

第2節 AIに関する各国の対応

こうした生成AIをはじめとするAIの急速な普及のなかで生じた倫理的・社会的な課題に対処するためには、国内のみならず、諸外国と協調した取組が必要である。

1 国際的な議論の動向

1 広島AIプロセス

AIについての倫理的・社会的課題に対する議論は2015年頃から活発化しており、我が国は、早期からG7/G20や経済協力開発機構（以下「OECD」という。）等における議論を先導し、AI原則の策定に重要な役割を果たしてきた。2016年4月に高松で開催されたG7情報通信大臣会合において、日本からAIの開発原則に関する議論が提案され、その後OECDで合意されたAI原則が2019年5月に公開されたことを受けて、同年6月のG20首脳会合にて、「G20 AI原則」が合意された^{*1}。2019～2020年には、AI原則については国際的なコンセンサスが形成されつつあり、同原則を社会に実装するための具体的な制度や規律の策定に関する議論に移行している。更には、2022年の生成AIの急速な普及により、G7等の国際協調の場においても、また各国においても、AIガバナンスの議論が活発化している。

2023年4月、群馬県高崎市でG7群馬高崎デジタル・技術大臣会合が開催され、生成AIの急速な普及と進展を背景に、「責任あるAIとAIガバナンスの推進」などについて議論が交わされた。同会合では、G7のメンバー間で異なる、AIガバナンスの枠組み間の相互運用性の重要性が確認され、「責任あるAIとAIガバナンスの推進」、「安全で強靱性のあるデジタルインフラ」、「自由でオープンなインターネットの維持・推進」等の6つのテーマからなる閣僚宣言が取りまとめられた。同宣言はその後、5月に広島で開催された主要7か国首脳会議（G7広島サミット）における議論に反映され、当該サミットの首脳コミュニケ（宣言）において、生成AIに関する議論のための広島AIプロセスの創設が指示された。具体的には、OECDやGPAI（後述）等の関係機関と協力し、G7の作業部会にて調査・検討を進めることとなった。

2023年9月には、7月～8月にOECDが起草したレポートや、生成AI等を含む高度なAIシステムの開発に関して議論すべく閣僚級会合が開催され、透明性、偽情報、知的財産権、プライバシーと個人情報保護等が優先課題であることが確認された。その後10月30日に「広島AIプロセスに関するG7首脳声明」^{*2}が発出され、まずは高度なAIシステムの開発者を対象とした国際指針と行動規範が公表された。更に同年12月には、AIに関するプロジェクトベースの協力を含む広島AIプロセス包括的政策枠組みや、広島AIプロセスを前進させるための作業計画が発表されている。

2 OECD/GPAI/UNESCOの動き

A OECD

OECD、GPAI、UNESCO等、多くの国際機関もグローバルな観点からAIガバナンス制度の

*1 経済産業省 AI原則の実践の在り方に関する検討会、「我が国のAIガバナンスの在り方ver. 1.1」, <https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_1.pdf> (2024/3/4参照)

*2 外務省, 「広島AIプロセスに関するG7首脳声明」, <https://www.mofa.go.jp/mofaj/ecm/ec/page5_000483.html> (2024/3/4参照)

検討を進めている。2019年5月にOECDのAI原則が公開されて以降、各種OECDレポートの発表やプロジェクトの推進等、G7との連携の下、積極的な活動が行われている。また、OECD、GPAI、UNESCOの3機関は、2023年9月に「生成AI時代の信頼に関するグローバルチャレンジ (Global Challenge to Build Trust in the Age of Generative AI)」*3を公表し、G7の包括枠組みを踏まえ、偽情報やディープフェイク等による社会的リスクに対し、イノベティブな解決策を進めるグローバルな連携プロジェクトを推進している。

2024年5月に開催されたOECD閣僚理事会では、生成AIに関するサイドイベント「安全、安心で信頼できるAIに向けて：包摂的なグローバルAIガバナンスの促進」において、岸田総理大臣から49か国・地域の参加を得て広島AIプロセスの精神に賛同する国々の自発的な枠組みである「広島AIプロセス フレンズグループ」*4を立ち上げることを発表した。

イ GPAI

「AIに関するグローバルパートナーシップ (Global Partnership on AI)」(以下「GPAI」という。)は、2020年、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に、OECDとG7の共同声明により創設された。同組織は、OECDが事務局を務め、価値観を共有する政府、国際機関、産業界、有識者等からなる官民間連携組織で、現在29か国が参加している。GPAIには、「責任あるAI」、「データ・ガバナンス」、「仕事の未来」、「イノベーションと商業化」という4つの研究部会が設置されており、専門家による議論と実践的な調査が実施されている。

GPAIの年次サミットである「GPAIサミット2023」においては、新たなGPAI専門家支援センターである、GPAI東京専門家支援センターの立ち上げが承認された。同センターでは、生成AIに関する調査・分析等のプロジェクトを先行的に実施する予定となっている。

ウ UNESCO

国連教育科学文化機関 (UNESCO) も、2021年にAIの倫理に関する勧告「UNESCO Recommendation on the Ethics of Artificial Intelligence」*5を採択し、各国における取組を支援している。2023年9月には、教育・研究に関する初の生成AIのグローバルガイダンスである「教育・研究分野における生成AIのガイダンス (Guidance for generative AI in education and research)」*6を公表し、生成AIの定義や説明、倫理的及び政策的な論点と教育分野への示唆、規制の検討に必要なステップ、カリキュラムデザインや学習等について紹介している。ほとんどの生成AIが主として大人向けに設計されていることから、教育現場での使用は13歳以上に制限すべきと提案し、各国政府には、データのプライバシー保護を含む適切な規制や教員研修等を求めている。

③ AI安全性サミット

2023年5月、OpenAIは、今後10年以内に人間の専門家のスキルレベルを超えるAIシステムが実現する可能性があるとして発表した。同社はこれを「フロンティアAI (Frontier AI)」と命名し、

*3 Global Challenge partners, "Global Challenge to Build Trust in the Age of Generative AI", <<https://globalchallenge.ai/>> (2024/3/21 参照)

*4 https://www.kantei.go.jp/jp/101_kishida/statement/2024/0502speech2.html

*5 UNESCO, "Recommendation on the Ethics of Artificial Intelligence", <<https://unesdoc.unesco.org/ark:/48223/pf0000381137>> (2024/3/13 参照)

*6 UNESCO, "Guidance for generative AI in education and research", <<https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>> (2024/3/13 参照)

核エネルギーや合成生物学等の人類の存在上のリスクに鑑みて、事後的対応ではなく国際的な規制を検討すべきとした。これを受けてスナク英国首相は、2023年11月1日～2日に英国ブレッチリーにて、「AI安全性サミット」*7を開催した。従来の人権や公平性といった「AI倫理」を超えて、AIによる「深刻且つ破滅的な危害」の防止を視野に入れた「AIの安全性」について議論されたことが特徴的である。

本サミットの成果文書として「ブレッチリー宣言」*8が採択された。また、英国はAIセーフティ・インスティテュートを設置することについても決定した。

2024年5月21日～22日には、韓国・英国共催により「AIソウル・サミット」が開催された(21日の首脳セッションはオンライン開催、22日の閣僚セッションはソウルで対面開催)。AI安全性の議論を深めるとともに、AI開発におけるイノベーション促進及びAIの恩恵の公平な享受について議論が行われ、首脳級の成果文書として「安全、革新的で包摂的なAIのためのソウル宣言」及び付録「AI安全性の科学に関する国際協力に向けたソウル意図表明」、閣僚級の成果文書として「安全、革新的で包摂的なAIの発展のためのソウル閣僚声明」が採択された。今後、2025年2月にフランスにて次回会合が開催される予定となっている。

4 国際連合の動向

前項のとおり、フロンティアAIに対する国際的なガバナンス体制への関心の高まりを受けて、2023年7月の国連安全保障理事会においては、英国主導でAIに関する議論が行われた。グテーレス国連事務総長は同年10月に、事務総長の諮問機関として、AIハイレベル諮問機関を立ち上げ、日本人の構成員も参加している。また、2024年3月21日、国連総会において、日本も共同提案国である、「持続可能な開発のための安全、安心で信頼できるAIシステムに係る機会確保に関する決議」*9をコンセンサスで採択し、同決議案は、安全、安心で信頼できるAIに関する初めての国連総会決議となった。同決議案は「持続可能な開発のための2030アジェンダ」の達成に向けた進捗を加速し、デジタルディバイドを解消するため、安全、安心で信頼できるAIを促進しており、加盟国に対し、安全、安心で信頼できるAIに関連する規制・ガバナンスアプローチの策定・支持を推奨している。さらに、加盟国及びステークホルダーに対し、AI設計・開発中のリスク特定・評価・軽減のためのイノベーション促進や、データ保全のためのリスク管理メカニズムの策定・実施・公表等の手段を通じて、AIシステムが世界の課題に対応できるための環境を整備するよう推奨している。また、AIシステムのライフサイクルを通じて、人権及び基本的自由が尊重され、保護され、促進されるべきことを強調している。

同決議案は、AIの国際ルールづくりに向け、広島AIプロセスをはじめ、G7やG20、OECD等で進めてきた議論を反映したものであり、国連総会決議には国際法上の拘束力はないものの、コンセンサスで加盟国が採択したということから、国際社会の総意としての政治的な重みを持つものである。

*7 GOV.UK, "About the AI Safety Summit 2023", <<https://www.gov.uk/government/topical-events/ai-safety-summit-2023/about>> (2024/3/12参照)

*8 GOV.UK, "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023", <<https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>> (2024/3/12参照)

*9 United Nations General Assembly, A/78/L.4 <<https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=0e5FKI9eh5r1MmYpD3&fe=true>> (2024/3/22参照)

2 各国における法規制・ガイドライン等の整備動向

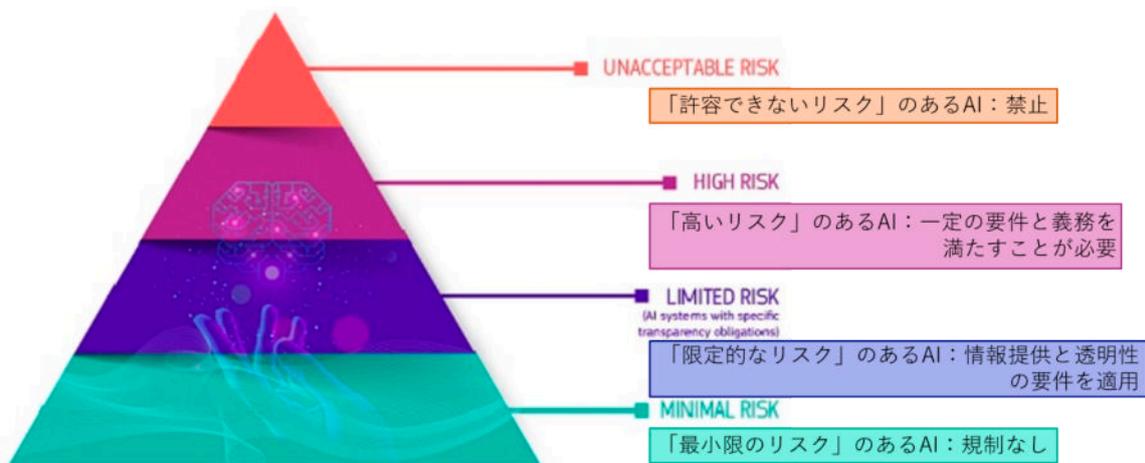
現在、AIに関する法制度や国際標準に関する議論が世界各国で活発に行われており、2023年はEUのAI法の欧州議会での採択、米国のAIの安全性に係る大統領令、日本のAI関連事業者向けのガイドライン案の公表など、AI政策にとっては大きな節目となる年となった。それぞれの国、地域におけるAIに関する規制の動きを見ると、生成AIに対する急速な関心の高まりを受けて、各国・地域ではそれまで検討してきたガバナンス制度の見直しが求められている。進化が速い技術に関する規制の整備においては、各国政府が主導しつつも、AI事業者側の自主的な取組も必要であり、官民両輪で進められているところである。

1 欧州連合 (EU)

域内発のビッグテック企業が無い欧州は、他の地域に先駆けて最も厳しい規制を志向し、2020年からAIの規制に関する議論を続けてきた。2024年5月21日には、欧州市場でAIシステムを開発・提供・利用する事業者を対象とする、法的拘束力を持つ世界初の包括的なAI規制法と位置付けられるAI法 (AI Act)^{*10} (以下「AI法」という。) が成立した。AIの規制に関する包括的な法律の成立は主要国・地域で初めてとされており、今後段階的に適用が開始され、2026年頃には本格的に適用される見込みである。

AI法は、リスクに応じて規制内容を変える「リスクベースアプローチ」という方針に基づいている。規制対象を、①許容できないリスク、②高いリスク、③限定的なリスク、④最小限のリスク、という4段階のリスクレベルのAIアプリケーション及びシステムに分類し、それぞれに対して異なる規制を課すこととしており^{*11}、上記の規制に違反した事業者には、最も重い違反の場合、最高で3,500万ユーロ (約56億円) の罰金、あるいは年間売上高の7%の制裁金が科される可能性がある^{*12} (図表 I-4-2-1)。

図表 I-4-2-1 AI法におけるリスクベースアプローチ



(出典) European Commission (2024)^{*13}を基に作成

*10 European Commission, "AI Act", <<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>> (2024/3/2 参照)

*11 European Parliament, "Artificial intelligence act", <https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI%282021%29698792_EN.pdf> (2024/3/12 参照)

*12 「EUがAI開発・運用を法で規制…学習データの著作権保護、違反事業者に制裁金56億円」, 『読売速報ニュース』2024年3月13日号

*13 European Commission, "AI Act", <<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>> (2024/3/15 参照)

2 米国

ビッグテック企業を多く保有する米国は、自国の企業保護に力を入れ、政府による規制よりも民間での自主的な対応を優先し、企業の取組に任せつつ必要の場合に政府が規制をかけるという立場をとってきた^{*14}。

民間側の取組として2023年7月、AI開発で先行する7社（Google、Meta PlatformsやOpenAI等）^{*15}がAIの安全な開発のための自主的な取組を約束したこと、更に9月には新たな8社（IBM、Adobe、NVIDIA等）^{*16}がそれに合意したことを米国政府が発表した^{*17}。各社は、自主的なコミットメント（Voluntary Commitments）として、安全性、セキュリティ、信頼性の3つの観点から原則を掲げている^{*18}。

ホワイトハウスは、強制力のある規制が導入されるまで、各社が上記の取組を続けるとしていたが、その3か月後となる2023年10月30日、バイデン大統領は、「安全・安心・信頼できるAIの開発と利用に関する大統領令（Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence）」^{*19}を発表した。対象とするAIの問題については、従来の倫理的観点から、安全保障問題に範囲を拡充しており、対象となる事業者はビッグテック企業に限らず、バイオテクノロジー企業等、国家の安全保障や経済に影響を及ぼす可能性のあるサービスや製品を取り扱う企業も含まれる。その内容については、AIに関する新たな安全性評価、公平性と公民権に関するガイダンス、またAIが労働市場に与える影響に関する調査等を義務付けるものであり^{*20}、AIの安全性とセキュリティのための新しい基準、米国民のプライバシー保護、公平性と公民権の推進等をその主要な構成要素としている^{*21}。

大統領令の発表に引き続き、同年11月には、ハリス副大統領が、先述の英国AI安全性サミットにて「安全で責任あるAI利用の新イニシアチブ（New U.S. Initiatives to Advance the Safe and Responsible Use of Artificial Intelligence）」^{*22}を発表し、その中で、大統領令の内容を具体化するべく、「米国AI安全研究所（AI Safety Institute）」（以下「US AISI」という。）を設置するとした。US AISIは、国立標準技術研究所（National Institute of Standards and Technology：NIST）内に設置され、危険な機能を評価及び軽減するためのガイドライン、ツール、ベンチマーク、ベスト プラクティスを作成し、AIリスクを特定して軽減するためのレッド

*14 「AI法制、産官学で世界と議論を」 専門家に聞く」、『日本経済新聞電子版』2024年1月1日号

*15 Amazon、Anthropic、Google、Inflection、Meta Platforms、Microsoft、OpenAI

*16 Adobe、Cohere、IBM、NVIDIA、Palantir、Salesforce、Scale AI、Stability

*17 The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI", <<https://www.whitehouse.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>> (2024/3/8参照)

*18 ①システム公開前の安全性確保：各社は、AIシステムをリリースする前に安全性のテストを行う。またAIのリスク管理に関する情報を、産業界、政府、市民社会、学术界と広く共有する。②セキュリティを確保したシステムの構築：各社は、独自のモデルや公開前のモデルの重要性を保護するために、セキュリティの確保では、サイバーセキュリティ対策やAI開発に係る知的財産の保護等を行う。また、第三者によるAIシステムの脆弱性の発見と報告を促進する。③国民の信頼の獲得：各社は、AIが生成したコンテンツであることを利用者が知ることができるよう、電子透かしの技術等、堅牢な技術を開発する。また、AIシステムの能力、限界、適切／不適切な使用領域を公表する。

The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI", <<https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>> (2024/3/8参照)

*19 The White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence", <<https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>> (2024/3/4参照)

*20 「AIガバナンスを考える (5) 社会・文化的背景で異なる対応」、『日本経済新聞』2024年2月8日付朝刊

*21 The White House, "FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence", <<https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>> (2024/3/10参照)

*22 The White House, "FACT SHEET: Vice President Harris Announces New U.S. Initiatives to Advance the Safe and Responsible Use of Artificial Intelligence", <<https://www.whitehouse.gov/briefing-room/statements-releases/2023/11/01/fact-sheet-vice-president-harris-announces-new-u-s-initiatives-to-advance-the-safe-and-responsible-use-of-artificial-intelligence/>> (2024/3/10参照)

チームを含む評価を実施する。また、人間が作成したコンテンツの認証、AIが生成したコンテンツの電子透かし、有害なアルゴリズムによる差別の特定と軽減や透明性の確保、プライバシー保護の導入等に係る技術的なガイダンスを開発する予定である。英国のAI安全研究所を含む国際的な同業機関との情報共有や研究協力、更には市民社会、学界、産業界の外部専門家との提携も可能となる。

一方、連邦議会でも連邦レベルでのAI規制に関する法案が議論されている。2023年6月には、上院が、AIの急速な進歩に連邦議会が対応するための包括的な枠組みである「安全なイノベーション枠組み (SAFE Innovation Framework)」を提唱し、同年12月までに産業界の代表や有識者を招いたテーマ別のフォーラムを9回にわたって開催した^{*23}。他方の下院は、2024年2月、AIに関する超党派のタスクフォースを設立すると発表し、AI政策の指針となる原則や政策提言を含む包括的な報告書を作成する予定となっている^{*24}。上下両院では、選挙等の個別分野でのAI利用を規制する法案が複数提出されているものの、未だ議会を通過したものは無い。2024年秋に大統領選挙を控える米国では、生成AIの普及に伴うディープフェイクによる情報操作等の課題に直面し、AIの規制に関する議論が益々活発化するものと予想される。

3 英国

英国は米国と中国に次いでAI研究が盛んな国とされており、AI分野への民間投資額においても、シンガポールの躍進により2023年に初めて4位に転落したものの、2019年以来、米国・中国に次いで世界3位を保ってきた^{*25}。現スナク政権は、法的拘束力のあるAI規制には消極的で、安全に配慮しながらAIシステムの開発を促し、経済成長に繋がりたいとする考えから、当面はEUのAI法のような厳格な規制を新たに整備せず、既存の枠組みで柔軟に対処する方針を表明してきた。同方針を踏まえ、英国政府が2023年3月に公表した政策文書「プロイノベーション型規制手法 (A pro-innovation approach to AI regulation)」^{*26}が、同国のAI規制の基本的な枠組みに位置付けられている。同文書では、セキュリティ、透明性、公平性、説明責任、争議可能性の観点から5つの原則が掲げられており^{*27}、AIガバナンスに取り組むに当たっては、「イノベーション促進型の、柔軟で法規制に縛られない、比例的で信頼できる、順応性があり、明確で且つ協力的な (pro-innovation, flexible, non-statutory, proportionate, trustworthy, adaptable, clear and collaborative)」アプローチをとるとしている。当面は既存の法規制の下、各政府機関の連携により、産業界に対して上記原則の実装を促しつつ、将来的には、原則について何らかの義務化を図る可能性があるとしている。

また、2023年11月27日、英国国家サイバーセキュリティセンター (National Cyber Security Centre : NCSC) と米国サイバーセキュリティ・インフラストラクチャー安全保障庁

*23 「米上院トップのシューマー議員、AI法案策定に向けた行動枠組み発表」、『ジェトロ・ビジネス短信』2023年6月22日号

*24 「米下院、AIに関する超党派タスクフォース設立」、『ジェトロ・ビジネス短信』2024年2月28日号

*25 Tortoise media, "The Global AI Index", <<https://www.tortoisemedia.com/intelligence/global-ai/#rankings>> (2024/3/21 参照)

*26 GOV.UK, "AI regulation: a pro-innovation approach", <<https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach>> (2024/3/19参照)

*27 ①安全性、セキュリティと堅牢性：AIシステムは、そのライフサイクルを通じて、堅牢、セキュア且つ安全でなければならず、リスクは常に特定、評価、管理されなければならない。②適切な透明性・説明可能性：AIシステムの開発者・実装者は、関係者に対してAIシステムがいつ、どのように、どのような目的で使用されているかの情報を十分に提供し、関係者に対して、AIシステムの意思決定プロセスの十分な説明を提供しなければならない。③公平性：AIシステムは、そのライフサイクルを通じて、個人あるいは法人の法的権利を侵害してはならず、個人を不公平に差別したり、不公平な商業的成果を生み出したりするために使われてはならない。④説明責任とガバナンス：AIシステムの供給と使用について効果的な監視を確保するガバナンス体制が構築されなければならない。AIシステムのライフサイクルを通じて明確な説明責任が伴わなければならない。⑤争議可能性と是正：AIによる判断や結果が有害であり、又は重大なリスクを伴う場合、それによって影響を受ける者に対して不服を申し立て、是正する機会を提供しなければならない。

(Cybersecurity and Infrastructure Security Agency : CISA) が中心となり、日本を含む18か国が共同で、AIシステムのセキュリティガイドラインである「セキュアAIシステム開発ガイドライン (Guidelines for secure AI system development)」^{*28}を公表した。同ガイドラインでは、AIの設計、開発、導入、運用とメンテナンスの各段階において、取り組むべき事項を取りまとめている。

4 日本

日本は、民主主義や基本的人権等の観点からは欧米と同様の立場である一方、文化や社会規範の差異により、AIに対する社会認識という点では、欧米とは異なる文化圏にある。これにより、AIガバナンスの方向性として、欧州が法的拘束力の強いハードローを志向しているのに対し、日本は現時点では、AIガバナンスに関する横断的な法規制によるアプローチではなく、民間事業者の自主的な取組を重んじるソフトローアプローチを志向しており、総務省と経済産業省を中心に取組が行われてきたところである。総務省のAIネットワーク社会推進会議による「AI開発ガイドライン」^{*29}が2017年に、「AI活用ガイドライン」^{*30}が2019年に公表され、また同年3月に内閣府の統合イノベーション戦略推進会議が決定した、「人間中心のAI社会原則」^{*31}を基にしたガイドラインが策定された。続いて2021年7月に経済産業省が公表した「AI原則実践のためのガバナンス・ガイドライン」(2022年1月に改訂)^{*32}では、AI事業者が実施すべき行動目標が実践例と共に示されている。同ガイドラインは、AIを開発・運用する事業者が参考にし得るよう、環境・リスク分析やシステムデザイン、運用等の項目毎にまとめられている。

2023年5月、政府は「AI戦略会議」を設置し、AIのリスクへの対応、AIの最適な利用に向けた取組、AIの開発力強化に向けた方策等、様々なテーマで議論を行い、「AIに関する暫定的な論点整理」^{*33}を公表すると共に、各省庁のガイドラインの統合に向けた作業を進めることとされた。同年9月には、同会議にて生成AIに対するガバナンスも含めて統合された「新AI事業者ガイドライン スケルトン (案)」が示され、そして12月、政府は「AI事業者ガイドライン案」を公表した。同案では、人権への配慮や偽情報対策を求め、安全性やプライバシー保護等の10原則を掲げ、人間の意思決定や認知・感情を不当に操作するものは開発させないとしているが、欧米のような一定の法的拘束力を持つものではない。同案はその後、一般からの意見の公募を経て、2024年4月19日に「AI事業者ガイドライン (第1.0版)」として公表された。

また、2023年12月のAI戦略会議において、岸田総理大臣は、AIの安全性に対する国際的な関心の高まりを踏まえ、AIの安全性の評価手法の検討等を行う機関として、米国や英国と同様に、日本にも「AIセーフティ・インスティテュート (AI Safety Institute)」(以下「AISI」という。)^{*34}を設立すると発表し、2024年2月14日、経済産業省所管の情報処理推進機構 (Information-technology Promotion Agency : IPA) に設置された。AISIは、英国・米国等の同様の機関と

^{*28} National Cyber Security Centre, "Guidelines for secure AI system development", <<https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development>> (2024/3/12参照)

^{*29} 総務省, 「AIネットワーク社会推進会議 報告書2017の公表」, <https://www.soumu.go.jp/menu_news/s-news/01_iicp01_02000067.html>

^{*30} 総務省, 「AIネットワーク社会推進会議 報告書2019の公表」, <https://www.soumu.go.jp/menu_news/s-news/01_iicp01_02000081.html>

^{*31} 内閣府, 統合イノベーション戦略推進会議決定, 「人間中心のAI社会原則」, <<https://www8.cao.go.jp/cstp/aigensoku.pdf>> (2024/3/12参照)

^{*32} 経済産業省, 「AI原則実践のためのガバナンス・ガイドライン ver. 1.1」, <https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20220128_report.html> (2024/3/12参照)

^{*33} 内閣府 AI戦略会議「AIに関する暫定的な論点整理」, <https://www8.cao.go.jp/cstp/ai/ronten_honbun.pdf> (2024/3/12参照)

^{*34} AIセーフティ・インスティテュート, <<https://aisi.go.jp/>> (2024/3/12参照)

も連携しつつ、AIの開発・提供・利用の安全性向上に資する基準・ガイダンス等の検討、AIの安全性評価方法等の調査、AIの安全性に関する技術・事例の調査などを行っていることとしている。