我が国の AI ガバナンスの在り方 ver. 1.1

AI 原則の実践の在り方に関する検討会　報告書

令和３年７月９日

AI 原則の実践の在り方に関する検討会

目次

1. はじめに

AI[1]を構成要素として含むAIシステム、AIシステムの機能を提供するAIサービス、その他付随的サービス、及び、これらを開発、利用、提供する者[2]に関するガバナンスのあり方が国内外で議論されている。国内では、AI戦略2019フォローアップや統合イノベーション戦略2020において、「AI社会原則の実装に向けて、国内外の動向も見据えつつ、我が国の産業競争力の強化と、AIの社会受容の向上に資する規制、標準化、ガイドライン、監査等、我が国のAIガバナンスの在り方を検討する」ことが掲げられている。欧米でも、AIシステムに対する規制のあり方について、基本的な考え方が公表されたり、具体的な規制が検討されたりしている[3]。2020年6月に設立されたAIに関するグローバル・パートナーシップ（GPAI）は、OECDのAI原則の実装に向けた国際的な動きである[4]。

今後、AIガバナンス[5]の議論は、国内外でさらに深まっていくと考えられるが、AIガバナンスの設計は容易ではない。説明可能性の不十分さといったAI特有の課題に横断的に対応することが考えられる一方で、様々な分野や用途に応用可能であるAIは応用ごとに留意点が異なりうるため、それぞれの分野や用途ごとの対応が必要になる可能性もある。また、ガバナンスを実効的にするためには、モニタリングやエンフォースメントについても適切に検討されなければならない。そして、AIシステムやサービスへの懸念に対応しつつ、イノベーションを阻害しないようにしながら、このような複雑で重層的なガバナンスを実現しなければならない。さらに、デジタル・トランスフォーメーションに横断的に関連しうるAIについて、そのガバナンスを整備することは、ウィズコロナ/ポストコロナ時代のデジタル活用を促進するものである。つまり、AIガバナンスは、様々な分野の有識者の知識と経験を結集しなければならないと解決できない喫緊の課題である。

そこで、AI社会実装アーキテクチャー検討会及びAI原則の実践の在り方に関する検討会

---

[1] 本報告書では、現在実用化が進められているのは、人間が知能を使ってすることを機械にさせようとする立場からのAI（「弱いAI」）であるという認識の下、「AI」という言葉を、「弱いAI」、中でも特に機械学習に関する学問分野（研究課題）を意味するものとして説明を行うこととする。経済産業省『AI・データの利用に関する契約ガイドライン 1.1版』を参考にしている。

[2] 総務省 AIネットワーク社会推進会議『AI利活用ガイドライン』の定義を参考にしている。

[3] たとえば、米国連邦政府と欧州委員会からは次のような方針が示されている。Executive Office of the President, Office of Management and Budget, Memorandum to the Heads of Executive Departments and Agencies, “Guidance for Regulation of Artificial Intelligence Applications” (November 17, 2020). European Commission, White Paper on Artificial Intelligence – A European approach to excellence and trust (February 19, 2020) , European Commission, “Proposal for a Regulation laying down harmonised rules on artificial intelligence” (April 21, 2021).

[4] 経済産業省ニュースリリース『AIに関するグローバルパートナーシップが設立されました』（2020年6月16日）。https://www.meti.go.jp/press/2020/06/20200616004/20200616004.html.

[5] 本報告書では、「Society5.0における新たなガバナンスモデル検討会」の定義を参考に、AIガバナンスを「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及び運用」と定義する。

では、憲法、民法、個人情報保護法、国際標準、官民共同規制等と AI との関連に精通した有識者だけでなく、説明可能な AI に関する専門家、AI 開発・利活用の経験が豊富な企業関係者、AI 原則の実践に関する取り組みに詳しい専門家、保険や監査と AI との関係に詳しい実務家、消費者団体の代表などとともに、AI ガバナンスのあり方を検討してきた。このあり方の検討にあたっては、経済産業省の Society5.0 における新たなガバナンスモデル検討会（以下、「ガバナンスモデル検討会」という。）において示された、ゴールベースのガバナンスを参考にしつつ、重層的なガバナンスの構造に関する議論を深めた[6]。

本報告書の第 2 章では、AI ガバナンスに関する国内外の動向をまとめた。まず、AI ガバナンスの議論の流れを説明するとともに、AI ガバナンスの土台となるリスクに関する議論をまとめた。その後、ガバナンスモデル検討会が提示するフレームワークを参考にして、ゴール、中間的なルール、個別分野等にフォーカスしたルールからなる、ユニバーサルな AI ガバナンスの構造を規定し、国内外の AI 原則、横断的な規制、ガイドライン、標準、特定の利用態様への規制、特定分野における規制、政府利用への規制の動向をまとめるとともに、モニタリング、エンフォースメントについても状況を整理した。

第 3 章では、国内外の動向を踏まえつつ、我が国のあるべき AI ガバナンス像を検討した。まず、ガバナンスモデル検討会の示唆等を参考に、Society5.0 時代に一般的に求められるガバナンスの観点から AI ガバナンスのあり方を検討した。次に、本検討会の委員を含むステークホルダーの視点から AI ガバナンスのあり方を議論した。そして、これらを踏まえて、我が国にとって現時点で望ましいと考えられる AI ガバナンスの全体像を提示した。最後に、本検討会では十分に検討できていない課題をまとめた。

AI ガバナンスについては、マルチステークホルダーが関与し、多様な視点から検討がなされなければならない。上述したとおり、AI 社会実装アーキテクチャー検討会は、様々なバックグラウンドを持つ有識者から構成されていたが、さらなる多様性と包摂性を求め、広く意見を求めるために、中間報告書を公表した。今年度は、AI 原則の実践の在り方に関する検討会と呼称し、パブリックコメントを踏まえて有識者の多様性を高めるとともに、議論を継続してきたところ、今般、我が国の AI ガバナンスの在り方 ver. 1.1 を公表する。

---

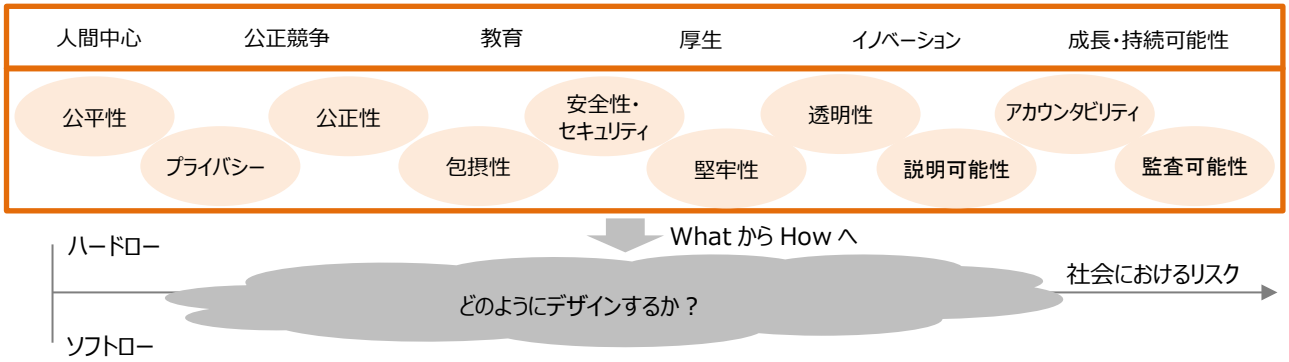
[6] Society5.0 における新たなガバナンスモデル検討会では、ビッグデータ、IoT、AI などデジタル技術が社会を急激に変えていく中で、「イノベーションの促進」と「社会的価値の実現」を両立する、新たなガバナンスモデルの必要性と、その在り方について検討が行われている。経済産業省ニュースリリース『「GOVERNANCE INNOVATION： Society5.0 の実現に向けた法とアーキテクチャのリ・デザイン」報告書を取りまとめました』（2020 年 7 月 13 日）、https://www.meti.go.jp/press/2020/07/20200713001/20200713001.html. 以下、この報告書を「ガバナンス・イノベーション報告書」という。なお、2021 年 2 月 19 日に「GOVERNANCE INNOVATION Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」報告書（案）が公表されている。https://www.meti.go.jp/press/2020/02/20210219003/20210219003.html. パブリックコメントを反映した「GOVERNANCE INNVOATION Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」は、2021 年 7 月中に公表される予定。

2. AI ガバナンスをめぐる国内外の動向

A. AI 原則からガバナンスの具体化へ

政府や国際機関・グループだけでなく、企業やマルチステークホルダーなどが、AI 原則を公表している。Jessica Fjeld らがまとめた AI 原則年表によれば、2016 年 9 月の Partnership on AI による信条（Tenets）に始まり、アシロマ AI 原則、Microsoft や Google による AI 原則などが続き、日本や欧州などからも AI 原則等が提示されてきた[7]。2019 年 5 月には、初めて複数国で合意された AI 原則が OECD から公表された。OECD の AI 原則は、包摂的な成長、持続可能な開発及び幸福、人間中心の価値観及び公平性、透明性及び説明可能性、堅牢性、セキュリティ及び安全性、アカウンタビリティからなる。同年 6 月の G20 首脳会合では、OECD の AI 原則を首脳宣言の附属文書とし、G20 AI 原則が合意された。

AI 原則の例



AI 原則については概ねコンセンサスが形成されつつあるところ、テーマは AI 原則から、AI 原則を社会で実現するためのガバナンスの議論に移行しているといわれている[8]。ガバナンスの要素には、法的拘束力のある法令、法的拘束力のないガイドライン、自主規制など様々なものがあるが、基本的な価値を守りつつ、イノベーションを促進するために、これらの要素をどのように組み合わせるのかが問われている。この点、ガバナンスモデル検討会が提示したゴールベースのフレームワークは、AI ガバナンスを設計する上で参考になる（ガバナンスモデル検討会の報告書の概要については、次頁のコラムを参照されたい）。

なお、AI 原則の議論が終了したわけではない。国際連合教育科学文化機関（UNESCO）では、グローバルスタンダードを目指し、包括的な AI 倫理に関する報告書がまとめられている。世界中から意見を受け付けるコンサルテーションや加盟国からのコメント、政府間協議を経て、2021 年の UNESCO 総会での採択に向けた手続きが進められる予定である[9]。

---

[7] Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar, “Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI” Berkman Klein Center for Internet & Society (2020).

[8] 江間有紗『AI 社会の歩き方：人工知能とどう付き合うか』（2019 年 2 月 28 日）、第 2 章 6 三（WHAT から HOW の議論へ移行）。

[9] UNESCO, “Elaboration of a Recommendation on the ethics of artificial intelligence,” https://en.unesco.org/artificial-intelligence/ethics#recommendation.

コラム：ガバナンス・イノベーション

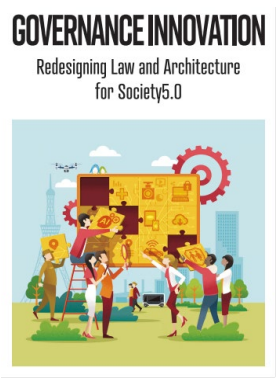
デジタル化を起点とする急速な変化の時代において、法規制を中心とする従来型のガバナンスモデルでは、イノベーションのスピードに追い付くことが困難である。そのため、一方で法がイノベーションのもたらす新たなリスクをコントロールできず、他方で法がイノベーションの発展を阻害してしまうという問題が指摘されてきた。こうした問題意識は、2019 年 6 月に大阪で開催された G20 の参加国にも支持され、G20 貿易・デジタル経済大臣会合の閣僚宣言では、「ガバナンス・イノベーション」の名の下に、各国が、イノベーションが起こりやすい政策を目指して努力すると共に、それに応じてイノベーションに対する障害を取り除くことを目指すことが宣言された。

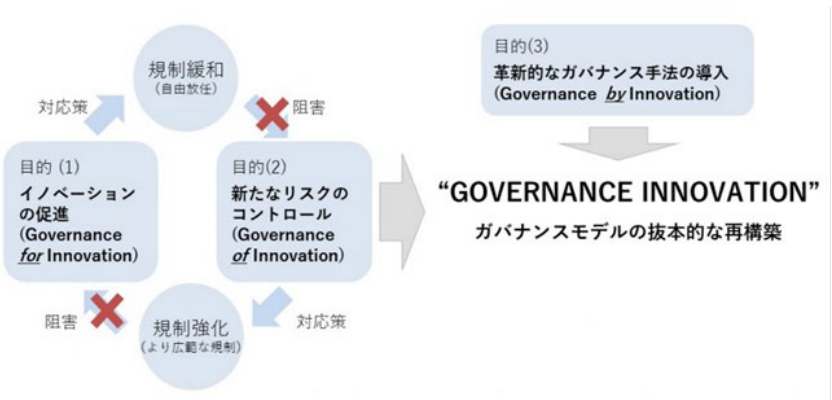
このような動きを背景として、ガバナンス・イノベーション報告書は、国がルールの設計から監督と執行までを一手に担う従来型のモデルから脱却し、企業がルール設計とモニタリング、エンフォースメントの中心的な担い手となっていくべきであり、ルール形成・モニタリング・エンフォースメントのガバナンスの各プロセスにおいて、サイバー空間及びフィジカル空間のアーキテクチャーを設計・運用している企業や、これらを利用するコミュニティ・個人による、ガバナンスへの積極的な関与を確保することが重要であると提言した。

ルール形成については、社会のスピードや複雑さに法が追い付けない問題を克服するために、規制を、細かな行為義務を示すルールベースから、最終的に達成されるべき価値を示すゴールベースにする。マルチステークホルダーの関与の下で、企業のゴール達成に向けて参照されるべきガイドライン等を整備する。このガイドライン等は、モニタリング段階で収集されたデータや、エンフォースメント段階における当事者の主張等を参照し、その効果や影響の評価を継続的に行い、頻繁に見直しの機会を設ける。その際は、ガバナンスに必要な情報が民間主体に集中していること（情報の非対称性）を踏まえ、企業自身による自主規制を促すため、企業が保有する情報をガバナンスに活用するようなインセンティブ設計を行う。市場や社会規範による規律を有効に機能させるため、情報開示に関する義務付けやインセンティブ設計（透明化ルール）を充実させる。

コンプライアンス・モニタリングおいては、企業による革新的な手法による法目的達成（コンプライアンス）を促進すると共に、自社の取組みに関する説明責任を重視する（コンプライ&エクスプレイン）。また、社会からの信頼を確保するために、自己チェック、ピアレビュー、内部監査、合意された手続、第三者によるレビューや監査等といった、リスクに応じた様々なアシュアランス（保証）の態様を活用する。企業、政府、個人といった各ステークホルダーが、リアルタイムデータへアクセスして効率的かつ実効的なモニタリングを実施できるような技術や仕組みについて検討する。ステークホルダー間でモニタリングの結果を報告・評価し、今後のルール改正やシステム改善に繋げられるような、定期的なモニタリング・レビューを行う。

＊エンフォースメントについては、紙幅の都合上、割愛する。報告書を参照されたい。





B. リスクベース・アプローチ

AI ガバナンスの設計にあたって、規制の程度をリスクの大きさに対応させるべきという考え方（リスクベース・アプローチ）は、国際的に広く共有されている。欧州委員会の AI 白書では、新しい規制の枠組みは、目的達成に効果的であると同時に過度に詳細な要件（excessively prescriptive）を定めるべきではないから、「欧州委員会は、リスクベース・アプローチにしたがうべきである」と述べている [10]。また米国は、規制導入にあたっては、潜在的な利益も考慮しつつ、受容できるリスクと受容できないリスクを見極めるリスクベース・アプローチで対応すべきであるという立場であって、予見される全てのリスクを緩和する必要はないとも考えており、詳細な要件を定める（prescriptive）規制に否定的である [11]。

産業界や消費者団体もリスクベース・アプローチを支持している [12]。日米財界人会議は「この分野における両国政府のいかなる取組も、既存のルールや規制に留意すべきであるほか、AI ガバナンスにリスクベースのアプローチを採用・・・すべきである。」と述べている [13]。デジタルヨーロッパは、アジャイルでエビデンスを基礎とするリスクベースの政策立案とマルチステークホルダーとの議論を通じて、EU の AI 政策の全体像の基礎とすべきであると述べている [14]。ヨーロッパのテクノロジー産業を代表する Orgalim も、EU の AI への規制アプローチは、リスクベース・アプローチと整合的であるべきだと述べている [15]。欧州の消費者団体の BEUC は、欧州委員会の具体的なリスク評価に異論はあるものの、リスクベース・アプローチ自体は否定していないと考えられる [16]。

C. リスクの評価と分類

リスクベース・アプローチという基本的な考え方を共有していても、具体的なリスク評価や分類については、国・地域、ステークホルダー間で、必ずしも共有されているとはいえな

---

[10] 前掲 3。

[11] 前掲 3。

[12] 日本経済団体連合会は、欧州委員会の AI 白書に対する意見書で、リスクベース・アプローチは重要であると述べている。

[13] 第 56 回日米財界人会議『共同声明の附属文書（デジタル経済）』（2019 年 9 月 18 日）。http://www.jubc.gr.jp/active/pdf/56%20JS%20Japanese.pdf.

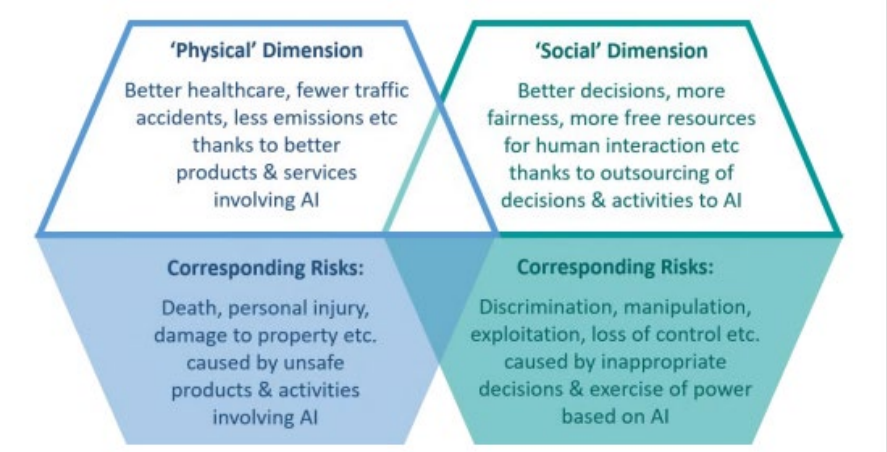
[14] DIGITALEUROPE, “DIGITALEUROPE’s Recommendations on Artificial Intelligence Policy” (November 13, 2019). https://www.digitaleurope.org/resources/digitaleuropes-recommendations-on-artificial-intelligence-policy/.

[15] Orgalim, “Orgalim Manifesto: a European Agenda on Industrial AI” (January 15, 2020). https://www.orgalim.eu/sites/default/files/attachment/Orgalim%20Manifesto%20for%20a%20European%20Agenda%20on%20Industrial%20AI%2015.01.2020.pdf.

[16] Bureau Européen des Unions de Consommateurs (BEUC), “BEUC’S RESPONSE TO THE EUROPEAN COMMISSION’S WHITE PAPER ON ARTIFICIAL INTELLIGENCE” (June 12, 2020). https://www.beuc.eu/publications/beuc-x-2020-049_response_to_the_ecs_white_paper_on_artificial_intelligence.pdf. “The proposed risk-based approach for the development of the new legal framework on AI and ADM should be revised and broadened”, p2.

い状況にある。たとえば、EU 域内でも、国ごとに意見が分かれている可能性がある点に留意すべきである [17]。

この論点について、いくつかの議論等を紹介する。まず、リスクの大きさの区分に関する提案や議論がなされている。欧州委員会の AI 白書は、ハイリスクであれば法的拘束力のある規制を課し、ハイリスクでなければそのような規制を課さないというバイナリーアプローチを採用している [18]。他方で、ドイツ政府が 1 年の時限的に設置したデータ倫理委員会は、リスクを 5 段階に分け、それぞれのリスクに対する規制のあり方について意見を述べている [19]。欧州の消費者団体は、AI 白書のようなバイナリーアプローチを採用すると、規制対象が狭くなり、規制が効果的でなくなってしまうと批判している [20]。



脚注 20 のウィーン大学等からの意見書の図を引用

用途に応じて AI リスクを分類する議論も見られる [21]。ウィーン大学、アルベルト・ルートヴィヒ大学フライブルク、European Law Institute は、安全性を欠いた製品や活動に起因

---

[17] Position paper on behalf of Denmark, Belgium, the Czech Republic, Finland, France Estonia, Ireland, Latvia, Luxembourg, the Netherlands, Poland, Portugal, Spain and Sweden, "INNOVATIVE AND TRUSTWORHTY AI: TWO SIDES OF THE SAME COIN" (October 8, 2020). https://em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf.

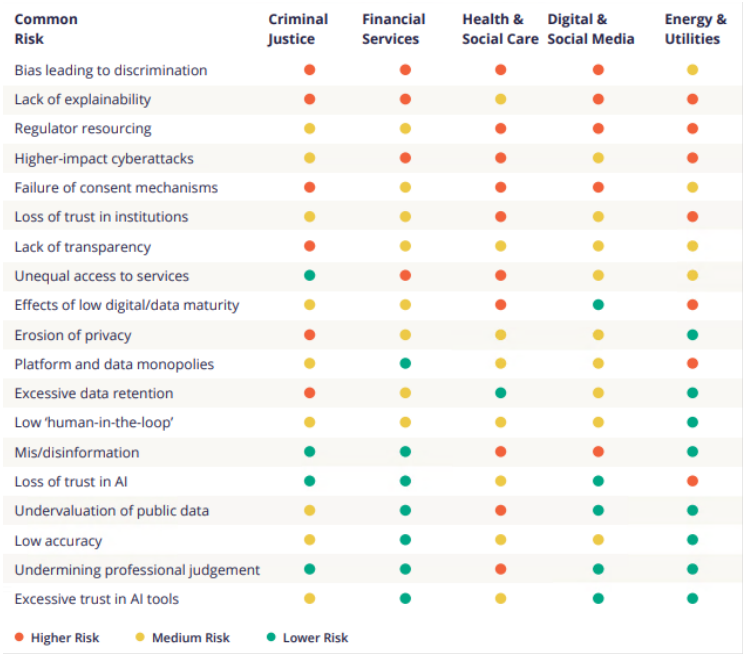
[18] 欧州委員会が 2021 年 4 月 21 日に公表した AI 規則案は、必ずしも相互に排他的なわけではないが、4 つのリスクレベルを提示していると評価されている。

[19] The Data Ethics Commission, "Opinion of the Data Ethics Commission" (December 2019). https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.pdf. 5 つのリスクレベルの概要は次のとおり。Level 1: 害がゼロ（実質ゼロ含む）→ 規制なし、Level 2: ある程度の害 → ニーズがあれば規制、Level 3: 定常的または無視できない害 → ライセンス、Level 4: 著しい害 → 監視や透明性義務、Level 5: 批判に耐えられない害 → 禁止。

[20] 前掲 16。

[21] The University of Vienna, the University of Freiburg, and European Law Institute, "Response to the public consultation on the White Paper: On Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final" (June 16, 2020). https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/News_page/2020/ELI_Response_AI_White_Paper.pdf.

する死傷や物損などを「物理的なリスク」に、AI の関与の下での不適切な意思決定や権力行使がもたらす差別、操作、搾取や、コントロールの喪失などを「社会的なリスク」にそれぞれ区分し、それぞれの区分に応じた規制のあり方を論じている。物理的なリスクに対しては、既存の Product Safety 指令、Product Liability 指令、分野別の規制の適用可能性を整理し、必要に応じて既存法令を拡張すればよいと述べる一方で、社会的なリスクに対しては新しい立法の必要性を提案している。



Common Risk	Criminal Justice	Financial Services	Health & Social Care	Digital & Social Media	Energy & Utilities
Bias leading to discrimination	●	●	●	●	●
Lack of explainability	●	●	●	●	●
Regulator resourcing	●	●	●	●	●
Higher-impact cyberattacks	●	●	●	●	●
Failure of consent mechanisms	●	●	●	●	●
Loss of trust in institutions	●	●	●	●	●
Lack of transparency	●	●	●	●	●
Unequal access to services	●	●	●	●	●
Effects of low digital/data maturity	●	●	●	●	●
Erosion of privacy	●	●	●	●	●
Platform and data monopolies	●	●	●	●	●
Excessive data retention	●	●	●	●	●
Low 'human-in-the-loop'	●	●	●	●	●
Mis/disinformation	●	●	●	●	●
Loss of trust in AI	●	●	●	●	●
Undervaluation of public data	●	●	●	●	●
Low accuracy	●	●	●	●	●
Undermining professional judgement	●	●	●	●	●
Excessive trust in AI tools	●	●	●	●	●

● Higher Risk ● Medium Risk ● Lower Risk

脚注 23 の AI Barometer より引用

このような抽象的・理論的な分析等のほか、具体的な応用類型別のリスクを検討した調査もある。英国のデジタル・文化・メディア・スポーツ省（Department for Digital, Culture, Media and Sport）に設置されたデータ倫理・イノベーションセンター（Center for Data Ethics and Innovation）[22]は、AI バロメータという報告書を公表している[23]。この報告書では、刑事司法、金融サービス、ヘルスケア・ソーシャルケア、デジタルメディア・ソーシャルメディア、エネルギー・公共事業について、差別につながるバイアス、説明性の欠如、サイバーアタックのインパクト、透明性の欠如、プライバシーの漏洩などのリスクを、高リスク、中リスク、低リスクの 3 段階で評価している。

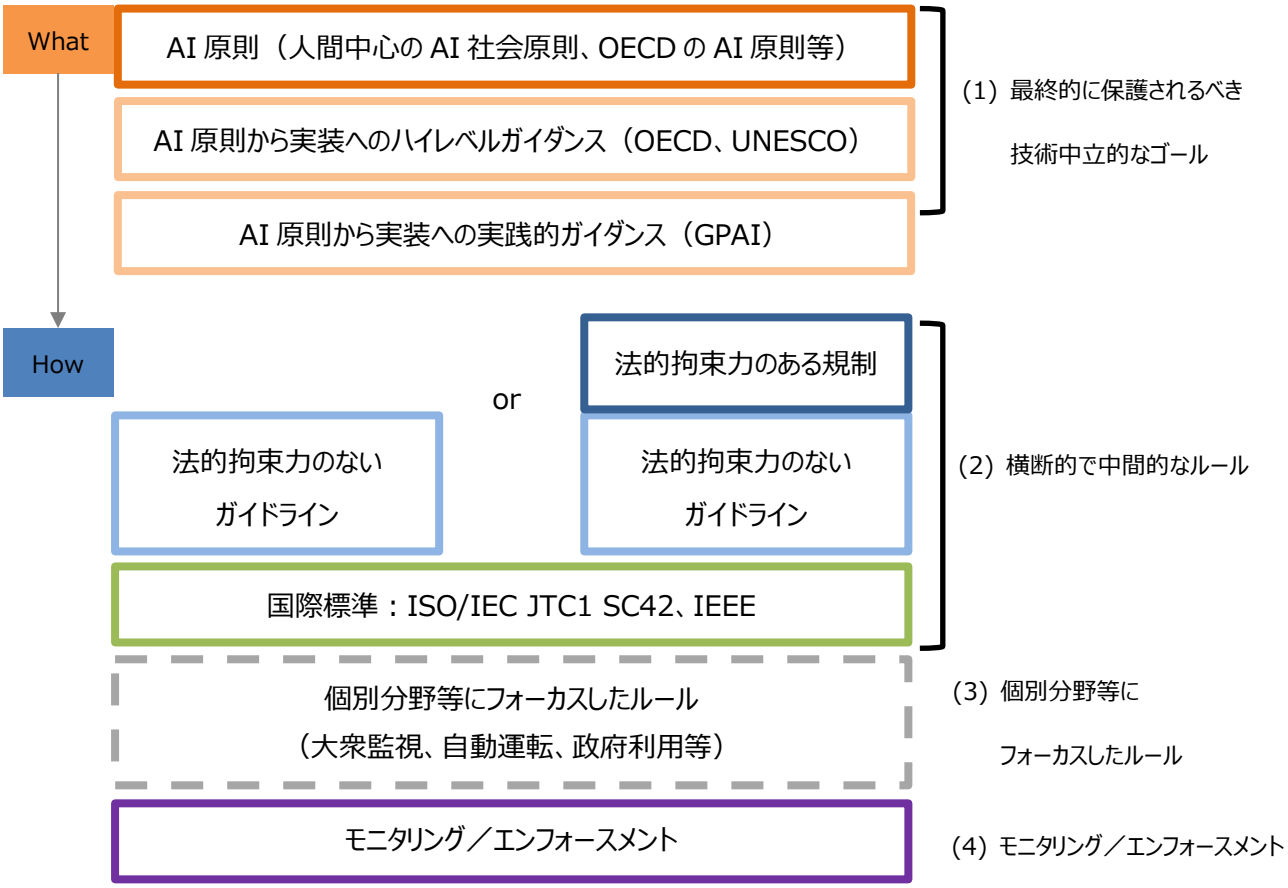
---

[22] 政策担当者、産業界、市民社会、パブリックをつなぎ、データ駆動型技術のためのあるべきガバナンスを発展させるための組織。

[23] UK Center for Data Ethics and Innovation, "AI Barometer" (June 18, 2020). https://www.gov.uk/government/publications/cdei-ai-barometer.

D. AI ガバナンスの構造

本報告書では、ガバナンスモデル検討会が提示している、Society5.0 時代のガバナンスの全体構造を記述するためのユニバーサルなフレームワークを参考に、各国・地域で議論されている AI ガバナンス要素を、（1）最終的に保護されるべき技術中立的なゴール、（2）ゴールを達成するための横断的で中間的なルール（横断的な規制、横断的なガイドライン、国際標準）、（3）個別分野等にフォーカスしたルール、さらには、（4）モニタリングとエンフォースメントからなるレイヤーに区分して整理した。



AI 原則については国際的にも緩やかにコンセンサスが形成されつつあるが、AI 原則を実践する AI ガバナンスの議論は各国・地域で比較的閉じており、AI ガバナンスに関するユニバーサルな全体構造については、少なくとも露わには、ほとんど議論されていない。これまでの国内外の議論を最大限尊重しつつ、AI 原則と同様に、国際的に望ましい AI ガバナンスの全体像を議論し、発信していくべきであると考える。もっとも、AI は国境を越えて利活用される技術であるから、このような議論の進め方は当然の流れともいえる。

(1) AI 原則（ゴール）

上述したとおり、AI 原則については、国際的に概ねコンセンサスが形成されつつある。我が国は、『人間中心の AI 社会原則』（統合イノベーション戦略推進会議決定）において、人

間中心の原則、教育・リテラシーの原則、プライバシー確保の原則、セキュリティ確保の原則、公正競争確保の原則[24]、公平性、説明責任及び透明性の原則、イノベーションの原則を掲げている。また、我が国を含む複数国間で合意した OECD の AI 原則では、包摂的な成長、持続可能な開発及び幸福、人間中心の価値観及び公平性、透明性及び説明可能性、堅牢性、セキュリティ及び安全性、アカウンタビリティが挙げられている[25]。G20 AI 原則は OECD 理事会勧告を引用して作成されたものであり、G20 貿易・デジタル経済大臣会合の閣僚声明の附属文書として合意され、その後、G20 首脳宣言の附属文書としても合意されている。ちなみに、欧州 AI ハイレベル専門家グループは、人間の自律性の尊重、危害の防止、公正性、説明可能性の 4 つの価値を信頼できる AI の基盤をなす AI 倫理原則として特定し、人間の関与と監視、技術的な堅牢性と安全性、プライバシーとデータのガバナンス、透明性、多様性・非差別・公平性、環境及び社会的幸福、アカウンタビリティの 7 つを信頼できる AI を実現するための要件に挙げている。

AI 原則を構成する諸要素のまとめ方や AI 原則の主たる名宛て人の違いがあるため[26]、AI 原則の要素同士が必ずしも対応するわけではないが、国、地域、国際機関・グループ、企業やマルチステークホルダーから公表された AI 原則を分類した研究によれば、世界の AI 原則は、プライバシー、アカウンタビリティ、安全性とセキュリティ、透明性と説明可能性、公正性と非差別性、人間による技術管理、専門家の責任、人間的な価値の促進からなる 8 つのテーマに区分されるという[27]。国内の AI ガバナンスは、国際的に調和の取れたものとすべきであるから、AI ガバナンスを議論する際には、このような世界的な AI 原則の動向にも留意する必要がある。

(2) 横断的で中間的なルール

(a) 法的拘束力のないガイドライン等

AI 原則の尊重を促す取り組みがいくつかの国で進んでいる。大きく分けて、AI 原則の要素別に解説等を加える「解説アプローチ」と、AI 原則の諸要素を企業等の実務に溶け込ませる「融合アプローチ」がある。解説アプローチの代表例が、欧州 AI ハイレベル専門家グル

---

[24] 公正競争確保の原則は、欧米の原則に対応しにくい例として挙げられることがある。前掲 7, p17.

[25] OECD の AI Observatory では、それぞれの原則に対する理論的説明（rationale）が示されている。https://oecd.ai/ai-principles.

[26] 我が国の『人間中心の AI 社会原則』には、「AI が社会に受け入れられ適正に利用されるため、社会（特に、国などの立法・行政機関）が留意すべき「AI 社会原則」と、AI の研究開発と、社会実装に従事する開発・事業者側が留意すべき「AI 開発利用原則」に体系化する（下線追加）」と記載されていることから、企業に求められている AI 開発利用原則では、7 つの要素を全て網羅する必要はないと解される。特に、公正競争確保の原則とイノベーションの原則については、国などの立法・行政機関に対する期待という側面が強いと考えられる。

[27] 前掲 7。世界の AI 原則をまとめたものとして、Linking Artificial Intelligence Principles という研究もある。http://www.linking-ai-principles.org/.

ープの試みである。彼らは 2019 年 4 月 8 日に公表した『信頼できる AI のための倫理ガイドライン』の中で、信頼できる AI を実現するためのアセスメントリスト案を企業等向けに提示し、2020 年 7 月 17 日にステークホルダーの意見を踏まえた最終アセスメントリストを公表している [28]。このアセスメントリストは、人間の関与と監視、技術的な堅牢性と安全性、プライバシーとデータのガバナンス、透明性、多様性・非差別・公平性、環境及び社会的幸福、アカウンタビリティからなる 7 つの要件別に構成され、質問形式ないしはチェックリスト形式でまとめられている。総務省 AI ネットワーク社会推進会議の『国際的な議論のための AI 開発ガイドライン案』と『AI 利活用ガイドライン』も、諸原則別に留意点を説明する形式を採用しており、解説アプローチに分類できる。米国国立標準技術研究所は、説明可能性を深掘りし、説明できること、人間が理解できること、説明が正確であること、自信を持って説明できる範囲で運用すること、という 4 つの細原則を提案している [29]。

融合アプローチの代表例は、シンガポールのモデルフレームワークである。このモデルフレームワークでは、まず本フレームワークを貫く原則 [30]に簡単に触れた後、企業内部のガバナンス構造と手段、AI が用いられた（AI-augmented）意思決定における人間の関与の度合いの決定、運用マネジメント、ステークホルダーとの相互作用とコミュニケーションの 4 つの観点から、実例を交えながら実務上の留意事項がまとめられている [31]。そして、説明可能性、透明性、公正性などの考え方はフレームワークの中に溶け込んでいる。さらに、世界経済フォーラムとともにまとめられたシンガポールの自己アセスメントガイドも、同じ 4 つの観点から実務上の留意点がリスト形式でまとめられている [32]。英国の『AI を用いた意思決定を説明する（Explaining decisions made with AI）』も融合アプローチの例として挙げることができる [33]。このガイドラインは、データ保護責任者（DPO）やコンプライアンスチー

---

[28] The High-Level Expert Group on AI (AI HLEG), “THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) for self assessment” (July 17, 2020). https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

[29] P J. Phillips, Amanda C. Hahn, Peter C. Fontana, David A. Broniatowski, Mark A. Przybocki, Four Principles of Explainable Artificial Intelligence (Draft), NIST Interagency/Internal Report (NISTIR) - 8312-draft (August 19, 2020). https://www.nist.gov/publications/four-principles-explainable-artificial-intelligence-draft.

[30] 本モデルフレームワークを貫く原則は次の 2 つ。a. 意思決定に AI を用いる組織は、その意思決定プロセスを、説明可能で、透明で、公正となるようにすべきである。b. AI による解決策は人間中心にすべきである。

[31] Info-communications Media Development Authority and Personal Data Protection Commission, “Model Artificial Intelligence Governance Framework Second Edition” (January 21, 2020). https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework.

[32] World Economic Forum, “Companion to the Model AI Governance Framework –Implementation and Self-Assessment Guide for Organizations” (January 2020). https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgisago.pdf.

[33] UK ICO and The Alan Turing Institute, “Explaining decisions made with AI” (May 20, 2020). https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/.

ム、技術部門、上級管理部門のそれぞれに照準を合わせた指針となっている。このガイドラインでは、透明性の確保、アカウンタビリティの確保、運営状況の検討、個人と社会全体へのAIシステムのインパクトの考慮という4つの方針（プリンシプル）が実務に溶け込むように工夫されている。

(b) 法的拘束力のある横断的な規制

特定のAI用途に対する限定的な規制は、後述するように散見されるが、現時点で、用途以外の基準によって複数の用途を包含する集合を特定し、それに対して法的拘束力のある横断的な規制を課している例はない。各国が横断的な規制に慎重になる中、欧州委員会が横断的な規制を検討している。欧州委員会は、2020年2月にAI白書を公表した後、同年7月に公表したAI法制初期影響評価（Inception Impact Assessment）の中で、1. ソフトロー、2. 任意ラベリング、3.a 遠隔生体認証の規制、3.b ハイリスクのAI応用の規制、3.c 全てのAIの応用の規制、4. これらの組み合わせ、からなる選択肢を提示した。3.aから3.cまでのいずれかの規制が採用される場合には、それぞれの対象に対して、AI白書に例示されていた、訓練データ、記録保持、情報提供、堅牢性と正確性、人間の関与などの要件が課されることになる。

なお、このEUにおける法的拘束力のある横断的な規制については、産業界や消費者団体等のステークホルダー間だけではなく、EU加盟国内でも意見が分かれているようである。デンマーク、ベルギー、チェコ、フィンランド、フランス、エストニア、アイルランド、ラトビア、ルクセンブルク、オランダ、ポーランド、ポルトガル、スペイン、スウェーデンの14カ国は、連名で発出されたポジションペーパーにおいて、ソフトローによる解決策に目を向けるべきであるという立場を表明している[34]。

欧州委員会は、2021年4月21日にAI規則案を公表した。この規制案は、受容できないAIシステム、ハイリスクAIシステム、透明性義務を伴うAIシステム、極小リスク・リスクなしAIシステムの4つのカテゴリーを規定し、ハイリスクAIシステムと透明性義務を伴うAIシステムに対していくつかの義務的な要件を課すことを提案している[35]。また、極小リスク・リスクなしAIシステムに対しては義務的な要件を課さず、自主的な取り組みを促している。この4つのカテゴリーはリスク別に分けられているため、ステークホルダーからはリスクベース・アプローチを採用したと受け止められている[36]。

---

[34] 前掲17。

[35] 前掲3。Centre for European Policy Studiesが主催したウェビナーにおける欧州委員会の説明資料がわかりやすい。https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf.

[36] たとえば、Orgalim, “European Regulation on Artificial Intelligence – Orgalim calls for legal clarity and workability” (April 21, 2021). https://orgalim.eu/news/european-regulation-artificial-intelligence-orgalim-calls-legal-clarity-and-workability. “Orgalim welcomes the proposed risk-based approach.”

欧州委員会が提案する要件の中には、義務か任意かの違いはあるが、様々な国や団体が実務的な留意事項として提示してきたものに対応しているものもある。たとえば AI 規則案の 9 条は、ハイリスク AI システムに対して、リスクマネジメントシステムの確立や実施を求めているが、欧州委員会の提案以前から、たとえば、ISO/IEC AWI 42001 において AI Management system の議論がなされている。10 条が要求する訓練、検証、試験用のデータセットの関連性や代表性は、国立研究開発法人産業技術総合研究所（以下、「産総研」という）の機械学習品質マネジメントガイドラインの A-1: 問題領域分析の十分性、A-2: データ設計の十分性、B-1: データセットの被覆性などに関連していると考えられる[37]。15 条が要求する適切なレベルの堅牢性はフェールセーフプランなどの技術的な冗長対策で満たすことが可能であることから、AI プロダクト品質保証ガイドラインが 2019.05 版から推奨している、「システムの品質」の項目における品質事故の回避性に概ね対応していると考えられる。今後の議論においては、国際的な調和の視点が欠かせないだろう。

(c) 国際標準

2017 年 10 月に、情報技術分野の標準化組織である ISO/IEC JTC1（Joint Technical Committee 1）に、AI にかかる国際標準を議論する SC42 が設置された。既に総会を 6 回開催するなど、AI の信頼性を含む様々なワーキンググループ（WG）に分かれて、標準化の議論が進められている。日本では、情報処理学会内の情報規格調査会に SC42 専門委員会が設置され、国内の意見のとりまとめや国際的な対応をしている。このような活動の中、産総研は、AI の研究開発だけではなく、SC42 にかかる実質的な取り組みを推進している。

SC42 は、AI のガバナンス、基礎的標準、データ、信頼性、ユースケースと応用、計算アプローチと計算的特徴の 6 つの WG にわかれており、そのうちの 2 つの WG のコンビナー（主査、とりまとめ役）を日本が確保している。その他にも、AI マネジメント、ビッグデータの品質、AI ライフサイクルなどのアドホックグループが設置されている。信頼性確保のための AI の品質評価手法の議論は日本が先行しており、2020 年 6 月に公表された産総研の機械学習品質マネジメントガイドラインが、その標準化提案の基礎的成果に位置づけられている。

AI 標準の国内委員会は、ISO/IEC JTC1 の活動とは別に、EU の標準化団体である CEN/CENELEC と協力関係を深めている。両者は、テーマを特定して、標準化の専門家同士で議論するとともに、2020 年 9 月には、政府の政策・標準担当者向けにワークショップ（EU-Japan Workshop on Trustworthy AI Standardization and R&D）を開催した。

---

[37] 国立研究開発法人産業技術総合研究所. 機械学習品質マネジメントガイドライン 第 2 版. 2021 年 7 月. デジタルアーキテクチャ研究センター・サイバーフィジカルセキュリティ研究センター・人工知能研究センター テクニカルレポート DigiARC-TR-2021-01 / CPSEC-TR-2021001.
https://www.digiarc.aist.go.jp/publication/aiqm/guideline-rev2.html.

ISO/IEC JTC1 の他に、IEEE でも AI 倫理を中心に標準化の議論が進められている。IEEE は、『倫理的整合性のとれた設計：AI と自律システムにおいて人の幸福を優先するためのビジョン（バージョン１）』を 2016 年 12 月に公表した[38]。そして IEEE は、このビジョンに基づき、IEEE P70xx シリーズの標準化の議論を進めている。具体的なテーマに、倫理的設計のモデルプロセス、自律システムの透明性、データプライバシーのプロセス、アルゴリズミックバイアスなどがある。日本は、自律システムの透明性のテーマの Secretary を確保している。

(3) 個別分野等にフォーカスしたルール

(a) 特定の利用態様に対する規制

欧米では、AI の個人識別能力やプロファイリング能力を活用した利用態様に対して、個別の規制が検討ないし導入されている。ここでは遠隔生体認証と採用時利用への規制を取り上げる。

欧州委員会は、AI 白書や AI 法制初期影響評価において、AI を用いた遠隔生体認証に対して規制を導入する可能性を示していた。AI 白書では、AI を用いた遠隔生体認証やその他の侵入的な監視技術については、欧州の基本的な権利に対するハイリスク応用であるとみなしている[39]。そして実際、欧州委員会は 2021 年 4 月 21 日公表の AI 規則案で遠隔生体認証システムをハイリスク AI システムに分類した。今後、43 条に関連する遠隔生体認証システムに対する整合規格や共通仕様等が整備されることで、横断的規制と特定の利用態様に対する規制のハイブリッドの枠組みが明らかになっていくだろう。米国でも、連邦レベルで、顔認識の商用利用について本人同意を義務づける法案（Commercial Facial Recognition Privacy Act of 2019）や、顔認識ツールを含む生体認証技術の使用を停止する法案（Facial Recognition and Biometric Technology Moratorium Act of 2020）が提出されている[40]。サンフランシスコ市などでは、市民のプライバシーの侵害への懸念から、警察による顔認証技術の利用を禁止する条例を定めている（サンフランシスコ市の条例名：Stop Secret Surveillance Ordinance）。ワシントン州でも顔認証に対する規制が導入されている[41]。一連の動きの中、IBM、Amazon、

---

[38] IEEE, “Ethically Aligned Design: Prioritizing Human Wellbeing with Autonomous and Intelligent Systems” (December 2016). https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf. その後、2017 年 12 月にバージョン２が公表されている。

[39] 前掲 3, European Commission, White Paper on Artificial Intelligence, p18.

[40] Security Industry Association（SIA）は、上記のモラトリアム法の導入について強い反対の立場を示している。SIA, “Security Industry Association Strongly Opposes the Facial Recognition and Biometric Technology Moratorium Act Citing Immeasurable Benefits of the Proven Technology” (June 26, 2020). https://www.securityindustry.org/2020/06/26/security-industry-association-strongly-opposes-the-facial-recognition-and-biometric-technology-moratorium-act-citing-immeasurable-benefits-of-the-proven-technology/.

[41] Reuters, “Washington State signs facial recognition curbs into law; critics want ban” (April 1, 2020),

Microsoft は、現在、警察への顔認識システムの提供を停止している。

米国イリノイ州は、AI を用いた面接に対する規制を導入した（Artificial Intelligence Video Interview Act）。面接を録画した動画に対して AI を用いる場合には、面接者に事前に告知し同意を得ることなどが義務づけられることとなった。連邦レベルでは、連邦取引委員会に対して、バイアスの問題に対処するためにインパクト評価を義務づける権限を付与する法案（Algorithmic Accountability Act of 2019）が、ニューヨーク市では AI を用いて採用候補者を絞り込む自動雇用意思決定ツールの販売を規制する法案が、それぞれ提出されている。また、欧州委員会の AI 規則案においても、採用目的の AI システムはハイリスク AI システムに指定されている。

(b) 特定の分野における規制

業界ごとの特有の事情を考慮して規制が整備されてきた分野では、AI 利活用に起因する課題に対応すべく、垂直的なルールを検討ないし導入する動きが見られる。我が国において、自動車分野では、SAE レベル 3 に対応すべく、道路交通法と道路運送車両法の一部が改正されており [42]、自動運行装置の安全基準も定められている [43]。国際連合欧州経済委員会（UNECE）の「自動車基準調和世界フォーラム（WP29）」では、自動運転に関連する国際基準が議論されており、2020 年 6 月 24 日に、高速道路等における 60km/h 以下の渋滞時等において作動する車線維持機能に限定した自動運転システム等について国際基準が成立した [44]。米国では、州ごとに自動運転のテストや実際の走行（deployment）に対する要件を定めたりするとともに [45]、連邦政府がガイドラインを示している [46]。欧州委員会の AI 規則案は、自動車等の Old Approach を採用する EU 規則の対象製品におけるハイリスク AI システムに対して直接的には適用されないが、これらの規則に関する今後の施行規則等の整備において、AI 規則案の要件が考慮されることになるという。

欧州委員会の AI 白書は、ヘルスケア分野もハイリスクとなりうる分野に挙げていたところ、AI 規則案によれば、医療機器規則及び体外診断用医療機器規則に AI 規則案の要件が反映されることが想定されている。我が国では、医師法第 17 条の規定により医業は医師しか

---

https://www.reuters.com/article/us-washington-tech-idUSKBN21I3AS.

[42] 道路交通法については、警察庁の自動運転に関するサイトを参照されたい。https://www.npa.go.jp/bureau/traffic/selfdriving/index.html.

[43] 国土交通省自動運転車に関する安全基準を策定しました！～自動運転車のステッカーのデザインも決定～』（2020 年 3 月 31 日）。https://www.mlit.go.jp/report/press/jidosha07_hh_000338.html.

[44] 国土交通省『自動運行装置（レベル３）に係る国際基準が初めて成立しました』（2020 年 6 月 25 日）。https://www.mlit.go.jp/report/press/jidosha07_hh_000343.html.

[45] California DMV, AUTONOMOUS VEHICLES, https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/.

[46] US Department of Transportation, USDOT Automated Vehicles Activities, https://www.transportation.gov/AV.

行うことができず、AI を用いた診療を行う場合でも、最終的な判断責任は、診断・治療等の主体となる医師が負うこととされている [47] [48]。英国では、保健省が、データ駆動型のヘルスケア技術のための行動指針を公表している [49]。

シンガポールでは、横断的なモデルフレームワークを補完すべく、金融業界における AI・データ分析の利用の際の公正性、倫理、アカウンタビリティ、透明性の促進のための指針が公表されている [50]。

(c) 政府の AI 利用に対する規制等

政府の AI 利用に対する規制等が提案されたり、公表されたりしている。欧州委員会の AI 白書では、規制対象となりうるハイリスクの AI 利活用の例として、亡命、移民、国境規制などを挙げていた。実際に、欧州委員会の AI 規則案では、たとえば、亡命、移民、国境管理はハイリスク AI システムの 1 つのカテゴリーとして特定され、いくつかの具体的な用途がハイリスクであると規定されている。英国では、政府や公的機関が責任を持って適切にデータを利用できるようにするために、2018 年 6 月に『データ倫理フレームワーク（Data Ethics Framework[51]）』が公表され、その後 2019 年 6 月には、このフレームワークを補完する目的で、アラン・チューリング研究所から『AI の倫理と安全性への理解（Understanding artificial intelligence ethics and safety[52]）』が公表されている。そして同月に、英国政府から、『AI の理解、アセスメント、計画、管理からなる公共部門における AI 利用へのガイド（A

---

47 厚生労働省医政局医事課長通知（医政医発 1219 第 1 号（2018 年 12 月 19 日））には、「人工知能（AI）を用いた診断・治療支援を行うプログラムを利用して診療を行う場合についても、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こととなり、当該診療は医師法（昭和 23 年法律第 201 号）第 17 条の医業として行われるものであるので、十分ご留意をいただきたい。」と記載されている。

48 その他、人工知能技術を利用した医用画像診断支援システムに関する評価指標等の取り組みについては、以下のホームページの資料に詳しい情報がある。https://www.mhlw.go.jp/content/10601000/000590652.pdf.

49 UK Department of Health and Social Care, “Code of conduct for data-driven health and care technology” (18 July 2019). https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology. データ利用に関して公正性、透明性、アカウンタビリティを確保するという要素だけではなく、利用者を理解する、成果とその成果を得るために技術の利用の仕方を決めるなどの要素も含まれており、いわゆる AI 原則というよりはマネジメントガイドに近いと評価できる。

50 Monetary Authority of Singapore, Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore’s Financial Sector (November 12, 2018). パラグラフ 1.4 については、2019 年 2 月 7 日公表のモデルフレームワークの内容を反映する形で更新されている。https://www.mas.gov.sg/-/media/MAS/News-and-Publications/Monographs-and-Information-Papers/FEAT-Principles-Updated-7-Feb-19.pdf.

51 The Government Digital Service, “Data Ethics Framework” (June 13, 2018). https://www.gov.uk/government/publications/data-ethics-framework.

52 The Alan Turning Institute, “Understanding artificial intelligence ethics and safety” (June 10, 2019). https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety.

guide to using artificial intelligence in the public sector[53])』も公開されている。さらに英国は、AI の政府調達のためのガイドラインも定めている[54]。カナダは、政府内での AI の利用の増加を見込み、自動意思決定システムが国民や政府機関に与える影響を低減させ、効率的で、正確で、一貫性があり、解釈可能な意思決定を目指すことを目的とする自動意思決定指令を定め、2020 年 4 月に施行した[55]。この指令によれば、アルゴリズム・インパクト・アセスメント、透明性の確保、品質の確保、異議申立ての機会の提供、実効性に関する報告が求められている。我が国では、政府 CIO 補佐官らが、「AI を政府情報システムまたは政府が提供するサービス等で活用する際に、リスク度に応じて考慮するべき AI の学習データに関する透明性の確保、偏り（バイアス）の排除、データ加工の来歴保管の必要性、権利関係について AI を用いたシステムに求める検討事項および考え方」を論じている[56]。なお、政府機関以外では、AI Now Institute が、公的機関のアルゴリズム・インパクト・アセスメントに関する実用的なフレームワーク[57]を提示したり、New Zealand Law Foundation が、AI を使った予測システムの政府による利用に関するレポートを公表する等している[58]。その他、ドバイ政府は、世界経済フォーラムの協力を得て、AI 技術の政府調達を支援するツールキットを提供している[59]。

(4) モニタリング・エンフォースメント

(a) モニタリング

欧州委員会の AI 白書の F. Compliance and Enforcement には、ハイリスク AI 応用について、監視当局が事後的に監視できるように、十分な文書の記録を要求すべきであると記載さ

---

53 UK Government Digital Service and Office for Artificial Intelligence, “A guide to using artificial intelligence in the public sector” (June 10, 2019). https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector.

54 UK Office for Artificial Intelligence, “Guidelines for AI Procurement” (June 8, 2020). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/890699/Guidelines_for_AI_procurement__Print_version_.pdf.

55 The Government of Canada, “Directive on Automated Decision-Making” (February 5, 2019). https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

56 政府ＣＩＯ補佐官　田丸 健三郎、満塩 尚史、西村 毅、梅谷 晃宏、楠 正憲、細川 努、株式会社 Ridge-i 代表取締役社長 柳原 尚史、総務省情報通信政策研究所調査研究部主任研究官　高木 幸一『AI システムにおけるデータ利用の特性と取扱い上の留意点』（2020 年 6 月）。https://cio.go.jp/dp2020_01.

57 AI Now Institute, “ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY” (April 2018). https://ainowinstitute.org/aiareport2018.pdf.

58 New Zealand Law Foundation, “GOVERNMENT USE of ARTIFICIAL INTELLIGENCE in NEW ZEALAND” (2019). https://www.data.govt.nz/assets/data-ethics/algorithm/NZLF-report.pdf.

59 Government of Dubai, “DEWA participates in launch of ‘AI Procurement in a Box’ toolkit in collaboration with WEF’s Centre for the Fourth Industrial Revolution” (July 8, 2020). https://www.dewa.gov.ae/en/about-us/media-publications/latest-news/2020/07/dewa-participates-in-launch-of-ai-procurement-in-a-box-toolkit.

れている[60]。この点は、欧州委員会の AI 規則案では、当局の求めに応じて要件を満たしていることを示すための文書等を提供するという、ハイリスク AI システム提供者の当局への協力義務として反映されており、さらに、ハイリスク AI システム提供者には上市後のモニタリングシステムの確立と文書化が求められ、重大なインシデントを市場監視当局に報告することも求められている。これは、法的拘束力のある規制の遵守状況を監視し、遵守していない場合に何らかの制裁を加えることで、当該規制の遵守へのインセンティブを確保しようとするものである。他方で、ガバナンスモデル検討会は、企業等にリアルタイムでのモニタリングの実施と、その結果としてゴールに沿った対応をしていることを説明させるメカニズムを設け、政府を含むステークホルダーからのフィードバックを通じて企業等に改善を促し、社会における企業・利用者・政府間の信頼を強めることを目指すべきであると提案している。

(b) エンフォースメント

欧州委員会の AI 白書では、ハイリスクの AI 応用に規制を課した場合の法令遵守確保の方法の例として、事前適合性評価を挙げている。事前適合性評価には、試験、検査、認証の手続きが含まれる可能性があり、このような手続きを設ける場合には既存の適合性評価を基礎にすると AI 白書には記載されていた。実際、欧州委員会の AI 規則案にも事前適合性評価の条項が盛り込まれている。また、第三者に対する損害の回復等について、欧州委員会の AI 白書には、AI システムによる負の影響に対して効果的な法的賠償を確保すべきであると述べられており、同日に公表された Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics において、この論点が整理されている[61]。現在、デジタル時代と AI に適合的な民事責任ルールについて、2022 年第 1 四半期に提案することを目指して議論が進められている[62]。さらに欧州委員会の AI 規則案には、当局による市場監視や罰則も規定されている。これらは、政府の強制力を通じて、法律が規定する状態を確保することを目指すものである。

他方で、ガバナンスモデル検討会が提示するエンフォースメントでは、行為がもたらした社会的影響やリスクの程度を考慮しつつ、対象企業にとって十分なインセンティブとなるような制裁を科せるような環境を整備することが重視されている。仮に、AI システムの不確

---

[60] 前掲 3, European Commission, White Paper on Artificial Intelligence, F. COMPLIANCE AND ENFORCEMENT, pp23-24.

[61] 民事責任について、欧州議会から欧州委員会への提案がなされている。European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)). https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.pdf. 欧州委員会には立法提案時にこの提案を考慮することが期待されているが、受け入れる義務はない。

[62] 欧州委員会は、2021 年 6 月 30 日から 7 月 28 日まで初期影響評価の開始に伴う意見募集を行っている。https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12979-Civil-liability-adapting-liability-rules-to-the-digital-age-and-artificial-intelligence_en.

実性が影響することで、企業がいかに努力してもリスクを回避できないのであれば、制裁を重くしてもリスク回避のインセンティブにはならないので、むしろ AI システムのもたらすリスクを最大限に回避しつつ、その開発を萎縮させないようにする方策が適当であると指摘している。たとえば、AI システムがどうしても回避することができなかった事故等については、原因究明とそれを踏まえた今後の予防の目的に特化した事故調査委員会を設置することや、事故調査委員会の設置や調査にあたって、訴追延期制度・条件付起訴猶予制度を導入することにより、関係者及び関係会社から確実に必要な情報を提出させるようにすることが考えられると述べている。

E. 国際的な調和とレイヤー間の協調

AI ガバナンスの設計に関して、国際的には、GPAI、OECD、UNESCO などのハイレベルな議論や、欧州評議会（Council of Europe）に設置された CAHAI（Ad hoc Committee on Artificial Intelligence）における人権、民主主義、法の支配の分野における欧州評議会スタンダードに適う AI の開発、設計、適用のための法的枠組みの検討[63]、ガイドラインや AI 標準などの中間的なルール形成などが同時並行的に進んでいる。AI 戦略 2019 フォローアップや統合イノベーション戦略 2020 に記載されているように、我が国では「AI 社会原則の実装に向けて、国内外の動向も見据えつつ、我が国の産業競争力の強化と、AI の社会受容の向上に資する規制、標準化、ガイドライン、監査等、我が国の AI ガバナンスの在り方を検討する（下線追加）」こと、つまり、国際的な調和を考慮して AI ガバナンスを検討することが求められている。また、AI 原則、中間的なルール、AI 標準などは、レイヤーが異なるが、それぞれが相互に関連しているため、AI ガバナンスの設計にあたっては、レイヤー間の協調も求められている。

国際的な協調に関して、TPP の土台となった P4 の国々の間で、興味深い動きが見られる。ニュージーランド、チリ、シンガポールは、2020 年 6 月にデジタル経済パートナーシップ協定に署名した[64]。この協定は、WTO の電子商取引交渉を補完するものであり、APEC や OECD 等で進行中のデジタル経済の取り組みを土台として発展させるものである。この協定の 8. 2 条が AI に関するものであり、同 3 項で、信頼でき、安全で、責任ある AI 技術をサポートするための倫理とガバナンスのフレームワーク（AI ガバナンス・フレームワーク）の採用の推進に努力することが定められている。また、豪州とシンガポールは、2020 年 8

---

63 https://www.coe.int/en/web/artificial-intelligence/cahai. 簡潔な日本語文献に、齋藤千紘『人権、民主主義、法の支配と AI 欧州評議会 CAHAI のアプローチ』（2021 年 5 月 11 日）がある。
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/2021_001_04_00.pdf.

64 New Zealand Foreign Affairs & Trade, Overview of the Digital Economy Partnership Agreement is a new initiative with Chile and Singapore, https://www.mfat.govt.nz/en/trade/free-trade-agreements/free-trade-agreements-concluded-but-not-in-force/digital-economy-partnership-agreement/overview/.

月にデジタル経済協定に署名した[65]。この合意は、環太平洋パートナーシップに関する包括的及び先進的な協定と豪星自由貿易協定下のデジタル貿易をさらに発展させるものであるが、AI 協力の MoU が附属しており、その目的の 1 つが、信頼でき、安全で、責任ある AI 技術の発展と利用のための倫理ガバナンス・フレームワークの発展と採択を支援することである。これらのデジタル経済協定や AI 条項ないし附属書の意義は現時点では定かではないが、GPAI や OECD のような多国間の枠組みや二国間の対話とは異なる国際協調であるといえる。

[65] Australian Government, Department of Foreign Affairs and Trade, Australia-Singapore Digital Economy Agreement, https://www.dfat.gov.au/trade/services-and-digital-trade/Pages/australia-and-singapore-digital-economy-agreement.

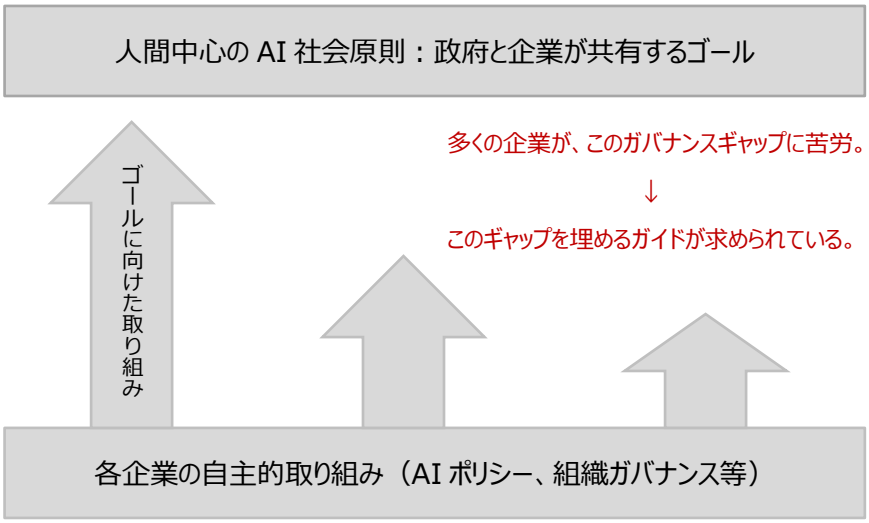
3. 我が国の AI ガバナンスのあり方

A. ガバナンス・イノベーションから得られる示唆

ガバナンスモデル検討会の報告書では、AI 技術の発展やその社会実装のスピードや複雑さに法が追い付けていない問題が指摘されている。そして、このような状況においては、細かな行為義務を規定するルールベースの規制がイノベーションを阻害しうることも同時に指摘されている。これらの問題を克服するためには、従来のルールベース型から、最終的に達成されるべき価値へと企業等を導くゴールベース型へと、ガバナンスの構造を変革することが求められている[66]。私たちの社会は『人間中心の AI 社会原則』を共有しているため、AI 利活用にあたって最終的に目指すべき価値はすでに示されており、さらに国際的なコンセンサスが緩やかに形成されてきていることから、ゴールベースのガバナンスを実現するための基盤は形成されつつあるといえる。

他方で、ゴールベースのガバナンスには、社会的に共有されたゴールと企業レベルでの達成手段との間に大きなギャップが生まれてしまうという短所があることも指摘されている[67]。実際、AI ビジネスの促進を目的とする民間団体では、有志のメンバーに AI 原則を社内規定に落とし込んでもらう試みをする中で、AI 原則に関する社内規定の具体化のためには、中間的かつ実践的なガイドラインが必要であるという結論に至ったという。

このような課題を解決するためには、ガイドラインのような中間的なルールをマルチステークホルダーの関与の下で策定していくことが望ましい[68]。他方で、このような中間的なルールに法令同様の拘束力を与えてしまうと、従来型のルールベースと同様に、イノベーションを阻害してしまう可能性がある点に留意が必要である。これらの示唆や留意点を踏まえると、AI のイノベーションと利活用の促進のためには、非拘束的で中間的なゴールベースのガイドラインが求められているといえる。



66 ガバナンス・イノベーション報告書、5.1.1。
67 ガバナンス・イノベーション報告書、5.1.1。
68 ガバナンス・イノベーション報告書、5.1.2。

B. ステークホルダーの意見

AI ガバナンスの議論では、マルチステークホルダーが対話しながら協力していくことが必要不可欠である[69]。このテーマについては、国内外で様々な議論がなされてきており、本検討会の議論もそのような文脈の中に位置づけられなければならない。たとえば、欧州委員会の AI 白書と AI 法制初期影響評価に対して我が国を含む世界中の企業、産業団体、消費者団体等から提出された意見については、本検討会や個別のヒアリングで得られた意見と合わせて考慮すべきである。また、消費者庁の「消費者のデジタル化への対応に関する検討会」の下に設置された「AI ワーキンググループ」では、我が国の消費者の声を踏まえた議論がなされたことにも留意しなければならない。

(1) 産業界の意見（欧州委員会に寄せられた意見）

欧州委員会は、AI 法制初期影響評価（2020 年 7 月）を公表し、その中で、1. ソフトロー、2. 任意ラベリング、3.a 遠隔生体認証の規制、3.b ハイリスク AI 応用の規制、3.c 全て AI 応用の規制、4. これらの組み合わせ、からなる選択肢を提示した。この影響評価に対するパブリックコメントは、AI ガバナンスのあり方に対する産業界の意見の把握に適した資料である[70] [71]。

まず、AI 応用に対してソフトローを含む何らかの規制が必要であるという方向性については、産業界全体で概ね一致しているようである。AI のような重要な技術には何らかの規制が必要であり、問題はそのアプローチであるというコメントが、産業界の総意を概ね適切に表現していると考えられる。

多くの産業団体や企業は、ソフトローには概ね賛同しているようである。ソフトローに賛同する産業団体や企業の多くは、その手法にもコメントしており、そこから 2 つのメッセージを受け取ることができる。1 つは、産業界の自主的な取り組みを支援する内容とすべきというものである。その中には、官民共同規制に言及している団体もある。もう 1 つは、AI 応用の多様性に配慮すべきというものである。産業別・分野別の取り組みや専門性を活かすことで、産業別・分野別のガイダンスの形成を支援すべきとの指摘がある。

他方で、任意ラベリングへの賛同はあまり多くないようである。否定的なコメントの中には、時期尚早である、対応のための事務的費用が負担になる、という意見が見られる一方で、付加価値がないという厳しい意見もある。欧州 AI ハイレベル専門家グループのアセスメントリストが AI 応用ごとの違いに配慮されていないことを理由に、ラベリング認定のチェッ

---

[69] 総合イノベーション戦略推進会議『人間中心の AI 社会原則』（2019 年 3 月）第 2 頁。

[70] 2020 年 7 月の AI 法制初期影響評価に対するパブリックコメントには、2020 年 2 月の AI 白書に対するパブリックコメント後の議論の進展が反映されていると考えられる。

[71] European Commission, Feedback received on: Artificial intelligence – ethical and legal requirements. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/feedback?p_id=8242911.

クリストとして用いられることに懸念を表明している企業もある。

ハイリスク応用への規制導入については、理論的には賛成するものの、規制対象を慎重に特定すべきという意見が多い。リスク分析の際に、リスクだけではなく機会損失にも目を向けるべきであると指摘する企業もある。また、人間の判断を AI に置き換えたものなどに限定すべきという指摘も見られる。義務的な規制オプションのうち遠隔生体認証への規制導入については、ルールの明確化の必要性や過度な規制への懸念が示されているものの、比較的賛同されているようである。なお、全ての AI 応用に対して規制を導入する提案については、賛同が見られない。

規制手法を組み合わせるべきという意見も見られる。たとえば、AI 応用によって潜在的なリスクは異なるところ、いくつかの規制手法を組み合わせて、リスクに応じて柔軟に対応することが常識的に思われるという意見も見られる。

自動運転と医療機器への AI 応用については、業界の事情を考慮すべきという意見が見られる。European Automobile Manufacturers' Association は、自動車業界には型式認証という厳格な事前規制等が存在する上に、国連欧州経済委員会自動車基準調和世界フォーラム（UNECE/WP29）において自動車分野での AI 応用へのルールに関する国際的な議論が開始されていることから、自動車業界の議論に委ねるべきという意見を表明している。MedTech Europe は、医療機器への AI 応用であっても、厳格な Medical Devices 規制と CE マーク規制が対応していることから、新たな規制は不要であるとの立場を表明している。

なお、欧州委員会が AI 規則案を公表した後に、欧州系の経済団体・産業団体から予備的なコメントが発せられているが、今のところ産業界の意見に大きな変化はないようである [72]。

(2) 本検討会とヒアリングにおける指摘

企業や産業団体からは、3. A.で述べたような AI 原則と企業の取り組みのギャップを埋める中間的なガイドラインの必要性が指摘されている [73]。その一方で、中間的なガイドラインをチェックリスト化することは避けるべきという意見もある。中間的なガイドラインが法的

---

[72] たとえば、Orgalim、DigitalEurope は、リスクベース・アプローチの採用を評価しつつも、ハイリスク AI システムに指定された分野には異論があるようだ。Orgalim, "European Regulation on Artificial Intelligence – Orgalim calls for legal clarity and workability" (April 22, 2021), https://orgalim.eu/news/european-regulation-artificial-intelligence-orgalim-calls-legal-clarity-and-workability, DigitalEurope, "New AI rules must be streamlined for Europe to become a global innovation hub" (April 21, 2021), https://www.digitaleurope.org/news/new-ai-rules-must-be-streamlined-for-europe-to-become-a-global-innovation-hub/.

[73] 我が国の『人間中心の AI 社会原則』の 1 つである「公正競争確保の原則」については、不公正な競争の是正の観点から他の法令等による取り組みが進められているところである。統合イノベーション戦略 2020 等では、AI 倫理の文脈で AI ガバナンスを検討することが要請されているところ、本検討会では、たとえば AI 自体の説明可能性の低さを補うためのガバナンスを主たる論点とし、AI を用いたサービスに関する情報開示の促進については関連分野として注視するにとどめることとする。後者に該当する他国の例に、Guidelines on ranking transparency pursuant to Regulation (EU) 2019/1150（December 8, 2020）がある。

に非拘束であるといわれても、リストを提示されてしまうとチェックが半ば義務であると受け止められ、その結果、形式的な負担増につながってしまうと懸念されている。むしろ中間的なガイドラインは、ゴールを達成するためのプロセスや組織体制の整備を支援する内容にすべきとの指摘が複数の委員からあった。たとえ企業が AI ポリシーを定めたとしても、それだけで AI 原則に沿った行動が促されるわけではないという問題意識がこの指摘の背景にある。

また、企業間取引における共通認識の形成の重要性も指摘されている。データの提供も含め、AI システムの開発や運用が 1 社で閉じることは少ない。そのため、AI システムの開発や運用のプロセスに関して、複数企業間で認識を共有することが求められている。この点について、経済産業省は、ソフトウェアの開発・利用に関する契約の主な課題や論点、契約条項例、条項作成時の考慮要素等を整理した『AI・データの利用に関する契約ガイドライン』を公表しているが、当該ガイドラインは AI 開発契約に関する一般的な留意事項や標準的な条項を扱うものであり、具体的な品質評価やマネジメントの方法を示しているわけではない。また、この品質評価に関して、そのブラックボックス性から AI システム自体の品質を定量化したり保証したりすることが難しいにもかかわらず、AI 開発者に対してそのような要求がなされることがあるという。品質の定量化は重要な課題であるが、それだけではなく、企業内ガバナンスやステークホルダーとの対話によって AI システムに関する広義の品質ないしは信頼を確保するという認識を企業間で共有することが重要である。日本のコーポレートガバナンスは、マルチステークホルダーとの対話を通じた価値の創造と持続可能な成長を重視しながら、従業員や取引先、地域社会などのステークホルダーにも同時にバランス良く配慮するという特徴も備えているとの指摘があるところ、むしろこの特徴を活かしながら、中間的なガイドラインを通じて、サプライチェーン全体で AI システムに対する信頼が確保されるように支援すべきである[74]。

BtoB 企業と BtoC 企業の違いに配慮すべきという意見もある[75]。BtoB 企業が取引相手の要求にしたがって AI システムを納品しても、運用の仕方が悪ければ他者に悪影響を及ぼす可能性がある。このようなリスクの回避を BtoB 企業だけに求めることは酷であり、そのよ

---

[74] 「株主だけではなく従業員の利害にも配慮し、さらに、社会とのつながりをも意識して経営するという日本企業のスタイルは、技術革新の時代にはむしろ適しているのではないか」とケンブリッジ大学のサイモン・ディーキン教授が論じたことを受け、この方向性を模索すべきことを説いたものとして、小塚壮一郎『AI の時代と法』岩波新書（2019 年 11 月 20 日）214 頁以下。サイモン・ディーキン『コーポレートガバナンスと企業のデジタル変革』NBL 1153 号（2019 年 9 月 1 日）41 頁以下。本検討会でも同様の指摘あり。

[75] 松本敬史、江間有沙『AI サービスのリスク低減を検討するリスクチェーンモデルの提案』（2020 年 6 月 4 日）。「米 Google 社や Facebook 社のようにデータの取得・AI モデルの開発・サービス提供の全てを自社で実現している場合には、複数のリスク要因に対して最適なアプローチを実施することが可能である。しかし、日本においてはサービスの提供者・AI モデルの開発者・実行環境の提供者・データの提供者が異なるケースが多く、AI サービス提供者が複数のリスク要因を包括的に検討するためには、関係者間で対話するための枠組みが必要となる。」

うなことを求める one-size-fits-all のガイダンスは望ましくないという指摘もある[76]。この点は、欧州 AI ハイレベル専門家グループのアセスメントリストが必ずしも AI 応用ごとの違いに配慮されていないという 3. B.（１）の指摘と同趣旨であるといえる。

知的財産への配慮が必要という意見もある。欧州委員会の AI 白書では、検討中の法的義務の１つとして、一定期間、データ・記録を保存すること及び当局の求めに応じた提出をすることが挙げられている。AI 白書が明示的に営業秘密の提出に言及しているわけではないが、ソースコード、アルゴリズム等の営業秘密の提出は適当ではなく、自主的な説明を促すことで説明責任への期待に応える方向を検討すべきである[77]。なお、欧州委員会の AI 規則案を見る限り、そのスタンスは AI 白書と基本的に同じである。

(3)　消費者の視点

AI に関する十分な情報が得られていない。そんな消費者像が統計調査から浮かび上がってきている。富士通総研が 2017 年に実施した「性別・世代別にみるロボットや AI（人工知能）への期待と不安」に関する調査では、「まだ具体的なイメージが湧かない」を選択した回答者が最も多い。欧州消費者団体（BEUC）が公表している AI に対する消費者の声によれば、21%の消費者が AI のことを聞いたことがないか、AI が実際に利用されていることを理解しておらず、43%の消費者が十分な情報提供を受けていないと感じている[78]。「ロボットや AI などのテクノロジーの具体的な内容や導入状況を、消費者（AI サービス利用者）が十分把握できていない可能性がある」という消費者庁の分析が現状を的確に表現している[79]。

また、消費者は AI に対して不安を感じている。富士通総研による上記調査では、「システ

---

[76] たとえば、THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) の中にある、”Could the AI system have a negative impact on society at large or democracy? Did you assess the societal impact of the AI system’s use beyond the (end-)user and subject, such as potentially indirectly affected stakeholders or society at large?”については、エンドユーザーあるいは BtoB の顧客の利用の仕方に依存するため、BtoB 企業には不適切な質問であるとの指摘もある。もちろん、このリストの活用は任意であるため、関係ない項目の検討を省略することが想定されているものと思われる。

[77] 営業秘密の強制開示に対する懸念について、公に示されている意見の例には以下のものがある。経団連『欧州委員会“White Paper on Artificial Intelligence: a European approach to excellence and trust”への意見』（2020 年 6 月 10 日）。「当局による事前適合性審査を実施する場合は、競争の源泉になりうる情報（例：アルゴリズム、データセットの詳細情報）の開示を安易に要求すべきではない。開示を要求する内容は、機密を含まない必要最低限の内容とし、根拠を明確にすることが必要である。」 JEITA『ポジションペーパー：欧州 AI 白書：優越と信頼に向けた欧州アプローチ」』（2020 年 6 月 13 日）。「製品・サービスの競争力の源泉になり得るソースコード、アルゴリズムなどは、当局からの要請による提出を義務付けるべきではないと考えます。当局からの要請で提出が必要とする場合があるのであれば、根拠を明確にし、必要最低限の内容にし、機密情報の取扱いには十分配慮することが必要と考えます。」

[78] 前掲 16。

[79] 消費者庁『消費者のデジタル化への対応に関する検討会 第 1 回 AI ワーキンググループ資料 4　AI をめぐる現状』（2020 年 1 月 31 日）。
https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/assets/review_meeting_004_200206_0007.pdf.

ムエラーによって被害が生じそうで不安」という不安に関する項目が上位にある。また、BEUC の調査でも、情報が操作されているのではないか、個人情報を適切に利用していないのではないか、といった声が上がっている。

その一方で、AI に対する期待も示されている。富士通総研による上記調査では、「医療・介護」「自動翻訳」など、サービスの普及が想像しやすい分野では比較的期待値が高い。BECU の調査でも、高い割合で交通事故の予測や病気予測に有用という回答が得られている。消費者庁の「期待と不安の混在」という分析が的を射ている。

このような状況を踏まえ、消費者庁は、事業者の適切な対応を期待するとともに、消費者のリテラシーを向上させて、AI をかしこく使っていくことが望ましいという結論に至り、この目的を達成するため、2020 年 7 月に、「AI 利活用ハンドブック～AI をかしこく使いこなすために～」を公表している。しかし、ここでの課題は、消費者のリテラシーが向上されるだけで解決するものではない。このリテラシー向上という方向性は、AI を利活用する企業等が、消費者に情報を適切に提供することを前提とするものであるから、AI 原則と企業レベルの取り組みのギャップを埋めるための中間的なガイドラインは、このような情報提供を後押しする意味でも必要であると考えられる。

C. 我が国にとって望ましい AI ガバナンス

本節では、これまでの分析を踏まえ、2.D.で示した AI ガバナンスの構造にしたがって、我が国にとって望ましい AI ガバナンスを提案する。

(1) 法的拘束力のない企業ガバナンス・ガイドライン

国内外で公表・検討されている原則、ガイドライン、論文等は、AI の利活用に付随する様々なリスクを例示するとともに、これらのリスクに対処するための留意事項や検討事項をまとめている。しかし、これらの留意事項や検討事項は、原則ごとにまとめられている場合が多く、企業ガバナンスへの変換が容易ではない。また、これらの原則、ガイドライン、論文等には、AI 利活用者すべてに一律に適用される「チェックリスト」を提示する意図はないと考えられるが、各項目の取捨選択は各利活用者の検討に委ねられているため、AI 利活用の経験が豊富ではない企業は全ての項目に対応せざるをえないと考える可能性がある [80]。さらに、国内外で公表・検討されているガイドライン等では、リスクの評価へのガイダンスが不足しているため、必要以上に潜在的なリスクに悩まされ、その結果、AI の利活用が阻害される可能性がある [81]。

リスク等への柔軟な対応を支援するためには、リスクの大きさ等に応じた対応や組織のあり方について、一定のガイダンスを提供することが有効である。このようなリスクベースのマネジメントは、企業ごとの柔軟な対応を前提とするゴールベースのガバナンスの自然な帰結でもある。小規模な AI 利活用におけるリスクへの対応や組織のあり方は、大規模な利活用におけるそれとは当然異なるであろう。また、具体的な AI の利活用の場面によってもリスクは異なる。このようなリスク評価やマネジメントのあり方も含めたガイドラインを提供することで、AI の社会実装をより促進できると考えられる。

ガイドラインの作成にあたっては、AI を利活用している企業を中心に、利用者・技術者・アカデミア・法律や監査の専門家といった幅広いステークホルダーによって議論されるべきである。政府は、そうした議論のファシリテーターとして機能することや、策定されたガイドラインを企業が満たしていることについて客観的に評価することで、当該企業に対する社会の信頼を醸成する役割を担うことが望ましいと考えられる [82]。また、AI の利活用に関する経験の違いに留意する必要がある。AI の利活用に関して経験豊富な企業もあれば、これ

---

[80] たとえば、AI のプロファイリング機能活用時の包括的な自主的チェックリストとして以下のものが広く知られている。パーソナルデータ＋α 研究会『プロファイリングに関する提言案付属 中間報告書』 NBL No. 1137（2019 年 1 月）。

[81] 経団連『AI 活用戦略～AI-Ready 社会の実現に向けて』（2019 年 2 月 19 日）、第 34 頁「例えば AI の信頼感確保のため、政府において AI 活用の領域、社会的影響等、さまざまな背景を考慮せずに一律に説明責任等を活用の主体に義務化するといったことは、AI 活用を阻害する懸念があり、不適当である」と記載されている。ガイドライン等が法的非拘束であっても、一定の影響力があることには配慮が必要である。

[82] ガバナンス・イノベーション報告書、5.1.2。

から積極的に活用していこうと考えている企業もあるだろう。AI を長年にわたり大規模に活用してきた企業の洗練されたガバナンスを、AI を小規模に活用し始めたばかりの企業に半ば強制してしまっては、AI の利活用を停滞させてしまうことになりかねない。また、経験が少ない企業には、事例レベルの具体的なガイダンスも有効である。

3.B.におけるステークホルダーの意見に加え、これらの要素を考慮すると、非拘束の中間的なガイドラインでは、特定の利活用経験レベルを基準にしない[83]、すべての企業に一律に適用されるものとはしない、先進的な取り組みをしている企業の自由な取り組みを妨げない、AI に関するリスク管理等の改善を支援する、企業間取引における AI システムの信頼評価で参照されうるものとする、AI 利活用を始めたばかりの企業でも役立つグッドプラクティスを含める、消費者等への説明を促すものとする、などの事項に配慮する必要がある。そして、非拘束の中間的なガイドラインのアウトラインとして、次のようなものが考えられる。

- AI 利活用の基盤作り：取組みの全社化、AI ガバナンスへの意識向上、AI リテラシーの向上
- AI システムの開発・導入：プリンシプルの作成、マネジメント体制の整備[84]、エスカレーションプロセスの確立、リスクマネジメントプロセスの策定[85]
- AI システムの運用：モニタリング、内部監査、外部評価の活用、ステークホルダーとの関係構築、改善と進捗管理

また、非拘束の中間的なガイドラインの作成にあたっては、諸外国の取り組み等を参考にしつつ、日本の企業ガバナンスの特徴にも目を向けるべきである。従業員や社会とのつながりを大切にしてきた日本企業は、欧米企業よりも AI リスクに対して優位にあるという研究もある[86]。当該ガイドラインには、日本企業の優位性を引き出す工夫が求められている。

---

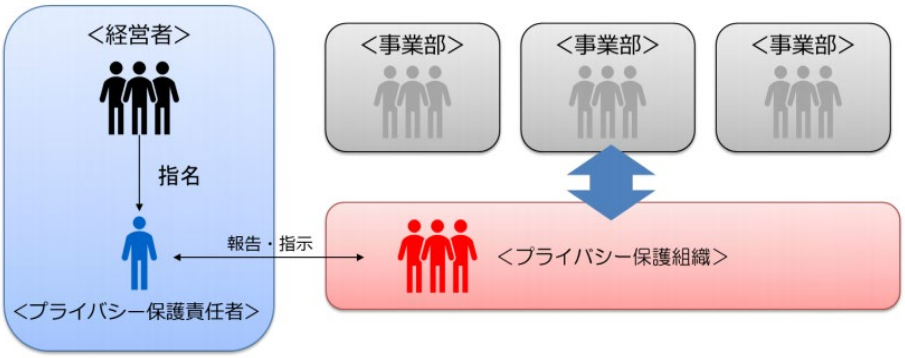
[83] AI の利活用水準の目安を提示したものに、経団連の AI-Ready 化ガイドラインがある。

[84] AI 利活用の多くはデータガバナンスの下流に位置づけられるものであるから（「AI 戦略は、大量の質の高いデータの存在を前提に、人工知能を活用して社会全体を高度化するものであり、データ戦略は AI 戦略の前提部分を担うものと位置付けられる。」（データ戦略タスクフォース第一次とりまとめ（案））、AI に関するガイドラインでデータガバナンスを扱わなければならないわけではないが、AI 利活用にはデータガバナンスが密接に関連するため、少なくともデータガバナンスの議論との整合性には留意する必要がある。

[85] AI システムでは、従来的な安全性だけではなく、意図した公正性を達成できない事態等もリスクに関連することから、諸目的に対する不確かさの影響（ISO31000）というリスクの定義が馴染むと考えられる一方で、安全性の文脈では、害とその発生する蓋然性を掛け合わせたものという従来的な定義を採用した方がマネジメントしやすいとも考えられる。

[86] サイモン・ディーキン『コーポレートガバナンスと企業のデジタル革命』NBL 1153 号（2019 年 9 月 1 日）。なお、この論考が指摘している、企業から見た 3 つの AI リスク（バイアスのリスク、透明性の欠如、技術的なオーバーリーチ）は、ガイドラインの作成にあたって参考となる論点である。

なお、AI に関するガイドラインの作成にあたっては、DX 化を進める企業の採用を促進するため、デジタル時代のガバナンスに関する他のガイダンスとの整合性に留意する必要がある[87]。



DX 時代における企業のプライバシーガバナンスガイドブックから引用

(2) 国際標準

ISO/IEC JTC1 SC42 国内委員会（情報処理学会の情報規格調査会に SC42 専門委員会を設置）は、様々な WG 等でコンビナー（主査、とりまとめ役）を確保するなど、主導的に交渉を進めている。前章で述べたように、国内委員会は、EU の標準化機関である CEN/CENELEC と良好な関係を築きつつある。この良好な関係は、欧州委員会の通信ネットワーク・コンテンツ・技術総局と経産省の間で設置された AI 共同委員会の活動にも波及している。この AI 共同委員会は、政策サイドと標準サイドをつなぐ対話の場として期待されている。今後も、SC42 国内委員会が日本政府と連携しながら、日本の提案が反映されるように活動していくことが重要である。

(3) 法的拘束力のある横断的な規制

産業界の意見や「AI 利活用ハンドブック」によるリテラシー向上の方向性を踏まえると、AI システムに対する横断的な義務規定は現段階では不要であると考えられる。将来、横断的な義務規定が議論される場合でも、リスクだけではなく潜在的な利益も考慮したリスクアセスメントを実施すべきである。そして、その際には、技術の発展によって特定のリスクが解消される可能性も考慮すべきである。

また、AI を用いた特定の技術自体を義務的規制の対象とすべきではない。義務的な規制

[87] 例えば、プラバイバシーガバナンスとの整合性を確保するために、ステークホルダーとの関係構築の項目において『DX 時代における企業のプライバシーガバナンスガイドブック』の対応箇所を引用したりすることが考えられる。今般の個人情報保護法の改正においては、いわゆる「プロファイリング」に対する懸念に対応して、①利用停止・消去等の要件の緩和、②不適正利用の禁止、③第三者提供記録の開示、④提供先において個人データとなることが想定される情報の本人同意、といった規律を導入された（個人情報の保護に関する法律等の一部を改正する法律（概要））。また、個人情報の利用目的の特定におけるプロファイリグの扱いの議論にも留意が必要である（第 155 回 個人情報保護委員会）。

が必要な場合でも、意図しない領域にまで規制が及ばないように、AI の応用分野や用途について慎重に範囲を定めるべきである。なぜなら、技術の具体的な使われ方等（利用分野、利用目的、利用規模、利用場面、影響を及ぼす対象が不特定か否か、事前周知が可能か否か、オプトアウト可能か否か等）によって、社会に与える利益や損害の可能性は異なるからである。たとえば、スマートフォンなどで AI を用いた顔認証技術が活用されているが、第三者による端末の不正ログインを防止できる効果がある上に、利用者本人は AI を利用していることを理解している場合も多く、性能が悪ければ利用しないことも可能である。それに対して、市街地のいたるところに設置されたカメラによって不特定多数の行動を監視する場合には、防犯という正当な目的であっても、プライバシー保護の観点を踏まえつつ、革新的技術の利活用による社会的便益を阻害しないよう、その運用指針等を検討する必要があるかもしれない。なお、ここでの議論は、次の「個別分野等にフォーカスした規制」の議論にも関係する。

(4) 個別分野等にフォーカスした規制

特定の分野においては、情報技術の側面からではなく、業法を所管している組織が規制に関わる方が望ましい場合もあると考えられる。AI を導入したからといって、これまで蓄積した安全に対する考え方を直ちに活用できないほど変わってしまうとは考えない方がよい場合もあるだろう。たとえば、自動車分野や医療分野では、これまでの規制に対する考え方や設計思想を活かし、それぞれの分野におけるルールメイキングを尊重した方が望ましいと考えられる。

D. 今後の課題

(1) 非拘束の中間的なガイドラインを利用するインセンティブの確保

非拘束の中間的なガイドラインは、法的に非拘束であるため、当該ガイドラインを利用するインセンティブが不十分となり、AI 原則を尊重した AI の社会実装の促進という目標を十分に達成できない可能性がある。

このガイドラインの利用を促進するためには、利用することによるビジネス上の意義を周知したり、利用することが利益につながるようなメカニズムを導入したりすることが考えられる。たとえば、AI システムの政府調達先を決める際に AI 原則を尊重している企業に加点するというアイデアが示されており、このような場面でガイドラインを活用できるかもしれない。

(2) 政府の AI 利活用に対するガイダンスの導入

国内外の動向で見たように、特に英国では、政府の AI 利活用に関するルール整備が進んでいる。我が国では、政府 CIO 補佐官らが、政府情報システムまたは政府が提供するサービス等で利用する AI システムにおけるデータ利用の特性と取扱い上の留意点を論じているが、この領域におけるルール整備が進んでいるとはいいがたい。政府のデジタル化促進に合わせ、政府内でも AI システムが積極的に導入されていく可能性がある。AI システムの最終利用者としての政府に対するガイダンスも必要になると考えられる。

(3) 他国のガバナンスとの調和

企業は AI システムの販路を世界中に求めることができることから、AI ガバナンスの構築においては国際的な調和が不可欠である。我が国は、GPAI、OECD、UNESCO、AI 標準の議論に参加しており、今後もこれらの議論をリードしていく必要がある。また、多国間だけではなく、効果的な場合には、日 EU の AI 共同委員会や CEN/CENELEC との議論など、二国間の協力も進めるべきである。これらの活動を通じて、我が国の AI ガバナンスが国際的な調和するものとなるように努めなければならない。

(4) 政策と標準の連携

AI ガバナンスの議論では、中間的なガイドラインや標準などの複数のレイヤーが相互に重層的に関連している。このような状況に対応するために、日 EU の AI 共同委員会では、政策サイドと標準サイドとの連携を探っている。EU の予算で ETSI が運営している INDICO プロジェクトは、政策サイドと標準サイドの連携の重要性を意識した活動である。今後も、効果的な領域では、複数レイヤーにわたって AI ガバナンスの一貫性を確保するために、政策サイドと標準サイドの連携を深めていくべきである。

(5) モニタリングとエンフォースメント

非拘束の中間的なガイドラインについては、その利用状況のモニタリングが課題となる。当該ガイドラインは法的に非拘束であるから、利用状況を緩やかに把握していくことが望ましい。当面は、たとえば、利用状況に関するアンケート調査を行い、利用が進まない場合にはその理由を特定することなどが考えられる。

また、エンフォースメントについても検討していく必要がある。たとえば企業の民事責任について、ガバナンスモデル検討会は、「AI システム等が生じさせた損害に関する過失の所在が明らかでない場合、過失責任を原則とする民法の不法行為責任のもとでは、被害者の救済を確保することができない。被害者の立証責任の負担を軽減した製造物責任法が適用される範囲にも限界があり、また、AI システム等の提供者に資力がなければいずれにせよ被害者は救済されない。」という一般的なガイダンスを示している。本報告書の AI ガバナンス体系に照らしながら、ガバナンスモデル検討会のガイダンスを活用し、横断的な対応が必要なのか、個別分野別あるいは使用態様別の対応が適切なのか等について、国際的な動向や議論も見ながら、より精緻な検討を行うことが求められている。

4. おわりに

本報告書では、AI ガバナンスに関する国内外の動向を整理するとともに、我が国の AI ガバナンスについて、現時点で望ましいと考えられる姿を議論した。現時点では、特定の分野を除き、AI 原則の尊重とイノベーション促進の両立の観点から、AI 原則を尊重しようとする企業を支援するソフトローを中心としたガバナンスが望ましいと考えられる。しかし、AI ガバナンスの具体的な議論は、国際的に見ても始まったばかりであるとともに、今後さらに議論が活発化すると考えられるため、引き続き国内での議論を継続していく必要がある。

5. 有識者会議名簿（AI 原則の実装の在り方に関する検討会）

渡部 俊也（座長）	東京大学未来ビジョン研究センター　教授
雨宮 俊一	株式会社 NTT データ　技術開発本部　本部長
生貝 直人	一橋大学大学院法学研究科 准教授
上野山 勝也	株式会社 PKSHA Technology　代表取締役
川上 登福	株式会社経営共創基盤　共同経営者 マネージングディレクター 一般社団法人日本ディープラーニング協会 理事
齊藤 友紀	法律事務所 LAB-01 弁護士
杉村 領一	国立研究開発法人産業技術総合研究所　情報・人間工学領域 人工知能研究戦略部　上席イノベーションコーディネータ
角田 美穂子	一橋大学大学院法学研究科 教授
妹尾 義樹	国立研究開発法人産業技術総合研究所　イノベーション推進本部 標準化推進センター　審議役
田丸 健三郎	日本マイクロソフト株式会社　業務執行役員 ナショナル テクノロジー オフィサー
土屋 嘉寛	東京海上日動火災保険株式会社　企業商品業務部　部長
中条 薫	株式会社 SoW Insight　代表取締役社長
原 聡	大阪大学 産業科学研究所 准教授
福岡 真之介	西村あさひ法律事務所　弁護士
古谷 由紀子	サステイナビリティ消費者会議　代表
増田 悦子	公益社団法人全国消費生活相談員協会　理事長
丸山 友朗	パナソニック株式会社 イノベーション推進部門 テクノロジー本部 デジタル・AI 技術センター　AI ソリューション部 主任技師
宮村 和谷	PwC あらた有限責任監査法人　パートナー
山本 龍彦	慶應義塾大学大学院法務研究科　教授

以下、昨年度の AI 社会実装アーキテクチャー検討会のみ参加

青島 武伸	パナソニック株式会社 イノベーション推進部門 テクノロジー本部 デジタル・AI 技術センター　データアナリシス部 部長（当時）