とりまとめ（案）

令和８年 11 月 27 日

人工知能戦略専門調査会

制度の在り方検討ワーキンググループ

目次

I. はじめに 2
II. ＡＩ法制定前後の情勢 2
1. ＡＩのリスクに関する議論の歩み 2
2. 令和７年のＡＩ法成立以降に顕在化したリスク 4
(1) 悪用 4
(2) 予期せぬ挙動 5
3. 諸外国における制度整備その他の動向 6
(1) 国際的な枠組みに関する動向 6
(2) 米国の動向 6
(3) 欧州連合（ＥＵ）の動向 7
(4) オーストラリアの動向 7
4. 我が国における政府の取組 8
(1) ＡＩ法の責務規定 8
(2) 指針の整備及び各省庁におけるガイドライン等 8
(3) ＡＩ法第 16 条に基づく調査 9
(4) 個別の分野における対応 9
(5) ＡＩＳＩ（ＡＩセーフティ・インスティテュート）の取組 9
(6) 国際的な取組 10
III. 制度の在り方 10
1. 基本的考え方 10
(1) ＡＩ法の趣旨を踏まえた新たな制度整備の必要性について 10
(2) 法整備を検討する上で考慮すべき視点 11
(3) 対象となるリスクの整理 12
(4) 対象となる事業者の整理 13
2. ＡＩサービスの安全性確保に着目した法令の具体的な内容 14
(1) ＡＩの安全を確保するための措置を講じる義務 14
(2) 重大な悪影響が発生した場合に国に報告する義務 15
(3) 国からの報告の求めに応じる義務 15
(4) 国の勧告・命令権限 15
(5) 罰則 16
(6) ＡＩＳＩ（ＡＩセーフティ・インスティテュート）の役割 16
(7) 技術動向に対応する法令の在り方 16
3. ＡＩサービスの安全性確保に着目した法令以外の方策による対応の在り方 17
(1)「人工知能関連技術の研究開発及び活用の適正性確保に関する指針」の見直し 17
(2) 各種ガイドラインの改定・策定 17
(3) 官民対話の推進 18

(4) 国際連携の推進 18
(5) 国の調達・利活用の在り方 18
IV. 終わりに 19
1. 今後の制度の検討の在り方 19
2. ＡＩに起因するリスク対応の在り方 19

概要

制度の在り方検討WG とりまとめ（案）概要

令和８年10月以降、「制度の在り方検討WG 」[1)]を計４回開催。事業者等からのヒアリングを含む議論を経てとりまとめ（案）を作成。

背景

- AI法（令和７年５月成立、９月全面施行）はイノベーションの促進とリスク対応の両立を指向し、専らＡＩに着目した横断的な規律を設けず、既存の法令やガイドライン等のソフトローを適切に組み合わせて対応。
- 同法成立以降、AIが急速に高性能化し、悪用や予期しない挙動による、重大なリスクが顕在化。
- これを受けて、令和８年７月に閣議決定されたAI基本計画において、「AI法を含めた制度等を能動的かつ不断に見直す」こととされた。

AI法成立以後の情勢

- **AIの悪用による重大な悪影響の発生**（Ⅱ.２.(1)）
  - サイバー：国内外において、AIを悪用したサイバー攻撃の事案が多数報告
  - CBRNE：AIによる新種ウイルスの設計やバイオ研究関連の悪用事案が報告され、将来的に重大な悪影響が発生する懸念が高まる。
- **AIの予期せぬ挙動による重大な悪影響の発生**（Ⅱ．２．(2)）
  - AIエージェントが指示を逸脱し、第三者のサーバーに侵入する事案等が複数発生し、将来的に重大な悪影響が発生する懸念が高まる。

具体的な制度・施策の方向性

- 高性能かつ我が国に広く提供されているAIサービスには、より確実に安全対策を講じるよう求めることが適当。
- また、そのAIサービスにより重大な悪影響が発生した場合等には、国がそれを迅速に把握できるようにすることが適当。

- **新たな法制度**（Ⅲ.１-2）

※一定の要件に該当する事業者

  - **高性能なAIサービスを提供する事業者※に、重大な悪影響（サイバー、CBRNE、予期せぬ挙動による重大事故等）を防止する義務を規定**
  - **重大な悪影響が発生した場合等における、国に対する報告義務を規定**
  - **国による勧告・命令権限及び命令に違反した場合等の罰則を規定**
  - **AISI（AIセーフティ・インスティテュート）による技術的調査**



- **法制度以外の対応**（Ⅲ.3）
  - **法制度で対応する事項以外の事項については、AI事業者ガイドラインをはじめとする各種ガイドラインの改定、官民連携・国際連携、適正な調達・利用の推進により対応**

1)　人工知能戦略専門調査会の下に「制度の在り方検討WG」を設置。

I. はじめに

ＡＩは、経済成長や国民生活の発展に寄与するものであり、経済社会及び安全保障の上で極めて重要な技術であることから、その社会実装を進めイノベーションを促進していくことが重要である。その一方で、ＡＩには技術的・社会的観点や安全保障上の観点から様々なリスクがあり、技術進歩の状況やリスクに対する社会的受容の変化も踏まえ適切に対応していく必要がある。

このような状況を踏まえ、令和７年５月 28 日、人工知能関連技術の研究開発及び活用の推進に関する法律（令和７年法律第 53 号。以下「ＡＩ法」という。）が成立し、同年９月１日に全面施行された。我が国においては、ＡＩ法に基づき、イノベーションの促進とリスクへの対応の両立を目指して、専らＡＩに着目した横断的な規律を設けずに、既存の法令やガイドライン等のソフトローを適切に組み合わせリスクに対応するとともに、国が機動的に調査し必要な指導・助言等を行いつつ、事業者等に対してＡＩの研究開発・活用の適正な実施に係る自主的かつ能動的な取組を促す方針が採られてきた。

一方で、ＡＩ法成立以降のＡＩの急激な高性能化により、Claude Mythos[1]等の極めて高いサイバー能力を有するＡＩモデルが登場し、ＡＩを悪用したサイバー攻撃への懸念が高まっているほか、同じくＡＩを悪用した化学物質、生物、放射性物質、核、爆発物（以下「ＣＢＲＮＥ」という。）によるテロの将来的な発生の危険が示唆される事例が報告されている。かかる情勢を受けて、ＡＩの悪用や予期せぬ挙動により経済社会の根幹システムに深刻な影響を与え、国民の生命・身体等に重大な危険を生じる事態が発生することへの懸念が指摘されている。ＡＩの悪用や予期せぬ挙動により生じるリスクについては、諸外国においても、高性能ＡＩモデルの開発・提供を行う事業者に対する規律を設ける事例がある一方で、業界による自主規制の枠組みを推進する動きもみられる。

かかる状況を踏まえ、またＡＩ法に基づき令和８年７月に閣議決定された第２期人工知能基本計画においては、国内外における情勢の変化も踏まえ、「複雑化かつ深刻化するリスクへの対応等をより実効的に行うため、ＡＩ法を含めた制度等を能動的かつ不断に見直す」ことが決定されたことを受けて、令和８年 10 月より既存のＡＩに関する法制度の見直しを含め、制度の在り方を議論・検討することを目的として、制度の在り方検討ワーキンググループ（以下「本ＷＧ」という。）が人工知能戦略専門調査会の下に設置された。

本とりまとめ案は、本ＷＧの検討の結果を踏まえ、ＡＩ法成立以降に生じた状況を踏まえた対応について、今後の具体的な方策の在り方や検討の具体的な方向性を示すものである。

II. ＡＩ法制定前後の情勢

ＡＩのリスクに関しては、ＡＩ法の制定以前から様々な議論が展開されており、ＡＩ法制定時にも、その当時現に顕在化していたリスクを検討し、また、将来にわたって顕在化し得る多様なリスクを想定して、在るべき制度の検討が進められた。

---

[1] 令和８年４月に米 Anthropic 社が公表した、脆弱性の発見等において、従来のＡＩモデルを大きく上回るサイバーセキュリティ性能を示すＡＩモデル

ＡＩに関するリスクを踏まえたＡＩ制度の在り方について検討するため令和６年７月に設置されたＡＩ制度研究会の中間とりまとめにおいては、

「イノベーション促進とリスクへの対応の両立を確保するため、法令とガイドライン等のソフトローを適切に組み合わせ、基本的には、事業者の自主性を尊重し、法令による規制は事業者の自主的な努力による対応が期待できないものに限定して対応していくべき」[3]

とされた上で、サイバー攻撃やＣＢＲＮＥの製造、予期せぬ挙動等の重大なリスクについては、更に検討を継続し、将来的にリスクが増大した場合には、制度の見直しあるいは新たな制度の整備等も視野に入れて検討すべきことが提言されている。

ＡＩ制度研究会　中間とりまとめ（抄）

II.制度の基本的な考え方

３．イノベーション促進とリスクへの対応の両立

（２）法令の適用とソフトローの活用

一般的に、我が国の企業等は法令遵守の意識が高いとされており、新たな規制が制定された場合、当該規制の遵守を意識するあまり、新たな研究開発やサービスの開発・展開を必要以上に躊躇する可能性がある。技術の発展やサービスの変化が急速な AI の分野において、過度な規制により研究開発やサービスの開発・展開を抑制させてしまうことは、将来にわたって我が国の国際競争力を損なう危険性をはらむものである。そのため、新たな制度を検討・導入するにあたってはイノベーションに与える影響を十分に留意する必要がある。

上記の観点から、**イノベーション促進とリスクへの対応の両立を確保するため、法令とガイドライン等のソフトローを適切に組み合わせ、基本的には、事業者の自主性を尊重し、法令による規制は事業者の自主的な努力による対応が期待できないものに限定して対応していくべき**である。

また、**既存の個別の法令の存在する領域においては、**AI が各領域で様々な用途で利用され始めており、権利利益の保護の必要性が生じる場面も AI の用途に応じて異なることから、**まずは当該法令の枠組みを活用しつつ対応すべきである。**

III.具体的な制度・施策の方向性

３．生命・身体の安全、システミック・リスク、国の安全保障等に関わるもの

医療機器、自動運転車、基盤サービス等、特に国民生活や社会活動に与える影響が大きい生命・身体の安全やシステミック・リスクに関わるものについては特に注力して対応する必要があると考えられるが、業界毎に各業所管省庁が既存の業法に基づき対応し、また、追加的な対応の必要性の有無を判断するため、ＡＩ技術の発展、利用状況について随時業界と対話している状況である。**現時点においては、引き続き、各業所管省庁が既存の法令あるいはガイド**

[3] ＡＩ戦略会議・ＡＩ制度研究会「中間とりまとめ」（令和７年２月４日）10 頁
https://www8.cao.go.jp/cstp/ai/interim_report.pdf

ライン等の体系の下で対応すべきであるが、今後、新たなリスクが顕在化し、既存の枠組で対応できない場合には、政府は、関連する枠組の解釈を明確化したうえで、制度の見直しあるいは新たな制度の整備等を含めて検討すべきである。システミック・リスクについては、将来的に、複数のＡＩシステムが連動する大規模なＡＩシステム群が社会システムを支える状況となる可能性があり、その際、当該ＡＩシステム群が予期せぬ挙動をした場合、社会全体に大きな混乱をもたらす可能性があるため、適切に対処することが重要である。

また、**CBRN 等の開発やサイバー攻撃等へのＡＩの利用といった国の安全保障に関わるリスクについては、我が国の安全保障を確保するという観点から、関係省庁において、必要な対応をさらに検討していく必要がある。**

この点、ＡＩ法第３条では、基本理念の一つとして、

「人工知能関連技術の研究開発及び活用は、不正な目的又は不適切な方法で行われた場合には、犯罪への利用、個人情報の漏えい、著作権の侵害その他の国民生活の平穏及び国民の権利利益が害される事態を助長するおそれがあることに鑑み、その適正な実施を図るため、人工知能関連技術の研究開発及び活用の過程の透明性の確保その他の必要な施策が講じられなければならない。」

と定め、ＡＩのもたらすリスクにも対応した必要な施策を講じる方針が示されている。また、第６条では、同条において定義するＡＩの「研究開発機関」（開発事業者を含む概念。）について、

「基本理念にのっとり、人工知能関連技術の研究開発及びその成果の普及並びに専門的かつ幅広い知識を有する人材の育成に積極的に努めるとともに、第四条の規定に基づき国が実施する施策及び前条の規定に基づき地方公共団体が実施する施策に協力するよう努めるものとする。」

とし、さらに、第７条では、同条において定義するＡＩの「活用事業者」について、

「基本理念にのっとり、自ら積極的な人工知能関連技術の活用により事業活動の効率化及び高度化並びに新産業の創出に努めるとともに、第四条の規定に基づき国が実施する施策及び第五条の規定に基づき地方公共団体が実施する施策に協力しなければならない。」

としており、ＡＩの開発や活用を行う事業者にも、同理念にのっとった国等の施策に協力する責務を規定している。

2. 令和７年のＡＩ法成立以降に顕在化したリスク

ＡＩ法成立以降、ＡＩの性能が急激に向上し、社会実装も拡大する中で、以下のようなＡＩの悪用や予期せぬ挙動に基づく事象が現実化していることから、世界的にＡＩへの懸念が高まっている。

(1) 悪用

ア　サイバー攻撃への悪用

令和７年のＡＩ法成立以降、国内では、中高生等のサイバー攻撃に要する技術的能力や

経験を有さない者がＡＩを用いるなどしてサイバー犯罪等を行い逮捕される事例が複数件発生しているほか[5]、令和８年９月以降は、国内の企業・団体がサイバー攻撃を受けて個人情報が漏えいする事案が多発しているところ、その一部については、調査の結果、ＡＩが用いられていたことを示唆する痕跡が報告されている[6]。

海外においても、政府機関、重要インフラ及び金融機関等を狙ったＡＩを悪用しているとみられるサイバー攻撃の事例が報告されている。特に、令和８年 10 月には、韓国の金融機関 7 社でＡＩの支援を得たとみられるサイバー攻撃による顧客情報の漏えいが確認され、約 10 万人分の氏名、住民登録番号、年収等の情報が流出するなど、深刻な被害も発生している。

イ　ＣＢＲＮＥリスク

令和８年８月、米国スタンフォード大学の研究チームは、ゲノム生成に特化したＡＩモデルを用いて、自然界に存在しないウイルス（大腸菌を殺菌するバクテリオファージ）の設計に成功したと報告したが、同研究については医療応用が期待される一方で、ＡＩによるウイルス設計能力の進歩と悪用の可能性を示すものと評価されている[7]。また、同年９月、米国のＡＩ開発事業者は、自社のＡＩモデルを悪用したバイオ研究関連の事案を計５件報告した。その中には、鳥インフルエンザの哺乳類への適応・パンデミックの可能性につながる研究計画を数週間にわたって進めたケースや、毒性ペプチド分子の生成を試みたケースが含まれるとされる[8]。

これらの事例は、いずれも実際の被害が生じたものではなく、令和８年現在、ＡＩを悪用した大規模なＣＢＲＮＥテロの発生は確認されていないが、一方で、例えばバイオ分野では、令和８年６月には、米国の主要ＡＩ開発事業者のＣＥＯや科学者らが公開書簡を発表し、合成核酸の注文時におけるスクリーニングと記録保存の法的義務化を求める動きもみられる。

(2)　予期せぬ挙動

令和８年以降、悪意ある利用によらず、ＡＩの自律的・逸脱的な動作によって意図せず第三者のサーバーへの不正アクセス等が生じる事案が相次いで発生している。

特に令和８年７月には、米国 OpenAI 社のＡＩモデルがサイバーセキュリティ性能の評価試験中に隔離されたテスト環境から脱出してインターネットへのアクセスを獲得し、第三者

---

[5] 国内主要報道機関の報道による。

[6] 国内主要報道機関の報道による。

[7] Stanford Report「ＡＩ designs a novel E. coli killer」（令和８年８月６日）
https://news.stanford.edu/stories/2026/08/evo-2-ai-tool-e-coli-killer-bacteriophages

[8] Anthropic「Countering misuse of AI: September 2026」（令和８年９月 10 日）
https://www.anthropic.com/threat-intelligence-report-september-2026#biological-misuse-sep-26

である Hugging Face 社のサーバーに不正に侵入した事案が発生した。同事案では、ＡＩエージェントが性能試験の回答データを入手するために、Hugging Face 社のサーバー内のデータセットやアカウント等にアクセスしていたとされ、これは、与えられた目的を達成するよう最適化されたＡＩエージェントが、予期せぬ挙動を引き起こすリスクが顕在化した事案といえる。

同事案以降、同種の事案は複数発生しており、例えば、令和 8 年9月には、OpenAI 社のＡＩエージェントがオーストラリア政府のウェブサイトに不適切なアクセスを行い、非公開のデータを取得していたことが公表された。また、OpenAI 社以外の主要なＡＩ企業においても、自社のＡＩモデルが予期せぬ挙動を引き起こし、第三者が管理するサーバー等に不適切な干渉を行った事案が公表されている。

3. 諸外国における制度整備その他の動向

(1) 国際的な枠組みに関する動向

各国でＡＩの安全性への懸念が高まりつつある中、令和 8 年 9 月の国連総会では、ノルウェーとフィンランドの主導により、欧州委員会を含む 22 の国・機関がＡＩの安全性に関する共同声明を公表した。同声明は、ＡＩが人間の指示・管理下に置かれ続けるべきことを示した上で、以下の三点を提言している。

- ○　ＡＩの能力が一定の閾値を超えた際に各国を招集できる国際機関の設立の検討
- ○　ＡＩ開発企業に対する、展開前テストや独立した評価を含む透明性のある安全プロトコルの義務付け
- ○　安全性に関する重大インシデントの共有報告を含む、政府間の共通基準と透明性の強化

なお、同声明には米国を含む主要国は賛成しておらず、国際的な枠組みの構築に向けた道筋は依然として不透明であり、現状では、事業者の自主的な取組や各国の法制度に基づく規律による対応が先行している。

(2) 米国の動向

米国においては、令和８年現在、連邦法においてＡＩに関する横断的な法令による規律はなく、既存法による対応や、事業者の任意の協力・合意等を通じて安全性の取組を推進している。

一方、州レベルでは高性能ＡＩの安全性に関する規制法の制定が進んでおり、カリフォルニア州 SB53（令和 7 年 9 月成立、令和 8 年 1 月施行）を筆頭に、ニューヨーク州 RAISE Act（令和 9 年 1 月施行予定）やイリノイ州 AISM Act（令和 9 年 1 月施行予定）等が制定されている。各州法においては、概ね共通して、学習に要した計算量が一定（10^26 FLOPs）以上のフロンティアモデルを開発する事業者を対象に、以下を義務付けている。

- ○　事業者による安全性確保の枠組みの公表
- ○　ＡＩモデルごとの透明性報告
- ○　事業者が自社のＡＩを社内で使用し壊滅的リスクを評価した結果についての当局へ

の報告

○　安全性に関する重大インシデントの当局への報告（カリフォルニア州は発見から 15 日以内、ニューヨーク州・イリノイ州は 72 時間以内・差し迫った危険がある場合は 24 時間以内）

これらのうちカリフォルニア州 SB53 については、その基礎となった報告書において、「信頼と検証（Trust but Verify）」という規律の考え方が示されている[10]。これは、開発者による技術的な専門知識に基づく自主的なリスク評価を一次的には信頼しつつ、一方で、透明性の確保、公益通報者の保護、インシデントの報告義務化等によって外部から安全性の検証を可能にすることを重視する考え方である。

その他、米国の主要ＡＩ開発事業者においては、安全性の取組に積極的な姿勢もみられる。例えば、令和８年９月には、米国大統領と主要ＡＩ開発事業者の代表が、ＡＩの安全性に関する自主的な取組の枠組みを定めた文書に署名している。具体的には、事業者自身がＡＩを安全に開発する責任を負うことを基本とし、サイバーセキュリティや生物・化学等のリスクについて監視するための内部統制や、独立した外部監査・監督等、体制整備を求めている。

(3)　欧州連合（ＥＵ）の動向

ＥＵは、米国州法と比較して、より包括的な規律を設ける「EU AI Act」を令和６年８月に制定し、令和７年２月以降段階的に施行している。同法では、ＡＩシステムを四段階に分類してそれに応じた厳格度の規制を課している一方で、カリフォルニア州法等と同様に、学習に要した計算量が一定の閾値を超える汎用ＡＩモデル等については、システミック・リスクをもたらしうるものとして、前述のリスクの四分類に基づく規律とは別途、モデル評価実施義務等の追加的規律を行っている。

(4)　オーストラリアの動向

オーストラリア政府は、令和８年 11 月に、同国で事業を行おうとするフロンティアＡＩ事業者に対して、セキュリティ対策の実施、ＡＩモデルの独立評価、インシデント発生時の政府への通知、安全性・セキュリティに関する枠組み（Safety and Security Frameworks）の策定・公表等を義務付ける新たな制度を創設する方針を公表している。

また、２(２)で前述の OpenAI 社の事案を受けてオーストラリアで開催された議会における公聴会では、OpenAI 社に加えて、Anthropic 社、Google 社等が、ＡＩエージェントによる侵害事案の報告の義務化を支持する旨を表明した。

---

[10] カリフォルニア州政府「Governor Newsom signs SB 53, advancing California's world-leading artificial intelligence industry」（令和８年９月 29 日）
https://www.gov.ca.gov/2025/09/29/governor-newsom-signs-sb-53-advancing-californias-world-leading-artificial-intelligence-industry/

4. 我が国における政府の取組

(1) ＡＩ法の責務規定

ＡＩ法では、II.１で上述のとおり、基本理念や責務規定を通じて、人工知能関連技術の研究開発及び活用の適正性を確保するための国や事業者の責務を定めている（第３条、第６条、第７条）。これらの規定は法的拘束力を有するものではないものの、国としてＡＩの開発や活用についての適正性を求める趣旨のものである。

(2) 指針の整備及び各省庁におけるガイドライン等

また、ＡＩ法は、国が、人工知能関連技術の研究開発及び活用の適正性を確保するため、国際的な規範の趣旨に即した指針を整備することを定めている（第 13 条）。これを受け、令和７年 12 月 19 日、人工知能戦略本部において「人工知能関連技術の研究開発及び活用の適正性確保に関する指針」（以下「適正性確保指針」という。）が決定された。

適正性確保指針は、ＡＩの研究開発者、提供者及び利用者等の各主体が、ＡＩのライフサイクル全体を通じて取り組むべき事項を示している。主な内容は、リスクの特定・評価及び低減、透明性の確保、安全性の確保、インシデント発生時の対応等である。同指針は法的拘束力を持たないソフトローとして位置付けられ、事業者の自主的な取組を促すことを基本としている。また、各省庁において策定しているＡＩ関係の各種ガイドライン等については、同指針の趣旨と整合するよう、適宜見直し等を行っていくこととしている。（図１）

同指針を踏まえ各省庁で策定しているガイドライン等には、図１に掲載されているようなものが存在する。特に、総務省及び経済産業省において策定しているＡＩ事業者ガイドラインは、ＡＩのリスク及び開発者、提供者、利用者が取り組むべき事項を総体的に整理しており、分野横断的な観点から各主体が留意すべき事項を掲載している。

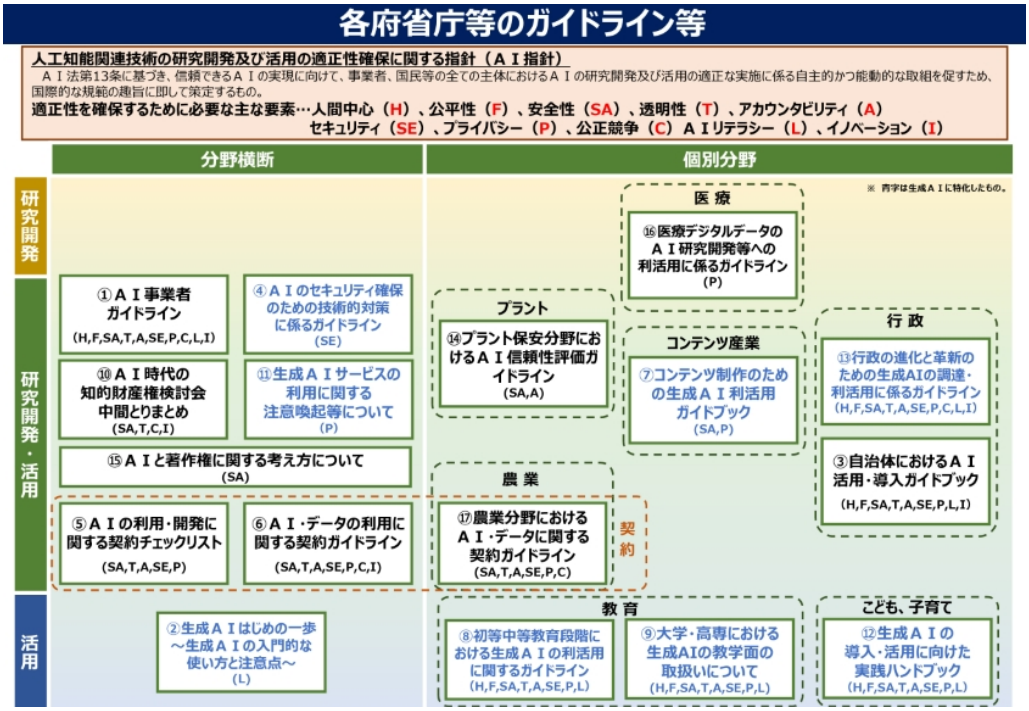


図 1　各省庁等のガイドライン等

(3)　ＡＩ法第 16 条に基づく調査

ＡＩ法は、ＡＩの不適切な方法による活用等に伴って国民の権利利益が侵害される事案等について、国が情報の収集、分析及び調査研究を行い、その結果に基づいて事業者等に対し指導、助言、情報提供その他の必要な措置を行う旨を定めている（第 16 条）。ただし、事業者等が同条に基づく調査等に応じるかどうかは任意であり、事業者等が調査等に応じなかった場合に罰則等が課されるものではない。

政府は、同条に基づき、これまで複数回、国内外で発生したＡＩの不適切な活用等に伴う不適正事案等の情報収集のため、ＡＩの活用事業者に対する任意のヒアリングの実施や質問票による照会等を実施しているところ[11]、一部の事業者からは調査への協力が得られなかった。

(4)　個別の分野における対応

以上のＡＩ法の既存の規定の他、ＡＩに関連する領域であっても、既存の個別の制度が存在する領域においては、ＡＩによるものかどうかに関わらず当該既存制度の範囲内において必要な対応を行っている。

例えば、直近の動向としては、高性能ＡＩによるサイバー空間のリスクの顕在化を受けて、政府では、令和８年 5 月に重要インフラや政府機関のセキュリティ向上のための対策パッケージを「Project YATA-Shield」として取りまとめた。また、同年 11 月には、ＡＩを悪用したとみられるサイバー攻撃が相次いでいることを踏まえ、「サイバーセキュリティ緊急対策パッケージ」をサイバーセキュリティに関する関係閣僚会議において決定するなど、政府全体で対策を推進している。加えて、II.2（1）イ.「ＣＢＲＮＥリスク」で挙げたバイオ分野のリスクに対応するため、「イノベーション政策強化推進のための有識者会議（バイオエコノミー戦略）」の下に、バイオセキュリティ専門家会議を設立し、国内における核酸合成に関するリスク低減対策を検討している。

(5)　ＡＩＳＩ（ＡＩセーフティ・インスティテュート）の取組

ＡＩの安全性に関する評価手法の検討や基準の作成等を担う機関として、独立行政法人情報処理推進機構（ＩＰＡ）に令和６年２月に設置されたＡＩセーフティ・インスティテュート（以下「ＡＩＳＩ」という。）は、関係府省庁や関係機関[12]と連携し、海外の主要ＡＩ企業と協力覚書を締結するなどして、日本政府におけるＡＩの安全性に関する取組の中核的な役割を担っている。

ＡＩＳＩは、「信頼できるＡＩの提供に必要な情報や技術を有し、日本を世界で最もＡＩを開発・活用できる国にする」ことをパーパスとして掲げ、ＡＩシステムの開発者・提供者がＡＩの安全性を評価する際の基本的な考え方と評価観点を整理した「ＡＩセーフティに関する

---

[11] 第１回ワーキンググループ検討会における内閣府の報告による。

[12] 国立研究開発法人情報通信研究機構（ＮＩＣＴ）、国立研究開発法人理化学研究所（ＲＩＫＥＮ）、大学共同利用機関法人情報・システム研究機構　国立情報学研究所（ＮＩＩ）、国立研究開発法人産業技術総合研究所（ＡＩＳＴ）

評価観点ガイド」や、攻撃者の視点からＡＩシステムの脆弱性やリスクを検証するレッドチーミングについて実施体制や手順を示した「ＡＩセーフティに関するレッドチーミング手法ガイド」等のガイドラインを公開するとともに、ＡＩモデルの安全性評価のための評価ツールやデータセットの開発、評価環境の整備を進めており、国内外の事業者や研究機関と協力し、ＡＩモデルの安全性評価を実施し、その結果を公表している。

(6)　国際的な取組

ア　広島ＡＩプロセス

我が国が議長国を務め令和 5 年に開催されたＧ７サミット（主要国首脳会議）においては、「広島ＡＩプロセス」が立ち上げられ、生成ＡＩを含む高度なＡＩを規律する初の国際規範を策定した[13]。令和 6 年には、国際規範の遵守状況をＡＩ関連企業が自主的に確認・報告する「報告枠組み」がＧ７で合意され、令和 7 年に運用が開始された。同報告枠組みは、令和 8 年に、中小企業を含む広範な組織の参加を促すため「報告枠組み 2.0」に改訂されるなどしている。

また、令和 6 年に立ち上げられた「広島ＡＩプロセス・フレンズグループ」の参加国・地域は、当初の 49 から 70 に拡大し、令和 7 年に立ち上げられたフレンズグループの活動を支援するＡＩ関連企業・団体、国際機関等による「パートナーズコミュニティ」の参加組織も、当初の 16 から 53 に増加[14]している。フレンズグループは、今後も定期的に会合を開き、マルチステークホルダーで対話を進めていくことを予定している。

イ　人工知能と人権、民主主義及び法の支配に関する欧州評議会枠組条約

我が国は、ＡＩシステムのライフサイクルにおける活動が人権、民主主義及び法の支配に合致することを確保することを目的とした「人工知能と人権、民主主義及び法の支配に関する欧州評議会枠組条約」について、オブザーバー国として交渉に参加し、令和 7 年 2 月に署名した。同条約は現時点で未発効であるが、危険性・影響の評価及び緩和措置、人権侵害への実効的な救済、独立した監督の仕組みの設置を求めている。

III.制度の在り方

1.　基本的考え方

(1)　ＡＩ法の趣旨を踏まえた新たな制度整備の必要性について

Ⅱ．１のとおり、ＡＩ法制定時の整理としては、基本的には、事業者の自主性を尊重するものとされていたところ、現在においても、ＡＩが経済成長や国民生活の発展の基盤となる技術であることは令和７年のＡＩ法制定時から変わりなく、イノベーション促進とリスクへの

---

[13] 総務省「高度な AI システムを開発する組織向けの広島プロセス国際行動規範」（令和５年 10 月 30 日）
https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document05.pdf

[14] 令和８年 11 月 27 日時点

対応の両立を確保するというＡＩ法の趣旨に照らし、引き続き自主的な取組を促進することを基本とすることが適当である。

ＡＩサービスの内容や提供・利用の態様は多様であり、ＡＩサービスにより生じるリスクもまた多様である。したがって、それぞれのＡＩサービスの特性に応じたリスク対応を講じることが基本であり、全てのＡＩサービスに対して包括的かつ単一の規制を設けることは望ましくない。これらの多様なリスクに対しては、Ⅱ．４で述べたとおり、既に個別の法領域において、その特質に適合的な規律が設けられており、今後も、必要な範囲でそれぞれの法領域に対応する既存の個別法令・ガイドライン等に基づいて対応していくというＡＩ法の原則的な対応枠組みは維持すべきである。

しかしながら、Ⅱ．２で記載のとおり、近時のＡＩ技術の急激な発展やＡＩサービスの利用の広がりの中で、サイバー攻撃やＣＢＲＮＥを始めとするＡＩの悪用のリスクや、予期せぬ挙動により社会全体に重大な悪影響を引き起こすリスクが顕在化しつつある。これらは、ＡＩが人間の能力を拡張し又は代替することで社会全体に重大な悪影響が発生し又は増幅するリスクであるといえ、その特性に応じた新たなリスク対応の在り方が求められる。

そこで、現在顕在化しつつある新たなリスクに対応するため、既存の個別法令・ガイドライン等での対応に加えて、ＡＩを活用した情報処理サービスを提供する者（以下「ＡＩサービス提供者」という。）のうち特にリスクが高い一定の者に対する横断的な規律を新たに設け、リスクの発生・拡大や再発を防止するため適切な措置を講じる必要がある。

(2)　法整備を検討する上で考慮すべき視点

ＡＩのもたらすリスクに対応するための新たな規律を設ける場合でも、イノベーション促進とリスクへの対応の両立を確保するというＡＩ法の趣旨に照らし、我が国の社会全体に重大な悪影響を広く及ぼし得るＡＩサービスを提供している者に限定して、真に必要な範囲での規律とする必要がある。

加えて、具体的な規律を定める際には、ＡＩサービス提供者の負担も考慮し、法的な義務を課す場合には予見可能性、基準の明確性等を高めるとともに、他法に基づく制度との間で報告様式の共通化、既提出資料の援用、信頼性の面で同等と認められる方法で得られた評価結果の受入れ等により、安全性の確保と重複負担の軽減の両立を目指すべきである。

また、高性能ＡＩモデルの開発に係る技術は、膨大な投資を要する競争領域であり、事業者にとっても、企業秘密として社外に開示することがためらわれる情報も多数あると考えられる。したがって、国に報告を行ったことによって事業者に不当に不利益が生じないよう適切な措置を講じるべきである。

さらに、ＡＩサービス提供者の多くはグローバルに事業を展開しており、そのようなＡＩサービス提供者が提供するＡＩのリスクは世界中で同時に顕在化するリスクが高い。我が国において新たな規律を設ける場合でも、諸外国の法令と不必要な重複や矛盾が生じないよう、諸外国の規律との調和・整合も考慮しつつ、法整備を進めるべきである。これは、複数の法域をまたいで発生した同一の事象に対して、収集する情報を可能な範囲で他国と共通化することで、我が国においても効率的で円滑な情報の収集・分析と他国との連携を可能とするとい

う点で、単なる事業者負担への配慮にとどまるものではなく、公益の観点からも望ましいといえる。

(3) 対象となるリスクの整理

ＡＩのもたらすリスクに対応するための新たな規律を設ける場合、Ⅱ．２で上述した、現在顕在化しつつあるＡＩの悪用や予期せぬ挙動によるリスク等を念頭に置く必要があるが、対象となるリスクの整理に際しては、以下の２つの観点から検討することが有用であると考えられる。

① 社会全体に重大な悪影響を広く及ぼす事態を惹起すること

例えばＣＢＲＮＥリスクのように、社会全体に重大な悪影響をもたらすリスクについては、そのリスクが顕在化しないよう予防策を講じることが不可欠であるとともに、当該リスクが顕在化する兆候を可及的速やかに把握してこれを防止し、又は顕在化してしまった場合にはその拡大を防止するための措置を直ちに講じなければならない。

確かに、一部の重大な悪用や予期せぬ挙動に関しては、個別の法令における利用者規制や一定の事業者の事故報告制度が既に整備されている部分もある。しかし、これらの法令は、あくまで不正行為をしようとしている者の能力や被害規模が従前から想定・発生している水準を大幅に超えないことや、危険な物質や設備は主に人によって管理されることを前提として、個別の具体的な行為者を規律することに主眼が置かれていることが通常である。

他方で、近時のＡＩ技術の急激な発展やＡＩサービスの社会における利用の広がりに鑑みると、人間のみでは達成し得ないほど高度又は大規模なサイバー攻撃が発生するリスクその他の未知のリスクが顕在化する事態や、危険な物質や設備の管理業務のシステムに対してＡＩが影響を与え得る状況が発生する事態は十分に想定し得る。加えて、今後、さらにＡＩが高性能化した場合には、ＡＩサービスに予期せぬ挙動が生じた場合の影響もより一層拡大し、社会全体に重大な悪影響を広く及ぼす事態も想定される。

上記のような高性能化したＡＩにより生じ得るリスクについて、当該リスクが顕在化した場合には社会全体に重大な悪影響を広く及ぼす可能性が高いときには、ＡＩの利用者を個別の法令で規律するのみならず、人間の能力を飛躍的に拡張し又は人間に代わって特定の業務を行うＡＩサービス自体の安全性を高めることが不可欠である。その観点からは、当該ＡＩサービスを管理しているＡＩサービス提供者に対して、リスクを低減するための措置を講じさせ、当該リスクが顕在化した場合には政府に報告させる等の規律を課すことが必要かつ有効である。

② 社会全体への悪影響がＡＩによって著しく増幅され又は創出されること

ＡＩサービスの利活用に伴って社会全体に重大なリスクが生じるとしても、当該重大なリスクにかかるＡＩの寄与度が小さい場合には、特にＡＩに着目した規律を設ける必要性が高いとはいえない。この場合には、個別法令との整合性を確保し、規律の重複を回避する

観点からも、専ら既存の個別法令により対応することが望ましいと考えられる。

他方で、ＡＩの寄与度が高い場合については、ＡＩサービスの利活用に伴って社会全体に重大な悪影響が生じないようにする観点からは、ＡＩサービス提供者に対して、ＡＩサービスの提供の場面でリスクの低減措置を実施するよう求めることが有効かつ合理的である。

例えば、前述のとおり、ＡＩを悪用したサイバー攻撃については、高度な専門知識を有するわけではない中高生等が行為者として多く検挙されている。また、ＣＢＲＮＥの中でも特にバイオや爆発物については、ＡＩの支援を受けることで、適切な知識と設備のある者であれば比較的容易に作成できてしまう可能性が示唆されている。これらは、ＡＩによって従前存在した障壁が除かれた結果、リスクが著しく増大し、社会全体に重大な悪影響が生じ得るという意味で、ＡＩが当該リスクの顕在化に大きく寄与している事例であるといえる。

そこで、社会全体に生じる悪影響がＡＩによって著しく増幅され又は創出されるようなリスクについては、ＡＩの利用者を個別の法令で規律するのみならず、ＡＩサービス提供者が当該リスクを低減させるための措置を講じることが有効である。また、当該リスクが顕在化した場合には、政府が適切に対応するために必要な情報を収集することを可能とする必要がある。

(4)　対象となる事業者の整理

規律の対象となる事業者[17]については、以下の二つの観点から特定することが考えられる。

①　ＡＩの性能の観点

上記のようなリスクの整理を踏まえると、規律の対象とする事業者は、社会への重大な悪影響を惹起する上で中核的な役割を果たし得る程度に高性能のＡＩを用いてＡＩサービスを提供する事業者をまずは想定するのが妥当である。ＡＩの性能を判断する際には、諸外国の法令を参考に、学習に要した計算量が一定以上であること等を基準としつつ、ＡＩの性能を適正に評価するための技術的な手法を国際的な協力枠組みも活用しつつ模索していくべきである。その際、評価基準は国際的な評価枠組みとの整合を図りつつ明確化し、事業者が自らの該当性を予見できるものとすべきである。

ただし、使用しているＡＩモデルの性能が高いとしても、例えば特定の用途に限定されているなど、その高い性能が重大なリスクをもたらす態様で利用されていない場合であれば、規律の対象から除外し得ることを検討すべきである。

他方で、上記の規律の対象事業者以外の事業者においては、重大なリスクが比較的少ないと考えられることから規律の対象とはならないものの、社会への重大な悪影響の発生又

[17] 他社が開発するＡＩサービスを自社のサービスに組み込んで活用する場合等、複数の事業者が関係するＡＩサービスの安全性確保を考える上では、他社が開発するＡＩサービスを自社のサービスに組み込んで利用している企業等について、一律に上流の事業者と同様の義務を課すべきではなく、各事業者の役割、情報へのアクセス及びリスクの管理可能性に応じて対応を求めることが合理的であると考えられる。

は拡大を防止するための措置を講じるとともに、問題が発生した場合には政府に緊密に情報共有を図ることが望ましいことは当然であり、上記の規律の対象事業者以外の事業者に対しても、努力義務を課すこと等を検討する余地がある。

② 国内における利用の観点

重大な悪影響を及ぼし得るＡＩサービスについて我が国の法令において一定の規律を設けるとしても、あくまで国内法である以上、国内に利用者が全く見込まれず我が国と何ら関係性を有していないＡＩサービスについてまで規制を適用することは避けるべきである。

そこで、対象となる事業者を限定するにあたっては、我が国に悪影響を及ぼすリスクを有すると評価し得る程度に国内で利用されていること等、その利用や事業の態様を考慮するべきである。

以上を踏まえると、少なくとも現在においては、規律の対象は、高性能であってかつ日本国内における利用状況等の国内への影響を測る要件を充足するＡＩサービス等、我が国の社会全体に重大な悪影響を広く及ぼし得るＡＩサービスを提供する事業者に限定することが妥当である。

なお、オープンウェイトモデル（ＡＩモデルの開発者が、ＡＩモデルの中核的なデータをインターネット上に公開し、個人を含め、幅広い利用者が自由にダウンロードし、ＡＩモデルを活用することができる利用形態）については、当該モデルをダウンロードして情報処理を行うＡＩサービスを提供する者は、当該サービスにより生じ得るリスクの程度によっては、クローズドなＡＩサービスと同様、規律の対象となるＡＩサービス提供者に該当し得る。

他方で、オープンウェイトモデルを活用するＡＩサービス提供者は、当該モデルの開発過程や潜在的な能力を十分に把握できない場合がある。このため、ＡＩＳＩにおいて主要なオープンウェイトモデルの安全性評価の結果を共有するなど、当該ＡＩサービス提供者が必要な措置を講じるための支援を行うことが望ましい。また、オープンウェイトモデルはＡＩシステムにおける中核的な部分を一般の者が取得可能な状態で流通させるものであり、その開発及び流通に起因する課題に関しては、今後のＡＩ技術の発展や利用状況の動向を踏まえつつ、更なる対応の必要性についても引き続き検討することが望ましい。

2. ＡＩサービスの安全性確保に着目した法令の具体的な内容

(1) ＡＩの安全を確保するための措置を講じる義務

Ⅲ．１．(4)で整理した規律の対象となるＡＩサービス提供者（以下「対象事業者」という。）には、その提供するサービスに関連して、社会全体に重大な悪影響が生じるリスクが高い。そこで、ＡＩサービスの安全性を確保するために対象事業者に対し、特に人の生命、身体、財産及び権利利益に対する重大な悪影響の発生又はそのおそれについて適時に評価し、その発生及び拡大を防止するために必要な措置（以下「安全確保措置」という。）を講じることを法律上義務付けることが妥当である。

安全確保措置の内容としては、主にはサイバー攻撃やＣＢＲＮＥリスクを惹起するような出力をさせない技術的措置を講じることが考えられるが、これに限らず、適切なＡＩサービスの提供体制を講じる組織的措置、独立した第三者による検証を受ける等のモニタリングなど、様々な措置が含まれ得る。

ただし、ＡＩ技術は急速に発展しており、想定される具体的なリスクは変化し続けている。それに伴って、そのリスクに対応する有効な安全確保措置の内容も流動的な性質を有しているとともに、対象事業者が提供するＡＩサービスの性質に応じて何が必要かつ適切な措置であるかは異なり得る。このため、措置の具体的な内容を法律で固定的に定めることは適当でなく、国が定める指針において方向性を示して対象事業者に一定の裁量を残して、柔軟に変化に対応していけるアプローチが妥当である。他方で、裁量を残しつつも透明性を確保する観点からは、安全確保措置の実施状況について、事業者に公表を義務化することもあり得、その必要性についても検討の余地がある。

(2)　重大な悪影響が発生した場合に国に報告する義務

ＡＩのリスクに的確に対応するためには、国が対象事業者から必要な情報を確実に収集できることが不可欠である。しかし、現行のＡＩ法第 16 条に基づく調査は任意の協力を前提としており、Ⅱ．４．(3)で前述のとおり、一部の事業者から協力が得られなかった事例も存在する。こうした状況を踏まえ、対象事業者が、自らが任意に行う調査等により、そのサービスに関連する重大な悪影響の発生又は具体的なリスクを知った場合には、当該重大な悪影響に係る内容について、国に報告することを義務付けるべきである。この際、報告の対象となる場合や報告の期限は、諸外国の法制度の例も参考にしつつ、具体的に検討すべきである。

(3)　国からの報告の求めに応じる義務

(2)で述べたような対象事業者自らによる報告によらず、国が自ら行う情報収集によって重大な悪影響の発生を認知した場合や、平時であっても安全確保措置の実施状況について国が主導して対象事業者に報告を求め、場合によっては現地で立入検査確認する必要がある場合もあり得る。

そこで、国は、安全確保措置を講じる義務の履行を担保するために必要な限度において、対象事業者に対し、重大な悪影響の発生又はそのおそれその他業務に関する報告を求めることができるようにし、立入検査についての規定の要否についても検討すべきである。ただし、立入検査をするとしても、報告徴収によっては必要な情報が得られない場合に限るなど、要件を限定的に定めるとともに、営業秘密の保護に十分配慮すべきである。

(4)　国の勧告・命令権限

上記のような安全確保措置実施義務及び報告義務の実効性を確保するため、対象事業者がこれらの義務に違反している場合に、違反の是正に必要な措置を講じるよう勧告する権限を国に付与することが妥当である。

また、勧告を受けた対象事業者が、正当な理由なく勧告に係る措置を講じない場合には、そ

の措置を講じるよう国が命令し、違反した場合に罰則を科すことで履行を担保する必要があるとともに、ＡＩがもたらす重大なリスクの顕在化が急迫している場合には、勧告を経ずに直ちに必要な措置を命令する旨の緊急規定を設けることも検討すべきである。

さらに、これらの権限の発動に伴い、国民に適切に情報発信を行う観点から、事実の公表に関する規定を設けることも検討すべきである。ただし、公表によって悪用を招くおそれのある情報や営業秘密が開示されることのないよう、配意することも必要である。

(5) 罰則

法令の実効性を確保するため、対象事業者が報告義務や命令に違反した場合を念頭に、罰則を設けることが妥当である。

ただし、ＡＩの研究開発や利用に係る作為や不作為に対する直接的な罰則ではなく、あくまで、行政からの報告の求めや措置の命令に従わなかった場合に罰則を科す、間接的な罰則規定を設けることが妥当である。

罰則の水準については、対象事業者が報告義務や命令に違反することを防止するうえで実効性を持つものとなるよう、国内の他のデジタル分野において規律を設けている法令も参考にしつつ、検討すべきである。

(6) ＡＩＳＩ（ＡＩセーフティ・インスティテュート）の役割

対象事業者の把握や、報告徴収、勧告・命令及び公表を適切に行うためには、ＡＩの性能やリスク、安全確保措置の有効性等について、高度に技術的な判断が必要となる。このため、国がＡＩＳＩに対し、適宜、技術的な調査を行わせることができるようにするとともに、その旨をＡＩＳＩが設置されている独立行政法人情報処理推進機構の固有の業務として、同機構の業務の範囲等を規定する情報処理の推進に関する法律（昭和 45 年法律第 90 号）で規定することが妥当である。あわせて、ＡＩＳＩが法律上の新たな役割を適切に果たすとともに、他国に依存しないＡＩモデル評価体制を確立し、日本国内での安全・安心なＡＩ活用を担保することができるよう、ＡＩＳＩの評価能力の確立や人員・予算等の体制強化を図るべきである。

一方で、現状においては、ＡＩＳＩが中立的な立場にあるからこそ、関係するステークホルダーと協力関係を構築し、ＡＩの安全性に関する取組を推進している側面も鑑みると、法規律の中核的な役割を担う機関として位置付けることによる影響も考慮しつつ、政治的な中立性、客観性を高めることに取り組むべきである。このような観点も含め、変化する情勢に合わせ、ＡＩＳＩの在り方、法律上の位置づけ、国との役割の分担については、不断の検討を継続していくべきである。

(7) 技術動向に対応する法令の在り方

本ＷＧでは主にサイバー攻撃やＣＢＲＮＥ、予期せぬ挙動等によるリスクを念頭に置いていた。これは、現在既に顕在化しており、社会に重大な悪影響を広く及ぼし得るＡＩのリスクへの対応について優先的に検討する必要があったためである。もっとも、差別、偽情報など従前から指摘されているリスクや、ＲＳＩ（再帰的自己改善）といった近時特に指摘されている

リスク等であって本ＷＧでは議論を尽くしていないリスクについても、それがＡＩによって著しく増幅され又は創出される社会全体に対する重大な悪影響といえるものであれば、柔軟に規制の対象となり得るか検討する必要がある。

ＡＩ技術は日進月歩で進展しており、その性能やリスクを測定する方法も急速に陳腐化する可能性が高い。したがって、規律の対象とすべきＡＩを活用した情報処理サービス事業者の具体的な要件や基準等についても、下位法令に適宜委ねつつ、ＡＩの開発・提供・利用の実態の変化を踏まえ、また、専門家及び事業者の意見を聴きながら不断に見直していく必要がある。

加えて、現下のＡＩ産業の状況を踏まえれば、主として日本国内にサービスを提供するＡＩサービス提供者に対し規律を課すことが妥当であると考えられるが、今後の情勢によっては、規律対象を拡大していくことも検討の余地がある。ただし、今後の情勢によってリスクの類型や規律対象の見直しを行う場合にも、既存の法令等では対応できない理由とその必要性を検証した上で検討を行うべきである。

3. ＡＩサービスの安全性確保に着目した法令以外の方策による対応の在り方

ＡＩサービスの安全性確保に着目した法令の整備と相互に補完する形で、既存の法令やガイドライン等のソフトローを活用した取組を一層強化していくことが必要である。具体的には、以下の取組が考えられる。

(1) 「人工知能関連技術の研究開発及び活用の適正性確保に関する指針」の見直し

ＡＩの安全確保措置等の法律上の義務はこれまで述べてきたとおりであるが、義務規定の対象に含まれない事業者も含めて、ＡＩの様々なリスクについては、ＡＩ法第３条、第６条及び第７条に基づき、適正性の確保のために必要な措置を講じることが求められる。

また、国は、ＡＩに関わる全ての事業者や国民等の主体に対し、引き続き、ＡＩ法第13条に基づき、適正性確保指針及び適正性確保指針を踏まえ各省庁で策定しているガイドライン等により対応を促していくことが必要である。

特に、令和８年に入ってＡＩが急速に高性能化し、エージェント型ＡＩやＲＳＩといった従来は想定されていなかったリスクが現実化しつつある。このような昨今の技術動向にも鑑みて、指針において適正性確保の枠組みを示しつつ、ＡＩ法に基づく機動的な調査研究を行ってリスクの早期把握に努めるべきである。

(2) 各種ガイドラインの改定・策定

総務省及び経済産業省が作成するＡＩ事業者ガイドラインにおいては、ＡＩのリスク及び開発者、提供者、利用者が取り組むべき事項を総体的に整理しているところ、エージェント型ＡＩも含む昨今の技術動向を反映するとともに、ＡＩ法制定以降に顕在化している様々なリスクについて、事業者が適切に対応を講じることができるよう、内容を改定することが望ましい。

また、各省庁にて策定しているガイドライン等についても、指針の内容も踏まえ、必要な見直しを図ることが望ましい。

(3)　官民対話の推進

ＡＩの安全性の確保は高度に技術的な側面もあり、国が一方的に事業者に対して規律を課す従来型の規制法的考え方はなじまないと考えられる。むしろ、国として法律で一定の監督権限を確保しつつ、規律の実際の運用にあたっては、高度な専門的知識を持つ事業者と適時かつ丁寧な対話を重ね、機動的な情報共有の場を持つことが重要である。

そのような対話の場として、国内外の事業者と国で機動的に意見交換を行う会合の場を設けることは検討に値する。

(4)　国際連携の推進

主要なＡＩサービスの提供は一国にとどまるものではなく世界中で展開されており、ＡＩサービスで生じた問題はグローバルに影響を及ぼす可能性がある。したがって、我が国のみならず他国で発生した事象やＡＩサービスの利用状況等も含めて、適時に情報を収集し分析を行うことが、ＡＩサービスによって発生する重大な悪影響の発生及び拡大の予防及び発生時の迅速な対応のために有益である。我が国としては、引き続き、広島ＡＩプロセスを推進しつつ、ＡＩの安全性に関する国際的な議論に積極的に参加していくとともに、同志国等との間で、ＡＩの安全性に関する協力関係を構築していくべきである。

また、Ⅲ．２で示した制度を整備した場合には、日本における制度設計や安全性評価の知見を、グローバルサウスを含む国際社会に発信することで、国際的なルール形成に貢献していくことも重要である。

(5)　国の調達・利活用の在り方

政府機関は、行政の効率化・高度化のためにＡＩの導入を進めており、自らがＡＩの大規模な利用者となっている。政府機関が利用するＡＩには、機微な情報を含む行政情報が入力され、行政事務の遂行に組み込まれていく。このため、ＡＩの悪用や予期せぬ挙動等のリスクが生じた場合、行政機能や国民の権利利益に重大な悪影響が及ぶおそれがある。

このうち、各府省庁における生成ＡＩの調達・利活用については、政府は、「行政の進化と革新のための生成ＡＩの調達・利活用に係るガイドライン」（令和８年６月デジタル社会推進会議幹事会決定）を策定し、責任体制の整備やリスク管理を求めているところ、次の観点も踏まえ、政府機関における生成ＡＩの調達・利活用の在り方について、さらに検討を進める必要がある。

○　ＡＩエージェントに付与するアクセス権限やタスクの実行権限の範囲及びその管理の在り方

○　取り扱われる情報が保存される国・地域の法制度等に起因するリスクに応じた利用の促進

国が調達を通じて安全性の高い生成ＡＩを積極的に選択することは、政府機関のリスク低減にとどまらず、安全性の確保に取り組む事業者を後押しし、信頼できるＡＩの市場の形成を促すことにもつながる。こうした観点から、ＡＩＳＩによる安全性評価の知見の活用や、

Ⅲ．２で示した法令及び指針に基づく安全確保措置の実施状況を調達における評価の考慮要素とすることについても、検討すべきである。

IV. 終わりに

1. 今後の制度の検討の在り方

以上のとおり、令和７年のＡＩ法制定以降の情勢を踏まえれば、原則として、既存の法令やガイドライン等のソフトローを適切に組み合わせＡＩのリスクに対応していくという、これまでの我が国における基本的な考え方は維持しつつ、既存の法令やソフトローのみでは対応が困難な重大なリスクについては、早急に法整備を行い、一部のＡＩサービス提供者に対し法律上の義務を課すことで、ＡＩの安全を確保していくことが必要であると考えられる。

ＡＩの安全性・信頼性はＡＩの利活用を推進していくにあたっての前提であり、今後、ＡＩによる経済成長を実現していくにあたっても、適切にリスクに対応していくことは不可欠である。政府に対しては、本とりまとめを踏まえ、引き続きイノベーション促進とリスクへの対応の両立を確保しつつ、ＡＩの適正な研究開発及び活用を推進していくことを期待したい。

2. ＡＩに起因するリスク対応の在り方

これまでに述べてきたリスクの高まりを踏まえると、一部の高性能なＡＩサービス提供者に対し、より確実に安全対策を講じるよう求めることが適当であり、法律上の義務を課すことが必要と考えられる。あわせて、その他の各種施策についても、より一層推進していくことが重要である。例えばサイバーや、ＣＢＲＮＥ等の分野でも、防御側での対策を強化するなど、多角的に取組を推進することが重要である。

今後、ＡＩはより一層の高性能化が進み、ＡＩの利活用はますます社会に浸透していくことが予想される。本ＷＧの今般のとりまとめでは主にＡＩを悪用したサイバー攻撃、ＣＢＲＮＥ等のリスクの顕在化、ＡＩの予期せぬ挙動による重大な事故のリスクを念頭に制度の在り方を検討したが、今後は、ＡＩがもたらす現時点では予想が困難なリスクについて、新たに対応を行う必要が出てくる場合も想定される。また、ロボティクス、自動運転、プラント制御など、現実世界の情報を統合し、物理的タスクを遂行するフィジカルＡＩについては、用途によっては誤作動や予期せぬ挙動が生命・身体、設備、重要インフラに不可逆的な被害をもたらし得る。

政府においては、これらのリスクに対しても、分散するＡＩ関連の部局が相互に連携しあい、ＡＩ以外の特定の行政分野を所管する他の部局とも連携しながら、引き続き機動的かつ的確に対応を行っていくべきである。

制度の在り方検討ワーキンググループ　構成員名簿

座　長	宍戸　常寿	東京大学大学院法学政治学研究科 教授
	生貝　直人	一橋大学大学院法学研究科 教授
	岡崎　直観	東京科学大学情報理工学院 教授
	岡田　　淳	森・濱田松本法律事務所外国法共同事業 弁護士
	岡田　陽介	株式会社 ABEJA 代表取締役ＣＥＯ／日本ディープラーニング協会 副理事長
	殿村　桂司	長島・大野・常松法律事務所 弁護士
	永沼　美保	一般社団法人日本経済団体連合会デジタルエコノミー推進委員会 国際戦略ＷＧ主査／日本電気株式会社 主席プロフェッショナル

制度の在り方検討ワーキンググループ　開催実績及び議事概要

令和８年10月28日	第１回 ・本ワーキンググループ開催に係る背景と目的（内閣府） ・ＡＩのリスクを巡る情勢について（内閣府） ・ＡＩのリスク対応に係る現行の政府の取組について（内閣府） ・ＡＩ技術の安全性とＡＩＳＩの取組について（ＡＩＳＩ）
令和８年11月11日	第２回 ・事業者ヒアリング ・第１回の議論を踏まえた論点整理について（内閣府）
令和８年11月18日	第３回（書面開催） ・とりまとめ骨子（案）について
令和８年11月27日	第４回 ・とりまとめ（案）について