

# UNESCOのAI倫理勧告案の現状

---

2021年9月17日  
総務省 国際戦略局

# UNESCOのAI倫理勧告案の現状

- 2019年11月の第40回UNESCO総会での決定を受け、2020年2月にAI倫理に関する勧告の作成を目的としたアドホック専門家会合を設置。2020年4～9月にかけて専門家による勧告案の起草が行われ、ユネスコAI倫理原則案を作成（日本から須藤修・東大名誉教授が参加）
- 2021年4月から6月にかけて特別委員会を開催し、ユネスコAI倫理勧告案を作成。
- 今後は、2021年11月（9日～24日）の第41回総会での勧告の採択を目指す。

## 勧告案の概要

### 価値及び原則

（AIシステムのライフサイクルにおけるすべての関係者によって尊重されるべき事項）

**価値**…人間の尊厳、人権及び基本的自由の尊重、豊かな環境と生態系、多様性と包摂性の確保、平和と共存

**原則**…①比例性と無害性、②安全・安心、③公正・無差別、④持続可能性、⑤プライバシーとデータ保護、⑥人間による監督と決断、⑦透明性と説明可能性、⑧責任とアカウントビリティ、⑨Awarenessとリテラシー、⑩マルチステークホルダーによる適応的ガバナンス

### 政策措置

（勧告に基づき加盟国が措置すべき分野等）

### 政策措置

①倫理的影響評価、②倫理的ガバナンスと管理（Stewardship）、③データ政策、④開発と国際協力、⑤環境と生態系（エコシステム）、⑥ジェンダー、⑦文化、⑧教育と研究、⑨コミュニケーションと情報、⑩経済と労働、⑪健康と社会的福利

### 監視及び評価

倫理影響評価と監視（モニタリング）

# UNESCOのAI倫理勧告案の現状

## ■ 価値及び原則

### 価値

人間の尊厳、人権及び基本的自由の尊重、豊かな環境と生態系、多様性と包摂性の確保、平和と共存

### 原則

#### ① 比例性と無害性

個人、人権、基本的自由や社会、環境に害を為す可能性がある場合、リスクアセスメントの手続きの実施と被害の事前防止措置の採択が確保されなくてはならない。

#### ② 安全・安心

望ましくない被害あるいは攻撃に関する脆弱性はAIシステムのライフサイクル全般を通じて対処、防止、除去される必要がある。持続可能でプライバシー保護可能なデータアクセスの枠組みが必要。

#### ③ 公正・無差別

社会正義を推進し、公平性、無差別性を遵守することが必要。あらゆる人による、現地の重要なコンテンツやサービス、文化的多様性を有したAIシステムへの包摂的アクセスを推進するべき。

#### ④ 持続可能性

AI技術が異なる開発レベルの国家間でどう適用されるかにより、持続的な目標の達成に有益になることも、障害となることもあり得るため、SDGsのような幅広い次元の目標に関しては、AI技術の意味合いについては十分に認識されて適用される必要がある。

#### ⑤ プライバシーとデータ保護

プライバシーはAIのライフサイクルを通じて尊重し、保護し、推進されなくてはならず、AIシステムのためのデータは国内・地域・国際の法的枠組みを尊重しつつ、国際法と整合し、本勧告の価値や原則に沿った手法で収集・使用・共有・保存・削除されなくてはならない。データ保護の枠組みやガバナンス機構はマルチステークホルダーアプローチで確立されるべきである。

# UNESCOのAI倫理勧告案の現状

## ⑥人間による監督と決断

AIシステムの全ライフサイクルにおいて、常に倫理的・法的責任を特定可能とするべき。AIシステムは究極的な人間の責任とアカウントビリティを代替することではなく、生死に関わる決定はAIシステムに委ねられるべきではない。

## ⑦透明性と説明可能性

AIアクターは製品やサービスがAIシステムを直接・間接に利用している場合、利用者に適時に適切な方法で通知するべきである。また、アルゴリズムの開発は説明可能性を確保するべき。

## ⑧責任とアカウントビリティ

AIシステムに基づく意思決定と行動の倫理的責任と法的責任は、最終的にはAIアクターに帰属するべきである。AIシステムのアカウントビリティを確保するために、適切な監視（oversight）、影響評価、監査とデュー・デリジェンスのメカニズムが開発されるべき。

## ⑨Awarenessとリテラシー

AI技術とデータの価値についての人々の認識と理解を促進するべき。AIシステムの影響についての学習は、AIシステムの人権、環境、生態系へのインパクトに立脚して行われるべき。

## ⑩マルチステークホルダーによる適応的ガバナンス

データの利用では国際法と国家主権が尊重されなくてはならず、領土内で生成されたデータや領土を通過するデータは、国家が国際法も踏まえて管理することができることを意味する。

AIの恩恵がすべての人に共有され、持続可能な開発に寄与するために、AIガバナンスには、AIの全ライフサイクルを通じて、様々なステークホルダーの参画が必要。

# UNESCOのAI倫理勧告案の現状

## 政策措置

### ①倫理的影響評価

加盟国は倫理影響評価のような、AIシステムの恩恵と懸念とリスクを特定し評価するための影響評価の枠組みを導入すべき。加盟国と民間企業は人権、法の支配、包摂的社会的尊重に対するAIシステムの影響を、特定し、防止し、提言し、いかに対処するかを説明するためのデュー・デリジェンスと監視のメカニズムを開発すべき。

### ②倫理的ガバナンスと管理（Stewardship）

加盟国はAIガバナンスのメカニズムが包摂的で、透明で、学際的で、多国間で（これは越境の被害についての減災や救済の可能性を含む）、かつマルチステークホルダーであることを確保すべき。

加盟国は、デジタル世界でも現実世界でも、人権と基本的自由と法の支配が尊重されるよう、強力な法執行メカニズムと救済措置の制定により、AIシステムを通じて発生した被害が、調査され、救済されることを確保すべき。

### ③データ政策

加盟国はデータ収集・抽出プロセスの十分性、適切なセキュリティと保護措置、フィードバックメカニズムを含む、AIシステムのための学習データの質の継続的評価を確保するデータ・ガバナンス戦略を開発すべき。

加盟国はサーベイランスのような懸念への対処を含め、国際法に沿ってプライバシー権を保護するために適切なセーフガードを設定すべき。

### ④開発と国際協力

加盟国と多国籍企業はAIに関連する倫理問題の議論を、関連の国際的、あるいは政府間の、そしてマルチステークホルダーのフォーラムに入れ込むことで、AI倫理の課題を優先課題にしていくべき。

加盟国は教育、科学、文化、コミュニケーションと情報、ヘルスケア、農業、環境、天然資源やインフラ管理、経済計画などの開発の分野でのAIの利用が、本勧告案で示されている価値と原則を遵守していることを確認すべき。

### ⑤環境と生態系（エコシステム）

加盟国と企業はAIシステムのライフサイクルを通じて、炭素排出量、エネルギー消費、AI技術製造を支える原料採取の環境インパクトを含む影響につき、直接、間接の評価をするべき。

加盟国は環境・生態系の観察、保護、再生を含め、対災害強靱性のためのAIを使ったソリューションの開発・採択を確保するようインセンティブを導入すべきである。

# UNESCOのAI倫理勧告案の現状

## 政策措置

### ⑥ジェンダー

加盟国はデジタル技術とAIが、ジェンダー平等の達成に最大限貢献し、女兒と女性の人権と基本的自由、そして安全性がAIライフサイクルのすべての段階で侵害されないよう確保すべき。

### ⑦文化

加盟国は絶滅危機にある言語や原住民の言語・知識を含め、文書遺産や有形・無形文化遺産の保存、改良、理解、推進、管理やアクセスにAIシステムを組み込むよう奨励される。

加盟国は例えば、AI技術を用いて創作された作品の知的財産権をいかに守る、あるいは守るべきか否かの決定など、AIと知財の交差点において新たな研究を促進すべき。

### ⑧教育と研究

加盟国は人々を力づけ、AIシステムの広範な利用によるデジタルディバイドおよびデジタルアクセスの格差を解消するために、すべての国のすべてのレベルの大衆にAIリテラシーの教育を十分に提供するため、国際機関や教育機関、民間・非政府団体と協力すべき。

### ⑨コミュニケーションと情報

加盟国はAIシステムを情報と知識へのアクセス向上に使うべき。AIアクターは自動化されたコンテンツの生成・監視・編集に関連して表現の自由と情報へのアクセスを尊重・遵守することを確保すべき。

### ⑩経済と労働

すべての国において、とりわけ労働集約的な経済を有する国において、加盟国はAIシステムの労働市場や教育の要件に対関する影響を評価し、対処すべき。

### ⑪健康と社会的福利

グローバルな健康リスクと不確実性と闘うための国際的連帯を構築・維持しつつ、病気の蔓延の抑制を含めて、人の健康の向上と生活の権利の保護のためのAIシステムの実装のために努力すべき。

# UNESCOのAI倫理勧告案の現状

## 監視及び評価

倫理影響評価と監視（モニタリング）

**加盟国は固有の条件や統治機構、憲法の規定に従ってAI倫理に関する政策、計画、メカニズムを監視（モニタリング）観察・評価すべき。**

**ユネスコは下記により貢献。**

- a) 厳格な科学的調査と国際的な人権法に基づく倫理影響評価（EIA; Ethical Impact Assessment)のための「ユネスコ方式（UNESCO methodology）」を開発。そのための政府職員、政策立案者、AIアクター向けの人材育成教材を提供。
- b) 加盟国のAI準備状況を特定する「UNESCO準備評価方式」を開発。
- c) AI倫理政策の有効性・効率性を事前・事後に評価する「UNESCO方式」を開発。
- d) AI倫理に関する政策の調査ベース・証拠ベースでの分析を強化。
- e) AI倫理政策に関する調査報告、科学的書籍、データや統計の収集と発信

# UNESCOのAI倫理勧告案の現状

## ■ 対処方針と結果

### ● 一般討論

本勧告案に関するユネスコの取組を評価した上で、法的な規制については技術の発展を阻害しかねないという懸念が有識者から示されており、本文書全体を通じて規制色が強い内容とならないように意識すべきであると主張。

### ● 「III.2. Principles」 (パラ38)

透明性を理由に営業秘密の開示を民間企業に迫ることを奨励すべきではないことを明確にするために、“other principles such as privacy, safety and security” の後ろに「and other legitimate interests such as intellectual property」の追加を提案。 ⇒ コンセンサスを得ることができず、受け入れられなかった。

### ● 「III.2. Principles」 (パラ46)

データの利用に関する国際法やデータ主権の定義について、国際的に広く合意された概念があるとは言い難く、またユネスコという場で議論することが適切と思われないことから、これらへの言及は削除するよう提案。  
⇒ 独、英、スイスも同様の趣旨の主張をしたものの、最終的には途上国（ブラジル等）の主張を考慮した事務局の修正提案がコンセンサスでまとまった（Data sovereignty の用語は削除。Data protectionのみが残った。）

### ● 「IV. Areas of Policy Action」 (パラ68)

”and that AI systems should not be given legal personality themselves”に関して、賠償責任分担等において法技術的に必要となる可能性があり、さらに国際的にコンセンサスが得られていないことから、削除を主張。AIに法人格を認めれば「人間が中心の社会」という前提が揺ぐという懸念につき共感しつつ、AIに部分的な法人格を与えることで、むしろ人間中心のAI社会を柔軟に構築できる可能性があることを指摘。  
⇒ 受け入れられなかったが、日本の意見は最終報告書に記載された。

### ● 「IV. Areas of Policy Action」 (パラ73)

データの利用に関する国際法やデータ主権の定義について、国際的に広く合意された概念があるとは言い難く、またユネスコという場で議論することが適切と思われないことから、これらへの言及は削除するよう提案する。  
⇒ (パラ46の議論も踏まえて、日本の主張したdata sovereigntyの削除、data security をdata protection とする修正提案が受け入れられた。

# UNESCOのAI倫理勧告案の現状 (参考)

## 【協議前の案文】

38. While efforts need to be made to increase transparency and explainability of AI systems throughout their life cycle to support democratic governance, the level of transparency and explainability should always be appropriate to the context, as some trade-offs may exist between transparency and explainability and **other principles such as privacy, safety and security**. People have the right to be aware when a decision is being made on the basis of AI algorithms, and in those circumstances require or request explanatory information from private sector companies or public sector institutions.
46. International law and national sovereignty should must be respected in the use of data. **Data sovereignty means that States, complying with international law, regulate the data generated within or passing through their territories**, and take measures towards effective regulation of data, including data protection, based on respect for the right to privacy and other human rights norms and standards.
68. Member States should develop, review and adapt, as appropriate, regulatory and legal frameworks to achieve accountability and responsibility for the content and outcomes of AI systems at the different phases of their life cycle. Member States should introduce liability frameworks or clarify the interpretation of existing frameworks to ensure the attribution of accountability for the outcomes and behaviour the functioning of AI systems. Furthermore, when developing regulatory frameworks, **Member States should, in particular, take into account that ultimate responsibility and accountability must always lie with natural or legal persons and that AI systems should not be given legal personality themselves**. To ensure this, such regulatory frameworks should be consistent with the principle of human oversight and establish a comprehensive approach focused on the actors and the technological processes involved across the different stages of the AI systems life cycle.
73. Member States should ensure that individuals retain rights over their personal data and are protected by a framework, which notably foresees transparency, appropriate safeguards for the processing of sensitive data, the highest appropriate level of data security, effective and meaningful accountability schemes and mechanisms, the full enjoyment of the data subjects' rights, in particular the right to access and the right to erasure of their personal data in AI systems, except for certain circumstances in compliance with international law, an appropriate level of protection while in full compliance with data protection legislation where data are being used for commercial purposes such as enabling micro-targeted advertising, transferred cross-border, and an effective independent oversight as part of a data governance mechanism which keeps individuals in control of their personal data respects data sovereignty and balances this with fosters the benefits of a free flow of information internationally, including access to data.