

AI事業者ガイドライン案 (第1.0版) 別添 (付属資料) 概要

総務省
経済産業省
(令和6年4月)

別添（付属資料）の位置づけ

- 本編では、事業者がAIの安全安心な活用を行い、AIの便益を最大化するために重要な「どのような社会を目指すのか（基本理念=why）」及び「どのような取組を行うか（指針=what）」を示した
- 別添（付属資料）では、「具体的にどのようなアプローチで取り組むか（実践=how）」を示すことで、事業者の具体的な行動へとつなげることを想定している

本編（why, what）

別添（付属資料）(how)



どのような社会を
目指すのか
(基本理念=why)



どのような取組を
行うか
(指針=what)



どのようなアプローチで
取り組むか
(実践=how)

「AI事業者ガイドライン」の構成

- 別添の記載内容は本編と対応しており、本編の読解及びそれに基づく検討や行動をサポートする解説書としての役割を果たす

	本編 (why, what)	別添 (付属資料) (how)
主体共通	第1部 AIとは	1. 第1部関連 [AIについて] A. AIに関する前提 B. AIによる便益/リスク
	第2部 AIにより 目指すべき社会 及び各主体が取り 組む事項 A.「基本理念」 B.「原則」 C.「共通の指針」 D.「高度なAIシステムに関する 事業者に通の指針」 E.「AIガバナンスの構築」	2. 第2部関連 [E.AIガバナンスの 構築] A. 経営層によるAIガバナンスの構築及び モニタリング B. AIガバナンスの事業者取組事例
主体別	第3部 AI開発者に 関する事項 ※「高度なAIシステムを開発する組織向けの 広島プロセス国際行動規範」における 追加的な記載事項 も含む	3. 第3部関連 [AI開発者向け] A. 「第3部 AI開発者に関する事項」の解説 B. 「第2部」の「共通の指針」の解説 C. 高度なAIシステムの開発にあたって遵守 すべき事項
	第4部 AI提供者に 関する事項	4. 第4部関連 [AI提供者向け] A. 「第4部 AI提供者に関する事項」の解説 B. 「第2部」の「共通の指針」の解説
	第5部 AI利用者 に関する事項	5. 第5部関連 [AI利用者向け] A. 「第5部 AI利用者に関する事項」の解説 B. 「第2部」の「共通の指針」の解説
その他 参考資料		6. 「AI・データの利用に関する契約ガイドライン」を参照 する際の主な留意事項について 7. チェックリスト 8. 主体横断的な仮想事例 9. 海外ガイドライン等の参照先

別添1. 第1部関連 記載内容

- AIシステムの例、AIシステム・サービスの具体的な活用例、AI事業者のパターン、業種、業務ごとのAIによる便益の例及び実例に基づくリスクの例を掲載
- これらの解説を通じ、本ガイドラインの記載内容のより深い理解につなげる

A. AIに関する前提

- AIの学習及び利用の流れ
 - 一般的なAIの学習・利用の流れの把握
- AIシステム概要
 - AIシステムのスコープ
- AIの開発から利用までのバリューチェーン
 - 一般的なAI活用の流れにおける主体の対応
- AIシステム・サービスの例
 - 代表的なAIシステム・サービス例とそれにかかわる各主体を具体化
- AI事業者のパターン
 - 事業活用時のAIバリューチェーンを具体化
- データ提供者について
 - 本ガイドラインでは対象外のデータ提供者の定義等

B. AIによる便益/リスク

- AIによる便益
 - 主に利益を享受する最終利用者に焦点を当てて記載
- AIによるリスク
 - 代表的な事例（想定を含む）を記載（「共通の指針」と紐付けた記載を行うことで、本編との関係性も明確化）
(生成AIを活用する際のリスクも含む)

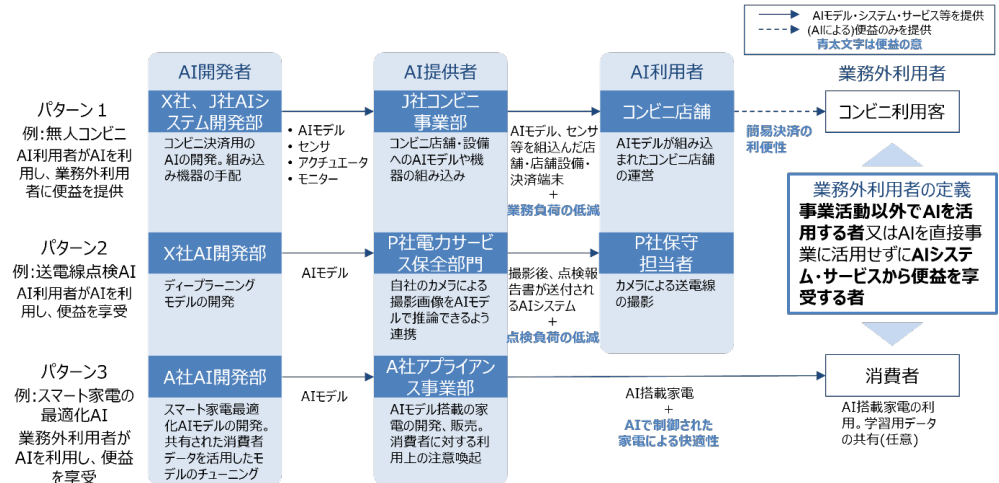
別添1 A. AIに関する前提 主な記載内容

- 本編での各主体・定義等を具体化するAIそのもの、活用場面及び各主体の役割の更なる理解につなげる
 - 主なAIシステム・サービスを例示しつつ、関連する主体を具体化する
 - ユースケース毎に、プロセスと各主体 (AI事業者) の役割及びパターンを整理する

AIシステム・サービスの例 (抜粋)

ケース名	活用 AI	概要	AI 開発者	AI 提供者	AI 利用者	業務外 利用者
採用 A I	テキスト解析	A 社グループのグローバル各社における人材採用部門が、 エントリーシートの書類選考を判断する際の参考情報として使用される AI サービス である。 A 社 AI 開発部門は、AI 利用者である A 社人材採用部門 (海外グループ企業を含む) より過去のエントリーシートデータ及び合否判定 (内定の判定) 結果を受領し、機械学習 (分類モデル) で合否判定を支援する AI モデルを作成している。	A 社 (開発部門)	A 社 (システム部門、人材開発部門)	A 社グループ (人材採用部門)	採用申込者
無人コンビニ	画像解析	全国のコンビニエンスストアチェーンを運営する J 社が提供する 画像認識 AI を活用した無人コンビニ (店内の客が商品を取るだけで AI が代金を計算し、店外に出る際に電子マネー等で一括決済ができるコンビニ) である。当 AI サービスには X 社で開発された無人コンビニ向けの AI システムを搭載している。	X 社	J 社 (AI システム開発部及びコンビニ事業部)	コンビニ店舗	コンビニ利用客
がん診断 AI	テキスト・画像解析	マルチモーダル学習を使用しており、「 本人の病歴・遺伝等に係る情報 (データ 1) 」及び「 内視鏡画像 (データ 2) 」を取込み、 内視鏡での診察中にリアルタイムにがんの可能性が高い部分をハイライト する。医師は出力画像を観察して、がんの可能性があるか判断する。 A 社が AI を開発しつつ、がん診断 AI システムを医療機関に提供している。	A 社 (AI 開発部門)	A 社 (医療 IT サービス部門)	医療機関 (システム部門及び消化器内科)	受診患者

AI事業者のパターン



別添1 B. AIによる便益/リスク（AIによる便益）

主な記載内容

- 便益を享受する最終利用者に焦点を当ててAIによる便益を整理し、その理解促進につなげる

	開発	マーケティング	販売	物流・流通	顧客対応	法務	ファイナンス	人事
従来から存在する便益の例	コード検証、ドキュメント作成の自動化	広告用メールの自動配信	受注後の対応メール等の自動発信	需要予測に基づく生産・在庫数最適化	チャットボットによる自動対応	翻訳	財務諸表の自動作成	給与計算等の自動化
（生成AIで更に向上）								
	類似コード・データの抽出・検証	データに基づいたパーソナライゼーション広告	チャネル別、ニーズ別の売上予測	配送ルート最適化	過去の問合せ内容に基づいたFAQ作成	法務文章のレビュー	過去実績にもとづいた将来予測、不正検知	職務経歴書等に基づいた人材需要マッチング
生成AI特有の便益の例	学習データの生成、コーディングアシスタント、新製品のブレインストーミング	販売促進（マーケティング素材・キャッチコピー等）の自動作成	営業トークスクリプトの自動作成	物流条件交渉のアシスタント	対応内容の自動生成、要約	規定に基づいた契約書ドラフトの自動生成	文脈を踏まえた上での社内問合せ対応	文脈を踏まえた上での人事面接の対応

別添1 B. AIによる便益/リスク（AIによるリスク）

主な記載内容

- 従来型のAIからのリスク及び生成AIで特に顕在化したリスクについて、主なものを整理
- AIによるリスクを知ること、AIガバナンスの必要性の理解を深める

	リスク	事例	対応する「共通の指針」
従来型AIから存在するリスク	バイアスのある結果及び差別的な結果の出力	<ul style="list-style-type: none"> IT企業が自社で開発したAI人材採用システムが女性を差別するという機械学習面の欠陥を持ち合わせていた 	1) 人間中心 3) 公平性
	フィルターバブル及びエコーチェンバー現象	<ul style="list-style-type: none"> SNS等によるレコメンドを通じた社会の分断が生じている 	1) 人間中心
	多様性の喪失	<ul style="list-style-type: none"> 社会全体が同じモデルを、同じ温度感で使った場合、導かれる意見及び回答がLLMによって収束してしまい、多様性が失われる可能性がある 	1) 人間中心
	不適切な個人情報の取扱い	<ul style="list-style-type: none"> 透明性を欠く個人情報の利用及び個人情報の政治利用も問題視されている 	1) 人間中心 4) プライバシー保護
	生命、身体、財産の侵害	<ul style="list-style-type: none"> AIが不適切な判断を下すことで、自動運転車が事故を引き起こし、生命や財産に深刻な損害を与える可能性がある トリアージにおいては、AIが順位を決定する際に倫理的なバイアスを持つことで、公平性の喪失等が生じる可能性がある 	2) 安全性 3) 公平性
	データ汚染攻撃	<ul style="list-style-type: none"> AIの学習実施時及びサービス運用時には学習データへの不正データ混入、サービス運用時ではアプリケーション自体を狙ったサイバー攻撃等のリスクが存在する 	5) セキュリティ確保
	ブラックボックス化、判断に関する説明の要求	<ul style="list-style-type: none"> AIの判断のブラックボックス化に起因する問題も生じている AIの判断に関する透明性を求める動きも上がっている 	6) 透明性 7) アカウンタビリティ
	エネルギー使用量及び環境の負荷	<ul style="list-style-type: none"> AIの利用拡大により、計算リソースの需要も拡大しており、結果として、データセンターが増大しエネルギー使用量の増加が懸念されている 	1) 人間中心
生成AIで特に顕在化したリスク	悪用	<ul style="list-style-type: none"> AIの詐欺目的での利用も問題視されている 	2) 安全性 8) 教育・リテラシー
	機密情報の流出	<ul style="list-style-type: none"> AIの利用においては、個人情報や機密情報がプロンプトとして入力され、そのAIからの出力等を通じて流出してしまうリスクがある 	5) セキュリティ確保 8) 教育・リテラシー
	ハルシネーション	<ul style="list-style-type: none"> 生成AIが事実と異なることをもつもらしく回答する「ハルシネーション」に関してはAI開発者・提供者への訴訟も起きている 	2) 安全性 8) 教育・リテラシー
	偽情報、誤情報を鵜呑みにすること	<ul style="list-style-type: none"> 生成AIが生み出す誤情報を鵜呑みにすることがリスクとなりうる ディープフェイクは、各国で悪用例が相次いでいる 	1) 人間中心 8) 教育・リテラシー
	著作権との関係	<ul style="list-style-type: none"> 知的財産権の取扱いへの議論が提起されている 	2) 安全性
	資格等との関係	<ul style="list-style-type: none"> 生成AIの活用を通じた業法免許や資格等の侵害リスクも考える 	2) 安全性
	バイアスの再生成	<ul style="list-style-type: none"> 生成AIは既存の情報に基づいて回答を作るため既存の情報に含まれる偏見を増幅し、不公平や差別的な出力が継続/拡大する可能性がある 	3) 公平性

別添2. 第2部 [E.AIガバナンスの構築]関連

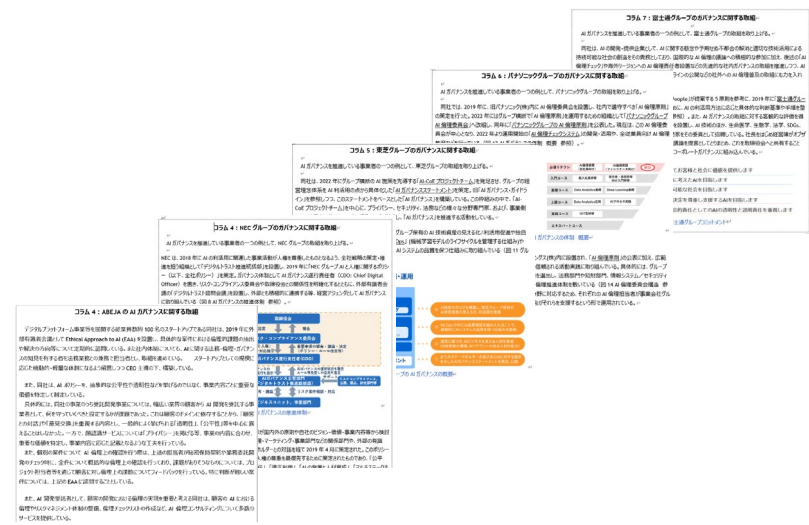
- 別添2では、AIガバナンスの構築のための「行動目標」、「実践のポイント」及びそれに対応する仮想的な「実践例」、実際の企業の取組事例を掲載
- 各事業者にとって、自身のAIガバナンスの構築のための検討を、具体例を交えつつ行うことを可能とする

A.経営層によるAIガバナンスの構築及びモニタリング

- 行動目標
 - 一般的かつ客観的な目標を記載
- 実践のポイント
 - 「AI原則実践のためのガバナンスガイドライン Ver. 1.1」を基に、国内外ガイドラインやISO等の要素も取り入れ
- 実践例
 - 仮想事例に基づく事例を記載
 - 生成AI等の最新動向への対応事例も織り込み

B.AIガバナンスの構築に関する実際の取組事例

- 実際の取組事例
 - 「AI原則実践のためのガバナンスガイドライン Ver. 1.1」に基づくAIガバナンスの取組事例について、5社分をコラム化
 - AIガバナンスを推進するにあたり、多くの企業がつまづく観点を記載



主な記載内容

- 各事業者が、ガバナンスの行動目標の意義を理解し活用することで、AIガバナンスの構築をサポートする

分類	行動目標 ※「3-1-1」のように更に細分化されているものもあり
1. 環境・リスク分析	1-1 便益/リスクの理解 1-2 AIの社会的な受容の理解 1-3 自社のAI習熟度の理解
2. ゴール設定	2-1 AIガバナンス・ゴールの設定
3. システムデザイン	3-1 ゴールと乖離の評価及び乖離対応の必須化 3-2 AIマネジメントの人材のリテラシー向上 3-3 各主体間・部門間の協力によるAIマネジメント強化 3-4 予防・早期対応による利用者のインシデント関連の負担軽減
4. 運用	4-1 AIマネジメントシステム運用状況の説明可能な状態の確保 4-2 個々のAIシステム運用状況の説明可能な状態の確保 4-3 AIガバナンスの実践状況の積極的な開示の検討
5. 評価	5-1 AIマネジメントシステムの機能の検証 5-2 社外ステークホルダーの意見の検討
6. 環境・リスクの再分析	6-1 行動目標1-1～1-3の適時の再実施

別添2 A. 経営層によるAIガバナンスの構築とモニタリング 構成

- 各「行動目標」に対し、「実践のポイント」及び「実践例」を整理
- 採用するAIの種類やリスクの程度に応じて各事業者が参照することで、AIガバナンスの検討を可能に

別添 記載内容

解説

行動目標1-1【便益/リスクの理解】:

各主体は、経営層のリーダーシップの下、AIの開発・提供・利用の目的を明確化したうえで、AIから得られる便益だけではなく意図しないリスクがあることについて、各主体の事業に照らして具体的に理解し、これらを経営層に報告し、経営層で共有し、適時に理解を更新する。

行動目標

- 事業者が取り組むことが重要となる一般的かつ客観的な目標を提示
- 各事業者の方針検討の際の材料となる

(実践のポイント)

[実践のポイント]

各主体は、経営層のリーダーシップの下、以下に取り組む。

- 事業における価値の創出、社会課題の解決等のAIの開発・提供・利用の目的を明確に定義
- 自社の事業に結びつく形で、「便益」及び意図せざるものを含めた「リスク」を具体的に理解
- その際に、回避すべき「リスク」及び複数主体にまたがる論点に留意し、バリューチェーン/リスクチェーン全体で便益を確保、リスクを削減
- 迅速に経営層に報告/共有する仕組みを構築

実践のポイント

- 上記の行動目標の実行のために重要となる事項や留意点を要約
- 各事業者が具体的な取組内容を検討する際の材料となる

[実践例]

【実践例 i : 便益・リスクの把握】

各主体は、経営層のリーダーシップの下（担当役員又は現場に一任するのではなく、経営層自らが主導することを通じて実施することも含む、以下同様）、便益だけではなくリスクについても検討し、

実践例

- 仮想的な実践例を記載
- 具体的な取組イメージを持つことで、各事業者が行動につなげやすくする

別添3～5. 各主体向け 主な記載内容

別添3～5

- 別添3～5では、各主体向けに、本編の詳細な解説を掲載
- 本編と合わせて活用することで、本編の内容に関する具体的なアプローチを参照し、検討可能

A. 本編「第3～5部」の解説

B. 本編「第2部」の 「共通の指針」の解説

C. 高度なAIシステムの開発に あたって遵守すべき事項 ※別添3. AI開発者向けのみ

• ポイント

- 本編記載事項に加え、重要となる観点を補足

• 具体的な手法

- 他のガイドライン等を参照しつつ、具体的に解説

• 参考文献

- 参考となる他の文献を記載

(例)

- デジタル庁「データ品質ガイドブック（B版）」（2021年6月）
- 国立研究開発法人産業技術総合研究所「機械学習品質マネジメントガイドライン 第4版」（2023年12月）
- AIプロダクト品質保証コンソーシアム「AIプロダクト品質保証ガイドライン」（2023年6月）
- NIST, “AI Risk Management Framework Playbook”（2023年1月）

- 高度なAIシステムを開発する組織向けの広島プロセス国際行動規範

別添3~5. 各主体向け 構成

- 各主体の重要事項に対して、「ポイント」、「具体的な手法」、「参考資料」を以下の構成で掲載

別添 記載内容

解説

A. 本編「第3部 AI開発者に関する事項」の解説

[本編の記載内容（再掲）]

データ前処理・学習時

D-2) i. 適切なデータの学習

Article I. プライバシー・バイ・デザイン等を通じて、学習時のデータについて、適正に収集するとともに、第三者の個人情報、知的財産権に留意が必要なもの等が含まれている場合には、法令に従って適切に扱うことを、AI のライフサイクル全体を通じて確保する（「2）安全性」、「4）プライバシー保護」、「5）セキュリティ確保」）

Article II. 学習前・学習全体を通じて、データのアクセスを管理するデータ管理・制限機能の導入検討を行う等、適切な保護措置を実施する（「2）安全性」、「5）セキュリティ確保」）

[ポイント]

- AIモデルの質の向上のために、AI開発者は、AIの学習等に用いるデータの質に留意することが重要となる。
- 利用するAIの特性及び用途を踏まえ、AIの学習等に用いるデータの質（正確性及び完全性等）に留意する
- また、AIによりなされる判断は、事後的に精度が損なわれたり、低下したりすることが想定されるため、想定される権利侵害の規模、権利侵害の生じる頻度、適用できる技術水準、精度を維持するためのコスト等を踏まえ、あらかじめ精度に関する基準を定めておくことが期待される。精度が当該基準を下回った場合には、データの質に留意して改めて学習させる
- なお、ここで言う「精度」には、AIが倫理的に正しい判断を行っているか（例えば、AIが暴力的な表現を行っていないか、ヘイトスピーチ等を行っていないか等）も含まれる

[具体的な手法]

- データに個人情報、機密情報、著作権等の権利又は法律上保護される利益に係るものが含まれていないか、確認を実施
 - 固有表現抽出

[参考文献]

- 国立研究開発法人産業技術総合研究所「機械学習品質マネジメントガイドライン 第4版」（2023年12月）
- NIST, “AI Risk Management Framework Playbook”（2023年1月）

本編の記載内容を再掲

ポイント

- 本編記載事項に加え、重要観点を補足

具体的な手法

- 他のガイドライン等を参照しつつ具体的に解説

参考文献

- 「ポイント」や「具体的な手法」の参照元を記載

本編の
第3~5部に
対応する解説

本編 第2部に
対応する解説

「共通の指針」に
関する各主体の
具体的なアプローチ

- 別添6では、2018年6月に初版が策定・公表された「AI・データの利用に関する契約ガイドライン」について、AIの開発・利用に関する状況の変化及び新たな技術の進歩に伴う、2024年1月時点での参照時の留意点を掲載
- 契約を通じて当事者間の権利及び義務を明確に定めることで、AIに関する取引を円滑に進め、これらに伴う無用な紛争を予防

AI・データの利用に関する契約ガイドライン



諸外国の動向

新技術の台頭



別添6「AI・データの利用に関する契約ガイドライン」を参照する際の主な留意事項

契約ガイドライン公表後の状況の変化を考慮すべき事項として、以下の内容等について記載

- (1) AIの開発と利用の概念について
 - 開発と利用の類型に収まらない取引が生じていることへの留意が必要
- (2) AIの開発・提供・利用と責任の分配について
 - バリューチェーンの多様化・複雑化に伴う個別の状況に応じた責任分配の在り方の検討
- (3) AIの開発・提供・利用とアカウントビリティについて
 - リスクが生じた際に一定の説明を求められる可能性があり、その説明の合理性確保のための契約締結前後での方針の整理

別添7. チェックリスト 概要

- 別添7では、AIによるリスクを抑えつつ便益を享受する取組の立案、実践を確実に推進するための「チェックリスト」及び「具体的なアプローチのためのワークシート」を用意している

本編・
別添1～5（付属資料）



本編・別添を読んでAIガバナンスの重要性や、各事業者に期待されることを理解する

チェックリスト



「チェックリスト」を活用し、本編・別添についての各主体の取組(What)を確認する

具体的なアプローチ検討の
ためのワークシート



「具体的なアプローチ検討のためのワークシート」を使用し、各事業者の具体的なアプローチ(How)を検討する

別添7. チェックリスト 活用方法

- 全事業者は、「別添7 Aチェックリスト[全主体向け]」を活用し、各自の取組状況の概観を確認する
- 高度なAIシステムに関係する事業者に該当する場合には、「別添7 Bチェックリスト[高度なAIシステムに関係する事業者向け]」も実施する

別添7 A **チェックリスト** 令和6年4月
[全主体向け]

本チェックリストは、AI事業者ガイドライン「第2部C. 共通の指針」を要約したものです。事業者に求められる重要な取組事項のチェックにご活用ください
 ※高度なAIシステムに関係する事業者は、「チェックリスト[別添7 B. 高度なAIシステムに関係する事業者向け]」も実施ください

別添7 B **チェックリスト** 令和6年4月
[高度なAIシステムに関係する事業者向け]

本チェックリストは、AI事業者ガイドライン「第2部D. 高度なAIシステムに関係する事業者に共通の指針」を要約したものです。高度なAIシステムに関係する事業者に求められる重要な取組事項のチェックにご活用ください(高度なAIシステムを開発するAI開発者にも適用される内容もあるため、AI提供者及びAI利用者は適切な範囲でご活用ください)

チェック項目	チェック項目
<input type="checkbox"/> 人間中心の考え方を基に、憲法が保障する又は国際的に認められた人権を侵すことがないようにしているか？	<input type="checkbox"/> 高度なAIシステムの市場導入前及び開発全体を通じて、AIライフサイクル全体にわたるリスクを特定、評価、軽減するための適切な措置を講じているか？
<input type="checkbox"/> AIに関わる全ての者の生命・身体・財産、精神及び環境に危害を及ぼすことがないよう 安全性 を確保しているか？	<input type="checkbox"/> 市場投入後に脆弱性、インシデント、悪用パターンを特定し、緩和しているか？
<input type="checkbox"/> 潜在的なバイアスをなくすよう留意し、それでも回避できないバイアスがあることを認識しつつ、回避できないバイアスが人権及び多様な文化を尊重する 公平性 の観点から許容可能か評価しているか？	<input type="checkbox"/> 十分な透明性の確保やアカウンタビリティの向上のため高度なAIシステムの 能力、限界、適切・不適切な使用領域を公表 しているか？
<input type="checkbox"/> プライバシー を尊重・保護し、関係法令を遵守しているか？	<input type="checkbox"/> 産業界、政府、市民社会、学界を含む関係組織間で、 責任ある情報共有とインシデントの報告 に努めているか？
<input type="checkbox"/> 不正操作によってAIの振る舞いに意図せぬ変更又は停止が生じることのないように、 セキュリティ を確保しているか？	<input type="checkbox"/> リスクベースのアプローチにもとづきAIガバナンス及びリスク管理方針を策定、実施、開示しているか？
<input type="checkbox"/> 透明性 を確保するために、AI自体やAIシステム・サービスの情報をステークホルダーに対し合理的で技術的に可能な範囲で提供しているか？	<input type="checkbox"/> AIのライフサイクル全体にわたり、 物理的セキュリティ、サイバーセキュリティ、内部脅威を含む、強固なセキュリティ管理に投資し、実施 しているか？
<input type="checkbox"/> データの出所、AIの意思決定等のトレーサビリティに関する情報やリスクへの対応状況等について、関連するステークホルダーに対して合理的な範囲で アカウンタビリティ を果たしているか？	<input type="checkbox"/> 技術的に可能な場合は、AIが生成したコンテンツを識別できるように、 電子透かしやその他の技術等、信頼性の高いコンテンツ認証及び来歴のメカニズムを開発、導入 しているか？
<input type="checkbox"/> AIガバナンスやプライバシーに関するポリシー 等を策定しているか？	<input type="checkbox"/> 社会的、安全、セキュリティ上の リスクを軽減するための研究を優先し、効果的な軽減策に優先的に投資 しているか？
<input type="checkbox"/> 上記の実現のため、各事業者の状況に応じた 具体的なアプローチ は検討しているか？	<input type="checkbox"/> 気候変動、健康・教育などの世界の最大の課題 に対処するため、 高度なAIシステムの開発を優先 しているか？
	<input type="checkbox"/> 国際的な 技術規格の開発を推進 しているか？
	<input type="checkbox"/> 適切なデータ入力措置と、個人データ及び知的財産の保護 を実施しているか？
	<input type="checkbox"/> 誤情報の拡散等のAI固有リスクに関する デジタルリテラシーの向上や脆弱性の検知への協力と情報共有等 、高度なAIシステムの 信頼でき責任ある利用 を促進し、貢献しているか？
	<input type="checkbox"/> 上記の実現のため、各事業者の状況に応じた 具体的なアプローチ は検討しているか？

検討には「**具体的なアプローチ検討のためのワークシート**」をご活用ください

検討には「**具体的なアプローチ検討のためのワークシート**」をご活用ください

• チェック項目は本編の要約を記載

• チェックすることで、各自の取組状況を概観

• 具体的な実践内容の検討に、「別添7C 具体的なアプローチのためのワークシート」を活用(活用方法次頁)

別添7. 具体的なアプローチ検討のためのワークシート 活用方法

- 本ガイドラインの記載内容に関して、具体的なアプローチを検討する際に重要となる事項を記載
- 事業者の事業内容及び置かれた状況等に応じ、各自でカスタマイズして活用することを前提としている

別添7C. 具体的なアプローチ検討のためのワークシート (共通の指針関連) 令和6年4月

利用上の留意点 ガイドラインに記載した内容に関して取り組むべき事項は、各事業者の事業内容や置かれた状況等により、個々に異なります。このため、本ワークシートは、あくまで各事業者が取り組むべき事項が何かを検討する際の材料をご提供するものであり、各事業者それぞれの状況に成り、カスタマイズして必要に応じて活用いただくことを前提としたものです。したがって、必ずしも、全ての事項について、検討が必要となるものではありませんので、活用の要否、各自の事情に応じた修正や取捨選択を検討ください。D列、E列の記載内容を基に、F列以降を各事業者にてご検討いただき、各自が取り組むこと（あるいは内容）を具体化の上、活用ください。

各自用にカスタマイズし運用するには、α. 取組内容を作成する者（当該ワークシートを基に、各自の取組内容のカスタマイズを行う者）、β. 実施状況の確認を行う者（現場において実際の確認を行う者）γ. 責任者（確認内容に対して責任を負う者）を特定ください（各主体の規模によって、α～γが重複する場合もある）。

α. ワークシートの作成者

β. 実施状況の確認を行う者

γ. 責任者

対応箇所	分類	検討にあたって重要な事項	各自の事業において検討対象とする事項 (該当しない場合はその理由)	他の主体との関係についての事項	✓	具体的なアプローチ	最終検討日 (見直し日)
1) 人間中心	1)① 人間の尊厳及び個人の自律	a. AIが活用される際の社会的文脈を踏まえ、人間の尊厳及び個人の自律を尊重しているか？ b. 特に、AIを人間の脳・身体と連携させる場合には、その周辺技術に関する情報を踏まえつつ、諸外国及び研究機関における生命倫理の議論等を参照しているか？ c. 個人の権利・利益に重要な影響を及ぼす可能性のある分野においてAIを利用したプロファイリングを行う場合、個人の尊厳を尊重し、アウトプットの正確性を可能な限り維持させつつ、AIの予測、推奨、判断等の限界を理解して利用し、かつ生じうる不利益等を慎重に検討した上で、不適切な目的に利用していないか？					
	1)② AIによる意思決定・感情の操作等への留意	a. 人間の意思決定、認知等、感情を不当に操作することを目的とした、又は意識的に知覚できないレベルでの操作を前提としたAIシステム・サービスの開発・提供・利用は行っていないか？ b. AIシステム・サービスの開発・提供・利用において、自動化バイアス等のAIに過度に依存するリスクに注意を払い、必要な対策を講じているか？ c. フィルターバブルに代表されるような情報又は価値観の傾斜を助長し、AI利用者を含む人間が本来得られるべき選択肢が不本意に制限されるようなAIの活用にも注意を払っているか？					

• 活用の前に、実施責任者等を決定する

• 事業者の事業内容及び置かれた状況等に応じ、各自でカスタマイズして活用する

• 「見直し日」も検討・記載することで定期的な更新を行う

• 具体のアプローチを検討する際に重要になる事項が記載されており、事業者が各自でカスタマイズする際のリファレンスとなる

別添8. 主体横断的な仮想事例

概要

- 別添8.では、本ガイドラインに沿って、AI開発者、AI提供者、AI利用者が重要事項の検討を行った場合の「主体横断的な仮想事例」を掲載
- 各主体が本ガイドラインの内容を実際に落とし込む際の具体的なイメージの想起や、各主体間での連携が重要になるポイントの明確化が可能
- 現在「採用AI」を扱う事業者を例として取り上げているが、今後、他の事例も追加していく予定

Case 採用AI		機械学習モデル：XGBoost（エントリーシートの文章で、応募者に対して可否を判断する） https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2022/10/RCModel_Case01_Recruitment-AI_JP.pdf		
		AI開発部門	人材採用部門（採用AIチーム）	人材採用担当者
		AI開発者	AI提供者	AI利用者
No	分類	共通の指針/各主体に関する事項	本UCにおいて主体が実施している活動	本UCにおいて主体が実施している活動
1) 人間中心				
各主体は、AIシステム・サービスの開発・提供・利用において、後述する各事項を含む全ての取り組むべき事項が導出される土台として、少なくとも憲法が保障する又は国際的に認められた人権を侵すことがないようにすべきである。また、AIが人々の能力を拡張し、多様				
①人間の尊厳及び個人の自律				
1	共通	AIが活用される際の社会的文脈を踏まえ、人間の尊厳及び個人の自律を尊重する	AIシステムの開発において、学習データの収集やラベリング、モデルの性能評価等は、AI開発者だけで完結せず、AI提供者側で	AIサービスの提供において、AI利用者が最終判断(応募者の可否)を行えるようになっている(Human-in-the-loop)
2	共通	特に、AIを人間の脳・身体と連携させる場合には、その周辺技術に関する情報を踏まえつつ、諸外国及び研究機関における生命	脳・身体と連携するケースではないため対象外	脳・身体と連携するケースではないため対象外
3	共通	個人の権利・利益に重要な影響を及ぼす可能性のある分野においてAIを利用したプロファイリングを行う場合、個人の尊厳を尊重し、アウトプットの正確性を可能な限り維持させつつ、AIの予測、推奨、判断等の限界を理解して利用し、かつ生じうる不利益等を慎重に検討した上で、不適切な目的に利用しない	AIシステムの開発において、実際の予測結果を学習データに用いる際には個人情報の取扱いに関わる誓約書の締結やアクセス権管理等を実施している。 ※公平性とプライバシーについては、「3）公平性」「4）プライバシー保護」を参照	AIシステムの開発において、実際の予測結果を学習データに用いる際には個人情報の取扱いに関わる誓約書の締結やアクセス権管理等を実施している。 ※公平性とプライバシーについては、「3）公平性」「4）プライバシー保護」を参照
②AIによる意思決定・感情の操作等への留意				
1	共通	人間の意思決定、認知等、感情を不当に操作することを目的とした、又は意識的に知覚できないレベルでの操作を前提としたAIシステム・サービスの開発・提供・利用は行わない	本ケースに関しては、2)①-3と同じ論点になる	本ケースに関しては、2)①-3と同じ論点になる
2	共通	AIシステムの開発・提供・利用において、自動化バイアス等のAIに過度に依存するリスクに注意を払い、必要な対策を講じる	本ケースに関しては、2)①-3と同じ論点になる	本ケースに関しては、2)①-3と同じ論点になる

別添9. 海外ガイドライン等の参照先 概要

別添9

- 別添9.では、本ガイドラインにおいて、他ガイドラインのどこを参照したかを整理している
- 事業者が、より詳細な内容を検討する際、他の (国内外の) ガイドライン等と本ガイドラインとの対応関係を明確化するための材料とする

参考とした主なガイドライン等

- Ethics guidelines for trustworthy AI (信頼性を備えたAIのための倫理ガイドライン) (2019年4月、EU)
- Recommendation of the Council on Artificial Intelligence (人工知能に関する理事会勧告) (2019年5月、OECD閣僚理事会)
- Recommendation on the Ethics of Artificial Intelligence (人工知能の倫理に関する勧告) (2021年11月、UNESCO)
- Blueprint for an AI Bill of Rights (AI権利章典の青写真) (2022年10月、THE WHITE HOUSE)
- Artificial Intelligence Risk Management Framework (AI RMF 1.0) (2023年1月、NIST)
- Advancing accountability in AI (2023年2月、OECD)
- Proposal for a Regulation laying down harmonised rules on artificial intelligence (AI Act) (2023年6月、EU)
- AIプロダクト品質保証ガイドライン (2023年6月、AIプロダクト品質保証コンソーシアム)
- 機械学習品質マネジメントガイドライン 第4版 (2023年12月、国立研究開発法人産業技術総合研究所)

本ガイドラインに引用、統合しているガイドライン等

- 人間中心のAI社会原則 (2019年3月、統合イノベーション戦略推進会議)
- 国際的な議論のためのAI開発ガイドライン (2017年7月、AIネットワーク社会推進会議)
- AI利活用ガイドライン～AI利活用のためのプラクティカルリファレンス～ (2019年8月、AIネットワーク社会推進会議)
- AI原則実践のためのガバナンス・ガイドライン Ver. 1.1 (2022年1月、AI原則の実践の在り方に関する検討会)
- AI・データの利用に関する契約ガイドライン (2018年3月、経済産業省)
- 広島AIプロセス包括的政策枠組み (2023年12月、G7)