

**(参考) 「国際的な議論のための AI 開発ガイドライン案」
と「AI 利活用原則案」**

国際的な議論のためのAI開発ガイドライン案とAI利活用原則案①

尊重すべき価値	国際的な議論のためのAI開発ガイドライン案	AI利活用原則案
適正な利用	<p>⑧ 利用者支援の原則 開発者は、AIシステムが利用者を支援し、利用者を選択の機会を適切に提供することが可能となるよう配慮する。</p> <ul style="list-style-type: none"> ・ 利用者にとって操作しやすいインターフェース ・ 利用者を選択の機会(デフォルトの設定、理解しやすい選択肢の提示等)を適時適切に提供する機能 ・ 社会的弱者の利用を容易にするための取組 ・ 利用者に対する適切な情報提供 	<p>① 適正利用の原則 利用者は、人間とAIシステムとの間及び利用者間における適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める。</p> <ul style="list-style-type: none"> ・ 開発者等からの情報提供や説明を踏まえた適正な範囲・方法での利用、利用者における必要な知識・技能の習得 ・ AIソフトのアップデート及びAIの点検・修理等の実施 ・ AIシステムによりなされた判断に対する人間の判断の介入 ・ 能力・知識・措置を講じる容易性等を踏まえた関係者間の役割・責任の分担 ・ 予防措置、事後対応(原因解明、再発防止措置等)における関係者間の協力
適正な学習(学習データの質)		<p>② 適正学習の原則 利用者及びデータ提供者は、AIシステムの学習等に用いるデータの質に留意する。</p> <ul style="list-style-type: none"> ・ 学習等に用いるデータの質(正確性や完全性など)
AI間の連携	<p>① 連携の原則 開発者は、AIシステムの相互接続性と相互運用性に留意する。</p> <ul style="list-style-type: none"> ・ 国際的な標準や規格への準拠 ・ データ形式の標準化、インターフェイスやプロトコルのオープン化への対応 ・ 標準必須特許等のライセンス契約及びその条件についてのオープン・公平な取扱い 	<p>③ 連携の原則 AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービス相互間の連携に留意する。また、利用者は、AIシステムがネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。</p> <ul style="list-style-type: none"> ・ 提供するAIシステム又はAIサービスの相互接続性と相互運用性 ・ データ形式やプロトコル等の標準化への対応 ・ AIネットワーク化により惹起・増幅される課題
制御可能性	<p>③ 制御可能性の原則 開発者は、AIシステムの制御可能性に留意する。</p> <ul style="list-style-type: none"> ・ 事前の検証及び妥当性の確認、サンドボックスにおける実験 ・ 人間や信頼できる他のAIシステムによる監督・対処 	
安全	<p>④ 安全の原則 開発者は、AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。</p> <ul style="list-style-type: none"> ・ 事前の検証及び妥当性の確認 ・ 本質安全や機能安全に資するための措置 ・ (生命・身体・財産の安全に関する判断を行うAIシステムについて)ステークホルダに対する設計の趣旨などの説明 	<p>④ 安全の原則 利用者は、AIシステム又はAIサービスの利活用により、アクチュエータ等を通じて、利用者等及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。</p> <ul style="list-style-type: none"> ・ (生命・身体・財産に危害を及ぼし得る分野での利活用において)AIシステムの点検・修理及びAIソフトのアップデートを行うことなどによる危害の防止 ・ 危害が発生した場合に備えた事前措置

国際的な議論のためのAI開発ガイドライン案とAI利活用原則案②

尊重すべき価値	国際的な議論のためのAI開発ガイドライン案	AI利活用原則案
セキュリティ	<p>⑤ セキュリティの原則 開発者は、AIシステムのセキュリティに留意する。</p> <ul style="list-style-type: none"> 情報の機密性、完全性、可用性に加え、信頼性、頑健性にも留意 事前の検証及び妥当性の確認 セキュリティ・バイ・デザイン 	<p>⑤ セキュリティの原則 利用者及びデータ提供者は、AIシステム又はAIサービスのセキュリティに留意する。</p> <ul style="list-style-type: none"> その時点での技術水準に照らした合理的な対策 侵害が発生した場合に備えた事前措置 セキュリティ対策のためのサービス提供、インシデント情報の共有 不正確又は不適切なデータ学習による脆弱性への対応
プライバシー	<p>⑥ プライバシーの原則 開発者は、AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する。</p> <ul style="list-style-type: none"> 事前のプライバシー影響評価 プライバシー・バイ・デザイン 	<p>⑥ プライバシーの原則 利用者及びデータ提供者は、AIシステム又はAIサービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する。</p> <ul style="list-style-type: none"> AIシステム又はAIサービスの利活用における他者又は自己のプライバシーの尊重、侵害が発生した場合に備えた事前措置 学習等に用いるパーソナルデータの収集・分析・提供における他者のプライバシーの尊重 AIシステム又はAIサービスを利用したプロファイリングを行う場合における対象者のプライバシーの尊重 第三者提供における本人同意の確保
倫理	<p>⑦ 倫理の原則 開発者は、AIシステムの開発において、人間の尊厳と個人の自律を尊重する。</p> <ul style="list-style-type: none"> (人間の脳や身体と連携するAIシステムの開発の場合)生命倫理に関する議論などの参照 学習データに含まれる偏見などに起因して不当な差別が生じないための所要の措置 AIシステムが人間性の価値を不当に毀損することがないように留意 	<p>⑦ 尊厳・自律の原則 利用者は、AIシステム又はAIサービスの利活用において、人間の尊厳と個人の自律を尊重する。</p> <ul style="list-style-type: none"> AIシステム又はAIサービスにより意思決定や感情が操作されるリスク、AIシステム又はAIサービスに過度に依存するリスク (AIシステム又はAIサービスを人間の脳や身体と連携させる場合)生命倫理に関する議論などの参照
		<p>⑧ 公平性の原則 AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービスの判断によって個人が不当に差別されないよう配慮する。</p> <ul style="list-style-type: none"> 学習等に用いられるデータの代表性やデータに内在する社会的なバイアス AIシステムよりなされた判断に対する人間の判断の介入

国際的な議論のためのAI開発ガイドライン案とAI利活用原則案③

尊重すべき価値	国際的な議論のためのAI開発ガイドライン案	AI利活用原則案
透明性	<p>② 透明性の原則 開発者は、AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する。</p> <ul style="list-style-type: none"> ・ 生命、身体、自由、プライバシー、財産などに影響を及ぼす可能性のあるAIシステムにおける入出力の検証可能性及び判断結果の説明可能性（※アルゴリズム、ソースコード、学習データの開示を想定するものではない。） 	<p>⑨ 透明性の原則 AIサービスプロバイダ及びビジネス利用者は、AIシステム又はAIサービスの入出力の検証可能性及び判断結果の説明可能性に留意する。</p> <ul style="list-style-type: none"> ・ 生命、身体、財産に危害が及ぼし得る分野で利活用する場合におけるAIシステム又はAIサービスの入出力の記録・保存 ・ 個人の権利・利益に重大な影響を及ぼす可能性のある分野で利活用する場合におけるAIシステム又はAIサービスの判断結果の説明可能性の確保（※アルゴリズム、ソースコード、学習データの開示を想定するものではない。）
アカウントビリティ	<p>⑨ アカ운タビリティの原則 開発者は、利用者を含むステークホルダに対しアカウントビリティを果たすよう努める。</p> <ul style="list-style-type: none"> ・ 利用者にAIシステムの選択及び利活用に資する情報の提供 ・ 開発原則①～⑧の趣旨に鑑み、AIシステムの技術的特性についての情報提供や説明、ステークホルダとの対話を通じた意見聴取 ・ AIサービスプロバイダなどとの情報共有・協力 	<p>⑩ アカ운タビリティの原則 AIサービスプロバイダ及びビジネス利用者は、消費者的利用者及び間接利用者を含むステークホルダに対しアカウントビリティを果たすよう努める。</p> <ul style="list-style-type: none"> ・ 消費者的利用者、間接利用者等に対するアカウントビリティ ・ AIシステム又はAIサービスに関する利用方針の通知・公表