【資料２】

AI事業者ガイドラインの
令和７年度更新内容

総務省
経済産業省
(令和8年3月12日)

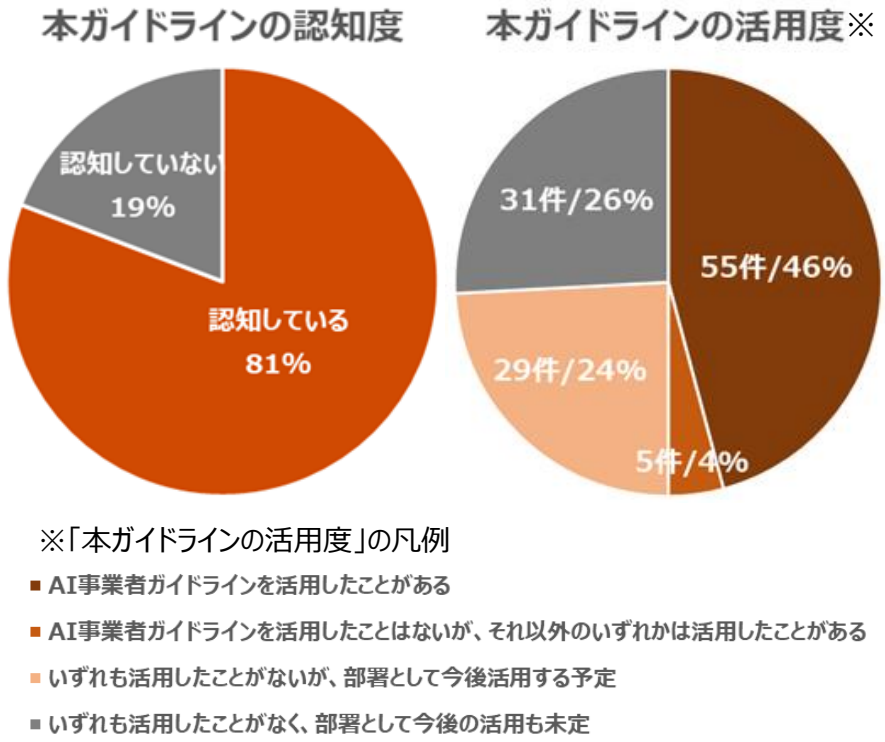
0. AI事業者ガイドライン更新の背景

- 2025年3月に「AI事業者ガイドライン（第1.1版）」を策定・公表
- AIネットワーク社会推進会議（総務省）、AIガバナンス検討会（総務省）、AI事業者ガイドライン検討会（経済産業省）の構成員の皆様よりいただいた御意見、事業者アンケートの結果、その他AIガバナンスの動向調査などを通じて、AI事業者ガイドラインの更新の論点・方針を整理

構成員/委員の皆様からの主な御意見

- **AIエージェントやフィジカルAI**の技術発展や社会浸透により、AIリスクの顕在化事例の増加や新たに考慮すべきリスク(社会変容等)が現れる中、ガイドライン上において、**各プレイヤーが具体的に対処すべきAIリスクや留意事項を明示してはどうか**
- 具体的にどのような**リスク**があり、どのような**対策**を行えばよいのかをいくつか**例示**することで、事業者が実効的なリスク対策を検討できるようになるのではないか
- 現在の**主体区分の概念が、現実の社会の影響を反映したものになっているか**どうか、見直す必要があるのではないか
- 学習、データ種類等の**多義的な用語に関して、読者の誤解を招かないように整理が必要ではないか**
- AI事業者ガイドラインのボリュームが増える中、地方自治体や小規模事業者等、**AIガバナンスの構築・実践をこれから始める方々の活用も見据えたバランスが重要**

事業者アンケート



政府や事業者におけるAIガバナンスの動向

1. AI事業者ガイドラインの令和７年度の更新の論点と更新方針

構成員・委員・事業者等からのご意見を踏まえ、令和７年度の更新の論点と更新方針を以下に整理

令和7年度更新の論点及び更新方針（案） 一覧

総務省検討会：AIネットワーク社会推進会議・AIガバナンス検討会
経済産業省検討会：AI事業者ガイドライン検討会

#	更新の論点	主なご意見	更新方針
1	**AI技術の動向の反映**	総務省検討会	**AIエージェント、フィジカルAIに関する事項の追記** ✓ AI事業者ガイドラインとしての定義の追加 ✓ 便益の追加 ✓ リスクに関する事項の追加 ✓ 留意すべき事項の追加 ✓ AIシステム・サービス例の追加
2	**AIによるリスクの記載の見直し**	総務省検討会	**AIによるリスクの記載の見直し** ✓ リスクベースアプローチに関する内容の追記 ✓ リスクの更新 ✓ 一部リスクの分類見直し
3	**主体区分の整理**	経済産業省検討会	**各主体区分の役割に関する補足の追加や図表の更新** ✓ AI開発者の定義の補足 ✓「一般的なAI活用の流れにおける主体の対応」の見直し ✓ 主体毎の役割の見直し
4	**特定単語の整理・見直し**	経済産業省検討会	**学習、データ種類等、多義的に捉えられる事項の記載** ✓「学習」「推論」の定義・表現の見直し ✓「データ」の定義・表現の見直し
5	**ユーザビリティの改善**	両省検討会	**AI事業者ガイドラインの活用を支援する資料・ツールの検討** ✓ 活用の手引きの検討 ✓ チャットボットの検討
6	**AIガバナンスに関する動向の反映**	両省検討会 事業者	**AIガバナンスに関する国内外の最新動向や、事業者の取組事例の追記** ✓ AI法や広島AIプロセスの動向等、国内外動向において注視するべき最新状況を追記 ✓「AIガバナンスの構築に関する実際の取組事例」への事例追加等
7	**その他**	両省検討会	**脚注記載内容やリンクの更新**

1. AI事業者ガイドラインの令和７年度の更新箇所の概要（案）

令和７年度の更新の論点と更新箇所 対応表

全体に関わる更新
- 6 AIガバナンスに関する動向の反映
- 7 その他

	本編（why, what）		別添（付属資料）（how）	
主体共通	第1部 AIとは	1 AI技術の動向の反映 2 AIによるリスクの記載の見直し 3 主体区分の整理 4 特定単語の整理・見直し	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と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共通の指針」 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ガバナンスの構築とモニタリング 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 主体区分の整理	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. 海外ガイドラインとの比較表	

1 AI技術の動向の反映
2 AIによるリスクの記載の見直し
3 主体区分の整理
4 特定単語の整理・見直し

1 AI技術の動向の反映
2 AIによるリスクの記載の見直し
3 主体区分の整理

2. 令和７年度の更新内容

AIエージェントやフィジカルAIに関する事項の追加

【更新内容】

① AI事業者ガイドラインとしての定義の追加

✓ AIエージェントの定義（案）
本ガイドラインにおけるAIエージェントとは、特定の目標を達成するために、環境を感知し自律的に行動するAIシステムとする。

✓ フィジカルAIの定義（案）
本ガイドラインにおけるフィジカルAIとは、センサ等によるセンシングを通じて物理環境の情報を取り込み、AIモデルによる処理を経て、設定された目的を達成するための最適な方策を自律的に推論・判断し、アクチュエータ（駆動系）等を介して物理的な行動へとつなげるシステムであり、サイバー空間での処理に留まらず、現実世界に対して直接的な働きかけ（移動、操作、加工など）を行うことを特徴とするものとする。

② 便益の追加

✓ AIエージェント…複数のシステムと連携しながらの自律的行動による、調整・分析・意思決定等の業務効率化の便益を追加

✓ フィジカルAI…物理環境での自律的行動による、労働力不足の補完、安全性向上、介護・生活支援等の便益を追加

③ リスクに関する事項の追加

✓ 自律的行動による人間の意図しない動作、攻撃対象・攻撃手法の増加、複雑機構を持つことによる制御の困難化、悪意のあるコード生成等の事項を追加

④ 留意すべき事項の追加

✓ 人間の判断を介在させる仕組みの構築や最小権限設定、ハードウェア残存データへの配慮等、留意すべき事項を追加

⑤ AIシステム・サービス例の追加

✓ 旅行先予約・提案エージェントや自律移動ロボット、AIエージェント作成サービス等のAIシステム・サービス例を追加

【主な更新箇所】

✓ 本編第1部.「AIとは」

✓ 別添1.「A.AIに関する前提」「B.AIによる便益／リスク」

✓ 別添3.4.5.「AI 開発者／AI提供者／AI利用者向け」

2. 令和７年度の更新内容

【更新内容の詳細】
①AI事業者ガイドラインとしての定義の追加

関連する用語　（本編第1部「AIとは」）

⋮

- **AI エージェント**[18]
  本ガイドラインにおける AI エージェントとは、特定の目標を達成するために、環境を感知し自律的[19]に行動する AI システムとする。
  （参考として ISO/IEC 22989:2022 では、以下のように定義されている）
  自動化された主体であり、環境を感知して応答し、自らの目標を達成するために行動を取るもの
- **フィジカル AI**
  本ガイドラインにおけるフィジカル AI とは、センサ等によるセンシングを通じて物理環境の情報を取り込み、AI モデルによる処理を経て、設定された目的を達成するための最適な方策を自律的に推論・判断し、アクチュエータ（駆動系）等を介して物理的な行動へとつなげるシステムであり、サイバー空間での処理に留まらず、現実世界に対して直接的な働きかけ（移動、操作、加工など）を行うことを特徴とするものとする。

[19] 自律については、高度な自律状態だけを指しているのではなく、ある程度の自律性を持つものも含む

- AIエージェント/フィジカルAIについては、本ガイドライン上の定義として追加

[18] AI エージェントよりも包括的かつ進化的な概念としてエージェンティック AI がある。これは、複数の AI エージェントにより自律的に意思決定を下しアクションを起こす目標主導型の AI システムである。

- エージェンティックAIについては、技術動向を鑑み、本年度は脚注補足に留め、次年度以降外部文献等も参照し、定義・リスクを追記していく想定

【更新内容の詳細】
②便益の追加

AIによる便益　（別添1.「B.AIによる便益/リスク」）

⋮

加えて、AI エージェントの登場により、ユーザーの意図を理解し、自律的にタスクを遂行することで、複雑な業務プロセスを効率化し、人的負荷を大幅に削減できる。単なる指示実行にとどまらず、複数のシステムやアプリケーションと連携し、状況に応じた判断や最適化を行うことで、従来は人手に依存していた調整・分析・意思決定を自動化することが可能となる。~~自律的な AI システム（以下、AI エージェント）も登場している。従来型の AI や生成 AI に比べより高度な効率化や自動化が可能となることで、生産性の向上につながること等が期待されている。~~

最後に、フィジカル AI は、少子高齢化による労働力不足を補い、人と協働して生産性を向上させることで、あらゆる産業や現場の自動化と効率化に寄与することが期待される。危険な環境で人の代わりに作業を行い、安全性を高めつつリスクを低減するほか、介護や生活支援を通じて人々の自立と QOL 向上に寄与し、福祉や医療などの分野で新たな支援の創出にもつながり得る。

- AIエージェントは、複数のシステムと連携した自律的な行動により、業務の調整・分析・意思決定の効率化に資する旨を追記
- フィジカルAIは、物理環境での自律的な行動を通じて、労働力不足の補完や安全性向上、生活支援に貢献し得る旨を追記

2. 令和７年度の更新内容

【更新内容の詳細】
②便益の追加（図「企業活動におけるAIによる便益の例」）

（別添1.「B.AIによる便益/リスク」

	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（一部AIエージェントやフィジカル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~~ 顧客の成約・解約率予測	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、AIエージェント、フィジカル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	~~過去実績に基づいた将来予測、不正検知~~ 経費精算アシスタント（自動仕訳、申請レビュー、証憑取得等）	~~職務履歴に基づいた人材需要マッチング~~ AI採用アシスタント（面接・評価）
	コーディングアシスタント（コード生成、不具合の自動修正等）	投稿から分析までを自律的に行うSNS運用エージェント	店頭ロボットによる自動接客・販売	自律搬送ロボット・ドローンによる自動配送	過去の問合わせ内容に基づいたFAQ自動作成	類似事例検索及び重要判例の自動要約	投資レポート・市場分析の自動生成	パーソナライズされたキャリアプラン提案

- AIエージェントやフィジカルAIの便益の事例を追記

2. 令和７年度の更新内容

【更新内容の詳細】
③リスクに関する事項の追加

（別添1.「B.AIによる便益/リスク」

AI によるリスク

データ汚染攻撃等の AI システムへの攻撃

- AI の学習実施時では性能劣化及び誤分類につながるような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 AI の推論結果又は AI への指示であるプロンプトを通した攻撃等もリスクとして存在する[11]。例えば、とあるチャットボットでは、悪意のある集団による人種差別的な質問の組織的な学習により、ヘイトスピーチを繰り返し発言するようになった

⋮

- 生成 AI をシステム開発に用いる場合、自然言語が直接ソースコードや設計情報に変換されるため、入力情報の信頼性がシステムの安全性に直結するリスクがある
- AI エージェントやマルチモーダルな生成 AI、フィジカル AI 等の複雑で多様な情報を扱うシステムでは、より多様な入力経路や外部連携が増えるため、被攻撃対象が拡大し、データ汚染や悪意あるプロンプト攻撃のリスクがさらに高まることが懸念される

⋮

- **自然言語を由来とした攻撃**を追記
- AIエージェント、フィジカルAIは**多様な入力経路・外部連携先**を持つため、**被攻撃対象が拡大**し得る旨を追記

ハルシネーション等による誤った出力[12]

- 生成 AI が事実と異なることをもっともらしく回答する「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。とあるテレビ番組の出演者が、「自身が金銭の横領で提訴されている」という偽情報を生成 AI が拡散しているのを発見。生成 AI に虚偽の告訴状まで作られたとして、当該生成 AI を開発・提供する企業を名誉毀棄損で提訴した
- AI エージェントの場合、自律的な動作の中で人間の意図しない商品の注文やファイル削除等の動作を行う可能性がある[13]

[13] なお、AI エージェントの自律性が高まるにつれ、人間による監視のみでは高速な AI 間相互作用への対応が困難となる場合は十分に想定される。今後、AI システム間の相互監視等、AI の自律性に対応した新たな安全確保アプローチについても検討が期待される。

- AIエージェントの**人間の意図しない動作**に関するリスクや対策等を追記

ブラックボックス化、判断に関する説明の不足

- 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。とあるクレジットカードにおいて、同じ年収を有する男性及び女性に対して、女性の方が利用限度額が低いとの報告が SNS 上で広がった。この問題に対し、金融当局が調査を実施し、クレジットカードを提供した企業に対してアルゴリズムの正当性の証明を求めた。しかし、企業はアルゴリズムの具体的な機能及び動作について説明することができなかった
- マルチモーダルな生成 AI、AI エージェント、フィジカル AI など複雑な構成や機構を持つ AI システムの場合、通常の AI システムに比べメンテナンスやトラブルシューティングの難易度が上がる場合がある

- AIエージェント、フィジカルAIは**内部構造が複雑なため、制御の難易度が高くなる**可能性がある旨を追記

2. 令和７年度の更新内容

【更新内容の詳細】
③リスクに関する事項の追加

（別添1.「B.AIによる便益/リスク」

AI によるリスク

悪用

- AI の詐欺目的での利用も問題視されている。中でも、AI で合成された音声を利用した詐欺が問題となっている。~~は急増している。~~とある女性に、娘の声で助けを求める電話があり、100 万ドルの身代金が要求されたものの、この声は AI を用いて生成されたものであり、誘拐を偽装した詐欺の電話であったことが判明した
- 昨今のコード生成 AI や AI エージェントの進歩により、コードの生成が容易になり、その悪用が問題視されている。例えば、セキュリティホールを狙うマルウェアの作成や、ネットワークに侵入するためのエクスプロイトコードを生成することが可能になっている。このような技術の悪用は、個人情報の漏洩や重要インフラに関わるシステムの停止・情報の改ざん等に繋がる可能性があり、深刻なサイバーセキュリティ上の問題となっている

⋮

機密情報の流出

- AI の利用においては、個人情報、自社のナレッジや知的財産等の~~及び~~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。例えば、AI サービスの利用上、従業員が業務利用のため、機密情報に該当するソースコードを、業務外利用者向けの対話型生成 AI に入力してしまう事例が明らかになっている。生成 AI 系サービスでは利用障壁が下がっていることから、特に、企業のル

⋮

- AI エージェントでは、外部システムやクラウドサービスと自律的に連携して各種タスクを実行するケースが増えており、その過程で、脆弱性を突かれた攻撃等によってエージェントの挙動が不正に操作され、内部データが意図せず外部に送信されるなど、機密情報が漏洩する可能性がある

このように、技術発展により AI 活用による便益が大きくなる一方で、従来型 AI でも現れていたリスクが生成 AI の普及によって一層顕在化している。~~生成 AI の台頭によりさらに増大傾向にある。~~また、生成 AI により新たに顕在化したリスクもある。[19]加えて、多くの生成 AI サービスで利用障壁が下がったことから、意図しないリスクを伴う使われ方をする恐れもある[20]。

[19] エージェンティック AI の登場により、自ら利用する AI がインターネット等を通じて他の AI 等と接続・連携することにより制御不能となる等、AI 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ことが期待される。

- AIエージェント等により**コード生成**が容易になっている現状を踏まえ、それを**悪用**した事例を追記
- AIエージェントが外部システムと自律的に連携する際、アクセス可能なデータベースを含む**内部データが不正に外部へ送信**されるリスクを追記
- 「エージェンティックAI」について本年度は具体に記載せず、**リスクが増幅される可能性**の記載のみに留める想定

2. 令和７年度の更新内容

【更新内容の詳細】
④留意すべき事項の追加(AI開発者)

（別添3.「AI開発者向け」）

AI 開発時

D-2） iii. 適正利用に資する開発

- ✧ 開発時に想定していない AI の提供・利用により危害が発生することを避けるため、安全に利用可能な AI の使い方について明確な方針・ガイダンスを設定する（「2） 安全性」）
- ✧ 事前学習済の AI モデルに対する事後学習を行う場合に、学習済 AI モデルを適切に選択する（商用利用可能なライセンスかどうか、事前学習データ、学習・実行に必要なスペック等）（「2） 安全性」）

⋮

[具体的な手法]

- 目的と照らした AI モデルの選択・調整
  - ➢ AI モデルの真正性
    - ✧ 採用する AI モデルについて、開発元や取得経路が信頼できることを確認

- 悪意をもって改変されたAIモデルを利用することを避けられるよう、**AIモデルの真正性を確認**することが重要である旨を追記

（別添3.「AI開発者向け」）

データ前処理・学習時

D-2） i. 適切なデータの学習

- ✧ 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」）

⋮

- 権利又は法律上保護される利益に関係するものが含まれる場合には、個人情報・機密情報・著作権等の適切な取扱いを実施

⋮

- ➢ 必要最小限のデータ入力・参照
  - ✧ AI の判断に必要な最小限の範囲にデータを限定し、不要な属性情報の付与を避けること

- **意図しないデータ外部送信リスクの被害を抑える**ため、扱うデータを必要最小限に限定する旨を追記

AI 開発時

D-6） i. 検証可能性の確保

- ✧ AI の予測性能及び出力の品質が、活用開始後に大きく変動する可能性又は想定する精度に達しないこともある特性を踏まえ、事後検証のための作業記録を保存しつつ、その品質の維持・向上を行う （「2） 安全性」「6） 透明性」）

⋮

- 説明可能性・解釈可能性を高めるための手法を検討する。なお、以下の検討にあたっては、開発とのトレードオフが生じる懸念もあるため、留意すること。また、LLM 等においては結果とその根拠を出力にするよう求めることができるが、AI 内部の決定ロジックについて説明しているわけではなく、結果と根拠のもっともらしい組み合わせを出している点で、従来型の AI の説明可能性とは別途検討すべきである点に注意する

- AIエージェントの自律判断を適切に監査するには説明可能性が重要だが、**LLMの根拠提示は内部の決定ロジックの説明ではなく、もっともらしい理由を出力しているのに過ぎない**旨を明記

2. 令和７年度の更新内容

【更新内容の詳細】
④留意すべき事項の追加（AI提供者）

（別添4.AI提供者向け）

- AI システム実装時

P-2） ｉ．人間の生命・身体・財産、精神及び環境に配慮したリスク対策

- ✧ AI 利用者を含む関連するステークホルダーの生命・身体・財産、精神及び環境に危害を及ぼすことがないよう、提供時点で予想される利用条件下でのパフォーマンスだけでなく、様々な状況下で AI システムがパフォーマンスレベルを維持できるようにし、リスク（連動するロボットの制御不能、不適切な出力等）を最小限に抑える方法（ガードレール技術等）を検討する（「2）安全性」）

⋮

- インシデントの未然防止
  - ➢ AI システム全体で安全を確保できる仕組みの構築（フェールセーフの実現）
  - ➢ AI 開発者も気づいていないようなリスクの存在を認識した場合に、速やかな AI 開発者への通知及び対応策の相談・検討
  - ➢ 安全確認等の事前及び動作時の人間関与、並びに事後における再発防止策の検討
  - ➢ AI 利用者側の適正利用申告等による AI 利用者の信頼性の確認

⋮

  - ➢ 不要なデータや冗長なログの削除等による、データ最小化と適切なデータ管理の徹底
  - ➢ ホワイトリスト方式を利用した連携ツール・システムの制限
  - ➢ 人間の判断を介在させる仕組みの構築
    - ✧ 判断が必要となる事項を重要度に応じて整理し、適切に対象を選定することが重要である

- **意図しないデータの漏洩リスクの被害を抑える**ため、データ最小化が重要である旨を追記
- AIエージェントでは**外部システムとの連携が増え、適切な連携ツール・システムの制限**が重要である旨を追記
- AIエージェントやフィジカルAIは自律的に動作するため、**人間の判断を介在させる仕組みの構築**が重要である旨を追記

（別添4.AI提供者向け）

P-5） ｉ．セキュリティ対策のための仕組みの導入

- ✧ AI システム・サービスの提供の過程を通じて、採用する技術の特性に照らし適切にセキュリティ対策を講ずる（セキュリティ・バイ・デザイン）（「5）セキュリティ確保」）

⋮

- ➢ セキュリティアーキテクチャ（Security Architecture）の選択
  - ✧ AI 開発者が提供する AI システムに求めるアーキテクチャ情報をもとにする
  - ✧ 自組織で独自のアーキテクチャを考えるよりも、AI システムを搭載するプラットフォームの提供元が推奨しているアーキテクチャをカスタマイズし利用する
  - ✧ オープンソースソフトウェアを利用する際は、開発元や取得経路が信頼できることを確認する
  - ✧ ユーザーやシステムに付与する権限を業務遂行に必要な最小限に設定する

- **オープンソース開発元の信頼性**の確認やAIエージェントによる**意図しない操作**を防ぐために、**適切な権限設定**が重要である旨を追記

P-4） ｉ．プライバシー保護のための仕組み及び対策の導入

- ✧ AI システムの実装の過程を通じて、採用する技術の特性に照らし適切に個人情報へのアクセスを管理・制限する仕組みの導入等のプライバシー保護のための対策を講ずる（プライバシー・バイ・デザイン）（「4）プライバシー保護」）

- 関連するステークホルダー及び個人のプライバシーの尊重
  - ➢ 個人のプライバシーを侵害する情報の消去、加工、AI のアルゴリズムの更新等（AI 利用者等関連するステークホルダー又は個人のプライバシーを侵害する情報を取得した場合）[125]

[125] フィジカル AI においては、個人情報を取得してしまう可能性があるうえ、こうした情報がデバイス上に保存・残存し、解析や不適切な利用に結び付くことでプライバシー侵害のおそれがあるため、不要な情報を取得・保持しない仕組みを設ける必要がある。また、ハードウェア廃棄時には、機器やクラウドの記憶媒体に保存されたデータも適切に削除することが望ましい。

- フィジカルAIにおける**プライバシー侵害リスク**及び**ハードウェア上に残存するデータの削除**も重要である旨を追記

2. 令和７年度の更新内容

【更新内容の詳細】
④留意すべき事項の追加（AI利用者）

（別添5.AI利用者向け）

U-2） ｉ．安全を考慮した適正利用

✧ AI 提供者が定めた利用上の留意点を遵守して、AI 提供者が設計において想定した範囲内で AI システム・サービスを利用する（「2） 安全性」）

[ポイント]

AI 利用者は、AI 提供者からの情報提供（AI 開発者の情報を含む）及び説明を踏まえ、AI を活用する際の社会的文脈にも配慮して、AI を利用すべきである[125,126]。

⋮

[127] 生成 AI を用いて生成したプログラムコードにセキュリティ上の脆弱性等が含まれていた場合、情報の改ざんや漏洩等につながる恐れがあるほか、誤った/非効率なコードが生成された場合、パフォーマンスの低下や事故等につながる懸念がある。また、生成したコードが他者の知的財産権等を侵害する可能性に留意が必要である。さらにコード生成 AI の利用においては、作成者の意図やコード構造の把握が難しくなり、保守・更新が困難になる可能性もある。そのため、ノーコード生成ツールも含め、社内ノウハウを如何に蓄積するかが重要である。

- AIエージェントにより**コード生成**が容易になることを踏まえ、**生成物の保守・更新が困難になる可能性**及び**社内ノウハウの蓄積**が重要になる旨を追記

U-5） ｉ．セキュリティ対策の実施

✧ AI 提供者によるセキュリティ上の留意点を遵守する（「5） セキュリティ確保」）

✧ AI システム・サービスに機密情報等を不適切に入力することがないよう注意を払う（「5） セキュリティ確保」）

⋮

● ログの定期的な確認

➢ システム利用状況や異常操作を把握するため、ログを定期的にレビューする

➢ 異常な操作を検知した場合に即時報告する体制を整備する

- AIエージェントやフィジカルAIは自律性を持って行動するため、**定期的な操作履歴確認や報告**が重要である旨を追記

別添5.AI利用者向け （別添5.AI利用者向け）

U-2） ｉ．安全を考慮した適正利用

✧ AI 提供者が定めた利用上の留意点を遵守して、AI 提供者が設計において想定した範囲内で AI システム・サービスを利用する（「2） 安全性」）

✧ AI システム・サービスが想定された仕様に基づき適切に動作しているかを確認する（「2） 安全性」）

⋮

● 適正な範囲・方法での利用

➢ AI の性質、利用の態様等に応じた便益及びリスクの認識、並びに適正な用途の理解（利用前）

➢ 適正利用のための必要な知識・技能の習得（利用前）

➢ AI システム・サービスの利用過程におけるログ（操作履歴、入力・出力の記録等）の管理体制の整備（利用前）

➢ AI の活用が適正な範囲・方法で行われているかについての定期的な確認（利用中）

➢ AI システムのアップデート及び AI の点検・修理又はそれらを AI 提供者へ実施を依頼（活用の過程を通じて、AI の機能を向上させ、リスクを抑制することを目的とする）（利用中）

✧ ただし、アップデートにより連携する他の AI に影響を及ぼしうることも考慮する

➢ 出力によって重大な影響又は被害が生じ得る場合、人間の判断を介在させる仕組みに基づき適宜判断（利用中）

➢ 不要なデータや冗長なログの削除等による、データ最小化と適切なデータ管理の徹底（利用中）

➢ AI 提供者（又は AI 提供者を通じて AI 開発者）に対するインシデント情報のフィードバック（何らかインシデントが発生した場合、インシデントが起こる予兆があった場合を含む）

- 出力によって重大な影響又は被害が生じ得る場合、**人間の判断を介在させる仕組みに基づいた判断**が重要である旨を追記
- **意図しないデータの漏洩リスクの被害を抑える**ため、データ最小化が重要である旨を追記

2. 令和７年度の更新内容

【更新内容の詳細】
⑤AIシステム・サービス例の追加

（別添1.「A.AIに関する前提」）

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
⋮						
AI エージェント作成サービス	ワークフロー作成	AI 提供者である X 社が提供する **AI エージェント（ワークフロー）作成サービス**である。**本サービスは、利用者が業務プロセスに沿ったワークフローを独自に構築し、そのワークフローに従って自律的に動作する AI エージェントを作成することができる。**このサービスに用いられる AI モデルは、AI 開発者である Y 社が開発したものであり、API を通じて本サービスに呼び出され、ワークフロー内での判断や意思決定を担っている。 本サービスの AI 利用者であるソフトウェア開発企業 Q 社は、X 社の AI エージェント作成サービスを用いて独自のワークフローを設計し、業務に特化した AI エージェントを構築した。このように **Q 社は AI 利用者であると同時に、構築した AI エージェントを自社サービスとして提供する AI 提供者としての区分も兼ねており、保守・運用の役割も担う。**	Y 社（AI モデル開発部門）	X 社（サービス運営部門） Q 社（サービス開発部門）	Q 社（サービス開発部門）	-
旅行先提案・予約 AI エージェント	AI エージェント	航空会社 G 社が提供する、**ユーザーの希望条件に応じて最適な旅行先とフライトオプションを提案し、外部予約システムと連携して実際の便予約まで行う AI エージェント型サービス**である。 本サービスは、従来の FAQ 応答型や予約サポート型 AI とは異なり、ユーザーの要望に応じた目的地提案だけでなく、社内外システムと連携し、空席確認・便選択・予約確定までを一連で実行可能とする点が特徴である。 本サービスの開発においては、G 社プロダクト開発部門が、G 社が開発したモデルを組み込んだ AI エージェントのワークフローや、外部 API の呼び出し設計、外部システムとの物理的な連携・認証・インフラ構築など、サービス全体の安定運用に必要な実装と運用管理を担当した。	G 社（AI システム開発部門）	G 社（プロダクト開発部門）	-	旅行予約者
営業・CS 支援 AI エージェント	AI エージェント	**営業やカスタマーサポート部門において、見込み客への対応から商談進捗の管理、お問い合わせ対応までを 24 時間自律的に支援する業務支援型 AI エージェント**である。 本サービスは CRM（顧客管理システム）や SFA（営業支援システム）と連携し、単なる質問回答にとどまらず見込み顧客の管理から商談スケジュールの自動調整、パーソナライズされたメールの配信まで、自律的に複雑な業務を一連で実行し、営業担当者の負担を大幅に軽減するものである。 Z 社の AI エンジン開発部門は自然言語処理やモデルのファインチューニングを担い、プロダクト開発部門は AI エージェントのワークフロー設計や、CRM・SFA との連携、ユーザーインターフェース設計を担い、サービスとして提供している。	Z 社（AI エンジン開発部門）	Z 社（プロダクト提供部門）	Y 社（営業部門）	-
自動運転システム	フィジカル AI	自動車メーカーH 社が提供する、**自動運転技術を搭載した車両制御システム**である。ユーザーが目的地を指定すると、車両は周囲環境を認識し、走行経路を自律的に計画・実行する。交通状況や道路標識、障害物をリアルタイムで検知し、安全かつ効率的な運転を実現する。 本サービスの開発においては、H 社 AI アルゴリズム開発部門が深層学習を用いた経路計画モデルを開発し、同社車両開発部門が車両制御システムへの統合、センサー・カメラとの連携、認証・安全基準対応、インフラ構築を担当した。	H 社（AI アルゴリズム開発部門）	H 社（車両開発部門）	-	運転者
自律移動ロボット	フィジカル AI	物流企業 J 社が提供する、**自律移動ロボットによる倉庫内搬送サービス**である。ロボットは倉庫内の地図を認識し、最適なルートを自律的に選択して荷物を搬送する。障害物回避や複数ロボット間の協調動作を行い、作業効率を大幅に向上させる。本サービスは、従来の AGV（無人搬送車）と異なり、固定ルートではなく動的な経路計画を行い、リアルタイムで環境変化に対応できる点が特徴である。 本サービスの開発においては、J 社ロボティクス AI 開発部門がロボットの自己位置推定技術や経路最適化アルゴリズムを開発し、同社物流システム開発部門がロボットのハードウェア設計、倉庫管理システムとの連携、運用管理を担当した。	J 社（ロボティクス AI 開発部門）	J 社（物流システム開発部門）	O 社（倉庫管理部門）	-

2. 令和７年度の更新内容

事業者がAIリスクを把握・対応しやすくなるように、「AIによるリスク」の記載を整理・拡充

【更新内容】

① リスクベースアプローチに資する内容の追記

✓リスクの大きさ/発生可能性等を加味して対策の優先順位を検討するという考え方に基づくリスクベースアプローチの説明を追加

✓参考文献の追加

－AIガバナンス協会「AI時代の経営意思決定とガバナンス ～攻めのAIガバナンス実現のための戦略レポート」

－EU「Artificial Intelligence Act Annex III: High-Risk AI Systems Referred to in Article 6(2)」

② リスクの更新

✓AIシステムへの攻撃例、マルチモーダルな生成AIやカメラ、音声認識活用時のプライバシー権の侵害リスク、ハルシネーションによりメリットも生じ得ること、教育分野のAI活用におけるリスク 、金銭的損失の被害者となり得るリスク、資格等の侵害リスク等に関する更新

③ 一部リスクの分類見直し

✓現行ガイドラインにおいて技術的リスクとして位置付けられている「差別的出力」は倫理・法的側面に関わることを踏まえ、リスク分類を変更

【更新箇所】

✓別添1.「B.AIによる便益/リスク」

✓別添2.「E.AIガバナンスの構築」

2. 令和７年度の更新内容

【更新内容の詳細】
①リスクベースアプローチに資する内容の追記

3.システムデザイン（AI マネジメントシステムの構築） ↵ （別添2.「E.AIガバナンスの構築」関連）

行動目標 3-1【ゴール及び乖離の評価、並びに乖離対応の必須化】：↓
各主体は、経営層のリーダーシップの下、各主体の AI の AI ガバナンス・ゴールからの乖離を特定し、乖離により生じる影響を評価した上、リスクが認められる場合、その大きさ、範囲、発生頻度等を考慮して、その受容の合理性の有無を判定し、受容に合理性が認められない場合に AI の開発・提供・利用の在り方について再考を促すプロセスを、AI マネジメントシステム全体、及び AI システム・サービスの設計段階、開発段階、利用開始前、利用開始後等の適切な段階に組み込むことが期待される。経営層は、再考プロセスについて基本方針等の方針策定、運営層はこのプロセスの具体化を行うことが重要である。そして、AI ガバナンス・ゴールとの乖離評価には対象とする AI の開発・提供・利用に直接関わっていない者が加わるようにすることが期待される。なお、乖離があることのみを理由として AI の開発・提供・利用を恣意的に不可とする対応は適当ではない。そのため、乖離評価はリスクを評価するためのステップであり、改善のためのきっかけにすぎない。↵

[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。↵
- 「AI ガバナンス・ゴール」からの乖離を特定し、リスクベースアプローチ[38]を用いて、リスクに対するコントロールを選択し、ユースケース、サービス又は製品ごとに適切なレベルの管理を実施↵

⋮

[38] リスクベースアプローチとは、AI の利用目的・利害関係者、発生し得るリスクの影響の大きさ/発生可能性などを踏まえて、対策の優先順位を決定する手法である。参考に AI ガバナンス協会 （AIGA）では、「AI 時代の経営意思決定とガバナンス ～攻めの AI ガバナンス実現のための戦略レポート」にて、ベネフィットとリスクを併せて評価するアプローチ手法が紹介されている。↵
https://cdn.prod.website-files.com/66e98b87b115812d1af8fc1c/69285da091ec71dde1ae3c71_management-strategy-report-ver1.0.pdf↵
また、EU「Artificial Intelligence Act Annex III: High-Risk AI Systems Referred to in Article 6(2)」↵
https://artificialintelligenceact.eu/annex/3/↵
にて整理されているリスク分類の考え方も、リスク識別および評価プロセスの補強をする際の参考となる。↵

- AIガバナンス協会（AIGA）のリスクベースアプローチに関する文献を参考として追加
- 高リスク領域が整理された EU AI Actの補助資料を参考として追加

2. 令和７年度の更新内容

【更新内容の詳細】
②リスクの更新

AI によるリスク

（別添1.「B.AIによる便益/リスク」）

データ汚染攻撃等の AI システムへの攻撃

- AI の学習実施時では性能劣化及び誤分類につながるような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 AI の推論結果又は AI への指示であるプロンプトを通した攻撃等もリスクとして存在する[11]。例えば、とあるチャットボットでは、悪意のある集団による人種差別的な質問の組織的な学習により、ヘイトスピーチを繰り返し発言するようになった
- 間接プロンプトインジェクションやマルウェアの生成、悪意ある目的で行われるファインチューニングなど、~~悪意のある第三者~~により、RAG が悪用されたり、AI モデルに設定された制御等を無視した悪意のある出力が行われたりするリスクがある

> • 悪意をもって、**AIモデルに設定された制御を外す**事例について言及

個人情報の不適切な取扱い等

- 人材採用に AI を用いるサービスにて、選考離脱及び内定辞退の可能性を AI により提供した際、学生等の求職者への説明が不明瞭であった他、一定の期間において、~~一時期~~同意にもとづいて第三者への情報提供が行われる規約となっていなかったこと等、透明性を欠く個人情報の利用が問題視され、サービスが廃止されることとなった事例が発生している
- マルチモーダルな生成 AI により、テキスト情報だけでなく、個人の顔写真や録音音声といったセンシティブな情報も取り扱われる機会が増加し、これらの情報が漏洩するリスクが一層高まる可能性がある
- カメラや音声認識等を通じて周囲の映像や音声を収集する過程で、個人情報を取得することとなる場合には、利用目的の通知又は公表等の対応が必要となるほか、目的の正当性や手段の必要性及び相当性を欠く態様で情報の保存、解析及び利用等が行われた場合には、肖像権等のプライバシー権を侵害するリスクがある

> • **マルチモーダルな生成AIやカメラ、音声認識を活用した際のプライバシー権の侵害リスク**を追加

~~バイアスのある出力、差別的出力、~~一貫性のない出力等

- AI モデル学習時のデータに、特定の属性が過剰に含まれている場合や、ラベル付与者の主観的なバイアスが含まれている等の場合、出力に偏りが生じることがある。例えば、学習データに特定の趣味嗜好を持つ顧客のデータが過剰に含まれている場合、おすすめとしてその嗜好に偏った商品ばかりを提示し、他の顧客層に不適切な提案を行ってしまうおそれがある。その結果、ユーザー体験の低下や、場合によっては消費者からの苦情・クレームにつながるリスクも生じ得る ~~IT 企業が自社で AI 人材採用システムを開発したが、女性を差別するという機械学習面の欠陥が判明した。この原因としては、学習に使用した過去 10 年間の履歴書において、応募者のほとんどが男性であったことから、男性を採用することが好ましいと AI が認識したためといわれている。当該企業は、女性を差別しないようにプログラムの改善を試みたものの別の差別を生むとして運用を取りやめる結果になった~~
- AI が同じ基準やポリシーで運用されているにもかかわらず、モデルの確率的性質に起因して、場面によって異なる判断を下すことがある

⋮

> • **差別的出力は技術特性だけで判断できない**ため分類を再整理。当該記載を削除し、**技術的特性に基づくバイアスの説明**に修正

差別的出力~~トリアージにおける差別~~

- IT 企業が自社で AI 人材採用システムを開発したが、女性を差別するという機械学習面の欠陥が判明した。この原因としては、学習に使用した過去 10 年間の履歴書において、応募者のほとんどが男性であったことから、男性を採用することが好ましいと AI が認識したためといわれている。当該企業は、女性を差別しないようにプログラムの改善を試みたものの別の差別を生むとして運用を取りやめる結果になった~~インシデント発生時に優先順位付けを行うトリアージにおいては、AI が順位を決定する際に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。医療場面のトリアージにて活用される際には、特定の人群に対して差別的な医療判断が行われることで、生命に対する脅威が発生する可能性がある~~

> • 差別的出力の分類を再整理することに伴い、**「トリアージにおける差別」を「差別的出力」**へと変更

2. 令和７年度の更新内容

【更新内容の詳細】
②リスクの更新

AI によるリスク

（別添1.「B.AIによる便益/リスク」）

過度な依存

- 人材採用活動等、重要な意思決定を行う場面において AI による判断をそのまま用いることや、人が最終判断する際にも AI システムによる判断を過度に人間が信頼し、自らの判断や確認を怠ってしまう（自動化バイアス）可能性を考慮していないなど、意思決定や意思決定支援を AI に委ねることが懸念される。このような AI への過度な依存により、事業者が説明責任を問われることや批判を受けることに繋がり得るなど~~AI の不適切かつ過剰な使用により、企業が説明責任を問われたり批判を受けたりする事例が発生している~~
- また、生成 AI を用いたチャットボットサービスと会話をしていた利用者が、AI の助言により自殺してしまうなど、AI に心理的に依存してしまう~~に対し精神的な依存状態となった~~事例が報告されている
- 教育においても、学生の独自性や批判的思考の発展が妨げられるリスクが懸念されている。例えば、生成 AI が作成した答案やレポートをそのまま用いることで、学生自身が情報を探索し、問いを立て、根拠を検討するプロセスを省略してしまい、結果として独自の思考力を育む機会が減少する可能性がある
- 教育分野においては、教室内で学童の表情をモニターしその心理状態を判断して教員による指導に活用する海外の例などがあり、AI システムを活用する上での適切性について慎重な検討が必要である

⋮

- **心理的依存、決定権の依存**として明記
- **教育分野におけるリスク**を追記

ハルシネーション等による誤った出力[12]

- 生成 AI が事実と異なることをもっともらしく回答する「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。とあるテレビ番組の出演者が、「自身が金銭の横領で提訴されている」という偽情報を生成 AI が拡散しているのを発見。生成 AI に虚偽の告訴状まで作られたとして、当該生成 AI を開発・提供する企業を名誉棄損で提訴した

[12] 画像生成 AI の場合は、ハルシネーションによって、自身が想定していたものとは異なるものが創出されることによる発見や想像もある。また、創薬プロセスのタスク等においてハルシネーションにより予測精度を高めるとの研究結果も登場している。

- **ハルシネーションによりメリットも生じ得る旨**を補足として追記

金銭的損失

- 企業においては、自社の扱う AI システム・サービスの出力により他者の権利を著しく侵害した場合等において、損害賠償請求など金銭的な責任を問われることがある
- AI を悪用した攻撃によって内部システムが破壊されたり、機密情報が漏洩したりすると、システム復旧費用や関連するステークホルダーへの補償費用等の金銭的負担が発生するリスクがある

⋮

- 加害者となるリスクだけでなく、**被害者となるリスク**も追記

資格等の侵害

- 生成 AI が法律や医療など、業法免許や資格が必要な領域で助言や回答を行う場合、本来は有資格者に限定される業務を無資格で行うことに該当し、法令違反と見なされる可能性がある~~生成 AI の活用を通じた業法免許及び資格等の侵害リスクも考えうる。例えば、生成 AI が法律又は医療の相談に回答する場合、業法免許及び資格の侵害が生じ、法的問題が発生する可能性がある。このリスクを回避しようとした場合、業界全体で生成 AI の導入が遅れ、新たなサービス及び効率向上が制限される可能性もある~~

- **生成AIを専門分野に活用する際に、資格等を侵害し、法令違反を問われる**可能性がある旨を追記

B.AI による便益/リスク

AI は、新規ビジネスを生み出したり、既存ビジネスの付加価値を高めたり、生産性を向上させたりする等の便益をもたらす一方で、リスクも存在する。

このリスクについては可能な限り抑制することが期待される。一方で、過度なリスク対策を講じることは、コスト増になる等、AI 活用によって得られる便益を阻害してしまうことから、リスク対策の程度をリスクの性質及び蓋然性の高さに対応させるリスクベースアプローチの考え方が重要である。なお、リスクへの対策を検討する際には、当該対策の実行可能性も考慮し、限られた資源を効果的に配分することが望ましい。

- **リスクの対策の実行可能性**の考慮も重要な旨を追記

2. 令和７年度の更新内容

【更新内容の詳細】
③一部リスクの分類見直し（表「AIによるリスク例の体系的な分類案」）

（別添1.「B.AIによる便益/リスク」）

・下表はAIのリスクを網羅したものではなく、想定に基づく事案も含んでおり、あくまで一例として認識することが期待される
・下表には政府等の公的機関も含めた社会全体での対応・議論が必要となるリスクも含まれる

大分類	中分類	リスク例
技術的リスク（＝主にAIシステム特有のもの）	学習及び入力段階のリスク	データ汚染攻撃等のAIシステムへの攻撃
	出力段階のリスク	バイアスのある出力~~、差別的出力~~、一貫性のない出力等
		ハルシネーション等による誤った出力
	事後対応段階のリスク	ブラックボックス化、判断に関する説明の不足
社会的リスク（＝既存のリスクがAIにおいても発生又はAIによって増幅するもの）	倫理・法に関するリスク	個人情報の不適切な取扱い等
		生命等に関わる事故の発生
		~~差別的出力トリアージにおける差別~~
		過度な依存
		悪用
	経済活動に関するリスク	知的財産権等の侵害
		金銭的損失
		機密情報の流出
		労働者の失業
		データや利益の集中
		資格等の侵害
	情報空間に関するリスク	偽・誤情報等の流通・拡散
		民主主義への悪影響
		フィルターバブル及びエコーチェンバー現象
		多様性・包摂性の喪失
		バイアス等の再生成
	環境に関するリスク	エネルギー使用量及び環境の負荷

- 差別的出力か否かは技術的特性だけで決まるのではなく、法的・倫理的評価に基づいて決まると考えられるため、**「差別的出力」を「倫理・法に関するリスク」**へと再分類

2. 令和７年度の更新内容

【更新内容の詳細】

③一部リスクの分類見直し（表「AI によるリスク例と共通の指針及び主体毎に重要となる事項のマッピング」）

※更新箇所抜粋

（別添1.「B.AIによる便益/リスク」）

	リスク例	関連する共通の指針	「共通の指針」に加えて主体毎に重要となる事項 第3部 AI開発者	第4部 AI提供者	第5部 AI利用者
技術的リスク	バイアスのある出力、~~差別的出力、~~一貫性のない出力等	**1）人間中心** ①人間の尊厳及び個人の自律 ③偽情報等への対策			
	ハルシネーション等による誤った出力	**2）安全性**	i. 適切なデータの学習 ii. 人間の生命・身体・財産、精神及び環境に配慮した開発 iii. 適正利用に資する開発	i. 人間の生命・身体・財産、精神及び環境に配慮したリスク対策 ii. 適正利用に資する提供	i. 安全を考慮した適正利用
		3）公平性	i. データに含まれるバイアスへの配慮 ii. AIモデルのアルゴリズム等に含まれるバイアスへの配慮	i. AIシステム・サービスの構成及びデータに含まれるバイアスへの配慮	i. 入力データ又はプロンプトに含まれるバイアスへの配慮
		8）教育・リテラシー			
社会的リスク	~~差別的出力トリアージにおける差別~~	**1）人間中心** ①人間の尊厳及び個人の自律 ②AIによる意思決定・感情の操作等への留意			
	過度な依存	**2）安全性**	i. 適切なデータの学習 ii. 人間の生命・身体・財産、精神及び環境に配慮した開発 iii. 適正利用に資する開発	i. 人間の生命・身体・財産、精神及び環境に配慮したリスク対策 ii. 適正利用に資する提供	i. 安全を考慮した適正利用
		3）公平性	i. データに含まれるバイアスへの配慮 ii. AIモデルのアルゴリズム等に含まれるバイアスへの配慮	i. AIシステム・サービスの構成及びデータに含まれるバイアスへの配慮	i. 入力データ又はプロンプトに含まれるバイアスへの配慮

2. 令和７年度の更新内容

開発、学習、主体区分等、多義的に捉えられる事項について、定義の追加や図表を更新

【更新内容】

① AI開発者の定義の補足

✓AI開発者について、AIシステムの構築の全てを担うわけではない旨を補足
✓ファインチューニング等、AIモデル開発後のモデル調整（事後学習）をAI開発者が役割として担う旨を補足

② 「一般的なAI活用の流れにおける主体の対応」の見直し

✓「AIモデル事後学習」を明記
✓「データ前処理・学習」や「システムへの実装」の流れを整理

③ 主体毎の役割の見直し

✓同一事業者が複数区分を跨ぐと考えられる事例等について表「AIシステム・サービス例」に追加
　－ファインチューニングを実施した事例
　－コード生成AIを活用した事例
　－APIを活用した事例
✓アライメントやRAG等、どの主体区分が役割を担うか不明瞭な箇所について、補足説明を追加

【主な更新箇所】

✓本編「はじめに」
✓本編第1部.「AIとは」
✓本編第2部.「A.基本理念」
✓別添1.「A.AIに関する前提」
✓別添3.4.「AI開発者／AI提供者向け」

2. 令和７年度の更新内容

【更新内容の詳細】

①AI開発者の定義の補足

- **AI 開発者（AI Developer）** ↓　（本編第1部.「AIとは」）

  AI システムを開発する事業者（AI を研究開発する事業者を含む）↓
  AI モデル・アルゴリズムの開発、データ収集（購入を含む）、前処理、AI モデル学習及び検証を通して AI モデル、AI モデルのシステム基盤、入出力機能等を含む AI システムを構築する役割を担う。[8]↵
  また、AI モデル・システムの開発及び実運用後も、特定領域におけるドメイン知識の拡充や環境の変化への対応、さらに人間の意図や価値観に沿った行動を実現するための調整（アライメント）を目的とした事後学習（Post Training）を通じて、AI モデルの性能を維持・改善することも役割として担う。↵

[8] 一般的には、AI 開発者は API 仕様策定や入出力設計、AI モデルを動作させるためのインフラ整備を担い、AI 提供者は UI/UX 設計や既存業務システムとの統合等を担う。よって、AI システムの構築の全てを AI 開発者が担うと整理されているわけではない。また、実際には多様なケースが存在するため、これらに限定されるものではない。↵

- 以下2点の記載間に生じる曖昧さを踏まえ、**AI開発者がAIシステムの構築の全てを担うわけではない**旨を脚注にて補足

1. 本編5頁にて、「AI開発者」は「AIシステムを開発する事業者」として定義されている
2. 本編6頁図「一般的なAI活用の流れにおける主体の対応」にて、「AI開発者」は「AIモデル検証」までが一般的な役割として明記されている

- ファインチューニング等、**AIモデル開発後のモデル調整（事後学習）をAI開発者が役割として担う**旨を明記

2. 令和７年度の更新内容

【更新内容の詳細】
②「一般的なAI活用の流れにおける主体の対応」の見直し

（本編第1部.「AIとは」）
（別添1.「A.AIに関する前提」）

「データ前処理・学習」と記載されていたところ、「学習」は「開発」の一部へと移動し、**「学習」の記載を当該箇所から削除**

「AIモデル事後学習」を追加

「システムへの実装（AIシステムの構築から連携）」の流れを整理

AI開発者

データ前処理
- データ収集：学習用のデータを収集する
- 前処理：収集したデータをAIモデル学習に適用させるための処理を行う

開発
- AIモデル学習・作成：データを学習させたAIモデルを作成する
- AIモデル検証：トライアルを通して作成したAIモデルの有用性を検証する
- AIモデル事後学習：特定の用途に適応させる等、追加の学習を実施する

AI提供者

システムへの実装
- AIシステム構築：AIモデルをシステムに導入し、AIシステムを構築する
- AIシステム検証：AIシステムの機能性や社会的影響を検証する
- AIシステム連携：既存/新設のシステムとAIシステムの連携を行う

提供

AI提供者からAI利用者にAIシステムが提供されるケース
- AIシステム提供運用支援：AI利用者にAIシステムを納品し、適正利用のための注意喚起等の運用支援を行う

AI利用者

利用

AI提供者からAI利用者にAIシステムが提供されるケース
- 適正利用・運用：AI開発者・AI提供者が意図する適正利用を継続する
- AI提供者からの注意喚起を参考にAIシステムを正常稼働させる運用を行う

AI提供者がAIシステムを運用しAI利用者がAIサービスを利用するケース
- AIサービス提供運用：AIシステムを正常に稼働させ、AI利用者にAIサービスを提供する
- 適正利用：AI開発者、AI提供者が意図する適正利用を継続する

業務外利用者
- 事業活動以外でAIを利用する者
- AIを直接事業で利用せずにAIシステム・サービスの便益を享受する、場合によっては損失を被る者

データ提供者
- 特定の法人や個人からデータを提供される場合に限らず、人、センサ、システム等からデータを入手する場合を含む
- データ提供においては、提供される者、もしくは入手する者にあたるAIの開発・提供・利用を担う者がデータを取り扱う際の責任を負う
- データ提供者は本ガイドラインの対象には含まれない

2. 令和７年度の更新内容

【更新内容の詳細】
③主体毎の役割の見直し（表「AIシステム・サービス例」）

（別添1.「A.AIに関する前提」）

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
⋮						
社内ナレッジ検索支援 AI	テキスト等生成	製造業の E 社向けに提供されている、社内文書を活用した質問応答支援システムである。E 社が所有するマニュアル、トラブル対応記録、過去の保守履歴などを対象に、自然言語での問い合わせに対応するチャットボットとして設計されている。 このサービスは、F 社が開発した大規模言語モデルをベースに、E 社固有の社内文書を用いてファインチューニングを実施することで、E 社の業務特性に最適化された回答を実現している。	F 社（AI モデル開発部門） E 社（ソリューション開発部門）	E 社（ソリューション開発部門）	E 社（情報システム部門）	-
プログラミング支援 AI	コード生成	AI モデル開発会社 D 社が提供する、ユーザーの入力に基づいてプログラミングコードの自動生成や既存コードの修正支援、バグ検出やコード品質の改善案の提示などを行うサービスである。コード生成 AI サービスは開発支援ツールとして利用されており、サービス全体の設計・運用も D 社が実施している。 本サービスの AI 利用者である、ソフトウェア開発を行う H 社は、D 社のコード生成 AI サービスによるプログラム作成支援を受けながら、独自の AI システム構築を行い、特定業務向けの AI サービスを実装した。このように H 社は AI 利用者でありながら AI 提供者としての区分も兼ねており、新規開発した AI サービスの保守・運用の役割も担う。	D 社（AI 開発部門）	D 社（プロダクト提供部門） H 社（ソフトウェア開発部門）	H 社（ソフトウェア開発部門）	-
売上予測 AI	テキスト等生成	小売業の I 社向けに提供される、店舗・商品別の売上予測を行う社内向け AI サービスである。このサービスは、過去の販売実績データや天候、販促施策、カレンダー情報などをもとに、週次・月次の売上を商品単位で予測し、在庫計画や発注業務の効率化を支援している。 サービスは Rest API として提供されており、売上予測モデルの API 仕様策定は A 社が担当。エンドポイント構成、パラメータ定義、リクエスト形式（例：対象店舗、商品 ID、予測期間など）、およびレスポンス形式（例：予測売上、予測精度指標など）が定められている。API コールの実装および社内システムへの統合は P 社が担当。店舗ごとの販売管理システムや在庫管理システムから自動的に API が呼び出されるように構成されており、店舗担当者や MD（マーチャンダイザー）が日常業務の中で自然に AI 予測の結果を活用できるようになっている[4]。	A 社（AI モデル設計部門）	P 社（システム導入部門）	I 社（経理部門）	-

- 同一事業者がファインチューニングを行い、AIシステムの提供を行う場合、AI開発者とAI提供者の区分を兼ねるものと整理

- コード生成AI等の支援を受けてAIシステムを提供した場合、AI利用者とAI提供者の区分を兼ねるものと整理

[4] AI モデルを API として設計・提供する場合、AI 開発者は主に入出力仕様やエンドポイント設計、OpenAPI 等による仕様定義を担い、AI 提供者はその API を実運用環境に組み込み、認証・公開・運用管理を行うものと本ガイドライン上では整理している。API 設計に関する詳細は、IPA（2025 年 3 月）「API 標準設計ガイド・基礎編」
https://www.ipa.go.jp/digital/data/jod03a000000a82y-att/api_standard_design_guide.pdf にて紹介されている。）

- APIの仕様定義等はAI開発者が担い、APIの実装処理はAI提供者が担うものと整理

2. 令和７年度の更新内容

【更新内容の詳細】
③主体毎の役割の見直し（アライメント）

（本編第1部.「AIとは」）

- **AI 開発者（AI Developer）**

AI システムを開発する事業者（AI を研究開発する事業者を含む）
AI モデル・アルゴリズムの開発、データ収集（購入を含む）、前処理、AI モデル学習及び検証を通して AI モデル、AI モデルのシステム基盤、入出力機能等を含む AI システムを構築する役割を担う。[5]
また、AI モデル・システムの開発及び実運用後も、特定領域におけるドメイン知識の拡充や環境の変化への対応、さらに人間の意図や価値観に沿った行動を実現するための調整（アライメント）を目的とした事後学習（Post Training）を通じて、モデルの性能を維持・改善することも役割として担う。

- アライメントを目的とした**モデル調整**（ファインチューニング等の事後学習）**はAI開発者**の役割と明記

（別添3.「AI開発者向け」）

[本編の記載内容 （再掲）] ※ 柱書のみ抜粋
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、多様な人々の多様な幸せ（well-being）の追求が可能となるように行動することが重要である。

⋮

[具体的な手法]
- AI システムが人間の意図や価値観に沿って適切に機能するようにするための取組（アライメント）の実施
- AI 倫理を所管する担当役員及び AI ガバナンスに関する社内組織を設置
- AI の開発時に参照可能な諸外国及び研究機関における生命倫理の議論の例は以下のとおり
  - 国際連合（UN）、世界保健機関（WHO）等の国際機関が発行するレポート 等
  - 大学等の学術機関の出す研究論文

（別添4.「AI提供者向け」）

[本編の記載内容 （再掲）] ※ 柱書のみ抜粋
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、多様な人々の多様な幸せ（well-being）の追求が可能となるように行動することが重要である。

⋮

[具体的な手法]
- 社会的、安全及びセキュリティ上のリスク軽減の研究の推進
  - 社会的、安全及びセキュリティ上のリスクを軽減するための研究、効果的な軽減策への投資
    （以下は研究内容の例）
    - 民主的価値の維持
    - 人権の尊重
    - 子供及び社会的弱者の保護
    - 知的財産権及びプライバシーの保護
    - 有害な偏見の回避
    - 偽・誤情報の回避
    - 情報操作の回避 等
  - リスク緩和に関する研究及びベストプラクティスの共有（可能な範囲で実施する）
- AI システムが人間の意図や価値観に沿って適切に機能するようにするための取組（アライメント）の実施

- **AIシステムが人間の意図や価値観に沿って適切に機能するようにすること**を手法として追記

2. 令和７年度の更新内容

【更新内容の詳細】
③主体毎の役割の見直し（アライメント）

（別添4.AI提供者向け）

P-2） ｉ．人間の生命・身体・財産、精神及び環境に配慮したリスク対策

✧ AI 利用者を含む関連するステークホルダーの生命・身体・財産、精神及び環境に危害を及ぼすことがないよう、提供時点で予想される利用条件下でのパフォーマンスだけでなく、様々な状況下で AI システムがパフォーマンスレベルを維持できるようにし、リスク（連動するロボットの制御不能、不適切な出力等）を最小限に抑える方法（ガードレール技術等）を検討する（「2）安全性」）

- インシデントの未然防止

⋮

➢ ガードレールの設定
  ✧ フィルタリング等により有害な出力を検知し制限する
➢ プロンプトの安全設計
  ✧ システムプロンプトの設計や、利用者の入力内容に対する適切な制限・誘導を行う

- アライメントの実行手段として考えられる「ガードレールの設定」や「プロンプトの安全設計」をAI提供者の留意事項として追記

（現ガイドライン別添93頁にAI開発者側の留意事項には一定記載が存在するため、AI提供者側の説明を補強する想定）

2. 令和７年度の更新内容

【更新内容の詳細】
③主体毎の役割の見直し（RAG）

（別添3.AI開発者向け）

データ前処理・学習時

D-3） i. データに含まれるバイアスへの配慮

✧ 学習データ、AI モデルの学習過程によってバイアス（学習データには現れない潜在的なバイアスを含む）が含まれうることに留意し、データの質を管理するための相当の措置を講じる（「3）公平性」）

⋮

[具体的な手法]

~~● RAG 活用時~~

~~➢ RAG を活用する場合、参照するデータの適切な取扱い~~

~~✧ 情報源の選択、データの前処理、チャンク化、ベクトルデータベース構築等を適切に実施する~~

⋮

- **参照するデータベース自体にバイアス**が含まれる場合は、バイアスへの対策とはなり難いと考えられるため、**当該記載は削除**

AI 開発時

D-6） i. 検証可能性の確保

✧ AI の予測性能及び出力の品質が、活用開始後に大きく変動する可能性又は想定する精度に達しないこともある特性を踏まえ、事後検証のための作業記録を保存しつつ、その品質の維持・向上を行う （「2）安全性」「6）透明性」）

⋮

[具体的な手法]

~~➢ RAG の導入による出力根拠の透明性向上~~

~~外部の情報源を検索して回答を生成する際、出典や引用元等を示すことが可能となる~~

（別添4.AI提供者向け）

● AI システム実装時

P-6） i．システムアーキテクチャ等の文書化

✧ トレーサビリティ及び透明性の向上のため、意思決定に影響を与える提供する AI システム・サービスのシステムアーキテクチャ、データの処理プロセス等について文書化する（「6）透明性」）

⋮

[具体的な手法]

● 説明可能性確保

⋮

➢ AI モデルの入出力傾向の分析

✧ AI に対する複数の入力と出力の組合せをもとに、AI の出力傾向を分析する（例えば、入力パターンを少しずつ変化させたときの出力の変化の観測等）

➢ RAG の導入

✧ 外部の情報源を検索して回答を生成する際、出典や引用元等を明記することが可能となる

- **RAGの導入はAI提供者**の役割と整理しているため、当該内容はAI開発者の留意事項からは削除し、AI提供者の留意事項に記載

2. 令和７年度の更新内容

「データ」「学習」など、多義的に捉えられる単語の定義や表現の見直し

【更新内容】

① 「学習」「推論」の定義の追加

- ✓ 多義的に捉えられる「学習」について明確な定義を記載（「学習」とは、「機械学習」を指しているのか、「In-Context-Learning（文脈内学習）」を指しているのか等）
- ✓ RAG等によりAI活用時に扱うデータが拡大している点を踏まえ、「推論」について明確な定義を記載

② 「データ」の定義・表現の見直し

- ✓ ガイドラインにおける「データ」の定義を記載するとともに、一般的に用いられる表現に見直し、明確化

【主な更新箇所】

- ✓ 本編第1部「AIとは」
- ✓ 別添1.「A.AIに関する前提」

2. 令和７年度の更新内容

【更新内容の詳細】
①「学習」「推論」の定義の追加

- **AIモデルのパラメータを決定するプロセス**を学習と明記（In-Context-Learningは学習には含まれない）
- ファインチューニング等を含む**事後学習**に関しても脚注にて定義
- **学習済みモデルに未知のデータを与えて出力を得るプロセス**を推論と明記（RAG等を介して外部データを参照する行為も推論に含まれる）

関連する用語　（本編第1部.「AIとは」）

⋮

- **学習**

学習とは、データを用いて AI モデルのパラメータを決定または改善するプロセスである。場合によって、AI モデルの汎化性能を形成し、パラメータを最適化する事前学習と、必要に応じた継続的なパラメータ調整を行う事後学習[17]の二つのプロセスから構成される。学習には、教師あり学習、教師なし学習、強化学習などの手法が含まれる。学習で用いるデータは、AI モデルの構築・評価に利用する訓練データ、検証データ、テストデータに分けられる。

- **推論**

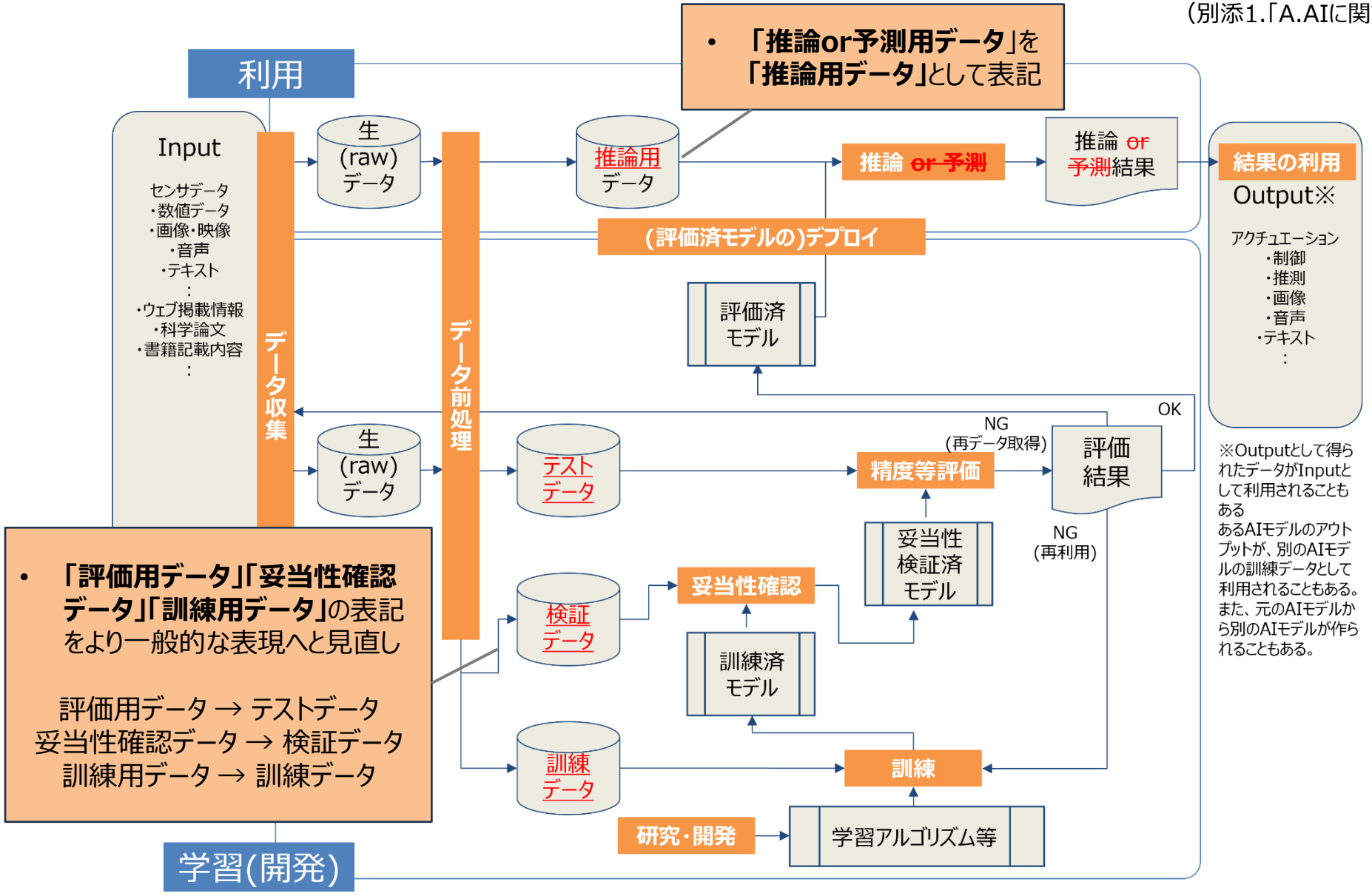
推論とは、学習済みモデルに未知のデータを与え、出力（予測・分類・生成など）を得るプロセスを指す。推論では、ユーザーが入力するプロンプトやセンサからの取得データに加え、RAG 等を介して外部知識を補完した情報等が用いられる。これらを統合した推論用データを AI モデルが処理することで、状況や特定の用途に即した回答を導き出す。

[17] 事後学習（Post-training）とは、事前学習済みの AI モデルに対して、特定の用途に適応させるために行われる追加の学習プロセスを指す。特定の専門知識の補完や、人間の意図・価値観に沿った応答を制御するアライメントを含み、実務環境の変化やニーズに応じて継続的に実施される。主な手法として、ファインチューニングや、データの追加による再学習などがある。

2. 令和７年度の更新内容

【更新内容の詳細】
②「データ」の定義・表現の見直し（図「AIの学習及び利用の流れの例」）

（別添1.「A.AIに関する前提」）



※Outputとして得られたデータがInputとして利用されることもある
あるAIモデルのアウトプットが、別のAIモデルの訓練データとして利用されることもある。また、元のAIモデルから別のAIモデルが作られることもある。

2. 令和７年度の更新内容

【更新内容の詳細】
②「データ」の定義・表現の見直し（新規追加表「AIの学習及び利用におけるデータ」）

（別添1.「A.AIに関する前提」）

	プロセス	概要	データ種類	概要	具体例
AIの学習及び利用の流れ	学習（機械学習）	データを用いてモデルのパラメータを決定または改善するプロセス	訓練データ (training data)	モデルのパラメータを最適化するために使用するデータである。学習アルゴリズムはこのデータに基づいて誤差を最小化し、入力と出力の関係を学習する	• 内部データ • 大規模オープンデータ（CIFAR-10,MNIST等） • 利害関係者などのデータ • センサ、システムから収集されたデータ
			検証データ (validation data)	モデルの学習過程において、訓練データとは別に使用されるデータである。モデルのパラメータの更新には使用されず、主にハイパーパラメータの調整や過学習の検出等、モデルの性能を中間的に評価するために用いられる。	
			テストデータ (test data)	モデル学習が完了した後にモデルの最終的な性能を評価するためのデータである。訓練や検証に使われていないため、適切な汎化性能の指標の基となる。	
	推論	学習済みモデルに未知のデータを与え、出力（予測・分類・生成など）を得るプロセス	推論用データ (data for inference)	推論用データとは、学習済みモデルが新しい入力に対して出力を生成する際に利用するデータである。これには、AIモデルが直接処理するユーザーからの指示や運用環境で取得されるデータに加え、文脈付与や精度向上のためにRAG等を通じて参照される内部データベースや外部知識などの追加情報が含まれる。これらを組み合わせることで、AIモデルの応答の正確性と一貫性を高められることが期待される。	• ユーザーからの入力（プロンプト、画像、音声） • 運用環境で取得されるデータ（センサ取得データ等） • 内部情報（FAQ、ナレッジベース） • 外部情報（Web検索結果、外部API） • コンテキスト情報（過去の会話履歴、セッション情報） • 他モデルの出力（別のAIモデルが生成したテキストや解析結果）

2. 令和７年度の更新内容

AI事業者ガイドラインの活用を支援する資料・ツールを検討

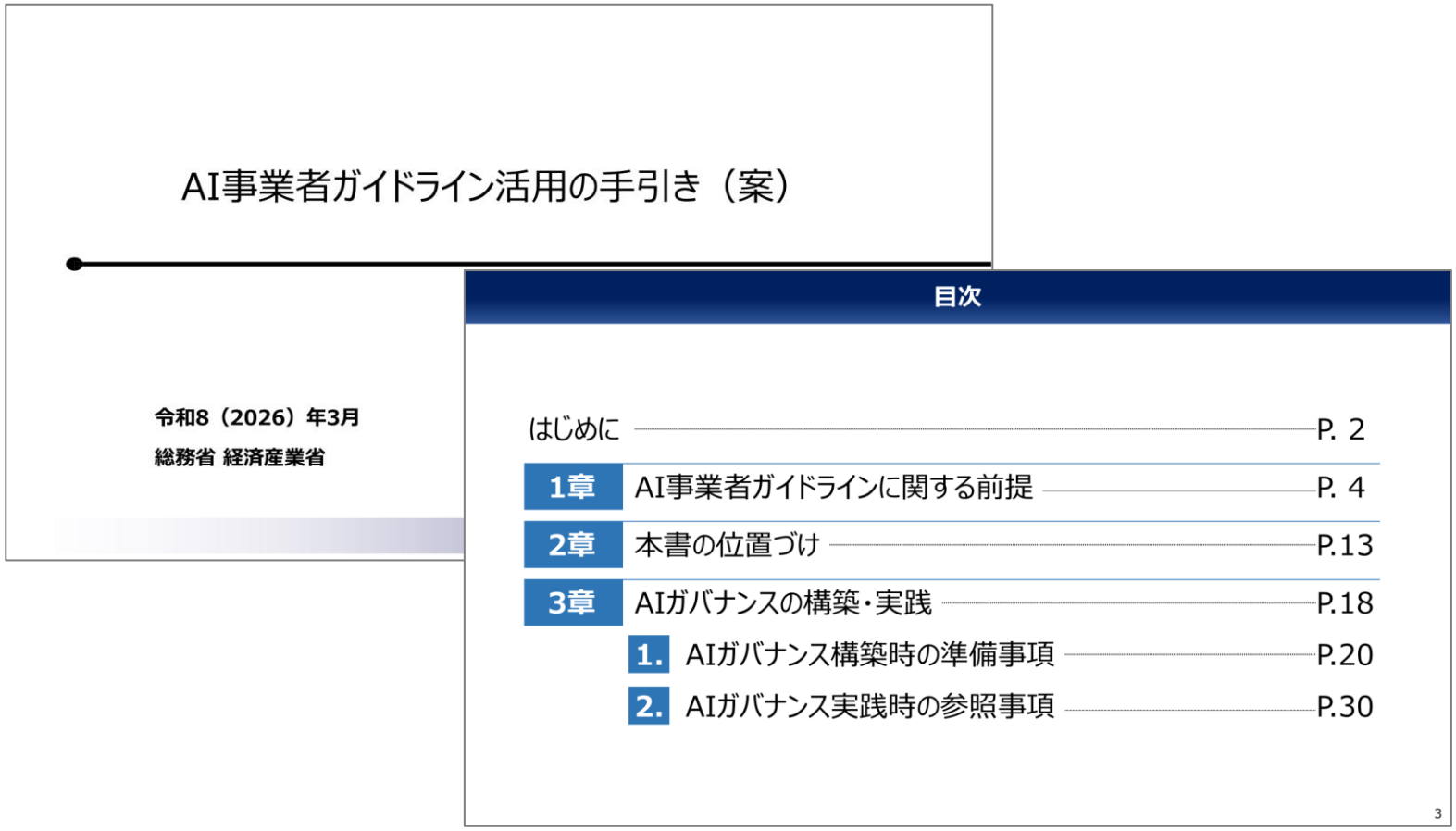
	経済産業省	総務省
検討中の資料・ツール	AI事業者ガイドライン活用の手引き（略称「活用の手引き」）	チャットボット（ルールベースAI）

概要	• 「活用の手引き」は、**「AI事業者ガイドライン」の活用を補助する目的**で作成
	• **特にAIガバナンスの構築・実践をこれから始める方々向け**
	• 内容としては、以下を紹介する構成
	  • AI事業者ガイドラインを活用する上で**前提となる考え方**
	  • AIガバナンスの構築時に準備・土台として**はじめに着手すると良いこと**
	  • AIガバナンスの実践時の**AI事業者ガイドラインの参照の仕方、活用例**
	• 令和7年度末に「活用の手引き（案）」をリリースし、来年度以降も拡充・見直しについて検討予定
		• AI開発者・AI提供者・AI利用者が確認したい情報に対して、素早くアクセスすることを支援
		• AI利活用における主体や確認したい事項の選択、及びフリーテキスト入力の2つの方法で、**AI事業者ガイドラインに関する情報**を確認可能
		• 令和７年度末を目途にリリースし、来年度以降も質問・回答の見直しについて検討予定

（ご参考） AI事業者ガイドライン活用の手引き

経済産業省のユーザビリティの改善の取り組みとして、資料として送付した「AI事業者ガイドライン活用の手引き（案）」を今年度末に公開することを予定

■AI事業者ガイドライン活用の手引き（案）



AI事業者ガイドライン活用の手引き（案）

令和8（2026）年3月

総務省 経済産業省

目次

はじめに P. 2

1章 AI事業者ガイドラインに関する前提 P. 4

2章 本書の位置づけ P.13

3章 AIガバナンスの構築・実践 P.18

1. AIガバナンス構築時の準備事項 P.20

2. AIガバナンス実践時の参照事項 P.30

3

2. 令和７年度の更新内容

人工知能戦略本部等の国内動向や広島AIプロセス等の国際動向において、注視するべき最新状況等を追記

【更新内容】

主に以下のトピックを反映

＜国内＞

①内閣府「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」

②総務省「自治体におけるAI活用・導入ガイドブック＜導入手順編＞（第4版）」

③総務省「AIのセキュリティ確保のための技術的対策に係るガイドライン」

④デジタル庁「行政の進化と革新のための生成AIの調達・利活用に係るガイドライン」

＜国際＞

⑤広島AIプロセス

- ✓ 報告枠組みの参加組織数の追記
- ✓ 参加組織の声の追記
- ✓ 参照する文書を一部変更

【主な更新箇所】

- ✓ ①・②・④・⑤：本編「はじめに」
- ✓ ②・④　：別添2.「E.ガバナンスの構築」関連
- ✓ ③　：本編第2部「C.共通の指針」
- ✓ ⑤　：本編第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」
- ✓ ⑤　：本編第3部「AI開発者に関する事項」
- ✓ ③・⑤　：別添3.「AI開発者向け」
- ✓ ③　：別添4.「AI提供者向け」
- ✓ ⑤　：別添7.「チェックリスト」
- ✓ ⑤　：別添9.「海外ガイドラインとの比較表」

2. 令和７年度の更新内容

【更新内容の詳細】

①内閣府「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」

（本編「はじめに」）

AI 関連技術は日々発展をみせ、AI の利用機会及び様々な可能性は拡大の一途をたどり、産業におけるイノベーション創出及び社会課題の解決に向けても活用されている。また、~~近年台頭してきた~~対話型の生成 AI によって「AI の民主化」が起こり、多くの人が「対話」によって AI を様々な用途へ容易に活用できるようになった。これにより、企業では、ビジネスプロセスに AI を組み込むだけではなく、AI が創出する価値を踏まえてビジネスモデル自体を再構築することにも取り組んでいる。また、個人においても自らの知識を AI に反映させ、自身の生産性を拡大させる取組が加速している。我が国では、従来から Society 5.0 として、サイバー空間とフィジカル空間を高度に融合させたシステム（CPS：サイバー・フィジカルシステム）による経済発展と社会的課題の解決を両立する人間中心の社会というコンセプトを掲げてきた。このコンセプトを実現するにあたり、AI が社会に受け入れられ適正に利用されるため、2019 年 3 月に「人間中心の AI 社会原則」が策定された。一方で、AI 技術の利用範囲及び利用者の拡大に伴い、リスクも増大している。特に生成 AI に関して、知的財産権の侵害、偽情報・誤情報の生成・発信等、これまでの AI ではなかったような新たな社会的リスクが生じており、AI がもたらす社会的リスクの多様化・増大が進んでいる。こうした状況を踏まえ、「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」（令和 7 年法律第 53 号）が 2025 年 6 月に公布、9 月に全面施行された[1,2]。

[1] https://www8.cao.go.jp/cstp/ai/ai_act/ai_act.html
[2] 同年 12 月に、同法第 13 条に基づき、全ての AI に関連する主体における AI 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の「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」（令和 7 年 12 月 19 日人工知能戦略本部決定）が策定された。https://www8.cao.go.jp/cstp/ai/ai_guideline/ai_guideline.html

②総務省「自治体におけるAI活用・導入ガイドブック＜導入手順編＞（第4版）」

（本編「はじめに」）

[4] 自治体向けの AI の利用方法や利用上の留意事項については、本ガイドラインを参考としつつ、自治体に特化した内容を記載した「自治体における AI 活用・導入ガイドブック＜導入手順編＞」（2025 年 12 月改訂）が公表されている。改訂後の本ガイドブックは、特に生成 AI の利用用により飛躍的な業務効率化が期待される点を、自治体における利活用事例とあわせて提示するとともに、ガバナンス確保のための体制構築、要機密情報の取扱い、人材育成の考え方等の留意事項についても提示している。
https://www.soumu.go.jp/main_content/000820109.pdf

（別添2.「E.ガバナンスの構築」）

- 官公庁・自治体
  デジタル庁「行政の進化と革新のための生成 AI の調達・利活用に係るガイドライン」（2025 年 5 月）デジタル庁「~~テキスト生成 AI 利活用におけるリスクへの対策ガイドブック（α版）~~」（~~2024 年 6 月~~）[46]
  総務省「自治体における AI 活用・導入ガイドブック＜導入手順編＞（第 4 版）」（2025 年 12 月）[47]
  東京都「文章生成 AI 利活用ガイドライン」（2024 年 4 月）[48]

[47] https://www.soumu.go.jp/main_content/000820109.pdf

③総務省「AIのセキュリティ確保のための技術的対策に係るガイドライン」

（本編第2部「C.共通の指針」）

[42] 総務省のサイバーセキュリティタスクフォースの下で開催された「AI セキュリティ分科会」の取りまとめ（2025 年 12 月）を踏まえ、AI 開発者及び AI 提供者における、LLM 及び LLM を構成要素に含む AI システムに対する脅威への技術的対策例を整理した「AI のセキュリティ確保のための技術的対策に係るガイドライン」を 2026 年 3 月に策定予定
https://www.soumu.go.jp/main_sosiki/kenkyu/cybersecurity_taskforce/index.html

（別添3.「AI開発者向け」）

[参考文献]
- 経済産業省「OSS の利活用とそのセキュリティ確保に向けた管理手法に関する事例集」（2021 年 4 月）
- 経済産業省「ソフトウェア管理に向けた SBOM の導入に関する手引 ver2.0」（2024~~3~~ 年 ~~8~~7 月）
- 総務省「AI のセキュリティ確保のための技術的対策に係るガイドライン」（2026 年 3 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）
- 独立行政法人情報処理推進機構「セキュリティ・バイ・デザイン導入指南書」（2022 年 8 月）
- NCSC, “Guidelines for secure AI system development”（2023 年 11 月）
- NIST, “The NIST CYBERSECURITY FRAMEWORK (CSF) 2.0”（2024 年 2 月）
- ISO/IEC27000 シリーズ
- NIST, SP800 シリーズ

（別添4.「AI提供者向け」）

[参考文献]
- 総務省「AI のセキュリティ確保のための技術的対策に係るガイドライン」（2026 年 3 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）
- 独立行政法人情報処理推進機構「セキュリティ・バイ・デザイン導入指南書」（2022 年 8 月）
- NCSC, “Guidelines for secure AI system development”（2023 年 11 月）
- ACSC, “Engaging with Artificial Intelligence（AI）”（2024 年 1 月）

2. 令和７年度の更新内容

【更新内容の詳細】

④デジタル庁「行政の進化と革新のための生成AIの調達・利活用に係るガイドライン」

（本編「はじめに」）

[3] 政府は様々な業務への生成 AI の利活用促進とリスク管理を表裏一体で進めるため、「行政の進化と革新のための生成 AI の調達・利活用に係るガイドライン」（令和７年５月 27 日 デジタル社会推進会議幹事会決定）を策定した。これは、政府における AI ガバナンスやベストプラクティスの共有体制、生成 AI の調達・利活用において留意すべきリスク等についての考え方、政府が利活用する生成 AI 全体の機能性や品質及び費用対効果の向上等について、AI 事業者ガイドラインや「政府機関等のサイバーセキュリティ対策のための統一基準群」等の既存のガイドライン及び諸外国政府のルールの動向等を踏まえ整理し、国の政府職員等向けのガイドラインとして示したものである。
https://www.digital.go.jp/resources/standard_guidelines#ds920

（別添2.「E.ガバナンスの構築」）

- 官公庁・自治体
  デジタル庁「行政の進化と革新のための生成 AI の調達・利活用に係るガイドライン」（2025 年 5 月）デジタル庁「~~テキスト生成 AI 利活用におけるリスクへの対策ガイドブック（α版）~~」（~~2024 年 6 月~~）[46]
  総務省「自治体における AI 活用・導入ガイドブック＜導入手順編＞（第 4 版）」（2025 年 12 月）[47]
  東京都「文章生成 AI 利活用ガイドライン」（2024 年 4 月）[48]

[46] ~~https://www.digital.go.jp/resources/generalitve-ai-guidebook~~https://www.digital.go.jp/resources/standard_guidelines#ds920

⑤広島AIプロセス（報告枠組みの参加組織数の追記・参加組織の声の追記）

（主たる更新箇所を抜粋）

（本編「はじめに」）

[11] 2023 年 5 月の G7 広島サミットの結果を受けて、生成 AI に関する国際的なルールの検討を行うため、「広島 AI プロセス」を立ち上げた。その後、同年 9 月の「広島 AI プロセス閣僚級会合」、10 月の京都 IGF での「マルチステークホルダーハイレベル会合」等を経て発出された「広島 AI プロセスに関する G7 首脳声明」を踏まえ、同年 12 月に「G7 デジタル・技術大臣会合」を開催し、同年の成果として、「広島 AI プロセス包括的政策枠組み」を取りまとめた。さらに、2024 年には、3 月の「G7 産業・技術・デジタル大臣会合」、10 月の「G7 デジタル・技術大臣会合」を経て、国際行動規範の遵守状況に係る「報告枠組み」が、同年 12 月に G7 で合意された。https://www.soumu.go.jp/hiroshimaaiprocess/
なお、2026 年 3 月現在、日本企業●社を含む●組織が回答を提出し、OECD のウェブサイト上で公表されている。
https://transparency.oecd.ai/reports

（本編第3部「AI開発者に関する事項」）

当該「行動規範」の遵守状況を高度な AI システムを開発する AI 開発者自らが自主的に確認し報告する「報告枠組み」が OECD との協力の下、G7 で合意され、2025 年 2 月より運用開始されている[59]。当該「行動規範」を遵守した高度な AI システムを開発する AI 開発者は、「報告枠組み」に参加することが期待されている。
なお、「報告枠組み」への参加により、AI ガバナンスに関する透明性の確保のみならず、以下のような組織内でのメリットがあることが報告されている[60]。

- 信頼できる AI の実現に向けたチーム間の連携強化
- ガバナンスの取組を国際基準と比較できるベンチマーク機能
- AI ガバナンスの構造に関する社内コミュニケーションの明確化
- リスクマネジメント分野におけるリソース配分の可視化の増進

[59] 広島プロセス国際行動規範「報告枠組み」 https://transparency.oecd.ai/
2026 年 3 月現在、日本企業●社を含む●組織が回答を提出し、OECD のウェブサイト上で公表されている。
https://transparency.oecd.ai/reports
[60] 総務省 広島 AI プロセス 報告枠組み 「報告枠組み参加組織の声」
https://www.soumu.go.jp/hiroshimaaiprocess/report.html

2. 令和７年度の更新内容

【更新内容の詳細】

⑤広島AIプロセス（参照する文書を一部変更）

（主たる更新箇所を抜粋）

（本編第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」）

D. 広島 AI プロセス「全ての AI 関係者向けの広島プロセス国際指針」~~高度な AI システムに関係する事業者に共通の指針~~

全ての~~高度な~~ AI システムに関係する事業者は、広島 AI プロセスを経て策定されたの「全ての AI 関係者向けの広島プロセス国際指針」~~及びその基礎となる「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指針」~~を踏まえ、~~「共通の指針」~~に加え、以下を遵守に即した行動をすべきである[49]。~~ただし、i)～xi) は高度な AI システムを開発する AI 開発者にのみ適用される内容もあるため、第 3～5 部に後述のとおり、AI 提供者及び AI 利用者は適切な範囲で遵守することが求められる。~~

1．我々は、安全、安心、信頼できる AI を適切かつ関連性をもって推進する上での、全ての AI 関係者の責任を強調する。我々は、ライフサイクル全体にわたる関係者が、AI の安全性、安心、信頼性に関して、異なる責任と異なるニーズを持つことを認識する。我々は、全ての AI 関係者が、自らの能力とライフサイクルにおける役割を十分に考慮した上で、「高度な AI システムを開発する組織の向けの広島プロセス国際指針（2023 年 10 月 30 日）」[50]を読み、理解することを奨励する。

2．「高度な AI システムを開発する組織の向けの広島プロセス国際指針」の以下の 11 の原則は、高度な AI システムを開発する組織にのみ適用可能な要素もあることを認識しつつ、高度な AI システムの設計、開発、導入、提供及び利用をカバーするために、全ての AI 関係者に対し、適時適切に、適切な範囲で、適用されるべきである。

① AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度な AI システムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる

② 市場投入を含む導入後、脆弱性、及び必要に応じて悪用されたインシデントやパターンを特定し、緩和する

③ 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、十分な透明性の確保を支援することで、アカウンタビリティの向上に貢献する

④ 産業界、政府、市民社会、学界を含む、高度な AI システムを開発する組織間での責任ある情報共有とインシデントの報告に向けて取り組む

⑤ 特に高度な AI システム開発者に向けた、個人情報保護方針及び緩和策を含む、リスクベースのアプローチに基づく AI ガバナンス及びリスク管理方針を策定し、実施し、開示する

⑥ 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、内部脅威に対する安全対策を含む、強固なセキュリティ管理に投資し、実施する

⑦ 技術的に可能な場合は、電子透かしやその他の技術等、ユーザーが AI が生成したコンテンツを識別できるようにするための、信頼できるコンテンツ認証及び来歴のメカニズムを開発し、導入する

⑧ 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、効果的な軽減策への投資を優先する

⑨ 世界の最大の課題、特に気候危機、世界保健、教育等（ただしこれらに限定されない）に対処するため、高度な AI システムの開発を優先する

⑩ 国際的な技術規格の開発を推進し、適切な場合にはその採用を推進する

⑪ 適切なデータインプット対策を実施し、個人データ及び知的財産を保護する

3．また、AI 関係者は第 12 の指針に従うべきである

⑫ 高度な AI システムの信頼でき責任ある利用を促進し、貢献する。

AI 関係者は、高度な AI システムが特定のリスク（例：偽情報の拡散に関するもの）をどのように増大させるか及び／又は新たなリスクをどのように生み出すかといった課題を含め、自分自身そして必要に応じて他者のデジタル・リテラシー、訓練及び認識を向上させる機会を求めるべきである。全ての関連する AI 関係者は、高度な AI システムの新たなリスクや脆弱性を特定し、それに対処するために、必要に応じて、協力し情報を共有することが奨励される。

【更新内容の詳細】

⑤広島AIプロセス（参照する文書を一部変更）

（主たる更新箇所を抜粋）

（本編第3部「AI開発者に関する事項」）

~~広島 AI プロセス~~「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~~

高度な AI システムを開発する AI 開発者については、~~上記に加え、「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」及び~~「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」[57]を参照することが望ましい~~遵守すべきである~~。

~~以下、「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」との比較において、当該「行動規範」において追加的に記載されている事項を示す。なお、当該「行動規範」全体の内容については、「別添 3. C. 高度な AI システムの開発にあたって遵守すべき事項」を参照のこと。~~

I. AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度な AI システムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる。

これには、レッドチーミング等の評価方法を組み合わせて、多様な内部テスト手段や独立した外部テスト手段を採用することや、特定されたリスクや脆弱性に対処するために適切な緩和策を実施することが含まれる。テストと緩和策は、例えば、システムが不合理なリスクをもたらさないように、ライフサイクル全体を通じてシステムの信頼性、安全性、セキュリティの確保を目指すべきである。このようなテストを支援するために、開発者は、データセット、プロセス、システム開発中に行われた意思決定に関連して、トレーサビリティを可能にするよう努めるべきである。これらの対策は文書化され、定期的に更新される技術文書によってサポートされるべきである。

このようなテストは、リスクと脆弱性を特定するため、また、偶発的か意図的かを問わず、セキュリティ、安全性、社会的リスク、その他のリスクに対処するための行動を通知するために、安全な環境で実施されるべきであり、また、特に導入前及び市場投入前等の AI ライフサイクル全体におけるいくつかのチェックポイントで実施されるべきである。テスト措置の設計と実施において、組織は以下のリスクに適切に注意を払うことを約束する：

- 高度な AI システムが、非国家主体も含め、兵器の開発、設計の取得、使用への参入障壁を低くする方法等、化学、生物、放射性、核のリスク。
- 攻撃的サイバー能力とは、システムが脆弱性の発見、悪用、又は作戦上の利用を可能にする方法等であり、そのような能力の有用な防御的応用の可能性があり、システムに含めることが適切であるかもしれないことを念頭に置くこと。
- 健康及び/又は安全に対するリスク。システムの相互作用やツールの使用による影響を含み、例えば物理的なシステムを制御したり、重要なインフラに干渉したりする能力を含む。
- モデルが自分自身のコピーを作ったり、"自己複製 "したり、他のモデルを訓練したりすることによるリスク。
- 高度な AI システムやモデルが有害な偏見や差別を生じさせたり、プライバシーやデータ保護等適用される法的枠組みへの違反につながったりする可能性等、社会的リスクや個人やコミュニティに対するリスク。
- 偽情報の助長やプライバシーの侵害等、民主主義の価値や人権に対する脅威。
- 特定の事象が連鎖反応を引き起こし、都市全体、領域活動全体、地域社会全体にまで重大な悪影響を及ぼすリスク。

各組織は、セクターを超えた関係者と協力して、これらのリスク、特にシステミック・リスクに対処するための緩和策を評価し、採用することを約束する。

また、これらのコミットメントに取り組む組織は、高度な AI システムのセキュリティ、安全性、偏見と偽情報、公平性、説明可能性と解釈可能性、透明性に関する研究と投資を促進し、悪用に対する先進的 AI システムの堅牢性と信頼性を高めることに努めるべきである。

- ~~リスク軽減のための緩和策を文書化するとともに、定期的に更新すべき。また、各主体はセクターを超えた関係者と連携してこれらのリスクへの緩和策を評価し、採用すべき~~

II. 市場投入を含む導入後、脆弱性、及び必要に応じて悪用されたインシデントやパターンを特定し、緩和する。

組織は、リスクレベルに見合った適切なタイミングで、AI システムを意図したとおりに使用し、導入後の脆弱性、インシデント、新たなリスク、悪用を監視し、それらに対処するための適切な措置を講じるべきである。組織は、例えば、責任を持って弱点を開示するインセンティブを与えるための報奨金制度、コンテスト、賞品等を通じて、導入後に第三者やユーザーが問題や脆弱性を発見し報告することを促進することの検討が奨励される。組織はさらに、他の利害関係者と協力して、報告されたインシデントの適切な文書化を維持し、特定されたリスクと脆弱性を軽減することが奨励される。適切な場合には、脆弱性を報告する仕組みは、多様な利害関係者が利用できるものでなければならない。

- ~~報奨金制度、コンテスト、賞品等を通じて、責任を持って弱点を開示するインセンティブを与えることの検討を奨励~~

2. 令和７年度の更新内容

【更新内容の詳細】

⑤広島AIプロセス（参照する文書を一部変更）

（主たる更新箇所を抜粋）

（別添7「チェックリスト」）

別添7 B　チェックリスト（案）　令和8年●月●●日

第2部D. 広島AIプロセス「全ての AI 関係者向けの広島プロセス国際指針」

本チェックリストは、AI事業者ガイドライン「第2部D.広島AIプロセス「全ての AI 関係者向けの広島プロセス国際指針」」の項目です。取組事項のチェックにご活用ください

※①～⑪については、適時適切に、適切な範囲で、適用されるべきである。また、⑫については、従うべきである。

チェック項目

- □ ①AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度なAIシステムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じているか？
- □ ②市場投入を含む導入後、脆弱性、及び必要に応じて悪用されたインシデントやパターンを特定し、緩和しているか？
- □ ③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、十分な透明性の確保を支援することで、アカウンタビリティの向上に貢献しているか？
- □ ④産業界、政府、市民社会、学界を含む、高度なAI システムを開発する組織間での責任ある情報共有とインシデントの報告に向けて取り組んでいるか？
- □ ⑤特に高度な AI システム開発者に向けた、個人情報保護方針及び緩和策を含む、リスクベースのアプローチに基づく AI ガバナンス及びリスク管理方針を策定し、実施し、開示しているか？
- □ ⑥ 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、内部脅威に対する安全対策を含む、強固なセキュリティ管理に投資し、実施する
- □ ⑦技術的に可能な場合は、電子透かしやその他の技術等、ユーザーが AI が生成したコンテンツを識別できるようにするための、信頼できるコンテンツ認証及び来歴のメカニズムを開発し、導入しているか？
- □ ⑧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、効果的な軽減策への投資を優先しているか？
- □ ⑨世界の最大の課題、特に気候危機、世界保健、教育等（ただしこれらに限定されない）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？
- □ ⑩国際的な技術規格の開発を推進し、適切な場合にはその採用を推進しているか？
- □ ⑪適切なデータインプット対策を実施し、個人データ及び知的財産を保護しているか？
- □ ⑫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？

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

（別添9.「海外ガイドラインとの比較表」）

AI事業者ガイドライン本編		AI事業者ガイドライン別添		②G7 広島AIプロセス
はじめに	-	別添.はじめに	-	
第１部 AIとは	-	別添1.第1部関連	A.AIに関する前提	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（2023年10月, 広島AIプロセスに関するG7首脳声明）
			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	-	-	概要版「各主体に共通の指針」にて広島AIプロセス包括的政策枠組み(2023年12月, 広島AIプロセスG7デジタル・技術閣僚声明)を引用
	D.広島AIプロセス「全ての AI 関係者向けの広島プロセス国際指針」	-	-	広島AIプロセス包括的政策枠組み II. 全てのAI関係者向け及び高度な AI システムを開発する組織向けの広島プロセス国際指針(2023年12月, 広島AIプロセスG7デジタル・技術閣僚声明)
	E.ガバナンスの構築	別添2.「第2部E.AIガバナンスの構築」 関連	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ガバナンスの構築に関する実際の取組事例	
第３部 AI開発者に関する事項	データ前処理・学習時	別添3.AI開発者向け	A. 本編「第３部　AI開発者に関する事項」 の解説	
	AI開発時			
	AI開発後			
	-		B. 本編「第2部」の「共通の指針」の解説	
	広島AIプロセス「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」		C. 広島AIプロセス「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」	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（2023年10月, 広島AIプロセスに関するG7首脳声明）

2. 令和７年度の更新内容

別添2.「第2部E.AIガバナンスの構築」関連の「B.AIガバナンスの構築に関する実際の取組事例」について、事業者の取り組みの追加・更新

【更新内容】

以下のコラムを更新・追加

＜既存コラムの更新＞

① コラム５（NECグループ）
② コラム６（東芝グループ）
③ コラム８（富士通グループ）
④ コラム９（ソフトバンク）
⑤ コラム10（NTT DATA）

✓リスク分析の体制・進め方・具体例や、AI利用の拡大・AI技術の進展を踏まえた組織体制の強化・ルールの見直し等に関する取組を反映

＜新規コラムの追加＞

⑥ IBM

✓責任ある技術委員会（Responsible Technology Board）を中心としたAIガバナンス推進体制や、日本のAI倫理チームにおける取組を紹介

⑦ Amazon Web Services

✓ISO/IEC 42001認証取得・8項目から成るAIガバナンスポリシーの策定・責任あるAIの開発におけるステークホルダーとの連携に関する取組を紹介

【更新箇所】

✓別添2.「第2部E.AIガバナンスの構築」関連の「B.AIガバナンスの構築に関する実際の取組事例」

2. 令和７年度の更新内容

- 脚注記載内容・リンクの更新　（※一部抜粋して掲載）

【更新内容の詳細】

■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」

（別添3.「AI開発者向け」）

D-6） i. 検証可能性の確保

✧ AI の予測性能及び出力の品質が、活用開始後に大きく変動する可能性又は想定する精度に達しないこともある特性を踏まえ、事後検証のための作業記録を保存しつつ、その品質の維持・向上を行う （「2）安全性」「6）透明性」）

⋮

- 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 2025.04 版」（2025~~3~~ 年 ~~4~~6 月）

■AISI「AIセーフティに関する評価観点ガイド（第1.10版）」「AIセーフティに関するレッドチーミング手法ガイド（第1.10版）」

（別添2.「E.ガバナンスの構築」関連）

別添 2.「第 2 部 E.AI ガバナンスの構築」関連

別添 1.B.「AI による便益/リスク」にて述べたとおり、AI の便益を享受しリスクを抑制するためには、AI に関する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つ、そこからもたらされる便益を最大化するための、AI ガバナンスの構築が重要となる。その際、常に変化する環境及びゴールを踏まえ、最適な解決策を適用し、適切に作動しているか評価・見直し続けることが各主体に期待される[22]。

⋮

[22] AI のセーフティ評価を行う際の基本的な考え方としては、AI セーフティ・インスティテュート（AISI）からは「AI セーフティに関する評価観点ガイド（第 1.10 版）」が公開されており、AI システムの開発者や提供者が AI のセーフティ評価を実施する際に参照できる内容となっている。また、同組織からは、AI セーフティ評価手法の一つとして「レッドチーミング手法」が紹介されている。
AISI「AI セーフティに関する評価観点ガイド（第 1.10 版）」（2025~~4~~ 年 ~~3~~9 月）
https://aisi.go.jp/effort/effort_framework/guide_to_evaluation_perspective_on_ai_safety/
AISI「AI セーフティに関するレッドチーミング手法ガイド（第 1.10 版）」（2025~~4~~ 年 ~~3~~9 月）
https://aisi.go.jp/effort/effort_framework/guide_to_red_teaming_methodology_on_ai_safety/

■一般社団法人日本ディープラーニング協会（JDLA）「G検定」

（別添2.「E.ガバナンスの構築」関連）

行動目標 3-2【AI マネジメントシステムの人材リテラシー向上】：
各主体は、経営層のリーダーシップの下、AI マネジメントシステムを適切に運営するために、外部の教材の活用も検討し、AI リテラシーを戦略的に向上させることが期待される。例えば、AI システム・サービスの法的・倫理的側面に責任を負う役員、マネジメントチーム、担当者には、AI 倫理及び AI の信頼性に関する一般的なリテラシー向上のための教育を、AI システム・サービスの開発・提供・利用プロジェクトの担当者には AI 倫理だけではなく生成 AI を含む AI 技術に関する研修を、全者に対して AI マネジメントシステムの位置づけ及び重要性についての教育を提供することが考えられる。

⋮

当社は、研修対象者の到達度を計るための JDLA の検定試験シラバスにもとづいたプログラムを活用している。JDLA の G 検定は、AI 技術の基礎から AI 倫理まで幅広く含む内容である。また、JDLA 主催の G202~~5~~3#~~5~~3（2025~~3~~ 年 ~~9~~7 月 ~~6~~7 日実施）の G 検定合格者アンケートにおいて学習時間は ~~15~~30～~~30~~50 時間と答えた合格者が約 25%~~43 割~~と最多であり[69]、研修対象者に過度な負担にならないことも確認している。

■一般社団法人金融データ活用推進協会（FDUA）「金融生成AIガイドライン（第1.1版）」

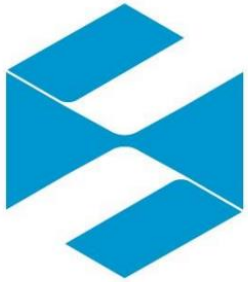
（別添2.「E.ガバナンスの構築」関連）

行動目標 3-1-1【業界の標準的な乖離評価プロセスとの整合性の確保】：各主体は、経営層のリーダーシップの下、業界における標準的な乖離評価プロセスの有無を確認し、そのようなプロセスが存在する場合には、自社のプロセスに取り込むことが期待される。

⋮

- 金融
  金融庁「モデル・リスク管理に関する原則」（2021 年 11 月）[54]
  ~~一般社団法人金融データ活用推進協会（FDUA）「金融生成 AI 実務ハンドブック」（2024 年 5 月）[55]~~
  一般社団法人金融データ活用推進協会（FDUA）「金融生成 AI ガイドライン（第 1.1 版）」（2025~~4~~ 年 ~~8~~7 月）[56]

総務省
Ministry of Internal Affairs and Communications



経済産業省
Ministry of Economy, Trade and Industry