AI 利用に伴うプロファイリングの諸問題

執筆者：弁護士 渡邊 満久

December 2018

In brief

プロファイリングというと、一般には、種々の情報から犯人像の推測を行う捜査手法を思い浮かべる方が多いと思いますが、ビッグデータを学習した AI による出力結果たる予測・推測結果と密接に関連するトピックです。

日本においては、他国と比較して、このプロファイリングによる問題点に関する議論は活発ではないように思われますが、2018 年 5 月 25 日に施行された EU 一般データ保護規則（以下「GDPR」といいます）において、プロファイリングに関する規定が設けられるなど、諸外国ではその問題点が意識され始めています。近い将来、日本においても、AI の利用が進むに従い、プロファイリングに関する議論が活発になることが予想されます。そこで、本ニュースレターでは、プロファイリングの問題点と対応策を検討致します。なお、本ニュースレター中、意見にわたる部分は全て執筆者の私見です。

In detail

1. プロファイリングとは

(1) 定義

「プロファイリング」という言葉を辞書で引くと、「犯罪捜査で、行動科学的分析により犯人像を割り出す方法。犯罪現場の状況や過去の犯罪のデータから犯人の特徴を導き出す。」と説明されています [1]。現在の日本においては、一般には、プロファイリングというと、この意味を指すことが多いと思われます。他方、GDPR においては、プロファイリング（profiling）は、以下のように定義されています。

> 「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。」（GDPR4 条 4 号）

ここでは、プロファイリングは犯罪捜査の場面に限定されていません。「個人データの自動的な取り扱い」の対象が特に限定されていないことから、これが何を意味するのかイメージが掴みづらいところであると思います。そこで、プロファイリングの具体例を紹介します。

(2) 具体例

日常生活においてもっとも身近なプロファイリング例として、広告を挙げることができます。例えば、消費者がインターネット通販サイトにアクセスすれば、様々な商品がお勧めの商品としてピックアップされて表示されます。あるいは、消費者がアクセスするポータルサイト等の端にも、様々な商品やサービスに関する広告が表示されます。このような広告は、無作為に表示されているわけではなく、多くの場合、その消費者の過去の購入履歴やインターネットの閲覧履歴等の様々な情報から、その消費者が興味を持っているであろう商品やサービスを AI が予測し、その予測結果に基づき一人一人の個人の嗜好に合わせてターゲティングされた広告が表示されています。

---

1 新村出編「広辞苑〔第七版〕」（岩波書店、2018）。

このほかにも、①個人の健康診断結果と入院や手術等の病歴データ等によって学習したAIを用いて、当該個人の将来の疾病予測を行い、それに従い保険加入審査を行うもの、②特定の企業の従業員データを含む様々な情報によって学習したAIを用いて、当該企業・ポジションに対する候補者の適性を予測し、候補者の中から最適な人材を企業に推薦する人材紹介サービス、③個人の家族形態や趣味嗜好等、一見、与信評価とは無縁とも思われる各種データを含む、個人に関するあらゆるデータから、当該個人の信用度(信用スコア)を予測するもの[2]、④過去の犯罪データ等によって学習したAIを用いて、特定の個人について犯罪を犯す傾向の有無、レベルの予測を行い、犯罪の予防を行うものなど、様々な分野で、AIによる個人の性質の予測・推測を利用したサービスが提供されています。

そして、このような、AIによって行う個人の性質の予測・推測のことを、プロファイリングと呼んでいます。

2. プロファイリング実施に関する諸問題

プロファイリングの普及により、例えば、事業者は、より効果的で効率的な広告を発信することができるようになり、広告費の削減及び売上の増加といったメリットを享受することができるようになります。また、一方で、消費者も、より効率的に、自分に合った商品を見つけることができるようになり、時間の節約やより高い満足度を得ることができるようになります。加えて、AIの与信審査により、それまで審査に落ちていた人が審査に通るといったことも起こり得ます。

しかし一方で、飛躍的に高度となったプロファイリングには、種々の法的問題が伴うことが指摘されており、とりわけ、個人の尊重やプライバシー等といった個人の人格的利益に関わる権利との衝突が指摘されています。したがって、プロファイリングを積極的に利活用するに当たっては、企業として、このような個人の重要な権利との衝突という問題[3]が存在することを認識し、必要な対策を講じておくことが重要となります。そうでなければ、思わぬところで炎上やネガティブキャンペーン等によるレピュテーションの毀損を招きかねません。

(1) 差別的プロファイリングが導かれ得る問題性

AIの学習用データセットの中に、差別的なデータが紛れていた場合、当該データで学習したAIによってプロファイリングされた予測結果は、差別的な内容になることが想定されます。

例えば、採用候補者の職務遂行能力をプロファイリングするAIが、女性差別的なデータによって学習されていたとすると、当該AIは、「女性」というカテゴリーに属しているというだけで、職能を低く予測し、その結果、多くの女性が採用候補から落とされるという事態が発生することが考えられます[4]。

このような結果は、平等原則に違反する可能性があるものであり、場合によっては、不法行為等による損害賠償請求の対象ともなり得るものです。しかし、AIのアルゴリズムが十分にオープンとなっていない場合、企業は、学習データや当該AIの出力が女性差別的となっていることを認識できない可能性があります。また、女性候補者は、AIのアルゴリズムが分からない以上、低く予測された職能結果に対して、効果的な反論を行うことが不可能となるおそれがあり、AIによる誤った評価(女性であることを理由とする低評価)を覆す機会が奪われることとなります(しかも、下記(2)のように、場合によっては、その評価が最終評価として定着し、当該女性に永続的につきまとうという事態も生じ得ます)。

(2) AIによる評価が最終評価として受け入れられることの問題性

AIによるプロファイリングは、データの質・量と、AIの性能自体の飛躍的向上により、その精度が高くなってきています。しかしその予測結果は、あくまで、当該AIによる統計的、確率的な予測・評価に過ぎません。

---

2 このような信用スコアは、融資の際に用いられるほか、賃貸住宅への入居審査や就職試験等でも用いられています。また、究極的には、あらゆるサービス業において、顧客の信用評価を用いてサービス提供の可否を判断するということもあり得ないことではありません。

3 以下で述べる問題点は、近時議論されているプロファイリングに関する問題点の全てではありません。特に、政治的活動に関連する諸問題については、ここでは一切触れません。

4 実際、米国のAmazon社において、AIを活用した人材採用システムを運用していたところ、最近、当該AIが女性差別的な評価を行う傾向にあることが分かったという理由で、当該AIを用いた採用を取りやめたと報道されています(https://newspicks.com/news/3376927/body/?ref=user_3595220 参照)。

例えば、[A, B, C, D]という属性を持つ人間は[X]という傾向があるという予測を行う AI があったとして、実際に[A, B, C, D]という属性を持つ個人が AI によってプロファイリングされると、統計的、確率的に当該個人は[X]という傾向があると評価されることになります。しかし、本当に当該個人が[X]であるかどうかは、[A, B, C, D]という属性だけではなく、その人個人を実際に評価しなければ明らかとはならないはずです。しかし、AI が飛躍的に高度化したことにより、AI に基づくプロファイリングは、高度に科学的、合理的な予測結果＝最終結論として、人々に広く受容されることが想像に難くありません。

このような事態は、個人の尊重原理や幸福追求権を侵害する可能性があるもので、やはり場合によっては、不法行為等に基づく損害賠償の対象ともなる可能性があるものです。

(3) プライバシー権を侵害する問題性

AI によるプロファイリングは、個人のセンシティブ情報をあぶり出すこともあります[5]。例えば、妊娠や鬱状態といった事柄や、通常人であれば私生活における秘密として秘匿しておきたいと考えるセンシティブ情報を、AI によるプロファイリングで予測することは可能です[6]。そして、このようなセンシティブ情報を、本人に無断でプロファイリングによって取得することは、プライバシー権に対する侵害となるとの主張があります。

この点については、上記(2)のとおり、プロファイリング情報は、あくまで統計的・確率的推測に過ぎないことから、特定個人のプライバシーに対する侵害とはならないという考え方もあります。一方で、プライバシー侵害に基づく損害賠償請求に関する先例的判決である「宴のあと」事件判決[7]は、プライバシー侵害の要件として、「公開された内容が…私生活上の事実または事実らしく受け取られるおそれのある事柄であること」と判示しています。したがって、科学的、合理的な予測結果であるプロファイリング結果は、「事実らしく受け取られるおそれのある事柄」としてプライバシーの侵害に当たるとする考え方もあります[8]。

このように、プロファイリングを行う事業者が、プロファイリングによってセンシティブ情報を取得する場合は、当該行為がプライバシー権の侵害に当たる可能性があり、あるいは取得したセンシティブ情報の保存、利活用行為がプライバシー権の侵害に当たるとされる可能性もあり、ひいてはプライバシー権侵害に基づく損害賠償請求がなされるリスクもあるということになります。

(4) プロファイリングが意思決定過程を不当に操作する問題性

上記(3)に加えて、過度なプロファイリングとそれに基づくターゲティング広告は、自己決定権に対する不当な干渉と評価される可能性があります。

前掲注 5 の例において、プロファイリングの結果、鬱状態にあると予測される女性に対して、化粧品のターゲティング広告を行うことは許されるのでしょうか。広告は、その性質上、消費者の心理状態に作用し、自社製品に対する関心を惹きよせ、もってその購入意思を生じさせるものです。しかし、鬱状態にある消費者は化粧品の購入傾向が高まるというデータに基づき、AI を利用して鬱状態にある消費者を探し、当該消費者にピンポイントな広告を配信し購入を動機付けさせるということは、個人の内心の自由を侵害し、個人の意思決定を外部から操作したものとして、自己決定権に対する不当な侵害ではないかということが議論されています[9]。

3. GDPR におけるプロファイリング対応

AI によるプロファイリングの積極的な利用は、事業者に大きな利益をもたらすだけでなく、消費者にも大きな便益をもたらすものです。しかし、一方で、上記 2.のとおり、個人の人格的利益に関わる基本的権利を侵害し得るものであ

---

5 実際に、アメリカの小売業者であるターゲット社が、無香料性ローションや特定のサプリメントといった一見妊娠との関連性が明確でない購入履歴から、顧客が妊娠していることや妊娠予定日までをもプロファイリングし、ターゲティング広告を行っていたことや、女性は鬱状態にあると感じるときに化粧品の購入傾向が高くなるとされているところ、既に臨床的兆候が出る前に鬱状態にあることを予測するアルゴリズムが構築されており、化粧品広告に利用されうる状況であることなどが日本にも紹介されています（山本龍彦「ビッグデータ社会とプロファイリング」論究ジュリスト 18 号（2016 年）35～37 頁参照）。

6 なお、妊娠や鬱といった医師による診断結果は、個人情報保護法上の「要配慮個人情報」に該当します。

7 東京地判昭和 39 年 9 月 28 日下民集 15 巻 9 号 2318 頁。

8 山本・前掲注 5）38～39 頁。

9 山本龍彦「ビッグデータ社会における『自己決定』の変容」NBL1089 号 30～31 頁。

ることも事実です。そこで、そのような法的問題点を踏まえた適切な対応を予め実施しておくことが有益です。この点、GDPR では、プロファイリングに関する規定が設けられており、日本における対応を検討する上で参考になります。以下では、GDPR におけるプロファイリングに関する代表的な規定を見ていきます。

(1) 異議申立権

GDPR では、管理者[10]又は第三者によって求められる正当な利益の目的のために個人データを取り扱う場合(この取扱いにプロファイリングが含まれます)等に、データ主体は、いつでもその取扱いに対して異議を述べることができます(GDPR 21 条 1 項)。そして、管理者は、当該異議が申し立てられた場合、やむを得ない正当な根拠を証明しない限り、当該個人データを取り扱うことができなくなります(同項)。

また、個人データがダイレクトマーケティング目的で取り扱われている場合、データ主体は、いつでも、当該マーケティングのための自己に関する個人データの取扱い(ダイレクトマーケティングに関するプロファイリングが含まれます)に対して、異議を述べることができます(GDPR 21 条 2 項)。そして、管理者は、当該異議が申し立てられた場合は、個人データを当該目的で取り扱うことができなくなります(同条 3 項)。

さらに、管理者は、このような異議申立権が存在することを、遅くともデータ主体に対して初めて連絡をするまでに、データ主体の注意を引くように明示的に、他の情報とは分けて表示しなければなりません(GDPR 21 条 4 項)。

(2) 自動処理のみに基づく決定をされない権利

GDPR では、データ主体は、自らに関する法的効果を発生させ、又は、同様の重大な影響を及ぼす自動処理(プロファイリングが含まれます[11])のみに基づく決定の対象とならない権利を有しています(GDPR 22 条 1 項)。但し、データ主体の明示的な同意がある場合等、一定の例外事由が定められています(同条 2 項各号)。

また、これらの自動処理に基づく決定が存在することや、当該決定に含まれる論理、重要性、データ主体に生じると想定される結果に関する情報は、データ主体に対して提供されなければならず(GDPR 13 条 2 項(f)、14 条 2 項(g))、さらにデータ主体は、これらの情報にアクセスする権利も有しています(GDPR 15 条 1 項(h))。

(3) GDPR における対応の性質

上記の規定を、語弊を恐れずに単純化すれば、個人は、プロファイリングの対象から除外してもらう権利を有し、さらに、重大な局面においてプロファイリングのみで自己を判断されない権利が確保され、それらの権利を実効化するために、管理者側が、個人に対して、権利の存在を十分に通知しなければならないということになります。

また、上記 2.で述べた問題点との関係性を若干敷衍すると、GDPR 21 条により、個人が自己決定によってプロファイリングの可否を判断する機会が確保されますので、プロファイリングに関する問題全般に有効であるということができます。加えて、GDPR 22 条は、上記 2.(1)や(2)で述べたような誤った評価を覆す機会が奪われたり、AI による評価が最終評価として永続的に受け入れられるといった問題点に対して、特に有効に作用するものといえます。

4. 日本において考えられる対応策

日本の個人情報保護法には、プロファイリングに対する規定は設けられていません。もっとも、先の 2015 年改正に向けて作成された、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大綱」16 頁(2014 年)では、「プロファイリングの対象範囲、個人の権利利益の侵害を抑止するために必要な対応策等については、…継続して検討すべき課題とする。」とされており、今後の個人情報保護法改正の際に、プロファイリングに関する規定を設けることが検討の対象となる可能性は否定できません。

また、これから更なるビッグデータ及び AI の利活用時代へ突入していくに際して、(データ)プライバシーへの積極的な対応そのものが、企業が取るべき戦略的なオプションとなっていくものと考えられます。したがって、法令というあ

---

10 「管理者」とは、GDPR 上で定義されている用語ですが、さしあたり、個人情報を取り扱う者という意味で理解しておけば問題ありません。

11 なお、プロファイリングを含まない自動処理もあります。例えば、オービスによるスピード違反取締は、個人データに基づき当該個人を評価しているわけではありませんので、プロファイリングを伴わない自動処理です。

る種のミニマムスタンダードに規定が設けられているか否かにかかわらず、積極的にプロファイリングへの対応を実施し、その状況を発信していくことが企業価値を高めることに繋がるといえます。

そこで、上記 3.の GDPR の規定を参考にしつつ、プロファイリングへの対応策について簡単に検討します[12]。

(1) 透明性の徹底

企業戦略としての(データ)プライバシーへの対応という観点からは、透明性を徹底することが重要となります。すなわち、プロファイリングを行う側は、どのようなデータを取得し、どのようなデータを利用して、プロファイリングによって被評価者のどのような側面を予測し、そのプロファイリング結果をどのように保管し、利用するのかといった情報を被評価者に対して提供することで透明性を確保することが重要です。被評価者が消費者であるにせよ、従業員であるにせよ、このような透明性が確保されていないことにより、不安感、不信感が醸成され、それが何かのきっかけで火が付き、いわゆる炎上といった事態が発生し得ます。こうした事態を未然に防ぐためには、透明性を確保し、プロファイリングを行う側で一体何を行っているのかということを、可能な限り明らかにしておくことが重要となります。

また、透明性の確保と併せて、プロファイリングを実施することで被評価者(あるいは社会)に対してどのようなメリットがあるのかということを明確に告知することで、その有用性を理解しプロファイリングの実施に積極的になる被評価者も存在すると思われます。そして、このような十分な情報提供を行った上でのプロファイリングであれば、被評価者の人格権が合理的に保護されていると評価できると考えられます。

(2) 異議申立権(GDPR21 条)と同種の対応

次に、被評価者がプロファイリングを望まない場合には、被評価者からの申出によりプロファイリングを停止する仕組みを整えておくことが有用であると考えられます。例えば、ターゲティング広告を行っている場合は、被評価者が、プロファイリングに基づくターゲティング広告の実施を停止させることができる仕組みを整備することが考えられます。

被評価者が、事後的に、プロファイリングから離脱する仕組みを整えることで、被評価者の人格権が合理的に保護されることになります。なお、その前提として、どのような情報からどのようなプライバシー情報をプロファイリングしているのかといった情報が被評価者に提供されている必要があり、上記(1)の対応がこれに資することになります。

(3) 自動処理のみに基づく決定をされない権利(GDPR22 条)と同種の対応

さらに、GDPR 22 条と同様にプロファイリング結果を最終評価として扱わないようにするということが一つの対応として考えられます。この対応は、特に上記 2.で述べた弊害のうち、(1)と(2)[13]に対して有効です。例えば、人事評価や与信評価に際して、AI によるプロファイリングを 1 次評価として導入することは、評価者毎のばらつきや大量の案件を即座に処理するという面で評価者側にも被評価者側にもメリットがあります。しかし、それを最終評価とするのではなく、必ず人間による最終的な評価を経るようにすることで、上述の法的問題に対応することができます[14]。

また、人間による最終評価に際しては、AI がそのように評価した理由を被評価者に対して説明できるようにすることが望ましいと考えられます。例えば、人事評価においては、そのような評価に至った理由を説明しなければ、被評価者も、評価を踏まえた自己の対応を検討することができません。そして、評価者において、被評価者に対し評価理由を説明できるようにするためには、プロファイリング結果を批判的に検討することが必要となり、更にそのためには AI のアルゴリズムの少なくとも根幹的な部分については、理解している必要があると思われます。

---

12 なお、近々、「パーソナルデータ＋α研究会」より、プロファイリングに関する提言案が公表されるという情報に接していますので、公表後、機会を設けてこの提言案についても、ニュースレターにてお届けできればと考えております。

13 上記 2.(2)で述べた、プロファイリング結果は確率的・統計的予測に過ぎないという点に対しては、被評価者の属性情報の細分化を推し進めることで、予測結果を限りなく被評価者本人の実際に近付けるという議論がなされることがあります(山本龍彦「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』(有斐閣、2018)92～93 頁)。しかし、属性を細かくするということは、被評価者の個人データを次々とインプットしていくということを意味し、これは被評価者のプライバシーを丸裸にするという別の問題を提起します。

14 前掲注 4 の Amazon 社の事例でも、Amazon 社は、AI による評価を最終評価とはせず、人間による評価を行った上で最終決定を行っていたと報道されています。

Let's talk

個別案件につきましては、下記の問い合わせ先までお問い合わせください。

PwC 弁護士法人
〒100-6015　東京都千代田区霞が関 3 丁目 2 番 5 号　霞が関ビル
電話 : 03-5251-2600（代表）
Email: pwcjapan.legal@jp.pwclegal.com
URL: https://www.pwc.com/jp/ja/services/legal.html

- PwC ネットワークは、世界 90 カ国に約 3,500 名の弁護士を擁しており、幅広いリーガルサービスを提供しています。PwC 弁護士法人も、グローバルネットワークを有効に活用した法務サービスを提供し、PwC Japan 全体のクライアントのニーズに応えていきます。
- PwC Japan は、PwC ネットワークの各法人が提供するコンサルティング、会計監査、および税務などの業務とともに、PwC 弁護士法人から、法務サービスを、企業の皆様に提供します。

弁護士
マネージャー
渡邊満久
mitsuhisa.watanabe@pwc.com