

## AI の責任ある展開に向けて

### 広島 AI プロセスへの政策提言

東京大学未来ビジョン研究センター 技術ガバナンス研究ユニット

## エグゼクティブサマリー

機械学習技術を中心とする人工知能（AI）技術が普及してきたことに加え、様々なコンテンツを生成する生成 AI が、多くの人に広く利用されることにより、生活や働き方に改善やイノベーションの期待が高まった一方、AI の適切な設計、開発、提供、利用の必要性が再確認された。AI は社会の中で学習し展開されるため、開発者、提供者、利用者、公的機関や社会全般を含む包括的な AI ガバナンスの構築が必要である。透明かつ公平な規律に基づくことで、イノベーションの促進とリスクの緩和を両立する枠組みの共有が求められている。

## 広島 AI プロセスの基本方針の提案

### 1. AI の責任ある展開のための議論の場の形成

基本的人権や民主的価値といった共通する価値を遵守するため、生成 AI がもたらす様々な可能性と課題を踏まえ、迅速かつ継続的にマルチステークホルダーで議論する場の形成を支援するべきである。

### 2. 枠組み間の「相互運用性」推進と AI 規律の相互尊重

G7 広島サミットで議論された「相互運用性」は、「標準」と「枠組み間の相互運用性」の二側面で考えられる。枠組み間の相互運用性とは、異なる国・地域や組織、分野の AI に関する規律を尊重しながらも、共通の目標を達成するための考え方である。規律には様々な分類があり、新技術にすべて新しい規律が必ずしも必要ではない。しかし AI サービスと既存法の適用関係が不明確な時や、社会的弱者を保護する等の目的のためには、各国・地域や分野において新立法を含めた対応を必要に応じて検討するべきである。

### 3. AI の展開における責任主体と対応策

AI の責任ある展開を推進するには責任主体を明確化し、各々に適切な対応策を整理すべきである。特に AI 設計から利用に至る一連のプロセスにおいて、異なる組織や個人が関与する場合、責任の所在が不明瞭になりかねない。そのため、AI の開発から提供に関わる事業者間取引では、契約による適切性確保が重要である。また、適切な取引が確保されるような監視する仕組みも構築するべきである。一方、AI 提供者と消費者間の取引では、提供者が適切な事前・事後対策を取るほか、AI 利用者も適切なりテラシーを身に着けることによって、法規制とは別の規律（市場・投資・評判等）によるガバナンスを利かせることも可能となる。また補償制度の構築等の救済措置を必要に応じて構築するべきである。AI ライフサイクルが国や組織を超えて展開することを踏まえ、このような各主体の責任と対応策の透明性を高める議論を推進していくべきである。

## 本政策提言策定のプロセス

本提言の内容は、東京大学未来ビジョン研究センター技術ガバナンス研究ユニット AI ガバナンスプロジェクトメンバーである江間有沙と、外部有識者である実積寿也（中央大学総合政策学部教授）、中川裕志（理化学研究所・革新知能統合研究センターチームリーダー）、加藤尚徳（次世代基盤政策研究所 (NFI) 研究主幹・理事) の議論に基づく。産学官の有識者に草案を送り、2023年8月3日にオンラインでフィードバックをいただく会合を開催したほか、メールや対面でいただいたフィードバックを盛り込んでいる。本稿を取りまとめるにあたってフィードバックを寄せていただいた方々のリストは附録を参照されたい。

## AI の責任ある展開に向けて

概ね 2010 年代半ば以降、深層学習でのブレークスルーをきっかけとした第三次 AI ブームにより機械学習技術を中心とする人工知能 (AI) 技術の普及が進展してきたことに加え、2022 年から 2023 年にかけて文章、音源、画像や動画等を含むコンテンツを生成できる多くの生成人工知能 (以下、生成 AI) が登場し、仕事だけではなく私的利用も含めて多くの人に身近に使われるようになった<sup>1</sup>。これにより生活や働き方に関する改善やイノベーションが期待される一方、偽情報、プライバシー、著作権、公平性等これまでに提起されてきた様々な課題が改めて認識されつつある。

AI 技術やシステム、サービスの適切な設計、開発、提供や利用に関するガバナンスの議論は、これまで AI 開発者や AI 提供者に主に焦点があてられてきた。しかし生成 AI を構成する機械学習技術は、社会の中で学習し展開するため、利用現場における人と機械、環境との相互作用がこれまで以上に重視される<sup>2</sup>。そこで AI 開発者、提供者だけではなく利用者や公的機関、社会も含めた包括的な観点から、責任ある AI の展開を議論する必要がある。

これは利用者に責任を押し付けるのではなく、誰もがアクセスできる透明かつ公平な規律の共通理解を推奨する提案である。これによりリスクを緩和しつつ、中小企業や新興企業の参入障壁を低くしてさらなるイノベーションの加速へとつなげる。このような点から、本政策提言では生成 AI も含む機械学習に基づく AI のガバナンスを適切に行うため、以下の 3 点の広島 AI プロセスの基本方針を提案する。

### 1. AI の責任ある展開のための議論の場の形成

国際機関、国や地域、市民団体や企業がこれまで策定してきた AI 原則には、人間の尊厳や人間中心、公平性、透明性、アカウンタビリティ、プライバシー等の価値が挙げられている<sup>3</sup>。これらの原則に共通するのは、基本的人権と民主的価値を守ることの重要性である<sup>4</sup>。

G7 広島サミットを受けて発足した AI の国際的な議論を行う広島 AI プロセスは、生成 AI の影響を議論し、AI の責任ある展開に向けた原則の確認を目的に掲げている。AI がもたらすリスクには透明性の低下 (いわゆるブラックボックス問題、誤回答等)、入力や訓練データの偏り、不完全な訓練に起因する不適切な出力結果 (公平性に係る問題) 等、AI 技術固有の特徴により生じるリスクがある。その他にも悪用、プライバシーリスク、秘密漏えい、セキュリティ等、既存の情報通信技術で生じているリスクが AI によって助長されかつ高度化されるリスクもある。特に、横断

---

<sup>1</sup> 生成 AI とは複数の機械学習技術の組み合わせによって構成されているが、基盤モデルをつくることと個別の目的や分野への調整を行うことで、人間に近い自然な文章等のコンテンツの生成能力を持ったことが特筆に値する。

<sup>2</sup> 本提言の前身として 2023 年 3 月に東京大学未来ビジョン研究センターで公開した政策提言においても、人間-AI 協調の研究、政策と実践の推進が含まれる。<https://ifi.u-tokyo.ac.jp/news/15309/>

<sup>3</sup> Anna Jobin, Marcello Ienca & Effy Vayena : The global landscape of AI ethics guidelines, Nature, Machine Intelligence, 1, 389-99, 2019

<sup>4</sup> G7 広島サミットのコミュニケにおいても、法の支配、普遍的な人権、ジェンダー平等及び人間の尊厳といった共通の価値を促進し、人間中心かつ包摂的で強じんな世界の実現が確認されている。

[https://www.mofa.go.jp/mofaj/ecm/cc/page1\\_001700.html](https://www.mofa.go.jp/mofaj/ecm/cc/page1_001700.html)

的な法制である著作権法、個人情報保護法に関しては、これまでも見直しが進められてきているが、生成 AI の登場により著作権や肖像権といった第三者の権利を侵害するリスクも社会的に議論されており<sup>5</sup>、知的財産に関わる検討が急がれる<sup>6</sup>。

一方、AI の利用を推進していくにあたり、現行制度が AI の設計・開発・運用・利用等の障害となる場合もある。例えば医師法等の業法や既存の安全規制制度等に関しては、引き続き安全の確保を図ることは重要であるものの、AI を活用した新サービスの適用関係が不明確であるため、イノベーションが阻害される可能性があることが課題となる。さらに、将来的な失業に対する不安やいわゆる AI 脅威論の観点から、特に海外においては、予防原則（Precautionary Principle）の視点を AI に適用し、リスクが明確でない段階から予め規制の枠組みに置くべきとの議論があるが、技術がもたらす将来的で不明確なリスク（懸念）に対して強すぎる規制は企業への参入障壁となりイノベーションを阻害する可能性があるほか、参入できても結果として市場の寡占化につながる恐れがある。法律上の課題の他にも教育現場、人材育成等での生成 AI 利用の在り方と既存体制や制度との調和が求められている<sup>7</sup>。

これらの論点を議論するためには、すでに確立している国際的かつ学際的なマルチステークホルダーの場を活用して、生成 AI がはじめとする AI がもたらす可能性やリスクについて迅速かつ継続的に議論する専門家や機関、市民を支援する体制を構築すべきである。人類や環境に対する新たなリスクや被害に備え、未来のために G7 として国を超えた協調の枠組みを構築することが必須である。

## 2. 枠組み間の「相互運用性」推進と AI 規律の相互尊重

G7 広島サミットのコミュニケ、G7 デジタル・技術大臣会合の閣僚宣言で言及されている「相互運用性」とは主に「標準」と「枠組み間の相互運用性」として議論される。

国内あるいは国際標準により、組織や国をまたいで開発、運用、利用が行われる AI システムにおいても相互比較、相互運用が可能となる。そのためには AI に関連する用語、基本概念や AI ガバナンス・マネジメント等に関する各種の標準の議論とコンセンサスが必要であり、現在 ISO/IEC<sup>8</sup>や IEEE<sup>9</sup>、NIST<sup>10</sup>、CEN-CENELEC<sup>11</sup>等の標準化団体によって標準化が推進されている。

一方、枠組み間の「相互運用性」は G7 デジタル・技術閣僚宣言の附属書 5 で「AI ガバナンスの枠組み間の相互運用性」との言葉で紹介される<sup>12</sup>。国内プロセスの相互調整を伴う相互承認や十

---

<sup>5</sup> 国内外において様々な議論が行われているが、日本では例えば生成 AI 開発に対して新聞協会ら 4 団体が著作権保護策の検討を求める声明を公開している他、クリエイターの団体からも声が上がっている。

<sup>6</sup> 日本においては内閣府知的財産戦略推進事務局が「知的財産推進計画 2023」を 2023 年 6 月に公開し、生成 AI 時代における知財の在り方の議論を開始している。 [https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/3kai/chizai2023.pdf](https://www8.cao.go.jp/cstp/ai/ai_senryaku/3kai/chizai2023.pdf)

<sup>7</sup> 文章生成 AI や画像生成 AI を用いて児童生徒/学生が学校の課題を行ってしまうことに対し、日本では各種ガイドラインが公開されている他、課題の出し方や評価の仕方についても工夫が必要であることが指摘されている。

<sup>8</sup> ISO/IEC JTC 1/SC42 では、これまでに AI に関連する用語・基本概念（ISO/IEC 22989）や AI ガバナンス（ISO/IEC 38507）等、すでに 20 の国際標準が作成・公表されており、また 31 の国際標準化が議論中である。

<sup>9</sup> IEEE の 7000 シリーズ他では AI の実務上の課題を対象とした規格が議論されている。

<sup>10</sup> 米 NIST は ISO/IEC のマネジメント基準や概念、OECD の AI 勧告と相互運用可能な AI の統一リスクベースフレームワークを議論し、関係性を整理している。

<sup>11</sup> CEN/CENELEC では欧州における AI の規格が議論されている。

<sup>12</sup> G7 群馬高崎デジタル・技術大臣会合の開催結果、附属書 5「AI ガバナンスのグローバルな相互運用性を促進

分性認定とは異なり、「枠組み」というレイヤーで相互運用性を促進することにより<sup>13</sup>、各国・各組織の AI に対する規律や対応が並存、協調できる。つまり、各国・地域、組織や応用分野による制度設計・運用上の自由度を与え、制度的、社会的、慣習的、文化的な文脈等にも配慮したうえで、グローバルな共通目的を効率的に達成することが期待できる<sup>14</sup>。

なお新技術に対して常に新しい規律が必要になるわけではない。例えば AI によってプロセスの効率化が起こるととどまり、AI システムやサービスプロダクトには本質的な変化がない場合に生じるリスクに対しては、既存の規律の解釈や拡張で対応可能である。むしろ不必要で過度な規制や、複数の規律による二重管理はイノベーションを阻害し、結果として社会全体としての損失につながる可能性がある。

ただし既存法の適用関係が AI を活用した新サービスに対して不明確でイノベーションを阻害している場合や、人が介在しないために責任主体が不明瞭化する等全く新しい特徴を持つ AI システムやサービスが提供される場合は、実証実験等で事前検討を行う仕組みを構築する等、社会全体としての対応を行うべきである。さらにはマルチステークホルダーの議論に基づいて、リスクの大きさや頻度、影響が及ぶ範囲等を考慮に入れながら、必要に応じて新たな立法を含めた対応を検討するべきである。また、何らかの新たな規律を導入する場合であっても、AI 技術の進展の速度を踏まえると、技術の変化に応じて規律を見直すアジャイルな体制を検討すべきである。

さらに、いずれの規律であっても、効果を発揮するためには誰でもアクセス可能かつ理解可能であるという透明性や公平な機会を提供するべきである。これにより中小企業や新興企業による参入障壁が低くなり、さらなるイノベーションを加速することにつながる。

---

等するためのアクションプラン」。<https://www.digital.go.jp/news/efdaf817-4962-442d-8b5d-9fa1215cb56a/#declaration>  
<sup>13</sup> 例えば、OECD が ISO や IEEE、NIST の標準、欧州 AI 法案、欧州評議会の影響評価法（HUDERIA）等を「相互運用枠組み（interoperability framework）」として項目比較を行っている。OECD.AI work promoting interoperability of AI risk management frameworks, IGF Policy Network on AI meeting #4, 18 July 2023, [https://www.intgovforum.org/en/filedepot\\_download/282/25999](https://www.intgovforum.org/en/filedepot_download/282/25999). OECD は他にも AI システムの分類や、AI システムのライフサイクルを分類し、比較できる「枠組み」を提供している、<https://doi.org/10.1787/2448f04b-en>. 日本では経済産業省が「AI 原則実践のためのガバナンス・ガイドライン」を公開している。<https://www.meti.go.jp/press/2021/01/20220125001/20220124003.html>

<sup>14</sup> ただし、AI ライフサイクルが国や組織をまたぐ場合、どこかの国や組織の法的枠組みに合わせる必要が出てくる。

## コラム：規律の多様性

各国・地域や各応用分野によって、採用される規律の在り方は多様である。表 1 に規律を形成する主体と、その規律のエンフォース（強制的な執行）の有無を整理した。国家はエンフォースしないが、市場や投資、評判等、国家以外が人々の遵守を駆り立てる規律もある<sup>15</sup>。進展が早く国や組織をまたいで展開する技術においてはこれらの規律の方が効率的な場合もあるが、そのためには人々のリテラシー向上等が必要となる。

表 1 規律の分類<sup>16</sup>

|               |                  | 国家によるエンフォース（強制的な執行）の有無 |                            |
|---------------|------------------|------------------------|----------------------------|
|               |                  | 国家がエンフォースする            | 国家はエンフォースしない               |
| 主体<br>規律を形成する | 国家               | 法令・一部ガイドライン            | ガイドライン <sup>17</sup> ・行政指導 |
|               | 産業界・個別企業         | 法令を背景とする標準化（強制規格）      | 業界ガイドライン・社内ポリシー・業界標準       |
|               | それ以外（市民団体・学術団体等） | 慣習法                    | 市場・投資・モラル・規範・学会基準・慣習・評判    |

### 3. AI の展開における責任主体と対応策

AI の責任ある展開を推進するためには、AI の開発から利用に係る一連のサプライチェーン/ライフサイクルにおける各主体それぞれが、自ら選択した規律を説明し、その帰結に責任を取る（アカウンタブルである：答責性がある）ことが重要である。そのため責任主体の明確化と、取るべき対応策を検討するべきである。なお、このようなサプライチェーンにおける各主体の責任関係は、AI システムに限らずこれまでの大規模工学システムでも同様の問題が存在し、これまでも既存法で対応されてきているが、そのような中、特に、ブラックボックス的性質など AI 特有の特徴を踏まえた新たな対応が求められることが考えられる。

本稿では、AI ライフサイクル（設計、開発、提供、利用、廃棄）の観点から、主な関係者の説明を表 2 に示し<sup>18</sup>、責任の向きを図 1 にまとめた<sup>19</sup>。

<sup>15</sup> 例えば市場による競争、ESG 投資、炎上リスク等。

<sup>16</sup> 藤田友啓「ソフトローの基礎理論」p.6 表の枠組みを利用して作成。

<sup>17</sup> 日本は強制力のないガイドラインでも事業者には影響力が強いとされる。

<sup>18</sup> 本稿では AI ライフサイクルに沿って関係者の責任を整理しており、関係者すべてを網羅しているわけではない。例えば本稿では触れていないが、業界団体や学術団体等による提言や社会発信も AI ガバナンスに重要な役割を担う。

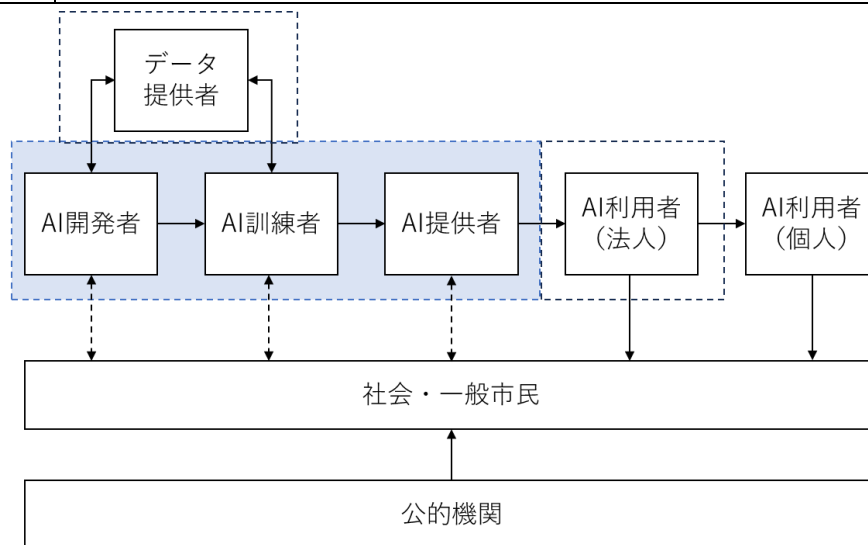
<sup>19</sup> AI の利活用に関与が想定される主体は、総務省「AI 利活用ガイドライン」

([https://www.soumu.go.jp/main\\_content/000624438.pdf](https://www.soumu.go.jp/main_content/000624438.pdf)) で分類されている。図 1 はガイドラインを参照しつつ、生成 AI の開発、利用を考慮に入れて作成した。また、「利活用ガイドライン」では利用者のうち最終利用者は「ビジネス利用者」と「消費者的利用者」に分類されるが、本稿ではビジネス利用者を AI 法案の「利用者」

（'user' means any natural or legal person, public authority, agency or other bodies using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity）の定義を参照し、利用者（法人）と利用者（個人）に分類した。

表2 主体の説明

| 主体         | 説明  |
|------------|---|
| AI 開発者     | AI アルゴリズムやモデルを開発する組織（あるいは個人）  |
| AI 訓練者     | 開発された AI アルゴリズムやモデルに前処理したデータを追加学習させる（ファインチューニングを含む）組織（あるいは個人）   |
| データ提供者     | AI アルゴリズムやモデル、あるいは AI をファインチューニングさせるために使うデータセットを提供する組織（あるいは個人） <sup>20</sup>  |
| AI 提供者     | 市場に出回る AI システムやサービスを提供する組織（あるいは個人）  |
| AI 利用者（法人） | AI 提供者から AI システムやサービスの提供を受け、自組織のサービスとして整え、自組織の構成員あるいは他組織/他者の利用を可能とする運用を行う組織   |
| AI 利用者（個人） | AI 提供者、あるいは AI 利用者（法人）が提供する AI システム、サービスを最終的に利用する個人。AI システム、サービスを業務で使う場合もあれば業務以外で使う場合もありうる                                  |
| 社会・一般市民    | 他組織/他者の利用する AI サービスやシステムにより付随的に自らの権利・利益に影響を受ける可能性のある組織（あるいは個人）。社会全体に AI サービスやシステム利用の影響が及ぶ場合、AI 開発者・訓練者・提供者・利用者やデータ提供者も含まれうる |
| 公的機関       | 国や地方自治体等、政策を策定、執行する組織 <sup>21</sup>   |



AI開発者・AI学習者・AI提供者は同一組織の場合もあれば、別々の組織の場合もある。データ提供者とAI開発者が同一組織の場合もある。自社製品を利用する場合は、AI開発者/AI提供者と利用者（法人）が同一組織となる場合もありうる。

—————▶ 責任の向きを示す。例えばAI開発者はAI学訓練とデータ提供者に対して責任を負う（表3）。

-----▶ 企業の社会的責任や、社会からのフィードバック等の間接的な責任の向きを示す。

図1 AI システムに関する主体と責任の向き

<sup>20</sup> ここでは作成、公開されたデータを収集して提供する組織や個人を想定しているが、他にも AI 開発者や訓練者がウェブ上をクロールして収集するデータや、AI 利用者の振る舞いから収集されるデータ等もある。

<sup>21</sup> 公的機関が AI 開発者、提供者、利用者（法人）の立場にもなりうる。



図1で示しているように、同一組織でAIサービスの開発から提供、場合によっては利用まで行う企業もあれば、異なる組織にサプライチェーンが複雑にまたがる場合や複数のAIシステムを組み合わせて提供されるサービスに組み込まれることもある(図2)。このように、一組織で複数の役割を担う場合は、すべての役割で求められる責任を果たす必要がある。

(1) AI開発から提供まで同一組織が担う例



(2) AI開発から提供まで異なる組織で担われている複雑な例

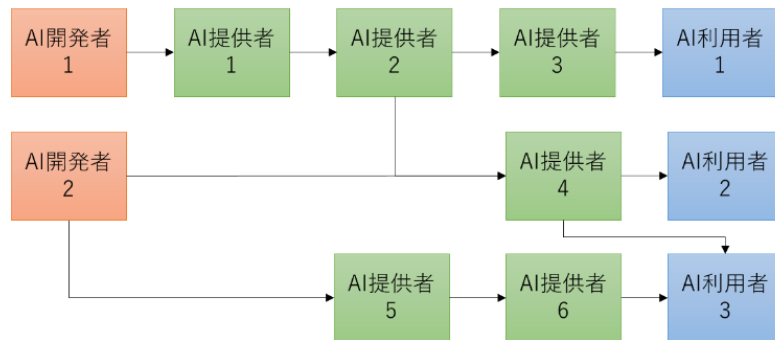


図2 AI開発から利用までのサプライチェーンの例

図1は各責任主体と、責任を負う対象を責任の向きとして矢印で示した。各矢印において考慮すべき項目例と責任なる対応の事例を表3に示す。なお、表3には示していないが間接的な責任として、点線の矢印で示しているように、AI開発者、訓練者、提供者等が社会や一般市民に対してビジョンの共有や社会発信等の情報提供を行うといった社会的責任や、逆に社会・一般市民側がAI開発者、訓練者、提供者等に対して適切なフィードバックや監視を行うことも期待されている。

表3 責任主体と主な責任ある対応の例

| 責任主体           | 責任対象              | 考慮すべき項目例   | 責任ある対応例  |
|----------------|-------------------|------------|--|
| AI 開発者         | AI 訓練者            | 透明性        | 合理的な範囲での情報開示や問題共有等 <sup>22</sup>                   |
|                | AI 提供者            | フィルタリング    | 不適切な入出力の抑止等  |
| AI 訓練者         | データ提供者            | 対価         | 適切な対価の支払い等 <sup>23</sup>                           |
|                |                   | 情報管理       | 著作権、個人情報等の適切な対応等                                   |
|                | AI 提供者            | 透明性        | 合理的な範囲での情報開示や問題の共有等                                |
| データ提供者         | AI 開発者<br>AI 訓練者  | 品質保証       | 提供するデータの品質保証等                                      |
| AI 提供者         | AI 利用者<br>(法人/個人) | 透明性        | 合理的な範囲での情報開示や問題の共有等 <sup>24</sup>                  |
|                |                   | 利用環境の周知    | AI 利用に適した環境やアップデート情報の周知等                           |
| AI 利用者<br>(法人) | AI 利用者(個人)        | 透明性        | 合理的な範囲での情報開示や問題共有等                                 |
|                |                   | 利用環境の周知    | ガイドライン策定や問題発生時の対応、AI 利用に適した環境やアップデート情報の周知等         |
|                | 社会・一般市民           | 適正利用       | 悪用・誤用等の不適切な利用をしない、適正な環境で利用する等                      |
| AI 利用者<br>(個人) | 社会・一般市民           | 適正利用       | 悪用・誤用等の不適切な利用をしない、適正な環境で利用する等                      |
| 公的機関           | 社会・一般市民           | 市場の監視      | 寡占・独占の防止、契約関係における公正性の担保、人権や社会的法益の保護等               |
|                |                   | セーフティネット   | 雇用の安定や社会的弱者の保護等                                    |
|                |                   | 制度設計       | 官民共同でイノベーション推進とリスク緩和の両立、リソースの乏しいスタートアップ企業のリスク対応支援等 |
|                |                   | 教育・リテラシー向上 | 官民共同で社会・一般市民の AI リテラシーの向上等                         |
|                |                   | 国際的整合性の向上  | 官民共同で健全な対外取引の保障、枠組み間の相互運用性の向上等                     |

表3ではAI開発者から利用者までの責任の向きと対応策を示したが、関係者間で採用される規律や重視される価値は、各国・地域や応用分野が抱える課題や目指す社会像に基づく多様性を

<sup>22</sup> 世界貿易機構（WTO）では取引の安心、安全を確保するために「ソース・コード／アルゴリズムの移転・開示要求の禁止」、「個人情報保護／消費者保護」、「ICT製品の暗号開示要求の禁止」を挙げている（[https://www.mofa.go.jp/mofaj/ecm/it/page22\\_003579.html](https://www.mofa.go.jp/mofaj/ecm/it/page22_003579.html)）。AIにおける透明性を考える際も、これらの国際ルールを考慮する必要がある。

<sup>23</sup> 生成AIに関しては特にデータ作成者に、生成された情報の価値が適切に還元されないことが問題となっている。

<sup>24</sup> あるAIの運用停止後も、そのAIの出力結果は随所で使われ続けることが想定される。そのため、AIの出力結果を用いる別のシステムが社会で利用されている限りは、AIの透明性を確保することが重要となる。

尊重すべきである<sup>25</sup>。しかし、各主体間でパワーバランスが異なる場合に備え、適正な取引が遂行されるかどうかに関する監視は必要となる。その際、AI 倫理や AI に関するガイドラインは公正な取引を進める根拠としても重要となる。

それぞれの主体の対応は、事業者間取引と、事業者と消費者間取引の場合で整理できる。

## 事業者間取引での場合

### 1. 事業者間で取るべき対応

AI 開発者から AI 提供者（図 1 における青枠）が同一組織ではない場合、主体間で契約<sup>26</sup>や社会的な約束（コミットメント）<sup>27</sup>を結ぶ。その際、AI モデル開発やデータ提供/訓練は、従来のソフトウェア開発とは異なる点を考慮に入れ締結することが重要である<sup>28</sup>。契約は主体間の責任の所在を議論するだけでなく、訓練データや AI モデルの知的財産権を考えるうえでも重要である。

### 2. 公的機関が取るべき対応

AI 開発者、AI 訓練者、AI 提供者間で結ばれる契約の多くは、スタートアップ企業（AI 開発者・AI 訓練者）と大企業（AI 提供者）間で行われることもある。そのため当事者間同士のインセンティブ対応で責任を担えない場合や、外部性が発生するため私的交渉によっては最適解が得られない場合、公的機関等は表 3 に示す市場の監視<sup>29</sup>、セーフティネット、制度設計、リテラシー向上や国際的整合性の向上等の対応を取る必要がある。

## 事業者と消費者間取引の場合

### 1. AI 提供者が取るべき対応

図 1 のうち AI 提供者と AI 利用者（個人）間の取引、さらに他者の AI 利用により付随的に権利・利益に影響を受ける可能性のある組織や人を保護、救済が必要になる部分は、必要な情報を AI 開発者や訓練者から得て、AI 提供者が対策を取るべきである。対策例を、事前・事後と技術・組織的対策に分類して表 4 に整理した。

---

<sup>25</sup> 例えば多様性を尊重する上でも人種、性別、年齢、学歴、ライフスタイル等の切り口があり、AI が応用される分野等によって重視される焦点も異なる可能性がある。また技術的な公平性の指標にも「一人の個人が集団属性に関係なく、他の個人と同様に扱われている状態を指す individual fairness」や「集団の中で男女等のセンシティブグループ間の公平性を指す group fairness」等がある。

<sup>26</sup> 経済産業省は AI・データの利用に関する契約ガイドラインを公開している

([https://www.meti.go.jp/policy/mono\\_info\\_service/connected\\_industries/sharing\\_and\\_utilization.html](https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization.html))。また経済産業省と特許庁が「研究開発型スタートアップと事業会社のオープンイノベーション促進のためのモデル契約書」も公開している。

<sup>27</sup> 2023 年 7 月にアメリカ政府と、OpenAI や Google 等生成 AI 開発を手掛ける 7 社は AI の安全性を確保する自主規制を導入することを約束（コミットメント）することで合意した。

<sup>28</sup> 例えばベンダーがクライアントの要求する機能の完成を保証する（性能保証）は深層学習においては困難である（モデルの内容や性能がデータの品質に依存するためであり、未知のデータが学習されたり、入力されたりすることにより開発時には想定しなかった挙動を示すことがあるため）。そのほか、AI モデルの判断や結果を事後的に検証することが困難な場合もある。

<sup>29</sup> 契約者間で契約が公正であるかを確認する内部監査や外部監査等も重要となる。

表4 AI提供者が取るべき主な事前・事後のリスク対策例

|             | 事前・予防的な対策例   | 事後・信頼性確保の対策例  |
|-------------|--|---|
| 技術的対策       | 不適切な生成物の悪用や誤用を抑制するAIの開発、AI実装部以外で安全性を確保する仕組みの構築 <sup>30</sup> | 事故や事件が生じた原因の解明  |
| 組織マネジメント的対策 | 社内ポリシーの策定・見直しや構成員に対するリテラシー教育の実施                              | データ、訓練方法、利用モデルや人によるモニタリング等を合理的な範囲で実施、記録して事後検証できる体制を構築 |

## 2. AI利用者（個人）が取るべき対策

AI利用者は適切なリテラシーを獲得することで、誤情報や悪用に惑わされずにAIサービスやシステムを適切に利用できる。さらにAI提供者を適切に評価できるようになることで、法的な規制とは別の規律（市場・投資・評判等）によってガバナンスを利かせることが可能になる。ただし適切な評価には、AI提供者からの情報提供も不可欠である。

一方、AI利用者がAIサービスやシステムによる事件や事故の被害者、あるいは無自覚的に加害者となってしまった場合には、補償や救済措置、サポートを行う組織への速やかな情報共有と被害拡大の防止に努める等の事後対応が求められる。

## 3. 公的機関等が取るべき対策

補償や救済措置は、AI提供者あるいは保険等の市場商品が提供する場合がある。しかし、そのためには損害や事故が生じる確率が理論的、経験的に推定できなければならない。

新サービスや新技術等のリスク推定ができない場合、実証実験等でリスクを検討する仕組みを構築する等、一組織、企業だけではなく社会全体としてのリスク対応が必要となる。また、事故時の原因究明の支援や公的専門委員会の設立等も必要となる。さらに一定の確率で損害や事故が生じるが、大きな社会的便益を見込めるようなAI利用については、被害者支援の基金の設立等の補償制度の構築も選択肢として考えられる。

AIライフサイクルが国や組織を超えて展開することを踏まえ、このような各主体の責任と対応策の透明性を高める議論を推進していくべきである。

<sup>30</sup> AIモデルやアルゴリズム等の実装部以外にも、基本制御や非常時に停止する他のシステムも含めてAIシステムとして安全確保していくといった仕組みも考えられる（「AI利活用原則の書く論点に対する諸説」、AIネットワーク社会推進会議 p.17 [https://www.soumu.go.jp/main\\_content/000637098.pdf](https://www.soumu.go.jp/main_content/000637098.pdf)）

今後の課題と展開、日本国内の AI ガバナンスに向けて

日本は第三次 AI ブームに端を発する AI のガバナンスに関する議論を早くから開始してきた。2016 年の G7 伊勢志摩サミットの香川・高松情報通信大臣会合では AI 開発の原則案<sup>31</sup>を提案し、この案はその後の AI ガバナンスの必要性に関する国際的な議論へと繋がった。

2023 年現在、日本は G7 広島サミットだけではなく、経済協力開発機構 (OECD) デジタル経済政策委員会 (CDEP)<sup>32</sup>、AI に関する国際的パートナーシップ (Global Partnership on AI)<sup>33</sup> の議長国、さらに国連が主催するインターネット・ガバナンス・フォーラム 2023<sup>34</sup> の開催国 (2023 年 10 月、京都) でもある。このような点も踏まえ、本提言は日本が G7 議長国として AI ガバナンスの議論を整理し、リードするにあたって提案すべき基本方針と論点を整理した。

G7 以外にも、AI の責任ある展開に関する議論は様々な国際的なフォーラムでも行われている。例えば、日本がメンバー国として参加している国際連合教育科学文化機関 (UNESCO) や OECD、GPAI 等の国際機関が提唱する原則や活動、日本がオブザーバ国として参加している欧州評議会 (Council of Europe) の AI 条約<sup>35</sup>、さらにはソニーや東京大学がメンバーとして参加している AI 研究の非営利団体である AI のパートナーシップ (Partnership on AI)<sup>36</sup> による提言や活動がある。広島 AI プロセスにおいては、これらの国際的な議論も取り込み、さらには日本国内の産学官民による原則や活動に関しても積極的に発信を行うことで、国際的な貢献を果たしていくことが期待される。

最後に、本政策提言は G7 広島 AI プロセスという国際的な議論に資することを目的と掲げているが、日本国内の体制やフレームワークの整備も喫緊の課題となっている。本提言は国内の論点の整理や方針の明確化にも応用でき、今後の議論の一助となれば幸いである。

#### 本政策提言公開主体とサポート

本政策提言に関する調査は東京大学未来ビジョン研究センターフラッグシッププロジェクト「AI 社会における未来ビジョンのデザイン」、「東京大学若手研究者自立支援制度 (卓越研究員)」のほか公益財団法人トヨタ財団 D18-ST-0008「人工知能の倫理・ガバナンスに関するプラットフォーム形成」の支援を得て実施した。また、東京大学未来ビジョン研究センター AI ガバナンスプロジェクトにおける企業や行政の方々との共同研究の成果も本提言の一部に含まれる。

---

<sup>31</sup> 国際的な議論のための AI 開発ガイドライン案、総務省 AI ネットワーク社会推進会議、[https://www.soumu.go.jp/main\\_content/000490299.pdf](https://www.soumu.go.jp/main_content/000490299.pdf)

<sup>32</sup> OECD-CDEP: <https://oecdgroups.oecd.org/Bodies/ShowBodyView.aspx?BodyID=1837&Lang=en>

<sup>33</sup> GPAI: <https://gpai.ai/>

<sup>34</sup> IGF 2023: <https://www.intgovforum.org/en/content/igf-2023>

<sup>35</sup> Council of Europe, Committee on Artificial Intelligence: <https://www.coe.int/en/web/artificial-intelligence/cai>

<sup>36</sup> PAI: <https://partnershiponai.org/>

附録：フィードバックを寄せていただいた方一覧（五十音順）

時間や所属組織の関係上、お名前の記載が叶わなかったが、有益なフィードバックを寄せていただいた方がいることも付け加え、フィードバックを寄せていただいた方々に感謝申し上げます。

荒堀淳一 富士通株式会社 AI 倫理ガバナンス室 室長  
市川 類 一橋大学イノベーション研究センター 特任教授  
江川尚志 産業技術総合研究所 特定集中研究専門員  
江間有沙 東京大学未来ビジョン研究センター准教授／理化学研究所客員研究員（編集）  
大屋雄裕 慶應義塾大学法学部教授  
岡田 淳 森・濱田松本法律事務所 弁護士  
落合孝文 渥美坂井法律事務所・外国法共同事業 プロトタイプ政策研究所所長・弁護士  
スマートガバナンス株式会社代表取締役共同創業者  
柿沼太一 STORIA 法律事務所 弁護士  
加藤尚徳 次世代基盤政策研究所（NFI）研究主幹・理事（\*）  
川原圭博 東京大学 大学院工学系研究科 教授  
北村 弘 CDLE（Community of Deep Learning Evangelists）AI リーガル Gr リーダー  
工藤郁子 大阪大学 社会技術共創研究センター 招へい教員  
栗林 諄 東京大学公共政策大学院 修士課程/リサーチアシスタント  
栗原 聡 慶應義塾理工学部 教授/共生知能創発社会研究センター センター長  
穴戸常寿 東京大学大学院法学政治学研究科教授  
城山英明 東京大学未来ビジョン研究センター教授  
実積寿也 中央大学総合政策学部教授（\*）  
杉村領一 産業技術総合研究所 上席イノベーションコーディネータ  
鈴木晶子 京都大学名誉教授/理化学研究所客員主管研究員  
高木幸一 KDDI 株式会社技術渉外部企画グループリーダー  
武田英明 国立情報学研究所情報学プリンシプル研究系 研究主幹/教授  
田丸健三郎 日本マイクロソフト株式会社 ナショナルテクノロジーオフィサー  
中川裕志 理化学研究所・革新知能統合研究センターチームリーダー（\*）  
成原 慧 九州大学大学院法学研究院 准教授  
羽深宏樹 京都大学法学研究科特任教授／スマートガバナンス株式会社代表取締役 CEO  
原山優子 東北大学 名誉教授  
福岡真之介 西村あさひ法律事務所  
舟山 聡 rinna 株式会社 Chief Legal Officer  
古川直裕 株式会社 ABEJA 弁護士  
松尾 豊 東京大学工学系研究科 教授  
渡部俊也 東京大学未来ビジョン研究センター 教授

※ 本政策提言の骨子作成に貢献いただいた方々は（\*）で示す。