



人工知能に関する理事会勧告

非公式翻訳

OECDによる法的規範

本書は、OECD事務総長の責任のもとで発行されている。本書で表明されている意見や主張は、必ずしもOECD加盟国の公式見解を反映するものではない。

本文書並びに掲載のデータ及び地図は、領土に関する地位或いは主権、定められた国境及び境界、またいかなる領土、都市、地域の名称をも害するものではない。

本文書は(無料)無償で提供される。それがいかなる点においても変更されない限り、本文書は追加の許可を必要とせずに(無料)無償で複製し配布することができる。本文書を販売することは許されない。

これはOECDの公式の翻訳ではありません。原文との整合性を確保するよう努めていますが、公式版は下記のウェブサイトに掲載されている英語版とフランス語版です。 <https://legalinstruments.oecd.org>.

理事会は：

1960 年 12 月 14 日の経済協力開発機構条約の第 5(b)に鑑み、

プライバシー、デジタル・セキュリティ、消費者保護及び責任ある企業行動の分野で OECD が策定した標準に鑑み、

国連総会で採択された持続可能な開発のための 2030 アジェンダで設定された持続可能な開発目標（A/RES/70/1）及び 1948 年の世界人権宣言に鑑み、

その他の政府機関及び非政府団体が国際場裏において人工知能（以下、「AI」という）に関する重要な取組を行っていることに鑑み、

AI が広範で多大な影響を世界的に及ぼし、社会や経済分野及び仕事の世界を変革させており、また、そのような影響が将来さらに強くなる可能性が高いことを認識し、

AI には、人々の福祉と幸福を増進し、確かで持続可能な世界の経済活動に貢献し、イノベーションと生産性を向上させるとともに、地球規模の重要な課題の解決の一助となる可能性があることを認識し、

それと同時に、これらの変革は、特に、経済のシフト、競争、労働市場の変化、格差、並びに民主主義と人権、プライバシーとデータの保護、及びデジタル・セキュリティに関して、社会及び経済の内部又は社会間及び経済間において本質的に異なる影響を与える可能性があることを認識し、

信頼がデジタル変革の成功要因であること、将来的な AI の利活用とその影響を予測することは難しいものの AI システムの信頼性が AI の普及と導入の鍵となる重要な要素であること、及びこのようなテクノロジーの有益な可能性を引き出しつつ関連するリスクを限定的なものにするためには十分な情報提供の下で社会全体で行われる開かれた議論が必要であることを認識し、

法律、規制、政策に係る国内及び国際的な既存の枠組については、人権、消費者及び個人データの保護、知的財産権、責任ある企業行動、並びに競争に関するものを含め、その妥当性の評価や新たなアプローチの開発が必要となる場合があるものの、既に AI との関連性を有することが示されていることを強調し、

AI の急速な発展と実装を踏まえ、信頼できる AI に対する人間中心のアプローチを推進し、研究を促進し、イノベーションのための経済的なインセンティブを保持し、かつその役割と状況に応じて全てのステークホルダーに対して適用される安定的な政策環境の必要性を認識し、

AI の利活用がもたらす機会を歓迎するとともに、それによって生じる課題に対処すること、及びステークホルダーがこれらに関与するために必要な能力を身に付けることは、社会における信頼できる AI の導入の推進や、AI の信頼性が世界市場での競争力のパラメータとなっていくために不可欠なものであると考え、

デジタル政策委員会の提案に対し：

I. 以下の用語については、本勧告の適用に当たり次のように理解されるべきものであることに同意する：

- **AI システム**：AI システムは、明示的又は黙示的な目的のために、実環境又は仮想環境に影響を及ぼす、予測、コンテンツ、推奨、又は決定などの出力をどのように生成するか、受け取った入力内容から推論する機械ベースのシステムである。展開後の自律性と順応性のレベルは、AI システムによってそれぞれ異なる。

- **AI システムのライフサイクル**：AI システムのライフサイクルは、通常、計画と設計、データの収集と処理、（モデルの構築及び/又は既存のモデルの特定のタスクへの適合、テスト、評価、検証、妥当性確認、使用/展開のための利用可能化、稼働とモニタリング、稼働停止/終了など、いくつかの段階から構成される。これらの段階はしばしば相互に作用し合い、かつ必ずしも順番に実行されるものではない。AI システムの稼働を終了する決定は、稼働とモニタリングの段階のどの時点でも行われる可能性がある。
- **AI アクター**：AI アクターとは、AI システムのライフサイクルにおいて能動的な役割を果たす者であり、AI の展開又は稼働を行う組織や個人が含まれる。
- **AI ナレッジ (knowledge)**：AI ナレッジ (knowledge) とは、データ、コード、アルゴリズム、モデル、研究、ノウハウ、トレーニング・プログラム、ガバナンス、プロセス及びベスト・プラクティスなど、リスク管理を含めた AI システムのライフサイクルを理解し、これに参画するために必要となるスキル及びリソースを指している。
- **ステークホルダー**：ステークホルダーには、直接的なものであるか間接的なものであるかを問わず、AI システムに関与するか、又は AI システムから影響を受ける組織及び個人の全てが含まれる。AI アクターはステークホルダーの一部である。

第1節：信頼できる AI の責任あるスチュワードシップのための原則

II. 本勧告を遵守する加盟国及び非加盟国（以下、「遵守国」という）に対し、全てのステークホルダーに関係する以下の「信頼できる AI の責任あるスチュワードシップのための原則」を推進かつ履行するよう**勧告**する。

III. 全ての AI アクターが、それぞれの役割に応じ、以下の「信頼できる AI の責任あるスチュワードシップのための原則」を推進かつ履行するよう**求める**。

IV. 以下の原則は相互に補完的なものであり、全体でまとめて考慮されるものであることを**強調**する。

1.1. 包摂的な成長、持続可能な開発及び幸福

ステークホルダーは、人間の能力の増強や創造性の向上、少数派の包摂の促進、経済・社会・性別その他の格差の改善、及び自然環境の保護などがもたらす包摂的な成長、幸福、持続可能な開発及び環境の持続可能性といった、人々と地球にとって有益な結果を追求することにより、信頼できる AI の責任あるスチュワードシップに積極的に取り組むべきである。

1.2. 法の支配、人権並びに公平性及びプライバシーを含む民主主義的価値の尊重

a) AI アクターは、AI システムのライフサイクルの全体を通じて、法の支配、人権並びに民主主義的及び人間中心の価値観を尊重すべきである。これらには、無差別及び平等、自由、尊厳、自主自律、プライバシーとデータの保護、多様性、公平性、社会正義及び国際的に承認された労働者の権利が含まれる。これには、適用される国際法によって保護された表現の自由及びその他の権利と自由を尊重しつつ、AI によって増幅された誤情報や偽情報に対処することも含まれる。

b) この目的を達成するため、AI アクターは、意図された目的外の使用や意図的な不正使用、または意図しない誤用などから生じるリスクに対処することも含め、人間の主体性や人間による監視の余地を残しておくことなど、状況に適した形で、かつ技術の水準を踏まえたメカニズム及びセーフガードを実装すべきである。

1.3. 透明性及び説明可能性

AI アクターは、AI システムに関する透明性及び責任ある開示に積極的に関与すべきである。これらを達成するため、AI アクターは、以下のために、状況に適した形で、かつ技術の水準を踏まえ、意味のある情報提供を行うべきである：

- i. AI システムの能力及び限界も含め、一般的な理解を深めること。
- ii. 職場におけるものを含め、AI システムが使われていることをステークホルダーに認識してもらうこと。
- iii. そうすることが実行可能かつ有益な場合には、AI システムに影響される者がその出力を理解できるようにするため、予測、コンテンツ、推奨、又は決定をもたらしたデータ/入力情報の出所、要因、プロセス及び/又は理論に関して平明で理解しやすい情報を提供すること。
- iv. AI システムの悪影響を受けた人々が出力に対する異議を申し立てることができるような情報を提供すること。

1.4. 頑健性、セキュリティ及び安全性

- a) AI システムは、通常の使用、予見可能な使用や誤用、又はその他の悪条件においても正常に機能するとともに、不合理な安全上及び/又はセキュリティ上のリスクをもたらすことがないように、そのライフサイクル全体にわたって頑健で、セキュア、かつ安全であるべきである。
- b) AI システムが過度の損害を引き起こす又は望ましくない動作を示す危険性がある場合であって、適当な場合には、必要に応じてそれらをオーバーライド、修復、及び/又は安全に廃止できるようにするためのメカニズムを整備しておくべきである。
- c) また、技術的に可能な場合には、表現の自由を尊重しつつ情報インテグリティを強化するためのメカニズムも整備すべきである。

1.5. アカウンタビリティ

- a) AI アクターは、その役割と状況に基づき、かつ技術の水準を踏まえた形で、AI システムが適正に機能していること及び上記の原則を尊重していることについて、アカウンタビリティを果たすべきである。
- b) この目的のために、AI アクターは、AI システムの出力及び問合せに対する応答の分析を可能とするため、データセット、プロセス及び AI システムのライフサイクルの中で行われた決定に関するものを含めて、状況に適した形で、かつ技術の水準を踏まえたトレーサビリティを確保すべきである。
- c) AI アクターは、その役割、状況及び能力に基づき、系統化されたリスクマネジメントのアプローチを AI システムのライフサイクルの各段階に継続的に適用し、適当な場合には、異なる AI アクター、AI ナレッジと AI リソースの提供者、AI システムの利用者、その他のステークホルダー間の協力を通じるなどして、AI システムに関連するリスクに対応するための責任ある企業行動を採用すべきである。リスクには、有害なバイアスに加え、安全、セキュリティ及びプライバシーなどの人権、労働者の権利、知的財産権に関連するリスクが含まれる。

第2節：信頼できる AI のための国家政策と国際協力

V. 遵守国に対し、国内の政策及び国際的な協力に関し、中小企業（SME）に対する特別の注意を払いながら、第1節の原則と整合させつつ以下の勧告を実行するよう**勧告**する。

2.1. AI の研究開発への投資

- a) 技術的に困難な課題や AI の社会的・法的・倫理的な影響と政策課題に焦点を当てた信頼できる AI のイノベーションを促進するため、各国政府は、学際的な取組を含め、研究開発やオープンサイエンスの分野で長期的な公共投資を検討し、かつ民間投資を奨励すべきである。
- b) また、各国政府は、有害なバイアスがない AI の研究開発のための環境を支援し、かつ相互運用性と技術標準の利用を向上させるため、代表性を有し、かつプライバシーとデータの保護を尊重するオープンソースのツールやオープンなデータセットに対する公共投資を検討し、かつ民間投資を奨励すべきである。

2.2. 包摂的な AI を推進するためのエコシステムの整備

各国政府は、信頼できる AI のための包摂的でダイナミック、持続可能で相互運用可能なデジタル・エコシステムの開発及びそれへのアクセスを促進すべきである。このエコシステムには、適切である限りにおいて、特に、データ、AI 技術、計算及び接続性のためのインフラ並びに AI ナレッジを共有するためのメカニズムが含まれる。これに関連し、各国政府は、データ・トラストのような、安全、公平、適法、かつ倫理的なデータ共有のためのメカニズムに対する支援を検討すべきである。

2.3. AI を推進するための相互運用可能なガバナンス及び政策環境の形成

- a) 各国政府は、信頼できる AI システムの研究・開発段階から展開・稼働の段階への移行を支援するための機動的な政策環境整備を促進すべきである。この観点から、各国政府は、適切な限りにおいて、AI システムのテストや規模の拡張を可能とする制御された環境を提供する実証実験の利用を検討すべきである。また、適切な限りにおいて、ガバナンスの目的達成に向けて柔軟性を提供する結果ベースのアプローチを採用するとともに、各国の管轄権の及ぶ領域の内外で協力して相互運用可能なガバナンスと政策環境を促進すべきである。
- b) 各国政府は、信頼できる AI のイノベーションと競争を奨励するため、適切な限りにおいて、AI システムに適用される政策や規制の枠組とその評価のメカニズムについて見直し、かつ状況に順応させるべきである。

2.4. 人材育成及び労働市場の変化への備え

- a) 各国政府は、仕事の世界と社会全体の変化に備えるために、ステークホルダーと緊密に協働すべきである。政府は人々に必要なスキルを習得させるなどして、広範囲に適用される AI システムを人々が効果的に利用し、かつそれとうまく関わることをできるようにすべきである。
- b) 各国政府は、AI の普及がもたらす労働市場の変化が労働者にとって公平なものであるようにするため、就労期間を通じたトレーニング・プログラム、社会的保護も含めた離職を余儀なくされた者の支援、かつ労働市場における新たな機会へのアクセス提供等を通じて、社会対話も含め、様々な措置を講じるべきである。
- c) また、各国政府は、職場における AI の責任ある利用の推進、労働者の安全及び仕事と公共サービスの質の向上、並びに起業家精神と生産性の向上のため、かつ AI の恩恵が幅広

くかつ公平に共有されることを目指して、ステークホルダーと密接に協働すべきである。

2.5. 信頼できる AI のための国際協力

- a) 開発途上国を含め、各国政府は、ステークホルダーとともに、本勧告が掲げる原則の推進、及び信頼できる AI の責任あるスチュワードシップの進展のために、積極的に協力すべきである。
- b) 各国政府は、OECD 及びその他の国際的並びに地域的なフォーラムを通じ、適切な限りにおいて、AI ナレッジの共有を促進するために協働すべきである。こうしたフォーラムは、AI に関する長期的な専門性を確保するため、国際的、分野横断的、かつ開放的なマルチ・ステークホルダーによる取組を奨励すべきである。
- c) 各国政府は、相互運用性を有し、また、信頼できる AI のため、マルチ・ステークホルダーの合意に基づく国際的な技術基準の開発を推進すべきである。
- d) 各国政府はさらに、AI の研究開発及び展開を計測するための国際的に比較可能な指標の開発と各国政府自身による利用とを奨励するとともに、本勧告が掲げる原則の履行の進捗をモニターするための土台となるエビデンスを収集すべきである。

VI. 事務総長と遵守国に対し、本勧告を広く普及させるよう求める。

VII. 非遵守国に対し、本勧告を十分に考慮し、かつ遵守するよう求める。

VIII. デジタル政策委員会に対して、AI ガバナンス作業部会を通して次の事項を行うよう指示する。すなわち:

- a) 本勧告を踏まえ、また、他の国際場裏における取組を考慮しつつ、本委員会における AI に関する重要な作業を継続し、エビデンスに基づいた AI 政策の計測に係る枠組の開発をさらに進めていくこと。
- b) 状況の変化と新たな政策優先事項に対応するため、本勧告の履行のためのさらに実用的なガイダンスを作成及び更新すること。
- c) AI に関する政策や取組について、本勧告の履行に係る実例を含め、情報交換のためのフォーラムを提供するとともに、AI の信頼と導入を推進するためにマルチ・ステークホルダーかつ学際的な意見交換を促進すること。
- d) 関係する他の委員会と協議した上で、本勧告の改訂後 5 年以内、及び少なくともその後 10 年毎に、本勧告の履行、普及、継続的な関連性について理事会に報告すること。