

〔論 説〕

税務行政における

AI (人工知能)・機械学習アルゴリズムの利用と法的問題

—調査選定システムの検討を中心として—

泉 絢 也

I 研究の目的

各国の税務行政がテクノロジー、とりわけ AI (人工知能)、その中心にある機械学習やデータマイニングの利用を進めている⁽¹⁾。日本の国税庁も、平成 29 (2017) 年 6 月に「税務行政の将来像～スマート化を目指して～」を公表した。これは、情報システムの高度化、外部機関の協力を前提として、現時点で考えられるおおむね 10 年後のイメージを示したもので、ICT の活用による「納税者の利便性の向上」と「課税・徴収の効率化・高度化」を柱として、スマート税務行政に進化していくことを宣言するものである。

具体的には、例えば、調査の必要性が高い大口・悪質な不正計算が想定される事案を的確に選定する観点から、過去の接触事績や資料情報の体系的なチェックに加え、統計分析の手法を活用することにより、納税者ごとの調査必要度の判定を精緻化するとともに、最適な接触方法や調査が必要な項目についても、システム上に的確に提示されるようになることが望ましいとされている⁽²⁾。また、AI のみならず、BI (Business Intelligence: 大量のデータを分析・可視化し、迅速な意思決定を補助するツール) 及び BA (Business Analytics: 統計学や機械学習等の技術を用いてデータ分析を行うツール) 等も駆使して、調査選定のほか、財産評価、各種パフォーマンス分析、架電コールリスト作成、申告審査・資力調査、最適な接触方法など各種の分析・検討を行うとしている。しかしながら、そこ

-
- (1) See e.g., OECD, *ADVANCED ANALYTICS FOR BETTER TAX ADMINISTRATION: PUTTING DATA TO WORK* (2016); OECD, *TECHNOLOGY TOOLS TO TACKLE TAX EVASION AND TAX FRAUD* (2017); OECD, *TAX ADMINISTRATION 2019: COMPARATIVE INFORMATION ON OECD AND OTHER ADVANCED AND EMERGING ECONOMIES* (2019); OECD, *TAX ADMINISTRATION 3.0: THE DIGITAL TRANSFORMATION OF TAX ADMINISTRATION* (2020) (hereinafter *TAX ADMINISTRATION 3.0*); Marcus Zackrisson et al, *AI and Tax Administrations: A Good Match*, 74 *BULLETIN FOR INTERNATIONAL TAXATION* 619 (2020). 望月爾「税務手続のデジタル化に向けた検討の動きと今後の課題」*税理* 64 巻 1 号 4 頁以下 (2021). 日本税理士会連合会国際税務情報研究会「主要国の税務行政の ICT/AI 化の展望と未来の税務専門家制度についての考察」(2021), [https://www.nichizeiren.or.jp/wp-content/uploads/doc/nichizeiren/business/taxcouncil/\(210126\)toushin.pdf](https://www.nichizeiren.or.jp/wp-content/uploads/doc/nichizeiren/business/taxcouncil/(210126)toushin.pdf) 参照。なお、本稿で引用する URL の最終閲覧日はいずれも 2021 年 6 月 14 日である。
- (2) <https://www.nta.go.jp/information/release/kokuzeicho/2017/syouraizou/pdf/smart.pdf>. https://www.nta.go.jp/information/release/kokuzeicho/2017/syouraizou/pdf/syouraizo_2906.pdf. 酒井克彦編著『スマート税務行政でこう変わる!! キャッチアップデジタル情報社会の税務』(ぎょうせい 2020), 同「税務行政におけるデジタル化と納税環境」*企業研究* 36 号 21 頁以下 (2020) も参照。

ではテクノロジーの利用による利点ばかりが並べられていて、マイナス面、とりわけ想定しておくべき法的問題や納税者の権利利益への影響に関して何ら触れられていない。国税庁は、令和3(2021)年6月に「税務行政の将来像」を改定し、「税務行政のデジタル・トランスフォーメーション」の実施を宣言しているが、状況は変わらない⁽³⁾。

このことを踏まえ、本稿では、納税者の調査必要度やコンプライアンスリスクを測定し、税務調査の対象者を自動的に選定するシステム(以下「調査選定システム」という)、とりわけ大量のデータから有用な情報を抽出するデータマイニングを行い、機械学習により、納税者のコンプライアンスリスクのスコアをはじめ出すモデルを生成するようなアルゴリズムを国税庁が採用する場合に想定される法的問題を明らかにする(ただし、国税庁が採用しようとしている調査選定システムの詳細は明らかではないことに留意)。具体的には、学習用データ、バイアス、ブラックリスト、ブラックボックス、偽陰性と偽陽性等に関する問題が存在すること及びこれらが個人の尊重、幸福追求権、平等原則を基本原理として掲げる憲法13条や14条、あるいは適正手続を保障する31条や84条等への抵触問題に発展する可能性があることを指摘した上で、税務行政のスマート化やデジタル・トランスフォーメーションを進めていくに当たり、国税庁はこれらの点に関して国民に見解や方針を示すべきであると主張する。

II 機械学習とデータマイニング

機械学習やデータマイニングの概要や特性について、簡単に確認しておく。

1 概要

機械学習とは、人間の学習に相当する仕組みをコンピューター等で実現するものであり、一定の計算方法(アルゴリズム)に基づき、入力されたデータからコンピューターがパターンやルールを発見し、そのパターンやルールを新たなデータに当てはめることで、その新たなデータに関する識別や予測等を可能とする手法である⁽⁴⁾。機械学習には大別して学習と推論の2つのプロセスがあり、基本的にそれぞれのプロセスで異なるデータを用いる。学習とは、入力されたデータ(学習用データ)を分析することにより、コンピューターが識別等を行うためのパターンを確立するプロセスである。この確立されたパターンを学習済みモデルという。推論とは、学習のプロセスを経てできあがった学習済みモデルにデータを入力し、確立されたパターンに従い、実際にそのデータの識別等を行うプロセスである⁽⁵⁾。

データマイニングとは、大規模なデータセットを統計的に分析し、新たなパターンや傾向を発見するものである。集計データを用いて特徴や傾向を特定し、個人ではなく人々の集団の識別や特徴付けを行ったり、事象間の関連性を把握したり、将来事象を予測したり

(3) 前掲注(2)の国税庁の資料のほか、次の資料も参照。

https://www.nta.go.jp/information/release/kokuzeicho/2017/syouraizou/pdf/syouraizo_r0106.pdf

https://www.nta.go.jp/about/introduction/torikumi/digitaltransformation/pdf/syouraizo2_r0306.pdf

(4) 総務省『令和元年版 情報通信白書』83頁(2019)参照。

(5) 総務省・前掲注(4)85頁参照。

することが可能となる⁽⁶⁾。税務行政の文脈でいえば、これまで見落としていたデータ上の相関を発見したり⁽⁷⁾、有意な相関関係が認められるものとして、不正取引や過少申告などのコンプライアンスリスクを判定する際に入力データとすべき疑わしい取引の特徴を導き出したり、要注意人物や要継続監視対象者を特定したりするために、データマイニングが行われる。

以上を踏まえて、本稿で想定するのは、確定申告の事績や過去の調査事績などを含む多種多量のデータを分析し、ここからコンピューターがコンプライアンスリスクの高い申告とそれ以外の申告を区別するパターン又はルールを自動的に発見し、統計的・確率的な予測に基づいてコンプライアンスリスクが高く、優先的に調査対象として選定すべき納税者を抽出するというような機械学習アルゴリズムである。その際、人間があらかじめ予測のための関数を用意する必要はない。場合によっては、リアルタイムで情報を収集し、調整を行い、予測の精度を高めていくようなシステムも見据えている。

2 特性

法的な観点からは、次に示す機械学習の主要な特性に注目しておきたい。かかる特性が組み合わさることで、機械学習技術は、他の統計的技術と比べて、質的に人間から独立したものであるかのように映ることや、機械学習が他の分析技術と区別され、行政機関の機械学習への依存度の高まりに対する潜在的な懸念を生じさせることが指摘されている⁽⁸⁾。

①自己学習

予測に際して各変数がどのように考慮されるかについて、人間による事前の指定に依拠するものではなく、機械学習アルゴリズムが効果的に自らパターンを探し出す。

②ブラックボックス

機械学習の分析結果は、直感的に説明可能なものではなく、伝統的に政府の行為を正当化するために提示されてきた理由の根底にあるような因果関係の説明を支援しうるものではない。

③高速化・自動化

機械学習は、高速かつ自動的なものであり、人間の熟議や意思決定を短縮したり、潜在的に回避したりできる結果をアルゴリズムが生成することに貢献しうる。

(6) See Luke Hutton & Tristan Henderson, *Beyond the EULA: Improving Consent for Data Mining*, in *TRANSPARENT DATA MINING FOR BIG AND SMALL DATA* 147, 149 (Tania Cerquitelli et al. eds., 2017).

(7) 税務当局に蓄積されている膨大なデータが十分に活用されていない可能性があり、データマイニングにより、タックスコンプライアンスの向上を図ることができるのではないかとという問題意識を示すものとして、Jani Martikainen, *Data Mining in Tax Administration: Using Analytics to Enhance Tax Compliance*, AALTO UNIV. SCHOOL OF ECONOMICS MASTER'S THESIS (2012), https://aaltodoc.aalto.fi/bitstream/handle/123456789/7398/hse_thesis_13054.pdf?sequence=1&isAllowed=y 参照。データが十分に活用されていなかった一例を示すものとして、VIKTORIA WÖHRER, *DATA PROTECTION AND TAXPAYERS' RIGHTS: CHALLENGES CREATED BY AUTOMATIC EXCHANGE OF INFORMATION* 377-378 (2018) 参照。

(8) See Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 *Geo. L. J.* 1147, 1167 (2017).

機械学習アルゴリズムの利用に対して注目される法的問題については後で具体的に考察を行うが、上記のような機械学習の特性を概観するだけでも、行政機関が機械学習による意思決定に大きく依存した場合に起こる問題を意識させられる。極論をいえば、①と②により、その意思決定の過程が不透明なものとなり、確たる理由・根拠が明らかではないまま、③により、高速かつ自動的に意思決定されていくのではないかという、漠とした不安を覚える⁽⁹⁾。

③に関して、ここでは、人間による監視が不十分なまま、税務処理を自動的に進めた場合に、質的にも量的にも重要な問題を引き起こす可能性があることを示す例（後述する自動化バイアスの例も参照）として、オーストラリアのRobodebt問題が想起される。オーストラリアでは、福祉給付金の過払いを防ぐために、AIを搭載したシステムを採用し、従前と異なり、人間の監視が不十分な状態になっていた。システムは、過払額を計算する際に、担当する社会福祉機関 Centrelink 側のデータ（受給者が申告したもの）とオーストラリア国税庁が保有するデータ（所得を平均化したデータ）をマッチングして所得確認を行い、ミスマッチがあると判断した場合には、受給者に対してオンラインアカウントにログインして所得情報の更新を促す通知を発行していた。そして、受給者が証明をしない（できない）場合には、債務を自動的に計算し、通知していた。しかしながら、収入金額のマッチングの際に期間的なズレが生じたり、1つの雇用主に雇われている場合でも両機関で登録されているその雇用主の名前が異なると2つの異なる仕事をしていると判断されたりするなどの欠陥があり、集団訴訟が提起され、結果的に、連邦政府は40万件、約12億ドルを返還することとなった⁽¹⁰⁾。

(9) かような不透明性と説明不可能性といった機械学習に対する最も直感的な不安を感じさせる弊害には、データ収集や変数の指定がなされるのではなく、アルゴリズムが選択され、開発されるときにもたらされるものもあるという指摘として、David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C.D. L. REV. 653, 668 (2017) 参照。

(10) See Jordan Hayne, *Robodebt Refunds Top \$220 Million as Social Services Boss Rejects Suicide Claims*, ABC NEWS (July 31, 2020), <https://www.abc.net.au/news/2020-07-31/robodebt-refunds-top-220-million-centrelink/12512310>; *Robodebt Class Action: Coalition Agrees to Pay \$1.2bn to Settle Lawsuit*, THE GUARDIAN (Nov. 6, 2020), <https://www.theguardian.com/australia-news/2020/nov/16/robodebt-class-action-coalition-agrees-to-pay-12-bn-to-settle-lawsuit>; Community Affairs References Committee, *Design, Scope, Cost-benefit Analysis, Contracts Awarded and Implementation Associated with the Better Management of the Social Welfare System Initiative* (June 21, 2017), https://www.apf.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/SocialWelfareSystem/Report; Community Affairs References Committee, *Centrelink's Compliance Program Second Interim Report* (2020), https://parlinfo.apf.gov.au/parlInfo/download/committees/reportsen/024338/toc_pdf/Centrelink'scomplianceprogram.pdf;fileType=application%2Fpdf; Legal Aid Queensland, *Have You Received a Debt Notice from Centrelink?* (2017), <https://www.legalaid.qld.gov.au/Find-legal-information/Factsheets-and-guides/Factsheets/Have-you-received-a-debt-notice-from-Centrelink>; Isabella Alberti, *The Double Side of Artificial Intelligence in the Public Sector*, 8 ACTA UNIV. SAPIENTIAE: LEGAL STUD. 151, 156-157 (2019); GORDON LEGAL HP, <https://gordonlegal.com.au/robodebt-class-action/robodebt-settlement-faqs/>. 石村耕治「AI 税務と税務専門職の将来像を展望する 第6回—税務のスマート化とタックスプライバシー—」税務事例 51 巻 8 号 68～69 頁以下 (2019) も参照。

Ⅲ AI ないし機械学習アルゴリズムを利用する場合に想定される法的問題

行政や司法が AI や機械学習を利用する場合の法的問題に関する研究が比較的進んでいるアメリカにおける議論を主な素材として、考察の手掛かりを得てみたい。

1 再犯リスク判定・予測的警察活動

特に議論の的になるのが、公判前釈放や仮釈放等の決定時に考慮される再犯リスクの判定や、予測的警察活動（predictive policing）の問題である⁽¹¹⁾。後者について、アメリカでは、定量化できる個人的特徴に基づいて犯罪傾向のある人物を特定することを目的とした統計モデルが、再犯リスクの判定の場面等において影響力をもつようになってきており、警察もいっどこで犯罪が最も起こりやすいか（ホットスポットが発生しやすい場所はどこか）、どのような個人が犯罪を犯す可能性があるかを予測するために機械学習や統計モデルを利用している⁽¹²⁾。

予測的警察活動とは、将来を見据えた犯罪予防のために、情報を利用し、高度な分析を開発・利用するあらゆる警察活動の戦略や戦術である⁽¹³⁾。統計的な予測によって、警察の介入が必要となりそうな対象を特定し、犯罪を防止し、あるいは過去の犯罪事件を解決するために、分析技術、とりわけ定量的技術を応用する。犯罪レベルを予測するための統計的・地理空間的分析は何十年も前から行われてきたものの、近年、非常に規模の大きいデータセットを利用して犯罪防止のための予測を行う分析ツールへの関心が急激に高まっている⁽¹⁴⁾。もっとも、予測的警察活動は比較的初期の段階にあって、その有効性についての証拠が混在しているといわれる⁽¹⁵⁾。犯罪活動に従事している可能性の高い個人を特定するために、データの裏付けにより、それ自体は無実の行為等と犯罪行為の間の相関関係を特定し、警察が犯罪を防止するのに役立つような自動化されたプログラムが組まれる。既存のデータを分析し、犯罪行為と相関関係のある要因を特定し、各要因の重みや他の要因との関係を分析し、統計的・確率的な予測のためのアルゴリズムを生成した後、個人と

(11) 予測的警察活動の（憲）法的問題について、差し当たり、山本龍彦「予測的ポリシングと憲法—警察によるビッグデータ利用とデータマイニング—」慶應法学 31 号 321 頁以下（2015）参照。

(12) See Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 875 (2016); Lehr & Ohm, *supra* note (9), at 658. 予測的警察活動に関する日本の状況については、次の資料等を参照。犯罪・交通事象・警備事象の予測における ICT 活用の在り方に関する有識者研究会「犯罪・交通事象・警備事象の予測における ICT 活用の在り方に関する提言書」（2018）。https://www.keishicho.metro.tokyo.jp/kurashi/anzen/anshin/ict_teigen.html、警察庁「令和 2 年 警察白書」85 頁（2020）参照。京都府警察における導入について、国立国会図書館調査及び立法考査局「科学技術に関する調査プロジェクト 2017 報告書 人工知能・ロボットと労働・雇用をめぐる視点」75～78 頁〔瀬戸崇志＝江間有沙〕（2018）参照。

(13) See Craig D. Uchida, *A National Discussion on Predictive Policing: Defining Our Terms and Mapping Successful Implementation Strategies*, NATIONAL INSTITUTE OF JUSTICE DISCUSSION PAPER NCJ230404, 1 (2009), <https://www.ojp.gov/pdffiles1/nij/grants/230404.pdf>.

(14) See WALTER L. PERRY ET AL., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS 1-2 (2013), https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf.

(15) See Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 114 (2017).

その行動等に関するデータを入力し、当該個人が犯罪行為に関与している可能性（予測）を出力する⁽¹⁶⁾。

ただし、データマイニングやアルゴリズムを利用することに関しては、次のような懸念ないし問題点が示されている⁽¹⁷⁾（実際、予測的警察活動技術の使用を禁止する自治体も出現している⁽¹⁸⁾）。

- ・細心の注意を払わずにデータマイニングを行った場合、既存の差別パターンを再現し、これまでの意思決定者の偏見を継承し、あるいは単純に社会に根付いている広範囲の偏見を反映するものとなりうる。歴史的に不利な立場に置かれてきた集団が実際に不利な待遇を受けるに値することを示唆し、このことが既存の不平等を悪化させるという歪んだ結果をもたらすことさえある。
- ・アルゴリズムは、手動でプログラムされていなくても、意図的であれ偶然であれ、かような差別的な傾向を示す可能性がある。差別は、プログラマーが特定の要因に不適切な重みを割り当てた結果ではなく、データマイニングプロセス自体の所産物である場合がある。
- ・問題となっている差別は意図的ではないため、データマイニングのプロセスの関与者側に偏見がないことを証明しようとする誠実な試みであっても、結果として得られる決定に公平性の承認を不当に与えてしまう可能性がある。
- ・データマイニングが法律等によって保護されている、保護を受ける資格があるとされる集団を不利に扱うメカニズムは、意図しない差別の場合にはあまり明らかなものでないため、不公正を識別して対処することが困難になる可能性がある。

ここで指摘されているような差別的ないし不公正な結果の具体例を挙げておく。

再犯リスクを評価するスコアについて、黒人の再犯リスクを過大評価し、白人の再犯リスクを過小評価する例がよく引き合いに出される。中でも、非営利の報道機関であるProPublicaによる調査結果の記事が有名である。この記事は、公表後、1年半で200以上の学術論文によって引用されたほどである⁽¹⁹⁾。ProPublicaは、公判前釈放等の場面で利用される全米で最もポピュラーな再犯予測アルゴリズムの1つであるNorthpointe社の再犯リスク判定アルゴリズムCOMPAS（Correctional Offender Management Profiling for Alternative Sanctions）を調査したところ、次の点が明らかになったとしている⁽²⁰⁾。

(16) See Rich, *supra* note (12), at 876.

(17) See Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674 (2016).

(18) See Keith Burbank, *Oakland 1st to Ban Predictive Policing, Biometric Surveillance Tech*, KTVU Fox2 (Jan. 15, 2021), <https://www.ktvu.com/news/oakland-bans-predictive-policing-biometric-surveillance-tech>; Nathan Sheard, *Oakland's Progressive Fight to Protect Residents from Government Surveillance*, EFF (Jan. 20, 2021), <https://www.eff.org/ja/deeplinks/2021/01/oaklands-progressive-fight-protect-residents-government-surveillance>.

(19) See Anne L. Washington, *How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate*, 17 COLO. TECH. L. J. 131, 135 (2018).

- ・白人よりも黒人の被告人を将来の犯罪者として誤ってフラグを立てる可能性が高い。黒人の被告人は、白人の被告人と比較して、高リスク者すなわち再犯リスクの高い者として誤って分類される可能性が2倍近くになる（黒人：45%，白人：23%）。
- ・しかしながら、高リスクと判断された黒人の被告人と白人の被告人の差はわずかなものであった（黒人：63%，白人：59%）。
- ・上記の格差は、被告人の前科や逮捕された犯罪の種類によって説明できるものではなく、人種の影響を犯罪歴や再犯率、被告人の年齢や性別から分離した統計的検定を実施したところ、やはり黒人の被告人の方が白人の被告人よりも45%程度、高いリスクスコアが与えられる可能性があった。

なお、銃撃事件に関連した逃亡未遂罪等で有罪となり、懲役6年、拡張監督（extended supervision）5年の刑を言い渡された被告人が、巡回裁判所が量刑判断の際にCOMPASリスク評価を考慮したことに対して、デュープロセスの権利を侵害している（COMPASの機密性により被告人が正当な情報に基づいて量刑判断を受ける権利を侵害し、個別的な量刑判断を受ける権利を侵害し、量刑判断においてジェンダー評価を不当に考慮している）などと主張した事件において、ウィスコンシン州最高裁（State v. Loomis, 371 Wis.2d 235 (2016)）は、制限や注意点を遵守して、適切に利用されているのであれば、巡回裁判所が量刑判断の際にCOMPASのリスク評価を考慮したことは被告人のデュープロセスの権利を侵害するものではない、と判示した⁽²¹⁾。ここでは詳述はできないが、この判決は、アルゴリズムの利用に対する法的問題を検討する上で多くの素材を提供する一方、法的問題に関する不安を必ずしも払拭しうるものではない。実際に種々の観点から、この判決に対して批判や懸念が示されている⁽²²⁾。

予測的警察活動の例として、オークランド警察から入手した薬物犯罪の検挙者数のデータに基づく調査がある。まず、警察のデータセットにどのくらいバイアスがあるかという点について、薬物犯罪は市全体にわたって均等に分布している一方で、薬物関連の検挙者は特定の場所、すなわち非白人や低所得者の人口が多い地域に集中しており、これらの地域では、他の地域と比較して約200倍の検挙者数が発生していることを指摘している。次に、警察の記録データが予測的警察活動モデルに与える影響を調べるために、米国で最大規模の予測的警察活動システムのベンダーであるPredPol社のアルゴリズム（過去の犯罪の種類、犯行場所、犯行時刻の3つのデータのみを用いて予測を行うもので、race-neutralなシステムであると開発者が説明するもの）をオークランドにおける薬物犯罪記録に適用し、その結果を検討している。結論として、アルゴリズムは、警察のデータに存

(20) See Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. 上記の調査結果を踏まえても、何をもって公正と考えるか、意思決定や政策判断にどのように取り入れるべきかなど、議論の余地は残されている。参考として、例えば、Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments* (2017), <https://arxiv.org/abs/1703.00056>; Virginia Foggo et al., *Algorithms and Fairness*, 17 OHIO ST. TECH. L. J. 123 (2021) 参照。

する明らかなバイアスを修正するのではなく、歴史的に過剰に取締りが行われてきたコミュニティに対して、ますます不均衡な取締りをもたらすものであり、むしろバイアスを強化していることが判明した。すなわち、オークランドで上記アルゴリズムを使用すると、黒人は白人の約2倍の割合で、白人や黒人以外の人種は白人の1.5倍の割合で取締りの対象となること（しかも、所得層別の割合を分析しても同様の結果となること）が明らかとなった⁽²³⁾。

予測的警察活動における機械学習アルゴリズムの利用に関して、商業的データと社会的データを参照して、個人の脅威レベルを決定しようとする時、肌の色が黒いことと、脅威レベルが高いこと又は特定の犯罪を犯した疑いが強いことを不当に結び付けることになりかねないし、過去のデータに基づいて犯罪マッピングを行うと、主に有色人種が居住する

(21) 最高裁は、要旨次のとおり判示した。

COMPAS のリスクスコアは、投獄するかどうか、どの程度の量刑とするかを判断する際に利用してはならないし、犯罪者が、地域社会で安全かつ効果的に監督できるかどうかを判断する際の決定要因としても利用してはならない。また、COMPAS リスク評価は、判断の際に考慮される数多く存在する要因の1つにすぎない（巡回裁判所は、COMPAS リスク評価に加えて、その判断を独立的に裏付ける要因を説明しなければならない）。

さらに、裁判所に提出される COMPAS リスクアセスメントを含む判決前調査報告書（Presentence Investigation Report）には、次の注意事項を書面で記載しなければならない。

- ① COMPAS の機密的性質（proprietary nature）からして、判断要素の重み付けがどのように決定されるか、リスクスコアがどのように決定されるかという点に関する情報は開示されないこと
- ② リスク評価では、被告人を全米のサンプルと比較しており、ウィスコンシン州の住民を対象としたクロスバリデーションは未了であること
- ③ COMPAS のリスク評価スコアに関する研究の中には、マイノリティの犯罪者を再犯リスクが高いと偏って分類しているのではないかという疑問を提起するものがあること
- ④ COMPAS のリスク評価スコアは、集団のデータに基づいているため、特定の高リスクの個人ではなく、高リスクの犯罪者の集団を識別することができるものであること
- ⑤ COMPAS は、量刑判断の際に利用するためではなく、矯正局が治療、監督、仮釈放に関する決定を行う際に利用することを目的として開発されたものであること

(22) See e.g., Katherine Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, 18 N.C. J. L. & TECH. 75 (2016); *Criminal Law - Sentencing Guidelines - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing - State v. Loomis* 881 N.W.2d 749 (Wis. 2016), 130 HARV. L. REV. 1530 (2017); Ellora Israni, *Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis*, HARV. J. L. & TECH JOLT DIGEST (Aug. 31, 2017), <https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1>; Han-Wei Liu et al., *Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability*, 27 INT'L J.L. & INFO. TECH. 122 (2019); Leah Wissner, *Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing*, 56 AM. CRIM. L. REV. 1811 (2019). この判決を取り上げる邦語文献として、緑大輔「判批」判時 2343号 128頁以下（2017）、山本龍彦「AIと『個人の尊重』」福田雅樹ほか編『AIがつなげる社会』331～332頁以下（弘文堂2017）、同「ロボット・AIは人間の尊厳を奪うか？」弥永真生＝宍戸常寿編『ロボット・AIと法』89～90頁（有斐閣2018）、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頁以下（2018）、柳瀬昇「AIと裁判」山本龍彦編著『AIと憲法』365～367頁（日本経済新聞出版社2018）参照。

(23) Kristian Lum & William Isaac, *To Predict and Serve?*, 13 (5) SIGNIFICANCE 15 (2016), <https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x>.

地域における nuisance crimes の逮捕者数の増加につながりうる上、予測的警察活動がどの程度、差別的な結果をもたらすのかは、一般市民や警察自身にはわからない、という指摘もある⁽²⁴⁾。

2 AI ないし機械学習アルゴリズムを利用する場合に想定される法的問題

前記 1 の議論との関連で指摘されることが多い AI ないし機械学習アルゴリズムを利用する場合に想定される法的問題について整理する。

(1) セグメントに基づく統計的・確率的予測

AI ないし機械学習アルゴリズムは完璧なものではなく、誤った予測を行うこともある。機械学習アルゴリズムは、データの範囲内でパターンと相関関係を利用し、統計的・確率的な予測を示すにすぎないのであるから当然である。

また、AI ないし機械学習アルゴリズムは、個人を、共通の属性をもった集団（セグメント）に細分化し、個々人の具体的な差異を捨象し、統計的・確率的な予測や評価をはじめ出すという側面を有する⁽²⁵⁾。AI による予測評価は、（これを個人の能力等に関する最終的な評価としない、セグメントをかなり微細化することで「個人」の実態に近付くといった留保を付しうることを念頭に）その方法次第では、前近代を否定して個人一人ひとりの事情をいいわば時間とコストをかけて一具体的に考慮することを要請するに至った近代憲法の原理規範すなわち個人の尊重原理と矛盾することが指摘されている⁽²⁶⁾。特に、予測や評価により何らかの不利益を被る場面においては、予測や評価の対象となった本人は、せめて、自分が属しているとされる集団ではなく、個人としての自分ないし自分自身をよく見てほしい、あるいは個人を十分に尊重してほしい、その上で予測や評価を行ってほしいと思うであろう。以上のことは、後述するバイアス問題やブラックリスト問題とも関連するものであり、プロファイリング⁽²⁷⁾がもたらす問題として捉えることもできる。

いずれにしても、予測が結果的に外れることがあることは当然であり、予測への行き過ぎた依存や個々人の具体的な差異を完全に無視するかのような対応は妥当でない。学習用データに誤りがあれば予測の精度は低くなる一方、学習用データの質と量が確保されてい

(24) Selbst, *supra* note (15), at 109.

(25) 山本・前掲注 (22) 「ロボット・AI は人間の尊厳を奪うか？」 87 頁参照。予測的警察活動の文脈でいえば、個人が示す集団的な特徴と結び付いて個人レベルではなく集団レベルでの疑惑に着眼している、認知的なショートカットが行われている、かようなバイアスがかがかかっていると捉えることもできる。See ANDREW G. FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT 128 (2017). 翻訳版として、アンドリュウ・ガスリー・ファガソン [大槻敦子訳] 『監視大国アメリカ』 199～201 頁 (原書房 2018) 参照。

(26) 山本・前掲注 (22) 「ロボット・AI は人間の尊厳を奪うか？」 88～93 頁参照。

(27) プロファイリングの定義については、EU の General Data Protection Regulation (一般データ保護規則。以下「GDPR」という) 4 条 (4) は、「自然人と関連する一定の個人的側面を評価するための、特に、当該自然人の業務遂行能力、経済状態、健康、個人的嗜好、興味関心、信頼性、行動、位置及び移動に関する側面を分析又は予測するための、個人データの利用によって構成される、あらゆる形式の、個人データの自動的な取扱いを意味する。」としている。本稿における GDPR の邦訳は、個人情報保護委員会による仮訳を参考としている。https://www.ppc.go.jp/enforcement/infoprovision/laws/GDPR/.

れば予測の精度は高まるが、いずれにせよ、次のバイアス問題やブラックボックス問題への配慮を要する。人間の権利利益に関わる場面では特に注意が必要であろう。

(2) 学習用データとバイアスの問題

「過少申告をした者」, 「犯罪を犯した者」, 「再犯をした者」などといったように、正解のラベルを付けた学習用データが用いられる機械学習を教師あり学習という⁽²⁸⁾(ただし、ここでの叙述は、実際に国税庁が教師なし学習ではなく、教師あり学習の手法のみを用いることまでを含意するものではない)。その際、大量のデータがあれば十分であるというわけではなく、データの質にも注意を要する。機械学習アルゴリズムの性能やパフォーマンス等は、学習のためにセットされる学習用データの量のみならず質に依存する。

補足すると、結果に影響を与える重要なデータがセットされていなければ性能が悪くなる可能性があるし、①データにバイアスが存在する場合と②母集団からのサンプリングに問題がある場合には差別的な影響を引き継いだアルゴリズムができあがる可能性もある(もちろん、入力変数の選択(特徴選択)の際に意識的又は無意識的に⁽²⁹⁾、差別が組み込まれることもありうる)。①について、収集したデータが既存の社会的バイアスを反映している場合に、バイアスが生じる可能性がある。上記のオークランド警察の例のように、人種グループ間の検挙率の違いが再犯リスクを計算するアルゴリズムによって再現されてしまう。COMPASの例も然りである。その原因について、アルゴリズムそれ自体に差別を生成するような装置が組み込まれていなかったとしても、基礎となる学習用データ、あるいはそれを生み出した社会や制度等から引き継いで再現してしまう可能性があるというのであるから、問題の根は深い。学習用データにバイアスが存在するか、そのバイアスは強度のものか、バイアスと入力変数はどのような関係にあるのかといった分析が必要となる。②について、データマイニングが母集団の偏ったサンプルから推論を導き出す場合や学習用データが実世界の環境を代表していない場合において、これらの推論に基づく決定は、パフォーマンスが低下しうるほか、データセットに十分に含まれていない集団や過剰に含まれている集団に不利益を与える可能性がある。例えば、人間の皮膚科医を凌駕する精度で写真から皮膚がんを発見するシステムについて、アルゴリズムが主に白い肌の写真を使用して訓練されたため、肌の色が濃い患者ではパフォーマンスが低下する可能性があるという⁽³⁰⁾。

しかも、上記のようなバイアスの再現にとどまらず、さらに事態を悪化させる可能性がある。例えば、予測的警察活動について、警察が戦術的な取締りの決定を下すために偏りのある予測を行っていることによって、初期の学習用データのバイアスがさらに悪化する

(28) 総務省・前掲注(4) 86頁参照。

(29) ここでは、人間による特徴選択や人間による学習用データの選択を想定しておく。

(30) See Rich, *supra* note (12), at 884-885; Barocas & Selbst, *supra* note (17), at 674; Sandra G. Mayson, *Bias in, Bias out*, 128 YALE L. J. 2218 (2019); Miriam C. Buiten, *Towards Intelligent Regulation of Artificial Intelligence*, 10 EUR. J. RISK REGUL. 41, 50-53 (2019), <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/towards-intelligent-regulation-of-artificial-intelligence/AF1AD1940B70DB88D2B24202EE933F1B>; Yavar Bathaee, *Artificial Intelligence Opinion Liability*, 35 BERKELEY TECH. L. J. 113, 161 (2020).

可能性がある。かような予測は警察が既に認識している地域を過剰に反映している可能性が高い。このため、警察官は同じ地域をパトロールし、犯罪行為の分布に関する事前の確信を裏付けるような新たな犯罪行為をそこで観察する可能性が高くなる。かようなターゲットを絞ったパトロールの結果がアルゴリズムにフィードバックされ、ますます偏った予測が生み出される。かように、選択バイアスと確証バイアスが重なっていくというフィードバックループが生じる可能性がある⁽³¹⁾。

アルゴリズムないしビッグデータのデータマイニングについて、逆に人間による意思決定がもたらす差別を防ぐために利用しうる面があるとしても⁽³²⁾、パフォーマンスの低下にとどまらずに、不当な差別や偏見をさらに悪化させる可能性があることは等閑視できない。

(3) ブラックボックス問題

予測の過程に着眼して、アルゴリズムによる予測の根拠や理由が不明瞭であるというブラックボックス問題もある。例えば、機械学習の手法の1つに近時の AI ブームのきっかけとなったともいわれる深層学習（ディープラーニング）がある。深層学習とは、多数の層から成るニューラルネットワークを用いて行う機械学習のことである。深層学習という概念は、あくまでもこの多層的なニューラルネットワークに着目したものであるが、深層学習により、コンピューターがパターンやルールを発見する上で何（どのような入力変数（特徴量、説明変数））に着目するかを自ら抽出することが可能となり、何に着目するかをあらかじめ人間が設定していない場合でも識別等が可能になったとされる⁽³³⁾。他方で、入力変数が多種多量であったり、ニューラルネットワークが多数の層から構成され、相互に影響し合ったりしている場合には、アルゴリズムがなぜそのような予測をはじめ出したのか、どのような変数をどの程度重視しているのか、変数同士の関係はどのようなものかなど、その根拠や理由を人間が完全には理解できないという事態が起りうる。複雑性や不可視性等によって、決定過程が見えづらい、判断理由が説明できない、理解できないという意味でのブラックボックス性である。

機械学習アルゴリズムが決定プロセスの一部として把握しているかもしれないパターンについて、膨大な数の相互に接続された層の中を透明化することや高次元のパターンを可視化することが困難であったりすることにより、人間が簡単に把握する明確な方法を見いだし難いといわれる⁽³⁴⁾。適当なモデルを選択したり、技術的に工夫したりすることで改善の余地がある一方⁽³⁵⁾、かような意味でのブラックボックス性に対する懸念の声も根強

(31) See Lum & Isaac, *supra* note (23), at 16.

(32) See FTC, *Big Data: A Tool for Inclusion or Exclusion?*, FTC REPORT 25-32 (2016), <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

(33) 総務省・前掲注 (4) 83 頁参照。

(34) See Bathaee, *supra* note (30), at 141-142. 例えば, Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J. L. & TECH. 889 (2018) は、ブラックボックスの理由として、ディープニューラルネットワークの複雑性とサポートベクターマシンの次元性を挙げる。

(35) See Lehr & Ohm, *supra* note (9), at 703-705.

い。人間が理解できるように、より単純なモデルを利用すると予測の精度が落ちかねないというトレードオフの問題もある。

さらにいえば、機械学習アルゴリズムが、過去のデータに基づいて相関関係を明らかにしたとしても、因果関係までを明らかにできるかは別問題である。過去のデータに基づく相関関係を導き出すアプローチから因果関係までも明らかにするのは困難であろう。かような意味でのブラックボックス性の問題もある（前記Ⅱ2参照）。

十分な規模のデータセットがあれば、一般に、意味のない相関関係を発見しうる。例えば、アメリカ大統領選挙において、ワシントンD.C.のプロフットボールチームが選挙前の最後のホームゲームで勝つと現職の政党が勝利し、負けると現職の政党が敗北するといったものである。このような例（疑似相関と呼ばれることもある）は数多く存在する⁽³⁶⁾。結局、発見された相関関係の取扱いには一定の注意が必要である。

ただし、ブラックボックス問題については、Explainable AI (XAI)⁽³⁷⁾やCasual AI⁽³⁸⁾と呼ばれる研究が進められており、今後の技術の発展により解決につながる可能性もある。

Ⅳ AIないし機械学習アルゴリズムを実装した調査選定システムを利用する場合に想定される法的問題

これまでの議論から得られる示唆を踏まえて、AIないし機械学習アルゴリズムを実装した調査選定システムを利用する場合に想定される法的問題を考察する。

1 IRSの調査選定システム

前記Ⅲでは、行政や司法がAIないし機械学習を利用する場合の法的問題に関するアメリカの研究を中心にみてきた。このこととの関連では、アメリカにおいては、Internal Revenue Service（アメリカの内国歳入庁。以下「IRS」という）の調査選定システムに対してどのような議論がなされているのかという点に関心が寄せられる。この点について、IRSは早くから納税者のコンプライアンスリスクの判定のためにAIを利用する研究を進めていたところ⁽³⁹⁾、予測的警察活動と、IRSがビッグデータを利用してどの納税者がコンプライアンスに違反しているかを予測することとの類似性が指摘されている⁽⁴⁰⁾。

(36) See FTC, *supra* note (32), at 9.

(37) See OECD, ARTIFICIAL INTELLIGENCE IN SOCIETY 91-95 (2019); Matt Turek, *Explainable Artificial Intelligence (XAI)*, DARPA, <https://www.darpa.mil/program/explainable-artificial-intelligence>. 原聡「機械学習における解釈性」人工知能 33巻5号366頁以下、同「説明可能AI」人工知能 34巻4号577頁以下、古川直裕「機械学習システムの法務・コンプライアンスリスク（下）」NBL1150号39頁以下（2019）参照。また、富士通研究所＝北海道大学「望む結果までの手順を導くことができる『説明可能なAI』を世界で初めて開発」（2021年2月24日）、<https://pr.fujitsu.com/jp/news/2021/02/4.html>も参照。

(38) See Sema K. Sgaier et al., *The Case for Causal AI*, 18 (3) STANF. SOC. INNOV. REV. 50 (2020), https://ssir.org/articles/entry/the_case_for_causal_ai.

(39) See Jeff Butler, *Analytical Challenges in Modern Tax Administration: A Brief History of Analytics at the IRS*, 16 OHIO ST. TECH. L. J. 258 (2020).

(40) See Kimberly A. Houser & Debra Sanders, *The Use of Big Data Analytics by the IRS: Efficient Solutions or the End of Privacy as We Know It*, 19 VAND. J. ENT. & TECH. L. 817, 855 (2017).

また、IRS が調査選定にデータ分析を使用することで、重大なコンプライアンスリスクが想定される申告書のみが選定され、調査範囲もリスクが高い問題に限定され、コンプライアンスを遵守している納税者にとって有益なものとなる可能性があることが認識されている一方⁽⁴¹⁾、次のような懸念も示されている。

- ・公開情報の信頼性の問題が IRS による誤認選定につながる可能性がある。本人がネットにアップした情報は自己のイメージの向上のために、誇張又は本当の姿を偽装したりすることが多いため、特に疑わしいという問題（airbrushing 問題）がある。本人以外が本人に関する情報を投稿できる Facebook について、本人が自身の情報を完全にはコントロールできないという問題もある。誤情報が本人の知らないうちにネット上に拡散される可能性がある。納税者によるものでも、他人によるものでも、誤った情報がコンプライアンス違反の特性を帯びる傾向がある場合には、調査対象として選定される可能性が高くなる。かような誇張された又は偏った情報が蔓延しているものとする、IRS の分析にかような情報が取り込まれた場合に誤認選定の可能性が出てくる⁽⁴²⁾。
- ・結果はアルゴリズムに左右されるのであり、特定の納税者や集団を不釣り合いに標的にする無意識のバイアスがアルゴリズムに組み込まれていないかという問題がある⁽⁴³⁾。マーケティングを行う企業は、ビッグデータ分析の精度が低くてもなお価値を見いだすかもしれないが、IRS はそのデータの信憑性に関して、より厳格な基準を求められる⁽⁴⁴⁾。ビッグデータ分析の大きな利点は、不注意の結果として、保護された集団の差別につながる場合には影を潜める。例えば、申告していない多額の現金収入のある個人は、新築住宅の購入代金を現金で支払う可能性が高いとする。IRS のアルゴリズムが、住宅ローン控除を受けていない新築住宅の取得と、申告されていない現金収入との間に、かような相関関係があると識別する。しかし、特定の宗教の信者は、借入金の利息を支払うことが禁止されているため、ローンを組まずに現金で購入することもある。このように、アルゴリズムが識別した相関関係は、意図的であるか否かにかかわらず、特定の宗教の信者について、コンプライアンスを遵守していない納税者としてプロファイル化しており、ひいては差別につながるものである⁽⁴⁵⁾。
- ・IRS は、調査選定に関する一般的な情報を進んで提供しているが、調査選定に用いられている DIF（Discriminant Inventory Function System）アルゴリズム、ビッグデータ、又は予測分析アルゴリズムが調査プロセスでどのように利用されているかという点につ

(41) See Carina C. Federico & David B. Blair, *Automation and Data Analytics to Drive LB&I Audit Selection*, BLOOMBERG TAX (June 5, 2019), <https://news.bloombergtax.com/daily-tax-report/insight-automation-and-data-analytics-to-drive-lb-i-audit-selection>.

(42) See Houser & Sanders, *supra* note (40), at 847-848; Kimberly A. Houser & Debra Sanders, *The Use of Big Data Analytics by the IRS: What Tax Practitioners Need to Know*, 128 (2) JOURNAL OF TAXATION 1,13-14 (2018), <https://ssrn.com/abstract=3120741>.

(43) See Federico & Blair, *supra* note (41).

(44) See Treasury and General Government Appropriations Act for 2001 (Pub.L. 106-554) § 515; 44 USC 3504 (d) (1) and 3516; Office of Management and Budget, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies* (Oct. 1, 2001).

(45) See Houser & Sanders, *supra* note (42), at 13-14.

いての詳細は提供しない。申告書の調査選定方法の詳細を開示することはなく、納税者からの情報公開請求も成功しないという透明性の問題もある（納税申告書のどの項目が不正を行っている兆候として取り扱われているかが納税者に正確に知られてしまうと、不正を行う納税者が行動を調整できてしまい、その兆候が課税庁にとって価値を失うという問題があるため、システムの完全な透明性を確保するのは困難である）⁽⁴⁶⁾。

- ・IRSが納税者の同意なしに納税者に関する電子情報を利用する場合、国民が、収集した情報を確認したり、調査選定の判断材料としてIRSの使用する情報に存在する誤りを修正したりすることができないことは問題である⁽⁴⁷⁾。

これらの懸念は、正確とはいえない情報に基づく誤認選定、バイアス、（前記Ⅲ2(3)のブラックボックス問題が技術的要因に基づくものであるのに対して）守秘義務や企業秘密等により情報公開請求が阻まれるという法制度的要因に基づくブラックボックス、あるいは課税庁の保有情報に対して納税者がアクセスし、訂正を行うような権利を含む自己情報コントロール権⁽⁴⁸⁾の欠如といった問題として、捉えうる。

(46) See Federico & Blair, *supra* note (41); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 658 (2017). 納税者からの情報公開請求に関して、Gillin v. IRS, 980 F.2d 819 (1st Cir. 1992) が参考となる。この事件の概要は次のとおりである。

Gillin氏は、IRSのJacksonville事務所の調査官が、自身がAndover Service Centerに提出した1985年分の連邦所得税申告書入手して調査していることを知っていた。IRSは同氏に法定の納税額不足通知を発行するための期間制限の延長に同意をするよう求めたが、同氏が拒絶したところ、IRSは申告是認を決定した（不足通知発行の理由なしと結論付けた）。その後、Gillin氏が、IRSに対して、FOIA（情報自由法）552条に基づき5種類の文書を公開するよう請求した。

これに対して、IRSは請求された①同氏の個人マスターファイルを公開し、②リーエン関連ファイルと③犯罪捜査部の関連ファイルは探したが彼に関するものはないとした。④申告書に不備があり、期限延長を要請可能と結論付ける根拠とされたすべての文書の写しについては、そもそも是認だから存在しないと狭義に解し、公開しなかった。⑤IRSの事務所間で「管轄権」を移すためにAndover Service Centerに使用された全文書と手続的規則の写しについては、担当部署の「管轄権」の問題ではないと狭義に解し、公開しなかった。⑤について、連邦地裁は文書の探索等を指示し、IRSは多数の応答文書を提示したが、一部はFOIAによる適用除外を主張し、墨消しの処置を行った。控訴審Gillin v. IRS, 980 F.2d 819は地裁の判断を支持し、墨消しの可否のみを判断し、次のとおり判示した。

数学的手法を用いて調査が必要な確定申告書を特定するDIFスコアリングについて、「開示された証拠によると、墨消しされた情報はすべてDIFスコアで構成されていた。DIFスコアリングは、調査の必要度が最も高い申告書を特定するために用いられる数学的手法である。IRSは、このDIFスコアリングの手法に関する情報を厳重に保護している。これは、その情報を知ることによって、あくどい納税者が申告書を調整してDIFスコアを下げ、調査の可能性を減らすことができるからである。かような情報は、他の法律によって特に開示が免除されているものにはFOIAは適用されない旨を定めるFOIA552条(b)(3)と、『調査のための申告書の選定に使用される又は使用されるべき、あるいはそのような基準を決定するために使用される又は使用されるべきデータ』の開示を免除するIRC6103条(b)(2)に基づいて、FOIAの開示の適用から除外される」(980 F.2d 819, 822)。ただし、IRC6103条(b)(2)(D)には「開示することによって国内歳入法に基づく査定、徴収又は執行を著しく損なうと長官が判断した場合」という条件が付されている。

(47) 1974年プライバシー法への抵触問題も含めて、Houser & Sanders, *supra* note (40), at 836 参照。この問題について、望月爾「デジタル化・グローバル化と納税者権利保護—税務行政のデジタル化の進展とその影響を中心に—」日本租税理論学会編『租税上の先端課題への挑戦』〔租税理論研究叢書30号〕142頁（日本租税理論学会2020）、石村・前掲注(10)63頁以下も参照。

2 AI ないし機械学習アルゴリズムを実装した調査選定システムを利用する場合に想定される法的問題

国税庁が AI ないし機械学習アルゴリズムを実装した調査選定システムを利用する場合に想定される法的問題として、次の点を指摘したい。

(1) 学習用データに関する問題

まず、バイアス問題にもつながりうる学習用データについては、次の点に留意すべきである⁽⁴⁹⁾。

- ・調査の必要性要件（税通 74 の 2 等）の存在などにより、国税庁に蓄積された過去の調査データは、ランダムに選ばれたものではなく、一定の基準により選定された調査に係るものとなる。過去に多額の非違・不正行為が把握された特定の業種や、期間損益計算に関係するもの（いわゆる“期ズレ”）など申告書類や帳簿書類から把握しやすい又は過去に多数把握された特定の非違・不正行為・取引等に偏る可能性がある。逆に、これまで発覚しなかった精巧な不正スキーム、新たな不正スキーム、あるいは新規性の高い取引等に関するデータが不足する。
- ・反社会的勢力、政治家等の権力者、税務調査に非協力の納税者など、種々の事情から調査件数が伸び悩んでいる又は（仮に、各集団に対して個別のノルマ件数を設定し、調査を実施しているとしても）深度ある調査ができていない特定の集団が存在する場合には、そのような集団に属する者に関しては、相対的に、過去の調査で把握したデータが質的及び量的に一般の納税者に対するものに劣る可能性がある。
- ・当局との争いを回避したい、税理士非関与や外国人であり税制に明るくないなど、種々の事情から、調査に協力的な又は国税庁に従順な納税者のデータに偏る可能性がある。
- ・税務調査を実施しても、不正行為や非違を発見できない、見逃してしまう場合があり、学習用データとして適さないものもある。ノルマをすでに達成している又は事務年度末であり事案を翌年度に繰り越したくないなどの理由から、実際には深度ある調査や課税処分が必要であるにもかかわらず、申告是認という拙速な処理を行っているケースもやはり適さないことになろう。
- ・組織風土や調査ノルマのようなものを背景として、（国税通則法 68 条 1 項所定の隠蔽又は仮装行為を行ったことを納税者自らが認めているいわゆる“自認事案”に係る重加算税賦課決定事案などを中心として）課税要件や重加算税の賦課要件を満たしていないにもかかわらず満たしているかのような処理が時折行われてきた事実を直視すべきである。課税処分後に裁判所等で処分取消となった情報をアップデートとするとしても、多くの事案が争訟にまで至らず、司法審査等を受けないことも踏まえると、過去の調査データについて、一律に、「利用可能な最良の代理物（the best available proxy）」⁽⁵⁰⁾である

(48) ただし、本稿では、自己情報コントロール権やプライバシー権の問題については深入りしない。

(49) セットされたデータ自体に誤りや偏りがある場合のほか、質的及び量的に必要なデータを収集していなかったり、手書きや紙ベースの記録しか存在していなかったりするなど、元となるデータに不備や問題があることも想定される。この意味で、国税庁がこれまで蓄積してきたデータを有効活用しようとしても、データの前処理など様々な壁に突き当たることが予想される。

(50) Selbst, *supra* note (15), at 134.

として、そのまま学習用データとしての利用を認めてしまってよいか。

- ・各国間の情報交換制度が整備されつつあることを踏まえても、情報源が、国税庁からみて入手しやすいもの、すなわち国内のもので、かつ、申告書類、法定調書、源泉徴収データ等に偏る可能性がある。

以上のような学習用データに関する問題が引き起こす影響について、次の(2)で考察するもののほか、例えば、調査対象者が国税庁からみて“税金を取りやすい”，すなわち調査に協力的な又は国税庁に従順な納税者集団に偏る一方、国税庁からみて“税金を取りにくい”，すなわち反社会的勢力、政治家等の権力者、税務調査に非協力的な者といった納税者集団は調査対象として選定されにくい、という結果をもたらさう。ともすると、“弱い者イジメ”，“誤認選定”，“特権階級優遇”，“弱腰行政”と揶揄される事態にもなりかねない。

また、過去の調査データに大きく依拠するようなアプローチは、“期ズレ”など古典的な非違の発見に力を発揮する一方で、不正行為や申告誤りへの対策としての限界に突き当たる可能性がある。例えば、新規性の高い業種・業態を営む納税者や新設の法令に関わる税務調査に関する限り、有用なデータが不足しているため、AIによる調査必要度の判定等が困難になる可能性がある⁽⁵¹⁾。直感やひらめきによって、手元にデータの蓄積がなくとも一応の判断を即座に行うことができるという人間の強みを生かした選定等も必要とされる場面があるかもしれない⁽⁵²⁾。差し当たりは、AIによる選定と人間による選定を組み合わせたり、後者の選定に基づく調査割合を確保したりするなどの対策が考えられよう。その意味では、AIの利用を進めつつも、AI任せにせず、熟練職員のノウハウ・経験を活用する、職員教育を充実化する施策を同時並行で実施する必要がある。

(2) バイアス等に関する問題

ア バイアス問題とエコーチェンバー現象

上記のような問題含みの学習用データを使用して生成された調査必要度を判定するアルゴリズムは、既存のバイアスを再現し、ひいては不当な差別や偏見につながる可能性がある。また、これに依拠して調査選定がなされ、バイアスのかかったデータ収集がなされることにより、バイアスが一層強くなる可能性がある。選択バイアスと確認バイアスのループである。学習用データに存在するバイアスが、アルゴリズムを通じて再現されるにとどまらず、延々と繰り返され、エコーチェンバー現象というべきかのように⁽⁵³⁾、バイアス

(51) タックスシェルターの例を挙げつつ、アルゴリズム執行ツールを使用している機関にとっての課題は、新しい形態の不正行為を捕捉するために、これをどのように継続的かつ反復的に更新していくかという点にあることを指摘するものとして、David F. Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REG. 800, 821 (2020) 参照。

(52) このように考えると、差し当たり、AI時代に適応する人材の育成や人間による最終判断を適切に行う態勢作りが肝要であるといえよう。なお、機械とは対照的に、人間は、わずかなデータで驚くべき予測能力を発揮することがあり、例えば、一度か二度しか見たことがない顔を、別の角度からでも認識できることについて、アジェイ・アグラワルほか〔小坂恵理訳〕『予測マシンの世紀』80～81頁（早川書房2019）参照。また、データがほとんどない状況での予測について、同書128頁も参照。

の程度が濃縮されていく可能性もあるのではないか。

例えば、税務調査において、特定の業種に係る事業を営む法人で、かつ、その代表者が特定の国籍を有しているものについて、法人税や消費税の申告に関して不正行為を行っている事例が、複数、把握されたとする。そこで、国税庁は、当該業種に係る事業を営む法人で、かつ、その代表者が当該国籍を有しているものについては、同様の不正行為を行っている可能性が高いとして、類似の法人に対して、全国的に優先的かつ深度ある調査を実施するよう指示を出し、多くの調査日数・人員を注ぎ込み、大量のデータを収集したとする。これによって、税務職員の中で、特定の国籍を有する個人や当該個人が代表者となっている法人は不正行為を行っているというイメージが固まる。このことから、その後も、これらの者は、調査対象として選定される確率が高まり、実際に多くの調査が実施されたとする。これらの調査においては、時として、担当調査官がノルマ的なプレッシャーなどから無理に不正取引の認定を行ったり、調書に多少の脚色を施したりすることもあったとする。

かように、ある種、偏りのあるデータが蓄積されて学習用データとしてセットされると、場合によっては、調査必要度の判定、コンプライアンスリスクの判定の際に個人や代表者の国籍を重視する、特定の国籍にフラグを立てて、重要視するようなアルゴリズムが生成される。当該アルゴリズムによる選定結果は、当然ながら調査官の経験とよく適合する。よって、調査官は当該アルゴリズムに共感し、これに従って調査選定し、調査を実施する。それなりに結果が伴うこともあれば、時には結論ありきで強引な認定・処分等も行われる。かように、構造上、アルゴリズムによる調査選定が採用されやすくなり、そこにますます限りある調査資源が集中投下され、その調査結果がアルゴリズムにフィードバックされ、取り込まれて、バイアスが繰り返され、より強くなっていくことになりはしないか。ここでは、後述する自動化バイアスの存在により、人間である調査官がシステムの判断に盲従するようになる危険性にも配慮すべきである。

歴史的にみて、仮に、国税庁に従順な納税者、税理士非関与の納税者、日本の税制に明らかな外国人の納税者など、国税庁からみて“税金を取りやすい納税者”から偏頗的徴税を行ってきたとすれば、そのような状況が延々と続くおそれがあることにも留意すべきである。とりわけ、外国人については、租税犯の刑に処せられた場合、重加算税を賦課された場合又は租税を滞納している場合に代表されるように納税義務を適正に履行していないことが、国籍法や出入国管理及び難民認定法との関係で外国人に多大な不利益をもたらさう上、このことに言語の問題等も加わって、外国人の立場は当局に対して弱いものとなりうることに注意が必要である⁽⁵⁴⁾。

学習用データに関して元々の調査選定や調査処理に問題がなかったかを十分に調査すべきであるし、そのような問題があることを十分に踏まえて、アルゴリズムの開発等を進める必要がある。アルゴリズムによる判断が統計的・確率的にみれば合理的な推論に基づい

(53) エコーチェンバー現象について、キャス・サンステイーン〔伊達尚美訳〕『#リパブリック』（勁草書房 2018）参照。

(54) 泉絢也「外国人労働者・移民の増加と租税—納税義務の履行状況が国籍法や入管法の法律関係に与える影響—」税法 583 号 2 頁以下（2020）参照。

ているとしても、元々のデータに誤りや偏見が含まれ、それがアルゴリズムによって再現ないし増幅され、個々の納税者がその属性等によりセグメント化ないし一般化・類型化された上で、調査必要度が高い者、コンプライアンスリスクが高い者であると判断され続ける可能性があることを度外視することはできない（後記（3）のブラックリスト問題への接続）。

イ 特定の属性への着目、プライバシー問題、情報収集・調査手法等の問題

納税者の特定の属性を重視して調査対象として選定することは不合理な差別として問題視されるべきか⁽⁵⁵⁾、どのような属性が問題となりうるのか、バイアスを取り除くことは可能なか⁽⁵⁶⁾、調査選定の段階で重視されるにとどまるのであれば許容されるべきかなど検討すべき点が見えてくる。かかる属性に関して、国税庁が比較的容易に入手可能なもので、かつ、議論の余地がありそうなものとして（ただし、ディープラーニングを利用する場合には、人間が思い付かないような選定理由の採用、特徴量の抽出が行われる可能性があることに留意）、法人名・個人名、本店所在地・居住地域、業種・業態、性別、年齢、趣味、人種、国籍、居住者・非居住者、身体や疾病の状況・健康状態、従順度（税務調査に対する協力度合い等）、宗教又は思想上の信条、家族関係・婚姻歴・離婚歴、交友関係・異性関係、取引先・取引金融機関（取引先に占める個人事業者の割合等）、関与税理士、などがある（納税者本人に限らず、親族や先代の不正行為の履歴⁽⁵⁷⁾を含むこれらに関する属性も同様である）⁽⁵⁸⁾。これらは、過去の調査資料、他の行政機関から入手した資料、尾行調査・内偵調査やSNS閲覧により収集した情報のほか、障害者控除、医療費控除又は寄附金控除の申告書類等から確認しうる。この辺りの議論については、納税者プライバシーや国税庁における情報収集・調査手法の観点から考察を進めることも考えられる⁽⁵⁹⁾。

データマイニングや機械学習アルゴリズムなどを前提とすると、予測の精度を希求する国税庁がより多くのデータを求める一山本龍彦教授の言葉を借りるならば“more data”⁽⁶⁰⁾を基本原則とする一ようになる可能性も否めず、かかる観点から研究を行うことの必要性が高まっている。今後、税務行政の将来は、納税者が行う税務申告よりも申告の元となるデータを重要視する道を進むのか⁽⁶¹⁾、申告の元となるデータよりも納税者が提出した税務申告又はこれを作成する納税者側のシステムを重要視する道を進むのか⁽⁶²⁾、いずれに

(55) 国籍を例にするならば、例えば、母国の納税風土や文化・習慣、日本との制度の相違により、コンプライアンスリスクが高いというように理由が付くのであれば、申告前の集団指導など上流で対策を講じることもありうるが、さりとて、下流での対策すなわち税務調査の実施の必要性が直ちに、大きく減少するものでもない。いずれにしても、国籍をどの程度重視するのか、当該納税者の国籍がコンプライアンスリスクの高さにつながる理由はどのようなものか、どの程度理由が付くのかなど困難な議論が待ち受けている。

(56) 調査対象者の選択でバイアスがないことを保証し説明することは困難かもしれないという指摘として、岡村忠生「租税手続のデジタル化と法的課題」ジュリ1556号58頁（2021）参照。

(57) なお、COMPASでは、親族の逮捕歴も質問項目に入っている。

(58) 前述した種々の事情から調査件数が伸び悩んでいる又は（仮に、これらの各集団に対して個別のノルマ件数を設定し、調査を実施しているとしても）深度ある調査ができていない特定の集団が存在する場合に、そのような集団に属していることのみをもって又はこのことを強調して調査選定することが、不合理な差別や偏見に当たるのか、調査の必要性要件（税通74の2等）を満たすのか、という問題を提起することも可能である。

しても国税庁が自ら収集するデータを重要視する道を進むのかといったように、いくつかの道に分岐している。どの道に進むかに応じて、柔軟な対策を講じる必要がある。

いずれにせよ、“more data”は端的に“less privacy”を意味するという指摘に耳を傾けるべきであろう⁽⁶³⁾。また、調査選定時に重視される情報や収集の対象となる情報の種類に係る問題、あるいは情報収集時点又は取得時点において想起される問題にとどまらず、広く国税庁による情報収集・調査手法のあり方と取得後の情報の取扱いについて、研究を深める必要がある

ここで想起しておきたいのは、IRSに関するものであるが、課税庁の調査権限が広範であるとしても、實際上、調査のリソースは限られていて、実調率は低調であるところ、この低い実調率が納税者のプライバシー保護に寄与してきたという指摘である⁽⁶⁴⁾。“more data”を基本原則とする課税庁がビッグデータを収集し、データマイニングやプロファイリングによって納税者を“丸裸にしていく”ようになれば、実調率が低いままであったとしても、換言すれば、量的には少ないとしても、質的に重大なプライバシー侵害を引き起こす可能性が高まる。もはや、税務職員の守秘義務（税通 127, 国公 100）⁽⁶⁵⁾や質問検査権の制限（税通 74 の 2 等）など既存のプライバシー保護装置のみではカバーできまい。この点については、分野が異なるものの、次の指摘も参考になる⁽⁶⁶⁾。

(59) 参考となる最近の文献として、Michael Hatfield, *Taxation and Surveillance: An Agenda*, 17 YALE J. L. & TECH. 319 (2015) (hereinafter *Taxation and Surveillance*); Michael Hatfield, *Privacy in Taxation*, 44 FLA. ST. U. L. REV. 579 (2017) (hereinafter *Privacy in Taxation*); Adam B. Thimmesch, *Tax Privacy?*, 90 TEMP. L. REV. 375 (2018) 参照。 測圭吾「日本の納税者番号制度」日税研論集 67 号 33 頁以下 (2016), 高橋祐介「納税者番号制度と納税者の秘密の保護」日税研論集 67 号 183 頁以下 (2016), 漆さき「国際的情報交換における納税者の権利保護」論究ジュリ 26 号 83 頁以下 (2018), 石村耕治「AI 税務と税務専門職の将来像を展望する 第 1 回～第 7 回・最終回—税務のスマート化とタックスプライバシー—」税務事例 51 巻 3 号 73 頁以下, 4 号 40 頁以下, 5 号 44 頁以下, 6 号 50 頁以下, 7 号 75 頁以下, 8 号 60 頁以下, 9 号 50 頁以下 (2019), 税研 216 号 (2021) の特集「プライバシー権と税制」掲載の各論稿も参照。

(60) 例えば、山本龍彦「個人主義とセグメント主義の相剋（覚書）—「パーソナライズド」の意味—」情報法制研究 2 号 54～55 頁 (2017) 参照。 See also Hatfield, *Taxation and Surveillance*, *supra* note (59), at 340-350; Hatfield, *Privacy in Taxation*, *supra* note (59), 610-614 (“Between the pressure of the tax gap and the efficiency potential of information technology, we should expect the IRS to collect more and more tax-relevant information over the next twenty-five years”).

(61) See Dmitri Jegorov, *Technological Advances in Administering Taxes: Will There Be a Tax Office in the Future?*, 74 BULLETIN FOR INTERNATIONAL TAXATION 598, 607 (2020).

(62) See OECD, TAX ADMINISTRATION 3.0, *supra* note (1)

(63) 山本・前掲注 (60) 54 頁参照。

(64) See Hatfield, *Privacy in Taxation*, *supra* note (59), at 610-614.

(65) 守秘義務の規定が納税者等のプライバシー保護規定及び税務情報の保護規定としての意味をもっていることを認めた上で、①調査担当職員以外のすべての税務職員の義務として規定すべきであること及び②情報化社会＝コンピューター時代における税務情報の保護という観点からみると、税務職員の守秘義務を通じて間接的に税務情報を保護するという手法はいかにも古く、上記時代の実情に適合した保護体制を考えるべきであり、税務情報開示禁止原則を規定すべきであり、長期的には自己情報管理権を保障すべきであることを主張するものとして、金子宏「税務情報の保護とプライバシー—納税者番号制度を視野に入れて—」『所得課税の法と政策』231 頁以下 (有斐閣 1996) [初出 1994] 参照。

(66) Elizabeth E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing*, 10 HARV. L. & POL'Y REV. 15, 17 (2016).

裁判官や法律学者は、伝統的な警察活動において警察による市民に対する監視の裁量権の行使にほとんど注意を向けてこなかった。それは、警察がそのような権限を保持すべきであるという仮定と、理論的には懸念があるとしても、監視の裁量権は、実際には相当に制限された権限であることに起因する。通常、リソースや技術上の制約によって、警察による捜査の対象となる人数は限定的なものにならざるをえない。しかしながら、警察が自動化された監視の裁量権を行使するためのツールを保持するときは、以上の想定は時代遅れのものとなる。

さらにいえば、近時において、国税庁は、個人のブログ、TwitterやFacebookなどSNSから情報を集め、これらと国税庁の保有する情報とを合わせて、調査選定の材料とすることができる。自身や他人によるSNSの何気ない投稿を通じて、自身の気付かないところで、プロファイリングをされて、コンプライアンスリスクが高いというレッテルを貼られることに一種の気味の悪さを感じる、あるいは納得できない納税者も多いのではないか。この点に関して、airbrushing問題が存在することについては既述のとおりである。

IRSはSNSから情報を入手する際に、職員が個人アカウントを使用したり、身元を偽ったりしないように指導しており、サイトへのアクセスやフレンドステータスを得るために架空のIDを使って登録することも禁じているようである⁽⁶⁷⁾。国税庁はこの辺りについてのどのような規律を設けているのであろうか。情報の誤りや偏り、(質問検査権の行使を伴わない情報収集も含めた)国税庁における情報の入手経路や入手手段の問題としてのみならず、納税者による国税庁保有情報へのアクセス・訂正の可否等の問題として、今後、議論を深化させる必要がある。

(3) ブラックリスト(ウォッチリスト)化に関する問題

バイアス問題とも関わるブラックリスト問題の存在も指摘しておく。例えば、AIが、過去の調査データに基づいて、特定の国籍を有する納税者に対して調査必要度が高いと判断するようなアルゴリズムを生成したとする。すると、不正計算を行う確率という観点から、特定の国籍を有する納税者が、調査必要度の高い者として継続的にリスト化され、ひいてはブラックリスト(ウォッチリスト)化されることになる。このこと自体は人間による選定でも起こりうるし、アルゴリズムの判断それ自体は差別や偏見に基づいているとはいえない面もあるが、それでもなお、上記のようなブラックリスト化に対する懸念が消失することはない。

(67) See Electronic Frontier Foundation, *IRS Training Course*, IRT-WBT CONTENT 2009 (2009), <https://www.eff.org/document/internet-research-tools>. IRSがFacebookから情報を得ていることの参考として、「IRSの税金詐欺の女王」と自称していたともいわれるRashia Wilson氏の事件がある。同氏は、他人のIDを不正利用するなりすまし詐欺により税金の還付を受けていたところ、同氏のFacebookの投稿が原因でそのことが発覚し、有罪とされた。IRSは、同氏のFacebookの公開アカウントを検索し、情報を得ていたといわれている。See Houser & Sanders, *supra* note (40), at 836, 839-840; Robert W. Wood, *Queen of Tax Fraud Gets 21 Year Prison Term: For the Second Time*, FORBES (Mar. 6, 2015), <https://www.forbes.com/sites/robertwood/2015/03/06/queen-of-tax-fraud-teasing-i-cant-be-caught-gets-21-years-prison-for-the-second-time/?sh=45a5c91b4d15>.

これは、前記Ⅲ 2 (1) のセグメントに基づく統計的・確率的予測の箇所です。述べたようなプロファイリングがもたらす不安やリスクに対する懸念といってもよい。統計的・確率的裏付けによって永遠に消すことができない、あるいはブラックボックス部分を視認できないような、例えば「遁脱を行っている確率が非常に高く、よって調査必要度の高い者又は還付を保留すべき者」といったスティグマを本人の関知しない、コントロールできないところで作り出す可能性があることに対する不安やリスクである⁽⁶⁸⁾。

上記の例でいえば、AI の判断に従って、特定の国籍を有する納税者に国税庁の調査資源が集中投下されることによって、偏ったデータがますます蓄積していく上、過去の税務調査で不正計算が把握されなかった納税者であっても、当該国籍を有している以上、確率的にみて不正取引を行っているという判断は容易には覆らない可能性がある。一度、ブラックリストに名前が掲載された場合、そのリストから削除されることは容易ではないことが想定される。

(4) ブラックボックスに関する問題

調査対象として選定された納税者は、無数に存在する納税者の中からなぜ自分が優先的に選定されたのかを知りたいと思うであろう。機械学習アルゴリズムを利用して選定している場合に、その結果や入力されたデータに誤りや偏りがありうるならば、なおさらである。しかしながら、納税者が次のような点を知りたいとしても、そもそも技術的に難しい可能性がある。技術的要因に帰するブラックボックス問題である。

- ・そのアルゴリズムは、どのようにして、調査必要度の判定を行うのか。どの入力変数（特微量）を重要視するのか。
- ・なぜ自分が、アルゴリズムによって、調査必要度が高いと判断されたのか。どの入力変数（特微量）が重要であったのか。
- ・入力変数（特微量）と調査必要度との間にどのような因果関係があるか。

例えば、確たる理由は不明であるものの不正行為を行っている可能性が高いと AI によって判断された納税者が調査対象として選定されたとする。この場合、調査の必要性要件（税通 74 の 2 等）を充足するといえるか、という問題がある。ここでは、ディープラーニングを利用する場合には、人間が思い付かないような選定理由の採用、特微量の抽出が行われる可能性があることに留意する必要がある（極端な例として、AI が「決算月の平均気温」、「納税者の生年月日」などを選定理由として重視する場合を考えてみるとよい）。AI の真の選定理由に合致しているか否かにかかわらず、選定理由について、人間による補充的な説明が可能であれば問題なしとするべきか。

(68) 参考として、河島茂生ほか「座談会 AI・ロボットの研究開発をめぐる倫理と法」福田雅樹ほか編『AI がつなげる社会』124～127 頁（弘文堂 2017）、山本・前掲注 (22)「AI と『個人の尊重』」320 頁以下、山本・前掲注 (22)「ロボット・AI は人間の尊厳を奪うか？」79 頁以下、山本龍彦「AI と個人の尊重、プライバシー」同編『AI と憲法』（日本経済新聞出版社 2018）、キャシー・オニール〔久保尚子訳〕『あなたを支配し、社会を破壊する、AI・ビッグデータ』（インターシフト 2018）など参照。

上記に列挙した点について、前記Ⅲ2(3)で述べたとおり、技術的な観点から、一定程度、判明する可能性はある。ただし、かように技術的要因に帰するブラックボックス問題が一定程度解消されるとしても、その情報が公開されない、あるいは当該納税者に開示されないという法制度的要因に帰するブラックボックス問題もある（もちろん、これらの点が明らかになったとしても、納税者自身が理解できるかどうかは別問題であるし、ブラックボックスがどの程度、明らかにされるべきなのか、常に因果関係まで明らかにされる必要があるのかなど議論の余地は残る）⁽⁶⁹⁾。

この点に関して、アメリカと同様に日本においても、アルゴリズムの開発を委託された企業の企業秘密に該当することや公開した場合に税務行政の適正な遂行に支障を及ぼす情報に該当することなどを理由に（行政情報公開5二～四、六柱書・イ、行政個人情報14三～五、七柱書・イ、平成17年4月付け国税庁「行政機関の保有する情報の公開に関する法律に基づく処分に係る審査基準」）、開示が認められない可能性が高い⁽⁷⁰⁾。

例えば、情報公開・個人情報保護審査会の答申（平成19年4月16日・平成19年度（行個）答申第1号）は、行政機関の保有する個人情報の保護に関する法律に基づく開示請求事案において、「所得税及び消費税の調査選定に当たっての具体的選定理由、事務処理上の必要事項等が記載される欄」に係る部分について、同法14条7号イの不開示情報に該当することを認めている。その理由として、「これらの記載情報は、国税当局が調査対象者を選定するに当たっての着眼点とも言うべき情報であることから、これらを明らかにすれば、国税当局が審査の材料を持っているかどうかという税務調査の手の内を明かす結果となると認められる。当該不開示部分を開示した場合には、所得税及び消費税の調査選定に当たっての着眼点等を知った一部の納税者においては、自らが調査の対象となる可能性の有無を予測し、税額計算上の不正手口の巧妙化を図るなど、租税の賦課又は徴収に係る事務に関し、国税当局による正確な事実の把握を困難にするおそれ又は違法若しくは不当な行為の発見を困難にするおそれがあると認められる。」と述べている（令和2年8月3日・令和2年度（行個）答申第58号も参照）。

また、同審査会の答申（平成31年3月11日・平成30年度（行情）答申第474号）は、行政機関の保有する情報の公開に関する法律（以下「情報公開法」という）に基づく開示請求事案において、「国税総合管理システムによる還付申告機械チェックにおける還付保留基準に関する情報」が記載されている部分について、同法5条6号イの不開示情報に該

(69) ブラックボックス性について、技術的要因と法制度的要因に加えて、アルゴリズムは「最小限の透明性義務の下で運営され、利益の最大化を図る民間企業」によって実装されていることがほとんどであるという組織的要因を挙げるものとして、Guido Noto La Diega, *Against the Dehumanisation of Decision-Making: Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, 9 JIPITEC 3, 9-10 (2018), <https://www.jipitec.eu/issues/jipitec-9-1-2018/4677> 参照。また、技術的要因は、当事者のリテラシーの問題とも関わる。“opacity as technical literacy”の存在を指摘するものとして、Jenna Burrell, *How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms*, BIG DATA SOC. (Jan. 6, 2016), <https://journals.sagepub.com/doi/full/10.1177/2053951715622512> 参照。

(70) 犯則事件の場合には、純粹の刑事司法とは異なるがそれと密接に関連するものであり、行政機関の保有する情報の公開に関する法律5条4号の公共の安全等に関する情報に該当する場合がありますが問題となること及び行政調査の手法に関する行政文書であって、その開示が法の潜脱を招くようなものは同条6号イに該当するという指摘として、宇賀克也『情報公開の理論と実務』168頁（有斐閣2005）参照。

当することを認めている。その理由として、「これを公にすると、消費税の還付を不正に受けようとする一部の納税者において、還付保留を免れるため、当該基準を下回る水準等での還付申告を行うことが容易となることから、租税の賦課又は徴収に係る事務に関し、国税当局による正確な事実の把握を困難にするおそれ又は違法若しくは不当な行為の発見を困難にするおそれがあると認められる。さらに、当該基準に該当する不正還付申告を提出した法人が、還付保留された場合には、実地調査等による接触が行われることを予測し、税額計算上の不正手口の巧妙化を図るなどにより、租税の賦課又は徴収に係る事務に関し、国税当局による正確な事実の把握を困難にするおそれ又は違法若しくは不当な行為の発見を困難にするおそれがあると認められる。」と述べている⁽⁷¹⁾。

さらに、同審査会の答申（令和2年9月7日・令和2年度（行情）答申第240号）は、情報公開法に基づく開示請求事案において、「移転価格調査に対する国税当局の問題意識や調査対象（法人）を選定する際の情報収集の方法並びに移転価格実態確認を含む移転価格調査における具体的な調査手法、着眼点、資料情報の収集方法及び参考資料の参照先並びに移転価格調査に関して法人から想定される質問事項やそれに対する国税当局の応答例等が記載されている」部分及び「移転価格調査において課税要件の充足を立証する等のための根拠資料を収集する具体的なケース及び当該ケースにおける留意事項や調査の方針、収集する根拠資料の具体的な資料名や資料収集の方法、過去の移転価格調査の具体的な事例並びに進捗段階ごとに検討すべき事項等が具体的かつ詳細に記載されている」部分について、同法5条6号イの不開示情報に該当することを認めている。その理由として、「当該情報は、国税当局において限られた人員の中で適正・公平な課税を実現するための重要な要素であり、さらに、税務調査に関する事項が納税者の重大な関心事であることからす

(71) 情報公開法に基づいて行われた行政文書開示請求に係る不開示処分（「消費税還付申告に係る事務処理手順等について（事務運営指針）（平成12年6月30日付課法第52号ほか1課共同）」の行政文書開示決定処分のうち、「還付保留基準、還付保留チェック表及び還付申告法人に対する接触体制について記載した部分」）の取消請求事案において、富山地裁平成17年1月12日判決（税資255号順号9887）は、同法5条6号イの不開示情報に該当することを認めている。まず、判決は、「還付保留基準は、不正還付の可能性がある法人を類型化したもので、還付保留基準に該当しない法人の還付申告については、その後何らのチェックが行われることなく、速やかに還付がなされることになり、還付保留チェック表は、還付保留基準により抽出された還付申告書をより細かくチェックするもので、チェック項目のいずれにも該当しなければ、実地調査を行うことなく、速やかに還付がなされるものである」ことを確認する。その上で、「還付保留基準及び還付保留チェック表が予め公になれば、不正還付を受けようとする法人があった場合、還付保留基準及び還付保留チェック表に該当しないような還付申告書を作成することも可能になり、その結果、何らの調査を受けることなく、不正還付を受けることが可能になる。また、還付保留基準及び還付保留チェック表に該当しないような還付申告書の作成まではしないとしても、自らが提出した還付申告書について、税務署による実地調査を受けるかどうかを事前に予測できるので、予め仕入れその他の伝票類などを偽造又は変造する等の不正な対策を取ることが容易になり、「したがって、還付保留基準及び還付保留チェック表が開示された場合には、消費税の還付原因事実があるか否かという正確な事実の把握を困難にするおそれがあると同時に、不正還付という違法な行為を容易にし又はその発見を困難にするおそれがあるといえる。この結果、消費税の還付という事務の適正な遂行に支障を及ぼすおそれがあるといえる。このおそれは、以上の検討内容からみて、単なる抽象的な可能性ではなく、法的保護に値する程度の蓋然性があるといえるべきであり、5条6号イに係るおそれが認められる。」と判示する。この判断は、控訴審である名古屋高裁金沢支部平成17年10月12日判決（税資255号順号10158）においても維持されている。

れば、これを公にすると、一部の納税者が調査対象等となることを予測し、税額計算上の不正手口の巧妙化を図る、あるいは、今後の税務調査への対策を講じるなどにより、移転価格調査事務に関し、国税当局による正確な事実の把握を困難にするおそれ又は違法若しくは不当な行為を容易にし、若しくはその発見を困難にするおそれがあると認められる。」と述べている⁽⁷²⁾。

AIや機械学習アルゴリズムを利用した調査選定システムに特有の問題ではないが、質問検査権の行使における調査の客観的必要性や調査理由の個別具体的な開示の問題場面⁽⁷³⁾においても法制度的要因に基づくブラックボックス問題が立ちはだかっているとみる余地もある。

また、質問検査権の行使に関して、課税庁における裁量の濫用の有無が問題となった場合には、納税者を不当に扱う課税庁の意図の有無や他事考慮の存在が論点となりうる。かかる場面において、ブラックボックス問題は、結果的に意図を有さないアルゴリズムが課税庁にとっての“隠れ蓑”になることになりはしないか、という懸念にも接続する⁽⁷⁴⁾。

なお、再述するが、ビッグデータの分析によって相関関係を示すことはできるかもしれないが、相関関係は因果関係ではなく、発見された相関関係が常に意味のあるものではない点に注意が必要である。

(5) 偽陰性と偽陽性に関する問題

偽陰性と偽陽性のどちらをどの程度重視すべきかという問題がある⁽⁷⁵⁾。実際にはコンプライアンスリスクが低い納税者について高いと誤判定する割合である偽陽性率が高い場合、“見込み違いによる権利侵害”を数多く引き起こすという問題がある。他方、コンプライアンスリスクが高い納税者について低いと誤判定する偽陰性率が高い場合、脱税犯や多額の過少申告者が、長期的に“野放し”となる可能性がクローズアップされるが、かように精度の低いアルゴリズムは、増差所得に結び付かず、成果がでないため、組織的に用いられないことになるから、自然淘汰されることが予測される。

(72) 参考として、情報公開法に基づいて行われた税務調査及び更正処分等において取得・作成された行政文書の開示請求に係る不開示処分の取消請求事案において、東京地裁平成19年8月28日判決（訟月55巻8号2764頁）は、移転価格税制に係る比較対象取引の選定過程を示す文書について、「同業他社の資本金、売上高、主力商品名、商品の主な販売先、販売組織の形態及び販売エリア、売上順位又は売上高の対前年比といった事業内容に関する事項につき、いずれの法人の取引を比較対象取引として選定するか、又は選定しないかに関する個々の同業他社に係る東京国税局職員の評価は、これらを公にすれば、東京国税局において、いずれの法人の取引を比較対象取引とするかという、その抽出過程及び比較対象法人が明らかになるものである」とした上で、「こうした情報が明らかになれば、移転価格税制上の問題を検討する際の税務当局の関心事項が明らかとなるため、納税者の中には、税務調査の対策として、関係する資料を隠ぺい・破棄する可能性も存し、その結果、税務当局において正確な事実の把握が困難になるなど、移転価格調査に関する事務の適正な遂行に支障を来すおそれがあることから、情報公開法5条6号イに該当するものと認められる。」と判示している。

(73) かかる問題場面については、酒井克彦『裁判例からみる税務調査』97～109頁（大蔵財務協会2020）参照。

(74) なお、行政機関が機械学習アルゴリズムを意思決定ツールとして利用するべきかどうかを判断する際の目安として、行政機関における裁量の広狭の程度を挙げるものとして、Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U.L. REV. 1277 (2018) 参照。

(75) 議論の参考として、Mayson, *supra* note (30) 参照。

予算を投じた以上、目に見える成果を求める国税庁（と委託を受けてアルゴリズムを開発する企業）には、偽陽性率を閉却して偽陰性率を重視するバイアスが生じるのではないかという懸念がある。一般論として調査選定の精度が高まること自体は歓迎されるが、だからといって、アルゴリズムの生成に当たって、偽陽性率やその弊害を軽視し、偽陰性率を低く抑えることのみを重視するような方針を採用すべきではない。例えば、施設内に放置された不審物を発見するためのアルゴリズムであれば偽陰性率を低く抑えることに傾注するような方針を採用することが認められるかもしれないが、権利侵害の危険がより深刻となる場面においては、偽陽性率を低く抑えるよう細心の注意を払うべきである。手続的保障の原則や納税者の権利保護を考慮して、偽陽性の割合を低く抑えることを重視するという方針を明示的に採用すべきであるが、この点に関する国税庁の見解は明らかではない⁽⁷⁶⁾。

3 対応策その他の検討

(1) 特定の属性の使用制限

バイアス問題について、不当な差別や偏見につながるような特定の属性を入力変数（特徴量）として使用することを禁止することで問題は解決するであろうか⁽⁷⁷⁾。仮に、バイアス問題に対処するために納税者の国籍に代表されるような特定の属性に係る情報を考慮しない、学習用データに組み込むことを禁止するような仕組みを採用するとしても、データマイニングが国籍による不公正な影響を与えるためには国籍を明示的に考慮する必要はないことに注意を要する。本人や家族の氏名のほか、業種・地域・取引先などから納税者の国籍を自動的に推定するか、推定したのと同じ結果になることが想定されるのである。

データマイニングのシステムは、その意図とは無関係に、差別的な結果を生み出したり、悪化させたりする可能性のある一連の人為的な決定を組み込んでいると考えておいた方がよいであろう⁽⁷⁸⁾。制限すべき特定の属性とはどのようなものか、特定の属性を考慮しないこと自体が妥当であるかという点に加えて、結局、他の事実から国籍が暗黙的に推測され、勘案される結果とならないか、など検討すべき課題は残る。

(2) 人間による判断の介在と自動化バイアスの問題

(76) IRS が現在使用している不正検知システムの 1 つである Return Review Program (RRP) の偽陽性率と他のシステムの偽陽性率を比較検討するものとして、TIGTA, *The Return Review Program Increases Fraud Detection; However, Full Retirement of the Electronic Fraud Detection System Will Be Delayed*, REFERENCE NUMBER: 2017-20-080 (2017) 参照。

(77) 租税法規の中に、これらの要素が直接的に課税要件として取り込まれていることは稀である。また、国家が収集することが許されないような情報（例えば、思想や信仰）を課税要件とすることは、そもそも租税実体法として憲法上許されない、という指摘もある。測・前掲注 (59) 59 頁参照。なお、国籍や人種などの差別や不平等につながりうる要素に係る統計データを公表等しないということは、現行の租税法規の中に直接的にこれらを要件として取り入れているものが少ないことを考慮したとしても、差別や不平等が存在する現実が存在している場合に、その事実を覆い隠し、検証を妨げ、現状の維持を招来しかねない、という視点もありうる。See Jeremy Bearer-Friend, *Should the IRS Know Your Race?: The Challenge of Colorblind Tax Data*, 73 TAX L. REV. 1 (2019). かような観点から国税庁による統計の取り方やその公表のあり方が検討されてもよい。

(78) 予測的警察活動との関係で、Selbst, *supra* note (15), at 116 参照。

バイアス問題に対して、調査選定の最終判断の過程に人間による判断を介在させる、最終的な判断を人間が行うこととすることで対処することできるか。バイアス問題については人間が機械学習アルゴリズムに潜むバイアスの存在に気が付くことができるかという問題がある上、自動化バイアスという問題があることも指摘しておかねばならない。

人間は自動化されたシステムをエラーに強いとみなし、コンピューターの解答を信頼する傾向があり、コンピューターが生成した解答と矛盾するような情報を探すことが少なくなってしまうという自動化バイアスの存在が指摘されている。アメリカのテキサス州で導入された社会福祉手当の受給資格の判断に利用される TIERS (Texas Integrated Eligibility Redesign System) など、機械と人間の混合システムの失敗が引き合いに出される。人間の資格審査担当者がコンピューターの判断をレビューし、最終決定を下すことになっていたにもかかわらず、誤った判定が数多くなされたというのである。かような混合システムの失敗は、完全に自動化されたシステムと混合システムとの間の実際的な区別は誇張されるべきではないことを強く示唆する。資格審査担当者のコンピューターシステムに対する直感的な信頼が混合システムにおける人間の参加の価値を低下させている⁽⁷⁹⁾。

このほか、ジョージア工科大学が行った実験では、被験者は、ロボットに案内されて廊下を進み、部屋に入ってアンケートに回答していると、火災報知器が鳴り、煙が部屋中に充満し始め、「緊急時誘導ロボット (Emergency Guide Robot)」と表記のあるロボットが動き出すという状況に置かれた。被験者は、ロボットに従って未知のルートを進むか、あるいは部屋に入って来たドアから自力で脱出するか一瞬の判断を迫られたが、30人の被験者のうち26人がロボットに従うという選択をただけではなく、明示されている出口から離れた方向に誘導されたとしても、ロボットの後を追い続けたという⁽⁸⁰⁾。また、オラクルは、金銭の管理に関して、人々は人間よりも、AIやロボットを信頼しているという調査結果を示している⁽⁸¹⁾。

かように自動化バイアスが存在する状況下では、何らの対策も講じないのであれば、人間による判断を介在させることに大きな成果を期待することはできない。ここでは、調査の必要性要件(税通74の2等)の充足を判断する裁判官においても自動化バイアスが働き、司法による救済を期待できない可能性も視野に入れておきたい⁽⁸²⁾。自動化バイアスへの対策として、例えば、人間の調査官による調査選定と機械学習アルゴリズムによる調査選定のいずれによって選定されたものを伏せて、候補者をリスト化し、最終的に人間によって判断させるなどの工夫を施す必要がある⁽⁸³⁾。

また、ブラックボックス問題についても対処が難しいことは明らかであるが、人間による判断や説明を介在させることで、因果関係の発見に向かうことや、納税者に対してわかりやすい、有益な説明がなされることはありえよう。

(79) See Danielle K. Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1270-1272 (2008).

(80) See Anna Green, *Humans May Put Too Much Trust in Robots, Study Finds*, MENTAL FLOSS (Mar. 2, 2016), <https://www.mentalfloss.com/article/76387/humans-may-put-too-much-trust-robots-study-finds>.

(81) See Oracle & Savanta, *Money and Machines: 2021 Global Study* (2021).

<https://explore.oracle.com/money-and-machines/oracle-money-and-machines-report/>. 要約として、日本オラクル「日本を含む14カ国、9,000人調査：67%が、財務管理において自分以上にロボット・AIを信頼すると回答」(2021年3月25日), <https://www.oracle.com/jp/corporate/pressrelease/jp20210215.html>.

なお、AI やアルゴリズムが関わる判断について人間の責任者を明確にするという対応策は一定の効果を発揮するであろう。ただし、個人や組織としての理解やコントロールが及ばないものに対して、組織が個人に責任を押し付けてはならないことに留意する必要がある。

(3) 人間による判断 vs アルゴリズムによる判断

そもそも、機械学習アルゴリズムによるバイアスある判断の遠因が人間の判断に求められることからわかるように、人間の判断に問題がないわけではないことは自明であろう⁽⁸⁴⁾。調査選定システムの結果を経て、結局は人間である職員が調査対象者として選定

(82) 裁判官を例に自動化バイアスを説明するものとして、Harry Surden, *Ethics of AI in Law: Basic Questions*, in THE OXFORD HANDBOOK OF ETHICS OF AI 719, 734 (Markus D. Dubber et al. eds., 2020) 参照。なお、現行の行政法規は、行政機関の決定が人間によってなされることを前提として作られているところ、職員による機械学習アルゴリズムへの過度の依存は、場合によっては、委任立法の問題として捉えることも可能であり、行政機関が決定を行う際に機械学習アルゴリズムを利用する権限を議会が行政機関に対して与えたものといえるか、あるいは議会から行政機関に対する特定の権限の委任は当該行政機関がその意思決定を機械学習アルゴリズムにさらに委ねることを妨げるものか、という議論にも接続することについて、Coglianese & Lehr, *supra* note (8), at 1178-1179 参照。

(83) 参考として、Engstrom & Ho, *supra* note (51), at 807, 849-853 参照。

(84) 裁判所や裁判官の判断にも問題があるものは存在すると考える向きもある。その判断が外部から理解できるものであるかという点については、理解の意義や理解の対象、一般市民や専門家など誰を基準にするかなど、種々の観点から検討の対象になる。

裁判所の判決は一般市民が理解できるものではないことなどを指摘した上で、従来の手動の法的プロセスで提供されるものよりもはるかに高いレベルの透明性を暗黙のうちに自動意思決定システムに要求することに対して、批判的見解を示すものとして、Jay Thornton, *Cost, Accuracy, and Subjective Fairness in Legal Information Technology: A Response to Technological Due Process Critics*, 91 N.Y.U. L. REV. 1821, 1838-1839 (2016) 参照。

裁判官が直感や専門的知識に基づいて法的な判断を下すことはあるものの、その判断の正当化理由を提供することは可能であるし、その事件の重要な事実—AI モデルに与えられるパラメータに似たもの—の多くとその重み付けはその説示から認識可能であるという指摘として、Bathae, *supra* note (30), at 152-153 参照。

裁判官による判断について、実際に判決文などにより明らかにされた理由で当該判断に到達したかどうかを検証する方法はないのに対し、AI システムを利用する場合には、理解可能な機械学習技術を使用するならば、少なくとも意思決定プロセスを完全に再構築できる可能性がある。このように、慎重に導入を進めるのであれば、AI の利用が増えることで、現行の法制度に比して、透明性を高めることができるという側面を有するという見解として、Surden, *supra* note (82), at 730-732 参照。

また、Michael E. Donohue, *A Replacement for Justitia's Scales: Machine Learning's Role in Sentencing*, 32 HARV. J. L. & TECH. 657, 664-665 (2019) は、次のような見解を示す。

裁判官は、アルゴリズムとほとんど変わらないブラックボックスである。いずれの場合も、当事者は、判決の判断過程を完全に検証することはできないとしても、その決定に係るインプットの正確性について、互いに確認しうる。また、アルゴリズムについてはオープンソースで精力的に検証されたツールセットがあれば不安は緩和される。システム全体のレベルでエラーが発見されたり、機能強化が行われたりすると、その修正をシステムのすべてのインストールに適用しうる。他方、裁判官のブラックボックスの性質は、システム全体のレベルで修正をかけることは難しく、特にそれを修正するためには、裁判官は、その理由付けについて、より広範かつ定期的に文書化することに取り組む必要がある。

参考として、笹倉宏紀「AI と刑事司法」弥永真生＝穴戸常寿編『ロボット・AI と法』233 頁以下（有斐閣 2018）も参照

するか否かを判断するのであれば⁽⁸⁵⁾、同システムの採用も従来の人間による意思決定の延長に位置付けることができるかもしれないが⁽⁸⁶⁾、この点は、自動化バイアスの存在により、人間がアルゴリズムの判断に盲従する危険性があるのではないか、人間による意思決定がブラックボックスでないといえるかという点も踏まえて、更に議論を要する。

これに対して、人間の判断にも問題があるのだとすれば、むしろ、アルゴリズムは修正が効く可能性が高く、公正なアルゴリズムを組めば、人間のように偏見が入ることなく、公正な結論が導かれる、よってアルゴリズムによる判断を支持すべきであり、アルゴリズムの利用を進めるべきであるといった声が上がること考えられる⁽⁸⁷⁾。アルゴリズムによれば、人間よりも、効率的に安定的な判断が下せるといふ点からも利するところがある。いずれにせよ、AIないし機械学習アルゴリズムの利用は、これまで人間が行ってきた営みを省察する機会を提供する。

(4) ブラックボックスと透明性

法制度的要因に基づくブラックボックス問題に関して、企業秘密の保護、守秘義務の存在又は納税者による悪用防止等の観点から、透明性を高めることへの限界、ブラックボックスをホワイトボックス化することへの限界が存在することを認識しうる。もっとも、企業秘密として保護されるべき、あるいは不誠実な納税者がアルゴリズムの情報を得てこれを悪用することを可能にするようなことは回避されるべきであるという観点は、アルゴリズムに関するあらゆる種類の情報に一律ないし一様に妥当するわけではない⁽⁸⁸⁾。

透明性には、国民に情報や決定過程を公開することを意味するもの (fishbowl transparency) と自己の行為の正当化理由を説明することを意味するもの (reasoned transparency) があるところ、両者は共存するものの、両者の区別を意識して議論することが必要である。国税庁としては、企業秘密や公務への支障を理由に前者の透明性に制限を設ける場合においても、後者の透明性要求を満たす方法を探るべきである⁽⁸⁹⁾。ここでいう透明性を「情報の公開かつ合理的な説明」に置き換えて、論じることもできよう。「完全な透明性 (full transparency)」が常に求められるわけではなく、現実的な範囲で、時には第三者機関による監視制度など他の制度を踏まえた上で、説明責任を全うすることに

(85) AIによる意思決定、AIの判断に大きく依拠した意思決定が行われる未来も見据えた議論が必要かもしれない。なお、政府の意思決定場面においては、人間による決定の代わりにアルゴリズムを使用するという考え方は、民主主義的ガバナンスにとって深刻な脅威となり、説明責任を果たせないコンピューター化された君主のイメージを思い起こさせるという指摘として、Coglianese & Lehr, *supra* note (8), at 1152 参照。

(86) 参考として、Coglianese & Lehr, *supra* note (8), at 1222 参照。麻薬の捜査等における麻薬探知犬に依拠した判断と何が違うのかという視点からの比較検討も興味深い。Rich, *supra* note (12), at 911-923。ただし、Lehr & Ohm, *supra* note (9) も参照。

(87) See Mirko Bagaric et al., *Erasing the Bias Against Using Artificial Intelligence to Predict Future Criminality: Algorithms Are Color Blind and Never Tire*, 88 U. CIN. L. REV. 1037 (2020)。

(88) 参考として、Rich, *supra* note (12), at 922-923 参照。

(89) 参考として、Cary Coglianese, *The Transparency President?: The Obama Administration and Open Government*, 22 GOVERNANCE 529, 537 (2009); Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1 (2019); Steven M. Appel & Cary Coglianese, *Algorithmic Governance and Administrative Law*, in THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 153, 176 (Woodrow Barfield ed., 2020) 参照。

つながる透明性が求められるのである⁽⁹⁰⁾。AIの開発に当たり、民間企業が主導権を握り、依頼主である行政機関がアルゴリズムを含む重要な点を理解できていなかったり、民間企業任せにしたりしていると、実質的に民間企業に説明責任が移るような事態が発生する可能性や民間企業の影響力が大きくなる可能性が高まる⁽⁹¹⁾。この点の是非も議論が必要であろう。

国税庁がアルゴリズムに関する情報の開示を行う場合には、意図的に情報を操作することなどを通じてアルゴリズムを騙してその判断を歪めた納税者へのペナルティの創設をセットで設けることも検討に値する。あるいは、調査の必要性要件（税通 74 の 2 等）の充足可能性や選定理由の説明可能性という問題に配慮する必要はあるが、調査選定システムの設計にランダムを要素を組み込んだり、同システムとは別にランダム選定の要素を含む調査選定を実施したりすることも議論されてよい。これにより、納税者に対して自分が選定されるか否かの確信をもてなくする、換言すれば、納税者を未知のリスクに直面させることで、戦略的に選定を免れようとするような納税者の試みを挫くのである⁽⁹²⁾。なお、リアルタイムで、データを取り込んで予測のためのモデルを更新するような調査選定システムを採用する場合にも、同様の効果を見込むことができるかもしれない⁽⁹³⁾。

(5) 憲法上の問題と対応策

これまでに挙げてきた種々の法的問題に対して、AIないし機械学習アルゴリズムを実装しているといっても、過去の調査データに大きく依拠した調査選定であり、そこで挙げられた法的問題は人間による調査選定の場合にも起こりうるものであって、何はともあれ選定の精度が高まるのであれば、それでよいのではないか、別途、課税処分の適法性が確保されるのであればよいのではないか、という見解もあるかもしれない。

一般論でいえば、調査選定の精度が高まること自体は歓迎される。しかしながら、納税者にとって税務調査は種々の負担が大きいものであるし、例えば、「国籍が〇〇であるから、AIが不正行為を行っている可能性が高いと判断した」という理由で、調査対象として選定することの是非は議論の余地がある。国籍の代わりに、前記 2 (2) イで挙げた特定の属性を入れて思考実験を進めてみてもよい。ディープラーニングを利用するには、人間が思い付かないような選定理由の採用、特徴量の抽出が行われる可能性があることを考慮すると、なぜそのような属性が不正行為を行っている可能性が高いという判断につながるのかを説明できない事態も想定されるが、それは常に許容されるべきか。

また、「理由は明確ではないが、AIが、あなたは不正行為を行っている可能性が高いと判断しているから、今回、調査対象として選定されました」と当該納税者に対して説明しても納得を得られないので、かような実際の選定理由を伏せておき、取って付けたような別の理由を用意して、当該納税者に対して説明するような調査官が出てくることも予想さ

(90) See Kroll et al., *supra* note (46), at 704-705.

(91) 参考として、Surden, *supra* note (82), at 734; Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. 633, 685 (2020) 参照。

(92) See Kroll et al., *supra* note (46), at 654.

(93) See Kroll et al., *supra* note (46), at 659-660.

れる。

これらの点も踏まえて、これまで指摘した法的问题というのは、調査の必要性要件（税通74の2等）に関係するばかりでなく、個人の尊重、幸福追求権、平等原則を基本原理として掲げる憲法13条や14条、あるいは適正手続を保障する31条や84条等に抵触する恐れもある重要な問題であることを強調しておきたい。ここでは、調査選定システムの利用と同システムへの依存が、上記で指摘したようなこれまでに類をみないほどのループないしエコーチェンバー現象を発生させるのではないかという懸念があることを改めて想起しておきたい。何はともあれ選定の精度が高まるのであれば、それでよいのではないかと、別途、課税処分の適法性が確保されるのであればよいのではないかと、簡単に片付けることはできない。

納税者の権利利益を侵害する可能性がある場面などのように、AIないし機械学習アルゴリズムによる予測が外れる可能性があること、バイアス問題やブラックボックスの問題などの問題が起こりうることを念頭に置いて、より慎重な対応を求められる場面があることを意識した制度作りや運営を行う必要がある。また、強力な信頼度を有するAIないし機械学習アルゴリズムの分析に基づくブラックリストが、不当な差別又は偏見を生み出す危険性について配慮すべきである。

差し当たり、（納税者本人でなくとも）第三者機関によるチェック制度の創設、国税庁のAI利用における倫理指針などに関するものも含めて、アルゴリズムの生成・利用に際し、国税庁が遵守すべき義務に関する規定等の制定も視野に入れて、議論を深めるべきではないか。調査選定システムの利用について、定期的に影響評価を実施し、分析結果を公表することも重要である。

また、privacy by design⁽⁹⁴⁾に倣い、taxpayer's rights by design 又は individual rights by design の観点から初期設定を考える、いい換えれば、あらかじめ制度設計の段階で納税者の権利利益を保護する装置を組み込んでおくべきではないか。例えば、国税庁によるアルゴリズムの開発方針や運用指針、影響評価等の公開にとどまらず、納税者が国税庁の保有するメタデータを含む自身の情報、アルゴリズムによる判定結果、判定の際に考慮された事情やその重み付けなどにアクセス、修正又は削除を要求したり、場合によっては積極的に正しい情報を提供したり、合理的な推論⁽⁹⁵⁾がなされることを求めたりすることができるように、技術的及び制度的側面からの仕掛けを講じておくことが考えられる⁽⁹⁶⁾。アルゴリズムに検証可能性、可視性、透明性をもたせるとともに、法制度として納税者が国税庁の保有する情報にアクセス等する権利や合理的推論がなされる権利を整備するのである（もちろん、アクセスが制限されるべき情報の種類については議論の余地がある）。

(94) Privacy by design については、堀部政明ほか編著『プライバシー・バイ・デザイン』（日経BP社2012）、新保史生「プライバシー・バイ・デザイン」論研ジュリ18号16頁以下（2016）参照。なお、OECD, TAX ADMINISTRATION 3.0, *supra* note (1) では、compliance-by-design という考え方がちりばめられている。

(95) 合理的推論がなされる権利について、Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494 (2019) 参照。

(96) 行政機関の保有する個人情報の保護に関する法律27条には訂正請求権が定められているが、訂正の対象となる自身に関する情報が開示されない場合には、この訂正請求権は実効性を失うのではないかと、という問題がある。

この点は、手続的な公正性を確保する伝統的な手続的デュープロセスという観点のほか、ソースコードの公開や監査のための証拠の生成プログラムなどを含むテクノロジー時代における技術的側面からの公正性を確保する技術的デュープロセス⁽⁹⁷⁾という観点で捉えることも可能であろう⁽⁹⁸⁾。納税者の権利利益の保護に実効性をもたせるために、納税者の諸権利を法律で明確にすべきであろう。

アルゴリズムの調査や理解に必要とされる専門的知識や能力を通常の納税者又は税の専門家に期待することはできない。この点からしても、支援のための独立的な専門組織、第三者機関の設立が必要となろう。アルゴリズムに関して、国税庁が意思決定をする際に、あらかじめ同組織から情報を得たり、ピアレビューを受けたりするなど、監視とレビューの機能も有する組織として設計することや、そのようなプロセスの制度化などが考えられる⁽⁹⁹⁾。

V 結びに代えて

調査選定システムが、これまで人間が気付かなかった又は暗黙知とされてきたような着眼点や選定手法等を発見し、実装し、国税庁という組織に標準装備・共有されることで、組織全体の調査選定の精度と調査効率の向上、コスト削減という効果をもたらすことは、一般論として、国民の利益にもつながりうる。よって、最新のテクノロジーを税務行政のために活用する国税庁の先進的な姿勢や試みには基本的に賛同する。

また、テクノロジーの活用自体は、差し当たり、既存の租税制度ないし法体系・価値体系に大きな変革を要求するものではないかもしれない⁽¹⁰⁰⁾。AIないし機械学習アルゴリズムを、既存の憲法や行政法、あるいは租税法の制度にうまくはめ込むことはできると考えるが、だからといって、その使用が常に正当化されるというわけではない。とりわけ、国税庁は、行き当たりばったりの又は無責任な方法でこれらを利用するようなことはせず、慎重に行動し、その設計と実装において潜在的な落とし穴と懸念事項に配慮しなければならない⁽¹⁰¹⁾。

本稿では、国税庁が税務行政のスマート化やデジタル・トランスフォーメーションの利

(97) 参考として、Citron, *supra* note (79); Danielle K. Citron & Frank Pasqual, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014) 参照。これに対して、Thornton, *supra* note (84), at 1840 は、人間の偶発的なバイアスを緩和するために法的な自動意思決定システムを使用することのベネフィットは、技術的デュープロセスの支持者が提起するバイアスの懸念をはるかに凌ぐものであり、証明済みの対策によってその大部分を解消しうる、という見解を示す。なお、納税者情報の収集(取得)・利活用(データ照合/情報連携)・頒布などの取扱プロセスに対する適正手続(due process)、もう少し的確に言えば手続的データ適正手続の権利(right to procedural data due process)をどのように強化するかという問題意識を提示するものとして、石村・前掲注(10)63頁以下参照。

(98) 別の視点から、学習用データのサンプリングに問題があるのであれば、統計的な観点からの調査、通常の税務調査とは内容や負担の異なる情報収集のための調査の実施も検討に値する。

(99) See Coglianesi & Lehr, *supra* note (8), at 1190.

(100) 機械学習の行政利用は法的観点からは例外的なものではなく、確立された価値観を維持し、バランスを再調整するために、法律や法制度を体系的に改変する必要はないという見解として、Coglianesi & Lehr, *supra* note (8), at 1213 n.275 参照。

(101) Coglianesi & Lehr, *supra* note (8), at 1215.

点ばかりを強調する中で、AIないし機械学習アルゴリズムを実装した調査選定システムを利用することにより想定される法的問題、具体的には、学習用データ、バイアス、ブラックリスト、ブラックボックス、偽陰性と偽陽性等に関する問題を指摘した。納税者にのしかかる税務調査の心理的負担も併せ考慮すると、利点ばかりを強調してその利用が推し進められることのないよう十分に警戒しておかなければならない⁽¹⁰²⁾。少なくとも課税処分より前の調査選定の段階では過度に問題視すべきではないという見解もあるかもしれないが、これらは、場合によっては、個人の尊重、幸福追求権、平等原則を基本原理として掲げる憲法13条や14条、あるいは適正手続を保障する31条や84条等への抵触問題に発展する可能性があり、看過できない重要な問題である（もっとも、技術的な専門知識のない筆者の目には、バイアスやブラックボックスという特性がより強調されて映ってしまう可能性があることに留意⁽¹⁰³⁾）。租税に特化した文脈ではないものの、AI、機械学習アルゴリズム、ビッグデータ、データマイニング、プロファイリングなどが引き起こす憲法上の問題については、既に国内でも議論が蓄積しつつある⁽¹⁰⁴⁾。しかしながら、国税庁の公表資料をみる限り、国税庁が、税務行政のスマート化やデジタル・トランスフォーメーションを進めていくに当たり、これらの点に関してどのような見解を有しているのか、そもそも国税庁がかような問題意識を有しているのか、明らかではない。まったくの無思慮であるのか、意識的に沈黙しているのか、判然としない。上記法的問題への対応策その他については前記IV3で検討したとおりであるが、少なくとも、国税庁は、バイアス問題やブラックボックス問題への対応についてどのように考えているのか⁽¹⁰⁵⁾、偽陰性と偽陽性のどちらを重視する調査選定システムを構築するつもりなのか、人間による判断の介在や透明性の確保などについてどのように考えているのか、納税者保護のあり方をどのように考えているのか⁽¹⁰⁶⁾といった点に関して、国民に見解や方針を示すべきである⁽¹⁰⁷⁾。

さらにいえば、AIや機械学習アルゴリズムの利用に伴う法的問題に対する最もシンプルな解決策は、これらを利用しないことである。重要な法的問題を解決できないまま利用

(102)ただし、タックスコンプライアンス遵守のための活動が、下流から上流へとシフトするならば、税務調査の負担感や臨場調査それ自体の頻度が減る可能性がある。See OECD, TAX ADMINISTRATION 3.0, *supra* note (1). この場合には、別途異なる観点からの考察が必要である。

(103)See Lehr & Ohm, *supra* note (9), at 661-663.

(104)例えば、山本龍彦編著『AIと憲法』（日本経済新聞出版社2018）、山本・前掲注（11）321頁以下、山本・前掲注（22）「ロボット・AIは人間の尊厳を奪うか？」79頁以下、山本龍彦「基調講演 プロファイリング等自動処理による人間疎外的決定への対応」法とコンピュータ36号45頁以下（2018）など参照。

(105)IRSのホームページをみると、IRSは納税者の市民権を保護することにも尽力しているとし、年齢、肌の色、障害、人種、報復行為、国籍、英語力、宗教、性別、性的志向、又は親としての地位に基づく差別を容認しないし、これには、IRSの職員、コミュニティサイトのスタッフやボランティアとの接触も含まれるとしている。https://www.irs.gov/newsroom/taxpayer-bill-of-rights-1.この点に関して、国税庁はどのような方針であるのか。

(106)納税者権利憲章を中心とした国際的な議論の参考として、望月爾「EUにおける納税者の権利保護の調和」立命385号38頁以下（2019）、同「第4回納税者の権利に関する国際会議報告—デジタル化時代における納税者の権利—」税制研究77号144頁以下（2020）参照。

(107)国税庁は、税務行政におけるAIの利用を進めるに当たり、統合イノベーション戦略推進会議決定「人間中心のAI社会原則」（2019）と、AIネットワーク社会推進会議「報告書2019」（2019）の「AI利活用原則」に定められている各原則を遵守すべきであることについて、税研218号に小稿を掲載予定。

を推進することのコストは、これらを利用することにより得られる利点を凌ぐ可能性がある。重要な法的問題を解決できないのであれば、利用しないこと、実戦配備しないことも選択肢の 1 つになるはずである⁽¹⁰⁸⁾。納税者の権利利益への影響がない場面に限定して、利用を開始するようなことも選択肢に入るであろう。

例えば、フィンランドの税務当局は、AI 倫理規則を定めて、公表している⁽¹⁰⁹⁾。そこでは、AI がどのような観点から自らの戦略目標に貢献するのかということに加えて、要旨次の点が定められている。

- ・ AI の利用のあり方について責任と倫理をもつこと
- ・ AI は信頼できるデータのみを使用すること（データの信頼性と目的適合性が確信できるまで、AI がデータにアクセスできないようにすること。データが使用されている間、これらの点をモニタリングしておくこと。データやアルゴリズムに欠陥や歪みが生じた場合には、可能な限り迅速に修正すること）
- ・ AI の運用には必ず人間が責任をもつこと（職員は AI の提案を採用しないことができること。AI は、意思決定過程におけるステップを後からトレースして、正当化できない限り、意思決定を行うことはできないこと。AI の運用責任者を指名すること）
- ・ AI は法律及び規則に従っていること（AI がどのように利用されているかを常に監視・評価し、逸脱があった場合には即座に対処すること。AI の利用は、税務データの機密性や納税者のデータセキュリティを危険にさらすものではないこと。パートナーを慎重に、かつ、責任をもって選ぶこと。下請けチェーン全体の運営に責任をもつこと）
- ・ 責任ある倫理的な AI アプリケーションに関する公開討論に参加すること

ここで定められていることは、国税庁が上記で指摘した見解や方針を示す際に参考となるであろうし、むしろ見習うべき点が大いにある。

最後に 3 点述べておきたい。第 1 に、本稿における考察は、税務行政において調査選定以外で AI ないし機械学習アルゴリズムを利用する場合に対しても、程度の差こそあれ、当てはまる（犯則事件に関係する利用であれば、なおさらである）。今後、個別の論点ごとに具体的な研究を進める必要がある。

第 2 に、本稿は、納税者プライバシーや国税庁による情報収集・調査手法の観点からの研究など多くの宿題を残すものである。本稿における AI ないし機械学習アルゴリズムの行政利用がもたらす法的問題等に関する考察（裁判官と自動化バイアスの問題やブラックボックスと行政裁量の問題に関するものを含む）を踏まえて、税務調査に関する過去の裁判所の判断を振り返ってみると、荒川民商事件・最高裁昭和 48 年 7 月 10 日第三小法廷決定（刑集 27 卷 7 号 1205 頁）や川崎民商事件・最高裁昭和 47 年 11 月 22 日大法廷判決（刑集 26 卷 9 号 554 頁）などが示した判断枠組みが果たして時代に適しているか、という興

(108) 前掲 ウィスコンシン州最高裁との関係で、Wisser, *supra* note (22), at 1826 参照。

(109) <https://www.vero.fi/en/About-us/finnish-tax-administration/operations/responsibility/finnish-tax-administrations-ethical-principles-for-ai/>. この点については、研究会において、立命館大学法学部の望月爾教授からご教示いただいた。この場をお借りして、感謝申し上げます。

味深い問題視点もみえてくる。

第3に、前記Ⅳ3(2)において、人間の判断にも問題があるのだとすれば、むしろ、アルゴリズムは修正が効く可能性が高く、公正なアルゴリズムを組めば、人間のように偏見が入ることなく、公正な結論が導かれる、よってアルゴリズムによる判断を支持すべきであり、アルゴリズムの利用を進めるべきであるといった声上がる可能性があることを述べた。アルゴリズムが本稿で示したような問題を克服した場合、あるいは人間(税務職員)による判断に対する人間(納税者)の信頼が崩れた場合には、人間による判断を受ける権利(人間の関与を得る権利⁽¹¹⁰⁾)ではなく、アルゴリズムによる判断を受ける権利の制定等に関する議論が高まることが予想される⁽¹¹¹⁾。ただし、人間の側に知見や経験が蓄積されないような仕組みができるとすれば、予想していなかった弊害が生じるかもしれない。

(2021.5.20 受稿, 2021.6.25 受理)

(110) 参考として、GDPR22条1項参照。

(111) この点に関する議論の参考として、Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611 (2020) 参照。

〔抄 録〕

各国の租税行政庁と同じように、国税庁も、平成 29 年 6 月に「税務行政の将来像～スマート化を目指して～」を公表し、AI、データマイニング、機械学習などを利用して税務行政のスマート化（令和 3 年 6 月には税務行政のデジタル・トランスフォーメーション）を推進することを宣言している。本稿では、国税庁が税務行政のスマート化の利点ばかりを強調する中で、納税者の調査必要度やコンプライアンスリスクを測定し、税務調査の対象者を自動的に選定する調査選定システム（とりわけ大量のデータから有用な情報を抽出するデータマイニングを行い、機械学習により、納税者のコンプライアンスリスクのスコアをはじき出すモデルを生成するようなアルゴリズム）を国税庁が採用する場合に想定される法的問題を明らかにする。具体的には、学習用データ、バイアス、ブラックリスト、ブラックボックス、偽陰性と偽陽性等に関する問題が存在すること及びこれらが個人の尊重、幸福追求権、平等原則を基本原理として掲げる憲法 13 条や 14 条、あるいは適正手続を保障する 31 条や 84 条等への抵触問題に発展する可能性があることを指摘した上で、税務行政のスマート化やデジタル・トランスフォーメーションを進めていくに当たり、国税庁はこれらの点に関して国民に見解や方針を示すべきであると主張する。