

総論 AI を多面的に考える

人工知能と共に生きる

市川 嘉裕

奈良工業高等専門学校



知能とはなにか？

知能とは、これこれこういうものである。といった具合に過不足なく説明できれば、それに向かって真っすぐに知能を作っていくことも可能であるかもしれない。しかし、知能は未だ何であるかはわかっていない。「人工知能」や「AI」という言葉がこれだけ世の中で聞かれるようになった現在、知能が何であるかわからないとはどういうことか？と思われるかもしれないが、それは人工知能がそもそも“知能とはなにか？”を探究することが最大の目的にある研究分野だからである。

本稿では、読者のみなさまが、人工知能とは結局どういうものであるか理解できるようになっていただけることを目指し、AIの仕組みや、AIの課題について出来るだけ簡単に紹介していきたい。読後には、みなさまが人工知能研究者と同じような視点から「知能」を考えられるようになっていただけたら幸いである。ただし、同様の内容は昨今のメディアの中にも溢れているため、客観的な事実については多方面の資料にも当たっていただきたいと思うとともに、本稿の差別化を図る意図で多分に私見を取り入れるため、すべては研究者視点の一意見として受け止めていただければ幸いである。

人工知能の誕生

知能に対する探究はもちろん、今に始まったわけではない。大昔から、それこそ人類が生まれたときから始まっていたかもしれない。それは知能という言葉すらなかったときからであろう。人工知能（Artificial Intelligence：AI）という言葉が生まれたのは1950年代で、コンピュータを使って知能を解明、実現しようという動きがその時代に始まったのである。

まずここで、いわゆる“AI”の前提について整理しよう。ほとんどのAIは、上記のことからもわかる通り、コンピュータ上で動くプログラムであるといえる。プログラムとは、コンピュータが処理すべき命令を記したものである。たいていのプログラムは人間の設計者によって人間にとって読みやすい言語（プログラミング言語）を用いて記述され、その後、翻訳機を通してコンピュータが読み取れる言語（機械語）に変換されて使われる。プログラムの処理を役割毎に大きく分けると、入力と出力、そしてその他の処理となる。例えば、電卓はプログラムで動いている小さなコンピュータであり、「1のボタン」を押せば、画面に1が現れ、続いて「+ボタン」、「2のボタン」、「=のボタン」と押すと、3が表示される。電卓において、ボタンを押すということが入力であり、入力（あるいは

入力との組み合わせ）に応じて画面の内容が変わることが出力である。入力から出力までの間には、電卓の中で様々な処理が行われていることになる。ただし、一般的なプログラムでは入力は必ずしも人間が行うわけではないし、人間には見えない処理だけをすることもある。

あなたは電卓の動作を見て“知的で賢い”と思うだろうか？電卓を賢くて知的と思うあなたにとって、電卓は立派な AI であるといってよい。しかし、世間一般ではそれが AI であるとは判断されにくい。何故だろうか？理由の一つは、「 $1 + 2 =$ 」を四つのバラバラな数字や記号として、それぞれに対応した特別なボタンを使って順に入力するという過程にあるだろう。知性の持ち主である人間はそのような入力をしないということである。言い換えると、電卓における入力が、人間の振る舞いに置き換えられにくいというところにある。仮に人間が「 $1 + 2 =$ 」を計算する場合、文字を一文字ずつ順番に頭にインプットすることはなく、文字全体を見て答えの 3 を導くことになる。その過程では、おそらくはまず「計算式であること」を認識し、そして「1 と 2 を足し算する」と意識した後、答えの 3 を導くであろう。導いた答えを声に出すのか、紙に書くのか、頭の中で考えて終わりのかは状況による。

初期の人工知能として、カウンセラープログラムの ELIZA が挙げられる。このプログラムはカウンセリングを目的とした対話をするために、短い文章（患者からの問いや訴え）を入力として、それに対する返答を文章によって出力する。文章という意味のある入力を受けて、文章で答えることは人間の振る舞いに置き換えることができ、AI であるといえるベースを備えている。これに加えて返答の内容が重要な鍵と

なる。この点については今日まで語られるものだけあり、当時のコンピュータのスペックで十分に動作できるよう最低限の処理で最大限にそれらしく返答するように工夫されたものだったのである。こういった文字や文章を入力とするプログラムをより人間らしくするために用いられる主な技術は自然言語処理と呼ばれており、現在でも研究が盛んな分野である。自然言語処理では、入力された文章を意味のある言葉の単位に分解したり、関連する言葉を導きだしたりすることができる。

VS 人間で進化する人工知能

ELIZA のように、巧みに人間らしく作られたプログラムが「人間的であるか」を検証する方法にチューリングテストがある。チューリングテストの一つの例は次の通りである。まず、テストの被験者となる人間がコンピュータを通じて隠れた他者と文字によってやりとりする。そのような状況で、被験者は二人の話し相手のうちどちらが生身の人間でどちらが対話プログラムであるかを当てる。多数の被験者がこのテストで間違えた回答をした場合、対話プログラムはチューリングテストに対して合格したこととなり、そのプログラムが確かに人間的であると結論付けられるというものである。このようなテストを通して、ELIZA も一定の条件においては、被験者に人間であることを信じ込ませることができたとされる。しかし、限られた条件において人間と区別のつかない動作をしているにすぎず、対話の内容に関して理解できているわけではないという批判に繋がることとなる。

一方でチューリングテストは、コン

コンピュータが人間の知的な振る舞いを模倣できるかを試す枠組みである。この対象となりうるコンピュータの動作は多岐に渡り、話し相手の文章に対して適切に返答することを始め、写真をみて何が写っているか答えたり、紙に書かれた文字をみて意味のある言葉として読み取ったりすることなど、様々なものが挙げられる。これを突き詰めると、人間にできることを、コンピュータが人間に劣らずに実行できるかどうかということが焦点となる。実際、以前はごく限られた人間の能力を実現するに留まっていたコンピュータも、最近では、囲碁のプロ棋士に勝利したり、翻訳の精度が飛躍的に向上したりしてきている。

今現在、人工知能技術の動きをリアルタイムに観測していると見えてくるのは、“VS 人間で進化する人工知能”という構図である。VS 人間で進化するとはつまり、新しい人工知能技術が生み出されると同時に、機械がまだ人間に劣っている点が指摘され、その改善のためにさらに新しい技術が生み出されるというサイクルである。例えば、機械が相手の感情を読み取ることを可能にしようという取り組みや、絵や物語を創作できるようにしようといった取り組みは近年の進化の例として想像しやすいだろう。新しい技術、あるいはその応用例に対して人間の能力を比較対象にできることは、評価のしやすさや研究目標のわかりやすさの面で優れている。筆者は残念ながら一昔前のリアルタイムを体験していないが、このようなサイクルはチューリングテストの提案によって、おおいに促進されたのではないかと想像できる。

「人工知能」の用法・用量

ここで、私たちの生活の中にも浸透してき始めている「人工知能」や「AI」という言葉遣いについて整理する。誤解を恐れずに言えば、前置きなしに人工知能と呼べるものはこの世にはまだないと筆者は考えている。そのため、筆者は日常で人工知能という言葉を目にするのはほとんどない。逆に、研究対象が人工知能分野と深く関わるため、人工知能という言葉を使って声をかけられることはよくあるという状況である。そんな中でも、人工知能や AI という言葉を使う場面がいくつかある。まず一つ目は、コンピュータゲーム等における NPC（人間が操作しないキャラクターや対戦相手）のことは AI と呼ぶ。これは、NPC がゲームという世界において、登場人物の一人としてそこに確かに存在していることや、人間のプレイヤーと同じ条件で対等に対戦する相手となっていることが理由である。あるいは、古くから通例として、それらの賢さとは関係なく AI と呼ばれていることも関係する。あるいは、これらのことを AI とは呼んでも人工知能とは呼ばないため、この用法の AI は既に独自の言葉になっているとも考えられる。二つ目は、教育（特にプログラミング）のための方便として人工知能という言葉を使う。例えば、プログラミング等の講義の導入として、「人工知能はプログラムです」や「簡単な人工知能を作ってみましょう」といった具合に用いる。ここで、人工知能の本質的な説明の一例を述べると、「現段階で仕組みはよくわからないけどなんだか知的ですごいもの」ということになる。ここでのポイントは、知的であると思う対象や、その仕組みが理解できるかどうかは個人の主観や能力によるというところである。プログラ

ミング初学者にとっては、最初の段階では仕組みがわからないことは当然であり、簡単なプログラムでも知的に感じるができる。幸いなことに、昨今の人工知能ブームの後押しもあり、このような方便は学習のモチベーションを保つことに効果的であると感じる。三つ目は、筆者がまだ仕組みを理解していない最先端の人工知能関連技術、特にその応用を指す場合である。四つ目に、本稿のこれまでのように、人工知能を要素技術に関してではなく総体として議論したり話題にしたりする場面である。

まとめると、人工知能という言葉は、総体を指す意味で使われることを除けば、ある知的な振る舞いをする技術に関して、聞き手が理解できていない場合、話し手が理解していない場合、またその両方を満たす場面で使われているといえる。言葉は、話し手と聞き手がどちらも理解をしている状態で使って初めてコミュニケーションを円滑する道具になりえる。そのため、ここで示したような用法をするには注意が必要で、前提をしっかりと整理した上で発信すべきである。こと人工知能については、言葉の乱用が一般消費者等の過度な期待や失望を生むことはすでに我々は経験済みなのである。

学習する機械

人工知能という言葉の乱用しないためにも、その仕組みを理解することが大切である。ここでは、今現在の主流である人工知能関連の技術について簡単に紹介していきたい。まず人工知能技術の中の大きなカテゴリとして「機械学習」というものがある。この技術の進展が現在のブームを牽引しているといつてよい。機械学習にもいくつも

の種類があり、少し広いところから順に説明していく。

機械学習 (Machine Learning) とはそもそも様々な技術の総称であり、技術ごとに様々な経緯で生まれてきている。人工知能における機械学習の役割は、入力に対して適切な出力を返すこと、あるいは人間らしい考え方をもってそれを実現すること、あるいは入力から新しい知識を見出すことなど多岐に渡る。通常のプログラムでは、入力に対して返すべき出力 (あるいはそれを導くための法則や方針) は全て人手で決めなければならないが、機械学習ではたくさんの情報を与えることによって、プログラムが目的に沿った適切な出力を自ら導くことができる。このようにプログラムが人間から直接的な指示をもらうのではなく、適切な動作を自ら獲得していく枠組みが学習と呼ばれている。他方で、人間の知的な活動の流れとして、情報を取り入れてから行動するまでの間に、状況理解、目標設定、計画といったステップを辿るという考え方がある。目に入ったものや聞いたものから、状況を理解することに対しては「認識」という技術が深く関わっている。目標設定や計画を立てるためには、自身がとろうとする行動が状況をよくするかどうか等を知るための「予測」や「推論」といった技術が深く関係する。いずれも機械学習が対象とする範疇であり、少々複雑なように思われるかもしれないが、それぞれの目的に沿った出力を、たくさんの情報をもって学習する点は共通している。ここで、学習のために利用できる情報 (以下、学習データと呼ぶ) の種類によって、機械学習は概ね「教師あり学習」「教師なし学習」「強化学習」の三つに分類される。

教師あり学習 (Supervised Learning) は、入力と出力のペアを学習データとして与え

ることによって、入力と出力の関係性を学習するものである。このように、入力に対する適切な出力を与えて学習させるため、教師ありと称される。教師あり学習では、例えば、入力されたものが何を表すか識別することや、入力された値からそれに関係する別の値を推測することが可能となる。具体的には、写真に写っているものが何であるかの認識（画像認識）や、人が何を喋ったのかの判断（音声認識）や、直近の気候情報等からの天候の予測等に利用される。

教師なし学習（Unsupervised Learning）は、学習データとして入力のみを与えることによって、データに潜む未知の知識を発見する学習である。教師あり学習と比較して、入力に対する適切な出力を与えないことから、教師なしと称される。教師なし学習は、明確な解答がない、あるいは解答の用意が困難な場合において、データ特有の新しい特徴やデータ間の関係性を見出すことができる。データマイニングといった技術と関連が深い。このような学習で得られた知識は、単体で用いるというより、他のプログラムと組み合わせて利用されることが多い。例えば、既存の文章を大量に読み込ませることで言葉を新たな体系で表現するといった学習は、それ単体では人間との比較はできないが、全く新しい文章を生み出すこと等に役立てることができる。

強化学習（Reinforcement Learning）は、学習の主体がある環境の中で自身の置かれた状態を知覚して、行動した結果、得られる報酬を頼りに適切な行動を獲得するという学習である。ここで、学習の主体とは、ロボットなどを思い浮かべてもらうとわかりやすく、強化学習の文脈においては一般的にエージェントと呼ばれる。エージェントからみて今の世界がどのような状況なの

かが入力であり、エージェントが実行する行為が出力に相当する。報酬はエージェントのとった行動に対する数値的なフィードバックであり、適切な行動にはプラス、適切でない行動にはマイナスといった具合に与えられる。あらゆる行動に対して適切な報酬を設定することはほとんどの場合は困難であるため、通常は明確に望ましい（または望ましくない）と考えられる状況になった際に、きっかけとなった行動に対して報酬を与えるように設定される。例えば、ロボットに適切な動作を獲得させることや、将棋等の盤上のゲームの手を改善することに利用される。具体的には、ロボットの動作では目的の作業を達成できたときに高い報酬を与えて途中の動きを学習させ、将棋等では対局に勝利すれば高い報酬、敗北すれば低い報酬を与えることで勝率の高い手を学習させる。

以上の説明中でいくつか挙げた例は理解を助けるためのものであり、必ずしも例に挙げたカテゴリの学習で解決されるわけではないことに注意していただきたい。例えば、教師あり学習でロボットの動作を学習させても問題はない。上記はあくまで学習データの違いによる分類である。それぞれの分類の中にはさらにいくつかの学習アルゴリズムが存在し、同じアルゴリズムであってもメカニズムの組み合わせやパラメータの設定により動作は異なり、最終的な学習結果が異なるものになる可能性がある。それぞれのアルゴリズムの内容は本紙面ではとても説明しきれないが、つまり「人工知能で問題を解決しよう」と同じくらい「機械学習で解決しよう」はまだ漠然とした提案であるということである。

機械学習による問題解決

さて、現実問題を機械学習で解決しようとするとき、問題を丸々入力しただけで答えを出してくれることはありえない。例えば、ある企業に敏腕な参謀がいたとして、彼は偉い人から悩みや起きている問題を聞いた後で、見事に問題を解決してくれるかもしれない。その際、問題解決に向けて、足りない情報の収集や、問題の整理を経て解決方法を提案するだろう。ここで最も大事な作業は問題の本質を見極めることである。一方で、現在の機械学習は優れた人間のように現実問題を一気に解決することは不可能である。前述した通り、機械学習が解決できるのは、たくさんの情報から得た学習結果を別の入力に適用して、予測や認識の結果を返すことだけである。たくさんの情報を処理して、別の有益な情報を見抜く能力は人間よりも機械学習にアドバンテージがある。しかし、機械学習による問題解決であっても通常の問題解決と同様、問題の本質を見極めることは人間が担う必要がある。人工知能はツール（道具）であるといった言葉を耳にされたことがあるかもしれないが、問題解決に使う場合はまさにその言葉の通りである。道具を有効利用するには、まずどんな問題を解消したいのか整理し、どういう方法で解消できるのか検討し、適切な道具を選び、正しい使い方で利用しなければならない。もちろん道具を使うことを目的としてはならないことは言うまでもない。ここで難しいのは、そもそもどのような道具があり、その道具がどのようなことに利用できるかわからない限り、それを使うという解決方法に至らないということである。また、仮に機械学習が動かせたとしても上手くいかない場合がある。その時、パラメータの設定が悪いのか、

学習データが悪いのか、実は上手くいっているが人間が納得できないだけなのか、それぞれの可能性を検証して修正していくことはさらに難しい。人工知能という道具は、他の IT 関連システム等と同様に目に見えづらく、どれも似たように見えるかもしれないが、実態は前述の通り多種多様であり、根幹の部分で異なることも少なくない。これらを使いこなすには、多くの学習コストが必要であり、その役割は専門家が担うことが適切である。付け加えると、昨今の人工知能技術の発展は目覚ましいため、専門家であっても常に学習することが欠かせない状況にある。

ここで、ある教師あり学習のアルゴリズムを用いて解決する手順の一例を示してみる。まず、機械学習を用いたシステムにしてもらいたいことを、入力と出力の形式で整理する。例えば、写真に写るキノコが安全かを判断してもらいたいなら、入力は画像で、出力は食べられるなら○、食べられないなら×ということに決める。次に、システムを学習させる。そのためには、既知のキノコの画像を大量に準備し、それぞれの画像のキノコが食べられるかどうか○か×でラベル付けする。用意する画像は、同種のキノコについて大量に、各々の種類についても同じように大量に必要となる。続いて、画像とラベルの組を学習データとして、学習モードのシステムに与えて学習させる。このとき、学習の進度を表すステータスを確認することで、データの量が十分であり学習が完了できたかどうかを判断する。最後に、学習が上手くいったかどうか確認するために、システムの判定精度を算出する。これには、学習データとは別に用意したテスト用データを、実際にシステムに入力して○か×を出力させて正解率等を求めることが必要である。これで、キノコ

の画像を入力することで食べられるかどうかを判定出来るシステムの核が概ね完成したことになる。このシステムによる判定精度はもちろん最後に確認した正解率に依存する。ただし、入力される画像が、学習データに用いた画像と似た条件の画像である場合に限ることに注意しなければならない。例えば、画像中に占めるキノコの大きさの割合や、明るさ、解像度、その他、様々な撮影条件により精度が大きく低下する可能性があるということである。またこの例では、写真に写るキノコが安全かどうかを判断することをスタートとしたが、次の段階で、画像のキノコが食べられるかどうか判断するという問題にすり替えられていることにも注意しなければならない。例えば、食べられるキノコの中にも、触れると皮膚がかぶれてしまうものや、呼吸器に影響があるようなものもあるかもしれないにもかかわらず、調理法や扱う際の注意点についての情報が全く欠けていることが問題である。ここで重要な点は、誰がどのような場面で使うシステムを構築しようとしているかを考慮することである。他にも、ラベル付けの作業は人間が手動で行うことが多いため、入力ミスや判断の相違によって不正確なデータの混入が起きてしまう可能性にも注意が必要である。

学習エージェントと共に生きる

機械学習の発展は目覚ましく、場合によっては人間を凌ぐようにもなってきた。しかし、人工知能研究の中ではそれら優れたプログラムは「弱い AI」と呼ばれている。端的に言えば、人間の知的な活動と比較して、解決できる問題が限定的であるということである。反対に、人間と遜色なく活動

できる知能をもつものが「強い AI」とされる。強い AI は人工知能研究の一つのゴールであると言える。しかし、そのゴールは明確であるとは言えない。弱い AI の適用範囲を広げていけばいずれ到達するという見方もできれば、全く異なるアプローチが必要になるという立場もある。

ここで、そもそも人間が知的であるということはどういうことかに立ち返る。例えば、AI は常識を備えることは困難であるが、常識テストをしてみれば、人間に勝ることも可能であろう。逆の見方をすれば、常識のない人間は知的ではないのかという問いも現れる。しかし、常識がなかろうが人間であるということは覆らないだろう。常識だけに留まらず、あらゆる機能的な側面のあるなしによって他者を排斥してはならないというのが人間共通にあるべき倫理観である。究極的に、知的であることが人間であるための必要条件にならないことは暴論になるだろうか。人工知能が倣うべき人間がそうであるならば、人工知能は何を目指せばいいのだろうか。筆者は、「道具」としての弱い AI と「人工知能という存在」としての強い AI に分けられると考える。前者は、これからより便利で強力なものが登場してくることが予想される。これについては、他の道具と同じように、利用や責任に関する取り決めが必要となる。後者は、実世界（VR 上でも SNS 上でも構わない）で一つの存在として他者から認められるようなものを指し、例えばペットのような存在になることが挙げられる。これは、目先のゴールとして、いわゆる動物愛護法と同様のものが適用される存在が生まれることを想像するとわかりやすい。

「学習エージェントと共に生きる」とは、最近、筆者がまとめた自身の研究室の世界

観である。これは、人工知能関連技術との関わり方について掲げた、主に学生や対外向けの方針である。これは一つに、道具としての AI であれ存在としての AI であれそれらと共に生きるというスタンスをとることで、必要以上に恐れず、必要に応じて利用する（協力する）という適切な距離感が保ちやくなるだろうという意図がある。他方で、そのようなスタンスを取りやすくする要素として、対象が主体性と学習する能力を備えていることが重要であるという考えがある。強化学習の文脈で主体的な存在のことをエージェントと呼ぶことは前で述べた。つまり学習エージェントとは、主体性と学習する能力を備えているもののことを端的に表した言葉として用いている。これは、知的かどうかはともかく、環境や経験によって変わる対象のほうが存在として認めやすいという考えをもとにしている。実際、社会に溶け込むような存在になるためにはそのような柔軟性が必要になるだろう。筆者は、対象（利用者や利用者集団）に寄り添い、生活や社会に溶け込むために、環境・ユーザの変化・多様性に対応できる知的なシステムを生み出すことを目指して研究を進めている。そして、読者のみなさまにお願いしたいことは、AI と共に生きるというスタンスで、身構えずに生活していただくことである。本稿を通じて、そのような関わり方を選択肢にもっていただけたなら幸いである。