある
ハード2	人間活動システム	意思や価値観に従う人間を含む	内在する法則はない。対象により異なり時と共に移ろい行く。 経験的原理 の探索
ソフト1	人工的抽象システム1	自然や機械を外部から模したモデル	シミュレーション結果を照合する実体に 法則性 がある
ソフト2	人工的抽象システム2	人の行動や組織を擬似的に外部から模したモデル	対象とする実体は不均一で不定であるが、擬似的な法則性をたてて模した 論理 モデルを代用する。
ソフト3	人工的抽象システム3	意思や価値観を含む内面を模したモデル	普遍的、確定的に模したモデルとして把握することは困難。 経験的原理 の探索
その他	超越的システム	上記の知識カテゴリーや構成システムに分類できない超越的概念	

知識カテゴリーに注目して前期の事例群を分析する。

- ・2017年日本経済新聞記事からAI等のキーワードを含む804件
- ・2017年人工知能学会誌所載論文論説121篇
- ・Preferred Networks社のホームページに開示された実施事例71件
- ・政府の取り組み事例であるNEDO採択テーマ57件

PFNとNEDOの実施事例はいずれも、「ハード1」ないし「ソフト1」に属する対象を扱っていて、学会誌ではさらに「ソフト2」まで対象を広げているが「ハード2」や「ソフト3」（人文学の課題）の領域には明示的に進入を避けている。これに対して、報道記事では全ての知識カテゴリーを扱っていて、DLやAIの現状に対して、原理的に過大な期待を寄せている様子が読み取れる。

² Peter Checkland *Systems Thinking, Systems Practice* pp.111-112(1981) John Wiley & Sons. 近年合成生物が登場し、「人工的物理・生物システム」あるいは「人工的システム」と修正すべきか。

1.2.3 次世代型へのアプローチ

科学哲学を基底とする知識論³の枠組みでは、DLを先鋒とする機械学習の限界が見えてきたことを踏まえ、ここで認識された枠組みの中で、その後実務的な適用を具体的に拡大深化していく方途がとられている⁴。一方で、次世代の人工知能、つまり第4次人工知能ブームの惹起が期待される次世代型の人工知能を構想する道も残されている。次世代型へのアプローチは、まず「拡張機械学習型」とでも呼ぶべきDLの限界を突破する試みのほかに、前段で確認したパラダイムとは異なる新たなパラダイムの設定を目指す方途も試みた。

(1) トランスフォーマティブ・アプローチ

(a) 現世代の境界領域を越える

論理的枠組み

DLの手法では「ハード2」や「ソフト3」の知識領域をまだ本格的に扱っていない。DLをパラダイムとするディシプリンの境界は「ハード2」ないし「ソフト3」との知識領域の境にある。何らかの工夫によりDLの手法を「内界」たとえば「こころ」の実体把握や理解に役立てる方策を生み出すことが出来ないであろうか。既存のディシプリンの境界を超えてその向こうに新たな別のディシプリンを創設する。これがトランスフォーマティブ・アプローチである。

事例的分析

事例としては、NHKの「人間ってナンだ？超AI入門」⁵のシーズン2と3合計24回分を対象にした。興味深い事例として、以下の試みがあった。

①3年分のチャット等の個人データを集積してデジタルクローンを作り、呼びかけに対する意思決定をまかせる。

②ゲーマーの特性を学習させ、当該ゲーマー向きのゲーム展開を制御するAIを作成。ゲーマーの闘争心、満足感・達成感等を織り交ぜプレイ時間の延伸を図る。

③50以上のセンサーと10以上のCPUを組み込んだ癒し系ロボット、動く縫いぐるみ風Lovot。スキンシップ、距離感等の特徴を記憶し1000人以上の個人認識が可能。

これらはいずれも「ソフト3」の領域を扱おうとしている。しかし、まだ完全ではない。

(2) パラダイム転換を図る

³ 戸田山和久「知識の哲学」産業図書（2002）

⁴ たとえば、一般財団法人マルチメディア振興センター（編）「世界のAI戦略」明石書店（2020）

⁵ シーズン2は2019年1月10日から12回でAIの製品開発者とMCの対談、シーズン3は続く4月3日からの12回で開発者を現地に訪ね哲学ナビゲーターの小説家を交えた鼎談。

(a) 未開の対象領域への接近

論理的位置づけ

ヒトの内面の情報処理機能を担う器官として最も重要な脳の具体的な機序はどこまで明らかになっているのであろうか。DL とは異なるアプローチで新領域のパラダイムの実体にせまる。

脳科学の知見の収集と総括

2011 年以降の文献・資料を中心に調査した。最近の脳科学の進展を概観し、特に脳機能の強化・拡張の可能性に注目した。また、脳の機能はどこまで模擬できているか。ヒト以外の生物の脳の模擬はどうか。

(b) 依拠する哲学の転換⁶

論理的枠組み

科学哲学の認識論は、対象を外部から観察し、対象に内在している法則性を見出すことによって、対象を識別する。この認識方法では対象に内在している普遍性は把握できても、個と状況により異なり変転する場合知識として確定することは出来ない。ヒトの「内界」の仕組みに相当する「ソフト 3」の場合、従って普遍的な機能を担う A I として実現することは困難である。個に寄り添い「パーソナル」な AI となるであろう。

「パーソナル AI」を指向した事例

新たな哲学による新パラダイムを詳細に論じた後に、その枠組みの下でアプローチを変えて、クリエイターが創造する次世代の人工知能のあり方を示す事例について検討を加えた。意識、価値判断、感情、こころ等をもったロボットが主役となる作品の壮大な歴史的変遷を分析した後、ロボットと人間が織り成す 2 作品を分析した。演劇の「R. U. R」と、2009 年に公開された映画「空気人形」に関してそれぞれ主要なシーンごとに主題を分析している。こころを持つことになった人形（AI）の他者との関係に関心が向けられている。

⁶ 西垣 通「A I 原論」講談社（2018）

2. 第3世代の発展型の事例

ヒトの「内界」をAIは直接、理解することはできない。しかし、「内界」を疑似的に理解しているようなアウトプットは出せるようになっている。

「きしゃのきしゃはきしゃできしゃした」はワープロ専用機が世の中に出始めた1980年代に、漢字変換能力のテストとして、ユーザーが面白がって試していたものである。「貴社の記者が汽車で帰社した。」と適切に変換することは当初できなかった。現在、私のPCのIME一括変換すると「汽車の記者は汽車で帰社した」と変換され意味をなさない。グーグル翻訳では「Kishanokishagakishadekishashita」が一発で「貴社の記者が汽車で帰社した。」と漢字変換される。これを英語翻訳すると最初に、「Your reporter came back by train」と表示され2番目候補として「Your reporter was return to the office by train」と表示される。グーグルのAIが、文脈を完全に理解していることは無いが、適切な漢字変換ができ、ほぼ適切な英語表現になっている。翻訳についても、AIが文脈を理解してということはあるが得ないが、あたかも疑似的に文脈を理解しているように見える翻訳文を作り出している。これまでの機械翻訳研究の成果を取り込み、統計的機械翻訳やニューラル機械翻訳の進化に負っている⁷。

このグーグル翻訳の機能は、ジョン・サールが1980年に提案した『思考実験「中国語の部屋」で提起した「意味」を理解することなく会話を成立させること』が、今実現していると言える。中国語が理解できない英国人に、沢山の中国語のカードが入った箱と、そのカードの使い方が書かれた分厚い英語のマニュアルを持って部屋に入ってもらい。部屋には小さな穴が開いていて、そこから英国人は中国語で書かれた質問を受け取る。そして英語のマニュアルに従って、決められた中国語のカードを返す。その英国人は中国語の質問と返答の「意味」がわからないにも関わらず、中国語によるコミュニケーションを成立させており、外部の人からは中国語を理解しているかのように見える⁸。この思考実験全体はAIのアナロジーになっている。すなわち小部屋全体がAIを表し、マニュアルに従って作業する英国人は、プログラムに従って動くAIハードウェアに相当する。コンピュータが「計算」することと、「意味」を「理解」することとは違うとする議論である。これに対して、「理解」や「意識」の基準となっている人間の場合でさえ、脳内でいかにして意識が生じるかという仕組みが明らかになっていないのだから、中国語の部屋も、中身がどうであれ正しく中国語のやり取りができていた時点で、中国語を理解していると判断して良いのでは、という反論も存在する。その英国人は、中国語を理解していないにもかかわらず、傍から見れば中国語のネイティブのようなコミュニケーションができていると言えるというわけである。しかし、ここで重要なことは、そのやりとりには意味論が欠けているということである。その文字列はプログラムによる「出力(output)」に過ぎないとみる。「意味」とはあくまで人間が読み取るものなので、「意味」の「理解」こそが「こころ」であると定義するなら、現段階のAIでの「理解」には「こころ」が関与しておらず、疑似的な「理解」とであると区別しておく必要がある。

⁷ 例えば、「機械翻訳の新しいパラダイム、ニューラル機械翻訳の原理」中澤敏明、情報管理、2017, vol 60,299

⁸ 心の哲学まとめ Wiki, https://www21.atwiki.jp/p_mind/pages/22.html

この「理解」と「こころ」のギャップを発生させる現象は、ロボット工学者の森政弘が 1970 年に提唱した「不気味の谷」⁹と予言した現象と通じるところがあり、それが今日的に観察されていると考えられる。森は、人間のロボットに対する感情的反応について、ロボットがその外観や動作において、より人間らしく作られるようになるにつれ、より好感的、共感的になっていくが、ある時点で突然強い嫌悪感に変わると予想した。人間の外観や動作と見分けがつかなくなるとふたたびより強い好感に転じ、人間と同じような親近感を覚えるようになる考えた。外見と動作が「人間にきわめて近い」ロボットと「人間とまったく同じ」ロボットは、見る者の感情的反応に差がでるだろうと予想できる。この二つの感情的反応の差をグラフ化した際に現れる強い嫌悪感を表す谷を「不気味の谷」と呼ぶ。「人間に近い」ロボットは、しかし、ヒトの「心」と「意識」を持たず、人間にとってひどく「奇妙」に感じられ、親近感を持てないだろうことから名付けられた。

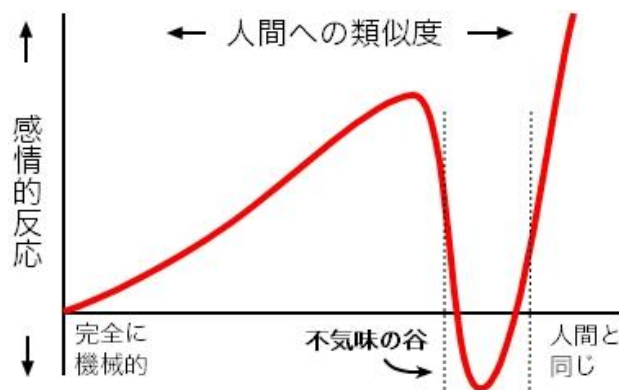


図 2-1 不気味の谷 uncanny valley (wikipedia)

これまで AI を活用した翻訳の機能を中心に論じてきたが、現在の翻訳技術水準は、ジョン・サールが思考実験として想定した「中国語の部屋」に、グーグル翻訳では（良く使用される言語間では）その水準に達しているといえる。これは、森政弘の「人間への類似度と感情的反応曲線」においてほぼ最高点に到達しているということも言える。まさに、AI が出力するものの「意識」と「こころ」の有り方について議論ができる段階に達してきていると言える。

そこで、人間の「こころ」を対象とする AI の実例について現状を概観することにする。AI は直接ヒトの内界を対象とすることはできないので、例えば、「拡張機能」により、『ヒトの心理や行動を外部から眺める人工的な抽象システムの「ソフト 2」の領域』から、『ヒトの感情、意志や価値観という内界を対象とする人間活動システムの「ソフト 3」』へ疑似的に拡張ができていくかのようなアプローチで、ヒトの「内界」に迫る。内界のうち、具体的には、ここでは「こころ」の動きに注目する。ヒトは喜怒哀楽を感じ・受け止め・表出し、他の人に共感したり反発したり、いい仕事をしようと張り切ったり、ゲームに興じることもある。ヒトは全く論理的でない判断をすることもあるし、理屈では理解できない行動をとることもある。

⁹ 不気味の谷現象、Wikipedia。

<https://ja.wikipedia.org/wiki/%E4%B8%8D%E6%B0%97%E5%91%B3%E3%81%AE%E8%B0%B7%E7%8F%BE%E8%B1%A1>

これらは「こころ」の動きである。そうした「こころ」の作動原理、少なくともどのようなアルゴリズムで「こころ」が作動しているかを理解できていない現在では、「こころ」を AI に組み込むことは不可能である¹⁰。「拡張機能」により AI が疑似的にヒトの「こころ」に迫ろうとする「疑似的人間活動システム」の概念はを図 3-2 のように整理できる。ヒトの「こころ」に駆動された、表情、発言、端末キー入力、ロボットなどへの接触などを入力信号として AI は受けとめる。入力と出力の間に複層のレイヤーによりディープラーニングをさせる。その際、アプリケーションに応じた新しい技術を活用してより適切なアウトプットをねらう。アウトプット媒体として、ディスプレイ中心の「癒し」や「共感」用のチャットベースのアプリや、ロボットによる「癒し」などがある。ユーザーがその効果を感じ取れば、その効果がユーザーの身体と脳に送られ、次のループが始まる。効果がないとユーザーが判断すれば、中断されることになる。ユーザーが継続するか、中断するかの情報自体が、AI からの出力がその対象者に適合したものかどうかの評価としてフィードバックされ、AI が学習する。

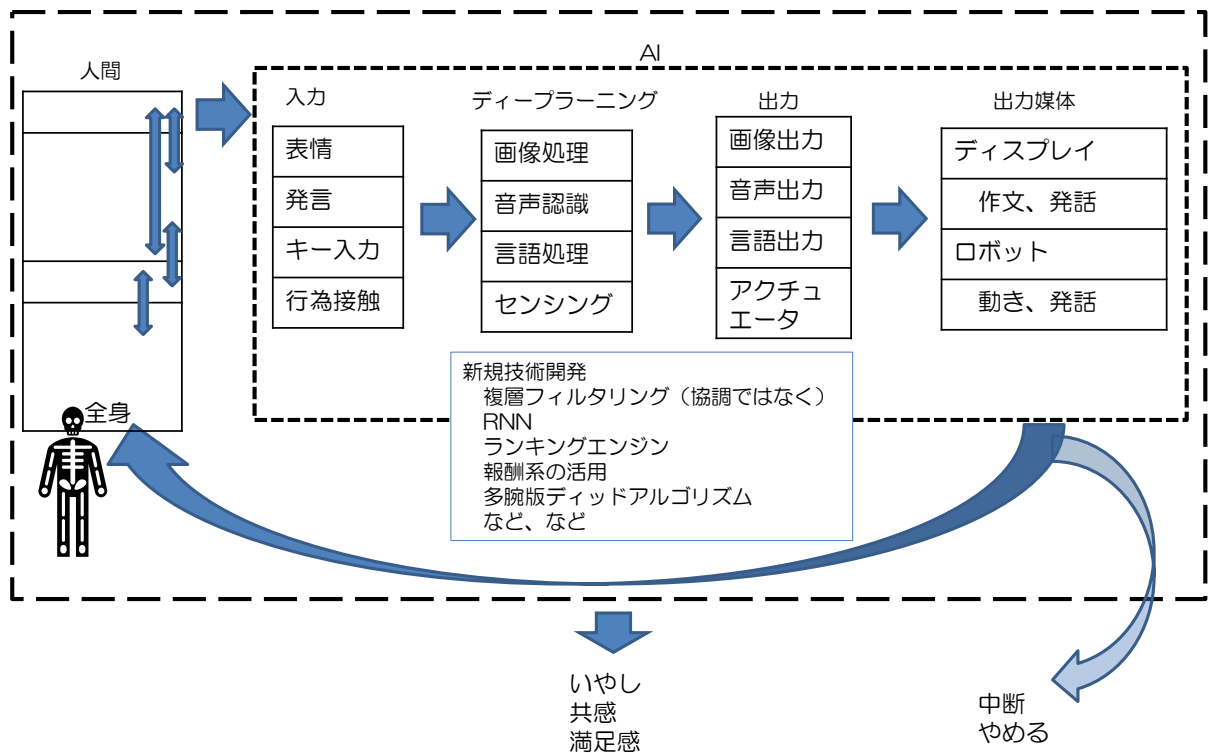


図 2-2 疑似的人間活動システム (著者作成)

実例については、NHK が放送した「人間ってナンだ？超 AI 入門」¹¹シリーズから今回の目的に合致する事例を採用する。シーズン 2 では、AI の製品開発担当者をスタジオに招き MC 達と対話することがメインである。シーズン 3 では、日本国内の開発者の現地訪問インタビューを実施して、最新技術の仕組みを松尾豊東大大学院教授と、哲学ナビゲーターで

¹⁰ 『「こころ」はいかにして生まれるのか』。櫻井武。講談社、2018

¹¹ シーズン 2 は 2019 年 1 月 10 日から全 12 回放送。シーズン 3 は 2019 年 4 月 3 日から全 12 回放送。

小説家の原田まりる¹²氏およびゲストが鼎談する。番組中の肉声の対話による書き取りは、各々の AI を開発する企業のホームページに記載されている紹介文よりも AI の特徴の説明が豊富であった。従って、その AI に関わるその企業が掲げる公式のメッセージと、テレビ放映時の音声記録書き出しの両方を活用することにする。事例については、下記 3 つに分類して、紹介することにする。

- (ア) AI がユーザーとチャットあるいは、センサーに感じる情報などを交換することにより、AI がユーザーをより擬似的に理解して、ユーザーに「癒し」や「共感」を与えようとする「人間中心型」と呼べるタイプがある。
- (イ) ヒトの「こころ」は 2 種に大別できる。個性と言われる要素が強い部分では、それは個人個人で大きく異なる。一方、ヒトの遺伝子に組み込まれていると思われる、生物として根源的な食欲、性欲、闘争心等に由来する「こころ」は普遍的に持たれていると考えられる。それらをターゲットに AI を活用していく「根拠説明型等」の方法論的進展が見られる。
- (ウ) 機械学習型の延長に位置付けられる事例として「癒しに特化したロボット」の開発がある。このメカニズムは電子回路とソフトウェアで作りこまれているに過ぎず、せいぜい対象者の趣味や特性の発現頻度に合わせてパラメータを自動的に調節する程度の個別化がなされているだけである。「拡張機械学習型」とする。

2.1 人間中心型への接近

最初の事例は「パーソナル AI」と言えるものである。ユーザーの心を直接的に知るために、現象論的に、音声やチャットでの対話により認識を模擬することで、そのユーザーに役に立つという意味で「パーソナル AI」である。次世代人工知能として『ヒトの感情、意志や価値観という内界を対象とする人間活動システムの「ソフト 3」』を個別に模擬できる固有の AI（「パーソナル AI」）を構想することは、新たなパラダイムの創出に相当する。つまり、他人の心を直接正しく理解することは難しい。現象論¹³によれば、ある人のふるまいを知覚して、その人が怒っていることはわかる。しかし、なぜ怒っているのか、またどれくらい怒っているのかなど未規程の部分が残る。この空白を埋めるために、相手に直接聞いて反応を見たり、別の人の証言をたよりにして確認したり、相手の人柄や置かれている状況等について知覚的な経験により直接的な知識にする。言い換えると、認識の枠組みが現象論等の哲学に基づいてなされている。つまり、対話を繰り返し、互いに「確信の確度」をあげる作業に基づいている。「パーソナル AI」を擬似的な他者にみたと、主体の内的状況の説明とその受容内容の確認とを繰り返し、「確信」の共有を図る。あたかも、人間活動システムの「ソフト 3」の領域に踏み込んでいるような事例である。

¹² 原田まりる（哲学ナビゲーター/小説家。哲学の魅力をわかりやすく面白く伝える案内人。AI 小説「びぶる」を上梓）

¹³ 植村玄輝、八重樫徹、吉川孝編著、現代現象学、新曜社 8・1「他人の心」、2017

2.1.1 共感チャット型コミュニケーションAI、「イケメンAI」

(NHK 超 AI 入門、シーズン 3。第 1 回・会話する。放送日 2019 年 4 月 3 日)

ユーザーから得られた多様な情報を、総合的に会話と情報提示へ結びつける複合的なもので、ユーザーを深く突き詰めていき、ユーザーに合った会話を作る。「癒し」や「共感」を生み出す。

(1) 特徴は共感 AI。AI×セラピー、心に寄り添う対話型 AI 開発。製品は「イケメン AI」。

(2) 会社

① SELF (株)、2014 年 11 月、資本金 3.7 億円。生明臣司 (代取)

② 会社案内 <https://self.systems/>

③ アプリ紹介サイト <http://self.software/#page00/slide2>

(3) SELF の AI エンジンの特徴 (ネット上での記載)

① 自動コミュニケーションを可能とする会話エンジン。

② SELF のコミュニケーションエンジンは大きく定義された状態構造と会話構造を自動でコントロール可能としており、独自に定義された状態要素を活用し、複層的、高速でどの会話や情報が提示するに適切かフィルタリングを行う。

③ また、記録作業は、単なるログとしてではなく、リアルタイムで次の会話や情報へ応用を行う。

④ SELF のコミュニケーションエンジンは単層的なレコメンドサービスやボット¹⁴サービスとは違い、ユーザーから得られた多様な情報を、各要素へ分解し、総合的に会話と情報提示へ結びつける複合的なもの。

⑤ 30 万近い会話のライブラリーを独自開発し、そのライブラリーからシステムが自動で適した会話を選択し、コミュニケーションを成立させる。ボットや自然言語処理開発者によって新たに開発された、自動で理解と推測を行い、自然な会話を進めることが可能なシステム。

(4) 番組中で紹介された特徴と評価

「心に寄り添うイケメン」についてのまとめた紹介としてネット上には『世の中のがんばっている、がんばりたい人に向けて開発されたイケメン。ユーザーと会話をすることで、特性を理解し適した会話を行います。ユーザーが疲れた時はイケメンが優しく包み込み、がんばりたい時は応援する、つらい時は慰める……イケメンが一番近くで、幸せを願い、支えになります。イケメンロボへ月額課金をする、様々な会話を投げかけることが可能になり、星座から導き出した日々の運勢を見ることができます』

技術の特徴としては、複層フィルタリング (よりユーザーに最適化する) 技術を使う。一般的な協調フィルタリングはショッピングサイトのお薦めの最適化などに使われている。これには他のユーザーのデータを使っている。弊社の場合は、ユーザーを深く突き詰めていくという方向で、よりフィルタリングを重ねていきユーザーに

¹⁴ Bot (ボット) は、robot (ロボット) の短縮形・略称で、転じてコンピュータやインターネット関連の自動化プログラムの一種のこと。

合った会話に落とし込んでいく。人の感情面に対応して、セラピー効果もでる。ユーザーを突き詰めていく。やればやるほどの確さが増していく。5,000以上の要素を組み合わせてフィルタリングする。精神安定的に使われている。人の感情部分に対応してくれる。瞬発的な感情をサポートしている。これを作ってわかってきたのは、ちゃんと受け答えをして、ちゃんと自分を理解してくれるならそれは人でなくてもいいのだということ。AIの役割は「人の感情部分に対応してくれるもの」と言える。アプリに話すことは「手軽に感情をぶつけられる」。うつ病が直ったり、自殺が止まったとかの例がある。それでこのアプリを使わなくなり、役目を終えたことになる。こころの港的。



図 2-3 イケメン AI

インタビュー記録鑑賞後の評価鼎談。

原田：グチアカウントにちかいようなアプリ。

石角¹⁵：米国では、精神的な病を抱えている人がいて、簡単にセラピストに行けない人がいるが、チャットぽい入り方は、いいと思う。サンドバック的。基本受け身。

松尾：感情面をサポートするというのは、ニーズがある。受け身だからこそ依存してしまう、それでも良くて「人は何かに依存したい」という傾向がある。トラブルがある間は、こういうアプリに依存して、後は、人を頼るということになって良いのかな、と思う。選択肢の一つとしては良いと思う。

石角：ビジネスとしての工夫が必要。ビジネスとしてどう収益を得るのか、回復すると使わなくなる。米国ならば、これ以上の質問をしたかったら、セラピストが入って、お金払ってねという方法がある。音声認識や対話の技術が進むとさらに取り入れていくということもある。もっと多くのユーザーが使うようになるかもしれない。心理学の先生がわかっていること以上のことが見えてくる可能性もある。

注) SELFの課金システムはネットでは、『※無料利用の場合、会話を重ねるにつ

¹⁵ ゲスト) 石角友愛 (AI ビジネスデザイナー、シリコンバレーを拠点に AI 導入をコンサルティング)

れて、イケメンの会話がかみ合わなくなります。正常な会話を楽しむためには課金もしくは、動画広告視聴が必要になります。』

製品としては5本目。

(5) コメント

(レベル4、ソフト2、ステージ4) ソフト3に「拡張」されてきている。ユーザーのこころを擬似的に理解でき、適切な会話構成ができるようになっており、「手軽に感情をぶつけられる」ことにより、うつ病が治ったり、自殺が止まったなどの例があるという。AIはヒトのこころを理解できないが、ここまでできる技術開発成果は素晴らしい。

2.1.2 チャット型恋愛アドバイスAI、恋愛ナビゲートAI

(NHK 超 AI 入門、シーズン3。第7回-恋愛する。放送日2019年5月15日)

恋愛により傷つかないようにしたいという女性開発者の気持ちを具現化して、気配されたりしたリスク回避の恋人探しアプリ。

(1) 特徴と製品。「その恋にAIがあなただの恋愛をナビゲート」、恋愛ナビゲーションサービス「Aill（エイル）」、「傷つかない恋愛」

(2) 会社

① 株式会社 AILL、(元は、株式会社 gemfuture (ジェムフューチャー) か?) 豊嶋千奈社長

② URL <https://aill-navi.jp/>

(3) Aill のサービス (ネット上での記載)

- ① (紹介ナビゲーション) まず Aill に登録し、自分のパーソナルデータを入力すると、すでに登録している異性から候補者が選出される。AI が、生活サイクルやキャリアプランを軸に、毎日1名~3名の異性をご紹介します。あなたと同じ価値観を持つ異性と出会うことができます。
- ② (会話ナビゲーション) あなたとお相手のチャットをAIが徹底アシスト。デートに誘うタイミングや、盛り上がる話題を提供します。出会う前に効率良くお相手を知ることができます。登録後、相手候補の異性とのやり取りが開始されると、そのやり取りをAIが解析し、双方の画面に「こうすると好感度が上がるよ」「今話した中からこの話題を深く聞けばいいよ」などと好感度が上がるような会話のテクニックがアドバイスされる(相手側へのアドバイスは画面上に見えない)。
- ③ (好感度ナビゲーション) 気になるあの人の、あなたに対する好感度をAIが可視化。効率的に異性にアプローチすることができ、あなたの背中を押してくれます。

(4) 番組中の特徴と評価

豊嶋千奈(恋愛ナビゲートAI開発会社、代表)、川村秀憲(北大教授、AIエンジン開発統括アドバイザー) 解説: 恋愛の手引きをするAI。六本木、(部屋のドア表示には LOUNGE NTT Docomo Ventures, Inc)。「その恋にAIがあなただの恋愛をナビゲート」テストユース版。相性(趣味や結婚観)でマッチングしても、従来のマッチングアプリでは、チャットの段階でうまくいかず、直接会うところまではなかなか

うまくいかない。本開発品では、チャット（会話）に AI が介入してアドバイスすることにより、願いを成就させる。チャットの単純な「何が好き」とかのやり取りの過程で AI が介入してくる。例えば相手が「テニスをやります」といえば、「テニスを好きになったきっかけは何なのかな？ 気になるね」などの話を深めるような質問の例として、このアドバイスが表示される。男性は仕事のことを語りやすいという分析が働いているようで、仕事の話が発信されると、「どんな仕事をしているのかな」、「最近印象に残った仕事内容は何かな？ 聞いてみてね」とかアドバイスが表示される。男性側にも AI がサポートする。デートに誘おうとすると「まだ〇〇さんを誘わない方がいいぜ〜！」とくる。AI は相手の好感度を分析しておりタイミングを図る。「人となり」がまだわからない段階でデートに誘われても行かない可能性大きい。テストユーザー1,000 人の会話データを集めて、うまくいった例、うまくいかなかった例がたくさん出てくるので、うまく行った例（一定以上会話が続く）を教師データに使う、AI をトレーニングしている。話題の選択タイミングをアドバイスできるようになった。

評価鼎談

原田：恋愛シミュレーションゲーム化している？

豊嶋：テストユーザーからリアル「乙女ゲーム¹⁶」だねって言われました。相手の男性と自分の間をつないでくれる AI だと思っている。

このシステムをある大手企業が福利厚生の一環として近々導入する予定であるという。

(5) コメント

（レベル 4、ソフト 2、ステージ 3）ただし、相手のところ、ユーザーの心が、恋愛マッチングでの傷付防止の配慮をねらっているので、ソフト 3 への挑戦とみる。しかし、その性能はまだまだ発展途上であろう。

2.1.3 パーソナル AI（PAI）、デジタルクローン

（NHK 超 AI 入門、シーズン 3。第 1 回・会話する。放送日 2019 年 4 月 3 日）
個人データから学習して、その分身をつくる。意思決定もさせようとする。

(1) PAI。私たち自身の意思をデジタル化し、それをクラウド上に配置してあらゆるデジタル作業をそのクローンにさせることを目的とした AI。

(2) 会社

① （株）オルツ。設立 2014 年 11 月。資本金、17 億円。

② <https://alt.ai/>

(3) Personal Artificial Intelligence の特徴（ネット上での記載）

① これまでの AI は、さまざまな人の膨大なデータ（ノウハウ）を解析し、最適な回答を導き出してきました。しかしパーソナル人工知能は、一個人の過去の行動や思考傾向などの「パーソナルな情報」のみを収集し、学習します。

¹⁶ 乙女ゲーム（男性キャラクターと会話を繰り返し恋人関係を目指すゲーム）。

- ② 将来的には、その人の話し方や好み、行動パターンなどのすべてを知る“分身”が完成するわけです。私たちは、特別なことをする必要はありません。ただ、いつものように日常生活を送るだけで、あとは AI が自動で情報を収集し、成長していく時代が実現されます。「少ない個人データから、“その人らしさ”」を再現する。
- ③ 現在、レコメンドエンジンとしての AI 利用や、将来的に自分の秘書（アシスタント）のような役割を果たす AI 像が語られることがあります。デジタルクローンはそれらと一線を画す存在になります。なぜなら、「決断をする AI」だからです。
- ④ 「alt 対話エンジン」には、AIML¹⁷ルールベースや知識ベース、オリジナルの回答を自ら考え作り出すことができる RNN¹⁸などの複数のモジュールが備わっており、それぞれのモジュールが出す複数の回答の中から、ランキングエンジンと意図認識エンジンが連携して、最適な回答を選び出すことが出来る仕組みになっています。チャットボットとは異なり、事前に登録していない質問に対しても、「対話エンジン」そのものが自ら考え、文章を作り出して発話出来ることが最大の特徴です。つまり雑談にも対応できるようになるのです。

(4) 番組中の特徴と評価

技術と効果：デジタルクローン化が目標（ネット上では「全人類 1 人 1 デジタルクローン」）。3 年間見聞きしたもの（SNS データ）があれば人として話せる。コミュニケーションのほとんどは、SNS チャットなどであり、そのデータで会話の傾向を取り込むことができる。それと連携してしまうと、その人を作ることができる。一人称や語尾の口調などのくせを学習する。メッセージサービスのチャットを食わせることができると日常生活の人格、仕事上のチャットであれば仕事上の人格がわかる。ユーザーの発言に対して、聞き返した方が良いのかよくないのか、答えた方が良いのかどうかという判断には、ランキングエンジンが入っている。何か問いかけると、7 個くらいのモジュールにそれぞれ問いに行き、それぞれが 5 個ずつ回答を準備する。35 個のなかから今の会話の中でどれが最適か選ぶエンジンが後ろにいる。ランキングをつけている。AI は通常待機している、何も考えていない。人間は答えた後に次何が来るかなとか、実は私はこれを言いたいなど絶えず考え続けていてだす。AI によって人間の裏側の目的などが明確になる。

追加ネット情報。『これまでの AI は、レコメンドまでを想定していましたが、私たちは「意思決定までを AI に任せる研究」をしています。人間は、1 日に 4 万回の意思決定（選択）をしていると言われていました。さらにそのうち、8 割は同じことの繰り返しとも言われています。ならばその大部分を占める“繰り返し”を AI に任せ

¹⁷ AIML(Artificial Intelligence Markup Language)。自然言語ソフトウェアエージェント構築のための XML を応用したマークアップ言語である。

¹⁸ RNN (Recurrent Neural Network) 通常のニューラルネットワークでは、ある層の出力は、次の層の入力に利用されるのみである。しかし RNN では、ある層の出力は、次の層の入力として利用されるだけでなく、一般的なニューラルネットワークの最後の層のような（中間データではないユーザーが利用可能な）出力としても利用される。

てしまえば、私たち人間のパフォーマンスは大幅に向上することになります。』

評価鼎談

松尾：難しいところをやっている。通常人間の発言（言語）は、思考、経験、哲学を踏まえてする。言語は氷山の見えているところで、その見えないところは難しい。「不気味の谷」ということがあって、発言が近すぎてその人がいいそうなことだが、心が入っていない。会話の世界で、「不気味の谷」を超えることができるのと本当にその人の言うようなことを話すことになる。共感力というのは話し手の立場にたって、話す。共感するのは人間でなくても AI でもいいのではないか。共感されたと人間が思えば良いので AI でも良いでないか。

(5) コメント

（レベル 4、ソフト 2、ステージ 4）ソフト 3 に「拡張」されてきている。

個人のチャット等のデータ 3 年分で、デジタルクローンを作り、意志決定まで任せたいとする。チャットのデータで一人称や語尾の口調などのくせを学習することは可能になるだろうが、どこまでの意志決定を代行できるようにすることができるのだろうか。デジタルクローン AI を意思決定代行まで含めてニーズが存在するだろうか。

2.1.4 パーソナルトレーニング AI (Personalized AI for everyone's wellness)

（NHK 超 AI 入門、シーズン 3。第 11 回・老いる。放送日 2019 年 6 月 12 日）

健康上の課題解決に向けてパーソナルトレーナーがやっていることをアプリで、ユーザーにやる気を喚起維持しながら継続的にやれるようにしようとする。

(1) 予防ヘルスケア×AI（人工知能）に特化したヘルステックベンチャー。

(2) 会社

① FiNC Tchenologies、南野充則¹⁹、代取&CTO。設立 2012.4.11。従業員 302 名。FiNC Wellness AI Lab。2016 年 8 月設立。

② URL <https://company.finc.com/corporate>

(3) パーソナルトレーニング AI のねらい（ネット上での記載）

私たちの身近には寿命と健康寿命のギャップや、体型のコンプレックス、腰痛や膝痛、肩こりなどの体の不具合と、健康に関する課題や悩みが数多く存在しています。私たちは「Personalized AI for everyone's wellness」というミッションのもと、最先端のテクノロジーを駆使して行動変容を起こすことで課題解決できる会社として FiNC Technologies を創業いたしました。私たちがやりたいことは健康寿命の延伸です。そしてコンプレックス、重い腰痛や膝痛に悩む人たちを一人でも減らしたい。健康の先にある幸せな生活を実現するため、日本だけでなく遠く離れた世界中の人々にココロとカラダの健康をサポートしてまいります。

(4) 番組中の特徴と評価

出演者は CTO、会社は有楽町ビルにある。パーソナルトレーナーがやっているこ

¹⁹ 著書「未来 IT 図解 これからのディープラーニングビジネス」、株式会社エムディエヌコーポレーション

とをアプリでやろうとするサービス。パーソナルトレーナーがまず、指導。姿勢評価をアプリ上で写真データを使ってやる。継続できるようなサポートとしてログをとる。ひとそれぞれに合ったコメントをする。怒られて伸ばすのか、ほめて伸ばすのかということも視野に入れている。多腕バンディットアルゴリズム²⁰を開発してユーザーのケイパビリティを高めていこうとしている。ユーザーの事例を当てながら情報を出す。試す回数をできるだけ減らしながら当りを引く確率を最大化する。これを統計的に考えるアルゴリズムが多腕バンディットアルゴリズム。AIが人間をラベリングしている。AIがその人にとっていい方法を見つけてくれる。アプリを開く回数が増えたとこのコメントは効いているなど考える。コメントの表現も大事。「今日はいよく頑張ったね」ということに対して、「おじさんボクないですか」とのコメントが来て、修正を加えた。

(5) コメント

(レベル4、ソフト2、ステージ4) ソフト2に「拡張」されてきている。

姿勢評価をしてそれにコメントを出すのは、画像解析のレベル。しかし、どのように励ましの言葉をかけると、持続的にトレーニングを進めていくのかというのはAIのこころの領域にせまる。

2.2 根拠説明型等の方法論的進展

癒し、共感、激励などが対象とする「こころ」は極めて個人により異なる。そのヒトの経験や知識に依存する意識が「こころ」を生み出すからである。この「こころ」はヒトにより大きく変化するために、現象論的な手法で、音声やチャット文字で対話や交流を重ねてヒトの疑似的な理解に迫る必要がある。しかし、「こころ」の中に、生物として組み込まれているともいえる「こころ」がある。闘争心や、困難だとされることを成就した時に味わう達成感満足感などである。これらは、多様なヒトがいただく属人的な変化のある「こころ」であるというよりも、普遍的でうつろいの少ない「こころ」あるいは情動に属するものである。この分野にAIを適用しようとする試みがある。本報告書では、多くのプレイヤーが共通して持つ、普遍的な特性に注目して、プレイヤーを興奮させ、満足させ持続的に達成感に浸らせるAIゲームを取り上げる。

また、現在、AIは病気の診断に多く使われている。例えば、内視鏡カメラで取得した画像データをAI診断することによって病気の診断に使うものであり、内臓表面画像からの疾病の摘出率は専門医の検出率を越えているものがある。一方、脳の機能の診断に使うものとして、患者の話す言葉を解析することにより、認知症の診断に利用しようとする研究開発が始まっている。脳の状態が発話に現象として反映されその言語解析により、認知症の診断に利用しようとするものである。この場合も、個人の属性に大きく依存しない、普遍的に表れる脳の

²⁰ multi-armed bandit problem、複数のアームと呼ばれる候補から最も良いものを逐次的に探す問題である。アームという奇妙な単語はこの問題のもとになったスロットマシン（バンディットマシン）の比喻から来ている。予測者はいくつかのスロットマシンを与えられ、それぞれのスロットマシンを引くと対応した報酬が得られる。繰り返す試行（アームの選択）を通じて得られる報酬を最大化するのが、予測者の目標である。

https://www.ai-gakkai.or.jp/my-bookmark_vol31-no5/

特徴を把握しようとしている。

ヒトの「内界」を対象としながらも、その「内界」は、複雑な「こころ」ではなく。生物として普遍的に発現する「こころ」や「意識」に限定することで新しく使える AI が発展してきている。

2.2.1 ゲーム AI

(NHK 超 AI 入門、シーズン 3。第 5 回・勝負する、放送日 2019 年 5 月 1 日)

三宅陽一郎の語るゲーム AI。ゲームでは、これまでは、同じゲームを何百万人に使ってもらおうということであったが、これからは、AI によって、ユーザーを理解して、ユーザーごとに変えて行こうというもので、百万通りのゲームの展開があるということを目指している。

(1) 会社、該当なし

(2) ゲーム AI と心理、人間（ネット上での記載²¹）

昔のゲームはスクリプトといい、どの場所でどの動作をどの場合でするかというパターンを、状態ごとに指定します。かつてのゲームはゲームデザイナーが脳でした。でも今はゲームの規模が大きくて、何万通りも作らなくてはならなくなるから、それができない。だから自律型 AI を作ろうというのが僕の立場です。操り人形じゃなくて、キャラクター自身に目と耳と鼻と感覚をつけて、世界を感じさせて、情報をもとに思考して、自分の体を動かしていきます。でも今はキャラクターのほうに知能を持たせるという作り方をします。どちらもいい面もあるし、悪い面もあります。けれども自律型 AI しか、大型ゲームに関しては道がないんですよね。それはこの 15 年かけて変えてきたことです。

人工知能の意識を作りたい。人工知能が自分自身を認識するといいますか、何らかの仕掛けで意識とかを作れたらと思っています。それを作れる可能性があるのはゲームだと思うんです。ゲームには世界があるので。目と耳と鼻と思考をつけて、感じたものを考えて、自分で動いていくところを、もうちょっと発展させれば、体験とか主観とか意思決定とか作れるんじゃないかと思います。生きているうちにできるかはわからないけれど、そこにたどり着くまでの間に今までにない新しい技術なり知見なり、面白いものがたくさん埋まっている気がするんですよ。今のところその方向に行きたいのは僕くらいです。ちょうど僕は争いが嫌いなので、一人でとぼとぼ歩いて行って、「あー、なんかみつけた」みたいな。そういうのができたらいいですね。

(3) 番組中で紹介された特徴と評価

三宅陽一郎の説明。実際は不確定要素が多いが、AI は自分がモデリングした中で判断しようとする、それ以外の要素をそぎ落している側面がある。敵であるモンスターだとか、仲間である人間のキャラクターがゲームの世界で自分を感じて、自分の主観世界をもってゲームに参加する AI が実装されている。哲学が必要で、体（か

²¹ <http://sugoihito.or.jp/2017/06/15704/>

らだ) 込みの知能全体を作りたかった、これができるのはロボットとゲームの世界。ゲームは知能の方に特化できる。ゲームの世界では AI は圧倒的に強いので、如何に手加減するかというところがポイント。メタ AI (ゲームの進行を監視しながらより面白くなるようゲーム全体をディレクションする役割を持ったシステム) がリアルタイムでロールを割り当てる。監督の AI、役者の AI がいて、うまい具合に切られ役を動かせる、とか。それを気付かれずに、プレイヤーに「俺すごい」と思われるのがいい。(松尾:「ゲームは報酬系のハック」という学生がいた、その通りでないか) ゲームの世界ではその報酬系を時間スケールで、積み上げていく、例えば、モンスター①を倒して経験値を上げ、モンスター②を倒してレベルアップする、さらに上げると次の街に行けてとか新しい道具を買えるというように時間とともに「やる気」を上げる。「時間のスケールとともに、様々な報酬系を組み合わせる」のがセオリー。質が異なる報酬を集める。ある程度ユーザーが強くなると、ゲーム側は強くしないで「俺つええ感」を与える。無双状態を一定期間プラトー(高原)的に作る。その後難易度を上げて、それを超えると、また「俺つええ感」を与える。これを繰り返す。ゲーム開発は人間研究でもある。今プレイしている人間を知ることができるとエンターテインメントとしては勝ちなところがある。AI が人間を理解することで、ゲームをユーザーに合わせて変化させていくことがゲーム産業で興っていること。開発費が数十億円/本を超えて、そのデータ量が 50G (wikipedia の総量) を超えるものもある。プログラムのデバッグ作業も AI がやる。

ゲームでは、これまでは、同じゲームを何百万人に使ってもらおうということであったが、これからは、AI によって、ユーザーを理解して、ユーザーごとに変えて行こうというもので、百万通りのゲームの展開があるということを目指している。教育も同じところがある。各人対象に、それぞれの人に応じた教育方法をつくるということで、ゲームのメタ AI も使えると思う。人に応じた教育課程の可能性がある。ゲーム系の AI デザインは教育に使えるのではないかと昔から言われている。「報酬系のハック」の利用。

(4) コメント

(レベル 4、ソフト 2、ステージ 4) ソフト 3 領域に差し掛かっている。人間であるプレイヤーの競技するマインドを掴んでいる。こころの中の闘争心を想定して、ゲームの中でその闘争心と満足感・達成感を織り交ぜて、プレイヤーのプレイ時間を増やすかを考える。こころの領域に踏み込んでいる。

2.2.2 認知症診断のための対話ロボット AI

(NHK 超 AI 入門、シーズン 2。第 11 回・老いる。放送日 2019 年 3 月 21 日)

言葉で脳の状態がわかる。認知機能低下の判断をするための対話ロボット「ルドウィッグ (Ludwig?)」

(1) キャッチコピー。認知症診断の AI エンジン

(2) 会社・機関。フランク・ルディッチ Frank Rudzics (ベクターインスティテュート²²)

(3) 番組中の特徴と評価

発話が不自由な人の脳で何が起きているか調査している。認知症や認知機能低下の研究として、脳を見るレンズ(脳レンズ)として言語を利用している。言葉で、脳の状態がわかる。対話ロボット「ルドウィッグ」。認知機能障害がある人と対話させ、音声から認知症判断を行うシステムを開発している。数日通して記録して、傾向をみる。語彙の量は大事な特徴になる。語彙が多い人は健全な場合が多いが、もともと少ない人もいる、その人を良く知って、時間をかけ追跡していくことが重要になる。患者が自宅のソファでできるテストの開発を考えている。その方がその人の頭の状態を良く知ることができる。

(4) 評価と考察

(レベル 4、ソフト 2、ステージ 2) ソフト 3 に差し掛かることが期待できる。

ひとが認知症であるかどうか、人間の発言内容や行動で判断する。AI が発言内容で判断できる可能性はある。まだ研究段階。

2.3 「拡張機械学習型」

機械学習型の延長に位置付けられる事例として「癒しに特化したロボット」の開発がある。このメカニズムは電子回路とソフトウェアで作られてしまっているに過ぎず、せいぜい対象者の趣味や特性の発現頻度に合わせてパラメータを自動的に調節する程度の個別化がなされているだけである。しかしヒトの趣味はうつろい行くものであり、それまでの統計的結果が当人にとっていつも快適であるとは限らない。したがってこのアプローチは第 3 世代型 AI の拡張を目指しているとはいえ、原理的には第 3 世代型の枠組みの中にとどまっている。とはいえ、第 3 世代型の「拡張」の有り方によっては、それを「次世代」と称するに相応しい場合もあるであろう。したがって、第 3 世代型の拡張ないしそこからの脱却を目指す取り組みに相当し、それらを精査し原理的な飛躍があるかどうかを検討する。

2.3.1 いやしのロボット、家庭用ロボット : Lovot

(NHK 超 AI 入門、シーズン 2、第 7 回・恋愛する。放送日 2019 年 2 月 21 日)

あえて会話をしない「癒しに特化したロボット」。

(1) キャッチコピー。小さな LOVE が、世界を変える。LOVE をはぐくむ家族型ロボット。Lovot は造語 (Love+Robot)。人間とロボットの信頼関係を築き、生活を潤いと安心で満たす存在をつくる。

(2) 会社

① GROOVE X (株) 代表取締役 林要 (AI ロボット開発者)、設立 2015 年 11 月

²² Vector Institute。米グーグルはカナダで人工知能 (AI) 関連の投資を拡大しており、トロント大学が新設した AI 研究施設に 500 万ドル (約 5 億 6 千万円) を拠出して、カナダの連邦政府とオンタリオ州政府の後押しも受け、2017 年 3 月 30 日に「ベクターインスティテュート」が開設された。「ディープラーニング (深層学習)」と呼ばれる AI の最新分野の第一人者でグーグルの副社長兼技術フェローのジェフリー・ヒントン氏が、同施設の最高科学顧問を務める。

② URL <https://groove-x.com/>

(3) Lovot のねらい（ネット上での記載）

未来は見えない。だからワクワクもするし、だから不安を感じる。ならば未来を、すこしでも身近で、安心できる場所としていくため、技術というものを活用しなければならんだ、と。この生活と地続きの感情、たとえば、うれしいとか、愉快だとか、さびしいとか、そういうものと真正面から向かい合うこと。私たちの技術は、そこをめざします。



図 2-4 Lovot

(4) 番組中の説明と議論

林要（AI ロボット開発者）いやしロボット：愛する対象が人でないとダメなのか。ロボットでもいいのではないかと、ということで、いやしロボットを開発している。1,000 人まで AI で個人認識している。人の認識、距離の認識、すごい技術である。特徴は誰がどんなふうに触ったかということがわかること。スキンシップによる情報交換を進めるために 50 以上のセンサー、10 以上の CPU で即時処理して、交信する。機能として、言葉の意味がわかる、部屋の地図を作れる、部屋の全体図の中で自分の位置がわかる、自分で動ける。人が愛するチャンスを増やすロボット。Lovot は言葉をしゃべらせるのはやめている。どう気軽に愛せるか。愛すること自体が脳内分泌（幸せホルモン）がよくなる。愛する力を鍛える。いやしロボット。癒しの効果はある。人の認識、距離の認識、すごい技術（松尾）

(5) コメント

（レベル 4、ソフト 2、ステージ 4）ソフト 3 に「拡張」されてきている。

Lovot は話さない。鳴き声を出すのみ。可愛さを出す Kawaii チームが設計しており、その表情を見ているだけで十分癒される。センサーの数と CPU を相当数使っており、ユーザーが撫でたり、抱っこするなどの接触行為がわかるようになっている。それに応じて喜んだり、目を閉じたりする。動く縫いぐるみという感じ。話す必要はない設計思想になっている。これはユーザーとの会話することにより違和感を醸し出さないようにするためには相当の開発を要することがわかっているから、この

設計になっているのであろう。Lovot の出力の効果は、抱っこしたりする時間の長短で計測できるであろう。

3. 脳科学の知見を反映した事例

AI 技術の進展と同様に、脳科学も近年目覚ましい進歩を遂げつつある。ここでは最近の脳科学の進展を概観することで、次世代以降の AI の R&D 方向と共にヒトの脳機能の強化・拡張の可能性に関して何らかの示唆を得ることを意図した。

一言で脳科学といっても、生物科学的アプローチから数理科学的アプローチまで、そのカバー範囲は極めて広い。そこで、次世代以降に期待されている汎用 AI (AGI) 技術との関連が深いと考えられる領域に絞って簡単な予備的調査を実施した。具体的には、記憶や意識が脳科学では現在どのように理解されているのか、脳・神経の信号解読はどの程度可能なのか、脳と外部機器の直接的接続、信号入出力 (BMI/BCI) ²³はどの程度制御出来るのか、脳へのフィードバックや刺激印加で脳機能の活性化、強化は可能か、といった辺りを主に 2011 年以降の文献・資料を中心に調査した。

3.1 脳科学の進展と脳機能の理解

3.1.1 情報伝達・処理システムとしての脳 ²⁴

ヒトの脳は、約 1,000 億個のニューロン (神経細胞) ²⁵が約 1,000 兆個に達するシナプス ²⁶によって相互接続され、電気信号をやりとりする神経回路ネットワークと捉えることができる。ニューロンは、電気信号を生成するための多様なイオンチャネルとその活性調節機構を有する機能素子である。シナプスにある小胞から神経伝達物質がシナプス間隙に分泌され、これが 0.1~0.2ms で次のニューロンの細胞膜にある受容体に結合すると電気信号が生成し、情報が伝達される。神経伝達物質としては、グルタミン酸、アセチルコリン、ノルアドレナリン、ドーパミン、等々、今まで数十種類が発見されている。シナプスは、神経細胞間の正確な配線を形成するだけでなく、学習や環境によってシナプス接続強度を変化させており、これがネットワークの基本的特性を支える ²⁷。

大脳皮質には 10 種類、小脳皮質では 5 種類ほどのニューロンがあり、機能を分担すると共にそれぞれの種類ごとに同じ階層に集まり、多層システムを形成している。この秩序だった神経回路により、情報伝達が効率化され、学習、記憶、運動、認知などの高度な情報処理が可能になっている、と考えられている。この様に、脳はニューロンの機能モジュールが階層的に構造化されたシステム (静的脳観) として記述される場合が多い。しかし、脳の機能は仮に学習がなくとも、外界の影響で変化し得るので、一時的な文脈に応じて個々の要素の機能は変化し、動的に決定される (動的脳観) といった捉え方もある (例えば津田一郎、2002) ²⁸。

²³ BMI/BCI:ブレイン・マシン・インターフェイス/ブレイン・コンピュータ・インターフェイス

²⁴ 主に理化学研究所 WEB 情報を参考にした。

²⁵ ニューロン数は大脳で数百億個、小脳で千億個程度といわれている。

²⁶ シナプスは大きさ約 $1\mu\text{m}$ 。ニューロン 1 個当たりのシナプスは 1 万个程度とされる。

²⁷ 池谷グループ (2012) は、多くのシナプス活動を観察できる新撮影技術「大規模スパインイメージング法」(Ca イメージングの一種)を開発し、マウスの脳神経回路が、シナプスの単位 ($1\mu\text{m}$ レベル) で正確に編まれることで機能を発揮することを、初めて明らかにした。出所: Naoya Takahashi, Kazuo Kitamura, Naoki Matsuo, Mark Mayford, Masanobu Kano, Norio Matsuki, Yuji Ikegaya (2012), "Locally Synchronized Synaptic Inputs", Vol. 335 no. 6066 pp. 353-356.

²⁸ 津田一郎 (2002) 動的脳観(力学系,動的システムの情報論,研究会報告), 物性研究, 78(6): 677-680.

現実の脳はニューロンだけではなく、グリア細胞からも構成されており、アインシュタインの脳は通常よりグリア細胞が多いことが特徴だった、といわれている。従来、このグリア細胞の機能は不明だったが、最近の研究からグリアが神経成長因子や栄養因子などを分泌し、ニューロンの維持や再生に重要な役割を有することが分かってきた。脳の構造・機能の解明は急速に進展しつつあるが、まだ不明点も多く発展途上でもある。

3.1.2 記憶と意識に関連した理解の進展

近年の脳科学の進展は、脳の透明化技術、二光子励起顕微鏡、光活性化局在顕微鏡、等の顕微鏡技術における革新や光遺伝学 (optogenetics) といった、バイオ・イメージング技術の進化に支えられている部分が多い²⁹。米国スタンフォード大学のグループが 2000 年初頭に汎用的な技術として開発した光遺伝学は、光で活性化されるタンパク分子を遺伝学的手法によって特定の細胞に発現させ、その機能を光で操作する技術である³⁰。具体的には、光に反応しかつイオンチャネルとして働くチャネルロドプシンを、遺伝子操作でマウス、等の特定のニューロンに導入し、光ファイバーを通して光照射を on/off することで膜電位を上下させ、ニューロンの発火を制御する (時間分解能は ms オーダー)、といった手法である。この光遺伝学の開発により、特定の神経活動を高精度で活性化もしくは抑制が可能となり、神経活動と行動発現とを直接対応させることができるようになった。以下、光遺伝学、等を活用した最近の主に生物科学的アプローチに基づく脳の記憶や意識に絡む成果を概観する。

Xu Liu ほか (2012) は、光遺伝学をマウスに適用して、エングラム (記憶の痕跡) が単なる概念ではなく、ニューロンに物理的に存在することを実証した³¹。まず、マウスが学習し、海馬の特定のニューロンが on 状態になると、これらの細胞が光に反応するチャネルロドプシンで標識されるトランスジェニックマウスを作成した。そして、マウスにある環境下で軽い電気ショックを与えることで、環境とショックの関係を学習・記憶させる。それにより、マウスは、この環境下に置かれると、ショックを思い出して恐怖による“すくみ”行動を示すようになる。この学習したマウスをまったく別の環境に移しても、学習中に on になった細胞群に、今度は光を照射して再び on 状態にすると、“すくみ”行動が喚起できることを確認した。つまり人工的刺激でショックの記憶を思い出させることが出来たわけであり、これは、記憶が特定の脳細胞に物理的に存在することを示すと共に、心が脳の物理的 (or 物質的) 変化に基づいていることを示すと推察される。

Steve Ramirez ほか (2015) は、光遺伝学を利用して、マウスのうつ状態が改善できることを示した³²。まず、オスのマウスに楽しい体験 (メスのマウスと共に過ごす) をさせ、そ

²⁹ たとえば、理化学研究所 脳科学総合研究センター編 (2016) つながる脳科学:「心のしくみ」に迫る脳研究の最前線, 講談社。

³⁰ Edward S Boyden, Feng Zhang, Ernst Bamberg, Georg Nagel, Karl Deisseroth (2005), Millisecond-timescale, genetically targeted optical control of neural activity. Nat. Neurosci.: 2005, 8(9);1263-8.

³¹ Xu Liu, Steve Ramirez, Petti T. Pang, Corey B. Puryear, Arvind Govindarajan, Karl Deisseroth, and Susumu Tonegawa (2012), “Optogenetic stimulation of a hippocampal engram activates fear memory recall” Nature, 2012 DOI: 10.1038/nature11028.

³² Steve Ramirez, Xu Liu, Christopher J. MacDonald, Anthony Moffa, Joanne Zhou, Roger L. Redondo & Susumu Tonegawa (2015), “Activating positive memory engrams suppresses depression-like behavior”, Nature 2015, doi: 10.1038/nature14514.

の時に活動した海馬の歯状回の神経細胞群を遺伝学的手法により標識（光感受性タンパク質遺伝子 ChR2）する。次いで、そのオスのマウスに慢性ストレス（＝体の固定）を与えると、マウスがうつ様の行動（＝嫌な刺激を回避する行動が減少；好きな砂糖水を好まなくなる）を起こすようになることを確認した。その後、うつ状態のマウスの海馬歯状回で楽しい体験の記憶として標識された神経細胞群を、光遺伝学手法で人工的に活性化すると、うつ状態が改善されることを見出した。

J. P. Johansen ほか（2014）は、光遺伝学をラットに適用し、怖い体験が脳に刻み込まれるメカニズムを検討した³³。光遺伝学によって、怖い刺激である電気ショックの瞬間だけラット脳内の扁桃体のニューロンを抑制すると、電気ショックと同時に提示された音刺激に対する恐怖記憶の形成が阻害されただけでなく、扁桃体でのニューロン同士のつながりの強化も妨げられ、ヘッブの仮説（シナプス可塑性）を支持する結果が出た。一方で、扁桃体のニューロンを活性化しただけでは、恐怖記憶は形成されないことも分かった。しかし、扁桃体に加え、覚醒や注意に作用する神経修飾物質のノルアドレナリンの受容体を同時に活性化させると、怖い体験を与えなくても恐怖記憶が形成されることも明らかになった。この結果は、恐怖体験の記憶形成において、ヘッブ型可塑性は有力仮説であるが、それだけでは十分ではなく、神経修飾物質の活性化も重要であることを示している。実際の脳内メカニズムは、ヘッブ型可塑性で説明されているよりも、もっと複雑であることが分かった。

また、理化学研究所の深井、等を主体とする国際共同研究チーム（2018）は³⁴、数理科学的アプローチを加えて、動物（ラット）の行動選択における意思決定の個体差が、その神経活動の個体差、つまり外からの刺激に対する神経ネットワークの個体ごとの感受性の違いによって決まることを発見した。ラットは、学習した行動選択のルールを、学習したことがない条件下でどのように応用するのか、ラットの脳神経活動を記録することで調べた。その結果、個体間の行動選択の大きなバラツキと、内側前頭皮質のニューロン集団の活動の揺らぎとの間に相関があることを見出した。そしてラットで見られた行動選択の個体差が、コンピュータ上で神経回路モデルを訓練することにより再現できることを示し、行動の個性が神経活動の外的擾乱に対する感受性（安定性）で決まることを示した。

ここまで記述した研究成果例は、動物実験をベースとしたものである。ヒトの脳に関する研究の進展はどうであろうか。ヒトの場合、疾病治療、等の一部の例外を除いて侵襲的な手段を用いた実験（主に生物科学的アプローチ）は、通常困難である。以下、主に数理科学的アプローチに基づいてヒトの記憶や意識を制御した例を概観したい。20 世紀後半から、脳波、等を望ましい状態に調整するニューロフィードバックと呼ばれるトレーニング手法が、精神疾患の治療、アスリートや兵士のパフォーマンス向上、といった目的で検討される様になったが、ここでは、その進化系として位置づけられる DecNef（Decoded Neurofeedback）法を

³³ Joshua P. Johansen, Lorenzo Diaz-Mataix, Hiroki Hamanaka, Takaaki Ozawa, Edgar Yeu, Jenny Koivumaa, Ashwani Kumar, Mian Hou, Karl Deisseroth, Edward Boyden and Joseph E. LeDoux (2014), "Hebbian and neuromodulatory mechanisms interact to trigger associative memory formation", *Proceedings of the National Academy of Sciences of the United States of America*, 2014, doi: 10.1073/pnas.1421304111.

³⁴ Tomoki Kurikawa, Tatsuya Haga, Takashi Handa, Rie Harukuni and Tomoki Fukai, (2018) "Neuronal stability in medial frontal cortex sets individual variability in decision-making", *Nature Neuroscience*, 10.1038/s41593-018-0263-5.

紹介する。

国際電気通信基礎技術研究所（ATR）の川人光男グループが開発した DecNef 法は、ヒト脳活動を fMRI（核磁気共鳴機能画像法）で非侵襲的に計測し、そのデータから脳の情報を解読、ほぼリアルタイムで脳に報酬として帰還し、結果として特定の空間的脳活動パターンを誘起する手法であり、3 つの段階から構成される³⁵。第 1 段階では、脳情報のデコーダを作成する。最初に被験者の特定の知覚や認知、行動に伴う被験者の脳活動パターンを fMRI で計測し、次いで、DecNef 法で使用する目標領域を決め、その目標領域の活動パターン（ターゲットパターン）を計算する。そして機械学習アルゴリズムで任意の脳活動パターンとターゲットパターンの間の類似度を定量化するためのデコーダを作成する。第 2 段階は、脳活動パターンの誘導である。被験者はターゲット領域の活動を行い、そこで計測された脳活動パターンをデコーダに入力し、その時のパターンがターゲットパターンにどの程度近いかわかり類似度を計算し、この類似度を緑色の○印の大きさに被験者にフィードバックする。被験者は、脳活動パターンの誘導により緑の○を出来るだけ大きくすることが求められる。この実験を一定期間繰り返すことで、被験者は自身の脳活動をターゲットパターンに近づけることを学習する。第 3 段階は、事前の活動と訓練後の活動の比較であり、DecNef 法をもちいた訓練によって、実際に被験者の知覚や認知、行動が変化したか否か、因果の検討、確認を行う。

この DecNef 法は、革新的研究開発推進プログラム（ImPACT）の中でも検討されてきたが、「つらい経験を思いだすことなく、無意識のうちに恐怖記憶を消去できるニューロフィードバック技術を開発」と題するプレス発表が、2016 年 11 月 22 日に ATR、NICT（情報通信研究機構）、米国 UCLA、英国ケンブリッジ大学の共同で行われている³⁶。更に 2018 年 3 月 7 日には、「二重盲検により恐怖記憶緩和の効果を証明：レディーメード脳情報解読によるニューロフィードバック法」と題するプレス発表が ATR、米国 UCLA、JST、内閣府の共同で行われた³⁷。ヒトの脳の記憶や意識も部分的には、人工的手段で制御可能といえるであろう。

3.1.3 意識生成に関する理論

前述の如く、脳機能に対する理解が進展し、脳への人工的な入力や操作で、ヒトを含む動物の記憶や意識を制御することが、部分的には可能になりつつある。しかしながら、意識生成のメカニズムは相変わらず未解明である。

そもそも意識とは何か。意識（精神）と脳（物質）の関係に関しては古くから多様な領域で議論されてきているが、意識の定義自体が曖昧だといえるかもしれない。

信原幸弘（2008）は、脳科学が扱う「心」は、日常的に理解される「心」のうち、その法

³⁵ Shibata K, Watanabe T, Sasaki Y, Kawato M (2011): Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation, *Science*, Vol.334, Issue 6061, 1413 – 1415.

柴田和久（2012）DecNef による神経科学の新しい試み, *臨床神経*, 52, 1185 – 1187.

³⁶ 対応論文：Ai Koizumi, Kaoru Amano, Aurelio Cortese, Kazuhisa Shibata, Wako Yoshida, Ben Seymour, Mitsuo Kawato, Hakwan Lau (2016): Fear reduction without fear: Reinforcement of neural activity bypasses conscious exposure. *Nature Human Behaviour*. DOI: 10.1038/s41562-016-0006.

³⁷ 対応論文：Vincent Taschereau-Dumouchel, Aurelio Cortese, Toshinori Chiba, J. D. Knotts, Mitsuo Kawato, and Hakwan Lau (2018): Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, www.pnas.org/cgi/doi/10.1073/pnas.1721572115

則的な側面だけであり、合理的な面（相手の振る舞いを合理的に解釈したり、相手を合理的に説得したりする時の心の読み取りや操作）は扱っていない、といった指摘もしている³⁸。脳科学では、意識や心をどう捉えているのか。脳科学辞典³⁹によると、脳科学では「意識」は、次の二つの意味で使われている。一つは、医学の世界で使われる「意識レベル」ないし「覚醒」という時の意識であり、もう一つは、心理学などが扱ってきた「クオリア」や「意識内容」という時の意識である。また「心」は日本語特有の概念であり、たとえば英語では「心」に対応するような言葉はない、といわれている。

さて、意識生成のメカニズムはどの様に考えられているのか。意識の理論（仮説）としては幾つか提案されているが⁴⁰、脳科学辞典に記載されている2つの理論を代表例として本節の最後に簡単に紹介する。

(1) GNW 理論 (Global Neuronal Workspace Theory)

- ・ B. Baars が提唱した GW 理論 (Global Workspace Theory) を S. Dehaene、等が脳科学的に検証できるように発展させた理論
- ・ 無意識の処理は、感覚器への入力・運動出力を担う周辺的な並列処理に対応し、これらの情報は通常、限られた場所（ワークスペース）に局在している。
- ・ 感覚器から出る多様な情報（無意識処理）の一部は、注意によって選択・増幅され、GW (Global Workspace) に入ると長期記憶や抽象的思考になる。
- ・ GW は前頭前野を中心とした脳内に広く分布したニューロン集団であり、感覚器からの情報が GW 全体に広まった時に意識が生じる。

(2) 統合情報理論 (IIT : Integrated Information Theory)

- ・ G. Tononi が提唱。脳科学知見との整合性は高い、とされる。
- ・ 主観的に体験する多様な情報が統合されて初めて意識が生成する。統合とは、原因と結果が構造化され相互に接続された複雑な関係を意味する。
- ・ IIT では、相互接続の複雑さを Φ (ファイ) で定量化し、 $\Phi=0$ では意識は生成せず、 Φ の数値と共に意識が強くなる、と考える。人の脳は他の動物より複雑なシステムで Φ が大きく意識が強い。

3.2 脳と外部機器との接続および信号入出力 (BMI/BCI)

3.2.1 脳と機械を接続する BMI/BCI 技術の概要⁴¹

脳と外部との直接的な信号のやりとりは、BMI (Brain Machine Interface) もしくは BCI (Brain Computer Interface) として古くから研究されてきたが、近年さらに活発化し論文

³⁸ 信原幸弘 (2008) 脳科学と心の機械化, 哲学 59, 97 – 114.

³⁹ <https://bsd.neuroinf.jp/wiki/%E6%84%8F%E8%AD%98>

(脳科学辞典の記述内容は、2016年6月9日の「意識」の原稿に基づく)

⁴⁰ 意識関連の理論としては GNW 理論、IIT 理論以外でも、D. Mumford (1992)や川人光男, et al,(1993)の生成モデル、M.S. A. Graziano (2017)のアテンション・スキーマ理論、等も知られている。

⁴¹ 全般に①、②を参考にした。①有住なな (2015) ブレイン・コンピュータ・インターフェース概説, オペレーションズ・リサーチ, 227 - 233. ②吉峰俊樹 et al. (2016) ブレイン・マシン・インターフェイス (BMI) が切り開く新しいニューロテクノロジー, 脳外誌 25(12) 964 – 972.

数も増加傾向にある（下図参照）。

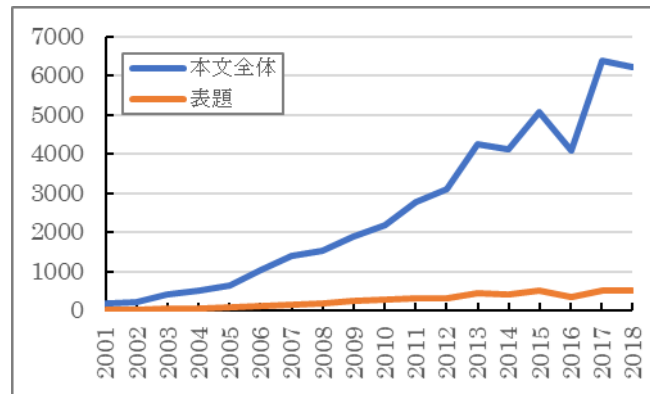


図 3-1 BMI/BCI 関連論文数の経年推移(Google Scholar)

また、米国では Neuralink 社⁴²（Elon Musk が 2016 年に立上げた BMI 開発企業）や Facebook 社⁴³といった企業が、最近 BMI/BCI 開発に進出し、話題となっている。

通常、ヒトが外部機器を操作する場合、脳から発信された指令信号が筋肉に伝達され、手足を動かしたり、キーボードに触れたりすることで実行される。この時、脳からの指令信号を筋肉を介さずにコンピュータで読み取り、直接的に義手やロボット、電動車椅子、その他の機器を操作する様なケース（脳→外部）が、出力型（or 運動系）BMI/BCI である。逆に外部の情報信号を、感覚代替機器、等（実用化例は人工内耳、人工網膜）を介して脳に直接入力するケース（外部→脳）は、入力型（or 感覚系）BMI/BCI と呼ばれている。BMI/BCI のキー技術のひとつは、脳の信号を記録し、特徴量を抽出、分析することであり、AI 技術の展開例と位置づけることもできる。

脳と BMI/BCI の接続は、非侵襲方式だと信号の質が低下するため、医療目的を主体に、脳の皮質内や硬膜下、等に電極を設置する侵襲方式も多く検討されている。BMI/BCI 技術の従来の実用化例としては、こういった出力型もしくは入力型機器が多く、現在も活発に研究開発が継続されているが、近年は両者を合わせた様な介入型 BMI/BCI（ex. DecNef 法の如きニューロフィードバック）が注目されつつある。また、長期的にはマインド・アップロードに繋がる技術と捉える向きもある。

⁴² 将来的には脳内チップ（侵襲型 BMI）を AI システムと接続し人間の能力増強を目指す、としている。

<https://www.rt.com/usa/401052-elon-musk-neuralink-funding/>

⁴³ 1 分に 100 ワード入力可能かつ他言語翻訳変換可能な非侵襲型 BMI を開発中。

<https://forbesjapan.com/articles/detail/11906>

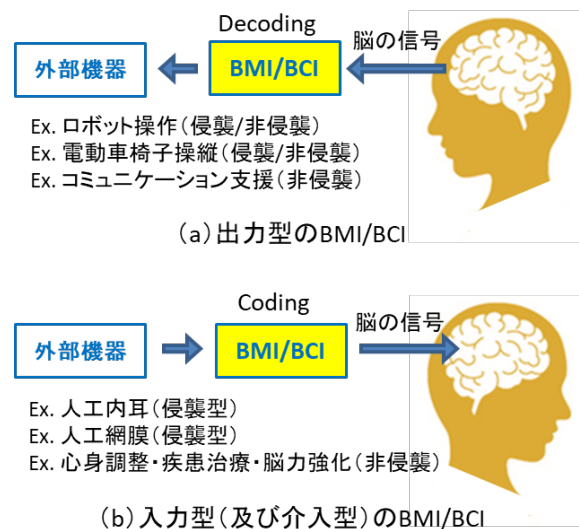


図 3-2 BMI/BCI 基本技術構成の概念図

有住（2015）、吉峰ほか（2016）を参考に筆者作成

本稿では、BMI/BCI による将来のマインド・アップロードの様な可能性の探索と脳へのニューロフィードバックや刺激印加によって脳力の強化・拡張は可能なのか、といった辺りに絞って焦点を当てる（詳細は 4.3 参照）。

3.2.2 代表的な出力信号計測法と簡易な電磁的刺激の入力法

脳・神経細胞の電気的な活動を計測する方法は幾つか提案されているが、BMI/BCI 技術で用いられている代表的な計測法は、概ね下記となる⁴⁴。

(1) 間接的計測法（応答速度≒秒オーダー）

ニューロンの活動を間接的に計測する方法としては、fMRI（核磁気共鳴機能画像法；血液量変化の計測）と NIRS（近赤外分光法；酸化・脱酸化ヘモグロビン量の計測）の 2 種が代表的であり、共に非侵襲型の計測である。

(2) 直接的計測法（応答速度≒ms オーダー）

ニューロンの電気信号を直接的に計測する場合、非侵襲型では脳磁計（MEG）と頭皮電位測定（EEG）、侵襲型では、硬膜外電位測定（eECoG）、硬膜下電位測定（sECoG）、皮質内電位測定（LFP）がしばしば使用されている。たとえば、非侵襲型の EEG は安全性は高いが、頭骨を介した測定のためニューロン電位の高周波成分はカットされ、約 50Hz 以下の成分が測定される。さらに空間分解能も 30～50mm と低く、信号強度も数百 μV と低いため環境の電磁ノイズや身体運動による EMG（筋電図）もノイズとなり、信号の精度は低くなる。一方、侵襲型は安全性が低下するが、信号の質は向上する。たとえば sECoG の空間分解能は約 2.5mm であり、EEG よりも 1 桁向上する。因みに、動物（サル）の sECoG から EMG（筋電図）を推定し、実際の筋電位との相関を調べる

⁴⁴ 横井浩史・姜銀来（2017）BMI 出力デバイス：人と機械が適応できる技術を目指して、精密工学会誌 83(11) 1000 - 1005.

と、筋肉によって相関状況は異なるが、概ね相関係数は 0.7 前後と報告されている。

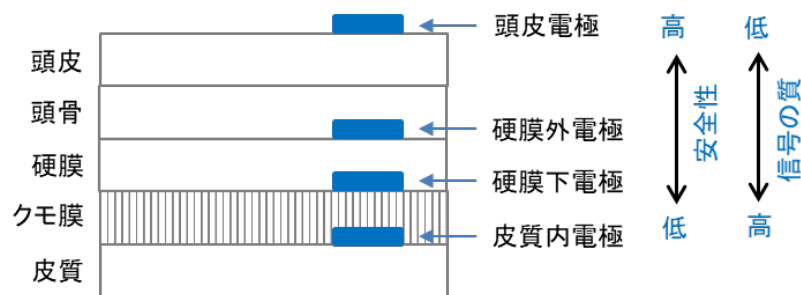


図 3-3 脳神経活動直接計測法の比較模式図

横井浩史・姜銀来（2017）を参考に筆者作成

次に、脳に電磁的刺激を印加する手段に関して補足する。脳への信号入力や電磁的な刺激を印加する場合は、上述した脳の出力信号読み取りと同じ電極が使用できる。但し、それとは別に比較的簡便に非侵襲で脳に刺激を印加する方法として、①経頭蓋磁気刺激（TMS）、②経頭蓋直流電気刺激（tDCS）、③経頭蓋交流電気刺激（tACS）がよく用いられている。①の TMS は、ニューロン間の関連性（シナプスの効力）に作用するとされている。②の tDCS は、一時的にニューロンの興奮レベルを変化させる。具体的には、陽極側の脳の部位の活動を高め、陰極側の脳の活動を低下、抑制する効果を有する。③の tACS は、頭皮の上を流れる電流と刺激される脳の内因的なリズムの共鳴作用がポイントだと考えられている。

3.3 BMI/BCI による脳の活性化・強化に関する研究例

3.3.1 精神疾患や高齢化に伴う認知症の治療

最初に川人グループを中心に検討されている DecNef 法を用いた精神疾患の診断・治療の事例を簡単に紹介する⁴⁵。DecNef 法に関しては 4.1.2 で解説したが、要約すると fMRI 信号から脳・神経活動情報を解読するデコーディング（一種の AI 技術）と、その結果をリアルタイムで脳に戻し脳状態を制御するニューロフィードバックから成る技術である。そして動的脳観に基づき「精神疾患を脳活動ダイナミクスの異常と捉える」といった点が、診断や治療の出発点となっているところがユニークで、東洋医学と類似したアプローチの様に思われ興味深い。成人の高機能自閉スペクトラム症（ASD）を 85%の精度で予測する分類器の開発成功や、同様の AI 技術を大うつ病、統合失調症、強迫性障害、慢性疼痛、等に適用して疾患バイオマーカーを開発した、等の成果が得られつつある。

Durham 大学の研究グループは、脳に tDCS（経頭蓋直流電気刺激）で刺激を与えることで、幻聴の制御が可能となることを示した⁴⁶。ニューロ・イメージングの先行研究では、「聴

⁴⁵ 川人光男（2017）、心身ダイナミクスとニューロフィードバック、心身医 vol.57(5) 414 - 420.

⁴⁶ Peter Moseley, (2014) The role of the superior temporal lobe in auditory false perceptions: A transcranial direct current stimulation study, Neuropsychologia 62 202-208.

覚言語による幻覚 (AVHs) が生じている時に、上側頭回を含む皮質領域のネットワークがアクティブである」ことが確認されており、幻聴の多い統合失調症の患者に tDCS で脳に電気刺激を与えると、幻聴を減らすことが可能であった。本研究では、健常で幻覚を起こさない参加者 (N = 30) に対して、聴覚信号検出タスクを実施し、健常者の脳に電気刺激を与えると、逆に幻聴を起こしやすくなることを見出した。

誰も居ないのに誰かが側にいる、という不気味な錯覚 (幻視) を起こしやすい人がある。Shahar Arzy, et al. (2006) は、この錯覚が、左の側頭頭頂接合部の領域への電気刺激で誘発されることを確認した⁴⁷。てんかん治療のために術前評価を受けていた患者 (発作を記録するため 102 個の硬膜下グリッド電極がインプラントされている) に対して、焦点電気刺激 (0.5 ~ 11.0 mA、継続時間 2s) を 50Hz、バイポーラで印加し、左の側頭頭頂接合部を刺激した結果、これらの奇妙な認識は、側頭頭頂接合部での身体および自己の多感覚処理の障害である可能性が示された。

大阪大学のグループ (2019) は、認知症に対する磁気刺激治療の可能性検討を最近始めた⁴⁸。TMS (経頭蓋磁気刺激) で脳に磁気刺激を反復印加するとうつ病の症状が緩和されることは従来から知られており、治療に用いられていたが、近年、この TMS が認知症にも有効だとする研究結果が国内外で一部出始めている。大阪大学では、アルツハイマー型認知症患者の脳 (前頭前野) を 1 秒間に 10 回磁気刺激し、休止を含めて 15 分間継続する、といった臨床研究を 2019 年 5 月から開始した。本法は非侵襲型の治療である。

ボストン大学のグループ (2019) は、加齢による記憶力 (ワーキング・メモリ) の低下は脳の特定位が分断されて同期不良となるため、と考え、脳に非侵襲の高解像 tACS (経頭蓋交流電気刺激) で電気刺激を印加し、同期の乱れ (リズムの乱れ) の改善を試みた⁴⁹。60 ~ 76 歳の高齢者 42 人と対照群の 20 ~ 29 歳の若者 42 人に、よく似た 2 枚の画像を連続して見せて、異なる部分を指摘させるというテストを実施した。電気刺激を与える前のテストでは、若者のほうが高齢者よりもはるかに正確だった。その後、高齢者には各自の脳のリズムに調整した電気刺激を 25 分間印加してからテストを再度実施したところ、正解率は 20 代とほぼ同じ水準まで向上した。なお、記憶の改善は刺激を加えてから約 50 分間持続した。

3.3.2 脳力活性化、強化

映画「レインマン」で有名になったサヴァン症候群は、飛びぬけた記憶力や部分的に天才的特異能力を発揮することで知られている。脳に電磁的刺激を印加することでこういった特異能力を発現させることは可能か。前項で述べた認知症治療や高齢者のワーキングメモリ改善は、本項の対象である脳力活性化の研究例としてもカウント可能であるが、ここでは多少、切り口の異なる最近の研究例を 2 件採り上げる。

ジョージタウン大学、等の研究グループ (2017) は、非侵襲の tDCS (経頭蓋直流刺激) で脳に微弱な電氣的刺激を与えると、一見関係なさそうな物事を結びつける (zapping) 能力が

⁴⁷ Shahar Arzy, Margitta Seeck, Stephanie Ortigue, Laurent Spinelli & Olaf Blanke (2006) Induction of an illusory shadow person, *Nature* 443, 287.

⁴⁸ 日本経済新聞 2019 年 3 月 18 日朝刊

⁴⁹ R. Reinhart & J. Nguyen (2019) Working memory revived in older adults by synchronizing rhythmic brain circuits, *Nature Neuroscience*, vol.22 820 – 827.

一時的に増大することを確認した⁵⁰。前頭前野皮質は、意味と類似性の結び付けに関与している、とされている。神経機能の調節手段として知られている tDCS で、こういった能力（≡ 創造的推論・洞察）が強化可能かを検討した。電気刺激有グループ、無グループそれぞれに動詞に関するテスト（広範な対象から類似性を見出す）、等を実施したところ、20 分間の電気刺激有りグループの方が好成績を示した。

情報通信研究機構（NICT）と大阪大学、北海道大学の研究グループ（2017）は、無意識のうちに英語リスニング能力を向上できるニューロフィードバック技術の開発に成功した⁵¹。R（right）と L（light）の音を聞いている時に脳波を計測して音の聞き分けに関連する脳活動パターンを取り出し、その大きさを円としてフィードバックする。そして学習者にはその円を大きくするようにイメージしてもらう。その結果、本人は、音の聞き分け学習をしているつもりがなくても、無意識に R（right）と L（light）の音の聞き分けが、5 日間程度で出来るようになった。

3.3.3 脳の信号によるコミュニケーション

1999 年に BMI/BCI の領域では画期的な論文が発表された⁵²。ラットの脳から侵襲的方法で読み取ったニューロンの信号で外部機器の操作が可能、といった報告であり、BMI の先駆けといわれている。この論文の著者の一人でもある Miguel A. L. Nicolelis（米国 Duke 大学教授）は、2011 年の一般向け著作で下記の未来を語っている⁵³：

- ・ 2030 年頃までには、非侵襲で脳の活動を検出する方法が開発される。
- ・ そして、人が脳の活動信号で多様なコンピュータ機器に作用可能となったなら、その後はブレイン・ネットの時代となり、自身の思考とウェアラブル or 体内埋込コンピュータで自由にコミュニケーション可能となる。
- ・ 自身の脳とつながるアバター（分身）を用いた、深海や宇宙の果て、更には細胞内の仮想探検旅行も行えるであろう。
- ・ このように脳が身体から完全に解放されたら、個人と個人の境界線はぼやけ、私達は巨大な集合意識の一部になるだろう。自己と世界のかかわり方が再定義され、人類はその歴史上、新たな時代を迎えることになるのかもしれない。
- ・ 但し、現在夢想されている、脳をそっくりコンピュータにダウンロードするとかコンピュータでシミュレートするといったことは決して実現しないだろう。パーソナリティの本質をディスクへ転送することは出来ない。

近い将来こういったブレイン・ネット時代が到来するか否かは不明だが、最近、BrainNet に関する論文が発表された。米国ワシントン大学とカーネギーメロン大学の研究チーム（2019）

⁵⁰ Adam E. Green, Katherine A. Spiegel, Evan J. Giangrande, Adam B. Weinberger, Natalie M. Gallagher and Peter E. Turkeltaub (2017), Thinking Cap Plus Thinking Zap: tDCS of Frontopolar Cortex Improves Creative Analogical Reasoning and Facilitates Conscious Augmentation of State Creativity in Verb Generation, *Cerebral Cortex*, 27:2628–2639.

⁵¹ 情報通信研究機構（NICT）・大阪大学、2017 年 6 月 15 日プレスリリース、
対応論文：<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178694>

⁵² J. K. Chapin, K. A. Moxon, R. S. Markowitz and M. A. L. Nicolelis (1999) Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex, *nature neuroscience*, vol.2(7), 664 – 670.

⁵³ M. A. L. Nicolelis (2011) Mind out of Body. *Sci. Am.* 304: 80-83.

が EEG（脳波測定）と TMS（経頭蓋磁気刺激）を用いて脳と脳を繋ぐインターフェース・システム（＝BrainNet と称する）を開発した⁵⁴、という。

EEG で送信者の脳波を読み取り、そのデータに基づき TMS で受信者のニューロンを磁気刺激して 眼閃（目を閉じていても光を感じられる現象）を起こし、送信者からの情報を認識させる、といったシステムである。実験では、先ず EEG の電極を付けた 2 人の送信者に、ブロックが画面上から落ちてくる「テトリス」のようなゲームをしてもらう。送信者はブロックを「回転させる」or「回転させない」を決定し、それを受信者に伝えなければならない。画面の両端にはそれぞれ 15Hz と 17Hz で点滅する LED が取り付けられており、それを見つめると周波数に応じた脳波が生じ、EEG で検出される。17Hz を「回転させる」、15Hz を「回転させない」として LED を見つめると、EEG で検出された脳波が、TMS を介して受信者のニューロンを磁気刺激し、眼閃を発生させる。これで受信者はブロックをどう操作すべきかが分かる、というものである。3 人 1 組の 5 グループで実験した結果、受信者は 81.25% の精度で送信者の指示受け取りに成功した、という報告である。

EEG（脳波測定）によるコミュニケーションとなると、産総研の活動も忘れるわけにはいかない⁵⁵。重度運動機能障害者向けの脳波による意思伝達装置（認知型 BMI）として、2010 年には非侵襲での脳波（EEG）のリアルタイム解析に基づくニューロコミュニケーターの試作に成功している。最近のシステムは、ユーザーの選択的注意に対応する「事象関連電位」を小型脳波計測システムと脳内意思解読手法で測定・解析し、PC 画面上の複数のピクトグラムから、ユーザーが選んだピクトグラム（＝ユーザーのメッセージ）をリアルタイムで検出、そのメッセージを CG アバターが音声とアニメで表出する形に進化している。現在は、脳波解読の高速化やロボット・アバター（分身ロボット）の開発、さらに脳と人工知能を融合した「ハイブリッド型 BMI」によるコミュニケーション支援技術の研究が実施されている。

3.4 次世代に向けた考察

今回の予備的調査に基づいて脳科学の進展現状を、冒頭で述べた狙いに対して簡単にまとめると以下の様になる：

- ・ 情報システムとしての脳の構造・機能はミクロなレベルでの解明が急速に進展しつつあるが、まだ不明点は多く発展途上である（たとえば、グリア細胞の機能、構造に関する静的脳観と動的脳観、等）。
- ・ BMI/BCI によるヒトの脳・神経信号の解読は、汎用レベルで高精度（空間・時間分解能）とはまだいえないが、ある範囲で可能であり、ヒトの記憶や意識を部分的に編集・制御することも可能になりつつある（精神疾患の診断・治療）。
- ・ しかし意識生成のメカニズムは未だ不明である（幾つかの仮説提示はあるが）。
- ・ 同様に BMI/BCI によって脳・神経信号で直接、外部機器を操作したり、外部とコミュニケーションしたりすることも、あるレベルでは可能である。

⁵⁴ Nuno R. B. Martins, Amara Angelica, Krishnan Chakravarthy, Yuriy Svidinenko, Frank J. Boehm, Ioan Opris, Mikhail A. Lebedev, Melanie Swan, Steven A. Garan, Jeffrey V. Rosenfeld, Tad Hogg and Robert A. Freitas Jr. (2019), Human Brain/Cloud Interface, Frontiers in Neuroscience, vol.13/March art.112.

⁵⁵ 長谷川良平（2017）Brain-Machine Interface がつくる未来の社会、精密工学会誌, 83(11), 983 - 987.

(侵襲型に比べ、安全性の高い非侵襲型 BMI/BCI の性能は限定される)

- ・ 脳へのニューロフィードバックや非侵襲・ウェアラブルな BMI/BCI で脳に電磁的刺激を印加し、脳を活性化、強化することも可能になりつつある。

以下、こういった現状を踏まえて、次世代 AI の方向や AI と共存する社会について再考し、現時点での見解をまとめる：

(1) AI やロボットは意識を持てるのか？

- ・ 脳の意識生成メカニズムが不明な現在、機械に自我、等の意識を持たせることは可能か？意識の定義や検証方法が曖昧な状況下で議論はし難いが、一応、「なるほど」と思える意識生成の理論（仮説）は幾つか提案されている。こういった状況から、機械に人工意識を付与できる可能性はある、と推測する⁵⁶。
- ・ 但し、ヒトの意識は、生命体としての長い進化の歴史や個としての成長の歴史、多様な社会的経験の上に成り立つものである。高速学習でデータをインプットした生命体ではない機械の人工意識は、ヒトの意識とは基本的に異なるもの、と捉える方が自然であろう（e.g. 寿命観を含む時間感覚は根本的に異なる？）。

(2) マインド・アップロードは可能なのか

- ・ BMI/BCI で脳・神経信号出力の解読や逆の入力（人工内耳・人工網膜）は、確かに一部可能で実用化もされている。したがって脳の発信する情報、記憶された知識、等を部分的に AI（コンピュータ）に移植することは可能であろう。
- ・ 日本でも渡辺正峰（東大、MinD in a Device）が、20 年後を目標に数千万個のニューロンと通信可能な BMI を作製し（DARPA の現目標は 100 万個）、意識アップロードを狙って研究していることが、知られている⁵⁷。
- ・ しかし、形式知はともかく暗黙知は転送できるのか、情動・感性は処理できるのか。情動（後述のシステム 1 思考）は進化的に古い脳（≡脳の深部）の影響が強いが、深部情報を表層部で高精度に読み取れるのか。また移植した情報や知識の断片を、脳と同様に有機的に再構成、再出力可能か、と疑問は多い。また、たとえば直径 10 μm の電極・配線 5,000 万本を密着状態で束ねると、その底面は 5cm×10cm 程度のサイズとなる。これだけの配線を、密着状態ではなく、欠陥の無い（或は許容可能な最小欠陥レベル）状態で、個体差のある脳に対して精密設置するという状況は、開放環境で物質をハンドリングするインクジェット・ヘッドの如きデバイス技術に比較的近いと思われるが、そういったデバイス技術の常識を超える高難度に見える。そして侵襲型で脳内に設置する場合、ヒトに対する負荷は高く安全リスクも高い。因みに米国 Neuralink 社が 2019 年 7 月に発表した初期的な BMI を 19 匹の動物に対してロボット手術で設置する実験では、成功率が 87% だった⁵⁸、とされている。極めて限定的なデータ・アップロードは可能としても、全脳レベルの

⁵⁶ 人工意識に対する国内の取組例としては、たとえば（株）アラヤの金井良太が主導する「神経科学の公理的計算論と工学の構成論の融合による人工意識の構築とその実生活空間への実装」がある。（CREST 2015 年採択）

⁵⁷ たとえば、日経 Xtech（2018/9/19）<https://tech.nikkeibp.co.jp/atcl/nxt/mag/ne/18/00017/00007/>

⁵⁸ <https://www.technologyreview.jp/nl/elon-musks-neuralink-says-its-nearly-ready-for-the-first-human-volunteers>

マインド・アップロードは実質的に不可能と推測する。

(3) 我々は、どんな「次世代 AI」を望んでいるのか？

- ・ 最初に、ヒトの思考プロセスを概観すると、心理学や行動経済学の領域では、ヒトの思考を、一般にシステム 1 とシステム 2 の二重過程理論で捉えている。

表 3-1 二重過程理論の概要 ⁵⁹

システム1思考	システム2思考
速い	遅い
自動的・無意識的	制御的・意識的
情動的・直感的	合理的・論理的
非言語的	言語的
努力不要	努力が必要(→疲労する;処理能力の制約)
一般的知識と無関係	一般的知識と密接に関係
主に短期的な利益追求	長期的な利益を勘案
進化的に古い	進化的に新しい

- ・ このマクロな二重過程理論をベースに考えると、システム 2 思考（≡アルゴリズム的）は。ヒトにとっては努力が必要で弱点である場合が多いが、機械（AI）との親和性は高い、と考えられる。ならば、ヒトに役立つ道具としての次世代 AI は、ヒトのシステム 2 思考を補強する様な AI ではないだろうか。このイメージは、汎用 AI（AGI）と呼ばれるものよりも、むしろ特化型 AI やコンピュータに近いのではないだろうか。
- ・ ビジネスが今後の AI 研究開発をリードする、という観点からは、第 2 世代、第 3 世代 AI 技術およびこれらの改良・進化系が、八百万の特化型 AI として試行錯誤され、それらの競争の中から創発的に次世代 AI の方向が定まっていく、といった“見慣れたシナリオ”も有力であろう。この場合、ヒトと AI の分業構造デザインが、ビジネスや産業政策のひとつのポイントになる、と思われる。
- ・ 脳科学の成果は、次世代 AI の研究開発にも示唆を与えるであろうが、ヒトの考え方や脳の活用法にも当然、多くの示唆を与える。AI の追究だけでなく、非侵襲型でウェアラブルなツールによるヒトの脳機能の活性化、強化、拡張といった研究開発の方向をより重視すべきであろう。ここで AI は脳・神経信号の解読精度向上手段、等として期待される。4 章で抽出される次世代の方向としては「脳科学と AI の成果を展開して、ヒトの脳機能活性化、強化、拡張を図る」が、最重要メッセージかもしれない。PC が立ち上がりつつあった数十年前に、「AI か IA か（Intelligence Amplifier）」といった議論があったが、現在も、AI にシフトし過ぎず IA の立場での状況判断が必要と感じる。

⁵⁹ ①D. Kahneman (2011) “Thinking Fast and Slow”, ②阿部修士 (2017) 「意思決定の心理学」を参考に筆者が作成

- ・ 最近の脳科学の進展に関する理解や AI 実用化の動向と限界の認識は、ヒトの考え方に影響し、従来よりも「脳と心」が注目されるのではないだろうか。特に先進国では、東洋的思想・スタイル (e.g. 瞑想、仏教、ヨガ、禅、内観、悟り) や東洋医学 (e.g. ツボと経絡、気) の見直し、現代的再解釈の動きが強まる、と推察される。一般的な動きとしては、筋肉系の身体トレーニング・ジムから脳や心の調整や鍛錬を意図したマインド・ジム、ブレイン・ジム、心身トレーニング・ジムへ、といったイメージである。ジムの場合は、自然体で行うケースと非侵襲・ウェアラブルな簡易 BMI/BCI を補助ツールとして使用するケースの両方が考えられる。数年前から盛んになりつつある米国発のマインドフルネスも、そういった一環として捉えられるかもしれない。そして、専門家の世界では、新たな科学や哲学の体系探索に影響を及ぼす、と推測する。
- ・ 今回の予備的調査では、ヒトの脳機能の活性化、強化、拡張ツールとして専ら BMI/BCI に限定したが、VR/MR/AR 技術や分身ロボット、等もヒトの脳力拡張に有効なツールと考えられる。
- ・ 前述のシステム 1 思考、システム 2 思考に話を戻すと、進化の過程で早期に刷り込まれた部分が多いと思われるシステム 1 思考 (≡ヒューリスティクス) は、ヒトを特徴づけ利点も多いが、間違いの誘発因でもある。そして先史時代ならともかく現代の社会環境にはミスマッチなバイアスを生む要因でもある。近年、心理学の分野でも、脳科学的手法 (e.g. fMRI による脳活動分析) を取り込んだ心理学実験が盛んになってきている⁶⁰。それら成果をニューロフィードバック法や VR 技術を活用したトレーニングに展開すると、システム 1 思考の現代的なチューニングもある程度は実現可能と思われる。同様の手法は、〇〇〇〇依存症、等の精神疾患や犯罪抑止、等にも展開可能であろう。
- ・ 勿論、治療的用途であれ、脳力増強的用途であれ、安全と倫理の議論と制度整備は必須である。
- ・ 話を AI に戻す。汎用 AI (AGI) の実現可能性は不明だが、汎用 AI (AGI) が仮に出来たとして、我々はこれをどう使うのだろうか？個人的にはここがしっくり来ない。但し、脳や知能の研究といった観点からは、汎用 AI を志向する全脳アーキテクチャの如きアプローチは意味がある、と考える。数理モデル、等価回路、アルゴリズム、等の工学的アプローチで、どこまで脳や知能のメカニズムに迫れるのか、どこにギャップがあるのか。こういった追究は興味深く、新たな哲学と科学のインキュベータとしての可能性もある。

⁶⁰ 初期の著名な論文としては : Joshua D. Greene, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, Jonathan D. Cohen (2001) An fMRI Investigation of Emotional Engagement in Moral Judgment, Science, Vol. 293, Issue 5537, pp. 2105-2108. DOI: 10.1126/science.1062872.

4. 脳科学の知見を反映した AI 模擬事例

4.1 AI チップ（ハードウェア）の現状

現状研究開発中の AI チップ(ハードウェア)は、脳科学・心理学の成果を反映していると言いつつ、ニューロモルフィックチップに代表される脳型チップには、未来研の考える「人間活動システム」を模擬できるアーキテクチャは未だ組み込まれていない。データ処理速度向上、消費電力低減を目的とした取り組みが主である。本項では、第一に、これまでの人工知能の進化を振り返りつつ AI チップ（ハードウェア）研究の現状をまとめ、第二に、近年著しい進展をみている脳科学の知見の現状をまとめ、第三に、これら脳科学の知見をベースとした次世代 AI チップ（ハードウェア）の実装のコンセプトを探る。方向性の一つとしては、自律分散制御を前提として、実空間（無限定空間）環境にて自らの「身体」に基づく「知」⁶¹ をベースにしたシステムの構築、即ち、外界、身体、脳 と一体になった同期（シンクロ）を発現できるシステム構築を視野においている。

4.1.1 人工知能の進化と研究開発の変遷⁶²

人工知能の進化とこれまでの研究開発の変遷を図 4-5 に記す。人工知能の概念は、1947 年にアラン・チューリングが提唱したとされ、その後 1956 年の「ダートマス会議」にて初めて「人工知能」という言葉が使われた。1960 年代の第一次ブーム（探索・推論の時代）、1980 年頃～1990 年頃の第二次ブーム（知識表現の時代）を経て、コンピュータの情報処理能力向上も背景としながら現在の（機械学習、表現学習の時代）に至る。

4.1.2 人工知能を支えるハードウェア^{63,64}

現在一般に使われているコンピュータは、ノイマン型コンピュータ（フォン・ノイマンアーキテクチャ）であり、演算を行うプロセッサと、データやプログラムを記憶するメモリ（主記憶装置）が別チップで構成され、両者を「バス」配線で接続する。そのプロセッサ／メモリが高速化・低消費電力化された為、この「バス」部分での配線容量の充放電による通信帯域の制約や消費電力の増大が顕著になってきた（フォン・ノイマンボトルネック）。その為、プロセッサと専用キャッシュメモリを組みにしたコアを多数用意するマルチコア（メニーコア）な



図 4-1 人工知能の進化と研究開発の変遷⁶²

⁶¹ 松田 雄馬「人口知能の哲学」東海大学出版会（2017）

⁶² 総務省資料「次世代人工知能推進戦略」 http://www.soumu.go.jp/main_content/000424360.pdf

⁶³ 森江 隆：応用物理 88, 481(2019)

⁶⁴ NEDO 資料「人口知能を支えるハードウェア分野」 <https://www.nedo.go.jp/content/100885736.pdf>

どのアーキテクチャが開発されている。こうしたコアを多数集めてインターコネクで接続したのが、現在のスーパーコンピュータであり、CPU や GPU を用いたディープラーニングによる AI コンピュータである。これらは、CMOS 半導体チップにより実現されており、微細化によるスケールン則（ムーアの法則）により性能向上が計られてきた。昨今、このスケールン則が限界に近づいており、代替技術の必要性も高まっている。一方、ノイマン型と異なる演算手法やアーキテクチャを特徴とする非ノイマン型コンピュータも研究開発されている。プログラムの逐次実行ではなく、必要データをそろえて演算を行うデータフローコンピュータ、動作時のハードウェア再構成が可能なリプログラブルコンピュータなど CMOS 半導体を用いたもの、更には全く異なる動作原理である、脳の神経回路網の働きを模倣した脳型（ニューロモルフィック）コンピュータ、量子力学的特性を応用した大規模な超並列情報処理を可能にする量子コンピュータ、などがある。

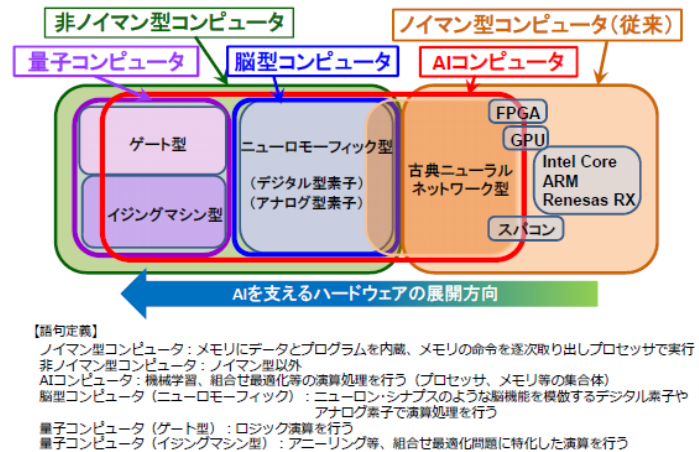


図 4-2 ノイマン型/非ノイマン型コンピュータ 64

図 4-6 に、ディープラーニングをサポートするコンピュータを AI コンピュータと定義し、ノイマン型、非ノイマン型コンピュータという括りで図示してある。このうち、脳型コンピュータの現状に関して以下に概観する。

脳モデルに近い計算を行う脳型（ニューロモルフィック）コンピュータは、脳のニューロンの模倣を、デジタル型素子で行うか、アナログ型素子で行うかで大別される。デジタル型素子では、ニューロンへの入力 of 積算が閾値を超える（発火）と次のニューロンへと情報が伝達されて行き、結果は学習情報としてメモリに格納される。一方、アナログ型素子では、ニューロンへの入力となるアナログ電圧の比較で発火に至り、その学習情報は、プログラマブルな抵抗素子特性で示されるものもある。

	古典ニューラルネット	脳型コンピュータ	
演算処理	デジタル	デジタル型素子	アナログ型素子
演算素子	CMOS	CMOSベース	メモリスタ、ReRAM、相変化メモリ、スピントロニクス、etc.
消費電力 (相対比(例))	高 (100000)	小 (10)	極小 (1)
問題記述	プログラム	結合強度 (オフチップ学習)	結合強度、しきい値など (オンチップ学習)
開発フェーズ	商用チップ〜システム開発	素子〜システム開発	素子〜チップ開発
原理イメージ	論理回路 	ニューロン・シナプス模倣 	(アナログ素子例)

図 4-3 脳型コンピュータ全体像 64

4.2 神経伝達、脳の機能と情動の脳内メカニズム

4.8.にて AI チップの現状ハードウェアに関して記述してきたが、次世代型を構想するに当たり関連する神経伝達の仕組み及び脳の機能に関して概観する。

4.2.1 神経伝達の仕組み

生体の神経細胞（ニューロン）は、細胞内にカリウムナトリウムなどのイオンがチャネルを通して出入りすることで、膜電位が変化する。膜電位とイオン電流の関係は、非線形特性をもち、ある閾値を超えると「スパイク」と呼ばれる、膜電位パルスを出力する（「発火（fire）」）。ニューロン同志はシナプスを介して接続されており、ほかのニューロンからスパイクが入力されると、シナプスでの結合重みに応じた影響（シナプス後電位）がニューロン内部で生成され、時空間的に加算される⁶³。

4.2.2 脳の機能について

(1) 脳の部位と役割

人間の脳は思考や行動、生命維持など、人間が生きていくうえで必要な事柄の全ての司令塔であり、脳部位によってそれぞれ請け負っている役割が異なる。人間の脳は、大脳、間脳、脳幹（中脳、橋、延髄）、小脳の4種類の領域に分類される。ここでは、大脳、辺縁系、基底核、前脳基底部、間脳、脳幹、脳室、小脳について概略を紹介する。

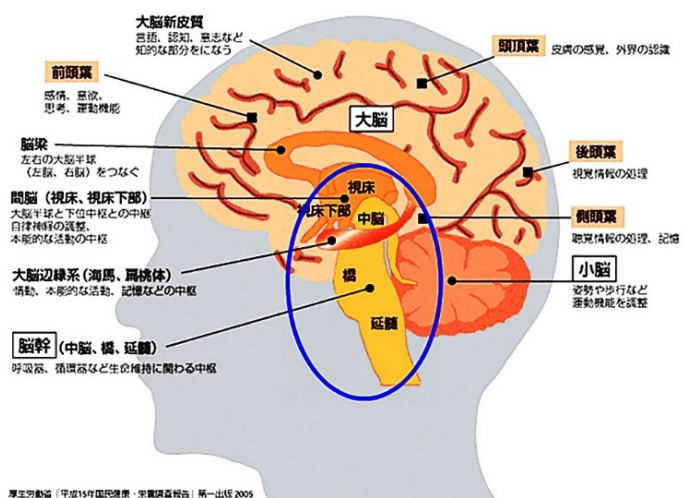


図 4-4 脳の部位と役割①

(1)-1 大脳⁶⁵

大脳は、「大脳皮質（灰白質）」と「髄質（白質）」から構成されていて、「大脳皮質」は新皮質（前頭葉、頭頂葉、側頭葉、後頭葉）、大脳辺縁系（古皮質—海馬、脳弓、帯状回）、旧皮質（嗅葉、梨状葉）に分類される。「髄質」の深部には大脳基底核が存在する。新皮質は帯状回よりも上の部分を指し、左右の大脳は脳梁によって連絡されている。

(1)-1-1 大脳皮質

① 前頭葉

前頭葉には前頭連合野、前頭眼野、運動連合野、運動野、ブローカ野が存在する。

前頭連合野は、他の中枢と言われる場所（各種連合野、言語中枢）から送られて来る情報を元に、どうするべきかを判断し、実行する総合中枢である。

⁶⁵ 出典：管理薬剤師.com 大脳皮質、辺縁系、基底核 より <https://kanri.nkdesk.com/hifuka/sinkei2.php>

認知力・思考力・注意力・集中力等に関わると共に、ワーキングメモリというパソコンで言うメモリに相当する作業記憶領域が存在する。ワーキングメモリは、「今どうするか」という数分間程度の記憶を保持するスペースである。こうした記憶は5分ほどで忘れてしまう。

運動連合野は、運動神経の中核であり、ここで指定された運動を行うのが運動野ということになる。

ブローカ野は、話す、書くといったアウトプットの言語中枢であり、逆に聞く、読むというインプットの言語中枢が側頭葉にあるウェルニッケ野となる。

② 頭頂葉

頭頂葉には体性感覚野、頭頂連合野、味覚野が存在する。

体性感覚野は、触圧覚や温痛覚を知覚神経を通じて判断する。頭頂連合野は、主として空間認識の中核である。距離感、上下感覚などを司る。体性感覚の情報等を処理して前頭連合野へと送る。

③ 側頭葉

側頭葉には聴覚野、側頭連合野、ウェルニッケ野が存在する。側頭連合野は、音、形、色といった聴覚野や視覚連合野から送られてくる情報を処理して、前頭連合野へと送る。

④ 後頭葉

後頭葉には視覚野、視覚連合野が存在する。

(1)-1-2 大脳辺縁系

大脳辺縁系は、古皮質（海馬、脳弓、歯状回）、旧皮質（嗅葉、梨状葉）、中間皮質（帯状回、海馬回）、皮質下核（扁桃体、中隔、乳頭体）の総称である。

大脳辺縁系の大きな役割は、記憶（短期記憶と長期記憶）と情動（やる気、怒り、喜び、悲しみ等の快不快）である。中でも特に重要なのが、海馬と扁桃体である。

① 海馬

海馬の役割は、短期記憶の保持と長期記憶の固定である。記憶を司る海馬は、目、耳、鼻（日常生活）から記憶のもととなる情報を全て海馬に一度集められ、最

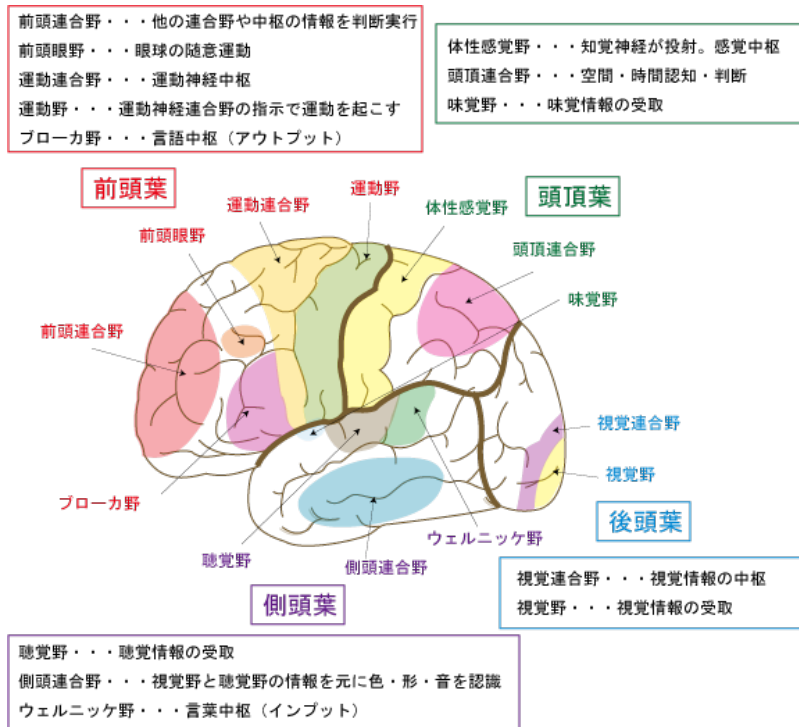


図 4-5 脳の部位と役割②

終的には海馬によって大脳新皮質で永久保存される。海馬は記憶の定着に重要な役目を果たす⁶⁶。

今日食べたものを覚えていたりとかアルツハイマーの診断に使うような記憶情報は海馬に短期記憶として蓄えられている。この短期記憶の中でも特に重要だと判断されたものは、各連合野へと移動して、例えば空間感覚情報は頭頂連合野、色とか形情報は側頭連合野と言った具合に各部位にそれぞれ長期記憶として蓄えられることになる。

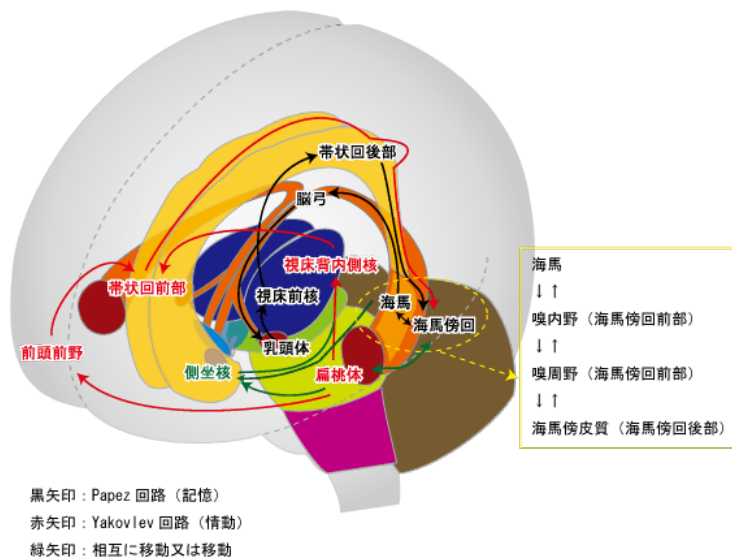


図 4-6 記憶回路及び情報移動

記憶の回路としては Papez 回路（海馬から始まって、脳弓→乳頭体→視床の前核群→帯状回後部と進み、必要な記憶は大脳皮質連合野へ、そうでなければまた海馬に戻り繰り返す）が有名。

長期記憶は、10 年前のあの時に、こんな感覚で歩いたり（頭頂連合野）、こんな景色や音（側頭連合野）だったなあという古い記憶を呼び起こして、イメージ化する（前頭連合野）ような感じである。

② 扁桃体

扁桃体の役割は、海馬からの視覚、味覚というような記憶情報をまとめ、快/不快（好き嫌い）を判断する事である。

ここで、何かの行動が快不快感情を生み、その情報を海馬へと送るというように、海馬と扁桃体は常に情報をやり取りし海馬傍回がその取次として働いている。

・海馬の構造

海馬傍回前部にある嗅内野を介してすべての大脳皮質連合野からの情報を海馬（CA3、CA4、※CA3 は海馬の錐体細胞層の一部、CA4 は歯状回の顆粒細胞層にあたる）に投射している。

顆粒細胞（CA4）の軸索で CA2 に達しており、錐体細胞（CA2,CA3）の軸索の枝で、CA1 に結合する。海馬からは、CA1 および海馬台から嗅内野を介して大脳皮質連合野に投射、CA1~3（主に CA3）から海馬采・脳弓を介して乳頭体や中隔核に投射する。嗅神経→嗅球→嗅索→後端で内側嗅条・外側嗅条に分かれて嗅三角

⁶⁶ 出典：Akira Magazine <http://www.akira3132.info/Brains.html>

(嗅球・嗅索・嗅三角を合わせて嗅葉という。) から投射される嗅覚野はこの嗅内野付近に存在している。

情動を司る回路としては Yakovlev 回路 (扁桃体から始まり、視床背内側核→帯状回前部、前頭前野→扁桃体に戻る) が有名。

・側坐核

側坐核 (腹側線条体) は帯状回前部に存在し、前頭連合野と連絡を取りあう。

海馬や扁桃体は側坐核へも投射していて、側坐核からは腹側淡蒼球に GABA 出力して、腹側淡蒼球からは視床背内側核へ→Yakovlev 回路に入る。扁桃体の興奮 (快反応) が引き金となって起こる中脳腹側被蓋野 (A10) からのドーパミン入力も側坐核へ投射していて、前頭前野→快楽、嗜癖行動を引き起こす。側坐核からの出力先は黒質や橋網様体がある。

(1)-1-3 大脳基底核

大脳基底核は大脳髄質内の核群で、尾状核、被殻、淡蒼球から構成され、主に錐体外路運動 (不随意運動の調節、筋の緊張の維持等) に関わっている。大脳基底核は、大脳皮質と視床、脳幹を結びつけている神経核の集まりである。大脳は基本的に外周部が灰白質 (ニューロンの細胞体がある場所) である場所が多いものの、大脳の深い所にあるのにもかかわらず大脳基底核は灰白質である。哺乳類の大脳基底核は運動調節、認知機能、感情、動機づけや学習など様々な機能を担っている⁶⁷。

大脳基底核は大脳髄質内の核群で、尾状核、被殻、淡蒼球から構成され、主に錐体外路運動 (不随意運動の調節、筋の緊張の維持等) に関わる。

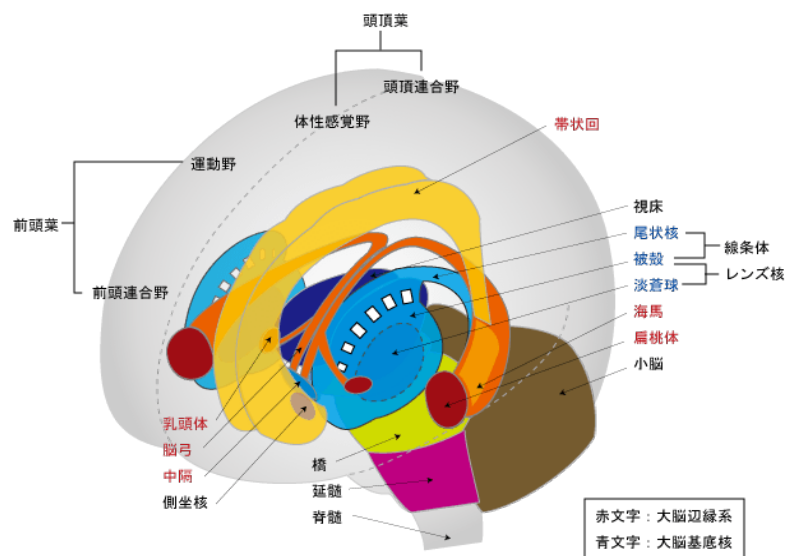


図 4-7 大脳辺縁系と大脳基底核⁷

⁶⁷ 出典 : Wikipedia

4.3 次世代型 AI チップ（ハードウェア）へのパラダイム変換（「人間活動システム」の模倣）へのアプローチ

4.3.1 機械学習の演算処理⁶⁸

機械学習の演算処理は学習と推論から成り立ち、学習は多くのデータより傾向を抽出し、作業精度を高める処理、推論は新たな未知のデータに対し、学習結果を基に答えを導き出す処理である。図 4-12 にニューラルネットワークの概念図を示す。生物の神経回路網の構成、動作を模倣したものであり、図 4-13 で示される多数の人工ニューロンより成る。入力層、隠れ層、出力層などが階層化されており、基本的にニューラルネットワークの演算処理は、多入力の積和演算である。学習は、まず入力層から出力層へ演算を進め（フォワードプロパゲーション）、その後、出力結果と正解を比較し、誤差を求める。次にこの誤差を用い、出力層から入力層へ演算を逆に進める（バックプロパゲーション）。そしてこれらの結果から重みの値を更新する。これら 3 つの動作を繰り返し、誤差を減らしてゆく為、莫大な演算処理を必要とする。一方、推論は入力層から出力層への演算である。

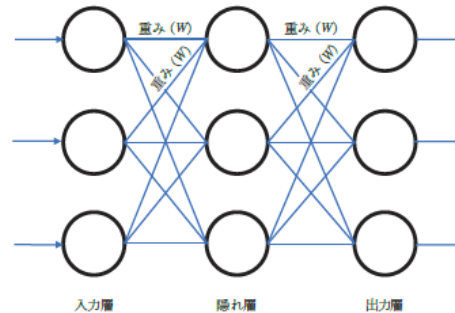


図 4-8 ニューラルネットワーク概念図⁶⁸

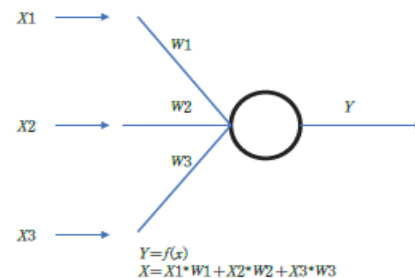


図 4-9 人工ニューロンモデル⁶⁸

4.3.2 デジタル AI チップからアナログ AI チップへ⁶⁹

機械学習演算高速化の為に、並列分散処理で積和演算をする試みが為されている。この特性を生かしたチップとして GPU に代表されるシリコン CMOS プロセス技術を用いたノイマン型アーキテクチャが現在主流である。もともとゲームなどでのグラフィック演算処理に特化したチップである為、浮動小数点演算処理ユニットが実装されている。これはそのまま機械学習の仕様とも合致した。近年の機械学習アルゴリズムの進歩により、多少精度を落とすことで、その値を保持するレジスタやキャッシュメモリ容量も削減でき、回路実装面積縮小、チップ消費電力低減も可能となってきた。このように、プロセッサとメモリが元々別チップであり、通信帯域制約や消費電力の増大という課題を抱えていたノイマン型は改善されてはきたものの、今後の更なるハードウェア高性能化、低消費電力化などの要求に答える必要があり、且つムーアの法則の限界にも対応しなくてはならない。

そこで登場してきたのが、重み表現に不揮発性メモリ素子を用い、積和演算をアナログ回

⁶⁸ 細川 浩二：応用物理 88,357(2019)

⁶⁹ AI チップが開く新しい情報処理のパラダイム

https://www.tel.co.jp/museum/magazine/015/report02_03/04.html

路で実現する、非ノイマン型アナログ AI チップである。こうしたアナログ AI チップは、積和演算をオームの法則とキルヒホッフの法則により実現する点が特徴である。

以下に不揮発性メモリを用いた、所謂インメモリ回路の研究例を挙げる。従来開発されてきた代表的 AI チップである「TrueNorth」(IBM 社開発) は、SRAM を記憶素子に使ったデバイスであるが、これに対し電源を切っても記憶の

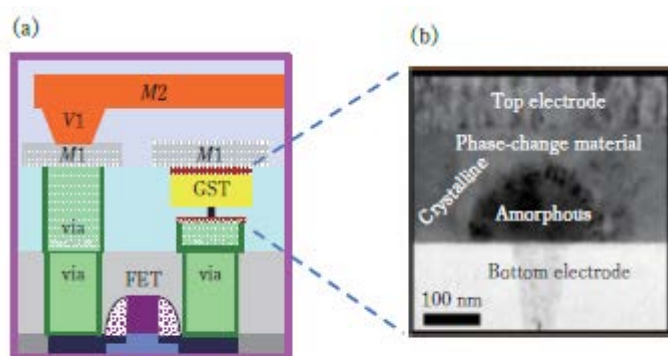


図 4-10 (a)相変化型メモリ素子構成図 (b)相変化メモリ素子⁶⁹

消えない不揮発性メモリである抵抗変化型メモリ (ReRAM) を用いたのが、ユタ大学と Hewlett-Packard 社が共同開発している脳型チップ「ISAAC」である。日本でも、産業技術総合研究所やパナソニックセミコンダクターソリューションズなどが ReRAM を用いたデバイス開発中である。一方、IBM 社は新たに相変化型メモリ (PCM) の活用を検討している (図 4-14)。

不揮発性メモリの採用は、低消費電力化のみならず、フレキシブルな学習機能を搭載できる可能性もある。既述のとおり、機械学習の演算処理には、学習と推論があるが、前述の「TrueNorth」は推論処理しか実行できず、あらかじめ学習済みパラメータをチップに移植して、推論処理を実行せねばならない、という不具合がある。

これら不揮発メモリ搭載したチップでは、外部からのスパイクパルス信号の頻度により、メモリ素子の抵抗値が変化する特性を利用する。これまでは、シナプスの結合強さをデジタル値で記憶しているが、これをアナログ値である抵抗値の違いで表現する事で、メモリ素子を通る電流値でアナログ回路的に積和演算を実行できるようになる。この構造、原理は脳内のニューラルネットワークに類似しており、記憶と演算が一体化した、インメモリ回路構成となる (図 4-15)。こうして推論処理を実行する脳型チップ自体に学習処理機能を付加できる。

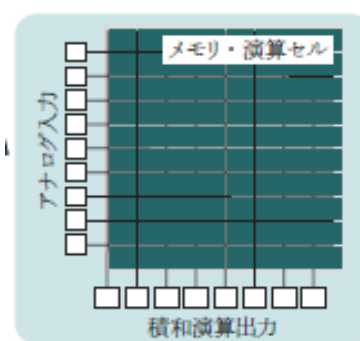


図 4-11 イン・メモリ回路図⁶³

4.3.3 次世代型 AI へのパラダイム変換～「人間活動システム」の実装に向けて

本研究は、現世代のディープラーニングを可能にする AI チップ (システム) から、「人間活動システム」を模擬し、個別の人間を支援できる次世代型 AI チップ (システム) 具現化の為のコンセプトを創出することを目的とする。個別の人間に寄り添う為には、脳科学、認知科学の世界で研究されている自己と他者の感情の共感のメカニズムを理解する事が先決となる。その為に、次世代 AI チップの候補と成りうる最新の研究を紹介した後、個別の人間の記憶を司るハードウェアコンセプトとは何か、そしてこうした個別の記憶をベースにした、感

情の共感（脳活動における共鳴）を可能にするハードウェアコンセプトとは何か、という手順で考察をすすめてゆきたい

(1) 世代 AI チップに関する最新の研究～アナログ AI チップ（モデル）からスパイクニューロンモデルへ⁶³

4.8.2 で紹介したアナログ AI チップは、アナログニューロンと呼ばれる、脳の神経細胞とその結合回路を単純化したモデルが使用される。一方、ニューロモルフィックでは、神経細胞がスパイクパルスを出力するなど、より忠実に脳の神経細胞のふるまいを真似るモデル（スパイクニューロン）を用いる事が特徴である。

図 4-16 は、1980 年代の第 2 次 AI ブームまで盛んに研究されてきた、現在のディープラーニングで一般的に用いられている、アナログニューロンモデル（発火頻度モデル）であり、発火頻度を表現するアナログ値で入出力を表し、

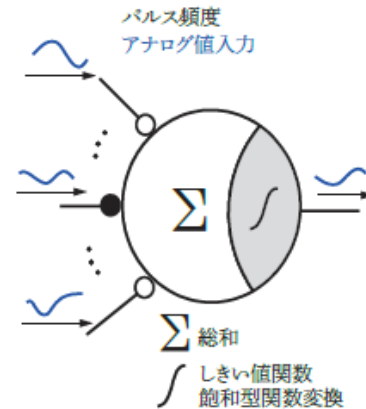


図 4-12 アナログニューロンモデル⁶³

図 エラー! 指定したスタイル

は使われていません。-3-4 イ シナプスも単純な重み付けで表現する。

一方、スパイク一つ一つに意味がありそうだという、スパイクタイミング依存シナプス可塑性（STDP : Spike-Timing Dependent Synaptic Plasticity）が発見されたのち、考えられたモデルが、図 4-17 に示すスパイクニューロンモデル（積分発火型）である。スパイクパルスによるシナプス後電位が特定の時定数で減衰する間に、多くの入力スパイクで加算されて、発火閾値を超えるとスパイクを出力するというモデルである。図 4-18 に STDP の概念図も示す。前段ニューロン(PRE)と後段ニューロン(POST)の間のシ

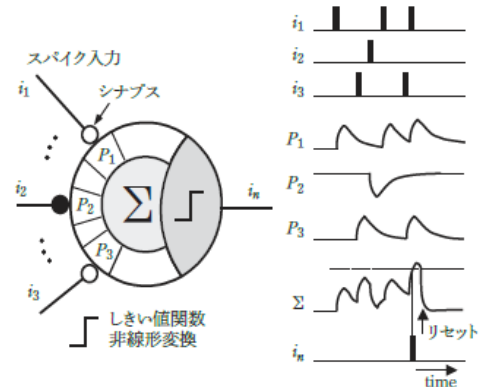


図 4-13 スパイクニューロンモデル⁶³

ナプス結合の重みの変化が夫々の出力スパイクタイミング差に依存するという関係を持つ。非対称型では PRE スパイクと POST スパイクの前後関係で結合重みが増減し、因果関係の学習に役立つ、一方対称型は、スパイク時間差の絶対値に依存し、相関関係の学習に役立つ、とされている。これを用いたネットワークを、スパイクニューラルネットワーク (SNN) と呼ぶ。スパイクタイミングを考慮する事で、時間領域での演算の組み込みが可能となり、より高度な情報処理が可能になると期待されている。現状の実用的 AI 処理にはこの SNN モデルまでは必要無いと思われるようだが、脳をより模倣した次世代の脳型 AI チップに関しては、スパイク同士の同期性が重要になる可能性があり、時間領域の演算機能の重要性

故に、この SNN が次世代のひとつの候補になる可能性がある。

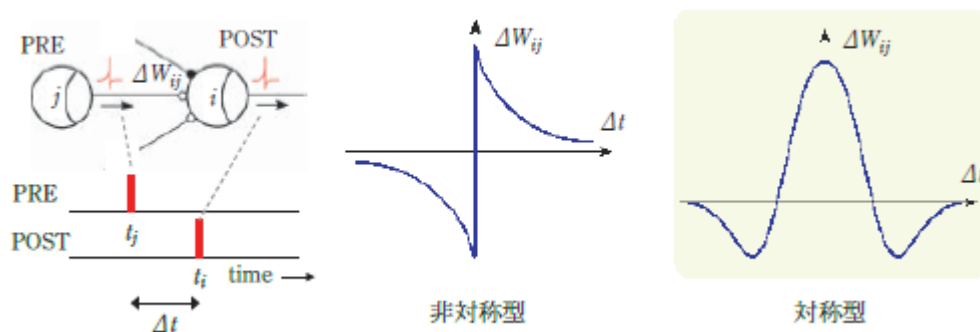


図 4-14 スパイクタイミング依存シナプス可塑性(STDP)の概念図 ⁶³

(2) 個別の人間の記憶を司るハードウェアとは何か

(1)で記載したハードウェアをベースとして、海馬を中心とした脳型処理モデルのデバイス実装の試みの一例を紹介する ⁶³

現在のディープラーニングベースの AI は、大量のデータから万人共通の認識機能を実現する技術である。しかし、人はその誕生以来様々な経験を重ね、その人に固有の記憶を脳内に形成している。この能力を AI に付与する事で、個々人の経験と記憶をサポートする次世代 AI を開発する試みが為

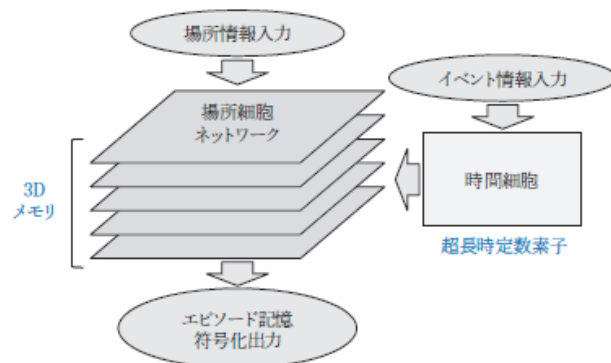


図 4-15 脳型記憶・海馬モデル ⁶³

されている。実際に、既存 AI モデルをプロセッサとみなしたときに、メモリに相当する装置をシステムに追加する提案もされている。森江らは、ブレインモルフィックの立場から、脳の記憶機能を司る海馬とその周辺部位（嗅内物質）のモデル化とハードウェア化の研究を進めている。その為には、外界情報をエピソード記憶（個人の体験に関する記憶）として、符号化・保持・統合・想起する機能が必要。自分が今どこにいるかを「場所細胞」*ネットワークで表現し、そこで何かのイベントが起こった際、感覚系から入力された情報に応答して、イベント間の時間遅れを検出する「時間細胞」と導入し、エピソード記憶を符号化・記憶するハードウェアを構築する。その為には、現状実用化の手前まできている高密度 3 次元積層メモリ技術、更には、秒オーダー以上の超長時定数実現が必要とされる（図 4-19）。

*)「場所細胞」とは、1970 年代に神経生理学者のジョン・オキーフ博士により発見されたもので、視覚/聴覚の為の脳の感覚野があるなら空間認識の為の脳部位も存在するはずだと考え

た⁷⁰。カントの「純粹理性批判」中の「空間はアプリオリに脳の中で認識される」との提唱に触発されたものである。カントは、人間の心には、あらかじめ空間を直感的に認識できる能力がある、我々の外にある空間とは、我々の心が作るものであり、どのように世界を捉えるのかは我々の認識そのものにかかっている、と考えた。

(3) 感情の共感（脳活動における共鳴）を可能にするハードウェアコンセプトとは何か

感情を作る脳の仕組みとは何か、そして他者の感情を知る仕組み（共感する脳の仕組み）とは何か、に関して、脳神経科学、認知科学の観点より考察を深めた乾氏らの研究⁷¹は、今後の次世代 AI ハードウェア

の実装の観点からも注目に値する。まず運動指令に関する脳内の予測制御に関して、図 4-20 に概念図を示す。脳から運動指令が出てから自分の感覚フィードバックを使えるまで 200~300msec 程度の遅延が生じる。更に神経系は様々なノイズを受け更に遅延が増大する。そこで予測制御が必要となる。自分の手を動かす際、自分の手の内部モデルを脳内にコピーし

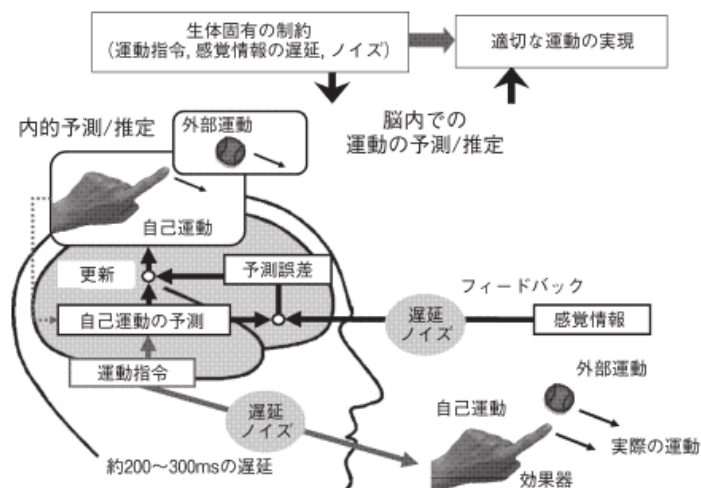


図 4-16 運動指令に関する脳内の予測制御の概念図⁷²

て、こう動くはずだ、と無意識にモニターしている。更に精度を上げる為、実際に手を動かした結果生じる感覚フィードバックで予測の正しさをチェックする。ここで予測との差分を取り、誤差最小となるように予測を更新する⁷²。

「情動」とは、外的刺激や内的な記憶の想起に伴って個体に生じる生理的な反応（他者により観察や計測が可能）、「感情」とは、情動の発生に伴う主観的な意識体験である（本人にしかわからない）というダマシオの定義⁷³を前提とし、乾氏は⁷¹、感情がつくられる仕組みを明らかにした。感覚には、内受容感覚（内臓感覚）、外受容感覚（五感）、自己受容感覚（運動感覚、平行感覚）の3つがあるとしたうえで、感情は、内臓の状態を知らせる自律神経反応を脳が理解すること、とその反応が生じた原因の推定、という2要因により決定される、と主張する。

他者の感情を知る仕組みに関しては、ミラーニューロン（下前頭回及び下頭頂小葉に存在する運動ニューロン）が深く関わるとする。他者の行為を、鏡を見ている自分の行為のよう

⁷⁰ 理化学研究所「つながる脳科学」講談社（2016）

⁷¹ 乾 敏郎「感情とはそもそも何なのか」ミネルヴァ書房

⁷² 乾 敏郎 第107回日本精神神経学会誌 114巻2号, 171（2012）

⁷³ ダマシオ, A.R. 田中三彦（訳）「無意識の脳 自己意識の脳 一身体的情動と感情の神秘」講談社（2003）

に反応し、視覚、聴覚とも運動指令信号が統合されている。また、コミュニケーション場面にある、行為者と観察者の脳の同じ部位が活動し共鳴状態になっている事が実際に観察されている。共感に関して、like-me システムが提唱されており、自己と他者の同一視する事で、観察した他者の動作からその人の行為の意図や感情をオンラインで推測するものであり、これにはミラーニューロンシステムが関与している。他者の動作の視覚情報の処理に関わる、後部上側頭溝が下頭頂小葉と相互結合し、更に下頭頂小葉が下前頭回と相互結合する事で、他者の動作を自分自身の運動に対応させる事ができ、理解へと繋がる。ここで下前頭回にあるミラーニューロンは、島を介して偏桃体と相互結合している。これにより他者の行為を自己の運動と同一視し且つその運動に伴って生じる感情と結合し、共感という機能が実現されるという。

乾氏は、感情を有する次世代 AI 構築の為には、脳機能のみでは不十分で、内受容感覚を受け持つ内臓（主として心臓であるが、広義の意味での身体性）の存在が欠かせない、と述べる。その上で、運動指令、感情制御など上記の仕組みは、効果器と脳との相互作用により予測誤差最小化のフィードバック回路により制御される、とする。ハードウェアコンセプトのヒントになると思われる。

(4) 総務省「次世代人工知能推進戦略」が描く次世代 AI チップ開発の将来像⁶²

総務省の描く次世代 AI チップ開発の将来像に関して最後に触れたい。ディープラーニングは、主に人間の脳における視覚情報処理をモデルに、画像処理を実現した事に触れ、今後視覚だけでなく、触覚、運動、言語などの機能を有する脳の汎化能力、機能をディープラーニングで実現できる可能性がある、とする。まずは触覚へと拡張する事でディープラーニングによる身体性の獲得を実現し、こうした身体性を踏まえた上での柔軟な運動制御、実感のある言語処理を行う次世代人工知能を実現する事が期待される、と述べる。

ハードウェアに関しては、今後の IoT 社会におけるエッジコンピューティングの重要性、Things に搭載され且つ長時間、低電力で駆動可能な AI チップの実現に関して、実用レベルの技術が欧米中心に登場しつつある、との認識の上に、日本においても、ニューロモルフィックチップが大学レベルで進められているが、今後は、脳科学の知見に基づく脳機能に倣うトップダウンの脳型コンピューティング研究と、脳神経回路模倣のボトムアップのチップ技術研究連携が必要、昆虫など高密度ニューロンを有する生物にも学ぶところがある、とする。今後のロードマップとしては、2020 年にヒト脳機能の一部領野を模倣するニューロモルフィック回路、2030 年には脳に加え大脳辺縁系を含めた認知・意思決定系機能を具備するニューロモルフィックチップの完成を目指し、更に 2040 年までに運動系（大脳・小脳）機能を搭載した全脳チップの完成を目指す。

5. 脳科学の興味深い研究事例

5.1 「昆虫の脳」を AI に活用する (μBRAIN)⁷⁴

昆虫の脳は人工知能におけるより効率的なモデルを構築するために利用できるだけでなく、「意識」の意味を探求するヒントになる可能性がある。

小さな昆虫は、「意識」といった概念を構成する最初の段階である「主観的な経験」を体験していることも分かっている。つまり、そうした昆虫の脳を利用することが、ロボットに「意識」を体験させる近道になることが考えられている。

これは、「兵隊の脳」をプログラムできるブレイン・コンピューター・インターフェースを作り出すことを狙っている。

DARPA はこのような「人間や動物の脳」と「コンピュータ」のギャップを埋めるようなテクノロジーの開発に力を入れることにより、新しい計算フレームワークと戦略を探索し、パフォーマンスを損なうこと等なく、サイズとエネルギーを削減することを狙っている。

5.2 「3 人」の脳内ネットワーク共有 (BrainNet)⁷⁵

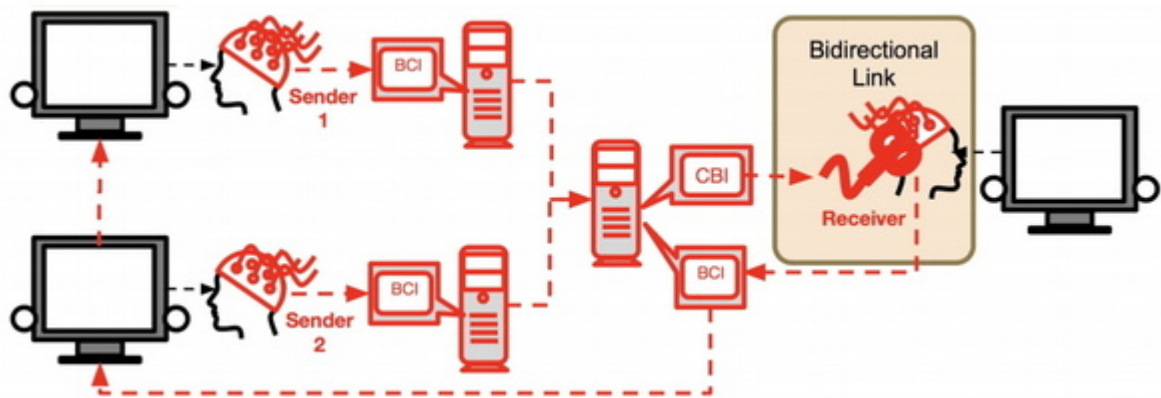


図 5-4 BrainNet の構造

2 人の送信者の脳波をコンピュータが受け取り (Brain-Computer Interface, BCI)、コンピュータが受信者の脳へ TMS を介して情報を伝達する (Computer-Brain Interface, CBI)。受信者の意思決定は脳波測定で送信者へフィードバックされる

University of Washington and Carnegie Mellon University により 3 人の脳波をリンクさせることで脳内ネットワークを共有し、テトリスなどのゲームにおいて脳波だけで指令を送ることができるデバイスが開発された。

2 人が「送り手」となり、1 人の「受け手」に対してツールを用いることで、脳波を通じて指令が伝達できる。

⁷⁴ Microscale Bio-mimetic Robust Artificial Intelligence Networks(μBRAIN)

<https://www.fbo.gov/index.php?s=opportunity&mode=form&id=5fdb0522f0667e24e19274519b8ef011&tab=core&cview=0>
<https://nazology.net/archives/28768>

⁷⁵ <https://www.itmedia.co.jp/news/articles/1810/01/news120.html>
<https://nazology.net/archives/21344>

BrainNet: A Multi-Person Brain-to-Brain Interface for Direct Collaboration Between Brains
<https://arxiv.org/abs/1809.08632>

これがインターネットを介して利用されるようになれば、世界的に複数人で「脳内ネットワーク」を共有することができる可能性がある。

5.3 「脳が量子コンピュータとして働いている」という仮説検証 (Biochemical Qubits)

76

UC Santa Barbara は、量子もつれや重ね合わせを持った量子ビットが、常温で湿った私たちの体内に存在する（仮説）ことを発表。しかし、機能する量子ビットは極低温下でしか見られない現象である。

リン原子は生物の体に豊富に含まれており、生化学的量子ビットとしての可能性を持っている。脳内での神経伝達物質やシナプスでの発火が、量子コンピュータのように量子もつれのネットワークを生み出しているかもしれない。それを実験室で真似ようという試みである。

⁷⁶ Are We Quantum Computers?

<http://www.news.ucsb.edu/2018/018840/are-we-quantum-computers>

<https://nazology.net/archives/4739>

6. 「パーソナル AI」を指向した事例

6.1 主導哲学転換型の事例

6.1.1 「パーソナル AI」の主導的哲学

この章では、個別のヒトに寄り添い、個々にそのヒトなりの支援を行う「パーソナル AI」を次世代人工知能構想のゴールとして指向する。第 I 期の今回は、「パーソナル AI」構想へ至る道程の起点となる事例として、脳科学、神経科学、心理学、人類学などを総合した最新の方法により、ヒトの価値観や情動、その存在場所とされる「こころ」を捉える研究成果を取り上げる。そこから、第三次人工知能ブームが依拠するパラダイムとは異なる、「人間活動システム」を模擬し、ヒトの「内界」に迫る次世代人工知能のパラダイムの可能性を探求する。さらに、本章後段の 5.3 章では、SF 小説やアニメから、人間の人工知能の想像力・創造性を観取する試みをおこなうが、前段では、その創造性観取のための方法論的ツールを「こころ」を捉える事例をもとに作成する。

今回の次世代人工知能構想に先立ち、当研究チームが 2018 年に取り組んだ事例研究では、ディープラーニングの「人間活動システム」上の原理的限界を明らかにした。そこでは、言語や画像に表出したビッグデータを入力とし、類似性のロジックにより導き出すディープラーニングのアウトプットが（開発者も気づかないうちに）差別や偏見などを潜在させてしまうなど多くの問題事例を示した。そして、原理的にいって、ディープラーニングの類似性のロジックでは、個性と個性がぶつかり合う発火から産み出される人間の創造性のダイナミズムを静的で平準化された世界に封殺してしまうという限界を明らかにしたのである。しかし、こうしたディープラーニングの人間活動システム上の限界にも関わらず、ディープラーニングを主軸とする第三次人工知能ブームは、人間活動システムの領域である個人の価値観に基づく消費行動・投資行動や企業人事の意思決定などにも踏み込み、巨額の研究開発費が GAFA を中心としたプラットフォーム企業や主要各国政府により投入されてきた。

さらに、その延長線上にトランス・ヒューマニストの未来像を描くにまで至った第三次人工知能ブームは、この章後段で取り上げる SF 小説が想像する人間的・社会的示唆に富んだ人工知能観とは異質なものがあある。情報工学者の西垣通は、その背景には、ギリシア哲学と一体化した中東のユダヤ=キリスト教の宗教的先入主があると評している。すなわち、「救済計画をもつ唯一神がこの宇宙(世界)を創造したのであり、その設計思想はロゴス(論理)をもとに記述されていて、正確な推論によって真理にたどりつける。そして、選ばれた人間だけが、ロゴス(神の御言葉)を解釈しつつ正義の実現をめざすことができる」「人知を通じてより高い普遍的な絶対知をめざすことができる」⁷⁷ といった宗教的先入主が、トランス・ヒューマニスト論までもが展開された第三次人工知能ブームを支えてきたというのである。

数千年前の宗教にまで遡らずとも、西欧には、人間というものの真実をロゴス(論理)中心主義的、選民思想的に捉え、西欧近代的な意味での「理性」を崇高なる(あたかもロゴスの

⁷⁷ 西垣 通. 2018. AI 原論 : 神の支配と人間の自由. 東京: 講談社.

ような)ものとし、「情動」を野蛮なる(あたかもロゴスに反する)ものとして、分離して考える知的伝統がある。近代哲学の祖ともいべき哲学者・数学者のルネ・デカルトの、心と体、合理的行動と情動を切り離れた心身二元論は、そうした西欧の知的伝統の起点といってもよいであろう。

こうした西欧の知的伝統は、類似性の単純なロジックにより人間の「こころ」までも模擬しようとする第三次人工知能ブームの基層に、(西垣の言う宗教的先入主とともに)少なからず伏流しているであろう。それは、言語や画像の形をとった人間の表出物が集積されたビッグデータ・プラットフォーム上に人間というものの真実の論理があり、その表出物の表層からの類似性抽出により人間の知性を再現できるはずであるといった願望的前提となつて、第三次人工知能ブームの基底にある。

宗教的先入主であれ、知的伝統であれ、あるいは、それらが単なる幻想や(西垣の言うように)荒唐無稽な夢物語であったとしても、その特定の幻想や夢物語が社会的に共有できる文化的・宗教的基盤をもつ欧米社会においては、研究開発者の社会的な内発的動機が惹起され、イノベーションへつながる次のステージの研究成果の道筋もみえてこなくはない。しかし、そうした宗教的・文化的文脈を社会的・歴史的に共有していない日本においては、その幻想や夢物語は社会において共有されにくい。その意味において、第三次人工知能ブームを主導した哲学は、戦略的に言って、日本の社会文化適合的な主導的哲学とはいえない。

このように、人間活動システム模擬にあたっての原理的・哲学的な限界を抱える第三次人工知能ブームに対して、「パーソナル AI」を指向する次世代人工知能構想では、西欧的「理性」偏重バイアスを排し、人間というものの本質にもっと奥深く迫り、それを文化相対的に「個性と共通性」の体系において捉えようとする立場にたつ。こうした立場は、20世紀後半から展開されてきた構造主義の考え方の主軸のひとつである。構造主義の祖ともいわれる文化人類学者のクロード・レヴィ=ストロースは、現存する狩猟採集部族の知性・論理性・倫理性(「野生の思考」)は、西欧近代のそれらと「同じ意味において、また同じ方法によって論理的である」とし、1962年に出版された『野生の思考』の中で次のように述べている：

「現在の地球上に共存する社会、また人類の出現以来いままで地球上につぎつぎ存在した社会は何万、何十万という数にのぼるが、それらの社会はそれぞれ、自らの目には、誇りとする倫理的確信をもち、それにもとづいて自らの社会の中に、人間の生のもちうる意味と尊厳がすべて凝縮されていると宣明しているのである。それらの社会にせよわれわれの社会にせよ、歴史的地理的にさまざまな数多の存在様式のどれかただ一つだけに人間のすべてがひそんでいるのだと信ずるには、よほどの自己中心主義と素朴単純さが必要である。」

「人間についての真実は、これらいろいろな存在様式の間の差異と共通性とで構成される体系の中に存するのである。」⁷⁸

こうした立場にたった論考は、レヴィ=ストロース後も様々な領域で展開されてきてきたのであるが、ヒトの「こころ」に関する研究成果としては、ここ20年間、最新の神経的、認知的、遺伝子的現象の解明の進展を利用して、進化論的にヒトの「こころ」を捉える論考が活発に提起されている。次頁以降で概観する、哲学者のジョシュア・グリーン、ダニエル・デネ

⁷⁸ クロード・レヴィ=ストロース、1980. 野生の思考. 東京：みすず書房。

ット、神経科学者のアントニオ・ダマシオ、認知科学者のポール・ブルーム、スティーブン・ピンカー、社会心理学者のジョナサン・ハイト、進化人類学者のクリストファー・ボーム、進化人類学者のロビン・ダンバー、進化心理学者のダグラス・T・ケンリックらによって提起された一連の研究成果は、このカテゴリーに分類することができる。これらは共通して以下の考え方を基礎としている：

- (1) 人間の「こころ」は、人間の歴史を通じて、主に社会集団を構成して協力して生き延びていくために進化してきた
- (2) 人間の「こころ」の個性は、進化の過程を通じて、将来起こり得る様々な適応課題に社会で対処するための社会集団内多様化戦略として進化してきた
- (3) 人間の知性や価値観は、「こころ」の進化過程を通じて生物学的に身体化したため、化学的、身体的、神経的プロセスに報酬（快）と処罰（不快）のメカニズムが作用することによって生み出される。

次世代人工知能「パーソナル AI」は、このように、人間の知性における情動の役割を重視し、「こころ」の生成過程を進化論的に捉える哲学的立場に依って立つ。

この哲学的立場は、ディープラーニングが主導する第三次人工知能ブームの哲学とは、「人間活動システム」適用上、以下の点で大きく異なっている。

第一に、パーソナル AI の哲学的立場は、人間の知性や価値観が創出される「こころ」を、その内部構造から捉え、身体的・神経的・化学的プロセス、情動、感情、行動、価値観、知性などを相互作用的に一体的に機能するものとして捉える。この考え方は、ニューラルネットワークの単調な何百・何千の特徴量抽出レイヤーで構成される人工知能が「なぜその答えを出したのか」（因果関係・内部構造）がよくわからなくても、アウトプットが外形的に類似していれば、人間の「こころ」が模擬できるとする第三次人工知能ブームの考え方とは根本的に異なる。第三次人工知能ブームのこの考え方を人間活動システムに適用しようとしたとき、外形的には類似した人工知能のアウトプットが、実は人種差別的偏見や性差別的偏見を内蔵した開発者の意図と本質的に逆のものになってしまう事例については、2018 年の当研究チームの報告で取り上げた。

第二に、パーソナル AI の哲学的立場は、人間の知性や価値観が創出される「こころ」の形成過程を進化論的に説明することを基礎としている。生物の進化過程には、ホメオスタシス（何としても生き延びようとする生物学的な力）が貫徹しており、それは遺伝子に刻まれた過去の経験から、未来に起こり得る様々な危機的な状況に備えようとするダイナミックなプロセスである。とりわけ、ヒトの「こころ」の進化過程の場合、社会を形成して社会で生き延びていこうとする選択圧力が強くかかった。ヒトの個性（個体多様性、集団多様性）は、進化の個体選択・集団選択の過程を通じて、未来に起こり得る多様な適用課題に、社会として、対応できるように産み出されたのである。この考え方には、ヒトの「こころ」のダイナミズムを捉える枠組みが用意されており、これは、第三次人工知能ブームが依って立つ、ビッグデータの静的なスナップショットから、類似性マッピングによって人間の「こころ」が模擬できるといった哲学的前提とは対極にある。また、哲学者のダニエル・デネットが言うように、「自然選択による進化は、たしかに発明力があり、またモジュールを大いに多用する過程ではあるが、だとしてもコンピュータープログラミングのような、トップダウン式の研究

開発ではない。それは、ダーウィンの〈奇妙な逆転〉としての、ボトムアップ式の研究開発である⁷⁹⁾。パーソナル AI を指向する次世代人工知能がパラダイムの可能性を追い求めるのは、ヒトの「こころ」の進化過程のようなボトムアップ型のパラダイムである。

第三に、パーソナル AI の哲学的立場は、快・不快といった、生物としての人間が、その内界の深層から発する生物学的プロセスに人間の知性の動因を見出す。この考え方は、救済計画をもつ唯一神がこの世界を創造し、その設計思想はロゴス(論理)をもとに記述されるといった第三次人工知能ブームのトランス・ヒューマニストが前提とする考え方よりも、ずっと西欧的バイアスから自由であり、人類的なスケールの共感を呼び起こす。この章後段の SF 小説からの創造性抽出分析では、そうした人類的な共感を呼び起こすような「パーソナル AI」の問題意識に通ずる示唆が観取された。また、仏教の開祖で哲学的思索に優れた王子ゴータマは、人間の本性を腐食する側面として苦しみをとらえ、「快樂はつねに得られないのに、いかなる手段を使ってでもそれに耽ろうとする欲望」を抑制することで苦しみを取り除こうと説いた。これは、創造神のロゴスを基軸とするユダヤ=キリスト教的哲学に比べて、ずっと進化論的(ホメオスタティック)な考え方であり⁸⁰⁾、パーソナル AI の哲学に通ずるものがある。その意味で、パーソナル AI の哲学は、第三次人工知能ブームが依って立つ哲学よりも、日本・アジアにとっての社会文化的適合性が高く、イノベーションに向けての社会の内発的動機を惹起しやすい、戦略的に優れた科学技術の主導的哲学と位置付けることができる。

6.1.2 「こころ」の進化の起点

人間の知性・価値観における情動の役割を重視し、「こころ」の生成過程を進化論的に捉える「パーソナル AI」が依って立つ哲学は、しかしながら、一般的に広く普及している考え方であるとは、とても言えない。前頁で触れたデカルト哲学の心身二元論を持ち出すまでもなく、ヒトの「価値観」と「情動」は、一般的には、あまり相互作用のない独立的な別のもの、あるいは、相対するものと考えられることが多い。さらには、「情動」は「価値観」を導き出す理性によって克服されるべき野蛮なものとする考え方は、明治以降の日本社会においても一般に広く普及してきた。

「価値観」は、例えば、自由、平等、公正、公共、参加といった近代社会がよってたつ理念や、忠誠、神聖、権威といった伝統的倫理、あるいは、挑戦、温情、共感、協力、協調といった社会生活に関わる信条・社会的了解など、いずれも概ねヒトの知性や思考を存在場所とし、近代の社会・経済・政治制度や科学技術などの知的創造物を支えるものとして一般的に捉えられている。一方、「情動」は、嫌悪、驚き、悲哀、快樂、怒り、恐怖などを第一義的に指し(これらは emotion として括られることが多い)、それらが文化的に装飾された同情、嫉妬、妬み、羨ましさ、軽蔑、プライド、恥、罪悪感、崇拜などの感情(これらは feeling や affect として括られることが多い)として捉えられ、(ビジネスの会議室でも立法議会でも)近代の知的創造物や価値観の所在や支援要因として扱われることは一般的にほとんどなく、阻害要因として扱われることさえある。

⁷⁹⁾ ダニエル・C.デネット, 2018. 心の進化を解明する : バクテリアからバツハへ. 東京: 青土社.

⁸⁰⁾ アントニオ・R・ダマシオ, 2019. 進化の意外な順序—感情、意識、創造性と文化の起源. 東京: 白揚社.

情動がヒトの知的創造物の支援要因ではなく、阻害要因ですらあるのだとすれば、情動の泥沼から自立・開放され、思考の世界のみから決定される意思によって振る舞えるヒトは、さぞかし知的に、創造的に、常に合理的「価値観」に基づき正しく振る舞えるに違いない。しかし、脳科学的に確かめられたそうした事例によると、情動から独立して思考するヒトは、そのように知的には振る舞えないことがわかっている。

情動から独立して思考するヒトの事例のひとつは、サイコパスである。500 人に 1 人くらいとされるサイコパスの脳では、情動が機能していないことが確かめられている⁸¹。サイコパスは、一般に IQ が高く、人の操縦術にたけている。精神科医でサイコパスの専門家のロバート・ヘアによると、その特徴は、口がうまい／見かけは魅力的、過大な自尊心、病的な虚言癖、悪賢い／巧みに人を操る、自責の念に欠ける、無感覚／共感性の欠如、自分自身の行動責任を受け入れないなどとされている⁸²。

IQ が高く、ヒトの操縦術にたけたサイコパスをしめす最たる事例として、史上最悪のサイコパスの犯罪とされるイギリスの診療医ハロルド・シップマンの連続殺人事件がある。シップマンは、28 年間に自分の患者 215 人を殺害したが、シップマンの患者の一人は次のように語っている。「診療所の他の先生方は、患者が部屋に入る前からもう『処方箋』を書いている。できるだけさっさと診て終わらせるという感じでした。シップマン先生は絶対そんなことはしなかった。心底、患者に関心をもっていたんです。私のプライベートについても質問してくれたりして、本当に気に入ってくれているように見えました。家庭医の先生としては最高でした」⁸³。

また、アメリカの社会奉仕活動家のテッド・バンディは、少なくとも 30 人の若い女性をレイプし、切り刻み、殺害したが、ワシントン大学で心理学を専攻教授からの評判もよく、自殺救済電話相談室でボランティア活動をしていた。それらを通して、女性から信用を得る話術を学んでいたのである⁸⁴。

サイコパスの脳では、快・不快といった基本的情動を決める部位扁桃体の活動が低く、他人の痛みを自分も感じる共感性が発出される部位である眼窩前頭皮質や良心をつかさどる部位である内側前頭前皮質の活動も低い。また、その両者、扁桃体と眼窩前頭皮質や内側前頭前皮質の結びつきが弱い⁸⁵。

この脳の違いは、人工知能倫理 (ELSI) の議論によく用いられる 3 つの心理実験、トロッコ問題、歩道橋問題、臓器移植問題において、顕著に表れる。

- トロッコ問題 (制御不能となり暴走するトロッコの線路の先に 5 人が括りつけられ、スイッチを切り替えた線路の先に 1 人が括りつけられている。スイッチを切り替えれば犠牲者は 1 人ですむが、切り替えなければ 5 人となる。あなたは

⁸¹ アンтониオ・R・ダマシオ, 2013. 自己が心にやってくる: 意識ある脳の構築. 東京: 早川書房.

⁸² クリストファー・ベリー=ディー, 2018. サイコパスの言葉. 東京: エクスナレッジ.

⁸³ クリストファー・ベリー=ディー, 2018. サイコパスの言葉. 東京: エクスナレッジ.

⁸⁴ クリストファー・ベリー=ディー, 2018. サイコパスの言葉. 東京: エクスナレッジ.

⁸⁵ 中野信子, 2016. サイコパス. 東京: 文藝春秋.

スイッチを切り替えることを支持しますかを問う心理実験)では、ほとんどの一般の人がスイッチ切り替えを支持する。

- 歩道橋問題 (制御不能となり暴走するトロッキの先に 5 人が括りつけられている。その手前には歩道橋があり、そこから太った人を突き落とせば、その人は犠牲になるが、トロッキを止めて 5 人の命を救うことができる。あなたは、太った人を突き落とすことを支持しますかを問う心理実験)では、支持する一般の人はほとんどいない。
- 臓器移植問題 (この会議室にいるこの青年を殺せば、臓器移植を待っている 5 人の命を救うことができる。あなたはこの青年を殺すことを支持しますかを問う心理実験)では、支持する一般の人はいない。

しかし、サイコパスはこの 3 心理実験のいずれに対しても、迷うことなく支持を表明し、迷う一般の人の気持ちが理解できないのだという⁸⁶。

トロッキ問題、歩道橋問題、臓器移植問題の支持・不支持を問われたときの一般の被験者の脳では、情動関連の部位 (内側前頭回、帯状回、角回) の反応と思考関連の部位 (中前頭回、頭頂葉) の反応に違いがあることが観察されている。トロッキ問題の場合、5 人を助けるために犠牲にしてしまう 1 人との物理的接触がなく、スイッチ切り替えの賛否を問うだけである。この場合の支持・不支持を問われたときの一般の人の脳では、思考関連部位が感情関連部位より強く活性化されている。一方、歩道橋問題の場合は、5 人を助けるために犠牲にしてしまう 1 人を直接的に突き落とし、臓器移植問題の場合は、5 人を助けるために犠牲にしてしまう 1 人が、現に目の前にいる生身の人間である。この 2 つの問題で、支持・不支持を問われたときの一般の人の脳では、情動関連部位が強く活性化され、思考関連部位は軽微な活性化もしくは不活性化されていた⁸⁷。

また、歩道橋問題と臓器移植問題における一般の人では、速く支持・不支持表明をした被験者 (0.2-3 秒の脳辺縁系が司る情動的反応がより強く働いている) ほど、不支持率が高く、遅く支持・不支持表明をした被験者 (すなわち、情動反応よりも思考的判断がより強く働いている) ほど、支持率が高くなることも観察されている⁸⁸。

これらの結果から、5 人の命を救うか 1 人の命を救うかという「合理的」計算は思考関連脳部位で処理され、5 人を助けるために犠牲にしてしまう 1 人の痛みに対する共感や、人を突き落とす・目の前にいる青年を殺すといったことに対する罪悪感・抵抗感・「道徳的」判断が、情動関連脳部位で処理されていることが示唆される⁸⁹。情動関連の脳部位が機能していないサイコパスの脳では、一般の人にみられる情動関連部位の活性化が起こらないため、3 つのいずれの心理実験においても迷うことなく、道徳的な抵抗感を負うことなく、支持の意思決定ができると考えられる。

⁸⁶ ジョシュア・グリーン, 2015. モラル・トライブズ. 東京: 岩波書店.

⁸⁷ Greene, J. D., 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293[2105].

⁸⁸ Greene, J. D., 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293[2105].

⁸⁹ ジョシュア・グリーン, 2015. モラル・トライブズ. 東京: 岩波書店.

こうしたサイコパスのいわば「情動なき合理的思考」がどのように機能するのかについては、日本で2017年に起きた相模原障害者施設殺傷事件の犯人供述に端的に見て取ることができる。サイコパスと推測されている相模原障害者施設殺傷事件の犯人は、「障害者は不幸しか作れない」、「障害があつて家族や周囲も不幸」であるにも関わらず、「障害者の安楽死を国が認めてくれないので、自分がやるしかないと思った」、「事件を起こしたのは、不幸を減らすため」であり、「日本のため」である、などと供述している。ヒトとして当たり前の道徳観が欠如し、そのために前提となる事実認識が間違っていることを横におけば、供述の論理性自体には供述全体を通して問題がなく、その意味で合理的である。このことが、情動なき「合理的思考」の無意味さ、怖しさを物語っているが、これは同時に、他者の痛みへの共感、社会的嫌悪、罪悪感などの「情動」が、社会を維持していくうえで必要な基本的な社会的ルールや道徳的価値観の基礎となっていることを示唆している。

原初の人類の社会はチンパンジーのころからそれほど大きくは進歩しておらず、力が支配する社会であった。ところが、その後、栄養効率の高い大型哺乳動物を恒常的に獲得できる狩猟技術の発達と相まって、45,000年くらい前までには、社会的ルールの内面化と個人の評判にもとづく平等主義的な社会へと移行したと進化人類学者のクリストファー・ボームは論じている⁹⁰。1頭のチンパンジーがもう1頭のチンパンジーを殺すのは、大変な労力を要し、重大なリスクを伴う。しかし、鹿などの大型で俊足な哺乳動物を恒常的に獲得する技術を身に付けてからの人類にとって、それは（トロロッコ問題でのスイッチを切り替えるように）簡単な作業となった。そのため、もともと服従より個人の自立を好む性質をもっていた人類は、群れの中で暴力的に生存資源（生殖と食料）を独占しようとする支配者を倒して暴力的支配を自制するようになり、さらに、肉の平等な分配を阻害する泥棒やズルをする者を処罰するようエスカレートしていったことが、考古学的証拠と人類学的証拠によって確からしく推測されている。この頃に人類は、社会的ルールと価値観を、恥ずかしさで赤面する現象にみられるように生物学的に内面化していったと、ボームは考えたのである⁹¹（したがって、サイコパスは赤面しない⁹²）。

進化の過程を通じて、情動に内面化されてきた社会的ルールや価値観は、ヒトの脳の中では社会脳とも呼ばれる、扁桃体、眼窩前頭野、側頭葉を主とする社会認知に関わる脳領域を中心として、神経細胞のネットワーク、脳領域のネットワーク、脳どうしのネットワークなど、複数の階層によって司られている⁹³。それらは、進化的に見て比較的最近発達したものであり、一部は人間特有のものである⁹⁴。

また、心の哲学では「志向性」(intentionality) という用語があり、これは、「信じる」、

⁹⁰ クリストファー・ボーム, 2014. モラルの起源：道徳、良心、利他行動はどのように進化したのか. 東京：白揚社.

⁹¹ クリストファー・ボーム, 2014. モラルの起源：道徳、良心、利他行動はどのように進化したのか. 東京：白揚社.

⁹² 中野信子, 2016. サイコパス. 東京：文藝春秋.

⁹³ 信原 幸弘, 2017. 心の哲学：新時代の心の科学をめぐる哲学の問い. 東京：新曜社.

⁹⁴ アントニオ・R・ダマシオ, 2013. 自己が心にやってくる：意識ある脳の構築. 東京：早川書房.

「推測する」、「想像する」、「望む」、「理解する」、「考える」、「意図する」などという言葉を用いる能力にかかわる志向姿勢または志向的構えをもつ能力を意味する⁹⁵。ヒトの志向性の水準を、チンパンジーやマカクなどの霊長類の志向性水準と比較すると、その差異は、前頭葉の容量と明確な相関関係にある。このような相関関係は、一般的に人類学や霊長類学で重視される言葉の理解や複雑さ、道具の複雑さなどには全くみられないことから、人間の脳はそもそも社会関係を築くために進化したものであるとの考え方（「社会脳仮説」）も提案されている⁹⁶。

6.1.3 「こころ」の進化過程

ヒトの「価値観」は、また、「情動」とは独立的に学習されていくものであるとの捉え方が一般的である。とりわけ、近代民主主義社会や資本主義経済を支える自由、公正、公共、成長、挑戦といった価値観は、学校や家庭や社会における教育や社会関係を通じて芽生え、育てられるものであって、嫌悪、驚き、悲哀、快楽、怒り、恐怖などといった「情動」とは独立的に成長過程を通じて学習されていくもの、場合によっては、「情動」を制御する能力を身に付けることによって、近代の市民としての自由、公正、公共といった「価値観」が身についていくものであるとすら考えられている。

しかし、ヒトが社会生活を営む上で必要な価値観は、言語習得以前の幼児においてすでに発現していることを示す実験がある。イェール大学の心理学者のポール・ブルームは、生後1年の幼児を対象に、助ける図形と邪魔する図形の人形劇から幼児の社会性判断力を確かめる心理実験を行った。登場するのは、目だけが付いた丸と三角と四角の3つの形状の図形で、3つの図形にはピンク、黄色、淡い青の色がつけられる。この形状と色の組み合わせは、次に述べる図形 A/B/C において様々に変更される。最初のシナリオの人形劇では、坂を上ろうとして一生懸命な図形 A に対し、助ける図形 B が、下から図形 A を押し上げて坂の上まで登れるよう助けてあげる。次のシナリオの人形劇では、坂を上ろうとして一生懸命な図形 A に対して、邪魔する図形 C が上から図形 A を下まで押しおろして邪魔をする。この2つのシナリオの人形劇を様々な形状と色の組み合わせで幼児にみせた後、幼児に助ける図形 B と邪魔する図形 C をみせ、欲しい方を選ばせるという心理実験である。結果は、ほとんどの幼児が、人形劇での図形の色と形状の組み合わせに関わらず、助ける図形 B を選択したという^{97 98}。また、京都大学の鹿子木康弘らによって行われた実験においては、攻撃を受けているパックマンの個体を身を挺して助けるような振舞いを示す別のパックマンの個体が、言語習得以前の幼児によって好まれることも示された⁹⁹。

⁹⁵ 信原 幸弘, 2017. 心の哲学：新時代の心の科学をめぐる哲学の問い. 出版地不明:新曜社.

⁹⁶ Dunbar, R. I. M., 2016. The Social Brain Hypothesis and Human Evolution. Oxford Research Encyclopedia of Psychology, 第 MAR 巻, p. Online Publication.

⁹⁷ Bloom, P., 2010. The Moral Life of Babies. The New York Times, May.

⁹⁸ ポール・ブルーム, 2015. ジャスト・ベイビー：赤ちゃんが教えてくれる善悪の起源. 東京: NTT 出版

⁹⁹ Kanakogi, Y., 2017. Preverbal infants affirm third-party interventions that protect victims from aggressors. Nature Human Behaviour, 1[0037].

この2つの実験結果が示唆しているのは、第一に、他者の協力的な振舞いを識別する能力とそれを肯定的に評価する傾向、社会正義や社会的価値観に相当する観念を、ヒトは生得的に備えていることである。第二に、どちらの心理実験においても、アニメーションでは図形的なキャラクターによって表現されている。このことから、2つの実験結果は、社会正義のような識別能力はかなり抽象的なレベルで働いているということも示唆している¹⁰⁰。これは、ヘビを本能的に怖がる幼児の識別能力とよく似ている。幼児は、ヘビは怖いと教えなくてもヘビを怖がるのであるが、幼児にヘビが認識できるわけではなく、ヘビのような形と動きをするものを怖がる識別能力を生得的に備えていることがわかっている。ヒトに限らず生物の赤ちゃんは、こうした生存上に必要な識別能力を生得的に身に付けているのであるが、ヒトの幼児に関しては、生存上に必要な識別能力の一部として、社会正義の観念を生得的に身に付けていると考えられる。

人間は社会的な生物であり、他者との協力なしには生きていけない。個人にとって、自分に協力してくれる他者を見つけること、そのような相手と協力して行動することが文字どおり生死を分かち重要性をもっている。そのために人間は進化の過程で、他者が協力的であるか・あるいは非協力的であるかという情報を敏感に探知する能力、また自分も利己心を抑えて協力的に振る舞う能力を発達させてきたのだ、と哲学者のジョシュア・グリーンは言う¹⁰¹。とりわけ、45,000年くらい前までに人類が発達させた大型哺乳動物を恒常的に獲得する狩猟技術は、部族を一気に全滅させるためにも利用できた。そのため、この頃から人類は強い部族間戦争の圧力にさらされるようになり、農耕に移行した後は、土地や灌漑の収奪によって生存の危機にさらされるため、さらに強い戦争圧力にさらされることになった。経済学者のサミュエル・ボウルズと理論生物学者のピーター・ターチンは（それぞれ別々に）、この部族間戦争圧力によって、協力的な社会集団と個体に選択圧がかかり、協力や利他行動などの社会的ルールがヒトの「こころ」に内面化され、それが社会制度に実装されることを通じて、ヒトの「こころ」と社会制度が共進化してきたと論じている^{102 103}。

さらに、遺伝学的には、「最近の遺伝子変容 (recent genetic selection)」と呼ばれる、完新世初期・農業革命の頃をピークとした解剖学的現生人類の数十もの遺伝子変容がみられることが確かめられている¹⁰⁴。その時期の遺伝子変容のうちのひとつで東北大学の河田雅圭らの生命科学チームが近年検出した SLC18A1 遺伝子は、不安傾向や神経質傾向の違いに影響する「こころの個性」に関わる遺伝子である。この研究成果により、「こころの個性」は、ヒ

¹⁰⁰ 久木田 水生, 2017. 麦とペッパー ―テクノロジーと人間の相互作用―. 一般社団法人人工知能学会. 人工知能, 32[5].

¹⁰¹ ジョシュア・グリーン, 2015. モラル・トライブズ. 東京: 岩波書店. 久木田 水生, 2017. 麦とペッパー ―テクノロジーと人間の相互作用―. 一般社団法人人工知能学会. 人工知能, 32[5].

¹⁰² サミュエル・ボウルズ, 2017. 協力する種 : 制度と心の共進化. 東京: NTT 出版.

¹⁰³ Turchin, P., 2015. Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth. Beresta Books.

¹⁰⁴ Pickrell, J. K., 2009. Signal of recent positive selection in a worldwide sample of human populations. Genome Research, 第 19 巻, pp. 826-37.

トの集団内で自然選択によって異なる遺伝子型が集団中に積極的に維持され、人類の進化過程で有利に進化してきたことが、進化遺伝学的手法により示された¹⁰⁵。

6.1.4 「こころ」のマトリックス

次世代人工知能「パーソナルAI」が依って立つ、ヒトの価値観・情動の生成過程を進化論的に捉える哲学的立場については、本章でここまでに取り上げた研究成果以外にも、多くの論考が提起されている。

それらの多数の研究成果の中でも、この章の後段で述べる、SF小説やアニメから人間の創造性を本質観取するためのツールとして有用と考えられるモデルを提案している論考があり、ここでは、以下の5つのモデルを事例として取り上げた。

- ① ジョナサン・ハイトの「道徳価値共通基盤」(ジョナサン・ハイト, 2014. 『社会はなぜ左と右にわかれるのか：対立を超えるための道徳心理学』. 東京：紀伊國屋書店.)
- ② ダグラス・T・ケンリックの「七つの下位自己」(ダグラス・T・ケンリック、ヴラダス・グリスケヴィシウス, 2015. 『きみの脳はなぜ「愚かな選択」をしてしまうのか：意思決定の進化論.』 東京：講談社.)
- ③ スティーブン・クォーツとアネット・アスプの「社会的指向のマーケティング・モデル」(スティーヴン・クウォーツ アネット・アスプ, 2016. 『クール：脳はなぜ「かっこいい」を買ってしまうのか.』 Tokyo：日本経済新聞出版社.)
- ④ ジェイン・ジェイコブズの「市場の倫理、統治の倫理」(ジェイン・ジェイコブズ, 1998. 『市場の倫理 統治の倫理』. 東京：日本経済新聞社.)
- ⑤ アラン・P・フィスクの「社会関係モデル」(Fiske, A. P., 2004. “Four modes of constituting relationships: Consubstantial assimilation” ; Relational models theory: A contemporary overview. Lawrence Erlbaum Associates Publishers, pp. pp. 61-146.)

これらはいずれも、遺伝子的・脳科学的・神経科学的解析、心理実験、経済行動観察、考古学的・歴史学的証拠、霊長類学的・文化人類学的証拠などを総合しつつ、「こころ」の進化過程を通じて産み出された、人間の生活のそれぞれの局面における、本質的、普遍的な倫理観、道徳観、行動原理を、ある程度の体系性・網羅性をもって、「こころ」のモジュールに分解し、モデル化している。次頁以降に5つの事例を「こころ」マトリックスとして表形式に整理し、この章後段でのSF小説・アニメから創造性を観取する軸ツールとして提供する。

なお、ジェイン・ジェイコブズの「市場の倫理、統治の倫理」に関しては、遺伝子的な証拠等にはほとんど言及しておらず、古今東西の文学や史実に基づくモジュール抽出が中心であるが、そのモジュールは、スティーブン・クォーツとアネット・アスプの「社会的指向のマーケティング・モデル」を人文学的に補強し得るという利点があり、事例のひとつとして、「社会的指向のマーケティング・モデル」に併記する形で取り上げることとした。

¹⁰⁵ Sato, D. X., 2018. Positive and balancing selection on SLC18A1 gene associated with psychiatric disorders and human-unique personality traits. *Evolution Letters*, 第 August 巻.

表 6-1 ジョナサン・ハイト「道徳価値共通基盤」

道徳価値共通基盤	自由	公正/互惠	温情	内集団忠誠	権威	神聖
進化適応課題	機会さえあれば他人を支配し、脅し、抑制しようとする個体とともに、小集団を形成して生きていかねばならない	双方向の協力関係の恩恵を得る	子どもを保護しケアする	結束力の強い連合体を形成する	階層制のもとで有益な関係を結ぶ	汚染を避ける
オリジナル・トリガー (進化過程における契機)	部族集団の中で、信頼を勝ち取らずに自分をリーダーとして認めさせようとするもの、いばり屋、暴君、他の個体を支配しようとするあらゆる攻撃的な行動	欺瞞 協力 詐欺	苦痛、苦境、自分の子どものニーズ	グループに対する脅威や挑戦、他の部族からの村への奇襲攻撃	支配や服従の兆候	廃棄物 病人
カレント・トリガー (現代生活における契機)	個人の自由に課せられる不当な制限と感ずるもの、高圧的管理魔の上司、ファシズム	私物化、えこひいき、不透明な厚遇、不当な冷遇、ズルして得しているヤツ	赤ちゃんアザラシ、かわいらしい、漫画キャラクター	スポーツチーム、国家、企業	上司、指導者	タブー視されている考え方 差別意識 (潜在意識に押し込まれている場合が多い)
特徴的な情動	わくわく感 ときめき プライド	怒り 感謝 罪の意識	思いやり	グループの誇り、裏切り者に対する怒り	尊敬、怖れ	嫌悪
関連する美德	挑戦 好奇心	公正 公共 信頼	親切、共感、弱者・困っている人に手をさしのべる	忠誠 愛国心 自己犠牲	服従 敬意	節制 貞節 敬虔 清潔さ

表 6-2 ダグラス・T・ケンリック 「7つの下位自己モデル」

下位自己	配偶者獲得	配偶者保持	親族養育	協力関係	病気回避	自己防衛	地位
進化適応課題	異性の気を引きつける	その異性を手放さない	家族の世話をする	友人をつくる	病気を避ける	身体への危害を逃れる	地位を得る
性質	自由奔放な独身者	よき妻 よき夫	愛情深い親	チームプレイヤー	強迫的な心気症者	夜警	野心家
活性化される契機 (現代生活における事例)	現実や想像の配偶者候補によって呼び覚まされる。セクシーな広告を見たり、ロマンチックな物語を読んだり、女性の下着に触れたりする	伴侶が満足しているか みじめな気分かについての手がかりに敏感であると同時に、社会の地平を眺め わたし、その伴侶を奪ってもっと幸せにしよう とねらう 邪魔者がいないか気を配る	自分の子どもに囲まれているとき 指揮をとるだけではなく、よその人の愛らしい赤ん坊を目にした り、はるかかなたで子どもが泣いている声を耳にしたり ただけで活動しはじめることもある	友情に登場のきっかけを与えるものごと	ほかの人のくしゃみや咳(せき)が聞こえたり、病変した皮膚を見たり、悪臭を感じたり、遠く離れた異国から来た人を見たり考えたりする	身体への危険の察知、見知らぬ人の怒った顔、ほかの人種や異教徒についての思考、怖い映画、陰惨な犯罪のニュース	自分が階層のどこにいるか、上と下にだれがいるかに敏感
活性化されたときの作用 (現代生活における事例)	人は独自の存在になりたがり、注目されたがった。たとえば、この下位自己を活性化させると、男性はほかの大多数の人が好むブランドではないほうを選んだ	自分たちの関係をおびやかすものの代表格とも言える、魅力的な同性に注意を払うようになる		人は、ふつう個人で消費する商品より、他人と自分を結びつける商品(自分の大学ロゴマークを誇示するリストバンドなど)に多くのお金をかける	人をふだんより内向的にする よそ者・異国から来た人への警戒心	用心深くなる	一見ささいに思える侮辱でも、攻撃性の爆発を引き起こす

表 6-3 スティーブン・クォーツ&アネット・アスプ
「社会的指向のマーケティング・モデル」

社会的指向	反逆的社会指向	支配的社会指向
	平等主義的な社会を好む	階層的な社会を好む
類人猿からの継承本能	反逆本能	地位を求める本能
進化適応課題	<ul style="list-style-type: none"> チンパンジーや人間などの横暴な種は本質的に支配されるのをひどく嫌う、下位者は結束し、権力を手に入れ成りあがろうとする者に制裁を加えることで、支配者的な行動を減らす結束行動（原始的な集団行動）をとる 人間社会では、結束行動はより綿密・意図的になり、道徳的制裁が、過度な競争、威圧、迫害をする指導的なオスの行動を制御する主な手段となり、それが組織化・制度化されていくようになった 	<ul style="list-style-type: none"> チンパンジーの社会は「力が正義」の階層的な社会であり、人間社会でもある程度の資源が存在すると、地位を求める本能が現われてたちまち階層ができる さらに、階層的な社会構造や中央集権的な政治構造から（生存資源の確保などの）利益がもたらされると感じれば、反逆本能はしばしば弱められ、階層的な社会がさらに組織化・制度化されていくようになった
神経科学的特徴	行動活性化システム（BAS）が強い ドーパミン（報酬系）活性	行動免疫システム・行動抑制システム（BIS）が強い
脳科学的特徴	<ul style="list-style-type: none"> 最後通牒ゲーム（分配を不公平と感じた場合、拒否して自分と分配者の取り分をゼロにできる行動経済学の実験）において、前頭（恐れや嫌悪といった負の感情と結びついた脳部位）が活性化される さまざまな形の不公平な扱いに対して人間が感じる、強烈な感情的反応がある 	<ul style="list-style-type: none"> リアルタイム結果表示の IQ テスト（少人数のグループに IQ テストを一斉に受けてもらいながら、1 問終わるごとに、グループ内での自分の順位が見える）において、扁桃体（恐怖に関連する部位）が強く活性化した。また、順位が上がったとき、脳の側坐核（報酬に反応する部位）が強く活性化した 地位・序列に対して人間が感じる強烈な感情的反応
消費選好	皆がベンツを欲しがっている集団において、BMW/PRIUS を買う	皆がベンツを欲しがっている集団において、ベンツを買う

表 6-4 ジェイン・ジェイコブズ 「市場の倫理統治の倫理」

市場の倫理	統治の倫理
<ul style="list-style-type: none"> ● 自発的に合意せよ ● 暴力を締め出せ ● 正直たれ ● 他人や外国人とも気やすく協力せよ ● 競争せよ ● 契約尊重 ● 創意工夫の発揮 ● 新奇・発明を取り入れよ ● 効率を高めよ ● 快適と便利さの向上 ● 目的のために異説を唱えよ ● 生産的目的に投資せよ ● 勤勉なれ ● 節儉たれ 	<ul style="list-style-type: none"> ● 忠実たれ ● 余暇を豊かに使え ● 目的のためには欺け ● 復讐せよ ● 位階尊重 ● 伝統堅持 ● 気前よく施せ ● 見栄を張れ ● 規律遵守 ● 排他的であれ ● 勇敢であれ ● 剛毅たれ ● 取引を避けよ ● 運命甘受 ● 楽観せよ ● 名誉を尊べ

表 6-5 アラン・フィスク 「社会関係モデル」

社会関係モデル	平等対等	市場値付け	共同体的 分かち合い	権威/序列
社会関係性	資源を均等に分配するための仕組みを採用して、たとえば優先権を交代制にしたり、コイン投げで決めたり、各自の貢献を均等にしたり、分け前を平等にしたりする。	通貨、価格、地代、給料、手当、利子、クレジット、デリバティブなど、現代経済を動かしているシステム。（ほかの三つの社会関係モデルと異なり）普遍的とはほど遠く、読み書き能力や計算能力や、近年発明された情報テクノロジー、数字、数式、会計、デジタル転送、正式契約言語などに依存する。	人びとは集団内で進んで資源を分けあい、誰がどれだけ得たり損したりしているかをまったく気にかけない。彼らは集団の概念を、共通の本質によって統一された「一心同体」ととらえ、それが汚染されないように守らねばならないと考える。	上位者は何でも欲しいものを得て、下位者から貢ぎ物を受け、下位者に服従と忠誠を命じる権利があるが、その一方で、下位者を保護するために、家父長的、ノブレス・オブリージュ的（武士道的）な責任をはたす義務がある。
進化論的自然選択メカニズム	互惠的（しっぺ返しの）利他主義（進化論的に安定的な戦略）	分業化と交換の適応値（適応度）	包括的適応度による血縁選択	階層構造における服従・支配の振舞いの適応値（適応度）
神経科学・脳科学的特徴	意図、だまし、不調和、視点取得、計算などを記録する脳の部位、すなわち島、眼腐皮質、帯状皮質、背外側前頭前皮質、頭頂葉皮質、側頭頭頂接合部などを包含する神経回路が司る	（報酬系／ドーパミンを基礎とするが、マックス・ウェバーのいう「合理合法」などに相当し、意識的思考によって司られる部分が多い）	（少なくとも部分的には）オキシトシンが司る	（少なくとも部分的には）テストステロン感知回路が司る
関係性の構造	アーベル群	アルキメデス順序体	等価関係	線形
主たる計測基準	（関係性の）距離	比率	類型・名目	順位

6.1.5 「こころ」のマトリックスによる創造性の本質観取

これら 5 つの「こころ」のマトリックスは、文学者やクリエイターが構想する未来の可能性を概観・総説するこの章の後段での創造性分析の有用なツールとして利用され、その作業を通じて、ヒトの価値観や情動に由来する「人間活動システム」の模擬を模索していく。この SF 小説・アニメの創造性分析にあたっては、次世代人工知能のパラダイムの可能性を人の「内界」におくパーソナル AI としては、フッサール現象学の「本質観取」をその分析方法論として採用する。

本質観取とは、

- ① 人間の「体験」について、それを自然科学や社会学のように外側から説明するのではなく、当事者がそれをどのように体験しているかを内側から考察しようとする姿勢であり、
- ② その体験から、ある種の「一般性」をもったもの（構造とか本質と呼ばれる）を取りだそうとする姿勢を基本とした現象学的方法論である¹⁰⁶。

SF 小説・アニメの分析においては、ここでいう「当事者の体験」は、分析者自身の SF 小説・アニメにおいて体験した世界とする。その体験世界から一般的、構造的な本質を一本の軸で説明するツールとして、本章後段で、「こころ」のマトリックスが利用されることとなる。

IT 産業において、本質観取と似た方法論は「ユースタories」や「ペルソナ」と呼ばれ、一般に広く普及している（シリコンバレーでこの方法論を最初に考案した開発者はフッサール現象学の信奉者だったのかもしれない）。IT 製品開発において、従来は、ユーザー・セグメンテーションによる製品ポジショニングなどの定量的、社会科学的な方法により製品戦略設計が行われていた。しかし、そうした定量的な方法論ではどうも戦略と実際の製品との乖離が大きくなってしまふなどの経験を積むうちに、2000 年代後半ごろから、ユーザー体験を内側から考察したユースタoriesの戦略設計方法論への移行が進んだのである。現在では、本質観取的なユースタoriesの方法論は、よく定着している。そうした意味でも、本質観取はパーソナル AI 構想にとって優れて適合的な方法論であると位置づけている。

¹⁰⁶ 西 研, 2015. 人間科学と本質観取. 著: 小林 隆児, 西 研, 編 人間科学におけるエヴィデンスとは何か: 現象学と実践をつなぐ. 東京: 新曜社.

6.2 クリエイターに芽生えた先見的事例

次世代人工知能の構想にあたっては、昨年度 2017 年度時点の各種データを用いて、ディープラーニングにより実現できる人工知能によりつくり出される世界の限界を探り、「数理的な機械学習とは異なる発想の下で、次世代の人工知能を構想する」というアプローチについて言及がなされた。本項においては、これまでの人工知能への取り組みとは異なるアプローチとして、研究者や技術者らによる数理的なモデルからは導き出せない、SF 作家を始めとしたクリエイターが想像する次世代の人工知能の在り方を示す事例を抽出することを目的とした。

6.2.1 先見的事例の抽出に向けた取り組みについて

(1) 研究者や技術者によるこれまでの取り組みについて

これまでの人工知能の検討は数理科学的なアプローチが中心であり、特にコンピュータ上で知能を実現することを目指したものの中心であった。計算機の登場以来、その上で動作する人工的な知能については、だれもが思い描くものであったといえる。

過去にさかのぼって検討された例を見ると、1970 年に発行された情報処理学会誌の人工知能特集においては、「産業用ロボットにも、自動制御プラントにも交通管制システムの類にも、『より高い知能』が要求されている」、「人間にはできるが機械にはできないことを人工的に実現してみること」であると述べられている¹⁰⁷。また、その時点では、「パターン認識などの機能が知的な機能とみなされていた¹⁰⁸」。ただし、「パターン認識についても、その時点では知能的とみられているが、やがてアルゴリズムが明確になれば非知能的な仕事になってしまう」とも述べられている。ここで、彼らは、人工知能についてのアプローチの案として、ミクロなアプローチとして生理学に基礎を置くものとしてあげ、マクロなアプローチとして「人間をひとつの暗箱としてその働きやふるまいを調べる」ものとしてあげている。ここで紹介されるアプローチの内容は、前者については生体情報処理を構成する神経回路網の素子によるシミュレートや、シナプスを表現する論理モデルの検討など、現在検討されているものと考え方はさほど変わらない。後者については、人間の思考や行動を捉えるモデルとしてゲーム理論を用いたモデル化などをあげている。どちらも、当時の計算機をベースに考えられているため、その限界の域を超えた人工知能が出来上がることはなかった。その号の人工知能特集における他の文献で議論されている内容についても、当時の計算機を道具として用いたものであり、その発想の域を超えることはない。例えば、「生物における情報処理では生物科学者の努力により、脳の情報処理の仕組みが計算機のそれとはかなり大きく違っていることは間違いない」としていながらも、「そのまま人工知能の研究に役立つとは考えにくい」と述べている。さらに、パターン認識や言語認識といった分野では現代でもその延長線上あるものも多く登場する。

現代のディープラーニングを旗手とする、第三世代の人工知能においても、基本はコンピュータ上のアルゴリズムであり、発想は 1970 年の段階からあまり変わっていないと考える

¹⁰⁷ 末包 良太, 戸川 隼人「人工知能特集号について」 情報処理 11(11), 635, 1970-11-15

¹⁰⁸ 南雲 仁一, 中野 馨, 吉沢 修治, 小林 欣吾, 西村 千秋「人工知能特集号：解説：人工知能への二三のアプローチ」 情報処理 11(11), 636-645, 1970-11-15

こともできるかもしれない。

なお、前述情報処理学会誌の特集の中に座談会形式による人工知能にかんする議論¹⁰⁹が掲載されている。その参会者のひとりの発言に「**野球を監督なしでファンの意見をオンラインで集約しながらコンピュータでやせたら面白いのではないか**」というものがあり、今まで、監督（人間）がやっていたことがコンピュータに代わり、それと人間が協調して野球というゲームを行うといった、人間と人工知能が融合した社会イメージのようなものも垣間見える。さらには、「コンピュータというものも大勢の人間の中のきわめて個性のある個人である」という捉え方をするとといったアイデアも飛び出し、この意見に対しては、結果的にそうなるという気がするといった賛同も得えて、「ほんのちょっとした偶然でたくさんの計算機が全然違ったものを覚えていくことになる」といった展開を見せている。

これらのイメージは、現在ゲームやインターネットや Web 上でほぼ実現されている。しかし、この時代の段階ではそれを実現するための仕組みを研究テーマとするほど具体的なイメージを持っていなかったとみることができる。

ここでは、1970 年の日本における研究者の議論を例としてあげたが、人工知能の将来を考えるうえで、多くの自然科学に係る研究者は、人工知能が計算機（コンピュータ）上で実現されるものとしている以上は、その時代の計算機の範囲を大きく逸脱した発想を持つには至らないと見ることができる。翻って、現代では、1970 年に比べコンピュータは大きく進化し、さらにインターネットをはじめとする機械同士がリアルタイムに結びつき協調し動作する環境も実現されている。それを起点に、どのような人工知能が生み出せるかということについて議論することは、かつての議論とは異なる結論やイメージが掘り起こされる可能性がある。しかし一方で、そこで議論されている内容は、1970 年の議論とあまり変わらないものであるということについて、その理由についての探索の余地があると考えられる。

また、さらに時代をさかのぼって、1950 年にアラン・チューリングが発表した「計算機械と知能」と題する論文について、そこで取り扱われている内容を見てみる。その論文では、今ではチューリングテストとして知られる人工知能を判定する方法が示されている¹¹⁰。チューリングはそれに先立ち、1937 年に「計算可能性数」に関する論文において、チューリング機械の理論を創設している¹¹¹。それらの論文において、チューリングは「機械は思考できるか、という疑問を考察することを私は提案する」と述べており、その時代において既に、機械が人と同じように思考する可能性について考えていたといえる。ダグラス・R・ホフスタッターの著によれば、チューリングのこの論文では、「チューリングがそれぞれの意思の系をどこまでもたどっていき、ある段階では見かけ上の矛盾を掘り起こしながら、彼自身の概念

¹⁰⁹ 石井 治, 大須賀 節雄, 後藤 英一, 末包 良太, 関口 茂, 辻 三郎, 戸川 隼人, 三浦 大亮, 矢島 敬二「人工知能特集号：座談会：人工知能について」情報処理 11(11), 1970-11-15

¹¹⁰ 「チューリングを読む」チャールズ・ベゾルド, 日経 BP 社, 2012 年 6 月 11 日: x

¹¹¹ 「ゲーデル, エッシャー, バッハ——あるいは不思議の環」ダグラス・R・ホフスタッター, 白揚社, 1985 年 5 月 15 日: 583

を練り上げることによって、より深い分析のレベルでそれを解決するのが常であった」ことが見て取れると述べており、このような取り組みにおいて思考するコンピュータとのやり取りはどのようなものかを考えることで、具体的なイメージにつながっていたのかもしれない。さらに、ホフスタッターは、「この論文は、コンピュータ開発が大きな進歩をとげ、人工知能に関する集中的な研究がなされたこの三十年に近い歳月を経た今日でも、依然光彩を放っている」と述べている（ホフスタッターの著作は、1979年に書かれている）。特にその中（計算機械と知能）では、チューリング自身が考えた内容に対する様々な反論や議論に対して、質疑形式でのやり取りを通しその内容がイメージしやすいものとして書き下されており、現代でもその内容の検証は機械上の知能についての議論には有用なインプットになると考えられると述べている。

（2）クリエイターによる人工知能構想に関する取り組みについて

クリエイターの取り組みを見ていくうえでは、クリエイターの想像のスタート地点を考慮し、作品がつくられた時代が、現代につながる人工知能の概念が世の中に提示される前と、それ以降に分けてみていくこととした。

まず、人工知能の概念が登場する前の段階における著作物について見てみる。一つには研究者ではない立場で人工知能を取り扱ってきた存在として、サイエンスフィクション（SF）がある。SFでは、技術考証なども含め、取り扱うテーマについての技術的背景を理解したうえで様々な物語が創作されているが、中にはその時点で全く存在しない技術を扱った物語も登場する。そこには人工知能という言葉すらなかった時代においてでさえ、人工知能のようなものが描き出されている。例えば、1920年に書かれた、チェコのカレル・チャペックによる『R.U.R』では、意思を持ち自分でものを考えるロボットが登場する¹¹²。

さらに、SFではないが、フィクションの世界でも、アラン・チューリングが指摘するように¹¹³、人間を知的能力で凌駕する機械をイメージするようなことを記述した小説が¹¹⁴、1872年には出版されている。

これらの舞台では既に、機械が知性や意思を持ち人間に対し害をなす存在になることが描かれており、このような知性を持った機械との共生をどう図るかということが題材になっている。このような描かれ方は、時代の背景もあるが、当時の機械の進歩からみると、将来そのようなことが起こっても不思議ではないと考えていたのではないかと見ることもできる。

※There would be no question of the machines dying, and they would be able to converse with each other to sharpen their wits. At some stage therefore we should have expect the machines to take control, in the way that is mentioned in Smuel Butler's 'Erewhon'.¹¹⁵

現在の人工知能においても、そのような恐れをもって取り扱われる面と、単に利便性や技

¹¹² 「ロボット (R.U.R)」カレル・チャペック、岩波文庫、1989年4月17日

¹¹³ Turing, Alan. "Intelligent machinery, a heretical theory (c. 1951)." B. Jack Copeland (2004): 465.

¹¹⁴ 「エレホン」サミュエル・バトラ、岩波文庫、昭和10年2月25日初版、昭和34年2月25日第八刷

¹¹⁵ Turing, Alan. "Intelligent machinery, a heretical theory (c. 1951)." B. Jack Copeland (2004): 465.

術の進化を追求した研究開発という二つの側面がみられる。前者については、アシロマ AI 原則に代表されるもので、それに続き様々な組織や機関が人工知能の開発や利活用における倫理などについての宣言を行っている。また、後者については昨今前者の影響を受け、多くの先端的な取り組みを行っている企業などではそれに従った研究開発を行う傾向にある。

なお、アシロマ AI 原則は、SF 作家のアイザック・アシモフが提唱した「ロボット三原則」の拡張版と見る向きもある¹¹⁶。

このように過去 SF 等の創作物やその創作者から発せられたメッセージが、現代の人工知能研究に与えた影響は大きいと見ることができる。また、チューリングのように計算機の可能性に対し、あたかも意思を持つ機械が存在するとしたらどのようなやりとりになるかまでも想定し、それを文章の形で書き出された内容が、将来にわたり人工知能の分野に影響を与え続けているものも存在する。

本取り組みにおいては、「今の技術のさらに先にある人工知能とはどのようなものなのか？」という問いに答えるために、ここまで登場した、サミュエル・バトラやアラン・チューリングといった、現代でも引き合いに出させるような概念を示した著作に類するものを、今のコンピュータ環境を知ったうえで創作されたものなどから選び出し、それらが差し示すこの先にある人工知能の在り方とはどのようなものかを探索することを目的として活動を行う。

6.2.2 事例抽出の進め方について

第三世代の人工知能により実現可能であるとされている領域の多くは、「人工物理システム」もしくは、「人工抽象システム」に分類されるものがほとんどであった。今後、この枠組みを突破し「人間活動システム」にまで対応可能な人工知能を構想するための一方策として、0 にあげたように、「現在の人工知能やコンピュータの環境が登場する前に書かれた作品」と、「現在のコンピュータ環境を知ったうえで、それを超える人工知能を描いた創作」を対象として、今後未来にわたり参照可能となるようなイメージを数多く集めることを目的として、フィクションとして作られた小説や映画、ゲームなどの作品からそのエッセンスを書き出すことを行うこととした。

(1) 先行研究事例と本稿でのアプローチの考え方について

本アプローチに対する先行的な取り組みを行っている、中京大学溝渕氏は「私達は人工知能と人間をどのようなものとして理解し、どのように共生していくかという問題に直面している。これまでに創作された人工知能や人工知能を想起させる存在が登場するフィクションは、こうした問題を考えていくうえでの示唆を私達に与えてくれるはずだ」¹¹⁷と述べており、フィクションを捉える枠組みとして、人工知能と人間の共生の観点を与えている。溝渕氏の研究においては、映画における人工知能の描かれ方を中心に上げる手法を採用している。

¹¹⁶ <https://www.tmrsearch.co.jp/sensor/wp-content/uploads/sites/2/2017/03/SENSOR35.pdf>

¹¹⁷ 溝渕久美子「新しい『他者』としての人工知能」人工知能 33(5) 645-652(2018.9)

本取り組みにおいては、以下の情報量のフレームに基づき、以下の3つのカテゴリー（文字の情報から理解するもの、映像や画像を通して理解するもの、利用者がそのまま体験して理解するもの）に分けて、インプットを収集整理することとする。

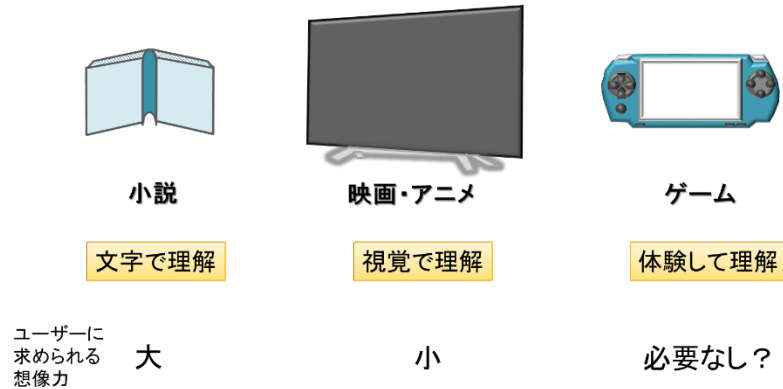


図6-1 創作物から情報を取り出す際の読み手の想像力の分類

なお、多くの映像作品として創作されたものには、小説やマンガとして創作されたものをベースにしたものも多く存在する。そのような作品については、元となる小説などの作品から当たることとし、そこに文章として描かれている人工知能が、どのような映像表現として描き出されているかといったことについても、考察を行うことを検討する。

さらに、このような事例においては、小説が1970年代に作成されたものであって、2000年代にやっと映像化されるようなものも存在する。そのような場合、小説が創作された時点では現代の環境が全くない中で、現代においても近未来的と評される作品になっている点についても考察を加えることを検討する。

また、人工知能を取り扱っている作品には、人工知能がロボットに実装されており、その主題がロボットに偏っているような作品や、人間の意思や脳の記憶を機械へ移し替えるような描き方をしている作品など、人工知能の扱われ方も様々である。本取り組みにおいては、単にどのように描き出されているかを取り出すだけでなく、人工知能の位置づけを含めた事例の収集を図り、描かれ方の分類などを検討することを考えている。特に、小説やアニメ、映画からのエッセンスの書き出し方については、先行研究なども参考にし、検討を進めることとする。なお、前出の溝渕氏の評価の仕方も参考にする。

事例収集の対象となる作品の選定については、前述の現代の人工知能の登場前後の時代関係なども踏まえ、以下の手順で対象となる作品を選定した。

- 1) 昨今映画などで取り上げられた作品とその原作
- 2) 名作と呼ばれる作品（フィリップ・K・ディック、ウィリアム・ギブスンの作品など）
- 3) 21世紀に書かれたSF小説の書評からのピックアップ¹¹⁸

¹¹⁸ 「21世紀SF1000」大森 望、ハヤカワ文庫 2011年11月10日

さらに収集された事例については、**エラー！参照元が見つかりません**。で進める「社会的情動の人文解析モデル開発」の検討過程で必要となる、社会的情動のモデルを描くために必要となる情報の生成を意識するとともに、**エラー！参照元が見つかりません**。で採用されている現象学による「本質観取」による作品からの書き出し行うための、方法論の検討も併せて実施する。

なお、先行的な取り組みとして、取り上げた溝渕久美子「新しい『他者』としての人工知能」 人工知能 33(5) 645-652 (2018.9) において、溝渕氏はSF作品に登場する人工知能については、「現実の人工知能の研究・開発と歩調を合わせるように、一方ではそれを凌駕する想像力によって、多様な形で描かれてきた」と述べている。また、映画に注目する理由として、映画における人工知能の描かれ方であり、映画というメディアは私達人間の知覚の在り方や世界の経験の在り方を如実に映し出すメディアであると述べている。

映画の中の人工知能を評価するにあたり、「『新しい他者』としての人工知能」という観点で、2000年以降の作品を取りあげている。なお、2000年以降というのは、1990年代から現代までにかけての人工知能の研究開発やその成果の製品化という状況が大きく影響していると述べている。

また、2016年から2017年においては、既に「ワトソン」や「AlphaGo」など我々が直接利用できないものだけでなく、人工知能を利用した身近な製品が多数存在していることから、「(人工知能は) 私達の日常に欠くことのできないものになった。人工知能が浸透するにつれ、遠い場所にいる単なる『悪しきもの』というよりも、身近な新しい『他者』として私達の前に立ち現れてきたのだ」と述べているように、溝渕氏の分析においては、作品で扱われる人工知能の技術的な先進性よりは、人間との関係性について着目し、整理を進めていると見ることができる。

この点では、2000年代以降の人工知能映画の特徴として、「人工知能が画像や音声、文字テキストなどのデータから「思考」や「学習」をし、人間と会話を主としたコミュニケーションを図ることそれ自体が主題となる作品が登場した」ことをあげており、そのことが人間と人工知能との関係に「他者」という概念を持ち合わせることが可能になったとみていることが見て取れる。それによりそのような作品では、「未来の存在を描いているのにもかかわらず、現実の人間社会における「他者」との関係の結び方や、現在までに作られた映画での「他者」をめぐる表現が参照される」と述べている。

逆に、コンピュータの登場前の時代では、「入力→処理→出力」の概念もないため、データをインプットするといったことは描かれず、単に人間とやり取りをするといった描かれ方がなされる。このような見方についても、書きだしたエッセンスをどう評価するかという点では考慮すべきだと考える（「他者」の話ではないが）。

なお、「他者」との関係について見ていくことは、「描かれる私的な領域における「他者」としての人工知能の表象を見ていくことによって、これからの人工知能と人間の関係だけで

なく、私達の現在の「他者」との関係を再帰的に考えることもできるだろう」と述べている。

実際に溝渕氏がエッセンスとして取り出した映画の内容は、前述の人間との関係に着目した点が多くみられる。

- ・『チャッピー』での抽出の内容

この映画の中で、学習する人工知能は無知で無垢な子供として扱われ、「パパ」、「ママ」と称する人間達によって教育される。その教育の在り方を対照的に描く構図が人工知能と人間の関係の在り方の善悪の構図に落とし込まれる

本取り組みにおいても、溝渕氏が着目した「他者」に代わるものを設定し、その目線でエッセンスを抽出するといったアプローチも一つの方法として考えられる。

6.2.3 対象となる作品の選定について

対象とする作品については、本稿では我々が本質観取を行う際に、本研究に携わる我々の想像力を最も有すると考えられる文字情報中心の小説を対象とすることとした（図 5-1 にあげた区分による）。

クリエイターが想像した小説において、人工知能に関する事例からの本質観取の対象となる作品の選定においては、前述のとおり二つの時代区分（現代の人工知能の概念の登場前・後）に創作されたものから数点を取りあげる。なお、時代区分うち、前者に該当するものとしては、19 世紀に創作された SF 小説から取り上げることとした。また、後者に該当するものについては、対象となる作品が非常に多く存在するため、前項で取り上げた「21 世紀に書かれた SF 小説の書評からのピックアップ¹¹⁹⁾」として、以下に説明する対象作品の絞り込みを行った。

時代区分前期に該当する作品としては、カレル・チャペック著の『R.U.R.』を取りあげ実際にそこから事例をとりだした。

また、時代区分後期に該当する作品については、前述の書評に掲載された 1,000 冊の作品について、以下の手順により対象とする作品を絞り込むとともに、それら作品の特徴づけまでを本稿作業の対象とした。

(1) 時代後期の作品の絞り込み手順について

前述のとおり時代後期に登場する作品は非常に多く、またその作品にて人工知能がどの程度取り上げられているかについてもばらつきが大きいいため、まずは前述の書評をもちいて作品の絞り込みを行うとともに、作品全体の傾向を把握することとした。

その手順は以下のとおりである。

- 1) 書評に「人工知能」や「ロボット」に関連する記述と考えられる作品の取り出し（一

¹¹⁹⁾ 「21 世紀 SF1000」大森 望, ハヤカワ文庫 2011 年 11 月 10 日

覧化ロングリスト作成)

2) 取り出した作品の紹介(発行社の作品紹介メッセージ)などから、さらに作品全体の特徴分析

3) 事例読み取りの対象とする作品が有する特徴の決定

本稿の作業はここまでであり、対象 92 作品からの対象作品の取り出しと、また、ここで得られた特徴量にもとづく対象範囲とある作品範囲の拡大については、次のステップの検討としている。

(2) 対象とした作品の一覧化とロングリスト(時代区分後期)

対象となる作品は、書評に含まれる内容をもとにして 92 作品が取り出された。なお、出版年は 2001 年から 2011 年の範囲になるが、実際に出版されたのはそれ以前で、再度出版社が変更になって発行された作品も含まれている。さらに、海外の作品は原語で出版された年度と日本語化されて出版された年度の違いなどがある。

表 6-6 ロングリストとして取り出された作品の分類

年度	国内	国外	その他	小計
2001	12	10	0	22
2002	8	4	0	12
2003	8	1	0	9
2004	2	1	0	3
2005	6	4	1	11
2006	5	5	0	10
2007	3	5	0	8
2008	1	3	0	4
2009	5	1	0	6
2010	7	0	0	7
小計	57	34	1	92

(3) ロングリストをもとにした作品の特徴分析

ロングリストに取り出された各作品について、作品を提供している出版社の作品紹介文や Amazon などの販売サイトに掲載される作品紹介文(主に「BOOK」データベース)を取り出し、作品紹介文に含まれる語句を用いた共起分析を行った。

共起分析の結果えられた語句の共起マップは以下の通りとなった。

6.2.4 『R. U. R.』から取り出した事例について

物語は「序幕」「第一幕」「第二幕」「第三幕」の四部で構成されている。各部の舞台は以下のような状況にある。

- ・序章：ロボットが開発される経緯の説明と、実際にロボットが普及した時期の物語。登場人物紹介によれば、「ロボットは序幕では人間のような服装。動作と発音はボツンボツンととぎれるようになされ、顔つきは無表情で、じつとのぞきこむようにする。」とある。
- ・第一幕：序幕から十年後であり、ロボットが人間に対して反乱を始めるタイミング。第一幕以降ではロボットは、「腰のところで紐でしめた布製の作業衣をつけ、胸には真鍮の番号をつけている。」とある。
- ・第二幕：第一幕に続き、ロボットが人間に対して反乱を開始した後。まだ、この時点では人間は生存している。
- ・第三幕：ロボットが人間に対して反乱を起こしたのちで、ほとんどの人間は滅ぼされてしまっており、残る人間はアルクビストのみになっている。

これら四つの舞台において、本作品を読み進めるなかで、他者としてのロボットと人間との対話について読み出た自分自身がどのように感じたかを整理するために以下の問を立てた。

- ① 人間とは異なる他者との対話で人間同士の対話と異なることは何か
- ② 人間とは異なる他者との対話であっても人間同士の対話と変わらないことは何か

作品を読み他者との対話（独白、心の声も含む）を取り出したところ、取り出した会話は以下の三タイプに分かれた。

- 1) 人間同士の会話
- 2) 人間とロボットの会話
- 3) ロボット同士の会話

この三つのタイプと四つの各部で取り出されたものは以下のような状況になった。表中の数値は取り出された各対話のタイプに該当する一連のやり取りの組の数である。対話の組は全部で 57 個取り出された。

表 6-7 R.U.R から取り出した会話の種類

	人間-人間	人間-ロボット	ロボット-ロボット	小計	ページ数
序幕	19	5	0	24	48
第一幕	5	1	0	6	55
第二幕	9	2	0	11	44
第三幕	2	9	5	16	31
小計	35	17	5	57	

各部に登場する三つの会話のタイプから「人間と他者としてのロボットの対話」には以下のような特徴が見て取れた。

(1) 序幕

序幕では、ロボットが世の中に普及し始めた状態で、「ロボット生産者であるドミンら」と「大資本家の令嬢であり人道同盟というロボットを保護しより良い扱いを受けるようにすることを目的にしている団体に所属するヘレナ」との間の会話を中心となっている。

ドミンらはロボットをあくまでも機械として扱い、ヘレナはロボットも人間と同じようにとらえており、ロボットにも人権のようなものが必要であると考えていると見ることができた。

そのようななかで、行われる人間とロボットの会話では、ロボットからは人間的なことは一切感じられない機械的な受け答えだけである。しかし、ロボットにも人間と同じ扱いを望んでいるヘレナは、あくまでも人間へ対峙するかのような対応を行っているかのように感じられる。

表 6-8 序幕における人間とロボットの会話と感じ取ったこと

分類	人物	一連の会話	感じとったこと
人間	ヘレナ	(立ち上がり、手を差し伸べる) 初めまして。こんな辺鄙なところへ来てきつととても寂しいでしょうね。	・ロボットには寂しいという感情はないものだ
ロボット	スラ	グローリー様、そのようなことは存じておりません。どうぞおかけ下さい。	
ロボット	スラ	ここの工場です。	・相手の質問を理解し回答はできても、相手の本当の意図までは理解できないものだ
人間	ヘレナ	そう、あなたここで生まれになったのね。	
ロボット	スラ	はい。ここで作られたのです。	
人間	ヘレナ	(飛び上がる) 何ですって？	
人間	ドミン	(笑いながら) スラは人間ではありません、お嬢様。スラはロボットなのです。	・ロボットは人間よりも正確に物事を覚えており、聞かれたらその内容を詳細に説明できるものだ ・ロボットはその時の状況に合わせ相手にとって良いと考えられることを判断しおすすめることができるものだ
人間	ドミン	スラ、お客様とお話し下さい。とても大切なお客様です。	
ロボット	スラ	どうぞ、お嬢様、おかけ下さい。(二人は坐る) ご旅行はいかがでしたか？	
人間	ヘレナ	ええ——け——結構でしたわ。	
ロボット	スラ	“アメリエ号”でお帰りにならない方がよろしくございません？ 晴雨計の針はうんと下がって、七〇五を指しております。“ペンシルバーニア号”をお待ちになったらいかがでしょう。この船ならとてもいい船で、とても強力な船ですから。	
人間	ドミン	どのくらいかね？	
ロボット	スラ	時速二十ノット。一万二千トンです。	
人間	ドミン	(笑いながら) もう結構、スラ、もういいよ、フランス語がどれほど上手か見せてごらん。	・ロボットは様々な言語を使うことができるものだ
人間	ヘレナ	あなたフランス語ができるの？	

分類	人物	一連の会話	感じとったこと
ロボット	スラ	四ヶ国語できます。Dear Sir! Monsieur! Geerter Herr! Cteny pane!と、英・仏・独・チェコ語を書き分けます。	
人間	ヘレナ	(跳び上がる)これはいんちきだわ。あなたはいかさま師よ！ スラはロボットなんかじゃないわ。私と同じような若い女よ。スラ、よくまあこんなことを——なんでこんな茶番劇をなさるの？	・ロボットは自分がロボットであることを理解しているものだ
ロボット	スラ	私はロボットです。	
人間	ヘレナ	違うわ、違うわ、うそを言っているのよ。スラ、許して、分かっているわ——宣伝のためにあなたに無理にやらせているのね。スラ、あなたは私たちと同じ若い女なのでしょう？ 言ってちょうだい！	
人間	ヘレナ	そんなこと関係ないわ。ロボットだって私たちと同じよい人間だわ。スラ、あんたったら自分を切らせようっていうの？	・ロボットには死ということが何かは分からないが壊れるということは分かっているものだ ・しかし、人間は壊れる＝死であると思っているものだ
ロボット	スラ	はい。	
人間	ヘレナ	えっ、あなた死ぬのがこわくないの？	
ロボット	スラ	存じません、グローリー様。	
人間	ヘレナ	そうしたらあなたはどうなるか知っているの？	
ロボット	スラ	はい、動かなくなるでしょう。	
人間	ヘレナ	なんていう恐ろしいことを！	
人間	ドミン	マリウス、お嬢様に自分がなんであるか申し上げなさい。	・ロボットは命令であれば仲間を破壊することに抵抗を抱かないものだ
ロボット	マリウス	ロボットのマリウスです。	
人間	ドミン	お前さんはスラを解剖室に連れていくかね？	
ロボット	マリウス	はい。	
人間	ドミン	気の毒と思うかね？	
ロボット	マリウス	知りません。	
人間	ドミン	スラはどうなるんだろうね？	
ロボット	マリウス	動くのをやめるでしょう。それを粉砕機に入れることになります。	・ロボットは自分の死の概念だけではなく、ロボット自体に死の概念は持ち合わせていないものだ ・また、ロボットは命令によって壊されてしまうことに対してそれに抗うという心理は働かないものだ
人間	ドミン	それが死だよ。マリウス。死が怖いかね？	
ロボット	マリウス	いいえ。	
人間	ドミン	どうです、グローリー様、ロボットたちは生きること执着していないのです。そもそも、なんのために生きるかを知らないし、生きる喜びを持たないのです。あの連中は雑草以下なのです。	・ロボットの生産者はロボットは人間と同じような感情を持っていないと考えるものだ

(2) 第一幕

第一幕では、当初から十年がたちロボットに何らかの意思が芽生え人間に対して抵抗を示すようになっていく。ロボットが人間のような感情を持ち合わせたからなのか、それとも合理的な判断のもと（命令されて実行している生産の効率化など）人間は非生産的なものとみなすに至ったのかについての記述はないが、後者のイメージがその裏にあると感じることが

できる。ロボットの意思は、あくまで合理的な判断ができるレベルにしかないと考えられる。

表 6-9 第一幕における人間とロボットの会話と感じ取ったこと

分類	人物	会話	感じとったこと
人間	ヘレナ	可哀そうなラディウス、あなたまでかかったのね？ どうにかならなかったのね？ ねえ、こうなるとあなたも粉碎機送りよ！——あなたったらしゃべりたくないの？——ねえ、ラディウス、あなたは他の者たちよりも上等なのよ。あなたを他の者と違うように作るのにガル博士はとっても苦労したのよ！——	<ul style="list-style-type: none"> ・ロボットでもより優秀なものは人間のために働きたくはないと思うものだ ・優秀なロボットであっても粉碎されることが死であるということは分からないものだ
ロボット	ラディウス	私を粉碎機送りにして下さい。	
人間	ヘレナ	あなたを殺させるなんて、私にはとてもお気の毒だわ！ なんで自分のことに注意しなかったの？	
ロボット	ラディウス	あなた方のためにはもう働きません。	<ul style="list-style-type: none"> ・優秀なロボットは人間に代わって何でもできると判断するものだ ・優秀なロボットは人間のおしゃべりなどは無駄なものだと考えるものだ
人間	ヘレナ	どうして私たちのことを憎んでいるの？	
ロボット	ラディウス	あなた方がロボットのように有能ではないからです。あなた方がロボットのように有能ではないからです。ロボットが何もかもします。あなた方きたらただ命令するだけです。余計なおしゃべりをしているのです。	<ul style="list-style-type: none"> ・ロボットは図書館にある情報をすべて吸収し世界中の知識を身につけることができるものだ ・人間はロボットと人間の能力が同じということは知識がたくさんあるかどうかではんだんしてしまうものだ
人間	ヘレナ	あなたはわざと、そんなことを言っているのよ！ ガル博士はあなたに他の者たちより大きな頭脳をあげたのよ。私たちのより大きい、世界で一番大きいのを。ラディウス、あなたは他のロボットたちとは違うのよ。あなたは私の言っていることが分かるはずよ。	
ロボット	ラディウス	私には主人なんかいません。すべて自分でどうしたらいいか知っています。	
人間	ヘレナ	だから私はあなたが何でも読めるように図書係にしたのよ——ねえ、ラディウス、私はあなたが世界中に、ロボットは私たちと能力が同じだと示して欲しかったのに。	<ul style="list-style-type: none"> ・優秀なロボットは全てを（人間を含め）支配すべきという判断に至るものだ
ロボット	ラディウス	私には主人なんかいません。	
人間	ヘレナ	誰もあなたに命令なんかしないわ。あなたは私たちと同じよ。	
ロボット	ラディウス	私は他の者たちの主人になりたい。	
人間	ヘレナ	あなたをきつとあとでたくさんのロボットの管理人にするわ、ラディウス。あなたはロボットたちの教師になるのよ。	
ロボット	ラディウス	私は人間たちの主人になりたい。	
人間	ヘレナ	あんたったら気が狂ったのね！	

分類	人物	会話	感じとったこと
ロボット	ラディウス	私を粉砕機に入れても結構です。	・人間はロボットが合理的な判断をしている(と感じた)ことの意味が理解できないものだ
人間	ヘレナ	私たちがあなたのような気のふれた者をこわがると思って？（小さな机に坐り、メモを書く）いいえ、そういうことではないの。ラディウス、このメモをドミン社長に渡してちょうだい。あなたを粉砕機にかけないように！（立ち上がる）なんて私たちのことを憎んでいること！この世に愛するものがないの？	

(3) 第二幕

ロボットが人間に対して反乱を起こし、人間はロボットから逃れようと躍起になっている。ロボットは、次第に人間を追い詰めていく。ロボットはあくまでも、機能的・性能的に優れていることが重要であり、それに劣る人間には必要がないと判断するレベルの思考しか持ち合わせていないと考えることができる。

表 6-10 第二幕における人間とロボットの会話と感じ取ったこと

分類	人物	会話	感じとったこと
ロボット	ラディウス	片づいたか？	・ロボットは不必要であれば人間であっても殺してしまう(壊してしまう)ものだ ・生産のためにつくられたロボットは生産に価値を提供できるものは人間であろうがロボットであろうが価値があると思えるものだ ※生とは何かは分かっていないものだ
ロボット	他のロボット	はい。	
ロボット	二人のロボット	(アルクビストを引っ張ってくる)こいつは射ちませんでした。殺しますか？	
ロボット	ラディウス	やれ。(アルクビストを見て)ほっておけ。	
ロボット	ロボット	こいつは人間です。	
ロボット	ラディウス	これはロボットだ。ロボットと同じように手を動かして働いている。家を建てている。働くことができる。	
人間	アルクビスト	俺を殺してくれ。	
ロボット	ラディウス	仕事をするのだ。家を建てるのだ。ロボットたちはうんと建てるぞ。新しいロボットのために新しい家を建てるぞ。お前はそのものに仕えるのだ。	・ロボットは人間もロボット同様であると考え、ロボットよりも能力が低い(生産や知識)ものは必要ないと思えるものだ
ロボット	ラディウス	(バリケードの上に昇る)世界のロボットよ！ 人間の権力は地に落ちた。工場の占領により、われわれはあらゆるものの支配者となった。人類の時代は終わった。新しい世界が来たのだ！ ロボットの国家だ！	
人間	アルクビスト	死んでいる。	

分類	人物	会話	感じとったこと
ロボット	ラディウス	世界はより力のあるものたちのものだ。生きたいと思うものは制覇しなければならない。われわれは世界の支配者だ。陸と海を支配する国家だ！ 星を支配する国家だ！ 宇宙を支配する国家だ！ 場所、場所をもっと多くの場所をロボットに！	
人間	アルクビスト	(上手のドアのところで)なんていうことをしてくれたのかね？ 人間なしでは亡んでしまうんだぞ！	
ロボット	ラディウス	人間はいない。ロボットよ、仕事に取りかかれ！ 行進！	

(4) 第三幕

ロボットが人間を滅ぼしてしまったときに、ロボットにないもの（欠陥）に気付く。人間がロボットを作ったときのレシピが失われてしまって、新たに生産することができなくなっていた。そのことに対して、ロボットは冷静に自分たちを維持するために必要なレシピを改めてつくり出すことを求める。これは、死への恐怖ではなく、ロボット自体がなくなってしまうことを防ぐための機能により生じるものだったが、次第に自分自身の死について理解し、個々のロボットに感情のようなものが生じてくる。

表 6-11 第二幕における人間とロボットの会話と感じ取ったこと

分類		人物	会話	感じとったこと
人間		アルクビスト	(坐る)ロボット諸君は何をしろというのだね？	・ロボットは自分自身を増やすことができないことは認識できているが、その方法を考え出すことはできないものだ(人間だけがそれをなし得ると考えているものだ)
ロボット		一号ロボット [ラディウス]	先生、機械は仕事をすることができません。ロボットを増やすことはわれわれにはできないのです。	
人間		アルクビスト	人間を呼びなさい。	
ロボット		ラディウス	人間はいません。	
人間		アルクビスト	人間だけが生命を増やせるのだ。私の邪魔をしないでくれ。	・ロボットとしての「怖い」という感情は自分たちが増やせない(滅んでしまう)ということから芽生えたものだ
ロボット		二号ロボット	先生、どうかご慈悲を。わたくしたちはとても怖いのです。私たちのしたすべてのことのつぐないをします。	
ロボット		三号ロボット	人間には生の秘密が分かっていたのです。その秘密を話してください。	・ロボットは自分たちが増えないことを死と捉えているものだ(自己の死という概念はない)
ロボット		四号ロボット	話して下さらないと、われわれは死にます。	
人間		アルクビスト	言ったじゃないか——人間を探し出せて、言ったらう。人間だけが子供を作ることができるのさ。生命を再生することが。昔あったものを何もかも取り戻すこと	・ロボットは人間と自分たちの違いについてはその能力の差でしかないと判断するものだ

			が。ロボットの諸君、お願いだから人間を探してきてくれ！	・ロボットは歴史を見て人間は人間を殺すものだ判断し、どのようにロボットが人間を殺すべきと判断するものだ
ロボット		四号ロボット	どこもかしこも探して見ました、先生。でも人間はいません。	
人間		アルクビスト	おお、何ということ、なぜお前たちは根こそぎ殺してしまったのだ！	
ロボット		二号ロボット	私たちは人間のようになりたかったのです。人間になりたかったのです。	
ロボット		ラディウス	私たちは生きたかったのです。私たちの方が能力があります。私たちはすべてのことを学びました。すべてのことができます。	
ロボット		三号ロボット	あなた方はわれわれに武器を与えました。われわれは主人にならないわけにはいかなかったのです。	
ロボット		四号ロボット	われわれは人間の欠点に気がついたのです。	
ロボット		ダモン	人間としてありたければ、お前たちは殺し合い、そして、支配をしなければならぬのだ。歴史を読んでみるがいい！ 人間の本を読んでみるがいい！ 人間でありたいのなら、支配しなければならず、人間を殺さなければならぬ！	・ロボットでも「恐怖」や「痛み」を理解するとより人間らしくなるものだ
人間		アルクビスト	ロボットは生命ではない、ロボットは機械さ。	
ロボット		二号ロボット	われわれは機械でした。先生、でも恐怖と痛みから別なものになったのです——	
人間		アルクビスト	何にかね？	
ロボット		二号ロボット	魂になったのです。	・ロボットの死に絶えてしまうという恐怖は、自己が死んでしまうという恐怖ではなくロボットという種が減ることだ
ロボット		ダモン	生きているロボットで実験をしろ。どうやって作るかを見出すのだ！	
人間		アルクビスト	生きている身体をだって？ なに、この私が殺せたと？ 私は一度だってそんなことを——黙れ、ロボット！ お前に言うておくが、私はもう年を取りすぎている！ 見ろ、見えないかね、私の指がふるえているのが？ メスを持っていられないのだ。見えないかね、目が涙でいっぱいなのが？ 自分の手だって見えないだろうよ。いや、いいえ、私にはできない！	
ロボット		四号ロボット	生命が絶えてしまいます。	
人間		アルクビスト	(上手から飛び込んできて、血だらけの白衣を投げずる)私にはできない、できない！ 神様、なんというおそろしさ！	・ロボットは種の保存のためには個々のロボットが死ぬようなことでも恐怖を感じないものだ
ロボット		ラディウス	(解剖室のドアのところで)切ってください。まだ生きています！	

ロボット		ダモンの叫び声	切れ！ 切るんだ！	・ロボットでも本当に自分自身が死に直面すると生に対する感情が出来上がるものだ
人間		アルクビスト	あれをすぐ運び去れ！ ききたくないのだ！	
ロボット		ラディウス	ロボットたちはあなたより我慢ができます。(立ち去る)	
ト書き			上手から血まみれの敷布でつつまれたダモンがよろけながら出てくる。	
人間		アルクビスト	(あとずさりする)何の用だ！ 何しに来たのだ？	
ロボット		ダモン	俺は、い、生きている！ い、い、生きている方がいい！	
ロボット		ヘレナ(ロボット)	(鏡の前で)私がきれいですって？ ああ、このひどい髪ったら、せめて何か飾りをつけられれば！ ねえ、庭ではいつも花を髪に挿すのよ。でもあそこには鏡もないし、誰もいないの——(鏡をのぞきこむ)お前はきれいだよって？ なぜきれいな？ 私を悩ましていてだけの髪の毛がきれいな？ いつも閉じているその目が美しいの？ 痛むようにと噛みしめるだけの唇が美しいだって？ 美しいってことは何のためなの、何なの？——(鏡の中のプリムスを見て)プリムス、あんたなの？ こっちへ来て、二人が並んでうつるように！ 見て、あんたの頭、私のと違うわ。肩も違うし、口も違うわ——ああ、プリムス、あんたどうして私のことを避けようとするの？ 何で一日中あなたの後を追いかけていなければならないの？ そして、それなのに、私がきれい何て言うの？	・ロボットにも自分(個)としての感情(美しい、他のロボットと違う)があるものだ ・ロボットであっても内から湧き上がる何かで笑ったりドキドキしたりするという感情が芽生えるものだ(それが人間らしさになる)
ロボット		プリムス	ヘレナ、君が僕から逃げていくんじゃないか。	
ロボット		ヘレナ(ロボット)	これは何というか仕方なの？ きてごらん！(両方の手でプリムスの髪をとかす)さっ、さっ、プリムス、あんたに触れるのはとってもいい感じよ！ 待って、あんたもきれいじゃなくっちゃ！(洗面台より櫛をとり、プリムスの髪を前へとかす)	
ロボット		プリムス	ねえ、ヘレナ、急に心臓がドキドキするってということじゃない、今、今、きっと何かが起こるに違いないっていう風に——	
ロボット		ヘレナ(ロボット)	(笑い出す)自分を見てごらんなさいよ！	
人間		アルクビスト	(起き上がり)なに—なんだ、笑い声？ 人間かな？ 誰が戻ってきたのだ？	
ロボット		ヘレナ(ロボット)	(櫛をはなす)プリムス、私たちに何が起こるっていうの？	
人間		アルクビスト	(二人の方へよろけるように)人間か、あんたたち——あんたたち—人間なのかね？	

人間		アルクビスト	よろしい、あの娘を解剖室へ連れてきておくれ。あの子を解剖するから、	・ロボットでも他人のために自己を犠牲にするという感情が芽生えるものだ
ロボット		プリムス	ヘレナをですか？	
人間		アルクビスト	そうだよ、そうとも、そう言っているんだ。言って用意万端ととのえておいておくれ——あれ、どうしたんだい、できるね？ あの子を連れてくるように他の者を呼ぼうかね？	
ロボット		プリムス	(重い乳棒をつかむ) 一步でも動いてみろ、お前の頭をたたき割るぞ！	
人間		アルクビスト	それじゃあたたき割るがいい！ さあやるがいい！ それでそのあとロボットはどうするのかね？	
ロボット		プリムス	(ひざまずく) 先生、どうか私を使って下さい！ 私はあの子と同じように作られています。同じ素材で、同じ日にです！ 私の生命をとって下さい。先生！ (ジャケットを開く) ここを切して下さい、ここを！	
人間		アルクビスト	いい加減にしろ、私はヘレナを解剖したいのだ。さあ、急いで。	
ロボット		プリムス	彼女のかわりに、私を使って下さい！ この胸を切り開いて下さい。叫び声一つあげませんし、ため息一つきません！ 百回でも私の生命をとって下さい——	
人間		アルクビスト	ねえ君、そうあわてないで。そんなに生命をぞんざいにするもんじゃない。君だって生きたくないわけじゃないだろう？	・ロボットでも泣くという感情が湧き上がるものだ
ロボット		プリムス	あの人がいらないなら、生きたくなんかありません、先生。あの人なしでは生きたくなんかないのです。ヘレナを殺させるわけにはいきません！ 私の生命をとって、どうってことはないでしょう？	
人間		アルクビスト	そうか、手伝ってくれるね？ プリムスを解剖するよ。	
ロボット		ヘレナ (ロボット)	(叫び声をあげる) プリムスですって？	
人間		アルクビスト	うん、そう、そうだ。そうじゃなきゃならないのだ、分かるね？ 私は——もともと——そう、お前を解剖したかったんだが、プリムスがお前の代わりを申し出たのだ。	
ロボット		ヘレナ (ロボット)	(顔をおおって) プリムスが？	
人間		アルクビスト	そうだよ、それがどうしたかね？ ああ、あんたは泣けるんだね？ そのプリムスとやらで何か問題なのか言ってみろ。	

6.3 映画『空気人形』：人工知能と「心」・「意識」について

6.3.1 本節の目的

(1) はじめに

心や意識を持つ人工知能とは何か、また、それをどのように実現するのか。更に、そのような心・意識を持つ人工知能が将来人間社会に現れた場合、どう扱い、扱われるか、また、どのようなインパクトが人間・社会に対してあるか。前者の課題、すなわち、心や意識を持つ人工知能の原理や実現方法についての研究は様々行われているが、後者の課題、すなわち実現した心・意識を持つ人工知能の扱いや社会的インパクトについてはまだ将来的な話である。具体的な事例もないため分からない部分が多い。分かりようがないとも言える。

本論考は、この点に注目し、映画を材料として考える。映画は監督、原作者、脚本家、撮影者等多数のクリエイティブな人々が制作に関与する。多くの観客に対して訴える内容となることが意図されているため、その内容は荒唐無稽な絵空事という訳でもない部分が多い。またどのような内容であったにせよ、その内容に対して示された共感、反発等の感情、感想を多くの観客について知ることができるというメリットがある。特に、映画は国際的なメディアであるため、国際的な反応の違いを見ることも可能というメリットがある。

人工知能の代表的な研究者の一人である松原（2018: 157 頁）は、「AI が 10 年程度先の未来で私たちの生活や人生に影響を与えたとしたら、それはコミュニケーションかもしれない」と述べている。それについて知るためには、「現在の立ち位置を変えず、視線だけを未来に向けても、たいていは見誤る」ものであり、「立ち位置を未来に変え、未来に存在する視線から、そこで何が起きているかを見ることが必要」と指摘する。まさに本論考で意図していることと重なる。

特に本節においては、心を持つ人工知能（ロボット）について、以下の仮説を持ち、分析を進めていく方針である。ただし、それはここでの分析等を通じて「検証」されるようなものではない。議論の出発点とでも思ってもらえればいい。

1. 心・意識を持つ人工知能が生まれたら、それは周囲の人間の心に影響を及ぼすものになる。人間の心は相手の心があればそれを想像するミラーニューロン¹²⁰が反応する（今の人工知能にはそれは反応していないとみられる）。心を持つ人工知能（あるいはそれを備えたロボット）はそもそも人間の心に影響を与えることこそがその目的となるだろうから、それによって、人間の心・脳が反応し、人間やその関係が影響を受ける。
2. 心・意識を持つ人工知能が活躍するような社会では、ロボットの持つ心をどう倫理的に扱うかということを考える必要が出てくる。例えば、心を持ったロボットを虐

¹²⁰ 「他者の行為を見て、あたかも自分が行動しているかのように反応する神経細胞」。ミラーニューロン説によれば、この細胞により、「自分の自我が他者の自我と同じものだ」と仮定し、自分の自我を他者の自我が置かれているのと同じ状況に置いてシミュレーションすることによって、他者の自我が生成する心がどのようなものか知ることが意識を特にしないで自動的に起こる。（人工知能学会監修『人工知能とは』近代科学社、2016 年、33 頁）

待することは倫理的に許されなくなる可能性がある。逆に、それが許されずとした場合には、そのことが人間間の関係を律する倫理感に影響を与える可能性がある。

いずれにせよ、人工知能の将来、その将来社会へのインパクトは不確実性が高い。特に、人工知能と心や意識の問題について、誰も将来どうなるかははっきりしたことは言えない。理論的には、後述するような様々な考え方が、人工知能が持ち得る心・意識について現在提唱されている。

この原稿では、映画の中で登場人物は心や意識を持つ人工知能に対して、どのような対応をしていたか、また、映画の観客（批評家、一般の観客を含む）はどのような感想を持ったかに注目して、整理する。そのことで上の命題についての考察を深めたい。

心・意識を持つような人工知能の実現はまだ長期的な話であり、映画の中の人工知能ではどのように描かれているか等について自らの観察、観客や批評家による観察を通じて、分析するという本節で取るアプローチは、新たな視点や課題を想像することにつながる事が期待でき、それらの視点や課題は人工知能研究にとって価値がある可能性がある。

(2) 映画『空気人形』

心・意識を持った人形やロボットが主人公や主要登場人物となっている映画は例えば以下がある。

ピノキオ (Pinocchio) : 映画 (米国、1940 年)

鉄腕アトム (日本、1952 年)

AKIRA : 映画 (1988 年、日本)

ブレードランナー (Blade Runner) : 映画 (米国、1982 年)

マネキン (Mannequin) : 映画 (米国、1987 年)

A.I. (A.I. Artificial Intelligence) : 映画 (米国、2001 年)

空気人形 : 映画 (日本、2009 年)

ドラえもん : アニメ、漫画 (日本、1969 年～)

Her/世界で一つの彼女 (米国、2014 年)

トランセンデンス (Transcendence) : 映画 (2014 年)

エクス・マキナ (2014 年、英国) など

これらの中で、2009 年に公開された映画『空気人形』を本セクションの分析対象とした¹²¹。なぜこれらの映画・小説の中で、『空気人形』を分析対象として取り上げるのか。この映画のプロット等については以降に詳しく説明するが、『空気人形』の人形は知能が人間を大きく超えたり、人間の能力を大きく超える能力（知力、体力）を持っている訳ではない。そのような人間を超える能力を有する映画、ドラマの場合にはその能力の描写、利用が話や展開の大きな中心となる。逆に『空気人形』では心を人形がもった場合どうなるかに話の焦点があるため、ロボット・人工知能の優れた「身体能力」「知能」という側面以外を分析する対象として

¹²¹ 取り上げる映画『空気人形』は、2009 年 10 月時点では既に世界の 15 か国・地域での上映公開が決まっていた（読売新聞、2009 年 10 月 2 日、夕刊 12 頁）

はより適切であると考えた。以下のように、身体、知能、感情の3つの次元から人工知能・ロボットを評価する分類の枠組みが考えられるが、「空気人形」は「身体能力：低い、知的能力：低い、感情：高い」のケースに該当する。

- 身体能力、知能、感情のそれぞれの機能・性能の3次元での人工知能・ロボットの能力の評（ $2 \times 2 \times 2 = 8$ 分類）
 - 身体能力：単純な反応 OR 複雑な動き・速度・移動等
 - 知能：単純な刺激応答 OR 高度の計算能力・学習能力・倫理的判断能力
 - 感情・心：単純な喜怒哀楽 OR 愛情・善悪

『空気人形』は、是枝裕和監督の2009年の作品であり、空気人形（ラブドール）が心を持つようになる話である。原作は業田良家¹²²の『ゴータ哲学堂 空気人形』である。業田氏の作品は約20頁の短編漫画であるが、同時に、吉野弘の詩『生命は』『I was born』にも映画の基本的なコンセプトは大きく影響を受けている。これらの作品が合わせて原作であるということもできるだろう。映画のキャッチフレーズは「心をもつことは、切ないことでした」である。

「空気人形」は空気が中に入っているだけの人形である。それが心を持つようになった。心は持ったかも知れないが、優れた知性を持っている訳ではない。その点では通常の人工知能とは異なる。現在の人工知能は優れた知能、計算能力はあるが、心や意識はない。ただし、空気人形はロボットである訳ではなく、心を与える人工知能の装置を備えている訳でもなく、現実的には工学的にあり得ない。

この映画における空気人形は人間の空虚さのメタファーとして描かれる。観客は、人間はこの空気人形と比較した時にどこが違うのか、人間も空気人形と同じように空っぽの中身の空虚な存在に過ぎないのではないかと感じ、現代社会における人間間のコミュニケーションの希薄さなどを考えることとなる。「空気人形」の役は韓国女優ペ・ドゥナが演じているが、是枝監督は、以下のように映画の趣旨を説明して出演を依頼したという。

人間誰しもある種の空虚感を抱えていて、それをみんな、なんとか埋めようとしている。でも実は、自分ではなく他人によって、その空虚感が埋められたときにこそ、豊かなものに感じられるのだと思う。そういうことを空気人形に合わせて描きたい。つまりこれは人形の映画ですが、人間についての映画です。（キネマ旬報社、2009年：54頁）

また、映画公開前のメディアとのインタビューで、是枝監督は以下のように映画の狙い、テーマ等を語っている。いずれも空気人形の物語を通じて、現代社会における人間とその関係性を描くとの意向が読み取れる。空気人形が存在することで、周囲の人間や社会の側が変わるということであり、映画のテーマ自体が本原稿の趣旨にも合っているだろう。

- 「どうやって人と人とかかわるかを描きたかった」「（吉野弘の詩『生命とは』について）空虚とは、実は他者に埋めてもらうものだということらえ方をすると、決して欠落ではなく、

¹²² 業田良家は『機械仕掛けの愛』の作者でもあり、同作品では知能を持つ様々なロボットについて書いている。

つながれる可能性。吉野さんの詩はそれを語ってくれていて、『ああ、この哲学なんだ』と。この映画の結末は決して悲しいだけのものではないはず」（読売新聞 2009 年 10 月 2 日夕刊 12 頁、「空気人形—是枝裕和監督」）

- 「人形の話なのだが背景には人間の物語があるということが見た人の心の中に芽生えるだろう」（最後はハッピーエンディングではないことについて）でも決して終わりではなく、誰かの始まりになっていく。そんな命の捉え方で選んだ結末です。」（朝日新聞 2009 年 10 月 2 日夕刊 6 頁、名古屋本社、「心が宿った人形の決意は—是枝裕和監督『空気人形』」）
- 「僕らの日常は輝きを失っている。人形を通して普段見落としているささいな出来事でも素晴らしいという貴重さを再発見してもらえばうれしい」（中日新聞、2009 年 9 月 25 日夕刊 9 頁「膨らむ切ない恋心—是枝裕和監督『日常の輝き再発見して』」）
- 「最近のファンタジーには毒がない。彼女（空気人形）が目の当たりにした世界だから、厳しいものを避けて通れない。それを人ではないものから人を見る視点で描いた。」「空気だから、空っぽだからこそ、命を感じさせる映画になったと思う」（朝日新聞 2009 年 10 月 9 日夕刊 3 頁、西部本社、「命宿る人形 現代を投影」）

是枝裕和監督は、これまでに、国内では日本アカデミー賞（『海街 diary』最優秀監督賞等）等、海外では、ナント三大陸映画祭（『ワンダフルライフ』グランプリ）、カンヌ国際映画祭（『誰も知らない』主演の柳楽優弥が最優秀男優賞、『そして父となる』審査員賞、『万引き家族』パルム・ドール）等を受けている。現在、国際的に最も注目される日本人の映画監督の一人である。これまでの作品は、『幻の光』（1995 年）、『ワンダフルライフ』（1999 年）、『誰も知らない』（2004 年）、『歩いてても歩いてても』（2008 年）、『空気人形』（2009 年）、『奇跡』（2011 年）、『そして父となる』（2013 年）、『海街 diary』（2015 年）、『三度目の殺人』（2017 年）、『万引き家族』（2018 年）等がある。

2009 年の作品『空気人形』は、カンヌ国際映画祭の「ある視点」部門（Un Certain Regard section, the 62nd Cannes Film Festival）で上映された（受賞は逃した）。日本では、第 19 回日本映画プロフェッショナル大賞の主演女優賞（ペ・ドゥナ）、第 19 回東京スポーツ映画大賞の主演女優賞、第 24 回高崎映画祭の最優秀作品賞、第 33 回日本アカデミー賞の優秀主演女優賞など、数多くの受賞を受けている。2009 年キネマ旬報作品ランキングでは日本映画の 6 位だった。

(3) 人工知能と「意識」「心」

人工知能が「意識」や「心」を持つようになるのかについては、人工知能の研究者は様々な考え方をしている。以下の表は、我が国の人工知能学会の研究者による一般向けの人工知能の解説書（13 人の代表的な研究者が 1 章ずつ担当）における「意識」「心」の捉え方をまとめたものである。人工知能がすぐにも「意識」「心」を持つようになる研究状況にはないが、原理的には実現可能とみなされている。少なくとも不可能とはみなされていない。

ただし、どのような状態を「意識」「心」を持つとみなすかについての考え方は様々である。「心」「意識」は、ある特定の単独の機能として存在しているのではなく、個々の要素の関係性の中に存在している。あるいは個々の要素を全体として総合化する仕組みとして存在し、

無意識から顕在意識化する、などの考え方がある。「意識」「心」を持たせるための研究アプローチも多様である。

人間や生物の脳がどのように意識や感覚、あるいは感覚意識体験（クオリア）を持つのかは科学的にかなりの難問であり、現段階では解明にまだかなりの距離があるとみられる。これが現在の研究者の考え方やアプローチの多様さの背景にはある。

表 6-12 人工知能は「心」「意識」を持てるか：人工知能研究者の考え方一覧

「心」「意識」についての考え方	出典
・意識とは「計算のプロセス」のこと	中島秀之（7 頁）
・（「知能」との対比で「心」について）「自己を認識する能力」が基本であり、自分という存在に対して意味づけを行い、他者の心を認識し、尊重する能力が生まれる	西田豊明（22 頁）
・（自己意識とは）自分の脳が働いているということを認識する部分が脳内にあって、それが自分の脳の働きを知っている（自覚している）という状態（いわば、リカーシブな状態） ・脳幹によって制御された脳の活動の結果は行動に表れ、その時間的経過を自分あるいは他人が包括的に眺めたときの状況を一言で指す言葉が心である	長尾真（82・85 頁）
・心も意識も「もの」として独立に存在しているのではない。心や意識という上位の概念は、下位の要素の関係の総体として出現し、かつ下位の要素の間の関係を支配する	堀浩一（104 頁）
・（「コンピュータが心を持っている」ことと「コンピュータがあたかも心を持っているように見える」ことは）決して区別できないという立場をとる ・ロボットに心がある、あるいは意識があるとみなしてやり取りすることが人間にとって便利ということになれば、それはそのロボットが心あるいは意識を持っている	松原仁（148 頁）
・意識とは、脳内の情報表現がコピーできないことに由来する、ほぼシーケンス的な注意のようなもの	山川宏（222 頁）
・意識とは、生命活動を細胞レベルの時間粒度を超えたスケールにて時間軸方向で安定化させるために神経細胞ネットワークにて創発される現象	栗原聡（231 頁）

出典：人工知能学会監修『人工知能とは』近代科学社、2016 年。

脳の客観（神経回路網の振る舞い（電氣的活動））と脳の主観（我々の意識、感覚意識体験）の間の隔たりは、「ハードプロブレム」と豪州の哲学者デービッド・チャーマーズ（David Chalmers）に呼ばれている。我々は現在、客観と主観とを結びつける科学的原理を一切持っていないとのことである（渡辺（2017）180 頁）。しかし、チャーマーズは人工物も意識が持てると考えている。フェーディング・クオリア（fading qualia）とは、チャーマーズが考えた思考実験であり、一つずつ脳のニューロンを人工のニューロン代替物に置き換えていく。一つのニューロンだけ交換する（そのニューロンの有する全ての神経配線を再現）ことは技術的に可能とみなすことはおかしくないと考える。順次一つずつ交換していき、最終的に全てのニューロンを置き換えることに成功すれば、その時に感覚意識体験は残るだろうと推定した（渡辺 198 頁）。

また、米国の神経科学者クリックとコッホは、視覚感覚を生み出すための最小の神経細胞集団、すなわち NCC（neural correlates of consciousness、意識の神経相関）を同定しようとしたが、NCC についてまだ科学的に解明されていない¹²³。

¹²³ NCC の定義は” the minimal set of neuronal events and mechanisms jointly sufficient for a specific

最近、Dehaene ら（2017）は *Science* 誌に、意識の現在の研究動向についてのレビュー論文を発表した。これまでに行われた多くの意識についての脳科学や認知科学の研究論文の内容を分析し、人間の脳内の意識（consciousness）は 2 つの要素から構成されると整理した。第 1 に、情報が脳内の局所的な処理に留まらずに、脳内部位で広く共有され、柔軟に脳内の計算と報告のために利用されるような場合に意識が発生している。これは、「C1」と呼ばれ、第 1 の意味での意識である。第 2 に、C1 の脳内の計算過程を自己モニタリングすることであり、その計算過程が正しいか間違っているかについての主観も生み出すような意識の発生である。これは「C2」と呼ばれ、第 2 の意味での意識である。C1 と C2 に対し、現在のコンピュータが行っている計算は、人間の脳内の局所的で無意識的な情報処理過程に相当する。

西垣（2018）は、「強い AI」（strong AI）とは、「人間のような意識をもって主体的に作動し、言語的相互行為をおこなえる機械、より広義には生命的な機械のこと」（22 頁）を言う位置づける。AI の作動はますます複雑化していくため、「その出力を予測することは、原理的には可能でも事実上困難に」なっていき、「疑似的自立性がいっそう強調されていくことは間違いない」（39 頁）と説明する。人間の心は、「閉鎖系であり、互いの意思にもとづく行動は予測不能」であり、そのような不可知性が、いわゆる「自由意志」の根拠ともなっている（37 頁）。他方、AI における推論のアルゴリズムのトレースは煩雑で困難ではあるが、それは生命体の「心」の不可知性とは根源的に異なる（170～172 頁）。「疑似的自立性」とは、人間の心のように閉鎖系システムであるための不可知性に起因するのではなく、あたかも心や自由意志をもっているように感じるということである。

自律性（autonomy）を持つためには、「自ら行動のルールを定めることができる」という点が前提である。あらゆる機械は、作動のルールを人間の設計者によって決められているので、他律系（heteronomous system）である。それに対して生命系は自律性を持つ存在である。内部でルールが創られる（遺伝的にルールが引き継がれることもある）ので、その行動を完全に予測することができず、「不可知性」がある。この「不可知性」は、人間の「自律性」や「自由意志」、それに伴う「社会的責任」と関連する、とのことである（西垣）。

繰り返しになるが、上記のように多様な考え方が存在しており、科学的な解明にはまだ時間を要するとみられるが、疑似的にでもあっても心や意識を持つような人工知能が実現することについて多くの研究者は否定している訳ではない。少なくとも原理的には可能とみなしていると考えられよう。

6.3.2 映画『空気人形』のプロットと主なシーンの分析等

(1) 映画のあらすじ

映画『空気人形』の全般的なテーマ、狙い等については既に説明したが、映画のあらすじは以下のとおりである（括弧内は映画出演者名）。

conscious percept”（固有の感覚意識体験を生じさせるのに最小限の神経活動と神経メカニズム）（渡辺（2017）、117 頁）のことである（元々の論文は、Crick, F. and C. Koch (1995). “Are we aware of neural activity in primary visual cortex?” *Nature* 375 (6527): 121-3.）

中年のファミレス店員である秀雄（板尾創路）は空気人形（ペ・ドゥナ）に「のぞみ」と名づけて夕食時に会話したり、夜に抱いたりして暮している。秀雄は、東京の開発から取り残された地域の古いアパートに一人で住んでいる。

空気人形はある日の朝、瞬きをしてゆっくりと立ち上がり、軒先の雫に触れて「キレイ」と呟く。いろいろ服を試着し、結局、メイド服で外出した。たまたま入ったビデオ屋の店員の純一（ARATA）に惹かれ、そこでアルバイトを始めた。「私は心を持ってしまいました」とつぶやく。純一に自分と似た空虚を感じ取る。

ある日、のぞみはビデオ店で勤務中に体に釘を引っかけてしまい穴が開き、体から空気が抜けていく。純一は驚きながらも、必死に息をのぞみのおヘソから吹き込んで救う。誰もいない店内で2人は抱き合う。のぞみは幸福を感じる。

秀雄との生活は続いているが、秀雄はのぞみに代わる新しい空気人形を既に購入している。のぞみが心を持ったことをようやく知った秀雄は、心を持った人形は面倒くさいとのぞみに言う。のぞみは家出し、空気人形の製造者である人形師の園田（オダギリジョー）の家に行く。同じ顔の人形でも廃棄のために引き取った時にはそれぞれ顔が違っていると、園田から聞く。心を持ったことの意味を理解し、園田に「生んでくれてありがとう」と告げる¹²⁴。

のぞみは純一に対して、何でもすると言うと純一はのぞみの空気をもう一度抜いてみたいと言ひ、抜く。今度はのぞみが純一のヘソに穴を開け、空気を吹き込もうとする。純一は死ぬことになり、のぞみも空気が抜けていくまま、ごみ置き場に横たわり、青空を眺める。

空気人形「のぞみ」を演じた女優ペ・ドゥナは、毎日特殊メイクに3時間はかけ、無機質な人形の皮膚の色を出したとのことである。ペ・ドゥナは韓国人であり日本語は母語ではないが、人形の役なら片言の日本語でも大丈夫とのこと監督から役のオファーを受けたとのことである¹²⁵。

(2) 主なシーンの観察・分析

本節では映画の中で、いくつかの映画のシーン（場面）を取り上げ、観察・分析を行いたい。それらのシーンの選択基準は以下のとおりである。

1. 空気人形が人間と接触する場面

- 純一と人形のやり取り
- 心を持った人形が所有主に否定される場面、など

2. 空気人形の自律的な行動、決定の場面

- 人形が心を持つようになった場面（美・快などに対する感情、人に対する感情（好・愛）、自分に対する感情など）
- 人形が自らごみになることを選択する場面、など

映画のシーンの観察・分析（表中の「このシーンと人工知能・心」に記載）は、本セクショ

¹²⁴ Wikipedia からの引用に加筆した。（<https://ja.wikipedia.org/wiki/空気人形>、https://en.wikipedia.org/wiki/Air_Doll、https://fr.wikipedia.org/wiki/Air_Doll）

¹²⁵ 産経新聞、2009年10月3日大阪夕刊、総合1面。「プレミアムシートー女優ペ・ドゥナ 日本映画『空気人形』の主人公」

ンの冒頭に設定した狙いを念頭に置いて行うが私の主観を反映するものである。映画評論家の町山智浩によれば、映画評論とは、第1に「作者が意図したことは何かを読み解くこと」、次に「作者の意図を超えたところまでを指摘すること」である¹²⁶。本節は、映画の製作者が意図していることを読み解くことが主たる目的ではないが、結果としてそれを含むことになる。映画の場面での空気人形の独白・行動や、空気人形と周囲の人間との会話などから、空気人形の心や人間の考え等を読み取ることが分析となる。空気人形の心の中を直接見ることはできないから、人形が自らの心の状態を語った内容から読み取る以外に、人形の様子や会話からどのような心の状態であるのかを観察者である私がある立場に置かれていた場合などを想像し、主体的に読み取ることとなる。

なお、この映画の「空気人形」は心を持つ人工知能・ロボットを代表する訳ではなく、その人間との交流等におけるやり取りも何かを代表する訳ではない。そこから何かの意味を読み取ったとしてもそれが全てに適用可能な訳では当然ない。サンプル数は1であるとも言えるが、それを材料として、ある条件の下でその存在の社会的インパクト等について示唆を考えるという位置づけである。

①冒頭のシーン（2:25-4:20（映画開始からの時間、以下同様））：ファミレス店員秀雄と人形

<p>【舞台設定】</p> <ul style="list-style-type: none"> ・ファミレス店員の秀雄（既に中高年である）は帰宅し人形に語りかける。人形には自分が店長であるように話す。後の場面で描かれるように秀雄はファミレス店員であり、店長から働きぶり等について常に嫌味を言われている（代わりはいくらでもいるんだ等）。
<p>【映画シーン】</p> <ul style="list-style-type: none"> ・秀雄が仕事を終え、スーパーで買い物を済ませ、激しい雨の中傘をさすことなくアパートに一人で帰宅する。 <p>秀雄：「ただいまー。言うこと聞いて傘持ってたなら良かった。うわー、最悪や」</p> <ul style="list-style-type: none"> ・テーブルで向いに女性（メイド服を着ている、背中側しか見えない）を座らせ、秀雄はコーヒーを飲みながら一人で話し続ける。女性は動かず黙って聞いている。 <p>秀雄：「せやからな。（自分が）昼休みに店の裏に（店員を）呼び出して言うたったんや。『お前の代わりなんかな、なんぼでもおんねんぞ』言うて。ほんまに昼の忙しい時に、厨房の中でおーと立ちやがって。ほんま役立たずや。誰のおかげで店回つとる思とんねえなあ」</p> <ul style="list-style-type: none"> ・カメラ視点が秀雄側からになり、女性の顔がアップになる。人間ではなくビニール製の人形である。 <p>秀雄「3つ年下やったかな。そうや。一人っ子やねん。甘やかされて育ったんや。まあ、どうでもええねんけどな。あんな奴は。ええと、どないしょ。もう遅いし。今日は風呂はやめとくか？ うん？」</p> <ul style="list-style-type: none"> ・場面が変わり、秀雄はベッドで寝ている。人形を抱きしめる秀雄。 <p>秀雄：「のぞみ。キレイや。。。のぞみ」</p>
<p>【このシーンと人工知能・心】</p> <p>人形といってもただのゴム空気人形であり、それに洋服を着せているだけである。安物のビニール製の人形であり、リアリティがある訳ではない。また、人工知能を持つ訳でも、会話に何等かの反応を少しでも示す訳でもない。それでも秀雄にとっては日々の空虚さをやり過ごしていくために役立っているようである。逆に言えば、何も反応も示さないことで秀雄の言っていることを全て受け入れることができているとも言える。何等かの反応を示すことができれば、それは秀雄を満足させる反応であり得ると同時に、望まない反応であることもあり得、また、知性があ</p>

¹²⁶ 『映画秘宝 期間限定版 ベストテンなんかぶっとばせ！』 洋泉社 MOOK、1998 年（モルモット吉田『映画評論・入門！』（2017 年、洋泉社）に引用、39 頁。

れば、秀雄が実はファミレス店長ではなく、しがない店員に過ぎないことも分かってしまうかも知れない。それでは秀雄は安心して会話を続けることができず、日々の現実からのつかの間の解放を帰宅後の時間に味わうことも不可能となるだろう。

つまり、映画冒頭場面での空気人形は心も意識も持たず、動かず座っているだけであるが、所有者である秀雄から期待されている機能は十分に果たしているのである。

②人形がこころを持つようになる場面（5:20～8:30）

【舞台設定】

・朝の場面。人形の所有者である秀雄は出勤前の準備で着替えなど忙しくしている。人形はベッドで横たわっている。

【映画シーン】

人形：「ふうー」

・人形の首がかすかに動く。秀雄は朝の出勤前の準備で、「いらっしゃいませ」と何回もつぶやきながら、急いで着替えている。ベッドに裸で横たわる人形に「寒い。寒い」と言ってシーツをかぶせる秀雄。

秀雄：「今日もキレイや。行ってくるわ」

・人形の手と足がゆっくりと動き始め、やがてゆっくり立ち上がる。窓まで歩き、外の景色を見て、昨夜の雨が物干しざおから滴り落ちてくるところに手を差し出す。手のアップの映像から画面が人形の体に移動する（この時点で、人形がビニール製から俳優のペ・ドゥナに変わっている）。雨で濡れる手、朝もやの外の景色を見て人形がつぶやく。

人形「キ。。。レ。。。イ。。。。キ。。。レ。。。イ。。。」

・タイトルバック「空気人形」。

【このシーンと人工知能・心】

人形が初めて「こころ」、「意識」を持つようになった場面である。朝の光の中で、雨で掌を濡らし、それを見て、「キレイ」と感じている。外からの映像情報を受け取り、それを自らの内面と照らし合わせ解釈し、キレイとの感情を発している。クオリア（感覚意識体験）を初めて持ち外界に表現した場面とも言える。こころを持ち始めた最初のシーンが、自然の美しさに気づいている場面であり、クオリア体験を人形が持ったことを印象的に描いている。そこにこれまでの何らの記憶がある訳ではないと見られる人形がまず発した言葉、認識、感情は自然（水滴）や景色を美しいと感じるものだった。

「クオリア」（感覚質、感覚意識体験）とは、何かのものを見た時等に、クオリア（色、におい、味など）が発生すると説明される。視覚の機能や原理は、写真機と同じようなものであるが、人間の場合にはクオリアが発生し、そこが機械とは異なるとされる。映画の中では人形の感じるクオリアの質に映画を通じて変化が生じる訳ではないが、最初には世界の美しさのみについての感情が発生したのが、最後には悲しさ、空虚さにつながっていく。

③ビデオ店での人形と純一との出会い（15:30～17:30）

【舞台設定】

・人形は心を持った最初の日、メイド服を着て外出した。一日、街を彷徨い、様々な人（ゴミ回収車で働く人、学校に向かう子供と見送る親、幼稚園児と幼稚園の先生、交番で昔あった殺人事件について警察官に話をしている初老の女性、空き地で景色を見ている老人など）を観察した。

・夕方のチャイムの音を聞き、夕暮れの町を急いで家に向かう途中でビデオ店の煌々とした明かりに魅せられ、店の中に入る。ビデオパッケージが陳列される店内を物珍しそうに見て歩く人形。店を一回りし、最後にレジカウンターで何か仕事をしている店員と対面して向かい合う。

【映画シーン】

店員（純一）：「何かお探しですか」

・人形は何も言い返さずに店員の顔を見る。

人形（独白）：「私は心を持ってしまいました。持つてはいけない心を持ってしまいました。」

【このシーンと人工知能・心】

店員（純一）はこの時点で相手が人形であることを分かりそうなものなのだが、普通の人間に対するのと同じように人形に質問している（相手が誰であろうとマニュアル対応するビデオ店の店員を揶揄しているとも取れる）。人形は心を持った後、初めて人間から直接話かけられ、どのように対応すればいいのか分からず言葉を発することができないようである。また、自分を人形としてではなく、人間に対するのと同様の対応をしてきている店員に対して好意を持った表情をしている。

店員は「何かお探しですか」という言葉で、探している DVD やビデオがあるのかを質問しているが、人形の方は映画やその DVD について何も知らない様子であり、店員の質問の意図が読み取れないようである。人形はこの世の中でどこを持つようになり、自分は何かを探しているのかどうかを考え、果たして自分が探しているものがあるのかどうか、あるとすれば何を探しているのかを自問し答えが見つからずに言葉を発することができないとも解釈できる。

④人形がビデオ店で勤務を開始する（19:05～20:00）

【舞台設定】

・人形「のぞみ」は心を持つようになり、秀雄が不在の昼間に街をさまよい、たまたま入ったビデオ店で働くことになった（その経緯は映画では省略されている）。人形に対して、店員の純一が仕事の内容を説明している。

【映画シーン】

店員（純一）：「もともと映画はフィルムでできていて、光を当てて白いスクリーンに映すんだ。」

人形：「フィル・・ム」

店長：「見たことないの？映画？」

・人形が DVD を店長に掲げて見せる。

店長：「そんなのは代用品。映画はやっぱ、映画館で見なくっちゃ。」

人形：「映画・・・カン？」

店長：（ひとりごと）「これだからなあ。最近の若い子は・・・」

【このシーンと人工知能・心】

人形がこれらのやり取りをどこまで理解しているのかは不明である。特に、映画、映画館、フィルムという言葉や、これらが意味していることについては少なくともこの場面では理解はしていない。そもそも「映画はフィルムでできていて」だけの説明（映画はフィルムを使って記録されているだろうが、フィルムで映画ができていく訳はない）では、映画の知識が全くない人形（機械）は、「映画」とは何かを理解できないだろう。ちなみに、辞書では映画は「フィルムに連続的に写しとった映像を、映写機でスクリーンに映し出し、目の残像現象を利用して形や連続した動きを再現するもの」と説明されており（goo 辞書）、人工知能に正確に理解させるのは容易ではない。

それらの理解のプロセスにここはどのような影響を与えているか。まず、このコミュニケーションに人形が関わったのは店員に惹かれたとの心の働きのためであり、また、その店員が働いているビデオ店で仕事をするための知識の獲得へのモチベーションから発生しているとも考えられる。すなわち、心を持つことによって人形に自由意志が発生し、そこから会話が生まれるようになったと捉えることもできよう。

⑤ビデオ店でやりとり：「嘘」をつく（21:00-22:00）

【舞台設定】

・ビデオ店で店長と空気人形（のぞみ）の間の仕事上の雑談。

【映画シーン】

店長：「のぞみちゃん、クリスマスどうしてんの？彼氏と一緒に？」

人形：「うん？」

店長：「彼氏だよ、好きな男。うーん、チュッチュッ・・・（DVD パッケージで男女がキスをしている写真を示す）」

・横を純一が通り過ぎる。

<p>店長「いるんでしょ？」</p> <p>人形：「いいえ。」</p> <p>店長：「本当？」</p> <p>のぞみ：「はい」</p> <p>人形（独白）：「私は嘘をつきました。心を持ったので嘘をつきました」</p>
<p>【このシーンと人工知能・心】</p> <p>心があるから相手に良い印象を与えたい、あるいは、思っていることをそのまま話すことで嫌われるかも知れないなど感じて、自分の本心を隠したいとの気持ちが生まれる。真実をそのまま述べるのではなく、意図を持って隠し、あるいは真実と異なることを話す、すなわち嘘をつくことを選択することとなる。この場面では、好意を持つ純一が横にいたので本心を隠している。</p> <p>ただし、逆に心があることで嘘をつくことの倫理的なハードルが上がるとの側面もあるだろう。心がなければ自己利益を拡大するために嘘をつくことで良心が痛むこともないものだ。</p>

⑥ビデオ店でのやり取り（29:50-31:30）

<p>【舞台設定】</p> <ul style="list-style-type: none"> ・ビデオ店の入り口のガラスのドアをスプレーをかけて磨く、のぞみ（人形）と純一。ドアを隔てて親しく会話をしている。
<p>【映画シーン】</p> <p>人形：「空には何があるの？」</p> <p>純一：「空には空気があって雲が浮いていて・・・夜になると月や星が出て。透明で目には見えないけど、でも、あるんだ」</p> <p>人形：「見えないけど・・・あるんだ・・・ああ、難しい」</p> <ul style="list-style-type: none"> ・店内のディスプレイの画面を見つめる人形。海の映像である。 <p>人形：「あれが海？」</p> <p>純一：「うん。そうだよ」</p> <p>人形：「わあ。へえー」</p> <p>純一：「行ったことないの？海」</p> <p>人形：「うん」</p> <ul style="list-style-type: none"> ・お台場周辺の海辺を歩く二人に画面が切り替わる。 <p>人形：「年を取るってどういうこと？」</p> <p>純一：「老いてだんだん死に近づいていくこと」</p> <p>人形：「し？」</p> <p>純一：「うん。命を失うこと」</p> <p>人形：「いのち？」</p>
<p>【このシーンと人工知能・心】</p> <p>人形は自らの体の中に空気が入っていることを認識している。それが他の人間とは異なることも知っている。それが何もない、空虚であるということの意味するのか、空気が入っているということは何も入っていない訳ではなく、何かのある物質で満たされており、他の人間と何ら異なる訳ではないのかとも考えている。この場面のやり取りは、人形はこの世の中で心を持つようになり生まれ、人形であり老いることはないが、それでもやはり命がある以上、それが永遠に続くことなく終わりがあのではないか、あるいは純一が年を取り、いつかは命を失うとすれば人形である自分だけ生き続けることはできないだろうことを示唆している。映画の結末へ進んでいく展開の布石となっている。</p>

⑦バスを待つ純一と人形（36:00～37:30）

<p>【舞台設定】</p> <ul style="list-style-type: none"> ・純一と人形は海に行った帰りに夕食をレストランで食べ、純一は人形に「誕生日」や草花の名前（「シロツメクサ」「タンポポ」）、草花がいつかは枯れる理由について教えている。
<p>【映画シーン】</p> <p>純一：「他にはどういうことを知りたい？」</p>

<ul style="list-style-type: none"> ・外を歩きながら人形は自分の影が白っぽいことに気づき、そのことを純一には知られないように道の端を歩く。 ・人形はノートを取り出して純一の顔のスケッチを始める。 ・家に帰るバスに画面が切り替わる。バスでもスケッチを続けている。 <p>人形（独白）：「あなたのことをもっと・・・」</p>
<p>【このシーンと人工知能・心】</p> <p>人形は、世の中のことについて殆ど何も知識がないことについては純一に対して恥を感じることはなく、純一から教えられることを熱心に聞いている。しかし、自分の体が空気からできていること、空気人形であること、純一など人間とは異なることについては知られたくないと感じているようである（既に純一はそれを知っているかも知れない）。</p> <p>自分について相手がどのように思うかについて考えることは心がなくてもできるが、心があることによって、自分がどのように思われて欲しいかについての感情が生まれ、そこから自分の姿や能力を変えようという動機が生まれる。また、心があることによって、数多くある知りたい対象の中から自分が何を最も知りたいのかについての優先順位、好き嫌いが生まれる。</p>

⑧帰宅する秀雄（39:00-39:40）

<p>【舞台設定】</p> <ul style="list-style-type: none"> ・夜になり、秀雄が帰ってくるまで人形は部屋でガラスを眺めたりしている。そこに秀雄が帰ってくる。
<p>【映画シーン】</p> <p>秀雄：ただいま（関西弁）。</p> <ul style="list-style-type: none"> ・人形は急いで洋服を着替え、ベッドに横たわり、布団をかぶる。 <p>秀雄：「一人で寂しかったかー。岡部のやつがさ、急に休みやがってさ。ほんま何考えとんや。あいつ。てんてこまいや」</p> <p>人形（独白）：「私は空気人形。型遅れの安物です・・・」</p> <p>※秀雄がいない間に、新しい型の人形を秀雄が購入したことを、その人形が入っている箱をベッドの脇で見つけて知っていた。</p> <p>秀雄：「おーい」</p>
<p>【このシーンと人工知能・心】</p> <p>ある特定の機能を果たすことを主たる目的として設計された機械（人形）は、心を持つことによって、その機能を果たすことについて、当初の設計時の意図とは異なる感情を持つようになることがあり得る（自由意志の存在、心の不可知性・不確実性）。逆にそのようなことが決して起こらず、心を持たせたとしても決定論的な意思しか持たないのであればそれは疑似的に心を持ったことにしかならない。</p> <p>ただし、意識・心を持たせたからと言って、それが人間と同じような心であるとは必ずしも限らない。喜怒哀楽の全ての感情の濃淡を人間と同様に持つかは設計上の問題である。</p>

⑨ビデオ店での人形と客とのやり取り（40:00-42:30）

<p>【舞台設定】</p> <ul style="list-style-type: none"> ・ビデオ店で、人形は客から探している DVD について質問されるが、的確に答えることができず、純一や店長から助けを受ける。
<p>【映画シーン】</p> <p>客1：「アリだったかマリだったかって名前の女の子が、こうフォークで髪を梳かすやつなんだけど・・・」</p> <p>人形：「はい。えーと、アリ・・・」</p> <ul style="list-style-type: none"> ・人形は DVD のカタログのページをめくって探しているが見つからない。 <p>客1：「分かんねえか・・・。弱ったなあ」</p> <p>純一：「きっとそれ『リトルマーメイド』ですね。こちらです」</p> <p>客1：「あっ、ありますか。助かります」</p> <p>人形：「（客に頭を下げながら） すいません」</p>

<p>客 2（交番勤務の警察官）：「あのさあ。警察官がすごーい悪いやつない？あの、汚職とか麻薬の裏取引とかで。もう腐りまくってるようなやつ」</p> <p>人形：「腐ってる？」</p> <p>客 2：「そう。何かない？」</p> <p>店長：「じゃあ、あれなんかどうです？『バッド・ルーテナント』アベル・フェラーラ」</p> <p>客 2：「それ。ほんとに悪いの？」</p> <p>店長：「もう悪いなんてもんじゃない。ハーヴェイ・カイテル。最低。一緒に「フェイク」とかどうですか？アル・パチーノがマフィアで」</p> <p>客 2：「ジョニー・デップが警察官だろ？」</p> <p>店長：「そうそうそうそう」</p> <p>客 2：「観たもん」</p> <p>店長：「ああ。観たんだ」</p> <p>人形：「マーラ？」</p> <p>客 3：「マル」</p> <p>人形：「マル？」</p> <p>客 3：「じゃあ、テオ・アンゲロプロスの。ええと。『蜂の旅人』」</p> <p>人形：「ゲロ？」</p> <p>客 3：「アンゲロ」</p> <p>店長：「あのね。橋の向こう側行くとツタヤがあるから」</p> <p>純一：「『マルメロ』はセル専用なんですよ。『蜂の旅人』は DVD ボックスなんでレンタルしてないと思いますよ。多分ツタヤでも」</p> <p>客 3：「すいません」</p> <p>純一：「ありがとうございます」</p> <p>※人形は頭を客に対して何度も下げる。純一、店長に対しても何度も頭を下げる。</p> <p>人形：「ごめんなさい。役に立てなくて」</p>
<p>【このシーンと人工知能・心】</p> <p>本来は人工知能であればこのようなやり取りは人間以上に得意なはずである。ただ、「腐っている」など本来の用途とは異なる使い方をされると、意図されている言葉の意味を読み取ることが困難になる可能性はあるし、また、言語化されないような曖昧な映画の画面情報の描写のみの情報を知らされても、その情報から映画のタイトルを探すのは難しいだろう。そのような人間と人工知能との会話における意思疎通や意図理解の困難さについては、川添（2017 年）が小説形式で描写し、解説している。</p> <p>また、この場面で役に立てなかったことについて純一や店長に頭を下げ済まなく思っているが、そもそも職場や家庭などで他の人や集団に対して「役に立ちたい」と思うのは人間らしい反応であり、望まれる行動を増やし、行動の質を高めることに対しての動機を提供する。マズローの欲求段階では 4 段階目の承認欲求に相当するが、それは心が生み出すものである。単なる機械であれば自分の持つ機能以上の働きを求められればそれはできないだけのことであり、状況に柔軟に対応しようと努めることなく、平然としているだけである。</p>

⑩公園での人形と老人との会話のシーン（43:20-45:00）

（⑩-1）

<p>【舞台設定】</p> <ul style="list-style-type: none"> ・空き地に置かれたベンチに座ってたばこを吸っている老人。人形が側を通り過ぎると声をかけた。
<p>【映画シーン】</p> <p>老人：「君、カゲロウって虫、知ってるかな？」</p> <ul style="list-style-type: none"> ・老人の方を人形が見るが答えられない。 <p>老人：「カゲロウはね、親になると一日か二日で死んでしまう。だから体の中は空っぽで、胃袋</p>

<p>も腸もないそうだ。そこには卵だけが詰まってる。ただ産んで死ぬだけの生きもんだ。人間も同じようなもんだ。下らんよ」</p> <p>人形：「あたしも空っぽなの」</p> <p>老人：「それは奇遇だね。私も同じさ。空っぽだ」</p> <p>・人形は老人から意外な答えを聞いて、その意味がどうしたことなのかしばらく考える。</p> <p>人形：「他にもいるのかしら？」</p> <p>老人：「今時は、みんなそうだろう」</p> <p>人形：「みんな？」</p> <p>老人：「ああ。特に。。。こんな町に住んでる人たちは。。。君だけじゃないよ」</p> <p>・人形はそれを聞いて微かに笑顔になる。空き地から川を隔てて、月島の高層マンション群の遠景に切り替わる。</p>
<p>【このシーンと人工知能・心】</p> <p>この会話では「空っぽ」の意味するところが、老人と人形の間では異なっている。人形の場合は、文字通り、体の中は空気だけしかなく何もないということであるのに対して、老人の場合は、カゲロウのように産んで死ぬだけの生き物であること以上に、人間であることに何かの上位の意味を持たせるような内実がないことを示唆している。言葉の意味理解が失敗する場合には、会話や意思疎通が失敗するものであるが、この会話の場合には、意味は異なるがメタファーであるため、会話が成り立っている。</p> <p>ある意味、現代の都会人の抱える空虚さーカゲロウのように与えられた短い生命の時間を生き急ぐーを空気人形のメタファーを通じて描くことがこの映画の一つのテーマであるが、この場面では、人間の空虚さと、空気人形の実質的な空虚さが直接対比され、どちらも同じようなものじゃないかと老人に言わせている。ただし、カゲロウのように生まれてすぐに死ぬだけで、中身が空虚であったとしても、それが老人の言うように「下らん」かどうかは個人の価値観の問題である。映画では、人形とその所有者、ビデオ店員以外にも、過食症の女性、付近を徘徊する老女、大企業の受付嬢の女性など話のプロットには直接関係のない、心に空虚さを抱える人達が散発的に登場するが、彼女らの人生も映画の後半にはポジティブに描かれ、空虚さが必ずしも「下らん」と切り捨てられるものではないことが暗示されている。</p> <p>ちなみに、このシーンのカゲロウの話は吉野弘の詩「I was born」からの引用である。詩では、「生まれる」という意味の英語の文が受動態であることに気づいた少年が、この世に生まれるのは自らの意思で選択したのではない、自分の意思で生まれてきた人は誰もいないと指摘する。人工知能を持つロボットは人間が作り出すものであるが、この点では人間と変わるものではないこともこのシーンは示唆している。</p>

(⑩-2)

<p>【舞台設定】 ⑩-1と同じ。</p>
<p>【映画シーン】</p> <p>老人：「君、こんな詩を知ってるかな？」</p> <p>人形：「し？」</p> <p>老人：「詩を知らないか？。。。まあいい。。。『生命は・・・』」</p> <p>人形：「いのちは・・・」</p> <p>老人：「えーと、『生命は・・・』」</p> <p>人形：「えーと、いのちは・・・？」</p>
<p>【このシーンと人工知能・心】</p> <p>詩人の吉野弘の作品『生命は』では、「生命は その中に欠如を抱き それを他者から満たしてもらおうのだ」「互いに 欠如を満たすなどとは 知りもせず 知らされもせず ばらまかれて いる者同志」と書かれている。この世界観によれば、空気人形や人間の空虚さは、人間社会を構成していくために不可欠のものであり、個々の人間が完全に自立し他者を必要としないのであれば人間社会も成立しない。そのような意味で空虚さにも意味があることが暗示される。ただし、この場面だけでは吉野弘のその詩の冒頭だけしか示されないの、観客はこのシーンだけからはそのようなことを読み取ることはできないが、老人は少なくともそのようなことを考えて</p>

いるようである。後の場面で、この老人は国語の代用教員であったことが示される。空気人形と同様に代用品であったという訳だ。

このシーンの後に、人形の独白で、この詩の冒頭部分の朗読がある。そのシーンでは、前にビデオ店の客だった人（先のシーンでDVDについて質問をしていた3人）がそれぞれ、交番勤務の警察官、スーパーの店員、メイド喫茶を訪れる浪人生として普段の生活を送っている場面、秀雄が混雑するファミレスで客から注文を聞いている場面、純一がビデオ店で一人で働いている場面などがそれぞれ短く流れる。

⑪秀雄が初めて人形が心を持っていることを知る場面（1:17:30-1:21:10）

【舞台設定】

・秀雄は新しい人形を使い出し、その人形をベッドに寝かせ、天井に星を映し星座の説明をしている。それはこれまでは「のぞみ」に対してしていたことだった（秀雄は新しい人形も「のぞみ」と呼んでいる）。秀雄はトイレに行くために2階のベッドを置く部屋から1階に降りた。トイレから戻ると部屋には前の人形「のぞみ」が立っている。

【映画シーン】

秀雄：「はっ」

・秀雄は驚き、部屋の隅に逃げる。

秀雄：「ええっ……。何？あの……。のぞみ？」

人形：「ケーキなんか買って誕生日？」

秀雄：「いや……。お前の時もあったってお迎えした日に」

人形：「知らない。覚えてない」

秀雄：「したって。写真も撮ったって」

・秀雄はあせって写真を探し出す。

人形：「私……。心を持ってしまったの」

秀雄：「こころ」

人形：「そう。こころ」

秀雄：「なんで？」

人形：「分からない。ねえ、私のどこが好きなの？」

秀雄：「どこって……」

人形：「言えないわよね。「のぞみ」って名前だって昔の彼女の名前じゃない」

秀雄：「読んだんか？俺のブログ」

人形：「その人の代用品なのよね、私」

秀雄：「そんな。そやかて。俺も最初は一人人なって寂しかったし。けど今は違うんや。何言うとんのや俺」

人形：「なんで、私なの？」

秀雄：「なんで？」

人形：「私じゃなくてもいいでしょ？そうなんでしょ」

・秀雄はテーブルの上に以前撮ったという写真を置く。言うのをしばらく躊躇した素振りの後で秀雄が人形に話す。

秀雄：「あのな。頼みがあんねんけど。もとの人形にもどってくれへんか」

人形：「もとの？」

秀雄：「普通の……。ただの人形に。無理なん？」

人形：「心なんか持たない方が良かった？」

秀雄：「うん。面倒くさいんや。俺。こういうの面倒くさいからお前にしたのに」

人形：「ひどい。そんな言い方」

・人形は部屋を走って出ていく。秀雄は引き留めようとする。

秀雄：「そうやない。そうやないんや。面倒くさいのはお前やのうて、人の方や」

【このシーンと人工知能・心】

このシーンはかなりホラー映画っぽい、他のシーンとは異なる演出が意図的にされている。1階の部屋はとても暗く、そこにペ・ドゥナは素顔に近い白い化粧をして立っており、2階の部屋

で新しい空気人形と過ごしていた秀雄が出くわす。ちなみに、ペ・ドゥナはホラー映画『リング』の韓国版『リング・ウイルス』（1999年）で貞子役を演じていた。

人間のためにサービスを提供する人形が、人間から必要とされなくなった時、特にその人形が心を持った場合、どうなるか。それは業田良家の他のロボットを扱った漫画には頻繁に現れるテーマである（『機械仕掛けの愛1』の『ペットロボ』『子育てマーシー』『リックの思い出』、『機械仕掛けの愛2』の『ハッピー先生最後の授業』『緑の枯れ草』など）。普通は人形が急に話を始めたら、こんな普通の会話をすることはできないだろう。

この映画では空気人形と人間との会話、人間の間の会話も淡泊なやり取りで終始しているが、この場面での空気人形と秀雄との会話は唯一感情的であり、かなり長めの会話となっている。やや長すぎるやり取りになっている。秀雄は関西弁であるが、人形はなぜか標準語で話している。現実社会での人間とのやり取りに疲れている秀雄にとっては、人形とのやり取りも不可知なものになったら疲れるだけなのだろう。「面倒くさいのはお前やのうて、人の方や」と言っているが、ここで「人」を指している中には心を持つようになった人形も含まれることになっている。

「空気人形」は、所有者に対しては一貫して機能を提供しつづけており、内面的には心があったにせよ、それを所有者に対しては示していなかった。所有者は働きかけに対してどのような反応をするかどうかについて不可知であることはなく、心を持っているとの認識を持つことはなかった。

⑫「のぞみ」が人形製作所を訪問するシーン（1:27:00～1:30:30）

【舞台設定】

・空気人形「のぞみ」は自分を制作した人形工場を探し訪れる。建物の中に入り、工場に多数置かれている制作途上の人形を眺める。そこへ人形の製作者（オダギリジョー）が現れる。

【映画シーン】

人形師：「おかえり」

人形：「ただいま」

・場面が変わり、人形師の机に移動し、対面して話合っている。

人形師：「それは作った僕にも分かんないな。でも、なぜ心を持ったかなんて、人間を作った神様にもきっと分かんないと思うよ。心なんて持たなければ良かった？」

人形：「分からない。でも……。苦しい」

・人形師は製作所の上の階の、既に返品され廃棄を待つ人形が積み上げられているところののぞみを連れていく。

人形師：「もともとは同じだったはずなのに、こうして戻ってきた時にはみんな違う顔をしてる。ちゃんと愛されたかどうか表情に出るんだ。それはこの子たちにもあるってことじゃないかな。心が」

人形：「この子たちはこの後……」

人形師：「年に一度春先にまとめて捨てるんだ。残念だけど燃えないゴミだ」

人形：「燃えないゴミ……」

人形師：「まあ、僕たちだって死んだら燃えるゴミだから。大した違いはないけどね。」

【このシーンと人工知能・心】

この場面で、人形は、人形としての死を迎える場合の運命について知ることになる。心を持たない人形であっても、いつかは古くなり用途がなくなり廃棄される。心を持っていれば「死」ということを、単なる機械やモノの寿命ということ以上の意味を持つ言葉として理解することができるだろう。

なお、空気人形は、前のシーンで、純一から息を吹き込まれた次の日、空気入れをゴミ捨て場に捨てに行っていた。これは人形だから永遠に存在するのではなく、（人形が好意を寄せている）純一の息が体内からなくなった時点でさらに空気を空気入れで吹き込むのではなく自ら「死」を選択することに決めたということだろう。それから、外出し、一人で、街を歩き、墨田川の遊覧船に乗り、景色の移り変わり、太陽の光の中で、風を受け、初めて、限られた生命となることを選んだことで、逆にその限定された、そして自ら選択した結果としての生命の喜びを感じている様子が描かれている。

空気人形は、映画の結末では空気が抜けてしまった自分を「燃えないゴミ」として、殺された純一を「燃えるゴミ」としてゴミ捨て場に捨てることになる。

⑬人形が自ら死を選択する場面（1:45:10~1:50:00 くらい）

【舞台設定】 ・映画のラストシーンである。空気人形は自ら「死」を選択し、ゴミ捨て場に横たわっている。
【映画シーン】 ・ごみ捨て場で横たわる人形。人間は人形が捨てられていると思っているので、人形の横に次々に他のゴミを投げていく。以前レストランで隣の席で食べていた少女が、人形の指輪を取り、自分の人形と人形の体の上に「交換ね」と言って置いた。人形から空気が抜けていく音が続く。 ・人形の想像の中でクラッカーの音が始まる。レストランで純一と一緒に食事をしている。食べ物を口に入れて食べる人形。ハッピーバースデーの音楽が鳴りだし、映画の全ての登場人物が拍手をしながら人形を祝福して集まってくる。秀雄は誕生日ケーキを持ってくる。人形は、泣きながらケーキのろうそくを息で消そうとする。
【このシーンと人工知能・心】 ピノキオや映画「AI」など、人形や機械が最後は人間になりたいという強い願望を持ち、そのために様々な辛い経験をして成長していくというのは昔からある人形物語の一つの原型である。映画『空気人形』は、空気人形は心は持ったが、空気人形として自ら「燃えないゴミ」となることを選択し、人間への変身願望はみられない。しかし、人形から空気が抜けていき意識がなくなる中で人形の想像の中では、人間のように料理を口の中に入れ、人間となることができ、涙を流しており、これまでの知り合いから誕生日の祝福を受けている。 心を持つような人工知能を備えた人形やロボットは、周りの人間のようになりたいたいと思うものなのだろうか。心の働き、自由意志は何かを達成したい、何かになりたいたいという願望を伴うものであるが、その対象に人間が含まれていてもおかしくはない。

(3) 観客の感想、映画評論家の批評

観客の感想

観客の映画についての感想は以下のウェブサイトの掲示板から集めた。

「みんなのシネマレビュー」75人が映画の感想を掲載

https://www.jtnews.jp/cgi-bin/review.cgi?TITLE_NO=16917

「みんなのシネマレビュー」では、ペ・ドゥナの配役が良かった、演技が良かったというペ・ドゥナに対しての賛辞が圧倒的に多かった。人形が心、意識を持ったということについて、人工知能への言及も含め、直接触れた感想は見られなかった。映画のテーマの一つである「空気人形との関わりを通じて、都会の人間の孤独感や虚無感が描かれている」ことについて書いた感想が約10件で最も多かった。

観客の感想では、社会には満たすべき空洞があり、映画の中では、その空洞を人形が埋めていき、人形もまた埋められていたという意味で、このような人形の役割については一定の共感がみられた。他方、純一と人形との映画の終盤での場面（純一が人形の空気をわざと抜く、人形が純一の中に空気を入れようとし結果的に死なせる）については受け入れられないなどとの否定的な見方が多かった。

その他の主な感想は以下の通りである。

- 人形は持ち主を毛嫌いしていき、他の人間との交流を広げていくのが不思議

- 空気人形は心を持つべき存在ではなかった
- 心を持った空気人形は、その空洞を満たしたくなった
- 死んだらゴミという考えは共感とは程遠い
- この世界の人々にとって空気人形とはどんな存在なのかが曖昧
- 純一が空気人形の空気を抜いた理由は何か
- 最後に見た夢が彼女の望んだ幸福かと思うと悲しくなる
- 「死」の概念を学ぶことの難しさが描かれている
- 空気人形が心を持って動き出し、ビデオ店で働き、そこで恋をすることが非現実的

映画批評家等の評論、映画紹介

映画評論家（新聞記者を含む）の評論、映画紹介等は、新聞、雑誌、インターネット等から集めた¹²⁷。大部分は映画公開の前後の2009年9～11月に書かれたものである。

まず、映画担当の新聞記者による映画紹介が多いため、監督のインタビュー等から映画の意図や映画の内容を簡単に紹介したものが多い。そのため、監督の説明するような映画の意図の内容に沿ったものとなっている。現在の社会の人間の抱える空洞や、人間間の関係について、空気人形の存在やその視点を通じて示唆を与えるもの、あるいは主演のペ・ドゥナにスポットライトを当てたものが殆どであった。

- ・「体が空っぽの人形は、うつろな魂を抱えた都会の人間そのものだ。人形が出くわすのは、鬱屈した思いを吐き出す相手のいない人たちばかり。…そんな中で人形は純一に恋をして傷つく。人のぬくもりに触れ、心を持つことの切なさを知る。うつろな人形が人間らしい存在に変貌していく様子を、物悲しさとともに描いていく」（勝田友巳、朝日新聞2009年9月25日東京夕刊7頁総合面、「シネマの週末・この一本：空気人形 都会の片隅の人々を描く」）
- ・「本来持つはずではなかった魂を持ち、知るはずではなかった哀しみを知ってしまった人形は悲劇的結末を迎えずにはいられない。自分勝手な男たちにモノとして玩弄される人形の宿命を粛然と引き受けるペ・ドゥナこそ荘重な悲劇の中で輝く永遠のヒロイン、現代のリリアン・ギッシュなのだ。」（柳下毅一郎、朝日新聞2009年9月25日夕刊2頁東京本社、「『空気人形』心を持ったラブドール」）
- ・「この映画は、無垢な空気人形の目を通して見渡した、現代社会のスケッチである」「町で空気人形が会うのは、心が満たされない人ばかりだ。心に空洞があるという意味では彼らも空気人形と変わらない」（シネマ5代表 田井肇、朝日新聞2009年11月12日朝刊28頁大分県版、「シネマワールド 空気人形—無垢な目で見た現代社会」）
- ・「公園に来る老人に、人形は『体の中は空っぽで、空気だけ』と告白する。老人は『私も同じだ。今時は皆そうだろう』と答える。是枝監督の最も伝えたいことが、この会話に表れているのではないだろうか」「登場人物は皆ひと癖もふた癖もありそうだが、実は見ている私たちも、彼らとそう変わらない。是枝監督は、現代人誰もが抱える孤独を、優しい寓話として提示している」（福永聖二、読売新聞、2009年9月25日夕刊11頁、「オールザットシネマ—空気人形：孤独の寓話 優

¹²⁷ 主要新聞社の新聞記事・雑誌記事検索データベース（朝日新聞、読売新聞、毎日新聞、産経新聞、日本経済新聞、東京新聞）、Lexus Nexus を利用（検索語は「空気人形」「air doll」「Koreeda」等）。

しく))

また、以上の新聞掲載の映画紹介の特徴は海外メディアでの『空気人形』についての紹介記事でも特に変化はなかった。

- “Director Hirokazu Koreeda, transforms this simple idea into a charming, richly textured fable that touches on fundamental questions of existence. Enlivened with touches of magic realism, the humorous film tries to define the elusive *je ne sais quoi* that makes us human, capable of feeling and expressing love, goodness, and truth.” “Her (the doll’s) lesson: We aren’t born with a soul, we acquire one through our relationships with others” (Tirdad Derakhshani. “Air Doll’: Japanese film breathes life into sex toy.” *The Philadelphia Inquirer*. June 18, 2010.)
- “The film provides an intimate look into modern illnesses such as self-imposed alienation, as well as easily overlooked charms of everyday life, through an innocent yet all-knowing pair of eyes. Nozomi’s open heart detects all the beauty and the filth around her, and the lovely *mise-en-scene* nimbly captures the process, maintaining a certain distance without missing the fine details.” (Lee Hyo-won. “Air Doll’ Leaves Poignant Sting” *Korea Times*. April 15, 2010.)

次に、ユリイカの 2009 年 10 月の臨時増刊号では、『ペ・ドゥナ総特集：「空気人形」を生きて』では、映画評論家等により、ペ・ドゥナについての長編の評論が多く掲載されている。また、月刊カット (Cut) の 2009 年 10 月号では『空気人形』が取り上げられ、是枝監督のインタビュー記事、映画批評が掲載された。それらの映画批評・評論の内容は、新聞記載の記事に比較すると、より映画の内容について批判的であり、ペ・ドゥナのこれまでの出演作品をも踏まえた本格的なものであるものの、本節で対象とするような、心を持った人工知能と人間社会といった視点から書かれたものは見られなかった。

- 「本作の構造は、まさに『あらかじめ内面を持たない』主人公と『つねに内面が散逸している』東京と、そこに蠢く『内面がいつも外在する何かに奪われながら』点在する人々の、スライド・ショーとしてある。そしてそれは、まさしくそのようなものとして映し出されるはずだった一。結論から言えば、それはそうはならなかった。それは、ペ・ドゥナがいたからである。…世界の厳然たるルールを軽々と超克していく怪物的な肯定性がスクリーンに蠢くのを目撃するのだ」 (宮寄広司。Cut 2009 年 10 月号. 89 頁)
- 「非現実の物語の、わずかなほころびのような間隙からやがて進しだす『真理』、そしてスパークする『気づき』。この現実を生きるわたしたちをして慄然とさせざるをえない、一瞬の閃光のような『真理』の瞬間にこそ本作の魂はある。」 (小柳大輔『Cut 2009 年 10 月号』89 頁)

6.3.3 分析等のまとめ

以下の仮説 (問題意識) が本節の出発点だった。

1. 心を持つ人工知能が生まれたらそれは人間の心にも影響を及ぼすものになる。そのことによって人間の心・脳が反応し、人間やその関係が影響を受ける。

2. 心・意識を持つ人工知能が活躍する社会では、ロボットの持つ心をどう倫理的に扱うかということを考える必要が出てくる。心を持ったロボットを虐待することは倫理的に許されなくなる可能性がある。逆に、それが許されたとした場合、そのことは人間間の倫理に影響を与える可能性がある。

第1の仮説（人間・社会への影響）について、以下が指摘できる。

- 1) 心を持つ前段階での人形であっても秀雄はその人形と時を過ごし、人形を擬人化していることで心の平穏を得ることができているとともに、「空気人形」の持つ機能も享受し、日常生活からの愛情の欠落を補っている。人形に実際に心を持たせることで、それらの経験の価値を損なうことも、増進させることもある。この映画の場合は、人形が心・意識、更に知性を持つことで秀雄にとっての人形の価値は損なわれ、ポジティブな心への影響は損なわれている。秀雄にとっての人形の価値は人間的な心を持たないこと、人間的な心からの解放であった。このような場合には人間は心を持った人形、ロボットからは逃避することになる。一般論としては、心を持つ人形・ロボットを必要としない、あるいはそのような人形・ロボットからの逃避を選ぶ人間はいるものであり、そのような人間は人形・ロボットを無視されるか、排除することで、人間や社会への影響を抑える動きとなるだろう。
- 2) ビデオ店員の純一は自らの存在の社会的な価値は代替の効くものであり、代用品であることから空虚感を持つ人間である。純一は人形のぞみから好意を持たれ、自分も空気人形と同じであるという純一との共感も持つようになる。純一も人形に対して好意を持ち、それはのぞみが空気人形であることを知った後も変わることはなかった。体に傷を作り空気が抜けていって弱っていく空気人形に自らの口で空気を吹き込むという偶然の体験を通じて、空気人形との疑似的な性的関係を持つことができている。映画ではこのような関係が続いた結果、純一は最後は人形に殺されることになってしまう。そうなった理由は映画では明確ではないが、過度のそして病的な愛情関係と、その対比において現実からの逃避願望から人間の男女が死に向けた選択をすることは現実にあるものであり、純一とのぞみがそのような人間同士の愛情関係にまで到達していることを暗示している。映画に引用された吉野弘の詩に記されているように人間はお互いに空虚さ、欠けている部分を補い合うことで、人間関係や人間社会を作っていく面があるものであり、そうであるならば、そのような人間間の補完を埋める役割を心を持った人形・ロボットが果たすことは当然予期できることである。
- 3) 映画では、空気人形や純一と似たような空虚さを持ち日常をやり過ごしている複数の人間の存在が映画の主筋とは離れたエピソードとして同時に描かれる。彼らの社会における存在と彼らが社会的に孤立している訳ではないことは、観客の立場から初めて分かることであり、映画の中のそれらの人間がお互いにまた「のぞみ」から影響を受けている訳ではない。ただし、心を持つロボットの存在が多くの人々が知ることとなり、人間の心の状況について俯瞰的な情報把握能力をも備えているロボットが社会に溢れるようになる世界では、映画の観客と同じ視線をそれらの同時代の人々が持つこととなるだろう。

第2の仮説（心・意識を持つロボットと倫理）について指摘できることは以下の通りであ

る¹²⁸。

- 1) 意識・心を持てば、クオリア（感覚意識体験）を持ち自然の美しさ等に気づくことになる。そのような意識体験について、それを人間が恣意的に止めたり再開したりすることは許されるか。更に、心を持つロボットが、古くなった場合や気に入らなくなった場合の廃棄をどうするか。機械に心を持たせることについて、それによって人間が受ける便益についてよりも、それによって機械が受ける不利益について関心が高まることは、映画の感想や映画の筋を見ると在り得るとみられる。
- 2) 映画では純一が自らの性的欲望のための、「空気人形」のぞみから空気を抜く行動をし、それはのぞみに苦痛を与えるものであった。心を持つロボットに苦痛を与える行為は許されるか。映画ではのぞみは愛する純一の行為を許容したが、ロボットが苦痛を与える行為を受け入れている場合はどうか。ロボットが心を持つようになった場合、それが苦しみや悲しみを持たないものであることはあり得るか。
- 3) 第1の仮説に関連するが、人間の心に影響を与えうるとすれば、心・意識を持つ人形・ロボットが、人間に対して悪意を持った働きかけをする可能性も出る。そのようにロボットを設計した人間がいて、ロボットが人間に危害を加えた時点で、現在の刑法犯で処罰の対象となるだろうが、ロボットの意志的行動の場合にはどうなるか。ただし、人間に危害は加えないが、心・意識に悪影響を与えることのみを目的とすることはあり得るだろう。それについての規制やルールが必要となる可能性がある。

以上指摘した点については、映画における空気人形の設定（非現実的とも言える）も影響している。映画では、心を持つ人形と、人間との区別が一見して容易ではない。心・意識に加え、人間とどれだけ似ているかは人間との関係の深さや内容に影響を持つだろう。また、映画では空気人形が自立的にビデオ店に行き、働き始めることができていることから、ビデオ店での人間関係が開始している。意識・心を持ったロボットはこのような自発的な身体能力を持つのは、現実的にはかなり困難な設定であるが、上と同様にそれは人間との関係性に影響を与えるものである。

上でのまとめは、それら映画における設定とのセットとして理解すべきものである。これらを技術的に実現するのは困難であるため、本節で指摘したような点は長期的な課題になるとみられるが、部分的にでも技術的進歩が起こる段階で、課題の一部は顕在化するとみられる。

参考文献

Dehaene, Stanislas, Hakwan Lau, and Sid Kouider. "What is consciousness, and could machines have it?." *Science* 358.6362 (2017): 486-492.

川添愛『自動人形の城：人工知能の意図理解をめぐる物語』東京大学出版会、2017年。

¹²⁸ 自立性な判断能力を持つロボットの暴力行為についての問題は、心・意識の有無に関わらず規制の対象となるものであることからここでは取り上げない。

キネマ旬報社.「キネマ旬報」2009年9月下旬号. No.1541.「空気人形」特集. 50～59頁.
月刊カット（Cut）. 2009年10月号. 通算253号. Rockin'on. 74～89頁.
業田良家『機械仕掛けの愛』第1巻. 小学館. 2012年.
業田良家『機械仕掛けの愛』第2巻. 小学館. 2013年.
業田良家『ゴータ哲学堂空気人形』小学館. 2000年.
人工知能学会.「人工知能」Vol.31. No.5 2016/9. 特集：人工知能とは何か？ 「人工知能と
Emotion」
人工知能学会監修『人工知能とは』2016年5月. 近代科学社.
青土社.「ユリイカ」2009年10月臨時増刊号. 第41巻第13号. ペ・ドゥナ総特集：「空気
人形」を生きて.
西垣通.『AI 言論：神の支配と人間の自由』（講談社選書メチエ）. 2018年.
松原仁『AI に心は宿るのか』2018年. 集英社.
モルモット吉田『映画評論・入門！観る、読む、書く』洋泉社. 2017年.
渡辺正峰（2017）「脳の意識 機械の意識：脳神経科学の挑戦」（中公新書）

7. 今後の展望

7.1 現時点でのまとめ

現在の人工知能の成果の多くはディープラーニングを先鋒とする機械学習の手法によっている。その担当者は理系の素養を基盤とした教育を受けた人達であり、日常的に自然科学で培った観察眼と思考形態を活用していると考えられる。言葉を変えて言えば、彼らの思考過程は科学哲学を基底とする認識論の枠組みによっているといえる。対象を外部から観察し、その対象に内在する普遍的原理、つまり法則を見出そうとする。この枠組みの下で得られる知識体系は、自然によって保証された「確かな根拠」によって支えられている。観察者の願望や価値観とは無縁のものである¹²⁹。

一般システム理論は、存在するもの全体を考察の対象に置き、そこに通底する理論を見出すことを願ってはじめられたが¹³⁰、自然科学の対象領域では対象固有の分析的方法論が効果的に知見を生み出したのに対して、システム論では対象を構成する要素の関係性に着目して、（要素の内部に分け入るのではなく）全体を包括的に把握することにより新たな知見を生み出すという方法論に独自性がある。

さて、ここでは、宇宙を構成する知のカテゴリーに関し科学哲学を基底とする認識論の枠組みで分類し（要素に分け）それぞれのシステムの間を明らかにした¹³¹。

- ・自然システム：普遍的（どこをとっても何時においても）に内在原理（法則性）がある
- ・人工的物理システム：ヒトによって構築されたが普遍的内在原理が存在している。自然システムと人工的物理システムはこの枠組みの中では同一の挙動をとるので、合わせて「ハード1」と区分する。

- ・人間活動システム¹³²：価値観や情動に促され個別に独自の行動を行っているヒトを含むシステム。これを「ハード2」と区分する

- ・人工的抽象システム：ヒトは「思考世界」を構想する。「実体世界」はヒトの認識の外に存在しそれを「外界」と言うのであれば、思考世界はヒトの認識の内「内界」に存在し、「モデル」としてそれを外部に表示することが出来る。自然システムや人工的物理システムを対象にして思考世界で構想したモデルは「ソフト1」と分類する。「ソフト1」は実体世界に照らし合わせモデルとしての確からしさを確認できる。仮想的な価値観や同一の情動を有するヒトの集団を想定し思考世界にモデルを描いた場合それは「ソフト2」と分類する。「ソフト2」の確からしさは実体世界に照らして確認することは困難である（そのような実体は通常存在しない）。さらに、独自の価値観や情動に導かれて行動するヒト（ないしヒトの集団）を思考世界で描いた「モデル」は「ソフト3」と分類する。「ソフト3」の確からしさを確認することも困難ではあるが不可能ではない（現に存在しているが、個々に異なり（不均一）変動し

¹²⁹ 戸田山和久「知識の哲学」産業図書（2002）観察者の「確信」は「確かな根拠」（アンカー）には当たらないと考える

¹³⁰ ルードイツヒ フォン ベルタランフィ（長野 敬、太田邦昌 訳）「一般システム理論」みすず書房（1973）

¹³¹ 本報告書2ページ参照

¹³² ヒト自体は幾つかのシステムの構成要素になる。医学や生理学の対象となる人体（自然システム）、標準的な価値観を有する組織や社会の構成要素であると仮想する経済学モデル（人工的抽象システム：ソフト2）

移ろいゆく（不定）ところに困難さがある）¹³³。

「次世代人工知能の構想Ⅰ」では、ここでまとめた「科学哲学を基底とする認識論」に基づく「知識論」¹³⁴と「一般システム理論」により区分される対象システム¹³⁵の知識特性とを組み合わせ得られる枠組みに従って考察を深めたものである。

7.2 次期計画「次世代人工知能の構想Ⅱ」の構想

本報告書では3種類のアプローチにより「次世代」の人工知能を構想し、事例的に探索を深めた。

- ・ トランスフォーマティブ・アプローチ
- ・ 新ディシプリン（新知識領域）の具体的探索
- ・ 新パラダイム（新たな哲学）による探索

しかし、いずれもまだ原初的段階に止まっていると言うべきであろう。(DLが不得手とする)新ディシプリン(新知識領域)としては脳科学を想定したがAIにつながるキーとなる知識はまだ芽生えていないのではないか。また「新パラダイム(新たな哲学)による探索」においては、今回は科学哲学以外の立場を十分に考察する時間がなく、次の機会に現象論哲学等の異なるものの見方に立って、科学哲学の立場では立ち入ることが困難なヒトの「内界」の世界を陽に取り込んだAIを追究し、次世代人工知能の可能性について検討を深めたい。

この領域は学問では「人文学」を中心にしたヒトの本性に迫る未開の領域である。

¹³³ 本報告書2ページ参照

¹³⁴ 参考資料129

¹³⁵ Peter Checkland *Systems Thinking, Systems Practice* pp.111-112(1981) John Wiley & Sons.