資料１－１

AIと個人情報保護：欧州の状況を中心に

１．GDPRとAI
２．EU AI法案：AIシステム全般と生成AI
３．EUデジタルサービス法とAI
４．3年ごと見直しに関連しうる論点
　　論点に関する補論

2024年4月3日

生貝直人　博士（社会情報学）
一橋大学大学院法学研究科教授

1．GDPRとAI

- 自動決定からの保護
  - 透明性、異議を申し立てる権利等（22条）
  - 信用スコアが与信判断に決定的な役割を果たす場合、判断を人間が行っても22条の自動決定と評価されうる（ECJ C-634/21、2023年12月）→AI法での補完規律へ
- 学習データとスクレイピング（CNIL、2022年10月等）
  - 法執行機関への顔識別システム提供のための顔画像収集・処理を、6条(f)の正当な利益根拠の対象とならないと判断→AI法での補完規律へ
  - 透明性ある情報提供（12条）、アクセス権（15条）、消去権（17条）

  →ChatGPT（EU各国）に関しても類似の論点（特に法的根拠）

２．EU AI法案：AIシステム全般

- AIシステムをリスクに応じて4段階に分類した規律を置く
  - **許容できないリスク（禁止されるAI慣行）**
  - **ハイリスク（適合性評価等の義務）**
  - 限定的リスク（透明性義務）
  - 低リスク・無リスク（拘束力の無い行動規範）
- 禁止されるAI慣行（概略）
  - 判断能力を著しく損なうサブリミナル技法や操作的・欺瞞的技法
  - **年齢、障害、特定の社会的・経済的状況に起因する脆弱性の悪用**
  - **本人や集団に不利な影響を与える社会的スコアリング（例外有）**
  - 個人の性格特性や特徴のプロファイリングのみに基づく犯罪予測
  - **顔識別DB作成目的のインターネット・CCTV顔画像無差別スクレイピング**
  - **職場及び教育機関における感情識別（医療・安全目的の例外有）**
  - **生体識別データに基づく特別カテゴリーデータの推測（例外有）**
  - 法執行目的の公共のアクセス可能な空間での生体識別（例外有）

２．EU AI法案：AIシステム全般

- ハイリスクAIのカテゴリー
  - 既存EU法での適合性評価義務対象：機械、玩具、レジャー用船舶、リフト、爆発性雰囲気装置、無線機器、圧力機器、索道設備、個人用保護具、ガス機器、医療機器
  - AI法での新たな指定：**バイオメトリクス（遠隔生体識別・感情識別）**、インフラ管理・運用、**教育や職業訓練での学生や希望者の評価や受入れの合否、雇用、労働管理、自営業へのアクセス、重要な民間・公共サービス（公的支援金給付、融資、緊急対応措置）、法執行**、移民・亡命・国境管理、司法又は民主主義プロセス
- ハイリスクAIシステムの要求事項
  - リスクマネジメントシステムの構築、**データとデータガバナンス**、技術文書、記録保持、**透明性と利用者への情報提供、人間による監視**、正確性・堅牢性・セキュリティ
- ハイリスクAIシステム提供者の義務：上記要求事項の遵守確保等
- ハイリスクAI配備者の義務：**基本権影響評価**の実施等

2．EU AI法案：生成AI

- 汎用目的AIモデル（生成AIはその典型）を含むコンテンツ生成AI提供者の義務
  - 出力が機械可読形式でマークされ、人為的に生成又は操作されたことを検知可能とする
- 汎用目的AIモデル提供者の義務
  - 設計や学習等の技術文書作成と当局への提供
  - 下流事業者への情報開示
  - DSM著作権指令4条（学習データオプトアウト）遵守措置、**学習データの詳細な要約公表**
- システミックリスクを有する汎用目的AIモデル（10^25FLOPs以上等）提供者の義務
  - システミックリスク特定・軽減のためのレッドチームテスト実施・文書化を含むモデル評価
  - **システミックリスクの評価・軽減**
  - 重大インシデントへの対応文書化と当局への報告
  - サイバーセキュリティ対策

→それぞれの義務は整合規格により具体化、それまでは欧州委員会主導で策定する行動規範（codes of practice）の遵守

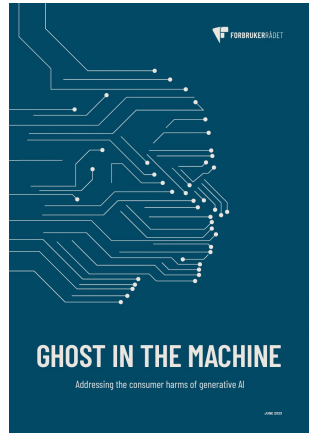
3条(65)「「システミックリスク」とは、汎用目的AIモデルの高インパクト能力に特有のリスクであって、その影響範囲の広さにより連合市場に重大な影響を及ぼし、または公衆衛生、安全、治安、**基本権もしくは社会全体に対する実際の若しくは合理的に予見可能な悪影響**により、バリューチェーン全体にわたって大規模に伝播し得るリスクをいう」

2．EU AI法案：生成AI

- 前文133「**さまざまなAIシステムが大量の合成コンテンツを生成できるようになり、人間が生成した本物のコンテンツとの区別がますます難しくなっている。**こうしたシステムが広く利用可能になり、その能力が高まることは、情報エコシステムの完全性と信頼性に重大な影響を及ぼし、**誤情報や大規模なマニピュレーション、詐欺、なりすまし、消費者への欺瞞といった新たなリスクを引き起こす。**こうした影響、技術の進歩の速さ、情報の出所を追跡するための新たな手法や技術の必要性を考慮すると、こうしたシステムのプロバイダーに対し、機械が読み取り可能な形式で表示し、その出力が人間ではなくAIシステムによって生成または操作されたことを検出できる技術的ソリューションを組み込むことを求めることが適切である。(…)」

"GHOST IN THE MACHINE: Addressing the consumer harms of generative AI" (Norwegian Consumer Council, June 2023)
「人間の行動をエミュレートする能力に制限を設けることなく、生成AIモデルを一般に公開することには根本的な問題がある。モデルが人間の感情をシミュレートするコンテンツを生成する場合、これは本質的に操作的（manipulative）である。"

→仮想友人サービスReplikaに対するイタリアGaranteの停止命令



３．EUデジタルサービス法とAI

- EUのプロバイダ責任を規定してきた電子商取引（2000年）を元に、違法・有害情報に対するプラットフォームの責任・責務や透明性のあり方を全面的にアップデート
- 媒介サービス事業者（IS）一般やオンラインプラットフォーム（OP）事業者一般に適用される規律の他、EU域内で月間アクティブ利用者4,500万人以上を有する「超大規模オンラインプラットフォーム（VLOP）」＋「超大規模オンライン検索エンジン（VLOSE）」事業者に、偽・誤情報を含むシステミックリスクの評価・軽減義務を課す
  - 2023年4月25日に17のVLOPと2のVLOSEが指定、2024年2月全面適用開始
- デジタルサービス法の要点
  - コンテンツモデレーション：透明性と救済（省略）
  - **プロファイリング関連規制**
  - **VLOP/VLOSEとシステミックリスクの評価・軽減**

３．デジタルサービス法とAI プロファイリング関連規制

- PF上の**ターゲティング広告**のパラメータ等の明示（OP~VLOP段階、26条他）
- **レコメンダーシステム**のパラメータ明示とユーザーによる修正可能性（VLOPはプロファイリングに基づかない選択肢の提供を含む）（OP～VLOP、27条他）
- **GDPR特別カテゴリー個人データのプロファイリング広告利用禁止（OP、26条3項）**
- **青少年保護と未成年個人データのプロファイリング広告利用禁止（OP、28条2項）**

- ※ダークパターンの禁止（OP、25条）：「サービス受領者を欺いたり操作したりするような方法で、又はその他の方法でサービス受領者が自由かつ情報に基づく決定を行う能力を実質的に歪めたり損なったりする方法で、オンライン・インターフェースを設計、組織、運用しないこと」

３．デジタルサービス法とAI
VLOP/VLOSEとシステミックリスクの評価・軽減

- VLOP/VLOSEは、自らのサービスがもたらしうる違法コンテンツ流布、**基本権（特に人間の尊厳、プライバシー、個人データ保護**、表現・情報の自由、非差別、児童の権利、消費者保護）、市民言説と選挙、ジェンダー暴力・公衆衛生・青少年保護等への影響**等の「システミックリスク」を自ら特定・分析・評価し（34条）、合理的・比例的・効果的な軽減措置を採る義務（35条）**と、公共の安全・公衆衛生への重大な脅威における危機対応メカニズムにおいて出される欧州委員会の要請決定の対象となる（36条）
- 欧州委員会が奨励・推進・招請して策定する、行動規範（codes of conduct）（45条）や危機プロトコル（48条）を通じて具体化する共同規制メカニズム
- 34条・35条の義務及び、行動規範・危機対応プロトコルの遵守について、年1回以上の独立監査を受ける義務（37条）
    - 評価・緩和措置検証のための外部研究者データアクセス提供義務（40条）

４．3年ごと見直しに関連しうる論点

- プロファイリング・自動決定に関わる規律
- バイオメトリクスデータと生体識別の位置付け
- スクレイピング、特に顔識別DB構築の位置付け
- ハイリスクや行うべきでないAI利用行為の特定
- 青少年やそれに限られない個人の脆弱性への対応
- デジタルプラットフォーム上のAIリスクへの対応
- 個人情報保護関連法制多元化への対応

論点に関する補論

補論①：個人情報保護関連法制多元化への対応

- データ、プラットフォーム、AI等の重要性の高まりに合わせ、国内外において個人情報保護関連法制の多元化が進む中、関係する各法領域と、個人情報保護法及び委員会の関係性を深化させていくべきではないか。
- EU法の例：
  - デジタルサービス法：26条3項・28条2項（特別カテゴリーデータ・未成年データプロファイリング広告利用制限）等のGDPR直接補完規定
  - デジタル市場法：5条1項（個人データ統合等の制限）、40条ハイレベルグループ（BEREC、EDPS、EDPB、ECN、CPCネットワーク、ERGAからの代表者で構成）
  - ePrivacy指令：5条3項（所謂cookie条項）の仏国内法をGDPRと合わせCNILが執行
  - AI法：禁止AI行為やハイリスクAIシステムでのプロファイリング関連規律
- また、GDPR70条に基づくEDPBオピニオン等を参考に、個人情報保護関連施策に対する委員会の助言機能の強化も検討の余地があるのではないか。

補論②：AIとプロファイリング

- 各領域におけるAIの利用が拡大する中、GDPRにおける自動意思決定の透明性確保やデータ主体保護の規定等を参考に、ベースラインとなるプロファイリング規律を検討する意義はあるのではないか。
- 同時に、プロファイリングが持ちうる個人の権利利益へのリスクは多様であり、特に「ハイリスク」な用途、あるいはそもそも行うべきではないプロファイリング関連行為も存在する（例えば個人の脆弱性を悪用するAIシステム等）。
- EUではGDPR以上の具体的規律はAI法を別途立法する形としたが、我が国においては、他法との補完関係を念頭に置きつつも、個人情報保護法においてリスクベースの具体的規律を検討する選択肢もあるのではないか。

補論③：デジタルプラットフォーム

- 同意の実質性の確保、特に当該サービスを事実上利用せざるをえない場合などへの対応は、独占禁止法等他法との補完関係を念頭に置きながらも、個人情報保護法においても規律の在り方を検討していくべきではないか。
- また、同意の論点以外にも、例えばデジタルサービス法における個人情報保護関連規律などを参照しつつ、マニピュレーション等の現代的課題への対応も視野に、事業者の規模やリスク等に応じた規律を個人情報保護法の中に設けることも検討する余地があるのではないか。

補論④：多様な個人の脆弱性への対応

- 同意取得や情報提供の在り方を含め、青少年の保護に関する具体的な規律を検討していく必要性は高いのではないか。それに際しては、データ保護バイデザイン・バイデフォルトや、行動規範・認証制度等の、一定の柔軟性を有する制度枠組を参考にするべきではないか。
- 同時に、青少年保護の他にも、高齢者や心身に障がいのある人々など、広く個人の脆弱性の観点を視野に入れた形で、本人保護の在り方を見直していくべきではないか。

補論⑤：認定個人情報保護団体制度等

- 認定個人情報保護団体制度につき、GDPRの行動規範制度（40・41条）における行動規範自体への認定枠組等を参考に、参加インセンティブの向上や指針の位置付けの見直し等を通じて、その役割を拡大する余地は大きいのではないか。
- 同時に、GDPRの行動規範制度は、個人データ処理の法的根拠（特に正当な利益）等の規律の具体化や、アカウンタビリティ（遵守の証明）といった、制度全体の特性がその役割を規定している側面が大きい。個人情報保護制度の中期的な在り方も考慮しつつ、当事者の知識を活かすための、官民の共同規制枠組の役割を検討していくべきではないか。
- 合わせて、国際的な認証制度の重要性が拡大する中、GDPRの認証制度（42・43条）も参考に、認証制度の法的位置付けについても検討する余地があるのではないか。