資料２－１

# 「AI 制度に関する考え方」について

令和６年５月

AI 戦略チーム

# 「AI 制度に関する考え方」について

**目次**

## １．はじめに

生成 AI[1]の登場により、AI の利用機会や活用可能性が拡大している。一方で、AI がもたらすリスクも多様化・増大化している。

このため、AI については、イノベーションの創出、社会課題の解決、国際競争力の確保、国民生活の向上等の観点から便益を最大化するとともに、リスクは可能な限り低減させ、様々な場面で利活用を推進していくことが求められる。

リスクへの対応については、AI 戦略会議が 2023 年 5 月、生成 AI を中心に「AI に関する暫定的な論点整理」を公表した。この中で、AI の開発・提供・利用の促進には、規律の明確化等を通じたリスクへの適切な対応が、いわば「ガードレール」として必要との認識が示された。AI に関するイノベーションとガードレールは対立概念ではなく、イノベーション促進のためにも、適切なガードレールが必要である。

2023 年 5 月の G7 広島サミットで岸田総理が提唱した「広島 AI プロセス」は、生成 AI を含む高度な AI システムに関する国際的な指針と行動規範を含む「広島 AI プロセス包括的政策枠組み」を 12 月の閣僚級会合でとりまとめ、G7 首脳に承認された。

上記のような国内外の議論も踏まえ、我が国における AI ガバナンスの統一的な指針として「AI 事業者ガイドライン」が 2024 年 4 月に公表された。また、個人情報保護法等における AI への対応についても検討が進んでいる。一方、EU においては AI 法案により、リスクの高い AI に対して透明性義務や報告義務などの法的規制も検討されているほか、米国は大統領令で既存法令を活用しつつ、大規模 AI モデル開発者から報告を受けるなどとしている。

こういった背景を踏まえ、我が国においても、

・自主的な規律ではリスク低減を十分に実現できないのではないか

・健全な競争環境のためには、むしろ適切な規制が必要ではないか

等の指摘も出てきたため、2023 年 11 月から、当時検討が進められていた AI 事業者ガイドライン等の行動規範の履行確保及び AI 利用の促進に向けた調査を行った。本書は、我が国における AI 利用の実態や国内外の制度整備の状況等について調査した上で、今後の AI 制度に関する考え方をまとめたものである。

なお、安全保障の観点からも、AI の開発・利用に関する検討が必要といった論点もある。前述の「AI に関する暫定的な論点整理」では、当該論点については、情報管理の必要性も勘案し、政府内で別途検討することとされている。本書においてもデュアルユースや偽情報など AI の民間利用に関わる範囲で一部検討しているが、今後の技術や国際的議論の進展に応じて、政府内における別途の検討の成果を AI 制度に関する考え方に反映していくことが考えられる。

[1] 文章、画像、プログラム等を生成できる AI モデルにもとづく AI の総称を指す。

## ２．多様なリスクへの対応と各国における制度整備の検討状況

生成 AI の登場により AI のもたらすリスクが多様化・増大しており、ここでは、リスクへの対応の一般的なアプローチと、各国等での具体的なリスク対応等についての調査結果を我が国の現状とともに述べる。

### （１）多様なリスクへの対応とリスクベースのアプローチ

AI は、その利用分野や利用形態等によって、多様なリスクを生じ得る。一方で、リスク対応として分野横断的に一律に対策を講じると、AI 利用自体あるいは AI 活用によって得られる便益を阻害する可能性がある。このため、AI の多様なリスクへの対応としては、分野や形態ごとに生じ得るリスクの大きさを把握したうえで、そのリスクの程度に応じて対策する、「リスクベースアプローチ」が適切と考えられる。

AI の利用拡大により懸念されるリスクとして、例えば、機密情報の漏洩・個人情報の不適正な利用、犯罪の巧妙化、偽情報等による社会の混乱、サイバー攻撃の巧妙化、知的財産権の侵害の助長など多様なリスクが「AI に関する暫定的な論点整理」でも指摘されている。

こうしたリスクに対しては、基本的には、AI 関連事業者が自らリスク管理を行い、「信頼性の高い AI」の開発・提供・利用を行う努力が必要であり、AI の開発・提供・利用にあたっての透明性の確保と事業者によるガバナンスが重要と考えられる。

### （２）各国等の制度整備の検討状況

#### ① EU

EU 理事会・欧州委員会・欧州議会は「AI 法案」について大筋合意し（2023 年 12 月）、欧州議会は最終案を承認した（2024 年 3 月）。

この法案では、主として人権侵害、差別・偏見の防止などを重大なリスクと捉え、センシティブな情報を扱う AI は禁止、製品事故等の危険性がある高リスクな AI にはリスク評価や基準遵守義務、誤使用等のリスクのある AI には AI を使用していることの表示義務等を定めている。

AI 法案は、汎用 AI モデルについて、透明性要件の遵守義務（技術文書の作成、EU 著作権法の遵守、学習コンテンツに関する開示など）、さらにシステミックリスクを伴う汎用 AI モデル[2]には、より多くの義務（モデル評価や敵対的テストの実施、サイバーセキュリティの確保、重大インシデントやエネルギー効率の報告など）を課している。

AI 法案には、高額の課徴金など罰則規定がある。施行は法律制定から 2 年後（例外あり）の予定である。

具体的な執行には、CEN/CENELEC（欧州標準）や ISO/IEC（国際標準）等の規格も活

[2] 「$10^{25}$FLOPs 以上」等。「FLOPs」は、コンピュータが処理する計算量で、AI の規模を示す指標として用いられることがある。なお、GPT4 は「$2\times10^{25}$FLOPs」と言われている。

用する可能性があり、いわば、広範なハードロー（法的規制）を用意しつつ、具体的な執行面ではソフトロー（規格、ガイドライン等）で補完する方針であると言える。

**② 米国**

米国は、アマゾン、グーグル、Meta、マイクロソフト、OpenAI など大手 AI 開発者から「ボランタリー・コミットメント（2023 年 7 月）」を確保するとともに、各省庁に対して大統領令（2023 年 10 月）を発出し、既存の法令・予算を活用しながら、イノベーションを促進しつつ、リスクにも対応することを指示している。

米国は、AI 開発大手による自主的な規律遵守を基本としつつも、既存の法令を活用し、主として国家安全保障の観点から、国防生産法上の大統領権限に基づき、デュアルユース大規模汎用モデル[3]の開発企業に報告を求めるなどとしている。

いわば、ソフトローをベースにしつつも、大規模汎用 AI の開発者には一定のハードローも課す方針であると言える。法目的や形態は異なるものの、リスクの高い大規模な AI に関して法規制を課す点については EU と共通している。

**③ 国際的議論（G7、OECD、GPAI、UN 等）**

G7 では、我が国が 2023 年の議長国として広島 AI プロセスを主導し、高度 AI システムに関する国際指針及び AI 開発者に対する国際的行動規範を含む「広島 AI プロセス包括的政策枠組み」が 2023 年 12 月に承認された。2024 年の G7 はイタリアが議長国として、広島 AI プロセスの更なる前進に向けてモニタリングツールの開発等に取り組むほか、G7 以外の国へのアウトリーチ活動も進めている。

2024 年 5 月には、OECD 閣僚理事会における生成 AI に関するサイドイベントにおいて、岸田総理から、広島 AI プロセスフレンズグループ（49 か国・地域が参加）の立上げや GPAI 東京センターの設立を発表した。同閣僚理事会では、松本総務大臣が AI パートの議長を務め、加盟国等による AI 政策に関する議論をリードする役割を果たした。本会合にて OECD AI 原則の改定が採択され、閣僚声明にも広島 AI プロセスに対する支持を明記した。

GPAI（Global Partnership on Artificial Intelligence）[4]、UN、UNESCO（国連教育科学文化機関）、欧州評議会等においても AI ガバナンスや AI 倫理に関する議論が行われており、2024 年 3 月には国連決議の採択、AI 条約案への妥結等の動きがあった。

---

[3] 大統領令での定義は、「広範なデータで学習され、一般的に self-supervision を使用し、少なくとも数百億のパラメータを含み、広範な文脈に適用可能であり、安全保障、経済安全保障、公衆衛生もしくは安全、またはそれらの組合せに重大なリスクをもたらすタスクにおいて、高いレベルの性能を示すか、または示すように容易に修正可能な AI モデル」。国防生産法に基づく報告義務の対象となるモデルの定義を商務長官が定めるまでの間は、$10^{26}$FLOPs（生物学的配列データを用いる場合は $10^{23}$FLOPs）以上の基盤モデルが対象とされている。

[4] 人間中心の考え方に立ち「責任ある AI」の開発・利用を実現するため設立された国際的な官民連携組織。

**④ 日本**

我が国は、2016 年 4 月の G7 香川・高松情報通信大臣会合における「AI 開発原則」に向けた提案を先駆けに、広島 AI プロセスをはじめ、G7・G20 や OECD 等の国際機関での議論をリードし、多くの貢献をしてきた。

一方で、我が国において、AI に関する諸原則を実践していくにあたっては、

- 少子高齢化に伴う労働力の低下等の社会課題の解決手段として、AI の活用が期待されていること
- 法律の整備・施行は、AI の技術発展及びその社会実装のスピード・複雑さとの間でタイムラグが発生すること
- 細かな行為義務を規定する「ルールベース」の規制を行うと、イノベーションを阻害する可能性があること

等の課題が指摘されてきた。

これらを踏まえ、「AI 事業者ガイドライン（第 1.0 版）（2024 年 4 月）」では、AI がもたらす社会的リスクの低減を図るとともに、AI のイノベーションと活用を促進していくため、関係者による自主的な取組を促し、非拘束的なソフトローによって目的達成に導く「ゴールベース」の考え方を採っている。

加えて、2024 年 2 月、IPA（独立行政法人情報処理推進機構）に AI セーフティ・インスティテュート（AISI）を設置し、米英の同種の機関とともに AI の安全性確保のための実務的なガイドラインや評価手法の検討、各種の技術的な調査研究等を行っている。

このように、我が国はソフトローによる対応を中心に行ってきた。一方、刑法、個人情報保護法など AI か否かに関わらず適用される法律はあるが、欧米のような大規模あるいはリスクの高い AI 等に適用されるハードローは存在しない。

### （3）対応を検討すべきリスク

法規制等により欧米で進むリスク対応も踏まえ、我が国においては、以下のようなリスクについて、特段の対応を検討する必要があると考えらえる。

**① 製品・サービスの安全性に関するリスク**

医療機器やモビリティ、人間が活動する環境で稼働するロボット等においては、AI が誤作動することで生命身体に直接関わるリスクがある。また、対人・対物サービスの中で利用される AI が、身体や財産に悪影響を及ぼす可能性もある。

これらのリスクに関しては、多くの場合、個別の製品・サービスの安全性に関する法令において、AI 利用に関する規律が導入されているか、規制が検討されている。

**② 人権侵害（プライバシーや公平性など）に関するリスク**

AI を用いたサーベイランスやプロファイリング等においては、偽情報・誤情報の流通とも相まってプライバシー侵害、AI 生成画像等を使った名誉・信用の毀損や誹謗中傷、司法・行政権の行使や雇用・採用に関する差別的取扱いなどのリスクが考えられる。

これらに関しては、UNESCO や GPAI 等において、AI の倫理的問題が議論されているほか、EU においては規制導入の議論が進んでいる。

**③ 安全保障、犯罪増加などに関するリスク**

AI に関しては、兵器転用やテロの高度化、サイバー攻撃や詐欺等の増加・巧妙化、民主主義を揺るがすような偽情報や誤情報の流通、企業の秘密情報の漏洩等のリスクも考えられる。

兵器における AI の利活用については、国連の「特定通常兵器使用禁止制限条約」（CCW: Conversion on Certain Conventional Weapons）の枠組みにおいて「自律型致死兵器システム」（LAWS: Lethal Autonomous Weapons Systems）の規制等が議論されている。

また、民主主義を揺るがすような偽情報や誤情報の流通については、各国ともに安全保障部局が情報収集や正確な情報発信等の対策を講じているほか、EU においてはデジタルサービス法に基づきオンラインプラットフォーマー等に対する対応要請も行われている。

**④ 財産権の侵害リスク**

投資詐欺などのような国内外からの AI 生成物を用いた偽情報・誤情報の流通や、AI の学習に用いた企業情報の意図しない流出等により、財産権が侵害されるリスクも考えられる。

**⑤ 知的財産権の侵害リスク**

生成 AI によって、誰でも容易に画像や記事を作成できる時代になっており、オリジナルコンテンツに類似したコンテンツが流通してしまうリスクがある。AIが学習しているデータや AI の出力方法にその一因があるとする見方もあり、権利者と AI 開発者等による議論が続いている。

**⑥ その他のリスク**

かつての AI は定型的な業務に適すると言われていたが、生成 AI の進化によって創造的な業務も AI が担える可能性が出てきた。このため、AI によって雇用が失われるリスクに対する懸念が見られる。

また、AI が人による管理を超えて暴走するリスク（人の指示に従わずに作動する、人を攻撃するなど）を警告する学識経験者もいる。

さらに、一部の AI 開発者のみにデータや利益が集中する、少数言語国では自国言語による高性能な AI が存在しないといった課題も指摘されている。

## 3．AI 制度に関する基本的考え方

### （1）AI 事業者ガイドラインの趣旨

「AI 事業者ガイドライン（第 1.0 版）（2024 年 4 月）」は、関係者による自主的な取組を促し、非拘束的なソフトローによって目的達成に導くゴールベースの考え方により作成されている。また、AI をめぐる動向が目まぐるしく変化する中、マルチステークホルダーの関与の下で、「Living Document」として適宜、更新を行うことを予定している。

具体的には、AI の開発・提供・利用に関係する事業者が取り組む事項について、①人間中心、②安全性、③公平性、④プライバシー保護、⑤セキュリティ確保、⑥透明性、⑦アカウンタビリティ、⑧教育・リテラシー、⑨公正競争確保、⑩イノベーションに関する共通の指針を示しているほか、開発者、提供者、利用者固有の取り組む事項についても記載している。また、高度 AI システムに関係する事業者に対しては広島 AI プロセスの国際指針・国際行動規範の内容に即した指針を特に明記している。

### （2）ソフトローの意義とハードロー適用の考え方

ガイドライン活用の背景には、イノベーションの阻害を懸念し、AI ガバナンスは基本的にはソフトローに委ねるべきという考え方がある。これは、技術進歩が速い AI は、ハードローでは適切に事業者等を規律できず、民間事業者や学識経験者等からの「常に改善・修正を繰り返すアジャイル・ガバナンスが有効である」との指摘を踏まえたものである。

一方で、私人による自主規制でもなく、政府による直接規制でもない、公私で問題解決に向かう新しい政策手法として「共同規制[5]」の考え方がある。これを踏まえ、ソフトローを補完するものとして、例えば、特にリスクの高い AI の開発・提供・利用に対して、体制整備・情報開示義務や技術基準遵守義務等の一定の規制を導入すべきとの指摘もある。特にリスクの高い AI の例としては、高リスク AI（EU ではリスク管理プロセスの確立、適合性評価が必要）、システミックリスクを伴う影響力の大きいモデル（EU では様々な義務）、デュアルユース大規模汎用モデル AI（米国では報告義務）などがある。

こうしたリスク対応のために必要な規制は、AI の利用にあたっての安心感を高め、AI の利用拡大を後押しするとの指摘がある。他方で、事業者にとってはコンプライアンスコストとなり、事業遂行に萎縮効果を生じるとの指摘や、規制を遵守しない国内外の AI 開発者や提供者に対して必要十分な履行確保措置をとることがサイバー空間においては難しいとの指摘も考えられる。ハードローを適用する場合には、ハードローを適用する必要性や公平性を整理するとともに、明確化を図ることが重要である。

[5] EU 市場に投入される製品やサービスの認証に使用される整合規格も、共同規制手法（co-regulatory instruments）の一種である。EC が発表した AI 法の提案は、CE マークを取得し AI ソフトウェアを市場に出すためのコンプライアンスツールとして整合規格に依存している。整合規格は「New legislative framework」（NLF）の一部であり、リスクの高い製品やサービスを特徴とする一部の技術・産業分野を規制するために採用されているアプローチである。NLF では、法律が一般的な要求事項のみを定義し、どのように遵守するかは法律の受け手に委ねられる。

### （３）国際整合性の確保

AI はボーダレスかつグローバルに利用される技術であり、グローバルに AI のリスクに対応していくための各国間の緊密な協力が必要である。

また、我が国で AI のイノベーションや利用拡大を目指すスタートアップや外国企業が安心して事業活動できるようにする観点からもルールの国際整合性や相互運用性（国際ルールとの偏差やその理由が明確になっている）が重要と考えられる。

我が国は、グローバルな共通ルールを策定する G7 広島 AI プロセスで全ての AI 事業者への国際指針や AI 開発者向けの国際規範を主導してきた。制度の検討にあたっては、広島 AI プロセスの結果や諸外国の制度を踏まえる必要がある。

### （４）リスクや技術進歩に応じた柔軟な制度

イノベーションと規制のバランスが重要であり、リスクベースアプローチが重要である。AI の利用促進に向け、AI ガバナンスが過剰規制とならないよう、ソフトローの最大限の活用を基本としつつ、リスクの高い使われ方をする AI や人権侵害や犯罪等につながり得る AI に対して必要な法的規制（ハードロー）のあり方を検討する必要がある。

また、技術の進歩や国際的な規制の議論に即応する必要があることから、規制を導入する場合には、制度に柔軟性を持たせる枠組みも重要である。また、分野別の多様なリスクに対応するため、既存の法制度の活用・見直しも同時に進めていくことも重要である。

### （５）AI 事業の主体及びリスクの高低に応じた考え方

我が国においても、安全保障、犯罪、人権侵害、製品・サービスの事故、知的財産権侵害など、AI に関する様々なリスクが懸念される（知的財産権に関しては４.（５）にて後述）。

これらのリスクに関して、AI の不適切な動作の抑止、必要な情報の開示、インシデント対応、リテラシー向上など広島 AI プロセスでも議論されてきた事項は、人権、安全保障等を含む複数のリスクを低減し、国民の安心・安全に寄与すると考えられる。

このようなリスクへの対応に関して、様々な種類のリスクが挙げられる中でリスクの高低の判断基準を検討するとともに、AI の開発・提供・利用等のライフサイクルに関わる各主体の役割は何か、ガイドラインや法制度はどうあるべきか等を、海外の制度等も参考にしつつ検討していく必要がある。

各主体の役割に関しては様々な考え方があるが、ここでは、海外の制度等を参考にした一つの考え方として、AI ライフサイクルを通した事業主体に関して、リスクの高低に着目して論点を整理した場合の一例を以下に掲載する。対応すべきリスクに応じて着目すべき AI や各主体の役割について、引き続き精査が必要である。

事業主体に関しては、AI 開発者、提供者、利用者という分類を用いるが、技術や業態の変化に応じて、分類・定義は引き続き検討が必要である。

リスクに関しては、安全保障等の観点から欧米においても大規模汎用モデル（高性能かつ広範に使用され、悪用された場合の影響が大きい）の規制が検討されており、少なくともここに着目するが、リスクの所在・内容・高低や各主体の役割等は今後も変わる可能性があり、

引き続き検討が必要である。

リスクへの対応に関しては、事前対応[6]と事後対応[7]があることも意識して検討する。

AI 開発者に対しては、リスクの高低を踏まえつつ、定期的に規制を見直していくことでイノベーションと規制のバランスを確保することが考えられる。例えば、当面はイノベーション促進の観点から、モデル開発やデータ提供等を支援し、その際に AI の概要や AI ガバナンスポリシーの開示など「AI 事業者ガイドライン」遵守を求めたり、安全性・脆弱性に関する政府等との情報交換の枠組みの構築も考えられる。一方、大規模汎用モデル等の開発者に対しては、体制整備や情報開示等を求める制度枠組みの要否も含め検討が必要である。EU では、必ずしも AI 開発者による対応だけでなく、AI 生成物への表示義務なども見られるが、大規模汎用モデルに関して開発者側に何らかの義務を課す点は米国と共通する。全国民が AI ユーザーになり得る状況の中で、大手 AI 開発者やオンラインプラットフォーマー等に対応を求めるのは一つの実効的な考え方である。

AI 提供者・利用者に対しては、リスクに応じた適切なリスク軽減措置を求めていくアプローチが考えられる。例えば、医療、自動運転、金融等の社会への影響が大きい重要分野は、技術の進展や利用状況に応じて制度の見直しの必要性等を検討することが考えられる。

また、リスクの高低に関わらず、AI 提供者・利用者による AI ガバナンスの自己宣言・開示の枠組み等を整備しつつ、民事法・刑事法違反行為に利用される蓋然性の高い AI に対する事後的な措置を検討することも考えられる。

オンラインプラットフォーマーも重要な関係者となる。リスクの一因である偽・誤情報の増加・精巧化に関しては、オンラインプラットフォーマーが不適切なコンテンツの削除等の役割を負うなど、情報流通全体の制度枠組みの中での対応も進めていくことが望ましいと考えられる。

[6] 許認可、第三者認証、自己宣言・開示等

[7] 安全性や脆弱性のリスクがある AI に対する対応措置

## （参考）AI 関係者を巡る制度検討のイメージ


※開発者・提供者・利用者の分類・定義等については引き続き検討が必要。

**① 影響大・高リスクの AI 開発者**

影響大・高リスクの AI（例.大規模基盤モデル等の高度 AI システム）開発者に対しては、EU、米国においてはハードローによる体制整備や情報開示の義務付けが検討されている。

我が国においても、セキュリティリスクやシステミックリスク等の防止、インシデント対応、広範な提供・利用段階でのリスク対応の難しさなどを考慮し、国民の安全・安心の観点から、AI 事業者ガイドライン等のソフトローを補完する法制度の要否の検討が考えられる。

制度を導入する場合、技術の変化の速さ、多様性などから、一律の規律の設定は容易ではなく、規制の大枠は決めつつも運用の詳細は官民連携型の第三者機関で検討するなど、共同規制型、ゴールベースの制度も重要であると考えられる。

**② 影響小・低リスクの AI 開発者**

リスクが低く、スタートアップや研究者等も多く、規制は避けるべきと考えられる。AI 開発者によってリスク等は異なり、AI 事業者ガイドラインなどソフトローによる対応が適切であると考えられる。

AI 開発者は、AI の目的等に応じて、安全確保、学習データへの配慮、情報提供、インシデント対応などが求められるほか、プロダクト認証制度の検討なども考えられる。

**③ 影響大・高リスクの AI 提供者・利用者**

業法・規制法がある重要インフラ等に関しては、AI 導入の基準等が必要な場合にはその法令（安全基準、設備基準等）で規制すべきと考えられる。例えば、AI を搭載した自動運転車、医療機器は、道路運送車両法、医薬品医療機器法の下で認可、承認例がある。AI の動向を踏まえ、どのような AI 利用に対して規制が必要かなど、業法ごとに検討が必要である。

業法・規制法がない重要インフラ等に関しては、技術の変化や利用状況に応じて機動的な対応が望まれるが、AI 提供者等が遵守すべき事項（例えば AI 搭載製品の安全基準）については、具体的な議論の積上げが必要である。このため、法令が整備されるまでの間、及び、法令が整備された後も、運用レベルで AI 事業者ガイドラインの活用もあり得ると考えられる。

**④ 影響小・低リスクの AI 提供者・利用者**

AI 事業者ガイドラインなどソフトローによる対応が適切であると考えられる。提供者・利用者は、AI の使途や組織の規模等に応じて、インシデント情報の把握や情報提供への協力、利用者のリテラシー向上等を含む AI ガバナンスポリシーの策定・公表も有効と考えられる。

提供者・利用者のガバナンスを第三者が認証する制度も考えられる。既に民間団体による認証サービスが検討されており、その発展や普及が望まれる。

**⑤ 政府による AI の適切な調達・利用、AI 利用に関するリスク情報の調査等**

政府は AI を適切に調達・利用する必要がある [8]。政府の調達基準や使用方法は他への波及効果を有する。

政府は、リスクの高低に関わらず、多様な分野で利用される AI について AI ガバナンスの自己宣言・開示の枠組み等を整備しつつ、人権侵害や法違反行為に利用される事例などリスク情報を調査・収集し、悪用の蓋然性が高い AI に対する事後的な改善・排除措置を検討することも考えられる。

[8] デジタル庁は、AI 事業者ガイドラインの施行を受けて、「ChatGPT 等の生成 AI の業務利用に関する申合せ（デジタル社会推進会議幹事会申合せ）」の改訂を予定。加えて、「デジタル社会推進標準ガイドライン」を改訂予定（国際的にも議論は未成熟であり、規則ではなく参考資料という扱い）。

## ４．具体的な AI 制度に関する考え方

具体的な制度検討にあたっては様々な論点がある。一つの考え方として３．（５）のような考え方を採った場合の主な論点等を以下に記す。

### （１）影響大・高リスクの AI 開発者に対する考え方

#### ① 目的

諸外国の例を見ると、EU は人権侵害や差別の防止、米国は安全保障を重視するなどの違いは見られるものの、細かな行為義務を規定するルールベースの規律は求めず、リスク対応のための体制整備や情報開示に力点を置いている点は共通している。

我が国において制度を整備する場合の目的としては、国民の生命・自由・財産・基本的人権を保護する観点から、AI の安全性や信頼性の確保による国民の安全・安心の向上が考えられる。加えて、例えばデュアルユース AI の適切なリスク管理、犯罪的行為への悪用の抑止、経済安全保障の観点も含めて検討することが考えられる。

なお、公正競争が確保されているか、引き続き注視が必要と考えられる。

#### ② 対象範囲

米国は AI モデルの規模、EU は規模に加えて使途やリスクレベルも考慮して規制対象を規定している。我が国も技術の変化や海外の制度も踏まえ、検討していく必要がある。

この点、今後の AI 技術の進化により、AI モデルの規模は必ずしも AI の性能と比例しない可能性があるなど、リスクレベルを定量的に定めるのが難しい面もあることから、AI モデルの規模や使途などを総合的に勘案した検討が必要であると考えられる。

なお、制度の目的を踏まえ、国外事業者が我が国に対して提供する AI についても対象とする必要があり、制度を遵守しない AI 事業者に対する AI の提供・利用段階の措置を含む実効的な対応のあり方についても検討が必要である。

#### ③ 制度の考え方

影響大・高リスクの AI は、国内外の事業者が開発・提供することが想定され、国際整合性の観点から、高度 AI システムの開発者を対象とする G7 広島 AI プロセスの国際指針や国際行動規範をベースに、リスクを評価し、緩和するための措置や関係事業者等に対する情報提供等の制度の内容を検討する必要がある。

また、技術革新の速さや不確実性に迅速・柔軟に対応するため、ハードローとソフトローの組合せによる「共同規制」の考え方を念頭に、例えば、リスク対応や情報提供に関する基本的な事項は法令で定め、細則については、AISI 等の専門機関において、官民共同でガイドラインや規格を策定する方法も考えられる。

## （2）影響大・高リスクの AI 提供者・利用者に対する考え方

### ① 人の生命・身体等に直接影響を及ぼすおそれのある分野

安全性確保の観点から法規制が存在し、例えば AI を利用した医療機器や自動運転車については、当該規制に基づき承認・認可等が必要である。概要を以下に示す。

#### ア．医療機器

医薬品医療機器等法により、安全性及び有効性を確保するため、リスクに応じて規制されており、特にリスクが高い医療機器は、構造、使用方法、性能等について、臨床試験データ等に基づいて審査された結果、厚生労働大臣によって承認されたものでなければ販売できない等とされている。

医療機器の審査に関する基本的な考え方は AI 活用の有無によって変わらないが、AI を活用した医療機器の評価法や評価指標が示されている。加えて、AI は、あくまでも医師が判断するための補助ツールであり、AI 診断、治療等を行う主体は医師であり、医師はその最終的な判断の責任を負う、とされている。

#### イ．自動運転車

自動車は、道路運送車両法により、その安全性を確保するため、構造、装置及び性能について保安基準が定められており、国により保安基準適合性が確認されたものでなければ運行の用に供してはならないとされている。特に、自動運転車は、自動運行装置が運行設計領域において使用されると仮定した場合の保安基準適合性について審査を受け、認可されたものでなければならないとされている。

#### ウ．その他の分野（重要インフラ分野等）

一部の重要インフラ分野では業法等によって、安全性を担保している。現状、例えば、設備の更新計画を検討するための劣化診断、自主保安における設備点検、効率的な設備運用に資する提案等に AI が利用されているが、人による判断を支援する一つのツールであり、AI の利用に伴う制度改正の検討はなされていない。

先進的な取組として、プラント保安分野で、経済産業省等は信頼性の高い AI の実装促進の観点から「プラント保安分野 AI 信頼性評価ガイドライン」を公表している。

### ② 権利侵害や差別的対応のおそれのある分野

各分野の法規制において、権利侵害や差別的対応が禁止されている。例えば、労働基準法や男女雇用機会均等法においては、国籍、信条等による差別や男女差別が禁じられている。採用の支援ツールとして、動画面接を AI でスクリーニングする事例があるが、選考の正確性を担保するため、AI 判定で不合格となったケースも人が最終的に確認している。このように、現状は人による判断を支援する一つのツールとして活用されており、AI の利用に伴う制度改正の検討はなされていない。

クレジット取引の分野では、包括信用購入あっせん業者は、支払可能見込額の算定において AI を含めた高度な技術的手法を利用することが可能であるが、「不当な差別、偏見その他の著しい不利益」のおそれがある方法は禁じられている[9]。

**③ その他分野**

生成 AI の利用が想定されるコールセンター分野では、業界別のセキュリティガイドラインを除けば、特段の規制は無い。

また、教育分野では、文部科学省は、初等中等教育段階の学校関係者が生成 AI の活用の可否を判断する際の参考となるよう「初等中等教育段階における生成 AI の利用に関する暫定的なガイドライン」を公表しているほか、大学・高専の対応の参考となるよう「大学・高専における生成 AI の教学面の取扱いについて」を周知している。

**④ 制度の考え方**

AI を利用する場面やコンテクスト（人の判断を AI に置き換えるのか、判断材料の一つとして AI を利用するのか等）により、リスクは異なる。人の判断を AI が代替するようになれば、リスク低減のための制度整備の検討が必要となる。

内閣府は内閣サイバーセキュリティセンター等と連携しつつ、関係省庁等の協力を得て、重要インフラ等を中心とした分野における AI 利用の実態やリスク管理の状況等について調査するとともに、その結果を AI 戦略会議等に報告する。

なお、権利侵害や差別的対応のリスクに関しては、現在は人が行っている行為を AI が代替した場合に、人が行う行為との同意性等をどのように評価するのかといった法的な議論も深めていく必要がある。

また、上記（１）の AI 開発者に対するものと同様に、AI 提供者・利用者についても、公正競争が確保されているか、引き続き注視が必要と考えられる。

### （３）悪用される蓋然性の高い AI に対する考え方

技術革新の進展により、多様な AI モデルが容易に開発可能となり、偽情報の拡散や詐欺などのリスクが顕在化しつつある。このような状況下において、政府あるいは関係機関は、AI の悪用による被害発生時、もしくは発生可能性が高いと判断される場合において、必要に応じて実態調査を行い、その結果を公表し、国民及び事業者に対して注意喚起を行うことが求められる。さらに、必要に応じて、改善命令や実名公表などの措置を講じるべきとの声があがっている。

ソフトウェアのサイバーセキュリティに関しては、IPA が調査を行い、必要に応じて事業者または利用者が講ずべき措置を公表する仕組みが情報処理促進法に基づき運用されてい

[9] 当該技術的手法を利用する場合には、「利用者の支払い能力に関する情報を当該利用者に対する不当な差別、偏見その他の著しい不利益が生じるおそれがあると認められる方法により利用していないこと」が必要とされている。（割販販売法施行規則第 62 条第 1 項第 2 号）

る。AI の安全性に関しても、AISI における AI の安全性に関する定義や基準の検討も踏まえつつ、ソフトウェアのサイバーセキュリティの仕組みと類似の仕組みを AISI に構築すること及び必要な制度的措置を講ずることについての検討が必要である。

### （４）AI を利用した偽・誤情報等の生成・拡散に対する考え方

AI に関するリスクの一つである偽・誤情報の生成・拡散に対する技術的な対策としては、①AI 生成物に電子透かし等[10]を付加し、生成されたコンテンツが AI 生成物であること等を受信者が容易に理解できるようにする方法、②コンテンツにその出所や来歴等に関する情報を付与する等の技術を普及させ、信頼性のある情報が受信側で偽・誤情報に紛れないようにする方法、③オンラインプラットフォーマーが AI 生成物を判別してラベリングし、受信者が AI 生成物であると見分けられるようにする方法等が考えられる。

また、EU はデジタルサービス法や官民が協調して作成した行動規範等により、不適切な情報への対処をオンラインプラットフォーマー等に求めている。

我が国においても、インターネット上の違法・有害情報の流通が社会問題化している中、総務省において具体的方策が検討され、その結果、大規模なプラットフォーム事業者に対し、①対応の迅速化、②運用状況の透明化の具体的措置を義務づける「特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律の一部を改正する法律案」（情報流通プラットフォーム対処法）が令和 6 年 5 月成立した。

また、総務省は、2023 年 11 月より、「デジタル空間における情報流通の健全性確保の在り方に関する検討会」を開催し、デジタル空間における情報流通の健全性確保に向け、制度面も含めた総合的な対策を検討しており、2024 年夏頃を目途にとりまとめが行われる予定である。

AI に関しては、EU の AI 法案でも検討されている、AI により作成されたディープフェイクの明示義務の検討や、世界の大手テック企業が合意したミュンヘンアコードのような我が国における事業者の取組の推進を含め、具体的な方策を打ち出すことが期待される。

なお、昨今なりすまし広告等真実でないインターネット上の情報を真実と誤認することに起因して様々な被害が発生している。こうした被害を防止するための対策に関し必要な施策を検討するに当たっては、AI 技術の発展・普及により情報の改ざんや偽情報の生成が今後より精緻化・巧妙化することを踏まえる必要がある。

### （５）知的財産権の侵害リスクに対する考え方

生成 AI による著作権等の知的財産権の侵害リスクについては、内閣府知的財産戦略推

[10] コンテンツが AI 生成物か否か、コンテンツがオリジナルか否かは、AI 開発者が提供する電子透かしの仕組みで証明可能だが、コンテンツ作成者の真正性も証明するためには、コンテンツ作成者の関与も必要だと考えられる。

進事務局において「AI 時代の知的財産権検討会」を開催し、4 月 22 日の検討会にて、中間とりまとめ案が議論された。また、知的財産権のうち、著作権については、文化庁の文化審議会著作権分科会法制度小委員会において集中的に議論を行い、2024 年 3 月、「AI と著作権に関する考え方について」がとりまとめられた。

**① AI と知的財産権等の関係やリスク対応の方向性**

AI 時代の知的財産権検討会では、知的財産法に係る法的考え方の整理のほか、技術による対応策、契約による対価還元策を取り上げている。同検討会では、法、技術、契約の各手段は、相互補完的に役割を果たす関係があることを確認するとともに、AI 技術の進歩の促進と知的財産権の適切な保護が両立するエコシステムの実現に向けて、AI 開発者、AI 提供者、AI 利用者等の関係主体に期待される取組事例について、検討されている。

これは、知的財産権の侵害リスクとして指摘されるものは、知的財産法では直接の保護対象として明記していない労力や作風、声、肖像等も含め、必ずしも知的財産法のルールのみでは解決できない点も複合的に関わることを踏まえ、そのようなリスク等への対応策については、AI ガバナンスの取組との連動が必要であり、生成 AI に関わる幅広い関係者が、法・技術・契約の各手段を適切に組み合わせながら、AI 技術の変化や契約慣行、AI を取り巻く社会の状況の変化等も考慮して、連携してアジャイルに取り組むことが必要であるという観点から検討されているものである。

**② AI と著作権の関係やリスク対応の方向性**

昨今の AI 技術の急速な発展に伴う生成 AI の一般への普及により、クリエイターや AI 開発事業者等から AI と著作権に関する懸念の声が上がってきた。そこで、生成 AI と著作権に関する考え方を整理し、周知するため、「AI と著作権に関する考え方について」（令和 6 年 3 月 15 日文化審議会著作権分科会法制度小委員会。以下、「本考え方」という。）を取りまとめた。

本考え方を参照の上、各関係者が生成 AI との関係における著作物等の利用に関する法的リスクを自ら把握し、著作権等の権利の実現を自ら図ることで、著作権者等の権利保護と AI の適正な開発及び利用の環境が実現できると考えられる。

今後は、本考え方の周知・啓発を行うとともに、侵害事例の蓄積や、技術の進展、諸外国における検討状況等を踏まえつつ、議論を継続していく。また、法解釈のみでは対応できない課題について、関係者間での相互理解が促進されるよう、適切なコミュニケーションを促していく。

## ５．今後に向けて

AI 技術の変化は依然として早く、利用形態やリスクなどの予測は難しいが、その影響の大きさや国際動向等を考えると、国民の安全・安心を守る観点から、あるべき AI 制度について検討を進めていくことが重要である。

AI 制度については、様々な指摘があることから、学識経験者や事業者、利用者など、幅広くマルチステークホルダーの意見を聴取しつつ、検討する必要がある。

また、仮に規制を導入する場合でも、ブラックボックス化する AI 技術により、規制当局が規制の虜（レギュラトリーキャプチャー）に陥る可能性もあることから、規制の執行段階においては、民間事業者の専門的能力の活用が必要になると考えられる。

この観点からは、規制の有無にかかわらず、AISI のような官民連携組織に専門人材を集め、AI の安全性や信頼性について検討・評価を行えるようにすることが重要である。また、米英の AISI 等の同種の機関との国際的な連携・協調を通じて、AI のリスク対応や安全性確保のための標準的な方策をとりまとめ、AI 開発者・提供者に対してその遵守を促していくことも必要である。