資料２

AI制度研究会
構成員提出資料

大阪大学

工藤構成員
提出資料

第5回 AI制度研究会

「政府によるAIの適正な調達と利用」に関する政策動向

大阪大学 社会技術共創研究センター 特任准教授
工藤郁子

2024年12月26日

大阪大学 社会技術共創研究センター
Research Center on Ethical, Legal and Social Issues

Osaka University
Research Center on
Ethical, Legal and
Social Issues

AI制度の在り方に関する4原則

本日の皆さんの御意見も踏まえ、制度の在り方を議論するに当たって、次の4点が基本原則だと考えています。
1つ目は、リスク対応とイノベーション促進の両立です。ガイドラインをベースとしつつ、リスクの大きさに応じて対策を講じ、AIの安全性を確保する必要があります。
2点目は、技術・ビジネスの変化の速さに対応できる柔軟な制度の設計です。
3点目は、国際的な相互運用性、国際的な指針への準拠です。
4点目は、**政府によるAIの適正な調達と利用**です。
政府の取組は、他への波及効果も大きいので、しっかりと検討を進めていきたいと思っています。

AI戦略会議・AI制度研究会合同会議
(2024年8月2日)
https://www.kantei.go.jp/jp/101_kishida/actions/202408/02ai.html

なぜ政府による調達？

- 公共調達（public procurement）はOECD諸国GDPの10-15%程度を占め、経済上のインパクトが大きい
  - 中央政府だけでなく、地方自治体や公益法人などを含む
  - ITシステム等の開発・運用に費やされる金額は増加傾向にある
- 政策推進の手段として公共調達を活用する「戦略的調達」という国際潮流がある
  - cf. 国連SDGs「12.7 持続可能な公共調達（SPP: Sustainable Public Procurement）」
  - 環境保護以外にも、ジェンダー平等、中小企業支援、イノベーション促進などの政策目的（Appendix 1 参照）
- AIに関しても、イノベーションを促進しつつ、責任ある利用を　目指す試みが、各国で進んでいる
  - 政府の策定する提案依頼書（RFP: Request for Proposal）や評価制度（参考例：政府情報システムのためのセキュリティ評価制度 [ISMAP] ）が、ベンチマークやソフトローとして機能することへの期待も

公共調達の適正性：生成AI「以前・以後」

- 生成AI「以前」：分析AIを念頭に、プロファイリングの権利侵害リスクに焦点
  - ユースケース（例）：法執行目的での公共空間におけるリアルタイム遠隔生体識別、移民審査業務の効率化、保育所入所選考業務の効率化、プッシュ型行政サービスの提供など
  - リスク（例）：プライバシー侵害や監視社会化、判断過程の不透明性や答責性低下、バイアスによる差別的決定など
- 生成AI「以後」：生成AIを念頭に、偽・誤情報やセキュリティに関するリスクに焦点
  - ユースケース（例）：チャットボットによる個別化された自動応答、自動翻訳によるアクセシビリティ改善、業務に必要なソフトウエアのコード生成など
  - リスク（例）：医療・公衆衛生等の分野での誤情報提供、プロンプト・インジェクション等による機密情報漏洩、個人データ漏洩によるプライバシー侵害など
- 上記に加えて、適時適切なAI導入ができないリスク、国外サービスの過度な依存に伴う経済安全保障リスク、特定企業との癒着・政治的腐敗に関するリスクなども
  - cf. パンデミックにおける「デジタル敗戦」、国際紛争によるサプライチェーンの寸断など

生成AI「以前」の政策動向

- 2019年4月、カナダ政府は行政機関におけるAI利用について「自動化された意思決定に関する指令（Directive on Automated Decision-Making）」を施行
  - カナダ政府が試行していた移民審査システムに対して、2018年秋、トロント大学の研究チームが懸念を表明
  - 4段階のインパクト・レベルに分類するリスク・ベース・アプローチを採用し、アルゴリズム影響評価（Algorithmic Impact Assessment）を実施（Appendix 2 参照）
- 2020年6月、世界経済フォーラムは「AI公共調達ガイドライン」、政府向けの「ワークブック」、英国政府と実施した「パイロットケース実施報告書」を公開
  - ブラジル、UAE、米国、**日本（デジタル庁）**など各国政府とワークショップを実施（Appendix 3 Appendix 4 参照）
- 2021年11月、英国政府は行政機関が利用するAIを対象に「アルゴリズム透明性記録標準（ATRS: Algorithmic Transparency Recording Standard）」を策定
  - ATRSでは、概要情報（一般市民向け）と技術情報（専門家向け）の2階層で、記録・開示する情報を整理（Appendix 5 参照）

生成AI「以後」の政策動向

- 2023年5月、**日本政府**の関連省庁が**「ChatGPT等の生成AIの業務利用に関する申合せ」**を実施
- 2023年8月、**東京都**が都職員向けの**「文章生成AI利活用ガイドライン」**を策定
- 2023年9月、カナダ政府が「連邦政府職員向け生成AI利用ガイド」を公表（ Appendix 6 参照）
- 2023年10月、米国大統領令において、政府によるAIの責任ある効果的な利用の保証を表明（ Appendix 7 参照）
- 2024年3月、米国カリフォルニア州政府が、 州政府機関による生成AIプロダクト調達のためのガイドライン等を公表（ Appendix 8 参照）
- 2024年3月、**神戸市**が神戸市職員のAI利用に関する**「AI条例」**公布
- 2024年4月、米国行政管理予算局が、連邦政府機関のAI利用指針を公表（ Appendix 9 参照）
- 2024年5月に成立したEU AI法において、「AI Officeが講じる措置」の一つとして、AIシステムに関連する公共調達手続のベストプラクティス促進等が記載
- 2024年10月、 G7デジタル・技術大臣会合において「公的部門におけるAIツールキット（Toolkit on AI in the Public Sector）」が議論（ Appendix 10 参照）

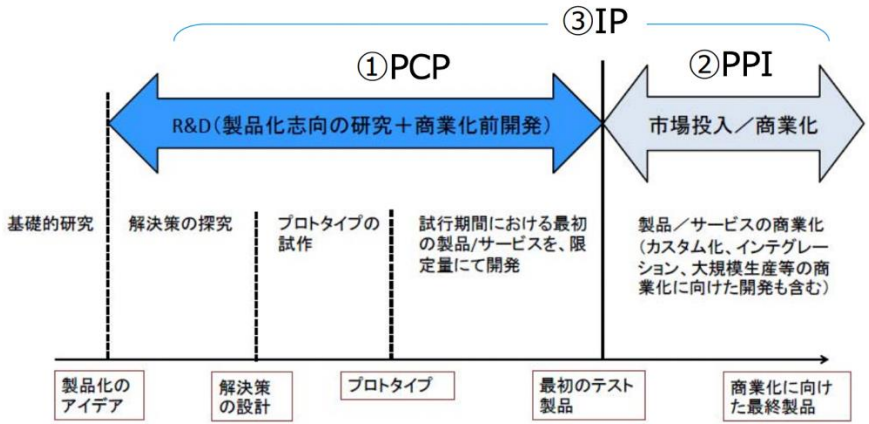
まとめ

- **政府によるAIの効果的利用**によって、業務効率の向上や行政サービスの応答性強化などが実現することが期待されている
- 一方、プロファイリングによる権利侵害リスク、偽・誤情報やセキュリティに関するリスク、国外サービスの過度な依存に伴う経済安全保障リスクなども指摘されており、**政府によるAIの責任ある利用**が求められている
- 諸外国では、政府によるAIの調達・利用に関する、**指針、ガイドライン、ロードマップ、職員向けの訓練プログラム等**が策定されており、それらは国家全体のAI戦略との整合性が図られていることが多い
- 日本は、国際的にみても比較的早期にセキュリティリスク等への対応を進めてきたと言えるが、今後、**具体的なユースケース**を念頭においた責任ある利用ルールの策定や、**インシデント対応**に係るルール整備が必要であると考える
- さらに、（AIに限らず）イノベーションを促進するための**戦略的調達のグランドデザイン**も必要ではないか

Appendix

Appendix 1 イノベーション促進目的の戦略的調達

- 諸外国においては（AIに限らず）イノベーション促進を目的とした**戦略的調達**のグランドデザインが行われている
  - 欧州（EU、英国、ドイツ）、米国、韓国では、イノベーションのライフサイクルを意識しつつ、法令・規則レベルから具体的な運用システムまで、多様なスキームが整備されている
  - 他方、日本では、会計法において調達手続きを規定しているが、諸外国のように広くイノベーションを促進することを目的とした規定・制度は設けられていない（ガイドラインに留まる）
- グランドデザインに加えて、報告・監査・統制等も担保する包括的制度が求められる



① PCP（pre-commercial procurements）：政府等公的機関においてニーズはあるが、市場に存在しないモノ・サービス等を「研究開発を経て試作品レベルのモノ・サービスとして公共調達する」

② PPI （public procurements of innovative solutions）：政府等公的機関にニー ズがあり、ある程度まで製品レベルに移行したモノ・サービス等の初期需要として「量産化に向けた一定規模を公共調達する」

③ IP（innovation partnerships）： ①と②を組み合わせて公共調達する

(via https://www.jst.go.jp/crds/report/TP20230601.html)

Appendix 2 カナダ・アルゴリズム影響評価

- 2018年、カナダ政府は、移民申請書類の処理を効率化するためのAIシステムの試行を開始
  - これに対して、不当な差別やプライバシー侵害などの問題点を指摘する報告書をトロント大学の研究者チームが公表
- 2019年、行政機関におけるAI利用について「自動化された意思決定に関する指令（Directive on Automated Decision-Making）」が発効
- 同指令において、政府職員の判断を支援・代替する取組みにつき、アルゴリズム影響評価（AIA: Algorithmic Impact Assessment）の実施が定められる
  - **リスクベース・アプローチ**：AIAでは、影響度を4段階に区分し、影響の大きさに応じて必要なリスク軽減対策を示す
  - **AIAツールの提供**：担当部署がアンケートに回答するとスコアが算出され、AIAの4段階のどこに該当するかがわかる仕組み
- AIAの実施結果は、カナダ政府のポータルサイトで公表

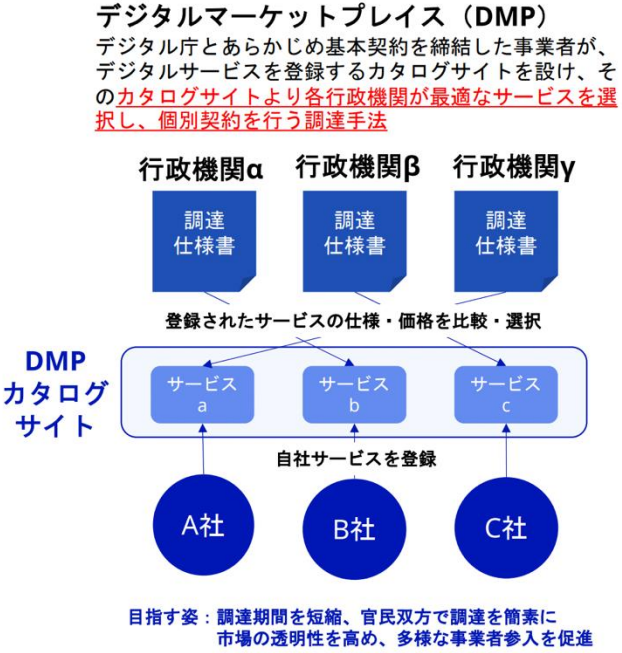
(via https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html)

Appendix 3 WEF・パイロットケース

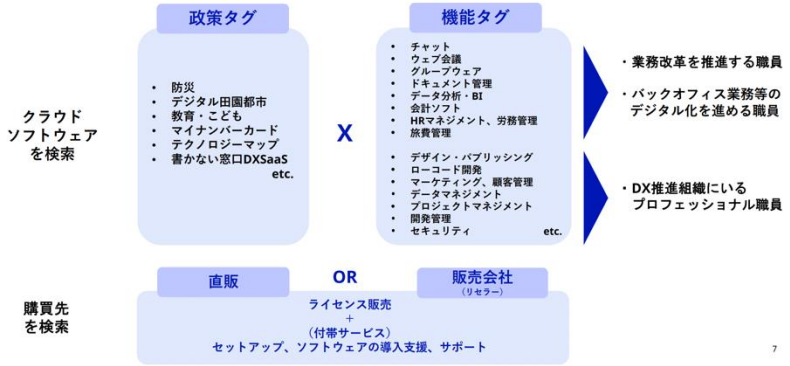
- 2020年、世界経済フォーラム（WEF）は、「**AI公共調達ガイドライン**」、政府向け「**ワークブック**」、英国政府と実施した「**パイロットケース実施報告書**」を公開
- 英国以外にも、ブラジル、UAE、米国、日本など各国政府とワークショップを実施し、利害関係者のコミュニケーション促進と責任ある技術利用への貢献を目指す

(via https://www.weforum.org/reports/ai-procurement-in-a-box)

Appendix 4 日本・デジタル庁の取組み



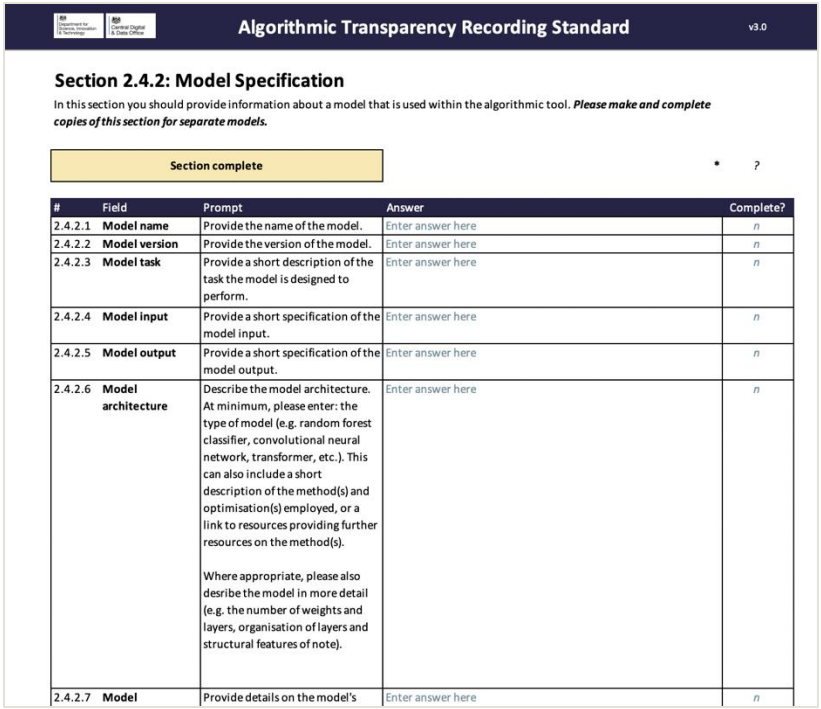
- 2022年6月、デジタル庁「情報システム調達改革検討会」
  - （AIの前に）SaaSなどについて**透明化・適正化・迅速化・多様化**を図る
  - 英国では、複雑な入札手続などの**参入障壁が解消**され、より多くのスタートアップ企業が公共調達に参加
- 2023年11月、「デジタルマーケットプレイス α版」が公開



(via https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/03735227-d301-4bec-a678-96e036d917ea/cdf02cda/20230920_meeting_administrative_research_working_group_outline_02.pdf)

Appendix 5 英国・アルゴリズム透明性記録標準

- 2021年、英国政府は、「アルゴリズム透明性記録標準（ATRS: Algorithmic Transparency Recording Standard）」を策定
- 第1階層で**一般市民向けの概要**、第2階層で**専門家向けの技術情報**を整理
  - 第1階層：アルゴリズムの名称、利用する理由と方法、詳細情報の入手方法
  - 第2階層：責任者情報、アルゴリズムの機能性・合理性、意思決定への活用方法、技術要件と利用データ、想定リスク・リスク軽減策、インパクト評価
- **雛形の提供**：担当部署がテンプレートシートに記入していくと、記録すべき事項が網羅できる
- ATRSの実施は任意だが、その結果は政府のポータルサイトで公表



Algorithmic Transparency Recording Standard v3.0

Section 2.4.2: Model Specification

In this section you should provide information about a model that is used within the algorithmic tool. ***Please make and complete copies of this section for separate models.***

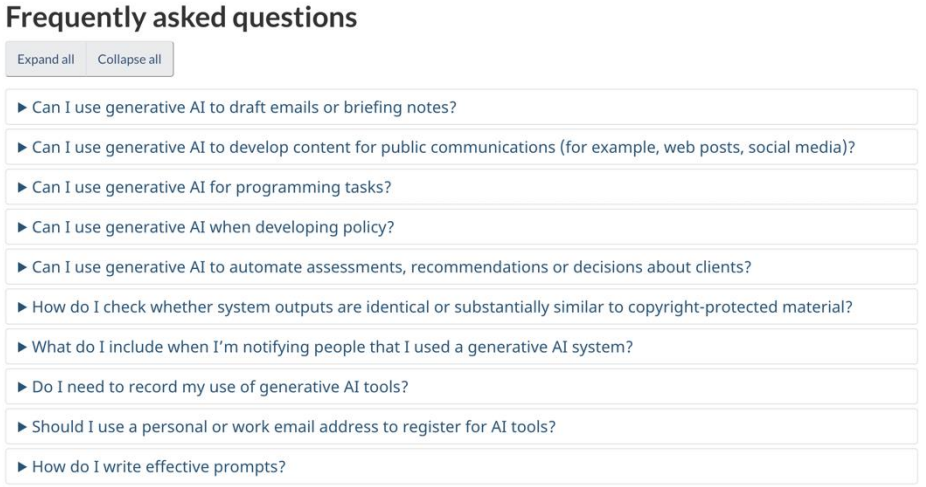
Section complete

#	Field	Prompt	Answer	Complete?
2.4.2.1	**Model name**	Provide the name of the model.	Enter answer here	n
2.4.2.2	**Model version**	Provide the version of the model.	Enter answer here	n
2.4.2.3	**Model task**	Provide a short description of the task the model is designed to perform.	Enter answer here	n
2.4.2.4	**Model input**	Provide a short specification of the model input.	Enter answer here	n
2.4.2.5	**Model output**	Provide a short specification of the model output.	Enter answer here	n
2.4.2.6	**Model architecture**	Describe the model architecture. At minimum, please enter: the type of model (e.g. random forest classifier, convolutional neural network, transformer, etc.). This can also include a short description of the method(s) and optimisation(s) employed, or a link to resources providing further resources on the method(s). Where appropriate, please also desribe the model in more detail (e.g. the number of weights and layers, organisation of layers and structural features of note).	Enter answer here	n
2.4.2.7	**Model**	Provide details on the model's	Enter answer here	n

(via https://www.gov.uk/government/publications/algorithmic-transparency-template)

Appendix 6 カナダ・生成AI利用ガイド

- 2023年、カナダ政府は、「連邦政府職員向け生成AI利用ガイド」を公表
- 「潜在的課題とベストプラクティス」の項目では、情報の保護、バイアス、品質、公務員の自律性、リーガルリスク、人間と機械の区別と並んで、**環境への影響**に関する記載も
  - 生成AIシステムの開発・利用に必要となる莫大な環境コストを課題として指摘
  - ベストプラクティス等として、ネットゼロまたはカーボンニュートラルなデータセンターでホストされる生成AIツールの使用や、AIサプライヤーが温室効果ガス削減目標を設定しているか確認することなどを例示
- **FAQ（よくある質問とその回答）**なども整備することで、政府職員の理解度を高める



(via https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html)

Appendix 7 米国・大統領令

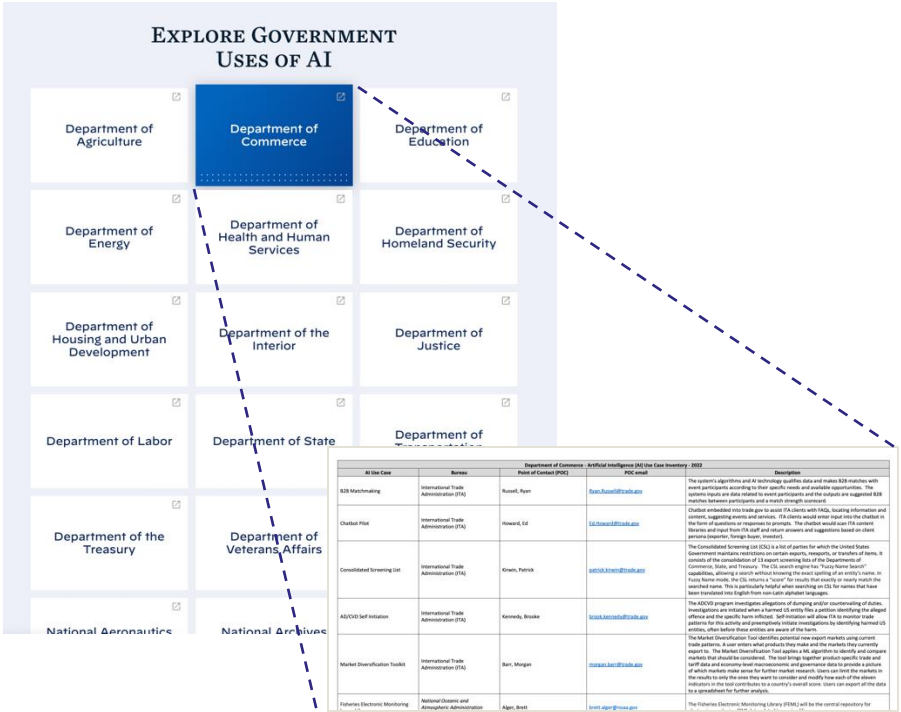
- 2023年、 米国のバイデン政権は「AIの安心安全で信頼できる開発と利用に関する大統領令」を発令
- 本大統領令の主要構成要素のうち8項目が、「政府によるAIの責任ある効果的な利用の保証」
  - 政府全体でAI専門家の迅速な採用を加速するとともに、権利と安全を保護するための明確な基準や各省庁がAIを利用する際の明確なガイダンスを策定するとした
  - 2024年3月には、米国行政管理予算局が、本大統領令を受けて、連邦政府機関のAI利用指針を公表（Appendix 9 参照）
- 公共調達は、**AIとCBRN**（Chemical [化学]・Biological [生物]・Radiological [放射性物質]・Nuclear [核]を用いた兵器等）のリスクとも関係するとの認識が示されている
  - なお、 2024年10月には「AIの開発や利用に関する国家安全保障覚書（AI NSM: Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence）」が発表され、 民主的価値観に沿った方法で国家安全保障の任務において先端AI技術を活用することなど、連邦政府機関によるAIの開発や利用に関する指針が策定されている
- 米国の**政権交代**に伴い、本大統領令が今後どのように取り扱われるか、注視が必要

Appendix 8 米国カリフォルニア州・ガイドライン

- 2024年、米国カリフォルニア州政府が、 州政府機関による生成AIプロダクト調達のためのガイドラインを公表
  - 生成AIのリスク評価と管理、生成AI製品の調達手続きなどについて記載
- **「意図的な調達」**だけでなく**「偶発的な調達」**もありうることを指摘
  - 意図的な調達（Intentional GenAI purchase）：生成AI利用が必要な場合に、特定の生成AIプロダクトを調達する場合
  - 偶発的な調達（Incidental GenAI purchase）：調達したものの一部に生成AI利用が含まれていた場合、提供されるサービスに生成AIが補助的に使用されている場合など
- ガイドラインに加えて、州政府職員のための**訓練プログラム**、リスク評価ツールキットも提供開始
  - 生成AIプロダクトの継続的なモニタリングと評価を担当する職員を指名することとし、当該職員は訓練プログラムを受講する必要がある

Appendix 9 米国・連邦政府機関のAI利用指針

- 2024年、米国行政管理予算局（OMB: Office of Management and Budget）が、2023年の大統領令を受けて、連邦政府機関のAI利用指針を公表
- リスク管理、透明性向上、イノベーション推進、人材拡充、 ガバナンス強化などについて記載
  - リスク管理について、連邦政府機関が、権利・安全に影響を与える可能性のあるAIを使用する場合、2024年12月1日までに、アルゴリズムによる差別を防ぐなどの具体的なセーフガード措置を講じることを義務付け
  - 透明性について、米国民の権利や安全に影響を与えるような**連邦政府AIユースケース一覧表（Federal AI Use Case Inventory）**を年次で公表



(via https://ai.gov/ai-use-cases/)

Appendix 10 G7・公的部門におけるAIツールキット

- 2024年、**G7デジタル・技術大臣会合**において「公的部門におけるAIツールキット（Toolkit on AI in the Public Sector）」が議論
  - 経済協力開発機構（OECD）と国連教育科学文化機関（UNESCO）が、G7に向けて作成し、G7デジタル・技術大臣会合共同声明で「歓迎」された文書
  - 公的部門で活用されるAIについて、G7各国における導入の傾向、ベストプラクティス、政策フレームワーク、AI活用を促進するための様々な施策などを記載
- G7各国政府のAIに関する取組みを比較・概観することができる

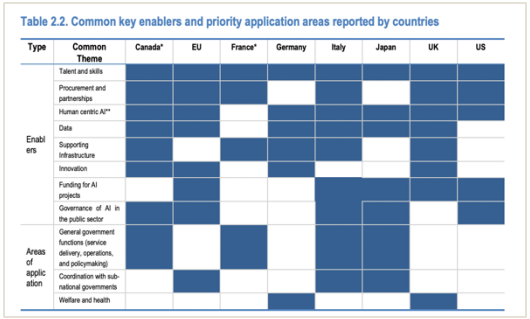
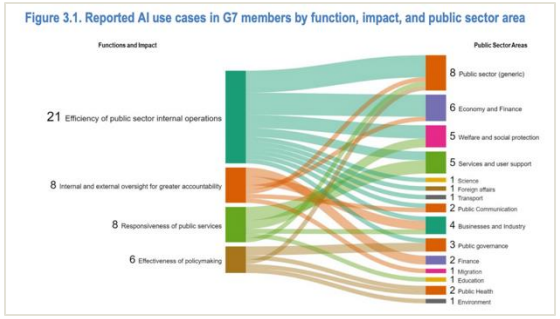


Table 2.2. Common key enablers and priority application areas reported by countries

Type	Common Theme	Canada*	EU	France*	Germany	Italy	Japan	UK	US
Enablers	Talent and skills	■	■	■	■	■	■	■	■
	Procurement and partnerships	■	■	■		■		■	■
	Human centric AI**	■	■		■	■	■	■	■
	Data	■	■		■	■	■	■	
	Supporting Infrastructure	■		■	■	■		■	
	Innovation	■	■		■			■	
Areas of application	Funding for AI projects		■			■	■	■	■
	Governance of AI in the public sector	■	■			■	■		■
	General government functions (service delivery, operations, and policymaking)	■		■		■	■		
	Coordination with sub-national governments		■			■	■		
	Welfare and health				■			■	



(via https://www.digital.go.jp/news/53ed2e40-a8be-4249-869d-e94f4f9a28fa)

Keidanren
Policy & Action

永沼構成員
提出資料

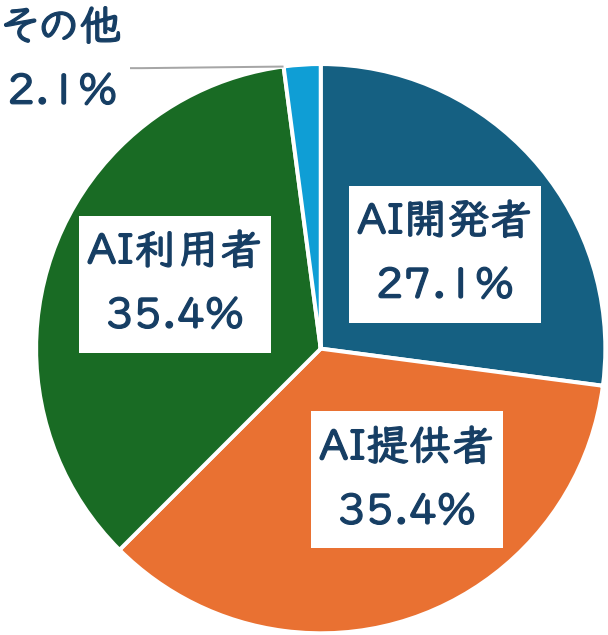
AI制度に関する基本的考え方

2024年12月26日

経団連デジタルエコノミー推進委員会
国際戦略WG主査 永沼美保

Ⅰ. 基本情報

- 経団連では本年8月30日～9月11日、関係会員を対象に、AI制度に関する簡易アンケートを実施（特に第1回AI制度研究会（8/2）において、岸田総理から指示のあった4つの「基本原則」を中心に意見照会）
- 本資料は、寄せられた回答を事務局にて整理したもの（必ずしも経団連として機関決定した意見に当たらないことに留意）
- 回答者の構成は以下の通り（重複回答有）



※その他＝調査研究

2. 設問と主な回答（1）

総論：ハードロー・ソフトローに関する基本的考え方

- 既存の法制度とソフトローを中心とすべき。ハードローは、現行法でカバーできないリスクに限定し、ガイドライン等のソフトローと組み合わせたハイブリッド型の規制が適切
- ハードローを適用すべきリスクの高い分野・機能等を明確化すべき

各論：4つの「基本原則」に関する基本的考え方

(1) リスク対応とイノベーション促進の両立

- AI開発者と提供者の間で責任分担を明確にし、AIの安全性を確保する仕組みを導入すべき
- AI製品や安全対策に係る認証制度等によって、組織的なガバナンスを強化する必要
- 適切なAI制度と併せて、産業政策としての支援策や、高いAIリテラシーを持つ人材育成等を進めるべき

(2) 技術・ビジネスの変化の速さに対応できる柔軟な制度設計

- 技術やビジネスの変化に係るモニタリングに応じた制度の見直しやガイドラインの提供が重要。AI事業者や技術者の意見を制度設計に取り入れる仕組みが必要
- 細かい規定は業界に委ね、政府は全体の目標を定めるなど、ゴールベースの制度設計を推進すべき
- 国際基準と整合性の取れた安全性評価基準を策定し、事業者の自主的な取組みを支援すべき
- 規制内容を分かりやすく設定するとともに、相談窓口やFAQを設け事業者の負担を軽減すべき

2. 設問と主な回答（2）

(3) 国際的な相互運用性や国際的な指針への準拠

- AI開発に必要なデータの国際的な流通を促進するため、データ保有者とAI事業者間のデータ流通を促進する制度設計やデータプール構築等に関する国際標準化が必要
- 各国の規制と整合性を保ちつつ、日本の法制度を柔軟に適用できるようにすべき
- 国際的なAI指針の策定に積極的に参加し、日本独自の強みを生かせる国際指針を形成すべき
- AI安全性基準の整備や、提供するAIサービスの国際標準化に向けて、AISI等国際的な連携に期待

(4) 政府によるAIの適正な調達と利用に関する考え

- 日本のAI技術の研究開発力の維持に政府調達が不可欠。国内技術の育成を通じて、外部依存を減らし、安定供給とセキュリティを確保することが極めて重要
- 調達・利用にあたっての責任の明確化・判断プロセスの透明化が必要。政府内でガバナンス体制を構築することによってAIのメリットを享受しつつ、国民に還元すべき
- 調達に際して、公正かつ多様な機会を提供し、日本のAI産業の振興と国際競争力の強化を図るべき
- AI調達基準やガイドラインの明確化、さらにISMAPのように、企業が参照可能な調達基準の整備が肝要

2. 設問と主な回答（3）

(5) その他意見

- ✓ **認証制度**：導入に際し、欧米との協調や、JIS/CE認証の相互承認のように国際的な枠組みの整備が肝要
- ✓ **業種ごとの対応**：業種共通のAI事業者ガイドラインに準拠しつつ、業種ごとの追加ルールを明確化すべき
- ✓ **中小企業への配慮**：新たな制度の導入に際しては、中小企業への支援策の展開も必要

3. まとめ

まずは、既存の法制度やAI事業者ガイドライン等を「ベースライン」と位置付けた上で、「リスクベース・アプローチ」によって、業種毎にリスクの高い分野を特定した上で、ハードロー導入の是非を議論すべきではないか

原山構成員
提出資料

欧州評議会 (Council of Europe) AI枠組条約 (Framework Convention) を読み解く

東北大学
GPAI東京専門家支援センター
原山優子

欧州評議会の位置付け

Statute of the Council of Europe (London, 5.V.1949)
To achieve a greater unity between its Members for the purpose of safeguarding and realising the ideals and principles which are their common heritage and facilitating their economic and social progress

欧州連合 (European Union)

■マーストリヒト条約 (1993/11)

■政治的・経済的統合

■27カ国

■共通外交・安全保障政策、司法・内務協力、単一市場

→AI規則 (AI Act)

欧州評議会 (Council of Europe)

■ロンドン条約 (1949/5)

■人権・民主主義・法の支配など共通価値の実現

■46カ国（オブザーバー５カ国）

■人権に関する条約策定、履行確保

→**AI枠組条約**
(Framework Convention on AI)

加盟国相関図

条約交渉参加57カ国

欧州評議会46カ国

OECD38カ国

EU27カ国
ベルギー
チェコ
デンマーク
エストニア
フィンランド
フランス
リトアニア
ルーマニア
マルタ
ドイツ
ギリシャ
ハンガリー
アイルランド
イタリア
ルクセンブルク
オランダ
ラトビア
クロアチア
キプロス
ポーランド
ポルトガル
スロバキア
スロベニア
スペイン
スウェーデン
オーストリア
ブルガリア

英国
スイス
ノルウェー
アイスランド
トルコ

セルビア
北マケドニア
ウクライナ
ジョージア
モンテネグロ
アルバニア
モルドバ
アゼルバイジャン
ボスニア＝ヘルツェゴビナ
アルメニア
アンドラ
モナコ
サンマリノ
リヒテンシュタイン

欧州評議会オブザーバー
日本
米国
カナダ
メキシコ
バチカン

交渉会合参加
イスラエル
コスタリカ
豪州
アルゼンチン
ペルー
ウルグアイ

韓国
NZ
チリ
コロンビア

出典
https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2023/07/20230711-Presentation.pdf

これまでの流れ

■Ad Hoc Committee on Artificial Intelligence (CAHAI) (2019-21)

■AI Co-ordination Group (2022-)

■Committee on Artificial Intelligence (CAI)(2022/4-)

→人工知能、人権、民主主義、法の支配に関する枠組条約 (Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law)

- 交渉締結（2024/3/14）
- 閣僚委員会にて採択（2024/5/17）
- 欧州評議会司法大臣会合にて**署名開放** (**2024/9/5**)

→同日署名国：EU (27カ国を代表）、**米国**、英国、**イスラエル**、アンドラ、ジョージア、アイスランド、ノルウェー、モルドバ、サンマリノ

- **日本**？

人工知能、人権、民主主義、法の支配に関する枠組条約

https://rm.coe.int/ai-convention-brochure/1680afaeba

- ■要点
  - ■初の**国際法的拘束力**のある条約
  - ■AIシステム（OECDの定義）の**ライフサイクル**を対象
  - ■**人権、民主主義、法の支配**を担保しつつ、**技術革新**と**イノベーション**に資する(Preamble)
    - ■ロンドン条約の精神
    - ■AIシステムの悪用、抑圧目的でのAIシステムの使用を懸念
- ■ドラフティング
  - ■メンバー国＆オブザーバー国
  - ←オブザーバー国間の連携
- ■プレナリーへの参加・意見聴取
  - ■68 国際組織・団体
    - ■OECD→改訂AI原則より「AIシステム」の定義を採択
    - ■標準機関 (e.g. ISO, IEEE)
    - ■市民団体 (CSOs)、学術界 (e.g. Alan Turing Institute)、産業界 (e.g. IBM)

枠組条約の内容 (1)

- 構成
  - 基本原則
  - 救済措置、手続き上の権利および保護措置
  - リスクおよび影響管理要件
- 条約履行手段(art.1-2)
  - 締約国は本条約の履行のため適切な立法上、行政上又はその他の措置を採用/維持
- 適用範囲(art.3)
  - **公的部門**への適用（公共調達先の民間企業も含む）
    - 対象となる民間企業のAIシステムのライフサイクルから生ずるリスクや影響に対処
    - 対象外の民間企業への対処方法（条約の適用、または、それ以外の適切な措置）につき、署名/批准時に宣言、国防は対象外
    - → 米国、英国は政府共通のAIガバナンスの仕組みを準備中（米国：大統領令を根拠にOMB覚書）
- 救済(art.14)
  - 人権侵害に対する実効的救済の確保
    - 人権に悪影響を及ぼしうるAIシステムについての関連情報を文書化、当該情報へのアクセスを許可された団体に提供、適切・可能な場合、被害者に通知、権限ある当局への申立の確保

枠組条約の内容 (2)

- 手続的セーフガード(art.15)
  - （救済に関する）効果的な手続の確保
    - 適切な場合、AIシステムが対応していることを通知
- リスク・影響評価 (art.16)
  - 締約国は、AIシステム・ライフサイクルから生じる人権、民主主義、法の支配に関するリスク・影響の特定
    - 評価、予防、緩和のための措置をリスクベースアプローチに基づき採用/維持
- 一時的な停止又は禁止 (art.16-4)
  - AIシステムの特定の使用が人権、民主主義、法の支配と合致しない場合、一時的な停止又は禁止等の措置の必要性を判断
- 報告 (art.24)
  - 公的部門における措置及び民間部門のAIシステムのライフサイクルから生ずるリスクや影響への対処について、締結から２年後、その後は定期的に報告
- 監督枠組 (art.26)
  - 履行の監督のため、独立・公平で、適切な権限・リソースを備えた枠組みを設置/指定

他のオブザーバー国を観察すると

■第5回CAI会議(2023/4)に総務省参与として参加

■デレゲーションの構成員
- 政府横断的なチーム形成

■欧州連合に対して
- すり合わせの機会

■国内のAI制度整備を念頭においた交渉
- 政府共通のAIガバナンスの仕組み
  - リスク影響評価（AIシステムの政府調達）
  - 救済に関する仕組み
  - 適切性の監督枠組
- 政府機関の所管事項への配慮

平野構成員
提出資料

M E M O R A N D U M

未校閲版

Dec. 26, 2024

中央大学 国際情報学部長・教授

平野 晋（すすむ）

Ⅰ．個人の性格、能力、成長性、感情等をAIに評価・予測・決定補助等させる問題[1]

[1] この（意思決定又はその補助の）問題は、別名「ADM」（自動意思決定）とも呼ばれている。「automated-decision making」又は「algorithmic decision making」等の略語である。*See, e.g.*, Sancho McCann et al., *Discretion in the Automated Administrative State*, 36 CAN. J.L. & JURIS. 171 (2023)． また、感情を認識する技術等は「emotion coding」や「affect recognition」等と呼ばれている。*See, e.g.*, Phoebe V. Moore, *Mirror for (Artificial) Intelligence: In Whoses Reflection?*, 41 COMP. LAB. L. & POL'Y J 47, 48 (2019)．なおAIが表情や声音等から性格、能力、成長性、又は感情等を評価することには科学的に根拠がなく、偽科学であると指摘する例として参考になる文献としては、Luke Stark & Jevan Hutson, *Physiognomic Artificial Intelligence*, 32 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 922, 926-28 (2022)． *See also* Maroussia Lévesque, *Analog Privilege*, 26 N.Y.U.J. LEGIS. & PUB. POL'Y 625, 654 (2024)(機械学習に基づく嘘発見器等々の"technologies might improve over time,[] but human judgment currently outperforms them by several orders of magnitude."と指摘)； Wayne A. Logan, *Policing Emotions: What Social Psychology Can Teach Fourth Amendment Doctrine*, 72 BUFF. L. REV. 685, 695 (2024)("A large and growing body of research . . . shows that there is no reliable evidence that humans can accurately and reliably detect emotional states from facial expressions.[]"と指摘)； Sandra Wachter, *Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the Europe Union, the United States, and Beyond*, 26 YALE J.L. & TECH. 671, 681 (2024)(感情認識ソフトウエアが客観性を欠くと指摘)．

- 例えば、〈雇用（含、採用）分野〉や[2]、〈教育分野〉[3]。
- そもそもヒトの判断よりも、AIの方が効率的であり、かつ客観的であるから中立的であると捉える前提が問題であると指摘されている[4]。
- しかしAIもヒトが創ったものだから、ヒトの偏見から逃れられない[5]。
- そして不公正・差別・不正確等の問題が、明らかに成って来た[6]。

---

[2] *See, e.g.*, OFF. OF SCI. & TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS 46 (Oct. 2022)(以下のように指摘："Automated systems with an intended use within sensitive domains, including, but not limited to, criminal justice, employment, education, and health, should additionally be tailored to the purpose, provide meaningful access for oversight, include training for any people interacting with the system, and incorporate human consideration for adverse or high-risk decisions."(emphasis added)).

[3] *See* Stark & Hutson, *supra* note 1, at 957-58 (コロナ禍時代に使用されたカンニング防止の為に目の動きを感知するソフトの使用が学生達から猛反発された事例を例示しながら、態度・表情等から学生を評価する「人相学的AI」の教育分野に於ける使用は、生物学的決定論や優生学や科学的人種差別を元気づけるとして批判している).

[4] *See*, Margaret Hu, *Critical Data Theory*, 65 WM. & MARY L. REV. 839, 862 (2024). なお、そもそもヒトの将来を予測することなどは不可能な事実を人々は常識では理解しているにも拘わらず、AIを用いた途端に人々が常識を失って、「見せかけの正確性」("veneer of accuracy")に騙されると示唆する文献例として、see Stark & Hutson, *supra* note 1, at 929-30.

[5] Houston Fed. of Teachers, *infra* note 24, 251 F.Supp.3d at 1171 ("Algorithms are human creations, and subject to error like any other human endeavor."と法廷意見が指摘). *See also* Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671 (2020)(AIはヒトが作るのだから、暴走したAIの責任をヒトが負わねばならない、あたかもフランケンシュタイン博士が造った人造人間の暴走の責任を同博士が負わねばならなかったのと同じである、と指摘); Danielle Keats Citron, *Technological Due Process*, 85 WASH. L. REV. 1249, 1253 (2008)(適正手続の保障等の政策に疎いプログラマが法を無視したプログラミングを行って法を歪める問題を指摘); Crystal Godfrey, *Legislating Big Tech: The Effects Amazon Recognition Technology Has on Privacy Rights*, 25 U.S.F. INTELL. PROP. & TECH.L.J. 163, 166 (2021)(プログラマの偏見が入り込むと指摘).

[6] *See* authorities cited in *supra* note 1. *See also* Yonathan Arbel et al., *Systematic Regulation of Artificial Intelligence*, 56

- **日本では（も）人事採用 AI のベンダ [7] が活発に売り込み、かつそれを利用する企業も少なくないように見受けるけれども、「やり過ぎ」に注意が必要ではないか[8]。**

ARIZ. ST. L.J. 545, 557 (2024); Godfrey, *supra* note 5, at 168(顔認識技術──**FRT**──が感情を読み取れるという主張が批判されていると指摘**)**. 雇用に **AI** を利用すると差別的結果が生じるリスクを指摘する文献としては、**see, *e.g.*, Sonderling et al., *infra* note 12.**

[7] 人事採用に AI を使うことが不適切であることは、日本でも、例えば平野が座長を務める「AI ネットワーク社会推進会議・**AI** ガバナンス検討会」の第２回会合（平成３０年[2018 年]１２月１０日）に於ける識者のご発表に於いても既に指摘されていた。「資料１ 早稲田大学 大湾先生 御発表資料：人事データ活用への関心とガイドライン作成に向けての議論」 **https://www.soumu.go.jp/main_content/000589116.pdf (last visited Sept. 12, 2024).** 平野自身の文献については、平野晋「**AI** に不適合なアルゴリズム回避論：機械的な人事採用選別と自動化バイアス」『情報通信政策研究』第 **7** 巻 **2** 号１頁（総務省,**2024** 年 **3** 月） **https://www.soumu.go.jp/iicp/journal/journal_07-02.html (last visited Sept. 24, 2024);** 「〔資料１〕**AI** の判断に対するヒトの最終決定権の限界：**Human-in-the Loop** の問題」***in*** 総務省「情報通信法学研究会 令和 **5** 年度」**2023** 年 **9** 月 **6** 日 **https://www.soumu.go.jp/main_sosiki/kenkyu/hougakuken/R05_siryou.html (last visited Sept. 24, 2024).**

[8] 「やり過ぎ」についての批判としては、例えば顔の表情から感情を読み取ることは困難であると指摘されているにも関わらず、大企業やスタートアップ企業等がこれを売り込んでヒトの一生を左右している問題が、連邦取引委員会（**FTC**）委員等による共著論文に於いて次のように指摘されている。この指摘が日本にも当てはまることがないことを、筆者は祈るばかりである。

> **A review that analyzed more than a thousand studies on emotional expression concluded that "[e]fforts to simply 'read out' people's internal states from an analysis of their facial movements alone, without considering various aspects of context, are at best incomplete and at worst entirely lack validity, no matter how sophisticated the computational algorithms."[] Nevertheless, large companies [] --plus a host of well-funded start-ups-- continue to sell questionable affect recognition technology, and it is sometimes deployed to grant or deny formative life opportunities.**
>
> **A striking example of the use of affect recognition is in hiring.**

Rebecca Kelly Slaughter et al., *Algorithm and Economic Justice: A Taxonomy of Harms and a Path Forward for the*

（教育分野に於いても、例えば予備校系／就活系役務提供等の事業者が行うAI性格分析等が本当に科学的に正確であるのか、ヒトの尊厳に反しないのか等々に注意が必要ではないか。）

- 例：〈リクナビ内定者辞退率予測事件〉。
- 例：アメリカの州法や条例がAI利活用の規制を始めている[9]。
- 例：感情認識AI/顔認識AI（emotional-recognition tools / facial recognition technology）の、〈教育〉や〈労働（含、採用）〉分野に於ける利用制限の検討が必要ではないか[10]。
- EUのAI法は既に、職場と教育機関に於ける、感情を推認するAIシステムの使用等も原則として禁止にした。同法第５条１項(f)項[11]。

Federal Trade Commission, 23 YALE J.L. & TECH. 1, 11-12 (2021)(emphasis added).

[9] *E.g.*, THE ILLINOIS ARTIFICIAL INTELLIGENCE VIDEO INTERVIEW ACT; MARYLAND CODE, LABOR & EMPLOYMENT §3-717; NEW YORK CITY LOCAL LAW 144.

[10] *See, e.g.*, Lucy L. Thomson & Trooper Sanders, *Human Rights Challenges with Artificial Intelligence*, 49 HUM. RTS. 24, 24 (2024)(EU AI Actが教育・労働分野に於けるemotion-recognition toolsの利用を制限している例をABA――全米法曹協会――が指摘); Wachter, *supra* note 1, at 716 ("Emotion detection AI should also be widely banned, ... due to the high levels of inaccuracy and lack of scientific evidence establishing the reliability of these techniques . . . ."(emphasis added)と主張); Leigh Harvis-Nazzario, Note, *It's the People: Preventing Bias in Automated Hiring Tools Starts with Humans*, 49 RUTGERS COMPUTER & TECH. L.J. 138, 149-50 (2022)(採用AI役務提供大手の某社が、採用候補者の顔の表情と性格との相関関係から仕事上の成功を予測可能であると主張していたところ、人権団体から連邦取引委員会―FTC―に提訴された事実を指摘しつつ、顔認識システムが偽陽性を生むという分析報告がNISTから公表された事実も指摘); Marylou Fabbo, *Companies' Online Recording Should Comply with Other State Laws*, 5 No. 5 NEW ENG. EMP. L. LETTER 3 (May 2024)(某社が、様々な質問に応答する候補者の顔の表情、アイ・コンタクト、及び声のイントネーション等の動画を分析して、顧客企業への適正を評価したことが、嘘発見器使用禁止法違反に当たるとして候補者が提訴した事例を紹介).

[11] EU AI Act, Regulation (EU) 2024/1689,

- 法執行の強化、及びガイドラインの遵守を強化すべきではないか。
  - 例えば採用活動に於ける AI 利用が、男女雇用機会均等法上の〈間接差別：disparate impact〉禁止義務や、〈公正採用義務〉を、遵守しているのかについての点検が必要[12]。
  - アメリカの大統領令が、AI 利用の問題領域を管轄する官庁に対しては具体的に法執行等の強化を命じていることに[13]、日本も見倣うべきかも。
- 『OECD AI 原則』第 1.3 条： 平易な説明責任と異議申立権を規定しているが[14]、日本の実務ではこれが遵守されていないのではないか。

---

https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html (last visited Sept. 21, 2024)("The following AI practices shall be prohibited: . . . the use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions"(emphasis added)と規定). *Cf.* Wachter, *supra* note 1, at 680-81 (感情認識ソフトがヒトの反応を客観的に測れるという科学的根拠は殆ど無いか又は全くない―"emotion recognition software has little to no ability to objectively measure reactions"―と指摘).

[12] *See* U.S. Equal Employment Opportunity Commission (EEOC), Artificial Intelligence and Algorithmic Fairness Initiative, https://www.eeoc.gov/ai (last visited Sept. 10, 2024)(積極的に法執行する意向を表明). *See also* Keith E. Sonderling et al., *The Promise and the Peril: Artificial Intelligence and Employment Discrimination*, 7 U. MIAMI L. REV. 1 (2022)(雇用主による AI 利用に対しても現行法規を積極的に執行する主旨で、AI 利用の諸問題を指摘した、EEOC 委員等による共著論文).

[13] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Oct. 30, 2024 (たとえば "Within 365 days of the date of this order, to prevent unlawful discrimination from AI used for hiring, the Secretary of Labor shall publish guidance for Federal contractors regarding nondiscrimination in hiring involving AI and other technology-based hiring systems."(emphasis added)と具体的に命じている).

[14] 本研究会第 1 回会合, 平野提出資料参照。*See also* Margot E. Kaminski & Urban Jennifer M., *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 1982 & n.146 (2021)(OECD AI 原則上の異議申立権を紹介しつつ、これ

- 候補者に告知と同意を求めたり、AI の判断を最終的にヒトが判断するという手続も、単に不適切な AI の利活用を正当化しているに過ぎないと批判されている[15]。
- input/output を定期的にモニタリング(内部監査)し修正する必要性があるけれども[16]、実行されているのかも不明。
- 独立した第三者による監査と開示の必要性が指摘されているけれども(次段参照)、実施されているのか。
  - arms-length ではない second-party による監査は[17]、信頼できないと批判されている[18]。
    - *Cf.*〈宝塚歌劇団いじめ事件〉の内部調査結果。
- second-party による「お手盛り」な監査では、「偽りの保証（false

が世界の立法者や実務に影響を与えるかもしれないと指摘)．なお、AI の決定に対する異議申立（告知聴聞）権のような適正手続は、AI の不正確なアウトプットを事後的に発見してこれを正す機能もある、という指摘(*id.* at 1989-90, 1998-99)には、説得性があろう。

[15] Stark & Hutson, *supra* note 1, at 930-31.

[16] Gary D. Friedman, *The Role and Influence of Artificial Intelligence in the Workplace*, 20221121P NYCBAR 38, Nov. 21, 2022 (City Bar Center for Continuing Legal Education, New York City Bar)(機械学習によって悪いインプットを学習する危険性があるので、定期的にインプットとアウトプットを企業が分析することが重要であると指摘)． *See also* Sonderling et al, *supra* note 12, at 79 (同旨)．

[17]「second-party audit」については、Ellen P. Goodman & Julia Trehu, *Algorithmic Auditing: Chasing AI Accountability*, 39 SANTA CLARA HIGH TECH. L.J. 289, 317 (2022-23)．

[18] *Id.* at 316-17. *See also id.* at 306-07, 308 & nn.70, 76 (EU の Digital Service Act に関連して独立した監査の重要性を指摘)．なおソフトウエア開発者から報酬を得て監査する会社にとっては、その開発者に不利な監査をする動機に欠けるとして批判する例としては、Robert Wennagel, *Dark Systems: Reprogramming Artificial Intelligence Regulations to Promote Fairness and Employment Nondiscrimination*, 39 SANTA CLARA HIGH TECH. L.J. 1, 57 (2022-23)．

assurance)」を与えてしまう、と批判されている[19]。

- 加えて「正確性」や「公正性」や「説明責任」等を実現させる方法としては、本来ならば外部の研究者による検証（peer review）が望ましいところ[20]、どこまでの妥協策なら許容され、かつ現実的であるのかも要検討。

---

[19] Goodman & Trehu, *supra* note 17, at 302-03 (本当は問題のある行動をしていても、これを「監査ロンダリング――"audit washing"――」して、法や社会規範を遵守しているような「偽りの保証――"false assurance" ――」を与えてしまうと指摘). なお、きちんと要件を規制されていない監査は、自社のプログラムを正当化させる虫の良い手段として使われてしまうと指摘する文献として、see *also* Wennagel, *supra* note 18, at 56.

[20] *See, e.g.*, Pauline T. Kim, *Data-Driven Discrimination at Work,* 58 WM. & MARY L. REV. 857, 879-80 (2017)("And in the absence of data sharing [between inside researchers and outsiders] or transparency about the choices made in constructing the model, others cannot test the robustness and validity of the results."と指摘して、透明性を欠く為に外部の専門家によって検証されない問題を指摘); Charlotte A. Tschider, *Beyond the "Black Box,"* 98 DEBV. L, REV. 683, 706 (2021)("lack of peer review"故にもたらされる危険性を指摘); Anastasia Konina, *Banks as Delegated Regulators of Technology*, 59 ALBERTA L. REV. 753, 768 (2022)(カナダ政府が調達する対象の規制実施用技術については、ベンダが実装する前にpeer reviewを実施せねばならない仕組みになっていると紹介); Goodman & Trehu, *supra* note 17, at 311(カナダ政府の指令"Algorithmic Impact Assessment Tool and the Directive on Automated Decision Making"に於いて、peer reviewを要求していると指摘); Teresa Scassa, *Administrative Law and the Governance of Automated Decision Making: A Critical Look at Canada's Directive on Automated Decision Making*, 54 U.B.C. L. REV. 251, 276 (2021)(カナダのDirective on Automated Decision-Making―DADM―とAlgorithm Impact Assessment―AIA― toolに関する論文に於いて、peer reviewが要求されていると指摘); Cary Coglianese & David Lehr, *Algorithmic Governance*, 71 ADMIN. L. REV. 1, 49 (2019)(営業秘密を尊重しながら透明性を実現する手段として、"independent peer reviews could be conducted under non-disclosure agreements"と指摘); Stephanie L. Lee, Note, *Clicking away Consent: Establishing: Accountability and Liability Apportionment in Director-to-Consumer Healthcare Artificial Intelligence*, 88 BROOKLYN L. REV. 1355, 1377 (2023)(AI開発者がモデルの弱点に取り組む為に、多様な研究者、教育者、及びデータ・サイエンティストに

- **オプトアウトの権利を付与すべきとされているけれども[21]、日本の実務では付与されていないのではないか（オプトアウトの権利を付与しなければ、告知しても無意味かも）（他方、就活生にとっては、力[ちから]関係故にオプトアウトの権利付与でも不十分かも）。**
- **AI ベンダはその役務を購入する顧客企業の為に行動するのであって、それを適用される個人側（*e.g.*, 就活生）の利益の為には行動しない[22]。従って、後者の利益や要望もステークホールダーとして、ルールに反映させるべきではないか[23]。**

よる peer review にアルゴリズムとソフトウエアをさらすべきであると指摘); Brendan Max, *SoundThinking's Black-Box*, 26 STAN. TECH. L. REV. 193, 238 (2023)(犯罪捜査に使われて冤罪を招いたアルゴリズムの問題に関する論文に於いて、peer-review に服していなかった点を批判). *See also Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993)(科学的に正しい専門家証人による証言認容考慮要素を示した代表判例である本件は、peer review にさらされていることも要素に入れている).

[21] AI BILL OF RIGHTS, *supra* note 2, at 7. *See also* Jonathan Ben-Asher, ALI-CLE Materials, *Artificial Intelligence and Employee Monitoring: Big Brother Is an Algorithm Named Julie, and She's Hiring, Watching and Rating You,* July 25-27, 2024 (コロラド州が 2024 年 5 月に制定した AI 法に於いては、消費者が AI 利用をオプト･アウトする権利が付与されていると指摘).

[22] *See* Brittany Kammerer, *Hired by a Robot: The Legal Implications of Artificial Intelligence Video Interviews and Advocating for Greater Protection of Job Applicants*, 107 IOWA L. REV. 817, 848 (2022)(採用 AI ベンダの関心事は顧客たる企業であって採用候補者ではないから、後者の声を代弁して後者を保護する為に立法府が規制立法すべきであると示唆している).

[23] *See* Margot Kaminski, *Voices in, Voices out: Impacted Stakeholders and the Governance of AI*, 71 UCLA L. REV. DISCOURSE 176, 191-92 (2024)(NIST の Artificial Intelligence Risk Management Framework の例を挙げながら、AI による影響を受けるステークホルダーの意見を反映させるべきと主張). *See also* Citron, *Technological Due Process*, *supra* note 5, at 1312 (rule making における公衆の意見反映の必要性を指摘).

II．政府調達に関連する問題

- 政府機関・地方公共団体がAIの評価や予測・推奨・決定等を用いて不利益処分を行う場合には、適正手続の保障が要請され、説明責任（告知聴聞等）を果たさねばならないのではないか[24]。
- その際、ベンダが提供するAIの不透明性ゆえに、政府機関・地方公共団体が負う説明責任を果たせなくならないような配慮が必要[25]。
- 例えば、政府調達の要件や契約に於いて、或る程度の情報開示や説明責任を義務化せねばならないのではないか[26]。ベンダ側の〈営業秘密〉との調整が必要[27]。

---

[24] *See, e.g.*, Houston Federation of Teachers, Local 2415 v. Houston Independent School District, 251 F.Supp.3d 1168 (S.D.Tex. 2017)(教員解雇の原因であるアルゴリズム上の低評価の理由を説明しないことは手続的適正手続の保障違反であるとされた事例)； Rebecca Crootof et al., *Human in the Loop*, 76 VAND. L. REV. 429, 454 (2023)(*Houston Fed. of Teachers* 事件を紹介)； Cahoo v. SAS Analytics Inc., 912 F.3d 887 (6th Cir. 2019)(アルゴリズムにより失業保険受給資格の有無を自動的に決定させていたところ、資格なしとの誤判断が93%にも達した上に、告知聴聞の権利も付与しなかった為に適正手続保障違反であったばかりか、その事実を知りながらも漫然とこれを使用し続けた怠慢ゆえに、関連する公務員が裁判所によって免責特権を否認された事件)． *See also* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014)(Technological Due Processの必要性を指摘)； Citron, *Technological Due Process*, *supra* note 5 (同旨)．

[25] *See Houston Fed. of Teachers*, *supra* note 24.

[26] *See* AI BILL OF RIGHTS, *supra* note 2, at 20 (third-party auditors等が評価を出来るようなアクセスを付与すべきと指摘)．更に、カナダ政府の例を紹介する前掲脚注20の出典も、政府調達の在り方を検討する上で有用かもしれない。

[27] *See* Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621, 651 (2021)(営業秘密を口実として開示を拒む問題を指摘)； Charlotte A. Tschider, *Legal Opacity: Artificial Intelligence's Sticky Wicket*, 106 IOWA L. REV. ONLINE 126, 131-32 (2021)(同旨)； John Villasenor, *Artificial*

III．倫理だけでは不十分という問題

筆者撮影@連邦最高裁 2017 年 1 月 10 日

- Oliver Wendell Holmes、Jr., *The path of the law*, 10 HARV. L. REV. 457, 457-458 (1897).
- *See also* 内閣府「第 3 回 人間中心の AI 社会原則検討会議 会議録」26 頁,平成 30 年[2018 年]7 月 5 日.
- 規制立法と自主規制を併用するガバナンスの場合には、<u>後者をきちんと遵守させる為の施策を併用する方法</u>も指摘されている。↓

Intelligence, Trade Secrets, and the Challenge of Transparency, 26 N.C. J.L. & TECH. 495, 530-31 (2024)(企業が提出情報内の営業秘密部分についての守秘扱いを要請した場合の、政府機関との調整の問題を指摘)．なお前掲注 24 の *Houston Fed. of Teachers*, 251 F.Supp.3d 1168 事件に於いては、裁判所が、原告側弁護士と鑑定人のみへのアクセスを許容する（"attorney eyes only"的な）守秘義務を負わせる裁判所命令（court order）を下した上で、ベンダのソースコードを含む営業秘密情報を原告側に開示させている。Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1282 (2020)． このような手続は、日本でも、今後の営業秘密の利益と適正手続の保障の利益との抵触問題の調整策を考える際の一助になるかもしれない。

- 例：規制省庁が自主規制を監視し、もし自主規制が機能しない場合には法の介入（*e.g.*，事業法等の法改正や政省令改正等々）を示唆することにより遵守させる[28]。

IV.「黒い白鳥」（"Black Swan"[29]）への備えを

- 予期せぬ危害への予防法務の必要性
  - 例：CORVID-19 パンデミック、ロシアのウクライナ侵攻[30]。
- AGI の危険性について、嘗ては〈SF〉等とレッテルされ揶揄された。しかし、今では．．．[31]

---

[28] *See* Margot E. Kaminski, *Regulating the Risk of AI*, 103 BOSTON U. L. REV. 1347, 1405 (2023)(完全な自主規制――"enterprise risk management"――に委ねる方法以外にも、規制官庁が介入するガバナンスの例もあると指摘しつつ、次のように述べている："While enterprise risk management can occur in the absence or shadow of law, regulators can also participate by nudging companies to conduct risk mitigation through oversight, through the threat of regulatory enforcement, by offering safe harbors, or by issuing best practices or other guidance."(emphasis added)).

[29] Noam Kolt, *Algorithmic Black Swan*, 101 WASH. L. REV. 1177, 1182 (2024).

[30] *Id.* at 1182.

[31] *E.g.*, Ethan Holland, *The AI Future*, 41-41-Fall DEL. LAW. 10, 14 (Fall 2023)("The speed at which experts predict AGI's arrival is daunting. Shane Legg, DeepMind co-founder, predicts 2028. Elon Musk, 2029. Geoffrey Hinton, the Godfather of AI, says five to 20 years. OpenAI's Sam Altman and DeepMind CEO Demis Hassabis both predict less than 10 years."); Glenn T. Melchinger, *The New Legal Code: Will AI Agents Bye-Back?*, 28-Jun HAW. B.J. 4 (2024)("the AI arms race is only accelerating and expanding, with new signs of AGI seemingly every day. . . . / Because of the disruptive implications AI has on human society, several high profile folks in the AI world have left their jobs and spoken about the risks, including the fear of "extinction

- 定点観測的にAI開発の進捗具合を政府が把握しておくことは、最低限度必要かも。
- 「国際的な議論のためのAI開発ガイドライン案」に於いても、熱心な議論の末に、汎用AIを対象とした[32]。
- 核物質の管理法規のような備えも、今後は必要ではないか[33]。

V. *ex ante* な事前規制のみならず、*ex post* な事後規制/救済も要検討

- 労働や教育分野のようなsensitive domainに於ける事後監査が必要かも。
- 今後の課題として、被害への賠償の在り方も要検討かも； *Cf.* EUのAI賠償責任指令案や製造物責任指令改定案、等々。不法行為法の目的の一つである「抑止」機能に重要な「内部化」（internalization）や、「矯正的正義」(corrective justice)[34] の実現が、不透明性ゆえに困難である問題の解決も要検討。
- ハードローである制定法等ではなくても、ソフトローとしての「標準」や「原則」や「ガイドライン」等々も、不法行為訴訟に於いては「過失」（注意義務）や「欠陥」（「通常有すべき安全性」）の基準として裁判所に認定されて、事後的に法のような効果が生じることがある。

---

level events.").

[32] AIネットワーク社会推進会議「報告書２０１７」平成29年7月28日，26-27頁＆脚注63-67.

[33] Kolt, *supra* note 29, at 1218.

[34] 不法行為法の原理である抑止機能や矯正的正義等については、see *generally* 平野晋『アメリカ不法行為法：主要概念と学際法理』（中央大学出版部，2006年）.

VI．まとめ

- AIの主な欠点である〈制御不可能性〉〈不透明性〉〈不公正な予測・推奨・決定等〉に照らして、以上をまとめつつ付言してみると、以下のようになろう。

  - アメリカのsensitive domainやＥＵのprohibited/high-risk AI systems領域については、<u>関係省庁による法執行とガイドラインへのコンプライアンス遵守や自主規制の監視を強化し、場合によっては法制定及び/又は法改正も視野に入れる</u>べき。

・【制御不可能性（予見不可能性）について】

- 安全性が重要な分野（例えば交通）に於いては、（完全自律型致死的兵器の実現不可能性の指摘も参考にして）事前のシミュレーションだけでは限界があることを認識した上で、利用を慎重にすべき。
- 個人を評価・予測するような分野（例えば労働/採用や教育）に於いては、定期的なinput/outputの内部監査、出自が明らかで更新された適切なデータ使用、及び差別的効果・間接差別や自動化バイアスが生じないように関与する人々のELSI的トレーニング＋α[35]等々が重要。

・【不透明性（説明責任）について】

- 特に個人の一生に関わる意思決定等に於いては、分かり易い（平易な）説明の義務化が必要。（現行法規だけでは不十分かも。）
- 〈営業秘密〉の壁を突破する開示義務の為の方策を要検討。（現行法規だけでは不十分かも。）
- 「お手盛り」な質保証や「偽りの保証」や「監査ロンダリング」にならないような、独立した第三者による中立的な監査と公開が必要。（現行

---

[35] debiasing（バイアス除去技法）を用いても効果には限界があると指摘する文献として、see *generally* Kevin Jon Heller, *The Concept of "the Human" in the Critique of Autonomous Weapons*, 14 HARV. NAT'L SEC. J. 1 (2023).

法規だけでは不十分かも。）

- 出来れば peer review に服させて検証することが望ましい。（現行法規だけでは不十分かも。）
- 事後的救済（民事賠償訴訟）に於ける原告側の立証困難性に対する措置の必要性も要検討。（「過失」や「欠陥」や「因果関係」の判例法上の推認のみで足りるか、又は不十分であれば立法も視野に。）

・【不公正な予測・推奨・決定について】

- 特に個人の一生に関わる意思決定等に於いては、オプトアウトの権利付与を要検討。（現行法規だけでは不十分かも。）（オプトアウト権付与でも不十分かも。）
- 国際基準（OECD AI 原則 1.3 等）と整合させ、かつ〈正確性〉を担保する為にも、異議申立権の付与、及びその為に必要な「説明責任」の履行が必要。
- 特に政府調達による政府の AI 利用に関しては、適正手続の保障義務を履行できるような仕組みも要検討。
- 不利益を被るステークホールダーに対して、ルール作成等への参加機会の付与も重要。

番外：その他

- 新興技術分野に於けるガバナンスの在り方論議は、サイバースペース法学に於けるガバナンス論（新たな対応が必要な分野の考察や判例法形成等）も参考に。（日本の事例がまだ無ければ米国に於ける具体的な問題事例から学ぶべき。）

++++++++ END OF THE TEXT ++++++++

福岡構成員
提出資料

AI制度研究会提出資料

AIルールの現状と対応

2024年12月26日

弁護士 福岡 真之介

課題	主要法律	現時点の検討・対応	備考、論点・対応
①AIの開発・利用全般	なし	「AI事業者ガイドライン」（2024/4/19）	【備考】EUではAI法が成立。
②AIの安全・安心	なし	AI向けの法律はなく、個別法で対応	【備考】EUではAI規則案、米国では「AIの安全・安心・信頼できる開発と利用に関する大統領令」が存在。個別法におけるAI対応にはばらつきがある。
③AIの開発・利用による著作権侵害	著作権法	「AIと著作権に関する考え方について」（2024/3/15）	【備考】AIに関する諸論点について現行法の解釈が記載されている。
④AIの開発・利用による意匠権・商標権侵害・不正競争防止法違反	意匠法、商標法、不正競争防止法	「AI時代の知的財産検討会中間とりまとめ」（2024/5）	【備考】AIに関する諸論点について現行法の解釈が記載されている。
⑤ディープフェイク（AI合成した肖像・声等の悪用）	民法（人格権・不法行為）、刑法（名誉棄損罪、わいせつ物頒布等罪）	「AI時代の知的財産検討会中間とりまとめ」（2024/5）	【対応】現在、肖像権・声の人格権の侵害に対して法律はなく、裁判例の積み重ねであり、AI非対応のため、①新規立法や、②ガイドラインによる対応が考えられる。
⑥偽・誤情報の拡散	民法（人格権・不法行為）、刑法（名誉棄損罪）	「デジタル空間における情報流通の健全性確保の在り方に関する検討会とりまとめ（案）」（2024/7/16）	【対応】「とりまとめ」には、偽・誤情報への対応として、プラットフォーム事業者による対応等が提言されている。

課題	主要法律	現時点の対応	備考、考えられる対応
⑦AIへの秘密情報の入力	不正競争防止法、民法（契約法）	「秘密情報の保護ハンドブック」及び「限定提供データに関する指針」の生成AIに関する記載部分改訂（2024/2）	【論点】生成AIへの秘密情報の入力が、①秘密情報の漏えいにならないか、②秘密管理性が失われて営業秘密の要件を満たさなくなるのではないか。
⑧個人情報の利用	個人情報保護法	「個人情報保護法 いわゆる3年ごと見直しに係る検討の中間整理」（2024/6/27）	【論点】①ユーザデータを学習しない生成AIへの個人データの入力が第三者提供にあたるか、②スクレイピングで取得したデータに要配慮個人情報が含まれている場合に本人同意が必要か。 【対応】個人情報保護法や同ガイドラインの改正による対応が考えられる。
⑨ハルシネーション	民法（不法行為、契約違反）	なし	【対応】チャットボット等のハルシネーションによる不法行為責任については、生成AI利用を明示した場合には、責任を軽減する法律の制定や免責条項を記載した利用規約が不当条項にあたらない旨の明確化が考えられる。
⑩ウイルスの作成	刑法（不正指令電磁的記録に関する罪）、不正アクセス禁止法	なし	【備考】2011年に不正指令電磁的記録に関する罪が新設され、ウイルス作成等が対象となった。なお、同罪に予備罪の規定はない。
⑪セキュリティ（安全・安心な利用）	サイバーセキュリティ基本法	なし	【対応】ユーザが安心してAIを利用するためにAIの認証制度を設けることが考えられる。なお、クラウドサービスには政府情報システムのセキュリティ評価制度のISMAPがある。

課題	主要法律	現時点の対応	備考、考えられる対応
⑫安全保障・武器	武器等製造法、生物兵器禁止法、化学兵器禁止法、経済安全保障推進法	なし	【備考】武器等製造法は銃器や火薬の規制が中心。経済安全保障推進法は戦略物資の安定供給確保が中心。AI関連の規定を設けることが考えられる。