JILISレポート
ディスカッションペーパー Date: 2025.01.14
https://jilis.org/report/2025/jilisreport-vol6no2.pdf

AI規制フレームワーク：日本型モデルの可能性

中央大学 総合政策学部 教授
実積 寿也

1. はじめに

人工知能（Artificial Intelligence、AI）の活用が急速に進みつつある。20世紀の半ばより長足の進歩をとげた情報通信技術により人類社会は大きな構造変化を経験し、急速な経済成長を実現してきたところであるが、AIはそれらを遥かに凌駕するインパクトを期待されている。特に、地球環境問題や所得格差の問題など人類社会が直面する諸問題の解決には、AIの活用によって得られる高度な予知能力が欠かせない。加えて、少子高齢化を伴う生産年齢人口の減少に直面することが確実な日本では、生産性を高めることが経済厚生を維持するためには必須であり、そのためにもAIの積極的活用は急務であると理解されている。

AIは大きなメリットを実現する一方で、これまでにないリスクを引き起こす可能性もあることが指摘される。その場合、AIは汎用技術（General Purpose Technology、GPT）として幅広い応用範囲が期待されるため、社会的リスクを引き起こす分野が他の技術と比較して桁違いに広範になりがちである。AIの利活用がもたらす生産プロセスや産業構造の変革は社会全体のDXを究極まで推し進める力をもち、従来の社会経済活動のあり方を一変させるため、AIリテラシー教育の充実や失業問題への対応など、AI導入初期には不可避な摩擦的事象への対応も必要である。そのため、AIという新技術については他の新技術とは質的に異なる規制が求められ、多くの国や国際機関で様々な試みがなされている。

その際、AIのもたらす便益や今後のイノベーションの余地を過度に制約することのないバランスのとれた規制水準であることが必要である。リスクに対して過剰に反応した結果、AI本来の利活用の余地を狭めると、期待した便益自体を得ることができない。自動車草創期の1865年に英国で導入された赤旗法（正式名称The Locomotive Act 1865）のような愚を繰り返すべきでないことは明らかであるし、リスクを無視した結果、19世紀初頭、産業革命初期の英国で発生した機械打ち壊し運動（Luddite movement、ラッダイト運動）に類する事態を引き起こせばさらに悲劇である。

本稿では、そうしたAI導入に際して必要となるルールづくりについて分析を行う。まず、次節において、AIがもたらす三種類のリスクを指摘し、続く第3節ではAIの本質的な不完全性を前提とした場合に必須となる社会全体のAIに対する信頼（トラスト）について説明する。第4節では、具体的なルールづくりにおいて大きな分岐点となるハードローアプローチとソフトローアプローチを比較し、日本における状況を解説する。第5節は、国境を超えて提供されることが多いAIサービスへの規制で発生する可能性がある「底辺への競争」（race to the bottom）の問題と、それに対する解決策について取り扱う。第6節で具体的な制度提案を示し、全体のまとめと残された課題を最後の第7節で提示する。

2. 三種類のリスク

AIがもたらすリスクには、技術的リスク、社会的リスク、人類存続リスクの3つがある。

まず、AIの技術的な特徴に由来する技術的リスクと、高性能なAIが社会に実装されることによって生じる社会的リスクについては、羽深（2023）が論じている。

技術的リスクは、外部から与えられたビッグデータを学習し、膨大な数のパラメーターを最適化することで自らのアルゴリズムやモデルを構築し、それを用いて出力を生成するという機械学習プロセスそのものに由来するリスクと定義されている。機械学習においては、不正確もしくは不適切な出力（生成

AIの場合は特にHallucinationと称される）が生まれることを完全には回避することはできない。この点はOECDの報告書でも「…there is always a chance that an AI will not perform as intended. Even an unbiased algorithm is unlikely to be 100% accurate.（…AIが意図したとおりに動作しない可能性は常にある。偏りのないアルゴリズムであっても、100%正確ではない。）［筆者訳］」と指摘されている（Berryhill et al. 2019, p.109）[1]。また、推論モデルが巨大で、パラメータ数が膨大である大規模言語モデル（Large Language Models、LLM）の場合は特定の出力が生み出された場合、事後的に、そのロジックを説明することは不可能であるため、その出力結果により不利益が発生した場合の原因究明が大きく妨げられ、是正措置が困難となることも技術的なリスクの一つである。

一方、社会的リスクは、高性能なAIが社会実装されることにより引き起こされる各種リスクとして定義される。プライバシー侵害、民主主義への介入、詐欺的活用、公正競争への悪影響、著作権などの既存の財産権への侵害、計算負荷による環境問題などがその具体例とされている。いずれも、AIを導入しようとする社会経済システムの側で必要な制度的準備を十分に整えていないことに起因している。

こうしたリスクの一部は利用者側のリテラシーが改善することで自ずから解決することが期待されるが、それ以外のものについては何らかの対応措置をとる必要がある。羽深（2023）は、技術的リスクについてはAIサービスの提供者側が最終的な責任を負う必要があり（取組例として内田他（2021））、社会的リスクについてはAIサービス提供者に限らず社会全体で責任を分担する必要があると主張している。しかし現状、AIのメカニズムにはサービス提供者自身にとってさえ本質的なブラックボックス性が残っており、どれほど努力したとしても無謬性や説明可能性の達成が見込めず、かつ発生するリスクの上限や発生可能性の予測が極めて困難である。その場合、サービス提供者側に最終的な責任を課すことは、技術開発のインセンティブを過度に損なう可能性がある。技術開発が滞り、AI活用によって期待される社会的利益を得ることができないという最悪の結果も予想される[2]。そのため、社会的リスクのみならず技術的リスクについても、提供側の合理的努力によって回避できない部分については、一般利用者を含む社会全体で担うことが社会厚生の観点からは適切である。

さらに、近年では、欧米を中心に、進化を極めたAIがもたらす人類存続リスク（existential risk）、もしくは人類絶滅リスク（extinction risk）が第３のリスクとして注目されている。これは、人間の能力を遥かに超えるAIが、操作者の不注意、もしくは悪意ある命令により、制御不能の状態に陥ることにより、人類社会そのものを存続の危機に立たせる可能性があるという主張である。これは、技術進歩により人間を超える知性が覚醒し、指数関数的な知能暴走が発生する起点としてVinge（1994）が定義する特異点（singularity）や、Bostrom（2014）が「any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest（ほぼすべての領域で、人間の認知能力を大きく上回る知性）［筆者訳］」（p.22）として定義する超知能（superintelligence）といった概念と類似点が多い。Singularityについては、その後、Kurzweil（2005）により大々的にフィーチャーされた結果、専門家以外にもその呼称や概念が広く知られている。こうしたAI脅威論については、多くの非現実的な仮定を前提としているため科学的根拠が乏しく、信用に値しないという評価も多い。ただし、これらが政策担当者やAI研究者の間で一定の懸念を惹起していることは事実であり、関連する研究開発の中断が提案され（Future Life Institute 2023）、2023年11月に英国で開催されたAI Safety Summit（AI安全性サミット）における議論に影響を与え、29カ国が署名したBletchley Declaration（ブレッチリー宣言）[3]にもその意図が反映されるなどしている。中川（2019）らは、こうしたAI脅威論が主に欧米で議論されている背景には、キリスト教に代表される一神教の考え方があり、知性をもつプレイヤーを神ならぬ人が創造してしまうことに対する根源的な嫌悪感があると指摘している。同時期のわが国での検討（総務省 2017）では、AIの開発に際して人間の尊厳と個人の自律を尊重すべきであり、国際人権法や国際人道法を踏まえるべきという倫理の側面は議論されたが、AI脅威論的な観点はない。その後の「人間中心のAI社会原則」（内閣府 2019）においても同様である。

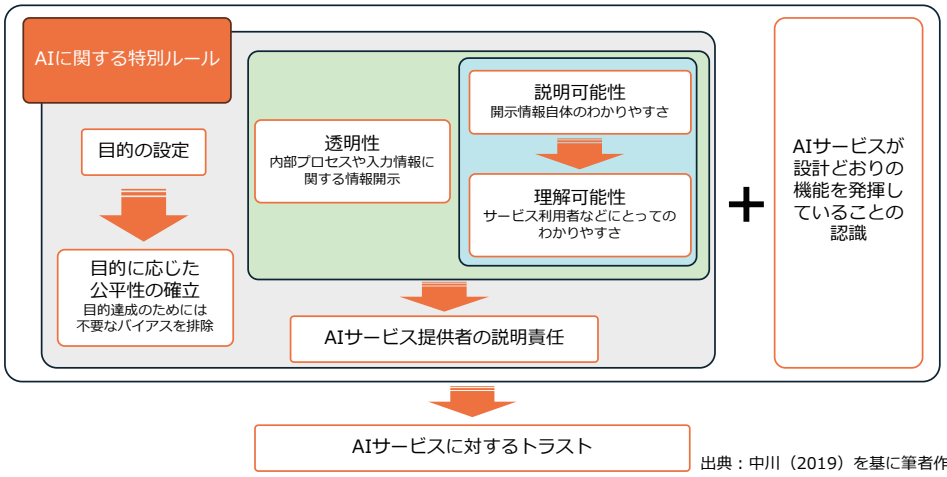
3.対策としてのトラスト

提供されるAIサービスのリスクを提供者側で完全にカバーできない場合、利用者を含む社会全体で発生す

1） 同様の指摘は、Bathaee（2018）、Kim and Routledge（2022）、Quin（2021）、Lee et al.（2019）、中川（2019、2020、2021）や荒堀（2020）などにも見られる。
2） この点はBathaee（2018）も指摘するところである。
3） https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

るリスクを分担する必要がある[4]。リスク発生確率が既知の場合は保険の設定が可能であるが、AIの場合はそれを期待できない。普及の初期段階にあり、技術開発が急速であるためにユースケースが常時変化しているAIは、リスクの存在は指摘できるものの、すくなくとも当面は、その確率が計算できない財・サービスの範疇に属するためである。その場合、「AIサービスが利用者利益を最大化する（そしてそれを通じて提供側の利益を最大化する）という目的で提供され、さらに関係者がリスクを最小化する努力を尽くしていること」を社会全体が信頼（trust、トラスト）[5]することができなければサービスの普及が進まない。サービス提供者側へのトラストを基盤として、発生不可避の被害に伴うコストを関係者間で分担する仕組みを構築することが求められる。AIの利活用に関わるすべての関係者、とりわけサービス提供者側が、自分たちの努力に関して市場からのトラストを得ることは、需要喚起の面からも重要である。市場からのトラストが需要にプラスの影響があるという点については、オンラインショッピングに関しNagy and Hadjú（2021）が実証的に示している。

市場からのトラストを得るためには、XAI（Explainable AI）[6]など内田他（2021）が列挙する各種の技術的対応策を導入して、その事実を公開するとともに、提供事業者自身が（さらには、関連する利害関係者に協調行動を促すような）特別ルールを設定し、外部から確認可能な形で自発的に遵守することが求められる。求められる特別ルールについて中川（2019）は、AIが設計通りの機能を発揮することを前提に、内部プロセスや入力情報に関する実質的な情報開示を行うこと、つまり十分な透明性（transparency）と説明可能性（explainability）を確保することで利用者の理解可能性（understandability）を保証することがAIサービス提供者の説明責任（accountability）の確立をもたらし、そのことでAIサービスに対するトラストが生まれると主張する（図表1）[7]。実際、Figueroa-Armijos et al.（2023）は人材採用プロセスにおけるAI活用についてアンケート調査を行い、機能に対する期待（performance expectancy）がトラストを向上させることを示し、一方、Grimmelikhuijsen（2022）は仮想的状況を設定したアンケート調査を使って、透明性だけではなく説明可能性が担保されないとAIトラストが獲得できないことを明らかにしている。中川（2019）は、AIを社会実装する際には、利用者間の公平性（fairness）の確保の重要性も指摘しており、透明性や説明可能性、説明責任の確保は、不必要なバイアスがアウトプットに入らないようAIを活用していることを利用者に保証する役割を果たすと指摘する。



出典：中川（2019）を基に筆者作成

図表1　中川（2019）の主張する特別ルールとトラストの関係

4）提供者側がリスクを完全にはカバーできないまま提供されている財・サービスはAIには限らない。例えば、自動車は故障や事故の確率をゼロにすることは今日までのところ達成できていない。顕在化したリスクはさまざまなコストを発生させるため、それをカバーする仕組みが求められる。

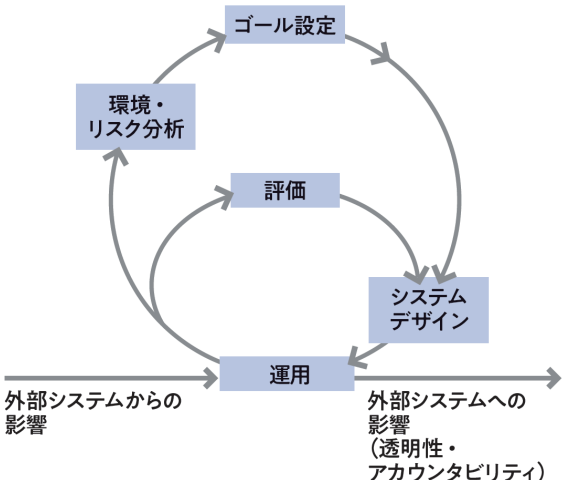
5）トラストは、利用者側が未知の事態に対応する提供者側の対処を、それを保証する確実な根拠がない状況下において、自らの楽観に依拠して事前に許容しておく行為として、Hosmer（1995）やPirson et al.（2019）が列挙する先行文献では定義されている。例えば、Kim and Routledge（2022）は、プレイヤーAとプレイヤーBの関係性が以下の三つを満たす場合、AからBへのトラストがあると定義している。
1) プレイヤーAは、特定分野に関してプレイヤーBの裁量に依存し、そのリスクを受け入れている。これにより、AはBの背信に対して脆弱になっている。
2) Aは当該分野に関してBに十分な能力があると楽観的に考えている。
3) Aは、Bが当該分野に関してはAに対し好意的なコミットメント（責任感）を持っていると楽観的に考えている。

6）AIによって作成された結果と出力をユーザーが理解し、信頼できるようにする技術の総称

7）ただし、満足のいく理解可能性の達成は現状困難であるため、現実解として採用されているのがAIサービスの提供段階に人間の関与を求める仕組み（human-in-the-loop）であると指摘する。

こうして設定された特別ルールを事業者自身が遵守することを維持する仕組み（組織統治、ガバナンス）[8]は対外的にもわかりやすい形で、かつ第三者からの確認を許容する形で構築することが望ましい。さらに、そうしたメカニズムは環境変化に対応して常に最適化を図ることが重要であり、さもなければ社会からのトラストを長期間確保することが望めない。経済産業省が提案し、企業に採用を呼びかけているアジャイル・ガバナンス（経済産業省 2021b）はその要請に応えるもので、外部環境の変化に対応した目標の再定義を前提にPlan-Do-Check-Actサイクルを回すことで、急速な社会の変化や技術の進歩に柔軟に対応するためのガバナンス方式となっている（図表２）。

図表２　アジャイル・ガバナンスのメカニズム



出典：経済産業省（2021b, p.iii）、許諾を得て転載。

4. ハードローアプローチとソフトローアプローチ

政府による具体的なルール作成においては、法的拘束力をもつ明文化されたハードローを用いる方法や、市場プレイヤーの自主的取り組みをベースとするソフトローを用いる方法、さらにはその両者のハイブリッドなど様々な政策アプローチがある。ハードローアプローチの代表例が2024年にAI法を定めたEUであり、ソフトローアプローチの代表例がAIサービス企業の自主的コミットメントや非拘束的ガイドラインを基礎に規律付けを行う米国や日本であるとされている。ただ、そういった国々で採用されているのは、実際には両者のハイブリッドである。EUのAI法は完全に非拘束的なガイドラインの活用を排除するものではなく、第56条で、それ自体には法的拘束力を欠く行動規範（code of practice）を用いて汎用AIを提供する事業者への規律を行うことを予定している。日本のソフトローアプローチの場合も、法的拘束力を持つ成文法の力を借りる箇所があり、羽深（2023）はその具体例として、AIを活用した与信審査を合法化した割賦販売法改正（2020年６月）、レベル４の自動運転を念頭においた道路交通法改正（2023年４月）を挙げる。現実に観察されるハードローアプローチとソフトローアプローチの差は、立法行為を含むか否かにではなく、AIを分野横断的に規制する広範かつ包括的な一般法をツールとしているか否かにある。

こうしたハードローアプローチでは、包括的な一般法が法的安定性を分野横断的に確保しているため、事業者にとっては法・規制環境の不確実性への対処に伴うコストは最小化されている。罰則の設定を適切に行えば、規律遵守も確保でき、リスク軽減効果も期待できる。一方で、具体的な規制水準は情報の非対称の影響を被る規制庁が詳細まで設定しなくてはならないため、技術やビジネスの実態と乖離することは避けられず、最適解を得ることは難しい。規制修正・改廃には、立法行為という時間とコストのかかるプロセスをその都度経る必要があるため、AI技術やビジネスプランの急速な進歩に追随することができず、長期的な効率性ロスも避けられない。また、後述するとおり、本アプローチは技術中立性を損なう可能性が高いため、長期的にはイノベーションの進展に歪みをもたらし、効率性ロスを生んでしまう。

ソフトローアプローチでは、正式な立法プロセスを経ないガイドライン等を主要ツールとするため、修正・改廃に手間はかからず、技術進歩等への追随は容易い。反面、成分法のような法的安定性には欠け、執行が担当行政庁の判断に委ねられる部分も多くなるため、対応を求められる企業にとっては不確実性が高い環境となり、コスト増要因となる。事業者が遵守を拒んだ場合は直接に対処する術をもたないため、自主的な遵守を促す必要があり、規律内容をデザインする際、インセンティブ両立性に配慮することが求められ、ハードローアプローチの場合と比べて、規制に一定の制約

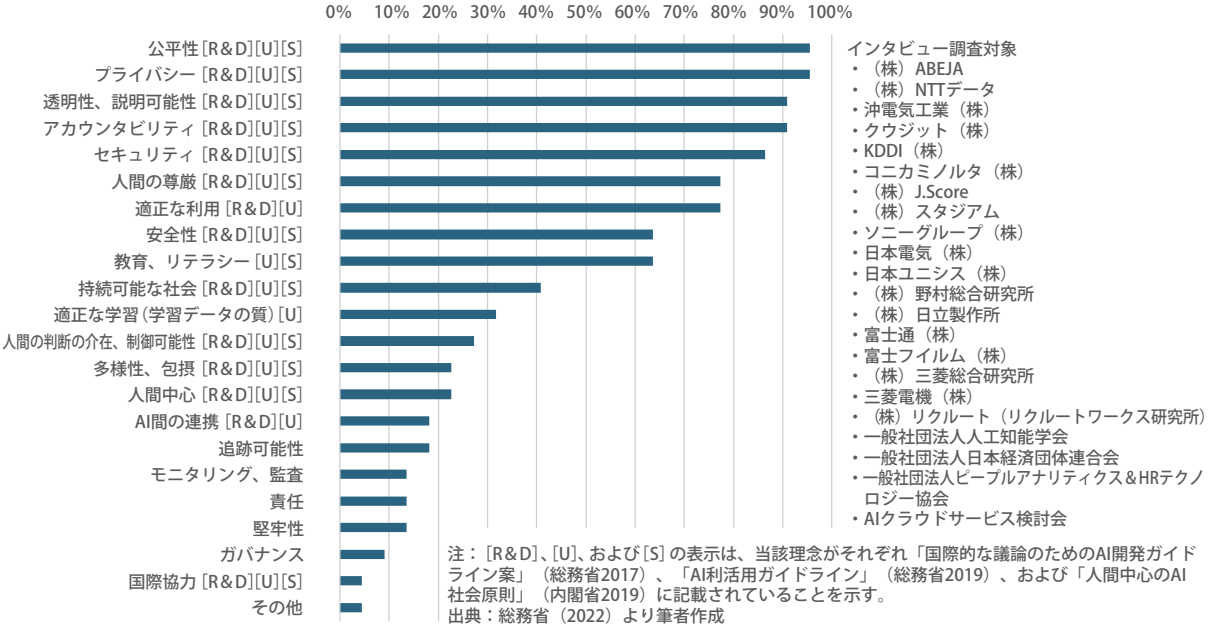
8）こうしたガバナンス（AIガバナンス）を、経済産業省（2021a）では「AIの利活用によって生じるリスクをステークホルダーにとって受容可能な水準で管理しつつ、そこからもたらされる正のインパクトを最大化することを目的とする、ステークホルダーによる技術的、組織的、及び社会的システムの設計及び運用」（脚注5）と定義している。

が課される[9]。この点で、現在問題となっているのが生成AIによるコンテンツ作成時の著作権帰属である。米国著作権法の運用に関し、DC連邦地裁は、生成AIは著作権主体となり得ないため、人間の関与がなく制裁されたコンテンツは著作権フリーであるという意見を2023年8月18日に表明している[10]。そのため、Noti-Victor（forthcoming 2025）が指摘するとおり、著作権許諾の対価としての収入獲得をビジネスモデルの基盤とするコンテンツ事業者にとって、生成AIの利用を開示することは、利潤最大化インセンティブと整合的ではない。そのため、AI活用の有無を開示することが法的拘束力のないガイドライン上に記述されていた場合、その遵守は期待できない。インセンティブ両立性を確保するためには、著作権法の仕組みを改めることが必要となる。あるいは、共同規制というやり方で、外部環境対応の柔軟性というソフトローのメリットに、法的強制力というハードローのメリットを組み合わせる方策も検討できる。例えば、新保（2020）が提案するように、情報開示を法定事項とし、実態と開示情報に乖離がある場合には政府が事後的介入を行うというハイブリッドなアプローチは考慮に値する。

AIという問題に取り組む各国の規制当局は、両アプローチの長所短所を比較検討したうえで、最適な対処を策定する必要がある。ブロードバンドを介することでAIサービスの提供は国境を超えて可能であり、かつ、AIは活用ケースが幅広い汎用技術としての性質を持つことを考えると、これはいわゆるAI先進国だけにはとどまらず、世界中の政策担当者の課題である。

ただし、そうした検討にあたっては各国の特徴を考慮することが重要である。日本の場合、以下に挙げる三つの事情がソフトローアプローチを採用することの比較優位性を高めた可能性がある。まず第一に、非拘束的な規律であったとしても、日本企業はそれらを自発的に遵守するであろうことを高い確率で期待できる点が挙げられる[11]。2020年と2021年に総務省が実施したインタビュー調査（総務省2022）は、日本の大手企業は、AI倫理、特にプライバシー、公平性、透明性、説明責任、セキュリティの原則といった総務省が先に発出した非拘束ガイドライン（「国際的な議論のためのAI開発ガイドライン案」（総務省 2017）、「AI利活用ガイドライン」（総務省 2019）、および「人間中心のAI社会原則」（内閣府 2019））に含まれる原則を自らの企業AIガイドラインに自発的に盛り込んでいる状況を明らかにしている（図表３）。

図表３　各種ガイドラインの採用率（N=22）



注：［R&D］、［U］、および［S］の表示は、当該理念がそれぞれ「国際的な議論のためのAI開発ガイドライン案」（総務省2017）、「AI利活用ガイドライン」（総務省2019）、および「人間中心のAI社会原則」（内閣府2019）に記載されていることを示す。
出典：総務省（2022）より筆者作成

9）ハードローアプローチの場合であればインセンティブ両立性を無視して良いわけではない。民間プレイヤーの利潤最大化動機と不整合な成分法を導入した場合、問題となる事業者が国外に拠点を移し、オフショアからサービス提供を継続する、あるいは、当該市場へのサービス展開を諦める、といった対応をとる場合がある。前者の場合は国内利用者を守る規律の枠組みに穴が空き、後者の場合には当該サービス利用で得られるメリットを放棄することになる。

10）Memorandum Opinion of the United States District Court for the District of Columbia, Case 1:22-cv-01564-BAH, Document 24 Filed 08/18/23. https://fingfx.thomsonreuters.com/gfx/legaldocs/lbvgooeoqvq/AI%20COPYRIGHT%20LAWSUIT%20thalerdecision.pdf

11）総務省での各種AIガイドライン策定に協力した筆者自身の経験に基づけば、多くの日本企業は政府が発出する拘束力のない指導文書やガイドラインを自社の事業環境における恒久的な変化と捉え、そうしたソフトなガイドラインに自社の業務を適合させようとする傾向がある。公表されたガイドラインに変更を求めることは、一部の大企業の場合を除き極めて稀である。

第二に、消費者やメディアからのプレッシャーの果たす役割の大きさを指摘できる。政府ガイドラインを遵守しないという事業者の選択は、日本市場においては企業ブランドに対し良い影響をもたらさない。不遵守が実際にトラブルを生んだ場合の悪影響は甚大であり、ネット上でコントロール不能な「炎上」をもたらす可能性さえある。今日のネット経済を支える注目経済（attention economy）のメカニズムにおいて、企業ブランドは最も重要な企業資産の一つであることを考えると、AIサービス提供事業者には非拘束的であったとしても政府ガイドラインを遵守する十分な理由がある。

第三に、ガイドラインに対するこういった対応は、「安全第一」という日本企業の伝統的な企業文化との整合性も高いことも重要な要因であろう。2023年にPwC Japanが実施した調査（PwC Japan 2023）は、AI投資に関し「米国はどの分野でも5割以上の企業が十分なROIを得ている一方、日本は総じて3割以下と、大きく水をあけられている実態が浮き彫り」という結果を得ている。効果の実感が不十分であるにも拘らず、日本企業が非拘束ガイドラインを尊重してAI活用を継続していることは、万が一への危機意識の高さを反映していると考えられる。

5. 国際競争と国際協調

多くの場合、AIサービスはブロードバンドネットワークを介することで国境を容易に越える。多言語対応の精度が近年大幅に改善されているため、日本語で利用しているAIサービスが国内事業者の提供しているものか否かは容易には判別できない。他方、AI規律の実効性という観点からみると、国内事業者が提供しているか否かの区別は重大である。ハードローアプローチの場合は司法管轄権の問題があり、たとえ域外適用を規定したとしても、実効性は国内事業者に対する場合と比較して明らかに劣る[12]。その実効性の水準が、政府への信頼感、市場での評価、企業文化などに左右されるソフトローアプローチについても同様である。消費者を十分に保護するためには、国内事業者と同様の制約を国外事業者にも課す必要がある。サービス提供者側の視点からすれば、規制の内外格差が発生し、海外事業者との間での公平な競争条件が保たれないために不利益を被る。

AIサービスを規律する国内ルールがグローバル市場における自国産業の不利を招く場合、産業政策の観点から、各国政府はより緩い規制を指向するインセンティブを持つ。規制の存在は、その水準に関わらず、提供者の行動を一部制約し、技術開発やイノベーションの可能性に対してはマイナスに作用するため、そうした規制に服さない外国企業との競争では不利益をもたらす。他方、適切な規制に服さないAIサービスは、消費者の利活用リスクを高めるが、より革新的な特徴を採用したり、生産プロセスの簡素化を通じてより高い費用対効果を実現したりすることが可能となるため、サービス提供者にとってはより多くの利益を生み出す可能性がある。そのため、経済成長を志向する規制当局には、消費者へのデメリットが提供者へのメリットよりも小さいと予想する場合、他国に先んじて規制緩和を行うこと、すなわち、より緩い規制水準を目指す「底辺への競争」（race to the bottom）を引き起こす理由がある。囚人のジレンマと同様の構造をもつ不毛な規制緩和競争を抑止する一つの方策が、国際協調である。関係各国が適切なAI規制水準について合意したうえで、抜け駆けを防止する実行力をもつ国際機関の設立が実現できれば、底辺への競争は発生しない。もう一つの解決策は、AIに対する過度の規制緩和の弊害についての情報を広く共有し、各国政府の規制緩和インセンティブを制御することである。底辺への競争の行く末がAIに対する社会からのトラストの喪失をもたらし、消費者からの支持を失うことになれば、サービス提供者が享受する利潤が減少し、AI産業全体に悪影響が生じる。この点を、政府を含む利害関係者が理解すれば、「底辺への競争」を行うインセンティブ自体が減少する。ただし、これらはいずれも膨大な資源を必要とするばかりか、協調に参加しない、もしくは、悪影響を理解しない国が一つでも残ると実効性が保てない点には注意が必要である。

また、仮に、多大な資源を費やして各国政府が合意に達したとしても[13]、そこではハードローアプローチとソフトローアプローチをとる国が存在するため、特に後者の国々において事業者の規律遵守を外部から確認するメカニズムが必要である。この点については、1970年にジョージ・アカロフが論じた「レモン市場」の場合と同じく、保険制度や第三者認証といった解決策が有効である。それら対策は、企業のAI利活用の透明性を高めることにもつながるが、それが機能するためには、開示された情報を正確に理

12）域外適用に関する様々な論点は田宮（2020）に簡単にまとめられている。

13）なお、各国が交渉を行う局面において、規制の理念を記述の中心とするソフトローアプローチは、より詳細を記述するハードローアプローチよりも、議論の叩き台として有効な指針を提供する可能性がある。わが国が、2017年に「国際的な議論のためのAI開発ガイドライン案」を公表し、その後のOECDにおける議論に大きな貢献を果たしたのはその一つの証左である。

解するためのリテラシーを消費者が身につけていることが前提条件となる。

仮に、国際協調により同程度の水準のAI規制が普及すれば、グローバル市場で規制遵守のためのコスト低下が期待できる。そうした環境が実現されなければ、各国毎のルールへの個別対応が求められる結果、国境を超えてサービスを提供する事業者にとって、規律遵守のコストが嵩む。当該コストは、小規模なAIサービス企業ほどインパクトが大きい。コスト高騰により、研究開発の遅延がもたらされかねないとともに、事業収益性への悪影響が生まれれば必要な投資を十分に集めることも困難になりかねない。同時に、潜在的参入者にとっての障壁として機能するため、長期的には、OpenAI、Meta、Facebookなどの大手テクノロジー企業による市場支配の強化がもたらされる。異なる規制が並立することは、各国政府にとって他国の経験に学ぶことを困難にし、規律最適化を阻害する可能性も高い。

6.求められる規律と支援策

どういったアプローチを採用するにせよ、新しいAI規律を導入する場合には、既存の法規制との整合性に十分注意する必要がある。AI以前の規律体系が十分に機能していたという事実[14]を踏まえた場合、AIの導入に伴う政策介入はAIが追加的にもたらしたリスクに対応するものに限定されることが望ましい。AIがもたらしたリスクが既存技術のそれと質的に変わらないものであれば、従来型の対応をまずは考慮すべきであり、既存の法制度で処理できる場合は新規介入の必要はない。AI技術を使っていることを理由に新たな規制を課すと、技術中立性が損なわれる結果、関連イノベーションが必要以上に遅延し、資源配分の効率性が損なわれる。また、AIが質的に異なる新たなリスクをもたらすとしても、政策介入に必要なコストが、政策介入よって実現される便益を上回る場合、介入を行わないことが経済厚生上は有利となる。政策デザインを行う主体が、最先端の技術や市場に関する情報をリアルタイムに知ることができず、情報の非対称性の影響を大きく受けるような場合、政策介入のためのコストが増大する点にも注意が必要である。同じAIであっても利用態様毎に発生させるリスクが異なるため、規制の必要性にばらつきがあるという視点も欠かせない。結果的に、AIに対するルールづくりの対象範囲はそれほど多くはなく、さらに包括的である必要もない。

ただし、上記は、市場メカニズムが十分に機能していることが前提である。AIサービスをめぐって市場メカニズムの機能不全が懸念されるのは、特に情報開示・透明性に関する部分である。一般的なAIリテラシーが低水準にとどまり、社会制度の対応も十分ではない普及初期においては、生産プロセスやアウトプットされるサービス自体にAIが関与していることは、最終利用時に漠然とした不安感を惹起することを通じて、あるいは、第4節で言及した著作権法の不備といった制度的メカニズムを通じて、需要（中間需要を含む）にマイナスの影響をもたらし、非効率な過小均衡を帰結する。AIを利用しているか否かの情報開示は、利用側の合理的な意思決定のために、ひいては、市場効率性の観点から必須となる。ただし、そうした情報開示は需要にはマイナスであるため、サービス提供者側の利潤最大化インセンティブとは矛盾するので、政府による義務付けが不可欠である。もちろん、その反面として、生産プロセスにAIを関与させない提供者にはAIを利用していないという開示を行うインセンティブが生まれることになるが、この場合は、先のレモン市場のケースと同様に、当該開示の信憑性を保証するための仕組みが要請される。アウトプットの作成にAIが関与しているか否かを第三者が正確に判定する手段が現時点では未確立であるため、AIモデルを提供している事業者自身に協力を求める必要がある（Knott et al. 2024）。

情報開示に関わる分野以外での政策介入はAI導入以前の法制度の状況、AIの活用態様、さらにはプレイヤーや政府のAIリテラシーに応じて異なる形をとる。この点、実積（2022）は、AIに関する専門知識やリテラシーがプレイヤーに十分備わっていることが期待されるB2B市場と、それらが十分とはいえない消費者も参加を排除されないB2C市場を分けて考えるべきと主張している[15]（図表4）。前者（B2B市場）については、政策担当者の情報非対称性の弊害を避けるために、市場メカニズムに可能な限り委ねることが適切であるが、後者（B2C市場）につい

14）Carrillo（2020）は「Although the questions posed by AI are numerous and significant, it is important to highlight that there is not a legal vacuum.（AIが投げかける疑問は数多く、かつ重大であるが、法的空白があるわけではないことを強調することが重要である。）［筆者訳］」（p.14）と指摘している。

15）実積（2022）の提案は、AIに適用すべき法的ルールや原則を、①AIの開発を含むすべての人間活動や社会活動に一般的に適用される、命令的な性質を持つもの、②既存法の類推適用によるもの、③AIの特性に対応するために導入されるもの、の三つに分類して議論しているCarrillo（2020）のアイデアと共通点が多い。なお、Carrillo（2020）は、これらの上位ルールとしてAI脅威論に対応するような「legal rules and principles of an imperative nature which apply generally to all human and social activity including the development of AI（AIの開発を含む、人間および社会活動全般に適用される、強制力のある法的規則および原則）［筆者訳］」（p.14）を含む。

B2B市場	B2C市場
B2B市場のプレイヤーはAIの利活用に関して必要となる基本的知識を備えている。 • 必要な知識を具備していないプレイヤーは、競争の結果、最終的には市場から排除される。 求められる政府介入は市場メカニズムを十分に機能させるためのものに限定し、基本的には民間プレイヤーの自発的努力に委ねることが適当である。 • 具体的な介入としては契約の履行を確保する司法システムや紛争処理システムの整備、不正競争や市場支配力の濫用を抑制する競争法上の措置が挙げられる。	一般消費者にとって日々進化を続けるAIに関して十分な知識をもつことは期待できない。そのため、B2C市場では、消費者保護のためのセーフティネットを政府の責任で整備する必要がある。外部性や価値財への対応のための対処も求められる。 • AIが生産プロセスの効率化のみを実現するプロセス・イノベーション型では、消費者保護の基本形は既存制度中にすでに存在している。必要となるのは、既存制度の解釈の明確化であり、AIに関する規制は他の生産技術に対する場合との同等性を保つ必要がある。 　• AIへの要求水準がそれを超える場合は、AI導入の均衡点が最適水準を下回り、逆の場合は最適を上回る。いずれの場合も、資源配分効率性の点から望ましくない。 • AI利活用によって本質的に新しいモノが創出されるプロダクト・イノベーション型の場合、既存の法制度では対処できない。関連するリスクの放置が大きな損害をもたらす場合は新規立法を含む政策介入が必須となる。これまで人間が提供に関わる責任を負うケースしか想定していなかった分野に、AIという新たな提供主体が登場するわけで、その対応はさらに以下の三つに分類できる。 類型Ⅰ：財・サービスの提供にそもそもAIが関与することが許されない類型 　• 宗教的もしくは文化的な要因などから人間のみが最終利用者に対応しなくてはならない分野が典型例である。EUのAI法第5条が定める「禁止される人工知能」はその具体例である。この場合は、当該AIサービスの提供禁止が唯一の選択肢であり、罰則を含む強力な新規立法が必須となる。 類型II：財・サービスの提供にAIの関与が許されるが、提供に関して人間が最終的な拒否権を持つ場合 　• 生成AIでさまざまな創作物を生み出し、それを事業に利用するケースは、この類型に対応する。ここでは、サービス提供に拒否権を有する担当者（人間）もしくは担当企業に一時的には全ての責任を担わせる方向で制度を構築することが適当である。一旦、責任を預かった担当者・企業は、その後、B2B市場における契約関係に基づき、バリューチェーンに関連する開発者や関連企業との間で最終的な責任分担を行う。責任充足には一定のコストが必要で、特定の責任にはプレイヤーごとに得意・不得意があるため、バリューチェーンに所属するプレイヤー全体で責任を分担することで、社会全体として効率的に望ましい状況を達成できる。 類型Ⅲ：財・サービスの提供にAIの関与が許されるが、人間の関与を必要としない完全自律型の場合 　• バリューチェーンに属する関係者、最終ユーザー、第三者を含む社会全体といった広範なステークホルダーの中で誰がどの範囲で責任を負うべきかを決定し、制度構築を行う必要がある。

図表4　最適な政策介入のあり方

ては消費者保護の視点が重要となるためである。

ところで、機械学習・深層学習によってAIを開発するためには、膨大な学習用データとともに、それを効果的に処理するための大量の計算資源が必要となる。特に、モデルの大規模化が進んだ今日では、演算を高速化する特殊なチップ及びそれを搭載したクラウドサービス又はスーパーコンピュータが必須である。加えて、高度なスキルを持つ研究者やエンジニアの投入も不可欠である。最新のAIモデルを開発するためには膨大な初期投資や運営費用を負担する必要があり、資本力のある一部のデジタルプラットフォーム事業者のみが負担可能となりつつある。そのため、特に巨大な投資を必要とする基盤モデルの場合、「将来、無数の基盤モデルが出現するのではなく、インフラ化した基盤モデルに収束し、それをAPIとして利用させる流れが主流となる可能性がある」（競争政策研究センター事務局 2023, p.4）という予測がある[16]。

AI産業を国際競争力の源としたい国々にとっては、学習データの利用可能性を拡充すると共に、高品質な計算資源を安価で利用できるよう民間企業による投資を促し、さらには、AI人材の拡充に注力する必要がある。例えば、わが国の著作権法では、機械学習・深層学習の目的であれば、原則として制約なしに著作物を利用することができると定めている（第30条の4）。また、人材については、2023年10月末に発した米国大統領令（The White House 2023）でAIの専門知識を持つ高度熟練労働者の米国への移住を奨励するなど、国を挙げた支援を行っている例もある。さらに、公正競争の観点からは、一部事業者による過度な市場支配を抑制するため、オープンソース型のサービス提供を支援するとともに、巨大事業者による市場支配力濫用を抑制することも必要である。

7. おわりに

AIという新しい技術は、技術的リスク、社会的リスク、および、人類存続リスクという三種類のリスクをもたらす。深層学習をベースとして構築されたAIが本質的に不完全であり、さらにその挙動がサービス提供者にとっても100％予測可能ではないため、それらリスクの一部は利用者を含む社会全体で引き受ける必要があり、その前提条件としてのトラストの醸成が求められる。そのためにはAIの開発、利活用をターゲットにした新たなルールの導入が必要とされる場合があるが、それにはハードローアプローチとソフトローアプローチ、およびそのハイブリッドのアプローチがあり、各国は長所・短所を考慮して最適な政策パッケージをデザインすることを求めら

16）現在、AIサービスの提供には二種類の形態がある。ひとつは、ソースコードが公開され、無償のライセンスによりライセンシーによる改良・改変、複製、再配布が可能なオープンソースソフトウェアとして提供されるものである。もう一つは、モデルのソースコードは公開されることはなく、利用者は、改良・改変等にも制約が課せられるプロプライエタリ・ソフトウェア（クローズドソース・ソフトウェア）として、APIによって提供されるものである。オープンソースによる提供では、様々な主体が関与して改良、改善などを行うことが想定されているため、多様なサービスが生まれやすく、品質改善が早い。

れる。わが国では、ソフトローアプローチをベースとするルール形成を最適とする条件が満たされているようである。また、各国が構築するルールは基本的に国内において最大の効力を発揮し、さらに、ルールはAIサービス事業者の行動を多少ともあれ制約する効果を有するため、国際社会では「底辺への競争」が発生する可能性があるが、その発生を抑止し、各国規制の実質的同一化を進める必要がある。具体的なAIルールづくりにおいては、既存制度との整合性を確保し、さらに市場毎の特徴を考慮する必要もある。

さて、AIサービス提供者の説明責任を確立してAIへのトラストを社会で醸成しようしても、利用者が提供者の説明を理解できる能力がなければ意味がない。あるいは、本稿で提案するような仕組みがサービス提供者を有効に規律し、望ましい品質のAIサービスが提供されたとしても、改善された品質が利用者に正しく評価され、高い支払意思額（willingness to pay、WTP）に帰結し、利潤拡大に貢献することがなければ、長期的に維持可能な最適均衡とはならない。そのため、本稿で取り上げたサービス提供者への施策とともに、利用者への施策を同時に展開し、AIに対する基礎的なリテラシーを国民の教養とする必要がある。その点については、2022年より高等学校において「情報I」を必修科目として設置したことはプラスの効果をもたらすであろう。今後は同様の基礎的な学習を、情報Iに触れることがなかった世代にも拡張することが必要かもしれない[17]。

（2024年11月23日脱稿）

追記

原稿提出後の2024年12月26日、AI戦略会議AI制度研究会が中間とりまとめ（案）を公表した。そこでは、多くの回答者がAI規制を必要とし、特にAIの悪用や犯罪に対する対策強化を求めているとする意識調査結果を紹介したうえで、多様なリスクに対応するための司令塔機能の強化や安全性・透明性を確保するための各種対応を政府に求めている。また、これとは別に政府部内ではAIの不正利用に対応する法律が準備されているという報道もなされている。このように日本政府は基本原則自体をソフトローで維持しつつも、それを補完するために特定分野でのハードローの充実を図っている。一方、EUは、AI法の準拠標準として機能するソフトローである行動規範（Code of Practice）の策定を、世界中の専門家の協力の下で進めており、ハードローとソフトローから構成されるEU流政策パッケージがその全貌を現しつつある。また、米国では、連邦レベルでは自発的コミットメントをベースとした施策展開が維持される一方、成原他（2024）が示すとおり州レベルにおいて関連立法が相次いでおり、全米的にはハイブリッド化が進んでいる。

こうした動きは、環境変化のペースが速く、さらに政府を超える力を持ちうる巨大プレイヤーが跋扈するAI市場に、各国政府が急ぎ対応していることの表れであり、スタート時のアプローチは異なっても、最終的な帰着点は非常に似通ったものになることを示唆しているように思われる。今後の展開に注目したい。

参考文献

荒堀淳一（2020）「AIの責任と倫理（第3回）AI倫理に対する企業の取組み（2）」『NBL』, 1172, 67–71.

内田尚和、鍛忠司、Blake, N.、間瀬正啓、大橋洋輝、Ghosh, D.、Gupta, C.、直野健、高田実佳（2021）「AIのトラストとガバナンスを支える研究開発」『日立評論』特別増刊号, 20–27.

競争政策研究センター事務局（2023）「生成AIを巡る独占禁止法上及び競争政策上の論点」第22回国際シンポジウム（2023年11月9日　）. https://www.jftc.go.jp/cprc/events/symposium/2023/231109sympo1.pdf

経済産業省（2021a）「我が国のAIガバナンスの在り方 ver.1.1 AI原則の実践の在り方に関する検討会報告書」. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_1.pdf

経済産業省（2021b）「Governance Innovation Ver.2 アジャイル・ガバナンスのデザインと実装に向けて」. https://www.meti.go.jp/shingikai/mono_info_service/governance_model_kento/20210730_report.html

鹿野利春、実積寿也、田中恵子、遠山紗矢香、豊福晋平、水越一郎、山形巧哉（2024）「すべての人に『情報I』の内容を！ DXのスタートラインとしての国民的素養」GLOCOM六本木会議. https://www.glocom.ac.jp/news/news/9538

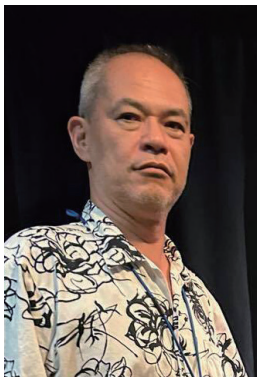
実積寿也（2022）「AIガイドライン、トラスト、ガバナンス―最適なルール策定に向けて―」『Nextcom』, 50, 4–12.

新保史生（2020）「AI原則は機能するか？―非拘束的原則から普遍的原則への道筋」『情報通信政策研

17）「情報I」の内容を対象世代以外にも拡張していくべきという主張については、鹿野他（2024）にまとめている。

究』, 3(2), 53–70.
総務省（2017）「国際的な議論のためのAI開発ガイドライン案」AIネットワーク社会推進会議. https://www.soumu.go.jp/main_content/000499625.pdf
総務省（2019）「AI利活用ガイドライン～AI利活用のためのプラクティカルリファレンス～」AIネットワーク社会推進会議. https://www.soumu.go.jp/main_content/000637097.pdf
総務省（2022）「報告書2022～『安心・安全で信頼性のあるAIの社会実装』の更なる推進～」AIネットワーク社会推進会議. https://www.soumu.go.jp/main_content/000826564.pdf
田宮寿人（2020）「行政法の域外適用に関する立法政策上の諸論点―個人情報保護法・令和2年改正の視点から―」『情報法制研究』, 8, 63–74. https://doi.org/10.32235/alis.8.0_63
内閣府（2019）「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. https://www8.cao.go.jp/cstp/stmain/aisocialprinciples.pdf
中川裕志（2019）『裏側から視るAI脅威・歴史・倫理』近代科学社.
中川裕志（2020）「AI倫理指針における課題」『人工知能』, 35(6), 845–854.
中川裕志（2021）「6. デジタル社会におけるAIガバナンス―倫理と法制度―」『情報処理』, 62(6), e34–e39.
成原慧、実積寿也、小熊美紀（2024）「AIをめぐる米国の政策と法的対応―連邦および州の動向―」『情報法制研究』, 16, 34-48. https://doi.org/10.32235/alis.16.0_034
羽深宏樹（2023）『AIガバナンス入門：リスクマネジメントから社会設計まで』早川書房.
PwC Japan（2023）「2023年AI予測　米国に離されるAI活用、挽回のカギは生成AI」, 2023年6月15日. https://www.pwc.com/jp/ja/knowledge/thoughtleadership/2023-ai-predictions.html
Bathaee, Y. (2018) The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology*, 31(2), 889–938.
Berryhill, J., Heang, K.K., Clogher, R., and McBride, K. (2019) Hello, World: Artificial intelligence and its use in the public sector. OECD, November 2019. http://oe.cd/helloworld
Bostrom, N. (2014) *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
Carrillo, M.R. (2020) Artificial intelligence: From ethics to law. Telecommunications Policy, 44(6), 101937. https://doi.org/10.1016/j.telpol.2020.101937
Figueroa-Armijos, M., Clark, B.B., and da Motta Veiga, S.P. (2023) Ethical perceptions of AI in hiring and organizational trust: The role of performance expectancy and social influence. *Journal of Business Ethics*, 186, 179–197. https://doi.org/10.1007/s10551-022-05166-2
Future Life Institute. (2023) Pause giant AI experiments: An open letter. March 22, 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/
Grimmelikhuijsen, S. (2022) Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, 83(2), 241-262. https://doi.org/10.1111/puar.13483
Hosmer, L.T. (1995) Trust: The connecting link between organizational theory and philosophical ethics. *Academy of Management Review*, 20, 379–403. https://doi.org/10.2307/258851
Kim, T.W. and Routledge, R.R. (2022) Why a right to an explanation of algorithmic decision-making should exist: A trust-based approach. *Business Ethics Quarterly*, 32(1), 75–102. https://doi.org/10.1017/beq.2021.3
Knott, A., Pedreschi, D., Jitsuzumi, T., Leavy, S., Eyers, D., Chakraborti, T., Trotman, A., Sundareswaran, S., Baeza-Yates, R., Biecek, P., Weller, A., Teal, P.D., Basu, S., Haklidir, M., Morini, V., Russell, S., and Bengio, Y. (2024) AI content detection in the emerging information ecosystem: New obligations for media and tech companies. *Ethics and Information Technology*, 26, 63. https://doi.org/10.1007/s10676-024-09795-1
Kurzweil, R. (2005) *The singularity is near: When humans transcend biology*. The Viking Press.
Lee, N.T., Resnick, P., and Barton, G. (2019) Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumerharms/
Nagy, S. and Hadjú, N. (2021) Consumer acceptance of the use of artificial intelligence in online shopping: Evidence from Hungary. *Amfiteatru Economic*, 23(56), 155-173. https://doi.

org/10.24818/EA/2021/56/155
Noti-Victor, J. (forthcoming 2025) Regulating hidden AI authorship. Virginia Law Review, 111. Cardozo Legal Studies Research Paper No. 2024-29. http://dx.doi.org/10.2139/ssrn.4909907
Pirson, M., Martin, K., and Parmar, B. (2019) Public trust in business and its determinants. *Business & Society*, 58(1), 132–166. https://doi.org/10.1177/0007650316647950
Quinn, R.A. (2021) Artificial intelligence and the role of ethics. *Statistical Journal of the IAOS*, 37, 75–77. https://doi.org/10.3233/SJI-210791
The White House. (2023) Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. 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/
Vinge, V. (1994) The coming technological singularity: How to survive in the post-human era [Paper presentation]. Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace, Westlake, OH, United States. https://ntrs.nasa.gov/citations/19940022856



中央大学 総合政策学部 教授
実積　寿也（じつづみ・としや）

郵政省、長崎大学経済学部、日本郵政公社、九州大学大学院経済学研究院を経て2017年より現職。総務省情報通信政策研究所特別研究員、Global Partnership on AI専門家委員。情報通信エコシステムの事象について主として経済学の観点からアプローチ。現在の研究テーマは、ネット中立性、AI政策、プラットフォーム規制。