

特集

AI時代の人材育成

「人間中心のAI社会原則」について

前 内閣府政策統括官（科学技術・イノベーション担当）付参事官
（現 大阪大学共創機構産学共創・渉外本部副本部長）

新田 隆夫



1. はじめに：AI社会原則の必要性

近年、人工知能（AI）に関する研究は、大量のデータへのアクセスと計算機能力の劇的な向上とともに、ディープラーニングに代表される人間の脳の情報処理の仕組みを模倣した手法を組み合わせることによって、飛躍的な進歩を遂げています。現在のAIは、画像や音声の認識、自然言語処理、自律システムの制御など広範な領域で人間と同等かそれを凌駕する能力を備え始めています。それによって、例えば、プロ棋士を打ち負かすAI囲碁ソフト、安心して車の運転を任せられる自動運転、流暢に通訳してくれる自動翻訳、熟練医の診察スキルと同等以上の医療画像診断など、AIは、様々な分野でその道のプロフェッショナルをしのぐレベルに達するビジネスやサービスを実現し提供しつつあります。

このように、AIは、人間の作業を効率化し、ものづくりの生産性を向上させ、単純作業、危険な作業から人間を解放し、交通事故ゼロの社会を実現してくれるなど、私たち人間社会に多大なる便益をもたらしてくれるものとの期待が高まっています。その一方で、応用される領域によっては人間をしのぐ能力を発揮するがゆえ、「AIに仕事を奪われるのではないかな?」、「人間がAIに操られるのではないかな?」、「個人データが漏洩することによって国家に監視されるのではないかな?」、「AIが人類に敵対するのではないかな?」など、将来のAIの発展に対して不安やおそれを抱く方もいるのではないのでしょうか。

社会全体がAIの便益を最大限に受けるためには、人々のAIに対するこれらの不安を解消しながら、AIのメリットを最大限に引き出す必要があります。そのためには、AIの社会実装を進めて

いく上で、社会全体で留意し、共通に理解しておくべき基本原則が必要です。

そのような観点から、政府においては、2018年5月より、「人間中心のAI社会原則検討会議」（議長：須藤修東京大学大学院教授）を開催し議論を重ね、その検討結果を踏まえた上で、最終的に、2019年3月に開催された第4回統合イノベーション戦略推進会議（議長：菅義偉内閣官房長官、議長代理：平井卓也内閣府特命担当大臣（科学技術政策））において、「人間中心のAI社会原則」を決定しました。これは、人々がAIを受容し社会全体でAIを使いこなしていく上で不可欠な政府全体としてのAI基本原則であるとともに、並行して検討してきた政府のAI戦略の中でも極めて重要な役割を担うものとして構成されるものです。

2. 人間中心のAI社会原則の概要

人間中心のAI社会原則は、AI-Readyな社会において尊重すべき3つの基本理念と、これを実現するための7つの基本原則により構成されます（図1、「人間中心のAI社会原則の全体構成」参照）。

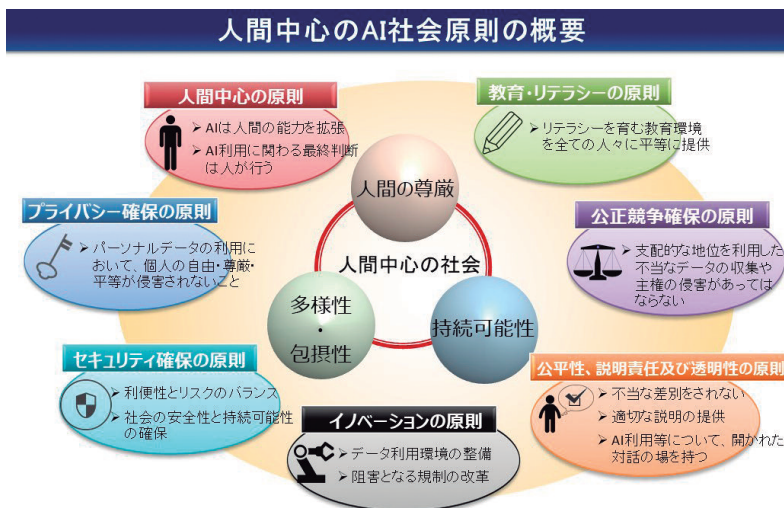


図1 人間中心のAI社会原則の全体構成

なお、「AI-Readyな社会」とは、私たちの社会全体が変革し、AIの恩恵を最大限に享受できる、又は必要な時にいつでもAIを導入しその恩恵を受けられる状態になることです。具体的には、個人のレベルで言えば、すべての人々が仕事や生活でAIを利用できるリテラシーを身に付けること、企業のレベルで言えば、AIの活用を前提とした経営戦略に基づいたビジネスを展開すること、イノベーション環境という意味では、あらゆる情報がAI解析可能なレベルでデジタル化、データ化され、AI開発やサービス提供ができる状態となることです。

(1) 基本理念

人間中心のAI社会原則においては、AIの社会実装により実現すべきAI-Readyな社会の基本理念として、「人間の尊厳」、「多様性・包摂性」、「持続可能性」の3つの価値を尊重すべきとしています。

「人間の尊厳」とは、単にAIを用いて社会、経済が効率性や利便性のみを追求したり、また、人間がAIに依存し過ぎるのではなく、むしろ人間がAIを便利な道具として使いこなすことによって、多様な能力や創造性を発揮し、やりがいのある仕事に従事できるような社会を実現すべきということです。

「多様性・包摂性」とは、男性／女性、若者／高齢者、健常者／障害者、日本人／外国人、都市住民／地方住民など、多様な背景や価値観、考え方を持った人々が活躍でき、幸せを追求できる社会を実現すべきということです。

「持続可能性」とは、地球規模の課題である気候変動、生物学的多様性の喪失、資源の逼迫などの地球規模のサステナビリティの脅威に対して、AIにより科学的・技術的蓄積を強化して、持続性ある社会をつくる必要があるということです。

(2) AI社会原則

我々は、AIを積極的に使いこなすことによって、AIから最大限の恩恵を受けられるAI-Readyな社会を創っていく必要があります。「人間中心のAI社会原則」では、そのために必要な、政府をはじめ産業界、アカデミア、ユーザを含めたすべてのステークホルダーが留意すべき7つの基本原則を定めています。以下にその概要を紹介します。

① 人間中心の原則

人々は、AIに過度に依存することで操られ

たりすることがないように、誰もがAIを正しく理解し使いこなせるようにならなければなりません。同時に、AIは、人々の労働の一部を代替するのではなく、人間にとって高度で便利な道具であるべきであり、これによって人間が持っている能力や創造性を拡大するものであるべきです。さらに、AIを利用するにあたっては、人が自らどのように利用するか判断と決定を行うことが求められます。

② 教育・リテラシーの原則

AIの利用者は、AIを正しく理解し、利用できる素養を身に付けていることが望まれます。また、AIに関わる政策決定者や経営者は、AIの複雑性や、意図的な悪用もありえることを理解した上で、AIの正確な理解と社会的に正しい利用ができる知識と倫理を持つ必要があります。さらに、AIの開発者側は、AIが社会においてどのように使われるかに関するビジネスモデル及び規範意識を含む社会科学や倫理等、人文科学に関する素養を習得していることが重要です。以上の観点から、教育・リテラシーを育む教育環境がすべての人に平等に提供される必要があります。

具体的には、人々の格差や弱者を生み出さないため、初等中等教育におけるリテラシー等の教育の機会や社会人、高齢者向けの学び直しの機会の提供などが求められます。

AIを活用するためのリテラシー教育やスキルとしては、誰でもAI、数理、データサイエンスの素養を身に付けられる教育システムが整備されることにより、すべての人が文理の境界を超えて学べる必要があります。また、リテラシー教育には、データにバイアスが含まれることなどAI・データの特性や、AI・データの持つ公平性・公正性、プライバシー保護、セキュリティ、AI技術の限界などを備えることも必要です。

③ プライバシー確保の原則

AI社会では、個人の行動データから政治的立場、経済状況、趣味・嗜好が推定できることがあります。このため、AI時代のパーソナルデータは、単なる個人情報という以上にその重要性や要配慮性に応じて慎重に扱われなければなりません。また、パーソナルデータが本人の望まない形で流通し、利用されることで、個人が不利益を受けてはなりません。

④ セキュリティ確保の原則

多くの社会システムがAIで自動化される時

代において、例えば政府機関や重要インフラ事業に関連する情報システムがサイバー攻撃を受けると、社会機能に深刻な悪影響を及ぼすおそれがあります。希少な事象や意図的な攻撃まで含めて常に完璧に対応することすることは不可能でしょうが、AI利活用による利益とセキュリティリスクのバランスに留意し、社会全体として安全性と持続可能性を維持する必要があります。

⑤ 公正競争確保の原則

特定の国や企業にAIに関する資源が集中した場合において、その支配的な地位を利用した不当なデータの収集や不公正な競争が行われる社会であってはなりません。

⑥ 公平性・説明責任及び透明性の原則

AIの利用者が人種、性別、国籍、年齢、政治的信条、宗教等のバックグラウンドを理由にして不当な差別を受けてはなりません（公平性）。

AIは、従来のプログラムと比較してブラックボックス性が高いですが、（必ずしもAIによる判断根拠を技術的に示さなければならないということだけでなく）AIを利用している事実、データの取得方法や使用方法、AIの動作結果が適切なものとするための仕組みなどについて説明しなければなりません（説明責任）。

AIの利用・採用・運用について、開かれた対話の場が適切に持たなければなりません（透明性）。

また、AIとそれを支えるデータについて、信頼性が担保されなければなりません（トラスト）。

⑦ イノベーションの原則

AI技術の健全な発展のため、プライバシーやセキュリティの確保を前提に、あらゆる分野のデータが独占されことなく、国境を越えて有効利用できる環境が整備される必要があります。また、AIの研究促進のため、国際的な連携を促進しAIを加速するコンピュータ資源や高速ネットワークが共有して活用されるような研究開発環境が整備される必要があります。

一方、政府は、AI技術の社会実装を促進するため、あらゆる分野で阻害要因となっている規制の改革等を進めなければなりません。

3. 世界におけるAI倫理に関する検討状況

AI・データは、国境を越えてつながるものであることから、AI倫理、原則は日本国内のみで策定、合意するだけでは不十分であり、国際的な共通理解を形成し、基本的なルールについて合意を目指すことが重要です。

世界各国がAIの研究開発・社会実装でしのぎを削る中、各国政府においても、AIによる経済成長を念頭に置き、国家戦略の策定を進めています。その中で、各国政府や国際機関においても、AIの社会実装を推進する上で、AIに対する人々の不安を払拭するためのAI倫理、原則の議論が進展しています（図2、「AI倫理に関する各国の動き」参照）。特に、EUが2019年4月に公表した「信頼できるAIのための倫理ガイドライン」の内容は、日本の考え方にも近く、また、OECDにおける議論については、傘下に設置された専門家会合の議論に日本の有識者も参加してきたこともあり、2019年5月に開催されたOECD閣僚理事会では、日本の考え方も反映された理事会勧告がとりまとめられました。

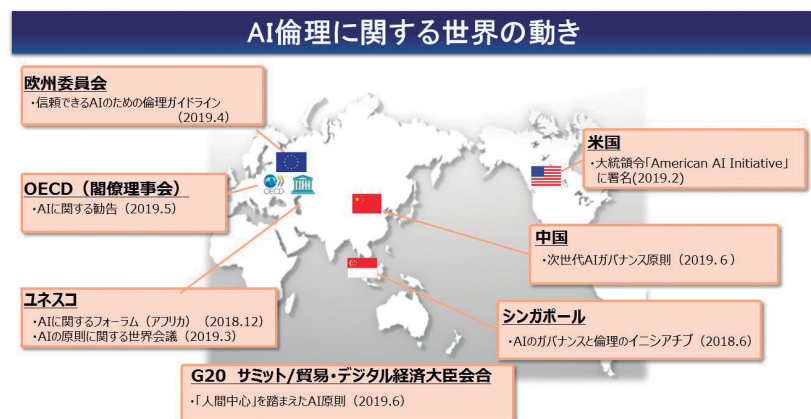


図2 AI倫理に関する世界の動き

2019年6月29～30日に開催されたG20大阪サミットの成果文書として、「G20大阪首脳宣言」が採択されました。同宣言において、OECD勧告を引用しつつ、G20加盟国間で初めてAI原則についてコンセンサスを形成できたことは、画期的と言えます。また、G20 AI原則では、「AIへの人間中心のアプローチにコミットし」との記述をはじめ、多くの点で我が国の「人間中心のAI社会原則」との共通項を見出すことができます（次ページ図3、「G20 AI原則の概要」参照）。

G20 大阪首脳宣言 (2019年6月28～29日)

強固な世界経済の成長の醸成

イノベーション:信頼性のある自由なデータ流通

12. ...人工知能(AI)の責任ある開発及び活用は、SDGsを推進し、持続可能で包摂的な社会を実現するための原動力となり得る。AI技術への人々の信頼と信用を醸成し、その潜在能力を十分に引き出すために、我々は、AIへの人間中心のアプローチにコミットし、経済協力開発機構(OECD)AI勧告から引用された拘束力を有さないG20・AI原則を歓迎する。...

G20 AI原則

1. 信頼できるAIのための責任あるスチュワードシップに関する原則

- 1.1. 包摂的な成長、持続可能な開発及び幸福
- 1.2. 人間中心の価値観及び公平性
- 1.3. 透明性及び説明可能性
- 1.4. 頑健性、セキュリティ及び安全性
- 1.5. アカウンタビリティ

2. 信頼できるAIのための国内政策と国際協力

図3 G20 AI原則の概要

4. 政府のAI戦略とAI社会原則との関係

AIの研究開発と社会実装は、米国、中国、欧州をはじめとして世界各国で熾烈な競争が繰り広げられており、これを国家的に推進するためのAI国家戦略についても、世界各国で検討が進められています。

一方、人口減少下の我が国にとっても、AIの社会実装は、経済成長を実現するためのエンジンとなり得るものですが、米国、中国などの後塵を拝していることから、早急なAI国家戦略の策定し、AI-Readyな社会を実現するための社会変革の方向性の明確化が求められてきました。

このような背景から、内閣府が政府全体の司令塔となり、2019年5月に開催された第5回統合イノベーション戦略推進会議において、政府としての包括的なAI戦略となる「AI戦略2019」をとりまとめたところです。

AI戦略2019においては、①教育改革、②研究開発、③社会実装、④データ、⑤AI社会原則を5本柱としており、AI社会原則は、AI戦略の一部を構成するものではあるものの、AI戦略に先立って検討が行われた経緯があります。その理由は、「人間中心」や「教育・リテラシー」など、まずAI社会原則において、目指すべき未来の社会ビジョンを明確にすることが戦略策定の前提となると考えられたためです。

AI戦略においては、AI人材の確保・育成が遅れているという我が国の産業界において顕在化する課題に対応するため、「教育改革」をAI戦略のベースであり、かつ最重要の柱として位置づけています。具体的には、未来の子供たちの誰もがデジ

タル時代の「読み・書き・そろばん」であるAIリテラシーを身に付けるという大きな目標を実現するため、小・中・高校や大学教育、さらには、リカレント教育に至るまで、教育システム全般にわたる教育改革を戦略として盛り込んでいます。同時に、世界に通用するグローバルトップのAI人材を育成することを目標としています(それぞれのレベルごとの具体的数値目標や人材育成施策については、図4「AI教育改革の概要」を参照のこと)。

教育改革に向けた主な取り組み

デジタル社会の「読み・書き・そろばん」である「教理・データサイエンス・AI」の基礎などの必要能力を全ての国民が育み、あらゆる分野で人材が活躍



図4 AI教育改革の概要

5. おわりに

AIの社会へのインパクトが多大となっていることから、シンギュラリティの言葉に代表されるように、将来のAIに対して不安や恐れを抱く人もいます。しかしながら、人々が不安や疑義を抱くAIは、社会実装ができません。AIの社会実装を進めるためには、人々がAIを受容し、社会全体でAIを使いこなせるようにするためのAI社会原則が必要となります。

我が国としては、「人間中心」の真の意味を理解した上で、実現すべき未来社会に向かうためのAI社会実装について検討することが必要です。そのために我が国が進むべき方向性としては、「人間中心」に配慮したバランスある研究開発、社会実装を推進することによって、持続性の高い「健全なAI社会」を実現することで、世界に先行する戦略を指向すべきではないかと考えます。