

人工知能関連技術の研究開発及び活用の
適正性確保に関する指針骨子

全体構成

1 我が国における適正性確保に関する基本的な考え方	2
(1) 本指針の位置付け.....	2
(2) 本指針における適正性確保の考え方.....	2
● 人間中心.....	3
● 公平性.....	3
● 安全性.....	3
● 透明性.....	3
● アカウンタビリティ.....	3
● セキュリティ.....	3
● プライバシー.....	3
● 公正競争.....	3
● AIリテラシー.....	4
● イノベーション.....	4
(3) 適正性確保のための基本方針.....	4
① リスクベースでのアプローチ.....	4
② ステークホルダーの積極的な関与.....	4
③ 一気通貫でのAIガバナンスの構築.....	4
④ アジャイルな対応.....	4
2 研究開発機関及び活用事業者が特に取り組むべき事項	5
(1) AIガバナンスによる俯瞰的な適正性の確保.....	5
(2) ステークホルダーとの信頼関係の構築に向けた透明性の確保.....	5
(3) 十分な安全性の確保.....	5
(4) 事業継続性確保による安全な環境の維持.....	6
(5) AIのイノベーションの基盤となるデータの重要性を踏まえたステークホルダーへの配慮.....	6
3 国及び地方公共団体が特に取り組むべき事項	7
(1) AIの積極的かつ先導的な活用によるイノベーションの促進.....	7
(2) 社会全体におけるAIリテラシーの向上.....	7
(3) AIガバナンスの在り方の検討.....	7
(4) 行政としてのアカウンタビリティを果たすこと.....	8
4 国民が特に取り組むべき事項	9
(1) 人間中心の原則に基づくAIの責任ある利用.....	9
(2) AIリテラシーに基づく適切な利用.....	9

1 我が国における適正性確保に関する基本的な考え方

(1) 本指針の位置付け

- 人工知能関連技術の研究開発及び活用の推進に関する法律（「A I 法」）第 13 条に基づく本指針は、信頼できる A I の実現に向けて、事業者、国民等の全ての主体における A I の研究開発・活用の適正な実施に係る自主的かつ能動的な取組を促すために、国際的な規範の趣旨に即して策定するもの。
- 本指針の構成としては、1 で全ての主体における A I の研究開発・活用の適正性確保に必要な主要素と基本方針を示し、2 以降で各主体が 1 を前提として特に取り組むべき事項を記載する。
- 全ての主体は、適正性確保に必要な主要素を認識、理解することが求められる。また、取り組むべき事項について、各主体の規模や立場、A I がもたらすリスクに応じて、その時点で適用し得る技術や知見を踏まえて適切な水準で対応することが求められる。
- なお、我が国は、信頼できる A I の開発、活用、普及に向けて、本指針を中心とした枠組みを、国際モデルとなるよう展開するとともに、A I に係る国際的なルール形成を行う枠組みである「広島 A I プロセス」を牽引してきた実績を踏まえ、引き続き A I ガバナンスにおける国際協調を主導する。

(2) 本指針における適正性確保の考え方

- A I は、経済成長や国民生活の発展に寄与するものであることから、A I の社会実装を進め、イノベーションを促進していくことが重要である一方、A I には様々なリスクがある。例えば、誤判断やハルシネーション等の技術的リスク、偽・誤情報の生成・拡散、偏見・差別の助長、犯罪への利用、過度な依存、プライバシー・財産権の侵害、環境負荷の増大、雇用・経済不安等の社会的リスク、さらにはサイバー攻撃等の安全保障リスクがあり、これらのリスクは A I の技術進歩とともに変化したり、未知のリスクが発生したりする可能性があり、リスクに対する社会的な受容水準も変化し得る。

1 ・ このため、本指針では、適正性確保に当たって、適正性についての一義的な定義
2 や絶対的な水準を定めるものではなく、各主体が研究開発、活用するA Iの特性、
3 用途、目的や、自身の立場、社会的役割等を踏まえて自主的に取組を進めるとい
4 う考え方の下、その際に考慮すべき主要要素を以下のとおり示す。

5 ● 人間中心

- 6 ☆ 人間の尊厳や基本的人権を尊重し、法令を遵守すること。
- 7 ☆ 誰もがA Iの恩恵を享受できるよう、多様性、包摂性を尊重し、多様な
8 人々の幸福の追求による包摂的な成長を目指すこと。
- 9 ☆ A Iの活用にあたっては、人間自らが意思決定を行うこと。

10 ● 公平性

- 11 ☆ A Iの活用によって、社会に不当な偏見や差別を生じさせたり、助長した
12 りしないこと。

13 ● 安全性

- 14 ☆ A Iの活用によって、生命、身体、財産に危害を及ぼさないようにするこ
15 と。

16 ● 透明性

- 17 ☆ A Iに対する信頼性が向上するよう、必要かつ技術的に可能な範囲での情
18 報の開示、事後的な検証可能性の確保等により、透明性を適切に確保する
19 こと。

20 ● アカウンタビリティ

- 21 ☆ A Iがもたらす社会的影響を踏まえ、責任の所在の明確化、責任を果たす
22 ための仕組みの構築等により、技術的、制度的、社会的観点からアカウン
23 タビリティを合理的な範囲で果たすこと。

24 ● セキュリティ

- 25 ☆ 不正な操作によるA Iの意図しない動作や停止をはじめとするA Iのセキ
26 ュリティ上のリスクを低減させるよう、A Iのセキュリティを確保するこ
27 と。

28 ● プライバシー

- 29 ☆ 取り扱うデータの重要性等に応じて、プライバシーを尊重し、適切に保護
30 すること。

31 ● 公正競争

- 32 ☆ 特定の者にA Iに関する資源が集中した場合においても、その有利な立場
33 を利用した不当なデータの収集を含む不公正な取引が行われないうにす
34 るなど、公正な競争の促進に貢献すること。

1 ● AIリテラシー

- 2 ☆ AIがもたらすリスクの社会的受容可能な水準は変わり得ることを認識し、
3 便益の最大化とリスクの抑制を図れるよう、知識・能力を身に付けるとと
4 もに、倫理観を保持すること。

5 ● イノベーション

- 6 ☆ 環境負荷の低減を含む持続可能性を確保しつつ、イノベーションの促進に
7 貢献するよう努めること。
8 ☆ 社会課題の解決に資するAIの技術開発に取り組むこと。
9 ☆ AIの活用を阻害する要因の改善を図ること。

10
11 (3) 適正性確保のための基本方針

12 (2)で示した考え方を踏まえ、適正性を確保するために取り組むべき基本方針は以下
13 のとおり。

14 ① リスクベースでのアプローチ

- 15 ・ AIがもたらすリスクを特定・評価し、AIを利用する分野や目的を踏まえた
16 影響度合いに応じて、適切な対策を講じる。

17 ② ステークホルダーの積極的な関与

- 18 ・ AIがもたらす便益、リスク等により影響を受ける主体（ステークホルダー）
19 がAIガバナンスに積極的に関与し、ステークホルダーと協働して課題解決に
20 取り組む。

21 ③ 一貫通貫でのAIガバナンスの構築

- 22 ・ AIがもたらすリスクをステークホルダーにとって受容可能な水準で管理しつ
23 つ、便益を最大化するため、研究開発から社会実装までが近接するAIの各段
24 階を一体的に捉えたAIガバナンスを構築。

25 ④ アジャイルな対応

- 26 ・ AIの技術進歩の速さや、予見可能性、説明可能性が十分でないことを踏まえ、
27 変動し得るリスクに対してPDCA（計画・実行・評価・改善）サイクルを回
28 しながら、柔軟・迅速（「アジャイル」）に対応し、AIガバナンスの成熟度を
29 高めていく。

2 研究開発機関及び活用事業者が特に取り組むべき事項

A I を活用した製品、サービスの開発、提供をする活用事業者（海外事業者も含む）は、その開発、提供したA I が多くの主体に影響を及ぼし得ることを踏まえ、国際的な規範、国際標準、各種ガイドライン等を活用しつつ、1(2)に示す適正性確保に必要なとなる主な要素に関して、特に以下の事項に取り組む。

また、A I の研究開発機関は、その開発したA I を第三者に提供する際は、1(2)に示す適正性確保に必要なとなる主な要素に関して、特に以下の事項に取り組む。

8

(1) A I ガバナンスによる俯瞰的な適正性の確保

- ・ A I の設計・開発・提供・実装等のライフサイクル全体で、リスクの特定・評価・対処をするための組織的なプロセス（経営層の関与したモニタリングや評価の仕組み、情報の適切な開示、教育・研修の実施等）を含むA I ガバナンスを構築・運用・継続改善。

14

(2) ステークホルダーとの信頼関係の構築に向けた透明性の確保

- ・ 学習データの出所と出力される生成物について、知的財産、プライバシー等の保護の適切な実施を含め、ステークホルダーとの信頼関係を構築するためにも、合理的な範囲で説明可能性を確保。
- ・ A I を提供する際、A I の適正な利用を可能にするための情報（A I の仕組み・限界、禁止事項、学習するデータの収集ポリシー、出力の信頼性に関する注意喚起等）をそのA I の利用者に提供。

22

(3) 十分な安全性の確保

- ・ A I を悪用したサイバー攻撃や詐欺をはじめとする各種犯罪等違法行為のリスクを特定・評価し、適切な対策を講じる。
- ・ ハルシネーションや偏見・差別の助長、偽・誤情報等（ディープフェイク技術によるフェイク動画、性的加工画像等）の拡散等につながるA I による不適切な出力の抑制、A I の意図しない動作や誤作動の防止をするため、最新の技術と知見を駆使して、解決、改善に向けて取り組む。

29

1 (4) 事業継続性確保による安全な環境の維持

- 2 ・ AIを用いたシステムの運用者やサービスの提供者は、これらに障害が生じた場
3 合に備え、損害を最小限にとどめ、早期復旧するために、平常時に行うべき活動
4 や緊急時の事業継続のための方法、手段等を定めた事業継続計画をあらかじめ策
5 定。

6

7 (5) AIのイノベーションの基盤となるデータの重要性を踏まえたステークホルダー
8 への配慮

- 9 ・ AIのイノベーションの基盤となる質の高いデータの確保と適正な活用が重要で
10 あることを踏まえ、データの整備と新たな創作活動等の好循環を実現するため、
11 知的財産等のデータ保有者等に対する利益還元のエコシステムや創作活動が促進
12 される環境の構築に向けた方策の検討、実施に努める。

1 3 国及び地方公共団体が特に取り組むべき事項

2 国は、1(2)に示す適正性確保に必要となる主な要素に関して、特に以下の事項に取り
3 組む。

4 地方公共団体は、1(2)に示す適正性確保に必要となる主な要素に関して、置かれた環
5 境や課題が多様であることを踏まえ、地域の実情に応じて、特に以下の事項に配慮しつ
6 つ、必要な対応を行う。

7 なお、A Iを開発、提供する際は、2に示す事項についても取り組む。

8
9 (1) A Iの積極的かつ先導的な活用によるイノベーションの促進

- 10 ・ 活用事例や留意点を周知することがA Iの普及促進に効果的であることを踏まえ、
11 国及び地方公共団体が自ら積極的かつ先導的にA Iの活用や公共調達における開
12 発実証機会の提供を進める。

13

14 (2) 社会全体におけるA Iリテラシーの向上

- 15 ・ 国及び地方公共団体は、A Iを適正に利用するための判断を全ての主体が適切に
16 できる環境を構築するため、国及び地方公共団体の職員を含めた社会全体におけ
17 るA Iリテラシーの向上を図ることが求められる。
- 18 ・ 常にA Iの最新の技術動向や活用実態を把握し、リスク及びその対応策を検討し
19 て、ステークホルダーの自主的な取組を促すための考え方を提示。
- 20 ・ 事業者、国民等におけるA Iの研究開発・活用における適正性確保に向けて、生
21 成A Iの基本的な使い方や注意点を学べるコンテンツの提供、社会人向けの生成
22 A Iスキル・知見の習得支援等、教育・ガイダンスを積極的に推進。

23

24 (3) A Iガバナンスの在り方の検討

- 25 ・ 国内外におけるA Iガバナンスを巡る動向を注視し、A Iガバナンスの在り方を
26 継続的に検討し、対応。また、A Iは国境を越えて展開されるため、国内だけ
27 なく、国際的なガバナンスが不可欠であることを踏まえ、A Iガバナンスの国際
28 協調を主導。

- 1 • 本指針及びガイドライン等も、A I の技術進歩による社会の変化をとらえ、アジ
2 ャイルかつ継続的に見直す。ガイドライン等は、本指針の趣旨と整合するよう、
3 適宜、点検・見直しを行う。
- 4 • 様々な局面におけるA I 導入の障壁を低減するため、A I を活用する際に想定・
5 発生し得る課題に対して、その責任の所在等に関する解釈適用上の論点及び考え
6 方を整理するとともに、判例等を踏まえ可能な限り解釈を明確化するよう努める。

7

8 (4) 行政としてのアカウントビリティを果たすこと

- 9 • 行政においてA I を利用する際、行政の信頼性を確保するため、求められる水準
10 を十分に考慮した適切なリスク対策等を徹底し、可能な限り判断の根拠が不明瞭
11 にならないよう国民へのアカウントビリティを果たす。
- 12 • 各府省庁において、A I ガバナンスの責任者を任命。地方公共団体において、A
13 I の適正な活用、リスク管理における責任者を明確化。

1 **4 国民が特に取り組むべき事項**

2 国民は、1(2)に示す適正性確保に必要な主要素に関して、以下に特に配慮した
3 対応を行う。

4

5 (1) 人間中心の原則に基づくA Iの責任ある利用

- 6 ・ 国民はA Iを利用する主体として、A Iの利用により法令への抵触、加害行為を
7 するおそれがあることを踏まえ、法令を遵守する。
- 8 ・ A Iの利便性のみならず、倫理、法令、人権、安全等に関する課題を理解し、責
9 任ある利用者としての自覚をもって行動するよう努める。

10

11 (2) A Iリテラシーに基づく適切な利用

- 12 ・ A Iの特性や仕組みを正しく理解し、能動的にA Iリテラシーを身に付けるよう
13 努める。
- 14 ・ A Iを利用する際、得られる情報の出所、正確性、生成過程等を理解し、人間の
15 判断、責任の下で意思決定を行うとともに、不当な差別、誹謗中傷、偽・誤情報
16 の拡散等を目的とした不適切な行為を行わない。
- 17 ・ A Iの生成物（文章、画像、音声、動画等）は、社会的、法的に適切な形で利用
18 する。