【特集　デジタルガバナンスの未来】

2　AI ガバナンスのあり方

Guest Speaker

佐藤 一郎（さとう・いちろう）
大学共同利用機関法人
情報・システム研究機構 国立情報学研究所（NII）
情報社会相関研究系 教授

1991 年慶応義塾大学理工学部電気工学科卒業。1996 年同大学大学院理工学研究科計算機科学専攻 後期博士課程修了。博士（工学）。1996 年お茶の水女子大学理学部情報学科 助手、1998 年同大 助教授、2001 年国立情報学研究所 助教授を経て、2006 年から現職。このほか、デジタル庁「政策評価有識者会議／行政事業レビュー（旧事業仕分け）」座長、経産省・総務省「企業のプライバシーガバナンスモデル検討会」座長他を歴任。テレビ朝日系列番組『仮面ライダー（ゼロワン）』（テレビ朝日系列、2019 年 9 月～2020 年 8 月）の AI 技術アドバイザーも務めた。

■　この章の問題意識　■

近年、AI 技術の目覚ましい進歩が進み、AI の性能が人類の知能を上回るシンギュラリティ（技術的特異点）を迎えるという 2045 年問題が話題となった。同時に、AI の開発・利用の両面から「AI のあるべき姿」が各方面で頻繁に議論されるようになってきた。ただし、これまでの議論はあくまで理念的なものが中心だった。ところが、生成 AI の登場により AI について人々が抱いてきた懸念が現実のものになったと受け止められ、欧州や中国では AI に関する法律が制定され、米国においても大統領令に基づく規律の検討が進められているなど、AI に向き合うための制度的な枠組みづくりが急ピッチで進んでいる。

しかしながら、現実に AI が我々の社会経済活動の一部として組み込まれるようになってきているからこそ、AI の利点を最大限生かすための環境整備と同時に、AI のリスクを冷静に見極めて対処していくためのルールの策定・稼働が求められる。また、こうしたルールを考える際、AI そのものが継続的に進化している moving target であることも忘れてはならない。我々は AI を今後どのように統治していくのか、また、できるのか。AI ガバナンスのあり方について骨太な議論が求められている。

聞き手＝**谷脇 康彦**　デジタル政策フォーラム 代表幹事

G7 伊勢志摩サミット以降、欧米から遅れ始めた

谷脇　AI ガバナンスのあり方が大きな問題となっています。例えばユーラシアグループは「2024 年 10 大リスク」[1]、とりわけ 10 年後を視野に入れた長期的なリスクの一つとして「**AI ガバナンスの欠如**」を挙げています。最近の AI 関連技術の進化の速さは AI ガバナンスのためのルール策定の動きを超えているとの懸念も出ています。そこで、そもそも AI を統治することが可能なのか、まず総論的に伺いたいと思います。

佐藤　AI ガバナンスは大切ですが、その前にガバナンスで何を守るかを決める必要があります。AI ガバナンスを確立することはとても難しいですし、AI ガバナンスを確立したところで AI に関わる様々な問題が一掃されるわけでもありません。「コーポレートガバナンス」を作っても企業の不祥事はいろいろなところで起きているわけで、ガバナンスを作っても問題は生じます。基本的なスタンスとして、ガバナンス万能論に偏るべきではないと思います。

「コーポレートガバナンス」は長年の積み重ねで確立されました。また、個人的には、ガバナンスは失敗によって整備されていくものだと考えています。AI のガバナンスを確立するには、我々は幸か不幸か、AI に関する失敗の経験が少な過ぎるのです。今ガバナンスを作ったとしても、すぐに失敗して、また作り直すことになります。ガバナンスというのは永遠に作り直していくものなのかもしれません。

従って、AI ガバナンスの確立を目指すことは大切ですが、現時点で AI ガバナンスの確立を前提にしていろいろな話を進めるのは少々危険だと見ています。

谷脇　日本は AI の議論への着手は他国よりも早かったのですが、EU（欧州連合）が AI 法をつくり規制の域外適用も予定しており、中国も生成 AI に関する法律を定めて“中華 AI”を推進しています。共同規制に軸を置いているアメリカも大統領令を出し、連邦機関において技術基準の策定などが進み、ガイドライン中心の日本はどんどん追い抜かれてしまっているように思います。

佐藤　2016 年の G7 伊勢志摩サミットの時点までは、日本、ヨーロッパ、アメリカは横並びでしたが、その後に OECD（経済協力開発機構）で AI に関する国際的な政策ガイドラインの検討・議論が進められ、2019 年 5 月に「**AI に関する OECD 原則**」[2]が採択されました。

私は、ちょうど同じ時期に OECD の別委員会（研究データ倫理）の日本代表を務めていたので、AI 原則に向けた議論を横目で見る機会があったのですが、率直に言って、日本はそこでの議論についていけていませんでした。ヨーロッパ各国もアメリカも専門家を送り込んでいましたが、日本は必ずしもそうではありませんでした。

「AI を規制する」ということが議論の主題だったはずなのに、日本国内ではいつの間にか「**AI の倫**

理」の話にすり替わっていました。日本では企業にも政府にも「規制があるとイノベーションは生まれない」という奇妙な思い込みが蔓延しているからかもしれません。日本の議論はちょっとふわふわした方向に傾いていました。

OECD のような国際的な場での実務議論では、自国の主張を通すには、他国の主張に対して理詰めで反論する必要があり、そのためには自国内での議論の積み重ねが極めて重要になります。EU は AI を規制する法律を作ることを目指していました。法律を作るにはかなり緻密な議論の積み重ねが必要であり、EU はそれを実践してきたと見るべきでしょう。アメリカは「規制反対」の基本スタンスをとった点では日本と歩調が合っているように見えましたが、舞台裏ではかなり深い議論を重ねており、「**AI 権利章典**（**AI Bill of Rights**）」[3]という指針を作り、それが「**責任ある AI**」に関する議論につながっていきます。

その頃から欧米との差が開き始めたように思います。

■AIに関するOECD原則の概要

（1）AIは、包摂的成長と持続可能な発展、暮らし良さを促進することで、人々と地球環境に利益をもたらすものでなければならない。

（2）AIシステムは、法の支配、人権、民主主義の価値、多様性を尊重するように設計され、また公平公正な社会を確保するために適切な対策が取れる－例えば必要に応じて人的介入ができる－ようにすべきである。

（3）AIシステムについて、人々がどのようなときにそれと関わり結果の正当性を批判できるのかを理解できるようにするために、透明性を確保し責任ある情報開示を行うべきである。

（4）AIシステムはその存続期間中は健全で安定した安全な方法で機能させるべきで、起こりうるリスクを常に評価、管理すべきである。

（5）AIシステムの開発、普及、運用に携わる組織及び個人は、上記の原則に則ってその正常化に責任を負うべきである。

■各国政府に対するOECDの提言

―信頼できるAIのイノベーションを刺激するために、研究開発への官民投資を促進する。

―デジタルインフラとテクノロジーでAIエコシステムとデータと知識の共有メカニズムの利便性を高める。

―信頼できるAIシステムの普及に道を開く政策環境を創出する。

―人々にAIに関わる技能を身につけさせるとともに、労働者が偏りなく転職できるよう支援する。

―情報を共有し標準を開発し、責任あるAIの報告監督義務を果たせるように、国際的、産業部門横断的に協力する。

生成 AI の楽観シナリオが浮かんでこない

谷脇　その点については、近著『**ChatGPT は世界をどう変えるのか**』[4]でも指摘されていますね。佐藤先生が一般の方々に最も伝えたかったことは何だったのでしょうか。

佐藤　**生成 AI が社会に与える影響について考えて頂きたかった**、ということに尽きます。POC（Proof of Concept、概念実証）の段階では「生成 AI って面白いな」「なんだか便利そうだな」と利便性ばかりを見ていたとしても、実用化するとなればリスクや影響に対峙しなければいけません。リスクは技術で低減できるものもありますが、生成 AI は社会に対してかなり大きな影響を与えるものなので技術だけでは解決しません。生成 AI を社会がどう受け入れるのか、あるいは拒絶するのかということについてスタンスを定めなければなりません。社会にどういう影響があるのか、予測も交えてお伝えし、皆さんに考えていただこうというのが執筆の狙いでした。

佐藤 一郎　国立情報学研究所（NII） 教授

谷脇　AI については、「生産性を上げて付加価値も上げる」とか「人間の仕事を奪う」とか、光と影の両面、様々な言説が唱えられてきました。そして最近では、AI が生成したコンテンツがネット上に増え、それらを AI が学習していくと「Model Collapse（モデル崩壊）」が起こり、情報の質が落ちるという指摘も出てきました。AI は情報空間全体にどのような影響を与えるのでしょうか。

佐藤　局所的にはともかく、社会全体として本当に生産性や利便性が上がるのかについてはかな

り疑問です。

AI が生成するコンテンツは増える一方ですが、読まれない無駄な情報がどんどん増えていきます。例えば、業務日誌のまとめに生成 AI を活用すれば、これまでポツポツと箇条書き程度だったものをきちんとした文章に整えてくれます。ただし、本質的に内容は何も変わっていないのに文章の量だけが増える。それを上司は毎日読まされるわけです。生成 AI は読み手側の負担を増やしてしまう技術なのです。ある程度まで利用が進むと、読み手側は生成 AI ではなく人間が編集した情報を選別して読むような方向に戻っていくような気がします。

また、生成 AI が生成するコンテンツの品質は、今後、上がるとは限らない。下がる可能性もあります。その理由は生成 AI が作ったコンテンツを生成 AI が学習するからです。品質も全体的に低下します。そうした質の低い、薄いコンテンツがネットに流れ、AI が学習し、また新しいコンテンツを生成するというサイクルが繰り返されていくと、「**AI の共食い**」のような状態に陥ってしまい、AI の学習モデルそのものが破綻してしまいます。

生成 AI は利用者が知りたいことに直接的に答えてくれるのでとても便利です。ウェブ検索は知りたいことに答えてくれるわけではなく、知りたい情報が載っているウェブページを教えてくれているだけ。あくまで間接情報であって、知りたいことにたどり着くためにはウェブを見にいって、そこの情報を読み込まなければなりません。生成 AI によってその手間を省けるとしたら、利用者は検索しなくなり、ウェブを見なくなっていくでしょう。ウェブコンテンツの質の悪化がその傾向に追い打ちをかけます。

すると、ネット広告の効果が下がります。当然、広告料、掲載料が下がります。ネット広告の代理店ビジネスが大打撃を受け、ネット広告を表示させることで収益を上げていたネットサービス事業者の収入も減ります。広告収入で成立してきた無料ネットサービスは SNS なども含めて立ち行かなくなるでしょう。これまでのように、誰もが無料で様々なネットサービスを享受できる時代は終わりを告げるかもしれません。一般の利用者の情報発信機会も減り、インターネット全体の利便性が失われ、活気がなくなっていくことになるかもしれません。

どう考えても、生成 AI に関して楽観的なシナリオが浮かんでこないのです。

プラットフォーマーの AI への取り組みに戦略的温度差あり

谷脇　マイクロソフトはオープン AI（ChatGPT）、グーグルは Bard、アマゾンは Titan、メタは Llama といずれも生成 AI に取り組んでいます。無料サービスを提供し、利用者の個人情報を収集し、広告収入で巨利を得てきた巨大プラットフォーマーです。自分たちのビジネスモデルを崩壊させかねない生成 AI を推進するのは、矛盾しているように見えます。

佐藤　巨大プラットフォーマーといっても、各社各様で事情が異なります。マイクロソフトはネット広告

ビジネスが必ずしもうまくいっていないので失うものはなかった。最も積極的に動いているのも頷けます。

他方、生成 AI が"**検索スルー**"をもたらせば一番打撃を被るであろうグーグルは、AI 開発を進めていることをアピールしながらも躊躇が見え隠れします。戦略的に様子見をしていると言った方がいいかもしれません。グーグルにとっては、ネット広告収入への影響を考えれば、生成 AI、対話 AI の普及が遅れるほうが望ましいと考えている可能性があります。

また、ネット広告への依存度が高いグーグルやメタが生成 AI の提供で先行した場合、「生成 AI の出力を操作した**ステルスマーケティング**ではないのか？」と疑われ、批判にさらされる可能性があります。例えば、グーグルの生成 AI で料理のレシピについて聞くと、答えにいつも特定のメーカーの、特定の調味料が含まれていたとしたら、それはステマなのではないかという疑念が生じるでしょう。

こうして見ると、巨大プラットフォーマーの生成 AI に対する向き合い方にはかなりの温度差があり、マイクロソフト系が生成 AI で先行したのは必然だったのかもしれません。ただし、その先行で生成 AI の勝負に決着がついたというわけではないと思います。

情報統制、言語・文化のデカップリングが進む恐れ

谷脇　AI による偽情報・誤情報の拡散リスクについてどう対処すべきでしょうか。ステルスマーケティングの話が出ましたが、AI の中立性や客観性はどうやって担保されるのでしょうか。

佐藤　生成 AI の出力の中立性・公平性をいかに確保するかについては、今後、議論になってくると思いますが、現状では良い方法がありません。出力を歪める方法の一つは**アルゴリズムを歪める**ことですが、これは技術的に検証することができます。しかし、**学習データ・訓練データの偏り**については判断が難しい。ですから、「偏っているかもしれない」「ステマかもしれない」という疑心暗鬼を抱きながら使うしかない状態が続くかもしれません。

皆が中立・公正なものを使いたいとも限りません。フィルターバブルが指摘されるように、自分にとって気持ちが良い情報空間にいたいと思う人はたくさんいます。自分にとって心地良い偏りは、すんなり受け入れられてしまうのです。

強権的な国家には、生成 AI は魅力的な技術です。国民がウェブをあまり見なくなって、生成 AI に頼るようになってくる、しかも国家にとって都合の良い情報を出力するように操作できるとなれば、国家にとって都合の良い情報空間に国民を閉じ込めておけるので、強力な情報統制装置になります。

この問題はさらに根深い危険性があります。生成 AI を構築・提供するのは技術的に簡単ではないので、すべての強権的国家が政権に都合の良い生成 AI をつくって国民に提供できるとは限りません。そうすると、偏った生成 AI を構築できる強権的な国から、個別にチューニングされた生成 AI の提供を受けることになるかもしれません。生成 AI を提供した強権国が、その提供を受けた強権国の情

報統制を担うことになり、その国の世論を操作できることになります。その生成 AI は、提供した国を批判するような情報は制限するでしょうから、ある特定の強権的な国が提供した生成 AI が多くの国に広まれば広まるほど、生成 AI を提供した国が望まない情報を与えられない人が他国にも増えていくことになります。このようにして生成 AI が国家統制の手段から国家覇権の手段になって、生成 AI ごとにブロック化された情報空間が形成され、**情報統制による世界のデカップリング**が進むことになるかもしれません。

もう一つの深読みは、生成 AI によって人類は**「バベルの塔」のように世界共通語を失う**かもしれないということです。ChatGPT は英語や日本語を含めて複数の言語で利用できます。ChatGPT に日本語で質問すると、それが海外に関する情報でも ChatGPT は日本語で回答してきます。ウェブ検索の場合は日本語以外のウェブサイトもリスト化されるのとは違います。

これまでは、何か知りたいことがあるとき、海外の情報なら英語や現地の言葉で読まなければならないということが多くありました。外国語、特に世界共通語である英語を学ぶ理由の一つがこの「知識の習得」でした。ところが、生成 AI を使えば母語で質問して母語で答えてくれるとなると、知識や情報を得るために外国語を勉強する必要性が減ってくることになります。少なくとも、世界共通語の英語の地位・重要性は下がると考えられます。

その状況は、塔を作って神に挑戦しようとした人類に対し、通じ合わない異なる言葉という罰が与えられたという旧約聖書の物語に通じるものがあります。

強権国にとっては都合が良いのです。「科学技術や先端知識を欧米先進国から学び取るために国民の英語学習をやむなく認めているが、なまじ英語の情報に触れるから国家にとって不都合な思想に染まる国民が出てくる」と考えているような国家であれば、英語を知らなくても欧米の先端技術を採り入れられるのなら英語教育をしないという判断をしてもおかしくありません。先ほどの情報統制によるデカップリングだけでなく、**言語・文化における世界のデカップリング**が進む恐れがあります。

谷脇　私も自著『**教養としてのインターネット論**』[5]に、インターネットの分断が進んでいるということを書きました。西側先進国と中国やロシア、それぞれの陣営のインターネットに対する考え方、国の関与の仕方が全く違う。両陣営の間に「デジタルのベルリンの壁」ができつつあり、それを乗り越えるのが難しくなっている。そこにグローバルサウスといった新興勢力も加わって、フラグメンテーション（断片化）が決定的となり修復不能になってしまうのではないかと危惧しています。

佐藤　私もそう思います。外国語を学ぶということさえしなくなってくると、世界がつながっていなかった中世以前に戻ってしまうような恐れがあります。また、英語が世界共通語になったのは第二次世界大戦以後であり、英米という主要戦勝国の言葉が世界標準語になったと言えます。生成 AI によってその英語の地位が下がるとしたら、それは第二次大戦後の世界スキームが終わることを意味するかもしれません。

アメリカも EU も AI 規制の実務レベルまで詰め切れていない

谷脇　危機感を新たにしました・・・。偏った AI を生み出さないために、EU やアメリカは法や自主規制の網をかけ、説明責任や透明性を事業者に求めようとしています。理念としては理解できるのですが、AI に対する第三者の検証が、真に客観的に行い得るのでしょうか。実効性に乏しいのではないかと感じるのですが。

佐藤　そうなる可能性は高いと思います。AI の中立性を検証できる範囲はとても限定的で狭いのです。アルゴリズムについては学習データと出力の相関を見ればある程度分かります。しかし、学習データそのものの中立性については検証が非常に難しい。日本の政権・与党における議論はアメリカの議論の影響を色濃く受けているのですが、当のアメリカの基準作りは難航必至なのです。

2023 年 10 月にバイデン大統領は「**人工知能の安心、安全で信頼できる開発と利用に関する大統領令**」[6]を発令しましたが、安全性とセキュリティの新基準づくりについては商務省傘下の**国立標準技術研究所**（**NIST : National Institute of Standards and Technology**）に丸投げしたかっこうになっています。NIST がどのような基準を作るのか全く見えてきません。アメリカの動きに追従するなら、まずは NIST の新基準がどうなるかを見極めてからということになるのですが、おそらく、NIST の担当者は頭を抱えていることでしょう。

ヨーロッパに関しては、EU の AI 規制の検討は初期段階では AI を組み込んだ「製品」の安全性から入ったのですが、最終的なまとめの段階では製品を前面に出さず「AI のリスク」に応じて規制を変えるという整理をして、結果的に自由度を上げました。おそらく検討段階で生成 AI は議論のスコープに入っていなかったはずですが、リスクの考え方でうまく対応したというか、うまく一時避難できたというところです。実際の法執行の段階になると、いろいろと悩ましい判断を迫られることになると思います。

アメリカもヨーロッパも、実務レベルのところについては詰め切れていないので、日本がその部分について検討を深められれば意見を聞いてもらえると思うのですが。現状、残念ながらそのような議論は行われていません。

まずは AI をしっかり定義することから

谷脇　AI の透明性や多様性を確保するためにオープンソースを指向するという可能性もあると思いますが、他方、オープン性を確保することで WormGPT や FraudGPT のような悪意をもった AI 活用が増える恐れはないでしょうか。

佐藤　現時点で、オープンソースが良いのか、悪いのかという結論めいたことは言えません。アメリカの巨大プラットフォーマーでも対応が分かれています。

マイクロソフトとグーグルは AI 技術の公開には慎重な姿勢を示す一方、メタは生成 AI を含めてオープンソースとすることで外部を巻き込んだ研究開発を指向しています。これまでオープンソースはソフトウェアの普及と発展に有効とされてきましたが、生成 AI のように多用途に使える技術の場合、フェイク情報や不正アクセス支援などに悪用される可能性も高いので、オープンソースの是非が見直される可能性はあると見ています。

マイクロソフトもグーグルも、オープンソースそのものに対して全く否定的なわけではなく、分野によってはかなり積極的に取り組んでいます。言い方を変えると、うまく使分けています。そうした事業者が AI のオープンソース化については懐疑的になっているという事実をしっかり受け止めるべきだと思います。AI のリスクや悪影響を止める手段がない以上、オープンソース化に一定の制限をかけるのは避けられないと思います。画像認識のような特定用途のものは別にして、汎用的なものについては慎重な対応が必要です。

余談を二つほど。一つは AI のリスクを考える場合、「AI 内部で安全性を高める視点」と「AI の外側で防御する視点」があるということです。AI を規制するということだけではなく、AI が何か問題を起こしたときにどのように被害を食い止めるか、**防御のための仕掛け**も含めて議論すべきだと思います。

もう一つは AI を恐れるのなら、**AI というものをまずしっかり定義すべき**だということです。AI とは何かをふわっとさせたまま議論していることが多いように感じます。多くの人が AI と呼んでいるものは、たいてい実用化前のものです。実用化されると AI ではなく「〇〇処理」とか「△△システム」という名前で呼ばれるようになります。20 年前の AI は「推論」とか「検索」でした。今、検索エンジンを AI と呼ぶ人はほとんどいません。AI という言葉が対象とするものは時と共に変わっていくものなので、制度設計する前に対象、つまり AI を定義づけるということをすべきだと思います。

お手本は EU の AI 規制

谷脇　各国では AI に課すべき規律として、「法的規制」「共同規制」「自主規制」などが検討されていますが、どのようなかたちの規律が適しているのでしょうか。

佐藤　AI の進歩の速さを考えると自主規制や共同規制的な方法が選択肢となりますが、必ずしも機能するとは限りません。例えば共同規制の場合、AI を担っている企業のスポンサーは大手プラットフォーマーであり、国の指導においそれとは従わない。そうなると法律に基づくハードローを組み合わせたエンフォースメント（行政上の強制執行）が必要になってきます。そのレベル感は課題に応じてとしか言えませんが、まずは AI の文脈以外においても国家と大手プラットフォーマーの関係性を見直して

いくべきだと思います。

日本にとってお手本になるのは EU の AI 規制でしょう。まずは EU の AI 規制をよく研究して、取り入れるところは取り入れ、問題があれば問題があるということを EU にフィードバックする。EU も EU 域内だけで AI 規制が有効に働くとは考えていないでしょうから、日本が仲間になると言えば聞く耳を持つと思うのです。EU の基本方針に合わせつつ、言うべきことははっきり言うというスタンスをとるのが現時点での最善策、というか、それ以外に打てる手がないように思います。

EU の AI 規制を手本と位置づけるのは、私は個人情報保護法の改正にも関わってきたのですが、その立場で見えたのは、**日本が何を言っても GAFA は聞く耳を持たない**ということです。EU との共同戦線を張るのが日本の現実解です。EU は大陸法なので法体系的にも近い。コモンローである英米とはやはり違います。

谷脇　サイバー空間には国境がありませんから、規律の国際的整合性が必要ですが、これをどのように確保していくべきでしょうか。AI に関する国際機関を設立すべきという意見もありますが有効でしょうか。

佐藤　国際機関を作って機能するかどうか、私にはよく分かりません。国際連合だって十分機能しているとは言い難い状況において、新たな機関を作ることにどれほどの意味があるのかという懸念はあります。仮に巨大プラットフォーマーにとって「規律を守らないこと」に利益があるのなら、国際機関を作って規律を定めても弱い抑止力程度にしかならないでしょう。グローバルにビジネスを展開する事業者の場合、国ごとの個別対応はコストが高くなるので、より厳しい国の規律に合わせるはずです。国際機関が運用する国際的規律にはそれ以上の厳しさを持たせないと形骸化します。現時点では、国際機関を作る意味を見出すことは難しいと思います。

人間が AI に介入する必要はある

谷脇　負の側面について、もう一つお聞きします。**データ駆動社会**（**Data Driven Society**）において AI が実装されることで、個別化・自動化・最適化が進みます。これは経済学的には効率的な資源配分に貢献すると考えられますが、他方、差別の助長、少数意見の切り捨て、説明責任の欠如といった問題が顕在化するのではないでしょうか。

佐藤　その通りだと思います。AI に限らず、データに基づくシステム全般に言えることですが、全体データを正確に反映した判断が社会的に適切とは限りません。AI の場合、過去のデータを学習してい

る以上、AI による判断は過去において多数・優勢だった対象に有利に働き、**現状の社会課題を固定・拡張**する可能性が高いのです。

例えば与信評価をデータに基づいて行うと、男性の方が女性よりもスコアが良くなります。それは総じて男性の方が女性よりも高収入という現実があるからです。その与信評価に従い男性を優先して融資すると、与信評価はますます男性有利になっていきます。男女は平等であるべきだという理念に基づいてそうした性差を減らそうとすれば、アルゴリズムを恣意的に改変するか、AI に学習させるデータを歪めて女性の評価を上げるような操作を加えなければなりません。

操作を加えてデータを歪めたうえでの出力が公平なのか、操作をしないで現実のデータに基づく出力が公平なのか――。**学習データに意図的に手を加えられた AI による出力は公平中立で信頼できるものと言えるか**という新たな問題が生じます。もはや技術だけの問題ではなく、社会としてどういう選択をするかという議論になります。

また、学習データにある程度以上の「量」がないと、AI には判断ができなかったり、誤差が多くなったりします。これが、少数データを足切りするような方向に働けば、人間社会における少数意見や少数派の立場を侵害する、あるいは、多数意見や多数派の立場を有利にした結果として相対的に少数派の立場を侵害するようなケースが発生し得ます。

AI が学習するのは過去のデータであり過去の状況を固定化するという特性がある以上、**人間が恣意的に介入する必要は「ある」**と思います。そうした介入をどこまで許すのか、それをどのようにガバナンスするのかが問われることになります。

人間が AI に順応するような社会にしないために

谷脇　デジタル政策フォーラム（DPFJ）[7]では、まさに AI ガバナンス、データガバナンス、セキュリティガバナンスの 3 つの要素で構成される「デジタルガバナンス」について検討を深めていこうとしています。そもそもこれからのデジタル技術を人類は制御することができるのかという問題意識をベースにして、真剣な議論を誘発したいと考えています。留意しておくべきことについてアドバイスいただけますか。

佐藤　2 点あります。

第一点は、**AI は完成された技術ではない**という認識を共有した上で議論すべきということです。

現在の生成 AI で使われている深層学習、ディープラーニングといったものがなぜうまくいっているのかといったことは、実は解明されていません。理論的なモデルも未完成なので、生成 AI に何ができて、何ができないのかも分からないし、予測もできない。AI 技術者も AI を制御できているわけではないのです。ですから、技術的に何が分かっていて何が分からないのか、どこまで制御できてどこから先は制御できないのかを整理した上で、その認識を共有し、その先に想定される問題と対応について早めに議論しておくのが良いと思います。

第二点は、**AI は規制対象であると同時に、規制の実現手段にもなり得る**ということです。

生成 AI によって、SNS やネット掲示サイトなどにおける誹謗中傷や差別的表現を、広範囲かつ高精度で監視することもできるでしょう。このため、AI の支援を前提とした現実世界またはサイバー世界に対する規制が行われる可能性があります。それは同時に盗聴や検閲の手段にもなり得ます。

フランスの哲学者ミシェル・フーコー（Michel Foucault）は『監獄の誕生』で、パノプティコン（人々を監視するために最適な建築構造）の中にいる人々は、常に監視されていることを意識し、規律化され、従順化すると指摘しました。現状は AI の処理能力の限界があって全市民を監視できないとしても、「AI に監視されているかもしれない」という意識が広がることによって市民が無意識のうちに AI に順応するような社会が形成されてしまうかもしれません。

AI の利便性を享受しつつ、国民の権利・便益の侵害が起きないことを担保しなければいけません。AI を規制や法執行の実現手段として使うときに、何が許されて何が許されてないのか、利用範囲と方法について議論すべき時期が来ていると思います。そういったことは規制する側からはなかなか言いだしにくいところでもあるので、デジタル政策フォーラムのような独立シンクタンクから問題提起していただくことはとても有意義だと思います。

谷脇　良い気づきをいただきました。デジタル政策フォーラムとしても建設的な議論をしていきたいと思います。ありがとうございました。

◇　　◇　　◇

【対談を終えて】

「AI ガバナンスは機能するか」をテーマに掲げた対談の話題は多岐にわたった。中でも特に示唆に富んでいた点を 5 項目に整理してみたい。

第一に、AI が常に進化し続けるものである以上、その制御可能性を考える AI ガバナンスも常にバージョンアップし続けなければならない。その際、「AI はどうあるべきか」という倫理的な議論にとどまるのではなく、AI が抱える課題を迅速に解析して「AI をどう制御するか」という具体のルールを考え、実行することが求められるステージに来ていることを認識しなければならない。

第二に、AI の普及はサイバー空間のデータ（情報）量を増やすとともにデータの質の大幅な低下をもたらし、学習データの劣化が AI そのものの劣化をもたらす AI 間の相互作用、いわば「共食い」による状態をもたらすことが懸念される。また、AI によってネット検索の地位が相対的に低下することで、最終的にはインターネットの利便性が低下するという指摘は興味深い。

第三に、AI による国内統制が進むこととなれば世界のデカップリングを生み出す可能性があるという指摘も重要だろう。国境を越えたデータ流通や知の共有が妨げられることで、インターネットの分断が深刻化する懸念をいかに払拭することができるだろうか。

第四に、オープン性の課題も重要な指摘だ。例えばインターネットはその内在するオープン性によって広く普及してきたが、オープンな AI が悪用されることによる負の影響の大きさを考えると、果たして AI のオープン性はどこまで有効なのかどうか、今後とも継続的に議論していくべき重要な論点の一つだろう。

第五に、AI が過去データを学習している以上、現状の社会課題を固定・拡張することになるという指摘は新鮮だった。社会的正義を実現するために入力データを歪めることがどこまで許容されるのか、それは誰が主体的に行うのかなど、さらに議論を深めていくことが必要だろう。（谷脇）

＜参考情報＞

[1] 2024 年 10 大リスク、ユーラシアグループ
https://www.eurasiagroup.net/siteFiles/Media/files/Top%20Risks%202024%20JPN.pdf
[2] 42 カ国が OECD の人工知能に関する新原則を採択、OECD
https://www.oecd.org/tokyo/newsroom/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence-japanese-version.htm
[3] 米国の AI 権利章典（AI Bill of Rights）について、内閣府
https://www8.cao.go.jp/cstp/ai/ningen/r4_2kai/siryo3.pdf
[4]『ChatGPT は世界をどう変えるのか』、2023 年 12 月、中央公論新社刊
https://amzn.asia/d/dFGsaip
[5]『教養としてのインターネット論　世界の最先端を知る「10 の論点」』、2023 年 9 月、日経 BP 刊　https://amzn.asia/d/6NI54pX
[6] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence、The White House　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/
[7] デジタル政策フォーラム　ホームページ　https://www.digitalpolicyforum.jp/