

# 「AIガバナンスを巡る国際的な潮流：AI原則をどのように実践に落とし込むのか」開催レポート

2023年3月7日、東京大学未来ビジョン研究センターは「AIガバナンスを巡る国際的な潮流：AI原則をどのように実践に落とし込むのか」と題したセミナーを開催した。東京大学国際学術総合研究棟4階SMBCアカデミアホールとオンラインのハイブリッド開催で、会場には30名強が集まった。

進化の著しい人工知能（AI）をめぐっては、ここ10年余り、社会との関わりを軸にAI開発の在り方が議論されてきた。特にいま、生成AI（Generative AI）や大規模言語モデル（LLM）などが登場し、議論は新たな局面を迎えたと言える。こうしたAIの最新技術が社会に与える影響の大きさが指摘されており、AIガバナンスの必要性に議論の焦点が当てられている。

本イベントは、グローバルなAIガバナンスの潮流を概観し、その中で日本が果たすべき役割について掘り下げていくことを主目的に、AIガバナンスの議論をリードするゲストを海外から迎えて実施された。

## 【パネリスト登壇者】

Gregory C. Allen: Director of AI governance project, Center for Strategic and International Studies

Rebecca Finlay: CEO at Partnership on AI

David Leslie: Director of Ethics and Responsible Innovation Research at The Alan Turing Institute and Professor of Ethics, Technology and Society, Queen Mary University of London

Karine Perset: Head of AI Unit and OECD.AI, OECD Digital Economy Policy Division - OECD

城山英明：東京大学未来ビジョン研究センター 教授

ファシリテーター：江間有沙：東京大学未来ビジョン研究センター 准教授

## AIガバナンスについての議論の潮流

最初に、東京大学未来ビジョン研究センターの城山英明氏より、なぜ今AIガバナンスを議論するのか、本イベントの狙いについて紹介があった。城山氏は、グローバルでのAIガバナンスの議論の潮流を掴むとともに、特にG7サミットが日本で開催されるタイミングで日本がどんな役割を担えるのかを考えていきたいと述べた。

パネリストからのプレゼンテーションはグレゴリー・アレン氏から始まった。アレン氏はCenter for Strategic and International Studies (CSCI) AIカOUNシルのディレクターである。

アレン氏はまず、どういうやり方をすれば堅牢な形でこれらの原則を実践できるのか、AI原則から実践、実際に利活用するために柔軟なアプローチを検討していくべきと自身の立場を述べた。

そもそも「AI原則を守ること」と「それらの原則を実践として示すこと」はまったく別物であり、その実践の場が、何千、何十万人の社員がいる大きな組織で何十億ドルもの様々なオペレーションを行っている場合と、より中小の組織では大きく異なる。

アレン氏はAIについて、その有用性とリスクの2つの面があることを忘れてはいけないと述べた。人口が減少する中、継続的に経済成長を続けるには生産性を高める必要があり、AIはそのためのテクノロジーとして大きく期待されている。反面、AIの利用における様々なリスクが存在するのも事実である。たとえば、オランダではアルゴリズムリスクの事例が起きている。児童手当の申請システムに人種差別を助長してしまうような不適切なアルゴリズムが使用され、何千という人たちに影響を及ぼした。

AIシステムを使用する場合、システムに入れておくべき価値をしっかりと見極めなければならない。特に民主国家においては共有されている価値は不可欠となる。その上で、AI原則を実践するために、原則を組織の手続きや技術的な手続き、ときには政府政策も変えていく必要がある。

アレン氏はここで「規制とガバナンス」「プラクティス」「技術戦略」の3つのキーワードを挙げる。必ずしもAI特有の規制を一から作らなければいけないということではない。たとえば商業用航空機のオペレートに機械学習システムを入れる場合、AI特有の規制はまだ存在しないかもしれない。しかし、ある部分では既存の安全規制が有効になるだろう。従来のソフトウェアとAIシステムでは異なる性質があるので、新たな規制を検討したり、法律を検討したりすることは重要だが、すでにあるソフトウェア向けの規制が有用だということは覚えておくべきだろう。

プラクティスの収集・普及にはいま様々な組織が取り組んでいる。これに関してはCSIS AIカウンスルもペーパーを出す予定になっている。また、様々なフェーズの技術戦略があり、すべての課題に対し、政府、民間企業を含めて議論をしていく必要があると述べた。

プレゼンテーションの最後に、アレン氏は日本に期待する役割として、「拘束力のないガイドライン」「拘束力のある規則」「規格そのものの開発」への貢献を挙げた。その中でも特に規格の開発が鍵となる。それには、企業、政府、学会などすべての機関、組織が規格開発のプロセスに参加し、サポートすることが重要である。

続いて、Partnership on AIのCEO、レベッカ・フィンレイ氏がプレゼンテーションを行った。Partnership on AIは大手テクノロジー企業6社のAIリサーチのトップにより2016年に設立されたグローバルなAIコミュニティであり、現在は100以上の非営利団体、学術界、産業界、メディアなどのパートナーが参加している。

健全なAIシステムの構築にはいくつかのアプローチがあるが、その1つとして、フィンレイ氏はPAIのマルチステークホルダーアプローチに注目している。このアプローチは、集団の中で作られる規範がベスト（もしくはベター）プラクティスに対して大きな影響を与えるという理解に基づいている。コミュニティとして規範を設定していく際に、カナダや日本のような小さな国、ミドルパワーが果たす役割も影響力が大きい。

マルチステークホルダーコミュニティとして影響力を持つためには、目標は「コミュニティの形成」「リテラシーの普及」「ポリシーイノベーションの推進」「変化の促進」の4つである。コアとなるのは、共同規格を開発し、コミュニティとして、台頭するAIの開発およびイノベーションのトレンドへの対応することである。また、これを個人や組織を含めた国際的な集まりを通じて行うことが重要である。マルチステークホルダーが集まって、台頭する技術に関して、その有用性とリスクの双方を見ていく。どうやってベストプラクティスで世界中の規範を策定していくのか、どうやって実践に落とし込むのか、どうやって開発していくのか、リソースをどうやってまとめていくのか、といった課題を専門家グループも含めて検討していくことになる。

また、LLM（large language model、大規模言語モデル）などの技術の安全性を慎重に見ていく必要がある。開発者、研究者、一般の市民を含め規格化に関して、責任のある形で考える。グローバルな舞台で日本がこの分野でリーダーシップを発揮することを期待すると述べた。

OECD AIユニットを率いるカリン・パーセット氏は、AIという技術が社会の中うまく広がっていくための環境を整備するタイミングはいまだという。OECDは2019年にAI原則を策定したが、この原則を実践に落とし込むための試みの1つとして、OECD AI Policy Observatoryを立ち上げた。これは、データ、トレンドの動き、政策、エビデンスなど、様々なリソースを集めるオンラインプラットフォームであり60カ国以上の事例をカバーする。

OECD AIユニットが考えるリスク管理のアプローチは4つの流れになる。まずリスクのコンテキストを定義し、次にAIリスクを評価、3つ目にAIリスクへの対処、4つ目がAIリスクのガバナンスである。このガバナンスの部分が一番重要で、インシデントのモニタリングツールなどが含まれる。つまり、AIシステムのリスクバリューチェーンを組み、その上で、相互運用可能な形でリスク評価をしていくということになる。

ここでいう相互運用性の効用は、実際にこの枠組みを実施する人たちの負荷を下げ、枠組みの有効性や効率性、執行可能性を高めることである。政策立案の上流部分で様々な国々に同意してもらう枠組みを作り、相互運用性をもって民主主義・市場経済の国々がお互いに話をできるようにする。ステークホルダーとして非加盟国も含め、大きなコンセンサスを作ることがOECDの役割である。

AIリスクのガバナンスという点では、国のレベル、地域のレベルにとどまらず、コーポレートガバナンスの重要性を指摘している。焦点は、単に法律の立案だけではなく、AIリスクの管理をオペレーションブルなものにする具体的な行動規範の策定が有用だということだ。パーセット氏はここで、OECDがリリースしている「責任ある企業行動のためのOECDデュー・ディリジェンス・ガイダンス」を例に出し、こうした既存のガイドラインをAIに特化させた形で展開することで、ガバナンスシステムをゼロから作らずとも短期的にAIリスクに対応できるだろうと述べた。

また、日本の役割として、個人ベースでも、もっと国際的な議論の場に参加してほしいと語った。

パネリストのプレゼンテーションの最後は、英のアラン・チューリング研究所のデイビッド・レズリー氏だ。レズリー氏はEthics and Responsible Innovation Researchのディレクターを務める。

アラン・チューリング研究所はこれまで様々な課題に関して原則から実践への枠組みの構築に取り組んできた。AIに関しても法務省および政府のデジタルサービスAIオフィスと協力し、2019年、公共部門におけるAIシステムの責任ある設計と実装を含むAI倫理に関するガイドラインをリリースした。その中で、実際のガバナンスの実践につなぐ試みとして、プロセスベースのガバナンスの枠組みを提示している。

レズリー氏が指摘するのは、AI倫理とガバナンスに関する議論における動きとして、道徳的な概念と社会的実践との結びつきが重視される傾向が強まっているという点だ。そして、そこで必要となるのは組織文化、技術的アプローチ、そして個人の意識の変革だとする。責任あるAIイノベーションの実現には、AIシステムを社会的なプロジェクトのライフサイクルを通して再認識することが重要となる。エンジニアリング、設計プロセスのあらゆる場面で、それらのプロセスがもたらす技術が現実社会に及ぼす影響について検討し、目的や価値、利益が果たす役割について考慮した上で倫理的に関与する。これが実践主導型プロセスベースのアプローチである。

また、レズリー氏は責任ある研究・イノベーション（Responsible Research and Innovation: RRI）の考え方と国内外の動向の習慣が、すべての研究イノベーションに組み込まれるべきという。このRRIの視点が、研究者やイノベーターに科学的発見と問題解決のすべてのプロセスには社会技術的側面と倫理的利害があるという認識を与える。アラン・チューリング研究所ではRRIツールとして「Care & ACT Framework」を作成し、これをサポートしている。

## パネルディスカッション：原則から実践へ、実践から原則へ

第2部では、アレン氏、フィンレイ氏、パートナー氏、レズリー氏に、城山氏も加わり、パネルディスカッションを行った。江間氏が司会を務めた。

AIガバナンスというと、通常、AIと共存する社会のためにどう準備を進めていけばよいのかという観点で「原則から実践へ（From principles to practice）」という流れで議論が展開する。しかし、今回のパネルディスカッションでは、逆に「実践から原則へと学ぶことがあるのか」という視点の議論も展開された。

いま生成AI、LLMなど、新たなAI技術が急速に普及している。特徴的なのは、原則に沿っているかどうかなど省みる作業を挟むことなく、いきなり実践のフェーズに入っていることである。となると、そうした実践を経て原則がアップデートすることもあるのではないか。こうした技術がどうやって 文明化するか、多くの人々がツールとして使うことで何が変わり、どのような原則を考えなければいけないのかといった議論も重要になるだろうと江間氏は論題を提起した。

議論ではChatGPT、あるいは生成AIの影響やリスクを検証する中、レズリー氏はこうした新たなAI技術について、ガバナンスという観点で事前のアプローチが必要だとする見方を示した。たとえばOpenAIは彼らの技術（サービス）を利用するユーザーに責任を持たせている。サービスを開発し、提供する側ではなく、ユーザーに責任を預けてしまっている。これは大きな問題で、政策レベルで見ていく必要があるのではないか。一方で、アレン氏はこうしたシステムの真のケーパビリティを評価することの難しさを課題として挙げ、まず適切に評価する方法を学ぶ段階なのだと述べた。

また、AIのグローバルガバナンスにおける日本のリーダーシップに関して、このタイミングで日本がG7の議長国となったことに期待する声のパネリストから多く上がっていた。数多くの国や地域がAI規制の準備を進めている今こそ、これら規制の相互運用性を担保するためにG7各国が一緒になってコミットメントをしなければいけない。高度なAIシステムに関して、ただ単に幅広くメリットがあるというだけではなく、これらのリスクを緩和する方法を探り、着地点を見つける必要がある。やり方によってはコラボレーションのバリアになる可能性があるため、G7がこれらの規制の枠組み、および規格の相互運用性を担保することは非常に重要となる。

会場から「AIはツールであってヒントを与えるもの、AIとAIを使う人間とのインタラクティブな関係を構築すべきで、最後は人間が判断の責任を負う。その際に一番問題となるのはファクトチェックではないか」という質問が出たが、AIはツールだという点にはパネリストもみな同意していた。システムにどのような限界があるかを現実的に理解することが必要である。ここでフィンレイ氏は、システムが何をするのか、何をしないのかを明確に文書化し、判断し、決定できるように、作業者がAIの使い方について教育される必要があると指摘した。

最後に城山氏から、実際に原則から実践へということを考える際、社会そのものが急速に変わっているということ認識しなければいけない、さらに運用においては民主主義との枠組みで考えなければならないところも出てくるという点も踏まえて、今後も議論を続けていくことの重要性が述べられてイベントは終了した。



(左より) 江間氏、アレン氏、フィンレイ氏、パートナー氏、レズリー氏、城山氏