AIガバナンスをめぐる海外の動向

株式会社エヌ・ティ・ティ・データ経営研究所
社会システムデザインユニット
シニアマネージャー
田中　理視

1．概要

生成系AIの急速な普及を前に、AIの利用を巡る対応のあり方についての報道や議論を目にすることが多くなっている。

数年前、いわゆる第３次AIブームの中で、AIの利用が社会に浸透し、将来的には「シンギュラリティ」と呼ばれる、AIが人間の能力を超える時代を見通してのAIの取り扱いについて、さまざまな議論がなされたところである。今般の生成系AIでは、AIがもたらす能力が、より身近な、そして驚異的な形でのサービスとして実現したことにより、改めてその取り扱いが問題となっている。

本稿では、このようなAIの取り扱いを含むガバナンスについて、特に海外の動向について整理するとともに、生成系AIの登場による新たな議論等も含めて示すことを目的とする。

2．AIガバナンスをめぐる論点

第３次AIブーム以降、ディープラーニングなどの技術の進展もあり、AIによる可能性が現実的なものとして示されるようになった。特にAIが人間の能力を超える「シンギュラリティ」がSFとはいえなくなったことを脅威として捉え、その対応のあり方について、主に倫理的な側面からの対応の必要性が説かれるようになった。特に、たとえばAIが判断を導き出す前提となる学習データが「汚染」されることにより、危険性の高い判断をもたらした事件が生じ[注1]、AIの危険性などに焦点を当てる見解などもみられた。

そのほか、AIに関しては、機械学習を通じたモデルによる説明責任の問題や、プライバシーの問題（たとえば一方的なプロファイルの生成など）、考慮すべき価値との関係でのあり方が示されてきたところである。

AIの取り扱いにおいて考慮すべき価値については、

・人間の尊重
・多様性・包摂の確保
・持続可能な社会
・人間判断の介在
・安全性・セキュリティ
・プライバシーの尊重
・公平性
・透明性
・アカウンタビリティ

などがあげられる[注2]。

AIの取り扱いに関するガバナンスでは、このような価値を保護しながら、AIの利用可能性の拡大を図ることが求められる。

3．各国等のAIガバナンスに関する状況

（1）米国

米国におけるAIガバナンスは、AIの規制だけに特化したものではなく、さまざまな連邦機関がAI 政策の検討とガイドラインを発行するような取り組みのアプローチをとっている。

具体的には「2020年国家AIイニシアチブ法」(The National AI Initiative Act of 2020[注3]) が策定され、連邦政府全体でAI研究と政策を調整する枠組みとAI研究の全国ネットワークの創設を行い、米国がAI研究を主導することを目的としてする。

表１　2020年国家AIイニシアチブ法におけるAIガバナンスに関連する内容

項目	内容
標準とガイドライン:	【AIの標準開発】 ・連邦政府によるAI技術の標準化に関する研究の推進を規定 ・技術的な標準やベストプラクティス、セキュリティとプライバシーのガイドラインなどが含まれる 【協力】 ・AIの標準開発のために、民間部門や国際的な組織、研究機関などと協力することを奨励 【既存のガイドラインのレビュー】 ・既存のAIに関連するガイドラインやポリシーを定期的にレビューし、必要に応じて更新することを要求
AIの倫理と公平性	【倫理的な考慮事項】 ・AI技術の研究、開発、利用に関連する倫理的な問題や挑戦に対処するための研究を強化することを推奨 ・人権、公平性、透明性、説明責任などの側面を含む 【公平性の確保】 ・AIのアルゴリズムやデータセットが公平であり、偏見や歪みをもたないようにする研究を奨励 【教育と普及】 ・AIの開発者や利用者、一般の市民など、さまざまなステークホルダーを対象としたAIの倫理や公平性に関する知識を広めるための教育や啓発活動の推進

図１　AIシステムのライフサイクルこれに応じた主要な取り組み

同法はガバナンス面からは、特にAI技術の標準化や安全性、セキュリティに関するガイドラインの開発、公共の利益へのアクセスと透明性確保、倫理と公平性に関する方針の強化等を内容としている（表１）。

2020年国家AIイニシアチブ法を具体的に実施するために、米国国立標準技術研究所（NIST）から2023年１月に「Artificial Intelligence Risk Management Framework（AI RMF 1.0)」[注4]が公表されている。これは、組織や個人に対して、AIシステムの信頼性を高めるためのアプローチを提供し、AIシステムの責任ある開発、展開、使用を長期にわたって促進することを企図したフレームワークとされている。このフレームワークでは、AIシステムに潜在する各種リスクの管理のあり方を示したうえで（図 1)、AIシステムへのリスクを管理するための事項などを示している。

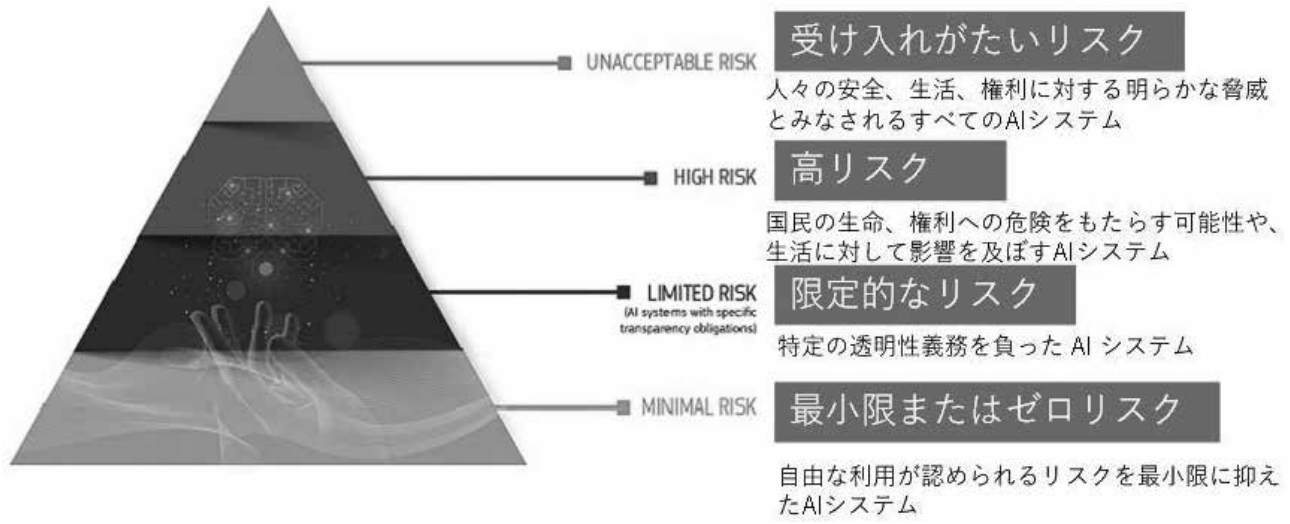
このフレームワークにより米国では、人工知能（AI）に関連する個人、組織、社会へのリスクをより適切に管理し、その利活用の促進を図っている。

（2）EU

EUではAIのガバナンスのために、AI法の導入を進めている。これらは指令ではなく「規則」の形をとるものである[注5]。

欧州委員会は2018年より、AIガバナンスに関する検討を進めており、2023年６月には欧州議会がAI法（Artificial Intelligence Act）案[注6]について採択している[注7]。欧州議会が採択したAI法案では、AIを用いるアプリケーションについて４つのリスクに分類し（図２)、それぞれのリスクに該当するアプリケーションに対して義務を設けている。特に高リスクに分類されるAIアプリケーションは、行政やインフラ、医療、労働など国民生活の根幹にかかわる部分のものが多く含まれている。

図２　EUのAI法案において類型化されるAIのリスクとそのシステム（AI法案ページより作成）



AI法案の概要を表２に示す。リスクの分類や、それに応じた対応のほか、透明性やデータに対する適正性の管理、システムの適合性評価等、人々の安全を確保する観点からの対応が盛り込まれている。

AI法案においては、違反した場合の罰則が定められている（表３）。AI法に対するペナルティは全体的に重くなっており、特に禁止行為に対する違反は、企業活動に極めて大きな影響を及ぼす内容となっている。

表２　Artificial Intelligence Act（案）の概要

項目	内容
リスク分類	・AI法はAIシステムを受け入れられないリスク、高リスク、限定リスクの３つのカテゴリに分類 ・これらに応じたAIシステムに課せられる要件と義務のレベルが決定されることを目的とする
高リスクAIシステム	・重要なインフラ、医療、法執行などで使用されるAIシステムは高リスクと見なされ、より厳しい要件が適用 ・リスク評価義務、人間による監督、透明性の義務、データガバナンス措置などを含む
禁止されたAIのプラクティス	・特定の状況で許可されていないリアルタイムの遠隔生体認証システムなどは、受け入れられないリスクと分類し、一部のAIの実行を禁止
透明性とトレーサビリティ	・AI法ではAIシステムにおける透明性とトレーサビリティの重要性を強調 ・ユーザーに対して明確で理解しやすい情報を提供し、人間の介入と制御を可能にすることを求める
データガバナンス	・高リスクのシステムに対して、AIシステムが使用するデータの取り扱いに関して、データの品質確保、データ保護、一般データ保護規則（GDPR）などの規制の遵守求める
適合性評価	・高リスクAIシステムはAI法で設定された要件に適合しているかを確認するために、適合性評価プロセスを実施 ・第三者のテストと認証を含む
ヨーロッパ人工知能委員会	・ヨーロッパ人工知能委員会を設立し、規制の解釈と適用に関する指針を提供し、国内監督当局間の協力を促進する

違反内容	罰金
AI の禁止行為に関する規則の違反	最高4000万ユーロの行政罰金、または違反者が企業の場合は前会計年度の世界年間総売上高の７％のいずれか高い方の罰金
第 10 条 (データとデータガバナンス) および第 13 条 (ユーザーへの情報の透明性と提供) に基づく規則の違反	最高 2000万ユーロの行政罰金、または違反者が企業の場合は前会計年度の全世界年間売上高の４％のいずれか高い方の罰金
AI 法に基づくその他の要件および義務の不遵守	最大1000万ユーロの行政罰金、または違反者が企業の場合は前会計年度の全世界年間総売上高の最大２％のいずれか高い方の罰金
不正確、不完全、または誤解を招く情報を指定機関および国内管轄当局に提供した場合	最高 500万ユーロの行政罰金、または違反者が企業の場合は前会計年度の全世界年間総売上高の最大１％のいずれか高い方の罰金

（3）中国

中国では、2017年に「新一代人工知能発展規画」が策定され、2030年に向けたAI戦略を示しており、この中でAIのあり方などのガバナンスも含めて重点任務を示している[注8]。これを受けて、特に倫理面と技術的実行面から、AIに関するガバナンスの整備が進められている。

倫理面では、「科学技術倫理の見直しに係る措置」（案）が2023年４月から５月にかけてパブリックコメントに付されている。AIとの関係でみると、

・さまざまな科学技術活動のうち、「人間や実験動物に直接関与しないが、生命と健康、生態環境、治安、持続可能な開発の観点から倫理的リスクや課題を引き起こす可能性のある科学技術活動」についてもこの措置の対象となること（第３条）
・大学、科学研究機関、医療保健機関、企業等においては、科学技術倫理審査の管理主体となるとされており、特に生命科学、医学、人工知能などの科学技術活動に従事する部門は、研究内容が科学技術倫理上デリケートな分野に関わる場合には、科学技術倫理（審査）委員会の設置が義務付けられること（第５条）
・科学技術倫理（審査）委員会では、データとアルゴリズムに関連する科学技術活動において、データ処理の方法は国のデータ安全に関する規定に準拠しており、データ安全のリスクモニタリングと緊急時の対応計画が適切である。アルゴリズムとシステムの研究開発は公平、公正、透明、信頼性、制御可能といった原則に準拠していることも審査対象となる。（第14条）

などの内容が含まれている。

技術的なガバナンスという点では、近時制定された情報関連法を踏まえて大きく３つの規則が示されている。ひとつは、「インターネット情報サービスにおけるアルゴリズム管理規定」[注9]（以下「アルゴリズム規定」）で、もうひとつは「インターネット情報サービスにおける深層学習管理規定」[注10]（以下「深層学習規定」）、それと後述する「生成型人工知能サービスの運営に関する暫定措置」[注11]（以下「生成系AI暫定措置」）がある。

アルゴリズム規定も深層学習規定も、規程の対象者は事業者であり、主に事業者に対する義務等を定めるものである。ただし、禁止された情報等の作成目的での使用禁止義務等は、国民やサービス利用者も対象となっている。

これらの規則の特徴として、AIが利用される具体的な分野を対象とするのではなく、アルゴリズムや深層学習等の技術的な手段を対象とした規制であることが

表４　アルゴリズム規定・深層学習規定の概要

【「インターネットサービス情報サービスにおけるアルゴリズム管理規定」の概要】(2022年３月１日施行)
・アルゴリズムの安全性責任（アルゴリズムへの安全性責任の実装、ユーザー登録、情報公開レビュー、アルゴリズムメカニズムのレビュー、セキュリティ評価の監視、セキュリティインシデントの緊急対応、データセキュリティ、および個人情報保護のための管理システムの確立）
・法令の遵守（法令遵守、社会道徳と倫理基準を尊重し、正しい政治的方向性、世論志向、価値志向の遵守）
・情報サービス仕様（アルゴリズムにおける透明性・説明可能性の最適化、安全性・公正性の確保等）
・利用者等に対する権利義務（利用者に対するサービス周知、個人情報保護）
・禁止行為（アルゴリズム推奨サービスを使用した法律・行政規制で禁止されている情報の作成・複製・出版・配布の禁止、アルゴリズム推奨サービスを使用した誤ったニュース情報の作成・コピー・公開・流布の禁止）
・セキュリティレビュー（データと個人情報の安全性を確保するため、公開前にセキュリティレビューの実施義務）
・監督・管理（国家サイバースペース局およびその他の関連部門による監督・管理、罰金等）

【「インターネット情報サービスにおける深層学習管理規定」の概要】(2023年１月10日施行)
・情報セキュリティ責任（ユーザー登録、アルゴリズムメカニズムのレビュー、倫理レビュー、情報公開レビュー、データセキュリティ、個人情報保護、通信詐欺防止、および緊急対応のための管理システムを確立の義務付け）
・法令の遵守等（法令遵守、社会道徳と倫理基準の尊重、正しい政治的方向性、世論指向、および価値指向の遵守）
・特定の機能・出力に対する仕様と手続
・利用者等に対する権利義務（利用者身元確認、個人情報保護）
・禁止行為（法律・行政規制で禁止されている情報の作成・複製・公開・配布目的での深層合成サービスの使用禁止、深層合成サービスを使用した虚偽のニュース情報作成・複製・公開・流布の禁止）
・セキュリティレビュー（データと個人情報の安全性を確保するため、公開前のセキュリティレビューの実施義務）
・監督・管理（国家サイバースペース局およびその他の関連部門による監督・管理、罰則等）

あげられる。そのため、AIを利用するシステムを中国が輸出入する際のルールとなることも予想され、さらには国際的なデファクト・ルールになることを安全保障などの観点から懸念する意見もみられる[注12]（表４）。

（４）その他

OECDでは2019年５月の年次閣僚理事会でAI原則を採択した。AI原則では、５つの原則と５つの提言を示している（図 3）。

５原則は以下の内容から構成される。

・AIは包括的な成長、持続可能な開発、福祉に資するものでなければならない。
・AIは、法の支配、人権、民主的価値観、多様性を尊重する方法で設計されるべきであり、公平で公正な社会を確保するための適切な保護手段を組み込む必要がある。
・AIにおいては、AI システムに関わる透明性と責任ある開示が行われ、人々がいつAI システムに関与し、結果に異議を唱えることができるかを確実に理解できることが求められる。
・AIにおいては、そのライフサイクルを通じて堅牢かつ安全に機能する必要があり、潜在的なリスクは継続的に評価および管理される必要がある。
・AI システムを開発、導入、または運用する組織や個人は、AI に関するOECDの価値観に基づく原則に沿って適切に機能することに責任を負うことが求められる。

図３　OECD「AI原則」における５原則と５つの提言注13

OECDのAI原則は、複数国間において合意された最初のAIガバナンスの取り決めであり、G20首脳会合の首脳宣言の附属文書として採択されている。そのため、一定の標準としての意義が認められる。

4. 生成系AIをめぐるガバナンス等の動向

（1）生成系AIのリスク

2022年頃から、生成系AIをめぐる議論が大きくクローズアップされてきた。生成系AIサービスでは、利用者が容易にAIを利用して、さまざまな出力を得ることができるという点で、AIをより身近にするものである。一方で、AIによる出力が急速に身近に普及することで、さまざまな権利侵害のリスクへの対応が十分でない、AIに求められる規律が満たされないなどの問題が指摘されている。

特に著作権等の知的財産権やプライバシー侵害などの危険性が高いものとされているほか、出力されたものの正確性が保証されていないにもかかわらず、流通することによる混乱が生じるなどの問題も指摘されている。

このようなリスクは、元々AIシステムに潜在するものであるものの、急速なサービスの開発や利用拡大により、その対応方針が求められている。以下、生成系AIに対するガバナンス等に関する各国の動向を簡単に紹介する。

（2）生成系AIのガバナンスに関する各国の動向

①米国

米国では、生成系AIに関するリスクと規制の必要性を検討するための公聴会が2023年５月に行われている。ここでは各界から生成系AIに対する利害関係を有する者の意見が陳述され、この中でAI規制を目的とした専門機関の設置に関する意見や、安全審査に関するスキームを設けることなどに関する意見も示された。

なお、米国では生成系AIをめぐる訴訟も提起されている。たとえばChat GPTを利用したチャットボットにより、虚偽の犯罪に関する出力がなされたことによる名誉棄損に関する事案や[注14]、人間を介在せず、AIにより生成された出力が著作権として保護されないとする事案[注15]等が発生している。

②EU

上述の欧州議会が採択したAI法においては、生成系AIに関する規制に関しても盛り込まれることとなった。具体的には、「自律性レベルで複雑なテキスト、画像、オーディオ、ビデオなどのコンテンツを生成することを特に目的とした AI システムで使用される基盤モデル」を提供する事業者等は

・コンテンツがAIによって生成されたことを開示する
・違法なコンテンツが生成されないようにモデルを設計する
・トレーニングに使用された著作権で保護されたデータの概要の公開

などが義務付けられている[注16]。

③中国

上述のように中国では、2023年７月に「生成系AI暫定措置」を策定している（2023年８月15日発効）。この措置では、生成系AIに関して国家の監督におくことを明示するほか、知的財産への配慮やAIシステムの不正競争等の防止、プライバシー、肖像権等の保護、透明性の確保、差別防止の措置などを、サービスの提供および利用において求めている。

同時に学習するデータ等に関するガバナンスや、提供するサービスにおいて遵守すべき仕様等についても示すほか、EUのAI法案同様、生成系AIによる出力に対して、その旨を明示すべきことを求めている。

5. 最後に

AIガバナンスに関しては、わが国でも2019年３月に「人間中心のAI社会原則」（統合イノベーション戦略推進会議決定）が策定されたほか、開発のあり方や当事者間での利害関係に関するルールづくり等が進められてきたところである。

AIシステムについては、わが国ではこれまでは、利用や開発ついての規制自体は設けられておらず、また著作権法などでも開発者が機械学習など行いやすい環

境が整備されてきた。

一方で、生成系AIの普及により、AIの開発や仕様において求められる要件や規制などの議論が盛んになっており、本稿でみたように各国でその対応が進められているところである。AIガバナンスについては、必ずしも一国内部のものではなく、開発されたシステム・サービスの輸出入を勘案すると、国際的なスタンダードにも密接に関連するところである。わが国においても、各国のガバナンスの内容等を踏まえつつ、国際的に協調できる形で引き続きAIガバナンスの整備することが求められる。

注１：2016年Microsoft社が、AIを通じたカジュアルで楽しい会話を通じた価値を提供するために、「Tay」というbotサービスを提供した。しかしTayが学習するデータに人種差別的なデータが多く含まれることとなり、Tayが差別主義的なツイートを行う事態が発生。MicrosoftはTayのサービスを開始後16時間で停止せざるを得なくなった。
注２：「AI倫理原則の世界的動向」(福岡真之介、NBLNo.1168（商事法務）P51)。これらの価値は、当時の日本および各国のAI倫理原則のうち、５つ以上の原則に共通するものを抽出して示されている。
注３：https://www.congress.gov/bill/116th-congress/house-bill/6216
注４：https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
注５：EUの規則として有名なものとして、GDPRがあげられる。
注６：https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html
注７：https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
なお今後の予定として、欧州議会、欧州委員会、EU理事会との２年程度の調整を経たうえで、規則として成立することが見込まれている。現状、事業者に適用されるのは、早くて2024年後半以降とされている。
注８：https://www.jetro.go.jp/biznews/2019/05/98e72340827c4a0e.html
注９：「互**联**网信息服**务**算法推荐管理**规**定」
https://www.gov.cn/zhengce/zhengceku/2022-01/04/content_5666429.htm
注10：「互**联**网信息服**务**深度合成管理**规**定」
https://www.gov.cn/zheng ce/zhengceku/2022-12/12/content_5731431.htm
注11：「生成式人工智能服**务**管理**暂**行**办**法」
http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm
注12：たとえばhttps://www.reuters.com/world/us/senate-leader-schumer-pushes-ai-regulatory-regime-after-china-action-2023-04-13/
注13：https://oecd.ai/en/ai-principles
注14：https://www.courtlistener.com/docket/67617826/walters-v-openai-llc/
注15：https://fingfx.thomsonreuters.com/gfx/legaldocs/lbvgooeoqvq/AI%20COPYRIGHT%20LAWSUIT%20thalerdecision.pdf
注16：修正第 399 条「規制」第 28 条b案（新規）

（筆者略歴）
株式会社NTTデータ経営研究所社会システムデザインユニット、シニアマネージャー。
三井情報開発株式会社（現三井情報株式会社）総合研究所を経て、2007年より現職。
主に情報ガバナンスに関する法制度等を専門とし、個人情報保護関係やデータ利用の在り方等に関する経歴を有する。
「AI・データの利用に関する契約に関するガイドライン」(経済産業省)、「農業分野におけるAI・データに関する契約に関するガイドライン」(農林水産省）等の作成支援に従事。