

クララは電気羊の夢を見るか？

廣安知之〔ひろやす・ともゆき〕
同志社大学生命医科学部教授

第1章

人工知能（Artificial Intelligence＝A I）をキリスト教主義的視点で捉えてみる

生命医科学部の廣安知之です。1998年から工学部の知識工学科にお世話になり、2008年に生命医科学部が立ち上がったタイミングで転籍し、現在、医情報学科に属しています。また、ヒューマンインフォマティクス研究室を日和悟准教授とともに主催しています。同志社大学に赴任してから、A Iと情報処理の技術を使ったアルゴリズムの開発にかかわる研究を行っています。

人工知能（Artificial Intelligence＝A I）は、毎日、その単語を聞かない日はないほどよく使われる技術単語となりました。A Iの定義は、Wikipediaによると「人工知能は、人工的にコンピュータ上などで人間と同様の知能を実現させようという試み、或いはそのための一連の基礎技術を指す。」とありますが、研究者によってその定義が異なっている状況にある余白のある技術です（文末〔参考文献〕1）。

本日は、Doshisha Spirit Weekの講演にあたり、A Iとキリスト教という2つのキーワードに関する話題提供をしたいと思います。

A Iとキリスト教についての話題提供は、無謀に見えます。今でも私も無謀だと思っています。私の専門である人工知能とキリスト教の関係について考えるきっかけになったのは、2016年度から始まった「同志社大学良心の実証的・実践的研究プロジェクト」への参加でした。プロジェクトリーダーの神学部の小原克博先生から、このプロジェクトを推進するにあたり、A Iと良心について議論してほしいというリクエストがあったのです。A Iと良心というテーマも無謀なテーマでした。途中は端折りますが、小原先生を始めとして多くの先生方のご協力を得て、A Iと良心というテーマで考えることができ、小原先生がまとめられた2冊の図書において結果を公表することができました（文末〔参考

文献] 2、3)。そしてさらに、小原先生よりA Iとキリスト教について講演してもらえないかというリクエストがあったのです。さすがに最初はお断りしたのですが、議論を続ける中で少しずつ重なる部分も見えてきました。そして、最後に小原先生が言われた一言で少し光明が見えました。「宗教学の一面は、人間とは何なのか?にあります。」という一言です。A Iは先に書いたとおり、「人工的に人間と同様の知能を実現させようという試み」です。すなわち、A Iを考えるという作業は「人間とは、人間の知能とはということを考える」ことになります。NHKでA Iの紹介番組がありましたがそのタイトルも「人間ってナンだ?」でした(文末〔参考文献〕4)。A Iも宗教学も「人間とは何なのか?」という共通点があることを教えてもらったのです。

同志社大学はキリスト教主義の大学です。本講演では、A Iをキリスト教主義的視点で捉えてみることを試みます。キリスト教主義的なやり方も様々あるとは思いますが、キリスト教・聖書から問いを立てA Iを検討することがその1つではないかと考えています。私は、キリスト教や宗教学については素人であり、本講演では、専門分野については多くを語らない内容になり、また、何か結論を導こうとするものではありません。それでも、私の話をきっかけにして、みなさんとともに考えることができればと思っております。これがConscienceであり、同志社の良心だと考えています。

第2章

第3次ブームを迎えた人工知能 (Artificial Intelligence = A I)

「人工知能 (Artificial Intelligence = A I)」という名前は1956年にダートマス会議でジョン・マッカーシーにより命名されたとされています。A Iは、これまでにブームと衰退を繰り返し2010年過ぎから第3次のブームを迎えています。今では多くの場面でA Iという単語が見られ、さまざまな分野で成果をあげています。

人の知的な作業の一つに探索があります。その探索能力を測るのにA Iが知能ゲームでの程度、人に勝つことができるのかをみる方法があります。A Iが初めてチェスの世界チャンピオンに勝利したのは、1997年でした。I B Mが開発したDeep Blueとよばれるスーパーコンピュータを利用しました。I B Mではこのシステムをさらに質問応答システムに拡張し、米国の人気クイズ番組「Jeopardy!」にチャレンジし、2011年に人に対して総合でWatsonが勝利しました。2017年になると、コンピュータが現役の将棋の名人位保有者に勝利し、将棋の世界でもA Iの実力が認められ始めました。囲碁は将棋より遙かに探索が困難とされていますが、2016年に、Deep Mind社が構築したAlphaGoが世界トップ棋士の李世 (イセドル) 氏に勝利し世界を驚かせました。このように年々、A Iは探索の分野で実力をつけつつあります。

画像探索、生成、識別、分類の分野でも非常に多くの成果をあげています。そこで中心になっているA Iの技術が機械学習と呼ばれる技術です。特にDeep Learning (深層学習) がコアな技術であり、革新的な成果をあげました。現在では、自動運転分野、農業分野、医療分野などでA Iは広く活用されています。

第3章

汎用A I クララ

このように様々な分野で成果をあげているA Iですが、人のように複数のことを同時にうまく処理するわけではありません。人はルールを学ぶことによりチェスも、将棋も、囲碁も遊ぶことができ、実力を上げていくことができます。また、人は音楽を楽しむこともできれば絵画を嗜むこともできます。食事やお酒を味わうこともできます。

一方、A Iは人工物であり、当初設計された目的にあった、特化した機能しか持ち合わせません。これを「特化型A I」と呼ぶことがあります。将棋や囲碁で名人に勝利したA Iは、それぞれの分野に特化したA Iというわけです。それに対して、人のように多くのことができる機能を有するA Iは「汎用型A I」と分類されます。特化型A Iは数多く作られてきましたが、汎用型A Iの出現はもう少し先になりそうです。また、「特化型A I、汎用型A I」という分類は、場合によっては「弱いA I、強いA I」ともあらわされます（文末〔参考文献〕5）。

まだ実際には、世の中に出現していない「汎用型A I」ですが、昔から多くの人々の関心を集めていました。映画「ターミネーター」では、汎用型A Iが未来社会に現れ人類絶滅の危機に陥れるというストーリーでした。『アンドロイドは電気羊の夢を見るか?』は、人と汎用型A Iとも言えるアンドロイドとの関係を描いた小説で、映画にもなりました（文末〔参考文献〕6）。最近では、ノーベル賞を受賞した、カズオ・イシグロが、汎用型A Iが主人公となる小説、『クララとお日さま』を公開しています（文末〔参考文献〕7）。A Iがお日さまに対して神ともいえる感覚を覚えて行動することや、A Iの終末、A Iの個性など非常におもしろい視点をいくつも披露している小説です。この小説の主人公が「クララ」でした。

本講演では、「汎用型A I」がどういう存在であるべきかの議論をしたいと思っています。最新型の汎用型A Iの「クララ」は、数十年前に問われたとおり「電気羊の夢を見るでしょうか?」テレビアニメの「アルプスの少女ハイジ」にでてくる「クララ」ではないですよ。

第4章

A I の倫理問題とキリスト教主義的視点で A I を考えるということ

汎用型 A I が出現すると、いくつかの問題が起こると杞憂されています。汎用型 A I を取り扱った多くの映画では、それらが反乱したり人に害を及ぼしたりします。その他、A I の知性のベースになっている知識にバイアスがあった場合、A I の判断や行動も問題になることが予想されます。性差別、人種差別などが生じ、就職の際に問題が生じる危険性です。また、問題が起こった場合に誰が責任をとるのかという課題もあります。技術的には自動運転は開発目前です。これは特化型の A I を使った例ですが、責任問題や法律の対応などが未完成のままです。汎用型 A I が出現すると人が従事していた仕事が奪われる危険性も指摘されています。このように多くの課題が汎用型 A I にはあるのです。

この問題に対して各国や関連する学会は努力し、対応しようとしています。内閣府は「人間中心の A I 社会原則検討会議」を立ち上げ、議論しています。米大統領は「米国の A I におけるリーダーシップの維持に関する大統領令」に署名しました。IEEE（米国電気電子学会）は、「倫理的に調整された設計第 1 エディション」を公表し、日本の人工知能学会も倫理指針を出しています。

しかしながら、科学は、倫理判断に対して言及ができるのでしょうか。その判断するのは人そのものであり、その判断を行うという行為こそがまさに人間らしい行動であり、判断を科学で規定することは難しそうです。ユヴァル・ノア・ハラリ氏はその著書『ホモ・デウス』で人間の法や規範や価値観に超人的な正当性を与える必要があると述べています（文末（参考文献）8）。科学は倫理判断に言及はできないが、事実には科学的手法により言及できる。そのため、宗教と科学が組み合わさることで、人間の行動の指針を与えることができる可能性があるとも述べています。このように「汎用型 A I」に対する倫理判断を行うためには、宗教と科学を組み合わせる必要があります。

同志社大学は、ミッション・スクールではありません。新島は、国を背負うべき日本の

若者を教育するために、キリスト教主義を徳育の基本としました。それは、新島が見た米国が驚くべき自由の国であり、その根幹となすのがキリスト教であると見抜いたからでありましょう。「汎用型AI」に対する倫理判断を行うためには、宗教と科学を組み合わせる必要があるとするならば、「キリスト教」と「AI」を組み合わせることは同志社大学だからこそできるのではないのでしょうか。この講演では、聖書から問いをたて、「汎用型AI」に関してどのようなことが考えられるかをキリスト教主義的検討手法にて行うことを提案します。具体的には、聖書にかかっている事項から次の3問を立て検討を試みます。

- (1) 世界はどのように作られたのか？
- (2) 創造主はどうしてこんな人間を造ったのか？
- (3) バベルの塔の物語からの学びは？

第5章

問1 世界はどのように作られたのか？

創世記には、創造主なる神が6日間かけて世界を創造されたとあります。最初に光を造られ、次に大地や生き物を造り、最後に人を造られました。

この話にはいくつか「汎用型AI」に関連した重要なポイントがあると思います。まず、創造主が世界を造られ、創造主は満足されたということです。造られた後、「見て、良しとされた」とあります。

次に、創造主はあらゆるものを創り出すことができ、そして、神に似せて人を創造されたというポイントです。また、その創り出されたものは無から造られ、材料を含めて一切合切を創造主が造られたということもポイントです。これらのポイントについては次の章で別の問いを立ててもう少し議論します。

創造主はあらゆるものを創り出すことができるにもかかわらず全てのものを自身で創り出されませんでした。1週間で作られなかったものは、人によって造ることにされたのです。システム工学では人工物・システムのできることを「機能」と呼びます。また、処理する手順はアルゴリズムと呼ばれます。コンピュータではアルゴリズムがプログラムによって実装され実行されます。人にインプットされた機能、アルゴリズムは何でしょうか？

創世記からわかることは、人も創造主と同じようにものを創り出す機能があり、それを実行するアルゴリズムが内在されました。人が作るものは「人工物」と呼ばれ、AIも人工物の一つです。次の問いでは人に内在されているアルゴリズムについて考えてみたいと思います。

第6章

問2 創造主はどうしてこんな人間を造ったのか？

創造主はどうしてこんな人間を造ったのかが次の問いです。

創造主は神に似せて人を造りました。しかし、人は創造主を裏切ってしまう。旧約聖書「創世記」にある失樂園の挿話では、蛇にそそのかされたアダムとエバが、神の禁を破って「禁断の果実」を食べ、エデンの園を追放されてしまいます。創造主に注意されたにもかかわらず、創造主のようになりたいという気持ちがあり、その欲望が勝ってしまったわけです。そして創造主の怒りをかかってしまいます。

創造主はなんでもできるはずですが、ところが、人は創造主を裏切ってしまう。そうになると、創造主はなぜ、創造主を裏切る人を作ったのかという大きな疑問が生じます。天地創造されたときには、創造主はすべてに満足されていたのです。創造主を裏切らない人を造ったほうがよかったのではないのでしょうか。創造主を裏切らない人を造っておけば、創造主は今でもおだやかに過ごせたのではないのでしょうか。

エデンの園の逸話からわかることは、もともと人には「創造主のようになりたいと感じる機能があった」ということです。しかしながら、人はこの機能を利用してはいけなかったのです。これは、人がある行為をするかしないかを決める際に、あらかじめ必然的に決まっているわけではなく、その人の意志で決定するという自由意志の問題が存在するということでしょう。そして、人には自由意志により問題を解決する機能とアルゴリズムが内在されているように思えます。すなわち、人は良きことも悪きこともできるような存在であり、創造主を裏切ることができる自由が創造主から与えられているのだと考えることもできるでしょう。創造主は、創造主に似せて人を造りました。そうすると、AIも自由意志を持つように造るべきではないのでしょうか。たとえAIが人を裏切るとしてもです。それは何故でしょうか。

現在のAIは第3次ブームの最中にあります。言い換えれば、これまで2回、AIの実

現を期待して挫折した歴史があります。挫折した理由はいくつかあります。その一つに、知識をすべて洗い出すのが非常に困難であるという問題があります。知識をすべて書き出して、機械に覚えさせ対処することができれば、その機械は人と同等の作業が可能であることは、論理的にはあり得ます。しかしながら、実際に知識の洗い出しをやってみるとすぐに破綻することがわかります。人が処理している知識はあまりに膨大で、洗い出す方法もよくわかりませんし、知識をどのようにうまく使うのか、実時間でデータにアクセスできるのか、莫大なデータをどのように保存するかなどなどの困難が伴います。

汎用型AIの実現の際に直面する課題の一つに「フレーム問題」があります。実世界ではさまざまな事象が起こりえますが、何かを実行する際には、当面の問題と関係するかもしれないかをふるい分け、関係しない問題を無視して関係する問題だけを対応しなければ無限の時間がかかってしまいます。すなわち、フレームを作って、そのフレーム内で考えることが必要なのです。しかしながら、当面の問題がどんなフレームに、どのフレーム内に対応しているのかを判断する必要があり、結局は無限の時間がかかってしまいます。人はこれをうまくやっているように見えます。これがフレーム問題です。この問題からもわかるように全ての知識を洗い出し、行動をルール化することは非常に困難です。

また、仮にある時点での知識の全てを書き出せたとしてもそれでは不十分です。創造主の造られた世界は刻一刻と変化します。親が子を作る仕組みを有することが生命の本質であるならば、世界は必ず時間が変わると変化します。システム工学では、問題に時間要素のパラメータを有する場合、その問題を動的な問題と呼びます。問題が動的であり、無限時間を対象にせねばならない場合、そこに存在する問題の数は無限になり、全て洗い出すことは不可能です。同志社大学の赤ちゃん学研究センターは、人が生まれる前から、生まれた直後、その発達を一貫して研究するというユニークな研究センターです。そのセンターの研究においても人の成長には外部からの刺激が非常に重要であることが明らかとなっています。

これに対応する方法、アルゴリズムがその時その時にうまく対応できるようにする方法ではないでしょうか。これをできるようにしている根源が自由意志なのではないでしょうか。汎用型AIを実現するためには、かつて創造主がなされたように自由意志を持つAIの構築が必要なのかもしれません。自由意志を持たせるためには想定した領域外の探索（外挿探索）をさせる必要がありますが、今のところ人が行っているような精度の高い外挿探索の実現は困難です。

ただし、人は、自由意志を持つ人工物を創る機能は有していますが、それを使うことが許されているのでしょうか。次の問いで考えてみたいと思います。

第7章

問3 バベルの塔の物語からの学びは？

最後の問いは、バベルの塔の物語です。

人は石の代わりにレンガをつくり、漆喰の代わりにアスファルトを手に入れます。手にいれた新しい素材を使って、人は天まで届く塔を建てて、有名になろうとし、その結果、神の怒りをかったという話だと思えます。人の行動のどこが悪かったのでしょうか。

天地創造の際に、創造主は材料を含めて一切合切無から造られ、造られなかったものについては人によって造られることを期待されます。そのため、創造主は人に材料から作り出す能力（機能）を人に与えていることとなります。その能力を使って人は材料をつくり人工物を作るのですが創造主によって叱られてしまうのです。この逸話からわかることは何でしょうか。

ここでも、人が持つ機能をすべて自由に使っても良いわけではないということに気づきます。作ってもよいもの、作ってはいけないもの、よく考えて作らなければならないということでしょう。また、人は人であるべきであり、創造主の如く振る舞う行為は好ましい行為ではないということでしょう。人が創造主ごとき力を得ようとする、人の能力の過剰なエンハンスメントも禁止されているように感じます。人の分を知る必要があり、身の程を知る、自分を過大評価しないことが重要でしょう。

自由意思の議論と同様に、人が何をすべきかを人自身が深く考えることを求められているのでしょう。AIを構築するときにも、作りたいという欲求に行動をまかせるのではなく、よくよく行動を吟味する必要があるのでしょう。心理学者のスタノビッチとウェストは、思考には早い思考と遅い思考の二つのモードがあることを示しました。そして、それぞれ「システム1」「システム2」と呼びました（文末〔参考文献〕9）。システム1は情動と連動し、直感的で速い思考モードのことです。システム2は論理的で遅い思考モードであり、論理的、統計的な思考はこのモードで行います。

良心に基づく行動もこれら二つのモードが存在するでしょう。良心（システム1）は、直面する問題に対応しようとする動機となります。良心（システム1）に起因する行動は直観的であり、迅速にものごとに対応できます。一方で、論理的に多くのことを検討して問題に対応しなければならない場合もあるでしょう。すなわち、良心（システム2）は、論理的に多くのことを検討して問題に対応しようとする動機となり、「熟考しようとする良心」のことを指します。

「特化型AI」はこれまでの人工物の延長上にある人工物だと考えられます。「汎用型AI」は、自由意志を有する必要があります。一步、創造主に近づいているのでしょうか。良心（システム2）に起因する行動は、時間を要します。「汎用型AI」を構築しようとしている時に、我々はシステム1ではなくシステム2を起動して行動しているのでしょうか。

第8章

まとめ

人工知能（Artificial Intelligence=A I）は「特化型A I」と「汎用型A I」に分類できます。本講演では、「汎用型A I」について、聖書に見られる逸話から検討を進めてみました。

明確な結論を得るには至っていませんが、一つの方向性は見出せたのではないかと感じます。また、今回お話しした聖書の言葉から、A Iや科学を語るキリスト教主義的な手法は見直されて然るべきだと思います。科学に倫理が必須な時代になりましたが人の行動指針を制約するには宗教の力を借りないとできないかもしれないからです。そしてこのキリスト教的な科学とのやりとりは、同志社大学が得意とするところでしょう。

最初にも申し上げましたが私の話をきっかけにして、みなさんとともに考えることができると思っております。これがConscienceであり、同志社の良心だと考えています。本日はありがとうございました。

〔参考文献〕

- [1] 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)
- [2] 同志社大学良心学研究センター 『良心学入門』(2018)
- [3] 同志社大学良心学研究センター 『良心学入門、良心から科学を考える：パンデミック時代への視座』(2021)
- [4] NHK 「人間ってナンだ？超AI入門」制作班 『超AI入門—ディープラーニングはどこまで進化するのか』(2019)
- [5] 鳥海 不二夫 『強いAI・弱いAI：研究者に聞く人工知能の実像』(丸善出版 2017)
- [6] フィリップ・K・ディック 『アンドロイドは電気羊の夢を見るか？』(浅倉久志訳 早川書房 1977)
- [7] カズオ・イングロ 『クララとお日さま』(土屋政雄訳 早川書房 2021)
- [8] ユヴァル・ノア・ハラリ 『ホモ・デウス：テクノロジーとサピエンスの未来 上・下』(柴田裕之訳 河出書房新社 2018)
- [9] ダニエル・カーネマン 『ファスト&スロー：あなたの意思はどのように決まるか？』(村井章子訳 早川書房 2014)

2021年10月26日 同志社スピリット・ウィーク秋学期オンラインによる「講演」記録