

AI 第三次ブームが先鋭化させる「テクノロジーと社会」の論点

— 文理を融合してこれから何を考えるか —

福留恵子*

Abstract

This article aims to preliminarily consider the meaning of “technology as a means”. First, the concept of “deciding purpose” is rethought with the analysis of the discourse about Artificial Intelligence (AI) and its autonomy. On the base of that, secondly, some of the factors which confuse the objective rationality in the social treatment of technology are discussed. And this discussion also suggests the influence of one of rhetoric, namely “personification of technology”. This suggestion requires us to give insight moreover on this problem with the integration of humanities and sciences.

1 はじめに～ AI への“怖れ”と AI の目的

第三次ブームの到来とされる昨今の AI（Artificial Intelligence：人工知能）の研究開発¹と実用拡大に伴い、さまざまな観点からの議論が続いている。その議論は新たな技術への期待とともに、しばしば怖れを帯びてもいる。その怖れの細部は多岐に渡るが、多くに共通するのは AI が人々の望むところと異なって作動することと、それがもたらす結果にかかわるように見受けられる。

一方で、AI の目的、あるいは AI が何を行うのかを決めるのは、機械ではなく人間である、または人間であるべきだとの意見もしばしば見られる。これが時に語られるように、議論の余地のない自明の理だといえるかどうかは、よく考えてみなければならない。実際に AI 研究開発の目指すところは、それ自体の設定した目的を遂行するように動作する機械（汎用 AI と呼ばれる）だと考える人々も少なくない。このように考える場合には、AI の目的を定めるのが人であることには議論の余地があるばかりか、むしろ自明に否定されさえもすることになる。そして、目的を機械自体が決定するとしたら、それが人間の望む結果と調和することは保証されず、その場合、上述したようにそれが怖れられることは合理的なのである。

それならば、そうあるべきだとの声にあるとおり、AI の目的を人が決めればよい。AI も一

* 東海大学経営学部・スチューデントアチーブメントセンター

つのテクノロジーである。「テクノロジーは使い方次第」、「テクノロジーはもろ刃の剣」とは AI に限らず、第 3 次ブームに限らず、繰り返されてきた言葉である。それが「社会」の「テクノロジー」に対する処方であり、テクノロジーの開発が配慮すべき社会的要因だったはずである。今こそ、まさに、AI（というテクノロジー）について、その徹底を図るべきときである。そしてそのことにより、テクノロジーは社会にとって有益で無害なものになるはずであり、AI も然り、である。「AI の開発は、それが役立つべき目的を上回ってはならず、あくまで人間の知能に資するだけにとどめられるべきではなかろうか。それならば、人間の知能を AI が凌駕するという「シンギュラリティ」などそもそも起りようがなく、杞憂（きゆう）であろう。」（酒井，2021）ともいわれるように、AI への恐れは AI の目的を人が決定しないことへの恐れである。しかし実際には、こと AI に関する限り、議論はそうように収束する様子を見せていない。誰（どのような人）が AI のどのような目的を決定するのか、あるいは決定しないのか、また、それによってもたらされると懸念される結果がどのようなものなのか。様々な想定のもとに、議論は収束よりはむしろ錯綜に向かう様相を呈しているようにも思われる。それは「目的を人が決める」ことの実現の難しさを、正確に言えば、その実現が容易だとは大方に考えられていないことを示している。人が目的を正しく望ましく設け、それ（のみ）を実現するようにテクノロジーの開発をリードする、そのことの実現可能性を信用することができていないのである。

それならば、そのように信用できない理由は何なのか。そしてそれは AI の場合に固有の不信感なのか、それとも実はテクノロジー一般についても潜在していたものなのか。AI をめぐる議論は多様だが、考え直してみなければならぬのは、そうした点である。そこで本稿では、既存の議論を照会することによって、「AI の目的を人が決定して、AI をその手段とする」ことの意味を再考したうえで、そのどこに信用しきれない要因が潜むのかを探ってゆく。

ここでは AI の社会におけるありかたが問題の中心となるだけに、参照すべき議論やその論者の背景となる専門分野は既に多岐に渡っている。そしてその多くが、AI 技術の研究開発分野と、倫理・規範などの人文・社会科学分野との対話の必要性を強調する。さらに、分野横断的な議論の場や議論のフレームワーク構築自体を志向する科学技術社会研究（STS）や、政府が先端技術活用に関する倫理的問題への対応のために必要とする「総合知」（「第 6 期科学技術・イノベーション基本計画」，文部科学省編，2021）も、これに呼応する。

ただ、そこで異分野が横断的に共有できる議論のフレームワークを特定することは容易なことではない²。そのようなフレームワークを特定するには、それによって探究すべきテーマ（問い）を（一つであれ複数であれ）特定することが必要である。そして、上述のような AI に対する不信感、より正確にはこの社会の AI についての取り扱いへの不信感を探究することも、そのような分野横断的な取り組み課題の一つを明確にするための、予備的な考察に当たる。

そうした趣旨に照らせば、特定の分野の知見のみを先行研究として参照し、これにのみ依拠することは、ここではむしろ適切ではないかもしれない。本稿では、第 4 節を中心として、「正当性」の概念をめぐる社会学的な議論の蓄積に着目するが、これも社会学の応用、あるいは

は AI という一事象に対する拡張的な当てはめとしてよりも、上述の分野横断的な議論のフレームワーク構築への援用と位置づけたい。重要なのは、共有すべきテーマの見極めと、これを議論するのになかったフレームワークの構築なのだと考える。

なお、以下では、とくに断りのない場合、「AI」とは、この深層学習による AI のことを指すものとする。

2 コンピュータ・ネットワーク・システムとその目的

深層学習による AI が、それ以外のコンピュータ・ネットワーク・システム（CNS）に対して特異なものであるか否か、という議論をここではひとまず保留して、AI もまた、ネットワークで相互接続したコンピュータをソフトウェアで動かすコンピュータ・ネットワーク・システム（以下、CNS）であることに着目する。

議論に先だって、そこで用いる主な言葉を簡単に整理しておきたい。

まず、CNS は、文脈や分野によっては情報システムと呼ばれることも多いが、ここでは情報システムという呼称は採用しない。それには理由がある。広義の情報システムは直訳である「情報体系」の字義の通り、データやプログラムの体系を指すことがあり、そのデータがデジタル化されていないものを想定する場合すらないとはいえない。そうした広義の情報システムにおいては、なにがテクノロジーで、なにがその用法や目的に当たるのかを、個々の具体例に応じて措定する必要がある、それが深層学習による AI についての想定と必ずしも整合しない。

一方、本稿では、AI を含めて、デジタル化されたプログラムとデータによって機能、つまり用途を特定される CNS を考察の対象とする。そして、それによって、そこに組み込まれるハードウェアとソフトウェアを「テクノロジー」、ソフトウェアに対応してハードウェアの動作を決定し、それによって得る出力をユーザが行為に反映することを「使用（使い方）」と一意に位置づけることが可能になる。すると、議論の焦点となる「目的の決定」は、その使用に予期・推測・期待する（使用の）結果を決めるということである。

一般的な CNS においては、具体的なユーザを特定した最終的なシステムの動作の決定は、プログラムを作成するプログラマ、より正確には動作を決定するパラメータ等の設定を行う者と、データを入力する者³が行っていることになる。設定されたパラメータや入力されたデータによって、システムの動作が一意に決定されるため、このように考えることができる。ただし、そこで設定・入力することのできるパラメータやデータ、あるいはそれぞれのパラメータやデータが設定・入力された場合のシステムの動作は、予めプログラムによって制約・決定されている。このことを反映するなら、エンジニアがシステムの動作、特にその外延を、したがってハードとソフトの使用法を決定していることになる。

したがって、このエンジニアが、このテクノロジーの使用法を、つまりその目的までもを独断で決定することも、それが望ましいと否とにかかわらず、原理的には可能である。勿論、実際の CNS の設計・開発・運用においては、ユーザの意思を理解して反映することが重要だとされる。たとえば開発についての工夫として広く知られたプロトタイピングにもアジャイル開

発にも、テクノロジーの目的の決定にユーザの意思を十分に反映しようとする趣旨は共通する。それは、ユーザによる目的の決定がそれだけ強く求められているということであると同時に、しばしば目的の直接的で最終的な決定者はユーザではなく、ユーザにとって決定者が他者であるということをも示す。だからこそ、ユーザの意思の反映が強調される。ユーザは多くの場合にプログラマではないことは勿論、CNS の設計者でもなく、そうしたエンジニアがユーザの意思を理解することは必ずしも容易でない。だからこそ、(ユーザの) 要求は慎重に理解・確認されねばならないのである。

AI がソフトウェアでハードウェアを動かすところから、こうした事情は、CNS に AI が組み込まれる場合も、基本的には変わらない。たしかに、しばしば言われるように、深層学習のソフトウェアには、それまでプログラムの原則であった、ハードの動作とその順番を逐一指定するようなアルゴリズムが用いられていない。そして、機械学習の過程で、入力と出力を関係づけるプログラムが、人の手を直接介さずに変化することが想定される。そのため、AI やこれを含む CNS が(「学習」などのチューンナップを経て) 実用に供される時点で、エンジニアにも、その AI に特定の入力を行ったとき、何が出力されるのかを事前に特定することができない。また、どのような経緯でその入力とその出力に対応づけられたのかを厳密に、かつその対応づけの理由を人が理解できるように再現することができない。その意味では、たしかに、目的は「誰にも決定できない」と言える側面はある。

ただ、そうした想定を行う場合でも、そのことと目的が無制限に拡散し得る、つまり人の手による決定が何もしなされないこととの間には、距離がある。極端な例を一つ挙げるならば、機械学習を用いようとも、動いて人に物理的に接触する出力装置が用いられていないところに、人に物理的・身体的な暴力をふるう(ことを目的に含む) AI も CNS も出現しようがないことは明らかである。あるいはテキストを出力しない限り、暴言を吐くことができないのも然りである。

つまり「何を出力するかがわからない(特定できない)」ことと「何を出力しないかがわからない」こととは別である。エンジニアは、設計・開発の時点で、その AI が「何をしないのか」ならば知っている。言い換えれば、何がその AI の目的に含まれ得ないかはわかることができる。さらに、開発した AI の機械学習によってユーザが不適切と考える出力が出現したり、そうした疑問が生じたりした際に、設計をし直すとか、少なくとも開発された AI の使用を停止するという選択肢を含んだ設計をすることもできる。

そのため、機械学習の過程で入力と出力の対応が変化するシステム、つまり本稿でいう AI の場合でも、そこに組み込むハードウェア、プログラム、学習データを特定するなどの設計のありかたによって、システムの目的は特定までされずとも、少なくとも限定される。そしてその限定を行うのは、やはり直接にはエンジニア、そしてこれを介して意思を反映するユーザだということになる。

それならば、ユーザは自らの利益や価値観に合わせて、いや、少なくともこれに反しないように、AI の動作と出力を決定し、エンジニアはそれに整合するようにテクノロジーを組み合わせ、動作させることはできる。仮に実際に開発された AI が、ユーザの利益や価値にそぐわ

ない場合には、設計をし直すこともできる。資金面の事情その他で設計のし直しが困難であるとしても、開発されたAIを使用しないことならできる。いずれにせよそこには、ここで問題となるような、ユーザが望まない出力を強制される要因は見いだせない。それならば、AIによる支配を怖れなければならない理由も見いだせない、はずである。

それにもかかわらず、AIによる支配への怖れがぬぐえていないのだとすると、ここまで述べて来たような目的の決定、あるいはその実現が危ぶまれるのだということになるだろう。AIの目的の「人による決定」は、そのどのような要因が危ぶまれるのか。次にこの点を考え直してみよう。

3 「人が目的を決定する」（「人による決定」）とはどういうことか

だがそもそも、「人がAIの目的を決定する」、あるいは本稿においてそれと同義の「AIを人の決定した目的のための手段とする」とは、われわれの社会において、どのような事態を指すのだろうか。

第一は、そのAIの機能が、そのユーザの予め定めた目的を実現する、また、そのようにAIを動作させ、そのように設計する、ということであろう。AIの動作と出力を、人が予め設けた目的に整合的に制御する、ということもできる。字義上、人の意に反する動作は封じられる。人の指示に反したり、人に損害を与えることも起こり得ない。勿論、そこで「人」とは、そのAIの目的を設けた人を指し、目的が合理的なものであることが前提である。ただ目的の意味は、それにとどまらない。

第二に、目的は、そのAIの動作の結果を評価する基準となる。目的と一致するように動作するAIは「成功」であり、それ以外は厳密には「失敗」だが、実際にはその一致の度合いに応じて、成否が相対的あるいは段階的に判断される場合もあるだろう。

この第二の意味によって、目的に一致するような動作結果を実現し続ける手段は、その価値を認定されることになるわけだが、実は手段はそうした結果を出して初めて価値を認められるばかりではない。期待による価値付与が成り立つこともある。

その手段に対応する目的に、高い価値が認められていけばいるほどに、手段は目的のその高い価値の実現に寄与する（と期待される）ものとしての価値を付与されることになる。勿論手段としての期待がそれほど大きくなければ、付与される価値も相応のものにはなるだろうが、そもそも全く期待ができなければ、それはその目的の手段と目されることはない。したがって、ある（有意義な）目的の手段と目されること自体が、目的の意義を根拠として、手段たるそのなにものかの価値を認めるということ、すなわち価値付与（という行為）なのである⁴。

たしかに「目的の決定」というとき、そこで決定されるのは、その目的を実現する意志（のみ）であって、そのための手段は別途定めるのだと理解されることもある。明示的または暗黙裡に、「手段を択ばず」との意を含む場合には、なおさらである。そのとき目的に高い価値がある（と認めている）ことは言わずもがなである。それに加えて、その目的が特定の手段、たとえばAIという手段を見出し、その手段による実現が想定されれば、そこでその「目的の決

定」は、その手段にも価値付与するという、もうひとつの決定行為ともなる。「人による AI の目的の決定」とは、そのような（目的の実現に必要な手段である）AI に対する価値づけでもある。同じ「目的の決定」であっても、手段を未だ定めていない目的の決定と、手段を特定しそれに対応した目的の決定の間には、この点に大きな違いがある。

特にその手段が AI である場合には、いわゆる「特化型 AI」と「汎用 AI」の分類との関係も整理しておく必要がある。AI は、単一の特定された目的の手段たる「特化型 AI」と、目的を特定しない「汎用 AI」との二分法によって論じられる場合も少なくない。後者の「汎用 AI」は現存しない「未来の技術」だとの説明もしばしば付記されるが、実体がないだけに、その意味は後述するように必ずしも明確ではない。

このように分類する場合の特化型 AI は、上述のような目的－手段図式の「手段」になり得る AI である。つまり、その目的を決定した人間による、目的の価値を根拠とした価値付与の対象となり得る。一方、汎用 AI は、その意味するところによって、位置づけが異なってくる。

汎用 AI が、正確な字義通りに、その動作目的に、したがってその社会的機能にも、何の制約も設けられていない機械を指すのであれば、それは既に目的－手段図式でとらえることのできない対象である。つまり「手段となり得ない AI」である。この意味での汎用 AI については、次節で改めて述べることにする。

ここで丁寧にみておきたいのは、複数の、そしておそらくは多数の特化された目的を有して、そのなかのいずれかの目的が機械的に（自動的に）特定されて、その手段として動作する AI のことである。これは厳密には「汎用」の名に反しており、いわば「多目的型 AI」とでも呼ぶべき存在であるが、この多目的型 AI と汎用 AI とは、必ずしも峻別されてはいない。汎用 AI の語が、実質的には多目的型 AI を指している場合もあれば、特化型 AI から多目的型 AI へ、そしてその目的の増大の先に汎用 AI を想定し、特化型と汎用を連続したものにとらえているように見受けられる議論もある。

多目的型 AI の場合、まず、その複数の目的のそれぞれと、その手段それぞれとしての AI の間には、前述のような目的－手段関係が成立する。ただしそれと同時に、その複数目的群からの目的の自動選択の根拠となる基準、すなわちメタ的な目的の存在や、メタ目的と下位の複数目的それぞれ相互の整合性（メタ目的と下位目的それぞれの間の目的－手段関係、下位目的相互の無矛盾性等）がなければ、その AI は目的を拡散させていることになる。これは結局、やはり手段たり得ない存在でしかない。

「人が目的を決める AI」というとき、したがって、そこからはこうした「手段たり得ない AI」は除外して考える必要がある。つまり、「人が目的を決める AI」とは、①特定された単一の目的に対応する手段として開発する AI、または②特定された単一のメタ目的と、これに整合する複数の下位目的を持ち、その下位目的のそれぞれに対する手段として開発する AI の集合、このいずれかを意味することになる。これらの条件を満たさない AI について、その目的を人が決定するということは、形容矛盾である。そうした手段たり得ない AI には、少なくとも目的の価値に根拠づけられた価値は、論理的には認められない。そして、「人が AI の目的を決定する」とは、少なくともそうした条件外の AI を、開発可能性の議論から排除することと

同義である。

ここで考察対象から排除しようとしている AIこそが、テクノロジーの可能性である、あるいは脅威なのだとする考え方はあるかもしれない。たしかに、そうした AI について、考える必要のある問題もおそらく残る。ただ、それを「人が目的を決める AI」の一部として論じることはいかなる。

4 「人による目的の決定」の実現を妨げ得る要因

このように「人が AI の目的を決定する」こと（以下、「人による決定」）の意味を整理したうえで改めて照合すると、AI の目的をめぐる見解の多様さ、また、より一般的にテクノロジーと目的の関係性のなかに、「人による決定」を困難に陥らせ得る要因が散見される。そうした要因を網羅的に述べることは本稿ではできないが、その一部であっても、“AI への怖れ”を理解し、その怖れへの合理的な対処を考える手がかりとして、注意を払う必要のあるものがある。

(1) 「人による決定」をめぐる平行線による無効化リスク

まず、AI に関する「人による決定」は、一つの規範としてしばしば断定的な口調で語られてもいるのだが、一方、これに明確に対立する主張が存在している。そして両者の対話は必ずしも継続していない。この点で、「人による決定」は異論に晒されているのであり、その実現が信用され得ない以前の問題として、実現はそもそも図られる保証がない。

「人による決定」の主張には、たとえば、AI の開発が、「人工知能をはじめとする情報技術はあくまでツールであり、汎用人工知能等が実現をしない限り（そしてそれはそう簡単ではないが）、技術の目的を設定するのはあくまで人間である。」（江間，2019）⁵、あるいは「まず、AI「が」と言うのでなく AI「で」と言うべきだと思います。AI「で」何かをするのは人間、つまり AI で行う処理が高度化しても最終的に判断をするのは人間です。」⁶（中島，2021）、などの言明がみられる。そこでは、「でなければならぬ」「（あくまで）人間である」など、それが広く共有された規範であるかのような修辞が用いられる。

ただ、その江間や中島も留保しているように、彼らがそうす「べき」だとまで強く主張する「人による決定」は、一方で誰しもに賛同されているわけではなく、むしろ明示的に対抗を示されており、彼らもそのことを承知している。「対抗」とは、端的に言って、人の決めた目的に従うことなく動作する AI の実現を、目指す人々も存在するということである。人と AI の関係性として、そこで想定すべき事態は多様である⁷が、その一つに、「将来、我々が進んでそのような方向に開発を進めれば、AI の判断に人間が従うということも起こりうるでしょう」（中島，前掲）と述べられる事態もある⁸。

ここで中島のいう「AI の判断に人間が従う」には、「人が目的とし得る出力を AI が行う」ことと、「その出力を人が目的とする（そして行為する）」ことの二つの意味が含まれている。この二つが共に成立した場合にのみ、「人が AI の目的を決定していない」と明確に言い得ること

とに注意する必要がある。両者は自動的に重ね合わされるわけではなく、そこにも人間の決定を介在させる余地はある。しかしいずれにせよ、両者が重ね合わせられ、AI の目的を人に決定させないという選択肢が存在することを、中島は述べているのである。

その選択肢を開発者の立場でより明示的にしたものの一つが、人工知能学会倫理指針の序文冒頭にある「人工知能研究は、人間のような知性を持ち自律的に学習し行動する人工知能の実現を目指している（人工知能学会倫理委員会，2017a）との記載である。勿論、現状において人工知能そのものの定義が画一的でないことに加え、知性の内実にも多様な解釈が併存する以上、学会のこの宣言の意味にも曖昧さが多分に残るのだが、「自律的な学習と行動」が目的の決定を含み得るとの解釈は常識の範囲内であろう。

AI の自律的行動、すなわち、AI の目的の AI による決定は、「人による決定」の対立事象であり、両者は並び立たない。上述の倫理綱領の上述の記載にも、これに対する疑問提示や反論が行われている。たとえばこの倫理綱領を作成した人工知能学会倫理委員会が行った公開討論では、第9条の「人工知能が社会の構成員またはそれに準じるものとなるためには、上に定めた人工知能学会と同等に倫理指針を遵守できなければならない」との記載に関して、Danit Gal が「第9条で疑問を感じます。技術自身が人間と同じ基準を順守するとはどういうことでしょうか？人間を真似する形で AI を開発することは良いことなのでしょうか？悪いことなのでしょうか？」（人工知能学会倫理委員会，2017b）と述べている。

しかし、この公開討論自体がそうである⁹ように、この対立は両論併記されつつ、それ以上解消に向けて議論されるのではなく、むしろいずれの主張を行う側もそれぞれ自覚的に、これを平行線のままに放置しているように見える。そしてそのことによって、「人による決定」は常にチャレンジに晒され続けていることになる。たしかに本稿では、「人による決定」に自覚的に反する汎用 AI を特化型 AI と峻別し、これが「人による決定」の議論の俎上にのせるべきでないことを述べた（第2節）。それにもかかわらず、究極の目標を自律性に置くことを AI 開発の前提として認定すれば、特化型 AI は未完成の汎用 AI と見なされることになり、連続的に汎用 AI に向かうこともそこに含意され得る¹⁰。それは、特化型を目的特化に留めないことを通じて、「人による決定」（人によって決定された目的）は無効化することができるということである。

（2）利害対立と目的決定プロセスの私的所有

AI の目的をめぐる議論される「人による決定」は、ともすると AI 自体による決定（AI の自律）に対置される。するとその「人」とは、誰のことを指すのだろうか。AI でない人であれば誰でもよいのであれば、実質的にそれは人類（全体）に近似するしかない。

しかし、現実には目的を決定するのは、「人類全体」ではあり得ない。技術開発など、それまで存在しなかったものを作り出す場合は特に明らかに、どの決定も、少なくとも現行では、いずれかの個人や集団が、つまり「誰か」が行うことになるのである。つまり決定プロセスは私行されている。そして、その「誰が決定するか」という点こそ、決定とその結果がどのようなものになるのかにとっては重要である。

そこで決定主体が自らに有益な目的を決定すると想定することは合理的だが、一方でその決定が他者に不利益を与える、あるいは他の決定の妨害に当たるとも、当然にあり得ることになる。社会に利害対立が偏在する以上、異なる「誰か」が決めた目的相互が、自動的に整合し調和することは考えられないからである。だからこそ、決定手続きの正当性が歴史的にも追求されてきたのだろうが、そのことをここで改めて繰り返す必要もないだろう。

ただ、それが自明のこととして認識できるのは、目的の決定が誰によって行われるのか（具体的に）わかる場合である。どのような個人や集団がどのような利害を背景にどのような目的を決定するのが具体的でなければ、それが誰にどのような不利益をもたらし得るかも具体的に認識することはできない。それにもかかわらず、目的の「人による」決定に焦点化し、その決定権を「人と AI の対立」に投じる議論の立て方は、決定者を匿名化し、それによって不利益や妨害を被る人々をも匿名化し、結果として利害対立の可能性自体を認識しづらくさせる。認識できない対立は、積極的に防止することも解消することもできない。人と人との間に生じるはずの不調和を必然的に隠蔽し、そこに議論を立ち入らせない効果を生じてしまうのである。

また、こうした議論とそれがもたらす構造に気づいてみれば、それは自身に不利益をもたらす AI を誰かが意図的に（＝そのような目的を設定することによって）作り出すかもしれないという怖れから、誰もが免れられなくなるのだともいえる。そうした構造のなかで決定された目的にもとづいて開発される AI は、その決定を行うのが人であるとしても、他者はそれを警戒せざるを得ない。それは「人による決定」それ自体にとって、一つの困難である。

（3）決定の正当性に対する厳格さの欠如

人が命令形の発話を行う場合には、そのどれもが必ずしも他者にとっての判断として、ましてや指示・命令として通用するとは限らない。無視や反抗にあうこともある。だからこそ、たとえば組織は意思決定と上意下達の正当化の機構としての階層を必要とする。そして、たとえば後から見て最善ではなかったと評価されるような判断や命令が、リアルタイムにはそれとして成員に受け入れられ、その言動に作用することもあるのだ。また、リーダーが優秀であることは望ましいとしても、優秀だというだけでリーダーにはなれないのが一般である。つまりその発話が判断や命令として機能するために、満たすべき要件が存在する。

それにもかかわらず、AI の出力が人にとっての判断や命令であるか、また、あり得るかが論じられるとき、どのような正当性のもとでどの人にとってそれが判断や命令であるのかが、議論の焦点となることはなぜか稀である。その出力が客観的事実に反していないという意味で、あるいは人体を傷つけるといった、誰の目にも明らかなダメージをもたらさないという意味で信頼できるものであるかどうか論じられることはある。また、出力を判断・命令として受け入れ、これに順ずることの心理的な可^レ否^クや、抵抗の有^レ無^クが問題とされることもある。けれども、その出力を判断・命令として信頼しなければならないのはなぜなのか、は、それらの議論によって説明されることはない。

人にとって AI の出力が無条件に従うべき命令だということは、人が自身の決めた目的によ

って、この命令から逸脱はできないということを意味する。AI が決めたのと異なる目的を人が決める余地もない。だが、出力をそうした存在として扱うか否かを決めるのは、受け手たる人の側でしかない。そして命令として扱えば、結果として、AI の目的の決定は人が行わなくてもよいと認めることになる。これもまた、「人による決定」を揺るがす要因となる。それも揺るがしているのは、人自身である。

ただ、これは不思議なことである。もしその命題が AI の出力ではなく、人の述べたものであったなら、それが人（々）の言動を方向づけたり制約したりするような判断や命令だと無条件に見なされるのは、考えられないことではないだろうか。AI の出力だからそこを問わないのだとすれば、逆説的にもまずその点で、そもそも AI の出力と人の判断は異なっていることになる¹¹。そのように出力を人の判断と根本的に違ってまで、人の判断と等値しようとする、そう考えると、むしろそこまですることの意味のほうを問う必要はないだろうか。

（４）「発展」「普及」による目的の代替

あるテクノロジーを手段として何らかの事態を自覚的に実現させようとするのと、そのテクノロジーの機能・性能を増大することや、それを普及させることは、行為として別のものであることは言うまでもない。論理的に言えば、前者にとっては後者はそのテクノロジーの目的ではなく、それを目的と位置づけることは自己目的化にすぎない。

それにもかかわらず、そのテクノロジーの研究開発において、上述のいずれもが目的として認識されることが、少なくとも日本ではこれまでもあった。そうした認識によって、後者の意味での目的（自己目的）さえ存在していれば、前者の意味での目的が不在であったり、不明確であったりしても、形式的にはその開発には目的が定まっていることになる。それは、いわば自己目的による本来の目的の代替である。たとえば、1990年代なかばのグループウェアの研究開発がそうであり（福留・刀川，1994）、また、1990年代のテレワーク（のための CNS）についても観察されている（福留，1998）。同様の事態が AI においても生じる可能性を否定する理由はない。

目的のこうした代替は、当該のテクノロジーの研究開発を多くの方向に推進し、多様な機能・性能と、それによる多様な目的の実現を可能にする、という可能性の拡大の面では、社会的機能を有するともいえるのかもしれない。新たなテクノロジーがイノベーションを励起することがしばしば期待されるが、それもこの社会的機能（への期待）への代表的な例である。

しかし一方で「目的の代替」は、本稿が関心をもつ「人による決定」にとっては、これを困難にする要因ともなる。代替によって、前者の意味での、本来の目的が不明確であっても、あるいは存在していなくても、そのこと自体から目を背けることができ、そのままでも研究開発は進展するからである。結果として、前者の意味での目的の決定は、それが人によってなされているか否かが問われる以前の問題として、目的不在のままで決定自体が無限に延期され得る。それは実質的に、人による決定を行わないのと同じである。

さらに、目的不在のままに多様な期待が増大すれば、そのテクノロジーは「なにか素晴らしい（価値のある）目的を達成する手段」ということになるために、前節で述べた神格化が相対

的に容易になることも推測できる。勿論、合理的には、この神格化には実体（となる具体的な目的の対応）がないにもかかわらずの神格化であるから、それが神格化の形骸化や硬直化にも結びつくと考えことに合理性がある。そして、どのような目的を定めるのかの議論を後回しに起きる硬直した神格化は、おそらくそのテクノロジーの崇拜といった観を呈することが想像される¹²。

上では、他のテクノロジー一般の場合と同様に、AI（というテクノロジー）についても、こうした目的の代替と、それによる「人による決定」の無限延期が生じ得ると述べたが、特にAIが置かれている状況は、ここにとどまらない。産業政策上の関心から、AIの研究開発に汎用目的技術を志向させるべきだとの議論があるが、本稿の議論との関係からいえば、これは上述の目的の代替を、一時的にであれ自覚的・明示的に推進することに相当する。

ここで汎用目的技術とは、テクノロジーのうちで、特にこれを活用する多様なテクノロジーを派生させ、それが新たな産業を成立させるもののことを指す。第一次産業革命における蒸気機関、第二次産業革命における内燃機関・電気モーター、第三次産業革命におけるインターネットがこれに当たるとされる。AIのうちでも、目的を特定しない汎用AIが、第四次産業革命をもたらす汎用目的技術として成長する可能性があり、経済・産業政策上、AIをそのように発展させることが望ましいとも論じられている（井上、2016）。

その経済・産業上の効果や意義はともあれ、この文脈ではAIの目的を当面は誰も決定する必要がない。それが将来的にどのようなテクノロジーを派生させるのかも、派生したテクノロジーの目的を何にするのか／何にできるのかも、論理的に定まりようがない。むしろ派生するテクノロジーを早期に限定することは、汎用目的技術としてのAIの可能性の芽を摘むことに当たるのかもしれない。目的はいずれ誰か、あるいは何かが決めるにせよ、成り行きに任されて決まるのにせよ、何でもあり得るままですでにできるだけ長く、多様で多くのリソースを呼び込むほうが、そこから何かを派生させようとする動機を高めることになると考えられるからである。

しかし、そのように考える限りは、人が目的を決定するのかしないのかを、アジェンダとして設定することはできない。そしてそのために、人の決定した目的とは別の効果を波及させるようなAIが出現する“怖れ”も残り続けることになる。そうした意味で、これはAIに誰がどのような目的をどのように決定するのかを自覚的に空白にしようとする企てともなる。

（5）「自由」という価値の硬直性と至上性

現在、AIを自由に研究開発することは、AIの研究開発者はもとより、これを越えた広範囲の人々に支持されているか、少なくとも容認されている価値だといってよいだろう。AIによるイノベーションや、さらにそれによる産業振興・経済成長のために、自由な研究開発が必須であるとか、望ましいといった議論も多い。前項で述べたAIの汎用目的技術化への指向もその一つである。

一方本稿で検討している「人による決定」も、第3節で述べたように、特定の価値と密接に結びつく問題である。その具体的な価値は、人によって決定される（AIの）目的に依存することになるが、つまり、ここで重要なのは、「人による決定」がそれと整合する価値の選択・

決定にほかならないという点である。そして目的の決定がそうした価値の決定でもある以上、決定される「その価値」と「研究開発の自由」という価値の整合性を問題として取り上げることが避けられない。

価値と価値の関係であるから、両者の整合を無条件に予期することはできない。決定する目的次第で、それが研究開発の自由にとって制約になることはいくらでもあり得る。このとき二つの価値は競合関係にあるということになり、それを解消するには、目的を変更するか、自由を変更するか、あるいはその両方が必要になる。ただし、自由な研究開発の名のもとに、研究開発者が自由に設定した（AI の）目的を変更するなら、それは“自由ではない”ことになる。「自由」は、選択肢の多様性を含意する反面、このような意味での硬直性を有している¹³。

そのために、研究開発の自由の尊重は、逆に人が決めようとする目的に強い制約を課す可能性に帰結する。したがって、ここで自由は、人が決める目的が依拠する（体现する）諸価値に対して至上性をもつことにもなる¹⁴。（研究開発の）自由と人が決める目的の依拠する価値が整合しないことが明示的になるときに、変更しやすいのは人の目的の価値であり¹⁵、それはつまり、「人による決定」が譲歩しなければならないということである。

それはまた、（研究開発の）自由を実現することを、開発の究極の目的にするということでもある。そしてそのようにみれば、それ以外の目的は、自由を損なわない限り存在しても構わないが、究極的には存在しなくても開発には差し支えない（自由という目的が存在しているため、開発は目的を見失ったことにはならない）ことになる。それも、「人によって決定される目的」の他の目的による代替の一種となる。

このように検討してゆくと、「人による決定」の自明性は大きく揺らぐ。「人による決定」は同時代人の一部が強く抱く願望であったり、強い規範性を想起させる修辞によって語られることであったりするかもしれないが、顕在的に、また潜在的に、いくつものチャレンジを受けてもいることを認めざるを得ない。そこには（1）のような「人による決定」への明示的な対抗もあるが、それ以外、特に（3）から（5）のような、「人による決定」と併存して語られる言説のなかにも、「人による決定」を揺るがす要因は含まれており、言うまでもなく後者はより認識しづらい。そうした認識しづらい要因により、当事者の気づかないうちに「人による決定」を損なうことも考えられる。

第1節で述べたように、「人による決定」が自明でない限りは、AI への恐れもまた、消し去ることはできない。善意から怖れる必要がないことをいわれても、怖れることのほうが合理的なのである。怖れを払拭するためには、「人による決定」を擁する立場によってであれ、これを揺るがす立場によってであれ、なんらかの形で、揺らぎの要因のそれぞれを取り除いて行くことが必要である。

5 「人による決定」が向き合わなければならないこと

AI の動作する目的を、同時に AI 開発の目的を、ユーザがその善意で決定することは、少な

くとも前節で述べただけの要因によって妨げられたり、代替されたりする脅威に晒されていることを述べた。それは、「AI（自体）による目的の決定」という選択肢を議論の前提から排して（第3節）も、なお、「人による決定」が覚束ないことを示している。AIが決定しないからといって、それだけで人が決定することにはならない。AIの目的の決定は、人によるかAIによるかの二項対立図式で、あたかも人とAIの主導権争いのように論じることのできる問題ではないのである。

ここで問われるのは、改めて、「目的の決定」とは、人にとって何なのかということであるように思われる。第3節で述べたように、これを一応定義することはできるが、前節で見たように、それだけでこれを確実な実行に導けるだけの十分な実効性はない。それならば、目的の実効性のある決定とは、改めて何なのか。

その決定のためには、少なくとも前節で述べた（「人による決定」を揺るがす）どの要因にも絡めとられない必要があるとはいえるとしても、そのために何を解決すればよいのかを、体系的にここで述べることはできない。ただ、これらの要因の複数に共通して欠ける論点、十分な議論のなされていない論点はいくつか存在しているように思われる。そうした論点が存在するのであれば、それを明確にしてゆくことは、「人による決定」を体系的に“防衛”するためにも、その“防衛”の是非や要否を問い直すためにも、必要な作業になるはずである。

以下でそのうちの3点を、断片的ながらも記しておきたい。

（1）決定する「人」とはどの人なのか——目的決定の社会性

「人による決定」の「人」、つまり決定主体は人類一般ではなく、特定の誰かであるはずだということは、前節の（2）でも述べた。しかし、当然といえば当然のこのことが、「人が決めるかAIが決める（自律する）か」と、目的決定の主体として「人」と「AI」を対置する立論では誰が決定者なのかという論点は潜在化する。決定主体としての人は匿名化せざるをえず、さらにAIと対置されることによって、「AIではない人」という意味で、あたかも人類全体やその代表のように表象されてしまう。そしてその結果、そもそも決定が、それを発話する（提案する、言い出す、求める）人と他者との間の交渉によって成立し、決定としての効果をもつのだという、決定という行為の社会性が不可視化される。つまりそのことについて議論・交渉・調整等ができなくなるか、少なくともしづらくなることが懸念される。誰と誰が、どのような利害についてやりとりすればよいのかが特定できず、やりとりを具体的に想定できなくなるからである。

そのように扱いつづらくなる論点の一つとして、実際には発生を避けることができないであろう決定主体間の利害対立（人が決定する目的と目的の間のコンフリクト）の見極めや、相互に譲歩したりさせたりするために決定に課す制約条件のありかたなどが考えられる。勿論、そのやりとりができないからといって、それだけで利害対立が防止できたり自然消滅したりするはずはない。単に潜在化するだけである。

あるいは誰のどのような決定から影響を受けるのが誰で、したがって決定が誰の認定によって受け入れられることが正当なのか、といった論点もあるだろう。また、決定に基づく他者の

行為なしに決定に実効性をもたせることも一般的にはできないことを想起すれば、決定が実効性をもつために、誰のどのような行為の接続を要するかという論点もある。さらに、こうした議論を抜きにすれば、これらにもとづいた決定の手続きの制度化なども、行いようがなくなるのも勿論である。こうした論点はいずれも、決定を公正なものにするための、基本的なそれでもあり、それが扱いづらくなることの問題は大きい。

われわれの社会においては、決定一般について誰がそれを行うのかが、その結果からなんらかの影響を受ける関係者にとって重大な関心事になる。つまり誰の行う決定も、他者にその影響が及ぶかぎり、無条件に自明に受け入れるべきこととはとらえられない。このことには、議論の必要はないだろう。AI の目的、つまり AI を手段として何を行うのかを決定するに際しても、その結果が人々に及ぶならば、この原則が損なわれる理由はない。誰かがそれを決定するなら、その人と対立する利害をもつ別の人は、当然のことながらその誰かの決定を妨げたり、協力を拒んだりしようとするだろう。しかし、それは「人による決定」の一つが損なわれることでもある。また、その結果として、人による決定が長期にわたって不在となったり、成り行きに任されるのと等しくなったり、あるいは「誰（どの人）によるのでもない決定」として、AI の出力を決定と見做すことが待望されることも考えられる。これは「人による目的の決定」の困難である。当然のことではあるのだが、しかし、「誰が」決定するのかということ自体を問題にしなければ、こうした決定の妨げを具体的に問題とすることはできなくなる。

逆に考えれば、人による決定が妨げられず、決定として他者に受け入れられるためには、満たすべき条件があることを想起できる。たとえばそうした実行性の条件の一つは、決定主体および決定手続きの特定と、関係者によるその共有である。そこで決定しようとする事項を特定せずに、いわば全権委任を想定したような決定手続きを設けようとすることは例外的である。それならば、「AI の目的の決定」に関しても、その目的が何を達成しようとするものなのか、それぞれに応じた決定手続きを定める必要があるはずである。われわれの社会に「責任能力」や「(決定) 権限」という概念があることが示すように、決定はそれを決定として受け入れる相手があって初めて実効性をもつ。そこが欠ければ、単なる「命令形の発話」に過ぎない。

目的の決定における人と AI の対置は、こうした認識と決定の条件の問題化を、図らずも不問に付すように作用する。これを防ぐという意味で、「人による目的の決定」を実現するために議論すべきは、その決定をするのが人か AI かではなく、誰であって誰でないのか、という点である。言わずもがなながら、「(目的の) 決定」は、人の関係性において生起する、つまり社会的な事象である。「人による決定」のためには、その社会性を視野のなかに回復し、決定手続きを含む決定の制度を（おそらくは決定を想定する目的ごとに）構築することが求められる。

（２）代替による決定の無効化

この決定の社会性を視野の内に回復することをもってしても、「決定」が容易にそれと整合しない別の決定によって覆されることになれば、「人による決定」の実効性は高められない。その懸念を呼び起こすのが、前節の、特に（４）と（５）で述べた「目的の代替」である。

目的の代替が生じ得るとき、そこには2つのレベルの目的が併存している。一つは、開発時点でそのAIの受益者が特定されるような目的である。たとえば従来人が目視によって行ってきた、X線写真による病理診断の一部あるいは全部をAIによって行わせる、つまりそのAIの目的とするといった場合の「目的」が、これにあたる。直接の受益者として想定されるのは、診断する人（医師等）とされる人（患者等）であり、間接的には診断にかかわるサービスを提供する組織（の関係者）が受益者に含まれる。その外部には、益を受けることのない人々の存在が想定される。そのなかに、目的によっては、むしろ損害を被る人々も生じ得る。

もう一つは、このような受益者が特定されない、あるいは明確には特定されない目的である。AIの「発展」「普及」や「自由な開発」はこれにあたる。現存しないAIの実現や、AIの普及範囲の拡大によって、現在はそうしたAIを自らの考える目的にとっての手段と想定し得ない人々がこれを手段と想定するようになった場合、そこでその当事者が受益者になる可能性は勿論存在する。けれども開発時点、普及以前の時点で、それが誰であり、どのような「益」なのかは特定できない。ただ、このレベルの目的には、例外が一つある。それは開発それ自体という「目的」である。自らの構想したAIの開発を望む人がそれを開発することは、その人にとっての「益」であり、これは開発時点、あるいはこの人を開発以前に、例外的に受益者として特定することはできる。

この例外的な目的を決定する開発者をのぞき、それ以外の「人」にとっては、二つの目的はその意味を大きく異にしている。二つめのレベルの目的は、将来的に一つめのレベル目的を発生させるかもしれない。X線写真による病理診断とは別の人々に特定の益をもたらすAIの応用がなされるかもしれないし、そのような益と受益者の多様性は、他の条件が同じであれば、AIの動作の多様性が増すほど大きなものになると推測できる。

その意味では、第二のレベルの目的も、第一のレベルの目的の実現に資する、つまりは第一レベルの目的の手段であり、第二レベルの目的は第一レベルの目的をも潜在的には内包する、ともいえる。ただ、内包はあくまでも潜在的である点が重要である。開発時点で、それが誰にどのような益をもたらすことなのか、つまり益と受益者の特定ができない。特定の益を期待していたのとは別の人に別の益をもたらし、それが前者と利益相反を生じるかもしれないのである。これは（1）で述べた決定の社会性の問題ともかかわる。第二レベルの目的は、そこに潜在的に内包される目的の具体的な利害関係者が誰なのかを特定し得ないために、上述の意味で社会的に扱うことができないし、その結果、成り行き任せや「AIによる決定」の待望を生じること否定できない。

こうした異なるレベルの目的が、しかし一方で、いずれも「AIの目的」と理解されている。また、「人によるAIの目的の決定」において、前節で述べたように無造作に代替されている。それが「成り行き任せ」に帰結する可能性を擁している以上、「人による目的の決定」に実効性をもたせるためには、目的の個々の代替の是非を問い、目的の外延をも「決定」することが欠かせないはずである。

(3) 擬人化によって不可視化されるものと可視化されるもの

目的の決定や決定主体について考えるに際して、AI には他のテクノロジーにない、擬人化という特異な事情が見られることも意味をもつ。たしかに、AI の出力がしばしば人の発話や動作と見分けにくいのは、自然言語で出力を行っているなど、形式的に両者を似せていることにもよっている。ただ、そうした出力を人のふるまいと分かちがたくする要因はそればかりではない。AI を擬人化するという慣例的な修辞によって、われわれは口述・記述のうえで、こうした出力と人のふるまいを、しばしば同じように扱おうとする。それによって、出力とふるまいは、表現形式の面ばかりでなく、認識上の取り扱い方の面でも、重ねて類似性を高めることになる。つまり、両者は一層似て見えることになる。

こうした擬人化の作用は、他のテクノロジー一般には想定しにくい、AI に特異な事情である。したがって、AI の出力は他のテクノロジーの動作よりも、人のふるまいに似たものとして認識されやすくなっていることが考えられる。またその認識傾向は、AI 開発が「人のような知性・学習・行動をもつ機械の実現」を目標とすることや、その AI の実現から産業振興や経済成長を派生させようを目指すことは、AI の出力と人のふるまいがより類似することを期待して望むことにより、助長されるかもしれない。そう考えることにも合理性がある。

すると、その反面、AI の出力と人のふるまいの差異は、見落とされたり過小評価されたりしかねない。認識のバイアスを抑えるためにも、重要な差異の見落としを防ぐためにも、AI の擬人化に改めて慎重になることが必要であるのに加え、擬人化によって既に見落としている重要な差異がないのか、再考する必要もあるだろう。差異の見落とし（＝類似性の過大評価）は、たとえば、AI の目的についての「人（全般・人類）による決定」対「AI による決定（AI の自律）」という対立図式を強化することや、それによって本節（1）で述べた決定主体の特定性の希薄化や、決定の社会性の希薄化にも接続している理論的な可能性がある。

一方、こうした問題設定自体が、擬人化という修辞に支えられている側面もある。既に述べたように擬人化によって、AI の出力が AI の目的になっている（＝AI が自律している）と認識したり、想像したりする傾向は強化されると考えられるが、そのことが同時に、その目的を（AI が決定しているのだから）人は（誰も）決定していないということを可視化し、「人によるのでない決定」が実現しているとか、実現するかもしれないとかいったことがそこに現実感をもつ。その現実感を前提としてはじめて、本稿のテーマとかかわりの深い、「人が決定すること」の可否、可否、成否といったことが問題化できるようになる。

つまり擬人化が、人以外の決定主体の現実感を介して、「人の決定」の可否、可否、是非といった論点をアジェンダとして設定する。この意味で、この問題設定は擬人化に支えられている。目的の決定主体を空白にしておくこと自体は、AI 以外のテクノロジーにおいても生じ得る。ただ、擬人化が「AI という決定主体」を想起させることによって、人が（誰も）決定しないという事態、すなわち決定の空洞化を、われわれが認識しやすくなる。そしてこれを認識すれば、そこから逆照射するように、他のテクノロジーにおける「人による決定」の可否を認識しなおすことも相対的に容易になる。それは、AI に限定しないテクノロジーの目的の決定一般において、その決定主体とその決定による利益は明確であるのかどうかを、見直す機会を

われわれに与える¹⁶。

近代以降の社会においては、目的が合理性の強力な根拠となる、また根拠とされるべきだとの通念はあるが、これに反して少なくともテクノロジーがかかわるとき、その目的の決定は実際にはこれほどに不確実である。AI 第三次ブームが先鋭化したのは、テクノロジーとの関係におけるわれわれの社会の目的決定をめぐる議論の現状と、その困難なのである。

ハードウェアの性能向上とコネクショニスト・モデル、深層学習の技術展開（いわゆる「理系分野」）、そして「人のような知性・学習・行動」という研究開発の目標設定（設定は理系分野を主な主体とした行為だが、社会的事象）のうえに、現在の AI が実現（理系）し、擬人化というレトリック（文系）が「人による決定」を揺るがすと同時に問題として明示化して、われわれにとって「決定する」とはなにかを問い返してくる（社会的事象であり、その言語化は文系分野を主な主体とした行為）。怖れを克服し、怖れる必要のないテクノロジーとの関係を合理的に獲得しようとする考察の先に出現するのは、「文理融合」よりもむしろ溶解してしまう文と理の境界である。

この溶解をひきうけて、われわれは先に進まねばならないのだろう。「文理融合」や「文理統合」のなかで考え得る問題、考える問題は一つとは限らないが、いずれにせよ、文と理の境界を溶解させるような問題をそこから排除してはならない。テクノロジーの目的やその決定の意味、「テクノロジーは手段である」というような常套句の含意や実効性の如何への問いが、その根底において擁されていないなければならない。

6 おわりに

「人による決定」によってわれわれが何を企図し、どこへ行こうとするのか。納得できるための議論の条件は何であって、そこで検討されるべき項目には何と何があるのか。それを全て明らかにするには、本稿は、いうまでもなく、あまりに多くの検討課題を残している。たとえば、第4節に挙げた要因の網羅性や体系性、第5節で述べた（1）から（3）相互の関連性など、議論が十分に尽くせていない点も多い。

それにもかかわらず筆を進めたのは、どのような形であっても、「（テクノロジーの目的の）人による決定」、換言して「テクノロジーを道具とすること」が何を意味しているのかが、決して自明ではないということを早く共有し、これをふまえてテクノロジーの目的の決定を再構築することは、おそらくは AI や他のテクノロジーと社会の関係を考えるさまざまな場面で必要である、そう考えたことによる。不十分な点はさらに検討を重ねてゆきたい。

また、AI と社会との関係を包括的に論じることが大括りに過ぎて、実質的でないとの見解¹⁷にも、耳を傾けるべき点が多い。実践上の有用さや有効性に照らして、どのように問題を分割することが適切であるかということも、理論化に残された課題である。

ただその一方で、この問題と対応策の完全な体系化は、そもそもきわめて見通し難いものでもある。開発が新たなものを生み出す性格を備えている以上、その過程で生じる決定という行為について、これを揺るがす作用の一つ一つを完全に予見することはできない。そのため、そ

こには開発や使用の現場で、それぞれ個別具体的に退けたり回避したりすることによって、「人による決定」の実現可能性を高めてゆくしかない、そうした部分が残るはずである。その意味で、「目的の決定」とは、そこから AI の活用や開発を始めるスタートラインを引くことではなく、そのスタートラインを引くために、またそれを無自覚に揺るがさないために、これを妨げるいくつかの行為、目的を決定しない／させないための行為に対抗するという行為の連続ととらえるしかない。

そしてそれならば、そうした連続的な行為は最終的にはその開発や使用の現場にある一人一人に託すことしかできない。勿論、著者自身もこれを引き受けなければならない一人である。AI の目的決定の要件を網羅的に体系化することは、その実践の助けになるだろうが、それが未完成であることを言い訳に上述の対抗を延期することは、少なくとも人による目的の決定にとっては適切ではないと思う。

引用文献・参考文献

- 江間有沙, 2018「人口知能社会のあるべき姿を求めて 特集にあたって」科学技術社会論研究第16号巻頭言, pp.9-14
- 福留恵子, 2011「技術とその社会的文脈—テレワークを事例として—」東海大学総合教育センター紀要第31号, pp.59-72
- 福留恵子, 1998「テレワークが拓げるワークスタイルの可能性」第6回情報文化学会発表予稿集, pp.54-57
- 福留恵子, 1997「組織にとって自由なコミュニケーションがもつ意味——情報と認知のマネジメント——」『組織科学』31 (1), pp.37-50
- 福留恵子・刀川眞, 1994「グループウェアが対象とする問題の構造分析」情報処理学会第48回全国大会予稿集, pp.141-142
- 福岡幸太郎・越川智瑛・矢野摂士・鈴木遊哉, 2021「文理融合、問われる本気度 科学立国背水の5年 ⑤」日本経済新聞朝刊2021年3月8日
- 伊藤智洋, 2016『人工知能と経済の未来』文藝春秋
- 人工知能学会倫理委員会, 2017a「人工知能学会倫理指針」, <https://www.ai-gakkai.or.jp/ai-elsi/wp-content/uploads/sites/19/2017/02/%E4%BA%BA%E5%B7%A5%E7%9F%A5%E8%83%BD%E5%AD%A6%E4%BC%9A%E5%80%AB%E7%90%86%E6%8C%87%E9%87%9D.pdf>
- 人工知能学会倫理委員会, 2017b「【開催報告（詳細版）】2017年度人工知能学会全国大会「公開討論：人工知能学会 倫理委員会」（2017/5/24）」, <https://www.ai-gakkai.or.jp/ai-elsi/archives/583>
- 神崎宣次, 2018「ロボットや人工知能に関する社会的議論の現状について」『科学技術社会論研究』第16号, pp.134-141
- Luhmann, Niklas, 1986, *Zweckbegriff und Systemrationalität: Über die Funktion von Zwecken in sozialen Systemen*, Tübingen: J.C.B. Mohr (=1990『目的概念とシステム合理性』馬場靖雄・上村隆広訳, 勁草書房)

松尾豊, 2015『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

文部科学省編, 2021『科学技術・イノベーション白書 令和3年版 Society5.0の実現に向けて』インパルスコーポレーション

中島秀之, 2021「人工知能技術と社会」(藤山知彦=編著 吉川弘之・日本産学フォーラム=監修『規範としての民主主義・市場原理・科学技術』第12講, 東京大学出版会), pp.257-281

酒井邦嘉, 2022「第1章 脳とAI 解説—人間と機械」(酒井編著 合原一幸, 辻子美保子, 鶴岡慶雅,

羽生善治, 福井直樹『脳とAI—言語と思考へのアプローチ』中央公論新社 所収), pp.64-69

第6期科学技術・イノベーション第6期計画 (令和3年=2021年3月26日閣議決定)

注

1 AIの第三次ブームについては、松尾(2015)。

2 分野横断的な探究を営みあるものにすることの困難を挙げた例は、枚挙に暇がない。さらに、第6期科学技術・イノベーション基本計画にうたわれる「総合知」についても、異分野をどのように融合するかに困難な課題が残ることが指摘される。たとえば、「(総合知は、)学問としての性質が大きく異なる自然科学と人文科学を機械的に組み合わせた印象がある」「人文科学は、(…)便利な道具のように使えるものではない」など、専門家の慎重な見方を受けた、領域横断のありかたについての「アカデミア全体での共通認識」の未達成の指摘(福岡・越川・矢野・鈴木, 2021)など。

3 この三者は、CNSの規模等に応じて、部分的に、あるいは三者の役割全体を、同じ者が担うことも考えられる。そのため、本稿において、設計・製造(プログラミング)・運用を特に区別する必要のない箇所では、これを「(CNS/AI)エンジニア」と総称する。

4 このことは、手段の「神格化」の語で、N.ルーマン(1986=1990)が精緻に論じている。また、手段の神格化とテクノロジーとの関係については、福留(2011)。

5 江間はこの「特化型AIについて」のことに限定して述べており、それと別に汎用AIを想定することを否定しているわけではないように思われる。本稿では検討対象を特化型AIに限定して(第3節)考察しているため、ここで参照するのも、江間の議論のうちの特化型AIに限定して述べられている部分(のみ)である。

6 その趣旨を中島は「AIは人間の道具なのだと思います。」と説明している。(中島, 前掲)

7 AIのいわゆる自律(性)、すなわち人が目的設定(AIを擬人化するなら、「指示」や「命令」)しなくとも動作することは想定される事態の一つだが、そこにはさらに、その動作が①人の意を汲んだ、「気の利いた」ものである場合、②人にとって望ましくないものである場合、③人にとってメリットもデメリットもない中立的なものであるが、その影響や波及効果が理解できるものである場合、④何の意味があるのかわからないのが理解できない場合、などが含まれる。また、そのAIの動作に対して、人間がどのように応じるか、人間をどのように応じさせるかにも、多様な事態が含まれる。一方、「将来、我々が進んでそのような方向に開発を進めれば、AIの判断に人間が従うということも起こりうるでしょう。」(中島, 前掲)と述べられるのも、ここに起こり得る事態の一つである。さらに、その「人間が従う」という事態にも細分化すれば、①AIの判断に人が従う(人の行為の目的をAIが決定する)という以外に、さまざまな状況が含まれる。たとえば、②AIの出力に基づき、人がさらなる判断をする事態。そこにもさらに、②'AIの出力を決定的な根拠として判断する場合もあれば、②''出力を仮説の一つとして検証を経た上で判断する場合もある。それぞれに、また、人のそうした判断の結果が、他者に及ぶ場合も及ばぬ場合もある。③AIの出力を人が無視する場合もある。AIについても、④AIがその「判断」を最終出力として停止する場合もあれば、⑤その「判断」にもとづいて、AIがさらに継続的に動作する場合もある。その動作が誰にどのように影響するかについても、また、多様である。いずれもが、「人による決定」に非ざる事態として考えられ得る。

8 中島自身は、そうした開発をすべきではないと述べる。(中島, 前掲)

AI 第三次ブームが先鋭化させる「テクノロジーと社会」の論点

9 記録によればこの発言は「議論も尽きないと思うのですが」と引き取られ、「倫理指針を出すまでは本当に大変で、(…)でも、修正をして苦勞をしたところを評価いただいて、(…)」とまとめられている（人工知能学会倫理委員会、前掲）。

10 勿論これは非合理的だが、なにごとにらず、非合理的な決定が起きることはある。だから、それを避けようとするなら、避けるための方策が必要なのだと思う。

11 AI の出力を人の判断や命令と容易に等値することができるのは、AI を擬人化しがちな修辞によるところもあるのかもしれない。人の言葉に対してであれば要求される正当性を、なぜか AI に求めることのない「AI の（人とは異なるものとしての）特別扱い」が、AI の擬人化を循環的に強化しているとも考えられる。つまり人との差異化（人にはなされない特別扱い）が擬人化を支えているのである。これもまた逆説である。

12 これは推測の域を出るものではないが、テクノロジー崇拜を成立させるメカニズムについての一つの仮説でもある。

13 ここで、「自由の硬直性」は、組織のマネジメントとコミュニケーションに関して論じたもの（福留，1997）を一般化している。同稿では、自由のこの硬直性ゆえに、コミュニケーションの自由が、組織マネジメントにおける他の目的に対して、手段となし得ないことを論じている。

14 だからといって、これを研究開発の自由に短絡する必然性はない。本稿で議論を十分に展開することはできないが、自由という概念は、無制約と同義のものとしてふりかざされるよりも、むしろ「制約の制約」をも含めた慎重な制約を、検討されてきた歴史をもつように思われる。研究開発の自由にも、同様の検討を行いつつ尊重する、必要性和余地があるのではないだろうか。

15 その変更のしかたは一つとは限らないが、たとえば、（目的の）決定者を（研究開発の自由に整合的な価値の賛同者へと）変更するのも一つのやりかたになる。すると、目的の内実を変更しつつ、「人による決定」であること自体は維持することができる。

16 決定という行為を媒介にすると、AI と他のテクノロジーが異質なものと思われやすい原因が、むしろ決定という行為のありかたや決定という概念の側にあることも考えられる。そこに、たとえば（AI の）自律性と呼応する（人の）他律指向性が存在してはいないか、といった検討課題が残る。

17 たとえば神崎（2018）。