特　集　科学技術と宗教

人工知能と宗教

―『AI原論』から見えてくるもの―

西垣　通[1]・島薗　進[2]（聞き手）

西垣通氏が2018年に上梓した『AI原論』は、従来のシンギュラリティ仮説を現代哲学やネオ・サイバネティクスの議論を用いて批判的に論じた書として話題となっている。同書では、ユダヤ＝キリスト一神教的な創造神話がAIの理論的基盤になっていると論じられている。本稿では西垣氏より、同書を読み解くキーワードについての解説をいただいた。

[1] にしがきとおる：東京大学名誉教授
[2] しまぞのすすむ：上智大学特任教授、（公財）国際宗教研究所理事長

島薗　このたび上梓された新著『AI原論』[1]にはいろいろ宗教的なことも書いてありますが、理解を深めるためにまず、西垣さんが自らの立場とする、基礎情報学について説明してください。あわせて「情報」とは何か、「階層的自律コミュニケーションシステム（HACS）」とは何かについても教えてください。

西垣　工学部出身でコンピュータ研究者として出発した私が、なぜ基礎情報学というやや哲学的な文理融合型学問の構築をめざしたのか、その動機から説明しましょう。現代は情報社会だといわれ、われわれはインターネットのなかに溢れる情報洪水の中に投げこまれています。けれども、情報とはいったい何でしょうか。学問的には現在、情報の定義をめぐって大きな混乱があるのです。たしかにインターネット内の記号つまり0と1のデジタル信号の量は膨大なのですが、それが表す意味内容という点ではどうでしょうか。もしかしたら、ひどく貧困で偏っているかもしれない。記号の量が多くても意味内容が豊かとは限らないのに、このあたりは、きちんと論じられていません。

情報というのは、「情況報告」の略で、もともと軍事用語だったという説があります。敵の様子を知ることには有用性がある。だから、広くいえば、われわれにとって価値／意味（significance）があるものが情報といってよいでしょう。しかし一方、コンピュータとは記号（デジタル信号）を形式的ルールにもとづいて扱う機械で、その処理は本来、意味や価値とは無関係なのです。だからコンピュータ情報処理が盛んになるほど、情報社会の実体を分析しにくくなりがちです。こういう混乱が生じたのは、20世紀半ばに提唱されたクロード・シャノンの情報理論[2]が誤って人々に受容されたからです。シャノンは通信工学者で、その理論は通信工学的には優れていたのですが、そこで定義された情報の基本概念はあまりに狭いものでした。つまり、シャノンが情報と呼んだのは「記号（機械情報）」だけであり、われわれが通常用いている「意味をふくむ情報」とはまったく別ものだったのです。

この点を改善し、情報という概念を、社会的な意味（社会情報）、さ

らにその根源にある生命的な価値（生命情報）にまで拡張したのが基礎情報学[3]に他なりません。基礎情報学においては、「情報伝達」という概念がとらえ直されます。コンピュータ同士なら記号を誤りなく伝送すれば終わりですが、人間同士では、たとえメールを正しく送信しても意図が通じるとは限らない。そこに「情報の意味解釈」というプロセスが入るからです。われわれはそれぞれ個別の閉じた心を持っていて、他者の心のなかは原理的にわかりません。では人間同士の情報伝達とは何でしょうか。むろんわれわれは、自分の知識や経験によって情報の意味を解釈しており、単純な事務連絡なら情報はたぶん伝わるでしょう。しかし、微妙な恋文なら伝達が成功するかどうか誰にもわからない。人間の閉じた心のあいだの情報伝達という難問をシステム論的に扱うのが「階層的自律コミュニケーションシステム（HACS：Hierarchical Autonomous Communication System）」というモデルで、これは基礎情報学の理論的中核をなしています。そこでは、人間の心のシステム間の情報伝達を、その上位にある社会システムの作動に関連づけてとらえるのです。

島薗　生命と機械の根本的な違いは何でしょうか。生命を特徴づける「オートポイエーシス」について、また、生命がもつ「自律性」、機械の「他律性」について教えてください。

西垣　基礎情報学は、生命システム論であるオートポイエーシス理論[4]をベースにしています。「生命とは何か」は古来の難問ですが、1970年代に生物学者ウンベルト・マトゥラーナとその弟子フランシスコ・ヴァレラによって提唱されたこの理論は、生命のもっとも洗練された定義といえるでしょう。「生物とは自己（オート）を創出（ポイエーシス）する存在だ」というわけです。単純な生物である細胞も、自分で自分を創り出しつつ生命活動をつづけている。人間の心も同じで、自分で自分の思考を紡ぎだすのです。とくに、環境のなかでの作動（行動）のルールを自分で創りあげる点が本質的です。ここは、コンピュータをはじめ、人

間の設計したプログラムにしたがって作動する機械との最大の違いといえるでしょう。自分の作動のルールを自分で創りあげるから「自律性（autonomy）」をもつわけです。

機械は作動のルールを他者（人間）が創りあげるので「他律性（heteronomy）」によって特徴づけられます。機械のなかには学習機械というものがあって、そこでは作動ルールが環境条件によって変更されます。しかし、メタルールである「変更の仕方」はあらかじめ人間が設計するので、他律性は変わりません。近年、人工知能の発展とともに「自律型機械」と称するものが出現していますが、これは正確な表現ではなく、環境に適応するよう学習する「適応型機械」と呼ぶべきでしょう。

自律性をもつからこそ、そこに自由意思や責任が生まれてきます。ロボットは自由意思をもっているように見えても、実は設計された通りに作動しているだけで、自由意思などもっていません。システム論的には、生命は閉鎖系で機械は開放系なのです。開放系だからこそ、ロボットの作動を人間が外部から自在に操作することができる。一方、人間のような閉鎖系ではそんなことはできません。では、それなのに、われわれが本来もっているはずの自律性や自由意思が阻害されるような気がする場合があるのはなぜでしょうか。実はこれは、人間同士の情報伝達という概念と関わってきます。

人間の心が閉鎖系なら、情報がお互いの間で伝わることなど本来ありえません。実際、オートポイエーシス理論においては情報という概念は除外されてしまいます。この点を理論的に工夫したのが基礎情報学のHACSモデルなのです。そこでは下位に人間の心の集団があり、その作動（発言など）を素材として上位の社会システムでコミュニケーションが実行されるのですが、上位の社会システムから見ると、人間はそれぞれ役割を果たしつつまるで他律開放系のようにふるまう。つまり社会的制約があると考えます。そこに「情報伝達」を可能にする余地が生まれ、また逆にいうと、人間がもっているはずの自律性や自由意思が阻害される余地も発生します。以上のようにして、HACSモデルにもとづき、情報社会における自由や責任といった問題を論ずることが可能にな

るのです。

島薗　人工知能（AI）の開発はどのような段階をたどって発展してきたのでしょうか。主要な学者の切り開いた地平とともに教えてください。そして、現代のAIブームの特徴について説明してください。

西垣　いまのAIのブームは第三次ブームです。第一次ブームはコンピュータが実用化された直後の1950〜60年代でした。1956年のダートマス会議でジョン・マッカーシーがAI（Artificial Intelligence）という言葉を使い、それが広まったといわれています。知能（intelligence）の実体とは論理的な計算処理であり、それを高速実行するのがAIだというわけです。もともと、コンピュータは数値計算だけでなく、論理計算をおこなう機械として誕生しました。初期のアイデアをつくった数学者のアラン・チューリングやジョン・フォン・ノイマンは、思考とは形式的な論理計算だと考えていたのです。だから第一次ブームが起きたのは自然な流れでした。しかし、形式的な論理計算で片付く問題はパズルやゲームくらいしかありません。応用範囲が狭すぎてすぐ下火になってしまいました。

第二次ブームが起きたのは1980年代で、そこでは「知識（knowledge）」がキーワードとなりました。主導したのは、私が留学したスタンフォード大学のエドワード・ファイゲンバウムです。基本的な考え方は、法律や医学など専門家の知識を論理命題として表現し、それらをコンピュータが自動的に組み合わせ、推論して結論を導くというものです。専門家の代わりをするのでエキスパート・システムとよばれます。一時は、弁護士や医者が失職するだろうという声さえ聞かれました。日本で第五世代コンピュータ開発という巨大な国策プロジェクトがおこなわれたのも、このときです。これは推論操作を高速化する機械で、技術的には成功したものの、まったく実用には供されませんでした。エキスパート・システムも同様で、一部で引き続き使用されているものもありますが、もはや下火です。弁護士や医者が失業しなかったこ

とはいうまでもありません。

こうして1990年代になると第二次ブームは失速してしまいました。理由は単純で、人間の思考というのは、形式的な論理計算だけではないからです。人間が知識（情報）をもとに推論をするとき、かならず自由意思にもとづく意味解釈というプロセスが入ってきます。そこに曖昧さや矛盾があっても、経験や直観にもとづき総合的に判断して結論を出すのです。そもそも、法律や医学の知識命題の内容は、純粋に形式論理的というよりむしろ多少の曖昧さを含むのが普通です。だから、コンピュータで厳密に自動推論しても結論は正確とは限らない。むろん、人間の専門家も間違えることもありますが、その場合、責任を問われる。コンピュータに責任はとれないので、信頼性が問題となって挫折したのです。

さて、現在の第三次ブームはいかに起きたのでしょうか。端的にいえば、「論理的に絶対正確でなくても、結論がだいたい合っていればよいだろう」と開き直ったからです。大量のデータを高速処理し、統計計算にもとづいて確率の高い解を自動的に出せれば便利だ、という考え方です。その背景には、インターネットに蓄積されていく膨大なビッグデータ、そしてまた、コンピュータ処理能力の急速な向上があげられます[5)]。

たとえば、外国語の文章を自動的に翻訳する「機械翻訳」というAI技術があります。これは第一次、第二次ブームのときも盛んに研究されたのですが、うまく行きませんでした。一生懸命に構文を解析しても、文法規則にも例外があるし、多義語を訳しわけるには文脈をとらえなくてはならない。正確な訳文をつくるのは困難だったのです。しかし、第三次ブームのいま、けっこう実用化されています。これはコーパス（用例データベース）を駆使して、原文と訳文のペアを検索し、もっとも頻度の高い訳文を出力しているのです。原文が簡単な文章ならそれなりに有用でしょう。でも難しい長文では間違いも少なくありません。とくに文学的な凝った表現などはもうお手上げになります。このように、第三次ブームは統計がキーワードで、応用範囲も広いのですが、誤りの可能性があることを忘れてはなりません。

島薗　人工知能の開発が進むと、人間にしかできないと思われていた「知性」の機能を機械が肩代わりするようになると考えられていますが、それにはどのようなものがあるのでしょうか。

西垣　近々、AIが人間の思考を肩代わりし、人間の仕事を奪うという声がよく聞かれます。さらには、2045年頃に人間より賢い超知性体の機械ができるという「シンギュラリティ（技術的特異点）仮説」[6]を信じている人もいます。しかし、私はこの仮説を支持できません。たしかにAI技術の進歩とともにわれわれの仕事のやり方は変わっていくでしょうが、全員が失業することなどありえない。少なくとも、現行のコンピュータをベースにするかぎり、AIが人間の思考をすべて肩代わりする日など来ないのです。

問題は、一部のAI学者がAIの能力を過大評価して宣伝し、マスコミが全能感を煽っているところにあります。技術的には、第三次ブームで一つのブレイクスルーがあったことは事実です。画像や音声、文章などのパターンを認識することは、昔からコンピュータにとって苦手な作業でした。論理処理だけでは片付かないからです。ニューラルネット・モデルといって脳神経網をまねた技術がパターン認識に有効だといわれてきましたが、あまりに膨大な計算を要するので、第一次／二次ブームの頃は実用になりませんでした。ところが、2010年代に入って、ジェフリー・ヒントンらがその実用化に先鞭をつけ、瞬く間に応用範囲がひろがっていきました。この技術は「深層学習（deep learning）」と呼ばれています。専門的な細かい工夫は別にすると、コンピュータの能力があがったことが実用化の最大の原因といえます。この技術の利点は、パターンの特徴を人間が詳しく教えなくてもAIが勝手にパターンを分類してくれることです。たとえば猫の画像を認識するとき、猫の顔の特徴の入力処理がほとんどいらない。そこで、一部のAI学者が「コンピュータが猫という“概念”を把握した」と宣伝しました。この延長で、AIが世界を自律的に認識しているという幻想がマスコミをつうじて振りまかれてしまったのです。

しかし、これは間違いです。AIはただ統計処理をして同じような画像パターンをグルーピングしただけで、猫という生物的概念を学習したわけではありません。要するに、現在のAIも、可能なのは数値計算／論理計算だけなのです。ビッグデータの分析には役にたつでしょう。しかし、最新のAIでも、人間のように世界から情報を受けとって、その意味を解釈しているのではないのです。

島薗　機械にできない事柄へのドレイファスやサールらの哲学的批判、また「フレーム問題」「記号接地問題」について教えてください。また、人工知能があたかも人間のように現出する可能性について教えてください。

西垣　ドレイファスやサールの批判の正確な内容は、AI技術だけでなく専門哲学的な議論をふくむので、詳細は参考文献[7]にあたってください。ここでは、あくまで情報学の観点から要点をかみ砕いて説明します。

もっとも決定的なポイントは、AIが記号のあらわす「意味」を理解できない、ということです。たとえば、猫とかcat（英語）とかchat（フランス語）とかいう記号から、人間はあの四つ足の可愛らしいペット動物のイメージを思い浮かべますが、それはコンピュータにはできません。「吾輩は猫である」という文章を夏目漱石という人名データと結びつけるAIプログラムはあるでしょうが、そこに、猫というペット動物のイメージは介在せず、ただ用例データの共起関係から形式的に操作しているだけなのです。このように「記号」をそれが表す意味内容に接合できないことが、「記号接地問題（symbol grounding problem）」にほかなりません。つまり、AIが扱うのは統辞論（syntax）だけで意味論（semantics）を扱えない、というのがドレイファスやサールの批判の骨子です。実際、AI研究者によって意味論を扱う試みは種々なされてきましたが、いずれも成功しませんでした。繰り返しになりますが、「グーグルの猫認識」でAIが猫という「意味」を把握できるわけではないのです。

サールがのべた「中国語の部屋」[8]という有名な例があります。漢字を読めず、中国語もまったく知らない英国人が部屋のなかにいて、外部の

人から壁の小穴をつうじて中国語の質問状を渡されます。漢字の文章を形式処理するためのルールブックは部屋のなかにあるので、英国人はルールにもとづき、質問状の記号を逐次処理して回答の記号列（文章）を書き、それを小穴から外部の人に返します。このとき、英国人は中国語の文章を「理解」しているといえるでしょうか？――いえない、というのがサールの考えです。AIのしていることは、所詮はこの英国人と同じであり、文章の意味とは無縁だというわけです。

いったいなぜ、機械には意味を理解できないのに人間にはできるのでしょうか。中国語のできない英国人でも、空港で迷うなど、どうしても必要な状況では、片言で中国人と「対話」できます。この理由は哲学的には、人間が身体をもって世界で生活しており、そのこと自体が、ある文脈（状況）のなかで志向性をもって世界の意味をとらえていることになるから、と説明できます。言い換えると、人間は身体をもって生きることをつうじて、環境世界から自分にとって意味（価値）があるものを選びとり、自分の世界を構成しつづけているのです。

このように、環境世界はつねに流動的で無限に変転していくものです。ゆえに、あらかじめその詳細を形式的（固定的）に記述しておき、論理操作によって問題の解をえることは難しい。これが「フレーム問題（frame problem）」なのです。つまり、AIの具体的応用の場面で、問題のフレーム（枠組み）を厳密に定義することには大きな困難があるのです。近くの店でハンバーガーを買ってくるといった、人間の子供には簡単な問題でも、AIロボットには難しい。なぜなら、店までの道が工事で塞がっていたり、お目当ての商品がたまたま品切れだったり、値段が変わったりと、状況はいつも無限に流動しているからです。この点は、囲碁や将棋のような有限状態ゲームとは本質的に違う点といえます。

したがって、AIが人間と当意即妙の対話をすることは不可能に近い。人間のように、状況の特殊事情を勘案して、リアルタイムで場面に即した行動をすることは非常に困難だということになります。状況（環境世界）が恒常的に安定しているときは、AIは人間よりも効率よく行動するでしょう。しかし、状況の流動性を考慮するなら、人工知能があたか

も人間のように現出する可能性は低いといわざるをえません。

島薗　カンタン・メイヤスーの理論がもつ意義と、思弁的実在論について教えてください。思弁的実在論は意味了解的存在としての人間という20世紀大陸哲学の考え方を凌駕するものなのでしょうか。

西垣　ドレイファスやサールに代表される現代哲学においては、世界のありさまはあくまで人間というフィルターを介して、人間の思考と相関的にとらえられる、とされています。最初にこれを明確に論じたのはご承知のようにカントでした。人間は絶対的存在である「物自体」を透明に絶対的に認識できるわけではない。このことは、人間という生物が、特有の知覚器官や脳神経系で限定づけられていることからも納得できます。こういう相対的世界観は、理性をもつ人間が世界を絶対的に分析できるという、西洋の古典的な形而上学の傲慢な独断を打破する思想として位置づけられるでしょう。20世紀後半に主流になったポストモダン哲学がとりわけ相対主義を主張していることは周知のとおりです。

しかし、こういう相対主義は、古典的な形而上学から派生した素朴実在論（人間と関わりなく、客観的に宇宙／世界が存在しているという常識的な思想）にもとづく、現代科学技術の営為の基盤を揺るがします。さらに、逆説的ですが、相対主義はその副作用として、絶対世界のありさまを語る神秘的な狂信主義を否定できなくなってしまうのです。具体的には、人類が出現する以前の出来事、たとえば地球の誕生といった出来事を科学的に語る言説の根拠はどこに求められるのか、という疑問が出てきます。メイヤスーの思弁的実在論[9]は、こういう問題と正面からとりくむラディカルな新しい哲学なのです。

メイヤスーは、世界／宇宙が神的な論理秩序にしたがって構成されているという古典的な形而上学に戻るわけではありません。相対主義をつきつめると何らかの絶対性を否定できなくなるということを緻密に論証し、その結果、世界／宇宙には絶対的な「事実」はあるけれども、そうなっているのはあくまで「偶然」だと論じます。それらの事実をむすび

つける何らかの絶対的な理由律などは無い、と断じるのです。メイヤスーは、科学的な論証からえられた「地球が45億年あまり前に誕生した」などの言説は客観的な事実性をもっていると論じたいのだと思います。彼は一種のデカルト主義者であって、数学的理論を信頼しているといえるかもしれません。なかなか面白い哲学であることは確かです。

しかし、事実をのべる科学的言説がいかにして客観性（絶対性）をもちうるかについて、メイヤスーは詳しくのべていません。さらに、「万象が偶然に変化しうる」というのは、科学技術にとって望ましい議論だとはいえないはずです。したがって私は、彼の議論が意味了解的存在としての人間という理念をくつがえすことに成功したとはいえないと思います。

島薗　知の追求には、人間が生存するための実践的動機によるか、絶対的・普遍的な真理に到達することを求めるという動機によるかの両面があると述べておられます。人工知能への情熱にこれがどう関わっているのでしょうか。

西垣　日本と欧米とでは、AI開発の動機や目的がかなり違うと思います。SFファンをのぞいて、この国では、AI開発にとりくんでいる関係者にとって、ビジネス面での実践的で短期的な成功が直接の主要目的です。たとえば、従来のような標準品の大量生産では労賃のやすい開発途上国に勝てないので、AIを利用してカスタムメイドの特注品を効率よくつくり、経済成長につなげるといったことです。

しかし欧米では、そういう短期的な動機ばかりではありません。国や企業の研究機関においても、より遠大な目標が掲げられ、そのために巨額の資金が投入されているのです。遠大な目標は、AIのような先端技術をもとに超知性体が出現し、それが世界／宇宙を進化させていく、といった一種の宗教的な夢想と一体になっています。私はそこに古来のユダヤ＝キリスト一神教の名残をみるのです。

これはトランス・ヒューマニズム（超人間主義）とよばれます。その

中にもいろいろありますが、典型例は、2045年にAIの知性が人間をしのぐという、前述のシンギュラリティ（技術特異点）仮説でしょう。そこでは、生物（人間）と機械の境界線はありません。ともに神の被造物だからです。こうして、脳をスキャンして情報をコンピュータ上に移すことで人間が不死になるというマインド・アップローディング説が述べ立てられることになります。

トランス・ヒューマニズムは人々を惹きつける魅力をそなえており、それがAI開発の情熱をうむ一因であることは間違いありません。ただ、そこには曖昧な点があります。トランス・ヒューマニストはどこから世界を眺めているのでしょうか。そこで語られる「知性」がはたして人間の生存のための知なのか、絶対的真理にいたる超越的な知なのか、判然としないところがあります。「人間を超える知」というと後者のような気がしますが、マインド・アップローディングは前者のような気もします。そして、この相違は、さきほどの相対主義と絶対主義という哲学的立場の違いとも関連しています。トランス・ヒューマニズムは生物と機械を同質ととらえるので、基本的には絶対主義に近い立場だといえるでしょう。つまり、AIの未来楽観論は、素朴実在論かせいぜい古典的な形而上学にもとづいているのです。ゆえに、この点をドレイファスやサールなど相対主義の現代哲学者から鋭く批判されることになります。そして、AI研究者はその批判をうまく論駁できません。

私が拙著『AI原論』のなかでメイヤスーの思弁的実在論をもちだしたのは、この論点を明示したかったからなのです。メイヤスーは形而上学の否定をきちんと踏まえた上で、現代哲学の相対主義に挑戦し、科学技術が絶対的事実にアクセスできると主張します。だからもし仮に「人間を超える知」があるとしたら、少なくともそれは、思弁的実在論のような「相対主義をふまえた絶対主義」にもとづかなくてはならないでしょう。しかし、拙著でのべたように、結論として、思弁的実在論はトランス・ヒューマニズムをささえる哲学的ベースにはなれないのです。要するに、トランス・ヒューマニストは、自分たちの主張の哲学的基盤が脆弱であることを認めなくてはなりません。

島薗　人工知能や情報工学は人工物を通して自然を捉え、ひいては生命を捉えようとするのですが、その力の増大は人間の倫理性をむしばみ、「責任」や「自由」の意義を見失わせる可能性はないでしょうか。

西垣　トランス・ヒューマニストのAI開発者にとって、人間と機械は同質です。自律型AIは思考でき、自由意思をもち、責任もとれるという極端な主張をするのです。そして情けないことに、この国でも、深い考えなしに欧米のトランス・ヒューマニストに追随する人もいますし、「自律型AI」という言葉がマスコミをつうじて流れています。こういう考え方が人間の倫理性をむしばむ可能性があるというご指摘に賛同いたします。

前述のように理論的には、あらゆる機械は人間の指示にしたがって作動する他律系であり、AIも例外ではありません。ですが問題は、いったいなぜ「自律型AI」という言葉が用いられ、機械と自由や責任を結びつける誤解がうまれるのか、ということでしょう。

ここで、動いたり話したりする眼前の対象が自律性をもつか否かをいかに判定するか、という問いを立てなくてはなりません。対象の行動が予測困難なとき、われわれは対象が自律性をもつのではないかと推測します。そして、自律性が選択の自由意思をもたらし、必然的に責任が生じてきます。だから、予測可能性がキーポイントになります。

生物は自分で時々刻々、自分の行動ルールをつくるので、習慣性による予測はできても、原理的に行動の正確な予測はできません。普通の機械は、所与のルールにしたがって行動（作動）するので、予測がつきます。しかし、AIでは、その内部ルールが非常に複雑なので、たとえ所与のルールにしたがっていても事実上はその行動（作動）を予測困難なことが多いのです。AIロボットに話しかけると、けっこう気の利いた回答が返ってくることもないではありません。こうして、「疑似的自律性」がうまれることになるのです。

このように、AIのブラックボックス化が「自律型AI」をもたらすのです。さらに加えてトランス・ヒューマニストたちはAIを神秘化し、

人間とAIの異質性を否定します。こうして、AIは自律性をもち、自由意思をもって選択行動をおこなっており、だから責任もとれるはず、という神話がささやかれることになります。

前述のように、第三次AIのもとでは、出力に統計的な誤りが発生するので、事故がおこったりすればその責任が問われることになります。いくら人間とAIが同質だと強弁したところで、コンピュータを刑務所に入れるわけにはいきません。そこでAIの製造者や使用者の責任が問われることになるのですが、実際には責任者の特定はなかなか難しいでしょう。AIプログラムが抽象的に記述されていることもあって、責任は分散してしまいがちです。さらに、「AIは公平中立な機械である」「機械は情報を正確に伝達し処理する」という建前があるので、責任のがれがおこなわれる可能性も大きいと考えられます。結果的に、社会の倫理性が損なわれる恐れがあるのです。これはAIの利活用において本質的な問題点となるでしょう。

島薗　人工知能開発が、人工物と道具的理性の生活世界に対する影響力を強める可能性はないでしょうか。身体的存在としての人間の、環境全体と接する経験、またスピリチュアルな経験の次元を軽視する可能性はないでしょうか。

西垣　以上のべてきたように、AIという技術をささえる思想は、とかく人間（生物）と機械（人工物）との同質性／連続性を仮定しがちです。これは、時々刻々、リアルタイムで生きている身体の重要性の無視につながります。私は脳科学の進歩は大切だとは思いますが、人間という存在をその脳メカニズムと同一視する議論には賛同できません。個人の脳細胞の詳細をスキャンして電子回路の基板にコピーすれば、個人のアイデンティティがそっくりコンピュータに移行できるというマインド・アップローディングは、まさにそういう脳中心主義の議論の典型といえるでしょう。

最近の医学的な研究によれば、内臓の諸器官のはたらきが記憶などの

脳の活動に強い影響をあたえているそうです。脳は身体という情報ネットワークの中枢ではあっても、脳だけを切り離してしまうと、うまく機能しないのではありませんか。身体は自然環境のなかで、時々刻々、自らを更新しつづけています。その活動は脳の論理構造に影響をあたえますが、逆に脳の論理構造だけから身体活動を演繹することはできません。

さらに、人間の身体は社会的存在でもあります。責任や自由といった概念は社会的な相互作用からうまれるのですが、そのあたりが脳中心主義の議論においては曖昧にされる恐れがあります。たとえば、人間の思考や意思は脳細胞の生化学反応によって生みだされており、自由意思などは幻想であって、脳細胞の生化学反応の因果関係が本質的だということになります。これを敷衍すれば、殺人をおかしても犯人に責任はなく、その脳細胞の活動のせいだということになるかもしれません。

しかし、こういう発想は、人間という存在を脳科学という非常に限られた側面からとらえた議論にすぎないと考えられます。ある対象を、一定の側面からとらえることは正確な議論の前提ですが、それ以外の側面からとらえれば全く違う議論が出現するのです。たとえば、音楽は物理的には空気の振動にすぎないし、絵画も化学物質の塊にすぎないけれど、それで芸術論がなくなるというのは暴論でしょう。

そういう意味で、私はメイヤスーと並んで現代の新実在論の旗手とされるドイツ哲学者マルクス・ガブリエルの議論[10]に賛同しています。ガブリエルは、メイヤスーと同様に相対主義を批判しますが、一方で科学万能主義に異をとなえ、われわれが多様な領域で客観的議論ができると主張するのです。AIと結託したトランス・ヒューマニズムには、科学万能主義の気配が感じられます。スピリチュアルな次元の議論も、人間が生きるうえで決して軽視すべきではありません。

島薗　人工知能の追求は実際的動機によって進められている側面が大きく、それはテクノクラートの支配、高度管理社会、新自由主義と深く関わっているように思うのですが、それについてはどのようにお考えでしょうか。

西垣　簡潔にまとめてみましょう。ユダヤ＝キリスト一神教の創造神話と古代ギリシアの哲学から、世界／宇宙がロゴスにもとづいて論理的に構成されており、理性をもつ人間はその探究と進化にむけて努力し、貢献すべきだという思想があらわれます。こういう方向性が啓蒙時代をへて世俗化し、現代の神話であるトランス・ヒューマニズムをうみ、AIの思想的母体となっているのです。

このこと自体は宗教的で神聖な感じがします。しかし、AI開発が実践的・利益追求的な性格をもつのは、ユダヤ＝キリスト一神教の世俗化のプロセスとも深く関わっていると思うのです。宗教改革において、カルヴァンの予定説は、教会の管理支配から一般の民衆を解き放つには有効だったのでしょう。しかしその一方で、あの世での救済をこの世での世俗的成功に短絡してしまう議論がうまれたことは、お金儲けを徹底的に是認する口実となってしまったのではありませんか。

私は信者ではありませんが、イエスが説いたキリスト教は本来、貪欲をいましめ、貧しい人々の魂を救うための清廉な宗教だったはずです。それが逆の方向に捻じ曲げられ、高度に発達した情報通信技術を用いて、多くの人々を管理し、経済的に搾取するために利用されるとすれば、非常に悲しむべきことではないでしょうか。

AIは他律的な機械です。AIの作動を細かく操作できるのは、一部の支配層にかぎられます。したがって、AIが社会的判断を下しつつ仕事をする社会とは、社会的制約がきわめて強化され、テクノクラートが自在に支配力をふるえる新たな情報社会になるでしょう。すると人間の自由な思考とその伝達の範囲はどんどん狭められてしまいます。マインド・アップローディングなどの高度な医学技術は、万一もし仮に実現できたとしても、一部の超富裕層のためのものです。一般の人々はけばけばしい娯楽を投げあたえられ、生活を高度に管理され、まるで取り換えのきく安価な機械部品のように隷属した生活をおくることになるでしょう。

そういう未来にならないためには、いったいどうすればよいのか？――これこそ、われわれが今、真剣に考えるべきテーマではないでしょ

うか。私が構築している基礎情報学は、その理論的検討のベースになると考えています。

注

1）西垣通（2018）『AI原論――神の支配と人間の自由』講談社（選書メチエ）
2）シャノン、C. E.＋ウィーバー、W.（2009）『通信の数学的理論』植松友彦訳、ちくま学芸文庫
3）西垣通（2004・2008）『基礎情報学（正・続）』、NTT出版
4）マトゥラーナ、H.＋ヴァレラ、F.（1991）『オートポイエーシス』河本英夫訳、国文社
5）西垣通（2016）『ビッグデータと人工知能』、中公新書
6）カーツワイル、R.（2007）『ポスト・ヒューマン誕生』井上健監訳、NHK出版
7）ドレイファス、H. L.（1992）『コンピュータには何ができないか』黒崎政男・村若修訳、産業図書
8）Searle, J. R.（1980）“Minds, Brains, and Programs”, *The Behavioral and Brain Sciences*, Vol. 3, No. 3. pp. 417–457.
9）メイヤスー、Q.（2016）『有限性の後で』千葉雅也ほか訳、人文書院
10）ガブリエル、M.（2018）『なぜ世界は存在しないのか』清水一浩訳、講談社（選書メチエ）