英国・フランスにおける監視カメラ・顔認識技術の利用と規制動向

2020年1月

国際社会経済研究所

小泉　雄介

y-koizumi@pd.jp.nec.com

欧米における顔認識技術の利用・規制動向の概要

- 近年の顔認識技術の精度向上により、空港、小売店、ショッピングセンター、ホテル、フィットネスセンター、交通機関、図書館、オフィスなど、様々な場面で個人を識別するために顔認識技術を利用する事例が増えてきている。
- ①スマホやPCのログイン、空港での入出国管理、テーマパークの入場管理など、本人の同意の下で個人認証（本人確認）の目的で行われる顔認識サービス（いわゆる顔認証：Facial Authentication）（多くは1-to-1）に対しては、懸念する声は少ない。
- ②また、警察機関が容疑者の顔写真と犯罪者データベース等の顔画像を捜査目的で照合する顔照合（Facial Matching）（1-to-many）に対しても、懸念する声は少ない。
- ③しかし、公道や店舗などで（本人同意なく）不特定多数の人々を対象に行われる顔認識、とりわけリアルタイムで個人を識別する自動顔照合（Automated Facial Recognition: AFR あるいは Live Facial Recognition: LFR）（many-to-manyまたはmany-to-1）についてはプライバシー等の問題が指摘され、欧米で同時並行的に規制化の動きが進んでいる。
- 欧米における規制化の動き
  - イギリス：　内務省の監視諮問委員会の設立、南ウェールズ警察に対する訴訟、データ保護監督機関（ICO）による調査など
  - アメリカ：　サンフランシスコ市の顔認識禁止条例制定、連邦法案など
  - EU：　一般データ保護規則（GDPR）の「ビデオ機器を通じた個人データ処理に関するガイドライン案」など

※「欧米におけるカメラ・顔認識サービスと規制動向」https://www.i-ise.com/jp/information/report/2019/20191029_facial_recognition.pdf もご参照ください。

1．英国（イギリス）の動向

2．フランスの動向

英国における顔認識技術の利用事例

- ①本人同意の下での個人認証（顔認証）
  - ヒースロー空港では、入国審査時に、ePassport gate（e-Gate）で顔認証を実施。従来は英国、EEAおよびスイス国民のみが対象だったが、2019年5月からオーストラリア、カナダ、日本、ニュージーランド、シンガポール、韓国、米国民に対象者が拡大された。（ヒースロー空港のみならず、エジンバラ空港、マンチェスター空港など国内主要空港でも同様。）
  - ブリティッシュ・エアウェイズは、ヒースロー空港のいくつかの出発ゲートに顔認識デバイスを導入。搭乗券とパスポートを見せることなく搭乗が可能に。
  - ロンドンのカジノ（Hippodrome Casino）での「自己除外リスト」登録者の顔認証による入店拒否。
- ②容疑者写真の顔照合
  - 犯行現場における容疑者の写真等と警察の保有する拘留者データベースとを捜査目的でマッチングさせる顔照合（Facial Matching）技術は、英国の警察で広く使われている。
- ③公共空間等での不特定多数に対する自動顔照合（本人同意なし）
  - 英国警察による自動顔照合（AFR：Automated Facial Recognition, またはLFR：Live Facial Recognition）
    - レスターシャー警察：　屋外音楽コンサート
    - ロンドン警視庁：　ノッティングヒル・カーニバル
    - 南ウェールズ警察：　欧州サッカー連盟チャンピオンズリーグの試合
    - グレーターマンチェスター警察：　ショッピングセンター
  - 英国での民間企業による自動顔照合
    - ロンドンのキングスクロス再開発地（不動産会社による公共空間での自動顔照合）
    - 小売店での万引き犯顔照合（Facewatch）：「最大15%のアラートは偽陽性（誤照合）」

③警察による自動顔照合（AFR、LFR）の実証実験（英国）

時期	実施主体	実施イベント	顔照合データベースの内容
2015年6月	レスターシャー警察	屋外音楽イベント（ロックフェスティバル）	レスターシャー警察の拘留者DB、およびユーロポールから得た国際犯の顔写真DB
2016年8月	ロンドン警視庁	ノッティングヒル・カーニバル	カーニバルへの参加を禁じられた人や、犯罪を行うためにカーニバルに参加する可能性があるとして警察が指定した人（組織犯罪者等）
2017年6月	南ウェールズ警察	欧州サッカー連盟チャンピオンズリーグの決勝戦（南ウェールズのカーディフ）	組織犯罪者・違法チケット販売者・フーリガンなど50万人のDB（スタジアムのみならずカーディフ市内全域で顔照合）
2018年	グレーターマンチェスター警察	グレーターマンチェスターのショッピングセンター →監督機関によって中止	30人の容疑者や行方不明者の顔写真データ

③警察による自動顔照合(AFR、LFR)の実証実験(英国)

時期	実施主体	実施場所	実験結果
2018年7月	ロンドン警視庁	Westfieldショッピングセンター	データベース件数:306 アラート発生数:1 逮捕者数:0
2018年12月	ロンドン警視庁	Westminster	データベース件数:2226 アラート発生数:5 逮捕者数:2
2019年1月	ロンドン警視庁	Romford High Street	データベース件数:2500 アラート発生数:10 逮捕者数:2
2019年2月	南ウェールズ警察	カーディフのCity Centre	データベース件数:830 アラート発生数:12 逮捕者数:3

③自動顔照合(AFR、LFR)に対する懸念

- 自動顔照合の対象となる市民、個人への透明性の欠如
  - 英国の監視カメラコミッショナーによれば、「レスターシャーの事例では、自動顔照合を行なうことに関する通知はチケットの裏面に小さな文字でなされたのみであり、それに気付いた参加ミュージシャンが反対声明を出すなど、かなり大きな問題になった」「自動顔照合の問題は、市民は撮影されていることには気付いても、データベースと照合されていることについてはわからないことだ」とのことである。
- 顔認識技術は他の個人データ取得技術に比べてプライバシー侵害リスクが高い
  - 英国議会の2018年の報告書は、「顔画像は本人が知ることなく容易に取得され保持される。また、顔写真データベース(パスポート、運転免許証、拘留者画像)は既に成人人口の90%をカバーしているため、顔認識技術は他の生体認証技術よりも重大な倫理的問題が存在する」としている。
  - 「犯罪とは無関係の一般市民に対して一律に顔照合をかけてよいのか」「スタジアムに入るときに全ての観客が指紋採取をされているようなものではないか」という批判も。
- 自動顔照合は行動の自由を萎縮させる
  - 「政府による顔認識の利用は、民主主義の自由と人権を侵害する可能性がある。人々が自由に集まり、意見を交換することによってこそ民主主義は成立する。顔認識の活用には人々の自由にリスクをもたらしうるものもある。政府は顔認識を利用して、特定個人の長期的監視を行うことができる。」(MicrosoftのCLO)

英国におけるカメラ・顔認識に関連した法令・ガイドライン・制度

- 法令
  - 2018年データ保護法： EUのGDPRおよび警察・刑事司法データ保護指令の下での新法
  - 2012年自由保護法： 地方自治体や警察によるカメラ設置を規制
- 第三者機関
  - 情報コミッショナー・オフィス（ICO）
    - 個人データ保護全般を監督。日本の個人情報保護委員会に相当。
  - 監視カメラコミッショナー（SCC）
    - 監視カメラに特化した監督機関。
- ガイドライン
  - CCTV行動規範（2014/2015年）: ICOが策定
  - 監視カメラ行動規範（2013年）： SCCが管轄

- 監視カメラに対する認証制度（2015年11月開始）
  - 監視カメラ行動規範の12原則を遵守していることを認証。
  - 認証マークはWebサイト等で使用可。
  - 40組織が認定取得（小売企業・病院・大学・警察等）。
    （2017年時点）

英国の2018年データ保護法

- 「所管官庁（competent authorities）」による「法執行目的（law enforcement purposes）」での個人データ処理は、2018年データ保護法（DPA）第3部によってカバーされる。
- とりわけ、法執行目的での自動顔照合（AFR、LFR）の利用は、個人をユニークに識別する目的での生体データの処理を伴うため、DPA第35条（8）bの「センシティブな処理」に相当する。
- そのようなセンシティブな処理においては、DPA第35条（第一のデータ保護原則）、第42条（保護措置：センシティブな処理）、第64条（データ保護影響評価）の要件にとりわけ注意を払わなければならない。
- DPA第35条（第一のデータ保護原則）
  - （1）第一のデータ保護原則は、法執行目的での個人データの処理は適法（lawful）であり、公正でなければならないというものである。
  - （2）法執行目的での個人データ処理は、それが法令に基づくものであり、かつ以下のいずれかを満たす場合に限り適法である。
    - (a)データ主体が当該目的での当該処理に同意を与えている場合、または
    - (b)所管官庁によって当該目的で行われる職務の遂行に当該処理が必要となる場合
  - （3）さらに、法執行目的での処理がセンシティブな処理である場合、当該処理は（4）項と（5）項で規定された2つのケースにおいてのみ許可される。
  - （4）第一のケースは、
    - (a)データ主体が法執行目的での当該処理に（2）項(a)にいう同意を与えている場合、かつ
    - (b)当該処理が実施される時点で、管理者が適切なポリシー文書（第42条参照）を用意している場合
  - （5）第二のケースは、
    - (a)当該処理が法執行目的で厳密に必要とされる場合、
    - (b)当該処理が別表8の条件の少なくとも1つを満たしている場合、かつ
    - (c)当該処理が実施される時点で、管理者が適切なポリシー文書（第42条参照）を用意している場合
  - （6）～（7）、（8）(a)(c)(d)省略
  - （8）「センシティブな処理」は、以下を意味する。
    - (b) 個人をユニークに識別する目的での、遺伝子データまたは生体データの処理

英国の2018年データ保護法

- 「別表8：第3部におけるセンシティブな処理の条件」で規定された条件は以下の9つ。
    - 法令上の目的（Statutory etc purposes）
    - 司法行政（Administration of justice）
    - 個人の生命に関する利益の保護
    - 子どもおよび危険にさらされている個人の保護
    - 既に公開されている個人データ
    - 法的手続きなど法律上の要求（Legal claims）
    - 司法行為（Judicial acts）
    - 詐欺の防止
    - アーカイブ目的等

最近の規制動向

- 内務省のバイオメトリクス戦略
  - 内務省は2018年6月28日にバイオメトリクス戦略「Biometrics Strategy: Better public services Maintaining public trust」を発行。内務省の取り組みとして以下を実施。
    - 内務省は法執行機関による顔画像と顔認識システムの使用についての検討を整合させるために、新たな監視・諮問委員会を設置。
    - 内務省や警察における新たなバイオメトリック技術の使用、あるいは既存のバイオメトリック技術の新たな適用に先立ち、関係機関に精査を求めながら、データ保護影響評価（DPIA）を実施。
    - 監視カメラコミッショナー（SCC）と連携して、監視カメラ行動規範を改定。
- 内務省の顔画像に関する監視・諮問委員会
  - 2018年7月に、「法執行機関における顔画像および新たなバイオメトリクスの利用に関する監視・諮問委員会」（Law Enforcement Facial Images and New Biometrics Oversight and Advisory Board）が立ち上げられた。
  - 同委員会の目的は、以下について、イングランドおよびウェールズにおける警察機関、並びに内務省およびその傘下機関による法執行目的での開発と利用を検討すること。
    - 顔画像の保存・照合システム
    - DNA、指紋、顔画像以外の新たなバイオメトリクス（音声、虹彩、指静脈、歩容認識を含む）
    - 警察機関が取得した顔画像の他機関との共有

最近の規制動向

- エセックス大学による評価報告書
  - 2019年7月に、ロンドン警視庁の自動顔照合（LFR）実証実験に対する批判的な評価報告書を発行。
  - 国内法でLFRを利用するための明示的な許可がないため、裁判所が異議を申し立てた場合、警察によるLFRの展開が違法と判断される可能性が高いと指摘。
  - 監視対象とする人物のリストに掲載される基準も明確ではなく、LFRで特定しようとしていた人々はカテゴリーもばらばらだった。リスト自体も正確さを欠き、すでに裁判が終わ（り無罪とな）った人物がリストに掲載されている事例もあったとのこと。
- また8月には南ウェールズ警察の自動顔照合（AFR）に対する訴訟の第一審があった。
  - 第一審は合法との判決。（→11月下旬に上訴された）
  - 監視カメラコミッショナー（SCC）は同判決に対し、「警察側がこの判決をAFRの一般的な展開に対するゴーサインと見なすことには注意を求める。AFRは、人権や国民の信頼に対する影響を伴う侵害的なツールである。適切な状況ではAFRの利用は合法でありうるが、法的枠組みの中で実証的に実施され、良いガバナンスと取組みの正当性を実証しなければならないという確信が高まってきている」という声明を公表。
- 情報コミッショナー・オフィス（ICO）は10月末に、警察による公共空間でのLFR（AFR）利用にする調査報告書、および意見書（→次頁以降参照）を公表。
- キングスクロス再開発地の事案に対しては情報コミッショナー・オフィス（ICO）が調査を開始。

【ご参考】英国ICOのライブ顔照合（LFR）に対する意見書

○「情報コミッショナーの意見：公共空間における法執行機関によるライブ顔照合技術の利用（Information Commissioner's Opinion: The use of live facial recognition technology by law enforcement in public places）」（2019年10月31日）

○サマリー

- コミッショナーは以前、ライブ顔照合（LFR）の比例的でない利用によって生じる個人の権利と自由へのリスク、個人の日常生活への不必要な侵入、警察の不当な介入等の潜在的な不利益に関する見解を表明した。また、コミッショナーはブログにおいて、そのような生体データの処理に対しデータ保護法令がいかに適用されるかについて述べた。
- 本意見書の目的は、法執行機関が公共空間で顔認識技術をデプロイする際の個人データ処理に関して、法執行機関をガイドすること。
- 本意見書の主たるメッセージは以下。
  - LFRの利用は、個人データの処理を伴うため、データ保護法（DPA）が適用される。これは、実証実験であろうと、日常的な運用であろうと同様である。
  - 「所管官庁」による「法執行目的」での個人データ処理は、DPA第3部によってカバーされる。
  - とりわけ、法執行目的でのLFRの利用は、個人をユニークに識別する目的での生体データの処理を伴うため、「センシティブな処理」（DPA第35条（8）b）を構成する。
  - そのようなセンシティブな処理は、LFRソフトウェアによって取得され分析された全ての顔画像に関係するものであり、DPA第35条、42条、64条の要件にとりわけ注意を払わなければならない。そのため、「データ保護影響評価（DPIA）」（第64条）や「適切なポリシー文書」（第42条）が実施されなければならない。

【ご参考】英国ICOのライブ顔照合(LFR)に対する意見書

○サマリー(続き)

- センシティブな処理は、当該画像がウォッチリスト上の人物とマッチングしたか、それともマッチングしなかった人物の生体データが短時間で削除されたかに関わらず、発生している。
- データ保護法は、デプロイメントの必要性や比例性の検討から、ウォッチリストの編集、生体データの処理、生体データの保持や削除まで、LFRのプロセス全体に適用される。
- 管理者は、LFRの利用の適法性の基盤(DPA第35条)を識別しなければならない。適法性の基盤は、行動規範のような他の利用可能な法的文書とともに、識別され、適切に適用されるべき。
- コミッショナーは、政府によって発行される、法令に基づく拘束的な行動規範(statutory and binding code of practice)によって、法的フレームワークを強化するという見解のもと、関連する機関と協働することを目指している。コミッショナーの見解では、そのような行動規範は、監視カメラ行動規範(2012年自由保護法の下で発行されたもの)で設定された基準の上に構築され、データ保護法制と整合的なものとなるだろうが、LFRや他のバイオメトリック技術の法執行目的での利用に明確で特別な焦点を当てたものになるだろう。それは、現行や将来的なバイオメトリック技術に適用可能なものであることを保証するように開発されるべきである。
- コミッショナーは、警察やその他の法執行機関が、南ウェールズ警察に対する高等裁判所判決で規定された義務を遵守するために、DPA第42条を遵守するために何が必要かについてどう裁判所が提供した勧告を考慮しながら、何が必要とされるかの詳細なガイダンスを提供することを目指している。

【ご参考】 英国ICOのライブ顔照合(LFR)に対する意見書

○「厳密な必要性」の基準

- データ保護法第35条(5)(a)「当該処理が法執行目的で厳密で必要とされる場合」の「厳密に必要とされる」の基準は何か。(→前述p.9参照)
- 管理者は各々のデータ処理とそのメリットについて注意深く検討し文書化する必要がある。コミッショナーは、管理者がDPIAや適切なポリシー文書を含め、なぜ法執行目的でのLFRを通じたセンシティブな処理が「厳密な必要性」の基準を満たしているかを明確に説明することを期待する。この基準を満たすためには、管理者はセンシティブな処理の「比例性」と、LFRの代替手段とを検討しなければならない。
- LFRがデプロイされる目的は、重要性の高いものであるべきである。一般的に、LFRを特定の重大犯罪や暴力犯罪を軽減する目的で利用することと、既知の万引き犯を識別する目的でLFRを利用することの間には、かなりの違いがある。軽罪の中にはより重大な犯罪や組織犯罪の一部であるものが含まれうることは認めるが、各ケースでそのメリットについて検討されなければならない。
- LFRが狭く定義された目的のために、ターゲット化されて、または小規模にデプロイされる場合には、厳密な必要性や比例性の要件を満たす可能性が高い。一例として、容疑者が特定の場所に特定の時間にいる可能性が高いことを示すインテリジェンス(情報)を警察が持っている場合である。他の例として、LFRが、空港などで、所轄官庁によって法執行目的で実施されるセキュリティ措置の一環である場合である。
- 換言すると、LFRのデプロイメントが以下である場合は、センシティブな処理であることを正当化することのハードルはより低いだろう。
  - (対象者が)ターゲット化されている
  - インテリジェンスに基づいている
  - 時間が限定されている

【ご参考】 英国ICOのライブ顔照合(LFR)に対する意見書

〇「厳密な必要性」の基準(続き)

- また、他のより侵害的でない選択肢の利用可能性がある場合、管理者はなぜLFRという侵害的な手段の利用が厳密に必要であるかを明確に説明できなければならない。
- ICOは、南ウェールズ警察による「厳密な必要性」と「比例性」の正当化が以下の点で十分でないと考えている。
  - なぜ目的を達成するためにより侵害的でない手段が考慮されていないかが十分に説明されていない。
  - LFRの利用のターゲット化が十分に保証されていない。
  - LFRを実施する場所の選択が特定の要因や合理的な疑いによって正当化されていることが十分に保証されていない。
- そのため、ICOは、センシティブな処理の厳密な必要性と、個人の権利の間の公正なバランスを両立させることを、SWPは保証していないとの見解である。

〇行動規範の導入

- コミッショナーは、LFRのようなバイオメトリック技術の利用によって生じる特定の問題に対処するためのさらなる保護措置を提供する、法令に基づく拘束的な行動規範(statutory and binding code of practice)を早期に導入することを政府に要求する。これは、データ保護法令を遵守しながら、どのように、いつ公共空間においてLFRを利用してよいかについて、法執行機関にさらなる情報提供を行うものである。これは、LFRの利用が比例的であり、必要であり、ターゲット化されていることを保証し、データ保護・プライバシー・人権に関する法令への遵守を保証するような監督を可能とする。

【ご参考】 顔認識技術に対する意識調査結果

- ICO：公共空間における警察のLFR利用に関する意識調査（2019年1月）
  - オンライン調査、回答者2,202人（18歳以上）
  - 82%の回答者が、警察のLFR利用に受容的。
  - 72%の回答者が、犯罪頻度の高いエリアでLFRを恒久的に利用することに同意。
  - 65%の回答者が、LFRは低いレベルの犯罪を防止するために必要なセキュリティ措置として同意。
  - 60%の回答者が、1人の興味ある人物を見出すためであっても群衆全員の顔を処理することについて受容的。
- ロンドン警察倫理パネル：ロンドン市民に対する意識調査（2019年5月）
  - 57%の回答者が、ロンドン警視庁によるLFR利用を受容。
  - しかし、アジア人の56%、黒人の63%は反対。
  - 16~24歳の52%、25~39歳の52%が反対。
- Ada Lovelace Institute：顔認識技術に関する意識調査（2019年9月）
  - 70%の回答者が、警察によるFR利用を受容。
  - 55%の回答者は、警察のLFR利用に対して政府が制限を設けるべき。
  - 29%の回答者は、以下の理由で警察のLFR利用に否定的。
    - プライバシーの侵害
    - オプトアウトの欠如、同意が不可能
    - ・サーベイランスの常態化
    - ・警察がLFRを倫理的に利用することへの信頼の欠如
  - 調査報告書の主たるメッセージ
    - 1. 市民のFRTに対する意識は高いが、知識は少ない。
    - 2. 同意は重要な保護措置である。
    - 3. 人々は監視が常態化することを恐れているが、公共の利益を証明できるならば、多数はFRTを支持。
    - 4. 警察がFRTをデプロイすることに対して、無条件の支持は存在しない。FRTへの支持は、制約下、適切な保護措置の条件下での支持である。
    - 5. 市民は、民間企業がFRTを倫理的に使用できるとは信頼していない。
    - 6. 企業や政府は今行動を起こす責任がある。

英国ICOへのヒアリング調査（2019年11月）

- Q1.1：顔認識技術にも様々な利用方法があるが、英国ではどのような分野でのどのような利用方法が問題視されているのか？
- Q1.2：それらの利用方法において、どのような点が懸念されているのか？

- A:ちょうどコミッショナーが、法執行機関（警察など）による顔認識（FR）利用について公式な意見書を出したところである。ICOでは民間のFR利用についても調査しているところである。

- メインな点は、FR利用には法的根拠が必要だということ。データ保護法（DPA）のみならず、様々な法律がある。警察利用の場合には、DPA以外にもコモンローや、その他の英国法が法的根拠となる。
- 2つ目に、最も深刻な点として、FR利用を国民にどうやって知らせるかということ。警察がどのように何をしているかを、効果的に知らせる方法が難しい。
- 3つ目の点は、技術のケーパビリティ。AIのトレーニングにおいて、人種・性別などの学習データが偏ると、結果に意図しないバイアスが発生する恐れがある。

- Q：バイオメトリクスなど他の個人データ取得技術に比べて、FRが特に侵害的な点は何か？
- A：DPAでは生体データに顔特徴データ、指紋データ、歩容データ、視線データなどが入っている。顔特徴データの場合は、外を歩いているだけで取られるということでリスクが大きい。本人同意を得ることが難しい。本人がデータを取られることに対して選択権がない。

英国ICOへのヒアリング調査（2019年11月）

- Q1.3：どのような条件を満たせば、警察による自動顔照合（AFR、LFR）の実施が認められるのか？

- A:南ウェールズ警察（SWP）に対する訴訟が、まさにその答えである。SWP訴訟の高裁判決では、SWPのリアルタイム顔照合（ライブFR、LFR）は法的根拠があるとされた。すなわち、SWPはコモンローの下で、LFR利用ができるとされた。政府の行動規範（監視カメラ行動規範）にも則っているとされた。しかし、高裁判決は上訴される見通しなので、最終判断はまだである。
  - （ICOは意見書において、高裁判決におけるSWPのLFR利用の正当化が不十分との見解を示している。）

- なお、SWP高裁判決は、SWPにおける特定ケースのLFR利用に限定された判決なので、全てのLFRに適用される訳ではない。すなわち、警察における全てのLFR利用が合法と判断された訳ではない。

- SWPに対する司法審査（judicial review）は高等裁判所によって拒否され、今後原告によって上訴審判所に申請され上訴される見込みである。（11月下旬に上訴された）

英国ICOへのヒアリング調査（2019年11月）

- Q1.4：空港やビル入館など、本人の明示的同意に基づく顔認証サービスについては問題はあるか？

- A:問題視はされているが、LFRとは状況が異なる。

- 空港などでのFR利用は、本人が同意したとなればリスクはより低いと考えられるが、データの保持期間や、テクノロジーのガバナンスが問題視されている。また、GDPR/DPAでは本人同意は却下ができるので、FR以外の代替方法も準備しないといけない。データの利用停止に対応しないといけない。

英国ICOへのヒアリング調査（2019年11月）

- Q1.5：警察における容疑者画像と犯罪者顔写真DBとの顔照合について問題はあるか？

- A:やはり問題はある。警察によるFR利用には2つある。
  - LFR（Live Facial Recognition）：　一般市民に向けてFRを行うもの。
  - Non Live：　過去のデータ（監視カメラに映った画像等）を加工しないで、DBと照合をかける。

- 質問はNon Liveの場合だが、DBのデータがどこから入手されたものか、DBのガバナンスは十分かが問題となる。つまり、DBの正確性が担保されないといけない。また、DBのデータとしてフェイスブックやインスタグラムの写真を使うのは避けるべき。

- Q:フェイスブックなどの写真を避けるべきなのはなぜか？

- A:データがフェイク写真の場合もあるし、本人でない（偽名の）可能性もある。データの正確性・信頼性が担保できない。また、ソーシャルメディアの利用者はまさか自分の写真が警察に利用されているとは思わない（個人の合理的期待がない）ため。

英国ICOへのヒアリング調査（2019年11月）

- Q2.1：民間における顔認証（Facial Authentication, Facial Verification）サービス（空港でのチェックイン・搭乗ゲート通過、ビル入退場、小売店での決済など）や顔認識（属性推定（Categorization））サービス（小売店での顧客の性別や年代などのカテゴライゼーションなど）の事例はどのようなものがあるか？

- A:民間の利用はまだ少なく、限定的である。空港、ショッピングセンター、ビル入館などである。属性推定して広告を出す用途では、まだ使われていない。リスクが高く、フレームワークがまだ決まっていない。

- Q:ショッピングセンターでの利用とはどのようなものか？
- A:ICOで調査しているところである。警察と同じようにLFR利用をしている。ICOは警察と民間企業（セキュリティ企業等）の関係についても調査している。警察が民間企業にデータをシェアしている。最近ニュースになっている。いちばん著名な例がキングスクロス事案。警察からの「行方不明者」の画像データをシェアしていた。
- Q:民間企業がLFRをやることと、警察からデータのシェアを受けていることのどちらに法的問題があるのか？
- A:どちらも問題と考えている。

英国ICOへのヒアリング調査（2019年11月）

- Q:小売店で酒類を買う際にFRで年齢確認をするトライアル事例を聞いたが、合法なのか？

- A:FRを使うという点で共通の課題となるのは、以下の3つである。
  - 法的根拠（適法性の根拠）があるか
  - さもなくば、本人から同意が取れているか
  - 技術に対するガバナンスが取れているか

- DPAでは、（本人同意がない場合は）FRを利用する管理者が、より侵害度の低い方法について検討することを義務化している。目的を勘案して、侵害度の低い他の方法がない場合のみFRの利用が許される。つまり、他の方法を考察して、FRしか目的を果たせないということを証明できないといけない。例えば警察は、FR利用の必要性、比例性を証明しないといけない。

- 18歳以上なのにFRシステムで未成年だと言われると問題（判断が不正確）であるし、逆に未成年が18歳以上と判断されても困る。

英国ICOへのヒアリング調査（2019年11月）

- Q:テロ対策などの場合は、LFRを使ってよいと言えるのか？

- A:民間企業が、LFRを利用する理由をきちんと説明することは難しい。本当に脅威がある場合でないと使えない。FR技術が使えるから使います、というだけでは駄目である。

- Q:2019年5月のGDPR適用後、監視カメラや顔認識技術に対する規制は以前より厳しくなったか？

- A:特に警察利用に対して厳しくなった。（GDPRと）DPAによって、個人データをどのように利用しているかを詳しく説明する必要が出てきた。以前は詳細に説明しなくてよかったが、新しい2018年DPAの下では、法的根拠を示す必要や、アカウンタビリティを示す必要が生じた。

英国ICOへのヒアリング調査（2019年11月）

- Q3.1：英国では今後、従来の法制度に加えて、顔認識技術に対するどのような規制や取組みがなされそうか？

- A:SWPの訴訟の結果、基本要件が判例として明らかになった。（上訴の結果）今後、もっと厳しい要件が課される可能性もある。10月のコミッショナーの意見書では、LFRに対する厳しい法的フレームワークやガイダンスが必要としている。

- Q:国民の意見はどのようなもので、コミッショナー見解は国民の意見を反映しているのか。国民が安全な社会が良いと言ったら、見解も変わるのか？
- A:コミッショナーの見解は国民の意見を直接反映したものではない。ICOで2019年1月に実施した世論調査では、警察がFRを使うのは良いが、民間が利用するのは好ましくないというものだった。

- Q:そうすると、警察にも厳しく義務を課すのは齟齬があるのでは？
- A:国民の意見は様々なものがあり、メディアの影響や個人の経験から意見も変わりやすいものでもある。現時点や来年に調査を行えば、結果はまた違ったものとなるだろう。

英国ICOへのヒアリング調査（2019年11月）

- Q:FRが犯罪を防いでくれるなら、プライバシーの権利を少し譲ってもよいという考え方もあるのでは。ICOとしてはプライバシーを守るべきであろうが、市民に対してどのような広報活動をしているのか？

- A:個人のプライバシーを守るのがICOの基本的スタンスである。個人の権利が国家安全や警察活動と干渉することもあることは認識している。LFRで実際に犯罪を防げた事例が少ない。Non LiveのFRで容疑者を逮捕したことはあるが。LFRにおいても、捜査活動に役立っているという根拠を示し、市民の理解を得ることは重要である。

英国ICOへのヒアリング調査（2019年11月）

- Q3.2：EU（EDPB）のビデオ機器ガイドラインは、BREXIT後も英国で有効なものなのか？

- A:EDPBのビデオ機器ガイドラインは民間企業のみを対象としている。警察関係の（ビデオ機器？）ガイドラインについてはEUで作っているところである。BREXIT後もICOとしてはEDPBのガイドラインを無視することはない。

- Q:しかし、GDPR自体は、BREXIT後は英国内での法的拘束力がなくなるではないか？
- A:ICOとしては、BREXIT後もEUと同じような（規制レベルで）考えたい。

- Q3.3：顔認識技術を規制する新たなEU規則の見通しはどのようなものか？（ファイナンシャルタイムズに掲載されたもの）

- A:詳細は知らないが、EUにそのような意図があるとは聞いている。欧州委員会が何を考えているかは分からない。将来的には、英国でも新しい法律ができるかもしれない。法律が不可ならば、ICOとしては法的拘束力のある行動規範を作ることをプッシュしていく。

英国ICOへのヒアリング調査（2019年11月）

- Q4.1：ギャンブル依存症対策としてカジノなどで顔認識技術が使われる場合、本人同意がなくても、家族の申請で依存症患者の顔写真を提出し、カジノの入口などで顔照合を行うことはGDPRの下で可能か？（日本の事例）

- A:他に、よりプライバシー侵害度が低い方法がない限り、使えない。第三者の同意を本人の同意をみなすことは難しい。その第三者に（当人の代理人としての）権限があることを証明できないといけない。この事例について、適法性の根拠を見つけることは難しい。またこの場合、本人には（顔照合をして立ち入り禁止にすることを）秘密にしなければならないのだろう。

- セキュリティガードがその人の顔を知っていて、入店を拒否するという方法で同じ目的を達成できるのではないか。

- この場合、本人の生命の保護（GDPR第9条2項c）の根拠を使うこともできない。本人の利益の保護よりも、本人に対する人権侵害が大きいケースとなるだろう。

英国ICOへのヒアリング調査（2019年11月）

- Q4.2：米国では、政府機関による顔認証技術の活用について、人権擁護団体がベンダー各社に対して、販売を停止することを要求している。どのように考えているか？

- A:興味深い内容である。その背景として、米国人が政府機関や警察を信用していないことがあるのではないか。それとも、FRシステムの性能がまだあまり良くないのか。現状、FRのポジティブな面よりも、問題やリスクの方が大きいと捉えられているのではないか。サンフランシスコ市でも市機関によるFR利用が禁止された。

- 英国では、FRはベンダーが開発中であり、まだ発展途上のものである。英国の人権団体は、強いスタンスでFR技術自体を禁止したいと言っている。SWPに対する訴訟をサポートしたLibertyなどである。LFRのみならず、FR全体を禁止したいと言っている。

- Q:空港など本人同意の上で用いるFRについては侵害度が低いのではないか。なぜ人権団体はFR全体を禁止したいと言っているのか？
- A:Libertyなどは基本的にはLFRをいちばん問題している。しかし、空港などでのFRの民間利用に対しても、（上記のような理由で）懸念を示している。LFRのリスクは、公共空間で大勢の市民のデータを取得できてしまうことである。

英国ICOへのヒアリング調査（2019年11月）

- Q:新しい技術にはベネフィットとリスクの両面があり、自動車やダイナマイトなども全面禁止するということはなく、有用性を生かすために法律で規制している。FRについてのみは全面禁止というのはおかしいのではないか？

- A:FRの場合はリスクが分かりやすい。誤用の可能性が国民にも見えてしまう。技術は進化が速いため、（全面禁止までは行かなくても）今からリスクを直視し早い手立てを取らないと、数年後にはコントロールできなくなり、取り返しがつかなくなる恐れがある。
- 2018年DPAでは、データ保護バイデザイン（DPbD）が追加され、設計の初期段階からデータ保護に配慮しなければならない。ベンダーは技術を用いた製品を作るが、自社の製品がどのように使われているかについて考えが及ばないのは問題だ。

- Q:製品に対するDPIA（データ保護影響評価）は、ベンダーで独自に（自分で）行うべきか？

- A:ベンダー側は技術製品を提供するだけであり、DPIAについてはベンダー側で行う必要ないかもしれない。ベンダーは自社製品の使われ方に対してはアカウンタビリティがなく、ユーザ企業側にアカウンタビリティがある。例えば、銃のメーカーには銃の使われ方に対するアカウンタビリティはなく、使う側の問題である。

【ご参考】EU:ビデオ機器を通じた個人データ処理に関するガイドライン案

- EUの個人データ保護に関する諮問委員会であるEDPB(欧州データ保護会議)は2019年7月10日に、「ビデオ機器を通じた個人データ処理に関するガイドライン(Guidelines 3/2019 on processing of personal data through video devices)」案を公表した。9月9日までパブリックコメントに付された。
- これはGDPR(EU一般データ保護規則)の下でのカメラ画像や顔認識技術の取扱いに関する指針案であり、事業者の立場から見ると非常に厳しい内容の規制も含まれている。
- EDPBが発行する指針はGDPRの法解釈を示すもので、EU各国の監督機関がGDPRの執行を行う際の根拠となる。
- 同ガイドラインの構成
  - 1. はじめに
  - 2. 適用範囲
  - 3. 処理の適法性
  - 4. 第三者へのビデオ映像の提供
  - 5. 特別な種類のデータの処理
    - 5.1 生体データを処理する際の一般的留意事項
    - 5.2 生体データを処理する際にリスクを最小化するための推奨措置
  - 6. データ主体の諸権利
  - 7. 透明性と情報提供の義務
  - 8. 保存期間と消去の義務
  - 9. 技術的措置と組織的措置
  - 10. データ保護影響評価
- 「欧米におけるカメラ・顔認識サービスと規制動向」(https://www.i-ise.com/jp/information/report/2019/20191029_facial_recognition.pdf)もご参照ください。

【ご参考】EU:ビデオ機器を通じた個人データ処理に関するガイドライン案

- 欧州での事業活動に影響を与える恐れのある規定

顔認識サービス例	GDPRガイドライン案における要件	理由
空港での顔パス認証	顔認識システムを専用ゲート内に設置し、顔認識に同意していない旅客の顔特徴データを取得しないようにしなければならない	顔特徴データはセンシティブデータであるため、取得に当たって本人の明示的同意が必要 （写り込みでの取得は不可）
コンサート会場での顔パス入場	顔認識システムの付いた入口と、そうでない入口（チケットをスキャンする等）の両方を明確に区別して設置しなければならない	
店舗でのリピーター分析	全ての来店客から事前同意を得なければならない	
ホテルでのVIP顔認識	登録済みのVIPか否かを判断するために入口で撮影を行う際、全ての入館者から顔認識に関する事前同意を得なければならない	
顔認証によるビル入退館管理	全ての入館者に顔認証を強いるのではなく、それ以外の入場方法（社員証の提示等）も提供しなければならない	強制的な同意は、有効な同意とみなされない

【ご参考】EU:ビデオ機器を通じた個人データ処理に関するガイドライン案

- ガイドライン案に対し、特に全ての顧客からの本人同意を必須とする法解釈に対しては、日本の電子情報技術産業協会（JEITA）からもパブリックコメントをEU側に提出している（https://home.jeita.or.jp/press_file/20190909170648_4dJDMVL5fG.pdf）。意見の骨子は以下である。
- 同ガイドライン案では、顔特徴データの取得・利用について同意していない利用者からの顔特徴データの取得は、GDPR第9条の特別な種類の個人データの処理に当たるとみなしているため、処理の適法性の根拠として企業側の「正当な利益」を用いることができない。そのため、本人同意を得ない限りはそのような利用者からの顔特徴データの（照合目的のみでの）一時的な取得も違法とみなしている。
- しかし、（同案の74項で規定されているように、）生体データがGDPR第9条の特別な種類の個人データに該当する条件の1つは、「自然人を一意に識別することを目的」として処理されていることである。しかるに、84項や82項の事例において明示的な同意を得ていない利用者から顔特徴データを取得することは、当該データが顔認識システムに登録されていないことを確認することが目的であり、個人を一意に識別することが目的ではない。したがって、このようなデータは第9条が適用される特別な種類の個人データではなく、一般的な個人データとみなすべきであり、その処理の適法性の根拠としては（GDPR第6条1項(f)の）「正当な利益」を許容するべきである。すなわち、本人同意を得ていない利用者からの顔特徴データの（照合目的での）取得も、「正当な利益」の根拠に基づき許容するべきである。

【ご参考】EUデータ保護指令とGDPRの規定の違い

- 顔特徴データを含む生体データは、EUデータ保護指令（1995年）では通常の個人データであるが、GDPR（2016年）では特別な種類の個人データに格上げされている。

	EUデータ保護指令	GDPR（EU一般データ保護規則）
生体データ（顔特徴データを含む）の扱い	通常の個人データ	特別な種類の個人データ（GDPR第9条） （「自然人を一意に識別することを目的とする生体データ」が、これに含まれる）
生体データの処理の適法性の基準	（EU指令第7条） ・データ主体の同意 ・契約の履行 ・法的義務の遵守 ・データ主体の生命に関する利益の保護 ・公共の利益/公的権限の行使における職務遂行 ・管理者等の正当な利益	（GDPR第9条2項） ・データ主体の明示的な同意 ・雇用及び社会保障並びに社会的保護の法律の分野における管理者やデータ主体の義務の履行や権利の行使 ・データ主体等の生命に関する利益の保護 ・政治、思想、宗教、労働組合の目的による団体の正当な活動 ・データ主体によって明白に公開された個人データ ・訴えの提起もしくは攻撃防御、裁判所の権能行使 ・重要な公共の利益 ・予防医学もしくは産業医学の目的 ・公衆衛生の分野における公共の利益を理由とする処理 ・公共の利益における保管の目的、科学的・歴史的研究の目的、統計の目的
生体データ処理に関する追加的規定	特になし	・加盟国は、生体データの処理に関し、その制限を含め、付加的な条件を維持または導入できる（GDPR第9条4項）
備考	―	自然人を一意に識別することを目的とする顔特徴データ（facial template）のみが上記の「特別な種類の個人データ」（GDPR第9条）に相当する。単なる顔画像や、属性推定用の加工データは、これに相当しないと考えられる。

【ご参考】欧州データ保護監察官（EDPS）の顔認識に関する見解

- EU機関に対するデータ保護監督機関である欧州データ保護監察官（EDPS）は、2019年10月28日の「Facial recognition: A solution in search of a problem?」という文書で、顔認識に関わるプライバシーおよびデータ保護の問題として、以下の5つを挙げている。
- 第一に、EUのGDPRは明確に、顔画像を含む生体データの処理をカバーしている。GDPRは基本的に、第9条2項に挙げられた10個の例外の1つを満たさない限り、個人を一意に識別する目的での生体データの処理を禁止している。
- 第二に、EU基本権憲章第52条にいうように、基本的人権に対するいかなる制限も、それが必要であることを証明しなければならない。人権に対する制限が大きくなればなるほど、必要であることを証明することのハードルは高くなる。顔認識技術が必要だとする証拠があるのか？同じ目的を達成するために、より侵害的でない他の手段が実際にあるのではないか？明らかに、「効率性」や「便利さ」は十分な理由とはなりえないだろう。
- 第三に、センシティブデータの大規模な処理を伴うような、顔認識技術の導入に、有効な法的根拠（legal basis）があるのか？同意は、明示的で、かつ自由に与えられ、事前に説明を受けた上での、特定のものでなければならない。人々が顔認識サーベイランスでカバーされた公共空間にアクセスする必要がある場合、オプトアウトすることも、ましてオプトインすることも難しい。GDPR第9条2項(g)の下で、加盟国やEUの立法者は、顔認識技術の利用が人権に対する比例的で必要な制限を保証するようなケースを決めることができる。
- 第四に、アカウンタビリティと透明性である。顔認識技術のデプロイメントは従来、不透明であることを特徴としてきた。我々は、誰によってデータが収集され、どのように利用されるか、誰がデータにアクセスし、誰に提供されるか、どれくらい保存されるか、どのようにプロファイルがつくられるか、自動意思決定に誰が責任を持つかについて基本的に知らない。さらに、インプットデータのデータ源をトレースすることはほぼ不可能である。顔認識システムは、我々の許可なく、インターネットやソーシャルメディアから得た膨大な量の画像で学習している。その結果、誰もがアルゴリズムの冷たい判断の被害者となりえ、カテゴライズされたり、場合によっては差別されうるだろう。
- 第五に、顔認識技術がデータ最小化などの原則やデータ保護バイデザインの義務を遵守していることは、極めて疑わしい。顔認識技術は決して完全に正確なものではなく、このことは犯罪者等と誤認識された個人に重大な影響をもたらす。「正確性」の目標は、アルゴリズムを完全なものとするために（センシティブな）データのエンドレスな収集に結びつく恐れがある。実際、バイアスや、偽陽性・偽陰性をなくすためにはデータがいくらあっても足りないだろう。

1. 英国（イギリス）の動向

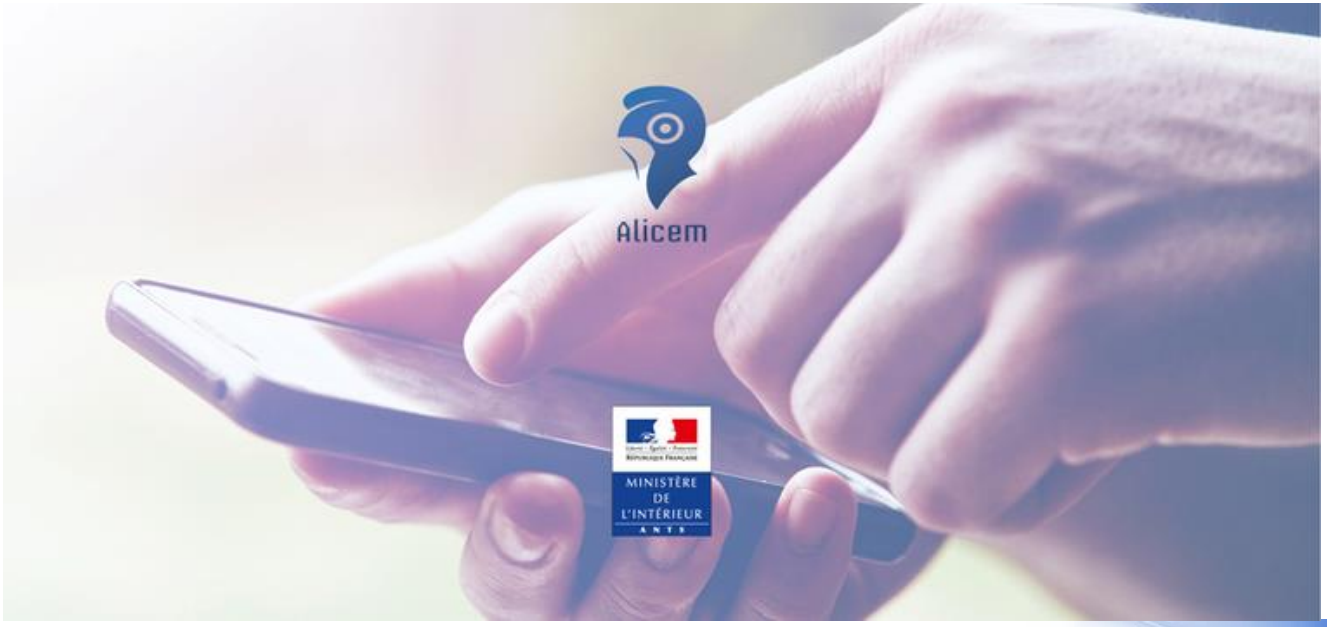
2. フランスの動向

フランスにおける顔認識技術の利用事例

- ①本人同意の下での個人認証（顔認証）
  - シャルルドゴール空港、オルリー空港、パリ北駅（ユーロスター搭乗口）の出国審査時に、EU市民は顔認証ゲートを利用可能。
  - エールフランスとパリ空港（ADP）は、2020年早期にオルリー空港で顔認証を利用して搭乗ゲートや手荷物預かりカウンターでの手続きを簡略化する実証実験を開始する計画。
  - フランス政府（内務省）は、国民向けに顔認証技術を用いたオンライン上の公的個人認証サービスの提供を2019年内に開始する計画（Alicem）。
- ③公共空間等での不特定多数に対する自動顔照合（本人同意なし）
  - ニースのカーニバルでの実証実験：　ニース市は2019年2月のカーニバルにおいて、ボランティア1000人を募って、自動顔照合（AFR）の実証実験を行った。6台の可動式監視カメラを通じて、群衆のリアルタイムの顔画像と、一定の対象者（迷子、脆弱なお年寄り、指名手配犯）の顔写真データベースが照合された。
  - その他、CNILは2010年に、スタジアムにおけるリアルタイムでの顔照合を目的とする、VESALIS社による個人データ処理の実験的な実施（スタジアムでフーリガンの暴動を防ぐための実証実験）を許可している。

①本人同意の下での顔認証の事例： Alicem

- フランス政府（内務省）は、国民向けに顔認証技術を用いたオンライン上のデジタルIDの提供を2019年内に開始する計画。
- 「Alicem」という内務省開発のスマホアプリを用いる。500以上のオンライン行政サービスへのアクセスが可能。
- 希望者は、バイオメトリックパスポート（または滞在許可証）のICチップ内の顔写真と、本人が動画撮影した顔画像とを用いて登録し、Alicemアカウントを作る。
- 登録後は、スマホアプリ上での顔認証により、個人認証を行う。
- Alicemはアンドロイド専用アプリで、非接触型リーダーを備えたスマートフォンが必要。



フランスEFUSへのヒアリング調査（2019年11月）

- Q1.1：フランスでは監視カメラは何台くらい設置されているのか。公道では、どのような場所に設置されているのか？

- A:CNILによれば、フランス国内には、公道、店舗、公共交通機関、オフィス、マンションなどに935,000台のCCTV（監視カメラ）が設置されている。近年、公共空間を撮影するCCTVの数は急速に増えており、とりわけ治安やテロに対処するために公共機関によって設置されたものが増えている。
- 公共空間や、一般市民に開かれた場所（駅や商業施設、市役所等）については、一定の理由とともに、当局によるCCTV設置の許可が必要。それ以外の私的空間では、CCTV設置に許可は必要ない。一般市民は、CCTVの存在について情報提供されなければならない。画像の保存期間は1ヶ月を超えてはならない。
- 公共空間や、一般市民に開かれた場所へのCCTV設置には、次のいずれかの理由が必要。
  - 公共の建物および施設とその周辺の保護
  - 攻撃や盗難のリスクに特にさらされている場所にある店舗のすぐ周囲の保護
  - 国防施設の施設　・交通規制　・交通違反の認識
  - 攻撃、盗難、麻薬密売のリスクに特にさらされている場所での人と財産のセキュリティに対する攻撃の防止
  - テロ行為の防止　・自然または技術的リスクの防止
  - 人々の救助と火災に対する防御　・遊園地の公共施設の安全性

※EFUS（European Forum for Urban Security）は欧州の250自治体から予算を得て、自治体におけるセキュリティの在り方について調査・助言を行う団体。

フランスEFUSへのヒアリング調査（2019年11月）

- Q1.2：英国のように、公道やスタジアムなどで、自動顔照合（AFR）技術の実証実験は行われているか？

- A:フランスにおける最初のAFRのパイロット実験はニースにおけるもの。ニース市は2019年2月、カーニバルの期間にAFRをテストした。6台のCCTVが、一定の対象者（迷子、脆弱な高齢者、指名手配犯）の顔写真に基づき群衆の中の個人を識別するために、可動な形で置かれた。実験では「モルモット」（サクラ）を演じるボランティアが雇われ、規制エリア内を歩かされたりした。

- A:その他、顔認識ではないが異常行動検知等の実証実験が何件か行われている。
  - パリ交通公団（RATP）：地下鉄構内での異常行動検知実証実験
  - イブリン県（パリ郊外）：デジタル地域整備の一環として、学校や消防署での異常行動検知
  - ニース市：トラムの中で、CCTVを用いて乗客の心理状態や情動を検知してパニックの予防などに役立てる

フランスEFUSへのヒアリング調査（2019年11月）

- Q1.3：フランスの警察は自動顔照合技術の導入を考えているのか？

- A:フランスでは現時点では、オルリー空港、シャルルドゴール空港、北駅（ユーロスター搭乗口）のみでFR（本人同意の下での顔認証）が許可されている。これらの例を除いて、フランスでのFR実証実験はまだ実用化に至っていない。にもかかわらず、多くの法執行機関や地方自治体はFRに関心を持っている。

- Q1.4：その場合、どのような反対意見が出ているか？

- A:市民団体の議論は基本的にはGDPRに基づいている。すなわち、事前の影響評価が足りない。FRのための法的フレームワークがない。追求する目的に対して生体データを処理することは比例的でない。FRは我々の自由にとってあまりに脅威である。

フランスEFUSへのヒアリング調査（2019年11月）

- Q2.1：フランスにおける民間での顔認証（Facial Authentication, Facial Verification）サービス（空港でのチェックイン・搭乗ゲート通過、ビル入退場、小売店での決済など）や顔認識（属性推定（Categorisation））サービス（小売店での顧客の性別や年代などのカテゴライゼーションなど）の事例はどのようなものがあるか？

- A:産業や開発者についてあまり情報を持っていないが、例えば、
  - 「Any Vision」（おそらくイスラエルの企業）は、「写真からその人が30歳だと認識する？（30年前の写真でもその人を認識する？）」ソフトウェアを開発した。
  - EvitechやFoxstreamのようなスタートアップ企業は、カメラを通じて、疑わしい叫び声や、ガラスの音、群衆などを検知するソフトウェアを開発した。
  - シスコ・フランスは、マルセイユとニースの2つの高校でFRゲートを導入する実証実験のシステムを提供する予定であったが、CNILによって却下された。

- Q:CNILはどのような理由で、マルセイユやニースの高校での顔認証ゲートの実証実験を却下したのか？
- A:学校が求めている目的（不審者等の侵入を防ぐ）と、顔認証という手段が不均衡であり、代替手段があるため、不適切と判断された。

フランスEFUSへのヒアリング調査（2019年11月）

- Q2.2：フランス以外の欧州における顔認識サービスの利用事例は？

- A:産業界や企業はこの領域におけるイノベーションを強力に推奨している。いくつかのEU加盟国における行政機関は、既に、例えば学校やイベント会場を保護するために顔認識技術（FRT）の実証実験を行っている（ドイツ、英国、スウェーデン）。また、テロリスト対策の文脈において、例えば群衆の中の疑わしい個人を識別する目的で、少なくとも警察利用においては、FRTのデプロイメントを進めたがっている。

- Q3.1：2018年5月のGDPR適用後、監視カメラや顔認識技術に対する規制は以前より厳しくなったか？

- A:はい。

- Q3.2：EU（EDPB）のビデオ機器ガイドライン案についてはどう考えるか？

- A:GDPRは個人データの保護にフォーカスした規則であるが、FRのケースにおいては、このような技術が発生しうるエラーを考慮できていない。技術進歩とともに、より深い法的フレームワークが開発されなければならない。

フランスEFUSへのヒアリング調査（2019年11月）

- Q4.1：フランスでは今後、従来の法制度に加えて、顔認識技術に対するどのような規制や取組みがなされそうか？

- A:FRTは新たなホットトピックであり、EFUSはFRTに対してバランスの取れたポジションを策定しようとしている。なぜなら、EU機関によって発表された最近の調査研究や意見によれば、法制度が急速に進化しうるからである。フランスや欧州では、市民を乱用から保護するためにFRの差別的な利用を制限する特別な規則（一定の例外と、データ活用のためのフレームワークも含む）を提案しようとする議論がある。

- Q5.1：フランスにおいて、公的機関が国民の顔情報を活用する事例があるか？

- A:フランス政府はAlicemというサービスの実証を行うと発表した。これはオンライン行政サービスに対して、スマートフォンを通じて顔認証でログインできるようにするもの。希望者向けのサービスである。

【ご参考】 CNILの顔認識に関するガイダンス

- フランスのデータ保護監督機関CNILは2019年11月15日に「Reconnaissance faciale : pour un débat à la hauteur des enjeux」という顔認識に関するレポートを公表した。
- フランスの公的機関による顔認識ソフトウェア利用に関するガイダンス。CNILは、FR利用には大きな政治的・社会的影響とリスクがあるとし、特に公的分野におけるFR利用のリスクを懸念。FRソフトウェアは100%の信頼性があるものでなくバイアスのかかった結果を生み出しうること、偽陽性（誤照合）の割合は個人の性別や民族により変化しうることを指摘。
- FRTの違法な利用の機会を最小化するために、CNILは以下の3つの要件を挙げている。実験フェーズにおいてFRを利用する公的機関に対し、これらを遵守するように勧告している。
  - （1）顔認識は、高いレベルの信頼性を持った認証メカニズムを実施することの確実なニーズがある場合、また、より侵害度の低い手段がない場合にのみ、実験的利用が行われるべきである。CNILは適法とみなす顔認識利用について例示している。これには、オンライン公共サービスへのアクセスのための顔認識の利用や、空港等での顔認証システムが含まれる。他方、CNILは学校でのセキュリティ・入退管理目的での顔認識利用に対しては反対している。
  - （2）管理者はあらゆる環境下で、記録される個人の権利を尊重しなければならない。FRで利用される各々の機器で同意を得なければならない、個人は自分のデータに対するコントロールを与えられなければならない、また情報通知の義務、利用と目的の透明性等である。
  - （3）実験的利用は、正確なスケジュールに従わなければならず、リスクを最小化するための厳格な手法に基づかなければならない。
- CNILは将来的には顔認識の利用にレッドラインを引くべき（越えてはいけない一線を決めるべき）としている。

【ご参考】 英国におけるAI倫理

英国のAI倫理・プライバシー関連の報告書類

〇英国のAI倫理・プライバシー関連の報告書類

- ICO（情報コミッショナーオフィス）：
  - 「Big data, artificial intelligence, machine learning and data protection version 2.2」（2017年9月）
  - 現在、「Explaining AI decisions guidance」「AI Auditing Framework」などのプロジェクトを実施中。
- 英国庶民院科学技術委員会：
  - 「意思決定におけるアルゴリズム（Algorithms in decision-making）」（2018年5月）
- 英国貴族院AI特別委員会：
  - 「英国におけるAI：英国はAIを活用し、そして活用できる準備ができているか（AI in the UK: ready, willing and able?）」（2018年4月16日）
- データ倫理イノベーションセンター（CDEI）：
  - 「オンラインターゲティング」中間報告書（2019年7月）
  - 「アルゴリズムを用いた意思決定におけるバイアス」中間報告書（2019年7月）

英国ICOのAI説明ガイダンス

○「Explaining decisions made with AI: Draft guidance for consultation」

- 「AIを用いた意思決定を説明する」というガイダンス案が、2019年12月2日にICO（Information Commissioner's Office：情報コミッショナーオフィス）とAlan Turing Instituteから公表され、2020年1月24日までパブコメに付されている。
- このガイダンス案は、事業者等がAIを用いた意思決定について、その影響を受ける個人に対して説明するための実践的アドバイスを意図したもの。
- 以下の3部構成。
  - 第一部： AIを説明することの基礎（The basics of explaining AI）（34ページ）
    - 主要な概念を定義し、様々な説明のタイプを概観。
    - ガイダンスの対象者は、AIシステムの開発に関係する全てのスタッフ。
  - 第二部： 実際にAIを説明する（Explaining AI in practice）（108ページ）
    - 事業者等が意思決定を個人に説明する際の実用上のサポートとなるガイダンス。
    - ガイダンスの対象者は、技術チーム、DPO、コンプライアンスチーム。
  - 第三部： AIを説明することが組織にとってどのような意味を持つか（What explaining AI means for your organisation）（23ページ）
    - 事業者が個人に「意味のある説明」を提供するために準備しうる様々な役割、ポリシー、手続き、文書について言及。
    - ガイダンスの対象者は、経営層、DPO、コンプライアンスチーム。

英国ICOのAI説明ガイダンス

〇主な内容

- AIを用いた意思決定に対しては、6つのタイプの説明がある。AIを用いる分野・ユースケースに応じて、これらの説明タイプの優先順位を変えるべき。

- AIには、より説明可能性（解釈可能性）が高いアルゴリズムと、より「ブラックボックス」的なアルゴリズムがある。AIを用いる分野・ユースケースやデータの種類に応じて、どのアルゴリズムを採用するかについて（設計段階で）検討するべき。

- 検討の結果、「ブラックボックス」的なアルゴリズムを選択した場合には、補足的な説明手法やツールを併せて用いるべき。

ICOガイダンス第一部：AIを説明することの基礎

〇AIとは何か

- AIとは、複雑なタスクを解決するために人間の思考を模倣しようとする広範な技術やアプローチの総称である。
  - ヘルスケア分野では、AIは病気の早期の兆候を見つけたり、病気を診断するために使われうる。
  - 警察分野では、AIは警察の介入をターゲット化したり、潜在的な犯罪者を識別するために使われうる。
  - マーケティング分野では、消費者に製品やサービスをターゲット化するために使われうる。

〇AIのアウトプット、AIを用いた意思決定（AI-assisted decision）

- 「AIのアウトプット」には3つの種類がある。
  - 予測（prediction）： ex. この人は債務不履行を起こさないだろう。
  - レコメンデーション： ex. この人はこのニュース記事を気に入るだろう。
  - 分類（classification）： ex. このeメールはスパムだ。
- 「AIを用いた意思決定」には以下の2つがある。
  - AIシステムのアウトプットや、その結果とられたアクション（意思決定）が人間の介在や監督なくして実施される場合には、AIシステムは完全に自動化されている。
  - 他方、AIのアウトプットを、人間が他の情報とともに吟味し、それらに基づいてアクション（意思決定）を行うプロセスもある。この場合、しばしば「ヒューマン・イン・ザ・ループ」と言われる。

ICOガイダンス第一部：AIを説明することの基礎

〇AIを用いた意思決定を説明することのリスク

- 利用者の不信感

  AIを用いた意思決定に関して情報を多く提供しすぎると、複雑であることから、利用者の不信感（distrust）が増大してしまうかもしれない。AIを用いた意思決定はしばしば複雑なものであるが、本ガイダンスで提示する様々な説明手法は、このような複雑性を理解可能なものとすることに役立つ。

- 商業的なセンシティビティ

  AIを用いた意思決定に関する説明が、当該AIシステムがどのように機能するかについて商業的にセンシティブな情報の開示につながるとの懸念もあるかもしれない。しかし、本ガイダンスで示すような説明によって、そのような開示の危険に晒されることはない。

- 第三者の個人データ

  AIモデルを学習させる方法や、個別の意思決定におけるインプットデータによって、他人の個人データが不適切に開示されるとの懸念があるかもしれない。本ガイダンスで提示する幾つかの説明タイプについては、このことは問題ではない。しかし、後述の「理由に関する説明」「公平性に関する説明」「データに関する説明」においては、当該個人と類似した他人がどのように扱われたかに関する情報や、個別の意思決定（複数個人が関連するもの）に対するインプットデータの詳細が開示される潜在的なリスクはある。事業者は、DPIAの一部としてこのようなリスクをアセスするべき。

- ゲーミング

  AIを用いた意思決定の背後にある理由付けを利用者が詳しく知り過ぎている場合、彼らがAIモデルでゲーミングしたり悪用したりするリスクから当該モデルを保護する必要があるかもしれない。AIを用いた意思決定の目的が不正行為や誤用の識別である場合（例えば詐欺検出など）、個人に提供する情報（とりわけ理由に関する説明）を制限する必要性は強いだろう。

ICOガイダンス第一部：AIを説明することの基礎

○6つの説明タイプ

- 理由に関する説明（Rationale explanation）：
  意思決定を導いた理由を、分かりやすく、非専門的な仕方で提供する。
- 責任に関する説明（Responsibility explanation）：
  誰がAIシステムの開発や管理、実装に関与したか。また意思決定に対する人間のレビューのために誰にコンタクトすればよいか。
- データに関する説明（Data explanation）：
  個別の意思決定において、どのデータがどのように用いられたか。またAIモデルの学習やテストにおいて、どのデータがどのように用いられたか。
- 公平性に関する説明（Fairness explanation）
  AIシステムがバイアスのかかっていない公平な意思決定を生み出すことを保証するために、設計や実装を通じて取られた措置。また、個人が平等に扱われてきたか否か。
- 安全性とパフォーマンスに関する説明（Safety and performance explanation）
  AIシステムの意思決定やふるまいの正確性、信頼性、セキュリティ、堅牢性を最大化するために、設計や実装を通じて取られた措置。
- 影響に関する説明（Impact explanation）
  AIシステムの利用や意思決定が個人や社会にもたらす影響。

ICOガイダンス第一部：AIを説明することの基礎

○理由に関する説明（Rationale explanation）

- AIの意思決定に対する「なぜ？」という質問に答えるもの。意思決定を導いた理由を、個人が分かりやすい仕方で理解することを助ける。
- この説明タイプは何の役に立つか。
  - 意思決定に異議申し立てすること

    個人がAIを用いた意思決定の背後にある理由を理解できることは極めて重要である。それによって、意思決定が本人の望むものや期待するものではなかった場合に、意思決定の理由付けに欠陥がありそうか否かを本人がアセスすることが可能となる。意思決定の理由付けを知ることにより、本人が異議申し立てをする場合に、その理路整然とした論拠を示すことが可能となる。
  - ふるまいを変更すること

    個人が（望ましくない）意思決定を正当なものと思う場合、将来望ましい判断結果を得るために自分のふるまいやライフスタイルをどのように変更したらよいかを考える際に、意思決定の理由に関する知識は役に立つ。

○責任に関する説明（Responsibility explanation）

- AIモデルの開発や管理に「誰が」関与しているか、また意思決定に対する人間のレビューのために「誰に」コンタクトすればよいかの理解を助ける。
- この説明タイプは何の役に立つか。
  - 意思決定に異議申し立てすること

    意思決定に異議申し立てをし、人間のレビューを求める際のコンタクト先が明確になる。
  - 情報提供（informative）

    AIシステムの設計やデプロイメントに関与する部門を明らかにすることで、情報提供の目的にも寄与しうる。

ICOガイダンス第一部：AIを説明することの基礎

○公平性に関する説明（Fairness explanation）

- AIの意思決定が基本的にバイアスがかかっておらず公平であることを保証するために取られた（そして取られ続けている）措置や、個人が平等に扱われてきたか否かについての理解を提供するものである。
- この説明タイプは何の役に立つか。
  - トラスト

    公平性に関する説明は、AIシステムに対する個人の信頼を高める上で鍵となる。AIを用いた意思決定においてどのようにバイアスや差別を回避しているかを個人に説明することによって、また個人が同じような他人に比べて差別的に扱われていないことを示すことによって、意味のあるトラストを促進することができる。
  - 意思決定に異議申し立てすること

    AIを用いた意思決定に対して、（それを不公平と感じる）個人が異議申し立てすることも可能とする。

○安全性とパフォーマンスに関する説明（Safety and performance explanation）

- AIシステムの意思決定の正確性、信頼性、セキュリティ、堅牢性を最大化するために設計や実装を通じて取られた措置。
- この説明タイプは何の役に立つか。
  - 安心感（reassurance）
  - 情報提供（informative）

    技術的な知識が豊富で熟達している個人が、AIモデルの適切性をアセスすることを可能とする。
  - 意思決定に異議申し立てすること

    AIの意思決定が不正確なものであるかもしれないという根拠で、あるいは安全でなかったり、危険であったり、信頼できない仕方で実施されたかもしれないという根拠で異議申し立てすることが可能となる。

ICOガイダンス第一部：AIを説明することの基礎

○影響に関する説明（Impact explanation）

- AIを用いた意思決定が個人にもたらしうる影響（すなわち意思決定が個人にとってどのような意味を持つか）について、事業者がどのように検討したかの理解を助ける。また、より広範な社会的影響についての理解も助ける。そのため、影響に関する説明は、AIを用いた意思決定が行われる前になされることが、しばしば適切である。
- この説明タイプは何の役に立つか。
  - 帰結（consequences）

    影響に関する説明の目的は第一に、個人にAIを用いた意思決定への一定の関与権限やコントロールを与えることである。意思決定の潜在的な帰結（ネガティブな、ニュートラルな、ポジティブな）を理解することによって、個人は当該プロセスに参加するべきかどうかを吟味することができ、意思決定が彼らにどのような影響をもたらすかについて予期することできる。
  - 安心感（reassurance）

    事業者が、AIシステムの社会的影響について検討することに時間をかけていることを知ることは、安全性、平等性、信頼性のようなイシューがAIモデルのコアな要素であると個人を安心させる助けになる。またAIを用いた意思決定のベネフィットとリスクに関して、個人がより多くの情報提供を受けることができ、それにより、AIシステムの開発や利用に関する議論において、よりAIを信頼しアクティブであることを促進する。

ICOガイダンス第二部：実際にAIを説明する

〇説明方法の優先順位付け

- 事業者は、AIシステムの利用分野、ユースケース、個人への影響を考慮することにより、説明方法に優先順位を付けるべき。
  - （0）様々なタイプ（6タイプ）の説明をよく理解する
  - （1）「理由に関する説明」と「責任に関する説明」を優先させる
  - （2）分野のコンテキストとユースケースを考慮する
  - （3）潜在的な影響を考慮する
  - （4）これらに基づき、その他のタイプの説明に優先順位をつける

- 説明方法の優先順位付けの例
  - AIを用いた雇用　→　次頁
  - AIを用いた医療診断　→　次々頁

ICOガイダンス第二部：実際にAIを説明する

○AIを用いた雇用

- ある企業でAIシステムを求職者の応募書類のフィルタリングツールとして用いている。同システムは、提出された履歴書から読み取れる個人の人間的属性や交際傾向に関連したデータを処理することによって、応募者を「不合格」または「面接試験に進める者」に分類する。懸念点は、データセットにバイアスが織り込まれている可能性（差別的な特徴やそのプロキシ情報がAIモデルの学習と処理に用いられている可能性）である。例えば、「男子中学卒」と「高収入の役職に就くこと」の間に強い相関関係が見られた場合、このデータセットで学習したAIモデルは、高収入の役職に対する応募者を判断するにあたって、女性応募者を差別的に扱うかもしれない。このようなケースで、事業者はどのタイプの説明を選択するべきか？

- （1）理由に関する説明と責任に関する説明を優先させる

  AIを用いた雇用意思決定によって影響を受ける個人に、当該意思決定の責任者は誰で、なぜそのような意思決定に到達したかを伝えるために、「責任に関する説明」と「理由に関する説明」を含める必要がある。

- （2）分野のコンテキストとユースケースを考慮する

  雇用や人事分野のコンテキストでは、バイアスが第一の懸念点となる。

- （3）潜在的な影響を考慮する

  意思決定が正当なものであると応募者が考えるかどうか、応募者が公平に扱われているかどうかの観点から、応募者に対するAIシステムの影響を考慮する。

- （4）その他のタイプの説明に優先順位をつける

  この事例では、応募者は差別されていないことを知りたいため、「公平性に関する説明」が求められている。このような差別は、従来からの差別や歴史的な不平等性に起因するものでありうる（これらはバイアスがかかった学習用データとしてAIシステムに影響しうる）。さらに応募者は、雇用主がAIツールの影響をどのように考えているかを理解するために、「影響に関する説明」を求めるかもしれない。「データに関する説明」もまた、面接試験に進める応募者を決定するためにどのデータが使われるかを理解するために、役立つかもしれない。

ICOガイダンス第二部：実際にAIを説明する

〇AIを用いた医療診断

- 放射線科医がAIシステムの画像認識アルゴリズムを用いて、患者のMRIスキャン画像における癌の検出を行っている。このAIシステムは患者のMRIスキャンからの何百万という画像を含むデータセットで学習を行っている。見知らぬデータパターンや予期せぬ環境異常（システムが認識できない対象物等）に直面した際に、このシステムが間違える可能性はある。このようなシステムの間違い（誤った判断）は、患者には致命的な身体的損害につながりかねない。このようなケースで、医師はどのタイプの説明を選択するべきか？

- （1）理由に関する説明と、責任に関する説明を優先させる

  AIを用いた診断によって影響を受ける個人に、当該意思決定の責任者は誰で、なぜそのような意思決定に到達したかを伝えるために、「責任に関する説明」と「理由に関する説明」を含める必要がある。

- （2）分野のコンテキストとユースケースを考慮する

  医療分野のコンテキストでは、AIシステムの安全性と最適なパフォーマンスが第一の懸念点となる。

- （3）潜在的な影響を考慮する

  AIシステムが誤った診断を下した場合、患者への影響は大きい。AIシステムの利用に伴うリスク、また事業者がそれらのリスクをどのように軽減したかを患者が十分に理解できるように、事業者の説明は包括的なものであるべきである。

- （4）その他のタイプの説明に優先順位をつける

  「安全性とパフォーマンスに関する説明」は、AIシステムが十分に堅牢で、正確で、セキュアで、信頼できること、またテストや検証の手続きがこれらを保証しうることの正当化を与える。

ICOガイダンス第二部：実際にAIを説明する

○「理由に関する説明」の構築方法

- 「理由に関する説明」は、AIシステムの理解にとって要となり、事業者がGDPRを遵守することの助けにもなる。これは「機械の中」を見ることを必要とし、「安全性とパフォーマンスに関する説明」、「公平性に関する説明」といった他の説明に必要となる情報を集めることの助けにもなる。
- 事業者の説明ニーズにとって適切なAIモデル（説明可能なモデル or ブラックボックス的なモデル）を選択するために、事業者は以下について考慮するべき。

  ①AIシステムを用いる分野

  ②AIシステムの個人・社会への潜在的影響

  ③現行システムを、（より説明可能でない）AIシステムに置き換えることのコストとベネフィット

  ④利用するデータが、説明可能なシステムを必要とするかどうか

  ⑤補助的な説明ツールがコンテキストにおいて適切かどうか

ICOガイダンス第二部：実際にAIを説明する

◇ 主要なアルゴリズム技術の一覧

- 解釈可能（interpretable）とみなされているアルゴリズム
  - 線形回帰（Linear regression（LR））
  - ロジスティック回帰（Logistic regression）
  - 正則化回帰（Regularised regression: LASSO and Ridge）
  - 一般化線形モデル（Generalised linear model（GLM））
  - 一般化加法モデル（Generalised additive model（GAM））
  - 決定木（Decision tree（DT））
  - ルール/ 決定リストとセット（Rule/ decision lists and sets）
  - 事例ベース推論/ プロトタイプと批判（Case-based reasoning（CBR）/ Prototype and criticism）
  - 超疎線形整数モデル（Supersparse linear integer model（SLIM））
  - 単純ベイズ（Naïve Bayes）
  - K近傍法（K-nearest neighbour（KNN））
- 「ブラックボックス」とみなされているアルゴリズム
  - サポートベクターマシン（Support vector machines（SVM））
  - 人工ニューラルネット（Artificial neural net）
  - ランダムフォレスト（Random forest）
  - アンサンブル学習手法（Ensemble methods）

ICOガイダンス第二部：実際にAIを説明する

○適切なAIモデルの選択：　①AIシステムを用いる分野の考慮

- AIシステムが適用される分野における個別の標準や慣習、要件を考慮する。
- 例えば金融サービス分野では、与信やローンの決定に対する厳格な正当化基準により、完全に透明で容易に理解可能なAI意思決定サポートシステムの利用が必要とされる。同様に医療分野では、厳格な安全性基準により、治療ツールや意思決定サポートツールには詳細なレベルのパフォーマンス試験・検証・保証が求められている。
- このような分野固有の要因は、AIモデルの複雑性や解釈可能性（interpretability）に関する選択にアクティブなインプットとなるべきである。

○適切なAIモデルの選択：　②AIシステムの影響の考慮

- AIシステムのアプリケーションの種類、また個人への潜在的な影響について考慮する。
- 例えば、コンピュータ画像認識システムと言っても、従業員の手書きのフィードバックを分類するためのシステムと、セキュリティポイントでの安全性リスクを分類するためのシステムとでは大きな違いがある。同様に、ランダムフォレストモデルでも、運転免許庁において申請者をトリアージするためのものと、病院の救急救命科において患者をトリアージするためのものでは違いがある。
- よりリスクの高い、あるいは安全最重視のアプリケーションでは、AIモデルが適切なアウトプットを保証しうることを徹底的に検討するべき。
- 安全性がそれほど重視されず、人々の生活に直接的な影響を及ぼさず、センシティブなデータを処理することのないようなリスクの低いAIシステムに対しては、最適なパフォーマンスを行う、高度に解釈可能なシステムを開発することに事業者が過大なリソースを費やす必要性は低いかもしれない。

ICOガイダンス第二部：実際にAIを説明する

○適切なAIモデルの選択：　③現行システムを新たなAIシステムで置き換えることの考慮

- 既存のデータ分析システムを、よりリソースを使い、より説明可能でないような新たなAIシステムで置き換えることのコストとベネフィットを検討すべき。
- 例えば、既存システムを新たなAIシステムで置き換えることに伴う、パフォーマンスの向上と、解釈可能性（interpretability）の低下とのトレードオフを検討すべき。

○適切なAIモデルの選択：　④利用するデータの考慮

- 適切なAIモデルを選択するために、事業者はどのようなデータをどのような目的で処理しているかを検討する必要がある。
- 検討の上では、以下の2グループのデータに注目することが助けになる。
  - ⅰ．人口統計的な特徴や、人間のふるまいの測定、社会的・文化的な特徴を示すデータ
  - ⅱ．研究や診断で用いられる生体医療データのような生物学的・身体的データ
- グループⅰのデータを処理する場合には、事業者はバイアスや差別のようなイシューに直面するかもしれない。この場合、事業者は最も解釈可能なモデルを選択することを優先させるべきであり、「ブラックボックス」システムは避けるべきである。
- グループⅱのデータを処理する場合、科学的洞察を得る目的（例えば、放射線診断）のみで、あるいは運用上の機能を果たす目的（例えば、車のナビゲーションにおける画像認識）のみで処理する場合には、より複雑な（説明可能性が低い）システムが適切であるかもしれない。しかし、ハイインパクトな、あるいは安全最重視なアプリケーションの場合は、事業者はモデルの選択においてAIシステムの安全性とパフォーマンス（正確性、セキュリティ、信頼性、堅牢性）を大きく優先させるべきである。ただし、グループⅡのデータの処理においてもバイアスや差別の問題が生じうることには留意すべきである（例えば、学習用データにおける代表性）。

ICOガイダンス第二部：実際にAIを説明する

○適切なAIモデルの選択：　④利用するデータの考慮（続き）

- 検討すべき、もう1つのデータ区別が以下である。
  - 伝統的データ：

    個人の購入履歴、職場での雇用期間など
  - 非伝統的データ：

    携帯電話や位置情報機器から収集されたセンサーデータや、ソーシャルメディアから収集されたテキストデータなど
- 非伝統的データが、個人に影響を与える意思決定に用いられる場合には、事業者は以下に留意すべきである。
  - 非伝統的データは、前述のグループｉのデータと同様に考えることができ、同じ仕方で扱うことできる。
  - 事業者は、ブラックボックスモデルではなく、解釈可能な（interpretable）結果を生み出す透明で説明可能なAIシステムを選択すべきである。
  - 例えば、信用リスク分析システムでGPS位置データ（非伝統的データ）が用いられる場合、事業者はその位置データによって個人のどのような重要な特徴が示されると想定しているのか、メタデータ等で示さなければならない。

ICOガイダンス第二部：実際にAIを説明する

○解釈可能なアルゴリズム（Interpretable algorithms）

- 可能かつ適切な場合には、標準的かつ最大限に解釈可能なアルゴリズム技術を採用するべき。
- ハイインパクトであったり、安全最重視であったり、その他の潜在的にセンシティブな環境においては、アカウンタビリティと透明性を最大化するようなAIシステムが必要とされる。
- このようなアルゴリズム技術には、決定木／ルールリスト、線形回帰やその拡張版の一般化加法モデル、事例ベース推論、ロジスティック回帰が含まれる。

○「ブラックボックス」AIシステム（'Black box' AI systems）

- ある種のデータ処理活動にとっては、解釈可能なAIシステムを用いることが困難であるかもしれない。例えば、画像の分類や、会話の認識、ビデオ映像での異常検知を行うようなAIアプリケーションにおいては、最も効果的な機械学習アプローチは不透明である可能性が高い。これらの種類のAIシステムの特徴量空間は、指数関数的に数十万や数百万の次元に増えている。このような複雑性のスケールにおいては伝統的な説明手法は適用できない。
- 事業者は、ブラックボックスモデルの潜在的な影響と将来的なリスクを徹底的に検討し、このようなシステムの責任ある設計と実装が可能と判断した場合に限り、ブラックボックスモデルを採用するべき。
- また、ブラックボックス的なアルゴリズム技術を採用する際には、分野・ユースケースに適切な、補足的説明ツールを用いるべきである。

ICOガイダンス第二部：実際にAIを説明する

〇補足的な説明手法・ツール

- 代理モデル（Surrogate models（SM））
- 部分従属プロット（Partial Dependence Plot（PDP））
- 個別条件付き期待値プロット（Individual Conditional Expectations Plot（ICE））
- 累積局所効果プロット（Accumulated Local Effects Plots（ALE））
- グローバル変数重要度（Global Variable Importance）
- グローバル変数交互作用（Global Variable Interaction）
- Sensitivity Analysis and Layer-Wise Relevance Propagation（LRP）
- Local Interpretable Model-Agnostic Explanation（LIME） and anchors
- Shapley Additive ExPlanations（SHAP）
- 反実仮想の説明（Counterfactual Explanation）
- Self-Explaining and Attention-Based Systems

英国のデータ倫理イノベーションセンター（CDEI）

- データ倫理イノベーションセンター（Centre for Data Ethics and Innovation：CDEI）は、英国政府が900万ポンド（約12億6000万円）を投資して、2018年11月に設立した諮問機関。AIを含むデータのガバナンス状況をレビューし、それらの倫理的・安全・革新的な利用を可能とし保証する方法に関して政府に助言を行う。
  - データ駆動型技術の導入は我々の社会のあらゆる側面に影響し、機会とともに倫理的課題を生み出している。CDEIは、データ駆動型技術のベネフィットを最大化する方法について調査しアドバイスするために、英国政府によって設立された独立諮問機関。
  - CDEIは、様々な分野や学界の人々を一堂に会し、政府に対する実践的な勧告を作成する。また、規制者や産業界に、責任あるイノベーションを支援し、信頼できるガバナンスシステムの構築を支援するようにアドバイスを行う。
  - CDEIの目的は、リベラル（自由主義的）な民主主義社会における倫理的・社会的な制約の中で、データ駆動型技術の潜在的なベネフィットを最大限享受できるようにする方法を特定すること。
- CDEIはデジタル文化メディアスポーツ省（DCMS）から2019/20年に250万ポンド、2020/21年に500万ポンドの予算を与えられている。
- 2019年度のレビュー領域
  - オンラインターゲティング（≒プロファイリング）
  - アルゴリズムを用いた意思決定におけるバイアス

英国のデータ倫理イノベーションセンター（CDEI）

- オンラインターゲティングのレビュー
  - オンラインターゲティングに焦点を当て、オンラインでのメッセージ・コンテンツ・サービスのパーソナライゼーションやターゲティングを通じて、データがどのように人々のオンライン環境を形成するかについて調査する。
  - 「オンラインターゲティング」は、「当人または類似した他者のデータやオンライン行動・嗜好を用いた、個人または集団の識別（＝プロファイリング）と、それに伴うオンラインでのパーソナライズされたメッセージ・コンテンツ・サービスの配信」を意味する。
  - オンラインターゲティングの例には、オンライン広告、検索結果最適化、ニュースフィード最適化、パーソナライズされたレコメンデーションが含まれる。
  - 中間報告書は2019年夏までに、政府への勧告を含む最終報告書は2019年12月までに発行予定。
- アルゴリズムを用いた意思決定におけるバイアスのレビュー
  - 様々な分野におけるアルゴリズムのバイアスの課題を調査。調査対象の分野は、犯罪司法（警察活動）、金融サービス、雇用者採用、地方政府の4つ。これらの分野を以下の理由で選定。
    - 1) アルゴリズムによる意思決定を使用する可能性がある
    - 2) 意思決定が人々の生活に重大な影響を与える
    - 3) アルゴリズムがバイアスのかかった意思決定を生み出したり悪化させるリスクがある
    - 4) アルゴリズムによって、意思決定における既存のバイアスに対処できる可能性がある
  - 中間報告書は2019年夏までに、政府への勧告を含む最終報告書は2020年3月までに発行予定。

CDEI：オンラインターゲティングに関する報告書

- オンラインターゲティングは、従来のターゲティングに比べて、3つの明確な特徴を持つ。
  - ①オンラインターゲティングは人々に関する前例のない量のデータの収集を伴う。また個人や集団に関するパワフルな洞察を生み出す洗練された分析を用いる。
  - ②オンラインターゲティングは比較的低コストで、様々な個人に対して大規模に、コンテンツや製品、サービスのターゲティングを可能にし、しばしば自分がターゲティングされていることや、どのような根拠に基づいてターゲティングされているかについて、個人に対する透明性を欠いている。
  - ③ターゲティングのアルゴリズムは、リアルタイムにアウトプットをモニターしたり、ターゲティングアプローチを改善することによって、我々の行動から学習している。
- オンラインターゲティングへの英国政府の取組み
  - Online Harm Agenda： 2019年早期に、英国政府はOnline Harm White Paperを公表。企業に利用者（とりわけ子どもその他の脆弱な集団）のオンラインでの安全に対する法的義務を設けようとするもの。企業は、広範囲にわたるオンラインの害悪（違法活動や違法コンテンツ、必ずしも違法ではないが有害なふるまい）に対処することに責任を持ち、独立的な規制者によって監視されることを提言。
  - オンライン広告規制レビュー： 2019年2月に、英国政府は、英国においていかにオンライン広告が規制されているかのレビューを行うとアナウンス。
  - GDPRとeプライバシー規則案／PECR（英国のプライバシー・電子通信規則）関連： ICOは、年齢に応じたオンラインサービスの設計規範の作業をしており、他にも、アドテックとリアルタイム入札、政治広告におけるデータ分析の利用、クッキー利用ガイダンス更新版などの作業をしている。
  - 競争とプラットフォーマーのイシュー： CMA（競争及び市場監督機関）はデジタル市場戦略の一環として、英国におけるオンラインプラットフォームとデジタル広告市場の市場研究を立ち上げた。

CDEI：オンラインターゲティングに関する報告書

1. オンラインターゲティングに対する一般市民の態度

- 市民との広範な対話活動におけるほとんどの参加者は、オンラインターゲティングは個人にベネフィットを与えていると感じており、良い顧客経験を創造する上で重要な役割を演じているとみなしていた。参加者たちは、オンラインターゲティングは、オンラインで自分たちが最も興味のある情報やサービス、製品に振り向け、識別する上で非常にベネフィットがあることを認識していた。

- しかし、ターゲティングに対する一般市民の態度は、それらがどのように機能し、どのように侵害的であるかを多く理解したときに変化する。ほとんどの人々は、ある種の形態のターゲティングは自分たちを不快にさせること、そしてターゲティングが実施されたり監視されたりするやり方に変更が必要であることに同意しているように見える。

- 最も一般市民が懸念する領域は、①個人の脆弱性を搾取する可能性や、②情報のトラストや社会における政治的議論に対するターゲティングの影響に関連するものであった。しかし、ターゲティング技術が、人々を脆弱性に基づいて搾取しうるならば、それは人々を保護する（例えば人々の注意を喚起するような支援サービスを提供する）ことにも利用しうるという理解も存在した。

- 多くの参加者は、オンラインターゲティングによってもたらされる害悪から保護される場合に、トレードオフがあるかもしれないことを理解していた。例えば、企業による個人情報へのアクセスを制限すれば、個人が享受するベネフィットは減少するかもしれない。

- 一般市民との対話の次のフェーズでは、我々は参加者たちが追加的な保護（個人としての保護と社会としての保護の両者）を求めているかどうかを調査する。参加者たちは、自分たちがオンラインで見るコンテンツや自分たちがターゲティングされる方法に対してどの程度までコントロールしたいかについて議論することとなろう。参加者たちはまた、個人、産業、政府、規制者の間の責任のバランスについても検討することとなろう。

CDEI：オンラインターゲティングに関する報告書

2. 規制とガバナンス

- 我々のスタート時点の立場は、法律や規制は、マーケットや規制メカニズムが失敗しているということ、またその結果として生じる害悪を低減する何らかの有効な手段が存在するという明確な証拠がある場合にのみ、強化すべき。
- オンラインターゲティングは、自律性および社会的結束（social cohesion）の観点で、機会と課題を提示するというのが我々の抱きつつある見解である。とりわけ課題のある領域には以下が含まれる。
  - ①脆弱な個人の保護： オンラインターゲティングは、様々な人々、例えば子ども、メンタルヘルスが弱い人々、その他オンラインのやり取りにおいて潜在的に脆弱なタイプの人々などが、その脆弱性やその他の要因に基づいてターゲティングされることを可能にする。このことは、脆弱な人々の搾取だけでなく、支援を与える機会をも提供する。
  - ②情報におけるトラスト（とりわけメディア、広告、政治的コンテンツ）： オンラインターゲティングは、個人に関連性があったり興味を持つ可能性が高いニュースコンテンツを人々に提示するかもしれない。しかし、「クリック・スルー・レート」等の測定基準を最適化するシステムを通じて、フェイクコンテンツやミスリーディングなコンテンツが増幅されることも可能とするかもしれない。このことは、「フィルターバブル」の発展を通じて社会的結束に影響を与えるかもしれず、市民や消費者として十分に情報を与えられた決定を行う人々の能力にインパクトを与えるかもしれない。社会的なレベルでは、このことは、民主主義制度の有効性や完全性にインパクトを与えるかもしれない。
  - ③マーケットにおけるトラスト： オンラインターゲティングは、自分に関連する適切な価格の製品を見出すことを手助けするかもしれない。しかし、自分に提示される製品と価格が、他の人のものと異なっているかどうかを知ることができない場合、オンラインマーケットにおける全般的なトラスト（およびその有効性）を減らすかもしれない。
  - ④有害な差別からの保護： ターゲティングは、その本性上、差別的なものである。このことは、人々に最も関連するコンテンツがターゲティングされる場合は、ポジティブなものでありうる。しかし、それがターゲティングクライテリアやアルゴリズム、データセットに埋め込まれた社会的バイアスを反映しているならば、有害なものでもあるかもしれない。
- これらの観点から見たオンラインターゲティングの現状の／潜在的な影響は、現在のガバナンスと監視の有効性、また改善の機会について我々がアセスメントする際の物差しとなるだろう。

CDEI：オンラインターゲティングに関する報告書

2. 規制とガバナンス（続き）

- オンライン世界には、マーケットの有効性や、個人が自分の利益を保護する能力を弱体化する恐れのある、いくつかの特徴がある。とりわけ、以下の3つの特徴である。
- ○限られた競争：
  - マーケットは、個人がサービスの選択を行い、自分のニーズに合うものや役立つものを正確に区別できる場合には、悪いプラクティスを駆逐する可能性が高い。しかし、巨大にネットワーク化されたオンラインプラットフォームや、アドテックやデータ交換エコシステムの侵害的な運用によって、データ駆動型技術が人々をターゲティングする方法においては、限られた競争しか存在していない。
  - ストリーミング音楽や映画レコメンドサービスのように、幾つかの領域では幅広い競合サービスが存在する。ソーシャルメディアプラットフォームのような他の領域では、競争は限られている。消費者が、自分たちに見せられているものは本当に自分たちが見たいものなのかどうかをアセスすることは困難または不可能である。消費者は、ターゲティングが自分たちの最も興味のあるものに役立っているのかどうかを知ることはできない。
- ○限られた透明性：
  - あるアルゴリズムは顧客の経験を改善することを目的とするかもしれないが、他のアルゴリズムは個人に影響を与えたり、操作したり、搾取したりしようとするかもしれない。同じアルゴリズムが、個人に応じて、両方の効果を持ちうる。保有されている個人情報やその利用方法に関する透明性は限定されているが、改善されつつある。しかし、どの程度個人が差別的に扱われているか、また、そのインパクトについてはほとんど透明性がない状況。
- ○消費者に対する害悪のリスク：
  - 企業が個人情報を利用し、個人の脆弱性につけこんで、製品やサービスをターゲティングする商業的なインセンティブがある。個人に対して自社のオンラインサービスを可能な限り長い時間使うように促進する商業的なインセンティブもある。この結果、アルゴリズムは個人の短期的な注意を捉え、個人のオンライン上での本能的な反応を悪用するように最適化される。人々の欲望に適合するように設計されたオンラインサービスが、より望ましいとされる恐れがある。

CDEI：オンラインターゲティングに関する報告書

3. ソリューション

- CDEIは以下の4つの領域における介入を検討している。
- ○アカウンタビリティと監視
  - プラットフォーム上で配信されるコンテンツに対するより大きな責任をプラットフォームに持たせたり、ターゲティングアルゴリズムの受容性を決定するために利用されるプロセスに関してオープンにしたりするためのメカニズム。
  - 個人や社会に、表示される広告やコンテンツ、またそれらが表示される集団に関する、より大きな透明性を与えるためのメカニズム。
  - ターゲティングのインパクトに関する独立の調査研究を行ったり、情報収集を可能とするためのメカニズム。
- ○コンテンツ配信に対するより強い規制
  - ターゲティングプロセスで利用される推論のタイプ、あるいは実施されるターゲティングの種類や範囲を制限するためのメカニズム。
  - 脆弱な人々を保護するために企業に対してより強い義務を導入するためのメカニズム。
  - 信頼できる情報源、あるいは多様な情報源をデフォルトとする、より強い義務を導入するためのメカニズム。
  - 仲介や苦情、自動モニタリングによって生じた懸念に対応してコンテンツを制限するためのメカニズム。

CDEI：オンラインターゲティングに関する報告書

3. ソリューション（続き）

- 〇個人のパワーや情報を強化する
  - より強力な同意ルールを求めるためのメカニズム。
  - 保有されている個人データ、それらのターゲティングプロセスでの利用方法、それらの情報源に関するより大きな透明性を導入するためのメカニズム。
  - 脆弱な人々からの搾取を防ぐために、個人が利用可能なツールを改善するためのメカニズム。
  - 信頼できる情報源、または多様な情報源。
  - 救済手段へのアクセス。
  - GDPRで規定されたデータポータビリティの権利の促進。
- 〇競争を促進する
  - 第三者が個人データを本人の代わりに管理するような、新たなビジネスモデルの開発を支援するような政策。

CDEI：バイアスに関する報告書

- アルゴリズムのバイアスに関する4つの調査分野

〇犯罪司法（警察活動）

- 機械学習アプローチをトレーニングするのに必要とされる警察のデータは、様々な品質のデータであったり、重大な歴史的なバイアスを含んでいる可能性がある。バイアスがかかっていたり、不完全だったりする歴史的なデータから学習したアルゴリズムは、一定の集団や個人に対してバイアスがかかった犯罪司法判断を継続したり助長したりする恐れがある。
- 警察活動において、今日では予測的分析技術が以下の方法で利用されている。
  - 予測的な犯罪マッピング： 近い将来に犯罪が生じやすい地理空間的な位置を予測し、最も必要な場所に人員リソース等を優先的に配備する。
  - アルゴリズムによる意思決定のサポート： アルゴリズムを用いたリスクアセスメントツールを、個人に関連した予測を行う。例えば、過去の行動の分析に基づくと近い将来における高い再犯リスクを示すような高リスクの犯罪者を識別する。

〇地方政府

- 地方政府は住民に関する重大な意思決定、とりわけ、リスクがあると判断され、支援や介入が必要とされる子どもに関する意思決定に責任を負っている。
- 大学や民間企業ではこの分野を目的とした予測的分析ツールの開発を始めている。いくつかの地方政府はこれらのポテンシャルを検討し始めている。この分野での入手可能な調査研究は比較的限られたものであるが、支援が必要とされる住民の脆弱性や、地方政府のデータの品質（の悪さ）に対する懸念に鑑みると、データ駆動型技術の利用がバイアスを悪化させるリスクが存在する。CDEIは、リスクがあると判断された子どもへの介入に関する意思決定をサポートするツールの利用に関心を持っている。

CDEI：バイアスに関する報告書

- アルゴリズムのバイアスに関する4つの調査分野

〇金融サービス

- 金融サービスには伝統的な差別の歴史が存在する。例えば、民族的マイノリティを郵便番号で赤線引きすること（居住地域に基づく融資差別）や、ローンにおいて女性には夫のサインを必要とすることである。
- このような酷い取扱いは現在では行われていないが、金融企業はいまだ、金融リソースが様々な集団間で均等に広がらないような社会経済環境において運用を行っている。結果として、これらの歴史的なバイアスがデータの中に反映されているかもしれない。
- 金融サービス分野におけるCDEIのレビューの焦点は、個人顧客に対する信用（貸付）や保険の意思決定である。金融サービス分野は長年にわたり統計的手法を用いる確固とした歴史を有している。例えば、高度に規制された市場における信用レイティング（信用スコア）である。
- CDEIは近年の技術的な進展、とりわけ、非伝統的なデータ源からのデータの利用（例えばソーシャルメディア）や最近の機械学習アプローチに焦点を当てている。

〇雇用者採用

- 雇用におけるデータ駆動型技術の利用は今後数年間のトレンドであると予測されている。相対的に簡単なテキストスキャン技術から、より複雑なコンテンツ分析技術（履歴書・申請書のスクリーニング）、そしてAIによる面接まで多岐にわたりうる。
- 歴史的に、特定の職業は全ての人に等しくアクセス可能なものではない。性別による賃金格差を報告するような最近のイニシアティブは、現在でもこれらのバイアスが残存することを示している。
- Amazonが性別バイアスの理由で採用アルゴリズムの使用を取りやめたような事例は、アルゴリズムを用いた技術は、埋め込まれたバイアスを増長する可能性を持っていることを示唆する。

CDEI：バイアスに関する報告書

〇公平性に関する見解

- 「「公平性」の概念は普遍的なものでも、常に一貫性のあるものでもない。公平性は様々なコンテキストにおいて、また同一のコンテキスト内であっても、様々な事柄を意味しうる。例えば、手続き的な公平性は人々に対する「公平な扱い」（いかに意思決定がなされるか）に関わるものであるのに対し、結果の公平性は人々に対する「公平な影響」（どのような意思決定がなされるか）に関わるものである。「公平」な手続きは「不公平」な結果を生み出すかもしれず、逆もまたありうる。」
- 「人間のシステムでは、公平性のどちらの定義が適用されているかについて、曖昧さを残すことが可能である。しかしアルゴリズムは曖昧でないことが求められる。これらの定義のいずれにも適合するようにアルゴリズムを設計することはできるが、同時にこれらすべての定義に適合させることはできない。人間は、これらの定義のどれにアルゴリズムを適合させるか選択しなければならない。」
- 「英国の2010年平等法は、個人を以下の特徴に基づいて差別することを違法としている。すなわち、年齢・障害・性転換・婚姻・同性婚・妊娠・出産・人種・宗教・信仰・性別・性的志向である。これら保護すべき特徴に基づいて差別をもたらすようなバイアスは直接的なバイアスと言える。しかし、アルゴリズムの利用はこれ以外の間接的なバイアスをも生み出す恐れがある。アルゴリズムの利用は、明白でない特徴や可視的でない特徴に対する差別を生み出す機会を増やす。例えば、アルゴリズムは効果的に金融リテラシーのない人々を識別したり、そのことを利子率の設定や返済条件の設定に用いるかもしれない。」

CDEI：バイアスに関する報告書

〇公平性に関する見解

- 「アルゴリズムにデモグラフィックな特徴（差異）や、それらの特徴（差異）のプロキシ情報を「見せない」ようにすることは、必ずしも公平なアウトプットに結びつくわけではない。例えば、犯罪者の再犯リスクを計算するように設計されたアルゴリズムが性別を考慮に入れないようにすることは、女性が男性よりも再犯する傾向が全体的に低いため、女性に対して反対に、不均衡に厳しい結果になる可能性が高いかもしれない。性別のデータを排除することにより、当該アルゴリズムは女性に対して不正確になり、結果的に公平でなくなる。
- 差異が「見えない」ことはまた、アルゴリズムに間接的なバイアスがかかっているかどうかを知ることを不可能にするかもしれない。例えば、人事採用で差別的な決定をするリスクを防ぐために、組織は意思決定プロセスから求職者の性別や人種（民族）を識別しうるデータを削除しようとするかもしれない。このことは、求職者をスクリーニングするためのアルゴリズムがこれらの特徴を直接的に参照できないことを保証するだろう。しかし、このアプローチは、機械学習アルゴリズムが性別や人種のプロキシ情報（例えば、人種と緊密に関連する郵便番号など）を用いてしまう可能性を排除しない。またこの例では、データセットから人種（民族）に関するデータを削除することは、このような間接的なバイアスが生じているか否かを評価することを不可能にするかもしれない。
- このことは、以下のような重要な緊張関係を明るみに出す。一方で、意思決定プロセスの一部として「異なる扱い」を避けるために、センシティブな属性はアルゴリズムによって参照されるべきではない。他方で、「異なる影響」を評価するためには、センシティブな属性はアルゴリズムが公平かどうかを確認する責任を有する者によって検査されなければならない。」

CDEI：バイアスに関する報告書

〇公平性に関する見解

- 「データ自体はしばしばバイアスの源となるが、同時に、データはバイアスの問題に対処する上でコアとなる要素である。データセットは頻繁に重大なバイアスを含む。それは、データセットが不完全で過少代表のものであったり、それらがバイアスの歴史的なパターンを正確に反映するものであったりするためである。例えば、警察活動における予測的分析ツールの利用に関する我々の初期の調査は、本技術の利用に関する主要な問題の1つは歴史的なデータセットに埋め込まれた潜在的なバイアスであることを示唆している。」
- 「意思決定アルゴリズムへのインプットとして、保護すべき特徴（あるいはそれらの特徴のプロキシ情報）に関するデータを利用することを避けることは、共通的なプラクティスである。なぜなら、そうすることは不法な差別を生み出す恐れがあるからである。しかし、意思決定で影響を受ける個人の中での保護すべき特徴の分布を理解することは、バイアスがかかった影響を識別するために必要なことである。例えば、各従業員が男か女かを知ることなく、ある企業における性別による賃金格差の存在を確かめることは不可能である。このような、保護すべき特徴をアルゴリズムに「見せない」ようにするニーズと、保護すべき特徴に対するバイアスがかかっていないかをチェックすることとの間の緊張関係は、組織がデータを責任ある仕方で利用しようとする上でのチャレンジとなる。」
- 「我々はどこまでバイアスを軽減すべきか、そしてそのために我々のアプローチをどのように統治すべきかを決めなければならない。これらの決定は、価値判断と、競合する価値間のトレードオフを必要とする。人間はしばしば、様々な考慮事項に対してどのような重み付けを行っているかについて明示的に言及する必要なく、これらのトレードオフを行うことに信頼を置かれている。アルゴリズムはそうではない。アルゴリズムはルールに則ってトレードオフを行うようにプログラミングされており、その決定は問いただすことができ、明示的にできる。このことは、人間とは異なるアカウンタビリティのアプローチを必要とする。」