

## 第 20 章

# 人工知能への招待 (道越秀吾)

### 1 はじめに

人工知能と聞くと皆さんは、何を思い浮かべるでしょうか。人によってそのイメージは大きく異なります。人工知能 (Artificial Intelligence) という言葉そのものは、1956 年に開催されたダートマス会議という研究会で計算機科学者のジョン・マッカーシーによって初めて使われました。その後、脈々と人工知能に関連した開発研究が続き、昨今の人工知能ブームに至ります。しかし、現在でも人工知能のはっきりとした定義は与えられていません。多くの方の人工知能への共通認識としては、「人間と同じような知的な処理能力を有する機械」ということができると思います。しかし、知性や知的な処理の定義が難しいため、人工知能の厳密な定義が難しいのです。

マスメディアで人工知能という言葉がよく使われています。しかし、専門家の間でさえ、人工知能は多義的であるため、専門家ではない一般の人々は様々なイメージをもって人工知能を捉えているでしょう。中には人工知能は万能であり、人間に代わってあらゆる問題を解決する魔法の杖であると思ってしまう人もいるかもしれません。しかし、実際の人工知能は、できることに限界があります。それでも、人工知能の社会での利用は着実に広がってきており、今ではビジネスの場でも使われている他、法的な問題など幅広い分野を巻き込んだものになってきています。

本稿は、主に技術面に焦点をあてて人工知能の全体像を明らかにすることが目的です。2 章で、人工知能とは何かを考えます。定義はないと上で書きましたが、いくつかの切り口によって、人工知能の意味を考えています。そして、最新の人工知能における中核的な技術である機械学習について説明します。3 章では、人工知能の活用事例を紹介します。人工知能で何ができるのか、それがどのように開発されているのかを説明します。4 章は人工知能の限界やそれを活用する際の注意点を述べます。5 章はまとめとなります。

## 2 人工知能とは何か

### 2.1 知能を感じさせる機械

ロボット掃除機は、周囲の状況を判断しながら部屋を自動で掃除してくれます。事前に人によってプログラムされたものではありますが、部屋を掃除するという目的のために、周囲の環境を元にロボット掃除機自身が行動を選択しており、人工知能であるといえるかもしれません。

ノートパソコンやスマートフォンで web サイトにアクセスするとたくさんの情報を得ることができます。通信販売サイトを閲覧したときに、開いた商品ページ中に推奨商品が出て思わずクリックしてしまった経験があるのではないのでしょうか。これは、閲覧状況の動向や類似ユーザーの購買履歴

データを基にして、その人が興味を持ちそうなアイテムを提示しています。これらは、かなり高度な知的処理といっても違和感がありません。

最近では、画像や音声をコンピュータが認識することも可能です。スマートフォンのカメラでも顔を自動で認識してフォーカスを合わせたり、笑顔になった瞬間に撮影を行う機能を持つ物もあります。カメラが顔の位置や笑顔になったかどうかを判断しているのですから、高度な知的処理といえるでしょう。また、スマートスピーカーというものを聞いたことがあるかもしれません。話しかけると、その内容に応じた答えを返してくれます。音声の認識からその返答の処理までコンピュータがやっているわけです。

単純なものから高度なものまで様々なものがありますが、人間の行なう作業を代替して代わりにやってくれる機械が、人工知能であるという見方ができるでしょう。

ここであげたものは、自動で動作したり、判断作業を機械が自動化するものです。自動的に動くものであれば、知的な処理をしていることにはなるのでしょうか。例えば、坂道を転がり落ちるボールを想像してみてください。自動的に動いていますが、知的だと感じられません。一方でロボット掃除機も同様に自動で動していますが、掃除をするという目的のもとで周囲の環境から取るべきこと行動を判断して自律的に行動しており、知的な処理とすることができます。

知的な処理とは、まず前提として目的があり、その目的のために周囲の環境に適合した行動を自律的に行うことと言えるでしょう。つまり、「周囲の状況（入力）によって行動（出力）を変えるエージェント（プログラム）」という捉え方ができるわけです (Russell and Norvig 2008)。状況に応じた行動する機械だけではなく、判断や解釈を自動的にする機械も同じ入力と出力の関係の枠組みで考えられます。例えば、図 1 にように猫の絵をみて「これは猫である」と言う機械は、入力が猫の絵であり、出力が「これは猫である」という言葉になります。

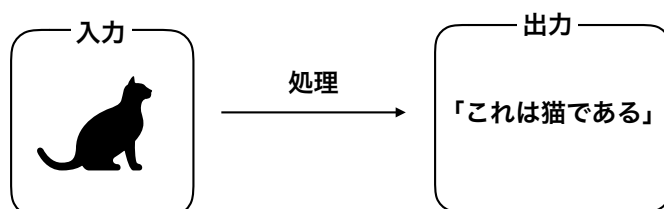


図 1 入力と出力の関係から考える人工知能。周囲の状況（入力）から行動（出力）を変える機械を人工知能と捉えることができると考えました。つまり、入力と出力の結びつきを与えているのです。たとえば、猫の写真をみて「これは猫である」と表示する機械では、入力が猫の写真で、出力が「これは猫である」という言葉になります。

### 課題 1

身の回りで人工知能が使われていると思われる機械の例をあげて、それがなぜ人工知能といえるのか理由を考えて下さい。

## 2.2 エキスパートシステムの発展と衰退

前の節では、人工知能とは、周囲の状況（入力）に応じて行動（出力）を与える機械として捉えました。では、この入力と出力の関係をどう獲得するのでしょうか。

人間の思考に当てはめて考えましょう。周囲の状況をどうやって判断するか。過去の経験や知識を

もとに判断しているのではないのでしょうか。知識を集積することで判断ができるようになるのです。

知識をコンピュータに獲得させることが人工知能の発展の鍵であると考えられました。1970 年代～1980 年代コンピュータの記憶容量の拡大や処理能力の向上から、コンピュータ上に有用な知識を取り込み、専門家が担っている現実の問題に役立てようという流れがありました。これをエキスパートシステムといいます。

初期のエキスパートシステムで特に有名なのが、1970 年代にスタンフォード大学で開発されたマイシン (MYCIN) というシステムです。血液中のバクテリアを診断し、感染した細菌を特定することにより、必要な抗生物質を推奨してくれるコンピュータプログラムでした。例えばこのプログラムを実行すると、「細菌の形は？」や「患者の痛みはひどいか？」などの質問が次々とプログラム上で行われ、それに答えていくと、菌の推定と必要な抗生物質を推奨されるものでした。マイシンの判断の正解率は 69% であり、感染症の専門家には及ばないものの、専門医ではない医師よりは良い正解率でした。分野が特化されているとはいえ、専門家の判断に匹敵する能力を持っており、確かに知的な処理を行っていると言えそうです。

では、このエキスパートシステムはどうやって構築されたのでしょうか。専門家が行うような判断をコンピュータにさせるには、専門家と同等の知識をコンピュータに獲得させる必要があります。その知識の元は、専門書などの文章や、専門家の経験などです。専門書から知識を抽出することはある程度機械的にできるかもしれませんが、しかし、専門家の持つ経験的知識は暗黙的なものであり、言語化されているものではありません。実際の専門家の高度な判断はそういった暗黙的知識に依存しているケースがあります。そういったものをデータ化することは容易ではありません。

このように分野を特化したとしても、知識のデータベースの構築や保守作業が困難であることがわかりました。実際、暗黙的知識のような定式化できない知識があるため、運用したエキスパートシステムは満足が得られないものも多く、その成功は限定的なものでした。

## 2.3 機械学習の時代

世界初のインターネットの web サイトは、1990 年 12 月に欧州原子核研究機構 CERN で作られました\*1。その後、web サイトは爆発的な発展を見せて、21 世紀に入る頃には世界中で広く使われるようになりました。インターネット上には大量のデータや文章が存在するようになり、インターネットの発展に連動して大量のデータが蓄積されていきました。このような大量のデータのことをビッグデータということもあります。コンピュータの処理能力や情報保存能力が大幅に向上したため、ビッグデータからコンピュータ自らが知識を獲得する機械学習が実用化されて広く使われるようになりました (図 2)。

機械学習は 3 つに分類されます。教師あり学習、教師なし学習、強化学習です。ここでは、理解しやすい教師あり学習を説明します。例えば、猫の写った写真と犬の写真があつて、それを人工知能に判定させることを考えましょう。機械学習では学習データが必ず必要です。教師あり学習では、犬の写真には「犬」、猫の写真には「猫」と事前に答えの書いた正解ラベルをつけたものを大量に用意しておきます。そして、人工知能に次々と写真と正解ラベルを提示していきます。つまり、教師が生徒に教えるように、「この写真は犬です」「この写真は猫です」と次々と教え込んでいくイメージです。このような画像による学習を繰り返して行くと、人工知能は、猫の画像と犬の画像のそれぞれの特徴を自ら獲得していきます。そして、十分な数を学習した後に、答えのラベルのついていない写真を見

\*1 CERN はスイスのジュネーブの近郊にある研究施設です。素粒子論という最先端物理学の研究のための大規模な研究所なのですが、このような情報技術の重要な成果もありました。

せると人工知能は学習した知識を使って猫か犬かを判定できるようになります。

前の節で説明したエキスパートシステムとの違いを比較してみましょう。エキスパートシステムでは、文献や専門家から知識を獲得してコンピュータにその知識を入力する手法でした。一方で、機械学習では知識をデータから自動的に獲得します。

犬、猫の写真の判別を自分でするときはどうやってやっているか、言葉で説明できるでしょうか。耳鼻口の形、胴体や足の長さ、様々なポイントで判断していると思いますが、言語化が難しいのではないのでしょうか。言語化ができないことをコンピュータに入力することができないので、犬と猫の判別プログラムを作ることは難しいことなのです。

一方で、機械学習では、この言語化しにくい特徴や性質の抽出を自動でやることができようになったため、人工知能の活躍できるフィールドが飛躍的に広がったわけです。大量のデータが手に入る問題であれば機械学習は非常に強力な手法となります。

機械学習には数多くの手法が存在します。中でも、ここ数年でとりわけ有名になったのがディープニューラルネットワーク (ディープラーニング) です (図 3)。ニューラルネットワークとは脳の神経網を模倣したコンピュータプログラムで、その原理は古くから知られており、1980 年代に機械学習として応用する手法<sup>\*2</sup>が発見されたことによって広く認知されるようになりました。しかし、複雑な問題に適用するためにネットワークの構造をさらに複雑にした、いわゆるディープニューラルネットワークは、概念そのものは以前からありましたが学習がうまくいかないという問題<sup>\*3</sup>があり、当時は使われていませんでした。それが 2012 年にコンピュータ科学者のジェフリー・ヒントンによってこの問題の解決法が発見されたため広く使われるようになったのです<sup>\*4</sup>。その後は、次々と改良するアイデアが提案されており、現在では世界中で研究されて、広く使われています。

ただ、昨今の人工知能ブームの中でディープラーニングが特に有名になったわけですが、ディープラーニングだけですべての問題を解決できる訳ではありません。ディープラーニングの登場以前から、機械学習には、サポートベクターマシン、ランダムフォレストなど様々な手法がありました。これらは、ディープ・ラーニングによって置き換えられたわけではなく、現在でも使われている有力なものです。それぞれの手法に向き不向きがあり、研究・開発の現場では使い分けられています。

### 3 人工知能の実際と応用

ここでは、人工知能で何ができるのか、何が行われているのか、実際の応用事例をいくつか取り上げて紹介します。

#### 3.1 人工知能の作り方

まず、具体的な応用例を説明する前に、人工知能がどのように作られるのか、説明したいと思います。

人工知能という特別なもののように感じるかもしれませんが、コンピュータ・プログラムの一種です。ですので、アプリケーションやソフトウェアと同様にプログラムを書くことによって作成されます。

プログラムは、コンピュータが理解できる言語、つまり、プログラミング言語を用いて記述します。

<sup>\*2</sup> 誤差逆伝播法

<sup>\*3</sup> 勾配消失問題

<sup>\*4</sup> 積層オートエンコーダと事前学習。現在では活性化関数というニューロン間の繋がりを記述の仕方を工夫する方法などが使われます。

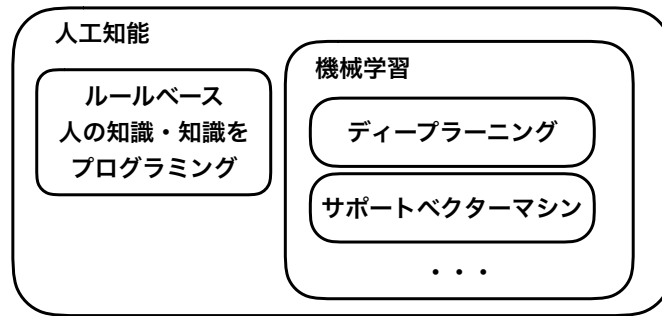


図 2 人工知能の概念図. 人工知能とは広い概念であり, 人間のような知的な処理をする機械のことですが, 明確な定義はありません. 最近, 言われる人工知能の応用例は機械学習という手法が用いられているものがほとんどです. 大量のデータからコンピュータが自ら学習しパターンや知識を獲得する手法のことです. 機械学習は人工知能の一分野とみなすことができます. さらに機械学習の中にも, 様々な手法があります. 最近, よく耳にするディープラーニングは機械学習の手法の 1 つです.

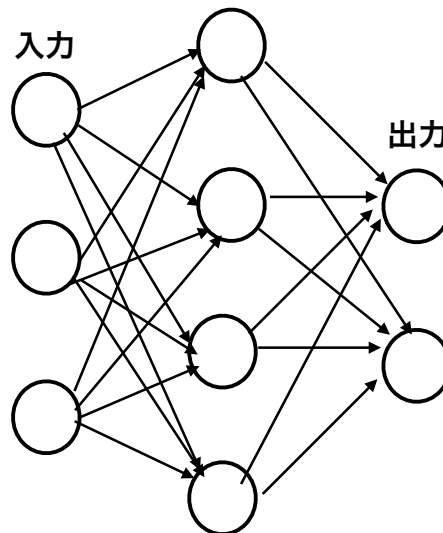


図 3 ニューラルネットワークの概念図. 人間の脳は, ニューロンとシナプスによる脳神経細胞網の結合・ネットワークによって働いているとされていますが, ニューラルネットワークはこのような人間の脳の働きに着想を得て考え出されたものです. 図の円で書かれた部分には数値データが入ります. 入力には学習させるデータを与えます. 図の矢印に従って, データが伝播・処理されていき, 出力を得ます. このネットワークの層構造を何層にも増やして, 複雑化したネットワークをディープニューラルネットワークと言います.

プログラム言語は沢山の種類があり, 用途に応じて使い分けられます. 例えば, 京都女子大学のプログラミングの授業では Ruby という言語を学びます. Ruby は Web プログラミングなどでよく使われる言語で, 分かりやすく学びやすいという特徴があります. 一方で, 人工知能のためのプログラミングでは Python というプログラミング言語がよく用いられます. Ruby と Python は考え方などはよく似ているので, Ruby をしっかり理解すれば Python もすぐに使えるようになります.

様々な人工知能の手法が提案されており, それらを完全に理解するには高度な数学や理論が必要となります. しかし, 実際に人工知能を作成する場合は, それらを全て理解しゼロからプログラミングする必要は必ずしもありません. 実際のプログラムの開発では, 多くの人が共通してよく使うプログ

ラムの部分は、予め共通化されて公開されています。ライブラリやパッケージ等と呼ばれる機能です。人工知能のプログラミングでは、詳細な人工知能の方法論の部分を記述する必要はなく、代わりに人工知能の専門家が開発したライブラリやパッケージを組み込むことで、人工知能に何をさせたいかという部分に集中することができます。

例えば、例に上げた犬や猫の判別させる人工知能であれば、Python の標準的な機械学習のパッケージを用いれば数十行程度のプログラムで書くことができます。さらに、それを数字の判別であったり他の用途のプログラムに書き換えることも容易にできます。1, 2 回生でプログラムをしっかり学んでおけば、十分に人工知能のプログラミングは可能です。

### 3.2 知性の象徴での人工知能の台頭

欧米ではチェスは人間の知能と知性の象徴とみなされることもあり、コンピュータの黎明期より人工知能で人間に迫る試みがなされてきました。しかし、チェスは複雑なゲームであり、コンピュータが人間に勝つのは容易ではありませんでした。1968 年に公開されたスタンリー・キューブリック監督の映画「2001 年宇宙の旅」では、想像上の最高の人工知能「HAL9000」が登場し、宇宙船の中で人とチェスをするシーンが印象的です。映画の中では、HAL9000 は人間に勝ちますが、当時はコンピュータが人間に勝つということは考えられないことだったのです。

転機が訪れたのは 1997 年でした。コンピュータの研究・製造の老舗である IBM の開発したスーパーコンピュータ「ディープ・ブルー」が当時の現役のチェスの世界チャンピオンであったガルリ・カスパロフを下しました<sup>\*5</sup>。手法としては、過去のチェスの対戦記録である棋譜を元に効果がある手筋を全て洗い出すものでした<sup>\*6</sup>。

チェスと似たような盤上のゲームに囲碁や将棋、オセロ (リバーシ) があります。これらのゲームとしての複雑さは、盤面に現れる局面の数で大まかに表されます。簡単なものから順に、オセロ (リバーシ)、チェス、将棋、囲碁とされており、将棋や囲碁はチェスよりも圧倒的に複雑であり、コンピュータが将棋や囲碁の名人に勝つことは当分はないだろうとされていました<sup>\*7</sup>。

将棋のプロの棋士達に 1996 年に「コンピュータがプロ棋士を負かす日は？来るとしたらいつ？」との質問がされました。多くの棋士が「棋士がコンピュータに負かされる日が来るなど、とても考えられない」と答えたほどでした<sup>\*8</sup>。しかし、将棋の人工知能プログラムの方も着実に力をつけて、2013 年にはプロの将棋棋士を初めて破りました。その後も進化を続け 2017 年には将棋の名人を下すほどまでになりました。

囲碁は将棋よりもさらに複雑であり、コンピュータが人間を超えるのが最も難しい分野であると考えられていました。しかし、Google Deep Mind 社が開発した Alpha Go が、2017 年に世界トップ棋

<sup>\*5</sup> なお、前年の 1996 年にもディープブルーとカスパロフの対局が行われて、そのときは 3 勝 1 敗 2 引き分けでカスパロフが勝利しました。1997 年の対局もカスパロフの 1 勝 2 敗 3 引き分けで、カスパロフが破れはしたものの僅差でした。この段階では、ディープ・ブルーがカスパロフを圧倒したというわけでもないようです。

<sup>\*6</sup> 中核の技術は探索のアルゴリズムであり、ミニマックス法とアルファ・ベータカットと呼ばれるもので、膨大な局面から先読みして良い局面を探す手法です。ここで問題になるのは良い局面とは何かです。良い局面とは何かを決めるのが評価関数というもので、昔は経験によるところが多かったのですが、最近は評価関数を決めるのに機械学習の手法が用いられています。

<sup>\*7</sup> ボードゲームの複雑さを表す指標として、そのボードゲームに現れる盤面のパターンが何通りあるかというものがあります。オセロは  $10^{28}$  と言われています。これは  $10000 \dots$  と 0 が 28 個並ぶ数です。簡単と言われるオセロでも途方も無い大きさです。チェスは  $10^{50}$ 、将棋は  $10^{71}$ 、囲碁は  $10^{160}$  と見積もられています。囲碁が文字通り桁違いに複雑なゲームであることがわかります。

<sup>\*8</sup> 将棋界でタイトルを独占したこともある第一人者の羽生善治氏はこの質問に 2015 年と答えたそうです。実際の結果と近いものになりました。

士である柯潔を下し、世界に衝撃をもたらしました。ニューラルネットワークや強化学習などの機械学習の手法が応用されました。

最難関の囲碁での人工知能の勝利は、人工知能の有用性や可能性を世の中に知らしめることになり、人工知能ブームを加速させるものとなりました。

## 4 人工知能の限界

人工知能の応用分野を見て、人工知能がいかに強力な手法であるかが理解できたかと思います。特定の分野では人間の能力を上回ることもあり得ます。しかし、決して人工知能が万能で全ての問題を解決するものであることを意味しません。最後にこの章では、人工知能の限界とそれに関わる問題について考えてみたいと思います。

### 4.1 人工知能が間違うとき

最新の機械学習の手法を使った人工知能でも間違うことがある、というみなさん驚くでしょうか。研究を重ねて正解率の高い人工知能を開発したとしても、実際の問題に使ってみると期待通りの性能が出ないこともありえます<sup>9</sup>。

前の節で示した、犬猫の判別の例で考えてみます。教師あり学習とは犬や猫の写真を正解ラベル付きで人工知能に画像と正解ラベルの関係を教え込ませるものでした。ここで、学習に用いた猫の画像の首に鈴がついていたとします。人工知能は首に鈴をつけたものが猫であるという判別方法を獲得してしまう可能性があります。その場合、例えば、首に鈴をつけてない猫を間違っただけで犬と判断したり、首に輪をつけた犬を逆に猫と判断してしまう可能性があります。これが過学習や過剰適合、またはオーバーフィッティングと呼ばれる現象です。

本来は、首の鈴は猫であるかどうかの判別にとっては本質的ではありません。首の鈴は本来は些細で無意味な情報、つまりノイズに近いものです。人間が判断する場合は、首に鈴をつけてない猫、首に鈴をつけた犬を間違うことがないでしょう。学習データに用いていない新しいデータで正しく判断できる能力のことを汎化能力といいます。過学習を避けながら汎化能力獲得させることが人工知能の開発における重要な課題といえます。

過学習をできるだけ抑える手法や過学習を起こしていないかどうか判定する方法は存在しますが、専門的であるためここでは詳細は省略します<sup>10</sup>。ただ、素晴らしい性能を誇示する人工知能があっても、過学習を起こす場合はその性能は期待できない可能性があるということを知っておいてもらいたいと思います。人工知能の判断だから信じよう、またテストデータで提示された性能が素晴らしいからこの人工知能の判断は信じようということではできません。

### 4.2 ブラックボックスの人工知能

機械学習のアルゴリズムによって学習させた結果は、ある種の数値データの集まりとして表現されます。例えば、ディープラーニングによる学習した結果は、各ニューロン間の結合の強さの数値デー

<sup>9</sup> 正解率 (accuracy) は、人工知能の予想の正しさを表す尺度です。似た概念の言葉に適合率 (precision)、再現率 (recall) などがあります。これらは、人工知能の何を評価したいかによって使い分けられています。

<sup>10</sup> ラッソ回帰、リッジ回帰やニューラルネットワークのドロップアウト法など精度を抑制しながら学習モデルを簡素化して過学習を防ぐ方法が開発されています。また過学習の確認は、ホールドアウト法やクロスバリデーションなど、学習データと評価データを分離する方法があります。このあたりの詳細は専門の講義やゼミで扱います。

タの集合です。人工知能の問題点の1つとして、この学習結果が何を意味するのか、どう判断しているのか人間には理解できず、いわゆる「ブラックボックス」になってしまうという問題です。ブラックボックスとは、中身の動作の原理がわからない機械の総称です。

犬猫の判別をどうやっているのか、人間でも明確に言葉にするのが難しいと説明しました。ディープラーニングを使うと確かに、この言葉にできない何かをうまく学び取って、判別できるようになるのですが、どう判別しているのかが、人間にはわからないのです。猫や犬の画像を人工知能に与えると、人間にはなかなか理解し難いプロセスを経て、結果が出てくるのです。まさにブラックボックスです。

犬猫の判別ならブラックボックスでも構わないかもしれませんが、また、囲碁や将棋のような盤上のゲームなどでは、人工知能が、なぜその手を打ったのか、その理由は分からなくても、対戦相手に勝つことさえできれば、その結果をもって価値を評価することが可能であり、意味があることです。しかし、実際の社会実装となると話は変わってきます。ブラックボックスでは結果を本当に信頼してよいのか、という疑念が残ったままになるのです。なぜその判断をしたのか判断・根拠の説明が求められることがあり、ブラックボックスのままでは人工知能にクリティカルな業務を任せることはできません。

現状では、判断・根拠を説明できないブラックボックスの人工知能が多いのですが、「説明可能なAI」に向けた研究が始まっています<sup>\*11</sup>。

### 4.3 フレーム問題

フレーム問題は、1969年にジョン・マッカーシーとパトリック・ヘイズによって提唱された人工知能における未解決問題であり、未だ本質的な解決はなされておらず、人工知能研究の最大の難問と言われることもあります。ここでは、哲学者ダニエル・デネットによって提示された洞窟の中の予備バッテリーを取りに行くロボットの例でこの問題を考えてみます。

バッテリーで動く人工知能を搭載したロボットを想像してください。この人工知能ロボット1号機はR1という名前です。洞窟の中には、ロボットR1が押せる程度の大きさの台車があり、台車にバッテリーが載せられています。同時に、この台車には時限爆弾が仕掛けられています。「洞窟から予備バッテリーを取り出してくること」とロボットR1に指示を与えます。ロボットR1は人工知能を搭載しているため、この指示を聞いて、洞窟の中に入り、予備バッテリーの乗った台車を押して持って帰ることに成功しました。ところが、洞窟から出たところで、時限爆弾が爆発してしまいました。ロボットR1は、バッテリーを洞窟から運び出すという目的は理解し行動しました。しかし、その行動を行ったことによって生じる副次的な結果、つまり台車だけを押しと爆弾も一緒に運んでしまうということを考慮していなかったため、爆発に巻き込まれたのです。

そこで、2号機のロボットR1-D1を開発しました。今度は「自分が行った行動の結果、副次的に発生することを全て考慮する」という機能を備えています。先ほどと同じ状況で、洞窟の中に台車に載せられた予備バッテリーがあり、時限爆弾も一緒に載せられています。ロボットR1-D1に「洞窟から予備バッテリーを取り出してくること」と指示を与えました。しかし、ロボットR1-D1は洞窟の中に入っていき、時限爆弾と予備バッテリーの乗った台車にたどり着きました。ところが、台車の前で動かなくなり、時限爆弾が作動してしまい、爆発に巻き込まれてしまいました。

これは、ロボットR1-D1は、自分の行動の副次的効果を全て考えるように作られているからです。

\*11 古典的な機械学習の手法としては、決定木・回帰木が学習結果の解釈がしやすい手法として知られていますが、汎化性能を保ったまま高精度化しにくいです。



「台車を動かすと時限爆弾が爆発してしまわないか」「バッテリーを動かす前に爆弾を動かさなくてはいけないか」「爆弾を動かすことで、洞窟の天井がおきないか」「爆弾に近づくと洞窟の壁の色は変わらないか」など関連する副次的に発生する可能性がある事態を全て考えてしまい、無限に思考を続けてしまったのです。確かに「洞窟の壁の色が変わるかどうか」は副次的効果の可能性の1つですが、本来は考慮する必要のないことです。本質ではないことまで含めてあらゆる事態を考えようとしたため、台車の前で停止してしまいました。

そこで、さらに改良を施します。3号機であるロボット R2-D1 は、「自分の行動の副次的な効果は考慮するが、目的と無関係なことは考慮しない」という機能を追加しました。今度は、洞窟に入る前の入り口で停止しました。目的と無関係なことを洗い出すために、無限に思考を続けてしまったのです。

似たようなことは人間の場合でも起きます。深く考えずに行動し失敗をしてしまうのは、1号機 R1 と同じです。逆に取りるべき行動を考えすぎてしまい、何もできなくなってしまうのは、2号機 R1-D1 や3号機 R2-D1 と同じです。しかし、人間はあらゆることを考えすぎてフリーズすることはありません。人間の思考にはフレームがあるからです。

フレームとは枠のことです。世界は広く常に様々なことが発生しています。全てを考慮すると無限の時間がかかるのは人間も同じです。人間の場合は、考えるべきことにフレーム（枠）を与えて、その枠の中で思考することができます。人工知能においても、フレームを与えればよいのですが、適切なフレームをどう与えるかという問題があり、結局同様の問題が発生します。人間のように判断できる「ロボット R2-D2」は難しいのです。

#### 4.4 強い AI と弱い AI

映画やテレビ番組などで意識と心を持った人工知能のロボットが登場するのを見たことがあるでしょう。例えば、未来から来た猫型ロボットのドラえもんは、怒りや悲しみを理解し、意識や心があるように感じられます。ここまでで紹介した人工知能とは、大きくイメージが異なります。

ここまでで紹介した人工知能は、確かに学習と推論をしています。特定の問題では、人を上回る性能を上げる事例も数多く報告されています。しかし、そこに意識や精神が宿っているように感じられません。

この議論に関連して、哲学者ジョン・サールによる強い AI と弱い AI という概念があります。弱い AI とは、コンピュータは人間のような心を持たない、便利なツールであるという立場です。確かに事前のプログラミングや機械学習の結果、高度な判断能力を有しています。しかし、そこには精神はないという考え方です。一方で、強い AI は、人間の心や意識は脳の情報処理であるので、適切にプログラムされたコンピュータは人間と同じように心を持つという立場です。

現在、作られている人工知能は全て弱い AI です。現在のところ強い AI が実現できる目処は立っていません。ジョン・サール自身も実現不可能だと主張しました。

ある機械が人工知能であるかどうかを判断するためのテストとして有名なものにチューリングテストというものがあります。イギリスの著名な数学者・コンピュータ科学者であるアラン・チューリングによって考案されました。チューリングテストとは、人間の判定者と機械が会話することによって行います。判定者は機械とは別の場所にて、キーボードやディスプレイを通して文字のみで交信・会話をします。交信する相手は機械であるか人間であるか分からないようになっています。会話をし、相手がコンピュータと見抜けなければ、そのコンピュータには知能があるという考え方です。

これに対応して、ジョン・サールの中国語の部屋という思考実験があります。ある小部屋を考えま

す。中にアルファベットしか理解できない人を閉じ込めます。この部屋には外部とやりとりするための小窓がついています。この小窓を通して部屋の中に中国語で書かれたメッセージを紙に書いて入れます。小部屋の中には、中国語の質問に答えるための完璧なマニュアルが置いてあります。中の人は、中国語を全く理解しませんが、このマニュアルに従って、答えを書いて部屋の小窓から外に返信することができます。この作業を繰り返します。部屋の外から見ると質問に対して自然な答えが返ってくるので、中には中国語を理解する人がいると思うでしょう。しかし実際は、中の人は中国語の内容を理解せずにマニュアル通りに作業を繰り返しているに過ぎないのです。

これは、会話ができるコンピュータのアナロジーとなっています。チューリングテストに合格する機械が出たとしても、コンピュータは記号操作を行っているだけで、心にとって本質的なものを欠いているとして、強いAIを否定しました。

強いAIの問題はコンピュータ科学以外の分野の研究者も議論しています。数学者・物理学者のロジャー・ペンローズは、意識は脳の中の微小管における量子効果に起因すると議論しました。量子効果とは、原子や分子などの小さなスケールで生じる現象であり、既存のコンピュータによる再現は容易ではありません。これも、コンピュータでは強いAIは再現できないという立場になります。

将来の技術革新の結果、強いAIが将来誕生する可能性はあるかもしれません。しかし、現状では強いAIは存在しません。人間の知性を完全に代替することができる万能機械は存在しないのです。

## 課題2

映画や小説などのフィクションに登場する強いAIの例をあげて、それがなぜ強いAIと言えるのか説明してください。

## 5 終わりに

以上、人工知能の概略、応用分野について見てきました。人工知能という言葉に触り回されずに、その限界を見極めた上で、可能性や応用を考えてほしいと思います。

本稿は人工知能の概念とテクノロジーとしての側面に焦点をあてて論じてきたため、人工知能にまつわる全ての話題に触れることはできませんでした。しかし、人工知能という分野は社会全体を巻き込んだ大きな流れになってきています。

人工知能は大学や企業の研究室の中だけの実験上のもではありません。社会に広く浸透し、どう使いこなすかという段階に入っています。IT企業だけに限らず、様々な業界の企業が人工知能の導入をすでに行っており、成果をあげています。ビジネスや経済の場に進んでいく場合でも人工知能は是非知っておいてもらいたい知識となります。

さらに人工知能にまつわる法律的な問題もあります。人工知能の応用の一つとして車の自動運転があります。自動運転というものは従来の法律では想定されていなかったため、これをどう扱っていくのか、今後の課題となっています。また、機械学習には大量の学習データが必要ですが、学習データの収集には著作権の問題も関係します。さらに人工知能が犯した間違いの責任を誰がとるべきか、法的・倫理的問題にもなっています。

人工知能は、現代社会における様々な事象を巻き込んだ大きな分野となっています。1つの事象を様々な観点の切り口から考察できる現代社会学部は、人工知能を学び・考えるには最適の場ではないかと思います。

最後により深く学びたい人は、「人工知能は人間を超えるか：ディープラーニングの先にあるもの」

(松尾 2015) や, 「深層学習教科書 ディープラーニング G 検定 (ジェネラリスト) 公式テキスト」(浅川他 2018), 「AI 白書 2020」を参照してください.

### 参考文献・推奨図書

- [1] S. J. Russell, P. Norvig (著), 古川康一 (訳), 「エージェントアプローチ人工知能」, 共立出版, 2008
- [2] 松尾豊, 「人工知能は人間を超えるか: ディープラーニングの先にあるもの」 KADOKAWA, 2015
- [3] 浅川伸一, 江間有沙, 工藤郁子, 巢籠悠輔, 瀬谷啓介, 松井孝之, 松尾豊, 「深層学習教科書 ディープラーニング G 検定 (ジェネラリスト) 公式テキスト」, 翔泳社, 2018
- [4] 独立行政法人情報処理推進機構 AI 白書編集委員会 「AI 白書 2020」, KADOKAWA, 2020