

## AI 倫理学のローマからの要求\*

### Rome Call for AI Ethics

#### 序

「人工知能」(AI) は人間の生活に深遠な変化を引き起こしており、今後もそれを続けるだろう。AI は、社会的共存と人間の福利を改善し、人間の未発達能力を高め、より効率的かつ効果的に実行しうる多くの仕事を可能にしたりは容易にするようになる時、とてつもない可能性を提供する。しかし、これらの成果は決して確約されたものではない。目下進行中の諸々の変容は、単に量的なものにとどまらない。その変容は何にも増して質的である。なぜならそれは、仕事の仕方や、我々が現実と人間性それ自体を知覚する仕方に作用し、その作用は、我々の精神的および個人的関係の習慣に影響を及ぼしうるほどだからである。新しいテクノロジーは、そのメンバー一人一人の生来的な尊厳とすべての自然環境を尊重しつつ、また非常に傷つきやすい人々のニーズを考慮しつつ、「人類家族 (human family)」(「世界人権宣言」前文) 全体に真に奉仕することを確実なものにするような規準に従って、研究され、生み出されなければならない。目標は、誰も排除されないことを確実なものにするだけでなく、アルゴリズムの条件付け (conditioning) 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でもある。

デジタル変容によってもたらされた諸々の問いの革新的で複雑な性質にかんがみると、巻き込まれるすべてのステークホルダーが共に進むことが、また AI によって影響を受けるすべてのニーズが表明されることが不可欠である。この要求は、共通の理解をもって成長し、我々が共有できる言語と解決を探索するという展望をもって前進するためのステップである。我々はこれに基づいて、一連の実践的なシナリオにおける真の貢献 (commitment) を奨励しつつ、設計の段階から配分と使用に至るまでの、技術革新の全プロセスを考慮する責任を認め、受け入れることができる。我々が、長期間にわたって徐々に AI に浸透させることのできる諸々の価値や原則は、人間社会と環境に役立てるために我々の活動を導き、テクノロジーの使用を促進するデジタル倫理の参照点として、規制し行為する枠組みを我々が確立することを助けるだろう。

今、我々は以前よりも、テクノロジーに焦点を合わせるのではなく、むしろ人間社会とその環境の善、すなわち我々に共通の、共有された家 [地球] と、その家に解きがたく接続されている人間という居住者の善のために、AI が発展させられる見通しを保証しなければならない。それは、言い換えると、デジタル・イノベーションがいかに発展させられるかの核心に人間と自然があって、人間と自然が、合理的アクターのようにふるまうものの決して人間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、反対にテクノロジーによって支援される見通しである。人間の生活において、機械がより重要な役割を持つよりテクノロジーな未来、しかしまたテクノロジーな進歩が人類の光輝を肯定し、その倫理的高潔性に依存し続けることが明白な未来を、準備し始める時である。

#### 倫理学

すべての人間は自由に生まれ、尊厳と諸々の権利において平等である。彼らは理性と精神を付与されており、相互に交友の精神（spirit of fellowship）をもってふるまうべきである（「世界人権宣言」第 1 条参照）。この自由と尊厳という基本的条件は、AI システムの産出と使用に当たっても保護され、保証されなければならない。これは、個々人が彼らの「人種、肌の色、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」のゆえに、アルゴリズムによって差別されることのないよう、個々人の諸々の権利と自由を保護することによって果たされなければならない（「世界人権宣言」第 2 条）。

AI システムは、人間と人間が住む環境に奉仕し保護するために想像され、設計され、実行されなければならない。この基本的見通しは、グループと個人のメンバーの双方が、可能な場所で、自らを完全に表現するために奮闘することを可能にする生活条件（社会的および個人的双方の）を創造するために、現実的な貢献に移行されなければならない。

真の人類の進歩および地球の尊重と緊密に連携するテクノロジーの前進のために、三つの要請に応じなければならない。[1] テクノロジーの前進は、誰一人差別することなく、どの人間をも包摂しなければならない。[2] それは、その核心に、人類の善とあらゆる人間の善を据えなければならない。[3] 最後に、それは、我々の生態系の複雑な現実に留意しなければならない。また、高度に持続可能なアプローチをもって地球—我々の「共有された家」—をケアし保護する仕方で特徴づけられなければならない。それは、未来における持続可能な食料システムを保障するための AI の使用も含む。さらに、各個人は、彼ら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 を基盤とするテクノロジーは、いかなる仕方においても、人々、特に非常に傷つきやすい人々を搾取するために、決して使用されてはならない。その代わりに、それは、人々が彼らの才能（能力／可能性）を発展させることを助け、地球を支えるために使用されなければならない。

## 教育

AI のイノベーションを通して世界を変容させることは、より若い世代のために、そして彼らとともに、未来の構築を計画することを意味する。この計画は、人文学、自然科学、そしてテクノロジーにおける様々な学問分野にわたる特殊なカリキュラムを展開しつつ、そしてより若い世代を教育する責任を引き受けつつ、教育への貢献に反映されなければならない。この貢献は、若い人々が受ける教育の質を改善するために働くことを意味する。それは、全員にとってアクセス可能な、差別のない、機会と扱いの平等を提供しうる方法論によって果たされなければならない。教育への万人共通のアクセスは、連帯と公正の原則によって、達成されなければならない。

生涯学習へのアクセスは、デジタル技術の過渡期の間、オフライン・サービスにアクセスする機会を与えられなければならない年配者にも保証されなければならない。さらに、

これらの技術は、障害を持つ人々が学び、より独立することを手助けすることにおいて、極めて有益であることが証明されなければならない。それゆえ、包摂的な教育は、社会参加のための助力と機会を提供することによって援助し補完するために、AI を使用することをも意味する（たとえば、可動性が限定されている者のための遠隔作業、認識能力のない者のための技術的支援）。

AI によって社会、仕事、教育にもたらされた変容のインパクトは、教育のモットー「誰も取り残さない」を現実のものにするために、学校のカリキュラムを綿密に検討し直すことを不可欠にした。教育のセクターにおいては、諸々の個別的成果を改善しうる高い客観的スタンダードを確立するために、改革が必要とされている。これらのスタンダードは、デジタル・スキルの発展に限定されるべきではなく、各人格が完全に彼らの未発達能力を実現することを確実なものにし、たとえそこから個人的利益が得られないときでも、コミュニティの善のために働くことに焦点を合わせるべきである。

我々が明日の社会を設計し計画するとき、AI の使用は、社会的に方向づけられた、クリエイティブな、接続的な、生産的な、責任のある、そしてより若い世代の個人的および社会的生活にポジティブなインパクトを持つ可能性のある活動形態に従わなければならない。

この教育の主たるねらいは、AI によって提出される機会と、また今後生じうる批判的争点についての認識を、社会的包摂と個人の尊重というパースペクティブから喚起することではなければならない。

### 諸々の権利

人類と地球への奉仕における AI の発展は、人々—特に弱く恵まれない人々—と、自然環境を保護する諸々の規制と原則に反映されなければならない。巻き込まれるすべてのステークホルダーの倫理的貢献は、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に、諸々の権利、原則、そしてある場合にはこのプロセスを指示し、構築し、導くために、法規制は絶対不可欠である。

国際平和を樹立し維持するツールとしての機能を果たすと同時に、人類社会と地球のためになる AI システムを発展させ実施するために、AI の発展は、確固たるデジタル方式の安全基準と相互に協力してゆかなければならない。

AI が人類社会と地球の善のためのツールとして作用するために、我々は人権保護のトピックを、デジタル時代における公共討論の核心に据えなければならない。オートメーションとアルゴリズムの機能の新たな形態が、より強い責任の発達を必要とするかどうかを問う時期が到来している。特に「説明義務 (duty of explanation)」の何らかの形態を考慮することが不可欠であろう。我々は理解可能な AI を基盤とするアルゴリズムのエージェント〔代行者〕の意思決定基準だけでなく、彼らの目的と目標の策定についても考えな

ればならない。これらのデバイス〔装置〕は、決定を下すために使用されたアルゴリズムの背後のロジックについて、個々人に情報を提供することができなければならない。これは、コンピュータの補助を受けた意思決定プロセスをより有効にする透明性、トレーサビリティ〔追跡可能性〕、および責任を増すことになる。

規制の新しい形態は、特に顔認識のような、人権にインパクトを与える、より高いリスクを持つ先進技術のために、諸々の倫理原則によって、透明性とコンプライアンスを促進することを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、我々は各々のアルゴリズムの発展のまさに最初から、「アルゴレシカルな (algor-ethical)」ヴィジョン、すなわち、倫理学のアプローチによって計画的に設計しなければならない。我々が信頼しうる AI システムを設計し計画することは、政治的意思決定者、国連関係機関や他の政府間組織、研究者、学術界および非政府組織の代表者の間でこれらのテクノロジーに組み込まれるべき諸々の倫理原則に関してコンセンサスを探求することを含む。このような理由から、この要求の後援者たちは、「アルゴレシックス (algor-ethics)」、すなわち以下の原則によって定められるような、AI の倫理的使用を促進するために、この脈絡において、また国内および国際レベルで、ともに前進する希望を表明する。

1. **透明性 (Transparency)** : 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 (Inclusion)** : 誰もが利益を得ることができ、すべての個人が自らを表現し、発展させるための可能な最善の条件を提供されるよう、すべての人間のニーズが考慮されなければならない。
3. **責任 (Responsibility)** : AI の使用を設計し、展開する者は、責任とトレーサビリティをもって前進しなければならない。
4. **公明正大 (Impartiality)** : バイアス〔偏見〕に従って創造または行為しない。かくして公正と人間の尊厳を保護するためである。
5. **信頼性 (Reliability)** : AI システムは、信頼性をもって作動しなければならない。
6. **安全とプライバシー (Security and privacy)** : AI システムは安全に作動しなければならない。ユーザーのプライバシーを尊重しなければならない。

これらの諸原則は、善良なイノベーションの本質的要素である。

ローマ、2020年2月28日

[ ] 内は訳者による。

(秋葉 悦子・訳)

\*【訳注】

2020年2月28日、教皇庁生命アカデミー（Pontifica Academia Pro Vita, PAV）年次大会最終日にサン・ピエトロ大聖堂近くのホールで開催された講演会「ルネッサンス：人間中心のAI（RenAIssance：A Human-centric Artificial Intelligence）」で公表された文書である。デジタル・イノベーションとテクノロジーの進歩が、人間の才能と創造性に次第に取って代わるのではなく、人間の才能と創造性に奉仕する未来を創造することに照準を定めて、AIへの倫理的アプローチを支援し、様々な組織、政府、機関の責任感を促進するために生み出された文書である。

講演会当日の最初の署名者は、PAV 会長ヴィンチェンツォ・パリア、マイクロソフト社長ブラッド・スミス、IBM 副社長ジョン・E・ケリー 3 世、世界農業機構（FAO）代表理事チャー・ドンユイ〔屈冬玉〕、イタリア政府イノベーション省大臣パオラ・ピサノの5名であるが、講演会には EU 議長デヴィッド・サッソーリも出席していた。パンデミックのさなかであったが講演会は一般公開され、会場はマスコミ関係者や多くの観衆で満席だった。教皇からはビデオ・メッセージが寄せられた。

2021年4月21日にEUが公表したAI規制案はこれをベースにしたものである。2023年3月31日にイタリアがチャットGPTの使用をスピーディーに規制した背景にもこの文書がある。

2023年1月10日にはユダヤ教とイスラム教の代表者がローマを訪れてこの文書に署名、日本では2022年10月20日に中央大学が署名している。

文書の原文、AI倫理学については、PAVのHPを参照。

<https://www.academyforlife.va/content/pav/en/projects/artificial-intelligence.html>