Feature 1

プロファイリングに関する最終提言

令和4年4月22日

パーソナルデータ＋α研究会

I 3つの提言

パーソナルデータ＋α研究会は、平成30年5月より「プロファイリングに関する提言案作成タスクフォース」を設置し、プロファイリングに対する（「規制ありき」ではない）適切な規律のあり方について検討を重ねてきた。その議論を踏まえ、本研究会は、平成30年12月19日付け「プロファイリングに関する提言案」NBL 1137号64頁（以下「提言案」という）を公表して世に問い、広く関係者の意見を求めた。その結果、提言案に対していくつかの意見が寄せられたほか、企業においてプロファイリングに関するコンプライアンス体制等の整備に用いられ（**提言案①**参照）[1]、いくつかの業界団体で提言案を参照した自主的取組みがなされるほか（**提言案②**参照）[2]、政府関係の文書においても参照され（**提言案③**参照）[3]、提言案公表により一定の成果を挙げることができた。

もっとも、提言案の策定後、国内外においてプロファイリングをめぐるさまざまな問題が噴出するとともに、個人情報保護法も令和2年に大規模な法改正がなされ、令和3年に個人情報保護委員会のガイドラインの改正が進み、プロファイリングに関連する規律が導入された。このため、パーソナルデータ＋α研究会は、このような環境変化をも踏まえて提言案の再検討を行い、環境変化に応じた一定の修正を加えた上で、以下の**提言①～③**のとおり最終提言を行うこととした。

提言①　企　業：プロファイリングに関するコンプライアンス体制等の整備

与信、採用・人事等にプロファイリングを用いる企業は、プロファイリングの実施、活用等に当たって、法令との抵触を避けるためのコンプライアンス体制を整備すべきである。その際、プロファイリングが、プライバシーのみならず、平等や公正概念との関係でも問題になり得る点に留意すべきである。また、上記企業は、プロファイリングの実施・活用等に伴う社会的・倫理的責任に配慮した体制づくりに努めるべきである。

なお、本提言に付属する最終報告書では、体制整備を支援すべく、法令遵守にとどまらない「Ⅲ　社会的責任等に関する留意点」をまとめている。

提言②　業界団体等：プロファイリングに関する自主的な取組みの支援

業界団体・認定個人情報保護団体等は、各企業のコンプライアンスや自主的取組みを支援すべく、ガイドライン等の作成を検討することを推奨する。その際、プロファイリングの対象となる個々の利用者・生活者等（以下「被評価者」という）と企業の間に立って、利害関係の調整や課題解決を目指すべきである。

提言③　政　府：民間主導の自主的な取組みの尊重および必要な措置の検討

政府は、民間による自主的な取組みを尊重すべきである。その取組みの実効性を慎重に見定めながら、企業・業界団体・認定個人情報保護団体等との対話を続けるべきである。また、政府は、プロファイリングに関する適切な規律が実現されるように、民間の自主的取組みを尊重しながら、技術進展および社会状況変化も踏まえつつ、個人の権利利益の侵害リスクを抑止するために必要な措置を検討すべきである。

●パーソナルデータ＋α研究会 参加者（順不同）　（令和４年２月１日時点）

	（所　　属）
荒井ひろみ	理化学研究所革新知能統合研究センター 研究員
市川　芳治	慶應義塾大学大学院法務研究科非常勤講師
大島　義則	弁護士（長谷川法律事務所）・慶應義塾大学大学院法学研究科特任准教授
大屋　雄裕	慶應義塾大学法学部法律学科教授
楠　　正憲	デジタル庁統括官 デジタル社会共通機能グループ長
栗田　昌裕	名古屋大学大学院法学研究科教授
小林慎太郎	株式会社野村総合研究所 ICT メディアコンサルティング部 上級コンサルタント
酒井麻千子	東京大学大学院情報学環准教授
坂下　哲也	一般財団法人日本情報経済社会推進協会（JIPDEC）常務理事
佐久間　淳	筑波大学大学院システム情報工学研究科教授
宍戸　常寿	東京大学大学院法学政治学研究科教授
数藤　雅彦	弁護士（五常総合法律事務所）
寺田　麻佑	国際基督教大学教養学部上級准教授・理化学研究所革新知能統合研究センター客員研究員
中川　裕志	理化学研究所革新知能統合研究センター チームリーダー
中川隆太郎	弁護士（シティライツ法律事務所）
中西　崇文	武蔵野大学データサイエンス学部准教授
原田　大樹	京都大学大学院法学研究科教授
藤田　卓仙	慶應義塾大学医学部特任准教授
松尾　　陽	名古屋大学大学院法学研究科教授
松前　恵環	駒澤大学グローバル・メディア・スタディーズ学部講師
望月　健太	ニューヨーク州弁護士／ FIP, CIPM, CIPP/E, CIPP/US（法律事務所 LAB-01）
吉川　徳明	株式会社メルカリ 執行役員 VP of Public Policy
横田　明美	千葉大学大学院社会科学研究院准教授
（事務局）	
山本　龍彦	慶應義塾大学大学院法務研究科教授
生貝　直人	一橋大学大学院法学研究科准教授
石井夏生利	中央大学国際情報学部教授
工藤　郁子	大阪大学社会技術共創研究センター招へい教員
成原　　慧	九州大学法学研究院准教授
小林　史明	明治大学法学部専任講師
菅谷　麻衣	拓殖大学政経学部助教
西村　友海	大阪大学社会技術共創研究センター特任助教

※　参加者の所属組織の意見ではなく、各個人の知見を持ち寄っての提言である。
※　本提言の元となる提言案・最終報告書の作成過程においては、一般社団法人ピープルアナリティクス＆HRテクノロジー協会、一般財団法人日本情報経済社会推進協会、株式会社メルペイにヒアリングのご協力をいただいた。

Ⅱ 提言の趣旨：プロファイリングの健全な実装に向けて

「プロファイリング」は、一般にはいまだ耳慣れない言葉かもしれない。しかしこの技術は、確実に我々の生活の中に入り込んでいる。たとえば、あなたがスマートフォンなどで受信するインターネット広告の多くは、あなたのウェブ閲覧履歴や購買履歴

1　たとえば、経済産業省AI原則の実践の在り方に関する検討会AIガバナンス・ガイドラインWG「AI原則実践のためのガバナンス・ガイドライン ver. 1.1」（令和４年１月28日）23頁には、提言案付属中間報告書の「自主的取組みに関するチェックリスト」の企業における活用方法が記載されている。

2　一般社団法人ピープルアナリティクス＆HRテクノロジー協会「人事データ利活用原則」（2020年３月19日）、World Economic Forum, “Good Data: Sharing Data and Fostering Public Trust and Willingness”, White Paper, 2021, pp.15-17,〈http://www3.weforum.org/docs/WEF_Good_Data_Sharing_Data_and_Fostering_2021.pdf〉（本稿記載のURLの最終アクセス日は2022年１月５日。以下同じ）.

3　シェアリングエコノミー検討会議・内閣官房情報通信技術（IT）総合戦略室「シェアリングエコノミー検討会議第２次報告書－共助と共創を基調としたイノベーションサイクルの構築に向けて－」（2019年５月）37頁・注21、内閣官房情報通信技術（IT）総合戦略室「デジタル社会構築タスクフォース～データ環境整備の方向性についてのとりまとめ～」（令和２年６月26日）30頁、経済産業省AI原則の実践の在り方に関する検討会「我が国のAIガバナンスの在り方 ver. 1.1」（令和３年７月９日）27頁・注80、AI原則の実践の在り方に関する検討会AIガバナンス・ガイドラインWG・前掲注(1) 10頁・注5、23頁等。

などからあなたの趣味嗜好やライフスタイルが「プロファイリング」され、そこで得られた予測結果を踏まえて個別的に送られてくるものである（いわゆる「ターゲティング広告」）。また、あなたがスマートフォンなどで受信するオンライン・ニュースの多くにも「プロファイリング」が介在している。スマートフォンを通じて日常的に読むニュースの多くは、「プロファイリング」によってユーザーの趣味嗜好や政治的傾向にマッチすると予測され、この結果に基づいて配信されたニュースなのである。こうして見ると、「プロファイリング」なるアルゴリズムを用いた予測技術は、すでに我々の生活の一部になっているといっても過言ではないだろう（「ユビキタス化するプロファイリング」）。

2018年5月に運用が開始された欧州連合（European Union, EU）の「一般データ保護規則（General Data Protection Regulation, GDPR）」は、このような「プロファイリング」を、「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」と定義した（4条(4)号）[4]。本提言では、わが国においていまだその定義に関するコンセンサスが形成されていないことを踏まえ、さしあたりこれを、「パーソナルデータとアルゴリズムを用いて、特定個人の趣味嗜好、能力、信用力、知性、振舞いなどを分析又は予測すること」と捉えておきたい。

もっとも、本提言では、かようなプロファイリングの中でも、個人の自由で主体的な生活に重要な影響を与え得るプロファイリングに焦点を当てている。それは、ここ数年、この技術が、マーケティング等の領域を超えて、与信、採用・人事、教育、行政の許認可などの場面でも積極的に利用されつつあるからである。たとえば中国では、アリババ・グループ傘下の芝麻信用が、アルゴリズムを用いて個人の社会的信用力を950点満点でスコア化し、このスコアが社会において広く共有、利用されるに至っている。信用力に関するプロファイリングの結果（社会的信用スコア）が、与信場面のみならず、プラットフォームを通じた取引（シェアリングエコノミーなど）や行政手続、さらには「婚活」といったさまざまな生活場面で利用され、中国人民の「第二の身分証」としての意味すら持ち始めているのである[5]。言うまでもなく、このような「Scoring Society」[6]では、プロファイリングの結果が個人の実生活に対して決定的に重要な意味をもつことになる。

こうした社会状況は、日本においても決して「他人事」ではない。日本でも、みずほ銀行とソフトバンクが共同開発した社会信用スコアリング・サービス（J.Score）が与信領域においてすでに運用を開始しているほか[7]、LINE[8]やNTTドコモ[9]といったリーディング・カンパニーも、同様のスコアリング・サービスを開始しており、マーケティング領域を超えたプロファイリングの利活用が進みつつある。また企業の採用活動や労務管理の場面でも、人工知能（Artificial Intelligence, AI）を用いた求職者や従業員の適性プロファイリングが急速に実装されつつある。こうして見ると、わが国も、プロファイリング技術を基礎とした「Scoring Society」に向けて歩みを進め始めており、プロファイリング結果（スコア）が個人の人生に及ぼす影響が以前にも増して大きくなってきているといえよう。

このことは同時に、企業が個人の社会的信用力等をプロファイリングすることに伴う社会的な責任が重くなってきているということを意味する。企業が、本来は考慮に入れるべきではない個人情報を用いて不公正なプロファイリングを行ったり、偏った、あるいはバイアスを含んだデータセットを用いて不公正なプロファイリングを行ったりした場合、当該個人の人生を大きく狂わせる可能性がある。また、プロファイリングの過程が著しく不透明であったり、不公正ないし不正確なプロファイリングに対して被評価者が異議を申し立てるような手続がなかったりすれば、個人が自由で主体的な生活を送ることは難しくなるだろう。この点で、プロファイリングを行う目的や領域によっては、企業がプロファイリングを実施するに当たって、一定の慎重さが求められるように思われる。実際、「Yahoo!スコア」については信用スコアリング事業開始後に、デフォルト仕様で信用スコアを算出する仕組み等について批判が集中し、最終的にサービス終了に至った[10]。また、内定辞退率スコアを算出し、顧客企業に提供するサービスが問題とされたリクナビ事案では、2度の個人情報保護委員会による勧告・指導と2度の

東京労働局による指導がなされるに至った[11]。

このような、「Yahoo!スコア」やリクナビ問題を経てわが国においてもプロファイリングの規律に関する議論が高まり、令和2年個人情報保護法改正および個人情報保護委員会のガイドラインの改正によりプロファイリングに関連する規律が導入されたが、この規律がプロファイリングにどのように適用されるかは必ずしも明確ではない。また、現状わが国において、どのような点に注意すればプロファイリングを実施している（または実施しようとしている）企業が社会的な非難（いわゆる炎上）を回避できるのか、また、何を行えば上述のような社会的な責任を果たしたといえるのかを示す具体的な指針もない。このことが企業にプロファイリングの実施に対する萎縮的な効果を与え、健全な「Scoring Society」への歩みを妨げているようにも思われる。他方で、プロファイリングは、今後の技術変革が見込まれる上、一部企業にとっては大きなビジネスチャンスにもなっている。企業等は個人情報保護法またはこれに関連する規律にも留意した上で、社会的責任を果たすための自主的な規律が求められているといえよう。本研究会は、このような認識の下で、前掲のとおり3つの提言をまとめた次第である。

本提言には、**最終報告書**が付されている。

本研究会は、この最終報告書の「**Ⅲ　社会的責任等に関する留意点**」において、企業が法令遵守の観点から留意すべき点のほか（前述のとおりプロファイリングに関連する一定の規律が設けられた）、不用意なプロファイリングの実施により社会的に非難され（いわゆる炎上）、レピュテーションを落とすことを防止するために留意すべき点を「**チェックリスト**」としてまとめた。2018年にはアマゾンが2014年頃から開発してきた人事採用プロファイリング（アルゴリズムを用いた採用プログラム）に女性差別の傾向が見られたことが大きく報じられたが[12]、仮にこのような採用活動が法的に違法の評価を受けないとしても、基本的人権への尊重を欠くものとして、企業の社会的信用を大きく傷つけることになり得る。そこで本研究会は、EUの動向なども踏まえつつ、プライバシーや平等といった基本的人権・憲法的価値に配慮したプロファイリングの実施・利用方法をチェックポイントとしてリスト化し、企業等の便宜に供しようと考えた。

なお、本提言は企業等の便宜を図るべく自主的な規律についての一定の方針や着眼点を述べるものであり、各組織において、本提言や最終報告書の内容を踏まえながら最終的なプロファイリングに対する規律を自主的に検討・構築する必要がある。リスト化されたチェックリストは当該リストをチェックすれば足りるというものではなく、また当該リストをすべて義務的にチェックしなければならないものでもない。むしろ各組織には、機械的・硬直的な判断から解放された柔軟かつ原理的なプロファイリング対応が求められている点に留意されたい。

4 GDPRの日本語訳に関しては、個人情報保護委員会ウェブサイト〈https://www.ppc.go.jp/files/pdf/gdpr-provisions-ja.pdf〉を参照した。
5 田中信彦「中国リポート　全国民の信用情報を政府当局が一元的に管理」週刊東洋経済6829号（2018）32〜33頁。田中によれば、中国政府は、このような信用スコアと治安維持システムとを結合させた全国民的な信用情報ネットワークの構築を進めている。実際、中国政府の国家発展改革委員会は、民間の信用情報機関と「共同賞罰措置の実施に関する覚書（MOC）」を締結し、信用情報と連動した賞罰メカニズムを構築している。
6 *See* Danielle Keats Citron & Frank A. Pasquale, The Scored Society: Due Process for Automated Predictions, 89 WASH. L. REV. 1（2014）.
7 J.Scoreウェブサイト〈https://www.jscore.co.jp/〉
8 LINE株式会社ウェブサイト〈https://line.me/ja/〉
9 NTTドコモウェブサイト〈https://www.nttdocomo.co.jp/info/news_release/2019/08/29_00.html〉
10 事案の概要につき、第二東京弁護士会情報公開・個人情報保護委員会編『令和2年改正　個人情報保護法の実務対応』（新日本法規、2021）290〜292頁〔奥田正夫〕。
11 事案の概要につき、第二東京弁護士会情報公開・個人情報保護委員会編・前掲注⑽ 252〜270頁〔白石紘一〕。
12 Reutersウェブサイト〈https://jp.reuters.com/article/amazon-jobs-ai-analysis-idJPKCN1ML0DN〉

プロファイリングに関する最終提言付属最終報告書

I プロファイリングの適切な規律を目指して

ICT技術の進展に伴い、いわゆる「プロファイリング」を法的・技術的にいかに規律していくべきかが、欧米を中心に大きな社会問題となっている。

たとえば、平成30年5月25日に適用開始されたEU一般データ保護規則（General Data Protection Regulation, GDPR）は、異議を述べる権利（21条）、プロファイリングを含む個人に対する自動化された意思決定（22条）等のプロファイリングに関する規定を置いている[1]。GDPRは日本に対しても域外適用され得るため（3条2項）、EU向けにサービスを提供する日本企業もプロファイリングに関する法的規制を受け得ることになる[2]。

わが国では、平成27年個人情報保護法改正ではプロファイリングに関する改正は見送られ、プロファイリングの問題は「民間主導の自主的な取組」等を勘案して検討していくこととなったが[3]、令和2年個人情報保護法改正に際して、一定のプロファイリングに関連する規律が設けられた。国会審議においても、プロファイリングの規律について議論がなされ、「いわゆるプロファイリングにつきましては、今回の改正で、利用停止、消去の要件の緩和、不適正利用の禁止、第三者提供の開示、提供先において個人データとなることが想定される情報の本人同意といった規律を導入した」との答弁がされている[4]。すなわち、利用停止・消去の要件緩和（令和3年改正反映後の個人情報保護法35条1項、5項、6項。以下、同法の記載については、デジタル社会の形成を図るための関係法律の整備に関する法律第51条関係改正後の条文表記による）、不適正利用の禁止（19条）、第三者提供記録の開示義務化（33条5項）、個人データとなる個人関連情報の規制（31条）の導入により、プロファイリングに関連する規定が設けられた。さらに、令和3年8月、個人情報保護委員会は、「個人情報の保護に関する法律についてのガイドライン（通則編）」（以下「GL通則編」という）3-1-1（利用目的の特定）を改正し、「例えば、本人から得た情報から、本人に関する行動・関心等の情報を分析する場合、個人情報取扱事業者は、どのような取扱いが行われているかを本人が予測・想定できる程度に利用目的を特定しなければならない」との解釈を追記し、「本人から得た情報から、行動・関心等の情報を分析する場合に具体的に利用目的を特定している事例」の事例2として「取得した行動履歴等の情報を分析し、信用スコアを算出した上で、当該スコアを第三者へ提供いたします」という具体例を挙げた[5]。

このようにわが国の個人情報保護法においてもプロファイリングに対する一定の規律が設けられたものの、たとえばいかなるプロファイリングが不適正利用の禁止（19条）の対象とされるかは、必ずしも明らかにはなっておらず、法規制の内容は極めて曖昧であり、個人情報取扱いの適正性を確保する措置・体制を各組織において検討していく必要がある。また、法規制以外の事項についても、各組織が社会的責任を果たすためには、引き続き自主的な取組みを継続していく必要がある。

特に近年、利用可能なデータが増大しディープラーニング技術等が進展したことに伴ってAI（Artificial Intelligence）が注目されており、機械的・自動的処理によるプロファイリングの社会的重要性が増してきている。そのため、プライバシーや公正性に配慮したプロファイリングを推進すべく、当研究会では「プロファイリングに関する最終提言」を作成した。

本最終報告書は、第一の提言である「企業：プロファイリングに関するコンプライアンス体制等の整備」を促進すべく、プロファイリングをめぐる論点や留意点を整理するものである。以下では、「**II プロファイリングの効用とリスク**」について概観し

た上で、「Ⅲ　社会的責任等に関する留意点」をまとめた。

Ⅱ　プロファイリングの効用とリスク

1.　プロファイリングの効用

近年、機械的・自動的処理によるプロファイリングに、大きな社会的期待が寄せられている。効率的・効果的に個人を評価したい企業は言うまでもなく、プロファイリングの対象となる個々の利用者・生活者等（以下「被評価者」という）にとっても、プロファイリングが有益な場合がある。たとえば、金融分野においては、従来の審査方法であれば与信を受けられなかった個人事業主等が与信を受けられるようになると期待されている。また、人事分野においては、応募者たる被評価者の素性を隠して偏見・憶測を回避できたり、能力の適正評価により「足切り」されずに審査を受けたりできると言われている。

プロファイリングは、オンライン広告配信やマーケティングだけでなく、金融、ヒューマンリソース、ヘルスケア、保険等の多くの分野で利用されるようになってきており、適切に活用されれば、個人の可能性を広げつつ、効率性の増大やリソースの節約といった社会的効用を増大させることができる。

2.　プロファイリングのリスク

他方、プロファイリングは主としてプライバシー権、平等原則、民主主義との関係で重要な問題を内包しており、これらのリスクに適切に対処する必要がある[6]。

なお、憲法の人権規定が私人間に適用されるかについては議論されているものの、通説では立法または解釈を通じて間接的に私人である民間企業の行為にも適用されると解釈されており（間接適用説）、民間企業といえども憲法と無縁ではない[7]。

(1)　プライバシー権

憲法13条前段は個人の尊重原理を定めており、これを受けて同条後段は「生命、自由及び幸福追求に対する国民の権利」(幸福追求権）を保障している。そして、幸福追求権の一環としてプライバシー権が保障されると解するのがわが国の通説である[8]。

平成27年改正により、個人情報保護法は「本人の人種、信条、社会的身分、病歴、犯罪の経歴、犯罪により害を被った事実その他本人に対する不当な差別、偏見その他の不利益が生じないようにその取扱いに特に配慮を要するものとして政令で定める記述等が含まれる個人情報」を「要配慮個人情報」(同法2条3項）と定義し、取得時および第三者提供時について原則的に事前同意を要する規制（オプトイン方式）を導入した（同法20条2項、27条2項)。

もっとも、要配慮個人情報ではない個人情報を取

1 *See* Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP251 rev.01（Adopted on 3 October 2017, as last revised and adopted on 6 February 2018).

2 *See* European Data Protection Board, Guidelines 3/2018 on the territorial scope of the GDPR（Article 3）– Version 2.1（adopted on 12 November 2019, as last modified on 7 January 2020).

3 高度情報通信ネットワーク社会推進戦略本部「パーソナルデータの利活用に関する制度改正大綱」（平成26年6月24日)。十分性認定に関してEDPB（European Data Protection Board）は、日本法には自動意思決定およびプロファイリングに関する一般的ルールが存在しないことを指摘し、関連する事例を監視するよう欧州委員会に勧告した。EDPR, Opinion 28/2018 regarding the European Commission Draft Implementing Decision on the adequate protection of personal data in Japan（adopted on 5 December 2018). その後、令和2年個人情報保護法改正で一定の法的手当てがなされた。

4 第201回国会参議院・内閣委員会会議録13号其田真理政府参考人答弁（令和2年6月4日)。

5 GL通則編3-1-1。

6 プロファイリングのリスクについて、①プライバシー権侵害、②不当な差別・選別、③行動の萎縮効果、④民主主義への影響、⑤自己決定権への介入の5つの問題を指摘するものとして、山本龍彦「プロファイリングと情報銀行」情報信託機能の認定スキームの在り方に関する検討会（第19回）資料19-3・2〜4頁〈https://www.soumu.go.jp/main_content/000764110.pdf〉。本最終報告書では、このうち①②④を特に検討しているが、他の問題の重要性を否定するものではない。また、プラットフォーム時代において変容するプライバシーの理念に関連する諸価値として、「人格」、「平等」、「民主主義」、「信認」、「公正な競争」を検討するものとして、大島義則「プラットフォーム時代におけるパーソナルデータ法制の理念と設計」NBL1181号（2020）22頁。

7 山本龍彦「はじめに」同編著『AIと憲法』（日本経済新聞出版社、2018）7〜8頁。

8 芦部信喜著（高橋和之補訂）『憲法〔第7版〕』（岩波書店、2019）123頁以下等。なお、近年、プライバシー権と個人情報保護法との関係性については議論が活発化している。曽我部真裕「自己情報コントロールは基本権か？」憲法研究3号（2018）71頁、山本龍彦「自己情報コントロール権について」憲法研究4号（2019）43頁、曽我部真裕＝山本龍彦「【誌上対談】自己情報コントロール権をめぐって」情報法制研究7号（2020）128頁参照。プライバシー権と個人情報保護とを「いったん独立の権利利益の問題として観念」する見解として西村あさひ法律事務所編・太田洋ほか編著『個人情報保護法制大全』（商事法務、2020）28頁〔宍戸常寿〕。

得し、当該個人情報からプロファイリングにより「本人の人種」等の要配慮個人情報に相当する情報を推知することは禁止されていない。プロファイリングを用いた場合には、妊娠、精神病、政治的信条等のセンシティブ情報を分析により獲得することも可能である。このように、本人の同意なしにプロファイリングによって要配慮個人情報を新たに生み出す場合、プライバシー侵害を生じさせる危険性がある[9]。わが国では視聴履歴に関して、「放送受信者等の個人情報保護に関するガイドライン」（平成29年4月27日総務省告示第159号）34条が、「受信者情報取扱事業者は、視聴履歴を取り扱うに当たっては、要配慮個人情報を推知し、又は第三者に推知させることのないよう注意しなければならない」と規定し、視聴履歴に基づき要配慮個人情報を推知することについて注意喚起を行っている。

プライバシー侵害がある場合、民事上の不法行為（民法709条）に基づく損害賠償請求も可能である。個人情報保護法違反と不法行為の関係について両者の関係は峻別され、個人情報保護法違反は直ちに不法行為ではないと解されるものの、個人情報保護法違反が同時にプライバシー侵害になる場合には不法行為に基づく損害賠償請求が認められることがある[10]。特に要配慮個人情報の同意なき取得、個人情報の目的外使用について、不法行為法上の違法性を認める見解も出てきている[11]。また、プライバシー侵害がある場合、人格権に基づく民事差止請求ができる場合もある[12]。

(2) 平等原則

要配慮個人情報の定義条項である個人情報保護法2条3項が「人種、信条、性別、社会的身分又は門地」の差別を禁止する憲法14条1項後段を参照して考案された経緯からすれば、上記「(1) プライバシー権」で述べたプロファイリングのリスクは、プライバシー権侵害リスクとともに平等原則違反のリスクも内包している[13]。また、令和3年8月、個人情報保護委員会はGL通則編を改正し、不適正利用禁止義務違反となる事例として「採用選考を通じて個人情報を取得した事業者が、性別、国籍等の特定の属性のみにより、正当な理由なく本人に対する違法な差別的取扱いを行うために、個人情報を利用する場合」を挙げるに至っており、個人情報保護法においても差別リスクが問題になり得ることが明らかになった。

個別のプロファイリングの差別リスクのみならず、プロファイリングは社会構造としても差別を再生産する可能性を秘めている[14]。たとえば採用や融資の局面においてAIを用いた予測評価システムが導入された場合、当該システムを利用しているあらゆる組織から排除され続けて差別が再生産される構造が形成される危険性がある。また、プロファイリングの内容を示すことが、特定のカテゴリーに属する人々に劣等なレッテルを貼るという危害を生み出すという懸念も存在する[15]。

このようなAIを用いたプロファイリングによる差別などの問題を念頭に置いて、総務省のAIネットワーク社会推進会議が2017年7月に公表した「国際的な議論のためのAI開発ガイドライン案」では、「開発者は、採用する技術の特性に照らし可能な範囲で、AIシステムの学習データに含まれる偏見などに起因して不当な差別が生じないよう所要の措置を講ずるよう努めることが望ましい」（⑦倫理の原則）としており、同会議が2019年8月に公表した「AI利活用ガイドライン」では「AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービスの判断にバイアスが含まれる可能性があることに留意し、また、AIシステム又はAIサービスの判断によって個人及び集団が不当に差別されないよう配慮する」としている（⑧公平性の原則）。

(3) 民主主義原理

わが国の憲法は議会制民主主義を採用しているが（前文、1条、15条1項・3項、41条、42条、43条1項）、議会制民主主義は、選挙権を行使する際の判断材料である「情報を自由かつ均等に取得して、選挙権を適切に行使することができることを前提としている」（最大判平成19・6・13民集61巻4号1617頁〔泉徳治裁判官反対意見〕）。プロファイリングはこうした議会制民主主義の機能条件である情報環境を掘り崩すリスクがある。

現実に、米国をはじめとした諸外国では、プロファイリング等の技術を用いた政治マーケティングの手法が洗練されてきており、議会制民主主義の機能条件である個人の政治的意思決定が誘導または操作されるリスクがあることが指摘されている。ビッグデータ分析を専門とするケンブリッジ・アナリ

ティカ社がFacebook上の性格診断アプリを利用した27万人とその「友達」の合計8700万人のパーソナルデータを不正に収集し、当該データが2016年の米国大統領選挙においてトランプ陣営の政治マーケティングに利用された疑惑が発生したことは記憶に新しい（ケンブリッジ・アナリティカ事件）[16]。この事件における政治的意思決定への影響については未解明な部分が多いが、ビッグデータの利活用に際して民主主義との緊張が生じることは確かであろう。

3. 小　括

以上のようにプロファイリングには効用とリスクの両面があり、適切な規律の下で運用されていくべきであろう。

Ⅲ 社会的責任等に関する留意点

1. チェックリストの意義

先述のとおり、令和2年個人情報保護法改正およびガイドラインの改正により一定のプロファイリングに関連する規律が導入されるに至ったため、個人情報取扱いの適正性を確保する措置・体制を各組織において検討しておく必要がある。また、不用意なプロファイリングの実施等が社会的・倫理的問題を提起した事例も現に生じており、企業のレピュテーションを維持するためにも社会的・倫理的問題を検討しておく必要がある。この点で、プロファイリングを行う（または行おうとする）企業が、その実施について一定の自主的取組みをすることは、レピュテーション・マネジメントの観点からも有益といえよう。Google[17]やソニーグループ[18]などのリーディング・カンパニーが、AIの倫理原則等を公にしていることも、その表れである。よりグローバルな観点に立っても、EUのGDPRがグッドプラクティスを推奨し[19]、米国のビッグデータレポート[20]がプロファイリングについて社会的・倫理的問題にまで立ち入った検討を行っていることが注目されよう。さらに、第40回データ保護プライバシー・コミッショナー国際会議において、AI開発の際の指導原則につき、「AIにおける倫理及びデータ保護についての宣言」（2018年10月23日）[21]が公表されたことも軽視できない（以下「AI指導原則」という）。

そこで以下では、プロファイリングやビッグデータをめぐる事例等を参照し、プロファイリングの実

9 プロファイリングが個人情報保護法20条2項（令和3年改正前の17条2項）の「取得」に該当するかについて肯定論を採るものとして山本龍彦『プライバシーの権利を考える』（信山社出版、2017）266頁、否定論を採るものとして宇賀克也＝藤原静雄＝山本和徳「《鼎談》個人情報保護法改正の意義と課題」行政法研究13号（2016）11頁〔藤原静雄発言〕。宇賀克也『個人情報保護法の逐条解説〔第6版〕』（有斐閣、2018）143頁も「本人の同意なしにプロファイリングによって要配慮個人情報を新たに生み出すことは、要配慮個人情報の『取得』に当たると解するべきかという重要な解釈問題が存在する」と指摘する。プライバシーの権利とAIとの関係に関する一般論については山本龍彦「AIと個人の尊重、プライバシー」同編著・前掲注(7)79頁以下。なお、これらの議論は令和2年・3年個人情報保護法改正前のものであるため、改正後における同法20条2項の解釈・運用については、今後注視する必要がある。近年では、要配慮プロファイリングと一般的プロファイリングを分けた上で、要配慮プロファイリングの取扱いについて(i)禁止カテゴリーの創設（政治的信条の予測等）、(ii)使用・提供禁止データの創設（遺伝情報等）、(iii)厳格な同意取得、(iv)説明責任・透明性の徹底、一般的プロファイリングについてプロファイリングの有無と目的＋例示の義務、という叩き台を打ち出す見解もある（山本・前掲注(6)10〜11頁、14頁）。

10 峻別説とその帰結について、松尾剛行『最新判例にみるインターネット上のプライバシー・個人情報保護の理論と実務』（勁草書房、2017）244〜246頁。

11 窪田充見編『新注釈民法(15)債権(8)』（有斐閣、2017）539〜542頁〔水野謙〕、千葉惠美子「判批」重判平成29年度（ジュリ臨増1518号）（2018）78頁。山本・前掲注(6)2頁。

12「石に泳ぐ魚」事件・最三判平成14・9・24集民207号243頁等。

13 山本・前掲注(6)3〜4頁は、不当な差別・選別の問題を(a)ブラックボックス問題、(b)確率の評価と自動バイアス、(c)不適正なデータの混入可能性と、検証困難性、(d)差別の再生産問題、(e)バーチャル・スラム問題に整理している。

14 山本龍彦『おそろしいビッグデータ――超類型化AI社会のリスク』（朝日新聞出版、2017）第3章。

15 *See* Mathias Risse and Richard Zeckhauser, Racial Profiling, 32(2) PHILOSOPHY & PUBLIC Affairs 131, 146（2004）. なおプロファイリングの是非については、いわゆる9.11以降にさまざまな検討が加えられるようになっており、本論文もそのような文脈を意識して執筆されている。

16 工藤郁子「AIと選挙制度」山本編著・前掲注(7)330頁。

17 Sundar Pichai “AI at Google: our principles”〈https://www.blog.google/technology/ai/ai-principles/〉

18 ソニーグループ株式会社「ソニーグループAI倫理ガイドライン（2021年4月1日改訂）」〈https://www.sony.com/ja/SonyInfo/csr_report/humanrights/AI_Engagement_within_Sony_Group_Ja.pdf〉

19 *Supra* note 1, pp.31-32.

20 Edith Ramirez et al., “Big Data: A Tool for Inclusion or Exclusion?”, FTC Report, 2016.1. Federal Trade Commission Website〈https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf〉.

21 個人情報保護委員会による仮訳として、「AIにおける倫理及びデータ保護についての宣言　第40回データ保護プライバシー・コミッショナー国際会議　2018年10月23日（火）ブリュッセル」〈https://www.ppc.go.jp/files/pdf/181114_shiryou2-2.pdf〉。

施に際して企業が法的・社会的責任を果たすための留意点（「自主的取組みに関するチェックリスト」）を抽出し、現在プロファイリングを実施している企業、今後実施しようと検討している企業等の便宜に供することとした。このリストではプロファイリングの工程を、(1)企画・設計、(2)データの取得、(3)データの加工・分析・学習、(4)実装・運用・評価、という4段階に区分し、企業が考慮すべき度合い（優先度）を3段階に分けて表示した（★★★>★★☆>★☆☆の順に優先度が高い）。

なお、言うまでもなく、自主的取組みに関するチェックリストすべてを遵守する必要はない。以下で留意点を縷々記述しているが、最も重要な視点は、「プロファイリングを行うことによって何を実現したいのか」という意図・姿勢を示して、ビジネス上の信頼を獲得することである。

2. 自主的取組みに関するチェックリスト

(1) 企画・設計段階

(1-1) プロファイリング導入の適否判断 ★☆☆

企画・設計の段階で、プロファイリングの導入により実現される価値・利益を明確に整理しておくこと

第一に、プロファイリングを導入する目的・動機・利益を明確にすることが望ましい。動作の安定性、経済的合理性、サンプル数などの理由で、プロファイリング以外の手法を利用するほうが有効な場合もあるからである（ケースA）。本研究会では、特に、AIの導入自体が目的化してしまっていないかを企業が慎重に検討することを推奨したい。

ケースA　ワシントンDCの教師評価ツール

ワシントンDC市は、「IMPACT」という教師評価ツールを導入し[22]、2009～2010年の学年度末に、評価スコアが学区内の下位2%に入る教師を全員解雇した。翌年にはさらに下位5%相当の206名を解雇した。IMPACTでは、数学と国語（英語）の教育能力を測定するために、生徒の学力の向上に対する当該教師の貢献度を算出する「付加価値モデル」によるアルゴリズムを採用していた。当該アルゴリズムによる評価スコアが評価全体の半分近くを占めており、学校経営陣や地域コミュニティから高評価を受けていたとしても、解雇される場合もあった。しかし、生徒の成績は数多くの変数が関係する上に、1年度かつ25～30名分のテスト結果ではサンプル数が少なすぎること、解雇によりシステムのエラーフィードバックがとれずアルゴリズムの微調整ができないことなどから、「付加価値モデル」の有効性を疑問視する声もある[23]。

第二に、被評価者に提供できる価値・利益についても組織内で明確にしておくことが望ましい。プロファイリングを導入する者（たとえば採用担当者）と被評価者（たとえば入社希望者）の利益は必ずしも一致しないためである。被評価者の価値・利益として、たとえば、金融分野において従来の審査方法であれば与信を受けられなかった者が、プロファイリングによって与信を受けられるようになること、人事分野において、これまで偏見等によって類型的に不採用とされてきた少数派が偏見を回避し、採用の可能性が新たに拓かれることなどが想定される[24]。なおリクナビ事案に照らすと、著しい背信性のあるスキーム設計はそもそもデータ主体の同意を得たとしても採用し得ない可能性がある点に留意しなければならない（ケースB）。

ケースB　リクナビ事案

2019年、株式会社リクルートキャリアが、就職情報サイト「リクナビ」に登録した学生のサイト閲覧履歴等をもとに、学生ごとの内定辞退率（選考離脱や内定辞退の可能性を示すスコア）を判定したデータを作成して、採用企業側に販売していたことが発覚した。学生の就職活動を応援するサービスを運営しつつ、その裏側で内定辞退率を作成し、企業に提供するという点において背信性を有するビジネスモデル[25]は、多くの批判を受けた。

(1-2) プロファイリングの実施の有無・利用目的の明確化 ★★★

どのようなプロファイリングを、どのような目的のために実施しているかを明確にすること

プロファイリングを実装した段階における通知・公表を想定すると、それらは、①そもそもプロファイリングを実施しているか否か、②実施している場合に、いかなる種別・内容のプロファイリングを実施しているか、の2段階に分けられる。本研究会は、これら①②を明示するとともに、かかるプロファイリングの実施目的を明確化することを推奨する（ケースC参照）。令和2年個人情報保護法改正を踏まえて、GL通則編3-1-1（利用目的の特定）が令和3年8月に改正され、利用目的の特定に関して、「例えば、本人から得た情報から、本人に関する行動・関心等の情報を分析する場合、個人情報取扱事業者は、どのような取扱いが行われているかを本人が予測・想定できる程度に利用目的を特定しなければな

らない」との規律が明確化されたところであり、本研究会の提言は、改正法の趣旨にも合致すると考える。

また、上記①②のほか、利用態様の例示（閲覧履歴、購買履歴等から性別・年代を分析・予測している等）を付記することにより、利用目的に関する本人の予測・想定を容易にする措置を採ることも検討に値する[26]。

ケースC　Yahoo! スコア事案

2019年、ヤフー株式会社は、Yahoo! JAPAN IDのユーザーの行動等（たとえば「ヤフオク!」の取引実績、「Yahoo! 知恵袋」での活躍度、宿泊・飲食店の予約キャンセル率、Yahoo! ショッピングの利用金額等）を分析し、ユーザーの信用度を数値化するサービス「Yahoo! スコア」を発表した。しかし、その説明としては、「お客様のサービス利用状況等に応じた特典等の付与、審査プロセスの簡略化、コンテンツ最適化、サービスの改善、広告の配信等のため」に作成されるとの利用目的しか記載されていなかった[27]。

(1-3)　エシックス・バイ・デザインの導入検討 ★★☆

法的・倫理的なリスクを抑制する設計を検討すること

本研究会は、プロファイリングの開発・設計段階において法的・倫理的な検討をすることは開発者または企業にとって負担になる面もあるが、他方で、法的・倫理的なリスクを管理し、社会的信頼を獲得できるという面もあると考える。近年では、法的・倫理的な観点に配慮した「デザイン」の規格化も検討されている（ケースD）。なお、AI指導原則4条でも、プライバシー・バイ・デフォルト（Privacy by Default）やプライバシー・バイ・デザイン（Privacy by Design）[28]を含む「エシックス・バイ・デザイン（Ethics by Design）」により、アルゴリズム等が責任をもってデザインされ、開発されるべきとの指摘がある。

ケースD　IEEE「倫理的に調和された設計」

学術団体であるIEEEのグローバル・イニシアティブは「倫理的に調和された設計（Ethically Aligned Design）」という報告書を公開した。報告書の著者らを中心として、「標準規格（P7000シリーズ）」の策定が進められている[29]。同規格には、データプライバシーの処理（P7002）、アルゴリズム上のバイアスに関する考察（P7003）、子供と学生のデータガバナンスに関する標準（P7004）、透明性のある雇用者のデータガバナンスに関する標準（P7005）、機械可読のパーソナルプライバシー条項の標準（P7012）など、プロファイリングと関連が深いものもある。

(1-4)　リスクベース・アプローチの検討 ★★☆

法的・倫理的なリスクに応じた体制・設計を検討すること

近年、リスクの大きさに応じて対応措置をとるリスクベース・アプローチが注目されている[30]

22 “IMPACT: The DCPS Evaluation and Feedback System for School-Based Personnel” DISTRICT OF COLUMBIA PUBLIC SCHOOLS〈https://dcps.dc.gov/page/impact-dcps-evaluation-and-feedback-system-school-based-personnel〉

23 キャシー・オニール著（久保尚子訳）『あなたを支配し、社会を破壊する、AI・ビッグデータの罠』（インターシフト、2018）10〜15頁。

24 本研究会では、楠正憲「データ活用と不気味の谷――プライバシーインパクトと社会的受容をめぐって」NBL 1100号（2017）27〜28頁にて指摘があった。

25 工藤郁子＝荒井ひろみ＝江間有沙「採用におけるプロファイリング・サービスの倫理的課題『リクナビDMPフォロー』事件を手がかりに」人工知能学会第34回全国大会論文集（2020）は、リクナビ事案で学生とリクルートキャリアの間に、利害の不一致と情報の非対称性があり、プリンシパル・エージェント関係と同様の関係が生じていたこと、サービスの設計・運用において学生がどう思うかを問う視点が失われていたことを指摘する。森亮二「デジタルプラットフォームの法的責任に関する近時の問題」現代消費者法46号（2020）28頁は、リクナビ事案が批判を受けた本質的な理由につき、「就活生にとって不可欠ともいえるサービスであり、就活生支援を銘打ったリクナビにおいて、就活生が知らぬ間に自分に不利な情報を採用企業側に提供されていたことの『背信性』にある」と指摘する。

26 山本・前掲注⑹14頁は、一般的プロファイリングに関して、「プロファイリングの有無と目的＋例示の義務」で対応することを提案する。

27 当時のウェブサイトの記載が変更または消滅しているため、事実関係については、ITmedia NEWS「『Yahoo!スコア』、10月からデフォルトで『オフ』に 『ユーザーの意見や社会情勢を踏まえた』」〈https://www.itmedia.co.jp/news/articles/1909/09/news075.html〉を参照した。

28 本研究会では、成原慧「パーソナルデータとアーキテクチャの関係をめぐる試論――プライバシー・ナッジとプライバシー・バイ・デザインを題材にして」NBL1100号（2017）9〜12頁および石井夏生利「カナダにおける最近のプライバシー議論」同号15〜18頁で、バイ・デザイン的アプローチの意義と限界を検討した。

29 “IEEE ETHICS IN ACTION in Autonomous and Intelligent Systems”〈https://ethicsinaction.ieee.org/〉

30 リスクベース・アプローチの国内外での議論状況について、経済産業省AI原則の実践の在り方に関する検討会「我が国のAIガバナンスの在り方 ver.1.1」（令和3年7月9日）6頁。なお、リスク評価等との関係ではプライバシー影響評価（Privacy Impact Assessment: PIA）も注目されている。PIAの実施については、個人情報保護法の令和2年・3年改正では導入されなかったが、令和元年に個人情報保護委員会が公表した「個人情報保護法　いわゆる3年ごと見直し制度改正大綱」（令和元年12月13日）19頁では、大量の個人データを扱う事業者における有用な手段として、PIAが推奨されていた。その後、個人情報保護委員会「PIAの取組の促進について―PIAの意義と実施手順に沿った留意点―」（2021年6月30日）が公表された。また、PIAに関する国際標準規格である、ISO/IEC 29134:2017は、2021年1月にJIS規格として発行された（JISX9251）。そのほか、リスクベース・アプローチに関しては、“CIPL Recommendations on Adopting a Risk-Based Approach to Regulating Artificial Intelligence in the EU”〈https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_risk-based_approach_to_regulating_ai__22_march_2021_.pdf〉も参照。

（ケースE）。もっとも、リスクベース・アプローチという大きな考え方は概ね受容されているとしても、具体的なリスクの分類・評価方法とそれに対する対応方法については必ずしも共通見解は確立していない。

個人情報保護法制においては利用目的を中心とした規律がなされており[31]、本チェックリスト(1-2)でも利用目的の規律を充実させることを目指している。もっとも、要配慮個人情報を推知するプロファイリング等の対象者に重大な不利益を与え得る可能性のあるプロファイリングに関しては、一般的プロファイリングよりも規律密度を深めていくことが必要な場合もある。具体的なリスクを測定した上で、そもそも社会的に禁止されるべきカテゴリーのプロファイリングに該当しないか、単なる包括的同意ではなくより厳格な同意取得の必要性はないか、説明責任・透明性を充実させる必要性がないか等のリスクに応じた対応措置の要否を検討することも有益である[32]。

（ケースE） 欧州委員会「AI規則案」[33]

欧州委員会は2021年4月、AI規則案を公表した。これはAIシステムをリスク別に4つのカテゴリーに分類し、カテゴリーに応じて義務的な要件を課す、自主的な取組みを促進する等の措置を定めるリスクベース・アプローチを採用している。すなわち①容認できないリスクのAI（汎用的なソーシャルスコアリング等）、②ハイリスクAI（一定の遠隔生体認証、雇用等）、③限定的リスクのあるAI（チャットボット等）、④最小限リスクのAI（スパムフィルター等）の4カテゴリーを設け、①については原則禁止、②は一定要件および事前適合性評価の遵守の下で許容、③は透明性義務の下で許容、④は自由に利用可、という仕組みを提案している。

（ケースF） AI法研究会「プロファイリングに関する指針案(骨子)」

令和3年7月開催のAI法研究会「AI×法・倫理シンポジウム」[34]で提案された「プロファイリングに関する指針案（骨子）」では、リスクベース・アプローチを基本とした指針案が提案された。

この指針案では、リスクベース・アプローチによりPF（プロファイリング）対象者に不利益を与える可能性がほぼない場合（レベル1）、PF対象者に軽微な不利益を与える可能性がある場合（レベル2）、PF対象者に重大な不利益を与える可能性がある場合（レベル3）に分けて同意取得の要否、対象者への説明および人間関与に関する基本対応方針を決定する案が提言されている[35]。

(1-5) ガバナンスの確立 ★★☆

個人情報の取扱いに関する責任の所在を明確化すること

各部門間の連絡・連携の不備によりプライバシーリスクが発生することがあるため、全社的なデータ保護の観点に責任を持つプライバシー保護責任者等の役職者の選任により責任の所在を明確にするなど、組織体制の確立を検討する必要がある[36]。現時点では、プライバシー保護責任者を指名した上でプライバシー保護責任者を中心としたプライバシー保護組織を設立している組織は稀であると考えられる（ケースG）。プライバシー保護責任者の指名で足りるのか、それなりの規模のプライバシー保護組織を設立すべきかは、会社の規模やプライバシーリスクの大きさによるところもあるが、プライバシー保護のために一定のコストをかけて体制を確立すべき場合もあり得ることに留意すべきである。

（ケースG） リクナビ事案

前掲（ケースB）のリクナビ事案では、部門横断で担当者が局所的に連携を繰り返す形でサービスの検討が進められた。また、サービスのスキーム変更に当たり、リクナビのプライバシーポリシー改訂に関する運用が社内で整備・明文化されておらず、関係者の認識が統一されないまま進行して、一部画面へのプライバシーポリシーの反映漏れが生じた[37]。

(2) データの取得段階

(2-1) データの適正な取得 ★★★

プロファイリングのために使用するデータ(インプットデータ)を適正に取得すること

インプットデータが「個人情報」に該当する場合には、法的な観点からも「適正な取得」（個人情報保護法20条1項）を行う必要がある。

プロファイリングの実施や利用目的について虚偽の情報（プロファイリングを実施しているのに「実施していない」と表示することなど）を与えてはならないことはもちろん、プロファイリングの実施等について曖昧な情報を与えることでインプットデータを取得することは避けるべきと考えられる。

また、本人の人種・信条・社会的身分等の「要配慮個人情報」（個人情報保護法2条3項）を、プロファイリングのためのインプットデータとして取得する場合、原則として本人の同意を要するオプトイン方

式を採用する必要がある（同法20条2項）。さらに機微情報（センシティブ・データ）など、取得自体が問題になるデータもある。たとえば、与信のためのプロファイリング（スコアリング）を行う場合には、「金融分野における個人情報保護に関するガイドライン」が適用される場合がある。当該ガイドラインにおける「機微（センシティブ）情報」（同ガイドライン5条1項柱書）は「〔個人情報保護〕法第2条第3項に定める要配慮個人情報並びに労働組合への加盟、門地、本籍地、保健医療及び性生活……に関する情報（後略）」とされ、原則として、取得・利用・第三者提供のいずれも禁止されている。個人情報保護法における「要配慮個人情報」よりも対象範囲が広い点に注意する必要がある[38]。

なお、機微情報の中でも特に遺伝情報をはじめとする生まれに関連する情報は、本人に帰責できず自己責任原則との関係で問題となり得る[39]。よって、遺伝情報等のデータを利用する際には、より慎重な検討が期待される。たとえば雇用分野では、厚生労働省の「雇用管理分野における個人情報のうち健康情報を取り扱うに当たっての留意事項」（平成29年5月29日付基発0529第3号）は、「色覚検査等の遺伝性疾病に関する情報については、職業上の特別な必要性がある場合を除き、事業者は、労働者等から取得すべきでない」と規定している（第3・8(3)。同規定は、遺伝性疾病に関する情報を取得し得る場合として、労働者の求めに応じて、就業上の配慮を行う必要があるときを挙げている）。さらに、遺伝情報は血縁者などで共通する部分もあるため、「グループ・プライバシー」[40]への影響を検討することが望ましい。アルゴリズム上の配慮も推奨する（後掲**(2-3)**も参照）[41]。

(2-2) データ取得時のユーザー・インターフェース等の工夫 ★☆☆

データ主体の意思決定を実質的に支援するための工夫を行うこと[42]

近年、データポリシーやプライバシーポリシーが難解で長文であり、IoTなどの進展も考慮すると、人間の認知限界を超えるとの問題意識が示されている[43]。こうした動向を踏まえ、プライバシーポリシー等を、簡潔で理解しやすく、容易にアクセスし得る形態で提供することを推奨する（ケースH）。また、プライバシーポリシーの読み手の年齢層や使用言語への配慮の要否（ケースI）や散在するプライバシーポリシーの統合の要否（ケースJ）についても検討に値する[44]。また、UI（ユーザー・インターフェイス）やUX（ユーザー・エクスペリエンス）

31 西村あさひ法律事務所編・前掲注(8)13頁〔吉田正彦〕は、「利用目的による制限」を「個人情報保護法における胆となる考え方」と指摘する。
32 山本・前掲注(6)10〜11頁、14頁。
33 Proposal for a Regulation laying down harmonised rules on artificial intelligence（Artificial Intelligence Act）.
34 AI法研究会ウェブサイト〈https://www.aiandlawsociety.org/event-details/ai-ho-rinnrishinpojumu〉
35「プロファイリングに関する指針案（骨子）」では、リスクベース・アプローチを基本とした「3ステップ・アプローチ」により3段階のステップで対応措置を判定する方針が提言されているが、「3ステップ・アプローチ」の具体的な内容については紙面の関係から割愛する。
36 総務省＝経済産業省「DX時代における企業のプライバシーガバナンスガイドブック ver1.1」(2021年7月）16〜17頁。Zホールディングスグループにおけるデータガバナンス等を詳細に検討したものとして、Zホールディングス株式会社「グローバルなデータガバナンスに関する特別委員会最終報告書」（2021年10月）〈https://www.z-holdings.co.jp/notice/20211018〉参照。
37 株式会社リクルート「本件の根本問題と今後のガバナンス強化に向けた取組み」〈https://www.recruit.co.jp/r-dmpf/06/〉
38 個人情報保護委員会＝金融庁「金融分野における個人情報保護に関するガイドライン」（平成29年2月）。なお、令和3年12月22日付けで同ガイドラインの一部改正(案)についてパブリック・コメントに付されている。
39 米国の遺伝子情報差別禁止法（Genetic Information Nondiscrimination Act: GINA）も参照。
40 LINNET TAYLOR, LUCIANO FLORIDI, AND BART VAN DER SLOOT, EDS., GROUP PRIVACY: NEW CHALLENGES of DATA TECHNOLOGIES. Vol. 126.（Springer, 2016）.
41 自己情報コントロール権のうち自己決定・同意の要素の手段的性格を指摘し、個人情報の社会通念上不当な取扱いおよびそれに起因する不利益防止のために他に適切な手段があれば必ずしも同意にこだわる必要はないとの見解を示すものとして、曽我部・前掲注(8)71〜78頁。
42 たとえば、GDPR12条「データ主体の権利行使のための透明性のある情報提供、連絡及び書式」12頁では、管理者は、データ主体に対し、簡潔で、透明性があり、理解しやすく、容易にアクセスできる方式により、明確かつ平易な文言を用いて、GDPR13条等に定める情法等（管理者が個人データ取得時に提供する情報等）を提供するため（特に子どもに向けて情報を提供するため）の適切な措置を講じなければならないとされている。また、近年、欺瞞的な方法で消費者を不利な決定に誘導するデザイン「ダークパターン」が問題提起されている。その倫理的課題について、慶應義塾大学グローバルリサーチインスティテュート「【開催報告】人を欺くサービス設計『ダークパターン』について考える」〈https://www.kgri.keio.ac.jp/news-event/083668.html〉などを参照。
43 Daniel J. Solove, Introduction: Privacy Self-Management and the Consent Dilemma, 126 HARV. L. REV. 1880（2012）.
44 株式会社野村総合研究所コンサルティング事業本部「実効性のある通知・同意取得方法の在り方に関する実証事業の報告」〈https://www.soumu.go.jp/main_content/000744405.pdf〉は、利用者情報の通知・同意にあたり利用者の理解や安心に資すると思われる工夫として、(1)階層的な通知、(2)サービスの利用タイミングに合わせたタイムリーな通知、(3)個別同意、(4)プライバシー設定、(5)同意の証跡の5つの類型に整理した上、(1)(3)(4)につき実証事業を実施した結果、いずれの工夫も利用者の理解や安心に資することを確認している。

における工夫をグッドプラクティスとして検討・共有していくことも期待される。

ケースH　設定画面のリニューアル

2018年3月、Facebook（現：Meta）は、モバイル上の「設定」メニュー全体のデザインを見直し、プライバシー、セキュリティ、広告など約20に分散していた設定画面の集約を行い、1つの画面からすべての設定をできるように変更した[45]。メニュー画面は、イラストなども用いながら簡明な説明を心がけたとされている。

ケースI　TikTok事件[46]

2021年7月、オランダのデータ保護当局（Autoriteit Persoonsgegevens）がTikTokに対し、プライバシーステートメントの提供に関してGDPRに違反したとして、75万ユーロの制裁金を科した（係争中）。違反事由として、TikTokは若年層ユーザーが多いが、プライバシーステートメントを英語のみで提供し、オランダ語で提供しなかったため、ユーザーである子供たちがパーソナルデータの収集・取扱い・利用について適切な説明を受けることができておらず、透明性のある情報提供を行う義務に違反したとしている。

ケースJ　Googleのターゲティング広告[47]

2019年1月、フランスのデータ保護機関がGoogleに対して、特にターゲティング広告に関連する個人情報の取扱いにおいてGDPRに違反したとして、5000万ユーロ（当時約62億円）の制裁金の支払いを命じた。違反事由の一つとして、データの取扱いの目的、保存期間、ターゲティング広告に用いる個人データの種類等に関する情報が複数の文書に分散しており、複数のステップを踏まないとアクセスできないため、透明性のある情報提供を行う義務に違反したとしている。近年、ユーザーが理解しやすい体裁に整える手段の一つとしてプライバシーポリシーの統合を進める企業も増えており注目される。

(2-3)　データセットの偏向に関する配慮 ★★☆

データセットの偏りをチェックすること

データセットの質は、プロファイリングの精度に大きく関わる。データセットは、ある事象や現象（フィジカルな世界）の一部分を切り取ったものでしかなく、現実世界を完全に反映したものではない。したがって、そこには偏りが生じたり、特定のコミュニティがデータセット上に過少または過剰に現れたりするリスクがある（後者は「過少代表・過剰代表（underrepresentation and overrepresentation）」の問題といわれる）。プロファイリングのためのアルゴリズムを設計する際には、その基礎となるデータセットに偏りや過少代表・過剰代表がないかをチェックする必要がある。なお、過少代表・過剰代表とは、たとえば、放漫な食生活を送る者たちが、自身のスコア低下を懸念して飲食記録の提供をしないことで、分析対象となるデータに占める生活習慣病予備軍の割合が実際よりも低下し、判断・決定に影響を与えることなどをいう（他の例として、ケースK）。

ケースK　過少代表による道路状況調査の地理的偏向

ボストン市は、「Street Bump」というアプリを開発し、市民のスマートフォンから得られるGPSデータ等を利用しつつ、市内の道路状況（路面のくぼみなど）の報告を受けることで、道路修補工事の予測に役立てる仕組みを構築した。しかし、この調査結果に依拠すると、工事が高所得者の居住エリアに集中してしまうことが判明した[48]。高所得者と低所得者でスマートフォンの所持率に差異があり、低所得者の居住エリアからの情報が過少に代表されてしまったためである。

なお、ボストン市は、このアプリをまず市職員に利用させ、一般市民からのデータは補助的に使うことで問題に対処しているという。

また、「現実」が偏ってきたためにバイアスが生じることもあり、「公正さ（fairness）」との関係で配慮が必要となる（ケースL）。たとえば、出世が見込める社員かどうかの判断を支援するAIプロファイリングの開発について、既存の人事評価データを用いる場合、このデータの中に人種・性別に対する過去の偏見や偏向が潜在しているおそれがあり、結果として既存の差別やバイアスが継承される危険性がある[49]。

本研究会は、開発者または企業が、データセットやアルゴリズムにこれまで存在してきた社会的差別やバイアスが承継されていないかをチェックすることを推奨する（後掲**(3-2)**も参照）。なお、AI指導原則1条b項において、「AIの使用が個人に与える影響の程度のみならず、グループ及び社会全体に与える集団的影響についても考慮すること」と指摘されているように、被評価者個人に対する否定的影響だけでなく、特定の集団に対する否定的影響も考慮しておくことが期待される。

ケースL　過少代表とバイアス承継問題による、人材採用システムの開発中止[50]

米国アマゾン社は、履歴書を審査して応募者を5点満点でランク付けするシステムの開発を2014年頃から進めていた。ところが、特定職種で女性に不利だったこと

が判明した。システムでは、同社に提出された10年間分の履歴書を学習対象としていたが、技術職は男性からの応募が多数だったため、男性を採用することが好ましいと判断された。逆に、履歴書に女性に関する単語（たとえば「女子大学」、「女性チェス部の部長」）が含まれていると評価が下がる傾向が見られた。

本システムの運用は中止され、2017年には開発チームが解散となった。なお、アマゾン社の採用部門は、AIが示したランクに目は通したものの、ランクのみで採用を決定していなかったとしている。

(3) データの加工・分析・学習段階

(3-1) 分析・評価段階におけるプライバシー侵害リスクの検討 ★★★

プロファイリング(分析・評価)の可否について、プライバシー侵害リスクを検討すること

近年、個人情報から「要配慮個人情報」（個人情報保護法2条3項）に相当する情報を分析により獲得した場合、要配慮個人情報の新たな「取得」と考えて本人同意を要求すべきとする見解も説かれている（注9参照）。この見解の当否は措くとしても、要配慮個人情報に相当する情報をプロファイリングにより獲得する場合には、プロファイリングの実施等についてすでに告知しているかなど（前掲**(1-2)**参照）、それによるプライバシー侵害リスクを慎重に検討することが求められよう。「放送受信者等の個人情報保護に関するガイドライン」34条は視聴履歴に基づき要配慮個人情報を推知することについて注意喚起を行っており、同様の趣旨と解される。

もちろん、実務上、分析・評価段階における過度の規制・制約は、意味のあるプロファイリングを難しくすることにつながり、開発者または企業にとっては受け入れにくいと考えられる[51]。他方で、一般の個人情報から要配慮個人情報に相当する情報を分析・評価を通じて予測することは、その予測精度が高ければ高いほど、第三者に真実らしく受け取られるおそれがあり、被評価者の法益との調整が要求される。両者のバランスについて、現段階では具体的な指針が示されておらず、今後の社会的議論と合意形成が目指されているところである[52]。

そこで、当面の間は、要配慮個人情報に相当する情報を分析・評価するか否かを慎重に検討し、仮に分析・評価を行う場合には、予測精度が高く、真実らしく受け取られるおそれがあるものについて、プライバシー侵害リスクに対する配慮を検討し[53]、妥当な規律レベルを探ることを推奨する（後掲**(4-5)**参照）。

(3-2) 公平性に配慮した学習技術の導入検討 ★★☆

社会的公正や反差別に配慮した技術的対応の導入について、プロダクトの目的と照らし合わせて検討すること

GL通則編は、不適正利用禁止義務（個人情報保護法19条）違反となる事例として「採用選考を通じて個人情報を取得した事業者が、性別、国籍等の特定の属性のみにより、正当な理由なく本人に対する違法な差別的取扱いを行うために、個人情報を利用する場合」を挙げている。GL通則編で挙げられている当該事例は必ずしもプロファイリングを直接的に対象とするものではないと考えられるが、差別的取扱いを規律する趣旨を含んでおり、差別的プロファイリングは不適正利用禁止義務違反になり得る[54]。個人情報保護委員会は、令和3年8月のGL通則編改正の際のパブリックコメントに寄せられた

45 Meta「より使いやすくするために、プライバシー関連ツールを更新」〈https://about.fb.com/ja/news/2018/03/privacy-shortcuts/〉
46 “Dutch DPA: TikTok fined for violating children’s privacy”〈https://edpb.europa.eu/news/national-news/2021/dutch-dpa-tiktok-fined-violating-childrens-privacy_en〉
47 “The CNIL’s restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC”〈https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc〉
48 Ramirez et al., *supra* note 20, at pp.27-28.
49 Ramirez et al., *supra* note 20, at pp.28-29.
50 「焦点：アマゾンがAI採用打ち切り、「女性差別」の欠陥露呈で」〈https://jp.reuters.com/article/amazon-jobs-ai-analysis-idJPKCN1ML0DN〉
51 本研究会では、工藤郁子「質疑応答と議論」NBL 1100号（2017）31頁〔「荒井報告への質疑応答」部分〕にて指摘があった。
52 この論点は「〈私づくり〉の主導権をめぐる対立」（阪本俊生『ポスト・プライバシー』（青弓社、2009）83頁）につながる。他者が有する「私」の自己イメージ像（三人称視点での「私」）を、一人称の「私」自らが管理できる（またはすべき）ものなのか、という問いである。これは、「アイデンティティ」が、被評価者たる個人と、評価者たる第三者や社会の境面において生み出されるという特性を背景としており、一意に定めることは困難であろう。
53 前記**(1-2)**で言及した透明性の確保や、（マーケティングでのプロファイリングの実施等については）オプトアウト手続の整備などが考えられる。
54 田中浩之＝北山昇「規制対象範囲・適正性の判断軸は？ 不適正利用の禁止義務への対応」ビジネス法務2020年8月号28頁参照。

意見に対して「プロファイリングに関連する個人情報の取扱いについても、それが『違法又は不当な行為を助長し、又は誘発するおそれがある方法』による個人情報の利用に当たる場合には、不適正利用に該当する可能性がありますが、プロファイリングの目的や得られた結果の利用方法等を踏まえて個別の事案ごとに判断する必要があると考えられます」と回答しており[55]、不適正利用禁止義務は一定のプロファイリングに対して及ぶことが明らかになっている。合理的区別か違法な差別的取扱いかの区別を明確にすることは困難を伴う部分があるが、憲法14条1項後段は特に差別を禁圧すべき類型として「人種、信条、性別、社会的身分又は門地」を挙げており、このような類型に基づく区別取扱いをするような場合には差別リスクが発生しないかをチェックしておく必要がある（ケースM参照）。

また、前述のように、現実社会の偏見・偏向がデータセットに潜在するために、結果として既存の差別が継承・再生産されてしまう可能性がある（前掲**(2-3)**参照）。本研究会は、この可能性に対して、データの加工・分析・学習段階において、差別につながる要素や機微に属する特徴（sensitive feature）を除去したり、その影響力を弱めたりするといった対策を講じることを推奨する。もっとも、単純な除去では、それらの特徴の間接的な影響を必ずしも排除しきれない。たとえば、「人種」という特徴を利用せずに与信スコアを分析したとしても、ある人種が特定の地区に集まっていると、「住所」という指標によって結果的に差別的な決定が行われてしまうことがある[56]。したがって、開発者または企業として、こうした間接差別（indirect discrimination）が生じていないかを検討することが必要な場合があるだろう。

社会的公正に配慮した分析技術については、近年研究が進んでおり、差別的な決定をすることのないようにモデルを修正したり、機微に属する特徴と決定との間の独立性を高めたりといった手法が検討されている。こうした公正配慮型学習（fairness-aware learning）[57]の導入なども一考に値する[58]。

ただし、公正配慮型学習を採用することにより、いわゆる「逆差別（reverse discrimination）」が発生するおそれもあり、その点についての留意が求められる場合もある。また、目指すべき目標が複数あると、それらのトレードオフが問題になることがある。たとえば、「アルゴリズムの予測精度を犠牲にしてでも、平等・公平を実現すべきか？」という倫理的課題に直面することが想定される。優先順位をつけて判断しつつ、技術的対応でカバーできなかった価値を社会的対応ですくい上げるなど、多面的なアプローチを検討することが推奨される[59]（後掲**(4-5)**参照）。

ケースM　HireVue社の事案

HireVue社はAIを用いた面接評価技術を提供している。同社の商品であるHireVueは、ビデオインタビューを分析し、言語選択や表情などからスコアを算出するアルゴリズムを用いて求職者を評価するAIツールであるが、これらのAIツールはしばしばジェンダーや人種のバイアスがかかりやすい点が指摘されている。2019年11月、権利団体の電子プライバシー情報センター（EPIC）は米国連邦取引委員会（FTC）に対して、HireVue社による顔認識技術を用いたツールの提供につき「不公正で欺瞞的な」行為があったとして調査を要請した[60]。HireVueはその後、顔分析による求職者の評価を中止したが、音声、イントネーション、行動などの生体情報による分析は続行している[61]。

こうした問題を受けて、採用時にAIツールを使用する際には求職者に対して事前の告知および同意を得ることを求める法律が2020年1月に米国イリノイ州で施行されている[62]。

(3-3)　アカウンタビリティ（答責性）に配慮したモデルの導入検討　★★☆

答責性（accountability）、説明可能性（explainability）、解釈可能性（interpretability）、透明性（transparency）などに配慮し、プロファイリングに利用したインプットデータを特定しておくことや解釈可能なモデルの導入を検討すること

ディープラーニング等によって、大量のデータに基づいた複雑な予測モデルの構築が可能になった。しかし、複雑になりすぎて、なぜそのような分析結果になったのかという理由を、開発者自身も説明することができないといった事態も生じている。こうした状況は、説明可能性・透明性等との関係で問題となり得る[63]。なお、こうした課題については、FAT（Fairness, Accountability, and Transparency: 公平性・答責性・透明性）という研究領域や研究コミュニティが解決に向けた取組みを行っており、参考に値しよう（ケースN参照。さらに後掲**(4-2)**も参照）[64]。

ケースN　複雑なモデルに解釈を与える試み

2017年の機械学習の国際会議ICMLのチュートリアルで例示された事例では、予測モデルが複雑すぎると、入出力の関係の解釈が困難になることが示された[65]。入力（input）に対してモデルが単純であれば、出力までの経路が比較的簡単に説明できる。しかし、予測モデルや与えられるデータが複雑になりすぎると、どの変数がより多くシステムに使われていて重要であるかなど、システム全体の説明が非常に困難になる。

こうしたいわゆる「ブラックボックス化」に対して、説明可能性を確保すべく、予測モデルの理由や根拠を示したり、複雑なモデルに解釈を与える試みが進んでいる[66]。たとえば、結果に効く変数（因子）のみを取り出したり、結果だけ活かす形でより単純なモデルに置き換えたりする研究がある。もっとも、こうした近似化の取組みも、正確性とトレードオフになることが多い。

(3-4)　その他の社会的価値への配慮 ★☆☆

公平性・答責性・透明性以外の社会的価値にも配慮すること

さまざまな社会的価値に配慮することを推奨する。民主主義的価値は、その一例である（ケースO参照）。

ケースO　ニュースのアルゴリズムとフィルター・バブル

ニュースや記事を収集・整理して推奨する「おすすめ機能」の開発を想定すると、利用者の閲覧履歴などを元に、嗜好データを取得・分析して「この記事を読んだユーザーは、この記事も読んでいます」と自動的に推論を行う「協調フィルタリング（collaborative filtering）」を採用することが想定される。しかし、これまでの好みに基づいて最適化するパーソナライズを強めすぎると、ライフスタイルの記事ばかりを好んで政治経済ニュースに注意を向けない傾向を助長したり、特定の政治傾向を強化し、自分と異なる見解に触れる機会が減ってしまったりする事態が懸念される。こうした現象は「フィルター・バブル（filter bubble）」と呼ばれ[67]、自分好みの情報だけに泡のようにとり囲まれてしまうことが問題であると指摘されている。

(4)　実装・運用・評価段階

(4-1)　人間関与原則の検討 ★★★

プロファイリングを利用した評価プロセスに対する人間の関与を検討すること

完全自動意思決定に関するGDPR 22条について、人生に重要な影響を与える決定には原則として人間が関与しなければならないという「人間関与原則」を定めたものであり、被評価者に対し、AIプロファイリングの統計的・確率的な判断に異議を唱える機会等を与えたものであるとする見解がある[68]。

AI指導原則5条c項でも、「個人の成長や見解に影響を与える技術への適用に異議を申し立てる又は

55 個人情報保護委員会事務局「『個人情報の保護に関する法律についてのガイドライン（通則編、外国にある第三者への提供編、第三者提供時の確認・記録義務編及び匿名加工情報編）の一部を改正する告示』等に関する意見募集の結果について」（令和3年8月2日）別紙2-1。「個人情報の保護に関する法律についてのガイドライン（通則編）の一部を改正する告示案」に関する意見募集結果57番。

56 Dino Pedreshi, Salvatore Ruggieri, and Franco Turini, "Discrimination-aware Data Mining", Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2008.

57 Toshihiro Kamishima, Shotaro Akaho and Jun Sakuma, "Fairness-aware Learning through Regularization Approach", Data Mining Workshops（ICDMW）, 2011 IEEE 11th International Conference on. IEEE, 2011.

58 なお、Google が公正性に関するベストプラクティスを公表している。"Responsible AI practices"〈https://ai.google/responsibilities/responsible-ai-practices/?category=fairness〉

59 この点については社会全体での対応が求められよう。

60 The Electronic Privacy Information Center（EPIC）, Complaint and Request for Investigation, Injunction, and Other Relief（Nov. 6, 2019）〈https://epic.org/privacy/ftc/hirevue/EPIC_FTC_HireVue_Complaint.pdf〉

61 HireVue, Facing FTC Complaint From EPIC, Halts Use of Facial Recognition〈https://epic.org/hirevue-facing-ftc-complaint-from-epic-halts-use-of-facial-recognition/〉

62 Artificial Intelligence Video Interview Act

63 AI指導原則3条a項では「説明可能なAI」への投資が要請されている。

64 AI指導原則3条でも「AIシステムの透明性及び明瞭性」が要求されている。

65 Been KimとFinale Doshi-Velezによる"Interpretable Machine Learning: The fuss, the concrete and the questions"と題する講演資料より。現在は要旨のみ確認できる〈https://icml.cc/Conferences/2017/Tutorials〉。

66 本研究会では、荒井ひろみ「プロファイリング規制に対する技術面からの一検討」NBL 1100号（2017）25〜26頁にて論点整理と概観を行った。

67 ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU（Penguin UK, 2011）. 翻訳として、イーライ・パリサー『フィルターバブル――インターネットが隠していること』（早川書房、2016）。

68 山本・前掲注(9) 103頁はGDPR 22条との関係で人間関与原則に言及し、同書105頁はGDPR 22条を「セグメントに基づくAIの統計的・確率的な判断からの自由を保障」するものと位置付ける。なお、AIの類型的・確率的な評価に実質的に反論を加えることの困難性について山本龍彦「AIと『個人の尊重』」福田雅樹ほか編著『AIがつなげる社会――AIネットワーク時代の法・政策』（弘文堂、2017）331頁。

抗議する権利を認めること、及び、適用できる場合には、もしそれが個人に重大な影響を与える場合には自動化された処理のみに基づく決定には従わない個人の権利を保証すること、また、適用できない場合にはその決定に異議を申し立てる個人の権利を保証すること」が要求されており、GDPR 22条と同様の原則がうたわれている。これらの考え方の背景には、自動化された決定が「個人の尊重」や「尊厳（dignity）」（日本国憲法13条、24条2項）を脅かすのではないか、との認識がある。

他方で、「人間関与を強制すると、かえって人間疎外が起きてしまう」との批判の声もある。たとえば採用活動において、（ときには応募者側がAIを用いて）大量の応募書類が送られてきており、その選抜に膨大なコストが必要になっているという現状のもとで、「人間関与」を原則とすれば、担当者の疲弊を招き、ひいては学歴や性別といった代理変数の利用を助長し、結果的に差別を誘発することになりかねないというものである。

そこで本研究会は、当面の間は、完全自動意思決定の内包するリスクに鑑みて、プロファイリングに基づく自動化決定による評価に不服が申し立てられた場合に限って人間による再審査を行うなど（その場合には合理的な金額の手数料を徴収することも考えられる）、常に人間が関与することを厳格に求めるのではなく、適切な手続や体制の整備を通じて、妥当な規律レベルを探ることを推奨する。

(4-2) 答責性への配慮 ★★★

答責性に配慮すること

答責性、透明性等への配慮は、実装・運用・評価段階において、より重要性を増す[69]。本研究会は、企業が、決定の過程（決定を導くデータを含む）について、被評価者や他のステークホルダーに非専門用語で説明できるかどうかを検討することを推奨する（ケースP参照）。

ケースP 再犯リスク予測プログラムとアルゴリズムの透明性

米国では一部の州で有罪確定者の再犯リスク予測プログラムが用いられている。ただし、このプログラムのアルゴリズムが明らかにされていないため、バイアスがあるのではないかとの批判もある[70]。

ある再犯リスク予想プログラムでは、アフリカ系を白人よりもリスクを高く評価するが、アフリカ系が実際に再犯をした割合は半分以下であったという調査も報告されている。

データやアルゴリズムのバイアスに関しての検証をするためにも技術システムの透明性や答責性が開発側には求められている。

	“White”	“Black”
再犯率が高いと予測されたが、実際には再犯しなかった率	23.5%	44.9%
再犯率が低いと予測されたが、実際に再犯した率	47.7%	28.0%

もっとも、説明内容や程度の妥当性については、いまだ社会的合意はない。また、AIのプロファイリングにのみ答責性を求めるのは均衡を欠くとの指摘もある。これはたとえば、人間の採用担当者であれば不採用理由を開示しなくても問題視されない現状をどう捉えるかという問題である[71]。他方で、プロファイリングの過程や評価理由が不透明であれば、何が能力や信用力に関する評価の基礎になるのかがわからず、将来の行動計画を練れなくなり、自律的・主体的に自らの人生を歩んでいくことが困難になるとの懸念も存在する[72]。

上記を踏まえると、おそらく当面のベンチマークにすべきは、事後（特に苦情対応時）に被評価者が理解し得るような「理由（justification）」を示すことであろう。ここで被評価者に事後提示される「理由」は、実際の決定過程を完全に反映したものである必要はない（別のアプローチとして、ケースQ）。関連して、決定理由を人間に提示できる「説明可能なAI」（Explainable AI, XAI）の開発動向にも関心を向ける必要があろう[73]。

ケースQ 信頼の構築

ディープラーニングの進展に伴う説明困難性を受け止め、「信頼（trust）」や「信用（fiduciary）」の構築を重視する立場もある。たとえば、医薬品の作用機序を十分に理解していなくても、医師や薬剤師の勧めに従って服薬することが一般的である。これは、高い能力・専門性を備えた医師等に任せたほうが合理的であるという考えや、国家資格による認証や専門職集団の行動規範（code of conduct）による制約などが制度的に担保されているから信じるに足るとの考えが背景となっていよう。このような信頼構築のための規律構造をAI化社会においても実現しようとする試みが存在する[74]。

(4-3) 安全管理措置の実施 ★★☆

プロファイリング結果(たとえば信用スコア)が個人に与える影響の重要性に鑑みて、その安全管理には細心の注意を払うべきである

プロファイリングを実施する場合には、プロファイリング結果（たとえば信用スコア）が漏えい、滅失、き損することによって本人の被る権利利益の侵害の大きさ等を考慮して、リスクに応じた安全管理措置を実施することが求められる。このことは、個人情報保護法23条が安全管理措置義務を規定した趣旨とも合致すると考える。

また、特に完全自動意思決定を行う場合には、個人に与える影響の重大性に鑑みて、プロファイリングのプロセスが外部から「攻撃」を受け、アルゴリズムが「き損」されたり、プロファイリング結果が歪められたりしないよう、安全管理措置を講ずることが求められる（ケースR）。

ケースR アルゴリズムに対する「敵対的な攻撃」への備え

近年、学習機能を利用してシステムの判断に歪みをもたらすことが可能であると指摘されており、そのリスクに備えた安全管理措置も一考に値する。2018年の機械学習に関する国際会議（ICML）では、画像認識アルゴリズムに対する「敵対的な攻撃（adversarial attacks)」の研究が紹介された[75]。誤った学習を引き起こすようなサンプルデータを意図的に読み込ませた結果、アルゴリズムが「騙されて」しまい、3Dプリントされた「カメ」を誤って「ライフル」とみなしてしまったというものである。

データセットを介して「攻撃」を受けたプロダクトは、セキュリティの観点から問題になる。このような「敵対的な攻撃」については、その攻撃方法と防御方法に関する技術的研究が進められており、その知見を取り入れるなどの対応が期待される。

(4-4) データ内容の正確性の確保 ★★☆

データ内容の正確性を確保すること

プロファイリングに用いるインプットデータの内容の正確性・最新性を確保することが求められる（個人情報保護法22条)。これは努力義務であるものの、プロファイリングの実施目的の達成に必要な範囲で正確性・最新性を確保することが望ましい。

正確性・最新性を確保するために、不必要になった場合に消去措置を行う手法が想定される。たとえば、クレジットカードの利用情報について、指定信用情報機関CICおよびJICCは、契約期間中および契約終了後から5年間と年限を設けている[76]。また、米国公正信用報告法（Fair Credit Reporting Act）では、情報の償却年限が定められており、ネガティブ情報は通常7年以上、破産宣告は10年以上経過したものは利用できないとされる。

なお、消去措置のほかに、情報の信頼性に関する情報である「情報のアノテーション（annotation)」[77]や「データ品質」[78]をメタデータとして示すというアプローチも想定される。すなわち、情報それ自体とは区別された注釈情報として、作成日・作成組織・最終更新日、さらには確実性・重要性などのメタデータを付随させることで、たとえば、犯罪履歴とおぼしきデータにつき、情報源が分析結果か、逮捕時の報道か、裁判例情報に基づくのか、また、どの程度「時の経過」があるか[79]、などが把握できるようになる。

69 なお、Googleが解釈可能性(interpretability)に関するベストプラクティスを公表している〈https://ai.google/education/responsible-ai-practices〉。
70 https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
71 「採否決定の理由を明示、公開しないことの自由」について東京高判昭和50・12・22判時815号87頁。
72 山本・前掲注(9) 79〜87頁。
73 たとえば、2016年8月、米国の国防高等研究計画局（Defense Advanced Research Projects Agency: DARPA）は、XAIへの投資プログラムを発表した〈https://www.darpa.mil/program/explainable-artificial-intelligence〉。
74 本研究会では、中川裕志「《パーソナルデータ＋α研究会シンポジウム記録》［第1セッション］人工知能の社会への影響と開発倫理指針」NBL 1130号（2018）67頁などで検討した。
75 Matthew Hutson, “Hackers easily fool artificial intelligences.” Science 361.6399（2018）. 215〈http://science.sciencemag.org/content/361/6399/215〉
76 株式会社シー・アイ・シー「CICが保有する信用情報」〈https://www.cic.co.jp/confidence/posession.html〉および株式会社日本信用情報機構「信用情報について」〈https://www.jicc.co.jp/credit_info/registration/〉
77 大屋雄裕「人民の、人民による、人民のための情報：個人情報の自由と範囲」社会保障研究3巻3号（2018）352頁。
78 内閣府の「分野間データ連携基盤の整備に向けた方針」でも「データの完全性、正確性、有用性、最新性等の品質を正確に知る必要があり、データそのものを入手する前に、データカタログのメタデータ等でデータ品質を知りたいというニーズがある。今後、基盤として評価の基準づくりに取り組み、それぞれのデータ品質を統一した基準のもとで表現することが望ましい」としており、注目される〈http://www8.cao.go.jp/cstp/tyousakai/juyoukadai/14kai/siryo2-1.pdf〉。
79 最三判平成6・2・8民集48巻2号149頁〔「ノンフィクション『逆転』」事件〕。

(4-5) 開示・訂正等の手続の整備 ★★★

被評価者からの苦情処理手続を整備すること

プロファイリングに用いるインプットデータについては、それが「保有個人データ」に当たる以上、開示・訂正・利用停止等の権利（個人情報保護法33条～35条）の対象となる。プロファイリングの結果情報（出力データ）については、それがそもそも個人情報保護法上の「保有個人データ」の内容を構成するかどうかが解釈上論点になると考えられるが、本研究会は、GDPRやカリフォルニア州消費者プライバシー法が、プロファイリング結果（後者については「推論（inferences）」）を「個人情報」として扱っていると解されること、現実にプロファイリング結果も、個人情報データベースを構成する「個人データ」になり得る上（個人情報保護法16条3項参照）、企業が開示、訂正等を行う権限を有していないとはいえないことから（同条4項参照）、企業は、可能な限りで、インプットデータと同様に、開示、訂正等の請求に応じるべきと考える。ただし、もちろん同法16条4項の「政令」で定める場合には保有個人データに該当せず、また保有個人データであっても同法33条2項各号に当たれば開示を要しない。

特に令和2年個人情報保護法改正により利用停止・消去請求の要件が緩和され（同法35条1項、5項、6項）、また35条に基づく請求権を実効化するために第三者提供記録の開示義務化（同法33条5項）がなされており、個人の権利利益保護の観点から各請求に対応していく必要性がますます増している。

なお、個社（特に新興企業）では、苦情処理手続の整備が難しい場合も想定されるため、業界団体や認定個人情報保護団体等に窓口業務を集約することも一考に値する。

(4-6) プロファイリングに関するリテラシー向上 ★☆☆

プロファイリング結果を直接取り扱う者（人事担当者等）が、適切にその結果と向き合い、同結果を評価できるための教育活動を行うこと

企業は、認定個人情報保護団体等とも連携しつつ、プロファイリング結果を直接取り扱う者のリテラシーを高めるための教育・研修・啓発活動を推進することが望まれる。

機械によって自動化された判断を人間が過信し、鵜呑みにしてしまう「自動化バイアス（automation bias）」も懸念されている。実際には、データセットの量、質、偏りなどから（前掲**(2-3)**参照）、アルゴリズムを用いた予測自体にバイアスが混入することもあり、当然ながら自動化判断も絶対ではない。プロファイリング結果（たとえば信用スコア）を取り扱う者が、その結果を盲信してアルゴリズムを「神格化」することがないようにする対応（研修など）が求められる。また、プロファイリング結果を安易に示して被評価者等を傷つけないよう表現や伝え方に配慮する対応も期待される。

(4-7) 被評価者等の反応の精査・検討 ★☆☆

被評価者等の行動変容も視野に入れて検討すること

評価・分析の結果が被評価者等の行動・選好に影響を与える可能性があるので、その構造も考慮に入れて設計することが望ましい。たとえば、認知件数に基づいて空き巣や痴漢が発生しやすい地域・時間帯を予測し、巡回ルートを提示するシステムを導入した場合、巡回頻度が増すことで警察への相談がしやすくなり、認知件数も増えた結果、当該地域・時間帯の巡回がさらに強化されるルートをシステムが提示するといったフィードバック・ループが発生する可能性がある。この際、もともと暗数が多かった地域・時間帯が見過ごされ、巡回リソースの配分が不適切になることが懸念される。

このほか、過剰適応やゲーミングを引き起こすおそれもある（ケースS）。

ケースS 大学ランキングと操作

米国の雑誌社が発表する大学ランキングでは、入学者のSAT（大学進学適性試験）のスコアなどを指標にしている。

そのため、米国のベイラー大学は、ランキングの向上を期待して、入学予定者に資金提供し、SATを再受験させた。スコアに反映される項目の評価を改善しようと努める大学は多いが、こぞって同じ目標を目指すようになれば、大学の多様性が失われるのではないかとの指摘もある[80]。

(4-8) 信用スコアを第三者提供する手続の確認 ★★☆

信用スコアの第三者提供等について慎重な配慮をなすこと

プロファイリング結果の中でも、信用スコアに該当するような情報については、第三者提供等の是非・方法について組織内で慎重な検討を行うことが推奨される。たとえば、信用スコアの第三者提供等を、包括的同意やいわゆるオプトアウト手続（個人情報保護法27条2項以下を参照）によって行うべきかについては、レピュテーションリスクとの関係で議論を尽くしておく必要があろう（ケースT）。また、提供先で個人データとなる個人関連情報の規制（同法31条）は、リクナビ事案で問題が顕在化した提供先での個人識別に関する問題を背景に導入されたものであり、このような態様での第三者提供の際には同意が得られていることの確認措置を検討する必要がある（ケースU）。

ケースT　Yahoo! スコア事案

前掲ケースCのYahoo! スコア事案では、スコア情報をYahoo! IDの連携先企業に提供する際に、ID連携画面で同意を取得する仕組みになっていたが、同画面では「サービス利用状況等に応じた特典等の付与のためのスコアを含む各種指標を提供します」と表示されるのみで、信用スコアの具体的な内容は明確に記載されていなかった[81]。

ケースU　リクナビ事案

前掲ケースBのリクナビ事案では、学生本人の同意なく個人データを顧客企業に提供していた点につき、個人情報保護委員会から勧告がなされた。また、内定辞退率の提供を受けた企業側において特定の個人を識別できることを知りながら、提供する側（株式会社リクルートキャリア側）では特定の個人を識別できないと判断していた点について、個人情報保護委員会は「個人データの第三者提供の同意取得を回避しており、法の趣旨を潜脱した極めて不適切なサービスを行っていた」と指摘した。

関連して、警察等の国家機関からの問合せや弁護士会照会等にどう対応するかの方針を定めておくことも検討に値する[82]。プロファイリングにより分析し推知した「機微情報」に相当する情報を、たとえば、捜査関係事項照会書に基づく要望などに応じて任意に提供するか令状を求めるかにつき、内規を作成しておくことを推奨する（個人情報保護法27条1項1号参照）。また、対外的にユーザーのデータが適切に扱われているのかに答えていくことも一考の余地がある（ケースV）。

ケースV　透明性レポート

Google[83]、Meta（Facebook）[84]、Twitter[85]などは、「透明性レポート」を公表している。当該レポートでは、ユーザーから寄せられる情報開示請求・削除請求、政府からのコンテンツ削除リクエストなどが公表されており、各種の請求にどう対応をしているかを一望できる。

(4-9) 監査可能性への配慮 ★☆☆

監査可能性（auditability）を確保すること

一定の情報開示を通じて、関心のある第三者（研究機関や個人情報保護団体等）が、アルゴリズムの動作を調査し（ケースW）、プライバシーや公正性を評価することを可能にする仕組みの構築を期待する[86]。

ケースW　ニューヨーク市のアルゴリズム監査

アルゴリズムが「ブラックボックス」化していることを前提に、モデルの背後にある前提を調べ、そのモデルの公平性を評価するといった監査を提唱する論者もいる。実際に、ニューヨーク市は、「市の機関によって利用される自動化決定システムに関する条例」を可決し、行政が使用するアルゴリズムの公正さを監視するタスクフォースを設置した[87]。

80 オニール・前掲注(23) 85頁。
81 事実関係については、ITmedia NEWS・前掲注(27)を参照した。
82 安全保障との関係については、山口いつ子「国家安全保障におけるアルゴリズムによる監視――憲法上の言論の自由・プライバシーとプラットフォーム事業者の役割」憲法研究3号（2018）47～54頁を参照。ポイントカード事業を展開する企業が、捜査関係事項照会に基づいて会員情報を捜査当局へ提供した行為が問題視された事例として、第二東京弁護士会情報公開・個人情報保護委員会編『令和2年改正 個人情報保護法の実務対応』（新日本法規、2021）247～251頁〔風早孝紀〕を参照。
83 Google「Google透明性レポート」〈https://transparencyreport.google.com/?hl=ja〉
84 Meta, “Transparency reports”〈https://transparency.fb.com/data〉
85 Twitter「Twitterの透明性に関するレポート」〈https://transparency.twitter.com/ja.html〉
86 AI指導原則3条c項。
87 The City of New York, “Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City.”（2018）〈https://www1.nyc.gov/office-of-the-mayor/news/251-18/mayor-de-blasio-first-in-nation-task-force-examine-automated-decision-systems-used-by〉. Rashida Richardson, “New York City Takes on Algorithmic Discrimination.” ACLU（2017）〈https://www.aclu.org/blog/privacy-technology/surveillance-technologies/new-york-city-takes-algorithmic-discrimination〉

Ⅳ まとめ

本最終報告書では、企業等の参考の用に供することを目的として、プロファイリングを実施する際に留意すべき事項を、法的・社会的責任の観点から検討した。本最終提言・最終報告書が、各企業・業界団体等のプロファイリングをめぐる悩みに関する解決の糸口となること、そして、自主的な取組みの「たたき台」として活用されることを期待している。

他方で、残された課題も多い。今回は、個人の自由で主体的な生活に重要な影響を与え得るプロファイリングについて全般的な留意事項を例示列挙するにとどまり、業界別（たとえば、与信領域、採用・人事領域、保険領域など）の個別事情を検討するには至らなかった。本研究会の提言案を元に人事分野や医療分野については個別ガイドラインが作成されるに至っており、業界別の個別事情を取り込んださらに具体的な民間主導の自主的取組みがなされることを期待する[88,89]。また、苦情処理手続などの具体的なガバナンスのあり方についてもさらに検討すべきであろう。

今後も関係各位との情報共有・意見交換を続け、プロファイリングの健全な実装に資する情報や議論を提供していきたい。

〔注記〕最終報告書の作成に至るまでの過程において、多くの方のご助言とご協力を賜った。ここに記して感謝したい。特に、ヒアリングにご協力いただいた、一般社団法人ピープルアナリティクス＆HRテクノロジー協会、一般財団法人日本情報経済社会推進協会、株式会社メルペイの皆さまに深く御礼申し上げる次第である。

88 こうした民間主導の自主的取組みとして、一般社団法人ピープルアナリティクス＆HRテクノロジー協会「人事データ利活用原則」（2020年3月19日）、World Economic Forum, “Good Data: Sharing Data and Fostering Public Trust and Willingness”, White Paper, 2021, pp.15-17, 〈http://www3.weforum.org/docs/WEF_Good_Data_Sharing_Data_and_Fostering_2021.pdf〉.

89 たとえば、人事分野ではニューヨーク市の条例（A Local Law to amend the administrative code of the city of New York, in relation to automated employment decision tools）が2021年12月に成立し、2023年1月1月に施行予定である。同条例では、バイアス監査結果のサマリーをウェブサイトに掲載するとともに自動雇用意思決定用に収集したデータの種類、入手先、データ保持ポリシーの保持が要求されているほか、自動雇用意思決定ツールの使用について従業員等への事前通知も要求されている。人事分野における自動意思決定の規律に関する立法例として参考になる。*See* Matthew Jedreski, Erik Mass, and K.C. Halm, “New York City’s Groundbreaking New Law Will Require Audits of AI and Algorithmic Systems That Drive Employment Decisions”, Davis Wright Tremaine LLP 〈https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2021/12/nyc-employment-ai-bias-audit-law〉

NBL
New Business Law