

# 人間中心の AI 社会原則(案)

## 目次

1 はじめに	1
2 基本理念	3
3 Society 5.0 実現に必要な社会変革「AI-Ready な社会」	4
4 人間中心の AI 社会原則	6
4.1 AI 社会原則	6
4.2 AI 開発利用原則	9
5 おわりに	10

本文書は人間中心のAI社会原則検討会議がとりまとめる人間中心のAI社会原則の草案である。今後、国内外から広く意見を募った上で2019年3月に本原則を策定する予定である。

## 1 はじめに

現代社会は地球環境問題、格差の拡大、資源枯渇等、人類の存続に関わる問題に直面している。AIはこれらの問題の解を導き、SDGs(Sustainable Development Goals)で掲げられている17目標を解決し、持続可能な世界の構築するための鍵となる技術と考えられている。

中でも我が国は、少子高齢化、人手不足、過疎化、財政支出増大等に直面し、社会課題のいわば先進国となっている。

我が国は、AIの活用により、経済発展と共に社会課題を解決するSociety5.0<sup>1</sup>の実現を通して、日本の社会と経済の活性化を実現し、国際的にも魅力ある社会を目指すと共に、地球規模でのSDGsへの貢献も果たしていく。

多くの科学技術と同様、AIも社会に多大なる便益をもたらす一方で、その社会への影響力が大きいがゆえに、適切な開発と社会実装が求められる。AIを有効に活用して社会に便益もたらしつつ、ネガティブな側面を事前に回避又は低減するためには、我々はAIに関わる技術自体の研究開発を進めると共に、人、社会システム、産業構造、イノベーションシステム、ガバナンス等、あらゆる面で社会をリデザインし、AIを有効かつ安全に利用できる社会を構築すること、すなわち「AI-Readyな社会」への変革を推進する必要がある。

この文書において、我々は、「AI」という用語について特定の技術を指すのではなく、広く「高度に複雑な情報システム一般」を指すもの<sup>2</sup>として捉え、社会に与える影響を議論した上で、AI社会原則の一つの在り方を提示し、AIの研究開発や社会実装において考慮すべき問題を列挙する。来るべきSociety 5.0がより良いものとなるためには、関係ステークホルダーが対話しながら協力していくことが必要不可欠である。本文書の全体構成を図1に示す。

<sup>1</sup> Society 5.0とは、情報社会(Society 4.0)に続く、我が国が目指すべき未来社会の姿である。Society 5.0で実現する社会とは、AI、IoT(Internet of Things)、ロボット等先端技術が社会に実装され、今までにない新たな価値を生み出し、多様な人々がそれぞれの多様な幸せを尊重し合い、実現でき、持続可能な人間中心の社会である。

<sup>2</sup> 「AI(Artificial Intelligence, 人工知能)」の定義については研究者によっても様々な考え方があり、現在のところ明確な定義はない。また一般に「AI」と呼ばれる様々な技術が単体で使われることは少なく、情報システムの一部として組み込まれて使われることが一般的である。本文書では、高度に複雑な情報システムに一般的にこの原則が考慮されるべきとの考えに立ち、本原則における「AI」は、「高度に複雑な情報システム一般」を意味する。

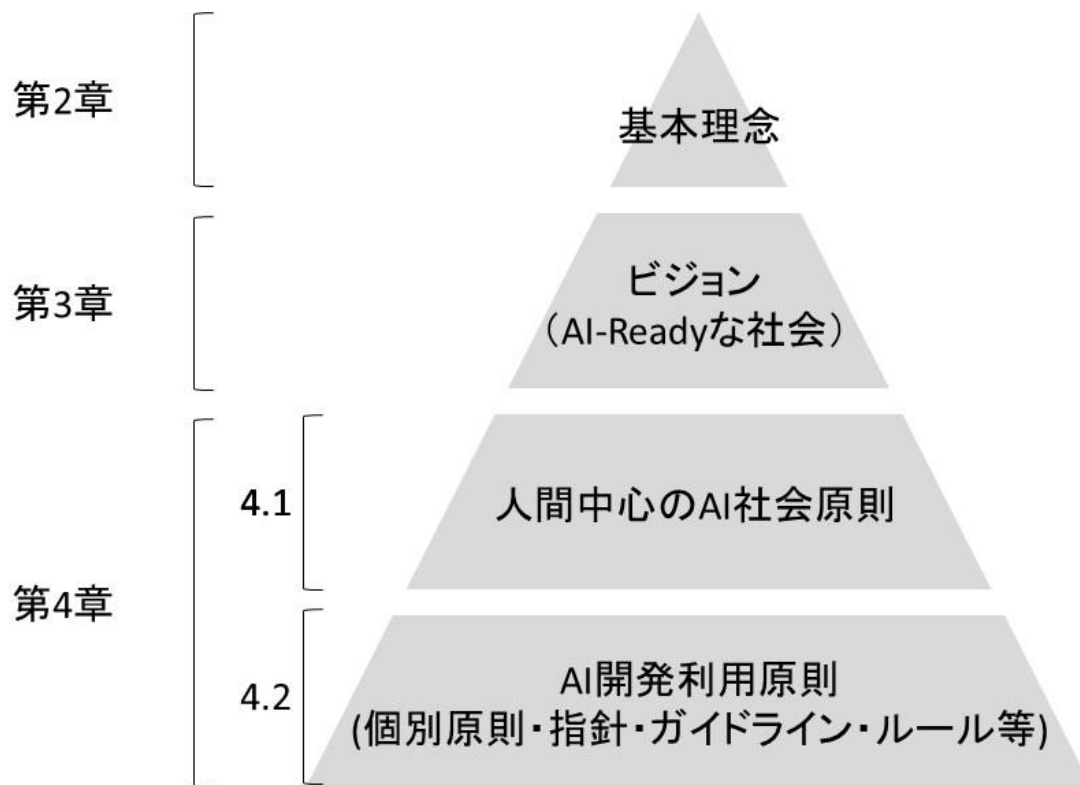


図1:本文書の全体構成

## 2 基本理念

AI は、Society 5.0 の実現に大きく貢献することが期待される。我々は、単に AI の活用による効率性や利便性から得られる利益が人々や社会に還元されることにとどまらず、AI を人類の公共財として活用し、社会の在り方の質的变化や真のイノベーションを通じて、SDGs など指摘される地球規模の持続可能性へとつなげることが重要と考える。

我々は、以下の3つの価値を理念として尊重し、その実現を追求する社会を構築していくべきと考える。

### (1) 人間の尊厳が尊重される社会 (Dignity)

我々は、AI を利活用して効率性や利便性を追求するあまり、人間が AI に過度に依存したり、人間の行動をコントロールすることに AI が利用される社会を構築するのではなく、人間が AI を道具として使いこなすことによって、人間の様々な能力をさらに発揮することを可能とし、より大きな創造性を発揮したり、やりがいのある仕事に従事したりすることで、物質的にも精神的にも豊かな生活を送ることができるような、人間の尊厳が尊重される社会を構築する必要がある。

### (2) 多様な背景を持つ人々が多様な幸せを追求できる社会 (Diversity & Inclusion)

多様な背景と価値観、考え方を持つ人々が多様な幸せを追求し、それらを柔軟に包摂した上で新たな価値を創造できる社会は、現代における一つの理想であり、大きなチャレンジである。AI という強力な技術は、この理想に我々を近づける一つの有力な道具となりえる。我々は AI の適切な開発と展開によって、このように社会のありかたを変革していく必要がある。

### (3) 持続性ある社会 (Sustainability)

我々は、AI の活用によりビジネスやソリューションを次々と生み、社会の格差を解消し、地球規模の環境問題や気候変動などにも対応が可能な持続性のある社会を構築する方向へ展開させる必要がある。科学・技術立国としての我が国は、その科学的・技術的蓄積を AI によって強化し、そのような社会を作ることに貢献する責務がある。

### 3 Society 5.0 実現に必要な社会変革「AI-Ready な社会」

Society 5.0 の実現への貢献が期待される技術には、IoT、ロボティクス、超高速広帯域通信網等と並んで AI がある。AI を用いて複雑な処理を機械にある程度任せられることが可能になっても、「何のために AI を用いるのか」という目的設定は、人間が行う必要がある。AI は、社会を良くするために使うことも可能であれば、望ましくない目的達成のために使われたり、無自覚に不適切に使われたりすることもありうる。そのため、我々は、「何のために AI を用いるのか」に答えられるような「人」、「社会システム」、「産業構造」、「イノベーションシステム」、「ガバナンス」の在り方について、技術の進展との相互作用に留意しながら考える必要がある。

#### (1) 「人」

AI が社会の隅々に浸透してくることに対応する「AI-Ready な社会」において、人間がどのように対応していくかが AI を十分に活用できる社会の実現の鍵となる。そのために人間に期待される能力及び役割は、以下のようなものになる。

- A) AI の長所・短所をよく理解している。特に AI の情報リソースとなるデータ、アルゴリズム、又はその双方にはバイアスが含まれること及びそれらを望ましくない目的のために利用する者がいることを認識する能力を人々が持つことが重要である。なお、データのバイアスには、主として統計的バイアス、社会の様態によって生じるバイアス及び AI 利用者の悪意によるバイアスの 3 種類があることを認識していることが望まれる。
- B) AI の利用によって、多くの人々が創造性や生産性の高い労働に従事できる環境が実現できることが望ましい。そのためには、出自、文化、趣向等の観点で、多様な人々が各々の目指す多様な夢やアイデアを AI の支援によって実現する能力を獲得できることが期待される。このことを実現するための教育システム及びそれらの達成に資する社会制度が実現されなければならない。
- C) データや AI の基礎教養から実装及び設計等の応用力を、幅広い分野の横断的、複合的及び融合的な枠組みで身につけた人材が十分に存在する。そのような人材は、社会のあらゆる活動の原動力となり、かつその人々の能力が AI を活用した生活環境の構成に寄与することが期待される。このような生活環境の整備によって、多くの人々がより豊かで充実した人生を送れるような社会制度が実現されている。

#### (2) 「社会システム」

AI を利用することで、個々のサービス・ソリューションの進化を促進し、効率化・個別化による多様なメリットを生み出すことが期待される。この変化から生じるメリットを社会の側において十分に受け止めるため、医療、金融、保険、交通、エネルギー等の社会システム全体が、AI の進化に応じて柔軟に変化し、対応できるようなものになっている必要がある。これには、社会的に受け入れられた既存の目的(利便性の向上や単純労働からの解放など)に照らした単純な効率化だけではなく、目的自体の多様

化・流動化によって生まれる新たな価値の実現や、AI の進化によってもたらされる可能性のある負の側面(不平等や格差の拡大、社会的排除等)への対応が含まれる。

そのためには、我々は、それぞれの社会システムのソフト面・ハード面の双方において、拡張性や相互接続性、発展的な秩序形成への仕組み等を備えた柔軟なアーキテクチャ設計を実現する必要がある。さらに、我々は、特に相互接続性・連携性を保証するために、様々な社会システムに共通のデータ利活用基盤を整備する必要がある。

### (3)「産業構造」

多様な人々が多様な夢やアイデアを実現できるよう、労働、雇用環境や創業環境が柔軟で国際的に開かれたものになっていることが必要である。そのために企業は公正な競争を行い、柔軟な働き方を促進していること、また人間の創造力が産業を通じても発揮され続けており、スタートアップへの投資が促進されていることが求められる。

### (4)「イノベーションシステム」

大学・研究機関・企業、さらに一般の人々に至るまで、分野や立場を超えて AI の研究開発、利活用及び評価に参加し、互いに刺激し合いながら、イノベーションが次々に生まれる環境ができていくことが必要である。

そのためには、リアル空間も含めたあらゆるデータが新鮮かつ安全に AI 解析可能なレベルで利用可能であり、かつ、誰もが安心してデータを提供でき、提供したデータから便益を得られる環境ができていくことが求められる。

研究開発者に加えユーザも含め、安心して AI を研究開発し利活用できる環境が整い、研究開発と利活用のサイクルが迅速に回ることによって、望ましい発展が加速していることが望ましい。また、AI の利活用によって、新たな発想やさらなる可能性が生まれ、イノベーションの地平が格段に広がっていることが求められる。

### (5)「ガバナンス」

社会情勢の変化や技術の進展に伴い、上記に挙げた「人」、「社会システム」、「産業構造」、「イノベーションシステム」で議論されるべき内容や目的設定は、常に更新し続ける必要がある。

そのため、政府、企業、大学、研究機関、一般の人々等、多様なステークホルダーが協働してルール、制度、標準化、行動規範等のガバナンスについて問題を設定し、影響を評価し、意思決定を行うと共に実装できる体制が整っていることが必要である。また、多様なステークホルダーだけではなく、社会的に声の挙げにくい人たちの声も拾い上げ、常に最先端の社会的、技術的な課題に取り組む体制を構築できていることが求められる。こうしたガバナンスの実装に際しては、法律によるのみならず、技術的手段を含む企業の自主的な取組によるなど、柔軟かつ実効的な方法がとられていることが求められる。国内におけるガバナンスに加えて、国境を越える問題に対処するための国際協力体制が整っていることが求められる。

## 4 人間中心のAI社会原則

我々は、「AI-Ready な社会」を実現し、AI の適切で積極的な社会実装を推進するためには、各ステークホルダーが留意すべき基本原則を定めることが重要と考える。

我々は、この基本原則について、AI が社会に受け入れられ適正に利用されるため、社会(特に、国などの立法・行政機関)が留意すべき「AI 社会原則」と、AI の研究開発と社会実装に従事する開発・事業者側が留意すべき「AI 開発利用原則」に体系化する。第2章に掲げた3つの基本理念を備えた社会を実現するために必要となる AI 社会原則並びに開発者及び事業者が考慮すべき AI 開発利用原則は、以下のとおりである。

### 4.1 AI 社会原則

AI 社会原則は、「AI-Ready な社会」において、国や自治体をはじめとする我が国社会全体、さらには多国間の枠組みで実現されるべき社会的枠組みに関する原則である。

#### (1) 人間中心の原則

AI の利用は、憲法及び国際的な規範の保障する基本的人権を侵すものであってはならない。

AI は、人々の能力を拡張し、多様な人々の多様な幸せの追求を可能とするために開発され、社会に展開され、活用されるべきである。AI が活用される社会において、人々がAIに過度に依存したり、AIを悪用して人の意思決定を操作したりすることのないよう、我々は、リテラシー教育や適正な利用の促進などのための適切な仕組みを導入することが望ましい。

- AI は、人間の労働の一部を代替するのみならず、高度な道具として人間を補助することにより、人間の能力や創造性を拡大することができる。
- AIの利用にあたっては、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る。AI の利用がもたらす結果については、問題の特性に応じて、AI の開発・提供・利用に関わった種々のステークホルダーが適切に分担して責任を負うべきである。
- 各ステークホルダーは、AI の普及の過程で、いわゆる「情報弱者」や「技術弱者」を生じさせず、AI の恩恵をすべての人が享受できるよう、使いやすいシステムの実現に配慮すべきである。

#### (2) 教育・リテラシーの原則

AIを前提とした社会において、我々は、人々の間に格差や分断が生じたり、弱者が生まれたりすることは望まない。したがって、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勘案して、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っていなければならない。AIの利用者側は、AIが従来ツールよりはるかに複雑な動きをするため、その概要を理解し、正しく利用できる素養を身につけていることが望まれる。一方、AIの開発者側は、AI

技術の基礎を習得していることが当然必要であるが、それに加えて、社会で役立つ AI の開発の観点から、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になる。

このような観点から、我々は、以下のような原則に沿う教育・リテラシーを育む教育環境が全ての人に平等に提供されなければならないと考える。

- 人々の格差や弱者を生み出さないために、幼児教育や初等中等教育において幅広くリテラシー等の教育の機会が提供されるほか、社会人や高齢者の学び直しの機会の提供が求められる。
- 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つけられる教育システムとなっているべきであり、全ての人々が文理の境界を超えて学ぶ必要がある。リテラシー教育には、データにバイアスが含まれることや使い方によってはバイアスを生じさせる可能性があることなどの AI・データの特性があること、AI の持つ公平性・公正性、プライバシー保護に関わる課題があることを認識できる内容を備えることも必要である。
- AI が広く浸透した社会において、教育環境は、一方的かつ均一的に教える教育の在り方から、個々人の持つ関心や力を活かす在り方へと変化すると考えられる。そのため、社会は、これまでの教育環境における成功体験に拘ることなく、常に最適な形へと柔軟に変化し続ける意識を全体として共有する。教育において、落伍者を出さないためのインタラクティブな教育環境や学ぶもの同士が連携できる環境が AI を活用して構築されることが望ましい。
- このような教育環境の整備に向けて、行政や学校(教員)に負担を押し付けるのではなく、民間企業や市民も主体性をもって取り組んでいくことが望ましい。

### (3) プライバシー確保の原則

AI を前提とした社会においては、個人の行動などに関するデータから、政治的立場、経済状況、趣味・嗜好等が高精度で推定できることがある。これは、単なる個人情報を超え、以上の慎重さが求められる場合があることを意味する。パーソナルデータが本人の望まない形で流通したり、利用されたりすることによって、個人が不利益を受けることのないよう、各ステークホルダーは、以下の考え方に基づいて、パーソナルデータを扱わなければならない。

- パーソナルデータを利用した AI 及びその AI を活用したサービス・ソリューションにおいては、政府における利用を含め、個人の自由、尊厳、平等が侵害されないようにすべきである。
- パーソナルデータを利用する AI は、当該データのプライバシーにかかわる部分については、正確性・正当性の確保及び本人が実質的な関与ができる仕組みを持つべきである。これによって、AI の利用に際し、人々が安心してパーソナルデータを提供し、提供したデータから有効に便益を得られることになる。



- パーソナルデータは、その重要性・要配慮性に応じて適切な保護がなされなければならない。パーソナルデータには、それが不当に利用されることによって、個人の権利・利益が大きく影響を受ける可能性が高いもの（典型的には思想信条・病歴・犯歴等）から、社会生活のなかで半ば公知となっているものまで多様なものが含まれていることから、その利活用と保護のバランスについては、文化的背景や社会の共通理解をもとにきめ細やかに検討される必要がある。

#### (4) セキュリティ確保の原則

AI を積極的に利用することで多くの社会システムが自動化され、安全性が向上する。一方、少なくとも現在想定できる技術の範囲では、希少事象や意図的な攻撃に対して AI が常に適切に対応することは不可能であり、セキュリティに対する新たなリスクも生じる。社会は、常にベネフィットとリスクのバランスに留意し、全体として社会の安全性及び持続可能性が向上するように務めなければならない。

- 社会は、AI の利用におけるリスクの正しい評価やそのリスクを低減するための研究等、AI に関わる層の厚い研究開発（当面の対策から、深い本質的な理解まで）を推進し、サイバーセキュリティの確保を含むリスク管理のための取組を進めなければならない。
- 社会は、常に AI の利用における持続可能性に留意すべきである。社会は、特に、単一あるいは少数の特定 AI に一義的に依存してはならない。

#### (5) 公正競争確保の原則

新たなビジネス、サービスを創出し、持続的な経済成長の維持と社会課題の解決策が提示されるよう、公正な競争環境が維持されなければならない。

- 特定の国に AI に関する資源が集中することにより、その支配的な地位を利用した不当なデータの収集や主権の侵害が行われる社会であってはならない。
- 特定の企業に AI に関する資源が集中することにより、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らない。
- AI の利用によって、富や社会に対する影響力が一部のステークホルダーに不当過剰に偏る社会であってはならない。

#### (6) 公平性、説明責任及び透明性の原則

「AI-Ready な社会」においては、AI の利用によって、人々が、その人の持つ背景によって不当な差別を受けたり、人間の尊厳に照らして不当な扱いを受けたりすることがないように、公平性及び透明性のある意思決定とその結果に対する説明責任（アカウンタビリティ）が適切に確保されると共に、技術に対する信頼性（Trust）が担保される必要がある。

- AI の設計思想の下において、人々がその人種、性別、国籍、年齢、政治的信

念、宗教等の多様なバックグラウンドを理由に不当な差別をされることなく、全ての人が公平に扱われなければならない。

- AI を利用しているという事実、AI に利用されるデータの取得方法や使用方法、AI の動作結果の適切性を担保する仕組みなど、状況に応じた適切な説明が得られなければならない。
- 人々が AI の提案を理解して判断するために、AI の利用・採用・運用について、必要に応じて開かれた対話の場が適切に持たれなければならない。
- 上記の観点を担保し、AI を安心して社会で利活用するため、AI とそれを支えるデータの信頼性(Trust)を確保する仕組みが構築されなければならない。

#### (7) イノベーションの原則

- Society 5.0 を実現し、AI の発展によって、人も併せて進化していくような継続的なイノベーションを目指すため、国境や産学官民、人種、性別、国籍、年齢、政治的信念、宗教等の垣根を越えて、幅広い知識、視点、発想等に基づき、人材・研究の両面から、徹底的な国際化・多様化と産学官民連携を推進するべきである。
- 大学・研究機関・企業の間に対等な協業・連携や柔軟な人材の移動を促さなければならない。
- AI を効率的かつ安心して社会実装するため、AI に係る品質や信頼性の確認に係る手法、AI で活用されるデータの効率的な収集・整備手法、AI の開発・テスト・運用の方法論等の AI 工学の確立が推進されなければならない。
- 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としつつ、あらゆる分野のデータが独占されることなく、国境を越えて相互利用できる環境が整備される必要がある。また、AI の研究促進のため、国際的な連携を促進し人工知能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べきである。
- 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らない。

### 4. 2. AI 開発利用原則

我々は、開発者及び事業者において、基本理念及び上記の AI 社会原則を踏まえた AI 開発利用原則を定め、遵守するべきと考える。

AI 開発利用原則については、現在、多くの国、団体、企業等において議論されていることから、我々は早急にオープンな議論を通じて国際的なコンセンサスを醸成し、非規制的で非拘束的な枠組みとして国際的に共有されることが重要であると考えます。

## 5 おわりに

「AI-Ready な社会」を世界に先駆けて構築していくため、我が国は、本原則を政府、関係企業、団体等で共有し、政策等に反映させるべきである。

また、国際的な議論の場において、我が国は、本原則を世界各国と共有した上で、国際的な議論のリーダーシップをとり、コンセンサスの形成を目指すべきであり、それによって SDGs の実現を支える Society5.0 の社会像を世界に示し、国際社会の協調的かつ創造的な新たな発展に寄与すべきである。

なお、本原則は、今後、AI 関連技術の進展、社会の変化、世界の情勢等に応じて、今後柔軟に進化・発展させるものである。