AI ガバナンスに関する提言 Ver 3.0

2026 年 3 月
デジタル政策フォーラム

基本的考え方

生成 AI を巡る技術開発は著しい速度で進んでおり、生成 AI を社会経済システムに実装する動きも加速化している。こうした中、生成 AI を制御するためのルールづくりは、各国における法的枠組みといった総論の段階から具体的な政策のあり方を巡る議論へと比重を移しつつある。

本文書[1]では、こうした生成 AI を巡る動向を念頭に置きつつ、検討の基本的な視点として、

1. AI のリスクの操作可能性の確保（リスクの最小化）
2. AI の利便性を最大限享受できる環境の整備（利便性の最大化）
3. 上記環境を自律的に実現する市場の創出（健全な市場環境の整備）

という 3 つの目的を均衡ある形で実現する AI ガバナンスに関する議論を可能な限り広範な視点で整理する。

AI ガバナンス、すなわち AI 技術の制御可能性を継続的に維持するための仕組み[2]については、AI がもたらす便益とリスクのバランスを常に比較しながら議論する必要がある。

AI は社会のあらゆる領域で生産性や創造性の向上に寄与するものであり、パーソナル化（インテリジェンスの分散化）を通じて個人のデータ利用に係るデータ主権（data sovereignty）を技術的に担保しつつ、利便性の高いサービスが享受できるようになるなど、多種多様な便益をもたらす。

他方、人権侵害などの被害を深刻化するリスク、人間による制御可能性が失われるリスク、さらには AI が人間を代替することで生じるリスクなどが考えられる。ちなみに、こうしたリスクについては可能な限り技術的解決を目指し、拙速に規制を導入することはイノベーションを促す観点から適当ではない。

デジタル政策フォーラム（DPFJ）は、2024 年 7 月、AI ガバナンスの枠組みの構築に向けた論点を提言 Ver 1.0 として公表した[3]。さらに、同年 12 月にはより詳細な論点の整理等を行い、特に AI 法を巡る議論が具体化していたことを踏まえ、法的枠組みのあり方や AI 総合戦略の推進の必要性などについて提言 Ver 2.0 として公表した[4]。

しかし、AI 技術が引き続き急速な進化を遂げ、議論の対象としての AI そのものが変化し続けている状況にある。

そこで、提言 Ver 2.0 以降も続けてきた有識者インタビューや企業関係者からのヒアリング、関連会合

1　本文書は谷脇康彦デジタル政策フォーラム（DPFJ）代表幹事が監修した。

2　総務省・経済産業省「AI 事業者ガイドライン（第 1.1 版）」（2025 年 3 月）総務省・経済産業省）において、AI ガバナンスは「AI の利活用によって生じ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正のインパクト（便益）を最大化することを目的とする、ステークホルダーによる技術的、組織的、及び社会的システムの設計並びに運用」と定義されている。
https://www.soumu.go.jp/main_content/001002576.pdf

3　https://www.digitalpolicyforum.jp/wp-content/uploads/2024/06/240701_AI01.pdf （以下、提言 ver1.0）

4　https://prtimes.jp/main/html/rd/p/000000011.000131931.html（以下、提言 ver2.0）

の開催[5]等を踏まえ、かつ最近の海外動向なども反映させつつ、改めてAIガバナンス全般にわたる論点整理を行い、「AIガバナンスに関する提言 Ver 3.0」として整理する。

ここにはAIガバナンスを確保するための法的枠組み、リスク管理手法、AIを巡る競争政策、産業政策、外交政策及び安全保障政策、その他AIがもたらす社会構造変化のインパクト分析を含む多種多様な項目が取り上げられており、かつこれらの項目が有機的に相互作用している点に注目する必要がある。AIガバナンスを論じるには俯瞰的な「鳥の目」が求められる。

なお、本文書では現時点で一般向けに提供されている生成AIを主として念頭に置きつつ議論を進めることとし、汎用人工知能（Artificial General Intelligence）は一部を除き対象としない。

1　リスクの最小化

(1) リスク管理のあり方

＜AIの段階別リスク管理の困難性＞

AI管理の手法としてリスク（人の生命や基本的人権に与える負の影響などを含む）を数段階に分けて管理する手法が存在する。例えば、EUのAI法ではリスクを4段階に分類[6]している。

これはAIモデルが持っているリスクを深刻度に応じて段階別に管理しつつ、これを規制の度合いにリンクさせるものである。しかし、この手法の場合、コントロールすべきリスクの範囲をどう画定するか、またリスクをどのような基準でランク分けするのかといった具体的なリスク管理の手法に加え、リスク判断の主体、当該主体の判断の的確性を第三者に明示する手法（説明責任）等が確立しているとは言い切れない。

この点、将来的にはAIのシステムログ（動作履歴）を記録・解析することでAIのリスクをスコアリングし、これを公表する仕組みが構築されることとなれば、自らが許容できるリスクに応じてAIを選択することができるようになる可能性がある。すなわち、当該AIから得られる便益（ベネフィット）とリスク（コスト）を比較考量し、各利用者が自らの用途に適した（パーソナル）AIを選択できる仕組みが構築できることを念頭に置きつつ、国際機関におけるAIの標準化（リスク評価の手法を含む）等の議論に積極的に参画していくことが必要である。

またAIモデルのリスク評価を規制の強弱とリンクさせるアプローチは、リスクそのものが変化し得る可能性を踏まえれば、規制適用の予見可能性を損なう可能性がある。この点、欧州においては高リスクAI

5　DPFJ主催オープンカンファレンス「デジタル政策の論点2026」（2026年2月）において、AIガバナンスが主要テーマの一つとして取り上げられた。

6　欧州AI法は、AIのリスクを①許容できないリスク（人の生命や基本的人権に対する直接的脅威を及ぼすものとして開発を禁止）、②高リスク（事前の適合性評価、データベースへの登録等の義務）、③限定リスク（AIとのやり取りであることを利用者に知らせる透明性確保の義務）、④最小リスク（規制なし）の4種類に分類している。

の技術要件を緩和する「デジタルオムニバス提案」[7] が行われているところであり、今後の議論の動向を注視する必要がある。

事実、AI の抱えるリスク源は多様であり、その全容を把握することが困難である。例えば MIT 調査 [8] によれば AI を巡り 700 を超えるリスクが存在するところであり、そのすべてを念頭に置いたリスク管理を制度として実装することには大きな困難を伴う。また、本調査では「開発後（post-deployment）のリスク」がリスク全体の 65% を占めるとの指摘があるなど、リスクが時間の経過とともに動的・質的に変化する。

無論、AI のリスク管理そのものは極めて重要であり、国内においても産学官の連携により AI リスクに関するレポジトリーの作成・分析等を積極的に推進すべきである。

＜主体別のリスク管理＞

上記を踏まえつつ、AI のリスク管理については、

- AI の開発者
- AI を実装したサービス提供事業者
- エンドユーザー

という 3 つの主体に分けて検討することが望ましい。ちなみに、AI のリスクを検討する際、開発段階において AI が内包する可能性があるリスクと、サービス提供段階において AI が有することとなるリスク（AI を実装したサービスの提供・利用の方法等によって顕在化する可能性があるリスク。例えば誤情報・偽情報の生成・流布などが含まれる）の 2 つが考えられるが、特に後者のリスクについては、当該リスクが AI によって初めてもたらされたものであるのか、それとも従来から存在しているリスクが AI によって顕在化・増幅したものであるのか等について慎重に議論する必要がある。

主体別リスク管理のうち、AI の開発者によるリスク管理は、開発時に留意すべき事項を限定的に列挙する“Do Not List”アプローチに止めることとし、今後、AI 開発において具体的な問題が発生した場合にはその時点で対処することを基本としつつ、定期的なモニタリングを行うことが考えられる。

具体的には、例えば欧州評議会の AI 条約（2024 年 9 月）[9] 等を参照しつつ、「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・法の支配”と十分な整合性を確保すること」などを原則とするこ

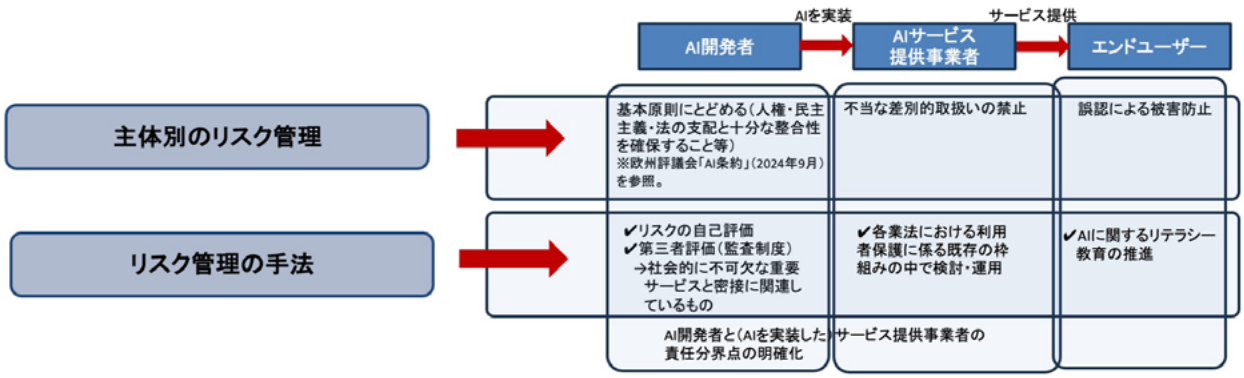
7　2025 年 11 月、欧州委員会は「デジタルオムニバス提案」を公表した。欧州域内における規制コスト負担を軽減することを目的とするオムニバス提案は１０の領域に及び、デジタル分野もその一つ（オムニバス７）とされている。この提案の中では一定の要件（再識別リスクの適正な管理、データ主体の異議申し立ての機会の確保、データの最小化等）に適合する個人情報を AI 学習データとして利用することを合理的利益として許容するほか、高リスク AI に関する技術要件の緩和を検討項目として挙げている。特に技術要件の緩和については AI 法の該当部分の施行を最大で 2027 年末までに延長（当初予定は 2026 年 8 月）することが提案されている。

なお、本提案は現時点では草案段階であり、欧州議会及び欧州理事会との調整を経て立法化がなされるものと見込まれている。

（出典）European Commission “Simpler EU digital rules and new digital wallets to save billions for businesses and boost investment”（November 2025）.
https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2718

8　P. Slattery et al. “Global AI adoption is outpacing risk understanding, warns MIT CSAIL”（MIT CSAIL News, August 14, 2024）
https://www.csail.mit.edu/news/global-ai-adoption-outpacing-risk-understanding-warns-mit-csail#:~:text=Global%20AI%20adoption%20is%20outpacing,it%20remains%20current%20and%20relevant.

9　2024 年 9 月、「AI 並びに人権、民主主義及び法の支配に関する欧州評議会国際条約」（Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law）に米国・EU を含む 15 を越える国・地域が署名（日本は 2025 年 2 月に署名（※））。
（※）: https://www.mofa.go.jp/mofaj/press/release/pressit_000001_01725.html



とが考えられる。

また、AI を実装したサービス提供事業者によるリスク管理についても、可能な限り限定的であることが望ましい。例えば電気通信事業法第 6 条[10] が規定するように、サービス提供における不当な差別的取り扱いを禁止する等の規律にとどめることが望ましい。

AI に関連して不当な差別的取り扱いを禁止するのは、AI を活用することで個別医療など個人情報を活用したきめ細かいサービスの提供が可能となる一方、個人の特性に応じて個別化（personalization）されたサービス提供ではなく、合理的な根拠に欠ける差別（discrimination）とならないようにすることが人権保護の観点から求められるからである。

なお、AI を組み込んだサービスを利用者に提供する場合、AI の開発者とサービス提供事業者との間の責任分界点についてもサービス提供の前段階において予め明確にしておくことが利用者保護の観点から求められる。

さらに、エンドユーザー（中小企業等を含む）におけるリスク管理については、AI のリスクについて正しく理解するためのリテラシー教育が求められる[11]（項目（5）を参照）。

＜リスク管理の手法＞

リスク管理は既述の AI リスクリポジトリーの作成・分析の結果等を踏まえて行うべきであるが、その際、リスクの自己評価あるいは第三者評価（例えば監査もしくは認証制度）の適用の可否について検討する必要がある。

10　電気通信事業法第 6 条は「電気通信事業者は、電気通信役務の提供について、不当な差別的取扱いをしてはならない。」と規定している。

11　例えば、2024 年 9 月、韓国において性的なディープフェイク画像や動画の所持・視聴を処罰する「性暴力犯罪処罰などに対する特例法」の改正法が成立している。また、Freedom House は報告書“Freedom on the Net 2025”（※）において、AI 生成コンテンツに関する規制が世界中で急速に導入されてきており、規制のアプローチは国によって大きく異なり、権威主義国家では検閲や監視の強化に力を入れる（国の方針に沿わない生成物の生成を禁止する）一方、民主主義国家では子供の保護と表現の自由の適正なバランスの確保に苦慮していると指摘している。
（※）Vesteinsson, Baker, Brody, Funk, Grothe, Slipowits eds. Freedom on the Net 2025, Freedom House, 2025 www.freedomonthenet.org

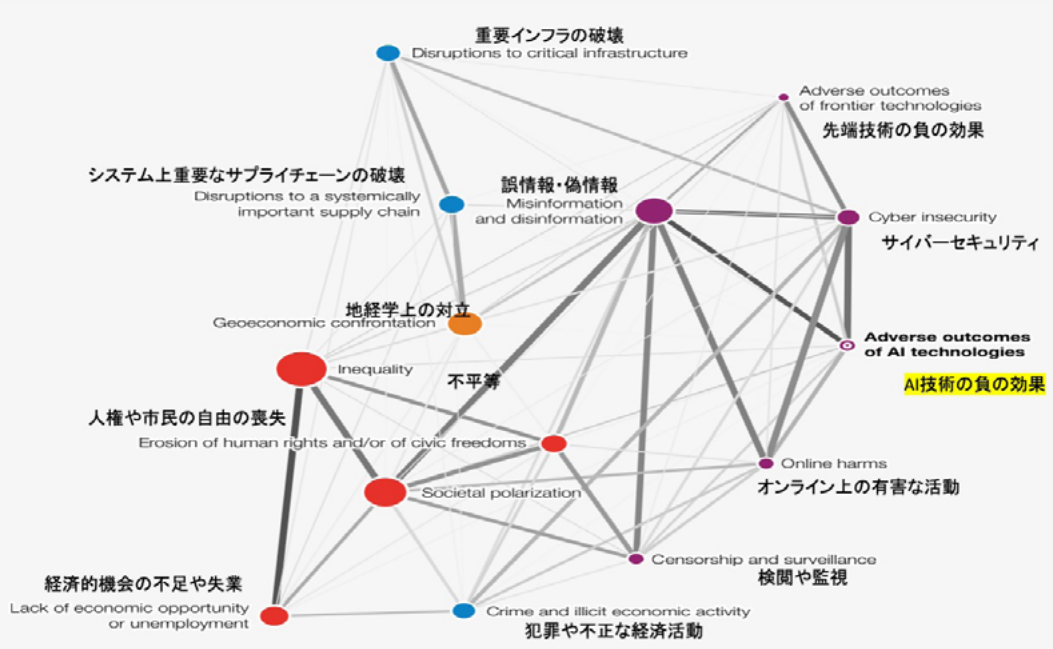
具体的には、AI の開発者においては開発者自らによる自己評価を基本とし、社会的に不可欠な重要サービスと密接に関連しているものについては第三者による監査制度を組み合わせることが考えられる。

またサービス提供事業者においては AI の機能だけを抜き出して評価することは困難であることから、各業法における利用者保護に係る既存の枠組みの中で検討・運用すべきであり、AI の活用を契機として規制の上乗せを行うことは適切ではない。

＜システミックリスク分析の必要性＞

AI のリスクを考える場合、AI そのものに内在しているリスク（risk of AI）と AI によって引き起こされるリスク（risk by AI）の 2 つの側面がある。後者のリスク（リスク要因が連鎖することでより大きく深刻な負の影響がもたらされる可能性）は AI のシステミックリスクであり、領域を超えた解析が重要になる[12]。特に AI を含むデジタル技術が汎用性の高い社会基盤となっていくことが見込まれる中、AI のもたらすシステミックリスクについての分析が今後重要になってくる。

AIのシステミックリスク



（出典）World Economic Forum “The Global Risks Report 2026” (January 2026)

（2）規制のあり方と実効性の確保

AI に関する規制の手法としては、ハードロー（法規制）とソフトロー（民間部門による自主規制）、さらにそれらを組み合わせた官民連携による共同規制（co-regulation）などの手法がある。例えば、欧州

12　世界経済フォーラム（World Economic Forum）は世界が直面するリスク（グローバルリスク）について年次評価を行っており、“The Global Risks Report 2026”（January 2026）では「AI がもたらす負の影響」が上位（”現時点”で第８位、”１０年後”で第５位）に位置付けられている。
https://reports.weforum.org/docs/WEF_Global_Risks_Report_2026.pdf

「AI 法」[13] や中国「生成 AI サービス管理暫定弁法」[14] はハードローを基本としている。

ただし、ハードローを志向する場合であっても基本法的な緩やかなアプローチと、具体的な行為規制を課す規律性の高いアプローチなど、規律のあり方について一定の幅が存在する。

また、共同規制とは、国がルールの基本方針を示し、その趣旨に賛同した事業者が基本方針に基づくルールを運用して運用結果を国に報告、国はこれを評価して必要に応じて基本方針を修正するという方式であり、欧州においてはプラットフォーム事業者の偽情報対策などで採択されている。共同規制は民間主導による柔軟な規律の適用という点で優れているが、他方、行政による規律が法的根拠に基づくことなく裁量的に行われないよう十分な透明性の確保が求められる。

急速な技術革新が進む中、過去の AI 関連の議論の中には市場の実態からかけ離れ、必要以上に議論が為念的・抽象的なものになる傾向も散見された。あくまで冷静な議論を前提とし、関係当事者の自主的な取り組みを基本としつつ、

- 必要な規律の確保
- デジタル産業の振興（規制と振興の適正なバランスの維持）
- 規律の国際的調和の実現

を三位一体で進めることを基本とすべきである。

＜主要国間で深まる対立＞

米国は、バイデン前政権による大統領令（2023 年 10 月）において、AI 法制定の可能性を示唆しつつ、AI に関する安全・セキュリティ基準の策定、AI のアルゴリズムによる差別禁止のためのガイダンスの策定等を進めることとされた。

しかし、2025 年 1 月にトランプ第二期政権が発足し、前政権による AI 政策全体を見直すこととされ、同年 7 月に「AI 行動計画」を策定・公表した。

この中では、AI イノベーションの加速化、米国内における AI インフラの建設（半導体製造の強化や電力グリッドの整備を含む）、AI を巡る国際外交・安全領域での先導という 3 本柱の下、30 項目の施策をリスト化した。さらに同年 11 月には AI イノベーションの加速化に向けた具体策として、新たな国家プロジェクトである“Genesis Mission”を立ち上げることを公表し、連邦政府が自ら整備する統合プラットフォーム ASSP（American Science and Security Platform）を活用して戦略的技術開発を推進する方針を公表した。

こうした米国のアプローチ（非規制を原則とし、産業振興を優先）は AI 法の施行に向けた欧州のアプローチ（規制は信頼できる AI 環境を実現する前提条件）とは鋭く対立している。

例えば、2025 年 2 月にフランスで開催された AI アクションサミットにおいて、フォンデアライエン EU 委員長は「AI が安全であるという人々の確信が必要であり、これこそが AI 法の目的である」とハードロー

13　2023 年 5 月、欧州理事会は「AI 法」を採択した。2024 年 5 月から段階的に施行されており、全面適用は 2026 年夏の予定であったが、現在は部分的に施行を遅らせる可能性がある（最近の動向は脚注 6 を参照）。
https://artificialintelligenceact.eu/

14　2023 年 8 月、中国は「生成人工知能サービス管理のための規則」を施行。本規則（第 4 条）では、法律や行政規則で禁止されているコンテンツの作成を禁止しており、「社会主義の中核的価値観を遵守」する生成物のみが認められている。
（出典）原田雅史「中国「生成人工知能サービス管理暫定弁法」の制定とその解説」企業法務ナビ（2023 年 7 月 21 日）
https://www.corporate-legal.jp/matomes/5362

のアプローチの正当性を主張したのに対し、バンス米副大統領は「AI 部門を過度に規制することは離陸途上にある革新的な産業を殺す（could kill）ことになる」と反論し、同サミットにおける共同声明[15]に署名しなかった（欧州や日本を含む 64 の国・地域が署名）。

また、米国国内においても連邦政府のアプローチと州政府のアプローチに乖離が生まれてきている。州レベルでは法案審議中の 2 州を除く 48 州で AI 規制が課されており、AI チャットボットの運用、偽・誤情報の生成、AI による医療行為などに一定の規制が課されている[16]。こうした連邦政府と州政府の AI 規制に関するスタンスの違いについて、連邦政府は過度の州政府の規制について訴訟を提起する等の体制の整備[17]を図っており、今後も様々な動きが出てくるものと想定される。

＜日本における AI 法の制定＞

日本においては、2025 年 9 月、「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」（AI 法）が施行された。

これに先立つ DPFJ 提言 Ver 2.0（前掲）においては、

- 日本において AI を巡る法制度を検討する場合、これまで政府において検討が重ねられてきた各種ガイドラインの内容を詳細に法制化するのではなく、ハードローとして AI 基本法を制定すること；
- AI 基本法においては、例えばサイバーセキュリティ基本法[18]を参照しつつ、AI を巡る政策の基本理念、国等の主体が果たすべき責務、AI 戦略の策定、政府における AI 戦略本部（及び本部事務局）の権能及び関係機関との連携等について規定すること；
- 開発者によるリスク管理など業態横断的な取り組みについては内閣官房に設置する本部事務局を中心に、またサービス提供事業者については業態ごとに主管官庁において行うこととし、特に AI の特性に応じて利用者保護の観点から業態横断的に確保すべき事項（統一基準）が必要と認められる場合は本部事務局が主導し、関係府省と連携しつつ統一的に施策を推進することが望ましいこと；

の3点を指摘した。

今般の法律はハードローという体裁をとりつつも、提言 Ver 2.0 のラインと同様に基本法的な項目で構成された非規制のアプローチを堅持している点は評価できる。

また、AI 法に基づいて策定された「人工知能基本計画」（2025 年 12 月閣議決定）においても、「イノベーション促進とリスク対応の両立」「アジャイルな対応」「内外一体での政策推進」の3項目を原則として掲げ、AI 戦略本部と各府省が連携しつつ関連施策の推進にあたる方向が明確化されている。

15　https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet

16　読売新聞「米 48 州 AI 規制ーー偽情報や自殺助長防ぐ」（2026 年 1 月 18 日）
https://www.yomiuri.co.jp/world/20260117-GYT1T00324/

17　White House, Executive Order “Ensuring a National Policy Framework for Artificial Intelligence”（December 11, 2025）
https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/

18　サイバーセキュリティ基本法では、サイバーセキュリティに関する施策に関する基本理念と各主体（国、地方公共団体、重要社会基盤事業者等）の責務、サイバーセキュリティ戦略の策定、基本的施策、サイバーセキュリティ戦略本部の設置等について規定している。
https://laws.e-gov.go.jp/law/426AC1000000104#Mp-Ch_1

＜ルールの実効性とプレーヤーの自律性＞

上記のとおり AI 法そのものは非規制を堅持しているが、今後、共同規制もしくはガイドライン等によって AI の開発・利用についての非拘束的なルールを整備する際、当該ルールの実効性を確保する観点から AI の開発者や利用者の自律性に実質的な制限（規制）が課されることのないよう慎重な対処が必要である。

例えば、2025 年 12 月に公表された「生成 AI の適切な利活用等に向けた知的財産の保護及び透明性に関するプリンシプル・コード案」（内閣府知的財産戦略推進事務局）[19] は AI 活用に係る知的財産侵害のリスクを回避するための行動規範と位置付けられる。その取り組み自体は評価されるが、「コンプライ・オア・エクスプレイン」の原則の下でルールが過度に詳細なものとなって、行動規範とはいえ実質的な規制として機能することがないよう、今後検討を深めていく必要がある。

また、AI のアルゴリズムの妥当性・透明性について、どこまで第三者が検証可能な情報開示を求めるか今後議論が必要である。

この点、従来の情報検索においては関連性の高い情報 URL が表示されるにとどまり、その情報を獲得した者自らが情報を読み解き解釈するリテラシーは引き続き要求されてきた。しかし AI の生成物はアルゴリズムによって整理された情報が提示されるのみであり、アルゴリズムの正確性について評価する局面がない。したがってアルゴリズムを客観的な第三者が評価する仕組みは今後の AI の社会実装において最も重要な課題の一つであると考えられる。

例えば、開発者に期待される自己監査及び第三者による外部監査についてはあくまで開発者による自

19　https://www.kantei.go.jp/jp/singi/titeki2/pdf/shiryo1.pdf

主的措置とし、必要に応じて政府（AI 戦略本部）と連携する仕組み（例えば政府等が監査指針を策定するとともに監査の実態を踏まえて当該指針の改定等を行う共同規制型の運用）の構築を検討することが考えられる。

特に大規模な AI の開発者について規制導入を検討する等の議論も散見されるが、前述のとおり規模の大きい AI は第三者による外部監査を自発的に導入することをオプションとするにとどめ、規模の大小が関連市場にどのような影響を与えるかという競争政策関連の議論（項目（6）を参照）とは区別することも選択肢として考えられる。

（3）外的リスクに対する脆弱性対策

AI が社会基盤となっていく中、AI のレジリエンス（抗たん性）を確保するための機能保証（mission assurance）[20] は極めて重要である。このため、特に AI の脆弱性等の外部リスクに係る対策については関係者が連携して取り組んでいく必要がある。

＜ AI に係るサイバー攻撃対策＞

AI モデルの外的リスクを管理する観点からは、AI の脆弱性調査（red teaming）について、監査（自己監査または第三者による外部監査）項目に取り入れることが適当であり、これを実施するためのガイドラインの策定を官民連携により行うべきである。この点、日本においては 2024 年 9 月に AI セーフティ・インスティチュート（AISI）が「AI セーフティに関するレッドチーミング手法ガイド」[21] を公表する等の対応が進んでいる。

なお、当該検討に際しては AI の有する特性が多岐にわたることを踏まえ、脆弱性調査の範囲・目的等について限定・明確化することが実効性を担保する観点から重要である。

また、AI の学習プロセスにおいてデータ汚染攻撃 [22] 等によって当該 AI が所期の機能を発揮しなかったり誤作動してしまうなどのリスクがある（AI に対するサイバー攻撃）。また、脆弱性の発見やマルウェアの作成、偽アカウントの生成や偽情報の配布等に AI を活用するリスクが顕在化している（AI によるサイバー攻撃）。こうした「AI に対するサイバー攻撃」及び「AI によるサイバー攻撃」への対処についても具体策を早急に検討する必要がある。

さらに、AI の普及が認知戦の激化を招いており、偽アカウントの大量生成、ボットによる反政府的言動の拡散などが行われているところであり、こうした認知戦に対応するためのデジタル技術の開発促進が求

20　機能保証とは「いかなる環境・条件であっても（DoD の）任務に不可欠な機能（MEFs: Mission-Essential Functions）--- 人材、機器、施設、ネットワーク、情報および情報システム、インフラおよびサプライチェーン --- に求められる能力や資産の継続的な機能維持や能力の抗たん性を防御・確保するためのプロセス」を指す。
（出典）US Department of Defense “Mission Assurance Strategy”（April 2012））
https://policy.defense.gov/Portals/11/Documents/MA_Strategy_Final_7May12.pdf

21　AISI は「AI セーフティに関するレッドチーミング手法ガイド」第一版を 2024 年 9 月に公表。その後、具体的な実施例を通してより詳細に理解できるよう改訂し、2025 年 3 月、第 1.10 版として公開。改訂版においては、RAG（Retrieval-Augmented Generation）の仕組みを実装した AI システムに対して実際にレッドチーミングを行い、その手順を詳細に解説するとともに、レッドチーミング実施の成果物を文書として取りまとめた。
https://aisi.go.jp/output/output_framework/guide_to_red_teaming_methodology_on_ai_safety/

22　データ汚染（data poisoning）攻撃では、学習データに間違った出力を生じさせる汚染データを挿入し、モデルが悪意をもって機能するように修正することを試みる。また、データ回避（data evasion）攻撃では、人間の知覚できないノイズ等を学習データに混入させて AI の判定結果を誤らせる。

められる[23]。

上記の検討に際しては、AI の透明性確保のために学習データや AI システムをオープン化することにより、脆弱性の発現、第三者による悪意ある模倣や犯罪行為への悪用が起きることが懸念されるため、オープン性の確保と AI が悪用される可能性の双方について同時並行的に検討を進める必要がある。

＜データ空間の健全性の確保＞

AI が学習データを数次にわたり学習する過程において、出現度の少ないデータを捨象するプロセス（クエリーに対する的中率を向上させることを目的）をとることが多い。この場合、前世代のモデルで出現確率が高い言葉が次世代において評価され、逆に出現確率の低い言葉が過小評価されることによりモデルの多様性が失われる（退化する）、いわゆる「モデル崩壊」（model collapse）が生じる可能性が指摘されている[24]。こうした状況を放置することは不正確で健全性に欠けるデータを拡散し、データ空間（data space）の汚染（contamination）を進行させることになる。

このため、AI の学習データを人間の作成したものに限定する、あるいは学習済み AI であることを対外的に明示する等の一定の規律として、例えば民主導の認証制度を設けることについて検討が必要である。

また、人間の作成したデータを増加させるという観点からは著作権の切れた文書や公的機関が作成した文書等を広く学習データとして活用可能とするオープンデータ化が有効である。

（4）生成物の取り扱い

AI は学習データを学習して言語モデルを形成し、これを活用した推論プロセスを経て生成物たるデータを出力するものである。そこで、第三者による改ざんが行われていないというデータの「完全性（integrity）」の確保という観点から見れば、上記（3）中の「データ空間の健全性の確保」は入力値（学習データ）の完全性を確保するという視点であるが、同時に出力値（生成物）の完全性を確保するための取り組みも必要になる。このため、生成 AI を用いて膨大な偽情報が既に流通している状況にある中、共同規制のアプローチを前提としつつ偽情報対策を効果的かつ具体的に推進する必要がある。

その際、AI の生成物であることを判別可能とする電子透かし（digital watermark）の導入が有効と考えられる。また、インターネット上の情報（コンテンツ）の作成者・発信者をユーザーが確認するためのオリジネータープロファイル（OP）技術の有効性についても、技術基準の国際標準化や OP を付与する主体のあり方等について関連して議論を深める必要がある。

2　利便性の最大化

23　AI の悪用の具体的事例については、例えば Open AI “Disrupting malicious uses of AI: an update”（October 2025）を参照。
https://cdn.openai.com/threat-intelligence-reports/7d662b68-952f-4dfd-a2f2-fe55b041cc4a/disrupting-malicious-uses-of-ai-october-2025.pdf

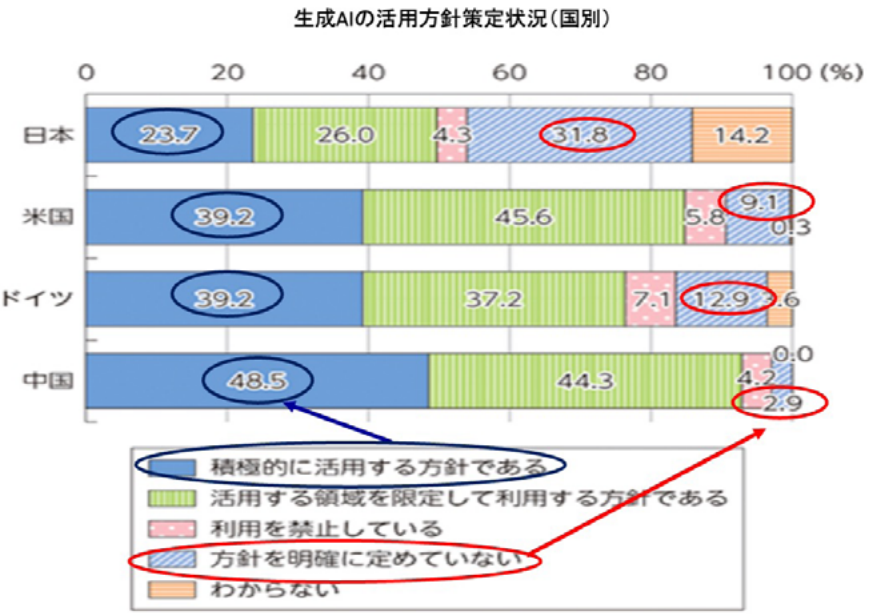
24　I. Shumailov et al. “The Curse of Recursion: Training on Generated Data Makes Models Forget” arXiv（May 2023）
https://arxiv.org/abs/2305.17493

（5）AI の積極的活用

日本においてはデジタル技術の活用が他国に比べて遅れており、AI 活用の分野においても立ち遅れが大きい。AI を導入している企業等においても米国発の AI を利用している事例がほとんどであり、今後、日本のデジタル赤字がさらに拡大していく可能性が高い（項目（9）を参照）。こうした中、AI を積極的に活用していくことは日本の社会経済にとって喫緊の課題である。

特にデジタル技術を自ら開発・運用する国家の力は「デジタル主権（digital sovereignty」と呼ばれ、AI はその中核を担うものであるという認識に立ち、政府は積極的に AI 領域の政策課題に取り組む必要がある。

AI活用に関する国際比較



（出典）総務省「令和7年版情報通信白書」（2025年）

＜課題解決のための AI 活用の推進＞

AI の活用については既に様々な取り組みが始まっているが、特に深刻な少子高齢化が進む中、教育分野と医療分野[25]においてデータ活用の取り組みが遅れていることを踏まえると、これらの分野における AI 活用を積極的に進める必要がある。

特に教育における生徒、医療における患者を起点として関連するデータを個人の許諾の下に紐づけて解析する仕組みは教育や医療の個別化に貢献することが期待される。

他方、こうしたデータ連携が過度のプロファイリングを招くことがないよう一定のセーフガード措置も併せて検討する必要がある。また、例えばカルテデータなど、地域や組織によってデータ様式が異なることからデータ連携が進んでいなかった事例についても、AI 解析による自動連携が実現する。

25　医療分野においては、例えば、個人のデータに基づく治療薬の処方、疾病リスクの予測や精度の向上、迅速かつ効率的な創薬の実現、医療事務作業の自動化などが期待される。

さらに、教育や医療の分野の他にも、地球的な課題である環境対策、人の生命財産を守るための防災・減災、豊かな暮らしを実現するための文化などの幅広い分野での AI の積極的な活用を図る必要がある。その際、これらの分野で AI を積極的に活用するために、留意すべき事項や開発すべき技術について検討を深める必要がある。

同時に、学習データとしての個人データの取り扱い、当該データを取り込んだ場合の出力に個人データが含まれる可能性の回避など、プライバシー保護の観点から所要の方策が必要となる。また、学習データや生成物の著作権法上の取り扱いについて明確化を図ることが求められる。

加えて、AI のリスクについて、既述のとおり一般利用者が正しく理解するためのリテラシー教育が重要になる。例えば官民連携による青少年インターネット利用環境の整備の取り組み事例と同様、AI のリスクについても広く周知啓発活動を行うことが重要である。

＜行政サービスにおける AI 活用の推進＞

国及び地方自治体における行政サービスの提供においては、引き続き少子高齢化が進む中、AI の活用や積極的なデータ連携により、限られた人的リソースを効率的に投入するとともに個別化によるきめ細かなサービスの実現を図っていく必要がある。

しかし、こうした行政サービスの提供における AI の積極活用については地域住民の理解を得ることが不可欠であることを踏まえ、兵庫県神戸市[26]の事例などを参照しつつ、必要な制度的枠組み（基本指針の策定ならびにリスクアセスメントの実施）の整備・運用を図るとともに、ベストプラクティスの共有を図るなどの取り組みが求められる。

＜ AI 活用と労働市場＞

AI の積極活用は社会の自動化を進め、雇用機会を喪失する（人間の仕事が奪われる）という主張がある。しかし、AI を既存の労働力の置換に充てることを目指すのではなく、あくまで労働生産性の向上及び新たな市場領域を創出するためのツールとして活用することを基本方針とし、政府もこれを実現する方向で所要の政策支援を行うことが期待される。

AI を含むデジタル技術は既存市場の効率化を進めることのみを趣旨とするものではない。むしろ既存の事業領域の壁を打ち破り新しい市場領域を生み出すことで新たな雇用を生み出すものであることを広く認識として共有していく必要がある。

3　健全な市場環境の整備

（6）健全なエコシステムの構築 ― 競争政策

AI の進化は基本的に民間の創意工夫によって行われるべきである。国はこれを積極的に支援するとと

26　神戸市における AI の活用等に関する条例（2024 年 3 月制定、同年 9 月施行）
https://www1.g-reiki.net/city.kobe/reiki_honbun/k302RG00001955.html

もに、公共の利益を確保する観点から必要なルール策定や政策支援を行うことを基本とすべきである。

その際、AI の開発者や利用者を含む多様な主体によるエコシステムを確保していくためには、健全な市場環境を確立するための競争政策が重要となる[27]。

そこで、AI 関連市場における参入障壁や、大企業による優越的地位の濫用などの反競争的行為を監視する仕組みを確立する必要がある。また、現在の大手有力 AI は既存の大規模プラットフォーム事業者が提供するものが主流となっているが、今後、AI 市場あるいは隣接市場（例えばプラットフォーム事業）において市場支配力が濫用される可能性及びこれに対する競争セーフガード措置について検討が必要である[28]。

特にプラットフォーム事業者のような複数レイヤーで事業展開を行う垂直統合型の AI 開発者は、それ以外の開発者と比して高い市場支配力を持ち、かつ隣接市場への市場支配力を行使する可能性が高いのではないかという懸念があり、競争政策としてどのように対処すべきか検討する必要がある。

加えて、市場支配力の濫用の有無について検証を加える場合の市場画定のあり方について、データの越境流通や AI のネットワーク化などを見据えつつ検討すべきである。

なお、欧州「AI 法」においては法律の域外適用の条項が盛り込まれているが、こうした域外適用が増加することで国外の規制が重畳的に国内で適用されることとなるなど過度の規制をもたらす負の可能性についても、その回避策に関する検討が求められる。

（7）産業振興とグローバル連携 ― 産業政策

AI 関連サービスを含むデジタル関連産業は、データの特性（限界費用ゼロ、非競合性など[29]）スケーラビリティを発揮することで寡占市場が形成されやすい一方、グローバルな接続性が求められる。このため、AI についてもコンピューティング資源を多様に組み合わせたり、ネットワーク化された AI の相互作用（interaction）によって機能を高めるなど、国境を超えて AI がネットワーク化される世界が想定される。こうした世界を前提に考えれば AI のオープン性の確保が必須となるとともに、ルールのグローバル化が求められる（項目（8）を参照）。

＜オープン性の確保と標準化戦略＞

インターネットが爆発的に普及した主因の一つは、「自律・分散・協調」を基本精神とする、そのオー

27　OECD “Artificial Intelligence, Data and Competition” OECD Artificial Intelligence Papers No. 18（May 2024）
https://www.oecd.org/daf/competition/artificial-intelligence-data-and-competition.htm

28　2024 年 10 月、公正取引委員会はディスカッションペーパー「生成 AI を巡る競争」（※）を公表した。その中で生成 AI の市場構造を①インフラ層（GPU などのハードウェアとデータ）、②モデル層（基盤モデル）、③アプリケーション層（各種サービス）の 3 層に整理し、レイヤーごとの競争リスクを分析した。本文書においては、アクセス制限・市場排除、自社優遇・抱き合わせ、アルゴリズムを用いた協調行為などの競争阻害行為の可能性を指摘し、本文書の公表後に意見招請を実施した。その後、2025 年 6 月に「実態調査報告書 Ver. 1.0」（※※）を発表した。
（※）https://www.jftc.go.jp/houdou/pressrelease/2024/oct/241002_generativeai_02.pdf
（※※）https://www.jftc.go.jp/houdou/pressrelease/2025/jun/250606_generativeai02.pdf

29　データの複製・配信にかかる追加コストがほぼゼロであることを「限界費用ゼロ」という。また、あるデータを誰かが利用（消費）しても、他の人がそのデータを同時に利用できなくなることがなく、データの量や質が減じられない性質を「非競合性（non-rivalry）」という。こうした特性は競争政策において少数のプレーヤーによる市場支配力の濫用を生じさせる可能性を高めるものであり、検索サービス等のプラットフォーム事業者への競争政策の適用に関する基本的な問題意識となっている。

プン性にある。同様に、AI についてもクローズドな私権型の AI（proprietary AI）とオープン型の AI（open AI）の2つのアプローチが考えられるが、健全な市場の発展を促すとともに AI 関連サービスの品質を維持する観点からは、十分な競争環境を創出するオープン性の確保が不可欠である。同様のアプローチは欧米でも見られる[30]。

こうした観点から、オープンソースの活用、異なる AI 間の相互運用性の確保をどのように実現するのか、こうした環境を実現するための標準化の促進、オープン型の AI 開発を促すことを前提とした研究開発支援などについて、政府は積極的に推進すべきである。

また、AI 関連の技術開発について日本はグローバル市場において既に遅れをとっている状況にある中、オープン型の AI を組み込んだソリューションの開発を国が支援するなど、オープン型の AI に対して積極的な振興策を講じることを検討すべきである。特に AI 系のベンチャー支援のための取り組みを強化するための議論が必要である。

その際、オープン性については形式的なオープン性と実質的なオープン性を区別し、政策としては後者のオープン性の確保を念頭に置くべきである。例えば AI の学習データや RLHF（Reinforcement Learning from Human Feedback）の過程における具体的なフィードバックについて公開されていない場合、技術仕様としてのオープン性は確保されているものの実効面における AI のオープン性が確保されない（立証できない）ことが懸念される。このように、実効性のあるオープン性を担保するためのセーフガードについても検討が求められる。

（8）国際的コンセンサスの醸成 ― 外交政策

AI は国内に閉じて開発・利用されるものではなく、ネットワーク化されサイバー空間で広く利用されることが前提となる。その際、上記の論点については国際的に緩やかなコンセンサスを形成しながら、各国の法制度などのルールに反映し、必要な調和を図っていくことが求められる。

その際、AI が戦略的分野であり、各国の産業競争力や課題解決に大きな影響を与えるものであることを踏まえ、産業、技術、外交など様々な領域の専門家による俯瞰的な取り組みが必要であり、政府部内及び官民連携による実効性のある体制整備が求められる。また、AI がグローバルサウスの抱える課題解決に貢献する可能性が大きいことを踏まえ、グローバルサウスの十分な参加を得た形で進めることが求められる。

30　欧州（EU）では、生成 AI を「単体のプロダクト」ではなく“生成 AI のバリューチェーン全体（半導体・クラウド・データ・基盤モデル・配布チャネル・アプリ）”として捉え、競争政策（独禁法・合併審査・規制）を組み合わせた議論が進んでいる。欧州委員会（DG COMP）は、2024 年 1 月、意見招請文書”Competition in Virtual Worlds and Generative AI: Calls for Contribution”（※）を公表。同年 9 月、整理文書”Competition Policy Brief”（※※）を公表し、競争上の論点を市場のボトルネック性の有無に置き、①チップ / 計算資源、②クラウド、③データ（ライセンスを含む）、④ AI 人材、⑤生産性ソフト / 配布面（OS・検索・SNS・メッセージング等）など、複数レイヤーにまたがる形で検証している。

ここでは、計算資源・クラウドの囲い込み（優遇条件、排他、クレジット設計）、データへのアクセス（学習用データの入手可能性、利用条件の公平性）、巨大プラットフォームによる優先表示 / 抱き合わせ / 自社優遇（配布チャネル支配）、パートナーシップ型の“実質的な支配”（持分ではなく契約でコントロールが生じるか）といった点が生成 AI 市場の競争を左右する要素として整理している。

このように、欧州における生成 AI を巡る競争政策としては、①計算資源（GPU/ クラウド）とデータのボトルネックの担い手、②提携・投資・契約が実質的な排他 / 囲い込みになっていないか（合併審査＋独禁法の両睨み）、③検索・SNS・メッセージング等の配布チャネルにおける自社 AI の優遇、④ AI 法等の実装が参入コストや競争条件に与える影響といった事項が論点として挙げられる。

（※）https://ec.europa.eu/commission/presscorner/detail/en/ip_24_85

（※※）https://competition-policy.ec.europa.eu/document/download/c86d461f-062e-4dde-a662-15228d6ca385_en

さらに、こうした国際的コンセンサスの醸成の中で特に急務なのが、AI の軍事利用に係る規範の形成である。2023 年 2 月にハーグで開催された「軍事領域における責任ある AI に関する会議」（REALM Summit）における提案「人工知能及び自律性の責任ある軍事利用に関する政治宣言」[31] にあるような、AI 利用に関する自主的なコミットメントを拡大していく必要がある。同時に国連の安全保障の枠組みの中で AI セキュリティ監査（査察）の仕組みを取り入れることも検討に値する。こうした AI と安全保障のあり方について、既に AI の軍事利用が現実化[32] していることを踏まえて議論を急ぐ必要がある。この点、米国が多数の国際機関から脱退するなど国際連携の共同歩調が大きく損なわれていることが深刻な影響をもたらすことが懸念される。

（9）AI がもたらす広範な影響に関する議論 ― 議論の拡張

＜ AI における集中と分散＞

コンピューターの歴史は、メインフレームと呼ばれる大型電子計算機の時代から始まる。インテリジェンス（頭脳）は中央にあり、そこに接続された端末を介して共同利用していた。しかしパソコンが登場すると、その普及にあわせてインテリジェンスの分散が進んだ。その後、仮想化技術や並列分散処理技術を使ったクラウドの登場により再度インテリジェンスの集中が起こり、世界各地にデータセンターを作る動きは一層活発になってきている。そして、近年はクラウド利用がさらに普及すると同時にエッジコンピューティングが登場し、インテリジェンスの分散も大きな潮流となっている。

このように、コンピューターの世界において集中と分散が繰り返されてきた。ただし、集中と分散は対立する概念ではない。技術革新やこれに伴うコスト構造の変化、運用・管理技術の進化、多様な利用者ニーズなどを踏まえながら、集中と分散の間で時々のベストミックスが選ばれてきた。

ネットワークの世界も上記と同様に、レガシーの電話網ではツリー状の階梯構造がとられ、電話会社が一元的に管理・運用していたが、インターネットの世界では点在する多様な主体により設置された無数のルータが相互連携しながら機能する水平分散型となっており、加えて近年のブロックチェーン技術を軸とする Web3 の世界も水平分散型の仕組みとなっている。

ただし、この変化は集中から分散へという一方向の変化ではない。集中型モデルと分散型モデルをどのように共存させ、どのような役割分担を実現するかという点が議論のポイントとなる。

AI においても同様であり、LLM（大規模言語モデル）という集中型モデルが唯一の選択肢ではない。エッ

31　本提案（US DoS “Political Declaration on Responsible Use of Artificial Intelligence and Autonomy”（February 2023））では、軍事 AI が国際法（特に国際人道法）の義務に合致した形でのみ使用されることを前提とし、軍事 AI に係る設計・開発・配備・使用に関する原則の公表、意図しない偏りを最小化する対策の実施、監査可能な軍事 AI の開発、軍事 AI の安全性・セキュリティ・有効性についてライフサイクル全体にわたる厳格なテストと保証を行うこと等について国が自主的にコミットすることをその内容としており、日本を含む５１か国が賛同した。
https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/

32　2024 年 4 月にイスラエルのネットメディア「+972 マガジン」による調査報道によれば、イスラエル軍は生成 AI「ラベンダー」を用いてガザ地区の 3 万 7 千人を抽出して工作員リスト化し、標的として攻撃する等の行為が行われている。
（出典）Yual Abraham “Lavender’ : The Ai machine directing Israel’ s bombing spree in Gaza” +972 Magazine（April 3, 2024）
https://www.972mag.com/lavender-ai-israeli-army-gaza/
上記調査の詳細については川上泰典「ガザの 3 万 7 千人を標的化：AI マシーン「ラベンダー」の存在明らかに」Yahoo ！ニュース（2024 年 4 月 9 日）も参照。
https://news.yahoo.co.jp/expert/articles/c72d4cbc32aa5577eac494dfd75b43652a20555f

ジにインテリジェンスを配備したパーソナル AI がネットワークで接続されて仮想的に統合運用される分散型モデルも考えられる。また、コア側で学習（training）を行い、エッジ側で推論（inference）を行う集中・分散連携モデルを構成することで低遅延、可用性向上（抗たん性の確保）、プライバシー保護などを実現することが期待される。

特に AI をリアル社会に実装するフィジカル AI（リアル空間とサイバー空間が一体化した CPS（Cyber Physical System）において AI とロボティックス等のアクチュエーターが連携する世界）の普及が見込まれる中、AI 振興における集中と分散のバランス、日本としての勝ち筋、標準化戦略等について議論を深めていく必要がある。

また、データセンターの地方分散は電力需要の地方分散とセットで考える必要があり、電力設備とデータセンター設備を連携させて分散配置するワットビット連携の取り組みが極めて重要な政策課題である。この取り組みは電力需要の分散化のみならず、電力供給管理に必要なインテリジェンスの分散にも貢献することが期待されるところであり、エネルギー政策とデジタル政策の連携強化が求められる。

＜ AI と安全保障＞

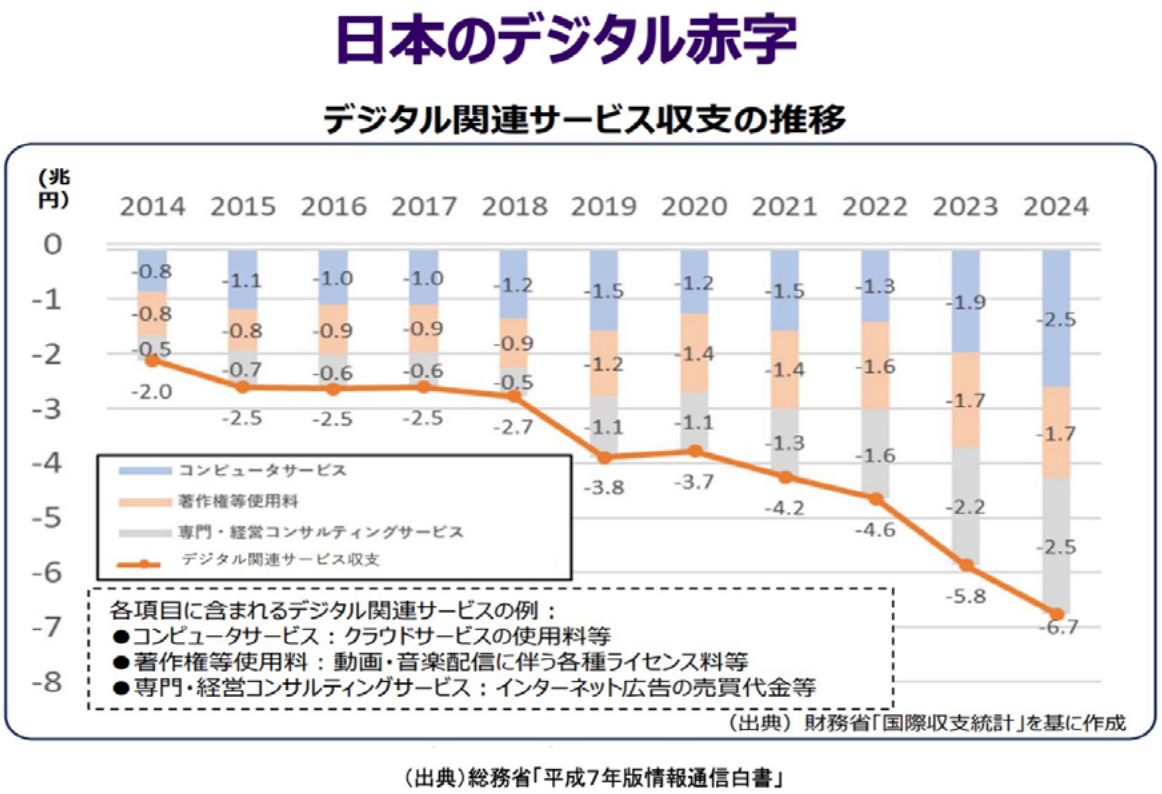
中国における AI 政策は国家の政治方針に合致する国内 AI のみを法律において許容しており、国が AI を通じて国民を統制するサイバー主権の一環と位置付けられる。これに対し、旧西側諸国では非規制を原則とする米国と AI 法による厳しい規制を前提とする欧州の対立はあるものの、表現や報道の自由、そして国家主権を確保する自由主義を旨とするデジタル主権を重要視しており、日本もその陣営に属している。このように、AI ガバナンスのあり方は一国の国家主権のあり方そのものにも密接に関連しており、安全保障戦略のあり方と切り離すことができない。

AI と安全保障の関連性については、認知戦の激化、サイバー攻撃の深刻化、兵器運用のあり方という3つの切り口で考えられる。このうち、認知戦の激化及びサイバー攻撃の深刻化については既に触れたところ（項目（3）を参照）であり、ここでは兵器運用のあり方について触れる。

AI は攻撃側・防御側のいずれも積極的に活用する方向感にあるが、自動火器管制や意思決定支援の領域に留まっており、国家間の安全保障領域での抑止戦略に重大な変化をもたらす状況には至っていない。しかし、今後とも武力行使における AI の活用については国際人道法を遵守する観点から議論を継続する必要がある（項目（8）を参照）。

AI を含むデジタル技術は平時と戦時の境目を曖昧にするグレーゾーン事態を生み、軍事・非軍事の境界線が曖昧なハイブリッド戦争が進展している中、AI ガバナンスのあり方は安全保障と密接に関連した政策課題であることを認識していかなければならない。

これに関連して、日本のデジタル赤字は 2024 年に 6.7 兆円まで拡大している。産業政策として日本の AI 産業を振興しなければ、デジタル赤字はさらに拡大し、安全保障の面でも大きな不安定要素となる。「デジタル主権（digital sovereignty）」を確立する上で AI は中心的役割を果たすものであることを踏まえ、AI を巡る産業政策と安全保障政策をリンクさせた経済安全保障の視点が重要である。



（出典）総務省「平成7年版情報通信白書」

＜AIと民主主義＞

AIが人々のコミュニケーションネットワークに実装されることにより、人と人の間だけではなく、人とAIの間あるいはAIとAIの間によるコミュニケーションの機会が飛躍的に増加する。通常の人と人とのコミュニケーションでは妥協点を探ることでコンセンサスを形成していく自己修正メカニズムをもっている。一般に、民主的な世界において人は議論すれば考えが変わったり、妥協や新たな気づきによって、より多くの人々が合意できる中間解（コンセンサス）が生まれる自己修正メカニズムが働く。

しかし、ボットのアルゴリズムは人々の議論の過程でも変更されることはない。つまり、ボット比率の高いネットワークでは自己修正メカニズムが働かない[33]。このため民主的な議論が有効に機能せず、議論が収斂しないばかりか議論における立場の違い（対立）が先鋭化するといった事態を招く可能性がある[34]。また、こうしたアルゴリズムに特定の偏りを持たせることで認知戦を効果的に進めることも可能となるとの指摘もある。

他方、AIを活用することで膨大の意見の中で傾向を見出すブロードリスニングなどの手法が効果を発揮し、数ある政策の選択肢の中から直接民主主義に近い形で意思決定が可能になるデジタル民主主義の

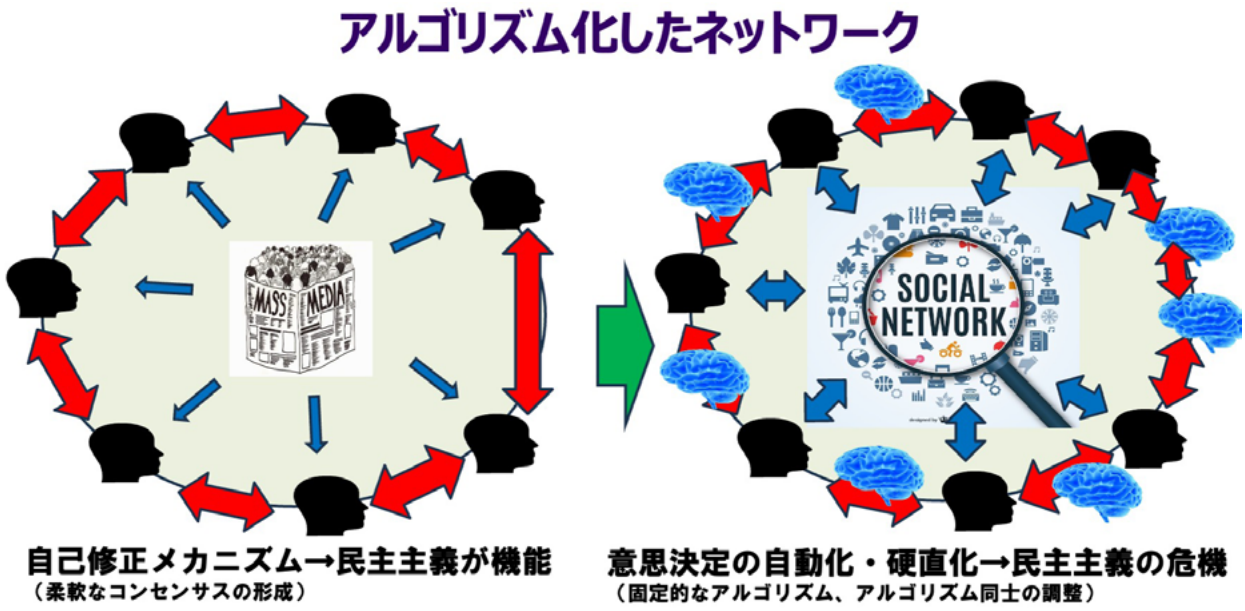
33　ユヴァル・ノア・ハラリ著「NEXUS 情報の人間史（下）AI革命」（2025年3月、河出書房新社））において、著者は民主的な議論が有効に機能しなくなる点について「人や世論を操るのがうまいボットや人知を越えたアルゴリズムが公の場での話し合いを支配するようになったら、私たちがこれまでにないほど民主的な討論を必要としているまさにそのときに、その討論がなりたたなくなりかねない。」と指摘している（同書 p199）。（谷脇康彦「アルゴリズム化されたネットワーク」（2025年11月、DPFJ コラム #30）http://bit.ly/4qlN1IT）

34　前掲（脚注11）“The Global Risks Report 2026”（January 2026）では、こうした状況を”Multipolarity without Multilateralism”（多国間主義が失われ、かつ国内外を問わず極端な意見の隔たり・対立による多極化）が進展している状況と記述している。

可能性も議論されるようになってきている[35]。

このように、AI が普及する中で民主主義にどのような影響がもたらされるのか広範な議論を継続していく必要がある。

また、これに関連して司法プロセスにおける AI 活用のあり方についても議論が求められる[36]。信頼できる司法プロセス（法の支配）を引き続き維持していくために AI はどのような要件が求められるのかについて議論が必要である。



（出典）谷脇康彦「アルゴリズム化したネットワーク」（DPFJコラム#30, 2025年11月6日）

＜総合的・俯瞰的な AI 戦略の推進＞

以上見てきたように、AI 政策はもはや技術的課題の克服やデジタル技術の利用促進という領域を大きく越え、多数の政策領域にまたがるクロスドメインな総合戦略が求められている。既に上記で触れてきたように、産業政策、競争政策、外交政策、安全保障政策などが有機的に連携したものであることが求められている。

日本においては、AI 法の施行を踏まえて 2025 年 12 月に人工知能基本計画が閣議決定された。従

35　例えば、米国建国 250 周年（2026 年）を記念する全国的な対話プロジェクトである”We the People”は、ナポリタン研究所）と Google 系のジャイグソーが協力して立ち上げたプロジェクトであり、「はい・いいえ」で限定的に回答する従来の世論調査等のアプローチとは異なり、AI を対話生成と要約に活用することで、個々人の言葉のニュアンスまで捉えることを試みる新しい公共的対話プロジェクトとなっている。具体的には、①自由記述形式の質問に答え、この回答に合わせてフォローアップ質問を AI で生成して思考を深める「個人表現（individual expression）」、②他の参加者の意見を AI で整理した意見セットを閲覧・評価し、この意見を比較しながら、自分の位置づけを理解する「集団的反映（Collective Reflection）」、③ AI が作成した宣言文（ステートメント）により、参加者の概ね共通する価値観や異なる点を抽出するとともに、自分の考えの位置付けを理解する「検証（Conversation Validation）」という 3 段階のプロセスを経て意見集約が行われる。（出典）Jigsaw “We The People’s first national conversation: Freedom an Eqality”（September 2025）https://bit.ly/49AhWLJ

36　弁護士法第 72 条は「弁護士又は弁護士法人でない者は、報酬を得る目的で訴訟事件（中略）その他一般の法律事件に関して鑑定、代理、仲裁もしくは和解その他の法律事務を取り扱い、又はこれらの周旋をすることを業とすることができない。ただし、この法律又は他の法律に別段の定めがある場合は、この限りではない。」とされている。法務省はリーガルテックの活用範囲について「AI 等を用いた契約書等関連業務支援サービスと弁護士法第 72 条との関係に関するガイドライン」（2023 年 8 月）を定めているが、生成 AI の普及を踏まえ、規制改革会議と連携しつつガイドラインの見直しについて検討を進めており、2026 年夏を目処に結論を出すこととしている。

来の枠を越えるこうした取り組みは評価できるが、計画に盛り込まれた施策は従来の省庁の枠内にとどまるものが多く、施策相互の有機的連携やその必要性に関する記述が盛り込まれる段階には至っていない。

特に、日本における生成 AI の活用は企業内において部分的なものにとどまっており、事業変革をもたらすような事例は未だ限定的である。AI を実装した産業とは、データ駆動型の新たな事業モデルの構築や産業競争力の強化につながる。

そこで、政府における AI 戦略の策定にあたっては、関連する先端性の高い技術開発、半導体の製造・流通、言語モデルの開発、データ流通のための環境整備[37]、知財・著作権などの権利処理の仕組み等、経済安全保障の視点を含む俯瞰的な AI 総合戦略としていくことが求められる。

（10）倫理的問題への対処

AI の急速な進歩に伴い、将来的に「自意識」を持つ AI の可能性も考慮に入れる必要がある。このため、生命科学分野と同様に、AI 研究に関する倫理的問題を検討し、具体的な研究倫理規定や研究承認プロセスを確立すべきである。例えば、「AI に自意識を持たせること」や「自己複製や改変能力をどこまで持たせるか」といった問題に対する倫理的指針を策定し、実装していく必要がある。この問題は「宗教とは何か」という人間にとって根源的かつ精神的な問題にも直結するものだとも言える。

今後の作業計画

冒頭に示したように、本文書の基本テーマは「AI 技術の制御可能性」である。人間と AI は対立する別個の存在ではない。AI はあくまで人間が作り出した道具（ツール）であるということを忘れてはならない。だからこそ、「AI がもたらす影響について人間が最終的なリスク判断を行い、自ら責任をとる環境の整備」を目指す AI ガバナンスが重要になる。その意味で AI を巡る議論は広範な領域に及び、社会的・経済的なガバナスルールのあり方だけでなく、社会構造そのものにどのような影響を与えるか見極めていくことが継続的に求められる。

こうした問題意識を踏まえ、DPFJ は本文書を基に引き続き関係者を交えたワークショップの開催などを通じ、本文書の更新を継続的に行う。併せて、本文書の更新機会などを捉えてオープンフォーラムを開催するなど、広く AI ガバナンスに関する議論を深めていく。その際、同様の議論を進めている他のフォーラムや学会などとの連携を積極的に進め、コンセンサスの醸成を図っていく。

37　データ社会推進協議会（DSA）・デジタル政策フォーラム（DPFJ）・デジタルトラスト協議会（JDTF）提言「データガバナンス戦略の推進」（2024 年 10 月）
https://prtimes.jp/main/html/rd/p/000000009.000131931.html
同「データスペース等に関する国際標準化の必要性」（2025 年 3 月）
https://www.digitalpolicyforum.jp/archive/2503_dss_jp/
同「官民連携によるデータガバナンス戦略の実現」～政府「デジタル社会の実現に向けた重点計画」の決定を受けて」（2025 年 6 月）
同「データ戦略の実現に向けた法制度見直しの方向性」（2025 年 10 月）
https://www.digitalpolicyforum.jp/archive/dsa%e3%80%81dpfj%e5%8f%8a%e3%81%b3jdtf%e3%81%8c%e6%8f%90%e8%a8%80%e3%80%8c%e3%83%87%e3%83%bc%e3%82%bf%e6%88%a6%e7%95%a5%e3%81%ae%e5%ae%9f%e7%8f%be%e3%81%ab%e5%90%91%e3%81%91%e3%81%9f%e6%b3%95%e5%88%b6/