

AIの〈自然〉、 AIへの〈作為〉

山本龍彦 慶應義塾大学法科大学院教授



1. 自然と作為

AI（人工知能）を中心とした先端的な情報技術によって、多様でインクルーシブな社会が創造されるという。確かに、使い方によってはそうだろう。私もその可能性に期待する者の一人である。しかし、こういうユートピア的世界を期待するならば、AIの〈自然〉を真剣に考える必要がある。つまり、あるがままのAIとは何か、あるがままのAIをあるがままに受けとめる社会とは何か、という思惟である。

近代社会とは、丸山眞男いわく、人間のあるがままを超克しようとする〈作為〉的空間である。人間が定めた目的のために、人間が主体的に選択・決断し、〈自然〉に引き戻されまいと無理をし続ける人為的な空間である。そういう煩わしさがあってもなお丸山が「近代」にこだわったのは、「何となく何かに押されつつ、ずるずると国を挙げて戦争の渦中に突入した」、あの戦争の悲劇的経験による。

本稿の目的は、AIの〈自然〉とは何かを紙幅の許す限りで考察し、「ずるずると」AI自然主義的世界に「突入」するのか、それとも、これに抗して「我々」が何かを選択し、その維持のために多少の無理をし続けるのかを熟議するきっかけを提供することにある。



2. More Dataという欲求

AIの〈自然〉を考える上で何より重要なのは、AIの存立基盤、成長条件、存在意義とは何かをイメージすることである。

AIの存立基盤、成長条件としてまず想起されるのは、データの量だろう。AIがより深くこの世界を学び、より賢くなるには、「より多くのデータ（more data）」が必要になる。データ量が限定され、AIを訓練する教師データのクオリティが落ちれば、AIの予測精度は落ち、その存在意義も失われる。だから、より多くのデータを欲し、より正確な予測を行おうとするのがAIにとっての〈自然〉と言い得る。

こうしたAIの〈自然〉と根本的に対立し得るのが、プライバシーである。近代社会は、「公」と「私」を切り分け、私的領域を見ぬもの、干渉せぬものとしてきた。私的領域では人間は本来的に自由であり、そのライフスタイル——何を「善い生き方」とするか——に関する個人の選択は尊重されなければならない、と。他方、公的領域では人間は共同体を支える「市民 (citizen)」として、道徳的かつ (H. Arendt的な意味において) 政治的に振る舞わなければならないとされた。しかし、「もっとデータを」というAIの〈自然〉的欲求は、近代社会が前提としてきた「公／私」の壁を打ち崩す。例えばAIが、個人Xの社会的信用力を正確に予測するよう命じられれば、公的領域におけるXの道徳的・演技的振舞いだけでなく、私的領域におけるXの自然的・動物的な姿までも捕捉しようとするだろう。その方がXのありのままの姿を捉えられるからである。近年、HRテック (Human Resources Technology) 領域で、労働者の職務適性や離職率などを予測するAIシステムが設計・利用されているが、本来的にAIは、その正確な予測のため、労働者のデータをシームレスに、すなわちそれがオフィスでの行動であろうと、トイレでの行動であろうと、住居内での行動であろうと食べたいはずである。

我々の私的居住空間を快適化するAIスピーカーやAI家電は、我々の私生活を監視するのに最適な装置ともなるのである。

AIの旺盛な食欲によって、かつ公私を隔てる境界が失効することで、我々のあらゆる行動が社会的評価の対象となり得る。これもまた〈自然〉である。例えば、従前、家の中で誰と何を食べるかが公的決定の基礎となることはなかった。しかし、例えばAIが、ある個人の将来の健康リスクを正確に予測するよう命じられれば、「家の中で誰と何を食べたか」までもが捕捉され、結果的に個人の極めて私的な行動が健康リスク評価と結び付けられることもあるだろう。前述のように、AIの存在意義はその高い予測力にある。したがって、我々のあらゆる行動——言動といった方が正確か——が「データ化」され、AIが学習できる言語へと「翻訳」され、公と私、あるいはサイバーとフィジカルの境界が徐々に融解されるという方向が〈自然〉なのである。

我々はこの〈自然〉をあるがままに受けとめるべきだろうか。

3. 開拓される心理

ディープニューラルネットワーク

(DNN:人間の脳神経系のニューロンを数理モデル化したニューラルネットワークを多層に重ねる深層学習の方法)など、AIの学習方法がより複雑・高度化していけば、無意識領域も含めた我々の心理状態までもが分析の対象となり、「社会化」されていくのが<自然>である。近代の法体系は、確立し、完成した「意思」(法学でいう「効果意思」)を前提に構成され、この「意思」が個人によって——他に視かれ、また他に影響されることなく——自律的に形成されるものと仮定されていた。したがって、「意思」が形成される以前の内心過程は法学的議論の対象とはならず、当然、保護の対象ともならなかった(民法の錯誤論も、『効果意思』という確固たる意思が存在するとの前提)に立ち、「それ以前の意思形成過程」は『動機』として捨象された。大村敦志『民法読解 総則編』(有斐閣、2009年)340頁。また日本国憲法は19条で「思想及び良心の自由」を保障するが、通説によれば、同条で保護されるのは既に確立している「信条」である)。しかし、情報技術の発達と相まって高度化するAIの予測力は、こうした前提を突き崩し、その対象を、意思形成以前の心理領域にまで拡大させていく。「いかにブックメーカーが、ギャンブラーをはまらせておくためにAIを使っ

ているか」というタイトルのガーディアン紙の記事が、「ギャンブル業界は客をプロファイリングして、新たな恐ろしい方法(in frightening new ways)でその客の行動を予測しようとAIを使っている」というマーケットの言葉を紹介しているのは、AIの発展によって心理領域がビジネスの新たな狩場になることを饒舌に物語っているように思われる(Mattha Busby, *Revealed: how bookies use AI to keep gamblers hooked*, *The Guardian*, 30 April 2014)。また、“Brexit”を問う2016年のイギリス国民投票や、トランプが勝利した2016年アメリカ大統領選に深く関与したとされるケンブリッジ・アナリティカ社(選挙コンサルタント会社)が、Facebookユーザーらの心理的プロファイリングを行い(フェイクニュースに対する脆弱性等の予測)、選挙運動に加担したことはよく知られている。もちろん、彼らは例外ではない。例えば、アメリカのターゲットスマート社も、SNSデータなどから有権者の支持政党のみならず、権威主義的な性格や同情的な性格などをプロファイリングしており(日本経済新聞2018年10月21日2面)、AIによる心理領域の開拓は、選挙運動の中心的手法となりつつある。

以上のように、AIの予測力は、嘘発見器を過去のものにする。とすれば、<自然>

に従えば、「意思」の前領域である心理状態が可視化され、社会化するのも時間の問題であるといえよう。

我々はこの〈自然〉をあるがままに受けとめるべきだろうか。



4. スティグマとしての「スコア」

前述した「more data!」というAIの自然的欲求と、近年よく指摘されるAIの判断過程のブラックボックス化が、社会的身分の固定化ないし新たな社会的カーストをもたらすのも、また〈自然〉かもしれない。例えば、個人の社会的信用力と人種・居住地等が関連するならば、AIはこれを考慮するのが〈自然〉だろう。しかし、人種や居住地が社会的信用力と関連するのは、過去に人種的マイノリティに対する偏見や差別があったからである。この偏見や差別のために、マイノリティは社会的・経済的な不利益を受け、特定の地域に暮らさざるを得ず、時に信用を損ねるような悪行をなさなければならなかったのかもしれない。こうした社会構造を捨象して、単に関連があるとの理由でAIがこれを考慮すれば、マイノリティの社会的信用力はマジョリティよりも一般に低く算定されることとなり、過去の差別が繰り返され続ける

ことになる。つまり、マイノリティはマイノリティとしてその社会的身分を「固定」されることになるわけである。

また、仮に、ある遺伝情報が社会的信用力や知性と関連するとされた場合でも、AIの〈自然〉からすれば、関連する以上それを考慮すべきとなるだろう。関連があるのに考慮しない方が、予測精度を敢えて下げる自殺行為なのであって、いかにも〈不自然〉である。しかし、遺伝情報なるギフトは、本人がその努力によって修正・変更できない「生まれ」に関する属性であるから、AIによるその考慮は「生まれ」による不利益をもたらすことにもなる。近代社会は、封建的身分制の否定から始まった。それはすなわち、「生まれ」によって個人の生き方が規定されることを否定し、個人がその生き方を自ら選び取ること、そしてその結果に対する責任を自らが負うことを肯定するものであった。それこそが、日本国憲法13条のいう「個人の尊重」の考え方である。「more data」に従い、その予測精度のために関連するものはすべて考慮するというAIの〈自然〉にゆだねれば、個人は自らが修正・変更できない属性によってその人生の可能性を狭められることにもなる。

もちろん、AIは人種や居住地、遺伝情報のみで人間を「格付け」するものではない。む

しろそれ以外の情報も含め、多元的・多角的に個人を評価することにより、こうしたセンシティブな要素が過度に強調された、偏った——人間による——判断を補正することができる。我々は、このことを忘れるべきではない。これがAIを用いたインクルーシブな社会の実現につながる。しかし、多元的な変数のそれぞれにどのぐらいのウェイトをかけたのかなどが「ブラックボックス」し得ることも忘れるべきではない。これも、AIの宿命、あるいは<自然>といえる。

この「ブラックボックス」には少なくとも2種類のものがあり得る (Guido Noto La Diega, *Against the Dehumanisation of Decision-Making*, 9 JIPITEC 3, 9-10 (2018))。1つは「法的ブラックボックス (legal black box)」である。これは、AIシステムないしアルゴリズムを設計した企業が、その設計に際してどのようなデータを考慮し、または排除したのか、当該データにどれぐらいのウェイトをかけたのか、どの情報を強調し、または過小評価 (deemphasize) したのかを「取引の秘密 (trade secret)」ないし知的財産として、または、これを開示すると被評価者によるゲーミングが起き (敢えてアルゴリズムに合わせた行動をとるようになる)、業務に支障が出るなどとして、秘密＝不開示にすることから生じる。ただ、

おそらくより重要なのは、「技術的ブラックボックス (technical black box)」である。これは、先述したDNNのような深層学習方法を用いると、AIによる判断のロジックに対するアクセスが極めて困難になることから生じる。「法的ブラックボックス」が仮に解けたとしても、AIの「思考」が高度化・複雑化することによって、AIがなぜそのような判断を下したのか、人間には理解できなくなるわけである。

そうすると、AIによって社会的信用スコアを低くつけられた者は、そのスコアが「生まれ」の影響をどの程度受け、他の無意識的な行動等の影響をどの程度受けたものなのかがわからず、何をすればスコアを上げられるのかをつかめずに、ロースコアのまま社会の下層を漂い続ける可能性もある。そもそも一旦低いスコアをつけられると、例えば就職で不利に扱われて定職に就けず、そのことがまた低いスコアへとつながり、そしてそれが住居を借りることを難しくし、そのことがまた低いスコアへとつながり、今度は……など、「負のスパイラル」に陥る可能性が高い (Daniel Keats Citron & Frank Pasquale, *The Scored Society*, 89 WASH. L. REV. 1, 32 (2014))。中国では既にアリババ・グループの信用情報機関 (芝麻信用) のスコアリングサービスが広く普

及しているが、実際、この信用スコアが低いと、融資や採用場面で不利に扱われるだけでなく、航空機チケットの購入が困難になったり、外国のビザ取得が困難になったり、移動の自由が制限されることもあるという（「婚活」などでも不利益を受けるという）。芝麻信用の事業でも、やはりスコアリングの具体的な基準は明らかにされていないため、結局、低いスコアは、現代的な「ステイグマ」ないし「緋文字」として機能する可能性がある。

こうして見ると、「more data」と「ブラックボックス」というAIの〈自然〉によって、スコアを新たな「身分」とした階層社会——スコアの高低により生き方が規定される社会——が生まれるのが〈自然〉の流れとも考えられる。もちろんそれは、中国の現在の状況が物語るように、安全な社会の実現につながるという積極的側面も有する。AIへの「給餌」のため、縦横に張り巡らされた監視ネットワークにより常に情報を取られ、それらの情報が信用スコアに何らかし連動するとなると、我々は起床から就寝まで——あるいは睡眠中までも（？）——スコアの上げ下げを気にしてお行儀よくしておかなければならないからである。しかし、我々はそれが、自らの行動の帰結に対する予測不可能性と、それに基づく萎縮と

緊張によって生まれる、「怖れ」に支えられた安全であることを忘れるべきではない。

我々はこの〈自然〉をあるがままに受けとめるべきだろうか。



5. AIへの「作為」——AIの「民主化」

以上、本稿は、AIの〈自然〉と、この〈自然〉をあるがままに受けとめた社会とは何であるのかにつき、ささやかな検討を加えてきた。また、その各検討の最後では、我々はその社会をそのようなものとして承認すべきかを問うてきた——我々は、AIの〈自然〉に身をゆだねるべきか。

もし、この〈自然〉に抵抗し、プライバシー（公私区分）、意思主義、責任主義、あるいは「自由」といった近代の主要なコンセプションを肯定し続けようとするならば、我々は〈作為〉を、つまり主体的で政治的な主権者的決断を行う必要があるだろう。そうでなければ、「何となく何かに押されつつ、ずるずると国を挙げて」、効率的で安全だが、予定調和的で固定的な階層社会へと「突入」していくことになるだろう。

では我々は何をすべきか。

紙幅の都合上、ここではAIの「民主化」ということだけを述べておきたい。AIの

<自然>は、「more data」であり、さらなる予測精度の向上であり、ブラックボックス化である。AIの「民主化」のためには、これらを近代立憲主義の主要なコンセプトと突き合わせ、両者のトレードオフを可視化すること、これを開かれた討議の俎上に載せることが重要である。例えば、予測精度の向上に対しては憲法上の「公正」概念をぶつけ、その両者の均衡点をどこに打つのかを民主的に議論すべきである。実際、アメリカのシカゴやフィラデルフィアなどで使用される犯罪予測システム（HunchLab）は、マイノリティに対する偏見・差別の固定化や助長を避けるため、マイノリティが特に手を染める傾向のある軽犯罪歴等を「過少に見積もる（deemphasize）」アルゴリズムを使用している（このアルゴリズムには、偏見への対抗のため多少のランダム性も組み込まれている）。またペンシルバニア州アレゲニー郡の児童虐待予測システム（AFST）にも、「人種は考慮しない」との政策的な判断が組み込まれている。このような予測精度と「公正」との調整・トレードオフは、単純な技術的問題ではなく、ハードな実体的価値判断を含んだ政治的な問題である。したがって、これをシステム設計にかかわる技術エリートのみで議論するのではなく、多様なアクターを包摂しながら民主的に議論することが重要

である（詳細は、山本龍彦「AIを『民主化』する——アルゴクラシー vs. デモクラシー」vivoice3月号（2019年））。ブラックボックスについては透明性、もしくは、透明性によって担保される近代の主要なコンセプト（責任主義や「自由」）がトレードオフの関係に立つ。これも可視化し、主権者的決断の俎上に上げるべきだろう。

AIシステム、アルゴリズム、コードは、今後は「法」と同じように、あるいはそれ以上に我々の行動に影響を与えていく。だとすれば、その設計プロセスをAIの<自然>にゆだねるのではなく、人為的に、また民主的にデザインしていくべきである。

AIのあるがままをあるがままに受け入れることがどのような社会を生み出すのかを強くイメージしながら。

プロフィール.....
やまもと・たつひこ 1976年東京都生まれ。慶應義塾大学法科大学院教授。慶應義塾大学グローバルリサーチインスティテュート（KGRI）副所長。博士（法学）。1999年慶應義塾大学法学部法律学科卒業。2001年同大学院法学研究科修士課程修了。2005年同大学院法学研究科博士課程単位取得退学。桐蔭横浜大学法学部専任講師、同准教授を経て現職。他に、司法試験審査委員（2014、15年）、ワシントン大学ロースクール客員教授（2017年）総務省AIネットワーク社会推進会議（AIガバナンス検討会）構成員（2018年～）。主な著書に『AIと憲法』（共編著、日本経済新聞出版社、2018年）、『おそろしいビッグデータ』（朝日新聞出版、2017年）、『プライバシーの権利を考える』（信山社、2017年）他。