資料１－４

第45回世界プライバシー会議クローズドセッション
2023 年 10 月
生成人工知能システムに関する決議（仮訳）

スポンサー:

・欧州データ保護監察機関 （ＥＤＰＳ）

共同スポンサー：

・公共情報アクセス庁（ＡＡＩＰ、アルゼンチン）
・ブエノスアイレス市オンブズマン事務所（アルゼンチン）
・連邦データ保護・情報自由監察官（ＢｆＤＩ、ドイツ）
・ラインラント・プファルツ州データ保護・情報自由監察官（ＤＲＬＰ、ドイツ）
・個人データ管理・保護委員会（ＣＮＤＰ、モロッコ）
・情報処理と自由に関する国家委員会（ＣＮＩＬ、フランス）
・欧州評議会 （ＣｏＥ）
・連邦データ保護情報コミッショナー（ＦＤＰＩＣ、スイス）
・個人データ保護機関（ＧＰＤＰ、イタリア）
・情報コミッショナーオフィス（ＩＣＯ、英国）
・透明性、情報アクセス及び個人データ保護に関する国立機関（ＩＮＡＩ、メキシコ）
・情報データ保護コミッショナーオフィス（ＩＤＰＣ、マルタ）
・オンタリオ州情報プライバシーコミッショナーオフィス（ＩＰＣ、カナダ）
・プライバシーコミッショナーオフィス（ＯＰＣ、カナダ）
・個人データプライバシーコミッショナーオフィス（香港、中国）
・個人情報保護委員会（ＰＰＣ、日本）
・個人情報保護委員会（ＰＩＰＣ、韓国）
・プライバシー保護当局（ＰＰＡ、イスラエル）
・データ保護機関（ＡＧＥＳＩＣ、ウルグアイ）

第45回世界プライバシー会議クローズドセッションは、

個人データを大規模に処理することが多い人工知能（ＡＩ）技術の急速な発展を認識し、

生成ＡＩ技術とアプリケーションの広範な導入、様々な領域での採用の増加、及び世界中での利用の広範な普及に留意し、

特にプライバシーと個人データの保護に対する基本的な権利に関連して、基本的な権利と自由に対する影響を含む、生成ＡＩ技術の倫理的及び法的な影響に関する公開討論で表明された様々な懸念を認識し、

開発及び規制が適切になされない場合には、データ保護、プライバシー及びその他の基本的人権にリスクや潜在的な危害をもたらす可能性がある生成ＡＩシステムが、多くの場合、導入前の評価が不十分なままに一般に公開されることを懸念し、

法的権限なしに、公的にアクセス可能な領域やソースから個人データが無差別に収集され、生成ＡＩ技術に投入されることに懸念し、

大量のデータを処理するために相当な計算リソースを費やす生成ＡＩモデルの開発と導入が環境に及ぼす影響を認識し、

様々な管轄区域がＡＩに特化した法律や政策の作成を継続していると同時に、データ保護・プライバシーの原則と現行法（法律、法案、法規及び規制を含む）が生成ＡＩ　製品・サービスに適用されることを想起し、

生成ＡＩを利用したサービス、特に個人データの処理と未成年を含むデータ主体の透明性に関して、様々なデータ保護当局が継続的な政策立案、調査、執行措置を行っていることを認識し、

2023年6月21日のG7データ保護・プライバシー機関（ＤＰＡｓ）ラウンドテーブルによる生成ＡＩに関する声明[1]（この声明は、プライバシー・データ保護のリスクが発生する可能性のある主要分野に特別な注意を喚起し、データ保護とプライバシーの原則が生成ＡＩツールの開発・利用に適用されることを示している）を歓迎し、

[1] https://www.ppc.go.jp/files/pdf/G7roundtable_202306_statement.pdf

生成ＡＩの文脈において共通の価値観を促進するには、法的な観点だけでなく倫理的、社会的及び技術的な観点からも、継続的なグローバルな議論と協力が必要であることを認識し、

データ保護・プライバシー機関（ＤＰＡｓ）を含む、世界中の公的機関が、生成ＡＩツールの起源となる基盤モデルが、データ保護・プライバシーを含む個人の権利と自由を完全に尊重し、人権法に反する不公平、非倫理的、又は差別的な扱いを防止するために、果たすべき重要な役割を担っていることを強調し、

自動化された意思決定や、特に脆弱な人々や子供に対するリスクの高い利用において、生成ＡＩシステムの持つ特有のリスクと潜在的な危害を認識し、

第40回世界プライバシー会議で採択された「人工知能における倫理とデータ保護に関する宣言」（この宣言は、人工知能システムがプライバシー・バイ・デザインやプライバシー・バイ・デフォルトの原則を適用して設計及び開発されるべき場合には、「倫理・バイ・デザイン」を求めるものである）を想起し、

第41回世界プライバシー会議で採択された「基本的人権としての、かつ、他の基本的権利を行使するための前提条件としてのプライバシーに関する国際決議」（この決議は、デジタル社会における信頼を構築し、イノベーションを加速し、かつ、人間の尊厳を保護するため、生成ＡＩは民主主義的価値観に基づいて人間中心であるべきであり、かつ、他の権利と自由の保護のために不可欠な基本的人権としてプライバシーを認識すべきである、とするものである）を再確認する。

第45回世界プライバシー会議は、次のことを強調する。

生成ＡＩは、他のＡＩシステムと同様に、データ保護・プライバシー、人間によるコントロール、透明性及び民主的価値観の原則に基づいて、責任がある及び信頼性がある方法で、設計、開発及び導入すべきである。

生成ＡＩシステムの開発者、提供者及び導入者は、「データ保護・プライバシー・バイ・デザイン」の原則に基づいて、生成ＡＩシステムを利用した新しい製品とサービスの構想、設計、運用及び管理にデータ保護・プライバシーを組み込み、データ保護・プライバシー影響評価において行った自らの選択と分析について文書化すべきである。

さらに、生成ＡＩシステムの開発者、提供者及び導入者は、倫理的で、信頼性のある、そして責任ある生成ＡＩを開発するための基礎として、影響を受ける個人と社会全体に対するリスク、危害及び潜在的な影響を理解すべきである。

また、生成ＡＩシステムの開発者、提供者及び導入者は、データ保護・プライバシーの義務を確実に遵守するための措置を講じる必要がある。これらのステークホルダーは、生成ＡＩシステムによって自分のデータが処理される個人がデータ保護・プライバシーの権利を確実に行使できるよう協力する必要がある。

生成ＡＩシステムの開発者、提供者及び導入者は、法的要件と、そのような法的要件を解釈する方法に関するＤＰＡのガイダンスに細心の注意を払うべきである。適切な場合には、ＤＰＡとの緊密なコミュニケーションを図ることが、生成ＡＩシステムに基づく製品とサービスの責任ある設計、開発及び導入に貢献し得る。

この点に関して、第45回世界プライバシー会議会は、生成ＡＩシステムの開発、運用及び導入の中核要素として､既存のデータ保護･プライバシーの原則を支持する。

１．個人データ処理の法的根拠

生成ＡＩシステムは、個人データが公的にアクセス可能な場合でも、適用される法律に従って法的根拠を有することが必要であり、合法的である必要がある。開発者は、システムを起動する前に、生成ＡＩシステムが合法であり、かつ安全であることを確立する必要がある。

関連法令で義務付けられている場合、生成ＡＩシステムの開発者、提供者及び導入者は、以下に関連する個人データの処理の法的根拠を最初に明らかにしなければならない。

a）生成ＡＩシステムの開発に利用されるデータの収集
b）生成ＡＩシステムの開発又は改善に利用されるデータセットの学習、検証及びテスト
c）生成ＡＩシステムと個人の関わり合い
d）生成ＡＩシステムによって生成されたコンテンツ

２．目的の明確化及び利用の制限

生成ＡＩシステムの開発者、提供者及び導入者は、特定の、明示的かつ正当な目的のために個人データを処理すること、また、それらと一致しない目的で、又は影響を受ける個人の合理的な期待を超えて、さらに処理しないことを確保する。これらの目的は、状況に応じて、適切、合理的又は正当なものでなければならない。

生成ＡＩシステムの開発者、提供者及び導入者は、その利用が違法である、又は、不公平、非倫理的若しくは差別的な扱いにつながる可能性がかなり高い、特に、これが基本的な権利と自由の重大な侵害につながる、生成ＡＩを開発し、又は運用してはならない。ＡＩシステムが個人に関する意思決定を行ったり支援したりするために利用される場合には、上記のことは、より基本的なことである。

生成ＡＩシステムの開発者、提供者及び導入者は、開発に利用される個人データが収集された目的との倫理的、法的、社会的及び技術的な整合性を慎重に評価するべきである。

３．データ最小化

生成ＡＩシステムの開発者、提供者及び導入者は、正当に特定された目的を達成するために必要な限度において、個人データの収集、共有、集約、保持及び更なる処理を行う必要がある。個人データの収集及び処理は、無差別に行ってはならない。

また、個人データが学習セットに含まれると、特に学習データ内の情報が生成ＡＩシステムの出力の一部として生成される可能性があるなど、プライバシーやその他のリスクが個人に生じる。したがって、個人データは、生成ＡＩシステムの意図された目的を達成するために必要な場合であって、データ保護・プライバシーの影響評価が実行された後にのみ、学習データとして利用するべきである。

生成ＡＩシステムの開発者、提供者及び導入者は、データ最小化、より効率的な計算方法、データ増加への依存度が低いアーキテクチャなどの戦術を通じて、エネルギー消費を削減する計算ターゲットを開発することにより、環境目標をサポートすることを目指すべきである。

４．正確性

開発段階では、生成ＡＩシステムは、個人データを含む、学習、テスト及び検証のための大量のデータを利用することが多い。生成ＡＩシステムの出力の正確性は、このような大規模なデータセットの品質と代表性に大きく依存する。影響を受ける個人を差別的、違法その他の有害な結果から保護するためには、生成ＡＩシステムの開発者、提供者及び導入者が、正確で、信頼性があり、かつ代表的であるデータに依拠することが最も重要である。生成ＡＩシステムの開発者、提供者及び導入者は、虚偽の又は誤解を招く情報を除外するため、データの内容を審査し、かつフィルタリングする措置を講じる必要がある。また、開発者は、自らのシステムの正確性に関して、裏付けのない又は時期尚早な主張を行うことを控えるべきある。

代表的な高い品質のデータを利用して学習された場合でも、生成ＡＩシステムによって生成されたコンテンツには、一般的に「幻覚」として知られる、個人データを含む、不正確な又は誤った情報が含まれる可能性がある。

生成ＡＩシステムの潜在的な正確性の欠如によってもたらされるリスクを軽減するため、生成ＡＩシステムの開発者、提供者及び導入者は、適切なデータ・ガバナンス手続（例：学習データセットのソースの記録）及び技術的な保護措置（例：入出力データにおけるフィルターの利用）を実施するべきである。さらに、導入後は、生成ＡＩシステムの動作と応答を監視するとともに、正確な応答を生成するように生成ＡＩシステムの動作を微調整するために、十分な連携を確保することが不可欠である。

５．透明性

利用される学習データに関する透明性の欠如により、個人のデータ保護・プライバシーへの影響についての懸念がすでに高まっている。生成ＡＩシステムの導入者は、生成ＡＩシステムの学習のプロセスにおける個人データの収集及び利用に関する方法、時期及び理由に関する情報を含め、生成ＡＩツールの開放性を確保するための適切な透明性の措置を実施するべきである。

生成ＡＩシステムを市場投入し又は運用する提供者は、導入者に対し、当該システムの利用による潜在的なデータ保護・プライバシーのリスクと、提供者が適切なポリシーと運用を通じてこれらの課題に対処する方法を通知する必要がある。これらのリスクとポリシーは、システムの利用前と利用中の両方で、導入者にとって、明確で、理解しやすく、すぐに利用可能なものとするべきである。

生成ＡＩシステムが意思決定や意思決定の支援に利用されている場合には、開発者、提供者及び導入者は、影響を受ける当事者にこれらの運用方法を明確に伝えるべきである。

生成ＡＩシステムの開発者、提供者及び導入者は、データセットのソース、データセットの法的権限とライセンス、データセットに対するいかなる変更、フィルタリング、その他のキュレーションの実践を含め、当該システムのデータセットに関して透明性のある文書を提供するべきである。

６．セキュリティ

生成ＡＩシステムの設計、構想及び運用において、開発者、提供者及び導入者は、特にシステムが外部データソースにアクセスできる場合には、効果的なセキュリテ

ィ対策を講じる必要がある。特に、これらの対策は次のことを目的とするべきである。

- 従来のサイバーセキュリティ・コントロールを、生成ＡＩシステムの脆弱性（例：間接プロンプトインジェクション攻撃）に合わせた特定のセキュリティ・コントロールと統合すること
- 攻撃者がモデルの学習に利用したデータセットに含まれる個人データを抽出して再現できる可能性があるモデルインバージョン攻撃を防止すること
- データ保護・プライバシーの要件の遵守を促進するために講じられた安全保護措置が損なわれることはないことを確保すること

生成ＡＩシステムの開発者、提供者、導入者は、ディープ・フェイクの作成やフィッシング攻撃用のテキストの生成など、そのシステムの悪用のリスクを評価し、軽減するべきである。将来の被害を防ぐために、これらのリスクの根本原因を特定して軽減することを優先するべきである。

７．プライバシー・バイ・デザイン及びプライバシー・バイ・デフォルト

生成ＡＩシステムの能力と限界は急速に進化している。例えば、以前はテキスト、画像、音声のみを入力として受け入れていた特定の生成ＡＩシステムは、最近ではマルチモーダルになり、様々な種類の入力を受け入れるように拡張されている。こうした技術の進化の結果、新たなリスクが生じることになる。プライバシー・バイ・デザイン及びプライバシー・バイ・デフォルトの原則に沿って、生成ＡＩシステムの開発者、提供者及び導入者は、データ保護・プライバシー影響評価を実施し、ライフサイクルのあらゆる段階でこれらのシステムによってもたらされるリスクを特定し、評価し、かつ対処するべきである。

プライバシー・バイ・デザイン及びプライバシー・バイ・デフォルトは、設計段階から始まるデータ処理のライフサイクル全体を通じてデータを保護することを目的とするものである。リスク指向のアプローチに基づき、この原則を遵守することにより、ＡＩにより生じうる脅威やリスクを事前に十分に考慮することで、最小限に抑えることが可能である。

ＡＩシステムの開発者、提供者及び導入者は、想定される処理活動、データ主体にもたらす可能性のあるリスク、データ保護原則の遵守と個人の権利の保護を確保するために利用可能な措置を慎重に評価する必要がある。プライバシー指向のアプローチは、特に強力なプライバシー・バイ・デフォルトのオプションやユーザー・フレンドリーなオプションとコントロールを含め、すべての段階で優先される必要が

ある。生成ＡＩシステムの機能に対する大きな変更は、ライフサイクルの各段階当たるものであり、データ保護・プライバシー影響評価が必要になる。

８．データ主体の権利

生成ＡＩシステムの開発者、提供者及び導入者は、特に個人データが様々なソースから取得される場合には、個人に対して、自分の個人データの収集及び利用について通知を受ける権利を確実に付与するべきである。また、生成ＡＩシステムの開発者、提供者及び導入者は、法律で規定されている場合には、影響を受ける個人が次に定める権利を確実に行使できるようにするために、適切な技術的及び組織的な措置を実施しなければならない。

・自分の個人データにアクセスする権利
・不正確な個人データを修正する権利
・個人データを消去する権利、及び
・個人に対して重大な影響を及ぼす自動化された決定に従わない権利

影響を受ける個人が権利を行使する能力は、生成ＡＩシステムが特別な種類のデータ又は未成年者の個人データを処理する場合には、特に重要である。

９．説明責任

生成ＡＩシステムの開発者、提供者及び導入者は、適用される国内法及び国際協定を遵守する責任があり、その遵守を証明しなければならない。説明責任の原則により､生成ＡＩモデルのサプライチェーンにおいて関与する様々な主体の間において、責任を明確に特定し、尊重することが必要である。

このような遵守の証明は、データ保護・プライバシー機関が生成ＡＩツールのデータ保護・プライバシー要件の遵守を評価できるようにするために、システムのライフサイクル全体における技術的な文書を利用可能にすることにより、行うべきである。生成ＡＩシステムの開発者、提供者及び導入者は、この文書の中に、モデルがどのように機能するか、モデルの学習のために利用されたデータは何か、そして、サービスを市場投入し又は運用する前における潜在的なデータ保護・プライバシーへの影響を含めるべきである。

生成ＡＩシステムの開発者、提供者及び導入者は、モデルがどのように機能するかを評価し、出力の不正確さやバイアスをテストし、潜在的なリスクを軽減するための効果的な対策を推奨することを独立の立場で行うことのできる外部監査を受け入れるべきである。また、説明責任を果たすには、生成ＡＩシステムの開発者、提供

者及び導入者が、健全なデータ・ガバナンスの手続やツールを実施することも必要である。

以上を踏まえ、第45回世界プライバシー会議は、以下のとおり決議する。

・この決議で定められた適用される原則と権利を含む、生成ＡＩ技術に関連したデータ保護・プライバシー法の適用と執行を確実に行うことにコミットすること、

・倫理的、法律的、社会的及び技術的な観点から、生成ＡＩに関連した個人データの保護・プライバシーを確保するために協力することにコミットすること、

・人工知能における倫理とデータ保護に関するワーキンググループ内において、生成ＡＩシステムのデータ保護・プライバシーのリスクに関する各管轄区域内での進行中の取組の展開状況を共有することにコミットすること、

・生成ＡＩシステムの開発者、提供者及び導入者に対し、データ保護・プライバシーを基本的人権として認識し、データ保護・プライバシー、人間の尊厳その他の基本的な権利と自由を保護する、責任ある、そして信頼性のある生成ＡＩ技術を構築するよう要請すること、

・生成ＡＩシステムの開発者、提供者及び導入者に対し、データ保護・プライバシー及びデータ主体の権利に関して、生成ＡＩシステムの開発と導入を理解するための研修を従業員や担当者に提供するよう奨励すること、

・ＧＰＡメンバーに対し、データ保護・プライバシーその他の人権に対するリスク及び生成ＡＩシステムに関連したデータ保護・プライバシーに適用される法的義務及び原則に関する啓発活動を奨励すること、

・生成ＡＩに関連して生じる新たなリスク及び基本的な権利と自由に対する潜在的な危害を継続的に監視すること、

・現在及び今後の立法上及び規制上のイニシアチブやアプローチについて、提唱し、かつ助言することを目指すこと、

・ＧＰＡメンバーに対し、生成ＡＩシステムに対する各自の執行取組を調整するよう要請すること、

・第46回世界プライバシー会議で、生成ＡＩシステムに関してＧＰＡ　ＡＩＷＧメンバーが実施した作業に関する中間報告書を提出することを検討すること、及び、さらに第47回世界プライバシー会議で追加の政策文書又は決議案を提出することを検討すること。