プロファイリングと情報銀行

慶應義塾大学法科大学院　教授
山本龍彦
yamamomo@wg8.so-net.ne.jp

1. プロファイリングの定義

例えば、以下のような定義がある。「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」（パーソナルデータ+α研究会「プロファイリングに関する提言案」（NBL1137号）。同様の定義は、AI法研究会・プライバシー部会「AIプロファイリングの論点」（2021年7月8日スライド）

2. プロファイリングのリスク

①プライバシー権侵害
→特に要配慮個人情報をプロファイリングする場合。元情報〔例えば脈拍情報〕については、情報銀行等と共有・提供することに同意していたとしても、例えば鬱状態にあるかどうかをプロファイリングされ、その結果情報を共有・提供することまでは同意していない場合、本人にとっては（ヘルスケアといった利用目的の範囲内としても）不意打ちとなりうる。推知情報だとしても、「真実らしく受け取られる」可能性がある〔宴のあと事件判決。東京地裁昭和39年9月28日〕
※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として、千葉恵美子・判解ジュリスト1518号78頁、窪田充見編『新注釈民法（15）』539、542頁（水野謙）。

②不当な差別・選別

（a）ブラックボックス問題：スコアの算出にどのような情報が使われたのか、どの情報にどれぐらいの比重がかけられて算出されたのかが明らかではない（いわゆる「ブラックボックス問題」）。

（b）確率の評価と自動バイアス：AIの評価は限定的データに基づく確率的な評価にすぎない。しかし人間には、コンピュータのはじき出した結果を信じるバイアス（自動バイアス）がある（vs. 個人の尊重）。

（c）不適切なデータの混入可能性と、検証困難性：本人の評価に本来使うべきではない情報が混入し、本人の信用スコア等に影響を与える可能性がある。
※リクナビ問題

（d）差別の再生産問題　これまで存在してきた社会の差別構造が、アルゴリズムを設計する際のデータセットの偏りや、性別等のセンシティヴ属性が算定に使用されることでスコアに反映し、差別構造が固定化ないし悪化する可能性がある。センシティヴ属性そのものを使わなかったとしても、それと密接に関連する情報（代理変数）が使われることで、結果的にマイノリティに「異なるインパクト（disparate impact）が生じることもある。
※遺伝的情報が算定に使われると、生まれによる差別が部分的に復活する可能性もある。

（e）バーチャル・スラム問題（結果の「ひとり歩き」によるスティグマ化可能性）　スコアの利用範囲が拡大すると、低スコアの者は社会の至る所で事実上の不利益を受けるうえ、スコアの算定基準が不透明であることでスコアアップの方法もわからず、その境遇が社会の下層で固定化してしまう可能性がある。スコアが独り歩きすることによるスティグマ化。

③行動の萎縮効果

④民主主義への影響：ケンブリッジ・アナリティカ事件、デジタル・ゲリマンダリング

⑤自己決定権への介入：マイクロターゲティング広告による強い誘導、ダークパタン。悪質なものでなくても、選好の固定化、選択肢の縮減、セレンディピティの縮減

3. 国内外の状況（→既に前回資料で紹介あり。ここでは一部）

（1）GDPR

〇22条：完全自動化決定の原則禁止

→データ主体は、当該データ主体に関する法的効果を発生させる、又は、当該データ主体に対して同様の重大な影響を及ぼすプロファイリングを含むもっぱら自動化された取扱いに基づいた決定の対象とされない権利を有する。

※原則として、プロファイリングやスコアリングの結果のみで、つまりコンピュータの自動処理のみで、個人の人生に重大な影響を与える決定を行っていけない。

→（明示的な同意がある場合などは、例外的に完全自動化決定を行うことも許されるが、その場合にも）データの管理者は、データ主体の権利及び自由並びに正当な利益を守るための適切な措置を実装しなければならない。少なくとも、人間の関与を得る権利、自らの見解を表明する権利及びその決定を争う権利を保障しなければならない。

→ガイドラインは、個別的文脈におけるデュー・プロセスを超えて、「適切な措置」として、システミックなアカウンタビリティの手段（監査制度と倫理委員会）を含むとしている。

○13条～15条:情報提供義務、アクセス権

→プロファイリングを含め、第 22 条第 1項及び第 4項に定める自動的な決定が存在すること、また、それが存在する場合、その決定に含まれているロジック、並びに、当該データ主体にとっての重要性及びデータ主体に生ずると想定される結果に関する意味のある情報。

ARTICLE 29 DATA PROTECTION WORKING PARTY, GUIDELINES ON AUTOMATED INDIVISUAL DECISION-MAKING AND PROFILING FOR THE PURPOSE OF REGULATION 2016/679, 17/EN. WP 251rev.01(Feb. 6, 2018)

→「アルゴリズムの複雑な説明」や「完全なアルゴリズム（full algorithm）の開示」を求めてはいない。「背後にある理論的根拠（rationale）、または決定に至るまでに依拠した基準（criteria）について、データ主体に対して伝えるシンプルな方法（simple way）を探すべき」とされる。具体的には、管理者は、①データ主体に対して、このようなタイプの活動にかかわっていることを伝えること、②ロジックに関して意味のある情報を提供すること、③この手続がもつ重大で想像される結果を説明すること

※少なくとも教師データの構成要素（包摂データ、排除データ）、アルゴリズムに反映させた政策的選択（調整基準も含む）、社会的影響評価、監査方法などが「意味のある透明性」にとっては重要であろう。

○35条:データ保護影響評価（DPIA）

→自然人の権利及び自由に対する高いリスクを発生させるおそれがある場合、管理者は、その取扱いの開始前に、予定している取扱業務の個人データの保護に対する影響についての評価を行わなければならない。

第1項に規定するデータ保護影響評価は、とりわけ、以下の場合に求められる:

(a) プロファイリングを含め、自動的な取扱いに基づくものであり、かつ、それに基づく判断が自然人に関して法的効果を発生させ、又は、自然人に対して同様の重大な影響を及ぼす、自然人に関する人格的側面の体系的かつ広範囲な評価の場合;

※DPIAは、自動意思決定「のみ」に依存しない意思決定についても義務的に要求される。この条文からは意識的に「solely」が削られている。作業部会は、繰り返し、35条の義務的なDPIAが、「完全に自動化されていない（not wholly automated）」場合でも要求されると述べている。

→ガイドラインによれば、DPIAの公表はGDPRの法的要求ではない。しかし、管理者は、「公表について検討すべきである」とされる。なお、「公表されたDPIAは、すべてのアセスメントを含む必要はない。特に、DPIAが、データ管理者にとってセキュリティリスクとなるような具体的な情報を含み、取引の秘密や商業的にセンシティブな情報を明らかにするような場合には……DPIAの主たる考えの要約でもかまわない」。

→「このようなプロセスの目的は、管理者の処理運営に対する信頼の醸成を助け、アカウンタビリティと透明性を実現することにある」。

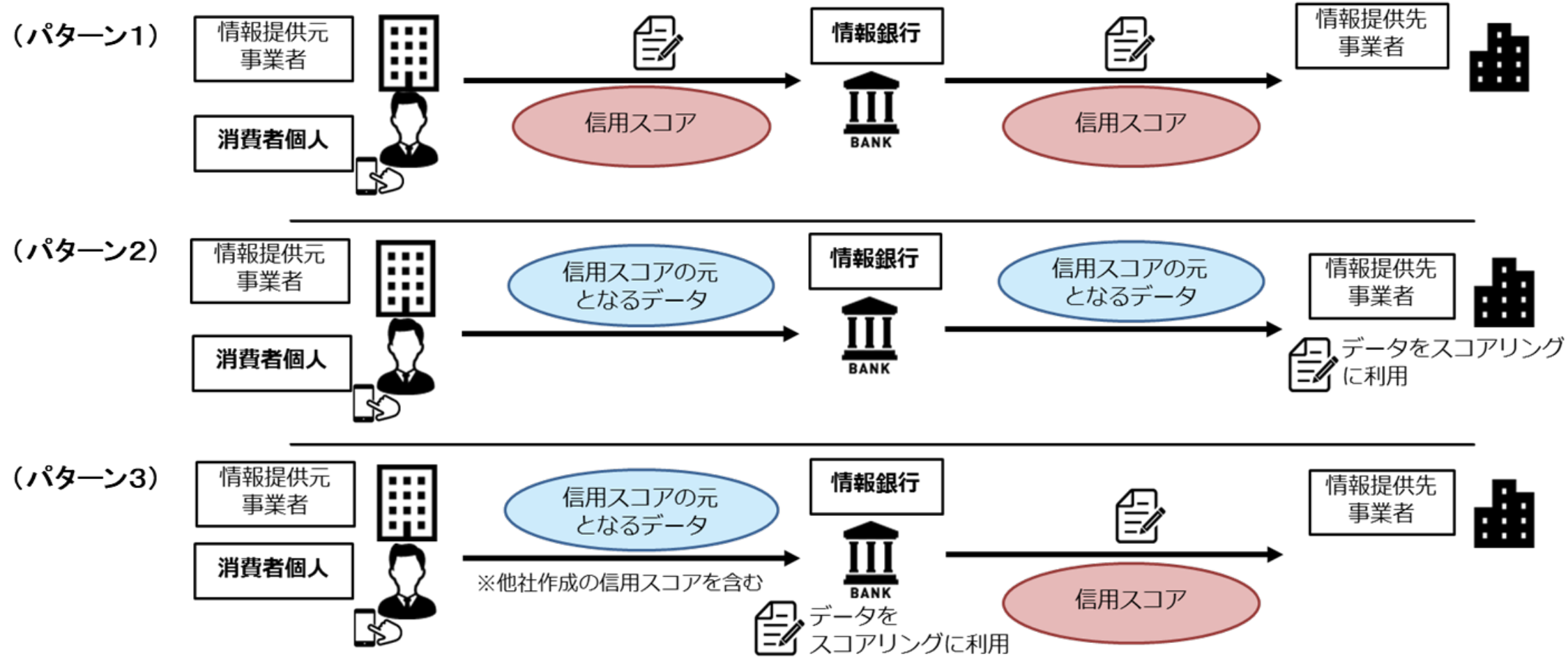
(2)欧州委員会「人工知能の利用を制限する包括的な規制案」(2021年4月)

→リスクベースアプローチを採用
①最小限のリスク
②限定的なリスク
③高リスク
④禁止
→採用・人事、教育、福祉などは③に分類(人の評価にAIを使う場合)
→第三者による事前審査制、リスク管理、ガバナンス、透明性、人間が監視できる仕組みetc.

※なお、AIを使ったサブリミナル的な行動変容は④に。

4. 情報銀行とプロファイリング

■ 情報銀行が「信用スコア」を取り扱う場合のパターン



（1）要配慮プロファイリング

・要配慮プロファイリングと一般的プロファイリングを区別すべきではないか。

①要配慮プロファイリング

→要配慮個人情報（※）を推知するプロファイリング等（対象者に重大な不利益を与えうる可能性のあるプロファイリング）

※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（個情法2条3項）

（例）疾患予測、センシティヴな身体および精神状態の予測、社会的信用力の予測、人事採用・人事考課のための適性・能力の予測、政治的信条の予測、犯罪傾向の予測

※キャンセル傾向の予測は？

②要配慮プロファイリングの取扱いに関する論点（叩き台）

（i）禁止カテゴリーの創設
→犯罪傾向の予測は本人にとって特に利益を生まない。
→政治的信条の予測は選挙や民主主義を不当にゆがめる可能性もあり、信頼される情報銀行として、同意があっても取扱うべきでないと解する余地もあるのではないか。
（ii）使用・提供禁止データの創設
→遺伝情報など、本人が努力しても変更・修正できない情報を要配慮プロファイリング・スコアリングに利用すべきではないと解する余地もあるのではないか（パターン3）。また、情報銀行は、これらの情報を要配慮プロファイリング・スコアリングを行う者に対して提供すべきではないと解する余地もあるのではないか（パターン2）
（iii）厳格な同意取得
→情報銀行による実施（パターン3）、提供先による実施（パターン2）について明示的に説明し（重要事項として強調し、リスク等についても説明することが求められる）、事前に同意を得ること。
（iv）説明責任・透明性の徹底

③データ倫理審査会の役割――ガバナンス体制

○情報銀行が行う場合（パターン3）

（i）事前審査（要配慮プロファイリングを実施する場合には、事前にデータ倫理審査会の審査を経ること。欧州委員会「人工知能の利用を制限する包括的な規制案」の事前審査制参照）
（ii）説明項目の検討（透明性確保）
（iii）定期的なHRIA（Human Rights Impact Assessment。アルゴリズムの公正さのチェック。差別的インパクトの査定とアルゴリズムへのフィードバック。なお、データ倫理審査会運用ガイドライン8.2.1参照）
（iv）個別審査（プロファイリング結果〔スコア等〕に対する異議申立てへの対応。アカウンタビリティ）
（v）提供先による利用状況の審査（提供先は情報銀行倫理審査委員会に対して報告義務を負う）
※必要に応じて、アルゴリズムの公正等を専門的に審査できるWGを設置することも考えられる（生貝構成員のコンフォミティ・アセスメント？）。
※必要に応じて、監査機関による監査を受けることも考えられる。
※苦情等を受けつける窓口の設置も検討されるべき（特に一定の決定に利用する場合、人間が関与する必要性も考慮）。

○情報提供先事業者が要配慮プロファイリングを行う場合（パターン2）

・提供先にデータ倫理審査会に類似する機関（同等機関）があり、上記機能を有しているかどうかを審査する（データ倫理審査会運用ガイドライン8.2.5参照）。
・提供先の同等機関からHRIAの報告等を受け、適切性を審査する。

○情報銀行がスコア等を利用する場合（パターン1、パターン3）

・要配慮プロファイリングの結果（スコア等）が目的に従って利用されているかを審査
・スコア等を踏まえて、個人に関する重要な決定を行う場合には、上記（iv）を行う（パターン1の場合、情報銀行が有効な個別審査を行うためには、アルゴリズム等に関する一定の情報を実施者〔提供元〕から得なければならない）。
・パターン1の場合、実施者（提供元）に対する報告義務

(2)その他のプロファイリング

①一般的プロファイリング(仮)
→レコメンドやターゲティング広告のために、性別、年代、趣味・嗜好などを予測することを含む(消費者に不利益を与える可能性が低いもの)。

②一般的プロファイリングの取扱いに関する論点(叩き台)

・本人に対しては、プロファイリングの有無と目的＋例示の義務(例えば、閲覧履歴や購買履歴等から、性別・年代を分析・予測している、など。パターン3、パターン2ともに)。
※ガイドライン(通則編)改正案＋α

③データ倫理審査会の役割
○パターン3
・プロファイリング・リストを作成させたうえで、そのなかに要配慮プロファイリングが含まれていないかを審査

○パターン2
・提供先からプロファイリング・リストを受け取り、そのなかに要配慮プロファイリングが含まれていないかを審査