

# AI倫理のための

## ローマからの呼びかけ

2020年2月28日、ローマ

### 導入

「人工知能」（AI）は人間の生活に大きな変化をもたらし、今後もそれは続いていく。社会的共存の促進や個人の幸福度の向上、および人間の能力の成長と作業の効率化という点において、AIは膨大な可能性を有している。しかし、それによって期待される結果は、決して保証されたものではない。昨今私たちが目の当たりにしているAIによる変化は、単なる量的なものではない。むしろ、AIによる変化は、仕事の仕方や現実と人間の本質の捉え方に影響を与え、精神的および対人関係における私たちの習慣にも影響をもたらすため、とりわけ質的なものと言える。新たな科学技術は、人類一人ひとりの生来的な尊厳とすべての自然環境を尊重し、また社会的に最も弱い立場にいる人々のニーズを考慮しながら、「人類社会」（「世界人権宣言」前文）全体に対して真に貢献できているかという基準に基づいて研究・開発されなければならない。そして、その目的は、誰一人排除しないようにすることだけでなく、アルゴリズムの調整によって脅かされる可能性のある自由の領域を拡大することにある。

デジタル変容（DX）による革新的かつ複雑な課題に鑑みると、すべてのステークホルダー（利害関係者）が協働し、AIの影響を受けるすべてのニーズを提示することが不可欠である。本文書は、共通の理解を深め、共有できる言語や解決策を模索するための一歩である。これに基づいて、技術革新の設計から流通および使用に至るまでのプロセス全体を考慮した責任を認識し、受け入れることができる。これにより、さまざまな実践的場面において真の参画を促進することが可能となる。長期的な視点で見ると、AIに組み込むことができる価値観や原則が、デジタル倫理を規制し、私たちの行動の手引きとなり、人類と環境の利益を目的とした科学技術の使用を促進する枠組みの確立に寄与するであろう。

技術に焦点を当てるのではなく、むしろ、AIは互いに密接に関連し合っている人類と私たちの共通の家（地球）のために開発されるべきである、という考え方を今こそ確実なものにしなければならない。言い換えれば、人間と自然がデジタル革新の開発の中心に存在し、理性的な行為者のように振る舞うが決して人間的ではないテクノロジーに次第に取って代わられるのではなく、むしろそれに助けをもらうという考え方である。今こそ、更に技術が進歩した未来に備える時である。その未来では、機械が人間の生活においてより重要な役割を果たすことになり、同時に、技術進歩によって人類の偉大さを確認し、その倫理観に依存し続けることになるだろう。

## 倫理

「すべての人間は、生まれながらにして自由であり、かつ、尊厳や権利とについて平等である。人間は、理性と良心とを授けられており、互いに同胞の精神をもって行動しなければならない」（「世界人権宣言」第1条）。AIシステムの開発および使用においても、人間の自由と尊厳という基本原則を遵守し、保証しなければならない。これは、いかなる個人も、アルゴリズムによる「人種、皮膚の色、性、言語、宗教、政治上その他の意見、国民的もしくは社会的出身、財産、門地その他の地位」（「世界人権宣言」第2条）を理由とした差別を受けることのないよう、個人の権利と自由を守ることによって果たされなければならない。

AIシステムは、人間とその生活環境を守り、支えるために考案、設計、実装されなければならない。この基本的な考え方は、集団および個人が可能な限り自分自身を十分に表現することができるような社会的および個人的な生活環境を創出するために適用されなければならない。

真の人類の進歩および地球への尊重と技術進歩を調和させるためには、以下の3つの条件を満たす必要がある。（1）誰一人差別することなく、すべての人に対して包摂的でなければならない。（2）人類全体の善とあらゆる人間の善を中心に据えなければならない。（3）そして最後に、生態系の複雑な現状に配慮をし、高度に持続可能な方法で、私たちの共通の家（地球）を保護するものでなければならない。これには、将来的に持続可能な食糧システムを確保するためのAIの使用も含む。さらに、各個人は、自身が機械と相互作用していることを認識していなければならない。

AIを基盤としたテクノロジーは、いかなる方法においても、人々、特に社会的に弱い立場にいる人々を搾取するために使用されてはならない。むしろ、人々の才能や能力を成長させ、可能性を広げることに役立て、地球を支えるために活用されなければならない。

## 教育

AIの技術革新によって世界を変容させることは、若い世代のために、そして、若い世代とともに未来を築いていくことを意味する。この取り組みによって、人文学や科学技術など、異なる分野をまたいだ特定のカリキュラムを開発し、若い世代を育成するという責任を担いながら、教育に対して貢献していかななければならない。これは、若者が受ける教育の質の向上に努めることと同義であり、すべての人にとってアクセス可能で、差別なく、平等な機会提供と対応がなされなければならない。教育への普遍的アクセスは、連帯と公平性の原則に基づいて確保されなければならない。

生涯学習へのアクセスは、高齢者にも保証されなければならない。この社会的なデジタル化および技術変化の時代において、オフライン・サービスへのアクセスも提供

されなければならない。さらに、これらの科学技術は、障害を持つ人々の学習および更なる自立支援において、非常に有用であることを証明し得る。したがって、包摂的教育というのは、移動制限のある方のためのリモートワークや認知障害を持つ人に対する技術支援など、AIを活用してすべての人に社会参加のための支援と機会を平等に提供することを意味する。

AIが社会、仕事、教育に及ぼす変容の影響により、「誰一人取り残さない」という教育の目標を実現するため、学校カリキュラムの徹底的な見直しが不可欠となっている。教育分野では、個々の成果向上を可能にする客観的かつ高い基準を確立するために改革が必要となる。これらの基準は、デジタル・スキルの発展に限定されるのではなく、むしろ、個々人が自身の能力を十分に発揮し、たとえ個人的な利益を得られずとも、地域社会の善に尽くすことにフォーカスされるべきである。

将来の社会を設計・構想する際、AIは、若い世代の個人および社会の生活に良い影響を与えるような、社会志向的、かつ創造的でつながりがあり、生産的で責任ある形で使用されなければならない。また、AIの社会的および倫理的な影響は、AIの教育活動の中心でなければならない。

この教育の主な目的は、社会的包摂および個人の尊重という観点から、AIがもたらす機会や潜在的な重要課題に対する意識啓発でなければならない。

## 権利

人類と地球のためのAIの開発は、特に社会的に立場が弱く恵まれない人々と自然環境を保護する規制および原則に反映されなければならない。すべての利害関係者の倫理的貢献は、極めて重要な出発点である。この未来を現実のものにするため、その過程を支援し、整理し、導く上で、価値観、原則、時には法的規制が不可欠である。

人類社会と地球に利益をもたらす、国際平和を構築および維持するためのツールとして機能するAIシステムを開発し、実装するためには、強固なデジタルセキュリティ対策とともにAIの開発を進めなければならない。

AIが人類と地球の善のためのツールとして機能するためには、デジタル時代における人権保護の問題を社会的な議論の中心に据えるべきである。AIによる新たな自動化機能やアルゴリズムによって、より強い責任感を醸成する必要性を問う時が来た。特に、「説明義務」について検討することが重要となる。つまり、AIを基盤としたアルゴリズムのプログラムの意思決定基準だけでなく、その目的と方針も理解可能なものにするよう配慮しなくてはならない。また、アルゴリズムの意思決定に使用される仕組みに関する情報を個人に開示できるようにしなければならない。これにより、透明性、追跡可能性、および責任感が向上し、コンピューターを利用した意思決定の過程がより正当なものになる。

特に、顔認証技術のような人権を侵害するリスクが高い先進技術については、透明性と倫理的原則の遵守を促進するための新しい規制が奨励されなければならない。

これらの目標を達成するためには、アルゴリズムの開発の初期段階から「アルゴリズムにおける倫理 (algor-ethical)」という考え方、つまり、倫理的アプローチによって設計されなければならない。信頼性のあるAIシステムを設計・構想する際には、政治的意思決定者、国連システム機関およびその他の政府間組織、研究者、学界、非政府組織の代表者の間で、システムに組み込むべき倫理的原則に関する共通認識を求める必要がある。このため、本文書の署名者は、「アルゴリズムにおける倫理」、つまり以下の原則によって定められるようなAIの倫理的 사용을促進するため、国家および国際レベルで協働する意思を表明するものとする。

1. **透明性**：原則として、AIシステムは、説明可能でなければならない。
2. **包摂性**：AIは、すべての人のニーズを考慮し、すべての人が利益を享受できるものでなければならない。また、すべての個人がAIを通して自分自身を表現し、成長していくための最高の条件を享受できるものでなければならない。
3. **責任**：AIを設計し、展開する者は、責任と透明性をもって開発を進めなければならない。
4. **公平性**：偏見に基づいた開発や行動をせず、公平性と人間の尊厳を守らなければならない。
5. **信頼性**：AIシステムは、正確に動作しなければならない。
6. **セキュリティとプライバシー**：AIシステムは、安全に動作し、利用者のプライバシーを尊重しなければならない。

以上の6つの原則は、適切な技術革新における基盤となる要素である。

(WCRP/RFP日本委員会訳)