

人工知能、ロボット、知性

久木田 水生

1 序

ロボット⁽¹⁾は古くから哲学的な考察の対象になっており、特に、機械が知性や思考能力、精神を持つことができるか、という問題は大きな哲学的論争を巻き起こしてきた。知的な作業を実行する機械が現実には作られるようになると、この主題は哲学の枠を越え、計算機科学者や工学者を巻き込んで議論されるようになった。知的な作業を行う機械を作ることは人間の知性について理解することを助ける一方で、知的であることを意図されたロボットを作るためには、製作者が「知性とは何か」ということについてあらかじめ何らかの理論を持っていることも必要である。このようにしてAIとロボットは知性についての私たちの理解と相互に影響を与えながら発展してきた。

AIやロボットを作る動機には異なる二つのものがある。一つはあくまでも実用上の役に立つ道具を作るという動機、もう一つは可能な限り人間に近い（あるいは人間を越える）知性を持つものを作るという動機である。実用的AIは今ではありふれた技術として利用され、日々進歩している一方で、後者のAIは初期のAI研究者が予期した成果にはそれほど近づいていないように思われる。

しかしながら実用的なAIを含む情報技術の進歩は後者のAIとロボットに関しても、また認知・知性についての哲学的理解に関しても、新しい方向性を示唆するように思われる。それはテクノロジーと人間の協働による知的活動の実現、という方向性である。現在、私たちの社会には多くの情報機器が浸透している。そして私たちは無意識のうちにコンピューターを利用し、そしてコンピューターに利用されている。この状況は、技術について、そして人間知性についての従来の理解を大きく変化させる可能性がある⁽²⁾。さらにこのような技術がペットロボット、

(1) ロボットとは何かということに明確な答えは存在しないが、本稿においては、自律的に行動する人工的なシステム、と大雑把に捉えておく。この意味でのロボットは物理的身体を持つ必要はなく、従って人工知能(AI)や「ボット」などもロボットである。

(2) Cf. Clark (2003).

医療・看護ロボットなど、私たちの生活に密着したロボットに組み込まれる時、ロボット倫理学において既に論じられている問題（例えばプライバシーの侵害や、ロボットの行為者性・被行為者性など）がますます深刻になることは確実である。

これらの重要な哲学的・倫理的問題について考えるためには、これまでのAI、ロボットの発展と哲学が相互に与えてきた影響を振り返ることは有益であるだろう。そこで本稿においては人工知能・ロボットの発展の歴史を振り返り、人工知能と哲学が互いに与えてきた影響について概観する。

まず2節ではAIの発展を、特にその背景となる設計パラダイムに焦点を当てて概観する。ここで注目するのは最も初期のAI設計のパラダイムである論理的AIと、それに対するオルタナティブとしてのニューラル・ネットワーク、遺伝的アルゴリズム、状況的ロボットである。また近年のウェブ・コンピューティングと新しいユーザーインターフェースが示唆する新しいAIのあり方についても考察する。3節では上記のAIが哲学とどのように関係してきたかを概観する。特に論理的AIと計算主義、機能主義の関係、そしてニューラル・ネットワークと状況的ロボットのそれぞれが支持する哲学的見解、すなわちコネクショニズムと身体化された精神説について考察する。最後に、これらのAIが直面する困難に対して、クラークが提唱した「拡張された精神仮説」を紹介する。

2 人工知能の発展

2.1 初期の人工知能—論理的AI

人工知能という言葉が使われるようになったのは1950年代であるが、人間の知能を機械によって実現するという考えはそれよりはるか以前に遡る。人工知能と呼べるものの元祖は19世紀にチャールズ・バベジによって設計された解析機関であろう。これは、パンチカードに書かれたプログラムと入力を受け取ると、蒸気機関を動力にした機械が、自動的に計算を行ってくれる、というものであった（ただしこの機械は当時の工作技術の限界と資金の不足によって実際に製作されることはなかった）。

論理的な推論を実行する機械の最も初期のものは19世紀にウィリアム・ジェヴォンズによって考案された「論理ピアノ」である。これは解析機関に比べればごく単純な機械で、人力で動作し、限定された形式の命題論理における推論を行うものだった。ここで注目すべきことは機械それ自体よりも、論理的推論が代数的に表現された命題の演算として捉えられていることである。このような方法はジョージ・ブールによって論理学にもたらされた。ブールの論理学は二つの点で画期的だった。一つは論理的な推論を代数的な記号の演算によって定式化したこと、もう一つは従来の三段論法に限定されない一般性を持つ推論の体系を作ったことである。

ブールの論理学は伝統的な論理学からの大きな進歩ではあったが、しかし完全に一般的では

なかった。例えばブールの論理学ではいわゆる多重量化を扱うことはできない。多重量化とは例えば「すべての人に愛する人がいる」のように「すべて」と「存在する」が入れ子のように使われることである。数学においては多重量化が不可欠である。量化を含めたより一般的な論理学は19世期末、ゴットロープ・フレゲによって作られた。フレゲの論理は、ヒルベルトらによる洗練を経て、現代の記号論理学の最も標準的な体系になっている。

記号論理学の体系が確立されたことは、人間が行う推論（の一部）を機械的に遂行するという試みへの最初の扉を開いたと言える。しかし記号論理学はあくまでも記号同士の静的な関係の理論である。機械による推論の実現には、記号操作の機械的遂行についての理論と、その工学的実装が必要だった。

人工知能における最も重要な理論的ブレイクスルーは1930年代、アラン・チューリングによってもたらされた。チューリングは、計算とは与えられた規則に従って記号を書き換えていく過程である、という考えに基づき、現在では「チューリング・マシン」と呼ばれている概念的な機械を考案した。チューリング・マシン(TM)は記号を書き込むマス目を持ったテープと実行ユニットから構成される。実行ユニットは、テープの上を移動する、マス目の記号を読み取る、マス目の記号を書き換えるという動作を行う。実行ユニットには「マス目の記号がこれこれならばこれこれの動作をせよ」という仕方で命令を与えることができる。命令には番号がつけられ、一つの命令の後には次にどの命令を実行すれば良いかということの指示も与えられる。マシンに与えられる命令の組み合わせをプログラムと呼ぶ。

通常算術において使われる関数はすべてTMに適切なプログラムを与えることによって計算させることができる。またTMが扱う記号それ自体にはそもそも何の意味も与えられていないため、記号の解釈の仕方によってTMは算術以外の様々な記号操作をしているものと考えることができる。例えば私たちはある種のパターンに従った推論をTMに遂行させることもできる。初期のAIの発想は一般にこのようなものであった。つまりTMと類似の機械が扱う記号に一定の解釈を与え、そしてその解釈のもとで、その機械に正しい推論と思われるような記号操作を実行させるプログラムを与える。このような発想に基づくAIを「論理的AI」と呼ぶことにする。

1950年代には、「人工知能」という言葉が使われ始め、1956年にはダートマス大学で始めて人工知能に関する集会が行われ、人間の知性的活動をシミュレートするという目標が掲げられた。この会議では、ハーバート・サイモン、アラン・ニューウェル、クリフ・ショーによって開発されたLogic Theoristというプログラムが紹介された。このプログラムはホワイトヘッドとラッセルの『プリンキピア・マテマティカ』における定理のいくつかを証明することができた。高度な知性を要すると考えられる論理的な証明を機械が遂行したことは人工知能の輝かしい未来を約束するよう思われただろう。

しかし人工知能に対しては当初から強い批判もあった。代表的なAI批判者は哲学者のヒューバート・ドレイファスである。1965年、ドレイファスは「錬金術と人工知能(Alchemy and

artificial intelligence)」と題された報告書を書き、それが非公式に回覧された。ドレイファスはAI研究者が予想するAIの可能性を真っ向から否定し、AI研究者をホラ話で資金を詐取する者のようにこき下ろした。ドレイファスの批判の要点は、人間の知性の本質は記号的な計算や推論ではシミュレートできない部分にこそある、ということである。ある技術に関して、初心者や中級者は頭の中であれこれ考えてから行動するが、エキスパートほどそのような明示的な思考なしに直観的に問題に反応し対処する。そして知的な人間とはそのような直観を持ったエキスパートを言うのである。従って明示的な計算と推論にもっぱら依存する人工知能は真の知性を実現することはできない、とドレイファスは論じた⁽³⁾。ドレイファスの批判には明らかに誤っている点もあった。例えば彼は、機械がチェスでエキスパートを負かすようには決してならないだろうと断言していた。

エキスパートが直観に従って下しているように見える判断も、実際には無意識のうちに何らかの規則に従っていることがある。そのような明示化されていない規則を抽出して、専門家の判断をシミュレートするAIが80年代には盛んに開発されるようになった。例えば患者の症状から病名を診断するAIは、実際の医者に引けを取らないほどよく病名を当てることができた。中には専門家によっても知られていなかった科学的事実を発見できたAIもあった⁽⁴⁾。しかし人間の知性には、規則として明示化できない部分が残されるということはおそらく真実である。機械によってシミュレートできるのはどこまでなのか、そして機械によってシミュレートできない直観的判断は一体いかなるプロセスを経て下されているのか、ということは人工知能や認知科学、哲学にとって興味深い問題であり続けるだろう。

例えば人間の顔を識別する、障害物をよけて移動するといった課題は、人間にとってはごく容易であるが、規則ベースの論理的AIのアプローチによって遂行することは極めて難しいと思われた。また論理的AIは自らの経験から学習をして、従うべき規則を修正することができなかった。論理的AIのこういった欠点を克服するためのアプローチがいくつか提案されてきている。

2.2 いくつかのオルタナティブ—ニューラル・ネットワーク、遺伝的アルゴリズム、状況的ロボット

論理的AIは数学的な計算や論理的な推論に注目して、それを規則に従った記号的な操作によって模倣することを目指す。その際、実際の記号操作の過程がどれだけ生物の認知と情報処理の過程に近いかは問題にされない。解析機関やコンピューターは大量のデータを高速で処理するという実用的な目的のために作られたものであり、正確なデータ処理を人間より速く行うことができるならば、その方法は問われない。初期の人工知能は基本的にこういった、結果と

(3) Cf. Dreyfus and Dreyfus (1986).

(4) Cf. Gillies (1999).

実用性を重視する伝統を継承していた。しかし人工知能と隣接する人工生命の分野では初期から生命的現象を生み出すメカニズムに焦点が当てられていた。論理的AIの限界に直面して、AI研究者たちもまた、システムの振る舞いだけでなく、知的振る舞いを生み出す生物的なメカニズムに注目するようになっていった⁽⁵⁾。ここではそのようなアプローチの代表的なものである、ニューラルネットワーク、遺伝的アルゴリズム、状況的ロボットに注目しよう。

ニューラルネットワークは人間の神経系を模したモデルで、多くの「ニューロン」が網の目のようにつながれた構造をしている。ニューロンには信号を受け取る端子と信号を送り出す端子がどちらも複数ついている。それぞれの端子は他のニューロンと「シナプス」によってつながっている。送信される信号はそのニューロンが発火したか否かということだけを伝えるものであるが、シナプスには異なる重み付けがしてあって、信号がそれに応じて増幅される。受け取った信号が一定の強さ、「閾値」を越えるとそのニューロンが「発火」し、送信用の端子から信号が送り出される。ニューロンは層をなしており、入力層から出力層へと順次信号が伝達される。各シナプスの重みを変化させることで、全体として入力と出力の組を変化させることができる。入力と出力はベクトルで表現され、各成分が入力層のニューロンに与えられる信号になる。例えば入力層には(A、B、C、D)というニューロンがあるとして、ここに(3、1、0、4)という入力信号を与えるとA、B、C、Dにそれぞれ3、1、0、4という強さの信号が伝達される、というように。

ニューラルネットワーク(NN)の一番の特徴は自動的な学習が可能だということである。例えば人の顔写真から性別を判定するという課題をNNに遂行させようとしたとする。最初は重み付けはランダムに設定してある。トレーニングとしてNNにいくつかの写真をみせる。NNの判断が間違っていた時には、間違っているということを知らせて重み付けを変更させる。一定期間のトレーニングの後、システムはトレーニングに使った写真だけではなく、初見の写真でも、かなりの成績で男性か女性かを正しく判定できるようになる。従来の論理的AIとは異なり、NNの重み付けは人間の設計者が意図的に与えたものではない。それは学習によってシステム自身が獲得したものなのである。

遺伝的アルゴリズムもまた人工的システムが学習することを可能にする方法である。何らかの問題に対して、その解とみなされうるものを例えば8桁の数字にコード化しておく。この8桁の数字を遺伝子型と呼ぶ。遺伝子型を解にデコードしたものを表現型と呼ぶ。まずランダムに遺伝子型の集団を発生させる。それらの表現型を正解と比べ、どれだけ正解に近いかを評価し、成績の良いものをいくつか選択する。選択された解の遺伝子型に対して、ランダムな変異や、二つの遺伝子の交叉(一方の遺伝子型の前半と他方の遺伝子型の後半を組み合わせる)の処理を施して、その結果から新しい遺伝子型の集団を発生させる。この作業を繰り返していくと、次第に正解に近い表現型を持つ遺伝子型が発生するようになる。遺伝的アルゴリズムは、特定

(5) Cf. 久木田 (2007)。

の問題に対する解を発生させることだけでなく、言語的コミュニケーションや協力行動など、生物のある行動が進化的にどのように発生してきたかを研究する方法として利用されている。

もう一つ従来の論理的AIに変わるパラダイムは状況的ロボティクスである。これは従来のAIを悩ませた問題の一つである「フレーム問題」に対処するためのアプローチである⁽⁶⁾。AIが現実の問題に対処しようとする時、それはその行動の帰結として何が起きるかを推論しなければならない。しかし現実世界においては一つの行動の帰結が無限に存在し、AIがそのすべてを推論することはできない。従ってAIは考慮すべき事柄に関して推論し、そうでない事柄は無視する必要がある。しかし何を考慮すべきで何を無視すべきかということ推論するためには、AIは結局すべての事柄を考慮しなければならない、推論が終わることがない。AIは解決すべき問題を前にして延々と推論を繰り返し、いつまでたっても問題に対処することができない。これがフレーム問題である。

状況的ロボティクスは、身体化されたロボットを特定の状況に置き、環境とロボットの間の力学的な相互作用（これを「カップリング」と呼ぶ）を通じて問題の解決を目指すというアプローチである。従来のAIは環境の特徴を内部の記号的表象に置き換え、それからその記号に対して計算を施し、行動の決定をするものだった。これに対して状況的ロボティクスでは、内部の記号を処理するという過程を迂回して、ロボットの身体と物理的環境との直接的なカップリングを通じて問題を解決することを目指す。例えばBrooks (1986) は環境との適切な相互作用による身体的な行動を実現するために、従来のような表象と計算に基づくAIとは全く異なる設計思想を開発した。それが彼が「包摂的アーキテクチャー」と呼ぶ設計である。ブルックスのロボットは、中枢の制御ユニットも外部世界の表象も持たない。その代わりにそれは外部世界の環境に置かれた時に、外部環境と直接的にカップリングすることによって環境に適応した行動を取れるように作られている。その内部では比較的単純な動作を行う複数のモジュールが階層状の構造をなし、そして隣接するモジュール同士が適切に相互作用する。このように作られたブルックスのロボットは、障害物を避けて部屋の中を移動するというような、従来の記号的AIでは困難だった課題を遂行することができた。このような、環境の中に置かれ、環境と相互作用するように設計されたロボットは、「埋め込まれたロボット」(embedded robots)、「状況的ロボット」(situated robots) などと呼ばれる。

これらの方法はそれぞれ論理的AIに対してアドバンテージを持つ。ニューラルネットワークは学習能力とパターン認識能力において、遺伝的アルゴリズムはやはり学習能力によって、状況的ロボットは現実世界で行動する能力において、論理的AIにまさる。またこれらのAIは、単に入力に対して正しい出力を返すということだけではなく、何らかの意味で人間（を含む生物）により近いメカニズムを模倣しているという点でも論理的AIとは異なっている。しかし他方で論理的AIが得意とする数学的な計算、および論理的な推論に関しては、これらのアプ

(6) Cf. Wheeler (2008).

ローチは絶望的である。これらが例えば足し算のような単純な計算をできるようになる可能性は限りなくゼロに近い。ここから私たちが得られるありふれた教訓は、人間の知性は多様であり、何か一つのパラダイムで人間の知性がすべて実現できる見込みはなさそうだ、ということである⁽⁷⁾。

次節では以上のようなAIの発展が知性と認知についての哲学的な理解にどのような影響を与えてきたかを振り返る。しかしその前に、現代の情報技術の進展がもたらす新しい人工知能のあり方について触れておきたい。

2.3 人工知能と自然知能の融合—ウェブ・コンピューティングと透明なコンピューティング

実際のな計算機科学者の多くは機械に自律的に行動させることではなく、コンピューターに人間の作業をサポートさせることにより大きな関心を持ってきた。その際に重要なのは情報の保存、検索、提示の方法、およびコンピューターと人間の間のインターフェースである。これらの技術の発展は新しい人工知能のあり方を示唆している。

コンピューターと人間の間のインターフェースは次第にコンピューター中心から人間中心に移っていつている。コンピューターは人間の認識と身体的動作に適合するように変化し、それに従って私たちのコンピューターを操作しているという意識は薄れていく。これからのコンピューターはますます継ぎ目なしに現実の世界へと浸透していこう。ドナルド・ノーマンはこのような技術を「透明なコンピューティング」と呼んでいる。

情報の保存、検索、提示においても現在、大きな変化が進行している。その最大のものがウェブ・コンピューティングである。現在、ほぼすべてのコンピューターがインターネットに接続され、互いに情報を交換しあっている。このことによって実質的に世界中のコンピューターが一つの巨大な情報プールになっているのである。この巨大な情報プールを活用するための技術が盛んに研究されている。この技術は単に情報工学的なものに止まらない。それはインターネットにつながっている人々をいかに利用するかという、社会工学的な次元を含んでいる。

さらに重要なことは、インターネットを利用するユーザーの集団は、明示化され言語化された知識以外の知識を絶え間なく生み出し続けている、ということである。インターネット上でユーザーの様々な活動、例えばウェブサイトの閲覧、情報の検索、商品の購入などが記録され、その膨大なデータがユーザーが利用できる仕方で提示されている。

こういった技術の発展は人工知能、あるいは人間の知能についての私たちの理解にとって大きな意味を持つ。例えばGoogleのようなサービスが私たちに何を提供しているのかを考えよう。それはまず私たちが知りたい情報を探すことを助けてくれる。その際Googleは単にその

(7) ただし複数のパラダイムが存在することを認めた上で、認知の基本的なアーキテクチャーが何であるかという問題は残り、AI研究者、ロボット工学者の間では論争が続いている。

語が含まれているページを探すのではなく、インターネット上での人々の行動とウェブページ同士のリンクから、検索語に関連する重要な情報を載せていると思われるページを推測して私たちに教える。Googleは私のメールを管理して、迷惑メールかもしれないものをより分ける。送受信されたメールの内容から判断して、私が興味を持つかもしれないサービスを紹介する。

こういった技術はある意味で非常に高度な知性を表現しているように思われる。第一にそれは膨大な、そして様々な種類の「知識」を持っており、ユーザーからの請求に応じて直ちにその知識を提供する。場合によってはユーザーからの請求がなくても、ユーザーの潜在的な需要を推測して情報を提供する。第二にそれはこの膨大なデータから新しい知識を推論することができる。その推論には演繹的なものもあればそうでないものもある。第三にそれは絶えず情報を収集しながら、その知識を更新している。それが有する知識の中にはWikipediaに書かれているような百科事典的なものもあれば、今日の名古屋の天気というような現実世界の現在の状態についての情報、インターネットを利用しているユーザーについての個人的な情報もある。さらに個人情報にはユーザー本人（あるいは他の誰か）が意識的に公開しているものもあれば、ユーザーの行動から自動的に抽出され記録されているものもある。

ここで問題になるのは前段落で言及されている「それ」が一体何を指しているのか、ということである。それはもちろん単独のコンピューターでも単独のプログラムでもない。それはインターネットを支える技術的インフラ、それによってつながれたコンピューター群、その上で協働する多数のプログラム群、そしてそれを利用するユーザー集団からなる巨大な共生体である。従ってこれは伝統的な意味での人工知能ではない。それは人工知能と自然知能の融合したものである。現在のコンピューターは単に人間の作業を模倣するだけでも、人間の活動をサポートするだけでもない。それは膨大な数の個人の持つ知識、そしてその集団の活動を利用して、これまでのコンピューターでは決して実現できなかった高度な知的振る舞いを可能にしている。

3 人工知能と知性の理解

3.1 論理的AIと計算主義、機能主義、行動主義

チューリング・マシンに始まる論理的AIは、形式的な記号体系とそれを操作する規則、そして記号に意味を与える意味論から成り立っている。論理的AIは実用においては大きな成功を収め、その成功が非常に顕著であったために、知性についての哲学的理解にも大きな影響を与えた。その影響の一つはジェリー・フォダーに代表される計算主義の台頭である。計算主義者は、人間（生物）には独自の内的言語があり、それを機械のように処理するのが精神活動である、という精神のモデルを立てる。

もう一つの影響は、心的状態をそれが果たす機能と同一視する機能主義である。TMは計算において様々な内部状態を遷移しながら特定の入力に対して特定の出力を返す。機能主義は人

間の心も同様のものであると考え。つまりある心的状態とは、入力と出力、そして他の心的状態との関係によって決定されるものである。このように考えると心を実現するものは生物学的な脳であっても、半導体と電線であっても構わないということが帰結する。

チューリングは自身は行動主義の立場に立ち、機械が考えているか否かは、外面的な振る舞いを見て判断すれば良い、と考えた。そこで彼が提案したのが今日「チューリング・テスト」として知られている、次のようなテストである。ある判定者に機械または人間と、パソコン通信のチャットのような仕方に対話をさせる。判定者は自分の相手が機械なのかそれとも人間なのかを知らされていない。しばらく対話をして、判定者に自分の相手がどちらだったかを判断させる。この過程を繰り返して、判定者が機械と人間とを見分けることができていなかったら、その機械は人間と同様に考えているとみなしても良い。

チューリングはAIが十分に発達すれば、このテストにパスするようなものが現れる可能性はあると考えていた。確かに原理的にこれが不可能であると想定する理由はないし、実際はかなりよく人間を欺くことのできるAIも作られている。一方、このテストの有効性に関しては多くの疑問が向けられた。Searle (1980) は「中国語の部屋の議論」を提示して、記号の意味の理解が伴わなければ、いかに入力された記号と出力された記号が適切に対応していても、その記号操作システムが知性を持っているとは言えず、それゆえチューリング・テストは知性のテストにはならない、と反論した。こういった反論以降、AIが本当の意味で思考していると言えるためには、単なる論理的AI、記号を操作する「構文論的機械」であるだけでは不十分で、それはまた言語の意味を理解する「意味論的機械」でなければならない、という認識が一般的になった。

また前節で見たフレーム問題のような、論理的AIによって実現できる機能の限界も論理的AIが知能と呼ぶに足らないとする批判を招いた。

3.2 ニューラル・ネットワークとコネクショニズム

初期の論理的AIの弱点の一つは、パターン認識など、記号操作の規則によって表現できないような推論だった。また論理的AIは経験から学習するというのも苦手であった。これをある程度克服したのがニューラル・ネットワーク (NN) である。論理的AIが人間によって書かれた明示的な規則に従って計算をしているのに対して、NNはそのような明示的な規則を持たない。計算機としてのNNの性質はニューロンとシナプスによって作られる構造と、そして各シナプスに対する重み付けによって決定されている。ここには複雑な文法も高度なアルゴリズムも存在しない。あるのは個々の要素が単純な規則に従って行う動作の集積である。十分に複雑なネットワークにおいては、個々の要素が単純な規則に従って行う動作の集積が、システム全体の非常に複雑な振る舞いを創発させる。

津田一郎は、生物の脳と神経系にしばしば見られるカオス的振る舞いが脳の機能において重

要な役割を果たしている可能性を示唆しているが、再帰的NN（出力が再びネットワークの中にフィードバックされるNN）においてもカオス的振る舞いが見られ、それが自律的な運動を円滑に行うことを可能にしていると考えられることは、NNが生物の脳のモデルとしての成功をより確からしくするものである⁽⁸⁾。

NNの成功は哲学においてコネクショニズムという立場を生み出した。コネクショニズムは、精神が内的言語のようなものを持つことを否定し、ニューロンからなるネットワーク全体の状態が精神の状態を表わすものとする。コネクショニズムの利点は様々である。特にコネクショニズムではパターン認識など、言語的に伝達できないが、経験から学習されるスキルをうまく説明することができる。またHarnad (1990) は記号接地問題へのアプローチとして、経験からの学習によって自律的に概念を形成することのできるNNを利用することを提案している。

人間の知的活動がかなりの部分、脳に依存していることは確かである。しかし上述のように人間の知的活動のすべてをNNによって実現することは容易ではない。というのもNNはパターン認識などにおいて大きな力を発揮する一方で、論理的な推論や数学的な計算は苦手としているからである。また私たちの心のすべてをNNのような仕組みで説明しようとするならば、私たちは自分の心について抱いている伝統的なイメージの多くを放棄しなければならない。このことは私たちの行動の説明を少なからず困難にする。コネクショニストはNNを認知の基本的なアーキテクチャーと考えながら、その上で様々な能力を組み合わせることによって、論理的な推論を含むより広範な人間の知的活動のモデルを考えている。

3.3 状況的ロボットと身体化された認知

古典的なAIの困難のもう一つは、生存にとって必要な行動を取る際、しばしば表象や計算が何の役にも立たないように思われる、ということである。計算主義の認知モデルによれば、生物は表象によって外部世界のモデルをその精神の内部に作り上げ、そして中枢の計算ユニットが規則に従って内部の表象を操作し結論を導き出す。しかし生物の行動の多くはこのような過程を経ないように思われる。例えばコオロギのメスはオスの鳴き声に反応してそちらに近づいていく。しかしコオロギはオスの鳴き声を聞き分け、そこからある方角にオスがいるという表象を抱き、それに基づいてそちらに向かうべきだという判断を下しているわけではない。メスの聴覚器官はオスの鳴き声だけに反応する仕組みになっており、そしてメスは左右の肢にある鼓膜に届く音の違いに直接反応して進行方向を修正している⁽⁹⁾。このようにしばしば生物は、内的な表象とそれに対する計算や推論を介することなく、環境に対する直接的な反応として行動をしている。環境との力学的な相互作用の結果として身体のある部分が反応し、その反

(8) 津田 (2002)、久木田 (2010)、Elman et al (1997)、Churchland (1995)。

(9) Cf. Clark (2001)。

応がさらに身体の別の部分の反応を引き起こす。中央で身体の各部分を制御しているようなユニットは存在せず、部分同士がお互いに対して適切に反応することによって、全体として調和した行動が可能になっているのである。このような行動にとって表象は単に不要であるというだけではなく、しばしば有害ですらある。目の前に飛んでくる石を避けるためには、表象の計算をしていたのでは遅すぎるのだ。

身体の制御という点において、動物（あるいは植物も）は自らの置かれた環境の状態と自らの身体の状態（姿勢、向き、移動の早さなど）を内部の計算的中枢の中で記号的に表象し、それを一定の規則に従って計算しているわけではない。私たちの身体は構造それ自体が、与えられた環境において適切に反応できるように作られているのである。もし環境の中で適切に行動することも知性と呼ぶことができるのであれば、その知性は論理的AIのような仕組みによって実現されるのではなく、環境とカップリングする身体に宿っているということができるだろう。このように表象を介さずに外部環境の情報を利用する方法は表象なき認知、身体化された認知と呼ばれる。

しかし状況的ロボティクスにもNNと同様の弱点がある。表象を持たないロボットは一定の環境において行動を制御することは得意であるが、しかし高度な計算や推論には全く向いていない。状況的ロボティクスは人間の知的活動のすべてを説明するものではなく、その基本的なある側面においては表象が不要だということを示すものである。ここからは人間の知性が多様な仕方で実現されているものだということが示唆される。実際、上述した異なるタイプのAI設計を組み合わせた様々なハイブリッド・システムが開発されている⁽¹⁰⁾。

3.4 拡張された精神仮説

ニューラル・ネットワークと包摂的アーキテクチャーに基づく状況的ロボット、そして遺伝的アルゴリズムはすべて一つの重要な設計原理に基づいている。それは比較的単純な要素を組み合わせ、それらを相互作用させることによって、システム全体として複雑で高度な振る舞いを実現するということである。このような設計の利点は、それが漸進的な進化によって作られたという説明をもっともらしいものにする、ということである。チューリング・マシンのような明文化されたプログラムに従って動作するAIは、その部品やプログラムの一部でも欠けてしまうと全体としての動作に大きな支障をきたす。従ってそれが漸進的な進化の産物として生まれたとみなすことは難しい。しかし人間の精神が生物進化のある段階で生じたということはかなり確かであるように思われる。従って、もし人間の知性がチューリング・マシンのようなものであるとするならば、それが進化の過程においてどのように誕生したのかを説明することが困難になる。その点、ニューラル・ネットワークや包摂的アーキテクチャー、遺伝的ア

(10) Cf. Sun and Alexandre, (1997).

ルゴリズムは漸進的進化という考え方とうまく調和するという利点がある。その一方でこれらのAIが人間に特有の高度な精神活動を行えるようになるとは思われない。表象を持たず昆虫のように床を歩き回るロボットがどのようにして高度な数学の定理の証明を行えるようになるのだろうか。前世紀の終わりに哲学と人工知能が直面していたアポリアはこのようなものだった。この問題に対して近年アンディ・クラークらが提唱しているのが、「拡張された精神仮説」(EMH)である⁽¹¹⁾。

進化生物学者のリチャード・ドーキンスは、生物の表現型はその生物学的な身体だけではなく、クモの巣やビーバーのダムのように、それらが作り出し、そして生存と繁殖のために利用する様々なものにまで拡張されている、と主張する。そして彼はそれらを「拡張された表現型」(extended phenotypes)と呼んでいる。これと同様にクラークらは、人間の精神的活動はその生物学的な脳や身体のみによって生じるのではなく、人間の生物学的身体と、人間が作り出し利用する環境を含めたシステム全体によって生じる、と主張する。彼がその典型として例に挙げるのが言葉（特に書き言葉）の使用である。私たちは紙とペンを使わずには複雑な計算をすることはできない。紙と鉛筆、そして脳を含む人間の身体のすべてが一つのシステムとしてこの計算という複雑な精神活動を可能にしている。クラークは人間の生物学的脳はパターン認識や身体の協調などに優れているが、入り組んだ推論を導くことや複雑な計画を立てることには向いていないと言う。人間がそれらを遂行することを可能にしているのは、それらを助ける環境を身体の外部に作り出しているからである⁽¹²⁾。

この仮説は発表以来、様々な議論を巻き起こした⁽¹³⁾。簡単には切り離すことができない脳や手足と違って、紙やペンは一時的にその人の活動を補助するに過ぎないと考える人々もいる。またペンや紙を操作する能力が知性であり、それはペンや紙に帰属させられるものではないと考える人々もいる。こういった反論から分かるのは、知性は生物学的身体に属する恒常的な能力であるという考えが根強いということである。しかし近年の認知科学や人工知能の研究によって、知性とは課題指向的で、分散した、特定の環境に限定されたものでありうるという認識が広がっている⁽¹⁴⁾。また上述したアポリアに対処する有力な仮説が他にないということも、EMHにとって有利な状況である。

EMHの観点から見れば、ニューラル・ネットワークや状況的ロボットは人間の生身の身体や脳のモデルであり、そして記号的AIは人間とスマート・ワールドからなるシステムのある一側面のモデルである。もし人間の知性がEMHが説明するようなものであるならば、これまでのAIのパラダイムのどれも十分に人間知性をシミュレートすることはできない、というこ

(11) Clark and Chalmers (1998).

(12) ただしこのことはクラーク以前にコネクショニストたちによっても主張されていた。Cf. Rumelhart, Smolensky, McClelland and Hinton (1986). この点を筆者に指摘してくれたのは呉羽真氏である。

(13) EMHをめぐる賛否両論については、EMHに関する主要論文を収録したMenary (2010)の序文を参照。

(14) Cf. Clark (2003).

とになる。AIが取り組むべき課題は、いかにして人間が外部の環境を自分たちの認知プロセスを促進するようなものにする技術を獲得したかを考え、その過程をシミュレートすることであろう。そしてそのようなAIの成功は、EMHに大きなサポートを提供するだろう。EMHはさらに、現在のウェブ・コンピューティングがもたらした新しい知性のイメージを、人間がこれまで持っていた知性の自然な拡張として説明することを可能にする。

4 結び

本稿では、人工知能の発展の歴史を振り返り、そして人工知能が哲学的な知性理解に与えてきた影響について概観した。

人工知能は、まず内的な記号の体系的な操作によって人間が行っている計算と推論と同等とみなしうる過程を遂行する、論理的AIから始まった。しかし論理的AIでは解決が困難な課題、特にパターン認識、学習、物理的環境の中での適応的行動に直面して、人工知能学者たちはいくつもの異なるパラダイムを生み出した。その中でも代表的なものにはニューラルネットワーク、遺伝的アルゴリズム、そして状況的ロボットがある。近年、発展の著しいウェブ・コンピューティングは人間とコンピューターを協働させることによって、従来の人間とは独立のAIとも、人間の道具としてのAIとも異なる、新しい知能の実現方法を提示している。

人工知能と哲学の関係には次のようなものがある。まず論理的AIは計算主義と機能主義を推進する大きな要因になってきた。コネクショニズムはニューラルネットワークの成功に強く依拠している。状況的ロボットは身体化した知性という考え方と密接に結びついており、身体と環境との相互作用を重視している。ニューラル・ネットワークと状況的ロボットは論理的AIと異なり、知性が進化的に発展してきたという説明を容易にする、という利点もある。しかしながらこれらは数学的な計算や論理的な推論を行うことは苦手であり、人間がこれらの能力をどのように獲得してきたのかを説明するのは困難である。アンディ・クラークらが提唱するEMHはこのようなアポリアを解決するものであると同時に、ウェブ・コンピューティングのもたらす新しい知性のイメージを、人間の従来の知性の自然な拡張として捉えることを可能にするという点でも利点がある。

謝辞

本稿の執筆に当たって呉羽真氏から心の哲学に関して多くの点で有益な助言をいただいた。西尾香苗氏にはドーキンスについての筆者の誤解を正していただいた。また神崎宣次氏には本特集に執筆するよう声をかけていただき、特集のテーマに沿うよう本稿の方向性を指示していただいた。ここに記して謝意を表す。ただしもちろん本稿に書かれているすべてに関しては筆者が責任を負う。

参考文献

- R. Brooks. (1986) A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, Vol. 1, pp. 14-23, 1986.
- P. M. Churchland. (1995). *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*, Bradford Books.
- A. Clark. (2001) Reasons, robots and the extended mind (rationality for the new millenium). *Mind and Language*, Vol. 16, No. 2, pp. 121-145, 2001.
- A. Clark. (2003) *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press, New York, 2003.
- A. Clark and D. Chalmers. (1998) The extended mind. *Analysis*, Vol. 58, No. 1, pp. 7-19, 1998.
- H. Dreyfus and S. Dreyfus. (1986) *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, The Free Press, 1986.
- J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi & K. Plunkett. (1997). *Rethinking Innateness: A Connectionist Perspective on Development*, Bradford Books.
- D. Gillies. (1999) *Artificial Intelligence and Scientific Method*. Oxford University Press, New York, 1999.
- S. Harnad. (1990) The symbol grounding problem. *Physica D*, Vol. 42, pp. 335-346, 1990.
- 久木田水生。(2007) 人工知能の人工生命への接近。『PROSPECTUS』、Vol. 10、pp. 45-57、2007。
- 久木田水生。(2010) 脳のカオスのモデルと心の哲学。『実証段階におけるカオス研究の哲学的考察（研究課題番号16520017）平成16・17年度科学研究費補助金（基盤研究（C）（2））研究成果報告書』、pp. 93-100、2010。
- R. Menary. (2010) *The Extended Mind*. The MIT Press, 2010.
- H. Rheingold. (2006) 『新・思考のための道具——知性を拡張するためのテクノロジー』。パーソナルメディア、2006。日暮雅通訳。
- D. E. Rumelhart, P. Smolensky, J. L. McClelland and G. E. Hinton. (1986) Schemata and sequential thought process in PDP models, in D. E. Rumelhart, J. L. McClelland & the PDP Research Group (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models*. The MIT Press, pp. 7-57, 1986.
- J. R. Searle. (1980) Minds, brains and programs. *Behavioral and Brain Sciences*, Vol. 1, pp. 417-424, 1980.
- L. Steels. (2008) The symbol grounding problem has been solved. So what's next? In M. de Vega, editor, *Symbols and Embodiment: Debates on Meaning and Cognition*, chapter 12. Oxford University Press, Oxford, 2008.
- L. Steels and P. Vogt. (1997) Grounding adaptive language games in robotic agents. In C. Husbands and I. Harvey, editors, *Proceedings of the 4th European Conference on Artificial Life*. The MIT Press, 1997.
- R. Sun and F. Alexandre (eds.). (1997) *Connectionist-Symbolic Integration*. Lawrence Erlbaum Associates, 1997.
- M. Taddeo and L. Floridi. (2005) Solving the symbol grounding problem: A critical review of fifteen years of research. *Journal of Experimental and Theoretical Artificial Intelligence*, Vol. 17, No. 4, pp. 419-445, 2005.
- 津田一郎。(2002) 『ダイナミックな脳—カオスの解釈』、岩波書店、2002。
- A. M. Turing. (1950) Computing machinery and intelligence. In A. R. Anderson, editor, *Minds and Machines*, pp. 4-30. Prentice Hall, 1964.
- P. Vogt. (2007) Language evolution and robotics: issues on symbol grounding and language acquisition. In Loula, Gudwin,

and Queiroz, editors, *Artificial Cognition Systems*, pp. 176–209. Idea Group Publishing, Hershey, 2007.

M. Wheeler. (2008) Cognition in Context: Phenomenology, Situated Robotics and the Frame Problem, *International Journal of Philosophical Studies*, Vol. 16(3), 2008.