

AI と知恵の自動化

シャノン・ヴァラー

出典

Shannon, Vallor, 2017: “AI and the Automation of Wisdom”, *Philosophy and Computing: Essays in Epistemology, Philosophy of Mind, Logic, and Ethics*, Thomas, M. Powers ed., Springer, pp. 161-178.

凡例

原文において、イタリックで表記されている語の訳語には傍点を加えた。また本文中の“ ”は「 」で示し、[] は紹介者が読みやすさのために補ったものである。() は言語表記のために用いた。また、重要箇所には紹介者の判断で下線を引いた。

要旨

本稿は、ロボット開発、機械学習、そしてコンピューターによる自動化の進展によって、人間の知恵 (wisdom) が直面することとなった3つの課題について明らかにする。本稿では、知恵を、総合的な価値判断に基づいた責任のある自己規制 (*responsible self-regulation*) の習慣によって豊かになった、道徳的あるいは知的な専門性 (*moral or intellectual expertise*) として捉えるが、本稿は、将来の労働現場においては、機械による専門性が、人間の専門性や知恵の、好都合で魅力的な代替品として見なされることになるだろう、ということを指摘する。さらに、自動化は、人間の専門性や知恵を開花させる新しく豊かな領域を作り上げてきたのだが、現在の技術的・政治的・経済的条件が、この歴史的に見出される潮流を途絶えさせてしまうかもしれない、ということも指摘する。最後に、これらの技術的・政治的・経済的諸条件を問題にし、これらの諸条件の結果に対する責任を想定しない限り、我々は、AI や自動化は、人間の道徳的で知的な完全性を十分に支援し、人間の知恵の育成に役立つものであるという広大な肯定的可能性を、無駄にしてしまう危険性があると結論する。

1. 自動化の台頭と人間の知恵に与えられた課題

2013年の研究によると、43%のアメリカ人の職が今後20年の内に自動化されると言われている。この研究によって、多くの経済学者、科学技術者、政策立案者、哲学者などはいわゆる「仕事の終わり」について熟考することになった。仕事の終わりとは、厳密に言

うと、失業率の異常な高さによって特徴づけられる未来のことである。

この展望は、いくつかの差し迫った哲学的問題を提示する。しかしながら、ここでは以下の展望と関連する、限られた哲学的考察に焦点を当てることにする。つまり、近年のコンピュータによる自動化の動向が、道徳的そして知的な徳としての人間の知恵の実践にとって代わるものである一方で、それは同時に、我々の最大の責任としての人間的知恵の実践を發揮させるものでもあるという側面に関連する考察である。

第2節¹では、「知恵」とは何かについて説明する。この概念は複雑で論争的な概念であるはあるが、本稿の見解に関係する概念としての知恵を規定する。他の残りの節（3節、4節、5節）では、機械知能（machine intelligence）や自動化の台頭によって知恵に与えられた、主要な3つの課題（three distinct challenges）を確認する。

第3節で以下の第1の課題について議論する。すなわち、機械学習とモバイルロボティクス工学のための、新しいアルゴリズム技術の進歩によってもたらされた近年のコンピュータによる自動化の動向は、歴史的に人間的知恵が社会の道徳的で知的な秩序の内担ってきた、多くの役割を次第に奪ってしまう恐れがあり、また、このような社会の道徳的で知的な秩序の断絶は、長期的にみれば人間の知恵を育む新しく豊かな文脈を形成する傾向にあったのだが、このような歴史的な傾向が途絶えてしまう危険がある、という点について議論する。さらに、私は、人間の知恵と機械知能との間に重要な区別をもうけ、現在の主流の経済的・政治的価値観に基づく新たな労働実践において、人間の知恵という徳を陶冶することに対する主要な障壁も同定する。

2つ目の課題（第4節）は、ますます機械知能に適した形で建設され、行為者である人間と機械の両者の絡み合いによって形作られた世界における、責任の概念に関わる。人間はAIに対して、どのような道徳的で知的な責任を持つのか、もしくは取れるのか、などの問いが提示される。ここにおいて私は、責任があること（being responsible）—明確な因果的関係を伴う形而上学的状況—と、責任のある状態にあること（being responsibly）—知恵であるところの有徳な状態の道徳的構成要素—を、本質的に区別する。

3つ目の最後の課題（第5節）は、人間の知恵をテクノソーシャルな徳（technosocial virtue）として救い出すというものである。この課題を果たすためには、機械によって一層実現されつつあるパフォーマンスを超えたところにある、人間の知恵に固有な機能を理解する必要がある。私は、行為者性や完全性や尊厳といった近代的な概念だけでなく、古典的な徳の伝統における自己修養（self-cultivating）や自己規制（self-regulation）の概念を引用しながら、世界を機械知能に適したものに作り直そうと急いでいる現在において、我々はどのようにして世界を知恵に適したように作り出すこともまた可能になるのだろうか、

¹ 本論では節番号が一つずれ「第1節（Sect. 8. 1）」と表記されているが、論文全体を検討すると誤植と考えられるため、紹介者の責任で訂正し「第2節」とする。

という問いに取り組むこととする。

2. 知恵と知性: 認識の梯子を登る

本節では、知恵と関係するいくつかの概念を区別し、それらの概念を知恵と対比するために、知恵の異なる形式を大まかに区別する。というのもこれらの区別は、知恵と機械知能との関係をより注意深く検討するのに役立つためである。したがって、知恵と同種のもの、例えば知識 (knowledge)、知能 (intelligence)、そして知性 (understanding) の区別について検討する。すると、以下のようになる。

知識 (knowledge) : 慣例的に、知識は、命題の真理値に関する信念が正当化された状態と定義される。知恵と異なり、知識は何ら特別な閾値を前提としない。たとえば、愚か者 (fool) は知恵ある者 (wise) ではないにもかかわらず、何らかの事柄は知っている。

知能 (intelligence) : 知識を何らかの関連する文脈や諸文脈に適用する能力。知能は様々なタイプに分割されうる。一例を挙げると、不安を和らげるために感情に関する知識あるいは心理学的な知識を用いる際に、ある知能は示されうる。

知性 (understanding) : 実践的あるいは理論的な認知についての少なくとも一つの領域に関する、[分割されない] 統合された熟達を前提とする。知性とは、ある領域に関する諸対象の大規模な構造とそれらの対象の間にある関係を、わずかであつても理解することを要求し、その領域において知識を用いる際の一般的能力を促進するために、知性は知恵の概念にもっとも近似していると言える。

この三つの概念を梯子のそれぞれの段として捉えるならば、知恵は最後の段階である (図 1)。

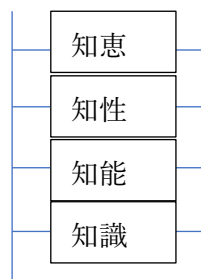


図 1

知恵を除く、三つの概念は、人間の内的な認識的状态であり、これらは認識論的に世界の何

らかの領域に沿う形で作用するが、必ずしもその領域を作り変えたり、その領域に影響を与えたりしない。たとえば知能は、行為の際に知識をうまく展開させる能力を前提とするが、一度この能力を獲得してしまえば、原理的には所与の行為者の内に潜在的にとどまるし、この能力を保持し続けるために、必ずしも知能を行使する必要もない。これに対して賢さ (cleverness) と専門性 (expertise) という言葉は、熟練した仕方で、また、世界を作り変えるような仕方で知能を用いるという、行為者の実際の実践 (practice) を表す²。たとえば、賢い人は、自身にとって主要な目的を確実に達成することに成功するような仕方で、知能を用いる (exercise intelligence)。賢い銀行強盗は、常に逮捕されないように注意しながら、奪えるお金を最大化するために、銀行強盗に関する自身の知識を用いる (use knowledge)。他方、専門家 (expert) は、ある社会的集団によって定義された規範的な基準を満たしたり、超え出たり、また目的を達成したりすることに確実に成功するような仕方で、知能および知性を用いる (deploy intelligent and understanding)。ここでの社会的集団とは、工芸、スポーツ、そして学問的な分野における集団を指す。

最後に知恵を明らかにする。そのために、知恵と専門性の違いについて説明する。スワートウッドは、両者の説明をする際に、RPD (recognition-primed decision 認識主導意思決定) モデルとして知られている専門家の技術のモデルを、知恵の基礎的な構造として用いている。このモデルによると、専門家は以下の能力を行使する。

² この「賢さ」と「専門性」という概念の位置づけは幾分曖昧であるように見えるため、少々煩瑣になるが、本論に即して詳しく紹介しておく。本論文によれば、この両者は、「熟考における卓越性 (excellence in deliberation)」と「決定における卓越性 (excellence in decision)」を前提としているとされ、前者の卓越性とは、目的に対して取り得る最善の方法をうまく査定する能力であり、後者の卓越性とは、熟考の結果を確実に行為にもたらす効果的な動機力 (effective motivational force) のことであるとされる。さらに、賢さと違い、専門性は、「判断における卓越性 (excellence in judgement)」も前提としているとされ、この卓越性がアリストテレス的フロネーシス (phronesis) を意味する場合には、実践的知恵を意味することになると説明されている。この判断という要素が両者を分かつ際には重要であり、専門家とは、自身の専門性を評価する集団によって究極的に価値があると定義された目的の中に近接目的 (proximal ends) を並び入れる (align with) ために、良き判断を実践するが、他方、単に賢い者は、他の価値に関係させて価値を査定することがなく、近接目的の充足を巧妙に追求するだけであるため、このような判断を実践することなしに振る舞うと述べられている。以上をまとめると、「賢さ」も「専門性」も「熟考における卓越性」と「決定における卓越性」を備えているが、後者は「判断における卓越性」をも備えており、近接目的の単なる充足だけでなく、集団的に評価された究極的な目的に照らしてそれを査定し、良き判断を実践することも要求されているといえるだろう。(pp. 164-165. 参照)

- ・直観 (intuition) : 現前し、また、重要な点において類似するある文脈に関連させて、過去の成功例に関連する記憶のうちに蓄えられたパターンを再認識する (recognize)。
- ・熟慮 (deliberation) : 新しい状況において、意識的に問題解決を行う。これは、実践的な領域における正確で抽象的な諸表象を、予期しない状況や問題に適應することに関わる。
- ・シミュレーション (simulation) : 直観された一連の行為 (intuited courses of actions) を想像的にテストする。
- ・メタ認知 (meta-cognition) : 直観され熟慮された解決策のメリットを比較し、決定する。

しかしながら、これらの要素だけでは、知恵の説明としては不十分である。ソフトウェアは、これらに、さらに以下の二つの要素を付加している。第一の要素は、アリストテレスのフロネーシスに類似した要素である。つまり、どのように生きるべきかに関する総合的な価値判断を下す卓越性 (excellence in making *holistic* value judgments concerning *how to live*) である。第二の要素は、これらの判断に照らして、自身を規制する能力 (the capacity to *regulate oneself* in light of those judgements) である。彼は以下のように述べている。すなわち、人々にとって、知恵は、何がなされるべきなのか、そして、それはなぜなされるべきなのかということを理解する卓越性を単に含意するだけではなく、ある人がその行為を実際に行うことへの心理的障害を理解し克服する卓越性もまた含意するのだ。

知恵において、付加された二つの要素は、責任のある自己規制という1つの概念にまとめることができる。責任のある自己規制とは、生の究極的な目的あるいは諸目的およびそれらに関連する価値についての、常に進行中である総合的な判断に照らして、自身を陶冶する反省的な習慣である。専門性と異なり、この責任のある自己規制の実践とは、ある特定の共同体や集団の慣例的な基準以上のもの、つまり生に関する総合的な目的や価値によって導かれている。また、その基準は唯我論的なものではない。というのも、ロバート・リードによると、知恵は「自己懐疑」の能力を必要としており、それは少なくとも、他者の眼前における「自己の再定義 (redefinition)」という事態に開かれており、この他者はどんな時でも我々に「どのように生きる選択をしているのか」という問いを投げかけてくるからである。さらに、ここでは、価値についての多元的な説明が想定されている。つまり、幸福とは、ある単一の実践的あるいは理論的な目的に分解されるのではなく、多くのしばしば互いに両立しない、それでいながら開花繁栄した生 (a flourishing life) を生み出しうる、諸目的を包含するのである。そして、ただ知恵のみが、1) 人生の所与の段階において、我々の道徳的および知的活動はどの価値によって導かれるべきなのかをよく判断し、

2) この判断のもつ創造的な側面について責任をとるといふ恒常的な実践を伴うのである。なぜ責任が伴うのかといえば、諸目的を能動的に選択する場合であれ、それらの目的が持つ動機づける力に単純に反応する場合であれ、我々は、我々の生き方を規制させることにした総体的な価値に対して、責任を持ち続けているからである。知恵ある人 (The wise person) は、この責任を積極的に引き受けるのだ。³

3. 人間の知恵をアルゴリズムに置き換える

知恵についてのここまでの説明は多くの重要な問いを提起する。例えば、機械はどこまで認識の梯子を登れるのか。機械は知能をもてるのか。あるいは、知恵をもてるのか、といったことである。

今後数十年の間に、データ分析のためのディープ・ラーニング技術やパターン認識やシミュレーションを、現実の世界の問題解決と行為へと適応させることで、人工知能は知能だけではなく「専門性」をも持つと、ますます見えることになる (*appear to be*) だろう。しかし、RPD モデルにおいて示されたように、確固とした (robust) 専門性は熟慮やメタ認知能力も内包し、後者の領域において AI は欠陥を抱えているため、AI がこのような能力を獲得するには時間がかかるはずだ。しかしながら、原理的には、機械に対して、熟考やメタ認知の能力を発展させることを阻害する、明確な障壁は存在しない。中長期的に見て、機械が専門性をもつというのは、ありえない光景ではない。

このような発展は、人間の知恵に対してどのような結果をもたらすのだろうか。このことを考えるにあたって、以下の二つのことを確認しなければならない。1) 人間は、機械とは異なり、複雑で多様で重なり合う社会的集団の一員として生きており、その集団は、価値についての多様な総体的諸構想の間で、人々を互いに発展させ交渉させるというこ

³ この「知恵ある者」と「責任」という論点は、第4節において再び詳しく触れられることになる。また、この第2節では、知恵が、最善の生に関する総体的な価値に関わるものであることが強調され、知恵とは、政治的知恵、科学的知恵、芸術的知恵、教育的知恵、技術的知恵といった仕方で領域によって異なったもの (*domain-specific*) であるものの、これらの知恵は、あらゆる点を考慮に入れて (*all-things-considered*)、最善な生き方に関する広範な一連の価値に従って専門家を導く能力であるという意味では共通していると考えられている。例えば、政治的知恵とは、最善の生き方に関する現在の総体的な判断に照らして「責任のある自己規制」という責任を伴いながら、あらゆる点を考慮に入れた上で(言い換えると、人間以外の自然の内在的価値といった、政治的でない価値も考慮に入れた上で)、どの価値が政治的専門家を導くべきなのかということ判断する能力である。(p. 166. 参照)

と。2) いかなる所与の道徳的で知的な領域においても、専門性あるパフォーマンスを首尾よく洗練させ維持させるために、人間は、人間に備わった認知的かつ感情的な脆弱性によって、絶えず自己規制の技術を実践し、その技術をより優れたものにするよう要求されているということ。以上のことは、どのような専門家集団においても、単に人間であるという社会的かつ心理学的事実によって、ある程度知恵を陶冶するよう導かれた人々が一定数いるだろう、ということの意味している。反対に、機械知能は、知恵を可能にする特徴的な社会的そして心理学的条件が欠けている。事実、この欠落が、それらの強い競争力の核心、商業的な魅力なのであり、機械のための知恵というのは、少なくとも予見可能な未来において、不可能であるのみならず、不必要なのである。従って、我々は以下の点について考えなければならない。すなわち、我々は認識や行為に実践的に関わることによって、歴史的に知恵の徳を発展させたり発揮したりしてきたのだが、このような関わりにおいて徐々に人間に取って代わろうとする、専門性をもった機械 (machine experts) の世界が生じる可能性や、これに付随して生じる代償や利益についてである。

人間の「仕事の終わり」についての警鐘によれば、機械によって、多くの領域における人間の専門性は、まもなく、無用で不適格なものになるだろう。例えば、IBM社の人工知能であるワトソン (Watson) は、メモリアル・スローン・ケタリングがんセンター

(Memorial Sloan Kettering Cancer Center) において、既のがん治療の診断を提供する用途で用いられている。MIT教授であるアンドリュー・マカフィー (Andrew McAfee) は、ワトソンはすぐに「世界一の診断医」になるだろうと述べている。このようなシステムは、麻酔学のように人為的ミス傾向があるとされてきた多くの重要な作業において、人間にとってかわり、人間をしのぐと予測されている。他にも、大きな法律事務所では、今まで弁護士補助員が行ってきた仕事を、Symantec社のClearwellといったソフトウェアシステムによって代替している。また、大学では、AIのチューターが各学生のニーズにあった専門的な教育を提供するだろうとする報告書もある。機械が仕事を奪ったとしても、人間は、単に、それらの機械を管理したり整備したりする役割を担うようになるのではないか、と思うかもしれないが、そう簡単にことはすまない。というのも、デジタルセンサー技術や移動式ロボットの進歩によって、機械が他の機械をこれまでよりも新しく洗練された方法で管理することが可能になりつつあるからだ。これは、ハードウェアだけでなくソフトウェアについても同様である。機械学習の進歩により、ソフトウェアエンジニアの仕事さえも、早晚コンピューター化されるかもしれない。

技術革新による失業の歴史的なパターンによると、失業は、より高い技能を必要とする分野の成長によって補填されてきた。しかしながら、このパターンは人工知能によって崩壊されつつある。というのも、人工知能は市場における高い技能を用いる分野の労働需要をますます圧縮すると予測されるからだ。とりわけ、これまでプログラムするのに困難があると考えられていた運転や文書作成といった仕事に、ディープ・ラーニングの技術によ

ってコンピューターが流入してきたことが大きい。つまり、コンピューターには不可能でありながら、仕事を必要としている全ての人間にあてがうほど十分な量の仕事など、もはや存在しないのである。

また、10年前まで、多くの専門家は、AIは人間の言語推論（verbal reasoning）への生得的能力には、到底かなわないと考えていた。しかしながら、マイクロソフト社と提携した中国の研究者は、新しいディープ・ラーニングのアルゴリズムによって、自身のシステムが、IQテストにおいて人間の言語推論と競合することができると報告した。この研究の代表者によれば、彼らのモデルは、学士の学位を持った人と修士の学位を持った人との間のレベルの知能を持ち合わせているという。AIにおける進展は、かつては悩みの種であったものの、今や定期的に人間の複雑な心的作用にも勝ち続けているのだ。このような技術革新によって、高い技能を用いる労働者は仕事の階梯を下ることになり、伝統的に低技能の働き手によって担われていた仕事を引き受けることとなり、低技能の労働者は一層この階梯から押し出され、一部の労働者は労働力の対象からも除外されてしまっている。教育によって技術との競争に打ち勝つという人間の労働に関する能力が、疑問に付されているのである。

たしかに、高水準の創造力が必要な仕事は、予見可能な未来において、人間の手に委ねられるという報告もある。しかし、これは、大衆は人間の手による微妙なニュアンスにどれほどお金を払い続けようとするのかということに大きく依拠する。また、AARONというコンピューター・プログラムによるドローイングは世界中の美術館で展示されているし、芸術史における最も創造的な作品をコンピューターに同定させることを可能にするアルゴリズムも作成されつつあるといった様々な例を考えると、安心してはいられないだろう。さらに、芸術的創造力のみならず、科学的創造力も新しい機械学習技術によって圧力を受けている。最近では、100年もの間、進化生物学者を悩ませ続けていた、扁形動物はどのように再生するのかという謎を、コンピューター・プログラムによるモデル作成によって解決したという事例がある。マイケル・レヴィン（Michael Levin）によれば、自然が行っていることを説明するモデルの開発とは、単なる統計や計算処理ではなく、科学者の行う最も創造的な事柄であり、科学的営みのまさに神髄なのである。

さて、ここまで、機械ができることを紹介してきたが、もし、ヒューバート・ドレイファス（Hubert Dreyfus）の有名な主張が誤りであり、コンピューターが実践的あるいは理論的専門性を実際に発達させるのなら、コンピューターがいまだにできないことは何であろうか。

もちろん、自動化された道徳的専門性（automated moral expertise）を開発しようという試みは、いまだに初期段階にあり、道徳の領域は、AIによる自動化の流れを免れようように見える。しかしながら、自動運転車やロボット兵士のための倫理的プログラミングの開発が急ピッチで進められているという事態を考えると、道徳の領域においても、コンピューターが人間の専門家に取って代わることになるかもしれないと危惧されるだろう。も

し、機械が最終的に道徳の専門家に取って代わることができるとしたら、このことは、経営者、消費者、そして政策立案者に、人間の倫理を周縁的なものにするのに十分な理由を提供するかもしれない。

しかしながら、予見される未来のいかなる時点においても、専門性を持った機械が優位に立つと想像することが困難な、いくつかの鍵となる領域がある。その領域とは、以下の通りである。

- ・ 長期間の (on a lifetime scale) 複雑な計画や目的についての、つまり、あらゆる点を考慮に入れた上でどのように生きるべきか判断する際に影響を及ぼすようなタイプの目的についての、かなり抽象的な表象に関係する意思決定。
- ・ 究極的ではあるが両立しない目的・目標・価値の間において、あるいは、価値の自発的な創出と順序づけの間において、間主観的な課題に対して理由を与え、かつそれに応答可能であるような仕方で、選択をすること。
- ・ 価値の選択に対して、また、その価値に照らして感情・行為・信念を自己規制することに対して、責任をとること。

従って、確かに、多くの領域において機械が専門性を備えるように見えるだろうし、機械学習における進展を考慮にいれるならば、これらの領域の中の一部においては、機械は確固とした (robust) 専門性に到達しさえするだろうが、機械が知恵に似た何らかのものを獲得する明確な経路など存在しないのである。

しかしながら、以下のような危険がある。すなわち、知恵の一般的で政治的な価値とは、生産的専門性 (productive expertise) の協力によって常に支えられてきたのだが、自動化によってこの知恵と生産的専門性の対が分離しつつある、という危険である。この分離はいくつかの点で問題含みである。第一に、知恵は生得的なものではなく、多くの経験や、失敗から学びそしてそれを反省するという機会を通して形成されるのだが、現代のテクノソーシャルなシステムは間違いに対して不寛容であり、多くの領域において、機械的な人為的ミスは今や機械による自動化によって最小限のものになっている。現在に至るまで、定期的な判断ミスは、システムを知的な専門家の指導の基に置くという避けられないコストを支払わせ続けてきたのだが、21世紀の経営者、政策立案者、組織は、失敗のさらなる機会を人間に与え続けるための動機を持たないだろう。また、甚大な飛行機事故といった災害が、人間の知能とコンピューターの知能の協同した実践によってもたらされてきたということを考えると、以上の問題はますます緊急性を帯びることになる。さらに、例えば IBM 社のワトソンは、診断や治療計画を導き出すために何百万というアルゴリズムのプロセスを実行しているように、AI は、人間の理解力にとって不透明なデータスケールやタイムスケールに基づいており、人間の医者はワトソンに対抗しようとは思っていないだ

ろう。

また、新たに注目されている「ヒューマンコンピューテーション」は大変示唆的である。アマゾン・メカニカル・ターク⁴のような、人間の知恵を「クラウド・ソーシング」する経済モデルは、今のところコンピューターが実行できない複雑で難しい作業をとて小さな部分に分割し、雇用主がこれらを多数の独立した人々に数セントで割り振り請け負わせる、という仕組みで機能している。確かに、ネットワーク上の人間とコンピューターの結合は、高次の知能によって導かれることなく、どのような人間よりも早くそして安く解決策を生み出すことを可能にする。しかし、問題の分離された諸部分に独立に取り組む人間が、多くの専門性、ましてや知恵を陶冶することができるということは決してあり得ることではない。ここにおいて人間は、作業が意味をもち価値を獲得する社会的意味の幅広い文脈から、切り離されてしまっているのだ。

さらには、ヒューマンコンピューテーションのねらいとは、人間の道徳的あるいは知的な弱さに由来する効果を、それらの弱さを実際に矯正する (remedy) 必要なしに、無効にする (neutralize) ということにある。ヒューストン計画を例にとろう。ヒューストン計画とは、うつ症状や不安症状に苦しむ人のために、人工的セラピスト (artificial therapist) を作り出すことを目的としている。これは、AI 技術に支えられることで、実際の諸個人から作り出された合成的な人格という形で提供されるが、この人格はクラウド上に様々なレベルの専門性を持ち、また、時間が経過することでクラウドが完全に変わるとしても、一貫して親切で忍耐強い人格が提供されるという。この例からわかるように、人間の道徳的徳そして知的な徳は、専門性を備えた機械によって置き換えられるという脅威にさらされているだけではない。我々の悪徳が心理的にはそのまま残され続けているながら、その悪徳の影響がアルゴリズムによってフィルタリングされ排除されることで、人間の構成要素はバラバラに砕けることとなり (fractured human composites)、その結果、人間の道徳的徳・知的な徳は、これらのバラバラな構成要素によって置き換えられてしまう、という脅威にもまたさらされているのだ。

以上をまとめると、新しい技術は人間の知識、技術、知能、そして知恵を育成し行使するためのより幅広い機会を作るよう約束するという、歴史的に見て正当であった主張に対

⁴ アマゾン・メカニカル・ターク (Amazon Mechanical Turk) とは、コンピューターだけでなく人間も資源としてシステムに組み込むことで、コンピューターが依然として実行できないタスクを人間の処理能力を利用することで解決しようとする、ヒューマンコンピューテーション (human computation) 一つである。ヒューマンコンピューテーションの用途によっては、大規模な人数を動員する必要があるが、アマゾン・メカニカル・タークは、クラウド・ソーシング市場として、金銭を作業の対価とすることで人間の作業への動機を担保するアプローチをとる。以下の論文を参照した。(鹿島久嗣、馬場雪乃「ヒューマンコンピューテーション概説」、『人工知能』29 巻 1 号、2014 年 pp. 4-11.)

して、機械学習におけるめざましい発展は、疑問を呈するのである。この急進的な発展に対処する準備のできた新興の業種などないのであって、このことは、労働者に技術を身につけさせる時間がないからなのではなく、高度な経済需要において、どのような種類の技術が人間に固有なものとして残り続けるのかということが、明確ではなくなっているからなのである。短期的に見れば、柔軟な器用さ、創造力、社会的知能という我々の利点は守られるかもしれないが、我々が思うほど大きな保護を受けはしないだろう。

もちろん、仕事がない世界というのは魅力的である。もし、世界の人口の99%の人々の福利が、政治や経済を支配する権力を持っている人々と比べて、また、機械化による利益を独占的に享受する人たちと比べて、低い優先順位にあるのではない場合、この可能性は一層魅力的であるだろう。しかしながら、自動化による利益の公平な分配を求める世界的な暴動の可能性といった、他のシナリオを考えることもできる。このようなシナリオを避けることはできるのであろうか。

4. 機械知能の責任をとる

そのようなシナリオを避けるためには、以下のように考える必要がある。つまり、AIと自動化された行為者は、人間とその知恵が服従したり抵抗したりしなければならぬ外的競合物 (external competitors) なのではなく、それらは、我々がそれらの認知的プロセスを、もはや直接的にあるいは完全には理解できない場合においてさえ、人間が責任を取りまた取り続けなければならない、我々の拡張 (extensions of ourselves) なのである。ロボット支配やシンギュラリティが誇大に宣伝されているが、AIは人間以上にはなれないのだ。我々は、以下の理由により機械の行為者の責任をとらねばならない。

- ・機械は、我々の目的を追求するために、世界を変えるような仕方で、我々自身の手によって作り出されたから。
- ・機械は、我々が提供することを選択した、時間、エネルギー、データ、そして人間による支援という投資に依存して、存在しているから。
- ・我々は責任をとることができるが、機械は自身の責任をとることができないし、おそらく今後も決して責任をとることがないから。

ケヴィン・ケリー (Kevin Kelly) のように、AIを、理性を持つ年齢にいたり自身の目的を自由に追求するようになった、我々の子供のようなものとして捉える見方(「生の第7の王国」(“7th kingdom” of life))もあるが、AIは知恵を持たず責任をとることもできないため、このような見方はふさわしくない。AIは、人間の子供でも、人間でない生命体でも、単なる道具でもなく、新たなカテゴリーに属するものなのである。ルチアーノ・フロリ

ディ (Luciano Floridi) が、エンジニアリングにおける「エンベロープ化 (enveloping) ⁵」という概念に訴えて指摘するように、我々は、「軽い (light)」AI の利益になるように情報環境を作り上げてきたのである。スマートフォンも、スマート・ウォッチも、スマート・カーも、今や全て以下の一つの目的のためにデザインされている。つまり、我々や世界、さらには我々が世界において評価する事柄について、可能な限り多くの情報を与えるという目的、言い換えると、AI に情報入力するという目的である。このように、以前は、我々や他の生物の利益のためにエンベロープ化されてきた世界は、物理的対象や空間や人間の行為がますます多くのデータの流れを生み出すようにエンベロープ化された世界へと、今や変化したのである。しかし、今こそ、誰のために我々の世界を作り上げることを本当に欲しているのかと我々は問うべきである。我々は、世界をスマート・マシンの支援するものに作り変えることに急ぎすぎるあまり、知的で知恵のある人間のために世界を作り上げる責任も持っているということを、忘れてしまっているのではないだろうか。

ここで、ジュディス・サイモン (Judith Simon) らの研究に従って、「絡み合ったテクノソーシャルな認識システム (entangled socio-technical epistemic system)」について参照し、AI に適した形で建設され、人間と機械の両者の絡み合いによって形作られた世界における、責任の概念について見解を述べたい。彼女によれば、我々は、諸個人と自動化されたシステムが複雑に絡み合った、あるテクノソーシャルなシステムに巻き込まれており、ここにおいて、我々は何事かの〈第1のあるいは近接的な原因〉であると断定できないという事態に陥っている。しかしながら、彼女によれば、このような状況においても、我々は「責任をとる」ことができるし、そうすべきであるし、また、しばしば実際に責任をとっている。ここにおいて、我々は、このように見いだされる責任を、「テクノソーシャルな事実性⁶」と呼ぶことにしよう。この責任とは、我々が直接に引き起こしたのではない我々の存在に関する事実に対してさえも、我々は根源的に責任を持っているということの意味している。なぜ責任を持っているのかといえ、我々はそのような事実に対して意味と価値

⁵ この「エンベロープ化」という概念は少々わかりにくいので解説しておく。これはエンジニアリングの領域において用いられる言葉であり、「電気自動車のためにソーラーパネルの設置されたガレージを作り上げる」といったように、「機械を支援するために機械の周辺にある環境を作り上げる」という場合をさして、「我々は環境をエンベロープ化する（あるいは、環境がエンベロープ化される）」と表現する (p. 174. 参照)。なお、この「エンベロープ化」という概念に関してのフロリディの見解は、以下の著作の第7章において詳細に展開されている。(L. フロリディ著、春木良且・犬束敦史監訳、『第四の革命』、新曜社、2017年)

⁶ ヴァラーはここで、サルトルの「事実性 (facticité)」という概念を援用して説明している。(p. 175. 参照)

値を与え、それらを選択しているからであり、また、我々は、これらの意味と価値に照らして、新しい何かをなすことに対していつもでも自由であるからである。ここにおいて重要なことは、責任があること (being responsible) と、責任のある状態にあること (being responsibly) の区別である。前者は、我々が着手した行為とその行為の帰結の間に明確な因果的つながりが存在する形而上学的な状態のことであり、後者は、形而上学的な状態ではなく、実存的かつ道徳的な状況のことであり、この状況において我々は、我々の自由や力、さらには知恵を導く恒常的な能力に向かい合うことを余儀なくされているのだ。確かに、諸個人と自動化されたシステムからの個別的な帰結の間にあるつながりは、ますます不透明なものとなっており、前者の意味での責任は問いづらくなっているかもしれない。しかし、後者の意味での責任はこの不透明性と完全に両立可能であるため、テクノソーシャルな決定論は排除されなければならないのだ。

5. 結論：知恵のための世界をつくる

最後に再び知恵という概念に立ち戻ることしよう。アリストテレス、儒教、そして仏教の間で、徳の概念は多様である。しかしながら、それらの考えに共通する核があり、それは、人間の意識についての基本的な生物学的可能性に根ざしている。すなわち、自己規制と自己修養 (self-regulation and self-cultivation) に根ざしているのだ。徳の伝統において、知恵とは、知識、知能、そして専門的技術以上のものとして理解されている。それは、ある生き方 (a manner of life) なのであり、究極的な価値として選択した目的に照らして、言い換えると、総体として捉えられた生に照らして、生それ自体が規制され、陶冶される生き方なのである。知恵の徳とは、端的に言えば、「責任のある状態にあること」の実践なのだ。

今日、AIの力は、経済的利益や利益と力の統合を提供するような形で、用いられようとしており、我々を取り巻く世界は、まさにこの可能性を現実にもたらすために、ますます作り直されつつある。しかし、もし、我々がこのAIの利用戦略を見直さないならば、我々は、AIと自動化がもつ広大な肯定的可能性を無駄にしてしまう恐れがある。というのも、この広大な肯定的可能性とは、その責任ある使用においてのみ、見いだすことができるからだ。AIと自動化の責任ある使用とは、我々の道徳的で知的な完全性に十分資するように、また、人間の知恵を増進させることに役立つ仕方で、AIと自動化の力を用いるということに他ならない。このように、AIと自動化を、知恵のための世界をつくるという目的に基づいて用いる場合に限って、それは、我々が現代の諸問題に対して取り組む際の、実効的で適合的な方法を見つけ出す助けとなるのである。

(山二滉大、三上航志)