



データ駆動型社会における「人間中心」に向けた課題

唐沢 かおり*

Challenges for the Data-driven Society with a Human-centered Concept

Kaori KARASAWA*

Abstract— This paper identifies two challenges for the data-driven society with a human-centered concept. The first challenge is that we need to regulate individuals' behaviors in order to foster cooperation and avoid social dilemmas. Such a regulation system should be designed to effectively take into consideration various cognitive and motivational biases which may influence the effect of interventions for behavioral regulation. The second challenge involves moral issues caused by the implementation of data utilization technologies. The issues include expansion of inequality, increase in hierarchical ranking of individuals, and the ambiguity of AI's responsibility when problems arise. To face these challenges, we must rigorously clarify the constraints that the data-driven society imposes on each of us and the society as a whole, and continue with the discussions to explore a more desirable future.

Keywords— A data-driven society, Human-centered, Functions of human mind, Morality

1. はじめに

「人間中心」の社会は、Society 5.0の基本的なコンセプトであるが、そのような社会の実現を支える仕組みの一つとして、データ活用に基づく「データ駆動型社会」の実現がうたわれている。多様なデータを大量に収集し、科学技術を駆使して、その分析結果を実社会の介入に用いることで、都市環境、生活環境、またその中での人の行動を「よい方向」に変えようとするのである。

その内実を、内閣府のホームページ [1] にもある記載に即してもう少し詳しく見てみると、次のようなことになるだろう。サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムにより、経済発展と社会的課題の解決が両立できる。IoTですべての人とモノがつながり、さまざまな知識や情報の共有やAIの活用が進み、新たな価値が創造される。これにより、社会の変革（イノベーション）がなされ、個人が持つ多様な必要も満たされる。その結果、希望の持てる社会、世代を超えて互いに尊重しあえる社会、一人一人

が快適で、活力に満ち、質の高い生活を送ることができる社会が実現する。これが Society 5.0 という構想のなかでうたわれている、人間を中心としたデータ駆動型社会の姿である。

このような社会が実現するなら、それは、もちろん素晴らしいことだ。技術の力やそれを背景にしたデータ活用により、個々人の様々なレベルでの必要、欲求が満たされ、満足感や幸福度が高まるとともに、今の社会が抱えている課題が解決していくのである。「内実」として書かれているビジョンを肅々と進めることが、今の社会が持つ閉塞感を打ち破り、無条件に私たちの明るい未来を描く力を持つなら、技術を信じて、それに私たちの将来をゆだねればよい。

しかし、少し冷静に考えるなら、このようなビジョンが示す社会の実現には、様々な障害があり、考えるべき問題点が多々あることに気がつく。誰もが快適で、質の高い生活を送ることができるとされているが、それは、どのような、どの程度の快適さや質の高さを意味するのだろうか。また、個々人の多様な必要というけれども、どこまで細かく個人の要求に答えたり、様々な選択肢を提示することが望ましいのだろうか。これらは、人間の持つ欲求と密接に結びついた基本的な問いであるが、実のところ、答えは明確ではない。

*東京大学人文社会系研究科 東京都文京区本郷 7-3-1

*Graduate School of Humanities and Sociology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo

Received: 7 January 2020, Accepted: 28 February 2020.

そもそも、Society 5.0を描く言葉の延長には、満足観、さらには幸福という、人間が求めてやまない価値を前提としたうえで、すべての人に対してより高い満足と幸福の獲得が約束されたかのような、いわばユートピア的な社会観が透けてみえる。しかし、快適さ、便利さ、豊かさ、健康など、満足や幸福につながることを追い求めることは、様々な問題をふくんでいる。一つには、より高いレベルで追い求めること自体が、後にも述べるように、主観的満足感の達成を困難にする危険さを含んでいることだ。私たちの要求水準は、満たされている程度に応じて上昇してしまうので、「これで終わり」という点に至ることが難しい。もちろん、より上のレベルを求めるのは、向上心の表れでもあり、進歩につながるので、必ずしも悪いことではない。しかし、満たされた状態に慣れてしまい、もはやその状態を、幸福だと受け止めないという結果も招きかねない。

また、多種多様な選択肢がある状態は、いっけん豊かさや自由につながるかにみえる。しかし、決めるためのコスト、迷うことのストレス、選ばなかったものがかえって魅力的に見えてしまうリスクなどがその背後にはつきまとう。自分の選択に満足を見いだすことをせず、より高い満足を与えてくれるべきものを、いつまでも求めてしまうことにもなりかねない。

快適で質の高い生活を、トップダウン的に与えるという社会の枠組みは、人間にとって決して望ましいことではない。下手をすれば、上記のようなモンスター的欲求者を生むという皮肉な結果をもたらすとともに、個人と社会全体の福利との調和を果たすことが困難になるという帰結にも至る可能性があるからだ。

今の社会が抱えている諸課題の多く、例えば、高齢社会、環境問題や、社会格差、労働人口の減少などは、人間が健康や長寿、豊かさや快適さといった価値を追い求め、またそれを通して満足や幸福を得ようとしてきた結果の副作用でもある。満足や幸福につながる資源は、無限大ではなく、エネルギーからヒューマン・サービスに至るまで、得られる資源には一定の限界がある。従って、一人ひとりが、自由に個人の欲求や満足を追求することと、社会全体の福利との間には、利害の対立がある。人間中心の Society 5.0は、この問題に立ち向かうことなくしては、手に入れることができないだろう。

資源の分配ということに焦点を当てれば、何らかの外的な規制をかけ、平等主義的な資源配分を行う、資源の価格を上げ、「買い取る」力を持つ個人が恩恵を享受できるようにするなど、その方略はいくつかあるだろう。これらは、対象によって使い分けられながら、すでに現実の社会の中に組み込まれている。しかし、それが本当に「正しい」方略の適用なのか、また、複数の方略を組み合わせるとしても、それらのバランスをどのようにとる

べきかなど、回答がすぐには見いだせない問いが残る。さらには、このような分配方略を実装した社会システムの妥当性を誰が評価するのも問題となるだろう。

技術とデータを活用し、「人間中心」の社会を実現するという構想自体は、否定されるものではない。データ、つまり、私たちがこの世界について知り得る知識を活用する技術を、社会課題の解決に資するように実装していくこと、また、それを未来の社会を設計する構想に組み込むことは、今、まさに社会が必要としていることだろう。ただし、実際問題として、個の欲求と社会の福利の対立は、個人の行動を一定程度、制御すること無くしては、解決に近づくことができない。技術の進歩やデータ活用に対する素直で楽観的な期待に甘えるのではなく、個と社会の福利の調和的な関係を目指した個人の行動の制御のあり方、さらには社会全体の善いあり方に関する議論を深めることが、私たちの未来を設計するにあたって不可欠な課題である。

そこでの議論においては、Society 5.0が掲げる「人間中心」というコンセプトを踏まえることが要求される。従って、様々な実社会への介入が前提となる行動の制御については、人の心や行動の特性を十分に考慮することが基本的な要件となるだろう。また、個人の満足や幸福と社会全体の福利との間に、調和的な関係を築くためには、技術が社会全体にもたらす影響の功罪を倫理的に検討する視点も必要になる。いずれも、データ駆動型社会を推進するにあたって、まず対応すべき課題の同定から始めることが求められている。そのような認識をふまえ、本稿では、人の心の特性や人と社会との関係を検討してきた社会心理学の観点から、これら二つに関わる課題を提示していく。

2. 行動の制御と心の特性

Society 5.0は、経済発展と社会課題の解決を両立することを目標の一つとしている。これは、一方では、データ活用に基づく社会介入により、個の多様性を踏まえた上で、その必要と欲求を満たすことのできる、豊かで快適な、ひいては幸福を得られる環境の構築を推進するというベクトルを持つ。しかし、他方、社会課題の解決という観点からは、人の行動に一定の制約をかけ、統制する試みが必要となることも、先に述べた通りである。例えば、環境問題という社会課題であれば、個人に快適さや便利さを提供しながらも、それを追求する行動をある程度、抑制することが必要となるというようなことだ。単純に個人が快適さや便利さを追求する行動をとると、社会全体の福利が損なわれるというジレンマ状況が存在するなか、両者の調和的な関係のために、個人に対してどのように働きかけ、その行動を制御していくかは、大

きな課題となる。

技術を提供する立場からは、技術の力でジレンマ状況を回避し、個人の快適さや便利さの追求を可能にするという主張がなされるかもしれない。消費電力が少なくても空調が開発されれば、その分、電力を使う量は増やすことなく、四季を通して、自宅やオフィスで快適な温度を保つことができるのは、確かにその通りだ。しかし、だからといって電気という資源の消費が止まるわけではない。消費電力が少ないことをいいわけに、より一層、夏は涼しく冬は暖かい環境を求めてしまうかもしれない。「限りある資源」という課題を念頭に置けば、個人の快適さが追及され続ける状況そのものに、本質的な問題が存在する。必要や欲求に従い、自己の行動を制御することなく振る舞うことにたいして、問題意識を持たない人たちが社会が構成されてしまうようなことになれば、社会課題の解決も果たされない。個人の快適さへの欲求、またそれを満たす行動に対しては、どこかで抑制をかけていく必要がある、たとえ、技術が進むとしても、個人の欲求や満足を追求する行動を、いかに制御できるのかという問題は、つねに私たちの社会につきまとうことなのである。

データ駆動型社会では、このような問題に対する解決策の一つとして、データを活用した環境への介入により、自己制御的行動を誘導するという手法があり得るだろう。例えば、リアルタイムで得られる人の動きとエネルギー消費のデータを用いて、より効率よくエネルギーを配分しつつ、そのときに有効な行動を促すように環境や情報を操作するなどである。また、エネルギー消費行動をモニターし、行動に応じて効率よく報酬を与えたり、ペナルティを課すことで、エネルギー資源の消費に抑制をかけることもできるだろう。このような例に見られる誘導が成功すれば、社会課題の解決に対して、一定の効果を持つので、アプローチを支える技術の進展は望まれるところである。

しかし、これらの事例も含め、さまざまな場での介入や操作の具体的設計に関して、人の心の特性という点からみると、有効性が減ることにつながる落とし穴がある。以下、行動の制御という試みに焦点を絞り、心の特性に由来することとして考慮すべき問題を取り上げていく。

2.1 報酬や罰を用いた行動の制御に関する問題

私たちの行動を統制する強力な要因は、報酬と罰である。これらの効果は、金銭や罰則など、個人の外にある要因に依存するものであり、外発的動機付けの効果と呼ばれている。外発的な動機を喚起するための操作は、態度や価値観といった「内的な」要因を変容させる手法に比較すると即効性があるし、報酬や罰の大きさを増すことにより、確実な効果が期待できる。

しかし、これらの効果は、報酬や罰が的確かつ継続的に提供されることにより維持される。従って、報酬や罰を提供するための資源や、人の行動を監視するというコストが発生する。また罰を与えたり、監視を行う存在は、敵意を持たれやすく、そのような責を担う人や組織に対する好意や信頼が損なわれるという問題にもつながりかねない [2][3]。

加えて、動機付け分野の研究では外的報酬が持つネガティブな効果が指摘されている。本来の自分の興味や関心と合致するような（したがって報酬がなくても行う）行動に対して、行ったことへの報酬を与える手続きを導入すると、「自分は、それを報酬のために行っているのだ」という（誤った）認識を持ってしまうことが起こる。そのため、もともと持っていた興味や関心が低下してしまい、報酬がない場面で当該の行動を行う可能性が下がってしまうのである [4]。例えば、省エネが重要だという価値観を持っていた人は、本来は報酬がなくとも、省エネ行動を行う。しかし、省エネ行動に対して報酬が与えられるような環境に置かれると、自分の省エネ行動は「報酬のためだ」と（誤って）考えてしまう。その結果、報酬がなくなった時に、かつてのように自発的に省エネ行動を行うことが見られなくなるのである。

もちろん、行動の統制にむけて、報酬や罰に頼らざるを得ない場面もあるだろう。しかし、望ましい行動をとることに対する価値の醸成や内的規範の形成を怠り、報酬と罰に頼ることは、目先の行動制御に陥るのみならず、望ましい行動へとつながる内的な関心や価値観を抑制してしまう可能性がある。

なお、近年、報酬や罰で人の行動を縛るのではなく、ナッジのように望ましい行動を引き出す環境設計によるアプローチも盛んである。ナッジは、ここで指摘した報酬や罰のネガティブな効果という問題を回避するには有効だろう。ただし、ナッジは、人が主体的に判断を発揮する場面では効果が薄く、人の内的な関心や価値を基盤にした意思決定力を涵養することの重要性を主張するプーストというアプローチも提起されていることを、ここでは指摘しておきたい [5]。

2.2 意図推論がもたらす問題

私たちは、自分に対して働きかける他者について、なぜそのようにするのか、意図を推し量る存在だ。とりわけ他者が何らかの操作や誘導により自分の行動を変えようとしていることに気が付くと、このような推論が働く。データ駆動型社会における人々の行動を一定の方向に誘導する試みも、もし、それが明示的になるのであれば、同様の推論が働くことになる（なお、そもそも、あからさまに操作や誘導を行っているとは気が付かせない介入もありうるが、行動を当人の意識が及ばないように操作してしまうという倫理的問題が生じる可能性がある）。

操作や誘導意図の推論において問題になるのは、それらが誰の利益のためになされているのかに関する認識である。古典的な説得研究で示されていることであるが、行動を変化させようとする説得的誘導の意図が、説得者の利益獲得にあるとみなされると、信頼性が下がり、メッセージの内容がいくら優れたものであっても効果がなく、誘導や介入を試みた者に対する拒否的な態度が強化される [6]。つまり私たちは、「他者が自己利益のために自分を動かそうとしている」と認識すると、それに対して反発を感じたり、態度を硬化させ、誘導や説得には応じなくなるのだ。

ここで注意したいのは、誘導や介入に従う行動が、社会の公益につながることで自体が明らかであっても、それと同時に誘導や介入を行う主体の利益にもなると見なされるのであれば、拒否的な態度が生じてしまう可能性がある。例えば企業などが主体となる場合、公益への配慮につながる行動を促されているとわかっているにもかかわらず、「企業は利益のために行動する」というステレオタイプの信念や不信感を強固に持っている人は、利益追求の意図を過剰に読み取り、反発してしまうだろう。また、誘導や介入の試みに対してこのような意図の推論を行うことで自体が、ネガティブなステレオタイプの信念や不信感を、一層強化してしまうという負のサイクルに陥る可能性も指摘できる。

加えて、知り置くべきことは、操作や誘導意図を過剰に読み取ることがもたらすマイナス面である。私たちには、自分の行動の自由を確保したいという基本的な欲求があるが、他者からの操作により動かされることは、そのような欲求とは反することである。従って、あたかも自分の自由が奪われたかのように思い、心理的な反発を経験することにつながる。このような心理的な反発は「リアクタンス」と呼ばれている [7]。リアクタンスが生じると、「自分は行動の自由を持つ」という感覚を回復するための行動をとる可能性が指摘されている。つまり、わざと操作や誘導に反する行動を行い、自分は自由に行動できることを確認したくなるのである。

2.3 自己の状態に関する判断にかかわる問題

一般的な自己制御のメカニズムは、自分が行った行動結果を把握し、それと自分が望む状態との差を知り、差を小さくする行動へと改めて自己を調整するという過程からなる。従って、行動の結果として得た状態に対して、それが、どの程度自分が望む状態なのか、満足できる状態なのかについて、自分の心の中をみつめ、把握することが求められる。しかし、このような満足度にかかわる自己認知は、得たものの絶対的な値ではなく、過去に得たものや他者の状態などとの比較に基づく、「差」により決まりがちである。自分の満足度を評価するとき、

「良い」から満足を得るというのではなく、「比べたうえで、なお良い」ことを過剰に求める心理が働くのである。

他者が自分よりも良い状態にあることを知れば、自分の置かれた状態が不当に思え、不満を抱き、より上を望むのは、どのような社会でも見られる人間の一般的な心理である。しかし、とりわけ、データ駆動型社会においては、データ活用が進むことで、多様なデータが簡単に得られるようになり、他者との比較が促進されてしまう。従って、他者より上だとか下であるということにより気にし、自分が得るに値するものは何なのかを考える際、自分自身ではなく、他者が参照点になる傾向が一層強くなる可能性がある。

また、選択の多様性が容易に目の前に示されてしまうことも、問題をもたらす。ある選択で十分満足できたかもしれないのに、他の選択と比べてしまい、そちらの方がより望ましいという可能性を考えたり、またはベストの選択を選びたいと考え、現状に満足せず、より上を求めて、選択肢の間をさまよったりという行動傾向も、強くなるだろう。

2.4 心の特性を踏まえた操作や誘導方法の設計にむけて

以上、行動の制御に関わることに焦点をあて、心の特性という観点から問題となる三点を指摘した。これらはいずれも、人々の反応が、制御のための操作や介入を試みる側の意図とは反したものになる可能性と、それが起こってしまう心のメカニズムについて示したものである。行動の制御がうまくいくこと（ひいては、それにより、人の行動が社会全体の福利と調和する社会が実現すること）を目指すのであれば、なるべくそのような可能性を減らすべく、事前に配慮しておくことが望まれることになる。

ただ、そのような配慮にあたって問題になるのは、人々の反応が実のところどうなるのか、高い精度で予測すること自体が困難であるという点だ。ここで示した各問題は、人の心と行動が一定の法則に従うことを示した心理学の知見に基づいているが、それは、いわば全体的な傾向としての知見であり、一般的な他者について、ある考えを持ったたり行動したりする「傾向が高まる」というものである。しかし、目の前の特定の人が、どのように考え行動するかを予想することは難しい。その背後には、心や行動の動きに影響する状況要因や個人差が、多様である一方、ある特定の場面と個人について、それらを事前に把握することが、困難だという事情がある。つまり、列記してきた三つの問題のそれぞれが、特定の場面でどの程度起こり得るのか、また誰に対して起こり得るのかを正しく知ることはできないし、何かしらの配慮を行ったとしても、確実に回避できるかどうかわからないということになる。

では、私たちはどうしたらよいのだろうか。人の心の働きや行動の制御に関わる試みは、今も、またこれからの社会においても必要なことである。確実な予測や統制が困難な中、簡単な解決は見いだせないにしても、「心の特性を考慮する」という手続きをどこかに取り込むには、どのような方略があり得るのだろうか。

そこには、効率の良い解決策はないが、事前のシミュレーションやパイロット的な実験・調査により、介入や操作の効果を検討した後に、柔軟にその具体的な設計を調節していくという、極めて地道な作業の組み合わせにより、改善をはかることが必要になるだろう。また、「他の成功事例や失敗事例」に関わるデータを分析・活用することも有効だろう。つまり、「実証的根拠を踏まえた政策」に近づけて行くのである。

事前の検討では、少なくともその時に見えている状況を基盤に、人々がどう考えふるまう可能性があるのかを吟味することになる。目指す効果が得られるのか、望まないネガティブな影響は無いのか、もしある場合、その深刻さについて検証するのである。その際、重要なのは、うまくいくシナリオにこだわることなく、うまくいかない場合についても、起こりえる結果を考えておくことだろう。

念のため述べておくと、これは操作や介入を止めるためではない（もちろん、「うまくいかない」可能性があまりにも高いものについては、実行すべきではないが、ここで想定しているのは、そのような状況ではない）。社会課題の解決は、着実に進める必要がある、個人個人の行動の変化を促すことについても、それは同様である。うまくいかない可能性について事前に検討しておくのは、それにより、「うまく行っているはずだ」と信じてしまうマインドを消去し、問題が生じたとき、それが問題であることに気が付くまでの時間を、全体を調整するような対策を見出すまでの時間を短縮するためである。これも心の特性の一つであるが、私たちは「見たいものを見てしまう」存在である。うまくいくことしか考えていないことは（そしてそうなると信じ込んでいることは）、問題を発見する目を曇らせる。深刻さが増す前に問題に気がつくこと、いったん動き出したことを改善することは、簡単なことではない。改善をはかるための柔軟性を必要とするのであれば、それを発揮しやすい仕組みを、あらかじめ織り込んでおくべきであるが、「うまくいかない場合」についても検討しておくことは、そのための方略の一つになるのである。

3. 倫理をめぐる問題

個人の欲求、満足、幸福の追求と社会の福利との調和のためには、個の制御に加えて、社会全体の倫理が保た

れる必要がある。ここでいう「社会全体の倫理の維持」とは、善悪、公正、正義などに関して、「こうあるべき」とされることに合致する人々の認識や価値観が維持されることを指す。

人々の素朴な認識や価値観に着目して倫理を議論するのは、社会的な要因が人の認識や行動に与える影響を考える社会心理学の切り口の一つである。多数の人々がともに暮らす社会が、円滑に運営されるためには、それを危うくするような、加害、裏切り、不正などに対して人々が否定的な価値観を共有し、そのような行為には罰則を与えることを是認し、実際にそれを可能にする社会システムの存在を認めることが必要になる。倫理（道徳）は、このようなことに関わる人々の認識のなかに存在する価値体系として、位置づけることができる。また、個人個人が、安心して、また安全に、他者と共生することができる社会の基盤となる点において、人間中心の社会を基礎づけるものとなる [8]。

データ駆動型社会における人間中心というテーマを考えるにあたっては、データ駆動型社会がもたらす帰結のなかに、倫理との齟齬、またさらには倫理を脅かす可能性が含まれないかを考察することが重要となる。本稿の後半は、技術の特性やデータの利活用に関わる問題を中心に、いくつかの可能性について論じていく。

3.1 技術・価値・倫理

最初に考えたいのは、技術に伴う価値と倫理との関係である。社会にとって価値あるものを生み出すことは、技術・工学の使命であるという点において、技術は常に一定の価値を組み込んでいる。例えば、便利な生活や楽しい時間を与えるための技術は、便利さや楽しさという価値が組み込まれている。従って、技術と倫理との関係については、技術が体現する価値の倫理性という視点から考察することが必要になる。

考察にあたっては、技術が組み込む価値の、中立性、透明性に注目したい。ここで、中立とは、誰にとっても同じように価値を持つ、またはそれを享受できることを、また、透明性は、価値が何であるのか、誰にとっても明示的に示されていることを意味する。誰にとっても便利さが実現できるのか、便利であるとわかるのかというようなことだ。（なお、技術が生み出す便利さ、楽しさなどではなく、中立性、透明性に注目するのは、便利さや楽しさの倫理性が、それらの定義や、置かれる文脈に依存するなど、一義的に定まらず、「善い便利さなら倫理的に正しい」というような議論の循環に陥る可能性を懸念するからである）。

さて、技術が生み出す価値を享受するには、多くの場合、経済的な対価やリテラシーが要求される。お金があり、技術がよくわかっている人は、そうでない人よりも、技術の恩恵を受けることができる。公的に導入されるも

のについては、個人の持つ対価やリテラシーに左右されないという論も成り立つかもしれないが、その場合でも、個人が所属する社会（地域や国）などが持つ経済力やリテラシーが、技術の恩恵を享受できる程度を決める。つまり、技術が組み込んでいる価値は、中立性や透明性が保証されているものではなく、それを享受する対象の偏り、不平等さと結びついているという側面を持つ。

誤解を招かないように確認しておく、経済的な対価やリテラシーを要求すること、またさらに言うなら、リソースを持ち、技術がもたらす価値を享受し得る対象を念頭に置いた技術導入が、倫理的に望ましくないと言いたいわけではない。ここで指摘したいのは、組み込まれる「価値」の性質を曖昧にした上で、中立性や透明性に対する幻想を与えてしまうことが問題となるという点である。

従って、技術の価値中立性・透明性の「神話」から脱却した上で、技術が実現する価値や規範（そしてその問題）について考える必要がある。これは、何を実現するための技術であるのかを「透明」にしたうえで、誰がよしとする価値なのか、恩恵を被るのは誰かということをはっきりとさせていくという説明責任が、技術側に課されることになる。技術の社会受容は、このような説明も含んだ技術導入の上に成り立つことを踏まえ、データの利活用についての諸技術も、まずは、この点を明確にしておくことが求められるだろう。

3.2 データ駆動型社会における不平等や格差

さて中立性、透明性は、不平等や格差という問題と直結することである。先に述べたように、人々に満足や幸福をもたらす社会といっても、そのためのリソースは無限ではなく、また、経済力やリテラシーにも格差があるのが現状だ。それを踏まえると、高い技術が実装された（快適な）環境が作られたとしても、それを享受できる人とそうでない人、また享受できるコミュニティと、そこから取り残されたコミュニティが出てくることになる。経済状態や学歴が、個人の努力を超えた要因でも決まることが指摘されている中であって、これらの差が、そのまま技術の恩恵の受益差につながることは、社会の不平等さの現れでもある。

では、データ活用は、このような不平等さを解消するベクトルを持つのだろうか。もちろん、その答えはデータの使い次第であり、不平等さへの配慮や格差解消を目指したプロジェクトの活性化など、活用の技術を開発し実装する側の問題意識が問われることになる。ただ、問題意識が問題の解決につながるためには、データ活用という営み自体から生じうる様々な可能性に、広く意識が向けられることも必要となる。ここでは、その一つとして、データ活用が「順位付け」を促進することで、さらに格差が深刻になる可能性について指摘しておきたい。

データ駆動型社会は、行動、身体状態、考え、交友関係、社会での立場など、さまざまな個人に関わるデータを収集する。データは個人を多数の数値からなる存在に置き換えるが、そのことは、個人が全体の中で、どこに位置づけられるのかをも明らかにすることになる。つまり序列がつけられ、その可視化が容易になるのである。

収集や公開には、個人の意思の尊重やプライバシーへの配慮などがなされるべきであるのだが（また、そのような配慮が実際になされてもいるとも思うが）、一方で、個人データ提供の機会があまりにも多数ある中、自分の情報が、どこに、どのように蓄積され利用されているのか、把握し統制することは難しく、意識・意図しない形で、データが蓄積されていく。例えば、Facebookの「いいね」を送信するパターンから送信者のパーソナリティが一定の確度で予測可能なことが明らかにされているが[9]、自分が「いいね」を送信する際、それ意識している人はほとんどいないだろう。このようにして集まった大量のデータは、利活用のために、あるパラメータ上で序列化されることにより、利活用の対象になっていく。

序列化された数値が、大量に集約されたうえで、何らかの選択のために利用される事態が生じると、サイバー空間上に存在する「人間を表す数値」を私たちは比較し、評価したうえで、選択する。多数の見ず知らずの人たちの中から、誰かを選ぶとき、数値は比較を容易にするだけでなく、ほぼ唯一の選択手がかりとして機能してしまう。当然数値が高い人を選ぶだろうし、その人が、数値通り「優れている」なら、選んだ人はみな満足し、再び、高く評価する。データ駆動型社会は、そのようにして集積されるフィードバック情報をも集めていく。そうすると、選ばれた人は、高い数値を維持するのみならず、それがより多くの人に支持されているという認証も得ることになる。

このようなシステムは、口コミ、噂、評判など、人間が古くから用いてきた信用格付けのための社会の仕組みを、サイバー空間上に、大々的に構築したものになる。それは、現実の顔の見える相互作用とは、けた違いの規模での比較を可能にしつつ、一方で、サイバー空間上での限定的な相互作用に基づく数値を固定化させていく。実際、インターネット上で他者がつけた評価を手掛かりにして自らの選択を決めることは、私たちの生活の中に深く入り込んでいる。

このような序列化の仕組みは、よく知らない他者との相互作用に伴うリスクを回避でき、選択を容易にするという点で、非常に便利である。ただし、実社会での顔の見える関係の中で交換される情報とは異なり、匿名性が高く、不利な評判を否定する機会が得にくいいため、いったん低い評価を受けてしまうと、上位に上がる機会自体が失われてしまうかもしれない。また、第三者の意図的

な介入がもたらすゆがみに気が付くことも難しい。さらには、もともと持っているさまざまな資源の差が、初期の評判の差を生み、そのまま序列が固定化したり、集約される情報量の多さと蓄積のスピードの速さゆえに、一気に序列化が進むことも考えられる。

もちろん、不平等、格差という問題は、いつの時代でも付きまとうことであり、人類がそれから完全に逃れることができた社会は、今だ、形成されてはいない。とはいえ、弱者がより弱者になるような不平等・格差の拡大や序列化の促進は、人間中心というコンセプトにも相反することであり、Society 5.0やデータ駆動型社会が、これらの問題を拡大することになるのは、本末転倒だと言えるだろう。

3.3 多数派と倫理

格差や序列化は、データ駆動型社会で、序列が上のデータが影響力を持つことを含意するが、それと並んで力を持ち得るのが、「多数派」であろう。データの分析とそれに基づく介入に際しては、なるべく多くのデータを集め、多数派のデータを中心に分析、介入を進めることが方略の一つとなる。しかし、人々の行動の集約、また多数のふるまいは、記述や予測の精度を高めることには貢献するであろうが、それ自体が倫理的に正しいことを保証するわけではない。そもそも、人のふるまいの集約は、多くの人がそうふるまうという意味で、規範となりやすいが、多数派に物事をゆだねることが、必ずしも倫理的に望ましい結果をもたらさないことは、人類が歩んできた歴史を見ても明らかだ。同様に、データをAIに学習させ、その結果としてAIが行う判断を社会的な場面に用いるという事例においても、倫理的な課題がむしろ明るみになるような事例が存在する。

たとえば「Beauty, Ai」という2016年にネット上で実施されたコンテストでは、参加者の顔画像をAIが評価し、美しさを判定するという企画がたてられたが、その結果は、受賞者の大半が白人となり批判されるというものであった。AIが依拠している多数派に基づく通俗的な美の価値体系が、私たちの倫理観とは相いれない差別的側面を持っていたのである。また、「Tay」というマイクロソフトが開発したチャットボットが、差別発言を繰り返すようになり運用が停止された事例も挙げることができる。

私たちが日常行う行動や、またその背後にある価値観や態度は、必ずしも倫理的に望ましいものばかりではない。もっとも、軽い倫理違反は、多くの場合、見逃されるし、一人の行動にのみ着目しているなら、その倫理性はあまり問題にならないようなことも多い。しかし、それらが多くの人の行動として集約されてしまうと、「皆がそうしている」ということ自体に、非倫理性が生まれてしまう。あるひとりの人がコンテストで白人の人を

選択したとしても、それ自体は非難されないが、多くの人がそうしてしまうことにより、そのような判断の非倫理性があらさまになるのだ。ここに挙げた例は、いずれもそのような性質を持つものであり、多数派のデータを活用することに潜む問題を、暴いたものだと言えるだろう。

3.4 責任の所在に関する問題

社会における倫理のある部分は問題を起こした際の責任を問うことを通して維持されている。他者に対する加害的な行為を行った場合は、その責任が問われ、罰を受けたり、補償する義務を負ったりする。これにより、被害を受けた側における精神的・実際の被害が回復され、正義が維持される。その際、責任を負うとされるのは、被害をもたらすに至った行為を「自律的に行ったエージェント」である。自らの判断により行った行為で、他者が被害を被れば、行為者がとがめられ、罰を受けたり補償の責任を負うという構図は、日常の相互作用から法に至るまで、社会の中に深く組みこまれている。

データ駆動型社会においては、データ活用の結果として生じた被害が問題となるが、そこに「自律的に判断を行う」能力を持つAIに関わることにより、被害を発生させたことに対する責任を論ずる枠組みの再考を迫られる可能性がある。

従来、技術がもたらした事故については、開発者やその技術導入・管理に関わる主体（企業や自治体など）が責任を負うことが通常であった。つまり、技術は「製品」というモノの中に組み込まれており、製造や実装を通して、モノの振る舞い方を決める力を持つ人や組織が「エージェント」として責任を負うという構図であった。しかし、自律的な意思決定能力をもつAIが関わると、そもそも誰が「エージェント」なのかという問題が起こる。開発者（企業）なのか、AIなのか、それともデータ提供者と恩恵の享受者である市民なのかが問われるのである。また、AIが与えた被害に対する補償のコストを誰が追うべきかについても同様である。

この問いについては、補償の責務を開発者から切り離れた実務的な法的制度設計の必要や、加害を行ったエージェントに対する応報的な動機に基づく責任の追及を放棄するとともに、補償を社会全体で引き受ける仕組みなどが提案されている[10][11]。また学術的な議論のみならず、小説、映画、漫画などの表象文化においてもAIの位置づけは様々に描かれており、一種の思考実験がなされているともみなせる。例えば、その中の一つである「AIの遺電子」では、自動運転車のAIに人格のようなものを持たせ、事故を起こした際、それを破壊処分することで、応報感情を緩和するストーリーが提示されている[12]。

私たちは、いまだ、AIの加害を理解するテンプレートを持っていない。自律的なエージェントの行為に責任を負わせるという、道徳的直感にもかなった枠組みは、AIには通用しない可能性があるとき、どのような責任体系であれば受容可能なのか、またそれが、既存の責任、倫理といった体系とどのように整合的に存在し得るのかについて、今後の議論が必要になる。

4. データ駆動型社会における人間中心とは

データ駆動型社会が人間中心であるための課題として、本稿では心の特性を踏まえた行動の制御、およびデータ利活用に関わる倫理的問題の二点から考察を行ってきた。そこで焦点を当てたのは、個人の欲求や行動の統制を通し、個と社会の調和をはかりつつ、社会全体の倫理を維持する必要であった。

行動の統制や倫理という言葉からは、ある種の不自由さを感じるかもしれない。しかし、人間を中心とするというのは、私と同時に他者も中心となることだ。ここでは、持てる資源、また考え方や価値観が多様な人間が、ともに生きる社会のあり方が真剣に問われる。個人の欲求が満たされることの延長に幸福があるというように Society 5.0 を描き、データの利活用がそれを実現すると喧伝することは、それに反することだ。未来が私たち個人、また社会全体に課す制約について、誠実に明らかにし、そのあり方を探る議論を重ねていかねばならない。

また、その議論は、データ駆動型社会において、一人ひとりの「私」がどのように位置づけられるのかという問題と切り離せない。「私」という各個人からデータが収集され、それがサイバー空間上で蓄積され活用される場において、「私」、ひいては、私や他者を含む「個人」がどのように扱われるのかということである。

実社会に生きる「私」という存在は、自分の歩んできた歴史の上に成り立つ、主体性を持つかけがえのない存在として、実空間に生きている。一方、データ駆動型社会が構築するサイバー空間における「私」は、データの集約としての存在である。サイバー空間上の「私」は、携帯端末などから集められた移動、商品の購入、SNSなどでの他者とのコミュニケーションに関わる履歴、また、これらから把握される身体の状態、考え方や価値観、性格などを表す大量のデータである。

このようなデータの集合体である私が、ビッグデータ内の一つのデータに過ぎないとされるのなら、実空間と同様のかけがえのないさや主体性を見いだすことは難しくなる。また、実空間では「私に関する知識」は、(その多くについて) 私が独占的に知るものであったが、そのような自己知に対する特権的な地位も脅かされ、サイバー

空間にあるAIのほうが、私よりも私のことを「知っている」事態が生じるかもしれない。

データ駆動型社会が必然的にもたらすこのような「私」のあり方は避けがたい現実であるが、それを踏まえてなお、「人間中心」を実現していくことが要求される。サイバー空間上の「私」(そして他者)は活用されるデータだが、実空間とのつながりの中で、どのように主体性を担保できるのかを考えることが、その鍵となるだろう。そのためには、技術が問題を解決するのだから、それを人々は享受すればよいという、パターンリスティックな目線の下に「人間」を置くのではなく、「駆動される」社会のあり方にも責任を持つ存在として、位置づける必要があるのではないか。

技術の革新が自分の満足や幸福を高めてくれるという期待を人々が持つことは、データ駆動型社会を推進する大きな社会的動機になる。しかし、その期待から主体性が失われ、個々人の欲求が満たされることを、保証してくれる社会への期待に転化してしまうと、人間中心から遠ざかる。行動の制御が心の特性から乖離しない、また、倫理が軽視されることのないデータの利活用により、満足や幸福の健全な着地点を探りながら、人間中心のデータ駆動型社会を構築していくことが求められる。

参考文献

- [1] https://www8.cao.go.jp/cstp/society5_0/index.html
- [2] J. R. P. French and B. Raven 'The bases of social power' in D. Cartwright (ed.) *Studies in Social Power*. Ann Arbor, MI: University of Michigan Press, pp. 259-269 (1959).
- [3] 山岸俊男: 社会的ジレンマ―「環境破壊」から「いじめ」まで, PHP 新書 (2000).
- [4] R. M. Ryan and E. L. Deci: *Intrinsic Motivation and Self-Determination in Human Behavior*, Springer US (1985).
- [5] R. Hertwig and T. Grüne-Yanoff: Nudging and Boosting: Steering or Empowering Good Decisions, *Perspectives on Psychological Science*, 12, pp. 973-986 (2017).
- [6] J. Mills and J. M. Jellison: Effect on opinion change of how desirable the communication is to the audience the communicator addressed, *Journal of Personality and Social Psychology*, 6(1), pp. 98-101 (1967).
- [7] S. S. Brehm and J. W. Brehm: *Psychological reactance: A theory of freedom and control*, New York, Academic Press (1981).
- [8] J. Haidt: The emotional dog and its rational tail: A social intuitionist approach to moral judgment, *Psychological Review*, 108, pp.814-834 (2001).
- [9] W. Youyou, D. Stillwell, H. A. Schwartz, and M. Kosinski: Birds of a feather do flock together: Behavior based personality assessment method reveals personality similarity among couples and friends, *Psychological Science*, 28, pp. 276-284 (2017).
- [10] 小林正啓: 自動運転車の実現に向けた法制度上の課題, *情報管理*, 60, pp. 240-250 (2017).

- [11] 戸田山和久：哲学の側から Let's 概念工学，戸田山和久，唐沢かおり（編著）：＜概念工学＞宣言，pp. 3-36. 名古屋大学出版会 (2019).
- [12] 山田胡瓜：第 76 話 ある AI の結末，AI の遺電子 07, pp. 175-190, 秋田書店 (2017).

唐沢 かおり



1992 年京都大学文学研究科博士過程退学。2010 年より東京大学人文社会系研究科教授。現在に至る。社会心理学，社会的認知，道徳的判斷，AI への態度などに関する研究に従事，Ph.D (Psychology)。日本社会心理学会会長。
