基礎研レポート

EU の AI 規則(3/4)

―適合性審査、汎用モデル

保険研究部 専務取締役 研究理事　松澤　登
(03)3512-1866　matuzawa@nli-research.co.jp

1――はじめに

EU の AI 規則（以下、本規則）は 2024 年 6 月 13 日に EU ジャーナル（日本の官報に相当）に掲載され、同年 8 月 1 日に発効した。本規則は AI システムにさまざまな規制を行っており、その概要は図表 1 の通りである。

前回までのレポートでは、総則・定義、禁止される AI の行為、高リスク AI システムの定義・規律、高リスク AI システムの提供者、配備者等の義務等について述べた。

今回のレポートでは、主に EU 基準への適合性の審査および透明義務・汎用 AI モデルについて述べる。図表 1 に色付きで示した項目があるが、ここが今回のレポートで解説する部分である。

【図表 1】今回解説する事項

総論・定義	適用範囲や定義
禁止されるAIの行為	許容できないリスクをもつAI
高リスクAIシステム	許容される高リスクAIシステムの果たすべき要件
適合性審査	EUの基準に合致しているか審査
透明義務・汎用AIモデル	汎用AIモデルのシステミック・リスクの防止
イノベーション支援	AIの革新を推進
EU及び域内国のガバナンス	AIシステムに係るEUレベルのガバナンス
市販後モニタリング	市場投入後の監視・是正
法律の執行・罰則	行動規範の作成・違反行為に対する罰則

2――前回までのレポートの振り返り

本稿を読んでいただくにあたっては、前回、前々回のレポートをまず読んでいただくことが望ましいが、ここで前回、前々回レポートを簡単に振り返っておく。

まず、AI システムの定義であるが、これは推論能力が鍵となる。簡単に言えば、本規則では、自律的に稼働する機械のシステムであって入力から出力を推論するものを AI システムと定義している。

これはすなわち人間の頭脳のような働きをするシステムと言い換えられる。本規則の適用対象は主に、提供者と配備者であって、提供者はAIシステムを開発の上（第三者に開発させる場合を含む）、EU域内市場に投入または稼働させる者をいう。配備者は自分でAIシステムを利用する者をいうが、事業として利用する者に限定され、私的に利用する場合を含まない。

本規則のルールの一つ目は「禁止されるAIの行為」である。本規則5条に規定されたAIシステムはEU域内市場への投入や稼働が禁止されている。いくつかあるが、たとえばサブリミナル技術を利用するなどして人間の行動を歪めさせるシステム、こどもや障がい者等を搾取するシステム、ソーシャリングスコアシステムなどが禁止されている。このうち、リアルタイム遠隔生体認証システム（街頭に設置された監視カメラでリアルタイムに個人を識別するシステム）を捜査に使用することは、司法機関の許可を得ることなど厳格な条件のもとでAIシステムの利用が可能である。

本規則のルールの二つ目は高リスク AI システムに関する規律である。人権や安全性に重大な影響を及ぼす可能性のあるAIシステムは、リスク管理措置の実施、データガバナンスの履行、技術文書の作成・ログの保管、配備者に対する使用説明書の作成、人的監視措置を可能とするシステムの構成、正確性および堅牢性をシステムに備えさせるなどの要件が課せられる。

そして高リスクAIシステムの提供者・配備者・輸入業者・販売業者はそれぞれ本規則の定める所定の義務が課せられる。

3――通知当局と被通知団体

本項3と次項の4は適合性審査に関する解説となる（図表2）。

【図表2】本項3と4で取り扱う部分

1｜通知当局と被通知団体とは

本項では通知当局と被通知団体について規定している。これらの定義は以下の通りである。

通知当局(notifying authority)とは、適合性評価機関の審査及び通知（＝認定）、並びにその監視に必要な手続の設定及び実施に責任を負う国家当局を意味する(3条(19))。

被通知団体(notified bodies)とは、本規則及び他の関連するEU調和法[1]に従って通知された適合性

[1] EUで製品を流通させるにあたり、特定分野の製品が満たすべき要件を調和（ハーモナイゼーション）させることとしているEU指令や規則（Union harmonization legislation）のことを指す（EU整合化法令と訳されることもある）。

評価機関を意味する(3 条(22))。適合性評価とは高リスク AI システムに関するⅢ章 2 節（高リスク AI システムの要件、前回レポートで解説）に規定された要求事項が満たされているかどうかを実証するプロセスを意味し（3 条(20)）、適合性評価機関とは、試験、認証及び検査を含む第三者適合性評価活動を行う機関をいう(3 条(21))。

（注記）通知当局とは国家の行政機関で被通知団体を監視する。被通知団体とは後述の通り、通知当局が適合性審査を行うことを認める通知（＝認可）を受けた団体である。通知当局、被通知団体というのはわかりにくいが、適合性審査機関を認可する当局、適合性審査実施の認可を受けた団体と言い換えるとわかりやすいかもしれない（以下、本稿では原文に従って、通知当局、被通知団体という)。

また、被通知団体の行う適合性評価とは前文によれば「高リスク AI システムの複雑性とそれに伴うリスクを考慮すると、被通知団体が関与する高リスク AI システムに対する適切な適合性評価手順、いわゆる第三者適合性評価（third party conformity assessment）を開発することが重要である」（前文 125）とされている。すなわち、被通知団体は複雑な高リスク AI システムについて、重大リスクの発生抑止機能などを確認する第三者機関であり、高リスク AI システムの市場投入に際しては、提供者と並んで最も重要な機能を果たしている。

2｜被通知団体となるための手続

(1)各加盟国は、適合性評価機関の評価、指定及び通知並びにその監視のために必要な手続の設定及び実施に責任を負う、少なくとも一つの通知当局を指定又は設置しなければならない。これら手続きは、すべての加盟国の通知当局が協力して策定しなければならない(28 条 1 項)。

(2)適合性評価機関（被通知団体となろうとする機関）は、その機関が設立されている加盟国の通知当局に通知申請書(application for notification)を提出しなければならない(29 条 1 項)。

(3)提出を受けた通知当局は、第 31 条に規定する要件を満たした適合性評価機関（被通知団体）のみを通知することができる(30 条 1 項)。

(4)被通知団体の要件について、詳細に定められている。要約すると、独立性、中立性、信頼性、専門性を有するものでなくてはならないとしている(31 条)。

(5) 被通知団体は、第 43 条に定める適合性評価手順に従って、高リスク AI システムの適合性を検証しなければならない（34 条)。

（注記）これら条文は被通知団体の要件と、適合性評価手順を遵守すべき旨を定めている。具体的な適合性評価については後述する。なお、32 条・33 条、35 条～39 条は技術的なことを定めた条文であり、省略する。

4――規格、適合性評価、認証、登録

1｜欧州標準と共通仕様

(1)欧州標準が存在する場合：欧州ジャーナルに掲載された欧州標準(harmonized standards)であって、Ⅲ章 2 節（高リスク AI システムの要件。前回レポートの 4)、Ⅴ章 2 節・3 節（汎用 AI モデル提供者の義務等、本レポートの 6、7）を満たす欧州標準が存在する場合において、その欧州標準に適合している高リスク AI システムと汎用 AI モデルは、これらの規定(Ⅲ章 2 節など)が要求する要件に

適合すると推定される(40 条 1 項)。

(2)欧州標準が存在しない場合：欧州委員会はⅢ章 2 節およびⅤ章 2 節・3 節を満たすための共通仕様(common specification)を定める実施法を採択することができる(41 条 1 項)。この場合において、共通仕様に適合する高リスク AI システムと汎用 AI モデルは、本規則の要求する要件に適合すると推定される（同条 3 項)。

（注記）前文では「欧州標準化を行うことにより、単一市場における競争力と成長だけでなく、技術革新を促進するために、技術水準に沿った本規則への準拠を確保するための技術的解決策を提供者に提供する重要な役割を果たすべきである」とし、欧州標準が存在しない場合「欧州委員会は、実施法を通じ、アドバイザリー・フォーラム(67 条、次回レポートで解説)の協議を経て、本規則に基づく特定の要求事項に関する共通仕様を定めることができるものとする」(前文 121）とある。高リスク AI システムおよび汎用 AI システムは、それぞれ独自機能を有しながらも、技術的に共通化が可能な骨格部分では、共通の仕様を有することが目指されている。適度な標準化は公正な競争を促進し、かつ AI システムの発展を促すものと考えられる。

2｜適合性評価

(1)付属書Ⅲ（前回レポート図表 5 に掲載）の 1.（生体に関するもの）[2]に列挙された高リスク AI システムについて、Ⅲ章 2 節（高リスク AI システムの要件）に規定された高リスク AI システムの適合性を実証するにあたっては、提供者が第 40 条に言及された欧州標準、又は第 41 条に言及された共通仕様を適用している場合、提供者は、付属書Ⅵ(下記図表 3)または付属書Ⅶ[3]に規定される適合性評価手続を実施しなければならない（43 条 1 項)。

(2)付属書Ⅲの 2.～8.[4]に該当する高リスク AI システムについては、提供者は付属書Ⅵ（下記図表 3）に従って適合性評価手続きを実施しなければならない（同条 2 項)。

【図表 3】付属書Ⅵ

①内部統制に基づき、②～④に従って適合性確認手順の実施
②提供者による 17 条（品質管理システム）に従って品質管理システムが確保されていることの確認
③Ⅲ章 2 節（高リスク AI システムの要件）に従い、技術文書が本規則を遵守していることの審査
④設計、開発過程、市販後モニタリングシステム(72 条。次回レポート)が技術文書と整合性していることの保証

(3)付属書Ⅰのセクション A（前回レポート図表 3）に列挙された EU の調和法が適用される高リスク AI システムについては、その提供者は、それらの法令で要求される適合性評価手順に従わなければな

[2] 該当するのは a)遠隔生体認証システム（本人確認のためのものを含まない)、b)生体のカテゴライズのために用いられる AI システム、c)感情認識を目的とした AI システムである。

[3] 付属書 VII は品質管理システムや技術文書の評価に基づく適合性を審査する（詳細は省略)。

[4] 該当する AI システムとしては、重要なインフラにかかわるもの、教育と職業訓練にかかわるもの、雇用・労働者管理・自営業の自己評価にかかわるもの、必要不可欠な民間サービス・公共サービスおよび給付を受けるためのアクセスにかかわるもの、法の執行にかかわるもの、移民・難民保護・国境管理にかかわるもの、司法行政と民主的プロセスにかかわるものである。

らない（同条3項）。

（注記）前文では「高リスクAIシステムの複雑性とそれに伴うリスクを考慮すると、被通知団体が関与する高リスクAIシステムに対する適切な適合性評価手順、いわゆる第三者適合性評価（third party conformity assessment）を開発することが重要である。しかし、製品安全分野における専門の市販前認証機関の現在の経験が、審査するリスクの内容をカバーできていないことを考慮すると、少なくとも本規則の適用初期段階においては、製品に関連するもの以外の高リスク AI システムに対する第三者適合性評価の適用範囲を制限することが適切である。したがって、このようなシステムの適合性評価は、生体に使用されることを意図したAIシステムを唯一の例外として、原則として、提供者が自らの責任において適合性評価を実施すべきである」（前文125）とする。適合性評価は、適合性評価機関に経験がないことを踏まえ、少なくとも規則制定後早期の段階では一部例外を除き、提供者自身が行うこととされている。したがって、本規則制定後しばらくは、被通知団体（適合性評価機関）の審査範囲は限定されている。

3｜EU適合宣言書

提供者は、高リスクAIシステムごとに、機械可読、かつ物理的または電子的に署名されたEU適合宣言書を作成し、高リスクAIシステムが市場投入または使用開始された後10年間、各国の所轄当局が自由に利用できるよう保管しなければならない。EU適合宣言書では、作成された高リスク AI システムを特定しなければならない。EU適合宣言書の写しは、要請に応じて関連する国家所管庁に提出しなければならない（47条1項）。

EU適合宣言書は本規則Ⅲ章2節（高リスク AI システムの要件）に定められた要件に合致していることを記載するものとする（同条2項）。EU宣言書を作成した提供者はⅢ章2節を遵守していることについて責任を負う（同条4項）。

（注記）EU適合宣言書は、Ⅲ章2節（高リスク AI システムの要件）を満たすことを提供者が表明する書類であるが、前文にEU適合宣言書の位置づけ等についての特段の説明はない。提供者は、高リスクAIシステムを市場投入し、サービスを開始する前に被通知団体の適合性審査を受け、これに合格したことを自ら表明することとされている。

4｜CEマーキング

CEマーキングは規則（EC）765/2008（製品がEU法令の要求する条件に適合していることを示すマークについて定めた規則）の要件に従って添付されなければならない（48条1項）。CEマーキングは、高リスクの AI システムについては、見やすく、読みやすく、消えないように貼付しなければならない。高リスクのAIシステムの性質上、それが不可能な場合又は困難な場合は、適宜、包装又は添付文書に貼付しなければならない（同条3項）。

（注記）CEマーキングとは、提供者が、AIシステムがⅢ章2節（高リスク AI システムの要件）に規定された要求事項、及びその他のAIシステムを規制するEU調和法であって、その貼付を規定するものに適合していることを示すマーキングをいう（3条(24)）。前文では「高リスクAIシステムは、域内市場内で自由に移動できるよう、本規則への適合を示すCEマーキングを付すべきである。製品に組み

込まれた高リスク AI システムについては、物理的な CE マーキングを付すべきであり、デジタル CE マーキングで補完することができる。デジタルでのみ提供される高リスクの AI システムについては、デジタル CE マーキングを使用すべきである」（前文 129）とする。CE マーキングは EU 規格を遵守していることを示すことにより、高リスク AI システムの域内での流通を円滑なものにすることを目的とする。

5｜登録

(1)付属書Ⅲ（前回レポートの図表 5）のポイント 2 で言及される高リスク AI システム（重要インフラ）[5]を除き、提供者、または該当する場合には認定代理人は、71 条で言及される EU のデータベース（次回レポートで解説）に自身とそのシステムの情報を登録しなければならない(49 条 1 項)。

(2)提供者が第 6 条 3 項に従って高リスクでないと結論付けた AI システムを市場投入する前、または使用開始する前に、提供者または該当する場合は認可された代理人は、71 条で言及される EU データベースに自身および自身のシステムの情報を登録しなければならない（同条 2 項）。

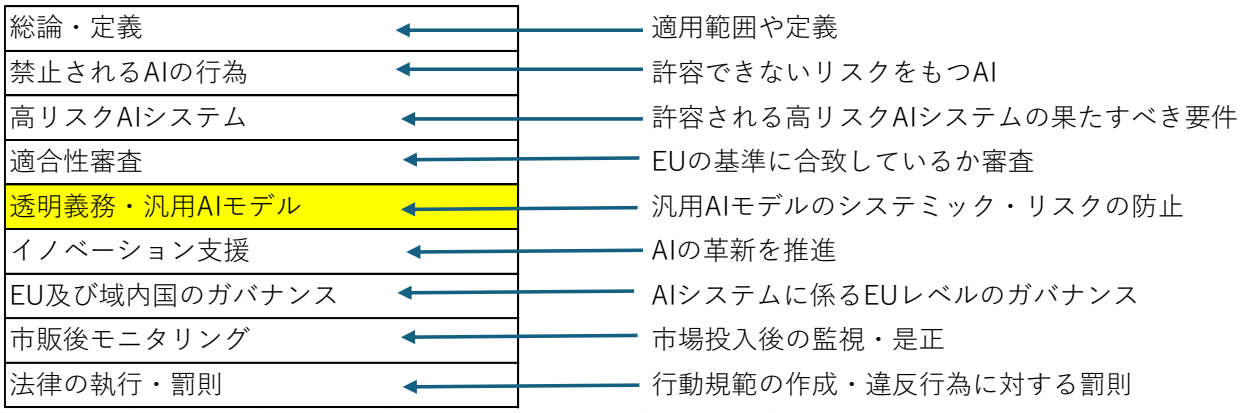
(3)付属書Ⅲのポイント 2（重要インフラ）で言及されている高リスク AI システムは、国家レベルで登録されなければならない（同条 5 項）。

（注記）前文では「AI 分野における欧州委員会および加盟国の作業を容易にし、国民に対する透明性を高めるため、関連する既存の EU 調和法の適用範囲に含まれる製品に関連するもの以外の高リスク AI システムの提供者、および、（中略）適用除外に基づき高リスクではないと考える提供者は、欧州委員会が設置・管理する EU のデータベースに、自らとその AI システムに関する情報を登録することを義務付けられるべきである」（前文 133）とある。AI システムを本規則によって構築されるデータベースに登録を行うのは、監視当局の作業の容易化および国民に対する透明性の向上のためである。

5――特定の AI システムの提供者と配備者に対する透明義務

本項 5 および 6、7 は以下の図表 4 に係る部分を解説するものである。

【図表 4】5、6、7 の該当部分（色塗りの部分）



[5] 該当するのは、重要なデジタルインフラ、道路交通、水道、ガス、暖房、電気の供給の管理運営にかかわる安全部品としての AI システムである。

(1)提供者は、自然人と直接対話することを意図したAIシステムが、状況及び使用の文脈を考慮し、合理的に十分な知識を持ち、観察力があり、思慮深い自然人から見て（AIシステムであることが）明らかな場合を除き、当該自然人がAIシステムと対話することを知らされるような方法で設計及び開発されることを確保しなければならない。この義務は、第三者の権利及び自由に対する適切な保護措置の下に、犯罪を検知、防止、捜査又は起訴することを法律で認められたAIシステムには適用されない(50条1項)。

（注記）前文では「自然人との対話またはコンテンツの生成を意図した特定のAIシステムは、それが高リスクに該当するか否かにかかわらず、なりすましや欺瞞の特定のリスクをもたらす可能性がある。したがって、(中略）特定の透明性確保義務の対象とすべきであ」るとする（前文132)。この条文は、たとえばチャットボット[6]などで人間ではなくAIシステムと対話している場合に、そのことが対話者にわかるようにしなければならないとするものである。自然人でなくAIシステムと会話していることを明確にすべきことは詐欺などの被害防止が念頭にある。なお、犯罪の検知などにはこの規定は適用されない。

(2)合成音声、画像、映像又はテキスト・コンテンツを生成する汎用AIシステムを含むAIシステムの提供者は、AIシステムの出力が機械可読形式で表示され、人為的に生成又は操作されたものであることが検知可能であることを保証しなければならない。提供者は、様々な種類のコンテンツの特殊性及び制限、実装のコスト、並びに関連する技術標準に反映され得る一般的に認められた技術状況を考慮し、技術的に実行可能である限りにおいて、その技術的検知策が効果的であり、相互運用可能であり、堅牢であり、信頼できるものであることを確保しなければならない（同条2項)。

（注記）前文では「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区別することがますます難しくなっている。このような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、誤った情報や大規模な操作、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」したがって「こうしたシステムの提供者に対し、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」（前文133）とする。AIシステムの出力が、人が録音・録画した自然な音声・動画等ではないことが明らかになるようにしなければならない。本項は提供者に対する義務であり、次項では配備者の義務として同様に規定されている。これら規定にかかわる深刻な問題として、ディープフェイクがある。この点については次項参照。

(3)ディープフェイクを構成する画像、音声または映像コンテンツを生成または操作するAIシステムの配備者は、当該コンテンツが人為的に生成または操作されたものであることを開示しなければならない。この義務は、犯罪の検出、防止、捜査または訴追のために法律で使用が許可されている場合に

[6] チャットボットとは、人工知能を活用した「自動会話プログラム」のことである。金融機関などにネット上で質問等をした場合、まずチャットボットが対応する会社も多い。

は適用されない。コンテンツが、明らかに芸術的、創作的、風刺的、フィクション的または類似の作品または番組の一部を構成する場合、本項に定める透明性の義務は、当該作品の表示または享受を妨げない適切な方法で、当該生成または操作されたコンテンツの存在を開示することに限定される（同条 4 項）。

（注記）いわゆるディープフェイク[7]に関する規制である。本項は合成された映像や音声が AI システムによって合成されたものについて透明性を求めるが、前文では「本規則に定めるディープフェイクの透明性義務は、作品の有用性と質を維持しつつ、通常の利用や使用を含め、作品の展示や享受を妨げない適切な方法で、そのような生成または操作されたコンテンツの存在を開示することに限定される」（前文 134）とされ、不当な目的に利用されることを規制することを目的とするが、芸術や表現の自由を阻害するものではないとされている。

6――汎用 AI モデル

1｜汎用 AI モデル等に係る定義

(1) まず汎用 AI モデル[8]の定義を述べる。それは、AI モデル（そのような AI モデルが大規模な自己監視を使用して大量のデータで学習する場合を含む）であって、有意な汎用性を示し、モデルが市場に投入される方法に関係なく、広範で明確なタスクを適切に実行することができ、様々な下流のシステムまたはアプリケーションに統合することができるものを意味する（3 条(63)）。

そして汎用 AI システムとは、汎用 AI モデルをベースとした AI システムであり、当該 AI システムを直接使用する場合、または他の AI システムと連動して利用する場合において、様々な目的に対応する能力を有する AI システムを意味する（3 条(66)）。

（注記）具体的にどのような AI モデルが汎用 AI モデルに該当するかどうかは本規則において重要なポイントである。この点、前文では「汎用 AI モデルの概念は、法的確実性を確保するために、AI システムの概念とは別に明確に定義されるべきである」とし、その定義は、「汎用 AI モデルの主要な機能特性、特に汎用性と幅広い明確なタスクを適切に実行する能力に基づくべきである。これらのモデルは、通常、自己監視のもとでの、学習あり、監視なし学習あり、強化学習ありなどの様々な手法により、大量のデータで学習される」（前文 97）とされる。そして「大規模な生成 AI モデルは、汎用 AI モデルの典型的な例であり、テキスト、音声、画像、映像などのコンテンツを柔軟に生成できるため、さまざまな特徴的なタスクに容易に対応できる」（前文 99）とされており、生成 AI モデルは近い将来汎用 AI モデルに該当することとなる（あるいは既に該当する）可能性が高い。また、汎用 AI モデルに該当するかどうかの判断基準としてパラメータ（変数）の数が重要とされ、前文では「モデルの汎用性は、特にパラメータ（変数）の数によって決定されることもあるが、少なくとも 10 億のパラメータを持ち、自己監視を使用して大量のデータで学習を行ったモデルは、重要な汎用性を示し、広範囲

---

[7] ディープフェイク（deepfake）は本来、機械学習アルゴリズムの一つである深層学習（ディープラーニング）を使用して、2 つの画像や動画の一部を結合させ元とは異なる動画を作成する技術である。現在、世間で言われているディープフェイクはフェイク動画、偽動画を指すことが多い。現実の映像や音声、画像の一部を加工して偽の情報を組み込み、あたかも本物のように見せかけて相手をだます方法として認識されつつある(NEC ソリューションイノベータの HP より)。

[8] AI モデルという意味であるが、AI システムの中核となる機能であり、入力から出力までの作業を行うシステムに組み込むことで、AI システムとして稼働するものを指す。

の特徴的なタスクを適切に実行すると考えられるべきである」（前文 98）とする。本規則において汎用 AI モデルの説明は上記に述べたところがすべてであるが、あえて簡単に言い換えれば、人間と同等の広範な判断能力を持ち、さまざまなタスク（仕事）を実施可能なものが汎用 AI モデルである。

(2)本規則において、システミック・リスクとは、汎用 AI モデルの保有する、他に影響を及ぼす大きな能力(capability)に特有のものであり、その影響力の大きさにより EU 市場に重大な影響を及ぼすリスク、または公衆衛生、安全、治安、基本的権利、社会全体に対する実際の悪影響も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大規模に伝播するリスクを意味する(3 条(65))。（注記）システミック・リスクについて、前文では「汎用 AI モデルは、①重大事故、重要部門の混乱及び公衆衛生と安全への重大な影響に関連する、実際の又は合理的に予見可能な悪影響、②民主主義のプロセス、公共及び経済の安全に関する実際の又は合理的に予見可能な悪影響、③違法、虚偽又は差別的なコンテンツの流布を含むが、これらに限定されないシステミック・リスクをもたらす可能性がある」（前文 110）とする（①②③の数字は筆者挿入）。すなわち、重大な事故(たとえば自動運転の致命的欠陥)、民主主義や公共の安全に関する悪影響（たとえば国政選挙におけるディープフェイク動画の拡散）などが、大規模に発生あるいは拡散されることがシステミック・リスクと考えられる。

2｜システミック･リスクを有する汎用 AI モデル

(1)汎用 AI モデルは、以下のいずれかの条件を満たす場合、システミック・リスクを有する汎用 AI モデルに分類される(51 条 1 項)。

(a)指標やベンチマークを含む適切な技術的ツールや方法論に基づいて評価された、高い影響力を持つもの、あるいは

(b)委員会の決定に基づき、職権で、または科学的パネルからの適格な警告に基づき、附属書ⅩⅢ（図表 5。モデルのパラメータ数、データセットの質又はサイズなど 7 つの基準）に定める基準に照らして、(a)に定める能力または影響に相当する能力を有すること。

【図表 5】付属書 XIIIの項目

モデルのパラメータ数
データセットの質またはサイズ
モデルの訓練に使用したもの―コスト、時間、エネルギー消費ーの総量
モデルの構築に要した入力と出力
モデルの能力のベンチマークと評価
モデルが到達できるインターネットの範囲により高い影響を及ぼせるかどうか
登録ユーザーの数

（注記）前文では「システミック・リスクは特に影響力の高い能力から生じることから、汎用 AI モデルであって、適切な技術的ツールや方法論に基づいて評価された高い影響力を有する場合、または、その影響力の大きさにより内部市場に重大な影響を及ぼす場合、その汎用 AI モデルはシステミック・リスクを呈すると考えるべきである。汎用 AI モデルにおける影響力の高い能力とは、最先端の汎用 AI モデルに記録されている能力と同等か、それを上回る能力を意味する」（前文 111）とし、また「汎

用 AI モデルのうち、高い影響能力の適用閾値を満たすものは、システミック・リスクを有する汎用 AI モデルであると推定すべきである。」（前文 112）とする。なお、現時点では図表 5 の項目が定められているが、閾値は未定な模様である。

(2)汎用 AI モデルが第 51 条 1 項(a)（上記(1)）の条件を満たす場合、その提供者は、当該条件が満たされた後、または満たされることが判明した後、遅滞なく、いかなる場合においても 2 週間以内に欧州委員会に通知しなければならない。その通知には、当該条件が満たされたことを証明するために必要な情報を含めるものとする。欧州委員会が、通知を受けていない汎用 AI モデルがシステミック・リスクを有していることを知った場合、欧州委員会はそのモデルを、システミック・リスクを有するモデルとして指定することを決定できる(52 条 1 項)。

（注記）システミック・リスクを有する汎用 AI モデルは欧州委員会への通知をかならず必要とする。前文では「潜在的な重大な悪影響を考慮すると、システミック・リスクを伴う汎用 AI モデルは、常に本規則に基づく関連義務の対象とすべきである」（前文 97）と強調されている。また、前文では「汎用 AI モデルの学習には、計算資源の先行割り当てを含むかなりの計画が必要であるため、汎用 AI モデルの提供者は、学習が完了する前に、そのモデルが閾値を満たすかどうかを知ることができる」（前文 112）ため、提供者は AI オフィス（64 条。次回レポートで解説予定）に通知しなければならないとする。また、通知を待たずとも、条件を満たした汎用 AI モデルを、システミック・リスクを有する汎用 AI モデルとして指定する権利を欧州委員会が有しているのは条文の通りである。

3｜汎用 AI モデル提供者の義務

汎用 AI モデルの提供者は、以下のことを行わなければならない(53 条 1 項)。

(a)要請に応じて AI オフィス及び各国所轄官庁に提供する目的で、最低限、付属書 XI（汎用モデルの一般的な説明と開発プロセスの関連情報。ここでは省略）[9]に定める情報を含む技術文書を作成、アップデートしなければならない。

(b)汎用 AI モデルをその AI システムに組み込むことを意図する AI システムの提供者に対して、汎用 AI モデルの最新の情報および文書を利用可能にすること。そして、EU 法および国内法に従い、知的財産権および業務上の機密情報または企業秘密を遵守し保護する必要性を損なうことなく、情報および文書について以下を満たさなければならない（図表 6)。

【図表 6】汎用 AI モデルを AI システムに組み込む場合に提供する情報

(i)AI システムの提供者が、汎用 AI モデルの能力と限界を十分に理解し、本規則に基づく義務を遵守できるようにすること
(ii)情報および文書には、最低限、附属書 XⅡ（汎用モデルの一般的な説明およびモデルの要素とその開発プロセス。ここでは省略）[10]に定める要素が含まれていること

[9] 概略（一部）だけ述べると①AI モデルの規模とリスクプロファイルに適合した技術文書、②AI モデルの四要素に関する詳細な説明、③評価戦略の詳細な説明、④該当する場合、外部及び内部からの敵性攻撃テストを実施するために導入した手段の詳細な説明が含まれる。

[10] 概略（一部）だけ述べると、①AI モデルが想定したタスクと統合できる AI システムの種類と性質を含む汎用 AI モデル

(c)著作権および関連する権利に関する EU 法を遵守し、特に、指令(EU) 2019/790 の第 4 条(3)[11]に従って表明された権利の留保を、最先端技術を通じて特定し、遵守するための方針を導入する。
(d)AI オフィスが提供する様式に従って、汎用 AI モデルの学習に使用したコンテンツ（著作物等）に関する十分に詳細な要約を作成し、一般に公開する。
（注記）本項は知的財産権、特に著作権との関係を規定する。AI モデルの開発と学習には著作物を利用する必要があるが、一般的には著作物の利用には著作権者の同意が必要である[12]。前文では「著作権で保護されたコンテンツの使用には、関連する著作権の例外や制限が適用されない限り、当該権利者の承認が必要である」（前文 105）としている。

しかしながら、EU においてはテキストマイニング[13]およびデータマイニングについては、権利者の権利留保（オプトアウト）がない限り、上記指令 4 条(3)によって、著作物の利用が可能となっている。この手法を利用することで、指令 4 条(3)により、著作物を著作権者の権利留保がない限りにおいて、汎用 AI モデルの学習データとして取り込むことができることとなる。

日本では、「当該著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合」を除き、AI システム学習用の著作物利用が認められており、著作者に権利留保の権利は認められていない点が相違する。

4｜汎用 AI モデル提供者の認定代理人

(1)汎用 AI モデルを域内市場に投入する前に、第三国で設立された提供者は、書面による委任により、域内に設立された認定代理人を任命しなければならない(54 条 1 項)。
(2)提供者は、その権限を有する代理人が、プロバイダーから受領した委任状で指定された業務を遂行できるようにしなければならない(54 条 2 項)。
（注記）汎用 AI モデル提供者にはさまざまな義務が課されるため、域外に所在する汎用 AI モデル提供者は法的な代理権限を有する EU 域内に所在する代理人を任命しなければならない（前回レポート P14 も参照)。

7――システミック・リスクを伴う汎用 AI モデルの提供者の義務

1｜システミック・リスクを伴う汎用 AI モデル提供者の義務の内容

第 53 条および第 54 条に記載された義務に加えて、システミック・リスクを有する汎用 AI モデルの提供者は、以下の義務を負う(55 条 1 項。図表 7)。

の一般的な説明、②汎用 AI モデルを AI システムに統合するための技術的手段を含むモデルの要素と開発のプロセスの説明等である。
11 本指令の 4 条(3)は、テキスト・データ・マイニングのための著作物の複製・抽出は、著作者が適切な方法で明確に権利を留保していなければ、著作権の例外又は制限の対象とするという規定である。
12 日本では訓練用のデータセットに著作物を利用することは原則として可能になっている（著作権法 30 条の 4)。
13 テキストマイニングとは、定型化されていない文章の集合からなるテキストデータをフレーズや単語に分解して詳細に解析し、有用な情報を抽出する分析手法を指す。

【図表 7】システミック・リスクを有する汎用 AI モデルの提供者の義務

(a) システミック・リスクの特定と軽減を目的としたモデルの敵対的テストの実施と文書化を含め、最新技術を反映した標準化されたプロトコルとツールに従ってモデル評価を実施すること
(b) システミック・リスクを有する汎用 AI モデルの開発、市場投入、使用から生じる可能性のあるシステミック・リスクを、その発生源を含め、EU レベルで評価し、軽減すること
(c) 重大インシデントおよびそれに対処するための可能な是正措置に関する関連情報を把握し、文書化し、AI オフィスおよび必要に応じて各国の所轄官庁に不当な遅延なく報告すること
(d) システミック・リスクを伴う汎用 AI モデルと、その物理的構造に対して、適切なレベルのサイバーセキュリティ保護を確保すること

ここで重大インシデントとは、直接的または間接的に以下のいずれかにつながる AI システムの事故または誤作動を意味する(3 条(49)。図表 8)。

【図表 8】重大インシデント

(a) 人の死亡、または人の健康への重大な危害が生ずること
(b) 重要インフラの管理・運営に深刻かつ不可逆的な混乱が生じること
(c) 基本的権利の保護を目的とする EU 法の義務に違反すること
(d) 財産や環境に重大な損害を与えること

(注記) 本項はシステミック・リスクを有する AI モデルの提供者の義務を列挙する。主なものとして、市場投入前の敵対的テスト[14]の実施と重大インシデントの報告である。前文では、敵対的テストの実施理由は「システミック・リスクを提示する汎用 AI モデルの提供者は、汎用 AI モデルの提供者に規定される義務に加えて、単体モデルとして提供されるか、AI システムや製品に組み込まれて提供されるかにかかわらず、これらのリスクを特定・軽減し、適切なレベルのサイバーセキュリティ保護を確保することを目的とした義務を負うべき」(前文 114) であるためである。また、重大インシデントの報告については「システミック・リスクの可能性がある汎用 AI モデルに関連するリスクを特定し、防止するための努力にもかかわらず、当該モデルの開発または使用が重大なインシデントを引き起こした場合、汎用 AI モデルの提供者は、過度の遅滞なくインシデントを追跡し、関連する情報および可能な是正措置を欧州委員会および各国の所轄当局に報告すべきである」(前文 115) とされている。

2｜システミック・リスクを伴う汎用 AI モデル提供者の義務遵守

システミック・リスクを伴う汎用 AI モデルのプロバイダーは、整合的な基準が公表されるまでは、55 条 1 項 (上記 1 ｜参照) に定める義務の遵守を証明するために、第 56 条にいう意味での実践規範に依拠することができる。欧州の実践規範への準拠は、当該基準が本条 1 項の義務をカバーしている限りにおいて、提供者に適合の推定を与える(55 条 2 項)。

(注記) 次項参照。

[14] AI システムに対して内外からのサイバー攻撃等を行うテストを指す。

3｜実践規範

AI オフィスは、国際的なアプローチを考慮しつつ、本規則の適切な適用に寄与するため、EU レベルでの実践規範の作成を奨励し、促進するものとする(56 条 1 項)。

AI オフィスおよび理事会は、以下の事項を含め、少なくとも第 53 条および第 55 条に規定される義務を実施規範がカバーすることを確保することを目指すものとする（同条 2 項。図表 9)。

【図表 9】実践規範の内容

(a)第 53 条第 1 項の(a)および(b)に言及される情報（AI オフィスまたは AI システム提供者に提供する情報）が、市場および技術の発展に照らして常に最新であることを保証する手段
(b)学習に使用された内容に関する要約の適切な詳細レベル
(c)適切な場合には、その発生源を含め、EU レベルでのシステミック・リスクの種類と性質の特定
(d)EU レベルでのシステミック・リスクの評価及び管理のための措置、手続及び方法（その文書化を含む）

（注記）AI オフィスとは、2024 年 1 月 24 日の欧州委員会決定で規定された、AI システムおよび汎用 AI モデルの導入、監視および監督、ならびに AI ガバナンスに貢献する欧州委員会の機能を意味する(3 条(47))。なお、詳細は次回レポートで解説予定。

本条でいう実践規範は AI オフィスが作成する。前文では「AI オフィスは、国際的なアプローチを考慮した実践規範の作成、見直し、適応を奨励・促進すべきである。AI オフィスは、実践規範が最先端の状況を反映し、多様な視点を適切に考慮することを確実にするため、関連する各国所轄庁と協力すべきであり、適切な場合には、市民社会組織、その他の関連する利害関係者、および科学パネルを含む専門家と、そのような規範の作成について協議することができる」（前文 116）とする。そして、「実践規範が発行され、AI オフィスにより関連義務をカバーするのに適切であると評価されれば、実践規範への準拠は、提供者に適合の推定を与えるべきである」（前文 117）とする。条文にある通り、AI オフィスは欧州委員会の機能の一つとされていて、したがって AI システムを監視する権限は、加盟国の管轄官庁および欧州委員会にあることになる。

8――小括

本稿で述べたのは、主に「適合性評価」と「汎用 AI モデル」であった。さらに本稿の範囲にはディープフェイクに関する対処の規定もあった。

適合性評価は、高リスク AI システムの EU 域内への市場投入や流通にあたって、適合性評価機関（被通知団体）からの審査を受けることを求めている。前回のレポートでは高リスク AI システム提供者が自社で実施すべきリスク管理システムを解説したが、今回のレポートで第三者の適合性評価について解説した。この適合性評価機関は EU レベルではなく、各国に設置されるものであることから、機関の間で意見が分かれる可能性もあり、本規則では、その場合の措置も規定されている（82 条。次回レポートで解説)。

また、汎用 AI モデルは欧州委員会の原案にはなかったものであるが、その後の欧州議会や理事会と

の折衝の過程で盛り込まれた規定である。汎用 AI モデルで問題視されているのは、システミック・リスクである。システミック・リスクはさまざまな場面で異なる意味で使用されている（たとえば銀行の連鎖倒産）が、ここでは重大事故、民主主義や公共の安全に関する悪影響が大規模に拡散することを指す(3 条(65))。システミック・リスクを有する可能性のある AI システムは汎用 AI モデルに限定されないとも思われるが、本規則では人と同じ能力を有する汎用 AI モデルが特に危険と判断したものであろう。

AI を用いた世論誘導疑惑は過去にも発生しているが、今後、ディープフェイク[15]などを用いて、ますますリスクとして高まっていくであろう。汎用 AI モデルの精度もますます高まる中で民主主義や公共の安全への危惧は増えこそすれ、減ることはない。汎用 AI モデルに係る条文は重要な規定であると考えられる。

15 ただし、本規則において、ディープフェイクについての規律は汎用 AI モデルにかかる規定とは異なる条文で行われている(50 条)。