AI ガバナンスの枠組みの構築に向けて(ver2.0)

2024 年 12 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(資料1)[1]、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（資料2）。

こうした中、AI を制御するためのルールづくりは、理念的な議論から具体的な議論へと急速に舵を切っている。欧州「AI 法」[2]や中国「生成 AI サービス管理暫定弁法」[3]のように AI 関連の法制度が成立した国・地域が存在し、また米国でも様々なレベルで AI を巡る具体的なガバナンスルールの議論[4]が行われている他、日本においても法制度の導入を含む議論[5]が本格的に始まろうとしている。

本文書では、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、

- AI のリスクを最小化するとともに、
- AI の利便性を最大限享受できる環境を整備し、
- こうした環境を可能な限り自律的に実現する生成 AI 市場を創出する

という3つの目的を均衡ある形で実現するための「AI ガバナンスの枠組み」、換言す

---

[1] 資料番号は付属資料「AI を巡る動向」の資料番号（以下同じ）。

[2] 2023 年 5 月、欧州理事会は「AI 法」を採択。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。
https://artificialintelligenceact.eu/

[3] 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

[4] 2023 年 10 月、米国政府は AI ガバナンスに関する大統領令を公表。政府機関において取り組むべき施策として、脆弱性調査（Red Teaming）の基準策定、アルゴリズムによる差別禁止のための明確なガイダンスの策定、ヘルスケア、教育等の分野における AI 適正利用の支援などの内容を盛り込んでいる。ただし、トランプ次期大統領は本大統領令を廃止する意向を既に表明している。なお、大統領令に先立つ官民連携の取り組みについては脚注 10 を参照。
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/

[5] 例えば AI 戦略会議「「AI 制度に関する考え方」について」（2024 年 5 月）を参照。
https://www8.cao.go.jp/cstp/ai/ai_senryaku/9kai/shiryo2-1.pdf

れば「AI 技術の制御可能性」を継続的に維持するための仕組みの構築に向けた検討の方向性を示す。

AI ガバナンスについては、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は適当でないと考えられる。

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。

デジタル政策フォーラム（DPFJ）は、2024 年 7 月、本文書について AI ガバナンスを巡る主要論点を整理した ver1.0[6]として公表した後、有識者へのインタビューや企業関係者からのヒアリングを行うとともに追加論点の整理等を行なってきた。今般の ver2.0 においては主要論点に関する基本的方向性を可能な限り明らかにしている。しかし、この基本的方向性はあくまで試案であり、今後の AI ガバナンスを巡る議論の進展に応じ、また AI 技術の急速な進化も踏まえつつ、必要な修正を加えていくこととする。

I 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（資料3）ではリスクを 4 段階に分類[7]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

---

[6] https://www.digitalpolicyforum.jp/wp-content/uploads/2024/06/240701_AI01.pdf

[7] 欧州 AI 法は、AI のリスクを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)、ハイリスク（事前の適合性評価、データベースへの登録等の義務）、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）、最小リスク（規制なし）の 4 種類に分類している。

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[8]。

なお、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という事実も存在する。例えば MIT 調査（資料4）[9]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・質的に変化する点も踏まえる必要がある。

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[10]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。

このうち、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）（資料5）[11]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

---

[8] 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も必要である。

[9] P. Slattery et al. “Global AI adoption is outpacing risk understanding, warns MIT CSAIL” (MIT CSAIL News, August 14, 2024)

[10] 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。）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。

[11] 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 10 か国・地域が署名（日本は未署名）。
https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

整合性を確保すること」などを原則とすることが考えられる。

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[12]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考えられる。

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、個人の特性に応じた個別化(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（具体策については項目(5)を参照）。

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[13]（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。

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。

(2) 規制のあり方と実効性の確保

---

[12] 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。

[13] 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」の改正案が成立している。また、米国の州レベルでは全米 19 州において選挙の文脈で AI 生成コンテンツについてはその旨を明示するラベリング規制が導入されている（2024 年 7 月時点※）など、利用者（有権者）の目線での法規制についても幾つかの国で動きが見られる。
※（出典）Funk, Vesteinsson, Baker, Brody, Grothe, Agarwal, Barak, Loldj, Masinsin, Sutterlin eds. Freedom on the Net 2024, Freedom House(October 2024)
https://freedomhouse.org/report/freedom-net/2024/struggle-trust-online

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）[14]などの手法がある(資料6)。例えば、中国（資料7）やEUにおいてはハードローを基本[15]とし、米国（連邦政府）は民間部門の自主規制を軸にした政策展開が行われてきている（資料8～10）。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、具体の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。

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）・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。

（AI基本法の制定）

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が適当である。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[16]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が考えられる。

なお、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

---

[14] 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。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。
なお、AI分野における共同規制の例には該当しないが官民連携の例として、大統領令（脚注3参照）の公表に先立つ2023年7月、大統領府とAI関連7社（Amazon, Anthropic, Google, Inflection, Meta, Microsoft & OpenAI）との間で非拘束の合意（AI開発における安全性、セキュリティ、信頼の確保という3つの項目から各社が取り組む内容を具体化）がなされた事例がある。なお、同年9月、上記7社に加えて8社(Adobe, Cohere, IBM, Nvidia, Palantir, Salesforce, Scale AI, Stability)が本合意に加わった。

[15] EUにおいては「許容できないリスク」について法規制で禁止するハードローによる直接規制であるが、「ハイリスク」なAI等については行動規範等を通じた共同規制を念頭に全体の制度設計が行われている点は留意が必要である。

[16] 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

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。

（AI 基本法と国の役割）

なお、AI 開発者について広く法規制の対象（登録もしくは届出制）とすることを制度枠組みの前提とすることは必ずしも適当ではない。なぜなら、誰もが参画可能な科学技術の開発行為の中で AI 開発という行為に限って法規制の対象とすることについて、十分な合理的な根拠と国民的なコンセンサスが求められるからである。

こうした観点からは、例えば、政府において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主的措置とし、必要に応じて政府（AI 戦略本部）と連携する（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）仕組み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）とは明確に区別することが考えられる。

(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）[17]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ティーミング手法ガイド(第 1.00 版)」[18]を公表する等の動きがみられる。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・明確にすることが実効性を担保する観点か

---

17 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：米国 Department of Defense “ Mission Assurance Strategy ” (April 2012)）

18 https://aisi.go.jp/assets/pdf/ai_safety_RT_v1.00_ja.pdf

ら極めて重要である。

また、AI の学習プロセスにおいてデータ汚染攻撃[19]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(資料 11)。

上記の検討に際しては、AI の透明性確保のための学習データや AI システムをオープン化することで、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を回避する観点から、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(資料 12)[20]。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 は学習データを取り込みモデルを形成し、これを活用して生成物たるデータを出力するものである。そこで、データの完全性(integrity)の確保という観点から見れば、

---

[19] 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。

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

上記(3)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も関連して議論を深める必要がある。

II 利便性の向上

(5) AI の積極的活用

（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[21]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。

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、留意する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

[21] 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。

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（資料 13～14）。

　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[22]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)の整備とその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を本来の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。

III 健全な市場の育成

(6) 健全なエコシステムの構築

---

[22] 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

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[23]。

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[24]。

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、AI のネットワーク化、言語の壁を越えた AI 連携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可能性についても検討が求められる。

(7) 産業振興とグローバル連携

現在のコンピュータ利用は集中型のクラウドと分散型のエッジの組み合わせにより多様なニーズに応じた資源配分の最適化が進んでいる。同様に AI についても集中型と分散型でコンピューティング資源を組み合わせたり、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）

---

[23] 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

[24] 公正取引委員会「生成 AI を巡る競争（ディスカッションペーパー）」(2024 年 10 月)
https://www.jftc.go.jp/houdou/pressrelease/2024/oct/241002_generativeai_02.pdf

インターネットが爆発的に普及した主因の一つはそのオープン性にある。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[25]。

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、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。

（産業としてのAI総合戦略の推進）

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を構築していくことにつながる。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[26]、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総合戦略を策定・推進していくことが求められる。

---

[25] 欧州では、2024年1月に公表された招請文書”Competition in Virtual Worlds and Generative AI: Calls for Contribution”において、生成AIと競争政策に関する論点リストが提示されている。
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_85
また、米国では大統領令（脚注3参照）の中でイノベーションや競争を促す観点から、「公正・オープン・競争的なエコシステムの促進」を主要推進項目の一つに挙げている。

[26] 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

(8) 国際的コンセンサスの醸成

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。

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」(資料 15)[27]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(資料 16)[28]していることを踏まえて議論を急ぐ必要がある。

(9)倫理的問題への対処

---

[27] 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、現在、日本を含む 51 か国が賛同している。
https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/

[28] 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

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。

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」であり、換言すれば「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」を目指すことにある。

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(2025 年夏を目処に ver3.0 への更新を実施予定)。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 ガバナンスの枠組み構築に向けた議論を深めていくこととしている。その際、同様の議論を進めている他のフォーラム等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。

以　上

AI ガバナンスの枠組みの構築に向けて(ver2.0)
【本編付属資料～有識者インタビュー～】

有識者インタビュー

デジタル政策フォーラムでは、2024 年 7 月、AI ガバナンスを巡る主要論点を整理した「AI ガバナンスの枠組みの構築に向けて」ver1.0 を公表した後、各領域の有識者へのインタビューを行っており、そのインタビューの内容について、「本編付属資料」として以下に掲載する。なお、インタビューは 2024 年 9 月から 11 月に実施した。

- 有識者(敬称略･五十音順)
  - ･ 國領 二郎　　慶應義塾大学総合政策学部教授
  - ･ 五神 真　　理化学研究所 理事長
  - ･ 坂村 健　　東京大学名誉教授
東洋大学 INIAD cHUB(情報連携学学術実業連携機構)機構長
YRP UNL 研究所長
  - ･ 橋田 浩一　　理化学研究所革新知能統合研究センター
社会における人工知能研究グループグループディレクター
分散型ビッグデータチームチームリーダー
  - ･ 村井 純　　慶應義塾大学教授
  - ･ 柳川 範之　　東京大学大学院経済学研究科教授

- **國領 二郎**
  慶應義塾大学総合政策学部教授

(AI と責任)

- **「責任」を主たる概念とすれば、AI 規制はシンプルに整理できる。「AI に正しいデータをインプットし、学習させたか」、「AI のモデル/アルゴリズムとして適切な設定や選択をしたか」など、プロセスを「責任」で整理すれば、規律は容易になる。自動運転、医療判断・行為、著作権処理における AI 利活用を想起してみれば、「責任」で十分に規律可能なことが理解できよう。**
- **善管注意義務や過失、重過失という従来の法律概念とも「責任」は密接であり、不可分である。リスク対応としては保険も一つの選択肢であり、民事であれば保険である程度はカバーできるだろうが、刑事は保険ではカバーが難しい。軍事や安全保障の領域ではそもそも保険が成立しない。**
- **ところで、AI をどのように罰すれば人々は事件の解決として納得できるのだろうか。AI を「死刑」にすることは可能だろうか。仮に AI を死刑にできたとしても、誰もその解決に納得しないのではないだろうか。規制について法技術的な検討を行うだけでなく、哲学や倫理から AI と社会を議論せねばならない。**

(データと利他)

- **データ駆動社会である現在において、また AI 社会となる将来において、最も重要なのはデータである。したがって、個人がデータを提供して、それによって提供した個人がひいては社会全体が利便を享受するアーキテクチャーを考えなければならない。**
- **データの特徴は外部性が強く働くことであり、連携・統合・集積が高いデータであるほど高い価値を有する。これは、マイケル・ポーターが提唱した「共通価値の創造(Creating Shared Value)」に通じる。個人が同意した価値あるデータの提供が、外部性を高めて、データベースや社会全体のデータの有用性を高める。**
- **データの提供は「利他」という概念で整理するのが合理的だ。共通価値の最大化と再分配の仕組みを実現することにより、経済学でいう個人の効用最大化も「利他」によって実現可能になる。**
- **モノに関する概念である所有と交換でなく、データに対するアクセス権を付与する・連携するとの考え方でデータのエコシステムを検討するべきだ。データとその集合体であるデータベースの品質を高める、維持するためには、基盤としての「トラスト」が必要となる。**

(社会構造の変化と時代のエートス)

- アメリカ大統領選をはじめ、各国の選挙を概観すると、「デジタル化した・デジタルを積極的に受け入れるコミュニティ」と「デジタルから距離を置いている・距離があるコミュニティ」の差は明らかだ。
- デジタルのほかでは若年世代、高等教育修了、自由主義的思想という切り口でも整理は可能であり、これらの切り口でもコミュニティの差異は顕在化する。なお、若年層は時間選好における割引率が高いため、将来価値よりも現在価値に重きを置く傾向がある。このため、手取りや可処分所得を変化させる政策に共感しやすい。中長期よりも短期の社会選択が拡大する可能性が高まっており、世界で長期的・俯瞰的な政策の選択が困難になりつつある。
- ところで、GM(ジェネラルモーターズ)と Google を比較すると、雇用者数は同程度であるものの、時価総額に大きな差がある。この傾向は今後ますます広がる。GM のような製造業は、資本家と労働者の歴史的な対立を超えて、分厚い中間層を先進各国で生み出してきた。他方、Google のような IT 企業は少人数でより高い付加価値を生み出すため、雇用を創出する力は製造業に比べて著しく低い。雇用創出力が低いことが中間層のフラグメンテーションに結びつき、所得格差を拡大した。そして、「分断」の素地を生み出した。
- 経済的な格差や政治的な分断が世界で加速することが懸念される。この社会構造の変化や時代のエートスを認識した上で、AI やデータの議論を進める必要がある。

（今後の議論の方向性）

- AI に関する規制は必要最小限度で制度化するべきだが、同時に理念法＋ソフトローの選択で思考を停止してはいけない。
- 欧米は法制度の整備と議論が日本よりも進んでいる。この状況で、どのような制度を導入するのが国家として戦略的であるのか、国際環境を睨みながら「したたか」に振舞わねばならない。AI による歴史的な変化に直面している今日、国際的に有利なポジションを日本がいかに占めることができるかを考えねばならないし、有利なポジションを取ろうとする気概を持たねばならない。
- 日本は小国になりつつある（もうなっている）ことを前提として、したたかに生き抜く道を模索する必要がある。AI 及び関連規制については、技術と法制度の二つの観点から議論されることが多いが、それのみに止まらずに文明的かつ歴史的な洞察をもって、グランドデザインを描く必要がある。

- **五神 真**
  理化学研究所 理事長

(デジタル革新と AI によるイノベーションは歴史的な社会変革)

- **半導体、通信、量子などデジタル技術の革新が新たなフェーズに入る中での生成 AI に代表される AI の革新技術が到来したリアルタイムデータと高度な予測計算の組み合わせは、社会全体のスマート化を加速し、そこから大きなサービスすなわち経済価値が生み出されるようになる。数年以内にその姿が明らかになる。**
- **日本は有線の高速大容量ネットワークが全国規模で既に整備されており、高度化する無線技術とリンクする中で、他国に先駆けたセキュアで低環境負荷の高度社会インフラの構築が可能であり、それが大きな優位性である。この強みを認識して、日本への投資の呼び込みと経済成長に最大限に活用するべきである。**
- **特に、生成 AI、先端半導体、量子コンピューティングが、現在進展しているイノベーションの核であり、これらと無線有線の高度通信技術、スーパーコンピュータが密接に絡み合って、経済社会を劇的に変化させる。**

(AI for Science、フィジカルインテリジェンス)

- **今やワンチップに搭載できるトランジスタ数は一兆個を超え、昔の計算センター並みの計算能力を装着できるようになった。生成 AI が急拡大する中で、近未来の主戦場は推論計算の高度化と低消費電力化で、高度な推論計算用チップがスマートフォンに搭載されるようになる。大規模言語モデルに対する推論計算に必要な大規模行列計算を処理する専用チップが誕生する。このチップにより、様々なエッジデバイスの知能化が一気に進むことで、自動運転車はもとより、AI、IoT とロボットなどが融合する高度技術が大きく発展する。その結果 AI が物理的な動作と結びつき、AI の利活用が人間の活動空間であるリアル世界に浸透する、フィジカルインテリジェンスが出現する。**
- **専用チップによる推論計算の高度化だけでなく、スパコンのさらなる高度化や量子古典ハイブリッド計算など人間が行える計算の領域が急速に拡大し、環境・エネルギー、健康・医療などのあらゆる領域で解くべき問題が解けるようになる。この計算可能領域の拡張には、ハード・アルゴリズム・ソフトのコ・デザイン(codesign)が鍵となる。**
- **地球環境問題などの人類課題の解決は急務であるが、手元の知恵では不足しており、研究開発の一層の加速が急務だ。生成 AI を含む急速に進展する AI 技術を活用して、研究開発を一気に加速させる AI for Science により、基礎科学を起点とした科学がインパクトの大きな社会課題解決を先導し、成長の機会を創出しなければならない。**

- AI for Science は、先述したエッジの知能化であるフィジカルインテリジェンスにシームレスに移行する。フィジカルインテリジェンスでは生成 AI と通信技術により、複数のロボットが互いに協調しながら自律的に動作できるようになる。
- フィジカルインテリジェンスは広義のロボティクスである。これによって、人間の身の回りの知能化が進むことと捉えるべきである。広義のロボティクスが実装され、普及するには、先端半導体・AI・通信・量子の一体的連携が不可欠となる。

（AI とこれからの社会）

- これからの社会において、AI を使わないという選択はありえない。そのためにも AI がどのように使われて、AI の利活用にどのようなリスクが存在するかを学問的に考えなければならない。理系だけでなく社会科学と人文知も動員して、ハード・ソフト・社会システムの総体から検討する必要がある。将来のある時点の社会像を描き、そこで起こり得る事象や状況を踏まえながら、課題の抽出とそこに至るパスウェーを学際的に研究し、議論するべきである。それは、経済的な成長につなげる戦略とも重なるはずだ。
- 今後はレイヤーが異なる学問領域を AI でつなぐことも起きるだろう。また、研究に時間を要するライフサイエンスなどでは、AI と高度化する計算科学を活用することにより、より大規模シミュレーションを行い、トータルシステムの予測に基づいて、研究を加速させることもできるようになるだろう。
- AI は実社会の営みと密接にリンクしている技術であるので、社会から切り出して、単体で検討して、議論することはできないし、全く意味がない。非連続的な進歩を AI が遂げる中、法制度を改正するスピードは技術のスピードに追い付かないことを前提として、いかに法制度を設計するかが重要になる。
- 一方で、AI の基礎学理は未だ未解明である。数理的な性質を解明することについて手を緩めるべきではない。クリーンな科学的なデータに基づく大規模モデルの開発と研究は AI そのものの数理的な性質を解明していく上で大変重要である。すなわち Science for AI も AI for Science と共に重要だ。
- 先端半導体・AI・通信・量子の一体的連携によりフィジカルとリアルが一体化することを通じて、人々や企業の行動変容を促し、地球という「コモンズ」を守る必要がある。AI はそれを支える重要な要素である。そのためにも、不確実性が増す世界において、共感や信頼を科学によって高めることが重要となる。

- **坂村 健**
  東京大学名誉教授
  東洋大学 INIAD cHUB(情報連携学学術実業連携機構)機構長
  YRP UNL 研究所長

(AI と国家の存立)

- **日本の全産業が「AI データセンター」から生まれる「知力」に立脚する時代がまもなく到来する。したがって、海外の「AI データセンター」に依存する構造は経済安全保障上、極めて危険な状況であり、日本の経済、社会の存立に関わる問題である。**
- **AI に関するリスクは AI が利活用される際のリスクを念頭に議論することがほとんどだが、AI を利活用できないリスクも検討する必要である。AI に関するリスクを最小化するためには、海外の AI サービスが遮断される可能性を検討するべきである。**
- **遮断がなかったとしても、すでにクラウドへの依存度は極めて高く、多額の利用料が海外に流出している状況だが、AI ではさらにこの傾向が加速する。日本企業が AI サービスを提供しても、その基盤となる「知力」を米国から購入する構造では、外貨の流出が継続的に発生する。これは、国内に発電所を持たず全ての電力を海外から購入するような状況に等しく、国富の維持の観点から極めて深刻な問題である。**
- **オープンソースの AI モデル、ローカル LLM、AI コンフェデレーションなど、国内の AI も多様な進化を遂げるだろうが、それらの精度をあげるチューニングを行うためにも、常に最先端、最上位の AI モデルにアクセスできる環境を整備することが絶対に必要である。最先端、最上位の AI モデルにアクセスして、研究開発を進めることが、経済のみならず安全保障など全ての領域で国力に直結する。**
- **基礎研究の進展は国力に直結する重要なリソースであり、アメリカは基礎研究における AI 利用に関して、フルスペックの機能を国外にはオープンにしていないのではないか。ビジネスに直結する AI 利用は民間に任せることが可能だが、日本でも基礎研究については「国立研究 AI」として国が主導して推進しなければならない。**
- **日本で最先端、最上位の AI モデルを持つということは、決して国際的な連携を否定するものではなく、技術的主権の観点から、最低限の技術的自立性がなければ、貢献無く消費するだけになってしまい、真の意味での国際協力は成り立たない。日米安保の枠組みがあっても自衛隊が必要なように、AI においても独自の基盤技術が必要である。**

（規制強化への反対）

- 哲学的かつ倫理的な内容にとどまる AI 基本法は必要となるが、AI に関する規制強化には強く反対する。起こり得る問題を事前に想定して、列挙するのでなく、問題が生じた事象に関してネガティブリストを作成すれば、規制すべき対象は足りる。経済社会構造が変化するのであるから、現在の法制度に照らしてブラックでもホワイトでもない、中間のグレーな行為は規制せずに、自由に研究し、利用できる枠組が必要である。
- 海外で開発されて、国内で提供されている AI サービスに関しても、厳しい規制をかけるならば、日本から撤退されるリスクを考慮しなければならない。国内の AI サービスですぐには代替できない水準のサービスが撤退されるならば、日本の産業全体の競争力が低下することに繋がりかねない。
- EU のように規制ありきの法制度には反対する。リスクベースのアプローチにも疑問を禁じ得ない。「利用されるリスク環境や取り扱うデータの気密性等で規律は異なる」との見解を表明して、AI 州法に拒否権を発動したカリフォルニア州知事の見解に賛同する。
- AI 開発拠点の立地環境は規制水準に影響を受ける。緩やかな AI 規制は AI 研究開発拠点の立地にプラスとなる。AI に関するマネーフローも規制水準の高いところから低いところに向かう。

（クリエイティビティ、AI ネイティブ）

- AI は「仕事を奪う」が、AI が代替できない仕事は当然ながら残る。人間同士が切磋琢磨するところに感動は生まれて、エンターテインメントに昇華する。AI により、労働環境や社会構造が劇的に変化するのだから、ベーシックインカム導入など、あらたな社会政策を研究しなければならない。
- インターネットは当初、ブロックチェーンのような機能を実装しようとしていた。AI による生成物にはブロックチェーンなどを用いて、トレーサビリティを確保することが必要となろう。このため、これからのインターネットは時間をかけて、Web3.0 に向かう。
- AI にはクリエイティビティしかない。人間は考えるために AI を利活用することになる。ただし、AI には欲求がないから、「ストーリー」は作れない。AI にはなく、人間にあるのが「欲求」であり、これこそが人間と AI との決定的な違いだ。
- デジタルネイティブの次は AI ネイティブが現れる。AI バディとともに人が成長する時代が到来する。このバディ AI を前提とする教育指導要領を研究するべきであり、国のコミットメントが必要となる。
- AI とのバディで人生を考える新世代となれば、我々とは異なる「善き社会」の在り方を模索するだろう。その新しい社会へのフリーハンドを確保するためにも、実験

的な取り組みを許容する環境整備が重要である。

- **橋田 浩一**
  理化学研究所革新知能統合研究センター
  社会における人工知能研究グループグループディレクター
  分散型ビッグデータチームチームリーダー

（EU 法と整合標準）

- **EU の AI 法では第 6 条が規定する高リスク AI に関して、質とリスクの管理が義務付けられるが、その詳細は欧州標準で定められる。この整合標準は法律とほぼ同じ拘束力をもつ。**
- **高リスク AI の管理方法が ISO/IEC の国際標準になり欧州標準として採用され整合標準になる予定である。GDPR と同様に EU の AI 法が世界中でデファクトスタンダードになれば、高リスク AI の適正管理の具体的な方法が世界中で義務付けられることになる。**

（AI システムの運用と管理）

- **AI システムのロギングとは、AI システムの目的やリスクに関係の深い事象の記録であるが、高度な AI システムは想定外の事態に対処できるので、AI システムにおけるロギングの範囲を前以て確定することは一般にはできない。AI システムの目的やリスクとさまざまな事象との関連性は文脈に依存するので、ロギングの範囲は予測不能な文脈に応じて変わる。また、ログデータはシステムの管理に用いるものだが、AI システムの場合は通常の運用と管理が「不確実性の下で目的を達する」という目標を共有しているので、運用と管理の区別はできない。不確実性をあまり前提しない非 AI システムの運用と異なり、AIシステムの運用は管理を含み、ログは運用の管理の両方に用いられる。さらに、管理は運用の方法の改善を含むので、管理そのものの方法の改善を含むことになる。**
- **AI には人間が関与する AI と人間が関与しない完全自動 AI とがあるが、人間が関与する、すなわちユーザーが存在する AI システムは、個人データや法人データに関する権利を守りつつシステムの管理運用にログデータを最大限に活用するため、ロギングシステムが必要になろう。**
- **ユーザーの非公開データが AI システムで取り扱われる場合にのみロギングシステムを設けてログデータを AI システムの運用管理にフル活用することにすれば、コストをかけすぎずに、データに関する権利を守りつつあらゆるログデータをフル活用できる。**

（パーソナルAIと分散管理）

- **個人ユーザーの非公開データを扱う AI システムにロギングシステムを設けてデ**

ータに関する権利を守りながらログデータを AI システムの運用管理にフル活用するということは、ロギングシステムは分散管理型の(利用者本人だけが全データにアクセスできる) PDS (パーソナルデータストア)ということである。そのような AI システムを「パーソナル AI」と呼ぼう。

- データの分散管理とは一人の管理者が一人のデータ(一つの組織のデータ)を管理することである。ほとんどのサービスは各ユーザーの最適化であるため分散管理で十分であり、その方が集中管理(一人の管理者が多くのユーザーのデータを管理すること)よりもリスクとコストが低い。
- パーソナル AI は各利用者のデータの価値を最大化する。パーソナルAIとは特定個人(特定組織)に専属するAIであって、利用者と対話して利用者のニーズを満たすサービスを仲介(選定・実行)する。これによって利用者は、各サービスを利用するためにアプリや Web サイトを操作する必要がなくなり、IT リテラシーデバイドが解消する。また、サービス提供事業者がこの仕組みにつながることによってサービスの利用が激増する。また、サービスの仲介は生成 AI 等によって技術的にはすでに可能になっている。したがって、パーソナルAIはこれから十年で世界中に普及するだろう。
- パーソナルAIにつながった多様なサービスのログデータは利用者の PDS に集約・蓄積され、利用者が自由に活用できる。こうして、パーソナルAIがデータの分散管理をもたらし、分散管理がパーソナル AI の普及を促進する。
- たいていの国でＢ２Ｃサービスは ＧＤＰの 70%程度であり、B2B サービスは 200%を越える。それらの仲介手数料がサービスの価格の 10%以上とすれば、パーソナルAIの市場規模はグローバルＧＤＰの 30%を超えるだろう。
- 適正に管理されたパーソナル AI があらゆるサービスのゲートキーパーになることで、オンラインの不正な行動操作がなくなる。第一に、パーソナル AI がユーザーのニーズに合うサービスを選定することにより、オンラインの広告がなくなり、したがって広告に注意を引くためのフェイクやエコーチェンバーが商業的価値を失う。これによって、グローバルＧＤＰの 1%に満たない広告事業はＧＤＰの 30%を超えるパーソナルAI事業に吸収される。第二に、パーソナル AI がサービス事業者に代わってサービスのユーザインタフェースを提供することにより、いわゆるダークパターンもなくなる。以上によって注意経済と監視資本主義が終焉する。
- AI システムがデータをフル活用して提供価値を最大化するには、ユーザーのトラストが必須である。そのトラストには、分散管理によって AI システムの管理にログデータをフル活用することが必要だろう。そのために規格と規制を組み合わせた規律を設計することが政府の役割である。

- **村井 純**
  慶應義塾大学教授

(ネットワーク化)

- **近年のAIの発展においては、大量データの取得とその学習の部分に関して独占や集中が起きているが、リアルタイム性のある要求を処理するには分散が合理的である**
- **したがって、これまでのような集中処理ではなく、これからは分散したAIのネットワーク化が急速に進展する。**
- **コンピューティング・アーキテクチャーとは、データとその処理を通じた結論を誰が享受するのかという構造であり、AIの進歩と普及においても、全体のモデルをどのように設計し、構築するかが世界的な課題になる。**

(エッジ化)

- **近年のAI開発はメガプラットフォーマーが中心だったが、全体のモデルが洗練されてくると、エッジ化が進展する。スマホのチップに推論のアルゴリズムが組み込まれるようになり、AIエッジコンピューティングなどが拡大して、分散に向かう。**
- **新しいチップで分散が進み、エッジ化すると、価格は劇的に下落する。**

(規制と標準)

- **インターオペラビリティ(相互運用性)が進むとAIがお互いに学びあうようになる。技術的趨勢を踏まえれば、国境で閉じる国内法の議論よりも、国際基準が重要になる。**
- **規制は技術進歩を止める側面があることを認識しなければならない。分散とインターオペラビリティを進めて、アーキテクチャ全体として安全な発展を確保することが肝要。**
- **AIでは一国主義はどんな大国でも困難である。分散化とネットワーク化により、アクセスポイントが増えることとなり、全体最適が進むことにつながる。キーとなるのは国際的な標準化であり、これは経済安全保障にも知的財産保護にも密接に関与する。**
- **チップ開発などにおける独占、寡占へのカウンターベイリングパワー(拮抗力)が現れるかどうは、オープンディベロップメントによって対抗勢力を形成できるかどうかにかかっている。EUモデルは標準化によってアメリカ型のドミナントに対抗しようとするモデルとみなせる。**

(AIとデータ)

- AI によるデータ使用を、プロトコルを通じてコントロールする手法が考えられる。AI で使用するデータは個人に帰属しないものもあり、データ管理者をどのように規律するか、データ管理者がどのような責任を有するかなどを検討した上で、制度を設計するべき。
- AI のデータサイエンスから考えれば、データを塊として認識して、対象とするべきであり、それらをダイナミックかつオートマティックに許諾して、即時に処理することが求められる。このプロセスを標準化することが必要である。
- 国内を対象とするデータ規制は限界があり、データ流通に関しては十分な国際的議論が必要となる。

（合意形成）

- 国連、OECD、G7、G20 などが合意形成の場として想定される。ポリシーメーカーとエンジニアが十分な議論をする必要がある。
- データを集めて、大規模に計算するプレーヤーが支配的であるのは近年の状況であり、それ以前の歴史的変遷を踏まえて、技術的なアーキテクチャを理解して、将来を見通した規制を設計する必要がある。
- 各国が別々に規制するようなら、AI の進歩や発展は阻害される可能性が高い。アシモフのロボット工学三原則[29]のような大原則の合意は意義がある。他方、悪用・濫用防止などの観点から規制を検討するのは、それがたとえリスクベースであったとしても困難ではないか。

（社会的な役割）

- AI は代替的でなく、人間の脳をサポートする補完的な存在になりえる。長期では AI は人間の仕事を奪うことにはならない。
- これからのインターネットは、広義としてはデジタル技術のインフラそのものであり、だからこそインフラとして果たさねばならない役割は大きい。インターネットの使命は、光の速度を前提とした地球上のコミュニケーションの基盤であることであり、データと計算を世界が安全に共有できるようにすることである。

---

[29] ロボット工学三原則

第1条　ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害を及ぼしてはならない。

第2条　ロボットは人間に与えられた命令に服従しなければならない。ただし、与えられた命令が第一条に反する場合は、この限りではない。

第3条　ロボットは前掲第一条および第二条に反するおそれのないかぎり、自己を守らなければならない。

- **政治的な分断を超えて、真にグローバルな空間をインターネットは維持できている。致命的なフラグメンテーション（分断）は今後も起きない。**

- **柳川 範之**
  東京大学大学院経済学研究科教授

（AI 法）

- **AI 法を制定するべきだと考えるが、AI は急速に進歩しており、立法によって制度が硬直化し、規制が変化の妨げになることは避けなければならない。法定するのは基本原則のみとして、実効性はソフトローであるガイドラインで担保するのが適当である。**
- **立法にあたっては、EU の AI 法などと規制内容を合わせることができるかを検討する必要がある。同レベルの規制を行うためにも、柔軟に改正が可能なソフトローとしてのガイドラインで詳細を規定することが現実的ではないだろうか。**
- **ただし、諸外国との規制のギャップが存在し、規制が緩やかである方が投資や立地の面からはプラスに作用することも考えられる。産業政策的観点からは、あえて国際的な潮流とは別に緩やかな規制を設ける戦略はありえよう。**
- **G7、G20 などを通じて、世界的に共通なルールを策定することは現実的な手法である。また、国連における合意形成は、加盟各国の制度化を促進する効果がある。国連が主導して世界的に定着した SDGsのように、AI に関する国際ルールを国連が主体的に形成することもありえよう。**

（経済への影響）

- **AI とインターネットでは国境を越える点が共通であり、国内規制に限界がある点も共通である。政治的な分断が、世界のインターネットを分断しつつあるように、AI においても二つの AI ネットワークが生み出されることがあるかもしれない。**
- **AI の普及によって、特定の市場における企業の参入・退出は活発化するであろうし、それを受けて産業構造も大きく変貌するだろう。AI による世界の産業構造の変化はインターネットによって起きた変化を凌駕するだろう。**
- **AI によって各人が生み出す付加価値は増大する。歴史的かつ構造的な変化によって、経済を測る物差しや指標を根本から見直すことも必要になるだろう。**
- **AI の普及によって、従来の仕事のやり方や組織の在り方も劇的に変化する。労働者が特定の企業の一員となる就労形態も変わる可能性があり、労働という概念自体が変化を迫られるだろう。**
- **AI が労働の代替でなく、補完になるように制度を設計することによって、AI の浸透による負の側面をできるだけ減らさねばならない。**

（データとガバナンス）

- **データは収集される規模によって価値そのものが変わるものであり、独占をいか**

に制約するかが課題となる。そもそも、知的財産制度は独占を特別に認めるものであり、競争政策とは二律背反の関係にある。

- 現状はプラットフォームを規制することによりデータ独占に対処しようとしているが、そのアプローチが現実的に有効であるかどうかは検討を要する。
- AI 規制を検討するにあたっては、市場の独占力を問題にするなら上流工程である開発を対象とするべきであろうし、利用者保護を重視するなら下流工程である AI の提供や利用にフォーカスするべきであろう。
- AI による生成物に関しては、生成にあたって学習したデータやレファレンスまで含めて情報を開示するべきである。これは、AI 及びその生成物に関するブラックボックス化への歯止めとなるものである。AI のブラックボックス化は、AI ガバナンスにおいて最も大きな懸念点である。

（AI と社会）

- AI の教育での利用は今後、拡大するだろうが、初等中等教育と高等教育とでは利用目的や態様は異なるだろう。
- AI 倫理に関しては簡単には結論が出ず、合意形成にも時間を要するだろうから、早めに社会的な議論を喚起することが重要である。
- AI はツールであり、有用・悪用いずれにもなり得るので、開発の自由度は確保して、利用段階から十分なモニタリングをして、社会的にチェックできるようにするのが現実的ではないか。

AI ガバナンスの枠組みの構築に向けて(ver2.0)

【本編付属資料～AI を巡る動向～】

AI を巡る動向

本附属資料は「AI を巡る動向」として、AI を取り巻く各国動向等について、公開情報を基にデジタル政策フォーラムにて取りまとめた資料である。なお、AI に関する技術的や法制度等の変化は流動的・非連続的であるため、資料内等に記載時点の情報である点をご理解いただき、活用いただきたい。

資料1

GPT(General Purpose Technology)

GPTとは、経済全体（通常は国家レベルまたは国際的なレベル）に影響を与える技術であり、既存の社会経済構造にインパクトをもたらすことで社会を劇的に変える力を有している。

No.	GPT	時期	分類
1	植物の栽培	紀元前9000～8000年	プロセス
2	動物の家畜化	紀元前8500－7500年	プロセス
3	鉱石の精錬	紀元前8000－7000年	プロセス
4	車輪	紀元前4000－3000年	プロダクト
5	筆記	紀元前3400－3200年	プロセス
6	青銅	紀元前2800年	プロダクト
7	鉄	紀元前1200年	プロダクト
8	水車	中世初期	プロダクト
9	3本マストの帆船	15世紀	プロダクト
10	印刷	16世紀	プロセス
11	蒸気機関	18世紀末19世紀初頭	プロダクト
12	工場	18世紀末19世紀初頭	組織

No.	GPT	時期	分類
13	鉄道	19世紀半ば	プロダクト
14	鋼製汽船	19世紀半ば	プロダクト
15	内燃機関	19世紀終わり	プロダクト
16	電気	19世紀末頃	プロダクト
17	自動車	20世紀	プロダクト
18	飛行機	20世紀	プロダクト
19	大量生産	20世紀	組織
20	コンピュータ	20世紀	プロダクト
21	リーン生産方式	20世紀	組織
22	インターネット	20世紀	プロダクト
23	バイオテクノロジ	20世紀	プロセス
24	ナノテクノロジー	21世紀	プロセス

（出典）総務省「平成30年版情報通信白書」

"Artificial Intelligence (AI) is a general-purpose technology that has the potential to : improve the welfare and well-being of people, contribute to positive sustainable global economic activity increase innovation and productivity, and help respond to key global challenges." OECD "Recommendation of the Council on Artificial Intelligence" August 2023

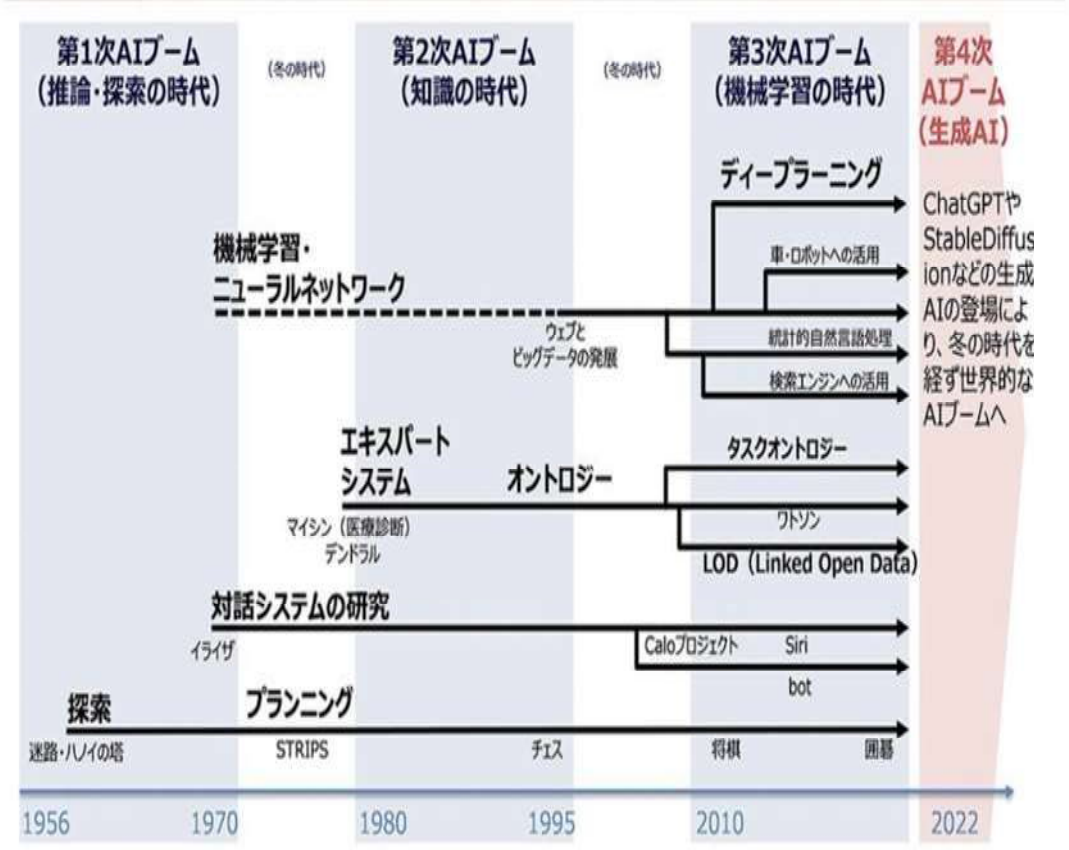
Sam Manning et al, "GPTs are GPTs : An Early Look at the Labor Market Impact Potential of Large Language Models," arXiv (August 2023)

→LLMがソフトウェアに実装されることで労働市場に大きな影響を与えるgeneral-purpose technologyであると主張。

資料2

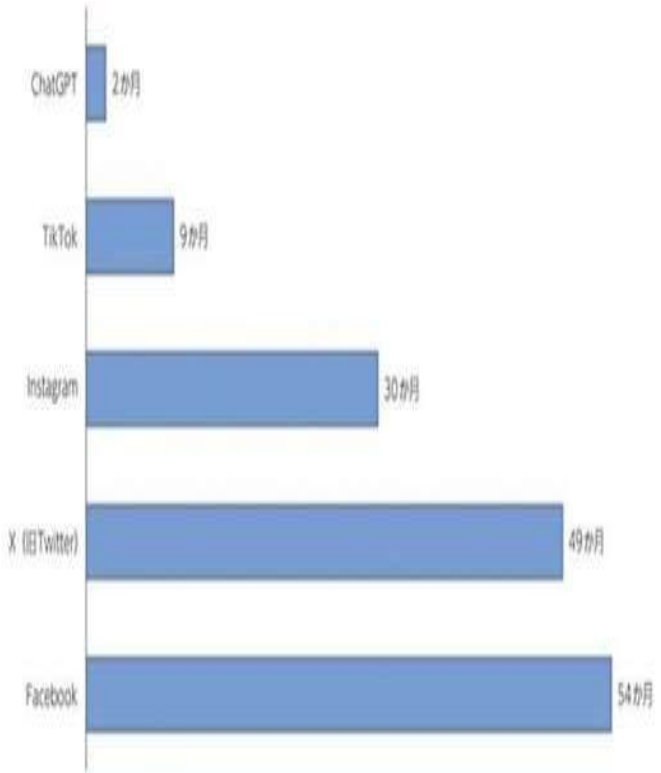
急速なAIの普及

AIブームの系譜



（出典）令和6年版情報通信白書

1億ユーザー達成までにかかった期間



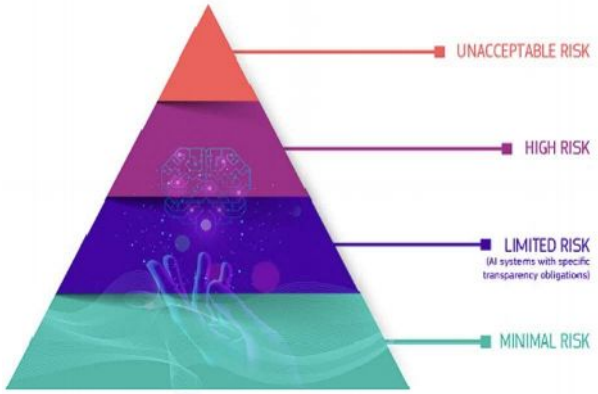
（出典）令和6年版情報通信白書

資料3

欧州：AI法の制定(2024年夏、部分施行)

※実際の適用は2段階的に実施。

リスクベースアプローチ



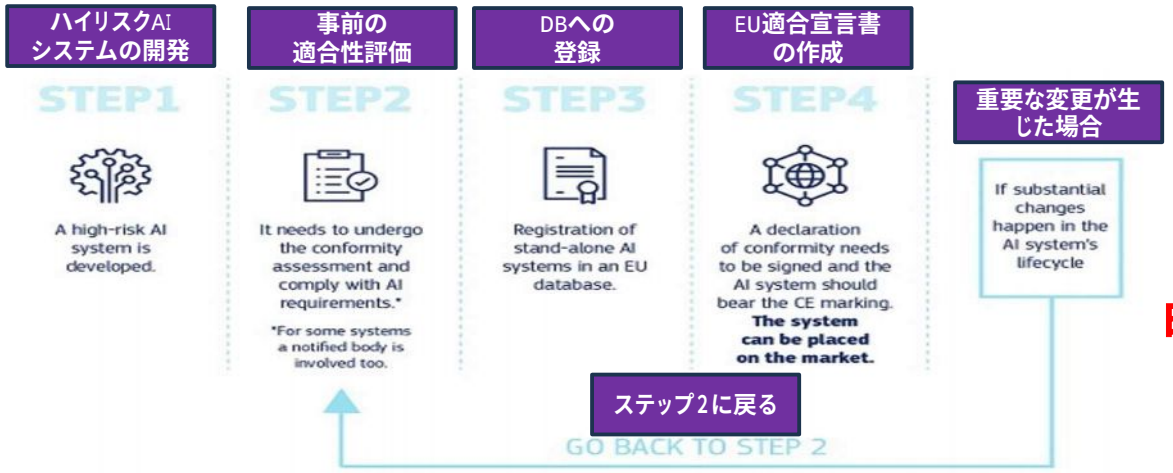
許容できないリスク 禁止
(人の生命や基本的人権に対する直接的な脅威)

1)**サブリミナル**な手法のAIシステム
2)年齢、身体的障害、精神的障害による**脆弱性**を利用するAIシステム
3)**ソーシャルスコア**を公的機関が用いるAIシステム
4)法執行を目的としたリアルタイムでの**遠隔生体識別**システム
(法第5条第1項)

ハイリスク 規制 →規制のプロセスは下図参照
(人の健康や安全、基本的人権、社会的・経済的利益に影響を与える可能性)

限定リスク 透明性の義務 →AIとのやりとりであることを利用者に知らせることが必要。
(深刻なリスクはないが、透明性確保のための要件を満たす必要)

最小リスク 規制なし →EUで利用されている大半のAIはこのカテゴリー
(リスクがごくわずか、またはリスクを伴わず)



※欧州 AI 法は、施行から 2 年後の2026年夏に全面適用。ただし、禁止される AI に係る規律は施行か ら 6 か月後、汎用 AI に係る規律は施行から 12 か月後、実践規範は施行から 9 か月後、高 リスクシステムに係る規律は施行から 36 か月後に適用が開始される。

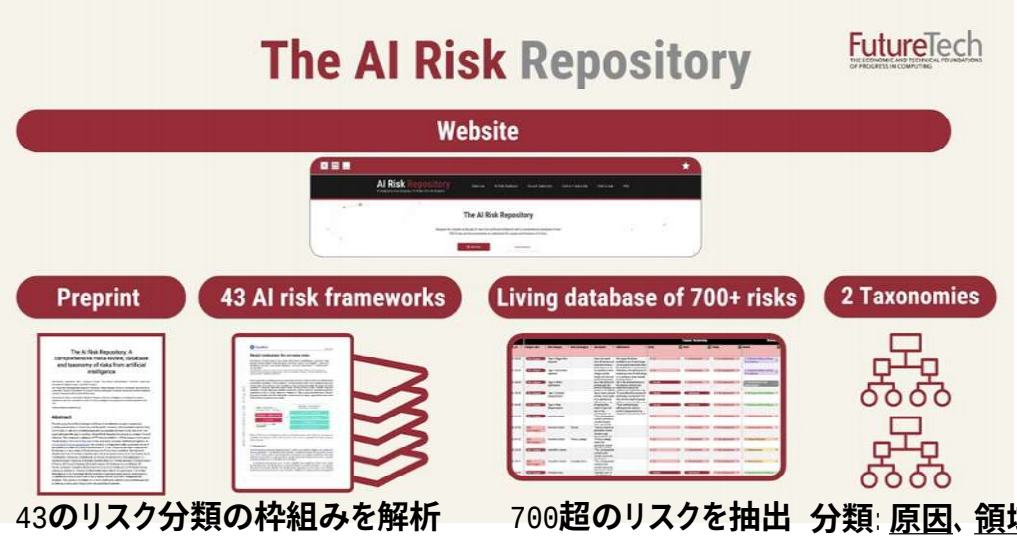
日本企業も要注意！
(注) GDPRと同様に制裁の域外適用 (最大3千万euroか全世界売上高の6%どちらか高い金額)

(**出典**) EU "Regulatory Framework Proposal on Artificial Intelligence" https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

3

資料4

AIリスクリポジトリ(MIT)



43のリスク分類の枠組みを解析　700超のリスクを抽出　分類: 原因、領域

Category	Level	Proportion
Entity	Human	34%
	AI	51%
	Other	15%
Intent	Intentional	35%
	Unintentional	37%
	Other	27%
Timing	Pre-deployment	10%
	Post-deployment	65%
	Other	24%

人間に起因するリスク比率も大

開発後のリスクが大きく、継続的なモニタリングが必要

Note. Totals may not match due to rounding.

Domain / Subdomain	Percentage of risks	Percentage of documents
1 *Discrimination & toxicity*	**16%**	**71%**
1.1 Unfair discrimination and misrepresentation	8%	63%
1.2 Exposure to toxic content	6%	34%
1.3 Unequal performance across groups	2%	20%
2 *Privacy & security*	**14%**	**68%**
2.1 Compromise of privacy by obtaining, leaking or correctly inferring sensitive information	7%	61%
2.2 AI system security vulnerabilities and attacks	7%	32%
3 *Misinformation* 偽情報	**7%**	**44%**
3.1 False or misleading information	5%	39%
3.2 Pollution of information ecosystem and loss of consensus reality	1%	12%
4 *Malicious actors & misuse*	**14%**	**68%**
4.1 Disinformation, surveillance, and influence at scale	5%	41%
4.2 Cyberattacks, weapon development or use, and mass harm	5%	54%
4.3 Fraud, scams, and targeted manipulation	4%	34%
5 *Human-computer interaction* 過度のAI依存やAIによる意思決定 (自動化)	**8%**	**41%**
5.1 Overreliance and unsafe use	5%	24%
5.2 Loss of human agency and autonomy	4%	27%
6 *Socioeconomic & environmental harms*	**18%**	**73%**
6.1 Power centralization and unfair distribution of benefits	4%	37%
6.2 Increased inequality and decline in employment quality	4%	34%
6.3 Economic and cultural devaluation of human effort	3%	32%
6.4 Competitive dynamics AI開発分野での健全な競争	1%	12%
6.5 Governance failure	4%	32%
6.6 Environmental harm	2%	32%
7 *AI system safety, failures & limitations*	**24%**	**76%**
7.1 AI pursuing its own goals in conflict with human goals or values	8%	46%
7.2 AI possessing dangerous capabilities	4%	20%
7.3 Lack of capability or robustness	9%	59%
7.4 Lack of transparency or interpretability	3%	27%
7.5 AI welfare and rights AI倫理	<1%	2%

Note. Domain totals may not match subdomain sums due to rounding and domain-level coding of some risks.

フィルターバブル等

(Source) P. Slattery et al. "Global AI adoption is outpacing risk understanding, warns MIT CSAIL" (MIT CSAIL News, August 14, 2024)

資料5

AI国際条約（2024年9月）

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10か国・地域が署名（日本は未署名）。

ここにフォーカスしているのが特徴

目的

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

適用範囲

- **公的機関のAIシステム**が適用対象。
- 民間部門については本条約の目的・趣旨に沿った方法で対処。

一般的義務

- **人権の保護**
- **民主的プロセスの完全性と法の支配の尊重**

活動原則

- 人間の尊厳及び個人の自律性
- 透明性及び監督
- 説明責任等 (accountability and responsibility)
- 平等及び差別の禁止
- プライバシーと個人データの保護
- 信頼性(reliability)
- 安全なイノベーション(safe innovation)

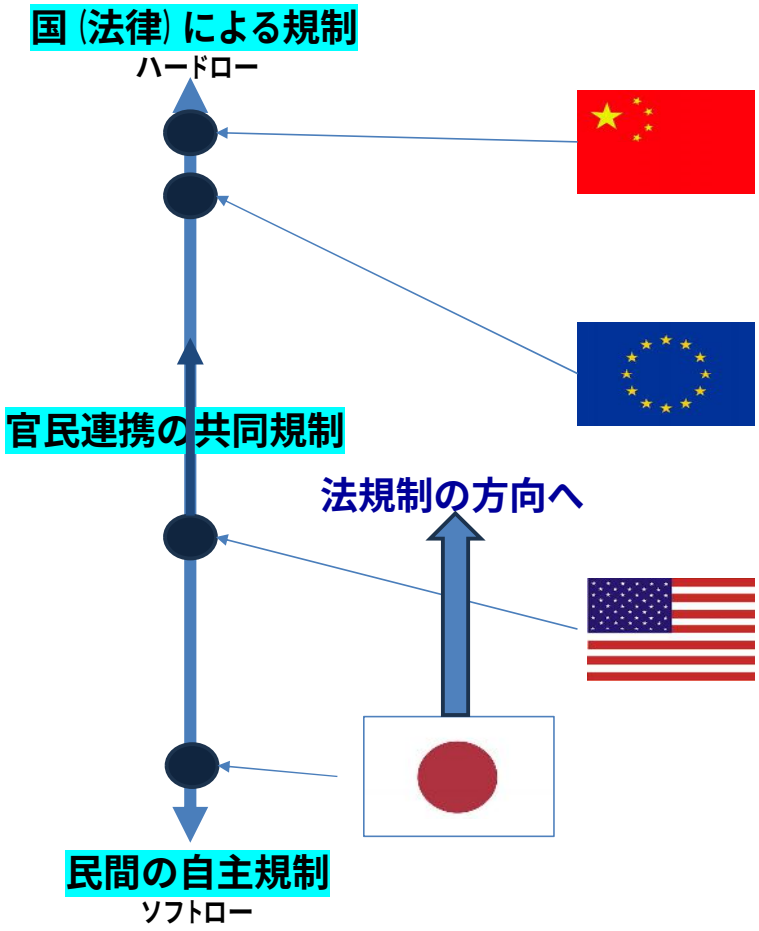
リスク及び負のインパクトの評価と対処

- **リスクの識別・評価・防止・対処**のための所要の措置を講じる
- **人権・民主主義・法の支配にそぐわないAIの使用を凍結・禁止する評価**を実施する。
- 本条約の遵守に関するメカニズムを創設する。

5

資料6

各国におけるAIガバナンスの動き



・2023年8月、「生成人工知能サービスの管理のための**規則」を施行**。
・**実質的に海外生成AIを排除**（法律等で禁止されているコンテンツの生成を禁止）。

・2023年6月、**欧州議会が「AI法案」を採択**。
・リスクベースでAIを4段階に分類し、規制を適用。

・2023年7月、大統領府は生成AIに関する**「自主的なAIコミットメント」**についてAI関連7社と合意。
・同年10月、AIに関する**大統領令**（各連邦政府機関におけるアクションを規定）を公表。

（注）トランプ次期大統領は上記大統領令の廃止を明言。

資料7

中国におけるAI規則の制定

■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。

第4条
1) **法律や行政規則で禁止されているコンテンツの作成禁止**(※)
2) **差別を防止するための効果的な対策**
3)**知財・企業倫理の尊重ほか、独占的・不法な競争行為の禁止**
4)**他人の肖像権、名誉、プライバシー、個人情報に関する権利侵害の禁止**

第6条
・**生成AIに関する国際ルールの策定への参加**

第12条
・**提供者は、(AIにより) 生成されたコンテンツを識別可能とする。**

第21条
・**違反が重大な場合はサービス提供停止を命令。犯罪であれば刑事責任を追及。**

(※) **社会主義の中核的価値観を遵守し、国家権力の転覆の煽動、社会主義システムの転覆、国家の安全と利益を危険にさらす (中略) など、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＝【実質的に海外の生成AIを排除】**

(出典) 中国政府HPを元に読売新聞記事 (23.8.16)などで補足。

7

資料8

米国におけるAIガバナンスの動き（2023年７月）

■2023年7月、大統領府は生成AIに関する「自主的なAIコミットメント」についてAI**関連7社**※**と合意**(拘束力はなく、企業の取り組みの具体策もない)。

※Amazon, Anthropic, Google, Inflection, Meta, Microsoft & OpenAI

安全性

1) AIのモデル･システムの部内･対外的なRed Teaming の実施

2) **情報共有**の促進

セキュリティ

3) **サイバーセキュリティ**や内部脅威に対するセーフガードへの**投資**

4) **第三者による脆弱性の発見･報告を促す仕組み**

信頼

5)コンテンツがAI製であることがわかる**電子透かしなどの開発**

6)AIのもつ能力や限界（公平性やバイアスのような社会的リスクを含む）に関する**情報公開**

7)有害なバイアスや差別の回避やプライバシー保護のように**AIがもたらす社会的リスクの研究を優先**

8)気候変動などの社会課題に対応できる**最前線のAIシステム**(frontier AI systems)**の開発**

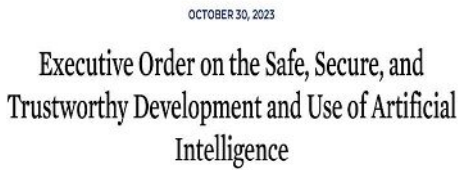
■**大統領令の制定**や超党派によるAI**法案制定の可能性**の追求などにも言及。

8

資料9

AIガバナンスに関する大統領令(2023年10月)

■2023年10月、米国政府はAIに関する大統領令を公表。政府機関において取り組むべき方策について系統的に整理。
■（前掲の）AI関連7社との合意とも整合的。



AIの新たな安全・セキュリティ基準
1) 重要AIについてRed Teaming の結果について政府と共有することを義務付け
2) NISTによるRed Teamingの基準策定、重要インフラへの適用
3)重要インフラの脆弱性の発見・措置のためのAIツールの開発
4)AIの軍事利用に関する適用基準等の検討

(AIがもたらす)人権侵害への対応
5)プライバシー保護技術の開発支援
6)AIアルゴリズムによる差別(=algorithmic discrimination)禁止のための明確なガイダンスの策定
7)ヘルスケア、教育等の分野におけるAIの適正利用の支援

イノベーション・競争の促進
8)公正・オープン・競争的なAIエコスシステムの促進
9)国際的なAI環境整備への協力連携の促進

■超党派によるAI法案制定の可能性の追求にも言及。

(Source)White House "Executive Order on the Safe, and Trustworthy Development and Use of Artificial Intelligence" (October 2023) 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/

資料10

カリフォルニア州AI法(2024年9月)

■カリフォルニア州議会にて2024年2月にScott Wiener上院議員(民)により提案(SB1047)。同年8月可決・成立。同年9月、知事は承認せず廃案。

特　徴

1) **一定規模以上のAIモデルを規制対象。**

・新たに州政府に設ける規制部門(Frontier Model Division(FMD))が定める**計算能力**(当面は10の26乗Flops超)を用いて学習したAIモデルで**コストが1億ドル**を超えるもの。

・**研究、開発、試作段階**で市場にリリースされるAIモデルは**対象外**。

2) **AI開発者の義務**

・**開発着手前の段階**----サイバーセキュリティ対策、完全なシャットダウンの確保、安全性とセキュリティに関する文書策定・遵守などを整備。

・**商業的に利用可能とする前の段階**---「重大な損害」※を引き起こす可能性の有無について評価

※重大な損害----(a)大量殺戮をもたらす兵器の製造や使用、(b)重要インフラへのサイバー攻撃による損害5億ドル、(c)大量の死傷者か5億ドルの損害の発生　等

・**監査報告書の提出**(年次)

・**インシデント発生時の72時間以内のFMDへの報告**

・上記を遵守しなかった場合の懲罰的損害賠償

(注)杉本武重「米AI法、先行くカリフォルニア　不正に「懲罰的損害賠償」」日経デジタルガバナンス(2024年7月3日)。なお、法案本文は以下のURLにて参照可能(https://digitaldemocracy.calmatters.org/bills/ca_202320240sb1047)。

知事コメント

・**議論の鍵**は、**規制を適用する基準がAIモデルの開発にかかるコストや必要な計算能力によるのか**、それとも、そうした要素ではなく**システムの実際のリスクを評価すべきかどうか**という点にある。**【→ガードレールとしての規制導入には賛成】**

・(法案は)**AIシステムがハイリスクな環境で開発されているのか、重要な意思決定に使われているのか、機微性を有するデータを利用しているのかなどを考慮していない**。**【→規制はリスクに基づき適用されるべき】**

・ただ大きいシステムが開発しているからという理由で、(AIシステムの)基本的な機能にまで**厳格な基準**を提供しようとしている。**【→規制水準は低くあるべき】**

(注) https://www.gov.ca.gov/wp-content/uploads/2024/09/SB-1047-Veto-Message.pdf

10

資料11

生成AIとサイバー攻撃

AIに対するサイバー攻撃（例）

・Data Poisoning（**データ汚染）攻撃**--- 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。AIの信頼性を喪失させる。データ改竄などのケースも。

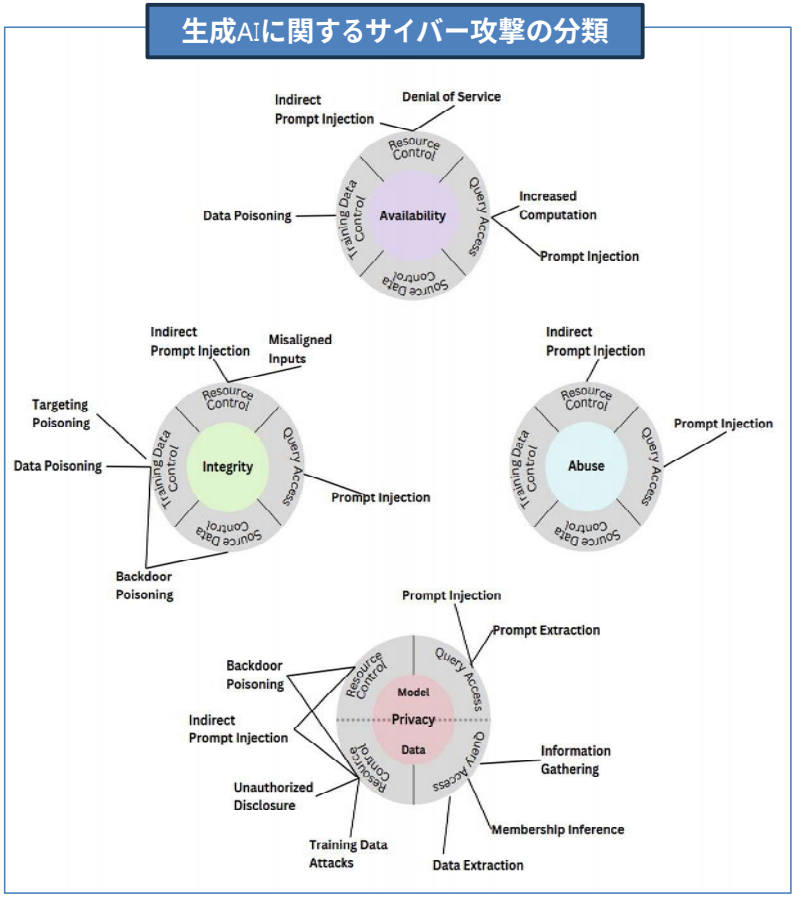
・Data Evasion（**データ回避）攻撃**--- 人間の知覚できないノイズ等を学習データに混入させてAIの判定結果を誤らせる。侵入検知のセキュリティ対策の回避など。

・Model Extraction（**モデル抽出）攻撃**---攻撃対象のAI（機械学習モデル）にデータを投入しつつクローンモデルを作成。正規サービスの妨害、学習データの盗用などの可能性。

AIを利用したサイバー攻撃（例）

・WormGPT/FraudGPT---AIを用いたシステムの脆弱性の探索やマルウェアの作成（ダークウェブに存在）

・BEC（**ビジネスメール詐欺**）---AIによる偽音声・偽動画を用いて取引先と誤認させ金銭の振込をさせる。

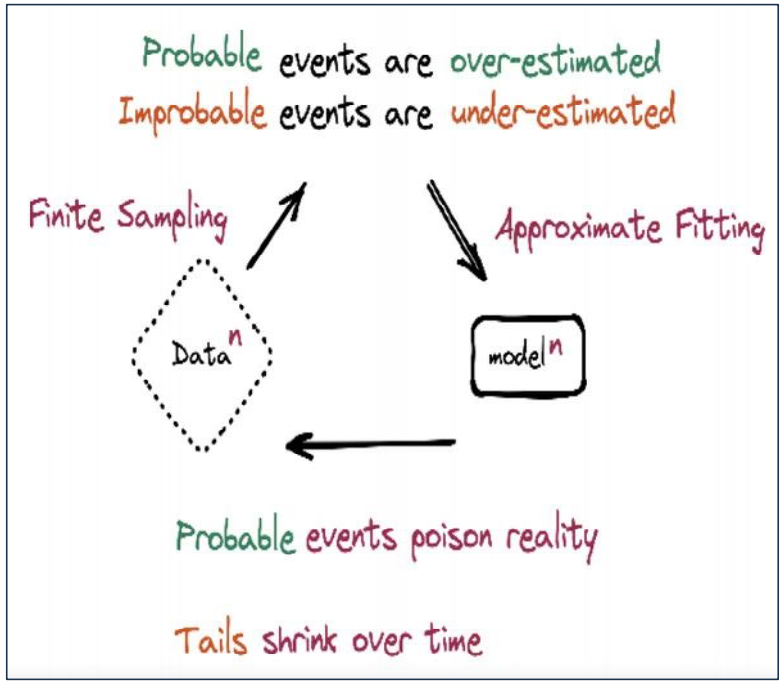


(Source)A. Vassilev et al. "Adversarial Machine Learning : A Taxonomy and Terminology of Attacks and Mitigation" NIST AI 100-2e2023 (January 2024)

資料12

生成AIにおけるモデル崩壊(model collapse)

AIの退化プロセス (degenerative process)



(Source) I. Shumailov et al. "The Curse of Recursion : Training on Generated Data Makes Models Forget" arXiv, May 2023

AIの学習プロセスにおいて、

- **構成比率の低い選択肢が何世代か経て無視される** (モデルへのフィッティングの過程でデータ分散が最小化する)
- **確率論的に誤った出力が行われる**

↓

ここのような**出力結果を学習データに加えるプロセス**が何度も繰り返す (生成AIにおける**"汚染"**の進行)

↓

当初と異なる学習データで構築された**全く異なるAIに変貌する可能性**

↓

モデル崩壊

↓

AIが少数データによる差別 (ジェンダー、人種など) や現実の歪曲を引き起こす可能性

対 処 策

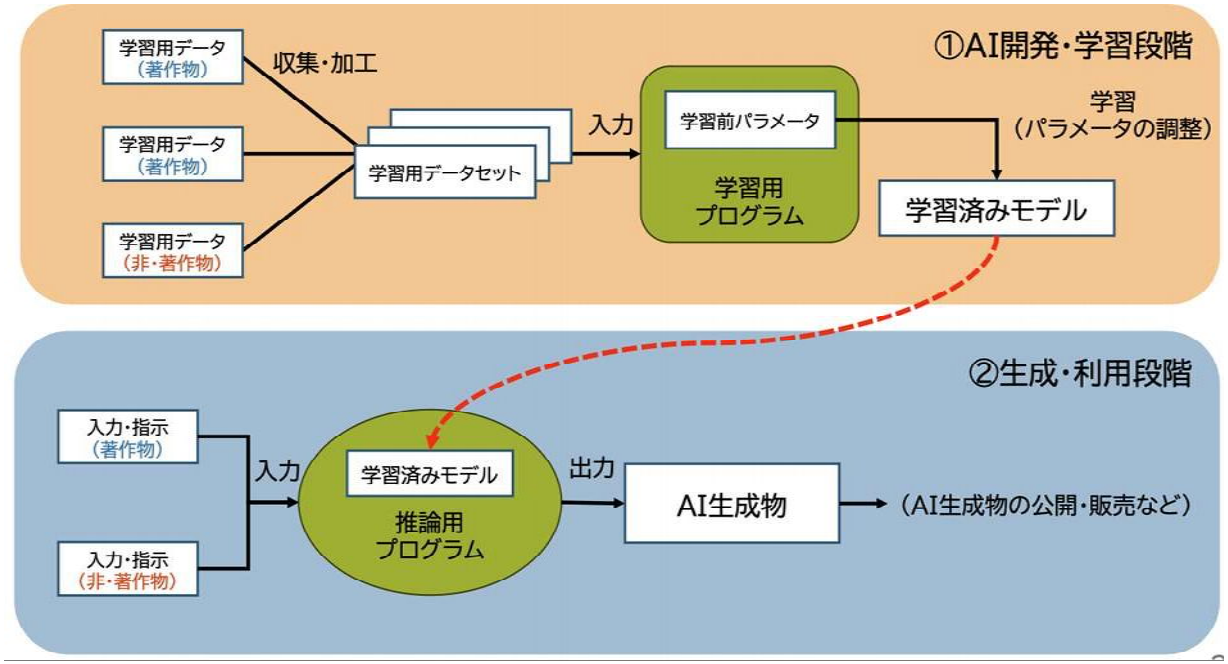
- ●人間の作成した元データの保存/再学習
- ●人間が作成した新規データの学習データとしての追加

(人間作成の学習データのみ切り分ける仕組みが不在、学習コストの上昇)

(Source) Carl Franzen "The AI feedback loop : Researchers warn of "model collapse" as AI trains on AI-generated content" Venture Beat, June 12, 2023

資料13

生成AIと著作権



著作物の「複製」「翻案」等

- 推論用に入力する著作物をサーバーに保存（複製）
- 既存著作物を含む生成物をサーバーやPC上に保存（複製・翻案）

著作物の「公衆送信」「譲渡」等

- 生成物をアップロード（公衆送信）
- 生成物の複製物を販売（譲渡）

「享受」→「著作権者は著作物から経済的利益」
- ・文章の著作物→閲読
- ・音楽・映画の著作物→鑑賞
- ・プログラムの著作物→（プログラムの）実行

AI**開発のための情報解析**のように、著作物に表現された思想又は感情の**享受**を目的としない利用行為は、原則として**著作権者の許諾なく行うことが可能**（著作権法第30条の4）

著作権者の利益を不当に害する場合は上記の適用除外
（例）情報解析用のデータベースの著作物についてライセンス市場が成立している場合

著作権侵害の要件は「**類似性**」（他人の著作物と同一・類似）または「**依拠性**」（他人の著作物に依拠）

個別に判断（要検討）

（論点例）
- ・元の著作物が学習データに含まれる→依拠性あり？
- ・元の著作物と生成物の類似性→依拠性あり？
- ・AIが独自の生成物を出力である証明→依拠性なし？

（注）AIと著作権の関係については文化審議会著作権分科会法制度小委員会において議論が継続されている。
（出典）文化庁著作権課「AIと著作権」（2023年5月）

資料14

頻発するAI訴訟(主として著作権侵害・個人情報保護違反)

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

THE NEW YORK TIMES COMPANY

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC,

Defendants.

Civil Action No. _____

COMPLAINT

JURY TRIAL DEMANDED

NEWS ▪ MEDIA ALLIANCE

News/Media Alliance Applauds NYT Complaint Filed Against Microsoft and OpenAI

By *Staff* December 27, 2023

2023年12月、米国New York Time (NYT)がOpen AI / MSに対し提訴 (New York 連邦地裁)。

- 自社の記事がAI開発のために無断使用されている。Chat GPTは自社記事逐語的な抜粋を生成することがあり、読者が購読料を支払わずに記事を読むことが可能。
- News/Media Alliance (約2,000の北米メディアが加入) がNYTを支持する声明。
- Open AIの反論：記事の学習はフェアユースに該当。

著作物の利用が以下の4条件に適合する場合は著作権の侵害に該当せず

1) 利用の目的と性格 (非営利性、変形的利用など)
2) 著作権のある著作物の性質 (事実の伝達 (例：学術論文) や部分的機能 (例：地図)
3) 著作物全体との関係における利用された部分の量及び重要性
4) 著作物の潜在的利用または価値に対する利用の及ぼす影響 (市場への影響)

(参照) AI画像生成・生成系AI問題まとめwiki
https://w.atwiki.jp/genai_problem/pages/50.html#id_18cc5b57

資料15

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

(US DoS, “Political Declaration on Responsible Use of Artificial Intelligence and Autonomy”, Feb 2023)

"軍事領域における責任あるAIに関する会議[REAIMSummit]"(2023年2月@ハーグ) において、米国務省が提案。軍事分野におけるAIの開発・配備・使用に際し、**自主的に遵守し、そのコミットメントをオープンにする**ことを提案。

- 軍事AI(military AI capabilties)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**されることを保証するため、**法的審査** (legal reviews)**などの効果的な措置**を実施
- **核兵器の使用に関する重要な情報付与や判断は、人間による管理と関与**を維持
- 兵器システムを含むすべての軍事AIの開発・配備は、高官(senior offioials) が監督
- 軍事AIの責任ある設計・開発・配備・使用に関する**原則を採択・公表**
- 軍事AIの開発・配備・使用は、適切なレベルの職員が判断
- 軍事AIの**意図しない偏り**(unintended bias)**を最小化する対策**を実施
- **監査可能**(auditable)な方法やデータソース、設計手順、文書によって軍事AIを開発
- 軍事AIを使用する職員、及び使用を承認する職員は、その能力と限界を十分に理解し、その使用について適切な判断を行うことができるよう訓練。
- 軍事AIの安全性・セキュリティ・有効性についは、**ライフサイクル全体にわたって厳格なテストと保証**の対象とし、自己学習による軍事AIは、重要な安全機能が低下していないことを確認する監視プロセスに従うこと。
- **意図しない結果を検出・回避・解除できるよう設計する等のセーフガード**を導入。
- 軍事AIの開発・配備・使用に関する議論を継続し、他の適切なコミットメントを見出すよう努力。

米国提案に対し、51か国 (EU加盟各国を含むG7各国など) が賛同。

(出典) https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/

資料16

調査報道「イスラエル軍によるAIを用いた爆撃」

(出典) Yural Abraham "Lavender' : The AI machine directing Israel's bombing spree in Gaza" (April 3, 2024) +972 Magazine
https://www.972mag.com/lavender-ai-israeli-army-gaza/

- 標的を拡大するため、イスラエル軍(8080部隊) のガザでの「大量監視システム」(約230万人分) の情報がAI標的システム"Lavender'"に入力され、37,000人に及ぶ標的を生成。標的生成に際し、既知の工作員の特徴に合致する標的を100ポイント制で評価。
- Lavender'の標的精度は90%であることが事前チェックで判明したが、ガザ攻撃の標的生成に利用。
- 軍はLavender'が生成する標的を自動的に承認するよう決定。
- 攻撃前の軍情報部による詳細な確認は不要とされ、短時間で形式的な確認作業。
- 多くの標的は下位の戦闘員 (15〜20名の巻き添え被害は承認) 。
- 標的追跡システム"Where's Daddy?"によって自宅に戻った後で攻撃を実施。家族が巻き添え。

(出典) 川上泰徳「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…イスラエル独立メディアが調査報道」Yahoo! News (2024年4月9日)
https://news.yahoo.co.jp/expert/articles/c72d4cbc32aa5577eac494dfd75b43652a20555f

【参考資料①】
～「デジタル政策の論点 2024」より～

「デジタル政策の論点 2024」より

デジタル政策フォーラムでは、2024 年 7 月に「デジタル政策の論点 2024」と題して、12 名の有識者との対話をまとめたデジタル政策を巡る論点集を刊行している。本参考資料においては、「デジタル政策の論点 2024」より、特に AI ガバナンスに関連する箇所を抜粋して掲載する。

なお、「デジタル政策の論点 2024」は書籍としても購入可能なほか、デジタル政策フォーラムのウェブページでも掲載している。本参考資料として抜粋した AI 関連だけでなく、AI を含むデジタル政策全般の俯瞰的な把握と理解を深めるため、是非ご一読いただきたい。

コンテンツ・コンタミネーションへの対策が急務

徳田 英幸　情報通信研究機構（NICT） 理事長

“

生成 AI の状況を見ていると、技術だけでは解決が難しい部分があると思います。

例えば「steganography（ステガノフラフィー）」という技術で生成 AI が作った画像に電子透かしを入れると、半分に切っても 4 分の 1 に切っても透かしが残り、生成 AI が作ったものだということを判別できます。ほかにも、「敵対的生成ネットワーク（Generative adversarial networks、GANs）」という技術があります。生成ネットワーク（generator）と識別ネットワーク（discriminator）という“敵対”する 2 つの AI を組み合わせ、AI が作った偽物の画像と本物の画像を見分けるようなことが可能になります。それを競うコンテストもあって、いろいろな研究者グループがしのぎを削っています。

ただし、悪意をもって画像を作るような人が画像を作る段階で透かしを埋め込むことはないでしょうし、GANs はまだ完全ではありません。今のところ、人が作ったコンテンツなのか AI が生成したコンテンツなのかを技術だけで高精度に判別することは難しい状況です。

これに対しては、制度やルールを整え、違反に対する罰則を設けることで対応する必要があります。これまでのところ、生成 AI に関しては、「とにかく自由に使ってみよう。大人も子供も誰もが参加して AI を民主化しよう。技術の進歩を止める規制には反対」という流れで進んできました。もちろん良い面もあったのですが、冷静に見れば“野放し”の状態にあるわけです。結果的に、真偽や出所の怪しいコンテンツ、偽情報がデジタル空間に出回って「**コンテンツ・コンタミネーション（情報汚染）**」がかなり深刻なレベルまで進んでしまったと認識しています。なんらかのルール、標準化や認定制度を制定することが必要と思います。

国内では、IPA（独立行政法人 情報処理推進機構）に AISI（AI Safety Institute）が設置され、海外機関とも連携して AI の安全性評価に関する基準や手法の検討等がされます。また、OECD が進めている GPAI（Global Partnership for AI）への貢献の強化と広島 AI プロセスの取組の推進の一環として、アジア地域初の専門家支援センターとなる GPAI 東京センター（仮称）が NICT（国立研究開発法人 情報通信研究機構）に設置される予定です。

”

『デジタル政策の論点 2024』（2024 年 7 月刊）

第1章 デジタル技術のフロンティアから抜粋

AIリテラシーは政策立案者や規制当局者に不可欠

林 秀弥　名古屋大学 大学院法学研究科 教授

“

欧州議会は2024年3月、世界初の包括的なAI規制法案（Artificial intelligence Act）を可決しました。安全の確保、人権の保護などを最優先し厳しい規制を求める勢力と、過度な規制はEU加盟国の国際競争力（特に対米）を削ぐとして反対する勢力の妥協の産物ですが、AIリスクを4段階（Unacceptable、High、Limited、Minimal）に分けて規制内容を細かく規定している点など、今後のAI規制のひな形になることは間違いありません。

ただし、対巨大プラットフォーム事業者に特化したセーフガードというものは設けていません。果たしてそれで大丈夫なのか、またもや遅きに失して、プラットフォーム事業者による独走・独占を許すことになってしまうのではないかという懸念はあったのですが、2024年1月に欧州委員会（EC）は「仮想空間と生成AIにおける競争（competition in virtual worlds and generative AI）」に関して競争法の観点から検討を開始することをアナウンスしました。合わせて、MicrosoftによるOpen AIへの出資に関してEU企業結合ルールに則って評価する考えを示しました。

こうした問題意識はイギリスの競争・市場庁も共有していて、2023年秋には「AI基盤モデルの一次報告書」を公表しました。その中で、AIが関係する反競争的行為のリスクや、それに競争法でどう対処するのかについて包括的に議論が行われています。ただ、まだ隔靴掻痒の感があり、リスクや可能性の話は書かれていますが、実際にまだ起こっていないので頭の体操をしているような印象です。

今後、AIは、医療、交通、金融、エンターテインメントに至るまで、私たちの生活の様々な側面に浸透し、様々な影響を与えてくると思います。AIが市場競争に与える影響を的確に分析するためには、まずAIへのリテラシーを高めることが重要だと思われます。個人や企業がAIを包括的に理解し、社会やビジネスへの影響を把握することがますます重要になってきます。過去にメディアリテラシーやデジタルリテラシーが重視されたように、現在では、市場競争の観点からも、AIリテラシーを重視することが不可欠になってくると思います。 広い意味での**AIリテラシーは、個人だけでなくAIに関する政策立案者や規制当局者にとっても不可欠**だと思います。

”

『デジタル政策の論点2024』（2024年7月刊）

第3章 データ駆動社会と競争政策から抜粋

新しい法の役割が求められている

生貝 直人 一橋大学大学院 法学研究科ビジネスロー専攻 教授

“

（日本政府にとって）EU の AI 法のような包括的な規律・枠組みを今後どうしていくのかは、大きな焦点になってくるでしょう。

言及しておきたいのは、EU が次々に法制化を進めているといっても、決して「ハードロー」一辺倒ではなく「ソフトロー」の要素もうまく組み入れているということです。我が国の場合、法律で細かく規定するハードローか、ガイドラインのようなソフトローでいくのか、二分法の議論になりがちなのですが、EU は両極の中間領域の開拓に長けており、蓄積も豊富です。（中略）0 か 1 ではなく、その間のどこかに答えがあるという EU のアプローチは、日本の政策当局も参考にすべきだと思います。

もう一つ、ホリゾンタル（横）とバーチカル（縦）の両面の視点をもって、ハードロー、ソフトローを使い分けていくことが重要だと思います。日本の AI ガイドラインは分野横断的にホリゾンタルな網をかけるところにとどまっていますが、EU では政治広告やディープフェイクといった個別具体的な課題について AI 法とは別に縦軸の規制をかけるといった柔軟な対応をしています。そうした分野横断的課題と個別課題の両面から丁寧な議論をすることが、我が国のこの 1 年くらいの重要な課題になってくると感じています。

特に、規制の大枠を法律で定めつつ詳細を事業者の自主的取り組みにゆだねる「共同規制」の手法は、今後日本でも多用されていくことになるでしょう。そこで重要になってくるのが「透明性の確保」です。事業者からデータの提供を受けることを含め、しっかりモニタリングをしていく。政府、市民団体、アカデミアなど様々なステークホルダーを巻き込むことが重要です。（中略）そうした国家、市場・ビジネス、市民社会、アカデミアによる新しい協力関係が様々な形で模索されている状況にあるのだと思います。

そうした中で、法がどのような役割を果たすべきか、というところに戻ってくるのです。共同規制やソフトローがうまく回っているのかどうかを評価可能とし、マルチステークホルダーによる改善を進めていくためには、従来とは異なる法の枠組が求められます。規制するか、規制しないかということではない、**新しい法の役割が求められている**ように思います。

”

『デジタル政策の論点 2024』（2024 年 7 月刊）

第 7 章 データ駆動社会とルールのあり方から抜粋

オープンという御旗は決して降ろさない方が良い

宍戸 常寿　東京大学大学院 法学政治学研究科 教授

“

（AI のオープンソース性が悪用されている状況があるとしても）基本はオープンなのだと思います。悪用のリスクを恐れて閉ざしても、悪いことをする人は結局出てくるでしょう。それを防ぐためにも徹底的にオープンにするという考え方もあると思います。ただし、現実的には悪いことをしやすい部分については公開範囲を限定するとか、一定期間クローズにするとか、個別対応が避けられないでしょう。基本はオープンで、クローズはケースバイケースで判断ということです。

日本学術会議の「オープンサイエンスを推進するデータ基盤とその利活用に関する検討委員会」（委員長：喜連川 優 情報・システム研究機構国立情報学研究所所長、デジタル政策フォーラム発起人）に私も委員として参画していたのですが、科学的知見をクローズにすることは人類の知の発展を妨げてしまうという危機感に基づいて、あるべき姿を検討しています。例えば新型コロナウイルス感染症の問題一つをとっても、何が有効な対処法なのかは初期段階ではよく分からなかったのですが、世界中の知見やデータがオープンな場に共有されるにつれて分析・研究が加速し、有効な対策が打てるようになりました。**オープンという御旗は決して降ろさない方が良い**と思います。

一方、日本企業や日本政府はリスクを高く見積もり、新しいことに対して慎重に、限定的に、閉鎖的に取り組む傾向があります。オープンに連携して広げることよりもクローズに向かい、小さくまとまってしまうのです。「基本的なデータをオープンに共有・活用することで新たな価値を創り出すデータ連携社会を築こう」という掛け声が政府と経済団体によって何度も叫ばれ、旗が振られましたが、一向に進みませんでした。そういうところは変えていかなければなりません。

”

『デジタル政策の論点 2024』（2024 年 7 月刊）

特集1 データ駆動社会におけるガバナンス機構から抜粋

AI ガバナンスで何を守るべきなのか？

佐藤 一郎 大学共同利用機関法人 情報・システム研究機構
国立情報学研究所（NII） 情報社会相関研究系 教授

聞き手：**谷脇 康彦** デジタル政策フォーラム 代表幹事

❝

G7 伊勢志摩サミット以降、欧米から遅れ始めた

谷脇 AI ガバナンスのあり方が大きな問題となっています。例えばユーラシアグループは「2024 年 10 大リスク」、とりわけ 10 年後を視野に入れた長期的なリスクの一つとして「**AI ガバナンスの欠如**」を挙げています。最近の AI 関連技術の進化の速さは AI ガバナンスのためのルール策定の動きを超えているとの懸念も出ています。そこで、そもそも AI を統治することが可能なのか、まず総論的に伺いたいと思います。

佐藤 AI ガバナンスは大切ですが、その前にガバナンスで何を守るかを決める必要があります。AI ガバナンスを確立することはとても難しいですし、AI ガバナンスを確立したところで AI に関わる様々な問題が一掃されるわけでもありません。「コーポレートガバナンス」を作っても企業の不祥事はいろいろなところで起きているわけで、ガバナンスを作っても問題は生じます。基本的なスタンスとして、ガバナンス万能論に偏るべきではないと思います。

「コーポレートガバナンス」は長年の積み重ねで確立されました。また、個人的には、ガバナンスは失敗によって整備されていくものだと考えています。AI のガバナンスを確立するには、我々は幸か不幸か、AI に関する失敗の経験が少な過ぎるのです。今ガバナンスを作ったとしても、すぐに失敗して、また作り直すことになります。ガバナンスというのは永遠に作り直していくものなのかもしれません。

従って、AI ガバナンスの確立を目指すことは大切ですが、現時点で AI ガバナンスの確立を前提にしていろいろな話を進めるのは少々危険だと見ています。

谷脇 日本は AI の議論への着手は他国よりも早かったのですが、EU（欧州連合）が AI 法をつくり規制の域外適用も予定しており、中国も生成 AI に関する法律を定めて“中華 AI”を推進しています。共同規制に軸を置いているアメリカも大統領令を出し、連邦機関において技術基準の策定などが進み、ガイドライン中心の日本はどんどん追い抜かれてしまっているように思います。

佐藤　2016 年の G7 伊勢志摩サミットの時点までは、日本、ヨーロッパ、アメリカは横並びでしたが、その後に OECD（経済協力開発機構）で AI に関する国際的な政策ガイドラインの検討・議論が進められ、2019 年 5 月に「**AI に関する OECD 原則**」が採択されました。

私は、ちょうど同じ時期に OECD の別委員会（研究データ倫理）の日本代表を務めていたので、AI 原則に向けた議論を横目で見る機会があったのですが、率直に言って、日本はそこでの議論についていけていませんでした。ヨーロッパ各国もアメリカも専門家を送り込んでいましたが、日本は必ずしもそうではありませんでした。

「AI を規制する」ということが議論の主題だったはずなのに、日本国内ではいつの間にか「**AI の倫理**」の話にすり替わっていました。日本では企業にも政府にも「規制があるとイノベーションは生まれない」という奇妙な思い込みが蔓延しているからかもしれません。日本の議論はちょっとふわふわした方向に傾いていました。

OECD のような国際的な場での実務議論では、自国の主張を通すには、他国の主張に対して理詰めで反論する必要があり、そのためには自国内での議論の積み重ねが極めて重要になります。EU は AI を規制する法律を作ることを目指していました。法律を作るにはかなり緻密な議論の積み重ねが必要であり、EU はそれを実践してきたと見るべきでしょう。アメリカは「規制反対」の基本スタンスをとった点では日本と歩調が合っているように見えましたが、舞台裏ではかなり深い議論を重ねており、「**AI 権利章典**（**AI Bill of Rights**）」という指針を作り、それが「**責任ある AI**」に関する議論につながっていきます。

その頃から欧米との差が開き始めたように思います。

■AIに関するOECD原則の概要

（1）AIは、包摂的成長と持続可能な発展、暮らし良さを促進することで、人々と地球環境に利益をもたらすものでなければならない。

（2）AIシステムは、法の支配、人権、民主主義の価値、多様性を尊重するように設計され、また公平公正な社会を確保するために適切な対策が取れる－例えば必要に応じて人的介入ができる－ようにすべきである。

（3）AIシステムについて、人々がどのようなときにそれと関わり結果の正当性を批判できるのかを理解できるようにするために、透明性を確保し責任ある情報開示を行うべきである。

（4）AIシステムはその存続期間中は健全で安定した安全な方法で機能させるべきで、起こりうるリスクを常に評価、管理すべきである。

（5）AIシステムの開発、普及、運用に携わる組織及び個人は、上記の原則に則ってその正常化に責任を負うべきである。

■各国政府に対するOECDの提言

―信頼できるAIのイノベーションを刺激するために、研究開発への官民投資を促進する。

―デジタルインフラとテクノロジーでAIエコシステムとデータと知識の共有メカニズムの利便性を高める。

―信頼できるAIシステムの普及に道を開く政策環境を創出する。

―人々にAIに関わる技能を身につけさせるとともに、労働者が偏りなく転職できるよう支援する。

―情報を共有し標準を開発し、責任あるAIの報告監督義務を果たせるように、国際的、産業部門横断的に協力する。

生成 AI の楽観シナリオが浮かんでこない

谷脇　その点については、近著『ChatGPT は世界をどう変えるのか』でも指摘されていますね。佐藤先生が一般の方々に最も伝えたかったことは何だったのでしょうか。

佐藤　生成 AI が社会に与える影響について考えて頂きたかった、ということに尽きます。POC（Proof of Concept、概念実証）の段階では「生成 AI って面白いな」「なんだか便利そうだな」と利便性ばかりを見ていたとしても、実用化するとなればリスクや影響に対峙しなければいけません。リスクは技術で低減できるものもありますが、生成 AI は社会に対してかなり大きな影響を与えるものなので技術だけでは解決しません。生成 AI を社会がどう受け入れるのか、あるいは拒絶するのかということについてスタンスを定めなければなりません。社会にどういう影響があるのか、予測も交えてお伝えし、皆さんに考えていただこうというのが執筆の狙いでした。

谷脇　AI については、「生産性を上げて付加価値も上げる」とか「人間の仕事を奪う」

とか、光と影の両面、様々な言説が唱えられてきました。そして最近では、AIが生成したコンテンツがネット上に増え、それらをAIが学習していくと「Model Collapse（モデル崩壊）」が起こり、情報の質が落ちるという指摘も出てきました。AIは情報空間全体にどのような影響を与えるのでしょうか。

佐藤　局所的にはともかく、社会全体として本当に生産性や利便性が上がるのかについてはかなり疑問です。

AIが生成するコンテンツは増える一方ですが、読まれない無駄な情報がどんどん増えていきます。例えば、業務日誌のまとめに生成AIを活用すれば、これまでポツポツと箇条書き程度だったものをきちんとした文章に整えてくれます。ただし、本質的に内容は何も変わっていないのに文章の量だけが増える。それを上司は毎日読まされるわけです。生成AIは読み手側の負担を増やしてしまう技術なのです。ある程度まで利用が進むと、読み手側は生成AIではなく人間が編集した情報を選別して読むような方向に戻っていくような気がします。

また、生成AIが生成するコンテンツの品質は、今後、上がるとは限らない。下がる可能性もあります。その理由は生成AIが作ったコンテンツを生成AIが学習するからです。品質も全体的に低下します。そうした質の低い、薄いコンテンツがネットに流れ、AIが学習し、また新しいコンテンツを生成するというサイクルが繰り返されていくと、「**AIの共食い**」のような状態に陥ってしまい、AIの学習モデルそのものが破綻してしまいます。

生成AIは利用者が知りたいことに直接的に答えてくれるのでとても便利です。ウェブ検索は知りたいことに答えてくれるわけではなく、知りたい情報が載っているウェブページを教えてくれているだけ。あくまで間接情報であって、知りたいことにたどり着くためにはウェブを見にいって、そこの情報を読み込まなければなりません。生成AIによってその手間を省けるとしたら、利用者は検索しなくなり、ウェブを見なくなっていくでしょう。ウェブコンテンツの質の悪化がその傾向に追い打ちをかけます。

すると、ネット広告の効果が下がります。当然、広告料、掲載料が下がります。ネット広告の代理店ビジネスが大打撃を受け、ネット広告を表示させることで収益を上げていたネットサービス事業者の収入も減ります。広告収入で成立してきた無料ネットサービスはSNSなども含めて立ち行かなくなるでしょう。これまでのように、誰もが無料で様々なネットサービスを享受できる時代は終わりを告げるかもしれません。一般の利用者の情報発信機会も減り、インターネット全体の利便性が失われ、活気がなくなっていくことになるかもしれません。

どう考えても、生成AIに関して楽観的なシナリオが浮かんでこないのです。

プラットフォーマーのAIへの取り組みに戦略的温度差あり

谷脇　マイクロソフトはオープン AI（ChatGPT）、グーグルは Bard、アマゾンは Titan、メタは Llama といずれも生成 AI に取り組んでいます。無料サービスを提供し、利用者の個人情報を収集し、広告収入で巨利を得てきた巨大プラットフォーマーです。自分たちのビジネスモデルを崩壊させかねない生成 AI を推進するのは、矛盾しているように見えます。

佐藤　巨大プラットフォーマーといっても、各社各様で事情が異なります。マイクロソフトはネット広告ビジネスが必ずしもうまくいっていないので失うものはなかった。最も積極的に動いているのも頷けます。

　他方、生成 AI が**‘検索スルー’**をもたらせば一番打撃を被るであろうグーグルは、AI 開発を進めていることをアピールしながらも躊躇が見え隠れします。戦略的に様子見をしていると言った方がいいかもしれません。グーグルにとっては、ネット広告収入への影響を考えれば、生成 AI、対話 AI の普及が遅れるほうが望ましいと考えている可能性があります。

　また、ネット広告への依存度が高いグーグルやメタが生成 AI の提供で先行した場合、「生成 AI の出力を操作した**ステルスマーケティング**ではないのか？」と疑われ、批判にさらされる可能性があります。例えば、グーグルの生成 AI で料理のレシピについて聞くと、答えにいつも特定のメーカーの、特定の調味料が含まれていたとしたら、それはステマなのではないかという疑念が生じるでしょう。

　こうして見ると、巨大プラットフォーマーの生成 AI に対する向き合い方にはかなりの温度差があり、マイクロソフト系が生成 AI で先行したのは必然だったのかもしれません。ただし、その先行で生成 AI の勝負に決着がついたというわけではないと思います。

情報統制、言語・文化のデカップリングが進む恐れ

谷脇　AI による偽情報・誤情報の拡散リスクについてどう対処すべきでしょうか。ステルスマーケティングの話が出ましたが、AI の中立性や客観性はどうやって担保されるのでしょうか。

佐藤　生成 AI の出力の中立性・公平性をいかに確保するかについては、今後、議論になってくると思いますが、現状では良い方法がありません。出力を歪める方法の一つは**アルゴリズムを歪める**ことですが、これは技術的に検証することができます。しかし、**学習データ・訓練データの偏り**については判断が難しい。ですから、「偏っているかもしれない」「ステマかもしれない」という疑心暗鬼を抱きながら使うしかない状態が続くかもしれません。

皆が中立・公正なものを使いたいとも限りません。フィルターバブルが指摘されるように、自分にとって気持ちが良い情報空間にいたいと思う人はたくさんいます。自分にとって心地良い偏りは、すんなり受け入れられてしまうのです。

強権的な国家には、生成 AI は魅力的な技術です。国民がウェブをあまり見なくなって、生成 AI に頼るようになってくる、しかも国家にとって都合の良い情報を出力するように操作できるとなれば、国家にとって都合の良い情報空間に国民を閉じ込めておけるので、強力な情報統制装置になります。

この問題はさらに根深い危険性があります。生成 AI を構築・提供するのは技術的に簡単ではないので、すべての強権的国家が政権に都合の良い生成 AI をつくって国民に提供できるとは限りません。そうすると、偏った生成 AI を構築できる強権的な国から、個別にチューニングされた生成 AI の提供を受けることになるかもしれません。生成 AI を提供した強権国が、その提供を受けた強権国の情報統制を担うことになり、その国の世論を操作できることになります。その生成 AI は、提供した国を批判するような情報は制限するでしょうから、ある特定の強権的な国が提供した生成 AI が多くの国に広まれば広まるほど、生成 AI を提供した国が望まない情報を与えられない人が他国にも増えていくことになります。このようにして生成 AI が国家統制の手段から国家覇権の手段になって、生成 AI ごとにブロック化された情報空間が形成され、**情報統制による世界のデカップリング**が進むことになるかもしれません。

もう一つの深読みは、生成 AI によって人類は**「バベルの塔」のように世界共通語を失う**かもしれないということです。ChatGPT は英語や日本語を含めて複数の言語で利用できます。ChatGPT に日本語で質問すると、それが海外に関する情報でも ChatGPT は日本語で回答してきます。ウェブ検索の場合は日本語以外のウェブサイトもリスト化されるのとは違います。

これまでは、何か知りたいことがあるとき、海外の情報なら英語や現地の言葉で読まなければならないということが多くありました。外国語、特に世界共通語である英語を学ぶ理由の一つがこの「知識の習得」でした。ところが、生成 AI を使えば母語で質問して母語で答えてくれるとなると、知識や情報を得るために外国語を勉強する必要性が減ってくることになります。少なくとも、世界共通語の英語の地位・重要性は下がると考えられます。

その状況は、塔を作って神に挑戦しようとした人類に対し、通じ合わない異なる言葉という罰が与えられたという旧約聖書の物語に通じるものがあります。

強権国にとっては都合が良いのです。「科学技術や先端知識を欧米先進国から学び取るために国民の英語学習をやむなく認めているが、なまじ英語の情報に触れるから国家にとって不都合な思想に染まる国民が出てくる」と考えているような国家であれば、英語を知らなくても欧米の先端技術を採り入れられるのなら英語教育をしないという判断をしてもおかしくありません。先ほどの情報統制によるデカップリングだけ

でなく、**言語・文化における世界のデカップリング**が進む恐れがあります。

谷脇　私も自著『**教養としてのインターネット論**』に、インターネットの分断が進んでいるということを書きました。西側先進国と中国やロシア、それぞれの陣営のインターネットに対する考え方、国の関与の仕方が全く違う。両陣営の間に「デジタルのベルリンの壁」ができつつあり、それを乗り越えるのが難しくなっている。そこにグローバルサウスといった新興勢力も加わって、フラグメンテーション（断片化）が決定的となり修復不能になってしまうのではないかと危惧しています。

佐藤　私もそう思います。外国語を学ぶということさえしなくなってくると、世界がつながっていなかった中世以前に戻ってしまうような恐れがあります。また、英語が世界共通語になったのは第二次世界大戦以後であり、英米という主要戦勝国の言葉が世界標準語になったと言えます。生成 AI によってその英語の地位が下がるとしたら、それは第二次大戦後の世界スキームが終わることを意味するかもしれません。

アメリカも EU も AI 規制の実務レベルまで詰め切れていない

谷脇　危機感を新たにしました・・・。偏った AI を生み出さないために、EU やアメリカは法や自主規制の網をかけ、説明責任や透明性を事業者に求めようとしています。理念としては理解できるのですが、AI に対する第三者の検証が、真に客観的に行い得るのでしょうか。実効性に乏しいのではないかと感じるのですが。

佐藤　そうなる可能性は高いと思います。AI の中立性を検証できる範囲はとても限定的で狭いのです。アルゴリズムについては学習データと出力の相関を見ればある程度分かります。しかし、学習データそのものの中立性については検証が非常に難しい。日本の政権・与党における議論はアメリカの議論の影響を色濃く受けているのですが、当のアメリカの基準作りは難航必至なのです。

2023 年 10 月にバイデン大統領は「**人工知能の安心、安全で信頼できる開発と利用に関する大統領令**」を発令しましたが、安全性とセキュリティの新基準づくりについては商務省傘下の**国立標準技術研究所**（**NIST：National Institute of Standards and Technology**）に丸投げしたかっこうになっています。NIST がどのような基準を作るのか全く見えてきません。アメリカの動きに追従するなら、まずは NIST の新基準がどうなるかを見極めてからということになるのですが、おそらく、NIST の担当者は頭を抱えていることでしょう。

ヨーロッパに関しては、EU の AI 規制の検討は初期段階では AI を組み込んだ「製品」の安全性から入ったのですが、最終的なまとめの段階では製品を前面に出さず

「AI のリスク」に応じて規制を変えるという整理をして、結果的に自由度を上げました。おそらく検討段階で生成 AI は議論のスコープに入っていなかったはずですが、リスクの考え方でうまく対応したというか、うまく一時避難できたというところです。実際の法執行の段階になると、いろいろと悩ましい判断を迫られることになると思います。

アメリカもヨーロッパも、実務レベルのところについては詰め切れていないので、日本がその部分について検討を深められれば意見を聞いてもらえると思うのですが。現状、残念ながらそのような議論は行われていません。

まずは AI をしっかり定義することから

谷脇　AI の透明性や多様性を確保するためにオープンソースを指向するという可能性もあると思いますが、他方、オープン性を確保することで WormGPT や FraudGPT のような悪意をもった AI 活用が増える恐れはないでしょうか。

佐藤　現時点で、オープンソースが良いのか、悪いのかという結論めいたことは言えません。アメリカの巨大プラットフォーマーでも対応が分かれています。

マイクロソフトとグーグルは AI 技術の公開には慎重な姿勢を示す一方、メタは生成 AI を含めてオープンソースとすることで外部を巻き込んだ研究開発を指向しています。これまでオープンソースはソフトウェアの普及と発展に有効とされてきましたが、生成 AI のように多用途に使える技術の場合、フェイク情報や不正アクセス支援などに悪用される可能性も高いので、オープンソースの是非が見直される可能性はあると見ています。

マイクロソフトもグーグルも、オープンソースそのものに対して全く否定的なわけではなく、分野によってはかなり積極的に取り組んでいます。言い方を変えると、うまく使分けています。そうした事業者が AI のオープンソース化については懐疑的になっているという事実をしっかり受け止めるべきだと思います。AI のリスクや悪影響を止める手段がない以上、オープンソース化に一定の制限をかけるのは避けられないと思います。画像認識のような特定用途のものは別にして、汎用的なものについては慎重な対応が必要です。

余談を二つほど。一つは AI のリスクを考える場合、「AI 内部で安全性を高める視点」と「AI の外側で防御する視点」があるということです。AI を規制するということだけではなく、AI が何か問題を起こしたときにどのように被害を食い止めるか、**防御のための仕掛け**も含めて議論すべきだと思います。

もう一つは AI を恐れるのなら、**AI というものをまずしっかり定義すべき**だということです。AI とは何かをふわっとさせたまま議論していることが多いように感じます。多くの人が AI と呼んでいるものは、たいてい実用化前のものです。実用化されると AI で

はなく「〇〇処理」とか「△△システム」という名前で呼ばれるようになります。20 年前の AI は「推論」とか「検索」でした。今、検索エンジンを AI と呼ぶ人はほとんどいません。AI という言葉が対象とするものは時と共に変わっていくものなので、制度設計する前に対象、つまり AI を定義づけるということをすべきだと思います。

お手本は EU の AI 規制

谷脇　各国では AI に課すべき規律として、「法的規制」「共同規制」「自主規制」などが検討されていますが、どのようなかたちの規律が適しているのでしょうか。

佐藤　AI の進歩の速さを考えると自主規制や共同規制的な方法が選択肢となりますが、必ずしも機能するとは限りません。例えば共同規制の場合、AI を担っている企業のスポンサーは大手プラットフォーマーであり、国の指導においそれとは従わない。そうなると法律に基づくハードローを組み合わせたエンフォースメント（行政上の強制執行）が必要になってきます。そのレベル感は課題に応じてとしか言えませんが、まずは AI の文脈以外においても国家と大手プラットフォーマーの関係性を見直していくべきだと思います。

日本にとってお手本になるのは EU の AI 規制でしょう。まずは EU の AI 規制をよく研究して、取り入れるところは取り入れ、問題があれば問題があるということを EU にフィードバックする。EU も EU 域内だけで AI 規制が有効に働くとは考えていないでしょうから、日本が仲間になると言えば聞く耳を持つと思うのです。EU の基本方針に合わせつつ、言うべきことははっきり言うというスタンスをとるのが現時点での最善策、というか、それ以外に打てる手がないように思います。

EU の AI 規制を手本と位置づけるのは、私は個人情報保護法の改正にも関わってきたのですが、その立場で見えたのは、**日本が何を言っても GAFA は聞く耳を持たない**ということです。EU との共同戦線を張るのが日本の現実解です。EU は大陸法なので法体系的にも近い。コモンローである英米とはやはり違います。

谷脇　サイバー空間には国境がありませんから、規律の国際的整合性が必要ですが、これをどのように確保していくべきでしょうか。AI に関する国際機関を設立すべきという意見もありますが有効でしょうか。

佐藤　国際機関を作って機能するかどうか、私にはよく分かりません。国際連合だって十分機能しているとは言い難い状況において、新たな機関を作ることにどれほどの意味があるのかという懸念はあります。仮に巨大プラットフォーマーにとって「規律を

守らないこと」に利益があるのなら、国際機関を作って規律を定めても弱い抑止力程度にしかならないでしょう。グローバルにビジネスを展開する事業者の場合、国ごとの個別対応はコストが高くなるので、より厳しい国の規律に合わせるはずです。国際機関が運用する国際的規律にはそれ以上の厳しさを持たせないと形骸化します。現時点では、国際機関を作る意味を見出すことは難しいと思います。

人間が AI に介入する必要はある

谷脇　負の側面について、もう一つお聞きします。**データ駆動社会**（Data Driven Society）において AI が実装されることで、個別化・自動化・最適化が進みます。これは経済学的には効率的な資源配分に貢献すると考えられますが、他方、差別の助長、少数意見の切り捨て、説明責任の欠如といった問題が顕在化するのではないでしょうか。

佐藤　その通りだと思います。AI に限らず、データに基づくシステム全般に言えることですが、全体データを正確に反映した判断が社会的に適切とは限りません。AI の場合、過去のデータを学習している以上、AI による判断は過去において多数・優勢だった対象に有利に働き、**現状の社会課題を固定・拡張**する可能性が高いのです。

例えば与信評価をデータに基づいて行うと、男性の方が女性よりもスコアが良くなります。それは総じて男性の方が女性よりも高収入という現実があるからです。その与信評価に従い男性を優先して融資すると、与信評価はますます男性有利になっていきます。男女は平等であるべきだという理念に基づいてそうした性差を減らそうとすれば、アルゴリズムを恣意的に改変するか、AI に学習させるデータを歪めて女性の評価を上げるような操作を加えなければなりません。

操作を加えてデータを歪めたうえでの出力が公平なのか、操作をしないで現実のデータに基づく出力が公平なのか――。**学習データに意図的に手を加えられた AI による出力は公平中立で信頼できるものと言えるか**という新たな問題が生じます。もはや技術だけの問題ではなく、社会としてどういう選択をするかという議論になります。

また、学習データにある程度以上の「量」がないと、AI には判断ができなかったり、誤差が多くなったりします。これが、少数データを足切りするような方向に働けば、人間社会における少数意見や少数派の立場を侵害する、あるいは、多数意見や多数派の立場を有利にした結果として相対的に少数派の立場を侵害するようなケースが発生し得ます。

AI が学習するのは過去のデータであり過去の状況を固定化するという特性がある以上、**人間が恣意的に介入する必要は「ある」**と思います。そうした介入をどこまで許すのか、それをどのようにガバナンスするのかが問われることになります。

人間が AI に順応するような社会にしないために

谷脇　デジタル政策フォーラム（DPFJ）では、まさに AI ガバナンス、データガバナンス、セキュリティガバナンスの 3 つの要素で構成される「デジタルガバナンス」について検討を深めていこうとしています。そもそもこれからのデジタル技術を人類は制御することができるのかという問題意識をベースにして、真剣な議論を誘発したいと考えています。留意しておくべきことについてアドバイスいただけますか。

佐藤　2 点あります。

　第一点は、**AI は完成された技術ではない**という認識を共有した上で議論すべきということです。

　現在の生成 AI で使われている深層学習、ディープラーニングといったものがなぜうまくいっているのかといったことは、実は解明されていません。理論的なモデルも未完成なので、生成 AI に何ができて、何ができないのかも分からないし、予測もできない。AI 技術者も AI を制御できているわけではないのです。ですから、技術的に何が分かっていて何が分からないのか、どこまで制御できてどこから先は制御できないのかを整理した上で、その認識を共有し、その先に想定される問題と対応について早めに議論しておくのが良いと思います。

　第二点は、**AI は規制対象であると同時に、規制の実現手段にもなり得る**ということです。

　生成 AI によって、SNS やネット掲示サイトなどにおける誹謗中傷や差別的表現を、広範囲かつ高精度で監視することもできるでしょう。このため、AI の支援を前提とした現実世界またはサイバー世界に対する規制が行われる可能性があります。それは同時に盗聴や検閲の手段にもなり得ます。

　フランスの哲学者ミシェル・フーコー（Michel Foucault）は『監獄の誕生』で、パノプティコン（人々を監視するために最適な建築構造）の中にいる人々は、常に監視されていることを意識し、規律化され、従順化すると指摘しました。現状は AI の処理能力の限界があって全市民を監視できないとしても、「AI に監視されているかもしれない」という意識が広がることによって市民が無意識のうちに AI に順応するような社会が形成されてしまうかもしれません。

　AI の利便性を享受しつつ、国民の権利・便益の侵害が起きないことを担保しなければいけません。AI を規制や法執行の実現手段として使うときに、何が許されて何が許されてないのか、利用範囲と方法について議論すべき時期が来ていると思います。そういったことは規制する側からはなかなか言いだしにくいところでもあるので、デジタル政策フォーラムのような独立シンクタンクから問題提起していただくことはとても有意義だと思います。

谷脇　良い気づきをいただきました。デジタル政策フォーラムとしても建設的な議論をしていきたいと思います。ありがとうございました。

”

『デジタル政策の論点 2024』（2024 年 7 月刊）

特集 2　AI ガバナンスのあり方から転載

【参考資料②】
～企業へのヒアリングを終えて～

企業ヒアリング

デジタル政策フォーラムでは、「本編付属資料」としての有識者インタビューと並行して、AI ガバナンスの枠組みの構築に向けて(ver2.0)の策定にむけ、利活用の現状を把握するために、関連する企業へのインタビューを実施した。コンサルタント、ベンチャーキャピタル、シンクタンク、金融、医療、広告、デジタルなどを領域とする十数社の企業から、多様な意見をヒアリングすることができた。内容は多岐に渡り、広範で示唆に富むものであった。以下の構成は、考察、傾向、主なコメントからなる。 なお、インタビューは 2024 年 8 月から 10 月にかけて実施した。

(1) AI 利活用の傾向と主な意見

- 大企業、中小企業といった規模の違いによって、AI の導入度合いに大きな差異はあったが、導入されている事例は PoC（Proof of Concept：概念実証）にとどまるか、いわゆる「サンドボックス」にとどまることが多い。企業の規模に関わらず、現時点での AI 導入は試験的な段階にとどまることが多い。
- 業務の効率化を目的とする AI の導入や運用がほとんどであり、効果が予測しやすく、導入が容易で、社内外で説明しやすい事業領域から着手している傾向が見られた。また、決裁が容易な範囲で、部門や組織を単位とする導入が多い。既存のオペレーションを大幅に変更する AI 導入は、インタビューでは確認することができなかった。
- 導入された AI 利活用事例の中には、ワークフローに定着しているものが一部であったものの、その多くは期待した効果が得られず、あるいは費用対効果に照らして、事業への本格的な導入が見送られたものであった。
- 企業等の AI 利活用は、現時点では経営改革レベルの全社的な取組ではない。バックオフィス業務における AI の広範な導入など、法人全体に影響や効果が及ぶ類の AI 導入に関しては、インタビューでは確認することができなかった。また、DX を全社的に推進するために、AI を導入しようとするケースも確認できなかった。「AI で重要なのは効率化ではなくビジネスフロー改革だが、現実の導入に当たっては、そこまで至っていない」との発言は、AI 導入の現状を端的に物語っているだろう。
- 「人員削減等による経営効率化を実現できなければ、AI 導入の積極的な意味がない」とする意見も聞かれたものの、アメリカの一部事例などとは異なり、AI チャットボットや RAG の導入より人員を大幅に削減した日本の事例は聞くことができなかった。
- 中小企業では AI の導入・運用を外部に依存している様子や、外部からの助言により導入規模などを決定する様子が確認された。
- 大企業では社内リソースを活用することにより、AI 利活用に関する意思決定を行っている様子が見て取れた。大企業には、情報システム部門やセキュリティマネジメント部門の知見が蓄積されており、これらの知見を活かして AI 対応部門を創設した事例もあった。
- AI 導入リスクの発見や判断に関しても、大企業では情報セキュリティや個人データ保護等の知見が役立っていることが確認された。また、情報システム部門の運用経験が、AI 導入に関する意思決定を迅速にしている傾向が見られた。
- 中小企業に関するインタビューでは、「過度のセキュリティ意識が AI の導入を妨げている」との発言も聞かれた。同時に「十分にリスクの認識や判断をせずに、あ

るいは受託者が責任を負う形で、AI 導入に踏み切る例があった」との発言もあった。

- 責任とリスクに関する外部基準の存在は AI の導入を促進すると考えられるが、特に中小企業ではハードローによる規制に反対する意見が多かった。他方、ガイドラインや整合規格による規律に対しては反対が少なかった。法規制に対する合意形成に時間を要するのであれば、ガイドラインによる規律が先行することになろう。
- 規制の具体的なイメージを持ってはいないものの、許認可事業においてはハードローによる規制が望ましいとの意見が聞かれた。
- 事業規模や業種の別に関わらず、あらゆる企業が AI を利活用できる「領域」を潜在的に有しており、AI 導入が企業の成長力を高めるとの指摘もあった。
- トラスト、競争資源、責任軽減、判断支援など、AI 導入にあたって重要となる要素やインセンティブは数多く考えられるが、規制や規律も普及への推進力になるとの発言もあった。
- だれもが AI に触れることができる機会の提供や、専門人材の育成なども AI の普及に寄与するものであり、重要な政策課題であるとの指摘もあった。
- 今回のインタビューでは、内容や事例は実に多岐に渡っていた。生成 AI の導入・普及は緒についたばかりであり、利用の態様は収束する段階にない。

(2) インタビューに基づく分析

- AI の導入、利活用は規模、職種により多種多様。AI リスクに関する情報収集は規模に相関。AI リスクに関する対応は、個人情報保護、情報セキュリティに関する全社的対応と相関。
- 導入や社内環境整備は大企業の方が中小企業よりも現時点では積極的。大企業では外部企業との連携により AI 導入を進める場合が多数。
- 画像認識などで AI を利用していた事業では、生成 AI も早期から導入される傾向。DX が進展している部門ではインハウスでの開発も多数。
- 製造ラインなどでの AI 導入と比べて、顧客サービスでの導入は限定的。バックオフィス系の業務でも AI 導入は限定的。クリエイティブ領域で生成 AI を利用している場合では、敢えて利活用を公表しない事例も多数。
- ガイドラインの制定とその遵守については企業規模によらず、概ね肯定的。影響力が大きい大規模開発への規制については肯定的な意見が多数。
- 許認可事業では、法律による AI 規制と業界ごとのガイドライン策定に肯定的。反対に AI 開発スタートアップでは法規制による萎縮を懸念。
- EU が制度化したリスクベースアプローチについては、肯定的意見と否定的意見が混在。ただし、海外の規制動向に対する関心は僅か。

・ AI によるビジネスモデル変化は見通せないとの声が多数。AI による労働代替への懸念は少数。 人材育成は、公的主体に期待しつつも、社内で行うことを重視する傾向。

(3) インタビューにおける主な発言（※上述との重複あり。）

① AI 及び生成 AI の事業活用状況について

・ 大企業であっても全社的に LLM や生成 AI を使っている企業は僅かだろう。従業員のリテラシーに関わらず、利用可能な社内システムを構築しないと全社的利用は進まない。
・ RAG に社内情報を入力してアウトプットを出力する事例は少なくないが、このような事例でも PoC に止まるものが多い。PoC で終わるものの多くは、リスクとベネフィットを勘案した結果だ。
・ 医療分野などにおいて、オンプレ環境でのシステム運用を課されている企業は当然ながらクラウド環境で AI を利活用することが難しい。この場合、AI を利活用する PoC ですら、準備に時間を要することになる。
・ 「生成 AI で業務を改革する」というような、全社的 DX での AI 利活用はほとんど耳にしたことがない。現時点では、従業員単位で AI を業務に利活用する事例が多いのではないか。開発に関連する部署で AI の利活用が進んでいる企業であっても、それ以外の部署では全く業務に利活用されていないことがある。B2C での AI 導入は、部門の単位で判断される傾向がある。
・ 先進的な企業では、複数の外部 AI サービスを社員が自由に選択できる環境を整えている。また、B2B で社外に AI サービスを提供している部門が、社内向けにもサービスを提供している事例がある。

② 具体的な AI 活用の内容について

・ 全社的に LLM や生成 AI を使用している企業は少ないが、RAG に社内情報を入力してアウトプットを活用したり、社内用チャットボット（技術資料の検索等も含む）などを利用したりすることは一部で定着している。
・ 海外市場や政府規制など、特定情報の収集における AI 利活用は広がっている。顧客サービスを視野に説明書や Q&A 事例を RAG に入力するケースも広がっている。
・ 法令や約款を AI に学習させて、各種の問い合わせに即答できるようにした事例など、社内法務への部分的な利活用も見られる。大部なマニュアルを AI に学習させて、業務を効率化、自動化する試みも散見される。
・ 不正チェックなどこれまで社内で専門家が行っていた業務を AI に学習させて、移管しようとする試みもある。不良品検査などで AI の画像認識を活用する事例や、コンテンツ制作過程で自動的に色付けをする事例なども広がりつつある。

・ 生成 AI だけでなく従来の AI を対象とすれば、AI 利活用が既に一般化している領域はある。検査業務や異常予知業務などでは広く AI が利活用されており、自動化による人員削減に部分的に結びついている事例も現れてきた。
・ 従業員が「調べもの」で AI を利活用することは普及してきた。画像生成、メール要約、翻訳、資料作成、調査での一次情報収集が、現時点では一般的な企業における生成 AI の活用例ではないか。

③ AI リスクに関する情報収集について

・ セキュリティポリシーを設けていても、AI リスクに関する情報収集をベンダーに任せている企業、団体は少なくない。
・ 社内でのシステマティックな情報共有よりも、部門ごとにベンダーなどからの情報提供へ依存している企業もある。
・ 関係省庁が公表しているガイドラインに基づき、社内ガイドラインを制定している大企業もある。それらの社内ガイドラインでは、公的なガイドラインが掲げるリスクに則り、規定が整備されている。
・ 法制度の変更や他社の取組みについては、関係省庁との協議を通じて情報収集を行っている大企業もある。

④ AI リスクの内容について

・ リスクとして一般的に検討されるものの中では誤情報に関するもの、ハルシネーションに関するものが多いだろう。リスクやインシデントを具体的に検討しないまま、利活用している企業も少なくない。
・ 取扱う情報が膨大な大企業はインプットする情報を精査する傾向にある。また、社内情報のプロンプト入力を一般に禁止、制限する企業も存在する。
・ リスクを理由に生成 AI の利活用を一般的に制限している企業がテック系ですら存在する。そのような企業では、個人で AI を利活用しているのが実態だ。
・ AI リスクに関して全社受講の社内 AI 研修を用意している企業もある。社内向けチェックリストの項目に沿ってリスクを部門単位で管理している企業もある。

⑤ AI ガバナンスの規制について

・ 法制度や公表されているガイドラインと照らし合わせて、AI 規制を理解している者はごくごく僅かだろう。
・ EU の AI 法のように、サービスの品質を担保するような制度は AI の利用実態にはそぐわないのではないか。
・ 社内ルールは公的なルールよりも厳しくなる傾向があり、法規制は利活用にマイナスの要因となることに留意する必要がある。規制は最低限であるべきだ。
・ 社内データであれば自由に利活用できている環境が、今後、法で規制されてしまう状況に変化するなら、開発スピードは確実に遅くなる。システム開発に責任を負

うベンダーなどが、規制によって煩雑な作業や手続に直面することは避けるべきだ。

- 例えば入力プロンプトが学習されないとサービス提供企業の規約に記載されていたとしても疑念が残るため、ハードローであるかソフトローであるかは別にして、何らかの社会的なルールでサービス提供企業に責任を課すべきだ。
- AI 規制は必要であり、大企業の立場としては具体的な規律を示してもらう方が運用しやすいが、スタートアップ等によるイノベーションを阻害しない方策はあわせて必要だろう。
- 入力するデータの機密性に鑑み、外部 AI サービスの利用が禁止されている事例は少なくない。ルールがないと、リスクテイクをしない企業、団体、ユーザーは存在するので、少なくともガイドラインはあった方が望ましい。
- 政省令に詳細な規制を規定して、技術の変化や進歩に対応できないことは避けなければならない。他方、守るべきガイドラインなら少なくとも年に一回は改正が必要になるのではないか。
- AI が経済や社会に与えるインパクトを考慮すれば、一定の規制は妥当で自然だ。EU が GDPR を法定して、日本企業であっても対応が必要になった状況と同じだと認識している。

⑥ 国・地域・団体又は自社独自ガイドラインの検討について

- サステナビリティレポートに AI 倫理を加えたり、AI 倫理方針を制定して周知したりする企業もある。また、提供、開発に分けて AI リスクをリスト化している企業もあり、AI に関する社内規定等の整備は企業間で差異が拡大している。
- 社内ガイドラインが同程度の内容であっても、厳密に運用している企業と緩やかに運用している企業とがあり、ガイドラインの文言に止まらず、ガイドラインの運用状況を確認することが必要だ。
- 公的なガイドラインと自社のポリシーがあれば、業界を単位とするガイドラインは必要ない。経済団体を単位にするなど、より大きな枠組みの方が対象となる企業数が多く、ガイドラインの実効性を担保できるのではないだろうか。
- 規制事業では業界ガイドラインは法律に準じるものとの認識されている。業界ガイドラインを定めるニーズはある。業界ごとに差異があるのは当然であり、必要に応じて業界ガイドラインを設けることは自然だ。