

「AIのアルゴリズムを説明してもらう権利」について

メタデータ	言語: jpn 出版者: 公開日: 2020-04-08 キーワード (Ja): キーワード (En): 作成者: 岡田, 安功 メールアドレス: 所属:
URL	https://doi.org/10.14945/00027262

「AI のアルゴリズムを説明してもらう権利」について

岡田 安功（静岡大学名誉教授）

要約：人工知能が普及して様々な用途で利用されている。本稿は人工知能が個人を評価するために利用される場合に論点を絞り、このような場合に人工知能のアルゴリズムによって評価された者がこのアルゴリズムについて説明を受けることが権利であり、少なくとも法律の保護に値する利益であることを主張している。本稿はこの主張の根拠を補強するために、日本の法制度や現状を検討するだけでなく、EU、イギリス、アメリカ、OECD 等が人工知能のアルゴリズムに対してどのように法的対応をしているかを検討している。このような比較法的検討から、「AI のアルゴリズムを説明してもらう権利」を議論する必要性と意義を明らかにしている。

キーワード：AI, 人工知能, アルゴリズム, 個人データ, プライバシー, プロファイリング, 行政手続

1 はじめに：問題提起

Artificial Intelligence の翻訳語である「人工知能」が日本語としてすっかり定着した。本稿では「人工知能」の原語を略した「AI」という単語を用いて議論する¹⁾。AI がどんな仕組みで機能しているのかわからない人でも、AI が最先端のコンピュータ技術であるというイメージを抱いている。ATM 等で生体認証をする技術、クレジットカードの発行や住宅ローンの可否について与信の判断をする技術、監視カメラで顔認証をする技術、乗用車を自動運転する技術、株価の変動を予測する技術、ネットで商品を物色する人をプロファイリングして商品を推奨する技術、内臓等の画像から病気を診断する技術、SNS を使ったイジメを発見して削除する技術等、誰もがいつでも気づいているわけではないが、現代人の日常生活には AI が入り込んでいる。このような AI を可能にしたのは、パーソナルデータと非パーソナルデータを含めたビッグデータを集積する技術の進歩とビッグデータを解析する技術の進歩である。上述の AI はビッグデータを前提として機能しているが、本稿ではビッグデータの中でもパーソナルデータに相当する個人データ²⁾を用いた AI に焦点を当てて、AI との向き合い方を考えてゆきたい。

私たちと AI の関わりは、私たちの自覚の有無に関係なく自ら AI を利用する場合がある一方、私たちが関係をもつ相手方が私たちを評価する手段として AI を利用する場合がある。

さて、AI は AI に実装されたアルゴリズムによってビッグデータを処理する。アルゴリズムが実装されない AI はありえない。ところが、自動運転に AI を利用する場合、AI のアルゴリズムは自動車会社ごとに違うが、自動車を購入して利用する者にとって最大の関心は自動車の安全性である。自動運転車を購入しようとするユーザーは自動運転のアルゴリズムを理解できなくても自動車が安全であればいい。自動運転の精度を高めるために、ユーザーがいつどこを走行したかが自動車会社のサーバへ逐次送られていても、個人を特定できない情報処理が施されていれば、多くのユーザーは自分のプライバシーを気にすることなく自動運転を利用するであろう。自動運転で AI が使われていても、多くのユーザーは AI のアルゴリズムにもプライバシーにも個人データにもほとんど関心をもたないと思われる。自動車会社もそのような自動運転車でなければ売れないので、そのような技術的対応をするはずである。

ところが、アルバイトの採用や正社員の採用に AI が使われる場合、応募する者はどんな AI が使われ

るのか気にならないだろうか。ある学生がスマホでアルバイトに応募して、スマホの向こうで学生に対応してプロファイリングをするのが AI で、AI がこの学生を不採用と判断すれば、人事担当者はこの判断に従う。アルバイトではこのような採用方法が普及し始めている。アルバイトだと割り切れれば、AI の判断だと知らされても、この学生はこの選考結果を許容するかもしれない。しかし、この学生が就活で正社員としての雇用に応募した場合、面接は人間が行なったとしても AI の判断で不採用になったら、この結果にこの学生は納得するだろうか。ただし、この納得できるかどうかという問題は、AI で判断されたことをこの学生が知っている場合にのみ発生する。人事担当者が AI を利用していることを学生が知らなかったら、この問題そのものが自覚されず問題になることもない。就活の学生ではなく、既に学校を卒業して生活のために働く場を必要としている者は、アルバイトであっても、AI のみの判断で求職を断られることに納得できないだろう。AI による雇用の決定は、労務管理費をカットできて、商品とサービスのコストをカットできるが、非正規雇用労働者の割合が高く、雇用形態が多様な社会で、AI の判断のみで不採用が可能であれば、社会的身分の固定化等、社会の在り方に格差や不公平等の深刻な後遺症を積み上げる可能性がある。

では、AI を使った不採用判断のどこが問題か。第一に、人間ではなく AI という機械の判断に頼った決定は、人事担当者も応募者も機械に支配されていることを意味する。人事担当者が AI の判断を参考にする程度ならいいのだが、本当に参考程度で済むだろうか。第二に、不採用者を含め応募した者は人事担当者から AI の利用を知らされない限り、自分自身の採否に AI が関与したことを知ることができない。第三に、これが最も重要なのだが、不採用者が不採用判断に AI が利用されたことを知ったところで、不採用の理由を知ったことにはならない上に、人事担当者が AI の判断理由を理解しているとは限らない。深層学習をする AI を使ったら、結論は出ても理由は闇の中である。

これらの問題点を少しでも解消するにはどうすればいいのか。知らないうちに AI が人間を支配する社会にならないために、AI の支配から人間の尊厳を守るために、AI を使って人を評価する業務を行う者は AI のアルゴリズムを相手に説明する必要がある。上記の人事担当者が AI を利用して雇用の採否を決定する場合、応募者は AI の利用だけでなく AI のアルゴリズムを説明されるべきである。現実の多くの募集では、落選の選考通知の記載が簡略で、落選の理由は書かれていないに等しい。落選の通知に AI のアルゴリズムが理解可能なレベルで説明されていたら、落選者は落選の理由を知る。落選者にアルゴリズムを説明する必要がなければ、アルゴリズムを理解しないまま AI を使い続ける人事担当者が増える可能性さえ存在する。他人を評価する業務で AI を使う場合、AI のアルゴリズムを評価される者に説明することは公正な評価を担保する。いうまでもなく、アルゴリズムの公開＝アルゴリズムの説明ではない。複雑なアルゴリズムを見せられると、理解できる者はごくわずかである。

自動運転車の購入者は自動運転のアルゴリズムを知らなくても気にしない。医療現場で AI を利用した画像診断を受けた患者は、AI のアルゴリズムを知らなくても気にしない。しかし、人事のように人を評価する場面で AI が使われると、AI のアルゴリズムの説明が問題になる。学生アルバイトと就活の例で示したように、アルゴリズムの説明の是非を判断する基準は決して明確ではない。

AI のアルゴリズムはどんな場面で説明が必要だろうか。説明が必要と思われる場面は確かに存在するが、必要な説明を確実に実現しようとしても、あらゆる場面を想定して要件を明確にするのは困難である。自分自身を評価した AI のアルゴリズムを説明してもらいたいと思ったとき、評価された者がこのアルゴリズムを理解する方法は法的に閉ざされているのだろうか。この問題を考える手掛かりとして、「AI

のアルゴリズムを説明してもらう権利」という比較的新しく主張されている権利がある。AI の評価によって不利益を受けた個人がそのアルゴリズムを知りたいと思ったら、このような権利が確立されているなら、この権利を行使するのが確実な方法である。この権利は、本稿が後に検討する EU の一般データ保護規則（略称、GDPR）の解釈として展開されるようになったが、権利概念そのものはまだ不明確である。しかし、AI が普及してアルゴリズムの判断が社会に大きな影響を与えるようになった現在、アルゴリズムそのものに対する何らかの規制を必要とする見解が広がりつつあるように思われる³⁾。アルゴリズムの規制は放置できない重要課題である。アルゴリズムに対する規制にはいくつものアプローチが可能であるが、本稿は権利という観点からアルゴリズムに対する規制の在り方を考察する。もちろん、このような権利アプローチのみではアルゴリズムの規制を十分に実現することはできないが、アルゴリズムに対する権利行使の限界をできる限り具体化しない限り、権利アプローチの限界を突破するアプローチは適切に構築することができないだろう。本稿の目的は AI のアルゴリズムが説明を求められる場合の法的根拠をできる限り明らかにすることだが、本稿の検討は結果的に権利アプローチの限界を明示することにもなる。権利概念がどれほど曖昧であっても、権利という問題意識のない分野には、権利以外の法的アプローチが生まれないし、権利の限界を克服するアプローチも生まれない。ただ、本稿が「AI のアルゴリズムを説明してもらう権利」について直接語るのは本稿のずっと後である。それまでこの権利と互いに支え合う様々な法的思考の装置を紹介する。

2 日本の法制度とアルゴリズム

2-1 個人情報保護法は AI とアルゴリズムを規制するか

機械学習をする AI が登場し、深層学習ができるようになり、AI の設計者が予想できないアルゴリズムの更新を AI 自身ができるようになり、AI は社会に大きな影響を与える存在になった。しかし、深層学習という技術が確立しても、ビッグデータがなければ AI は学習できないので、それだけでは AI は機能しない。ビッグデータには個人データが含まれる。では、日本の個人情報保護法は AI について何か規定しているのだろうか。

個人情報保護法は AI そのものについては規定しないが、個人データを利活用するという観点から個人データを取扱う方法を規定して、AI に大量の個人データを処理させることを可能にしている。特に、2015 年の改正で「匿名加工情報」という概念が導入され、匿名化した個人情報を大量に第三者へ提供することが可能になった。「匿名加工情報」は「特定の個人を識別することができないように個人情報を加工して得られる個人に関する情報であって、当該個人情報を復元することができないようにしたもの」と定義されている（2 条 9 項）。「加工」方法として、個人情報に含まれる記述の一部を削除するか個人識別符号の全てを削除するか、またはこれらを他の記述に置き換えることが規定されている（2 条 9 項 1 号及び 2 号）。この「匿名加工情報」は個人情報に復元できないと定義されているが、実際には AI を使えば復元できるといわれている⁴⁾。この可能性を懸念したのか、本人を識別するために匿名加工情報を他の情報と照合する識別行為が匿名加工情報取扱事業者に対して禁止されている（36 条 5 項、38 条）。このため、匿名加工情報取扱事業者はこの識別行為を可能にするアルゴリズムを AI に実装できない。この限度で、明文化されていないが、個人情報保護法は AI とアルゴリズムを規制している。

36 条 5 項と 38 条は、AI を使って、匿名加工情報を個人情報に復元することなく解析し、ある個人情報と照合して、この個人について評価することを禁止していない。この禁止がないのは、個人情報取扱

事業者も同様である。では、この AI のアルゴリズムを AI で評価された個人に説明する必要があるだろうか。例えば、このアルゴリズムが医療で使う画像診断用のものか、就活における内定を判断するためのものか、用途によってアルゴリズムへの望ましい対応が異なることは自明であろう。この問題について、個人情報保護法は沈黙している。そもそも、AI のアルゴリズムを説明する必要があるかという問題は、AI の利用が評価対象者に告知されなければ、発生しない。このどちらも個人情報保護法は規定していない。

個人情報保護法に AI の利用に関する告知の規定がなく、当然ながら、データ主体は AI のアルゴリズムの内容と利用方法について適切かどうかを判断する機会がない。もちろん、AI のアルゴリズムを丁寧に説明してもらったからといって、データ主体がそれを理解できるとは限らないし、深層学習をする AI は判断の根拠を示せない。個人の権利義務に関わる重要な決定をする場合は、決定に至る過程と根拠を明示できる仕組みを法的に保障することが、これまでも行われてきた。行政手続法や各種の訴訟法はその典型である。AI についてもこのような配慮が必要である。

データ主体が提出した個人情報を AI で処理して個人評価をしたデータ管理者に対して、データ主体が AI に実装されたアルゴリズムについて説明を求める規定は、個人情報保護法に存在しない。そもそも、個人評価に AI を利用するかどうかについてデータ主体に告知する規定が個人情報保護法に存在しない。この問題が典型的に顕在化するのはいわゆるプロファイリングの場面である。個人情報保護法はプロファイリングを明文化していないが、17 条 2 項が個人情報の「適正な取得」を定めており、同項が規制する要配慮情報の「取得」にプロファイリングが該当すると読み込む学説が存在する⁵⁾。このような踏み込んだ解釈をしても、この解釈からは、AI を使ったプロファイリングに伴うアルゴリズムの説明の是非について、現在の個人情報保護法を根拠とする規制の可否を論じる余地は発生しないように思われる。

28 名の気鋭の研究者で構成される「パーソナルデータ + α 研究会」は 2018（平成 30）年 12 月 19 日に中間報告書として「プロファイリングに関する提言案」⁶⁾を公表した。この最終報告書は 2019（令和元）年中に公表される予定だが、本稿の締切日にはまだ公表されていない。この中間報告書は、EU 法とアメリカ法の影響を受け、AI によるプロファイリングがアルゴリズムによって行われることを明確に意識しており、アルゴリズムへの言及が幾度も行われているが、アルゴリズムに対する直接の規制には触れていない。この中間報告がアルゴリズムを権利の観点から規制するアプローチを取らなかったのは、「行為ベースの規律（行為統制型規律）からガバナンス・ベースの規律（構造統制型規律）へ」⁷⁾という問題意識があったからだと思われる。最終報告書は個人情報保護法の改正に大きな影響を与えらると思われる。一方、2019（平成 31）年 4 月 25 日に個人情報保護委員会が発表した「個人情報保護法 ～いわゆる 3 年ごとに見直しに係る検討の中間整理～」もアルゴリズムの公開について言及していない。

2-2 AI とアルゴリズムに対する日本のソフトロー

内閣府の統合イノベーション戦略推進会議は、「人間中心の AI 社会原則会議」に検討させた「人間中心の AI 社会原則」を、2019（令和元）年 3 月 29 日に決定している。「人間中心の AI 社会原則」は三つの価値、即ち、「人間の尊厳が尊重される社会(Dignity)」「多様な背景を持つ人々が多様な幸せを追求できる社会(Diversity & Inclusion)」「持続性ある社会(Sustainability)」を基本理念としている。「人間中心の AI 社会原則」は「AI 社会原則」と「AI 開発利用原則」で構成され、前者は社会や国が留意すべき原則で、後者は開発・事業者が留意すべき原則とされている。「AI 社会原則」は「人間中心の原則」「教育・リテ

ラシーの原則」「プライバシー確保の原則」「セキュリティ確保の原則」「公正競争確保の原則」「公平性、説明責任及び透明性の原則」「イノベーションの原則」で構成される。「AI 開発利用原則」は、上記の基本理念と「AI 社会原則」を踏まえて定めるものとされ、「早急にオープンな議論を通じて国際的なコンセンサスを醸成し、非規制的で非拘束的な枠組みとして国際的に共有されることが重要である」と規定されている。実際に、「人間中心の AI 社会原則」は後に紹介する OECD の AI 政策に大きな影響を与えており、両者の規範構造に大差はなく、「人間中心の AI 社会原則」は OECD 加盟国の AI 政策と調和している。これは主要国が OECD 加盟国である EU の AI 政策と調和していることをも意味する。

内閣府の「人間中心の AI 社会原則」は「非規制的で非拘束的な枠組み」でありながら「国際的に共有される」ものとして構想されている。これは AI の規制にとって極めて重要な観点である。現在の AI にはビッグデータが不可欠でネットに接続されることが多いので、AI の規制方法は基本的な部分が世界共通でなければ意味がない。「非規制的で非拘束的な枠組み」といっても、これを遵守しない AI の開発や利用が具体的な権利侵害や利益侵害を発生させる場合がありうる。更に肝心なことは、これを遵守しても権利侵害や利益侵害としか思えないことが発生する場合がありうることである。本稿はこの問題に対応する観点から、本稿の最終部分で、AI のアルゴリズムを説明してもらえないことが法的利益の侵害になり、権利侵害になる場合があることを論じる。

内閣府の「人間中心の AI 社会原則」は AI のアルゴリズムに言及していないが、「公平性、説明責任及び透明性の原則」を遵守するために事業者等に AI のアルゴリズムを説明する必要性が発生する場合があると思われる。

総務省情報通信政策研究所は「AI ネットワーク社会推進会議報告書 2019」（以下、「報告書 2019」）を 2019（令和元）年 8 月 9 日に公表している。この報告書は、AI ネットワーク社会推進会議がまとめたもので、「AI 利活用ガイドライン」を示している。このガイドラインは、「AI 利活用原則」を実現するために講ずべき措置を解説している。AI 利活用原則の基本理念の一つとして「人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること」が挙げられている。この理念は後述する EU の「信頼できる AI のための倫理ガイドライン」の倫理原則と共通する。

AI 利活用原則は次の通りである。i 適正利用の原則、ii 適正学習の原則、iii 連携の原則、iv 安全の原則、v セキュリティの原則、vi プライバシーの原則、vii 尊厳・自律の原則、viii 公平性の原則、ix 透明性の原則、x アカウンタビリティの原則。このうち、ii は AI の学習に提供するデータの質に留意することによって、AI のアルゴリズムによってバイアスが発生しないようにしている。AI のバイアスを防止するという点では vii から x はいずれも重要な原則である。各原則は相互に関連している。特に、vii は憲法 13 条が保障する基本的人権の根拠となる思想を表明している。

さて、上記の「報告書 2019」は「viii 公平性の原則」の論点として「アルゴリズムによるバイアスへの留意」を指摘している。また、「報告書 2019」は「ix 透明性の原則」を「AI サービスプロバイダ及びビジネス利用者は、AI システム又は AI サービスの入出力等の検証可能性及び判断結果の説明可能性に留意する」と規定している。ガイドラインは「本原則は、アルゴリズム、ソースコード、学習データの開示を想定するものではない。また、本原則の解釈に当たっては、プライバシーや営業秘密への配慮も求められる」と解説している⁸⁾。このガイドラインはアルゴリズムの透明性が営業秘密によって制限されることを容認している。しかし、この原則はアルゴリズムの開示を禁止しているわけではない。また、AI の予測・認識プロセスがブラックボックス化しているアルゴリズムを開示しても、結果に対する十分な

説明にはならない。ガイドラインはこの点を配慮して「AI の判断結果の説明可能性を確保すること」を目指している⁹⁾。この「説明可能性」の対象としてアルゴリズムの説明が検討課題になると思われる。

「AI 利活用原則」にも強制力がない。しかし、AI と人の多様な関係を一律に規定する具体的で適切なルールが見つからない状況において、強制力のない規制を採用することは極めて賢明である。

3 EU とアルゴリズム

3-1 一般データ保護規則 (GDPR)

EU は 2018 年 5 月に「一般データ保護規則」(日本での通称 GDPR)¹⁰⁾を施行している¹¹⁾。GDPR は欧州委員会が欧州議会の関与を経て制定したもので、EU 加盟国にそのまま適用される。EU 加盟国は、自国の個人データ保護法が GDPR と同じ規定を設けているときは自国の規定を適用するが、自国の法律に GDPR と同じ内容の規定がないときは GDPR を国内に直接適用できる。GDPR にはアルゴリズムについて直接の規定がないが、EU の 29 条作業部会(Article 29 Working Party)¹²⁾が「自動化された個人に対する意思決定及びプロファイリングに関する規則目的のためのガイドライン」¹³⁾を 2018 年 2 月 6 日に採択して、GDPR の実施におけるアルゴリズムへの対応の在り方を示した。このガイドラインが想定する AI は機械学習をする AI である。

GDPR は 5 条で、個人データの取扱いについて、合法性(licéité)、公正性(loyauté)¹⁴⁾、透明性(transparence)を義務づけている。また、GDPR は、個人データの取得が本人 (personne concernée) からでも (13 条 2 項 f 号)、本人からでなくても (14 条 2 項 g 号)、個人データを収集したデータ管理者が「公正 (équitable) かつ透明な処理を保障するために必要な情報」を本人に提供することを義務づけている。この情報の一つに、「プロファイリングを含む自動化された決定の存在」に関する情報があり、自動化された決定を行う場合には、「根底にある論理 (logique sous-jacente) に関する有益な情報」と「処理の重要性及び予想される結果」が本人に提供される。

これらの規定を前提として、GDPR は 22 条 1 項で「個人データの自動処理のみで決定されない権利」を次のように規定している。

GDPR 22 条 1 項:「データ主体」¹⁵⁾は、自動処理 (プロファイリングを含む) のみに基づいてなされる決定のうち、法的効果又は同等の重大な影響を自分自身にもたらすものに服さない権利をもつ。」

なお、上記規定の「自動処理」にいう「処理(traitement)」は日本の個人情報保護法で規定する個人データの「取扱い」に相当する。また、自動処理の一つとして例示されている「プロファイリング」は、GDPR 4 条 1 項 4 号で「プロファイリングとは、自然人に関する特定の人格的側面を評価するための個人データの利用に該当する個人データの自動処理のすべての形式であり、特に、当該自然人の労働遂行能力、経済状況、健康、個人的嗜好、関心、信頼性、行動、位置又は移動、に関する側面を分析又は予測するために個人データを利用することをいう」と定義されている^{16) 17)}。

ただし、GDPR 22 条は「個人データの自動処理のみで決定されない権利」に例外を設けており、データ主体とデータ管理者間の「契約の締結又は履行に必要な場合」(2 項 a 号)、「データ管理者側に人を介在させる権利」(3 項)を実施すれば、この権利は適用されない。

GDPR の 22 条が保障する「個人データの自動処理のみで決定されない権利」は主に透明性と公正性を確保するための権利である。上記のガイドラインは 22 条について、この透明性を確保するために、AI の利用に責任を負うデータ管理者に対して、AI を利用して行なった決定の論拠(rationale)と基準(criteria)

を、簡単で偽りのない方法(simple ways)で本人に説明することを求めている。ガイドラインによると、アルゴリズムについての複雑な説明や、アルゴリズムを完全に公開することは必ずしも必要ではないとされ、15 条 1 項に規定する「関係する論理についての意味のある情報」の提供が「十分包括的(sufficiently comprehensive)」に行われることが求められている。データ主体に提供される「根底にある論理 (logique sous-jacente)」の中にアルゴリズムが含まれることは確かだが、その提供のされ方は GDPR にもこのガイドラインにも明確には書かれていない。

このガイドラインが目指すことは AI を利用して行う決定について論拠と基準を明確にすることである。論拠と基準が明確になれば、GDPR 22 条 1 項の「同等の重大な影響」に該当するかどうか判断の難しい事例についても、決定そのものへの客観的な評価が可能になる。このガイドラインはアルゴリズムに対する理解が容易ではないことに配慮して、アルゴリズムを正確に解りやすく説明することを求めている。GDPR は AI だけで判断することを禁じているので、ガイドラインは AI のアルゴリズム以外に使った論理も含めて本人に理解させようとしている。GDPR は AI の用途の代表例としてプロファイリングを想定しているが、プロファイリングをした者が相手に結果の理由を説明する際、AI のアルゴリズムとそれ以外の論理をどのように関連づけて説明するのだろうか。AI が自ら出した結論について理由を示せなくても、AI の利用者がその結論を支持する場合、どのような説明をするのだろうか。この説明は必ずしも容易ではない。GDPR の 22 条について、AI の使用を説明するだけではアルゴリズムによる危害に対して完璧な救済を与えることができないという批判がある¹⁸⁾。

さて、アルゴリズムについて GDPR とガイドラインの関係をどのように理解するべきであろうか。AI で評価された者が AI のアルゴリズムについて説明を受けることを、GDPR そのものが保障しているのか、GDPR はこれを保障していないが、ガイドラインがアルゴリズムの説明を受ける余地を規定したのか。アルゴリズムに対して強制力を伴う規制を模索すると、このどちらかの解釈が考えられる。しかし、GDPR がアルゴリズムについて明確な規定をしていないことに積極的な意味があるとすれば、GDPR はガイドラインとの組み合わせでアルゴリズムが強制力のない規制方法、いわゆるソフトローによる規制を規定していることになる。この問題については 3-3 で観点を変えて考察する。

3-2 信頼できる AI のための倫理ガイドライン

EU の欧州委員会に設けられた「AI に関する高度専門グループ」(Groupe d'experts de haut niveau sur l'intelligence artificielle、略称 GEHN IA)は、2019 年 4 月 8 日に「信頼できる AI のための倫理ガイドライン(Lignes directrices en matière d'éthique pour une IA digne de confiance)」を公表した。GDPR を前提としたこのガイドラインがアルゴリズムについてどのように考えているかを探してみよう。

このガイドラインは信頼できる AI の特色として、合法性 (licite)、倫理性 (éthique)、健全性 (robuste) を要求している。ここにいう合法性は EU 条約、EU 憲章、国際人権法等が保障する基本権や上記の GDPR 等のその他の EU 法を守ることであり、基本権に基づいた倫理原則(principes éthiques)によって倫理的で技術的にも社会的にも健全な AI の保障が目指されている。そのため、このガイドラインは、まず四つの倫理原則を示し、次にこれらを満たすための七つの要件を示した。

四つの倫理原則は次の通りである。i 「人間の自律への尊重 (respect de l'autonomie humaine)」、ii 「危害防止 (prévention de toute atteinte)」、iii 「公正 (équité)」、iv 「説明可能性(explicabilité)」。

倫理原則を満たすための七つの要件は次の通りである。i 「人的関与と人的監視 (action humaine et

contrôle humain)」、ii 「技術的健全性とセキュリティ (robustesse technique et sécurité)」、iii 「プライバシーとデータガバナンス (respect de la vie privée et gouvernance des données)」、iv 「透明性 (transparence)」、v 「多様性、被差別と公正 (diversité, non-discrimination et équité)」、vi 「社会的かつ環境的幸福 (bien-être sociétal et environnemental)」、vii 「説明責任 (responsabilité)」¹⁹⁾。

「透明性」の要件は「追跡可能性、説明可能性、コミュニケーションを含む」と説明されている。この「コミュニケーション」はAIシステムと交流する者が自分の相手をAIシステムだと知らされることを保障する。また、「説明責任」は次のように説明されている。説明責任は「監査可能性、負の影響を最小限に削減すること、本人に対する仲裁及び争訟の通知、を含む。」²⁰⁾「監査可能性はアルゴリズム、データ、設計過程を評価する可能性を意味する。」²¹⁾。このガイドラインが説明責任を果たす手段としてアルゴリズムの監査を提案していることは注目に値する。アルゴリズムは企業秘密等の理由で知的財産として開示を拒否される時があるが、このガイドラインに従うと、アルゴリズムは内容次第で監査により倫理に反するものと認定される可能性がある。

このガイドラインは、現在または未来の政策や規則に代わるものではなく、「ヨーロッパにとって信頼できるAI」について議論の出発点となることを意図するとされている。また、このガイドラインは、ヨーロッパを超えて地球規模で、AIシステムの倫理的フレームワークについて、研究、熟慮、議論が促進されることを意図するとされている。

3-3 「説明してもらう権利 (right to explanation)」

GDPR が「説明してもらう権利 (right to explanation)」を規定しているかどうかについて多くの議論がある。本稿が「AIのアルゴリズムを説明してもらう権利について」というタイトルで議論を進めるのも、この議論の影響である。AIのアルゴリズムを説明する必要があるかどうかは、「説明してもらう権利」の説明範囲を議論する際に、「透明性」を確保するために必要かどうかという観点から議論されている。「説明してもらう権利」を承認する議論がある²²⁾一方で、この権利概念が曖昧だという批判もある²³⁾。この批判こそが「説明してもらう権利」の性質を正確に表現しており、「説明してもらう権利」を正面から否定しないものの、この権利概念を追求することに実益を見ない論者が多いと思われる。AIの判断に人間が不当に支配されないために、AIの利用について透明性を確保して説明責任を果たし、結果的に公正な社会を実現することが「説明してもらう権利」の目的である。この目的を達成する手段がGDPRに幾つも規定されているのであれば、「説明してもらう権利」にこだわる必要はない、というのが「説明してもらう権利」に批判的な論者のスタンスである²⁴⁾。「説明してもらう権利」に肯定的な論者も、GDPRにこの権利を実現する手段が幾つも規定されていることを指摘することに熱心で、この権利を理論的に構成する作業には熱心でない²⁵⁾。

「説明してもらう権利」は上述のGDPR 22条を根拠にして主張されている。22条は1項で「個人データの自動処理のみで決定されない権利」を規定している。自動処理にはプロファイリングが含まれるので、EUはEU法の主要な規定を説明するRecitalでGDPRのプロファイリングについて説明している²⁶⁾。Recital 71には「決定に関する説明を手に入れる権利」が保障されると書かれている。ここから、Recitalの法的効力も含めて、「説明してもらう権利」をGDPRが保障しているかどうか問題になった。Recital 71によれば、「決定に関する説明を手に入れる権利」は「人間の関与を手に入れる権利」「自分自身の見解を表明する権利」「決定に異議を申立てる権利」等、GDPRに規定されている三つの権利とともに保障

されている。しかし、「説明してもらう権利」そのものは GDPR に明文化されていない。興味深いのは、「説明してもらう権利」に消極的な論者も積極的な論者も「説明してもらう権利」は上記の三つの権利と密接な関係にあると理解していることである。GDPR に規定された権利があるから「説明してもらう権利」はいらないと考えるか、「説明してもらう権利」は曖昧だから三つの権利で補うと考えるか、または、上記の三つの権利は「説明してもらう権利」を前提とすると考えるか、議論のスタンスはこのように分類できる。どの議論も、GDPR が規定しているかどうかは別として、「説明してもらう権利」に相当する考え方を承認しているように思える。

また、Recital 71 は、公正かつ透明なデータ処理を保障するために、管理者に対して「個人データについて不正確な結論を導く要因が修正され、かつ、誤りのリスクが最小限になることを保障すること」等について、「技術的かつ組織的で適切な対策」を実施することを求めている。この対策にアルゴリズムに対する監査制度が含まれていると考えられている²⁷⁾。AI に対する監査制度は「説明してもらう権利」が実施されない場合に透明性を高める代替措置になる。この監査制度等の個人データの濫用を防ぐ制度的な仕組みが GDPR に詳細に規定されていることを根拠にして、「説明してもらう権利」を不要と考える立場と、監査制度が「説明してもらう権利」の曖昧さを補うと考える立場がある。

GDPR が「説明してもらう権利」を保障しているかどうかについて論争が続いている。しかしこの論争において、アルゴリズムに対するデュー・プロセス (algorithmic due process)、アルゴリズムに対する監査 (algorithmic audit)、アルゴリズムに対する影響評価 (algorithmic impact assessments) 等の実施を通して、アルゴリズムに対する説明責任に対応することは共通認識である。「説明してもらう権利」が GDPR に規定されているとしても、その実現にこれらの仕組みが必要だと考えられている。この思考は権利の実現にはガバナンスやシステムが必要だという発想に基づいている。

4 イギリスとアルゴリズム

イギリスのデータ保護法 (Data Protection Act 2018) は EU からの離脱に備えて GDPR を実施するための改正を受けている。イギリスは世界で最も厳格に個人データを保護するデータ保護法を 1984 年に制定して以来、この分野で世界をリードしてきた。イギリスは EU からの離脱後も AI が使用する個人データの保護について EU と同等以上の保護を与え、OECD を通じて EU やアメリカに影響を与え続けると思われる。AI のアルゴリズムに対する対応は今後も基本的には EU とほぼ同じだと思われる。

イギリスでは 2019 年 4 月 1 日から「データ倫理・イノベーション・センター (Centre for Data Ethics and Innovation ; 略称 CDEI) が活動を始めた。CDEI は、政府が設立した独立行政機関で、調査と助言の権限をもち、AI 及びデータを駆使した技術のために最適のガバナンス体制を発展させるために、政府の政策立案者、産業界、市民団体、国民と連携することが課されている。CDEI の活動は官民の全部門に及んでいる。CDEI の最初の仕事は機械学習をする AI のアルゴリズムが行なった決定のバイアスを調査して助言することだった。CDEI は 2019 年 7 月 19 日にアルゴリズム的意思決定とオンライン・ターゲティングについてそれぞれ中間報告²⁸⁾を発表している。この中間報告が設定した目標はアルゴリズムによる決定を公平だと信用できる環境である。

AI のアルゴリズムに対してイギリスは法規制ではなく CDEI という独立行政機関による調査と助言という手法を選択した。CDEI にはアルゴリズムに対する常設の監査機関という側面がある。CDEI の調査と助言は、公開されて社会的に共有されやすく、その蓄積が強制ではない規範を形成する。

5 アメリカの判例とアルゴリズム

アメリカにはアルゴリズムの規制を考える上で参照すべき重要な判例がある。これは2016年のウィスコンシン州最高裁判所の判決 *State v. Loomis*²⁹⁾である。この判決は、2017年に連邦最高裁判所がこの判決に対する裁量上告を認めなかったので、確定している。銃が発砲された自動車を運転していた容疑で逮捕された被告人の *Loomis* に対して、裁判所は COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)³⁰⁾ というソフトウェアを使って再犯可能性等のリスク評価をして刑期を決めた。COMPAS を用いて量刑したことがデュー・プロセスに対する被告人の憲法上の権利を侵害したことになるかどうか、大きな争点になった。当然ながら、再犯可能性等を評価したのは COMPAS に実装されたアルゴリズムである。COMPAS を開発して提供している Northpointe 社は、企業秘密であることを理由に、法廷への COMPAS のアルゴリズムの提出を拒否した。裁判所はアルゴリズムを審査することなくアルゴリズムが予測した再犯可能性等の予測結果に基づいて量刑した。

ウィスコンシン州最高裁判所は結論として次のように述べている。「この判決で述べたように適切に使われるのであれば、控訴裁判所が判決において COMPAS のリスクアセスメント (COMPAS risk assessment) を考慮することは被告人のデュー・プロセスに対する権利を侵害せず、控訴裁判所は本件において誤った裁量の行使をしなかった。」この判決は、アルゴリズムを直接の審査対象にしていなかったが、COMPAS の適切な利用方法を示すことによって、アルゴリズムの法的規制の在り方について大きな示唆を与えている。

次に、判決が COMPAS の利用方法についてデュー・プロセス違反にならないと判断する論理を判決文から拾ってみよう。以下の箇条書きの冒頭の数字は本稿が参照して注記した FindLaw に掲載された判決文の段落に振られた番号である。

9 「COMPAS のリスクスコア (COMPAS risk scores) を考慮することが他の独立した要因 (factors) によって支持されるのなら、COMPAS の利用は決定的な力をもたないので、誤った裁量の行使にならない。」

88 「COMPAS は決定的なものではあり得ないが、判決を言い渡す裁判所は COMPAS のリスクアセスメントを次のような問題に関連する要因として利用することができる。(1) 再犯リスクが低いが刑務所に拘束される犯罪者を収監しない犯罪者にすること。(2) 犯罪者がコミュニティで安全かつ効果的に監視されることができかどうかを評価すること。(3) 保護観察及び監視の期間及び条件、並びに違法行為に対する対応を課すること。」

98 「リスクスコアは、(1) 犯罪者が投獄されるかどうかを決定するために、又は (2) 刑罰の厳しさを決定するために、利用されてはならない。更に、リスクスコアは、犯罪者がコミュニティで安全かつ効果的に監視されることができかどうかを決定する際に、決定的な要因として利用されてはならない。」

上記の 9 と 88 はいずれも COMPAS の利用を是とする一方で限界を規定している。98 は、COMPAS のアルゴリズムが利用しているデータを考慮しての判断だが、COMPAS の利用について制限を設ける一方で、制限を守れば利用を是とすることが含意されている。これらの論理は COMPAS のアルゴリズムの利用を間接的に規制していると言えるだろう。特に、因果関係を論理的に証明して犯罪の成立を判断する刑事司法においては、因果関係ではなく相関関係で判断するアルゴリズムを使う COMPAS で犯罪の成立を判断することは理論的に無理である³¹⁾。COMPAS のアルゴリズムは、判断過程が技術的にブラックボックスで説明できないだけでなく、アルゴリズム自体が企業秘密を理由に法廷で公開されないので法的

にもブラックボックスである。このようなアルゴリズムに依拠した判決を出せば、判決もまた AI 倫理が要求する透明性を欠くことになる³²⁾。この意味で、上記 9 の論理は GDPR 22 条の「個人データの自動処理のみで決定されない権利」と共通する思想を含んでいる。しかし、この判決は COMPAS にバイアスがないことを前提として出されているが、バイアスを指摘する研究が発表されている³³⁾。

ところで、国際学会の ACM (Association for Computing Machinery) は社会技術システム(socio-technical systems)の FAT (公平性 fairness, 説明責任 accountability, and 透明性 transparency) について国際学会を開催している。FAT という三つの要素は相互に関連しあっている。この FAT を基本的権利として理解するのがアメリカの電子プライバシー情報センター(Electronic Privacy Information Center ; 略称 EPIC)である。EPIC が設立した The Public Voice は 2018 年 10 月 23 日に「人工知能のユニバーサルガイドライン(Universal Guidelines for Artificial Intelligence)」³⁴⁾を発表している。12 項目で構成されるこのガイドラインは、「透明性への権利(Right to Transparency.)」と「人間が決定する権利(Right to Human Determination)」を提唱している。「透明性への権利」は次のように説明されている。「すべての個人は自分自身に関して AI が行なった決定の根拠について知る権利をもっている。これには結論を生み出した諸要素、論理(logic)及び技術へのアクセスが含まれている。」アルゴリズムがこの論理(logic)に含まれることは明らかである。EPIC は知る権利を根拠にしてアルゴリズムの公開を要求して、AI の透明性を実現しようとしている。

6 OECD の AI 政策とアルゴリズム

2019 年 5 月に OECD は「人工知能に関する勧告(Recommendation on Artificial Intelligence)」³⁵⁾を採択した。この勧告は AI を信頼して扱うために「人工知能に関する OECD 原則 (OECD Principles on AI)」³⁶⁾と呼ばれる五つの原則を提唱した。i 「包摂的成長、持続可能な開発、幸福(Inclusive growth, sustainable development and well-being)の原則」、ii 「人間中心の価値と公正(Human-centred values and fairness)の原則」、iii 「透明性と説明可能性(Transparency and explainability)の原則」、iv 「強靱性、セキュリティ、安全性(Robustness, security and safety)の原則」、v 「説明責任(Accountability)の原則」。このうち、iii と v を実施すると AI のアルゴリズムが公開または説明の対象になるのか、アルゴリズムにどのように対応するべきかという問題に直面する。

この勧告は、策定の段階で日本が「人間中心の AI 社会原則」、「AI 開発ガイドライン案」、「AI 利活用原則案」について検討の背景や議論の状況を紹介していたので、これらの原則やガイドラインと整合がとれている。

7 AI のアルゴリズムを説明してもらう権利

7-1 「AI のアルゴリズムを説明してもらう権利」の意義

AI の用途は多様で、AI と人の関係も多様である。このような状況で、AI のアルゴリズムを説明してもらうことを権利として承認する意義を考えてみよう。この意義の中心になる思想は「人間の尊厳」を確保するという思想である。AI については、人間が機械に支配されないという、「個人の自律性」を守るための思想が必要になる。ここから AI の利用者が AI の利用について責任をもつという思想が導出される。これは人権思想を前提にすれば当然である。「人間の尊厳」に由来する「個人の自律性」を尊重するという思想は日本国憲法では 13 条に「すべて国民は、個人として尊重される」と規定されている。AI と人間の関係は個人ごとに異なるので、個人という個性を無視して人一般と AI の関係を規定して、これに個人

を従わせるのは危険である。本稿がこれまで紹介してきた AI の規制は大部分が原則やガイドライン等の倫理的規範であり、AI の規制が強制力のある法律に馴染みにくいことには相応の理由がある。この意味において、自民党の改憲草案が現在の 13 条に規定する「個人として尊重される」を、個人の個性や個別の事情を無視した「人として尊重される」に変更しているのは、人権無視というだけでなく AI 時代に必要とされる倫理に逆行する。日本国憲法 13 条前段が規定する個人を尊重する思想はすべての人権の基礎である。

AI の在り方について様々な考え方が法制度やガイドライン、指針等で示されてきたが、これらは全て国に加えて AI の開発者や利用者が遵守すべき個別の価値を表明したものである。AI を開発・利用することに責任があるなら、AI を利用して評価される者が AI の利用について知らされなければ、この責任に対するチェックが効きにくいのは自明である。しかも、AI の利用関係は当事者ごとに異なるにもかかわらず、個別の AI 利用関係において遵守されるべき価値には社会が共有すべき基本的な価値がある。この価値は公正と呼ばれ、AI のアルゴリズムのバイアスを解消するものとして議論されることが多いが、人間の尊厳を確保するための価値である。AI を使わなくても人間の判断にはバイアスがある。AI のアルゴリズムにアルゴリズムを設計する者の無意識のバイアスがかかることは広く指摘されてきた³⁷⁾。AI のアルゴリズムについてバイアスの増幅を懸念する議論が多いが、AI のアルゴリズムは仕様次第で人間の判断のバイアスを減少させることができる³⁸⁾ので、AI はそのために導入されることが多い。しかも、AI のアルゴリズムからバイアスを除去することは不可能である³⁹⁾。ここからアルゴリズムの公正性を定期的に評価する必要性が発生する。公正さが失われると差別が生まれる。ただし、公正も差別も明確な定義が困難で内容は文脈に応じて多様である。この公正性を維持するために、AI の利用についてこれまでデュー・プロセスと呼んできた適正手続の一環として透明性と説明責任が要求される。この適正手続という思想も日本国憲法の 13 条と 31 条に由来する。2017（平成 27）年 2 月に発表された「人工知能学会倫理指針」も第 4 条で人工知能の開発と利用について「公正性」を求めて、次のように規定している。「人工知能学会会員は、人工知能の開発と利用において常に公正を持ち、人工知能が人間社会において不公平や格差をもたらす可能性があることを認識し、開発にあたって差別を行わないよう留意する。人工知能学会会員は人類が公平、平等に人工知能を利用できるように努める。」何が公正であるかについては具体的な場面を想定して常に議論を続ける必要があるが、抽象的な理念としての公正の重要性を否定する根拠はないと思われる。

AI の規制は法律によらない規制が望ましいと一般に考えられている。それにもかかわらず、AI を規範する倫理は日本国憲法の基本的な価値から導くことができる。日本国憲法が自然法思想に基づいている⁴⁰⁾ので、これは当然である。これは EU も同じ事情である。アルゴリズムの問題を「AI のアルゴリズムを説明してもらう権利」として言語化すれば、この権利の範囲と規範性について議論が進むに違いない。GDPR の紹介で指摘したことだが、「説明してもらう権利」は他の権利や監査等の仕組みとセットになって権利が実現されやすくなる。「AI のアルゴリズムを説明してもらう権利」は、GDPR を根拠として主張されている「説明してもらう権利」の一部でやや具体化されているが、それでも同様の指摘が可能である。それゆえ、「AI のアルゴリズムを説明してもらう権利」は法律要件がそれほど明確ではない抽象的な権利として理解されても構わない。抽象的な権利であるにせよ、権利としての存在を承認されるところから、これを実現するための法的思考や法的施策が工夫されることになる。この点は「自己に関する情報の流れをコントロールする権利」という定義で代表される情報プライバシー権が、その不明確性にもかかわらず、否

それゆえにこそ、個人情報保護制度を立法化させる原動力になったことを振り返れば明らかである。「説明してもらう権利」の説明で紹介したように、ガバナンスアプローチやシステムアプローチは必要だが、これらのアプローチを具体的に構築するにはそれを支える法的理念が必要である。この法的理念に該当するものが「AI のアルゴリズムを説明してもらう権利」である。アルゴリズムが人々の気づかないところへ普及して、人々は気づかないままアルゴリズムに支配されている。我々はこんな社会に生きている。「AI のアルゴリズムを説明してもらう権利」が権利に値しないと批判したり、この権利の主張を根拠のない戯言だと批判したり。こんな批判が起こるだけでも、アルゴリズムの透明性と説明責任について社会的な規範が形成されて行き、アルゴリズムの公正さが担保されるための社会的な仕組みが形成されて行くはずである。

7-2 「AI のアルゴリズムを説明してもらう権利」の根拠

EU では GDPR 22 条が「個人データの自動処理のみで決定されない権利」を保障しているが、AI のアルゴリズムをアルゴリズムで評価された者に説明することについて、GDPR には明文の規定がなく、GDPR のガイドラインはアルゴリズムをどの程度まで説明すれば良いのか不明確である。日本には GDPR 22 条のような規定がない。そのような日本で AI のアルゴリズムを説明してもらうことが法的な請求の対象になりうるのだろうか。この問いに対して、本稿は現行法の下でも法的請求の対象になる場合があるという立場で、私法上の根拠と公法上の根拠を以下で検討する。本稿のここまでの議論で明らかだと思われるが、AI についての倫理的規範はそれぞれ説明の仕方に多少の違いがあるものの世界共通といえるレベルに達している。日本、EU、イギリス、アメリカは、OECD を通じて、AI や AI を稼働させる個人データの保護、セキュリティについて共通の政策を採用している。人間の尊厳を守るために個人の自律性を保障し、不当な差別を避けるために AI のアルゴリズムの内容に公正性を求め、こっそり AI を使われないために AI の利用について透明性を保障すること、その手段として AI で個人を評価する者が説明責任を負うこと、AI で利用される個人データを保護するためにセキュリティを確保すること等は、世界共通の倫理である。AI を使って決定される個人への評価を形成する過程で、これらの価値を実現することは法の適正手続（due process of law）の要請に合致する。

7-2-1 私法上の根拠

AI を利用して他人を評価する業務について、評価される本人にアルゴリズムを説明することを事業者が義務付ける明文の法的根拠は存在しない。しかし、このような説明が倫理的に要請される場面が現代社会に存在することは、これまでの議論で明らかであろう。特に、EU に対しては、AI について GDPR と同等の規制を日本で実施しなければ日本と EU 間の人とモノの交流が困難になる。

AI のアルゴリズムを説明しないことが民法の不法行為に該当すれば、アルゴリズムの説明を義務づける法律がなくても、AI によって評価される者に対して評価で使われたアルゴリズムの説明が必要になる。民法 709 条は不法行為の要件として「他人の権利又は法律上保護される利益」の侵害を規定している。「法律上保護される利益」は、判例上、法律の保護に値する利益なので、「AI のアルゴリズムを説明してもらう」ことが民法 709 条の保護する「法律の保護に値する利益」になる場合があることは想像にかたくない。AI で評価される者が AI のアルゴリズムを知っていたら避けることができた不利益を受けた場合、AI のアルゴリズムについて説明を受けることが「法律の保護に値する利益」に該当するかどうかを検討

されるべき問題として浮上する。では、どのようにして「法律の保護に値する利益」の有無を判断するのか。現行法の下では、現実が発生した不利益を評価して、AI 倫理に対する侵害の程度が著しいと判断できる場合、法的利益として保護する必要性が発生すると考えられる。この判断にとって、先に紹介した総務省情報通信政策研究所の「AI ネットワーク社会推進会議報告書 2019」が提唱する「AI 利活用原則」は有力な指針である。「AI 利活用原則」には法的拘束力がないが、この原則に含まれる「viii 公平性の原則、ix 透明性の原則、x アカウンタビリティの原則」が現実の事例において AI のアルゴリズムについて説明を求める場合があると思われる。このような事例が法廷で争われると、裁判所が「法律の保護に値する利益」への該当を認定する可能性があると思われる。「AI 利活用原則」は先に紹介した欧州委員会の「AI に関する高度専門グループ」が公表した「四つの倫理原則」とそれを実現するための「七つの要件」と内容が重複するので、「AI 利活用原則」の考え方を裁判所が法源の一つである条理として適用すれば、AI のアルゴリズムについて日本が EU と共通の政策を採用したことになる。そして、この共通政策は日本と EU がともに加盟する OECD の「人工知能に関する勧告」の趣旨に反しない。

また、GDPR 22 条は「個人データの自動処理のみで決定されない権利」の対象を「法的効果又は同等の重大な影響を自分自身にもたらす」自動処理に限定している。この規定が存在しない日本において、「法的効果又は同等の重大な影響を自分自身にもたらす」自動処理のすべてについて、民法 709 条の不法行為に該当しないと予め決めつけることはできないであろう。

ただ、AI のアルゴリズムを説明してもらうことが「法律の保護に値する利益」になるとしても、裁判所に訴えて勝訴判決をもらわない限り、これを認めてもらえないかもしれないというのは、AI のアルゴリズムを説明して欲しい者にとって法的地位が不安定で酷である。それに、民法 709 条でこんな争いをする場合、賠償請求金額は何円だろうか。仮に裁判所が原告勝訴の判決を出すとしても、驚くほど低い金額の賠償を命じるはずである。訴訟費用を予想した場合、争うメリットがあると判断できるのはどんな人だろう。名目的損害賠償制度でも創設しない限り、アルゴリズムを説明してもらえなかったことを法廷で争う人は篤志家以外にいないだろう。

そこで更に、一步踏み込んで問題提起をしたい。「AI のアルゴリズムを説明してもらう」ことは「権利」であり、「利益」にとどまるものではない。本稿がこのように主張するのは、「AI のアルゴリズムを説明してもらう権利」がプライバシー権から派生していると理解しているからである。データ主体にとって自己のデータを AI のアルゴリズムでみだりに評価されることはプライバシー侵害になる。日本ではプライバシー権といえば個人情報の流通を保護する権利だと理解している人が多いが、プライバシー権が生まれたアメリカでは最高裁判所が憲法上のプライバシー権を個人の自律を確保するための権利として理解している⁴¹⁾。1970 年代に日本で生まれて通説になった情報プライバシー権というプライバシー権概念は、プライバシー権が個人情報の流通に対して個人の自律を発揮する局面を描いたものとして理解されるべきである⁴²⁾。AI に対して個人の自律を確保するためにアルゴリズムを説明してもらうことを権利として要求する。これはプライバシー権の行使そのものである。GDPR は個人の自律を尊重するためのものである。

このような主張に対して、アルゴリズムは知的財産として法的に保護されるという反論が成立する場合があることは考慮しておく必要がある⁴³⁾。確かに、アルゴリズムがプログラムとして表現されると、アルゴリズムへ直接の保護は及ばないものの、表現されたものについて著作権が成立する。プログラムに自然法則を利用した技術的思想として高度の創作性があれば発明になり、出願すればプログラムに特

許権が成立して、アルゴリズムに実質的な保護が及ぶことになる。しかし、著作権も特許権も本稿のテーマである個人を評価するために作られたアルゴリズムの保護手段としては機能しないだろう。プログラムの表現を変えれば著作権を侵害することなくアルゴリズムを真似ることができるし、特許権の審査期間は長すぎる上に、出願に伴い発明が公表されるのでアルゴリズムの構造が公になる。本稿のテーマにとって、アルゴリズムを知的財産として保護しうる現実的な唯一の手段は不正競争防止法である。不正競争防止法は「秘密として管理されている生産方法、販売方法その他の事業活動に有用な技術上又は営業上の情報であって、公然と知られていないもの」(2条6項)を「営業秘密」として保護している。この営業秘密に該当するアルゴリズムを民間の事業者が、任意に公開するのではなく、不正競争防止法を根拠にして公開を拒否したら、この拒否は基本的に正当な拒否である。では、アルゴリズムの公開が請求される場合、現行法で営業秘密に該当するアルゴリズムの全てについて非公開にすることが立法政策として適正であろうか。希望者に対して不利益な判断を下した者が相手に対して説明責任を果たすことが正義に叶う場合があると思われる。このような場合にアルゴリズムの開示が求められたら、アルゴリズムを開示するか、アルゴリズムについて理解可能な説明をする必要があると思われる。ただし、この説明を実現するためには不正競争防止法の改正が必要であり、改正のために関係業界の反対を押し切ったとしても、開示又は説明する必要があるアルゴリズムを明確に類型化する必要がある。この類型化には困難が予想される。

不正競争防止法上の営業秘密に該当しそうなアルゴリズムについて、開示又は説明を義務づける基準を一律に規定することは容易ではない。これは現状ではケースバイケースの判断に頼らざるをえない。不正競争防止法で保護されるアルゴリズムの透明性を開示又は説明という手段で実現するとしても、保護されるアルゴリズムの全てがその対象になるわけではない。日本の研究者にも事業者によるプロファイリングの実施について第三者機関によるアルゴリズムの監査を義務づけるという提案があるが⁴⁴⁾、現状ではこれを一律に義務づけることは困難で、ある程度は自主規制に頼らざるをえないだろう。第三者機関の監査を受ける制度は、強制でもない限り、事業者に自主規制の意思と経済的余裕が必要であり、第三者機関の設立母体が業界団体であれば、第三者機関に対する公的機関の監督が必要になる。これらの問題も現実を踏まえてGDPRを参考にしたより前向きな議論が必要である。

現行法を前提とすると、AIのアルゴリズムを公開してもらうことも説明してもらうことも、現実には事業者の判断次第ということになりそうである。このような要求のために法廷で争う国民は限られている。この原因は、個人に対して法的効果や重大な結果が及ぶ決定をAIの自動処理だけで行うことについて、日本の個人情報保護法が何も規定していないことにある。つまり、GDPR 22条のような規定が日本にはないことが原因である。現在のAIがパーソナルデータを含むビッグデータによって稼働可能になっている以上、個人情報保護法にAIを規制する明確な規定が存在しないのは問題である。

ここまでは私法関係について具体的に適用される可能性がある条項についての議論である。最後に、理念ともいえるべきやや抽象的な根拠を紹介しておきたい。本稿がここまで展開した主張は私法の一般原則である民法1条の諸原則に合致するのではないだろうか。「公共の福祉」(1条1項)の観点からAIの利用について制約を承認し、「信義誠実の原則」(1条2項)の観点からAIのアルゴリズムについて説明する必要があることを承認し、「権利濫用の禁止」(1条3項)の観点から当事者の対等でない関係を前提とした権利行使を抑制すると考えれば、民法の基本原則は本稿の主張を背後で支えるものになる。また、民法の規定のうち、公序良俗に違反する法律行為を無効にする90条、不法行為に損害賠償請求を

認める 709 条を通して、私法関係に憲法の「個人の尊重」(13 条)、「法の下での平等」(14 条)及び「適正手続」(31 条)が間接的に適用されると考えれば、個人の尊重と自律を守るために、AI を利用して個人を評価する者と評価される者を対等な関係にして、AI のアルゴリズムを説明してその内容を相互に共有するという手続的な正義が憲法によって要請されることになる。

7-2-2 公法上の根拠

行政手続法 5 条によって、行政庁は審査基準を定めるだけでなく、この基準を「できる限り具体的なものとしなければならない」上に、「公にしておかなければならない」。この審査基準は法律で定められた許認可等の法律要件を行政裁量によって詳細にしたものである。多数の申請に対して数少ない許認可を与える場合、公平な判断を確保するために AI の導入が検討される。この場合、最大の問題はアルゴリズムをどのように設計するかである。不公平を排除しようとするほどアルゴリズムは複雑で分かりにくくなり、しかもバイアスから免れることができない。行政処分について、EU の GDPR 22 条のように、AI のみの判断で行ってはならないと義務づけるのは、現場の行政担当者にとって行政裁量を与えられるのではなく押し付けられるに等しい。

行政庁へ利益処分を申請する者が多数いて、不公平な審査を避けるために行政庁が AI を利用する場合、行政庁は法律の定める法律要件を具体化する要件事実を詳細に検討して、要件事実に優先順位等の配慮を加えて審査基準を定めてアルゴリズム化する。この審査基準は行政手続法 5 条によって公にされる。この場合、審査基準としてのアルゴリズムは行政裁量の範囲内にあり、行政庁が公平に審査をしようとする限り、行政庁の判断は AI の判断と同じにならざるをえない。こんなことにはお構いなく、AI の判断を唯一の根拠として申請を承認された者はこの審査に不服をもたないであろう。不利益処分には理由が付記される(行政手続法 8 条、14 条)が、申請を承認されなかった者が不服を申立てた場合、行政庁は AI の判断と行政担当者の判断が一致したことを説明する以外に、アルゴリズムを説明する必要が生じると思われる。行政法学で蓄積された理由附記の法理によってアルゴリズムは決して新しい問題ではない。

多様な行政処分の個々の事例を本稿で検討することはできないが、行政処分が裁量に基づく処分であるにせよ、行政処分が法律を根拠に行われる以上、AI を利用して行われた行政処分は国民の求めに応じてアルゴリズムの説明を伴うと考えられる。そのためにも、個人情報保護制度の改正が望ましい。

8 結び

人工知能(AI)と呼ばれる技術が社会生活のいたるところに浸透している。この傾向は今後も増えることがあっても減ることはないだろう。この技術は「知能」という名称がついている上に、人間には不可能な瞬時の判断ができるため、過剰な期待とシンギュラリティ論に代表される過剰な不安を伴っている。AI へ過大な信頼を寄せてしまうと、AI の判断のままに生きる人間が登場する。本稿は個人が自律性を確保して AI を使いこなすために必要な法制度の一端について議論した。AI を利用して生じる最大の問題は AI の判断を左右するアルゴリズムを当事者が理解しない状況が生まれることである。本稿は AI を利用して個人が評価される場合に論点を限定して、AI のアルゴリズムが AI によって判断される者に対して必要に応じて説明されるべきことを主張した。本稿はこの主張の根拠を比較法的な観点から展開し、AI のアルゴリズムを説明してもらうことが「法律の保護に値する利益」であり「権利」に該当する場合があることを主張した。また、現状では「AI のアルゴリズムを説明してもらう権利」がプライバシー権

に由来する抽象的な権利であることも指摘した。

個人はそれぞれ多様な関係を形成しながら生きている。この関係を本稿は私法関係と公法関係に分けて議論したが、私法関係にせよ公法関係にせよ、AI のアルゴリズムを説明してもらうことに法的根拠があるかどうかは、具体的な個々の法律関係の中で判断されることになる。この意味において、本稿は未開拓に近い「AI のアルゴリズムを説明してもらう権利」について総論的・包括的な議論を行ったことになる。

アルゴリズムの規制は人工知能社会の最大の問題である。アルゴリズムを下手に規制すると人工知能の開発が遅れてしまう。これは日本の国際競争力を阻害することにもなる。しかし、AI の利用状況を見て誰もが不安を感じている。一人一人の人間が自律と尊厳を維持して生きて行ける社会を実現するためには、アルゴリズムの規制を国際的な協調体制の中で実現するしかないと思われる。ネットで繋がった社会では、アルゴリズム規制の国際的協調体制だけがアルゴリズムの不当な支配を避ける唯一の選択肢である。このような状況下で、日本が AI のアルゴリズムを開示したり説明したりすることに消極的であれば、アルゴリズム規制の国際ルールに日本の意見が反映されにくくなる。AI のアルゴリズムを規制する是非についてもっと議論してほしいというのが自戒を込めた本稿の願いである。

注

- 1) 私は「知能」に値する人工知能が開発されているとは考えていない。本稿はアルゴリズムの利用に対する権利が主題なので、人工知能を厳密に定義することはしない。総務省も、平成 28 年版「情報通信白書」以来、人工知能の定義を断念している。総務省は、知性や知能の定義ができないので人工知能の定義も困難、という立場である。名人位の棋士が将棋の AI に勝てなくなったが、このように目的を特化した「弱い AI」ではなく、開発中の「強い AI」や「汎用 AI」も名前こそ人工知能だが、私はそれらの技術を知能に値するとは考えていない。AI はどんなに進歩してもただの道具である。私はシンギュラリティを信じない。参照、ジャン=ガブリエル・ガナシア（著）、伊藤直子・他(訳)『そろそろ、人工知能の真実を話そう』（早川書房、2017）、西垣通『AI 原論 神の支配と人間の自由』（講談社、2018）。
- 2) 個人情報保護法 2 条 6 項は個人データを「個人情報データベース等を構成する個人情報」と定義している。
- 3) キャシー・オニール著、久保尚子翻訳『あなたを支配し、社会を破壊する、AI・ビッグデータの罠』298 頁以下（インターソフト、2018）はアルゴリズムの出力結果を監査する必要性を力説している。しかも、監査を自動化する場合は、監査が公正になるように監査方法自体を事前に人間が検証することを主張している。この主張はアルゴリズムの開発者を規制することによってアルゴリズムによる不当な支配を避けようとする思想を前提にしている。AI の開発者を倫理的に規制するという思想は後に紹介する各種の AI 倫理で承認されているが、いずれも抽象的なルールでアルゴリズムの規制に具体的には踏み込んでいない。本稿は、開発者に対する規制を必要だと考えるが、AI で評価される者の立場からアルゴリズムの規制を考えてゆく。
- 4) ネット上の匿名化された情報の大部分から本人を特定できることが証明されている。Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye, *Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models*, 10 NATURE COMM. 1 (2019).
- 5) 山本龍彦『プライバシーの権利を考える』266 頁（信山社、2017）。

- 6) NBL 1137 号 64-85 頁 (2019)。
- 7) 山本龍彦「『完全自動意思決定』のガバナンス —行為統制型規律からガバナンス統制型規律へ?」情報通信政策研究 3 巻 1 号 25 頁 (2019)。
- 8) AI ネットワーク社会推進会議「AI 利活用ガイドライン ~AI 利活用のためのプラクティカルリファレンス~」10 頁 (令和元年 8 月 9 日)。
- 9) AI ネットワーク社会推進会議「AI 利活用原則の各論点に対する詳説」38 頁 (令和元年 8 月 9 日)。
- 10) GDPR は英語表記で、フランスとイタリアでは RGPD、ドイツでは DSGVO と略称される。イギリスの EU 脱退に対応して、規則の名称を検索の便宜のためフランス語正規版の表記で紹介しておく。RÈGLEMENT (UE) 2016/679 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) .
- 11) GDPR について既に多くの紹介がなされているが、全体の概説として、宮下 紘『EU 一般データ保護規則』(勁草書房、2018)、小向太郎・石井夏生利『概説 GDPR』(NTT 出版、2019)。
- 12) 29 条作業部会は GDPR の施行とともに廃止され、欧州データ保護委員会(European Data Protection Board ; 略称 EDPB)が後継の機関として活動を引き継いだ。
- 13) ARTICLE 29 DATA PROTECTION WORKING PARTY, GUIDELINES ON AUTOMATED INDIVIDUAL DECISION-MAKING AND PROFILING FOR THE PURPOSES OF REGULATION 2016/679, 17/EN. WP 251rev.01 (Feb. 6, 2018).
- 14) GDPR のフランス語版で表記されている *loyauté* は、英語版では *fairness* だが、ドイツ語版では *Glauben*、イタリア語版では *correttezza* である。国ごとにニュアンスが異なり適切な訳語を決めがたい。総合すると「公明正大で正しくて信じることができる」というニュアンスがあるようなので、あえて「公正性」と訳した。
- 15) ここでは、条文の翻訳であることを配慮して、フランス語の *personne concernée* の訳語に「本人」ではなく、我が国で定着している訳語の「データ主体」を当てた。本来「データ主体」は *data subject* の訳語で個人情報保護法の「本人」に該当する。私は「データ主体」という訳語を、プライバシー権の積極的権利としての側面を強調するための、日本的風土に配慮した意識で、意図的な誤訳だと推定している。
- 16) EU は GDPR に先立ち 2010 年に「プロファイリングにおける個人データの自動処理に関する個人の保護のための、閣僚委員会の加盟国への勧告(CM/Rec(2010)13)」(Recommendation CM/Rec(2010)13 of the Committee of Ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling)を採択して、加盟国に勧告している。この勧告は AI のアルゴリズムに言及していない。もし、日本が EU の加盟国であれば、AI でプロファイリングをする事業者がプロファイリングに用いる個人データのセキュリティを厳格に管理し、データ主体の権利を保障するための措置を、個人情報保護法に規定する法改正を求められるところであった。2015 (平成 27) 年に行われた個人情報保護法の改正では EU のこの勧告に相当する規定が入らなかった。
- 17) GDPR の「プロファイリング」に関する規定は我が国でも注目を集めて紹介されてきた。山本達彦「ビッグデータ社会とプロファイリング」論究ジュリスト 18 号 34-44 頁 (2016)、石井夏生利「プロ

ファイリング規制」ジュリスト 1521 号 32-37 頁 (2018)、石井夏生利「GDPR におけるプロファイリング規制の概要」自由と正義 70 巻 6 号 15-20 頁 (2019)。

- 18) Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' is Probably not the Remedy You Are Looking for*, 16 DUKE L. & TECH. REV. 18, 84 (2017).
- 19) この要件は「アカウントビリティ」と訳されることがあるので、誤解を避けるために、以下に補足しておく。explicabilité の原則を実現する要件なので responsabilité に「説明責任」という訳語をつけた。説明責任に対応する英語は responsibility ではなく accountability である。英語の accountability のフランス語訳は responsabilité である。フランス語版の responsabilité はイタリア語版なら responsabilità となるはずだが、イタリア語版ではここだけ英語版と同様に accountability と表記している。英語圏では responsibility と accountability を区別しているが、日本語と同様にフランス語とイタリア語には accountability に相当する言葉がないようだ。ちなみにドイツ語版では Rechenschaftspflicht と表記されている。これは responsabilité とほぼ同義である。
- 20) Groupe d'experts de haut niveau sur l'intelligence Artificielle, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, 2019, p.18.
- 21) Groupe d'experts de haut niveau sur l'intelligence Artificielle, supra note 20, p.24.
- 22) Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT'L DATA PRIVACY L. 233 (2017).
- 23) Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76, 97 (2017)は、「説明してもらう権利 (right to explanation)」が曖昧であることを批判して、これに代わり説明責任を果たす法的対策を模索している。
- 24) Bryan Casey, Ashkon Farhangi & Roland Vogl, *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L. J. 143, 149 (2019).
- 25) Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L. J. 189 (2019).
- 26) Recital 71 Profiling (<https://gdpr-info.eu/recitals/no-71/>)
- 27) Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L. J. 189, 205 (2019).
- 28) <https://www.gov.uk/government/publications/landscape-summaries-commissioned-by-the-centre-for-data-ethics-and-innovation>
- 29) State v. Loomis, 881 N.W.2d 749 (Wisc. 2016), cert. denied, Loomis v. Wisconsin, 137 S.Ct. 2290 (2017). (<https://caselaw.findlaw.com/wi-supreme-court/1742124.html>) . See State v. Loomis, 130 HARV. L. REV. 1530 (2017).アルゴリズムの観点からこの判決を検討するものとして、山本龍彦、尾崎愛美「アルゴリズムと公正 : State v. Loomis 判決を素材に」科学技術社会論研究 16 号 96-107 頁 (2018)。
- 30) Northpointe 社による COMPAS の宣伝用の解説として “COMPAS Risk & Need Assessment System” (http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf) が web 上に残されている。
- 31) 同旨、緑 大輔「アルゴリズムにより再犯可能性を予測するシステムの判断結果を考慮して裁判所が量刑判断を行うことが、適正手続保障に反しないとされた事例」判例時報 2343 号 129 頁 (2017)。
- 32) Han-Wei Liu, Ching-Fu Lin & Yu-Jie Chen, *Beyond State v. Loomis: Artificial Intelligence, Government*

Algorithmization, and Accountability, 27 INT. J. OF L. & INFO. TECH. 122 (2019)は COMPAS のアルゴリズムが法的にも技術的にもブラックボックスであることを批判する。この論文は COMPAS のバイアスが誤った判断を導くことも指摘する。

- 33) Panel for the Future of Science and Technology (STOA), *A governance framework for algorithmic accountability and transparency* (2019). Anne L. Washington, *How to Argue with an Algorithm: Lessons from the COMPAS- Propublica Debate*, 17 COLO. TECH. L.J. 131 (2018).
- 34) <https://thepublicvoice.org/ai-universal-guidelines/>
- 35) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- 36) <https://www.oecd.org/going-digital/ai/principles/>
- 37) 江間有沙『AI 社会の歩き方 人工知能とどう付き合うか』35 頁以下 (化学同人、2019)。
- 38) Cass R. Sunstein, *Algorithms, Correcting Biases*, <<https://ssrn.com/abstract=3300171>> (2018) はアルゴリズムでバイアスを除去できると主張する。
- 39) Ronald Yu & Gabriele Spina Ali, *What's Inside the Black Box? AI Challenges for Lawyers and Researchers*, 19 LEGAL INFO. MGMT. 2, 4 (2019).
- 40) 日本国憲法前文の「国政は、国民の厳粛な信託による」、「人類普遍の原理」、「政治道德の法則は、普遍的」等は自然法思想を前提とした文言である。
- 41) 代表的な判例として、*Roe v. Wade*, 410 U.S. 113 (1973)がある。橋本公亘「プライバシーの権利」芦部信喜・奥平康弘・橋本公亘編『アメリカ憲法の現代的展開 1 人権』3 頁以下 (東京大学出版会、1978) は *Griswold v. Connecticut*, 381 U.S. 479 (1965)に始まる憲法上のプライバシー権に関するアメリカの判例を分析して、アメリカの憲法上のプライバシー権を自律権として捉えている。このような観点は現代でも、芦部信喜著、高橋和之補訂『憲法 第七版』124 頁 (岩波書店、2019) に見られる。芦部自身は消極的であるように見えるが、芦部も指摘するようにプライバシー権の議論は情報をコントロールするという観点が重視されるようになっていく。このような傾向を決定づけたのは、自己情報のコントロール権をプライバシー権として保障することを明確にするために、自律権である自己決定権をプライバシー権から分離して体系化した佐藤幸治の一連の業績であり、佐藤幸治『憲法』454 頁 (青林書院、1995) にその一つを見ることができる。情報プライバシー権もプライバシー権の一つであるが、情報コントロールという観点だけでプライバシー侵害の有無を考えると、コントロール自体が自律行為であるにも拘らずこの側面が忘れられてしまい、身体 of プライバシー (憲法 33 条)、場所 of プライバシー (憲法 35 条) 等の空間を保護するプライバシー権という側面も忘れられてしまう。サーバに対するパケット攻撃についてセキュリティの観点から議論されても、プライバシーの観点からの議論がされていないのは、情報プライバシー権の影響だと思われる。
- 42) 情報社会に直面した 1970 年代の日本人が個人の尊厳をもっと強く自覚していたら、プライバシー権の確立を目指して、憲法学がプライバシー権を個人情報に対する権利に狭小化することはなかっただろう。その結果、日本のプライバシー権は自律という概念から遠ざかり空間に対する規律を失った。
- 43) 石井夏生利「EU 一般データ保護規則におけるクレジットカード情報の取扱い」日本クレジット協会クレジット研究 7 号 33 頁 (2018)。
- 44) 竹地 潔「ビッグデータ時代におけるプロファイリングと労働者への脅威」富山大学紀要富大経済論集 63 巻 1 号 19 頁 (2017)。