

## AGI と社会と人類と

大屋雄裕<sup>1</sup> (慶應義塾大学)

### 要 旨

AGI が実現した社会において想定すべき問題について、第一に AGI を法的人格として承認するか、した場合にその存在としての性格の違いからどのような問題が想定できるかという側面と、第二にその存在が我々人類の教育という営為にどのような影響をもたらすかという側面に分けて論じる。その際、人間の条件をめぐる議論（パーソン論）においては義務論的に論じられる本質的人格性と帰結主義的に正当化される手段的人格性を区別して論じる必要があることを指摘し、またネットワーク化された AGI において「個体」の概念が想定可能かを検討した。また、教育においてしばしば例示と反復というモデルが用いられているところ特に単純・容易な段階の作業が AGI に代替されることからその可能性が失われる危険性があることを指摘した。仮に反復を超える価値観のアップデートが人類の課題になるとすれば、その能力を習得するための明示的手法が必要になることを結論としている。

キーワード：AGI、人格性、人類、教育

### 1. はじめに

人工知能 (AI; Artificial Intelligence) 技術の急速な発展は、我々の社会のあり方をすでに変貌させている。個人の日常でいえば AI を活用したリコメンデーションを抜きにしてさまざまなストリーミング・サービスが提供する膨大なコンテンツから適切なものを選択することは困難になっているし、産業界でも AI によってトラックによる輸送ルートや運転手の労働強度を最適化するような活用が試みられている。さらに 2022 年から急速に発展した生成系 AI (generative AI) は楽曲・画像・動画の自動生成によって音楽産業や映画産業に影響を与えはじめ、アメリカ・ハリウッドではこれによって人間の仕事が奪われることを懸念した俳優・脚本家などの労働組合が 4 ヶ月にわたるストライキを行なう結果になった<sup>2</sup>。また、2024 年のアメリカ大統領選挙においては生成系 AI により合成されたバイデン大統領の音声が発話に使われるという事件が実際に生じ、選挙の公正が脅かされる危険が生じるなどその影響が我々の民主政プロセス全体に及ぶ懸念が生じている<sup>3</sup>。

文章・画像・音楽・映像といったメディアの垣根を越えるマルチモーダルな入出力を生成

<sup>1</sup> 慶應義塾大学法学部法律学科教授。

<sup>2</sup> 報道の一例として、「ハリウッド俳優のストライキが終了、将来的な AI 利用や配信ビジネスにとって重要な転換点になる」WIRED (2023 年 11 月 10 日)、<https://wired.jp/article/hollywood-actors-strike-ends-ai-streaming/>

<sup>3</sup> 「「偽バイデンからの電話」は始まりにすぎない。生成 AI による偽情報が米大統領選に与える影響」WIRED (2024 年 1 月 29 日)、<https://wired.jp/article/biden-robocall-deepfake-danger/>

系 AI が可能にしてきていることは、従来のように特定の用途を前提として開発される AI を超えた行動の汎用性や自律性を持つ汎用人工知能 (AGI; Artificial General Intelligence) の到来を予感させるかもしれない。その実現時期 (あるいは実現可能性それ自体) に対する考え方はさまざまだが、たとえばそれを 2045 年に措定することも不自然ではないと思われる。

本稿では、そのように AGI が実現した社会が到来したとして、そこに向けて我々が想定しておかなければならない問題について述べることにしたい。ただすでに他の機会に論じている内容も多いので適宜重複を避け、主にこれまで明確に述べていなかった点を重点的に扱うことを試みる<sup>4</sup>。

さて、さしあたり AGI の定義については、『AI 事典』の記述に従い「設計時の想定を超えた未知の多様な問題を解決できる知能」としておこう<sup>5</sup>。そこで目指されているのは汎用性の実現であり、なかでも人類レベルの汎用性を持つ AGI の実現が当面の目的となっていると言ってよい。もちろん可能性の問題と言うならば、実現された汎用性が人類と同等のレベルであってもこれまで人類が負ってきたさまざまな制約、典型的には特定の課題に投入し得る時間・労力の制限を超えた探索が可能になることによって従来のレベルを超えた知の実現が可能になることや、さらに人類を超えた**超知性**が実現することも展望されていなくはない<sup>6</sup>。だがそれらは将来的な課題として、当面はいま述べたように人類に相当するレベルの汎用性の実現された時代のことを考えることにしよう。

そのような AGI の我々にとって近いイメージは、たとえばロビタやドラえもんのような創作物で描かれてきたロボットの姿かもしれない。もちろんこれら (彼ら) のように物理的な肉体 (を代替する機械) を備えている必要はおそらくないし、これらのキャラクターは自律性だけでなく粗忽さやトラウマのような一定の性格的な**ゆらぎ**を帯びており、その故に「人間らしい」と表現されてきたのだが、このような性格を必然的に帯びている必要もないだろう<sup>7</sup>。するともう一つの可能性は『2001年宇宙の旅』(スタンリー・キューブリック監督、1968年)に登場する AI・HAL9000の姿かもしれない。作品中で HAL9000 の取った選択は矛盾する命令を受け取ったことによる異常動作だとしばしば理解されているが、それは我々の生存が尊重されるべきことを当然視する人類的な倫理を前提した評価であり、実際にはそれを**前提しない合理性**によって最善の選択 (乗組員の物理的排除) を導いたと考えることもできるだろう。AGI のいる社会へのイメージは、この両極端のどちらに立脚す

---

<sup>4</sup> 以下、homo sapiens という種とそれに属する個体を意味する場合は「人類」または「ホモ・サピエンス」、社会の成員として認められる主体を意味する場合には「人間」と書き分けることとする。

<sup>5</sup> 山川宏「汎用人工知能とは」中島・浅田他編『AI 事典 [第3版]』(近代科学社、2019年) 50頁。

<sup>6</sup> 代表的な例として、Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014 = 倉骨彰 (訳) 『スーパーインテリジェンス—超絶 AI と人類の命運』(日本経済新聞出版社、2017年)。

<sup>7</sup> ロビタについては、失敗する・感情らしいもので機嫌が変わるといった特徴があり、それが人間らしさとして受け止められたと作中では描写されている (手塚治虫『火の鳥 復活編』(1970~71年))。またドラえもんについては、かつてネズミにかじられて両耳を失ったことから病的な恐怖を覚えている描写がある (藤子・F・不二雄『ドラえもん』(1969~94年))。

るかによって変わってくると考えることもできる。

いずれにせよこのような存在が我々の社会に立ち現れてきたとき、そこにはどのような問題が起きるのだろうか。以下ではそれを、第一に我々の社会における AGI の位置付けの問題——すなわち我々が AGI をいかなる存在として社会の内部で扱おうとするのかという問題——と、第二に AGI のいる社会における人類の問題に分けて議論することにしよう。

## 2. AGI と人間のあいだ

第一の問題は、端的に言えば我々の社会において人権享有主体として認められている人間を定義する条件のいくぶんか、あるいはすべてを AGI が備えるに至ったとき、我々は彼らを人間として取り扱わなければいけないのではないかということである。また、その帰結を引き受け、AGI に人権享有主体性を認めた場合にはその存在としての性質が我々人類と大きく異なることをどう処理するかということも議論しなければならないが、いずれもすでに論じたことがあるのでここで長くは繰り返さず、一定の補足的な議論のみをしておくことにしよう。

### 2. 1. 人間の条件

まず人間の条件をめぐる問題は、倫理学においてこれまで主に動物を対象とした「パーソン論」——人格 (person) として扱われるべき根拠をめぐるものとして問われてきた。そしてまずそれを理性や言語使用のように動物にはない条件に設定すると、人類の特権性が守れる一方でそれらを欠くホモ・サピエンス個体 (典型的には遷延性意識障害の患者や重度心身障害者) が排除されてしまうという過少包摂の問題が生じる。他方、受苦性 (痛みや苦しみを感ずる能力) のように彼らを排除しない条件にすると、ある程度以上発達した中枢神経系を持つ生物も対象となる結果として、肉食や畜産といった動物の利用が不可能になるという過剰包摂の問題を抱えることになる点が課題とされてきたのである。もちろんどちらについても**問題だと思ふか自体が問題**であり、歴史的に言えば非白人がしかるべき理性を備えていないという理由から人権享有主体性を否定されてきた (あるいは限定的に承認されてきた) ような例もあれば、動物の利用すべてが道徳的に正当化されないので禁止されるべきという立場の論者も現実に存在する<sup>8</sup>。

AI を対象とすると問題の構図がまさに逆転する点——AI は理性や言語使用において人類 (の少なくとも一部) より優位にある一方、受苦性や可傷性を持たない——が重要だという指摘はすでにしたのでここでは措く<sup>9</sup>。以下では将来的な議論のために、法的人格をめぐる議論においては本質的人格性と手段的人格性の相違を十分に踏まえる必要があることの

<sup>8</sup> たとえばゲイリー・L・フランシオンはこのような観点から動物の所有と制度化された利用すべてを否定している。Gary L. Francione, "Animals --- Property or Persons?" in: Cass R. Sunstein & Martha Nussbaum (eds.), *Animal Rights: Current Debates and New Directions*, Oxford University Press, 2004. ch. 5 = 土屋裕子 (訳) 「動物は財産か、人格か？」安部圭介・山本龍彦・大林啓吾 (監訳) 『動物の権利』(尚学社、2013 年) 第 5 章。

<sup>9</sup> 大屋雄裕「外なる他者・内なる他者—動物と AI の権利」論究ジュリスト 22 号 (有斐閣、2017 年) 48-54 頁。

みを指摘しておくことにしよう<sup>10</sup>。

## 2. 1. 1. 本質的人格性

本質的人格性とは、当該対象が人格を有すべき条件を備えている以上必然的に人格性が承認されるべきであるという議論を意味している。たとえば動物の権利論においては、犬や猫といった動物が人類と同じように苦痛を感じる能力（受苦性）を備えている以上、理由なく苦痛を与えられない権利を人類と同じように持つべきだといった形で議論が展開されることが多い。一定の能力と関連性（relevance）を持つ条件を対象が有しているかという点のみが注目されていることが重要であり、その結果としてどのように問題のある帰結が導かれるとしても、それが倫理的である以上やむを得ないということになる。たとえば前述のように受苦性を人間の条件として採用すると動物の利用が完全に禁止され、その結果として一部の人類の生存には不都合が生じるかもしれないが、しかしその結論が倫理的なのだから我々はそれを甘受するべきだというわけだ。別の言い方をすれば、本質的人格性をめぐる議論は義務論（deontology）に基づいており、帰結を問わず一定の「正しさ」がそこに備わっているために我々はその結果を引き受けなければならないということが含意されているのである。

## 2. 1. 2. 手段的人格性

これに対し手段的人格性とは、一定の対象に人格性を認めることが我々にとって何らかの便宜を提供するという理由に基づいて正当化される議論のことである。法人一般の実在性については特に民法領域で伝統的に論争が行なわれており、かつ明快な結論を得ることにつながらなかったのだが<sup>11</sup>、たとえば数人の共同出資により一定の事業を営むことを目

---

<sup>10</sup> 法律学においては一般的に、我々ホモ・サピエンスの個体を「自然人」と呼び、生来的・必然的に人権を保有する主体と位置づけたうえで（たとえば民法3条1項「私権の享有は、出生に始まる」）、それらと一定の法令に基づいてその存在が認められる「法人」とが法律上の人格、すなわち権利能力・意思能力・行為能力を備えた主体であると整理している。また、そこで法人が有する人格のことを「法人格」と呼ぶ。以下の議論では、混乱を避けるため自然人と法人が共通して持つに至る法律上の人格性のことを「法的人格」または「人格性」と呼んでいる。

<sup>11</sup> たとえば内田貴『民法I—総則・物権総論〔第2版補訂版〕』（東京大学出版会、2000年）は法人学説——法人の本質をめぐる擬制説・有機体説・組織体説などのあいだの争いについて、「今日のように、法人が重要な経済主体として活動し、それに関する法技術的装置が完成している法体系のもとでは、現実の問題を解決するための解釈論には直結しなくなっており、その意味では、このような論争を行なうことの意味自体が失われている」（212頁）と冷淡に述べている。また水津太郎は団体の実在性をめぐる安藤馨と筆者の議論に関連して、法人本質論に対する関心が現代の民法学で高くない理由を「ある意味において社会的実体がなければ、その法的取扱いはそもそも問題とならない。他方で、法人格が法によって与えられたものであるのは、当然のことである。つまり、法人本質論に関する二つの学説は法人の異なる側面を捉えたものにすぎず、実にか擬制かを問うのには意味がない」からだと説明している（水津太郎「コメント」安藤馨・大屋雄裕『法哲学と法哲学の対話』（有斐閣、2017年）84頁）。

的としている場合、法人を設立しない手法が存在しているにもかかわらず<sup>12</sup>会社という法人形態が発展してきたのには、典型的には営業の安定性・持続性・発展可能性などにおいて**その方が便利だから**という理由が挙げられるだろう。ここでは、一定の制度を通じて法的人格を想定することから得られる**社会的利益**がその正当化根拠として想定されていることになる。

もっとも極端な例の一つである相続財産法人について考えてみよう。これはある人（被相続人）が相続人の有無がわからない状態で死亡した場合に民法 951 条の規定に基づいて自動的に成立する法人であり、その構成要素は被相続人の財産のみである。所有者であった被相続人の死亡によってその財産の所有者がいなくなったと考えると、動産については無主物先占（同 239 条 1 項）の対象となりいわば早い者勝ちで獲得できることになるが、当然ながら大きな混乱を招くことになるしのちに相続人がいることが判明するとそれらの動産を返却しなければならないことになるだろう。このような不都合を避けるために必要だと考えられたのが相続財産法人であり、別に選任される清算人の手によって管理されることになる（同 952 条・953 条）。最終的に、相続人が判明した場合には被相続人の死亡にさかのぼって相続が発生することになり、当法人は「成立しなかったものとみなす」ことになっている（同 955 条本文）<sup>13</sup>。社会的不利益を回避するために設立され、順調に相続人が発見された場合には**最初からなかったことになる**相続財産法人は純粋に法技術的な装置であり、手段的人格性によって正当化されているものと考えられることができるだろう。

これも別の言い方をすれば、そこで問題になっているのは当該制度がもたらす帰結のみであり、対象が備える本質ではない。一般的な株式会社や学校法人であればそれを構成している複数のホモ・サピエンス個体があり、彼らが織りなす精神的・肉体的活動によって擬制される法人の意思や行為が存在するだろう。その意味でそこに意思や行為の実体があると考えるのも当然である。しかし相続財産法人のように内的な活動を一切持たない財産だけの存在であっても、一定の必要性があれば我々はそこに法的な人格性を措定することができる。本質的人格性とは逆に、ここでは制度の社会的価値のみを考慮し対象の本質を無視する帰結主義（consequentialism）が採られているのである。

### 2. 1. 3. AI と法人格をめぐる議論

これまでも AI に法人格を付与するとか認めるという主張はたびたび行なわれてきたが、それらがこのどちらの視点に立っているものかについては区別して議論する必要がある。2016 年には欧州議会法律問題委員会からロボットに対して「電子人」(electronic person)としての地位を与えることが提案されたが (31(f)項)<sup>14</sup>、将来的にロボットの自律性が向上

<sup>12</sup> 民法上の組合（667 条以下）、商法に定める匿名組合（535 条以下）が典型例として挙げられる。

<sup>13</sup> 相続人がいないことが確定した場合には、相続債権者・受遺者に対する精算を行ない（民法 957 条）、被相続人の療養看護に努めたなど特別の縁故があったものから請求があった場合には相続財産の全部または一部を分与したのち（同 958 条の 2）、残余の財産については国庫に帰属することになる（同 959 条）。

<sup>14</sup> European Parliament Committee on Legal Affairs, "Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))", May 31st, 2016, PR - PE582,443v01-00.

した場合に「既存の法的カテゴリー——自然人、法人、動物又は物体——に照らしたその本質、あるいは新たなカテゴリーが作られるべきかという問題」が浮上することを認めてはいるものの（前文 T 項）、現時点では課税の便宜やロボットが**自動的に動作する**という意味においての自律性を得たことにより人間に損害を与える可能性が出てきたことへの対応として損害賠償等の権利問題を解決する手段と位置付けられたものであり、基本的には手段的人格性に関するものと理解すべきだろう。

他方、ロボットについてローマ法上の奴隷の位置付けを参照した限定的な行為能力を認める構想についてウゴ・パガロが論じている際には、ロボットの性質が動物とも通常の自然人とも異なる中間的な領域にあるという根拠が挙げられており、本質的人格性に関する議論と位置付けることができる<sup>15</sup>。

「法人」という言葉のあいまいさ——ここまで論じてきた法的な人格性のことを意味しているのか、人格性を制度的に認められた**法人**という部分集合のことを問題にしているのか——もあり、特に法律家以外の人々がこの問題を論じる際には混乱が生じがちであるように思われる。手段的人格性は現時点の AI に対しても一定の便宜があれば想定できるようなものであり、今回のように来たるべき AGI について論じる際に問題になるのは本質的人格性だということは確認しておく必要があるだろう。

## 2. 2. 主体としての性格

そして AGI に本質的人格性とそれによる人権享有主体性が認められた場合に問題になるのが、その存在に関する性格が我々人類と大きく異なることである。第一に、人類と同様に AGI が断絶によって構成される**個体**という概念を持つのか、第二に、人類と同様の可傷性を AGI が備えているかという問題を考えることができるだろう<sup>16</sup>。

### 2. 2. 1. 断絶と個性

順に検討しよう。人類の個体のあいだには癒やしがたい断絶があり、私は他者の痛みを自分のものとして知覚することができないし、知覚可能な領域もみずからの肉体に大きく規定されている。私にとってみずからの痛みは直観的であり、何らかの観察や推論を経て知る対象ではないが、他者の痛みは一定の情報収集——足の小指をぶつけたとかその箇所が赤くなっているといった事実の観察、苦しそうな顔をしているといったような非言語的コミュニケーション、「痛い」と叫ぶような言語的コミュニケーションなど——を経て知るよりほかない。また、手に持ったナイフを自在に操ることができるとか車を運転しているときの車幅感覚のように肉体の境界線の外側にあるものを知覚するように思われる事態は存在するが、基本的には肉体の延長として——肉体の受けた知覚から推論した動作を知覚のよう

<sup>15</sup> Ugo Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts*, Springer, 2013 = 新保史生（監訳）『ロボット法』（勁草書房、2018年）。

<sup>16</sup> 第一の問題については大屋雄裕「人格と責任—ヒトならざる人の問うもの」福田雅樹・林秀弥・成原慧（編）『AI がつなげる社会—AI ネットワーク時代の法・政策』（弘文堂、2017年）344-361 頁、第二の問題は大屋雄裕「AI における可謬性と可傷性」宇佐美誠（編）『AI で変わる法と社会—近未来を深く考えるために』（岩波書店、2020年）45-62 頁において扱った。

に感じているという形で——生じているものに過ぎない。逆に言えば、人類はこのような断絶を抱えているからこそ言語的・非言語的なコミュニケーションを発展させ、互いに配慮するような社会的実践を作り出したと考えることもできるだろう。

しかし AGI はここで見られたような個性を持ち、互いのあいだで不透明なコミュニケーションを行なうように発展させられるのだろうか。むしろそれぞれの状態が互いに透明であり、確定的に知り得るようなものとして、ネットワークを介して接続されることになるのではないだろうか。

たとえば、現在でも入手できるカーナビゲーション・システムの多くには移動体通信システム（テレマティクス）が搭載されており、リアルタイムの道路情報や渋滞状況などを受信しながら最適なルートを選択を行なうといった機能を備えている。もちろん通信は双方向であり、車速やワイパーの動作状況を送信することで混雑状況や降水の有無をサーバ側もリアルタイムで把握することができるようになってきているわけだ。

このとき、我々にとって自然な見方によれば車内にあるカーナビの装置が個々の存在であり、それがテレマティクスによってネットワーク化されているということになるだろう。しかし、さらに進んでこのシステムが自動運転車に搭載され、サーバ側から送られた情報をもとに選択されたルートへと自動的に向かうような事態を考えよう。しかもそのとき、単に個々の車にとって最適になるルートではなく、システムに接続されている車全体の動作を最適化するアルゴリズムによって**社会全体の便益**の実現が目指されるとしよう。このとき、目の前のカーナビ装置が**個体**でありそれがネットワークによって接続されていると考えるべきなのだろうか。すべてのカーナビ装置とネットワークが形成する最適化システムが個体であり、目の前の装置はその端末——我々に見える**あらわれ**にすぎないと考えるべきなのだろうか。

繰り返すが、我々人類が個体へと区切られているのはそのあいだに乗り越えがたい不透明性があるからなのであった。だとすればそのような性質を持たない透明なコミュニケーションによって構成される存在は**個体を持たない**、あるいはその**全体が一つの個体である**と考えるべきだろう。このような存在と我々人類とが同じ社会のなかで共存することができるかという問題を扱った古典的作品として我々は、『ソラリス』を想起するべきだろう<sup>17</sup>。

具体的かつ卑近な例を挙げておけば、たとえば選挙権を AGI に対してどのように配分することができるかという問題を考えればよい。全体が 1 個体なので 1 票しか配分しなければ人類と比較して非常に微弱な影響力のみが与えられることになり、実質的な参政権の保障とは認められないだろう。他方、端末ごとに 1 票を配分すれば膨大な数になるだけでなくそれらが一斉に同一の投票行動を取ることにより、人類を排除できるだけの影響力を得てしまうことになると予想される。このどちらでもない制度は可能だろうか？

## 2. 2. 2. 可傷性とゆらぎ

また、人類の個体は他者による配慮を受けることなしには生存を維持できない無力な存

<sup>17</sup> Stanisław Lem, *Solaris*, MON, 1961 = 沼野充義（訳）『ソラリス』（早川書房、2015 年）。またこの問題は以下で取り扱っている。大屋雄裕『自由か、さもなくば幸福か？—21 世紀の〈あり得べき社会〉を問う』（筑摩書房、2014 年）第 4 章 3。

在として生まれ、やがて成長して独り立ちし、老いて死んでいくという運命に規定されている。前述したコミュニケーションの不透明性によってこの間に得た経験や能力を別の個体へと完全に継承させることもかなわないので、人類は絶えず学び、得たものを変容させ、次世代へと不完全に受け継ぐというサイクルを歴史的に重ねてきた。逆に言えばコミュニケーションが不確実だからこそ我々の文化や社会には常にゆらぎが生じており、そこから新たな習慣や技術が生まれてくることになったとも考えることができるだろう。

これに対し AGI には（意図的にそう作らない限り）「死」が存在せず、限りある人生の期間という概念もないだろう。人類が殴れば傷付き、病めば死ぬといった可傷性（vulnerability）を持つからこそ、一方ではその故に他者を故意に傷付ける行為が刑事法を通じて禁止され、他方ではその禁止に実効性を持たせるために限りある人生の一部ないし全部についての自由を奪う制裁が違反者に対しては加えられることになっているのであった。ベンサムが指摘したとおり刑罰とは犯罪から生じる苦痛が現実化することを避けるために加えられる人工的な苦痛であり、その両面ともが人生を失うことに対する苦痛の感覚、すなわち可傷性に支えられている<sup>18</sup>。

すると問題は、この要素を共有しない AGI と人類とが同じ制度の下で共存できるかということになるだろう。少なくとも高度の自律性を実現した AGI が人権享有主体として認められるに至ったとき、その行為によって他者に被害が及んだときに加える制裁としていかなるものを想定するかという課題にはそのときまでに解決を与えておく必要があるのではないだろうか。

### 2. 2. 3. もう一つの可能性

ただし、これらの問題がいずれも「個人」という単位を基礎として採用した近代法体系において特に問題になるという点には注意する必要がある。我々人類の個体も特定の社会・文化のなかに一定の属性（人種・性別など）を帯びて生まれ落ち、そのことによって存在の基礎を規定されているという指摘は、共同体論（communitarianism）の論者たちによってなされてきた。その典型であるマイケル・サンデルは、人類が持つのは「位置付けられた自我」（situated self）であり、社会の受け継いだ価値観や慣習から自由ではあり得ないと主張している<sup>19</sup>。この立場からは、人類も社会や文化のようなネットワークの一部であり「個体」としての見方は本質的でないとの主張があり得るだろう<sup>20</sup>。

また、人生の始まりと終わりにおいて典型的にそうであるように、他者による配慮によつ

---

<sup>18</sup> Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, T. Payne and Sons, 1780 = 中山元（訳）『道徳および立法の諸原理序説〔上下〕』（筑摩書房、2023年）。

<sup>19</sup> Michael Sandel, *Democracy's Discontent: America in Search of a Public Philosophy*, Harvard University Press, 1998 = 金原恭子・小林正弥（監訳）『民主政の不満—公共哲学を求めるアメリカ〔上下〕』（勁草書房、2010年）。

<sup>20</sup> なお実際の共同体論者たちはそのような立場はあまりにも保守的であり、個人の自由や自己決定を無視していると批判されることを恐れ、ここまで強い主張はしないのが一般的である。たとえばサンデルは生まれ落ちることによって当初のあり方は規定されるものの、人類は自分が受け継いだ文化や習慣を批判的に再検討し次世代に受け継ぐものを変えていくことができると主張している。

て支えられることなしには生存できない状態を人類は誰もが経験している。あるいはさらに、他者によるケアに支えられ他者をケアする関係のなかにあることが人類にとって通常の状態なのであり、そのような関係を見捨てて個々人を独立した自律的存在と想定すること自体が誤っているという指摘が主にキャロル・ギリガン以降の文化的フェミニズムからなされている（ケアの倫理）<sup>21</sup>。この立場からは、近代法がすべての個人を独立した自律的存在だとしてきたことは**事実**に基づいているのではなく「そうでなければならぬ」という前提（方法的個人主義 *methodological individualism*）にすぎないし、その正当性自体が問い直されなければならないということになるだろう。

近年の消費者法領域でも、自律的な個人が正常であるのに対し一定の理由で脆弱性（*vulnerability*）を抱えているカテゴリー——たとえば未成年者や高齢者——を**例外的な存在**と位置付けたうえで保護を加えるのが消費者法であるという従来の理解を改め、現代的な情報環境のなかではすべての個人が特定の環境において浅慮を引き起こしてしまうような脆弱性を抱えていると考えるべきではないかとの議論が行なわれているが<sup>22</sup>、このような立場からも問題はむしろ近代法の側にあり、AGI の出現を契機としてその本質的な再検討が必要となるという立論が可能なのかもしれない。

ジョン・ダンが「なんびとも一島嶼にてはあらず。／なんびともみずからにして全きはなし。／人はみな大陸（くが）の一塊（ひとくれ）。本土のひとひら。」と書いたように、全体の一部としてのあり方こそが人間の本質だという見方にも伝統的な蓄積があることは意識されるべきだろう——「ゆえに問うなかれ、誰（た）がために鐘は鳴るやと。／そは汝（な）がために鳴るなれば」<sup>23</sup>。

### 3. AGI の存在と人類

さて、AGI のいる社会における人類の問題に進むことにしよう。この論点は前述した寿命という限界のなかで我々人類が社会・文化を維持するために行なってきた**教育**の手法とその可能性という問題に関するものである。

#### 3. 1. 例示と反復

現代まで続く法律学の端緒の一つが、530 年前後に東ローマ帝国で編纂された『ローマ法大全』（*Corpus Iuris Civilis*）が 1070 年頃に北イタリアで「発見」された時点に求められ

<sup>21</sup> このような観点からの社会構想については、たとえば以下を参照せよ。池田弘乃『ケアへの法哲学—フェミニズム法理論との対話』（ナカニシヤ出版、2022 年）。

<sup>22</sup> たとえば消費者庁「消費者法の現状を検証し将来の在り方を考える有識者懇談会」（2022～23）とその問題提起を受けた消費者委員会「消費者法制度のパラダイムシフトに関する専門調査会」（2023～）の検討内容がこれにあたる。また学界においてもたとえば「（小特集）デジタル・AI 実装社会における法とガバナンス—情報法・競争法・消費者法の交錯・変容〔上下〕」法律時報 1207・1208 号（2024 年）はそのような問題関心に立つものと位置付けられる。なお参照、大屋雄裕「意思・契約・個人—変容と消費者法の未来」法律時報 1207 号（2024 年）55-59 頁。

<sup>23</sup> John Donne, "Mediation XVII", *Devotions Upon Emergent Occasions, and severall steps in my Sicknes*, 1624. 翻訳はアーネスト・ヘミングウェイ『誰がために鐘は鳴る〔上下〕』（新潮社、2007 年）の大久保康雄訳を用いた。

ることは広く知られている。その後、法学教育の中心の一つとなっていたボローニャ大学における 1350 年頃の講義風景を描いた写本挿絵があるのだが<sup>24</sup>、それが現代の大学法学部におけるものほとんど変わらないことに驚く人もそうでない人もいるだろう。前方に座った教員が口述する内容を学生たちが聞き、ノートを書き写すこと。おそらく印刷技術のなかった当時は講義内容をそのまますべて書き写すことで知識の複製を実現していたのに対し、現代では単純な複製はコピーやファイル共有に委ねることができるので要点や強調箇所の筆記に留められる程度が主な違いである。我々が行なってきた教育とはこのように、不透明なコミュニケーションを超えて知識と経験を継承していくための努力だったと言ってよい。

その際、重要な手段として教師の行為を反復させ、そのなかから重要なポイントを学生に気付かせるという例示と反復による訓練があることに注意しよう。特に言語化しにくいコツやノウハウについては、先人の行為を観察して反復しようとする努力を通じて獲得させるという手法が採られてきた（小学校の体育で二重跳びのやりかたを教える場面などを想定するとよい）。さらにそこでは、単純な作業からより複雑で複合的なものへと段階的に技術を習得させるようなプロセスがしばしば選択されたとも言うことができるだろう。

これはむしろ反省すべき点だと思われるのだが、現代における法学教育においても多くの要素はそのように非形式的な習得に委ねられている。たとえば項と号の差、「9 条 2 項」と「9 条の 2」の違い、あるいは「及び」と「並びに」の使い分けといった事項を考えたとき、もちろんそれらを明示的に言語化して説明しようとした試みは存在してきたもののカリキュラムに明示的に組み込まれているわけではなく、その存在に気付いた一部の学生を除けば、授業で配られるレジュメや教科書・論文などの記載を見てそれを真似るうちにできるようになったというケースが大宗を占めているだろう<sup>25</sup>。

それは職業生活に入ってからと同様であり、たとえば法律家として行なう契約審査のような作業を考えても、まず単純・簡潔なものを任せられ、その結果について上司や先輩から指導を受け、次第に複雑・困難なものが単独で担当できるように成長していくというプロセスを想定することができる。我が国でも特に法律学に関する領域ではこのような OJT (on the job training) 的手法が広く受け継がれてきたと言うことは許されるだろう。

そしてその点が、AGI がいる社会における我々の問題をもたらすように思われる。たとえば典型的な税務や登記申請に関する書類作成は、すでに専用のソフトウェアや情報システムによってその大部分が自動化された状況にある。契約審査についても、現時点の AI によって少なくとも単純・簡潔なものを代替することが可能になり、実際に商用サービスとして提供されているものもある。さらに発達した AGI が形式性と反復可能性の高い作業の多

<sup>24</sup> Laurentius de Voltolina, Liber ethicorum des Henricus de Alemannia, c. 1350, [https://upload.wikimedia.org/wikipedia/commons/f/fc/Laurentius\\_de\\_Voltolina\\_001.jpg](https://upload.wikimedia.org/wikipedia/commons/f/fc/Laurentius_de_Voltolina_001.jpg)

<sup>25</sup> 古典的には林修三『法令用語の常識』（日本評論社、1975 年）などが挙げられる。なお、弥永真生『法律学習マニュアル』（有斐閣、2001 年）が嚆矢と思われるが、おそらくはこの点についての反省を踏まえて法律学における思考法や基礎知識を明示的に述べ、可視化する試みが近年には現われていることを注記する。その例として、横田明美『カフェパウゼで法学を—対話で見つける〈学び方〉』（弘文堂、2018 年）、横田明美・小谷雅子・堀田周吾『法学学習 Q&A』（有斐閣、2019 年）、田高寛貴・秋山靖浩・原田昌和『リーガル・リサーチ&リポート—法学部の学び方〔第 2 版〕』（有斐閣、2019 年）、白石忠志『法律文章読本』（弘文堂、2024 年）。

くを代替してくれる社会がやってきたとして、我々は上述したような教育プロセスを維持し、次の世代を自律的な能力を備えた「個人」へと育て上げることができるのだろうか。

### 3. 2. 気付きを超える教育

もちろんそのような例示と反復を通じて得られる**気付き**に頼るのではなく、困難な課題への対応法を習得するための形式的・明示的な学習方法を開発することが一つの対応として想定し得るかもしれない。自身も芸術家である中ザワヒデキによれば、かつて絵画は対象の忠実な描写を目指すものだったが、写真技術が発明されることによってそのような絵画への需要は消滅してしまい、画家たちはそれと異なるあり方を追求していくことになったという<sup>26</sup>。それがたとえば細部の描写ではなく情景全体の与える印象を把握することを目指した印象派（19世紀後半）や事件の衝撃を描写したピカソ「ゲルニカ」（1937）だということになるだろうか。そこで目指されているのは実物を規準としてそれにどこまで近付けるかという価値観ないし評価関数のもとで**より高い評価**を目指すのとは異なり、新たな価値観・評価関数を提案しそのアップデートを目指すような営みだと位置付けることができるかもしれない。

そして中ザワによれば、少なくとも現代美術においてそのような批評的能力を身に付けるための教育手法は例示と反復ではない。これまでの美術史を学ぶことはその一つの材料だと位置付けられているが、そこで採用されてきた手法を真似ることや自分でも使えるようになることは目指されていない。実作と批評のあいだには教育方法の違いがあるというのである。

法律学においても、高いレベルの実践にはそのような価値観の書き替えという側面があることは多くの法律家が同意するところだろう。アメリカにおける典型例として、連邦最高裁によるブラウン対教育委員会事件（1954年）を考えよう<sup>27</sup>。それまで、白人と黒人を別の学校に通わせる人種別学（segregation）についてはプレッシー対ファーガソン事件（1896年）において示された「分離すれども平等」（separate but equal）という考え方に基づいて肯定されていた<sup>28</sup>。言い換えればそこでは施設やカリキュラムといった教育条件が判断基準であり、その点で同じ取扱いを受けていれば平等だという価値観が示されていたと言える。しかしブラウン事件において最高裁は、人種を基準として異なる取扱いをすること自体が「いかなる州も（……）法の平等な保護を否定してはならない」と定める連邦憲法修正14条に違反するものであり、人種別学を定めたカンザス州法は無効だと判示したのである。ここでは分離した後の取扱いではなく**分離すること自体**が問題とされており、社会に対して新たな理解が提示・強制されていると考えることができよう。

我が国においても、従来は一人一票という形式的な基準を意味すると考えられてきた選

<sup>26</sup> 2023年度人工知能学会全国大会・企画セッション「アートにおいても敗北しつつある人間一人の美意識もAIにハックされるのか？」（<https://www.youtube.com/watch?v=QVFrZdJsD1A>）における中ザワヒデキの報告内容を元にしてはいるが、要約は筆者によるものであり、あり得る誤りについても筆者に責がある。

<sup>27</sup> *Brown v. Board of Education of Topeka*, 347 U.S. 483 (1954).

<sup>28</sup> *Pressy v. Ferguson*, 163 U.S. 537 (1896).

挙権の平等について、選挙区（あるいは選出される議員数）に対する有権者数という実質的な価値を基準として採用し、当時の衆議院における定数配分が憲法に違反している旨を判示した最高裁判決（昭和51年4月14日民集30巻3号223頁）が同様に価値観のアップデートを試みたものと位置付けることができるだろう。また、最近の旧優生保護法違憲判決（最高裁判所大法廷判決令和6年7月3日）における草野耕一裁判官補足意見が、「善きあり方を構想しその実現を図る自由を妨げられることのない利益」（自己実現を妨げられない利益）が個人の人生についても国家のあり方についても想定できることを前提として、現実の立法がそこで想定されている「善き国家」のあり方、すなわち「憲法に適合した国家」に反する場合には司法が「為政者が憲法の適用を誤ったとの（……）判断を歴史に刻印し、以って立憲国家としての我が国のあり方を示すこと」が最善の対応であり、それによって「善き国家の構想・実現という理念を積極的に推進する」ことができると位置付けていることは注目に値する。それは法解釈の機能として、社会あるいは立法府の持つ価値観（あるいは憲法に示された「善き国家」のあり方に対する理解）をアップデートすることを積極的に引き受けているからである。

判決、あるいは法解釈一般においても、そこで提示することが求められている「理由」（reason）とは現実に一定の判断に至る過程を示す**因果関係**ではなく、そのように判断することが規範的に正当であるということを主張する**意味関係**であり、「原因」（cause）とは異なるという点については、すでに論じてきたので繰り返さない<sup>29</sup>。少なくとも現時点においてはこの要求にAIは応えられておらず、「説明可能AI」（eXplainable AI）のような技術も模索されているが、しばしば単に帰結に至る因果関係を可視化する試みに留まっており、一定の価値観によって判断を正当化する論理的関係を明示する、さらには価値観自体をアップデートするような創造的判断を行なうという水準には達していない。あるいはこのような機能が、当面は我ら人類の役割として残されるのかもしれない<sup>30</sup>。

仮に我々が、AGIのいる社会において人類の果たすべき使命をそのような批評的能力を駆使することによる価値観のアップデートに求めるとするならば、そのための能力を涵養するための教育手法を開発することが必要になると考えられる。仮に2045年を目途としてAGIの利活用が実現するとすれば、その時代に向けた課題はおそらくそこに求められることになるだろう。

---

<sup>29</sup> 大屋雄裕『法解釈の言語哲学—クリプキから根元的規約主義へ』（勁草書房、2006年）。また最近この点について、司法におけるAI利用の可能性という観点から再び論じている。Takehiro Ohya, "AI and Judiciary Decisions", in: Catherine Régis, Jean-Louis Denis, Maria Luciana Axente, and Atsuo Kishimoto (eds.), *Human-Centered AI: A Multidisciplinary Perspective for Policy-Makers, Auditors, and Users*, Chapman and Hall / CRC, 2024, ch. 13, pp. 141-150.

<sup>30</sup> 中ザワヒデキは、「人間が人工知能を使って創る芸術」（したがってそれを支える価値観である美学は人間のもの）ではなく「人工知能が自ら行う美学と芸術」、すなわち人工知能自体が新たな美学を作り出し価値観をアップデートする芸術の可能性を追求している（中ザワヒデキ「人工知能美学芸術宣言」（2016年）

[https://www.aloalo.co.jp/nakazawa/2016/0501b\\_j.html](https://www.aloalo.co.jp/nakazawa/2016/0501b_j.html)）。これが実現した際には理由の提示や価値観の想像をめぐるAGIと人類の差も消滅することになるだろう。

## 参考文献

- Bentham, Jeremy (1780) *An Introduction to the Principles of Morals and Legislation*, T. Payne and Sons = 中山元 (訳) 『道徳および立法の諸原理序説〔上下〕』(筑摩書房、2023年)。
- Bostrom, Nick (2014) *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press = 倉骨彰 (訳) 『スーパーインテリジェンス—超絶 AI と人類の命運』(日本経済新聞出版社、2017年)。
- Donne, John (1624) "Mediation XVII", *Devotions Upon Emergent Occasions, and severall steps in my Sicknes*.
- European Parliament Committee on Legal Affairs (2016) "Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))", May 31st, 2016, PR - PE582,443v01-00.
- Francione, Gary L. (2004) "Animals --- Property or Persons?" in: Cass R. Sunstein & Martha Nussbaum (eds.), *Animal Rights: Current Debates and New Directions*, Oxford University Press, ch. 5 = 土屋裕子 (訳) 「動物は財産か、人格か？」安部圭介・山本龍彦・大林啓吾 (監訳) 『動物の権利』(尚学社、2013年) 第5章。
- Lem, Stanisław (1961) *Solaris*, MON = 沼野充義 (訳) 『ソラリス』(早川書房、2015年)。
- Ohya, Takehiro (2024) "AI and Judiciary Decisions", in: Catherine Régis, Jean-Louis Denis, Maria Luciana Axente, and Atsuo Kishimoto (eds.), *Human-Centered AI: A Multidisciplinary Perspective for Policy-Makers, Auditors, and Users*, Chapman and Hall / CRC, ch. 13.
- Pagallo, Ugo (2013) *The Laws of Robots: Crimes, Contracts, and Torts*, Springer = 新保史生 (監訳) 『ロボット法』(勁草書房、2018年)。
- Sandel, Michael (1998) *Democracy's Discontent: America in Search of a Public Philosophy*, Harvard University Press = 金原恭子・小林正弥 (監訳) 『民主政の不满—公共哲学を求めたアメリカ〔上下〕』(勁草書房、2010年)。
- 池田弘乃 『ケアへの法哲学—フェミニズム法理論との対話』(ナカニシヤ出版、2022年)。
- 内田貴 『民法 I—総則・物権総論〔第2版補訂版〕』(東京大学出版会、2000年)
- 大屋雄裕 『法解釈の言語哲学—クリプキから根元的規約主義へ』(勁草書房、2006年)。
- 『自由か、さもなくば幸福か?—21世紀の〈あり得べき社会〉を問う』(筑摩書房、2014年) 第4章3。
- 「外なる他者・内なる他者—動物と AI の権利」論究ジュリスト 22号 (有斐閣、2017年) 48-54頁。
- 「人格と責任—ヒトならざる人の問うもの」福田雅樹・林秀弥・成原慧 (編) 『AI がつなげる社会—AI ネットワーク時代の法・政策』(弘文堂、2017年) 344-361頁。
- 「AI における可謬性と可傷性」宇佐美誠 (編) 『AI で変わる法と社会—近未来を深く考えるために』(岩波書店、2020年) 45-62頁。
- 「意思・契約・個人—変容と消費者法の未来」法律時報 1207号 (2024年) 55-59頁。
- 水津太郎「コメント」安藤馨・大屋雄裕 『法哲学と法哲学の対話』(有斐閣、2017年) 84頁。
- 山川宏「汎用人工知能とは」中島・浅田他編 『AI 事典〔第3版〕』(近代科学社、2019年) 50頁。