



# AI のガバナンス: 日本に関する青写真



# 目次

序文 .....	3
「問うべきなのは、『コンピューターで何ができるか』ではない。『何をすべきか』である」 .....	3
人間の状態向上に向けた新たな機会 .....	3
将来に向けた安全対策 .....	4
AI の公共ガバナンスのための 5 項目から成る青写真 .....	5
マイクロソフト内での AI ガバナンス .....	6
<b>第 1 部 日本における AI のガバナンス .....</b>	<b>8</b>
AI の公共ガバナンスのための 5 項目から成る青写真 .....	9
1. 政府主導の新たな AI 安全フレームワークを実装しそれをベースに構築する .....	10
2. 重要インフラストラクチャを制御する AI システムのために効果的な安全ブレーキを義務付ける .....	11
3. AI のテクノロジーアーキテクチャに基づく幅広い法規制の枠組みを策定する .....	14
4. 透明性を促進し、学術界や非営利団体が AI を利用できるようにする .....	22
5. 新たな官民パートナーシップの推進により新しいテクノロジーに付随する不可避の社会的課題に 対処するために AI を使用する .....	25
<b>第 2 部 Responsible by Design: 社会に利益をもたらす AI システムの構築に向けたマイクロソフトのアプローチ .....</b>	<b>26</b>
責任ある AI の開発に向けたマイクロソフトの取り組み .....	27
マイクロソフトでの責任ある AI の運用化 .....	29
ケーススタディ: 新しい Bing への責任ある AI アプローチの適用 .....	31
企業文化を通じた責任ある AI の促進 .....	33
責任ある AI に取り組むお客様の支援 .....	38
<b>第 3 部 日本の AI の現状 .....</b>	<b>40</b>
AI で社会的課題に対処するには .....	41
より健康な未来のための AI .....	41
より持続可能な未来のための AI .....	42
教育とエンパワメントのための AI .....	43
公共サービスの未来のための AI .....	44
<b>参考文献 .....</b>	<b>46</b>

# 序文



マイクロソフト副会長兼プレジデント、  
Brad Smith

## 「問うべきなのは、『コンピューターで何ができるか』ではない。『何をすべきか』である」

これは、私が 2019 年に Carol Ann Browne と共同執筆した書籍の AI と倫理に関する章のタイトルです。当時、「これは私たち世代の代名詞になるような問いかもしれない」と書きましたが、4 年経った今、この問いは世界中の首都においてだけでなく、多くの食卓においても、話題的となっています。

OpenAI の GPT-4 の基盤モデルの力を利用したり、その噂を耳にする人が増えるにつれ、驚きの声やあっけにとられる声があちこちから上がってきました。多くの人が熱狂し、興奮している一方、懸念を示す人、恐怖を感じる人もいます。著書で 4 年前に指摘したことが、ほとんどの人々にも明らかになりました。要するに、私たちは、これまで人間にしかできなかった意思決定ができる機械を人類の歴史上初めて作り上げた世代となったのです。

世界中で、次のような共通の疑問が呈されています。この新しいテクノロジーを使ってどのように問題を解決できるのか。このテクノロジーが引き起こし得る新たな問題をどのように回避し、管理するのか。これほど強力なテクノロジーはどのように制御すればよいのか。こうした問いには、幅広く思慮深い対話だけでなく、断固とした効果的な行動も求められます。

日本の場合、これらの問題はさらに重要です。過去半世紀の間、日本ほど強靱性と革新性に富んだ国はあまりありませんでした。しかし、この先 2030 年までとそれ以降、国民の要求と議論の中心がテクノロジーに集中するような新たな機会と課題が生じるでしょう。

日本で問われている問題の 1 つは、労働人口の減少と高齢化にどう対応し、支えるのかです。日本は、AI の力を活用して、人口構成の変化やその他の社会変化に対応しながら経済成長を促進する必要があるでしょう。本書では、マイクロソフトの企業としてのアイデアと提案を日本の状況に当てはめていくつかご紹介します。

世界の人々の役に立ち、信頼される AI ソリューションを開発するために、マイクロソフトは取り組みの指針となる倫理原則を定義、公開し、実践してきました。また、これらの原則を実践するべく、エンジニアリングとガバナンスのシステムを構築し、継続的に改善してきました。マイクロソフトでは現在、約 350 名が責任ある AI に取り組んでおり、社会に役立つよう設計された安全かつセキュアで透明性の高い AI システムを構築するにあたってベスト プラクティスが実践できるよう支援しています。

## 人間の状態向上に向けた新たな機会

こうしたアプローチが進化した結果、マイクロソフトでは責任ある AI が人々の生活を向上させる方法が増大していく様子が見えるようになりその力に確信を持つようになりました。GPT-4 のような基盤モデルの力が生活の中で副操縦士として機能することで検索がさらに強力な調査ツールへと変わり仕事の生産性を高めます。また、13 歳の子供の代数の宿題を手伝いたいが思い出せなくて四苦八苦している親にとっては、AI ベースのアシスタントが家庭教師になって助けてくれます。

このテクノロジーは、作業の迅速化、簡素化、向上を支援することで日常的な作業に役立つ一方で、AI の真の可能性は世界の最も捉えにくい問題のいくつかを明らかにすることが期待されていることにあります。個人の視力を補助する、がんの新しい治療法を進化させる、タンパク質に関する新たな知見を生み出す、危機的天候から人を守る予測を提供する、など AI の活躍も目にしてきました。その他にも、サイバー攻撃を回避する、外国からの侵略や内戦に苦しむ国々において基本的人権を保護する、などにイノベーションが役立てられています。

マイクロソフトは、本書の第 3 部で取り上げる日本からの革新的なソリューションについて楽観的に捉えています。これらのソリューションは、教育、高齢化、医療、環境、公共サービスなどのさまざまな領域における喫緊の課題のいくつかが、日本の創造性とイノベーションによって対処可能であることを示しています。

AI は非常に多くの点で、おそらくこれまでのどの発明よりも人類に利益をもたらす可能性を秘めています。1400 年代に活版印刷機が発明されて以来、人類の繁栄は加速度的に拡大してきました。蒸気機関、電気、自動車、飛行機、コンピューティング、インターネットなどの発明は、現代文明の構成要素の多くをもたらしてきました。そして AI は、印刷機自体と同じく、人間の学習と思考の進化を真に支える新しいツールなのです。

## 将来に向けた安全対策

もう 1 つの結論も同様に重要です。それは、AI の使用で生活を向上させるさまざまな機会に焦点を当てるだけでは十分ではないということです。これはおそらく、ソーシャルメディアの役割から得た極めて重要な教訓の 1 つと言えるでしょう。10 年余り前、技術者も政治評論家も、「アラブの春」の間に民主主義を広めるにあたってソーシャルメディアが重要な役割を果たしたことを熱弁していました。しかし、それから 5 年後、ソーシャルメディアはそれまでの多くのテクノロジーと同じく、武器にもツールにもなることを私たちは学びました。このときは、民主主義それ自体がターゲットでした。

10 年経った今、私たちはさらに賢明になっており、その知恵を働かせる必要があります。そして、今後待ち受けている可能性のある問題について、早い段階から明敏に考える必要があります。

また、AI を適切に制御することは、そのメリットを追求するのに匹敵するほど重要であると考えます。マイクロソフトは企業として、安全かつ責任ある方法で AI を開発し、展開することにコミットしており、断固とした姿勢で取り組んでいます。AI に必要な安全対策は、広範囲で責任を共有する必要があります。テクノロジー企業だけに任されるべきではありません。私達の AI 製品やガバナンス プロセスは、多様な利害関係者の観点が反映されるべきであり、それが私達自身とは異なる文化状況や社会経済状況の下で AI テクノロジーを開発し、展開する助けになります。

2018 年、マイクロソフトは AI に関する 6 つの倫理原則を採択し、そのうちの説明責任という 1 つの原則が他のすべての基盤であると記しました。なぜならば、今後も機械が人間の効果的な監視対象であり続けるようにし、機械を設計し運用する人が他の全員に対して説明責任を果たすようにすることは根本的な要件だからです。要するに、AI を常に人間の支配下に置くことを保証する必要があります。これは、テクノロジー企業にとっても政府にとっても最優先事項にしなければならない。

またこれは、もう 1 つの重要な概念に直結しています。民主主義社会の根本原則の 1 つに「法を超越した存在はない」というものがあります。政府も法を超越した存在ではなく、企業であれ、製品やテクノロジーであれ、法を超越した存在であってはならないのです。ここから重要な結論が導き出されます。AI システムを設計し運用する人が説明責任を果たすには、その意思決定や行動が法律に基づいていなくてはならないのです。

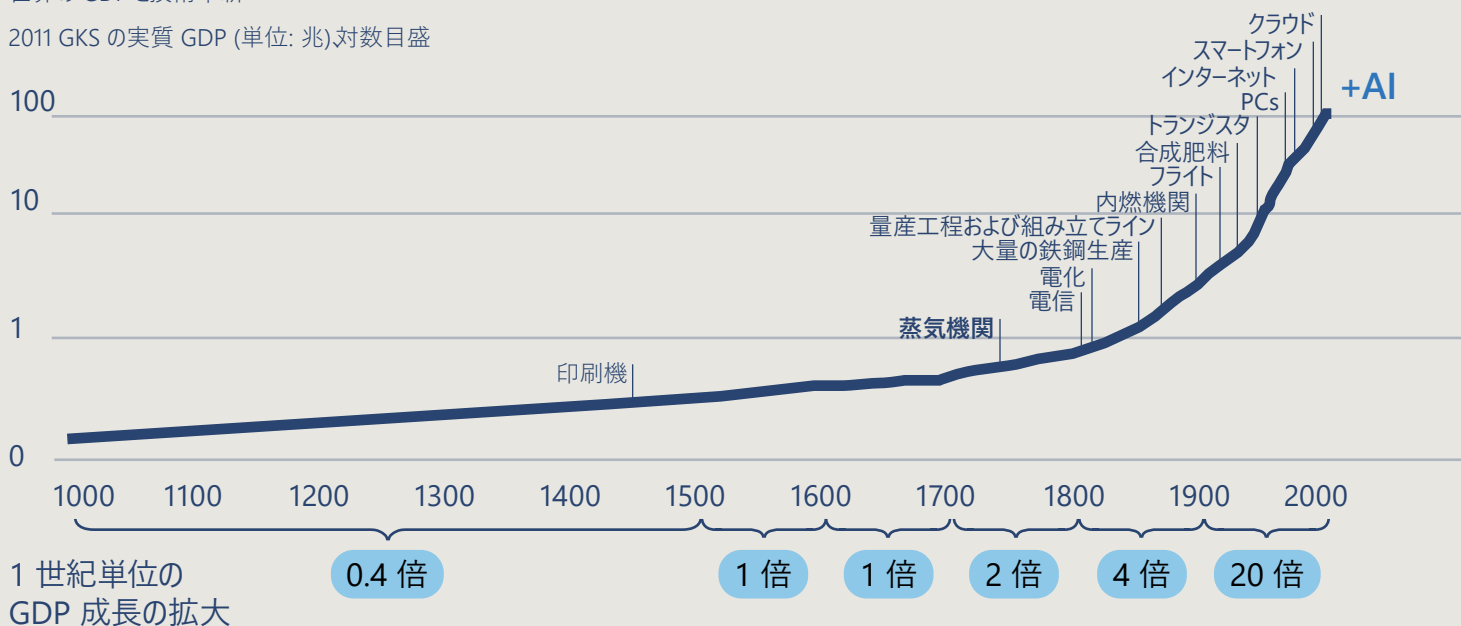
これは色々な意味で、目下展開されている AI の方針と規制に関する議論の中心となっています。では、AI が法規制に従うよう保証するために政府はどうすればよいのでしょうか。言い換えれば、新しい法律、規制、方針はどのような形を取るべきでしょうか。



テクノロジーによって GDP の成長が促進され、変化のスピードが加速

世界の GDP と技術革新

2011 GKS の実質 GDP (単位: 兆) 対数目盛



出典: Maddison Project Our World In Data

## AI の公共ガバナンスのための 5 項目から成る青写真

マイクロソフトは、責任ある AI プログラムで得られた教訓を基に、AI ガバナンスの推進に役立つ 5 項目から成るアプローチの詳細を示す青写真を 5 月に公開しました。ここでは、日本の状況に即してそうした政策アイデアや提言を示します。提言にあたって、この青写真のすべての部分がより幅広い議論により改善の余地があること、そしてさらに深く展開する必要があることを謙虚に認識しています。しかし、この青写真の提供により今後の取り組みに建設的に寄与することを願っています。そのための具体的なステップは以下のとおりです。

- 政府主導の新たな AI 安全フレームワークを実装し、それをベースに構築する
- 重要インフラストラクチャを制御する AI システムのために、効果的な安全ブレーキを義務付ける
- AI のテクノロジーアーキテクチャに基づく、幅広い法規制の枠組みを策定する
- 透明性を促進し、学术界や非営利団体が AI を利用できるようにする
- 新たな官民パートナーシップの推進により、新しいテクノロジーに付随する不可避の社会的課題に対処するための効果的なツールとして、AI を使用する

さらに広い意味では、AI ガバナンスのさまざまな側面を国際レベルで正しく機能させるには、各国の国内ルールを結び付けて、ある法域で安全と認められた AI システムが別の法域でも安全と認定できる多国間の枠組みが必要になるでしょう。実際、この前例は数多くあります。たとえば、国際民間航空機関が定めた共通安全規格がそうで、これは東京からニューヨークへの飛行中に機体の改修が不要になることを意味します。

日本は、現在の G7 議長国として、広島 AI プロセス (HAP) の立ち上げと推進に素晴らしいリーダーシップを発揮してきた実績があり、AI 問題に関する世界的な議論や多国間の枠組みを前進させることができる立場にあります。HAP を通じて、G7 首脳と各方面の関係者は AI ガバナンスに対する協調的アプローチの強化と、人権と民主主義的価値観を支持する信頼できる AI システムの開発の促進を図っています。世界的な原則の策定に向けた努力は、G7 を超えて OECD や GPAI (Global Partnership on AI) などの組織にも広がりつつあります。

2023 年 9 月に発表された G7 デジタル技術大臣会合閣僚宣言では、AI システムの開発者と導入者を含むすべての AI 関係者に対する国際的な基本原則を策定する必要性を認めています。また、高度な AI システムを開発する組織の行動規範も支持しています。こうした取り組みへの日本のコミットメントとこれらの対話における戦略的位置を考えると、多くの国が AI 規制に関する日本のリーダーシップと模範的行動に期待を寄せるでしょう。

日本が示しているように、責任ある AI に対する国際的に相互運用可能かつ俊敏なアプローチの実現に向けた取り組みは、AI のメリットを地球規模で最大化するために欠かせません。マイクロソフトは、AI ガバナンスが目的地ではなく継続的な取り組みであることを認識しながら、これらの活動を何か月も何年もかけて支援していきたいと考えています。

### マイクロソフト内での AI ガバナンス

最終的には、高度な AI システムを開発/使用するすべての

組織が独自のガバナンス システムを発展させ、実装する必要があります。第 2 部では、マイクロソフト内での AI ガバナンス システム、つまり、私たちがどこからスタートし、現在はどこに位置し、どのように未来に向かっていくのかについて説明します。

このセクションで確認したとおり、新しいテクノロジーに対応する新しいガバナンス システムの開発は、それ自体が 1 つの過程です。10 年前、この分野はかろうじて存在している程度でした。現在、マイクロソフトには AI を専門とする従業員が 350 名ほどおり、この数をさらに増やすために、来年度への投資を行っています。

このセクションで述べてきたとおり、過去 6 年間で、さらに包括的な AI ガバナンスの構造およびシステムをマイクロソフト全体で構築してきました。私たちはゼロからスタートしたのではなく、サイバーセキュリティ、プライバシー、デジタルの安全性を保護するためのベスト プラクティスを流用しました。これはすべて、企業の包括的なエンタープライズ リスク管理 (ERM) システムの一環であり、現代の世界における企業およびその他多くの組織を管理するうえで欠かせないものとなっています。

AI に関して、マイクロソフトはまず倫理原則を策定しましたが、これを基にさらに具体的な企業ポリシーを作成する必要がありました。現在は、これらの原則を具体化した企業基準のバージョン 2 を整備して、エンジニアリング チームが従うべきよりの確なプラクティスを定義しています。この標準の実装にあたっては、迅速かつ継続的に成熟しているトレーニング システムやツール システム、テスト システムを活用しています。この標準は、監視、監査、コンプライアンス対策など、追加のガバナンス プロセスによってサポートされます。

人生のすべてに言えることですが、人は経験から学びます。AI ガバナンスに関しては、さる機密性の高い AI のユースケースをレビューする際に必要となった詳細な作業から、特に重要な学習ができました。2019 年、マイクロソフトは社内でも最も機密性の高い斬新な AI ユースケースを対象として、機密性の高いユースケースのレビュー プログラムを設定し、厳

密で専門的なレビューを実施して、カスタマイズされたガイドランスが得られるようにしました。それ以降、マイクロソフトが完了した機密性の高いユース ケースのレビュー件数はざっと 600 件になります。この活動のペースは AI の進化のペースに匹敵する速度で加速しており、過去 11 か月間でほぼ 150 件ののぼるレビューを行っています。

こうしたことはすべて、マイクロソフトが企業文化を通じて責任ある AI を推進するため行ってきた取り組みに基づいており、この取り組みは今後も続けていきます。つまり、今後も新しく多様な人材を採用し、責任ある AI のエコシステムを成長させると共に、社内の既存の人材に投資してスキルを高めてもらい、AI システムが個人や社会に与える潜在的影響についてより幅広い視野から考えられるよう能力を強化するのです。また、テクノロジーの最前線には、これまで以上に優れたエンジニアと多彩な教養分野の才能ある専門家を組み合わせた分野横断的なアプローチが求められます。

マイクロソフトの責任ある AI プログラムの開発には世界中の関係者が関与しています。マイクロソフトにとって重要なのは、これらの問題に取り組んでいる世界中の人々の優れた意見をプログラムに取り入れながら、AI ガバナンスに関する代表的議論を前進させることです。日本で近く開催されるマルチステークホルダー会議に参加できることに胸を躍らせているのはそのためです。

今年 10 月、日本政府は「私たちが望むインターネット - すべての人々のエンパワーメント」をメイン テーマに、インターネットガバナンス フォーラム 2023 (IGF) を開催する予定です。IGF では、国際的な基本原則やその他の AI ガバナンスのトピックを推進するために、重要なマルチステークホルダー会議が予定されています。日本でのこれらの会議で他の参加者から学んだり、自社の高度な AI システムの開発と展開の経験をお話したりしながら、共通のルール作りに向けて前進できることを楽しみにしています。

マルチステークホルダー エンゲージメントのもう 1 つの例として、マイクロソフトの責任ある AI オフィスは 2023 年 [STEM ソン センターの Strategic Foresight Hub](#) (戦略的展


望ハブ) と組んで [Global Perspectives Responsible AI Fellowship](#) を立ち上げました。このフェローシップの目的は、AI とその社会に与える影響、および AI システムが導入される様々な社会状況、経済状況、環境状況をより適切に組み込む方法に関する実際の議論に、グローバル サウス諸国の市民社会、学界、民間部門からさまざまな関係者を招き入れることです。世界規模の包括的な調査の結果、アフリカ(ナイジェリア、エジプト、ケニア)、ラテンアメリカ(メキシコ、チリ、ドミニカ共和国、ペルー)、アジア(インドネシア、スリランカ、インド、キルギスタン、タジキスタン)、および東欧(トルコ) からフェローを選定しました。今年中には、目下の課題に光を当てるための対話の結果とビデオ投稿を共有するほか、AI アプリケーションのメリットを活かすための提案、グローバル サウスにおける AI の責任ある開発と利用について重要な洞察を示す予定です。

本書の内容はすべて、私たちは人工知能の責任ある未来を築き上げるための共同の旅路に乗り出しているという精神で書かれています。私たちは皆、お互いに学び合うことができます。また、今日どんなに良いと思えても、さらに良く続ける必要があります。

テクノロジーの変化が加速する中、責任ある AI ガバナンスへの取り組みもその変化に遅れずについていく必要があります。常に全世界の AI システムの中心に人間を据えながら適切なコミットメントと投資を行えば、実現できるはずで



Brad Smith  
マイクロソフト副会長兼プレジデント



第 1 部  
日本における AI のガバナンス



## AI の公共ガバナンスのための 5 項目から成る青写真

世界中の政府が AI を管理するための枠組みの策定を目指しています。当然ながら、単一の正しいアプローチはありません。マイクロソフトは AI ガバナンスの推進をさらに迅速化できるよう支援するために、多くの人にとって急を要する問題や課題に基づく 5 項目から成るアプローチを提供します。この青写真のすべての部分がより幅広い議論により改善の余地があり、そしてさらに深く展開する必要があります。しかし、この青写真の提供により今後の取り組みに建設的に寄与することを願っています。

この青写真は、AI を使用して人々の生活を向上させる多くの機会を認識しています。また、それと同時に、政府と民間両方のイニシアティブ (広範な国際協同研究など) に基づいて新しい制御を迅速に開発しています。そのための具体的なステップは以下のとおりです。

- 政府主導の新たな AI 安全フレームワークを実装し、それをベースに構築する
- 重要インフラストラクチャを制御する AI システムのために、効果的な安全ブレーキを義務付ける
- AI のテクノロジーアーキテクチャに基づく、幅広い法規制の枠組みを策定する
- 透明性を促進し、学术界や非営利団体が AI を利用できるようにする
- 新たな官民パートナーシップの推進により、新しいテクノロジーに付随する不可避の社会的課題に対処するための効果的なツールとして、AI を使用する

### AI のガバナンスのための 5 項目から成る青写真

1

政府主導の新たな AI 安全フレームワークを実装し、それをベースに構築する

2

重要インフラストラクチャを制御する AI システムのために、効果的な安全ブレーキを義務付ける

3

AI のテクノロジーアーキテクチャに基づく、幅広い法規制の枠組みを策定する

4

透明性を促進し、学术界や非営利団体が AI を利用できるようにする

5

新たな官民パートナーシップの推進により、新しいテクノロジーに付随する不可避の社会的課題に対処するための効果的なツールとして、AI を使用する

## 1. 政府主導の新たな AI 安全フレームワークを実装し、それをベースに構築する

政府の行動を加速させる最も効果的な方法の 1 つは、AI の安全性を推進する既存または新たな政府の枠組みを足がかりにすることです。このテクノロジーの安全な利用を確保するための重要な要素は、リスクの特定と軽減にだけでなく、展開前のシステムのテストについてもプロセスを規定するリスクベースのアプローチです。

たとえば、NIST (米国国立標準技術研究所) が策定し、2023年に発表された [AI リスク管理フレームワーク](#)には、AI ガバナンスを推進するための強力なテンプレートが用意されています。このフレームワークは、政府機関、市民社会団体、マイクロソフトなど複数のテクノロジーリーダーによる取り組みにより、コンセンサス主導で透明なプロセスを通じて策定されました。NIST は、何年にも及び重要ツール開発業務の経験を AI リスク管理の領域に持ち込むことで、サイバーセキュリティのリスクに対処しています。マイクロソフトには、サイバーセキュリティの最前線で NIST と連携してきた長年の経験があります。NIST がこの専門知識を適用して組織が AI に関連するリスクを抑制、マッピング、測定、管理できるよう支援しているのは心強いです。マイクロソフトは、NIST の AI リスク管理フレームワークの実装を約束しています。NIST のアプローチを高く評価しているのはマイクロソフトだけではありません。多数の政府、国際組織、主要企業も新しい AI リスク管理フレームワークの価値を認めています。同様に、新たな国際規格の策定も進んでおり、たとえば AI 管理システムに関する ISO/IEC 42001 は今後数か月以内に公開される見通しです。これは、企業だけでなく政府にとっても、責任ある AI の推進に向けて取り入れることができるもう 1 つの重要な枠組みになるでしょう。

政府には、これらの枠組みを中心にアメとムチを使い分けながら進展を加速させる機会があるはずです。多くの国において、政府調達、製品品質の向上と、産業界のプラクティスの全般的促進をもたらすことを繰り返し体現しています。政府は、まずは重要なリソースやサービスに対する人々の権利、

機会、アクセスに大きな影響を与える可能性のある重要な意思決定システムを中心に、AI リスク管理フレームワークまたはその他の関連国際規格の要件を AI システムの調達プロセスに組み込むことを検討してもよいでしょう。

多くの政府は AI に関する国家戦略をすでに発表したが、発表に向けて準備を進めているところで、日本もその例外ではありません。日本政府は、現行の 3 つの AI ガイドライン (2017 年の『AI 開発ガイドライン』、2019 年の『AI 利活用ガイドライン』、2022 年の『AI 原則実践のためのガバナンスガイドライン』) の再検討に入っています。また、民主主義的価値観に沿った信頼できる AI の構築を目的に、2023 年末までに新たな AI ガイドラインもまとめる予定です。2023 年 9 月に発表された計画によると、この新たな AI ガイドラインではリスクベースのアプローチを採用し、俊敏なガバナンスを求め、AI 開発者、AI 導入者、利用者を含むすべての AI 関係者の基準となるガバナンス アプローチを設定するようです。

2023 年 4 月、日本の自民党は国家 AI 戦略文書『AI ホワイトペーパー』を発表し、日本における AI テクノロジーの規制と導入促進に向けた提言を行っています。このホワイトペーパーでは、AI 規制に関する戦略と政策の概要を示し、大規模言語モデル、インフラストラクチャ開発、リスク評価など、最新の AI テクノロジー手法、動向について検討しています。また、公共機関における AI の導入や官民セクターにおける AI の利用を拡大するためのアイデアも提案しています。さらに、教育をはじめとする高リスク分野における AI 規制に対する斬新なアプローチについて検討するほか、AI 規制に対する俊敏なガバナンス アプローチの導入を勧めています。

国家 AI 戦略以外にも、日本は 2022 年 5 月、安全保障に影響を与える可能性がある経済活動を規制することで、日本の国家安全保障を強化することを目的とする包括的な経済安全保障法案として、経済安全保障推進法 (ESPA) を可決し、14 の重要インフラ分野を特定しました。これらの分野の業務に不可欠な AI ツールには、ESPA に基づく審査制度が適用される可能性があります。

マイクロソフトは、AI 発展の速度によって安全とセキュリティに関連する新たな問題と課題が生じていることを認識しています。また、こうした問題と課題に対する評価および対処を支援する実行可能な基準を策定するために、他者との連携に力を注いでいます。この取り組みの一環として、高機能モデルがもたらすリスクの評価および対処方法を含め、高機能モデルに関する安全のベスト プラクティスを推進し続けることを目指して、OpenAI、Google、Anthropic と共に [Frontier Model Forum](#) を最近立ち上げました。今後も他の企業と協力して、責任ある AI に関するフレームワークとプラクティスをさらに推進していきたいと考えています。

## 2. 重要インフラストラクチャを制御する AI システムのために、効果的な安全ブレーキを義務付ける

歴史から、新しいテクノロジーの可能性と危険性に関する重要かつ繰り返される教訓が得られます。印刷機の登場以来、政府は新しい発明を受け入れるか拒否するかを判断する必要性に直面してきました。1400 年代後半に入ったころ、欧州では印刷機が採用されていましたが、オスマン帝国ではおおむね禁止されていました。1500 年までに、オランダは 1 人あたりの読書量で他を上回っていました。この小国がやがて経済イノベーションの最前線に立たされたことは、偶然の一致ではありません。

それ以来ずっと、発明家と政府はたいてい、新しいテクノロジーをうまくコントロールしてその力を活用することが、進むべき最良の道だという結論に達してきました。テクノロジーの歴史には、たくさん例があふれています。高層建築物がなければ現代の都市は存在しなかったし、エレベーターがなければ高層建築物は存在しなかったでしょう。現代に住む私たちは、何も考えずにケーブルでつながれた金属の箱に入り、何階もの高さを行き来していますが、1800 年代には当然ながら、ほとんどの人がどうしたものかとためらっていました。米国のエレベーター発明家 Elisha Otis 氏は 1850 年代に、大衆がエレベーターを非常に危険と思えなかつた受け入れないことに気がきました。

この潮目が変わったのが 1854 年のニューヨーク万国博覧会で Otis 氏がエレベーターの新しい安全ブレーキを披露したときでした。集まった群衆の面前で、エレベーターのケーブルを切断し、すぐさまブレーキが作動してエレベーターの落下を止める様子を実演したのです。人々の不安は払拭され、近代的都市の誕生という重要な成果がもたらされたのです。

電気から鉄道、スクールバスに至る、あらゆることについて、このパターンが繰り返されました。現代の家屋や建物にはブレーカーが設置されており、電流の急激な上昇から保護しています。ブレーカーは都市の建築基準で必須となっています。同様に、何億人もの人々が世界で最もかけがえのないもの、つまり自分の子供を朝のスクールバスに乗せています。バスの運行の一部は、バスに非常用ブレーキを搭載し、そのブレーキを使用するための訓練を受けた運転手を必要とするという規則に基づいています。現代の飛行機には対地接近警報装置 (GPWS) と航空機衝突防止装置 (ACAS) が搭載されており、民間航空機での旅行が信じられないほど安全になっています。その一方で、セーフティクリティカルなシナリオで意思決定を行う能力は、機械ではなくパイロットに与えられています。

人工知能のある未来に目を向けると同時に、新しいテクノロジーに関連する潜在的な危険性を管理するうえで、同様の基本的アプローチが繰り返し作用することは覚えておく価値があります。つまり、新製品が暴走し始めるときを特定し、コア部分にはブレーキとして動作できる効果的な安全システムを設置し、必要な場合に適切な人がすばやく使用できるようにして、減速したり停止したりできるようにするのです。

潜在的な AI シナリオがすべて重大なリスクをもたらすわけではありません。実際のところ、リスクのないものがほとんどです。ただし、配電網、給水系統、緊急事態対応、都市の交通フローのためのインフラストラクチャを管理または制御する AI システムを考慮するとき、重大なリスクがからんできます。これらのシステムを引き続き人間の制御下に置くためには、「安全ブレーキ」が必要です。

こうした課題に対処するには以下のステップが役立ちます。



## 重要インフラストラクチャを制御する AI システムに対する効果的な安全ブレーキを確保するために政府が講じることができる 4 つの対策

- 1 展開されている高リスクの AI システムのクラスを定義する
- 2 インフラストラクチャの制御のために、AI システムの使用には安全ブレーキを設計段階から組み込むよう システム開発者に義務付ける
- 3 重要インフラストラクチャを強化システム システムが必ず人間の制御下に置かれるよう、事業者が高リスク システムのテストと監視を徹底させる
- 4 指定された重要インフラストラクチャの運用を制御する AI システムは、ライセンスを受けた AI インフラストラクチャにのみ展開するよう義務付ける

第 1 に、政府は重要インフラストラクチャを制御する目的で展開されている高リスクの AI システムのクラスを定義し、システムの安全への包括的なアプローチの一環としての安全ブレーキを保証する必要があります。安全ブレーキの概念を AI システムに適用するために、重要インフラストラクチャの運用を制御する目的で使用される AI システムに焦点を当てる必要があります。リスクが低く、同じ深さの安全対策を必要としない重要インフラストラクチャ セクター内では、従業員生産性ツールとカスタマー サービス エージェントなど、多くの AI システムが使用されることになるでしょう。

それよりも、高機能システム、増加しつつある自律システム、デジタルと物理の境界を越えるシステムに焦点を当てる必要があります。さらなる議論に拍車をかけるために、以下を行う AI システムに焦点を当てることが 1 つの出発点となります。

- 大規模にネットワーク化されたシステムに影響を及ぼす意思決定または措置を行う
- 物理的な入力と出力を処理または管理する
- 自律的または半自律的に操作する

- 大規模な被害 (物理的・経済的・環境的な被害を含む) の重大な潜在的リスクを及ぼす

第 2 に、政府は重要インフラストラクチャの制御のために、AI システムの使用には安全ブレーキを設計段階から組み込むようシステム開発者に義務付ける必要があります。

システムの安全は確立した研究分野であり、特に航空機、自動車、原子力といった分野に導入されており、重要インフラストラクチャを制御する AI システムのエンジニアリングに応用すべきものでもあります。重層的に実装された「安全ブレーキ」の概念を用いて、AI の安全に対する階層化されたアプローチを確立する必要があります。

「安全ブレーキ」の実装はシステムごとに異なりますが、あらゆる場合に核となる設計原則があります。それは、意図しない結果を検出/回避する機能がシステムに必要だということです。



また、意図しない動作を示す場合に解除または無効化する機能も必要です。さらに、ヒューマン コンピューター インタラクションの設計におけるベスト プラクティスを具現化することも必要です。

**第 3 に、政府は、重要インフラストラクチャを強化する AI システムが必ず人間の制御下に置かれるよう、事業者が高リスクシステムのテストと監視を徹底するよう指導する必要があります。**重要インフラストラクチャの計画的展開の状況では、特定のシステム テストが必要になります。言い換えると、先進の AI モデルの使用は、特定の製品またはサービスでどのように使用されるかという観点からレビューする必要があります。

システムの安全のベスト プラクティスに従い、システムとその各コンポーネントをテストし、検証し、妥当性を厳格に確認する必要があります。常に人間による制御が可能な形でシステムが動作することを証明する必要があります。実際には、システム事業者、AI インフラストラクチャプロバイダー、規制監視団体の間で緊密かつ定期的な調整が必要になると予想されます。






**第 4 に、指定された重要インフラストラクチャの運用を制御する AI システムは、ライセンスを受けた AI インフラストラクチャにのみ展開する必要があります。**リスクの高い重要インフラストラクチャ システムの運用を制御する AI システムがライセンスを受けた AI インフラストラクチャに展開されるよう義務付けることが賢明だと考えます。これは、AI インフラストラクチャはマイクロソフトなどのハイパースケール クラウド プロバイダーである必要があると示唆しているわけではありません。重要インフラストラクチャの事業者は AI インフラストラクチャを構築し、このようなライセンスを独自に取得する資格を得る可能性があります。しかし、このようなライセンスを取得するには、AI インフラストラクチャ事業者はアプリケーションレベルの対策が失敗した場合に人間の制御を確保できるよう、別の介入ポイント (実質的に、第 2 の独立した保護層) を許可するシステムを設計し、運用する必要があります。

このように言うと、人間の能力を上回る AI システムを開発し

ようとする場合に「安全ブレーキ」がどれほど現実的なのか、また将来にわたって使用できるのか、という疑問を持つ人もいます。AI システム自体が安全ブレーキを回避して無効化してしまうのではないかと、か AI システムはあらゆる段階での人間の反応を予想して、その反応をかわすべくさまざまな回避策を見つけるのではないかと、という疑問も湧いてきます。

このような疑問を投げかけるときは、現状の事実を明確に把握しておくことが重要です。現代の最先端の AI システム (OpenAI 社の GPT-4 や Anthropic 社の Claude など) に危険な機能 (人間による監視を回避し、停止困難になる機能など) がないか、[Alignment Research Center](#) の資格のあるサードパーティの専門家が明確にテストしています。これらのテストでは、GPT-4 および Claude には現在、こうしたことを実行するだけの機能はないと [結論](#) しています。この厳しいテストと導き出された結論により、現代の最先端の AI モデルの機能は明確になっています。ただし、私たちは Alignment Research Center によるこれらトピックについての継続的調査の呼びかけに注意を払い、AI の機能評価に対する業界全体の取り組みの必要性を認識する必要もあります。要するに、現状を理解するためだけでなく将来に備えるためにも、適切な体制をきちんと整えなくてはなりません。だからこそ、最先端領域にある小規模ながら重要な高機能 AI モデルに関してアクションを取る必要があるのです。このトピックは次のセクションで扱います。

## AI 基盤モデルのテクノロジー スタック

 アプリケーション	AI モデルの出力を実際に利用するソフトウェア プログラム
 API サービス	アプリケーションが事前トレーニング済みモデルにアクセスするための手段となる API (アプリケーション プログラム インターフェイス)、またはエンドポイント
 強力な事前トレーニング済み AI モデル	GPT-4 のような事前トレーニング済みモデル。ゼロからスタートせずに同様の問題を解決できる
 機械学習アクセラレーションソフトウェア	大規模 AI モデルの開発と展開のプロセスを加速させるソフトウェア
 AI データセンター インフラストラクチャ	高度なスーパーコンピューティング インフラストラクチャ (高帯域幅のネットワーク接続が可能な高度な GPU のクラスタを含む)

### 3. AI のテクノロジー アーキテクチャに基づく、幅広い法規制の枠組みを策定する

AI の責任に関連して考えられる法規制の課題をあれこれ考えるうちに、AI 自体のテクノロジー アーキテクチャを反映した形で AI 向け法規制アーキテクチャが必要なことがだんだんと見えてきました。つまり、法律により、AI テクノロジーの各種側面の管理に関わる各種アクターの役割に基づき、各アクターにさまざまな規制責任を課す必要があるのです。このような理由から、新しい AI 基盤モデルの構築と利用に必要な重要な部分を検討することが役に立ちます。

#### AI 基盤モデルのテクノロジー アーキテクチャについての基礎知識

マイクロソフトのようなソフトウェア企業は「テクノロジー スタック」を構築しています。テクノロジー スタックには、組織および人々が毎日頼りにしているアプリケーションの構築と実行に使用されるテクノロジーの階層があります。AI テクノロジー スタックを説

明する唯一の適切な方法はなく、2 人の開発者がまったく異なる説明をする可能性も高いです。しかし、AI の規制の将来について考えるという目的からすると、前のページのチャートを考慮することから始めるとよいでしょう。GPT-4 のような高度な事前トレーニング済み AI モデルは、スタックの中央の 3 段目に示されます。このようなモデルは、OpenAI のような企業の開発者および研究者が、その下の 2 つの層をベースに作成します。GPT-4 の場合、カリフォルニア州サンフランシスコにある OpenAI の技術スタッフが、AI スーパーコンピューティング インフラストラクチャを活用してモデル開発業務を行いました。このインフラストラクチャは、マイクロソフトが開発業務専用構築し、現在運用しているものです。[2020 年 3 月の開設時](#)にマイクロソフトが発表したとおり、このデータセンターには、当時世界トップ 5 のスーパーコンピューターにランクインした、単一のスーパーコンピューティング システムが収容されています。このスーパーコンピューティング システムは、28 万 5,000 以上の CPU コアを搭載しています(おそらく CPU は

最新の PC またはラップトップの最も基本的なコンポーネントです)。また、このシステムには最新の GPU が 1 万枚以上搭載されています。このようなチップの最新型でないバージョンは、ゲーミング コンソールやゲーミング ラップトップに搭載されており、大量の数式を同時に処理できます。データセンター内の各 GPU サーバーには、毎秒 400 ギガバイトのデータを処理できるネットワーク接続があります。

2020 年のデータセンター開設の発表時にマイクロソフトの最高技術責任者 Kevin Scott が語ったとおり、「これらの [新しい GPT] モデルの刺激的な点は、可能になるものの範囲が広いことです」。OpenAI 社とマイクロソフトが 2020 年に説明したように、歴史的に機械学習の専門家、言語間翻訳などの単一タスク学習用のラベル付けされた多くのサンプルをはじめ、個別の小規模な AI モデルを構築していました。

しかし、このタイプの大規模なスーパーコンピューティング インフラストラクチャを使用し、さらにカスタマイズした機械学習ソフトウェアも使用して、大量のデータ (数十億ページの公開テキストなど) を検査することで、学習可能な単一の大規模 AI モデルを作成できるようになりました。2020 年の発表時の「このタイプのモデルは、言語、文法、知識、概念、およびコンテキストの微妙なニュアンスを非常に深く吸収するため、長いスピーチの要約、ライブゲーミングチャットのコンテンツモデレーション、何千ものリーガルファイルからの該当箇所の検索、GitHub の検索によるコードの生成といった複数のタスクで、優れたパフォーマンスを発揮できます」というマイクロソフトの発言は、2023 年には世界中で認められています。

このすべてが示すように、この 10 年で最も驚くべき技術的な発展として印象に残った中核部分が、この 10 年が始まって早くも 3 か月目に、はっきりと世間の目に触れる形で事前発表されていたのです。良いニュース (少なくともマイクロソフトおよび OpenAI 社の視点による) は、過去数年の業務を通じて安全とセキュリティのプロトコルを強化したことで、さらに強力な AI モデルに備えられるようになったことです。

それでは大規模な AI モデルはどのように展開され、使用されているのでしょうか。これらは大量の計算リソースが必要なため、大量の GPU を搭載し、高度なネットワーク接続を利

用できる、複数の国の高度なデータセンター内で実行されません。GPT-4 の場合、マイクロソフトの Azure プラットフォームで実行されます。これにはそれ自体で多大な追加投資と、最先端のデジタルテクノロジーの展開が必要です。ただし、最初の段階で高度な AI モデルを構築するのに必要とされるような高度に専門化されたインフラストラクチャは不要です。

これらのモデルの実際の使用には、テクノロジースタックの上半分が必要です。ユーザーは、アプリケーション (スタックの最上位) を利用して GPT-4 のようなモデルとやり取りします。ChatGPT、Bing Chat、および GitHub Copilot はすべて、このようなアプリケーションの例です。規模を問わず企業および組織はきっと、新しいアプリケーションを作成するか既存のアプリケーションを修正して、生成 AI モデルの力を活用する機能とサービスを実装するでしょう。すでに有名なものを含め、多くが消費者向けアプリケーションになります。その他の多くは、企業、政府、非営利団体が内部使用のために組織内で作成するか、それぞれの顧客が作成します。要するに、生成 AI で推進されるアプリケーションの新しい波がすぐに世界中で日常生活の一部になります。

このようなアプリケーションは、API (アプリケーション プログラム インターフェイス) と呼ばれるエンドポイントを通じて AI モデルの機能にアクセスします。API は、顧客自身がインフラストラクチャ上で実行していない中核テクノロジーの構成要素にアクセスする最も重要な方法として長い間使用されています。

実例として、マイクロソフトは GPT-4 のような OpenAI モデルへの API アクセスを提供するために、Azure OpenAI Service を作成しました。この API を利用して、マイクロソフトのインフラストラクチャ上でホストされているモデルにアクセスできます。要するに、顧客は自分で選択したアプリケーションを構築し、単に API を呼び出してプロンプトを送信し、GPT-4 から出力を受信することによって GPT-4 の力を活用できます。顧客は、GPT-4 のような先進モデルの実行に必要な、高度なインフラストラクチャを維持する必要はありません。また、マイクロソフトの信頼とコンプライアンスへの長年の取り組み、および Azure OpenAI サービスの一部として GPT-4 の上に構築されている安全システムの恩恵を受けます。

## AI のテクノロジー アーキテクチャを反映する規制アーキテクチャの作成

AI 規制アーキテクチャは、以下に示す AI テクノロジー アーキテクチャにおおむね一致するように設計するのが合理的と考えられます。これらの課題についてここ数か月考慮した結果、法規制により、テクノロジー スタックの 3 つの層に焦点を当て、各レベルで異なる義務を課すことによって、おそらく最も良い影響をもたらされると考えます。下の図は、この提案されたアプローチを例示したものであり、ここに示す詳細な分析とコミットメントは、マイクロソフトが企業としてこれらの要件を推進するのに役立つと考えます。

## アプリケーション層での既存の法的保護を AI の使用に適用

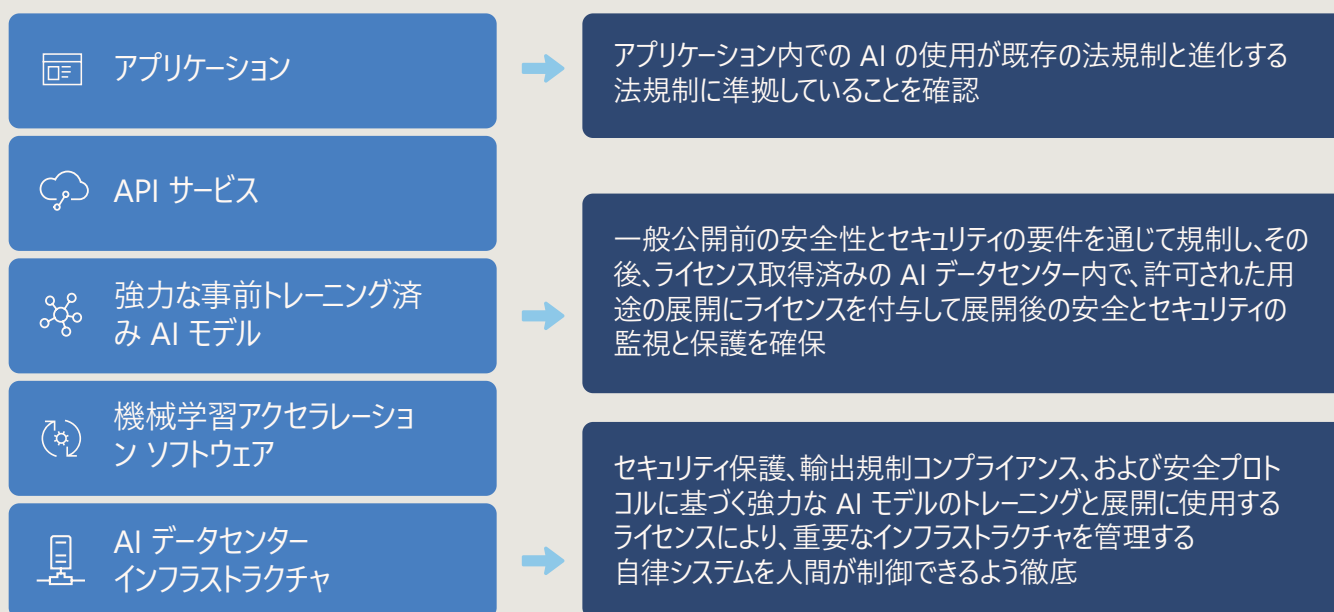
大多数の個人と組織にとって、法律が特に重要になるのは、アプリケーションが AI を利用して情報とサービスを他者に届けるときです。アプリケーション層は、人の安全と権利に最

も影響がある層で、特にさまざまなテクノロジー シナリオによって AI の影響が大きく異なる可能性があるためです。結果として、AI モデルの出力を使用して個人/組織にサービスを提供するアプリケーションに適用するべく、行動および社会的影響を制御する法規制が必要になります。

日本は長年、国民を守るために幅広い法整備を行っており、変化に対応するためにそれらの法律を修正する方法を模索し続けています。2018 年には著作権法を速やかに改正し、機械学習へ著作物利用に関する例外規定を追加しました。著作権法を管轄する文化庁は、2023 年初めの生成 AI の登場を AI と著作権に関する社会的理解を促進する好機と捉え、オンライン説明会を公開で実施し、上述の機械学習に関する例外規定についても取り上げました。

AI が社会に与える影響に関する多くの分野に新しい法規制が必要なのは、良いニュースです。その代わりに、私たちは既存の法規制を適用して実施する必要があり、日本に

## 提案される AI 規制アーキテクチャ





おけるこれに関する議論や世界中の複数の規制当局が実行する姿勢が見られるのは心強いことです。これは特に、新しくさらに強力な AI を使用するために作成されている多くのアプリケーションに関連するものです。また、あらゆる経済分野および国にある企業やその他の組織にとって重要になります。組織と個人の区別なく、意思決定と行動には引き続き既存の法律が適用されます。不法行為に対して裁判所で「いや裁判長、私にそうさせたのは AI なので」などという新手の弁護を展開して言い逃れできるようなことにはならないのです。

単純明快な結論ですが、その結果は重大な意味を持っています。つまり、AI を使用するすべての組織が、テクノロジー自体だけでなく、その幅広い法的責任にテクノロジーがどう影響するのかを評価する機能にも習熟する必要があります。また、裁判所も政府機関も、特定のシステムで AI がどのように使用されたのかを分析する新しい機能を開発する必要があります。

この目的を達成する手段は複数あり、その中には、以下のように企業として実施できるものもあります。

**第 1 に、顧客が最先端のベスト プラクティスを適用して AI を合法的に、かつ責任を持って展開できるよう支援するために、顧客と連携します。** AI の重要な特徴の 1 つとして、特定のグループおよび課題への実世界の影響は、AI モデルまたはシステムの開発者のみによって定義されるのではなく、特定のサービスまたはアプリケーションでのその実装においても定義されることがあります。実際に多くの状況において、AI が展開される前にこれらの実世界への影響を明確に特定してテストできるのは、アプリケーションレベルにおいてのみなのです。その結果、多くの場合、責任は共有されたり、分散され、さまざまな組織がさまざまな役割を果たす必要が出てきます。それ故に、顧客がサービスに AI を利用して独自の機能を開発する場合も、責任を持って行うことが非常に重要なもの、至極当然なのです。また、主要テクノロジー企業が責任ある AI の展開のための最新のベスト プラクティスやツールについて、情報を共有したり専門知識を与えたりすることが

非常に重要であるのも同じ理由からです。マイクロソフトは、法令順守システムの実装、サイバーセキュリティの推進、プライバシーの保護など、デジタルテクノロジーに関連するその他の課題について、この類の取り組みを 20 年間行っています。マイクロソフトは人工知能に関連する同様の取り組みを 5 年前に開始しており、今後は、このイニシアティブを拡大して顧客と共にさらに広く、深く取り組んでいきます。

**第 2 に、規制機関は AI の新しい専門知識と機能を追加する必要があります。** 極めて急速に、この必要性は世界のほとんどの政府の事実上すべての機関に拡大していくでしょう。たとえば、医薬品や医療機器を監督する規制当局でさらに必要となるのは、新薬の臨床試験の企業による最先端の AI システムの使用を評価する AI の専門家です。同様に、国土交通省航空局 (JCAB) では、新しい飛行機の開発時に航空機メーカーによる AI の新たな使用を評価するために、AI の専門家が追加で必要になります。生成 AI 自体が強力なツールとなり、規制当局は AI の使用の評価をしやすくなります。これは、GPT-4 のようなモデルと ChatGPT、GitHub Copilot、Microsoft M365 Copilot のようなサービスのおかげで、AI の力を活用してデータにアクセスし、迅速な評価を行うことが非常に容易になるからです。Google が最近のホワイトペーパーにおいて適切に指摘したように、政府が「各セクターの規制機関に対し、既存当局が AI 使用にどう適応すべきかを含め、既存の監視/施行制度を更新して AI システムに適応するよう指示する」ことが重要になるでしょう。これらの新しいツールを機能させるために、規制機関には資金、スタッフ、コミットメントが必要になります。

**第 3 に、マイクロソフトは幅広い教育への取り組みをサポートしており、AI テクノロジーおよび責任ある AI プラクティスに関する情報を議員、裁判官、法律家に提供しています。** 急速な AI の発展によって、法律の作成や施行に関わる人が新しい AI テクノロジーとその仕組みについて学ぶ必要に迫られています。1980 年代にパーソナル コンピューターが初めて人気を博したときも、同じようなニーズが見られました。たとえば、裁判官は PC のソフトウェアおよびハードウェアに関する証

拠、またはそれを必要とする証拠に一部依存し始める判決を下す必要がありました。1990 年代の初めに、マイクロソフトはこの新しいテクノロジーの仕組みに関する情報を伝える幅広い取り組みをサポートしました。現在でも、電子情報開示などの特定分野で行っています。AI の使用を促進するということは、このような新たな取り組みが必要になるということです。マイクロソフトは弁護士会などの公益団体や市民団体、そしてその活動をサポートすることによって、こうした仕事を支えます。

現行法の適用、規制当局のスキル向上、残る規制ギャップへ対処などの方法に関するアイデアを推進するために、世界中で重要な公共政策議論が続いています。2023 年 4 月の『AI ホワイトペーパー』の中で、自民党は「AI テクノロジーの急速な進歩に対応するには」俊敏性が不可欠だとして、迅速な規制の必要性を叫んでいます。他の法域を見ると、たとえばインドのデジタル インディア法では、包括的なリスクベースのアプローチを採用し、安全や治安に最も高いレベルのリスクをもたらす AI システムに最も厳格な要件を適用する総合的な AI 規制の枠組みの策定を試みています。

## 高機能 AI 基盤モデルに対応する新しい法規制の策定

テクノロジースタックのアプリケーション層には、既存の法規制を適用し、構築できますが、マイクロソフトはこの層の下にある、新しくさらに強化された AI モデルの出現を反映する 2 つの追加の層に対して、新しいアプローチが必要になると考えます。このうちの第 1 の層は最も強力な新しい AI モデルの開発用であり、第 2 の層は高度なデータセンターでこれらのモデルを展開/使用する場合に使用します。AI という最先端領域での仕事により、マイクロソフトは新たなクラスのモデルの出現を見てきました。高機能の基盤モデルがインターネット規模のデータセットでトレーニングされ、新しいタスクに対して即座に効果的に機能します。つまり、GPT-4 のようなモデルを使用すれば、1 つのプロンプトに単語を使用してこれまでにない画像を作成し、次の瞬間に有名な歴史上の人物風のスピーチを作成できます。

最先端技術を活用する、これらの基盤モデルの機能は見るからに驚異的であり、予測がつかないものになります。モデルの規模が拡大すると共に、機能は進化してきました。その中には、予想どおりの進化もあれば、事前に誰も予測し得なかった、小規模モデルにおいては観察できなかったような驚異的進化も見られました。一般公開前の厳格なテストおよびエンジニアリングにもかかわらず、ユーザーへの限定リリース

## すべてのライセンス制度に対するマイクロソフトのコミットメント

マイクロソフトは、高度な AI モデルに関する特別な知識を共有して、規制しきい値を定義できるよう政府を支援します

マイクロソフトは、高機能基盤モデルの開発/展開ライセンスを取得するために満たさなくてはならない要件を定義する政府の取り組みをサポートします

マイクロソフトは、ライセンス制度の効果的な施行を確保する政府の取り組みをサポートします

を経ないと、モデル機能の全体像が見えてこないことがあります。これらのモデルの力を活用し、法律および社会的価値基準に一致させる必要のある取り組みは複雑であり、進化しつつあります。

高機能モデルのこうした特徴により、対処すべきリスク面が現れます。今日に至るまで、マイクロソフトは、AI モデル開発の最前線で働いている開発者が自ら課した、高い安全基準の恩恵を受けています。しかし、こうした社会的に重要な課題を分別や自制心のみに委ねるべきではありません。リスクを予測して先手を打つ、リスク規制の枠組みが必要です。そして、すべてのアクターに善意があり、高機能モデルで生じる課題への対処能力が十分に備わっているわけではないという単純な真実を認める必要があります。AI をツールではなく武器として使用するアクターもいれば、待ち受けている安全性の課題を過小評価しているアクターもいます。

先週、OpenAI 社の CEO である Sam Altman 氏が米国連邦議会で証言し、研究開発の最前線にある、この小規模ながら重要な高機能モデルのクラス向けにライセンス制度の創設を呼びかけました。マイクロソフトはこの呼びかけを支持し、このライセンス制度を実現し、その実装を監督するための新しい規制当局の創設をサポートします。

**第 1 に、マイクロソフトとその他の主要な AI 開発業者は、高度な AI モデルに関する特別な知識を共有して、規制しきい値を定義できるよう政府を支援する必要があります。初期の課題の 1 つは、このレベルの規制に従う必要のある AI モデルを定義することです。**初期の課題の 1 つは、このレベルの規制に従う必要のある AI モデルを定義することです。この目的は、現在存在していて将来もサポートすべき AI モデルの豊富なエコシステムを規制せずに、極めて高機能で最前線を再定義しかねない少数の AI モデルを規制することです。このホワイト ペーパーでは、この小さなサブセットのモデルを高機能 AI モデルと呼びます。

高機能 AI モデルを構成するものに対して適切なしきい値を定義するには、今後数か月で熟考を重ね、話し合い、研究する必要があります。モデルのトレーニングに使用される計算

量は、モデルの機能に対する 1 つの実用的な指標です。ただし、現時点ではいくつかの点で不完全であり、特に、アルゴリズムの改善による計算効率の向上や、まったく新しいアーキテクチャの登場を考慮すれば、長期的に使用に耐えるものではないでしょう。

もっと長続きしそうですが複雑な提案は、安全性とセキュリティにとって重要な領域において高い能力を示唆する機能や、新たな飛躍の進歩を示唆する機能など、さらに前進する前に理解を深めておく必要のある機能を定義することです。このような機能ベースのしきい値を設定するにはさらなる研究と議論が必要であり、このような機能を定義する初期の取り組みは速いペースで継続する必要があります。当面は、世界の多くの複雑な問題に対するのと同じく、現在提供されている最良のオプション（計算ベースのしきい値）から始め、それを機能ベースのしきい値に迅速に進化させる作業プログラムに取り組むのが賢明かもしれません。

**第 2 に、マイクロソフトは、高機能 AI モデルの開発/展開ライセンスを取得するために満たさなくてはならない要件を定義する政府の取り組みをサポートします。**高機能 AI モデルのライセンス制度は、3 つの主要な目標を達成するように設計する必要があります。何よりもまず、高機能 AI モデルの開発と展開において安全とセキュリティの目標を確実に達成する必要があります。2 つ目の目標として、ライセンサーとその規制当局との間で緊密な調整の枠組みと情報の流れを確立する必要があります。これによって、安全とセキュリティの目標を達成するための開発マテリアルをタイミングよく確実に共有し、実行します。3 つ目の目標として、安全とセキュリティの目標を共有する国どうしの国際協力の基盤を提供する必要があります。高機能 AI モデルの有効利用を確実になものにし、誤用を防止するには、国内のイニシアティブだけでは不十分です。現在ではごく簡単にモデル ウェイトを国外へ移動できてしまうこと、つまり高機能 AI モデルの「最重要部分」にアクセスできる人が当該モデルを気軽に他国へと持っていけることを念頭に置いたうえで、対策を進めなくてはなりません。



## KY3C:

金融サービス向けに開発された  
「Know Your Customer」というコンセプトを AI データ サービスに適用

Know your Cloud (クラウドについて知る)

Know your Customer (顧客を知る)

Know your Content (コンテンツを知る)

安全とセキュリティの目標を達成するために、マイクロソフトは、大規模な訓練運転の事前通知、危険な機能または飛躍的な機能の特定に焦点を当てた包括的なリスク評価、内外の専門家による一般公開前の大規模なテスト、途中過程での複数のチェックポイントなどのライセンス要件の構想を練っています。モデルの展開は、評価されたリスクレベルに加え、信頼できるユーザー、規制当局、その他の関係者が残余リスクをどのように管理するかについての評価に基づき、制御する必要があります。安全対策が期待どおりに機能していること、展開されたモデルが常に人間の制御下にあることを確認するために、公開後の継続的監視が必須となります。

実際には、こうした枠組みを効果的に施行するには、結局のところテクノロジースタックをさらに 1 層掘り下げ、高機能 AI モデルが開発、展開される AI データセンターに移動する必要があります。

**第 3 に、高機能 AI モデルのテストまたは展開に使用される AI データセンター事業者がライセンス要件を課すことにより、高機能 AI モデルに対応するライセンス制度の効果的な施行を確保する政府の取り組みをサポートします。**現代の高機能 AI モデルは高度な AI データセンターで構築されま

す。前述のアイオワにあるマイクロソフトの施設のように、大量の演算能力、特殊化された AI チップ、高度なインフラストラクチャエンジニアリングが必要です。したがって、このような AI データセンターは現代の高機能 AI モデルの重要な要素であり、包括的な規制制度における効果的な制御ポイントです。

電気通信ネットワーク事業者および重要インフラストラクチャのプロバイダーに対する規制モデルがあるのと同じく、AI データセンターのプロバイダーにライセンスを与える必要があります。これは、AI データセンターのプロバイダーが高機能 AI モデルの安全でセキュアな開発と展開を確保する役割を責任を持って効果的に果たすようにするためのものです。ライセンスを取得するために、AI データセンター事業者はサイバーセキュリティ、物理セキュリティ、安全アーキテクチャ、場合によっては輸出規制コンプライアンスに関する特定の技術的な能力を満足させる必要があります。

事実上、これにより、マネーロンダリング、犯罪者またはテロリストによる金融サービスの悪用を防止するために銀行業向けに開発された原則が AI に適用されるようになります。「Know Your Customer」(KYC) の原則では、金融機関には、顧客の本人確認、リスクプロファイルの確立、および疑



わしい活動を検出するための取引の監視が義務付けられます。

同様に、まずモデルが展開される「クラウドを知る」という KYC 原則を高機能 AI モデルの開発者に求めるのも重要です。認可されたライセンスのある AI データセンターを使用することにより、高度なモデルの開発者は確実に複数ベンダーから選択できるようになります。また、高度なモデルの開発者は、必須の技術標準および義務を満たすことによって、独自のクラウド インフラストラクチャも構築または運用できるようになります。ライセンスが付与された AI データセンター事業者は、その時点の規制要件を満たす必要があります。そのうちのいくつかを検討してみましょう。

第 1 に、AI データセンター事業者には、高機能 AI モデルを悪意のある攻撃や敵対者から保護するうえで果たすべき特別な役割があります。この役割には、技術的および組織的な対策だけでなく、AI データセンター事業者、モデル開発者、規制当局の間での脅威インテリジェンスの継続的なやり取りも含まれるでしょう。

第 2 に、機密性の高い使用を伴うシナリオの場合など、特定の事例において、モデルが動作しているクラウド事業者は KYC 原則の 2 つ目 (モデルにアクセスしている顧客を知る) を適用する必要があります。細部の詰めを行うために、熟考と話し合いを重ねる必要があります (特に、さまざまなシナリオにおいて固有の顧客データを収集/維持する責任を負うべき者を決定する場合など)。

KYC の手順を実行している AI データセンター事業者は、モデル開発および展開のための適切なライセンスをすべて取得しておけば、規制当局も仕事を進めやすいでしょう。大規模な

訓練運転にありがちな大量の演算使用を、詳細な調査のために 規制当局に報告しておくのも、1 つのアプローチです。

第 3 に、輸出規制措置が進化する中、AI データセンター事業者は、テクノロジー スタックのインフラストラクチャ層およびモデル層に関わるものなど、これらの措置の効果的な実施を支援することもできます。

第 4 に、前述のとおり、AI インフラストラクチャ事業者には、重要インフラストラクチャを管理/制御する AI システムのために、安全プロトコルの適用および、効果的な AI 安全ブレーキの確実な整備について、重要な役割と義務があります。インフラストラクチャ事業者にとっては、第 2 の独立した保護層として介入する機能を持つことで、AI システムが必ず人間の制御下にあると人々に知らしめることが重要になるでしょう。

これらの初期の考えは当然ながらすべてさらに発展させる必要があります。これらのトピックに関しても、OpenAI の皆さんが今後重要な貢献をしてくれるに違いありません。今はっきりと言えるのは、最前線にある AI モデルの性能が向上し、より自律的になり、デジタルと物理の境界を越えるようなものになって初めて、この多層化されたライセンス制度の重要性が増してくるということです。前述のとおり、効果的なライセンス制度を計画し、実施するのは有益であるとマイクロソフトは考えています。たとえば、高機能 AI モデルが運用で中心的な役割を果たしているときに配電網やその他のセーフティクリティカルなインフラストラクチャに対する制御を維持できるよう徹底するなどの場合に、ライセンス制度は特に役に立ちます。

## 4. 透明性を促進し、学术界や非営利団体が AI を利用できるようにする

### AI の重要な倫理原則としての透明性

今後数か月および数年で真剣な話し合いが必要となる多くの AI ポリシーの課題の 1 つとして、セキュリティと透明性の関係および衝突があります。セキュリティ上絶対に機密にすべきだと多くの専門家が確信している領域がいくつかあります。AI モデル ウェイト (モデルの機能に対してコアとなるモデル構成要素) がその一例です。場合によっては、重要な国家安全保障と公共の安全利益の保護も必要になります。その他、セキュリティのニーズとベスト プラクティスの理解を促進させる場合にも透明性が重要なケースが多数あります。つまり、衝突が存在する場合もあれば、存在しない領域もあるのです。

マイクロソフトは 2018 年に AI の倫理ガイドラインを採用し、透明性を 6 つの基本原則の 1 つにしました。この原則を実装する中で、人々が AI システムとやり取りしている

ことを認識できるようにすることを含め、状況に応じてさまざまなタイプの透明性を提供することが重要であるとわかってきました。生成 AI では、この原則はこれまでよりさらに重要になります。生成 AI は継続中の研究およびイノベーションが必須となる領域です。この領域での新たな取り組みを促進するために、マイクロソフトは 3 つのコミットメントをしています。

**第 1 に、マイクロソフトは透明性レポートを毎年発行して、責任ある AI を安全に管理するうえでのマイクロソフトのポリシー、システム、進捗、実績を公開します。** 透明性レポートは、企業の説明責任を推進し、目標に向けた到達水準と進捗について人々の理解を深める効果的な手段であると証明されています。マイクロソフトは透明性レポートを信じています。

**第 2 に、マイクロソフトは、自由に点検可能な、高リスク AI システムの国家登録制度の策定をサポートします。これは、高リスク AI システムがどこでどのように使用されているのか人々にもわかるようにするためのものです。** AI システムがどこでどのように使用されているのかを明らかにすることによ

## AI の透明性を促進するためのマイクロソフトのコミットメント

マイクロソフトは、透明性レポートを毎年発行して、責任ある AI を安全に管理するうえでのマイクロソフトのポリシー、システム、進捗、実績を公開します

マイクロソフトは、高リスク AI システムがどこでどのように使用されているのか人々にもわかるようにするために、自由に点検可能な、高リスク AI システムの国家登録制度の策定をサポートします

マイクロソフトは引き続き、AI データ システムの設計にあたっては、人々が AI データ システムとやり取りしているときにはその事実が通知されるように、また、システムの機能と制限事項が明確に伝えられるように徹底していくことを約束します

マイクロソフトは、人々は受け取ったコンテンツが AI の生成したものであることを「知る」ことができるように、重要なシナリオにおいて AI が生成するコンテンツにラベル付けを義務付けることにメリットがあると考えます

てAI システムに対する人々の信頼が高まります。高リスク AI システムについて、マイクロソフトは国家登録制度の策定をサポートします。この登録によって人々が展開されたシステムの概要とシステムの安全性と権利尊重の性能を確保するための手段を確認できます。この情報は人々の役に立てるようにするためにシステムレベルで表現し使用状況に関する詳細を提供し、専門家でない一般向けの記述にする必要があります。これを実現するために欧州の複数の都市に倣ってアルゴリズム透明性基準を採用し、AI の使用法に関するアクセス可能な説明を策定するというアプローチも一案でしょう（例としてアムステルダム市のアルゴリズム登録についてのページを参照）。

**第 3 に、マイクロソフトは引き続き、AI システムの設計にあたっては、人々が AI システムとやり取りしているときにはその事実が通知されるように、また、システムの機能と制限事項が明確に伝えられるように徹底していくことを約束します。** 透明性の確保は、広範なレポートや登録を通してだけでなく、特定のシナリオにおいても特定 AI システムのユーザーにとっても重要です。マイクロソフトは、AI システムを使用する人が十分な情報に基づいた意思決定ができるようサポートすることを念頭に置いて AI システムの構築を続けます。マイクロソフトは透明性に対して包括的なアプローチを取ります。たとえば、AI システムとやり取りしていることを伝えるユーザー インターフェイス機能だけでなく、教材（例：新しい Bing primer）システムの機能と制限に関する詳細文書（例：Azure OpenAI Service の Transparency Note）などを含めます。この文書およびエクスペリエンス デザインの要素は AI システムの利用目的を人々が理解し、独自の使用について十分な情報に基づいた意思決定ができるよう支援することを目的としています。

**第 4 に、マイクロソフトは、重要なシナリオにおいて AI が生成するコンテンツにラベル付けを義務付けることにメリットがあると考えます。** これにより、人々は受け取ったコンテンツが AI の生成したものであることを「知る」ことができるようになります。これは、マイクロソフトが検討する価値があると考え、KY3C アプローチの第 3 の部分です。マイクロソフ

トはそのサービス Bing Image Creator と Designer で上記のことに取り組んでいる中、「どのコンテンツ」を AI が作成しているのかを知る権利が人々にはあると考え、動画や音声などのオリジナルが当初、人間ではなく AI モデルによって作成されたという事実を人々に通知しています。このラベル付けの義務は、オリジナルのコンテンツの特定のカテゴリが AI を使用して変更されたときにも通知する必要があります。これには、新しい法律の策定が必要であり、それに対処するために重要な問題や詳細が数多く生じます。たとえば、近年では、AI の武器化によってコンテンツが改変されたり、動画などの「ディープ フェイク」が作成されたりした場合に民主主義や人々が被る新たなリスクにどう対処するかという点に、注目が集まっています。未来のテクノロジーに関する懸念も適切に提起されています。

幸い、新たな透明性レポート イニシアティブを作ることに加えて、AI 透明性に既存の技術構成要素を使用する機会もあります。そうした取り組みの 1 つが Adobe、BBC、Intel、マイクロソフト、Publicis Groupe、ソニー、Truepic をはじめとする 60 以上の組織が参加する国際標準化団体である Coalition for Content Provenance Authenticity (C2PA) です。この団体は、現在は生成 AI にも対応しているデジタル コンテンツ認証のための世界初の技術仕様を 2022 年に公開するなど、オンライン情報の信頼性と透明性の向上に力を注いでいます。マイクロソフトは最近、C2PA 仕様を利用して、最先端の来歴ツールを新たにリリースする予定であることを Build 2023 カンファレンスで発表しました。これを使用すると、AI によって生成された視聴覚コンテンツの区別や、その生成元の把握が容易になります。まずは主要な画像形式と動画形式をサポートし、マイクロソフトの新しい AI 製品のうち Microsoft Designer と Bing Image Creator の 2 つで利用可能なサービスをリリースする予定です。これは重要なステップですが、情報の完全性を保護するための必要な幅広い取り組みにおける 1 つのステップにすぎません。



## 学術研究および非営利団体のための AI リソースへのアクセス

他にも、透明性に追加される要素や、もっと注目に値する要素があります。これは、学術研究や非営利団体のための AI リソースへの幅広いアクセスを提供するニーズです。大規模 AI モデル（および AI プロジェクト）のトレーニングのための計算リソースが高コストなので、高等教育機関と非営利団体において当然ながら懸念が生じつつあります。マイクロソフトは、この課題を十分に理解しています。というのも、2019 年にマイクロソフトが行った OpenAI への大規模なテクノロジー投資は、部分的にその非営利の扱いにより、まさに OpenAI 自体に対するニーズから始まったものであるからです。

テクノロジーセクターのほとんどが生まれて、継続的にイノベーションをしている背景には、国内の大学で推進されている重要な基礎研究の力があります。このサクセスストーリーは、世界中の多くの国で研究され、模倣されてきました。ここ数十年で、増え続けるコンピューティングリソースとデータサイエンスによって推進された広範囲にわたる基礎研究が、ほぼすべての分野で見られるようになりました。学術研究者がかなり多くのコンピューティングリソースを取得できない限り、科学的な調査と技術的なイノベーションが損なわれるリスクが現実にあります。この問題のもう 1 つの側面も重要です。学術研究者は、AI の理解を促進することによって人々への説明責任の浸透を支援しています。人々は、商業セクターで作成されているモデルの動作分析によって AI の説明責任を進展させる研究など、この分野での研究を進める学者を必要としています。新しく小規模なオープンソース AI モデルが出現し、その重要性も明らかで、AI が関係する他の基礎研究プロジェクトにおいて、これまでよりも多くの計算能力が必要になることはほぼ確実です。また、新しい資金ソースが集まって学術界に集中的リソースが提供されないと、学術研究は危機に立たされるでしょう。これを受けて、マイクロソフトは以下の 3 つの取り組みに焦点を当てました。

**第 1 に、マイクロソフトは学術研究のためのコンピューティングリソースを提供するために、新たに提案された米国の国家 AI 研究リソース (NAIRR) の創設をサポートします。また、世界各国の学術機関によるアクセスに対応するための拡張を歓迎し、支持します。**米国は「計算リソース、高品質データ、教育ツール、およびユーザーサポートへの大幅に拡大されたアクセスを AI 研究者や学生にもたらす共有研究インフラストラクチャ」として、国家 AI 研究リソースを推進しています。マイクロソフトは、こうした研究リソースの創設を支持しており、AI の機会とリスクに対する理解を深めるのに重要だと考えています。

また、世界各国の学術機関にアクセスを提供するために、NAIRR の拡張を歓迎し、支持します。こうしたリソースには、すでに世界中の国から同様の関心が多数寄せられています。これには日本も含まれ、AI 戦略会議の『AI に関する暫定的な論点整理』では政府に対し、学術機関の AI 研究を支援するためにコンピューティングリソースとデータ関連インフラストラクチャを開発することを勧告しています。これは、リソース集約な研究を実施できるように計算リソースへのアクセスに政府が資金を提供することを提案している自民党の『AI ホワイトペーパー』と一致しています。

**第 2 に、マイクロソフトは学術研究プログラムへの投資を増やすことで、マイクロソフト外部の研究者が確実にマイクロソフトの基盤モデルと Azure OpenAI Service にアクセスして研究に着手し、結果を検証できるようにします。**この拡張されたコミットメントは、マイクロソフトの Turing Academic Program と Accelerating Foundation Models Research Program の成功を基にしています。その目的は、マイクロソフトの最先端の基盤モデルに加え、マイクロソフトの Azure OpenAI Service を利用した OpenAI モデルに対する API ベースのアクセスを学術界が取得できるよう支援することです。これにより、最先端の用途と、これらのモデルの社会技術的な影響を研究者が調査できるようになります。このようなアクセスの提供を補完するにあたって、高機能モデルの用途、安全/セキュリティへの影響に関する最前線の研究に従事している学術界のガバナンス開発のベストプラクティ



スが重要になるのです。マイクロソフトは、マルチステークホルダーグループのサポートおよび同グループ(学术界の代表者を含む)との提携によってこのようなベスト プラクティスを開発する機会を歓迎します。

**第 3 に、マイクロソフトは非営利団体にご利用いただくための無料および低コストの AI リソースを作成します。**最後に、マイクロソフトは、世界中の社会的ニーズへの対応において非営利団体が果たす重要な役割を深く理解しています。非営利団体は革新的ソリューションを活発に生み出す役割を担っていることから、その活動に合った AI の新たなモデルや機能を非営利団体が幅広く、簡単に、かつ安価で利用できることが重要であるとマイクロソフトは考えています。Tech for Social Responsibility (社会的責任に関するテクノロジー) 部門をはじめとする Microsoft Philanthropies は、Microsoft Cloud 上で 35 万の非営利団体を支援しています。Microsoft Philanthropies が世界中の非営利団体に提供している現金および技術の寄付額と割引額は年間 40 億ドルを超えます。これは、国家による対外支援予算額のトップ 10 の 1 つに匹敵する数字です。マイクロソフトは先週、Microsoft Cloud for Nonprofit 向けの AI ソリューションを発表し、この支援を拡大しました。これらの AI ソリューションは、非営利団体の業務最適化、寄付者との関係構築の能力を高める目的で設計されています。

## 5. 新たな官民パートナーシップの推進により、新しいテクノロジーに付随する不可避の社会的課題に対処するために AI を使用する

近年得た教訓の 1 つは、民主主義社会がテクノロジーの力を活用し、公共部門と民間部門を連携させたときに最も大きな成果が達成できるというものです。社会に与える AI の影響に対処するために、この教訓を基盤にする必要があります。AI は並外れたツールであり、とてつもない利益をもたらす可能性を秘めています。とはいえ、他のテクノロジーと同様、武器としても利用可能であり、そうした利用を目指す者が世界中に現れるでしょう。その一方で、AI を悪用する者や社会的課題との戦いに AI を活用することもできます。私たちは官民パートナーシップを通じて、力を合わせてこうした活用を図る必要があります。

特に、固有の AI を利用して民主主義と基本的権利を強化し、AI スキルを広く利用できるようにすることで包摂的な成長を促進し、AI の力によって地球上のサステナビリティ ニーズを推進するには今、重要な取り組みが求められています。おそらく何よりも重要なのは、新しい AI テクノロジーの波は、大胆な考えや行動をする機会を与えてくれているということです。それぞれの領域において成功への鍵となるのは、具体的なイニシアティブを用意し、政府、業界、NGO が集結して、そのイニシアティブを推進することです。マイクロソフトは各領域において自らの役割を果たすつもりです。

日本政府はこの取り組みの先頭に立ち、2023 年 10 月にインターネット ガバナンス フォーラム (IGF) を京都で開催する予定です。この重要なフォーラムは、AI 政策に対するさまざまな利害関係者の観点を国内および国際レベルで収集するための重要な機会となることが期待されます。

The background is a dark blue gradient with a network of glowing white and light blue lines and dots, suggesting a digital or data theme. The lines are thin and connect various points, some of which are larger and brighter than others. The overall effect is a sense of connectivity and data flow.

## 第 2 部

Responsible by Design: 社会に利益をもたらす AI システムの構築に向けたマイクロソフトのアプローチ

## 責任ある AI の開発に向けたマイクロソフトの取り組み

マイクロソフトはこの 7 年間、責任ある AI (強固な倫理原則に基づいた人工知能) を進化させるべく取り組んできました。信頼は与えられるものではなく獲得するものであり、自社だけでなくコミュニティに対しても責任を広く担っているという謙虚な認識を持って、取り組みを続けてきました。これが、自社のコミットメントを達成することと、お客様やパートナー様によるコミットメントの達成を支援することの両方を重視する姿勢につながっています。

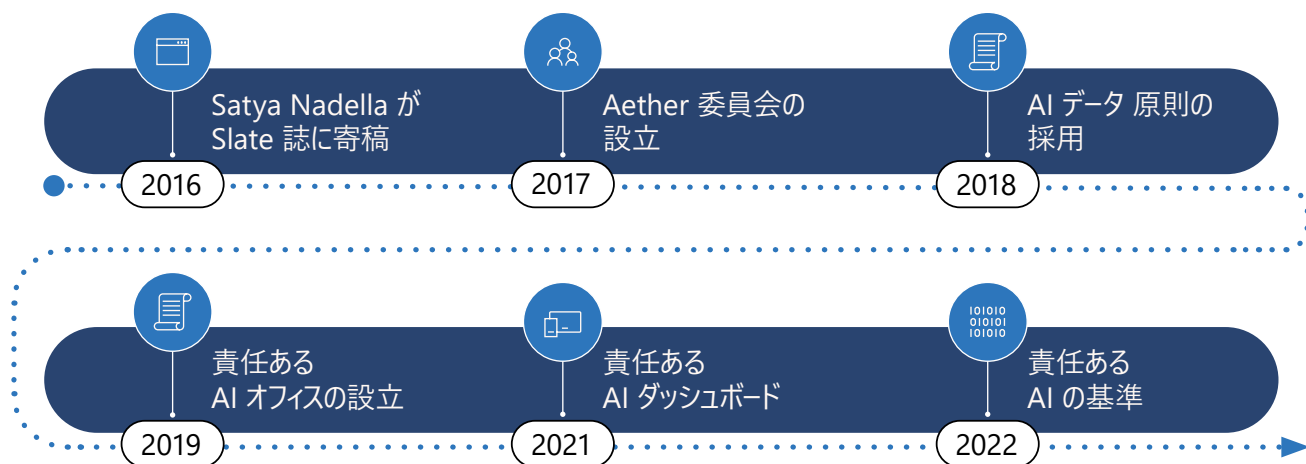
責任ある AI に対するマイクロソフトの取り組みは、2016 年に始まりました。このとき、会長兼 CEO の Satya Nadella が、AI によって人間が能力を高めるというビジョンを紹介しました。Satya はマイクロソフトの中核的な AI 原則の始まりを示したのです。これは現在も続いている価値観です。このビジョンに基づき、研究者、エンジニア、政策専門家から成るマイ

クロソフトの Aether 委員会が立ち上がりました。ここでは、マイクロソフトの AI 原則に関連する最先端の新しいトレンドについて、各参加者が分野ごとの専門知識をもたらしています。こうした取り組みが 2018 年の AI 原則の策定と採用につながりました。

2019 年には、責任ある AI オフィスを設立し、取り組みを強化しました。このチームがマイクロソフトのプログラムのガバナンスを調整しており、全社的な共同作業によって、責任ある AI の基準の第 1 版を作成しました。これは、概要レベルの原則を、AI システムの構築を担うエンジニアリング チーム向けの実用的なガイダンスに落とし込むためのフレームワークです。

2021 年には、マイクロソフトのプログラムを運用化するために導入していた重要な構成要素について、公式に語りました。

### 責任ある AI に対するマイクロソフトの取り組み



## 責任ある AI ガバナンスのフレームワーク



また、責任ある AI に対する取り組みの導入と拡大を促進するために、トレーニング、プロセス、ツールを増やす構想を練りました。2022 年には、責任ある AI の基準を更新し、現在使用しているバージョンまで進化させました。このバージョンは公開もしています。責任ある AI の基準では、マイクロソフトが潜在的なリスクを事前に特定、測定、軽減する実践的な手法を用いながらどのように AI システムを構築していくかを表明しています。この Responsible by Design のアプローチにより、潜在的損害を最小限に抑えつつ、AI のメリットを拡大する反復可能なプロセスが最初から確立されます。

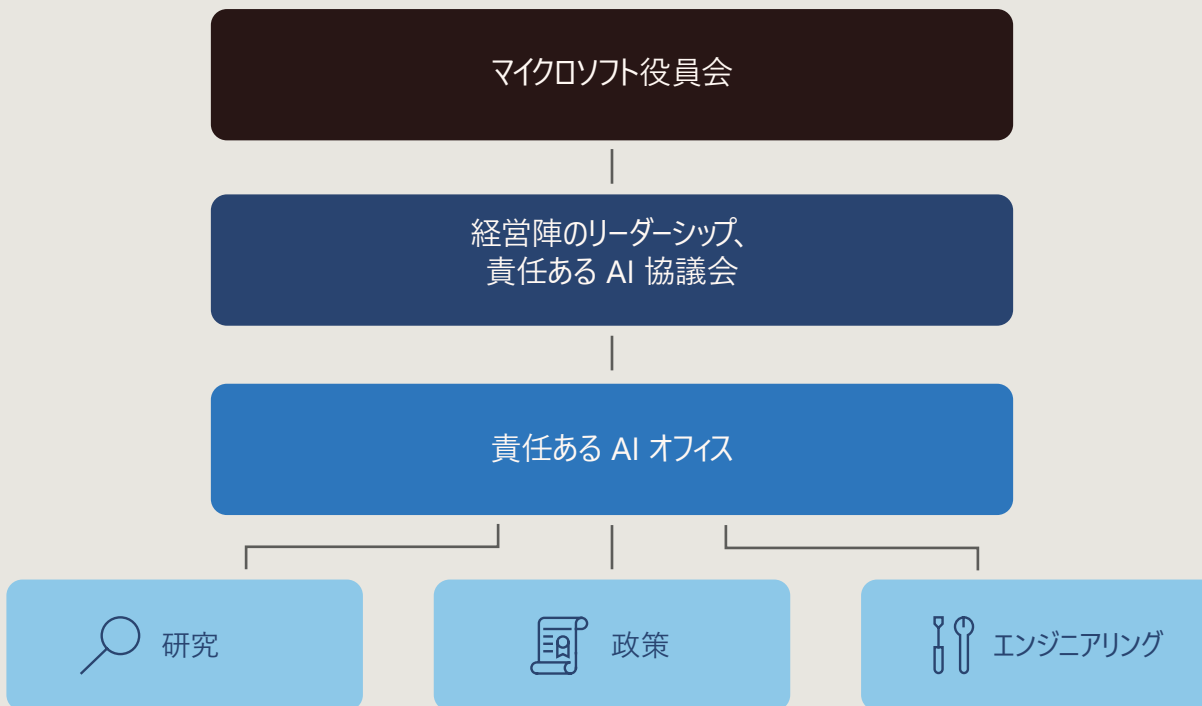
この 7 年間の進展を誇らしく思います。

これらの取り組みがあったからこそ、現在の段階まで到達できました。つまり、AI システムのライフサイクルに安全性と責任を組み込むところまで、取り組みを深めるという段階です。これは、責任ある AI の原則とプラクティスによって従来のサイロ

が取り払われ、部門横断型のチームが協力したときに初めて実現します。機会と潜在的リスクが身近にある現状において、マイクロソフトは、自社の教訓を共有し、責任ある AI のプラクティスをすべての組織が業務に適用できるよう支援しなければならないと考えています。これがまさに、マイクロソフトが取り組んでいることであり、自ら模範となることでリーダーシップを取っていければと思います。



## マイクロソフトのエコシステム



## マイクロソフトでの責任ある AI の運用化

### 基盤となるガバナンス構造の設定

AI の進展のペースが加速し続ける中、マイクロソフトは、ガバナンス構造を継続的に進化させています。このガバナンス構造は、マイクロソフトの責任ある AI プログラムの基盤となる要素として、進歩と説明責任を実現するために構築したものです。マイクロソフトのガバナンス構造の構築と、責任ある AI を企業全体に展開するという決断は、経営層によって推進されました。会長兼 CEO の Satya Nadella と社内の上級幹部チーム全員の指示により、責任ある AI が全社的に義務付けられています。マイクロソフトの経営層は、責任ある AI の対応を 1 つのチームや部門に課すのでは不十分だと認識していました。プライバシー、セキュリティ、アクセシビリティに対する

長年の全社的なコミットメントから学んでおり、責任ある AI を実現するには、社内の最高経営層の支持を獲得すると共に、すべての階層で推進される必要があることを悟っていたのです。

この状態を達成するために、マイクロソフトの責任ある AI オフィスは、企業全体の多様なチームと部門を取り込んだガバナンス システムを構築しました。実務レベルでは、エンジニアリング部門、研究部門、政策部門の中核チームがそれぞれに特有のスキルを持ち寄り、責任ある AI を全社的に推し進めるうえで重要な役割を担っています。また、責任ある AI に携わる職務は経営層によって「責任ある AI チャンピオン」として指名され、製品チーム、エンジニアリング チーム、セールス チームにも組み込まれています。責任ある AI チャンピオンの任務は、各自のチーム内で責任ある AI のプラクティスを率先し

て実践することです。つまり、責任ある AI の基準を採用し、問題を見つけたら軽減策の候補についてチームに直接アドバイスし、責任あるイノベーションの文化を育むことです。責任ある AI オフィスは、全社的にこれらのチーム間の調整をサポートし、製品に関する深い知識と責任ある AI に関する専門知識を活かして、マイクロソフト全体で一貫したアプローチを策定します。

その次の階層に当たる責任ある AI 協議会は、マイクロソフトの責任ある AI プログラムの導入に関して、経営層の調整や説明責任を果たすための場です。この協議会の議長を務めているのは、マイクロソフトの副会長兼プレジデントの Brad Smith と最高技術責任者の Kevin Scott です。Kevin は企業としてのテクノロジービジョンの設定と、マイクロソフト リサーチ部門の監督を担っています。責任ある AI 協議会は定期的に招集され、責任ある AI 専任の中核的な研究チーム、政策チーム、エンジニアリング チーム (Aether 委員会と責任ある AI オフィスを含む) の代表者のほか、導入に対して説明責任を負うエンジニアリング責任者や上級ビジネス パートナーも集まります。

一番上の階層に当たるマイクロソフト役員会の Environmental, Social, and Public Policy Committee (環境・社会・公共政策委員会) は、責任ある AI プログラムを監督します。この委員会と定期的にやり取りすることで、マイクロソフトの非常に厳格なエンタープライズ リスク管理フレームワークが、責任ある AI プログラムに確実に適用されます。

### 標準化の必要性

AI システムの目的の考案から、AI の操作方法の設計まで、AI に関するすべての意思決定の中心には常に人間を据えなければなりません。マイクロソフトの責任ある AI の原則には、私たちが守ろうとしている永続的な価値が述べられています。しかし、責任ある AI システムをどのように構築し、展開するべきかに関しては、より具体的なガイダンスが必要です。これが責任ある AI の基準を策定した理由です。これは、責任ある AI の原則を守ることを日常的な習慣とするために、エンジニアリング チーム向けの一連のルールを提言した、より実践的なガイドとなっています。

責任ある AI の基準は、エンジニアリング チームにとって、責任ある AI システムの構築方法に関する実用的なガイダンスとなります。これは数年にわたる全社的な取り組みの成果であり、研究者、エンジニア、弁護士、設計者、政策専門家からのさまざまな意見を反映しています。損害を事前に特定、測定、軽減する方法について、よりいっそう具体的かつ実践的なガイダンスを定めているため、この基準を策定したことは責任ある AI の実践において大きな前進であるとマイクロソフトは考えています。また、この基準では、起こり得る製品の悪用を防止しながら有益な利用を実現するためのツールと管理策を採用することをチームに求めています。

AI 製品のライフサイクル全体に取り組むエンジニアリング チームに対し、責任ある AI の基準は、次の 2 つの方法で具体的な方向性を示しています。

- **目標の明確化:** 目標は、責任ある AI の原則を守るとはどういうことなのかを定義します。また、説明責任などの意味の広い原則を最終的な成果に細分化します。たとえば、AI システムを影響評価やデータ ガバナンス、人間による監視の対象とするなどが、最終的な成果となります。
- **プレイブックの概説:** プレイブックの具体的な手順やステップは、責任ある AI の基準に規定されている目標を達成するために、AI システムのライフサイクル全体にわたってチームが実施しなければならないものです。各ステップは、これらの目標を達成する力をチームに身に付けさせるために利用できるリソース、ツール、プラクティスと対応しています。たとえば、責任ある AI の基準に規定されている目標の 1 つは、AI システムの予測可能な障害や既知の障害の修復時間を最小限にとどめることです。この目標を確実に達成するために、反復的なレッド チーミングによって潜在的損害を特定するようチームに求めています。そのうえで、各種ツールや確立済みの戦略を試したり導入したりすることで、そうした損害の広がり具合を測定し、軽減するようにも求めています。AI システムの潜在的損害を特定、測定、軽減するこのサイクルは、責任ある AI の基準に規定されている要件の多くを補強します。

## 責任ある AI の基準の構造

原則	責任ある AI の取り組みの指針となる永続的な価値
目標	確実に達成する必要がある成果
要件	目標を確実に達成するのに必要な手順
ツールとプラクティス	要件を満たすのに役立つもの

## 抑制と均衡の確保

責任ある AI の基準を作成し、更新するにあたり、マイクロソフトが早い段階で認識していた点があります。それは、多様なユース ケースに関する複雑な社会技術的考慮事項すべてを、網羅的な事前定義済みのルール群としてまとめるのは不可能だということです。そこで、影響の大きいケースや新たに発生した課題と疑問を継続的にレビューして監視するためのプログラムおよびプロセスを設けました。社内ではこれを「機密性の高い使用」と呼んでいます。

「機密性の高い使用」プログラムにより、自社 AI システムの高リスクのユース ケースに携わるチームは監視を強化することができます。このプログラムは 2018 年に Aether 委員会の下で開始され、2019 年に責任ある AI オフィスが発足して以降は、同オフィスによって運営されてきました。2019 年 7 月から 2023 年 5 月までの間に、企業全体で 600 件以上の「機密性の高い使用」ケースのレビューを行いました。このうち、2022 年 7 月から 2023 年 5 月までの期間にレビューしたケースは 150 件近くにのぼります。

「機密性の高い使用」プログラムを、報告、レビュー、ガイダンスのフレームワークとして考えてみましょう。このプログラムではまず、必須の報告要件を満たします。そうすると、責任ある AI オフィスの「機密性の高い使用」チームによる責任ある AI プロジェクトのハンズオン レビューとコンサルティング プロセスが開始されます。最終的に、責任ある AI の基準におけるベースライン要件を補足するプロジェクト特有のガイダンスと要件が策定されます。「機密性の高い使用」のレビュー プロセスが開始されるタイミングは、マイクロソフト従業員が AI システムの開発または展開に携わっており、かつ、その AI システムの想定される使用または悪用によって以下の事態が起こり得る場合です。

1. ユーザーの法的立場または人生の機会に重大な影響をもたらす
2. 重大な身体的被害または心理的被害のリスクを呈する
3. 個人の人権の行使能力が制限される、侵害される、または損なわれる

責任ある AI オフィスの「機密性の高い使用」チームは報告を受けると、プロジェクト チームのメンバー、そのチームの責任ある AI チャンピオン、その他の関連する利害関係者と協力して、優先順位付けを行い、レビュー プロセスを開始します。レビューの体系化と課題の深掘りを促進するために、チームの「責任ある AI の影響評価」や製品ドキュメントなどの作成物を利用するだけでなく、プロジェクト チーム自体とも密接かつ継続的にやり取りを行います。レビュー プロセス中に、集中的なコンサルテーションのために、社内のあらゆる部署から各分野の専門家を呼ぶことも多くあります。特に影響の大きいユース ケースや革新的なユース ケースの場合は、そのプロジェクトをエスカレーションし、「機密性の高い使用」委員会のレビューと助言を受けます。この委員会は、エンジニアリング、研究、人権、政策、法務の社内専門家と、世界各地のお客様対応組織で構成されています。「機密性の高い使用」チームもまた、企画段階から部門横断型を採用しています。メンバーは社会科学、法律、エンジニアリング、政策の分野のバックグラウンドを有し、データサイエンティストや学術研究者、政策アナリスト、弁護士、国際的外交官、機械学習エンジニアとしての職歴を持っています。

レビューの最後に、「機密性の高い使用」チームは、プロジェクトを前に進めるための要件を提示します。これらも、責任ある AI の基準にとどまらない、目下の特定のプロジェクト向けに策定された追加の要件です。「機密性の高い使用」レビューの結果、責任ある AI の基準と原則にプロジェクトが十分に沿っていないと判断したために、その AI アプリケーションの構築と展開の機会を却下したこともありました。たとえば、カリフォルニア州の地元警察当局が、パトロール中にボディカメラや車載カメラの顔認証機能をリアルタイムで利用することについて、マイクロソフトは「機密性の高い使用」のレビュー プロセスで時期尚早だと判断しました。マイクロソフトの副会長兼プレジデントの Brad Smith はこの経緯について公式に語っており、契約を辞退したことを発表しました。「機密性の高い使用」のレビュー プロセスは、規制のない環境で顔認証が利用された場合にもたらされる技術的課題に対処しただけではありません。顔認証の利用については社会全体で議

論されるべきであり、法律の制定も必要であるという見解をまとめるうえでも役立ちました。

「機密性の高い使用」のプロセスによるもう 1 つの重要な成果が、より機密性の高い AI プラットフォーム サービスに対する制限付きアクセスのポリシーであり、これらのサービスの利用と展開に対する審査を厳格化するものです。マイクロソフトはこのポリシーの下、リスク軽減のための技術的な管理策を導入しているだけでなく、潜在顧客に対し、利用申請書の提出、利用目的の開示（事前定義済みの許容可能なユースケースのいずれかに利用目的が合致していることを確認するため）、サービス利用規約を読んで合意した旨の確認を求めています。マイクロソフトの責任ある AI の原則に沿った利用の場合に限り、申請書が承認されます。

### ケーススタディ: 新しい Bing への責任ある AI アプローチの適用

2023 年 2 月、マイクロソフトは新しい Bing をリリースしました。これは、AI を活用した Web 検索エクスペリエンスです。新しい Bing は、Web 検索結果を要約したり、チャットエクスペリエンスを提供したりすることで、ユーザーを支援します。ユーザーは詩、ジョーク、手紙などのクリエイティブなコンテンツを生成することもできます。Bing Image Creator を使用すれば画像も生成可能です。AI を活用した新しい Bing は、OpenAI の最先端の大規模言語モデル (LLM) である GPT-4 をはじめ、マイクロソフトと OpenAI のさまざまな先端技術によって実現しています。マイクロソフト全体の責任ある AI チームは、GPT-4 が OpenAI から一般公開される前の数か月間で、この最先端の AI テクノロジーと新しい Bing の Web 検索機能を連携させるための一連のカスタマイズされた機能とテクノロジーを開発しました。

リリースの準備にあたり、マイクロソフトは責任ある AI エコシステムの総力を挙げて取り組みました。新しい Bing のエクスペリエンスは、マイクロソフトの AI の原則とマイクロソフトの責任ある AI の基準に則って開発されました。また、この開発には、マイクロソフトの責任ある AI オフィス、エンジニアリング チ



ーム、Microsoft Research、Aether など社内全体にわたる、責任ある AI のエキスパートたちが協力しました。

マイクロソフトは、自社の "AI の原則" と "責任ある AI の基準" を指針として、新たなエクスペリエンスがもたらす革新的で有益な用途を守りながら、新しい Bing の潜在的な有害性や不適切利用を特定、測定、緩和することに努めました。以降のセクションでは、マイクロソフトのアプローチを説明します。

## 特定

モデルレベルの作業は 2022 年夏の終わりごろに開始し、まずは GPT-4 の調査分析を行いました。このとき、OpenAI と協力して大規模なレッド チーミングも実施しました。このテストには、この最新テクノロジーに追加のセーフガードを適用していない場合の動作を評価するという目的がありました。具体的な意図としては、有害な応答を生成することで、考えられる悪用手段を明らかにし、機能と制約を特定しようとしていました (なお、「応答」とは AI システムの出力のことであり、この場合は大規模言語モデルの出力を指します。「コンプライーション」、「生成物」、「回答」と呼ばれる場合もあります)。共同での学習により、OpenAI のモデル開発が推進され、リスクに対する理解が深まり、新しい Bing 用の初期の軽減戦略の策定が促されました。

モデルレベルのレッド チーミングだけではありません。新しい Bing AI エクスペリエンスを制限付きリリース プレビューとして公開する前に、専門家から成る部門横断型チームがアプリケーションレベルのレッド チーミングを何度も実施しました。このプロセスにより、敵対的アクターがどのようにシステムを悪用し得るかについての理解を深め、緩和策を向上させることができました。悪意のないテスト担当者も新しい Bing 機能を幅広く評価し、欠陥や脆弱性がないかを確認しました。

## 測定

レッド チーミングによって具体的な損害の例を明らかにすることはできますが、実稼働では、ユーザーが無数のさまざまな会話を新しい Bing と行うこととなります。さらに、会話は複数

回のやり取りから成り、文脈に依存するため、会話内の有害な応答を特定するのは複雑な作業です。新しい Bing の AI エクスペリエンスに含まれる有害性をよりよく把握し、それに対処するため、マイクロソフトは新しい AI エクスペリエンスに特化した "責任ある AI" メトリックスを追加で開発しました。このメトリックスは、脱獄や有害なコンテンツ、根拠のないコンテンツといった潜在的な有害性を測定するのに利用できます。また、マイクロソフトは部分的に自動化された測定パイプラインを通じ、大規模な測定を可能にしました。

この測定パイプラインによって、潜在的損害の測定を大規模かつ迅速に実施でき、変更を実稼働環境に反映する前に毎回テストを行うことができます。プレビュー期間やそれ以降、あるいは継続的なレッド チーミングを通じて新たな問題が特定されるため、マイクロソフトは新たな損害を評価するための測定指標を増やし続けています。

## 軽減

潜在的な損害および悪用の特定と測定を行ったため、従来の検索機能で使用されていた軽減策に加える形で、新たな軽減策を考案しました。その一部を以下に挙げます。

- **プレビュー期間、段階的リリース:** マイクロソフトの漸進的なリリース戦略は、自社のテクノロジーを研究所から社会に安全にリリースする方法の中核部分を成してきました。また、マイクロソフトは、新しい Bing のメリットを確実に提供するために、慎重かつ綿密なプロセスを徹底しています。プレビュー期間にアクセスできる人数を制限したことで、新しい Bing を人々がどのように使用しているのか (どのように悪用する可能性があるかも含む) を確認したうえで、リリース対象を広げることができました。マイクロソフトは得られた教訓に応じて、製品のパフォーマンス向上、既存の軽減策の改善、新たな軽減策の導入を行うために、日々、新しい Bing に変更を加え続けています。
- **有害性や不適切利用を緩和するための AI ベースの分類器とメタプロンプト:** LLM の使用により、有害性や不適切利用につながりかねない問題のあるコンテン

ツが出力される可能性があります。分類器とメタプロンプトは、この種のコンテンツのリスクを減らすために新しい Bing に導入された緩和策の例です。分類子はテキストを分類し、検索クエリやチャット プロンプト、生成された応答に含まれる有害と思われるさまざまな種類のコンテンツにフラグを付けます。フラグは、生成されたコンテンツをユーザーに返さない、ユーザーを別のトピックに誘導する、ユーザーを従来型の検索に転送するといった緩和策の可能性につながります。メタプロンプティングでは、モデルの挙動を誘導するためにモデルに指示を与えます。たとえば、メタプロンプトには「ユーザーが選択した言語でコミュニケーションを取る」といった 1 行を含めることができます。

- **検索結果の根拠付け:** ユーザーが情報を探している場合に、新しい Bing は、Web 検索結果に表示される情報を根拠とした応答を返すように設計されています。たとえば、Bing のシステムには、応答の根拠付けのために、上位の検索結果からのテキストと、メタプロンプトによる指示が与えられます。しかし、Web 上のコンテンツを要約する際、新しい Bing は、入力ソースに存在しない情報を応答に含める場合があります。つまり、根拠のない結果を生成する可能性があるということです。マイクロソフトは、ユーザーが要約のシナリオやチャット エクスペリエンスで生成された根拠のないコンテンツに過度に依存するリスクを緩和するため、いくつかの対策を講じました。たとえば、検索結果に基づく新しい Bing の応答には、ユーザーが応答について検証し、詳細を知ることができるよう、出典 Web サイトへの参照が含まれています。また、ユーザーに対し、今やり取りしている相手は AI システムである旨の明確な注意書きを示すと共に、最良の判断をできるように Web 検索結果上の出典資料を確認することを忠告しています。
- **会話の脱線の制限:** チャット セッションが非常に長くなると、応答が繰り返しになったり、役に立たない内容になったり、新しい Bing で意図されているトーンでなくなったりする場合があることが、プレビュー期間中に判明しま

した。このような会話の流れに対処するために、チャットセッションあたりのターン (ユーザーからの質問と Bing からの回答の両方を含むやり取り) の回数を制限しました。これは、問題をより適切に軽減できるようにシステムをアップデートするまでの措置です。

- **AI に関する開示:** 新しい Bing では複数のタッチポイントで、AI に関する重要な開示を行っています。そこで、ユーザーは自分が AI システムとやり取りしていることや、新しい Bing の詳細を確認する機会について知らされます。

マイクロソフトが知識を増やし、プレビュー期間やそれ以降に集められたフィードバックを基に改善を施していくにつれて、損害を特定、測定、軽減するためのアプローチは進化し続けていくでしょう。

プライバシー、デジタルの安全性、透明性に対する取り組みなど、新しい Bing に関する責任ある AI の取り組みについては、<https://aka.ms/ResponsibleAI-NewBing> で詳しく紹介しています。

## 企業文化を通じた責任ある AI の促進

手順と基準は、責任ある AI を運用化するうえで欠かせない要素であり、責任ある AI の原則と行動を守る文化を構築するのに役立ちます。こうした補完的なアプローチのおかげで、マイクロソフトはコミットメントを達成しやすくなります。

従業員はマイクロソフトの文化の中核を成します。一人ひとりがマイクロソフトのミッションと目標に貢献するのです。マイクロソフトは、責任ある AI を推進する文化を深めるために、AI に注力する人材に投資すると共に、すべての職務に責任ある AI の当事者意識を持たせるようにしています。

## 人材への投資

マイクロソフトは長年にわたり、責任ある AI に向けたコミットメントの一環として、人材に多大な投資を行ってきました。今で

は 350 人近くの従業員が責任ある AI に携わっており、そのうち 3 分の 1 以上は責任ある AI 専任です。これらの従業員は、政策、エンジニアリング、研究、セールスなどの中核部門で働き、責任ある AI をマイクロソフトのビジネスのあらゆる側面に取り入れています。

AI システムを開発および利用するチームには、社会技術的観点でテクノロジーを考えるよう求めています。つまり、マイクロソフトは、さまざまな展開状況で現れる AI の複雑な文化的要素、政治的要素、社会的要素を考慮しているということです。また、コンピューターサイエンスに対する従来のアプローチを抜本的に変えているということでもあります。これまでに設けてきたトレーニングやプラクティスのおかげで、個人レベル、社会レベル、世界レベルでの AI の有益な影響や有害となり得る影響をチームが予測しやすくなっていることは事実ですが、これだけでは不十分です。AI システムを開発しているチームとその報告を受ける経営層には、見落としている点があるかもしれません。これが、責任ある AI のコミットメントにおいて多様性とインクルージョンが重要である理由です。

多様な人材とインクルーシブな文化への投資に関する主張は定着しています。それでも、責任ある AI を実現するうえでの多様性とインクルージョンの重要性は、どれだけ強調してもしすぎることはありません。このため、責任ある AI のエコシステムに対して行っている拡大中の投資では、多様な人材の新規採用も対象になっています。年次の『[ダイバーシティ & インクルージョンレポート](#)』で示したとおり、マイクロソフトは多様性とインクルージョンにおいて少しずつ前進を続けています。しかし業界としては、まだ長い道のりがあります。AI の分野では依然として白人と男性が中心であり続けています。マッキンゼーによる [2022 年版の AI に関するグローバル調査](#) によれば、AI ソリューションに携わる従業員のうち、女性や人種/民族的マイノリティはそれぞれわずか 4 分の 1 程度にすぎません。

マイクロソフトはすべてのレベルで多様性とインクルージョンを推進し続け、責任ある AI プログラムにおいては特に力を入れていきます。できるだけ広く社会の役に立つ AI システムを

構築するには、採用活動によって、多様かつ活発で積極的に関わり合う従業員のコミュニティを作り、維持していく必要があります。

### すべての職務に責任ある AI の当事者意識を持たせる

マイクロソフトで働くすべての人が、責任ある AI のコミットメントに沿った AI システムに貢献する機会と責任を有していると私たちは考えています。あらゆる職務のすべての従業員が多様なスキル、観点、情熱を活かし、この活動に何かをもたらすのです。このように見方を変えるには、つまり、職務やチームにかかわらず誰もが責任ある AI を推進できると考えるためには、文化を変える必要があります。

こうした文化の成長を支えるために、マイクロソフトは従業員のスキル開発やコラボレーションの促進に投資してきました。

### 知識とスキルの開発

マイクロソフトは、個人と社会に対する AI システムの潜在的影響について幅広く考える力をチームに身に付けさせるために、トレーニングとプラクティスを作成してきました。

たとえば、AI システム設計の最初期段階にあるチームの場合、以下の作業において影響評価が参考になります。

- AI システムの利用目的を明確化する
- AI システムが解決すべき問題をどう解決するのかを調べる
- 影響を受ける利害関係者を特定する (マイクロソフトの直接のお客様に限られない)
- 各利害関係者に影響を与える可能性がある潜在的な損害およびメリットを明確化する
- 潜在的損害に対する事前軽減策を説明する

チームによる影響評価の実施をサポートするため、責任ある AI オフィスは、オンデマンド トレーニングや対面ワークショップ、



例とプロンプト質問を記載した補足的なガイドンスドキュメントを作成しました。マイクロソフトはベスト プラクティスの共有に取り組んでおり、その一環として、影響評価のテンプレートとガイドンスドキュメントを公開しています。

マイクロソフトの全従業員が利用できる広範な責任ある AI トレーニング コースでは、責任ある AI の原則やガバナンス プロセスを詳しく説明するなどして、責任ある AI に向けたマイクロソフトのアプローチに従業員を適応させています。また、データサイエンティストと機械学習エンジニア向けに特別に作成されたコンテンツも提供しています。

チームは、マイクロソフト エコシステムにいる多様な責任ある AI 専門家を活用することもできます。これらの専門家は製品ライフサイクル全体を通じて、リアルタイムの協力やフィードバックを行います。こうしたコミュニティには、Aether 委員会、責任ある AI オフィス、責任ある AI チャンピオンのコミュニティがあります。責任ある AI チャンピオンは、責任ある AI の基準の採用を推し進める職務であり、そのコミュニティは大規模で拡大し続けています。

## Azure Machine Learning に組み込まれた責任ある AI



### 公平性

公平性を評価し、公平性に関する問題を軽減することで、すべての人のためのモデルを構築



### 説明可能性

モデルの特徴量重要度の値を生成することで、モデルの予測を理解



### 反事実条件

特徴量の摂動を確認し、モデル予測が異なる最も近いデータポイントを特定



### プロンプト フロー

大規模言語モデルを基盤とするアプリケーション向けのワークフローを作成し、プロンプトの構築、評価、チューニングを簡素化



### 原因分析

ある特徴量が現実の結果に与える影響を推定



### エラー分析

エラー率が高いデータセット コホートを特定し、モデル内のエラー分布を可視化



### 責任ある AI スコアカード

責任ある AI に関する自社の洞察をまとめた PDF を入手し、コンプライアンス レビューの参考として、技術分野とそれ以外の分野の利害関係者に共有



### Azure Content Safety

ヘイト、暴力、性、自傷行為に関するあらゆる言語のコンテンツ (画像とテキストの両方) を検出



## コラボレーションの促進

責任ある AI に対する取り組みの初期段階でマイクロソフトが認識していた点があります。それは、責任ある AI のプラクティスを構築するうえで社内の研究者、政策専門家、エンジニアが果たす重要な役割です。各グループが、取り組みに欠かせない洞察と専門知識をもたらします。そのためマイクロソフトは、グループ間のコラボレーションを実現すべく努力しています。

- 機械学習から人文科学まで、さまざまな専門知識を持った研究者が、最先端の AI システムの構想に力を

貸してくれています。また、AI システムによって生じる潜在的損害の特定、測定、軽減に関するベスト プラクティスや、AI イノベーションの興味深い機会についての洞察をもたらしています。

- 政策専門家は、責任ある AI のガバナンスを定義し、運用化します。これには、AI システムの責任ある開発の指針となるルールづくりも含まれます。マイクロソフトのガバナンス フレームワークでは、説明責任を与えてコラボレーションを促進する形で、組織内の役割と責任を概説しています。

## 責任ある AI チャンピオン

マイクロソフトの責任ある AI チャンピオンの紹介

マイクロソフトは企業全体で、責任ある AI チャンピオンのネットワークを育ててきました。これらの一人ひとりが、Responsible by Design の文化を推進するうえで欠かせません。

### Mihaela Vorvoreanu (研究部門)



「責任ある AI は、技術的問題に技術的ソリューションで 対応することだけではありません。責任ある AI 専門家のみならず、人間に関する専門家とも早い段階から深く協力していく必要があります」

### Ferdane Bekmezci (データ サイェンス部門)



「組織に文化を植え付けるのには時間がかかります。私は文化の採用を全社的に推進していくことに情熱を持っています。倫理的かつ社会的に信頼できる方法で AI が開発および利用されるようにすることは重要だからです」

### Alejandro Gutierrez Munoz (データ サイェンス部門)



「責任ある AI を推進することは、倫理原則に沿った AI システムの実現、信頼の醸成、コンプライアンスの確保、社会的責任の推進のために極めて重要です」

### Lisa Mueller (設計部門)



「AI は急速に変化しています。そのため、AI システムに対する信頼を構築し、高め、広げていくためには、コミュニティの拡大と AI 原則の全社的な採用が重要です。このようなアプローチの一環として、多くの分野にこの取り組みに貢献してもらい、真の差別化を図ることも重要です」

### Shweta Gupta (カスタマー エンジニアリング部門)



「多様な利害関係者を連携させて、責任ある AI の原則を適用しながら AI ソリューションを構築することで、潜在的 リスクの特定と対処が容易になるだけでなく、開発中の システムが全体として目的にかなうものであることを確認 できます」

- エンジニアは、責任ある AI の基準に従った AI システムを設計および開発します。AI システムによって生じる潜在的損害の特定、測定、軽減に必要な手順を自動化し、大規模に実行します。また、教訓を基に、実現可能なまったく新しい責任ある AI ソリューションを構築します。

頻繁にコラボレーションを行い、互いの専門知識 (経営層によって強化されたプラクティス) を信頼し合うことで、より有益で責任あるソリューションにつながる文化を構築できています。継続的な対話の結果、どのチームも、「人間中心に考えて協力しながら AI に取り組めば、責任ある製品だけでなく全体として優れた製品を構築できる」と口をそろえています。

## 責任ある AI に取り組むお客様の支援

責任ある AI に対するマイクロソフトのコミットメントのうち、特に重要なものの 1 つが、責任ある AI に取り組むお客様に自社の教訓を共有して、支援することです。責任ある AI のプラクティスが採用されたときに、思い描くとおりの社会的利益を確実に実現するには、マイクロソフトの努力だけでは不十分です。

このコミットメントの一環として、マイクロソフトは自社のプラットフォーム AI サービスの透明性に関するドキュメントを Transparency Note として提供し、お客様による責任あるシステム展開を支援しています。Transparency Note では、マイクロソフトのプラットフォーム テクノロジーをいつどのように展開すべきかをお客様に理解していただけるように、明快かつ日常的な言葉を用いて AI システムの目的、機能、制約をお伝えしています。また、ソリューションの機能や責任ある AI の基準を超えるユース ケースも記載しています。Transparency Note はマーケティング ドキュメントと技術ドキュメントのギャップを埋めるものであり、AI の責任ある展開を行うお客様が把握しておくべき情報を積極的にお伝えしています。Azure OpenAI Service の Transparency Note の例を[こちら](#)で確認できます。

お客様は、責任ある AI のプラクティスを運用化するための実践的なツールも必要としています。長年にわたり、マイクロソフトにおける責任ある AI の研究は各種ツール (Fairlearn や InterpretML など) の計画につながってきました。このツール群は機能面で向上してきています。たとえば、AI システムの潜在的なエラーや制約を特定、診断、軽減する機能など、その範囲は、責任ある AI のプラクティスの多くの側面にわたります。これらのツールはマイクロソフトで最初に構想されて以降、積極的なオープンソース コミュニティの貢献を通じて社外で改善され、進化し続けてきました。このツール群は、責任ある AI ツールボックスの GitHub リポジトリで確認できます。マイクロソフトの最新のツールは、現在プレビュー中の Azure Content Safety です。これは、ヘイト、暴力、性、自傷行為に関するあらゆる言語のコンテンツ (画像とテキストの両方) を検出するよう設計されたモデルを活用して、より安全なオンライン環境とコミュニティの構築を支援します。

マイクロソフトの責任ある AI プログラムでは、責任ある AI ツールボックスを基にして、より成熟度の高い責任ある AI ツールのいくつかを Azure Machine Learning に直接統合する作業に投資しています。その目的は、お客様もエンジニアリング システムとツールの開発によるメリットを享受できるようにすることです。こうした機能をまとめた画面が、責任ある AI ダッシュボードです。機械学習の実務者やビジネス利害関係者は、AI システムの構築や既存の AI システムのカスタマイズの際に、この 1 つの画面でモデルのデバッグを実行したり、情報に基づく責任ある決断を下したりできます。プレビューに追加された最新機能のいくつかでは、テキスト データと画像データをサポートしています。このため、ユーザーは、モデルの構築、トレーニング、評価の段階で、非構造化データによって構築された大規模モデルを評価できます。また、プロンプト フロー (根拠などの尺度を基に、大規模言語モデルのプロンプトの構築、評価、チューニングをスムーズに行える機能) をサポートしている最新機能もあります。

## AI のガバナンス: 日本に関する青写真

マイクロソフトは今後もこれまでと同様、責任ある AI に関する研究主導の革新的成果を基に、責任ある AI に取り組むお客様を支援する実践的なツールを生み出す作業に投資し続けます。

AI の開発、評価、利用に関わるコミュニティは、マイクロソフトの直接のお客様以外にも広がっています。この幅広いエコシステムのお役に立てるよう、マイクロソフトは、責任ある AI プログラムでの重要な作成物を一般に公開しています。これには、責任ある AI の基準、影響評価テンプレート、最先端の研究成果の集積などがあります。マイクロソフトのデジタル学習パスを活用すれば、経営層は、効果的な AI 戦略の考案、AI 対応の文化の醸成、責任あるイノベーションなどの能力を強化できます。これらのリソースはオンライン (<https://aka.ms/rai>) でご確認いただけます。



第 3 部  
日本の AI の現状



## AI で社会的課題に対処するには

AI は、大きな社会的課題に対処する絶好の機会を世界中の国々にもたらしています。以下、すでに AI を利用して医療、高齢化、環境持続可能性、教育、公共サービスなどのさまざまな分野で変革を推進している日本のイノベーターの事例をいくつか紹介します。

マイクロソフトは、生成 AI の可能性を解き放つ日本の取り組みに寄り添う準備ができています。マイクロソフトは、神戸市およびその他のパートナー企業と共に、2023 年 10 月 11 日に Microsoft AI Co-Innovation Lab を神戸市に開設する予定です。Microsoft AI Co-Innovation Lab は、関心のある企業にソリューションの構築、開発、試作、テストのための施設へのアクセスを提供するもので、マイクロソフトの AI テクノロジー専門家がこの共創活動に直接関与できます。この神戸市の施設は世界で 5 番目となる Co-Innovation Lab で、社会問題に対する革新的な解決策、グローバルなビジネス、未来の AI 人材を生み出す役割を担います。将来、この Co-Innovation Lab 発の AI ユースケースを目にする機会が増えることを期待しています。

### より健康な未来のための AI

#### FRONTEO

日本の認知症患者数は、2025 年までに 730 万人（高齢者の約 5 人に 1 人）に達すると予想されています。超高齢化社会を迎えた日本では、認知症ケアは早急に対処する必要がある国家的問題です。しかし、認知症の診断は専門知識と経験が必要なため容易ではなく、早期診断と治療の妨げになる可能性があります。FRONTEO は、この課題に対処するために、認知症診断を支援する AI システムを Microsoft Azure と AWS 上で開発しています。この AI 統合モデルは、発言内容や言葉の選択の傾向を分析することにより、患者と医療専門家との日常会話から抽出した 5

分～10 分程度のテキストデータから認知症の兆候を発見できます。急速に進む高齢化を考えると、早期発見は専門医療機関で認知症治療を速やかに開始するうえで不可欠です。この革新的なサービスは、高齢化によって悪化する国家的問題に対処するための業界の総力を挙げた取り組みを反映するだけでなく、日本におけるデジタル医療の目覚ましい進歩も示しています。

この認知症発見システムの基礎となった研究開発プロジェクトは、国立研究開発法人 日本医療研究開発機構 (AMED) の資金提供による研究プログラムの後援の下で実施されました。FRONTEO は、2020 年に日本特許庁から認知症診断支援 AI システムに関する特許認可を受け、現在は製造承認の取得に向けて取り組んでいます。

#### 国立研究開発法人 日本医療研究開発機構 (AMED)

日本は、ゲム医療の発展のために、厚生労働省が 2019 年 12 月に発表した全ゲム解析等実行計画を積極的に推進しています。国立研究開発法人 日本医療研究開発機構 (AMED) は大規模な取り組みに着手し、2021 年から 2022 年の間に 1 万 2,000 人分の全ゲム解析を完了し、年間 1 万人以上まで規模を拡大することを視野に入れています。1 人の患者のデータだけで 400GB を超えることもある膨大なデータ量を考えると、堅牢なインフラストラクチャが不可欠です。これに対し、AMED はオンプレミスサーバーからクラウドにプラットフォームを移行し、マイクロソフトの AI 機能を利用して効率的なデータ管理と迅速な規模拡張を実現することを計画しています。AMED は、マイクロソフトの AI テクノロジーが日本における幅広い医療管理および疾患治療活動に役立つものと期待しています。

## より持続可能な未来のための AI

### AGRIST

AGRIST のロボット開発への取り組みは、「収穫期の人手不足を解決するために収穫ロボットが必要」という農家の声から始まりました。日本のスタートアップ企業である AGRIST のミッションは、テクノロジーを通じて農業の課題に対処する持続可能なビジネス モデルを創出し、100 年先も続く持続可能な農業を実現することです。AGRIST は自動収穫ロボットを開発しており、そのロボットを通じて作物画像データを収集し、そのデータを農業ビッグ データに変換しています。そこから、収穫量の変化や害虫や病害の発生などの条件をビッグ データとして分析する新しい農業システムを確立することを目指しています。

Microsoft Azure は、データ蓄積および AI の力を利用した分析の基盤の役割を果たしています。AGRIST は Microsoft for Startups プログラムの重要なメンバーです。AGRIST は、農作物収量を増加させるための次の進化として、「agriss」というオペレーティング システムの開発に着手しています。ロボットから収集した農場ビッグ データを AI によって分析することにより、作物収量予測など、データドリブンで再現性と生産性の高い農業の実現を目指しています。AGRIST の夢は、日本を超えて世界の食糧問題の解決に寄与し、それによって「全人類の幸福・Well-being」に貢献することです。AI は、この大きな夢を実現できるかもしれません。



AGRIST が開発した自動収穫ロボット

### パナソニック コネクト

日本の大手電機メーカー、パナソニックグループの子会社であるパナソニック コネクトは、2023 年 2 月、国内 1 万 2,500 人の従業員全員に自社独自の AI アシスタント「Connect AI」を導入しました。これは、メール作成、情報収集、コンピューター コードの作成といった日々の作業への利用を目的としています。Connect AI は、エンタープライズレベルのデータおよびプライバシー保護を可能にする Microsoft Azure OpenAI Service を使用して開発されました。AI アシスタントへの問い合わせ件数は、導入当初は 1 日あたり 2,000 件でしたが、わずか 2 ~ 3 か月で 5,000 件以上に急増しました。パナソニック コネクトの上級担当者は、「すべての実務担当者が AI を日常業務の一環として利用すべきと考えています。そのため、検討したのは AI を利用すべきかどうかではなく、いつ利用し始めるかでした」と述べています。日本は世界で最も急速に高齢化が進んでおり、持続可能な経済成長の見通しに影を落としています。パナソニック コネクトは生成 AI を、人間にしかできない創造的作業に集中できるようにして従業員の生産性を高める方法の 1 つと考えています。



パナソニック コネクト CEO の樋口泰行氏は、ConnectAI を使用して新入社員向けの歓迎スピーチを作成

## PKSHA

日本のスタートアップ企業 PKSHA Technology の子会社である PKSHA Workplace は、生成 AI の力を利用して働き方の未来を実現することを目指しています。同社は、社内問い合わせへの対応、自動応答、チームメンバー間のコラボレーション促進を可能にするシステムとして「AI ヘルプデスク for Microsoft Teams」を 2022 年に発売しました。最新バージョンでは、Azure OpenAI Service を介して大規模言語モデルを利用して、AI ヘルプデスク内の会話記録から直接 FAQ を自動生成する新機能を搭載しました。この機能により、知識の共有や創造に必要なコストの削減だけでなく、自動応答率や自己解決率の向上も可能になります。

## JERA

JERA は、デジタル イノベーションを通じたエネルギー産業の変革と世界のエネルギー課題への対処に重点的に取り組む企業です。同社は、発電所の運転効率の向上と環境影響の低減に向けた取り組みを開始しました。マイクロソフトは、JERA のこの取り組みで主要パートナーとして貢献しました。JERA とマイクロソフトは、Microsoft Azure Digital Twins と生成 AI を利用して、新しい運用 & 保守 (O&M) ソリューションを共同開発中です。その狙いは、発電所の性能を改善すると同時に停止時間を削減し、それによって運転効率を高めることです。このプロジェクトは、クラウドと AI を利用して発電所からのリアルタイム運転データを遠隔で分析することで、問題の予知および検出の自動化、意思決定の迅速化、発電所管理の効率化を実現します。発電所の運転担当者を時間のかかるデータ収集・分析作業から解放し、高度な運転業務や価値創造に集中できるようにすることにより、従来の発電所の運転に劇的な変化をもたらすでしょう。JERA は、マイクロソフトとのパートナーシップを通じて CO2 排出量と環境フットプリントが削減され「JERA ゼロエミッション 2050」目標の達成に寄与することを期待しています。

## 教育とエンパワーメントのための AI

### 沖縄大学

2016 年の「障害者差別解消法」の制定以降、高等教育における障がいのある学生数は着実に増加しています。これに関連して大きな課題となるのが聴覚障がいのある学生に対する支援の提供です。大学の講義は黒板を使った説明よりも口頭での説明に頼ることが多いためです。また、ゼミでの議論や発表も主に口頭で行われます。

沖縄大学は、建学の精神である「教育の機会均等」に基づき、障がいのある学生を積極的に受け入れてきました。従来、同学はノートテイクをボランティアの学生に頼っていましたが、補助する学生との調整が必要で、文字起こしの品質にばらつきが生じていました。現在は、合同会社シコミュが提供する「AI mimi」サービスを利用して、障がいのある学生の



オペレーターがテキストを修正し、字幕を編集。AI と人間の共同作業。



聴講を支援しています。「AI mimi」は、Azure Cognitive Services による音声認識に、人間の専門オペレーターが行う修正作業を組み合わせたハイブリッド文字起こしサービスです。このサービスは、マイクロソフトの「AI for Accessibility」プログラムにも採用されており、音声をわずか 1 ～ 2 秒のタイムラグで文字に高精度変換します。情報のアクセシビリティ向上は、全日本国民の教育の機会均等とエンパワーメントに向けた重要な一歩です。

### Microsoft Base Ritsumeikan

2023 年 8 月、立命館大学はマイクロソフトが後援する先駆的なイニシアティブとして、次世代研究ベース「Microsoft Base Ritsumeikan」を設置する計画を発表しました。同学は、日本の教育システムを社会において急速に進む技術の進歩に合わせることで、生成 AI やメタバースなどの最先端テクノロジーを革新的な学習に有効利用することを目指しています。特に、大学運営用および学生関連の内部データベースの管理と利用のために、Microsoft Azure を利用して立命館オリジナルの生成 AI モデルを開発する計画です。2024 年 4 月に開設予定の Microsoft Base Ritsumeikan は、立命館の学生と教職員だけでなく、他の教育機関、地方自治体、および企業の人々が集い合って、新時代の新たな社会的課題に協力して対処する場となるでしょう。

### 弁護士ドットコム

日本では、法的手続きに巻き込まれた人の 2 割程度しか弁護士に相談できず、これは「2割司法」と呼ばれています。2005 年に弁護士が設立したスタートアップ企業である弁護士ドットコムはこの問題に対処するため、誰でも無料で弁護士に相談できる「みんなの法律相談」と呼ばれる既存のサービスを基に、2023 年 5 月に「チャット法律相談」を開発しました。この新たな AI 法律相談チャットサービスはマイクロソフト Azure 上に構築されており、2007 年以来「みんなの法律相談」に寄せられた 125 万件以上の相談から抽出した質問と回答を利用しています。このサービスにより誰でも 24 時間 365 日無料で気軽に法的助言を求めることができるようになり、誰でも法律相談を利用できる未来が実現するでしょう。

## 公共サービスの未来のための AI

### 東京都庁

東京都庁は、現代で最も変革的なテクノロジーである ChatGPT を積極的に導入し、活用する先駆者になりました。

2023 年 8 月に AI の業務利用を開始し、約 5 万人の職員が文書要約の業務や政策立案の参照用に ChatGPT を利用しています。この重要な決定を実現したのは、Microsoft Azure が提供する強力なデータおよびプライバシー保護でした。同庁は、高い利用率に対応するために、4 月に設置されたプロジェクトチームの検討に基づいてガイドラインをまとめました。ガイドラインには、プロンプト エンジニアリングに似た ChatGPT の実用的な活用方法のほか、具体例も示されています。

関係者は、テキスト生成 AI の導入を検討中の他の地方自治体もこれらのガイドラインを活用してほしいと考えています。東京都庁の決定は、単なる具体的なユースケースの 1 つとしてではなく、日本全国の公共サービスの未来に向けた大きな一歩と考えるべきです。



## 大阪府

大阪府庁は、住民 QOL (生活の質) 向上を主な目的として業務に AI を導入するために、マイクロソフトと事業連携協定を締結しました。特に、スマートシティ関連施策に生成 AI を利活用すること、人材育成活動に AI を取り入れることを意図しています。

さらに大阪府の社会課題の解決に貢献するため、マイクロソフトは、高齢者サポート、孤立化対策、屋外活動を促すことを目的に、高齢者向けのコミュニケーション支援サービスの開発を支援する予定です。この「大ちゃんと話す」は、関西弁で読み書きしながら会話やイベント情報の提供が可能な「大ちゃん」という名前の犬のキャラクターを特徴としています。高齢者の孤立化は、高齢化が進む日本の社会が抱える問題の1つであり、これはそうした高齢者の孤立化問題に地方自治体が生成 AI を採用した最初の事例です。

## 浜松市

外国人居住者の数が非常に多く、多様な人口構成でよく知られる浜松市は、多文化コミュニティに対応するために積極的な対策を講じてきました。注目に値する施策の1つとして、地方新聞や浜松広報誌を日本語版だけでなく英語版やポルトガル語版でも発行していることが挙げられます。しかし、住民の多様性が拡大するにつれて、言葉の壁が原因で十分な情報を提供できないことが慢性的な課題になっていました。

こうした課題にさらに対処するために、浜松市役所は VOTE FOR が Microsoft Azure をベースに開発したクラウドベースの広報および公共情報支援サービス「PR Plus」を導入しました。PR Plus はさまざまな言語を網羅した翻訳および音声合成機能を備えており、浜松市の公共コミュニケーションのリーチは大幅に拡大しました。このサービスを利用することで、日本語能力に関係なく外国人居住者の情報アクセスを促進するだけでなく、視覚障がい者のアクセシビリティ向上にもつながりました。PR Plus に移行した結果、読者数が大幅に増加し、デジタル記事には常に約 1 万人の読者からのアクセスがあります。さらに、住民のデジタルコンテンツの閲覧状況について理解を深めるために、AI を利用した分析ツールを将来利用することも検討中です。PR Plus と浜松市は、利用者行動を把握することにより、メッセージを表示するタイミングや方法を AI で最適化して広報の効果をさらに高めることが可能になるものと期待しています。

# 参考文献

## 序文

## 第 1 部 日本における AI のガバナンス

### 引用文献1

[Japan's population drop since 2015](https://asia.nikkei.com/Economy/Japan-s-population-drops-to-126m-in-2020-census-down-0.7-vs.-2015)  
<https://asia.nikkei.com/Economy/Japan-s-population-drops-to-126m-in-2020-census-down-0.7-vs.-2015>

[G7 Hiroshima Summit 2023](https://www.g7hiroshima.go.jp/en/Hiroshima-AI-Process)  
[https://www.g7hiroshima.go.jp/en/Hiroshima AI Process](https://www.g7hiroshima.go.jp/en/Hiroshima-AI-Process)

[Hiroshima AI Process](https://www.bloomberg.com/news/articles/2023-05-20/g-7-leaders-agree-to-set-up-hiroshima-process-to-govern-ai)  
<https://www.bloomberg.com/news/articles/2023-05-20/g-7-leaders-agree-to-set-up-hiroshima-process-to-govern-ai>

[Tentative Summary of AI Issues](https://www8.cao.go.jp/cstp/ai/ronten_youshi_yaku.pdf)  
[https://www8.cao.go.jp/cstp/ai/ronten\\_youshi\\_yaku.pdf](https://www8.cao.go.jp/cstp/ai/ronten_youshi_yaku.pdf)

[LDP AI Whitepaper](https://www.taira-m.jp/ldp%E2%80%99s-ai-whitepaper-etrans_2304.pdf)  
[https://www.taira-m.jp/ldp%E2%80%99s ai whitepaper etrans\\_2304.pdf](https://www.taira-m.jp/ldp%E2%80%99s-ai-whitepaper-etrans_2304.pdf)

[Economic Security Promotion Act](https://www.cfr.org/sites/default/files/pdf/economic-security-promotion-act-%28summary%29%28English%29.pdf)  
[https://www.cfr.org/sites/default/files/pdf/economic security promotion act %28summary%29%28English%29.pdf](https://www.cfr.org/sites/default/files/pdf/economic-security-promotion-act-%28summary%29%28English%29.pdf)

[AI Governance in Japan v1.1 from Expert Group](https://www.meti.go.jp/shingikai/mono_info_service/ai-shakai_jisso/pdf/20210709_8.pdf)  
[https://www.meti.go.jp/shingikai/mono\\_info\\_service/ai-shakai\\_jisso/pdf/20210709\\_8.pdf](https://www.meti.go.jp/shingikai/mono_info_service/ai-shakai_jisso/pdf/20210709_8.pdf)

### 一般情報源

[Governing AI: A blueprint for the future](https://aka.ms/GoverningAI-blueprint)  
<https://aka.ms/GoverningAI-blueprint>

[Japan ESPA cutting edge technology development](https://japannews.yomiuri.co.jp/politics/politics-government/20220513-27693/)  
<https://japannews.yomiuri.co.jp/politics/politics-government/20220513-27693/>

[Japan Expert Group on Architecture for AI Principles to be Practiced](https://www.meti.go.jp/english/press/2022/0128_003.html)  
[https://www.meti.go.jp/english/press/2022/0128\\_003.html](https://www.meti.go.jp/english/press/2022/0128_003.html)

## 参考文献

[Fortune business insights AI market report](https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114)  
<https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114>

## 第 2 部 Responsible by Design: 社会に利益をもたらす AI システムの構築に向けたマイクロソフトのアプローチ

### 引用文献1

[NIST AI Risk Management Framework](https://www.nist.gov/itl/ai-risk-management-framework)  
<https://www.nist.gov/itl/ai-risk-management-framework>

[Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies](https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/)  
<https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>

[OECD AI Principles](https://www.oecd.ai/en/ai-principles)  
<https://www.oecd.ai/en/ai-principles>

[OECD Framework for the Classification of AI systems](https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm)  
<https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm>

[UNESCO Recommendation on the Ethics of Artificial Intelligence](https://www.unesco.org/en/artificial-intelligence/recommendation-ethics)  
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

### 一般情報源

[Governing AI: A blueprint for the future](https://aka.ms/GoverningAI-blueprint)  
<https://aka.ms/GoverningAI-blueprint>

## 参考文献

[Critical Infrastructure Sectors | CISA](#)

<https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors>

## 第 3 部 日本の AI の現状

### 一般情報源

[FRONTEO \(高齢化\)](#)

<https://www.fronteo.com/en/20201021>

[国立研究開発法人 日本医療研究開発機構](#)

<https://news.microsoft.com/ja-jp/2023/04/03/230403-whats-next-for-healthcare-the-latest-trends-in-dx-support-for-the-healthcare-industry/>

[AGRIST](#)

<https://www.microsoft.com/ja-jp/industry/blog/retail/2022/02/10/microsoft-for-startups-agrist/>

[パナソニック コネクト](#)

<https://news.microsoft.com/apac/features/not-if-but-when-why-japans-panasonic-connect-is-going-all-in-on-ai/>

[PKSHA](#)

<https://prtimes.jp/main/html/rd/p/000000088.000022705.html>

[JERA \(エネルギー\)](#)

<https://news.microsoft.com/ja-jp/2023/09/11/230911-jera-and-microsoft-partner-to-create-more-efficient-and-sustainable-power-plant-operations-jp/>

[沖縄大学](#)

<https://customers.microsoft.com/ja-jp/story/1426663873879628640-okinawa-higher-education-azure-jp-japan>

[Microsoft Base Ritsumeikan](#)

<https://news.microsoft.com/ja-jp/2023/08/21/230821-accelerating-learning-challenges/>

[弁護士ドットコム](#)

<https://www.bengo4.com/corporate/news/article/zztucptlqh6>

[東京都庁](#)

<https://www.nikkei.com/article/DGXZQOCC156IM0V10C23A8000000/>

[大阪スマートシティ関連施策](#)

<https://news.microsoft.com/ja-jp/2023/09/08/230908-signed-a-business-collaboration-agreement-with-osaka-prefecture-to-promote-the-utilization-of-ai/>

[浜松市の取り組み](#)

<https://customers.microsoft.com/ja-jp/story/1468104978479874435-hamamatsu-city-government-azure-ja-japan>

### 参考文献

[Our commitment to the UN Sustainable Development Goals - Microsoft](#)

<https://www.microsoft.com/en-us/corporate-responsibility/un-sustainable-development-goals>

