基礎研レポート

AI 事業者ガイドライン

総務省・経済産業省のガイドライン

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

2024 年 4 月 19 日に総務省と経済産業省は「AI 事業者ガイドライン（第 1.0 版）（以下、ガイドライン）」[1]を公表した。ガイドラインは、AI の安全安心な活用が促進されるよう、我が国における AI ガバナンスの統一的な指針として策定された。またガイドラインの内容を詳述する別添が策定・公表されている[2]。

経緯としては、これまで総務省の「国際的な議論のための AI 開発ガイドライン案」「AI 利活用ガイドライン」及び経済産業省の「AI 原則実践のためのガバナンス」が策定・公表されてきた。これら 3 つのガイドラインを現状に照らして統合・見直しし、昨今の AI 技術の発展、また AI の社会的実装に関する議論を反映し非拘束的なソフトロー[3]として、今回のガイドラインとして取りまとめられた。

ガイドラインとして取りまとめられた意義としては、「AI の利用は、その分野とその利用形態によっては社会に大きなリスクを生じさせ、そのリスクに伴う社会的な軋轢により、AI の利活用自体が阻害される可能性がある[4]」ためである。このようなリスクを特定・把握し、対処することで AI の進展を阻害しないためにもガイドラインが必要となる。

本稿ではガイドラインを簡単に紹介し、一定の評価をすることを目的とする。その際、EU における AI 規則（以下、EU 規則という）を参考にする。

なお、ガイドラインおよび EU 規則の規定は多岐にわたるので、本文で重要な部分には下線を引いている。

---

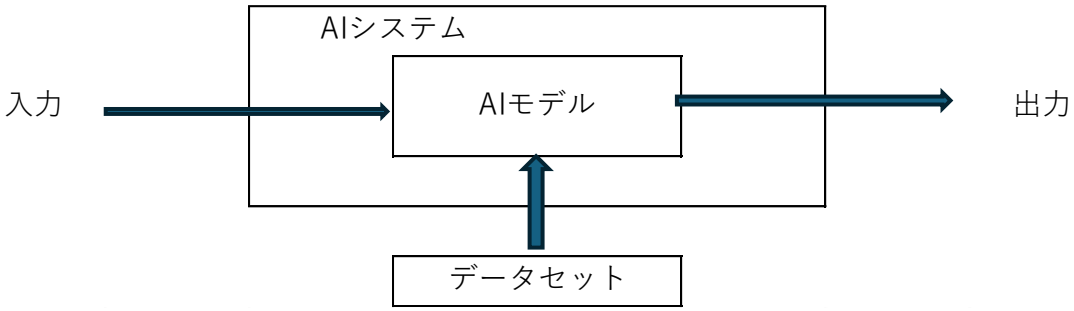
[1] 総務省・経済産業省「AI 事業者ガイドライン」
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_1.pdf
[2] 別添　https://www.meti.go.jp/press/2024/04/20240419004/20240419004-2.pdf
[3] 法律のような要件・効果のはっきりしているルールをハードローと言い、自主的に遵守する類のルールをソフトローという。
[4] 前掲注 1　p3 参照

2――AIシステムの構造

1｜AIシステムの構造

ガイドラインが定義するAIシステムは一読してもわかりにくい。ガイドラインによると、「『AIシステム（以下に定義）』自体または機械学習をするソフトウェア若しくはプログラムを含む抽象的な概念」とする。そしてガイドラインが定義する「AIシステム」とは「活用の過程を通じて様々な自立性をもって動作し学習する機能を有するソフトウェアを要素として含むシステム」としている。また、AIシステムは「明示的又は暗黙的な目的のために推測するマシンベースのシステム」であり、「受け取った入力から物理環境又は仮想環境に影響を与える可能性のある予測、コンテンツ、推奨、意思決定等の出力を生成する[5]」ものである。ガイドラインの定義を図にすると図表1のようなものである。簡略化していえば、AIシステムとは、①自律性をもって動作・学習し、②特定の目的のために推測し、③現実等に影響を及ぼす出力を生成するものである。

【図表1】AIシステム



ここで図表中央にあるAIモデルは「AIシステムに含まれ、学習データを用いた機械学習によって得られるモデルで、入力データに応じた予測結果を生成する[6]」ものとされている。例示的に言えば、AI利用者がインプットしたデータ（たとえば内臓の画像データ）から想定される出力結果（たとえば特定の病状）を推測するようなものを指す。出力結果を出すためにAIモデルには多量のデータセット（＝一定の目的のために一定のフォーマットで作成されたデータ群）が学習のため読み込ませてある。

これは機械学習というもので、コンピューターに大量のデータセットを読み込ませ、データ内に存在するパターンを学習させることで、新規のデータの性質等を判断するためのルールを獲得することを可能にする技術である。さらに、機械学習の一種にディープラーニングがある。これは人間が物（たとえば犬と猫）を判断する際に見分ける鍵となる特徴量（たとえば耳の形状や鼻の形状）をAIモデルに与えなくとも、AI自体が自動的に特徴量を獲得する技術である。この技術はヒトの脳の仕組みをベースに開発されたものである。最近の機械学習はディープラーニングを利用している。

ちなみにEU規則ではAIシステムのことを「AIシステムとは、さまざまなレベルで自律的に動作するように設計され、配備後に新たな状況に適応することができる機械ベースのシステムであって、かつ、明示的または暗黙的な目的のために、予測、コンテンツ、推奨、または決定（これらは物理的ま

[5] 前掲注1　p8参照。
[6] 同上

たは仮想的な環境に影響を与える）などの出力をどのように生成するかを、受け取った入力から推論するもの」と定義している(3条1項)。EU規則の定義とガイドラインの定義の相違の有無やポイントを明確に示すことは難しいが、ポイントとして、EU規則では「推論する機械ベースのシステム」がAIシステムであるとする一方で、ガイドラインでも「明示的又は暗黙的な目的のために推測するマシンベースのシステム」とするので、大きな相違はないものと考えられる。

2｜AIシステムの関係者

ガイドラインでは対象となる事業者として3種の事業者が挙げられている。

【図表2】AIに関係する事業者

(1)AI開発者：AIシステムを開発する事業者（研究開発者含む)。
(2)AI提供者：AIシステムをアプリケーション、製品、既存のシステム、ビジネスプロセス等に組み込んだサービスとしてAI利用者、場合によっては業務外利用者に提供する事業者。
(3)AI利用者：事業活動において、AIシステム又はAIサービスを利用する事業者（ただしAI利用者には個人が私的に利用するケースにおける利用者は含まない)

EU規則では、AI開発者という用語の定義はなく、開発（他社に委託して作成する場合を含む）し、市場投入および/または運営する者を提供者(provider)と呼び、自社が事業目的で利用する者を配備者(deployer)と呼ぶこととしている。EU規則では、禁止されるＡＩの行為や高リスクAIシステムにかかわる規定などに加え、提供者と配備者の義務を中心に規定されている。

ガイドラインは各主体全体の責務を規定するとともに、AI開発者、AI提供者、AI利用者別の事項として、各主体の役割を規定している。AIシステムに関係する各主体別に責務を規定するという点では、構造的にEU規則とガイドラインで大きな相違はない。

3――AIにより目指すべき社会及び各主体が取り組む事項(総論)

1｜基本理念[7]

AIにより目指すべき基本理念は以下の3点である。

【図表3】基本理念

(1)人間の尊厳が尊重される社会(Dignity)
(2)多様な背景を持つ人々が多様な幸せを追求できる社会(Diversity and inclusion)
(3)持続可能な社会(Sustainability)

EU規則の目的は「域内市場の機能を向上させ、人間中心の信頼できる人工知能（AI）の導入を促進することである。同時に、域内におけるAIシステムの有害な影響に対して、健康、安全、民主主義、法の支配、環境保護など、憲章に謳われている基本的権利の高水準の保護を確保し、イノベーション

[7] 前掲注1　p10参照

を支援することである」(1 条)とされている。要するに EU 規則が目指すのは基本的人権の尊重とイノベーションの支援である。

EU 規則 1 条を見ると、(1)～(3)は含んでいるように思われる。ただ、ガイドラインでは権利の保護を超えて「多様な幸せを追求できる社会」といった権利を積極的に追及していくということが理念として目指されている点に特徴がある。

なお、EU 規則に目的として掲げられているイノベーションの支援はガイドラインにおいて基本理念には含まれていないが、次項の「原則」に含まれている。これは内容が異なるというよりも、整理の仕方が異なっているだけで大きな相違ではないだろう。

2｜原則[8]

ガイドラインでは、各主体が「基本理念」より導き出される人間中心の考え方をもとに、AI システム・サービスの開発・提供・利用を促進し、人間の尊厳を守りながら、事業における価値の創出、社会課題の解決等 AI の目的を実現していくことが重要である。このためには、以下の価値を実現する必要がある（図表 4)。

【図表 4】ガイドラインにおける原則

・安全性・公平性
・プライバシー保護・セキュリティ確保
・透明性・アカウンタビリティ
・教育・リテラシー
・公正競争の確保・イノベーションの促進

ここで挙げられている原則に関しては、EU規則の各条文にも規定されていることから、EU 規則とガイドラインとで大きな相違はないと考えられる（次の「４－共通の指針）を参照)。

ただし、ここで留意しておくべき点として、EU 規則の各条文は、AI システム一般ではなく、規則上、高リスク AI システム[9]に分類されているものに対して、適用されるものがほとんどであることである。この点、ガイドラインではこのような区分をせず、一般に AI システムの適用される原則・指針を網羅的に規定したうえで、高度な AI システムに関係する事業者に共通の指針を別途定めている（後述）という形式をとっている。

4――共通の指針(各論)[10]

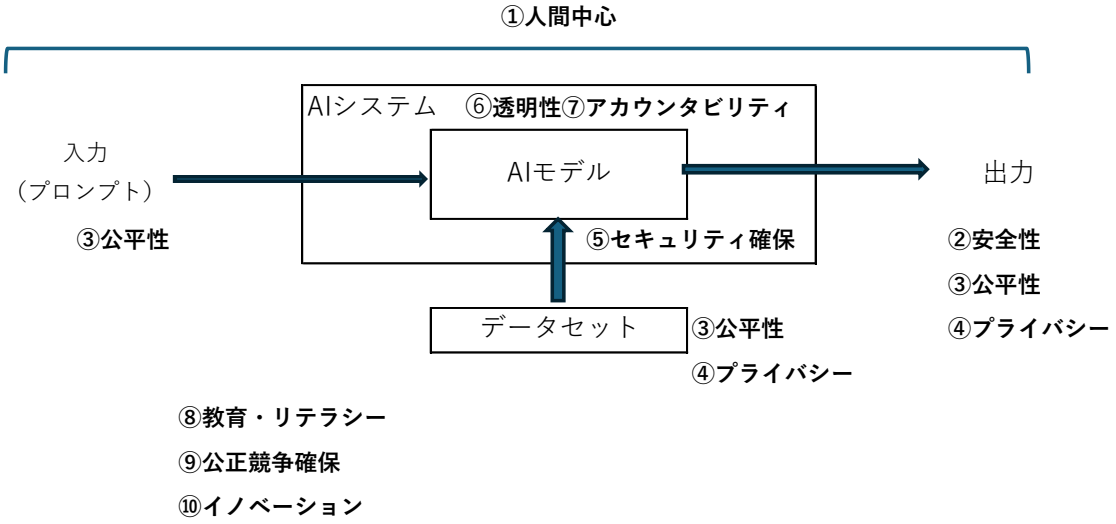
1｜共通の指針

各主体は以下の「共通の指針」に照らして各業務に取り組むことが必要とされる（図表 5)。

---

8 同上　p11 参照
9 EU 規則では人の権利や安全にリスクをもたらす AI システムなど一定の AI システムを高リスク AI システムとしてカテゴリー化している(6 条 4 項、49 条)。
10 前掲注 1　p12～ｐ20 参照

【図表 5】共通の指針



①人間中心：ガイドラインは、AI が人権を侵すことがないようにすべきであるとする。また AI が人々の能力を拡張し、多様な人々の多様な幸せ（well-being）の追求が可能になるように行動することが重要であるとする。具体的には、以下の項目が挙げられている（図表 6）。

【図表 6】指針―人間中心

ア）人間の尊厳及び個人の自律
イ）AI による意思決定・感情の操作等への留意
ウ）偽情報等への対策
エ）多様性・包摂性の確保
オ）利用者支援
カ）持続可能性の確保

本項目はいずれも重要な点に触れているが、そのうちア）～ウ）をピックアップする。

ア）では、ガイドラインは、たとえば個人のプロファイリングについて言及している。具体的に、プロファイリングを行うにあたっては、AI の限界を認識し、生じうる不利益を慎重に検討したうえで、不適切な目的に利用しないとしている。この点、たとえば英国では Children’s code では過去に不幸な事故が起こったことを教訓として、こどものプロファイリングを原則として禁止している[11]。また EU 規則ではプロファイリングを行う AI システムは必ず高リスク AI システムに分類され、特別な規定（リスク管理システムの導入など）が適用されるようになっている（6 条 3 項）。

つぎに、イ）においては人の意思決定を不当に操作することを目的とした AI システム・サービスの開発・提供・利用は行わないとする。著名な事件としてはケンブリッジ・アナリティカ事件がある。

11 https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/12-profiling/

この事件ではFacebookの個人情報を利用して、米国の大統領選に介入したもので、有権者の政治的な思考を不当に操作したとの疑惑がもたれている。またEU規則では「その目的または効果が、十分な情報に基づいた意思決定を行う能力を著しく損なわせることにより、人または人の集団の行動を実質的に歪め、その人、他の人または人の集団に重大な損害を与えるか、または与える可能性が合理的に高い意思決定を行わせること」(5条1項（a)。条文は一部省略)は禁止されている。

さらにウ)の偽情報への対策として、ガイドラインではリスクの高まりを受けて必要な対策を講じるべきものとされている。このリスクは生成AIの登場とともに深刻化しており、本物と区別のつかない偽動画等により、本人を架空のスキャンダルに陥れる事例がいくつも発生している。EU規則では「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」(50条4項)としており、AIシステムが作成した映像等を作りものであることと表示しないで利用することは禁止されている。

②安全性：ガイドラインでは、各主体が、AIシステム・サービスの開発・提供・利用を通じ、ステークホルダー（＝AI利用者や、出力結果が適用される個人など）の生命・身体・財産に危害を及ぼすことがないようにすべきであるとする。加えて、精神及び環境に危害を及ぼすことがないようにすることが重要であるとする。具体的には、ア)人間の生命・身体・財産、精神及び環境への配慮、イ)適正利用、ウ)適正学習といった項目が挙げられている。このうち、ア)に関しては以下の項目が挙げられている（図表7)。

【図表7】人間の生命・身体・財産、精神及び環境への配慮の

ⅰ）信頼性（出力の正確性など）
ⅱ）堅牢性（無関係な事象に対して著しく誤った判断を発生させないようにする）
ⅲ）必要な場合における人によるモニタリング及びコントロール
ⅳ）適切なリスクの分析、および監視・対策
ⅴ）危害を加える可能性がある場合に、講ずべき措置を整理し、ステークホルダーに情報提供
ⅵ）安全性を損なう事態の対処方針の検討

ここで、EU規則で高リスクAIシステムが満たすべき要件の一覧を挙げてみる（図表8)。

【図表8】EU規則における高リスクAIシステムが満たすべき要件

a)リスク管理システム及びリスク管理措置―8条、9条
b)データガバナンス―10条
c)技術文書および記録保存－11条、12条
d)使用説明書－13条
e)人的監視措置－14条
f)正確性および堅牢性など―15条

本項目では、

・(図表 7) の、 ⅰ)信頼性、 ⅱ) 堅牢性は、(図表 8)の f)正確性および堅牢性などに、また、

・同じくⅲ)モニタリング・コントロールは、e)人的監視措置に、

・同じくⅳ) リスク分析等およびⅥ) 事態発生時の対処方針は、a)リスク管理システム及びリスク管理措置に、

・同じくⅴ)ステークホルダーへの情報提供は、d)使用説明書にそれぞれ該当する。

また、イ) の適正利用については、EU 規則では c)技術文書および d)使用説明書に従って利用されることで確保される。

さらに、ウ) に関して、データの正確性、最新性は、EU 規則では b)のデータガバナンスでカバーされる。

そうすると細部はともかく、概ねガイドラインと EU 規則の高リスク AI システムに対する実体的な規定は一致しているように思われる（なお、記録(ログ)保管については下記⑥も参照)。

③公平性：ガイドラインでは、各主体が、AI システム・サービスの開発・提供・利用において、特定の個人ないし集団への人種、性別、国籍、年齢、政治的信念、宗教等の多様な背景を理由とした不当で有害な偏見及び差別をなくすよう努めること（具体的には図表 9) が重要であるとする。

【図表 9】偏見及び差別をなくすための取組

ア) AI モデルの各構成技術に含まれるバイアスへの配慮
イ) 人間の判断の介在

本項目では出力結果におけるバイアス、特に差別につながるようなバイアスに対する対処を規定している。ガイドラインは、バイアスは学習データ、AI モデルの学習過程、AI 利用者の入力するプロンプト（入力)、AI モデルの推論時に参照する情報等から発生することがあり、バイアスの要因を特定することを求められるとする。そして、結果の公平性確保のため、AI による単独判断ではなく、適切なタイミングで人間の判断を介在させることを求めている。

EU 規則では、バイアスのかかった出力をリスクとして、様々な対応手段が設けられているが、特に、データガバナンス（上記図表 8 の b)) で学習用、検証用、試験用のデータセットに偏りがないことを求めている(10 条 1 項)こと、また学習し続ける高リスク AI システムが偏った出力を出すリスクについて適切なリスク軽減措置を取るよう求めている(上記図表 8 の f)、15 条 4 項)。

④プライバシー保護：ガイドラインでは、各主体が、AI システム・サービスの開発・提供・利用において、その重要性に応じ、プライバシーを尊重し、保護することが重要であるとする。その際、関係法令を遵守すべきであるとする。

本項目は、個人情報保護法遵守およびプライバシー保護の重要性を指摘している。EU 規則では、個

人情報保護について直接触れた条文はないが、プライバシーの流出等も考慮すべきリスクの一つと考えられている。また、プライバシーに関してEUでは個人情報保護規則であるGeneral Data Protection RegulationがAIシステムに直接適用されるとともに、EU規則の前文で「プライバシーおよび個人情報保護の権利は、AIシステムのライフサイクル全体を通じて保証されなければならない」（前文69）と宣言されている。

⑤セキュリティ確保：ガイドラインでは、各主体は、AIシステム・サービスの開発・提供・利用において、不正操作によってAIの振る舞いに意図せぬ変更又は停止が生じることのないように、セキュリティを確保することが重要であるとする（図表10）。

【図表10】セキュリティ確保

ア）AIシステム・サービスに影響するセキュリティ対策
イ）最新動向への留意

本項目で触れられているのは、AIシステム・サービスの機密性・完全性・可用性・安全性を維持するために、その時点での技術水準に照らして合理的な対策を講ずることを求めている。それと同時に、AIシステム・サービスの脆弱性を完全に排除できないことを認識すべきものとしている。

本項目に関連して真っ先に想起されるのは、AIシステムに対するサイバー攻撃である。この場合に、AIシステムの停止だけでなく、プライバシー情報を含む情報漏洩や、偽情報の出力（ハルシネーション）などのリスクが発生するおそれがある。この点に関係し、EU規則では「高リスクのAIシステムは、適切なレベルの精度、堅牢性、サイバーセキュリティを達成し、ライフサイクルを通じて一貫した性能を発揮するように設計・開発されなければならない」（15条1項）としている。

⑥透明性：ガイドラインでは、各主体は、AIシステム・サービスの開発・提供・利用において、AIシステム・サービスを活用する際の社会的文脈を踏まえ、AIシステム・サービスの検証可能性を確保しながら、必要かつ技術的に可能な範囲で、ステークホルダーに対し合理的な範囲で情報を提供することが重要であるとする。具体的には、以下が挙げられる（図表11）。

【図表11】透明性

ア）検証可能性の確保
イ）関連するステークホルダーへの情報提供
ウ）合理的かつ誠実な対応
エ）関連するステークホルダーへの説明可能性・解釈可能性の向上

本項目では、透明性、言い換えるとステークホルダーへの説明責任について述べている。まず、AIの判断にかかわる検証可能性を確保するためログを保管することを求める（上記②参照）。そのうえで、AIの性質、目的等に照らしてAIのデータ収集や学習・評価の手法などの情報をステークホルダーに

提供する。さらにはステークホルダーの積極的な関与を促し、その意見を収集することを求めている。

事例として挙げられているのは、クレジットカードの審査において、利用限度額が同年収で男性より女性の方が低い査定がなされていたケースである。金融当局が調査に乗り出したが、クレジットカード企業はアルゴリズムの正当性について説明ができなかったというものである[12]。

EU 規則では、まず「高リスク AI システムは、配備者がシステムの出力を解釈し、適切に利用できるよう、その運用が十分に透明であることを保証するような方法で設計・開発されなければならない」(13 条 1 項)とされている。そして、透明性の観点から作成されるのが「使用説明書」の作成であり、「高リスク AI システムには、適切なデジタル形式またはその他の方法で、配備者に関連し、アクセス可能で理解可能な、簡潔、完全、正確かつ明確な情報を含む使用説明書を添付しなければならない」(13 条 2 項)とされている。さらに、ステークホルダーへの開示として、欧州委員会と加盟国で EU データベースを構築することとされている(71 条 1 項)。利用者等は EU データベースを参照することで、AI システムの使用説明書や、その他の参考情報を参照することが可能である（同条 4 項)。なお、開示項目については図表 13 を参照。

⑦アカウンタビリティ：ガイドラインでは、各主体は、AI システム・サービスの開発・提供・利用において、トレーサビリティの確保、「共通の指針」の対応状況等について、ステークホルダーに対して、各主体の役割及び開発・提供・利用する AI システム・サービスのもたらすリスクの程度を踏まえ、合理的な範囲でアカウンタビリティを果たすことが重要であるとする。具体的には、以下が挙げられている（図表 12)。

【図表 12】アカウンタビリティの具体的項目

ア) トレーサビリティの向上
イ) 共通の指針の対応状況の説明
ウ) 責任者の明示
エ) 関係者間の責任の分配
オ) ステークホルダーへの具体的な対応
カ) 文書化

本項目で挙げられているのは、AI 開発者や AI 提供者などの「共通の指針」の対応状況や、具体的な責任の分担などをステークホルダーに説明することなどである。

この点、上述⑥の通り、EU 規則では高リスク AI システムの提供者等が EU データベースに登録をすることでアカウンタビリティを確保している。具体的に EU 規則のもとで提供者が登録すべき内容は図表 13 の通りである。

[12] 同上　p15 参照。

【図表 13】EU 規則の登録事項（提供者）

1. 提供者の氏名、住所、連絡先
2. 提供者に代わって情報を提出する者の氏名、住所、連絡先
3. 該当する場合は、正式な代理人の氏名、住所、連絡先
4. AI システムの商品名、および AI システムの識別とトレーサビリティ（追跡可能性）を可能とする曖昧でない追加的な参照項目
5. AI システムの意図された目的、AI システムを通じてサポートされる部品（要素）と機能の説明
6. システムが使用する情報（データ、入力）とその動作ロジックの基本的かつ簡潔な記述
7. AI システムのステータス（市販中または使用中、市販/使用中止、リコール）
8. 被通知団体（＝適合性判定機関）が発行した証明書の種類、番号、有効期限、および該当する場合はその被通知団体の名称または識別番号
9. 該当する場合は、上記 8 で言及されている証明書をスキャンしたコピー
10. AI システムが市場投入され、使用開始され、または利用可能となった加盟国
11. EU 適合宣言書の写し
12. 電子的な使用方法書
13. 追加情報の URL（オプション）

図表 13 で特に説明が必要なのは、8～11 であろう。ここは図表 12 のイ）共通の指針の対応状況の説明に該当する。EU 規則では、高リスク AI システムに関する EU 規則の各種規定への適合性を審査することとされている。具体的には加盟国ごとの被通知団体（＝適合性審査機関。上記図表の 8. 参照）が適合性を審査して、合格した場合に証明書を発行することとなっている。高リスク AI システムの提供者は当該証明書、及び自ら EU 規制に適合していることを宣言する EU 適合宣言書の写しを登録することとされている。このようにステークホルダーは EU データベースを参照することで必要な情報を得ることができるようになっている。対して、ガイドラインではこのような適合性の審査制度および開示制度がないことに留意が必要である。

⑧教育・リテラシー：ガイドラインでは、各主体は、主体内の AI に関わる者が、AI の正しい理解及び社会的に正しい利用ができる知識・リテラシー・倫理感を持つために、必要な教育を行うことが期待される。また、各主体は、AI の複雑性、誤情報といった特性及び意図的な悪用の可能性もあることを勘案して、ステークホルダーに対しても教育を行うことが期待されるとする（図表 14）。

【図表 14】教育・リテラシー

ア）AI リテラシーの確保
イ）教育・リスキリング

本項目では、例えば偽情報に関するリテラシーの向上などを目指している。事例としては AI 合成音声による詐欺などが挙げられている。また生成 AI が事実と異なることをもっともらしく回答する「ハ

ルシネーション」の被害も存在する[13]。

EU 規則でも、「AI システムの提供者および配備者は、技術的知識、経験、教育および訓練、ならびに AI システムが使用される文脈を考慮し、AI システムを利用する人を考慮し、そのスタッフおよびその代理として AI システムの操作および使用に対処するその他の人の十分な AI リテラシーを、最善の範囲で確保するための措置を講じなければならない」(4 条)とされている。

⑨公正競争確保：ガイドラインでは、各主体は、AI を活用した新たなビジネス・サービスが創出され、持続的な経済成長の維持及び社会課題の解決策の提示がなされるよう、AI をめぐる公正な競争環境の維持に努めることが期待されるとする。

EU 規則では、公正競争の監視が各国の行政当局である市場監視当局の権限とされており、「市場監視当局は、市場監視活動の過程で確認された、競争規則に関する EU 法の適用に潜在的な関係を持ちうる情報を、欧州委員会および関連する各国の競争当局に毎年報告しなければならない」(74 条 2 項)とされている。この報告を受けた欧州委員会又は各国競争当局は競争法に照らして、違法かどうか、違法ならばどう是正措置を取るかを検討することになる。

⑩イノベーション：ガイドラインでは、各主体は、社会全体のイノベーションの促進に貢献するよう努めることが期待されるとする。具体的には以下の通り（図表 15)。

【図表 15】イノベーション

ア)国際化・多様化、産学連携及びオープンイノベーション等の推進
イ)相互接続性・相互運用性
ウ)適切な情報提供

EU 規則では、イノベーションの支援として、加盟国ごとに AI 規制のサンドボックス制度を設けることとされている(57 条 1 項)。その目的は以下のものとされている（図表 16。同条 9 項)。

【図表 16】EU 規則における規制のサンドボックスの目的

(a)EU 規則、または関連する場合に適用される他の EU 法および国内法への規制を遵守していることの法的確実性を向上させること
(b)AI 規制のサンドボックスに関わる当局との協力を通じて、ベストプラクティスを共有するための支援をすること
(c)イノベーションと競争力を育成し、AI エコシステムの発展を促進すること
(d)エビデンスに基づく規制の向上に貢献すること
(e)特に新興企業を含む中小企業が AI システムを提供する場合、EU 市場へのアクセスを促進・加速すること

[13] 前掲注２　p 15 参照

日本では既に規制のサンドボックスが分野に限らず認められているが、ガイドラインにおけるイノベーションの支援が簡素な記載にとどまっていることが若干気になるところではある。

2｜高度な AI システムに関係する事業者に共通の指針

高度な AI システムとは最先端の基盤モデル及び生成 AI システムを含む、最も高度な AI システムを指す。ガイドラインでは高度な AI システムに関係する事業者に向けた共通の指針を定めている。この指針では、たとえば「AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度な AI システムの開発の全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる」など、一般の共通の指針よりも一歩踏み込んだ指針内容となっている。具体的な内容は以下の通り（図表 17）。

【図表 17】高度な AI システムの満たすべき要件

Ⅰ）AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度な AI システムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる（リスク管理措置）。
Ⅱ）市場投入を含む導入後、脆弱性、及び必要に応じて悪用されたインシデントやパターンを特定し、緩和する（インシデントの緩和措置）。
Ⅲ）高度な AI システムの能力、限界、適切・不適切な使用領域を公表し、十分な透明性の確保を支援することで、アカウンタビリティの向上に貢献する（アカウンタビリティの向上）。
Ⅳ）産業界、政府、市民社会、学界を含む、高度な AI システムを開発する組織間での責任ある情報共有とインシデントの報告に向けて取り組む（各界との情報共有）。
Ⅴ）特に高度な AI システム開発者に向けた、個人情報保護方針及び緩和策を含む、リスクベースのアプローチにもとづく AI ガバナンス及びリスク管理方針を策定し、実施し、開示する（AI ガバナンスとリスク管理方針）。
Ⅵ）AI のライフサイクル全体にわたり、物理的セキュリティ、サイバーセキュリティ、内部脅威に対する安全対策を含む、強固なセキュリティ管理に投資し、実施する（強固なセキュリティの実施）。
Ⅶ）技術的に可能な場合は、電子透かしやその他の技術等、AI 利用者及び業務外利用者が、AI が生成したコンテンツを識別できるようにするための、信頼できるコンテンツ認証及び来歴のメカニズムを開発し、導入する（AI 作成識別技術）。
Ⅷ）社会的、安全、セキュリティ上のリスクを軽減するための研究を優先し、効果的な軽減策への投資を優先する（効果的なリスク軽減への研究・投資）。
Ⅸ）国際的な技術規格の開発を推進し、適切な場合にはその採用を推進する（国際規格の採用）。
Ⅹ）適切なデータインプット対策を実施し、個人データ及び知的財産を保護する（個人データ保護）。
ⅩⅠ）高度な AI システムの信頼できる責任ある利用を促進し、貢献する（責任ある利用）。

上記図表 17 は、主に前述の「共通の指針」をさらに厳格化したものと考えられる（リスク管理措置やインシデントの緩和措置、アカウンタビリティなど）。

他方、EU 規則では、高リスク AI システムとは別に汎用 AI モデルについての規定がある（高リスク

かつ汎用 AI モデルも存在する)。汎用 AI モデルとは、人間と同じようなタスクがこなせる AI で、高度な生成 AI などが該当するとされている。ただ、EU 規則における汎用 AI モデルは、システミック・リスク（大規模に悪影響を及ぼすリスク）を生じさせるものとして特別な規定が適用されるようになっており、ガイドラインの高度な AI システムの満たすべき要件と同じものかは不明である。さらに規制の具体的な中身も上記図表 17 とは大きく異なっており、ここではガイドラインと EU 規則の汎用 AI モデル規制との比較は行わないこととする。

5――AI ガバナンスの構築

ガイドラインでは、各主体間で連携しバリューチェーン全体で『共通の指針』を実践し AI を安全安心に活用していくためには、AI に関するリスクをステークホルダーにとって受容可能な水準で管理しつつ、そこからもたらされる便益を最大化するための、AI ガバナンスの構築が重要となるとする[14]。

具体的には、

①AI システム・サービスがライフサイクル全体においてもたらしうる便益/リスク、開発・運用に関する社会的受容、外部環境の変化、AI 習熟度等を踏まえ、対象となる AI システム・サービスに関連する「環境・リスク分析」を実施する。

② これを踏まえ、AI システム・サービスを開発・提供・利用するか否かを判断し、開発・提供・利用する場合には、AI ガバナンスに関するポリシーの策定等を通じて「AI ガバナンス・ゴールの設定」を検討する。なお、この AI ガバナンス・ゴールは、各主体の存在意義、理念・ビジョンといった経営上のゴールと整合したものとなるように設定する。

③ 更に、この AI ガバナンス・ゴールを達成するための「AI マネジメントシステムの設計」を行った上で、これを「運用」する。その際には、各主体が、AI ガバナンス・ゴール及びその運用状況について外部の「ステークホルダーに対する透明性、アカウンタビリティ（公平性等）」を果たすようにする。

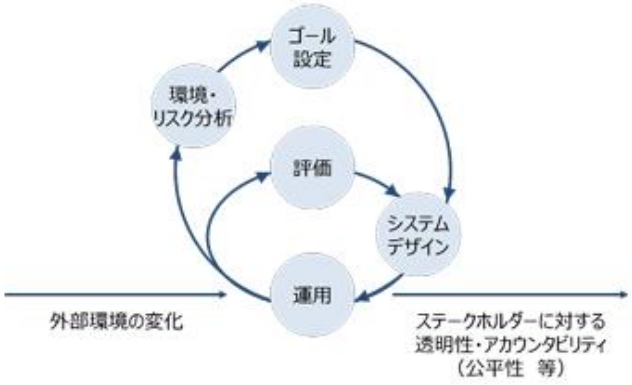
④ その上で、リスクアセスメント等をはじめとして、AI マネジメントシステムが有効に機能しているかを継続的にモニタリングし、「評価」及び継続的改善を実施する。

⑤AI システム・サービスの運用開始後も、規制等の社会的制度の変更等の「外部環境の変化」を踏まえ、再び「環境・リスク分析」を実施し、必要に応じてゴールを見直す。

このような AI ガバナンスは、ガイドラインによれば、サイバー空間とフィジカル空間を高度に融合させたシステム（Cyber-Physical System、CPS）の社会実装を進めるために必要なものとされている。なお、この AI ガバナンスの主語は「各主体」であることから、AI 開発者、AI 提供者、AI 利用者それぞれが AI ガバナンスを実施することが期待されているようではある。しかし、各主体ともそれぞれの立場に起因する限界があり、適宜連携を取りながら、AI ガバナンスを実施していくほかはないように思える。

[14] 前掲注１　ｐ24 参照

【図表 18】AI ガバナンス



出典：ガイドライン

EU 規則でガバナンスとして触れられているのは、EU レベルで設けられる組織と加盟国レベルで設けられる組織についてである。EU レベルでは、AI オフィス（欧州委員会の組織の一部）、欧州人工知能理事会（加盟国の代表で構成される会議体）、アドバイザリー・フォーラム（産業界や消費者団体で構成される会議体）、科学パネル（独立専門家による会議体）が設立される。また加盟国では市場監視当局（具体的に AI システムを監視する組織）が設置される。

EU 規則において、ガイドラインでいうところの AI ガバナンスに近しいものは、市販後モニタリングであろう。主な規定だけを抜き出すと、以下の通りである。

すなわち、提供者は、AI 技術の性質と高リスク AI システムのリスクに見合った方法で、市販後モニタリングシステムを確立し、文書化する(72 条 1 項)。市販後モニタリングシステムは、配備者から提供されるか、又は他の情報源を通じて収集される、高リスク AI システムの性能に関する関連データを、その耐用期間を通じて、積極的かつ体系的に収集、文書化及び分析しなければならない。そしてこれにより、提供者は、AI システムがⅢ章 2 節（高リスク AI システムの満たすべき要件）に定める要件に継続的に適合していることを評価できるようにしなければならない(72 条 2 項)とする。

ガイドラインでは、AI システムとして、社会的受容等も含めて、PDCA を回して機能改善を目指していくこととされている。これに対し、EU 規則の市販後モニタリングでは、高リスク AI システムの要件を満たしているかどうかをモニタリングして、適宜修正を行うというものである。ガイドラインの方は改善を目指すという意味付けが強いという相違点はある。しかし、いずれもモニタリング⇒修正の過程で改善を図ることが期待されているものであり、類似の性質を有するものと考える。

6――AI 開発者・AI 提供者・AI 利用者に関する事項

1｜AI 開発者に関する事項[15]

AI 開発者は、AI モデルを直接的に設計し変更を加えることができるため、AI システム・サービス

[15] 前掲注 1　p26～p28 参照

全体においてもAIの出力に与える影響力が大きい。また、イノベーションを牽引することが社会から期待され、社会全体に与える影響も大きい。このため、自身の開発するAIが提供・利用された際にどのような影響を与えるか、事前に可能な限り検討し、対応策を講じておくことが重要となる。

具体的な対策を項目のみ挙げる（図表19）。

【図表19】

ア）データ前処理・学習時 ・適切なデータの学習 ・データに含まれるバイアスへの配慮
イ）AI開発時 ・人間の生命・身体・財産、精神および環境に配慮した開発 ・適正利用に資する開発 ・AIモデルのアルゴリズム等に含まれるバイアスへの配慮 ・セキュリティ対策のための仕組みの導入 ・検証可能性の確保
ウ）AI開発後 ・最新動向への留意 ・関連するステークホルダーへの情報提供 ・AI提供者への「共通の指針」の対応状況の説明 ・開発関連情報の文書化
エ）その他期待される事項 ・イノベーションの機会創造への貢献

まず、ア）であるが、EU規則における「データによるAIモデルの学習を行う高リスクAIシステムは、学習用のデータセットを使用する場合は、（中略）品質基準を満たす学習、検証、テストのデータセットに基づいて開発されなければならない」（10条1項）に該当する。

つぎに、イ）であるが、これは人の生命や身体あるいは環境などに危害を及ぼさないように設計されるべきということがポイントとなっている。そうすると、EU規則では高リスクAIシステムの満たすべき要件である「リスク管理措置は、各ハザード（危険）に関連する残余リスク、及び高リスクAIシステムの全体的な残余リスクが許容可能であると判断されるものでなければならない」（9条5項）と同様の位置づけを有することと考えられる。

また、ウ）であるが、これはEU規則では使用説明書の作成及びEUデータベースへの登録（13条2項、71条1項、付属書Ⅷの12）義務によって確保される。

エ）については期待される事項であることから、省略する。

このようにガイドラインが定めるAI開発者に関する事項（＝求められる事項）はEU規則でも規定されているものと言える。AIシステムが人の権利や安全に悪影響を与えたとしたら、まずAIシステムの機能や動作についてAI開発者に一種の製造物責任が求められることとなることから、各種の義

務が課せられることは当然のことと考えられる。

2｜AI 提供者に関する事項[16]

AI 提供者は、AI 開発者が開発する AI システムに付加価値を加えて AI システム・サービスを AI 利用者に提供する役割を担う。AI を社会に普及・発展させるとともに、社会経済の成長にも大きく寄与する一方で、社会に与える影響の大きさゆえに、AI 提供者は、AI の適正な利用を前提とした AI システム・サービスの提供を実現することが重要となる。そのため、AI 提供者は AI システム・サービスに組み込む AI が当該システム・サービスに相応しいものか留意することに加え、ビジネス戦略又は社会環境の変化によって AI に対する期待値が変わることも考慮して、適切な変更管理、構成管理及びサービスの維持を行うことが重要である。具体的には、以下の通り（図表 20）である。

【図表 20】AI システム提供者が行うべきこと

ア)AI システム実装時 ・人間の生命・身体・財産、精神および環境に配慮したリスク対策 ・適正利用に資する提供 ・AI システム・サービスの構成及びデータに含まれるバイアスへの配慮 ・プライバシー保護のための仕組み及び対策の導入 ・セキュリティ対策のための仕組みの導入 ・システムアーキテクチャ等の文書化
イ)AI システム・サービス提供後 ・適正利用に資する提供 ・プライバシー侵害への対策 ・脆弱性への対応 ・関連するステークホルダーへの情報提供 ・AI 利用者への「共通の指針」対応状況の説明 ・サービス規約等の文書化

図表 21 を見ると概ね、共通の指針で定められたことを行ったうえで、各種の安全対策、セキュリティやプライバシーの確保、利用者やステークホルダーへの情報提供、技術等の文書化などを行うべきとしている。

比較して EU 規則(下記図表 21)を見ることとしたいが、EU 規則における提供者は、ガイドラインの AI 開発者と AI 提供者を併せた概念である。したがって、図表 19 と図表 20 とを、図表 21 と比較することとなる。

ここで目立つ相違点としては、EU 規則では EU 規則準拠の公式な適合性審査制度があるため、適合性宣言書や CE マーキングなどにかかわる規定があること程度であろうか。

[16] 前掲注 1　p31～p33 参照。

【図表 21】EU 規則における提供者の義務

(1)高リスク AI システムが 8 条～15 条の要件(高リスク AI システムの満たすべき要件)に準拠していることを確保すること
(2)高リスク AI システムに氏名、登録商号（商標)、連絡可能な住所を明記すること
(3)品質管理システム(17 条)を策定・文書化すること
(4)技術文書等(18 条)を作成すること
(5)自己の管理下にある場合、自動的に生成するログを保管(19 条)すること
(6)新規投入前に適合性評価手続き（43 条）を受けること
(7)EU 適合宣言書(47 条)を作成すること
(8)本規則への適合を示す CE マーキング(48 条)を添付すること
(9)EU データベースへ登録(49 条)すること
(10)本規則に不適合の場合、必要な是正措置を講じ、情報を提供(20 条)すること
(11)所轄官庁の合理的な要請があれば、高リスク AI システムが 8 条～15 条に定める要件に適合していることを証明(21 条)すること
(12)EU 指令（アクセシビリティ指令）に従って障がい者等の利用を容易にすること

そのほか、所轄官庁からの監督を受けることや、障がい者等の利用を容易にすることなどの相違点もある。

3｜AI 利用者に関する事項

AI 利用者は、AI 提供者から安全安心で信頼できる AI システム・サービスの提供を受け、AI 提供者が意図した範囲内で継続的に適正利用及び必要に応じて AI システムの運用を行うことが重要である。これにより業務効率化、生産性・創造性の向上等 AI によるイノベーションの最大の恩恵を受けることが可能となる。また、人間の判断を介在させることにより、人間の尊厳及び自律を守りながら予期せぬ事故を防ぐことも可能となる。

具体的には、以下の通り（図表 22)。

【図表 22】ガイドラインにおける AI 利用者の責務

・安全を考慮した適正利用 ・入力データ又はプロンプトに含まれるバイアスへの配慮 ・個人情報の不適切入力及びプライバシー侵害への対策 ・セキュリティ対策の実施 ・関連するステークホルダーへの情報提供 ・提供された文書の活用及び規約の遵守

ガイドラインでは、AI 利用者が果たすべき責務を記載している。すなわち安全性、プロンプトにおけるバイアス排除、プライバシー・セキュリティ確保、およびステークホルダーへの情報提供など常

識的にAI利用者が遵守すべきことが規定されている。

他方、EU規則では、配備者の義務として以下（図表23）が定められている。使用説明書遵守のための措置を講ずることや人的監視措置を講ずること、重大インシデント発生の場合に当局へ報告することなどが定められており、ガイドラインと比較して、重たい規制が課せられている。

ここからわかることは、EU規則では、提供者(日本では開発者および提供者)の開発・提供行為だけでなく、配備者が実際にAIシステムを運用する際の運営体制を重要視していることである。

【図表23】EU規則における配備者の義務

1. 使用説明書(13条)に従うための技術的・組織的措置を講ずること
2. 能力があり、訓練を受け、権限を有する者による人的監視措置(14条)を講ずること
3. 適切で十分に代表的であるデータを入力することについての管理を行うこと
4. 使用説明書に基づき監視し、重大インシデント発生時には提供者および市場監督当局に報告するとともに使用を中止すること
5. 高リスクAIシステムによって自動生成されたログを、システムの意図された目的に照らして、少なくとも6か月は保管するものとすること
6. 職場で高リスクAIシステムを使用する場合は、事前に労働者に対して、使用の対象となることと、およびその影響を通知しなければならないこと
7. 公的機関、またはEUの機関、団体、事務所もしくは機関である高リスクAIシステムの配備者は、登録義務（49条）を遵守しなければならないこと
8. 犯罪者捜査のために事後遠隔生体識別のために高リスクAIシステムを使用する者は遅くとも48時間までに司法当局の使用の認可を要請すること

7――検討

さて、ここでガイドラインとEU規則の全体像を比較してみたい。日本でもAI規制の法制度化の動きがある[17]ので、法制化にあたっての課題（すべてをカバーしているわけではないものの）と言ってもよいかもしれない。

(1)法的拘束力の有無

ガイドラインには事業者に対して法的拘束力がないのに対して、EU規則には法的拘束力がある。したがって、規定違反についてガイドラインでは何らペナルティがないが、EU規則では巨額の制裁金が科せられるおそれがある。そのため、ガイドラインでは各事業者によって遵守していない項目があってもおかしくはない。また是正措置を講ずるかどうかは当該事業者が自主的に決めることであり、外部より求めることは想定されていない。AIシステムによる社会や人権に与える悪影響がどの程度であるかを想定するかにもよるが、AIの利用範囲の急速な拡大を踏まえると、たとえば違反行為に当局が是正命令を出すことができるようにすべきであり、日本でもAI監督法の立法を必要とする事実は十

[17] 内閣府AI制度研究会　https://www8.cao.go.jp/cstp/ai/ai_kenkyu/ai_kenkyu.html　参照。

分あると考えられる。

(2)監督当局や審議機関の存在

EU レベルの監督機関として、欧州委員会の一機能である AI オフィスがある。AI オフィスは全体の調整を行うと同時に、汎用 AI モデルの監督を行う。また、各加盟国代表による構成される欧州人工知能理事会（規制執行について助言を行う会議体）、産業界・市民社会などからの代表者で構成されるアドバイザリー・フォーラム（欧州人工知能理事会及び欧州委員会に助言を行う会議体）、独立した専門家による構成される科学パネル（規則施行を支援する会議体）、および各国の市場監視当局（各国で AI システムを監視・監督する行政機関）が存在する。AI のガバナンスという場合にはガイドラインでは、事業者（開発者など）が自社内で行うガバナンスのことを意味するが、EU 規則では、事業者の行うモニタリングだけではなく、EU レベル、国家レベルで高リスク AI システムを監視することを意味する。

そのほか適合性を審査する制度があるが、それは次項(3)で述べる。

(3)適合性審査制度

EU 規則では、高リスク AI システムの満たすべき要件（リスク管理システムの導入、データガバナンスの実施、技術文書作成および記録保存、使用説明書作成、人的監視措置導入、正確性および堅牢性など）を、当該高リスク AI システムが満たしているかを審査する制度が存在する。EU 規則上、通知当局（notifying authority）と呼ばれる加盟国の行政機関が、適合性を審査する団体である被通知団体（notified bodies）を通知（＝認定）する。被通知団体は高リスク AI システムが規則上の要件を満たしているかどうか審査し、満たしている場合は証明書を発行する。証明書の発行を受けた提供者は、EU 適合宣言書を作成し、高リスク AI システムに CE マーキングを付するという一連の手続がある。

上記(2)(3)で述べた通り、EU 規則では AI システムに関するガバナンスを国家レベル、EU レベルで実現することを定めている。上記(1)でも述べた通り、AI 監督法を立法するのであれば、AI のリスクを継続的に監視・監督するという視点から、国レベルで行うべき監督体制の構築、あるいは国レベルのガバナンス体制をどうするかを決める必要がある。また特に、EU 規制の柱となっている適合性審査制度を設けるかどうか検討すべきである。

(4)禁止される AI の行為

ガイドラインでは、人の意思決定や感情を不当に操作することを目的とした AI システム・サービスを行わないという記載がある。

他方、EU 規則では、ガイドラインと同様の規定のほか、こどもや障がい者等を搾取する AI システム、ソーシャルスコアリング、予測取締システム、生体データの無差別収集、感情認識システム、機微な特徴を利用した生体分類システム、リアルタイム遠隔生体識別システムが禁止されている。

日本でも EU 規則を参考としつつ、リスクが高すぎて社会的に認容できない AI システムがガイドラインに定めるもの以外がないのか検討を行う必要があろう。

(5)高リスク AI システム

EU 規則では、主に AI システムを高リスクかどうかに分け、高リスク AI システムには必要な要件（リスク管理システムの導入、データガバナンスの実施、技術文書作成および記録保存、使用説明書作成、人的監視措置導入、正確性および堅牢性など）を満たすことを求めている。逆に、高リスクでない AI システムにはディープフェイク禁止など一部の規定が適用されるに過ぎない。

ガイドラインは適用される AI システムがどの程度およびどのようなリスクを有するものかを規定していない。単純な機器(例ー炊飯ジャー)にも AI が使われていることから、どこから規制対象となるかは日本で法制化するときに論点となるところであろう。

なお、規制対象となる AI システムが満たすべき要件については、本文で見てきた通り、EU 規制とガイドラインで大きく異ならないと考えられる。

(6)登録制度

EU 規則では高リスク AI システムおよびその他の一部 AI システムへデータベースに登録することを求めている。これは、AI の提供者（特に開発者）とステークホルダーの間に接点がないことから、一般の人でも参照できる登録制度として構築されている。この点は日本でも法制化にあたって同様の取組を行うことが考えられる。

(7)汎用 AI モデル

欧州委員会が起案した EU 規制の当初案にはなかったが、最終案に盛り込まれ立法されたのが汎用 AI モデルへの規制である。これは加速度的に進化する生成 AI などが大規模な悪影響を及ぼすことへの危惧から AI モデルの提供者（開発者含む）に、たとえば敵対的テストの実施や最新鋭技術の適用などの義務が加重したものである。

ガイドラインではこのような汎用 AI モデルを想定した規定はない（ただし、高度な AI システムに関する指針は存在する）ので、EU 規制を踏まえて導入の要否を検討するべきであろう。

8――おわりに

本文に書いた通り、ガイドラインは網羅的に遵守事項が記載されている。ただし、法令ではないので、「７－検討」で述べたように、EU 規則にあって、日本にも導入を検討すべき事項（監督組織や是正命令など）が今のままでは取り入れることができない。

欧州では包括的な規制ができたが、米国では州や連邦に立法の動き[18]があるが、包括的な規制としては立法されていない。この一つの理由としては、AI が驚異的な進歩を示すとともに、問題の外縁がどこなのかを決定しづらいところにもあると考えられる。本稿で比較対象とした EU 規則も現状の知見をもとに、いったんの規制として必要なパーツをそろえたというようにも見える。この規制で十分かはわからないが、EU 規則では欧州人工知能理事会など各種の会議体を設け、不断に検証していく体制が盛り込まれている。

他方で、問題となるのは、権利保護のため厳格な規制を引きすぎて、AI の発展を阻害するのではないかというところにもあるだろう。EU 規則でそのために用意されているのが、AI 規制のサンドボックスである。

人の権利・安全の保護と、AI 技術の発展促進という難しいバランスをどう確保していくか、日本に遅滞なく判断することが迫られている課題と言えよう。

18 一例としてカリフォルニア州ではディープフェイクについての法律が成立している。