ChatGPT

日本の「AI事業者ガイドライン（第1.2版）」実務分析

エグゼクティブサマリ

日本の「AI事業者ガイドライン（第1.2版）」（2026-03-31公表）は、従来の“生成AIの利用”中心の論点に加え、**AIエージェント**や**フィジカルAI**といった「自律性・外部連携・物理世界への作用」を前提にしたリスクを、用語定義・リスク整理・対策の観点で明確に押し出した更新と位置づけられる。特に、**「学習」「推論」**の定義にRAG等の外部知識連携を含めて記述し、AIの意思決定・入出力・データ来歴を追跡できる状態（トレーサビリティ）をアカウンタビリティの土台に据えることで、実務上は「PoC→本番」移行の要件が上がる。 [1]

実務インパクトは以下に集約できる。
第一に、AI適用範囲が**「単体モデル」から「システム（ワークフロー）＋外部API＋権限（ID/認証）＋運用（監視・更新）」**へ拡張され、セキュリティ・事故時対応・責任分配を含む運用設計が不可欠になる（AIエージェント/フィジカルAIの定義が、そのまま“設計チェック項目”になる）。 [2]
第二に、説明可能性は「モデル解釈」だけでなく、**誰に何を説明するか（説明の受け手の必要性の共有）**を含む“対ステークホルダー設計”として書かれており、契約・調達・顧客説明（FAQ/免責/注意事項/利用制限）と一体で整備する必要が生じる。 [3]
第三に、**サプライヤーを含むステークホルダーへの説明**や、価値連鎖/リスク連鎖全体でのリスク削減が前提化され、ベンダ管理・委託先統制・SBOM/モデル来歴・監査対応の“準備コスト”が顕在化する。 [4]

推奨アクション（実務担当者向けの最短ルート）は、次の3点を最優先に据えるべきである。

- **AI棚卸し（AIシステム/サービス単位）＋役割（開発者/提供者/利用者）マッピング**：社内の生成AI導入、外部SaaS利用、業務委託（開発・運用）まで含めて整理し、責任分配と説明責任の“相手”を確定させる。 [5]
- **トレーサビリティ実装（データ出所・意思決定・入出力・変更履歴）**：ログ方針、データ来歴、外部API呼出、権限行使（誰が何を実行させたか）を“監査可能”にする。ガイドラインのチェックリストにも直結する。 [6]
- **インシデント対応の型（AI特有の攻撃・誤動作・悪用・出力事故）**：プロンプトインジェクションやDoS等の脅威と対策を含む、AIセキュリティ運用を組み込む（総務省の技術ガイドラインも参照される領域）。 [7]

（オンライン参照用の図：内閣府「AI法」概要画像URL例）

```
https://www8.cao.go.jp/cstp/ai/ai_act/main.png
```

[8]

入手先と文書体系

公式入手先

本版（第1.2版）の **一次配布元（公式）** は、経済産業省の「AI事業者ガイドライン」ページに集約されている。同ページから、本編、別添（溶け込み版）、チェックリスト、ワークシート（Excel）、および見え消し版（第1.1版からの差分）にアクセスできる。[9]

総務省側の掲載ページは、内閣府[10]が公表する「AI指針」ページ上の「各府省庁等のガイドライン等一覧」からも参照でき（総務省リンク、経産省リンクが並記）、政府横断の位置づけが明示されている。[11]

（公式URL・主要ファイル直リンク：実務用）

```
【第1.2版（最新版）・配布ページ（経済産業省）】
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20260331_report.html

【第1.2版 本編（PDF）】
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf

【第1.2版 本編（第1.1版からの見え消し版PDF）】
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_10.pdf

【第1.2版 別添（溶け込み版PDF）】
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_3.pdf

【第1.2版 チェックリスト（別添7A/7B）（PDF）】
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_5.pdf

【ワークシート（Excel）】
（配布ページから取得：同ページにExcelリンクあり）

【旧版（第1.0/1.01/1.1）まとめ（経済産業省）】
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20240419_report.html
```

[12]

本編と別添の役割分担

別添（付属資料）は、本編と対応しつつ、**「具体的にどのようなアプローチで取り組むか（実践）」**を示す“解説書/実務リファレンス”として位置づけられている（本編＝why/what、別添＝how）。[13]

また、本編は「リスクベースアプローチ」の重要性と、Living Documentとしての継続更新を明示しており、企業側は“固定的な準拠”ではなく、更新を前提にガバナンス設計を行う必要がある。[14]

改訂履歴と差分の原文特定

公表日ベースの改訂履歴

経済産業省の検討会ページおよび旧版まとめページにより、少なくとも以下の版管理が確認できる。 [15]

- 第1.0版：2024-04-19（旧版ページにPDF一式） [16]
- 第1.01版：2024-11-22（旧版ページにPDF一式、見え消し版あり） [17]
- 第1.1版：2025-03-28（旧版ページにPDF一式、見え消し版あり） [18]
- 第1.2版：2026-03-31（最新版ページにPDF一式、見え消し版あり） [19]

差分（見え消し版）を“原文レベル”で追う方法

第1.2版の差分は、**「第1.1版からの見え消し版」**が本編・別添ともに提供されており、条文・注釈・用語定義の追加/修正を原文で追跡できる。 [20]

また、第1.0→1.01、1.01→1.1 についても、旧版ページに見え消し版が用意されているため、初版からの“更新の連鎖”を追える。 [21]

第1.2版で象徴的な追加の原文抜粋

第1.2版の改訂で最も象徴的な点は、AIエージェント/フィジカルAIを**定義として明確化**したことにある（見え消し版に掲載）。 [22]

> 「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。」 [22]

> 「本ガイドラインにおけるフィジカル AI とは…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…現実世界に対して直接的な働きかけ…を行う…。」 [22]

さらに、透明性・説明可能性等の概念については、海外定義（NISTやEU等）との差を注釈で整理している（見え消し版の脚注）。 [23]

> 「NIST…透明性…説明可能性…解釈可能性…に分類…。」 [23]

更新内容の全体像（検討会資料）としては、用語定義（学習/推論/データ）やユーザビリティ改善（活用の手引き、チャットボット）を含めて更新が説明されている。 [24]

主要な新規項目・変更点と実務対応

以下は、第1.2版の改訂点として実務影響が大きいものを、**「目的」「想定リスク」「企業の具体的対策（技術/組織/手続）」「実装課題と対応案」**に分解して整理した。

主要変更点サマリ表

変更点（第1.2版での強調/追加の核）	目的	想定されるリスク	企業が取るべき具体的対策（技術・組織・手続）	実装上の課題と対応案
AIエージェントの定義明確化	自律的に行動するAIを前提に、責任分配・監督・安全設計の出発点を揃える	外部API/社内システム連携により、権限逸脱・誤作動が実害化（自動発注・予約・送金等）／説明不能な行為連鎖	**技術**：権限最小化（スコープ制限）、ツール実行のガードレール、重要操作のHITL、実行前シミュレーション、ツール呼出ログ。**組織**：オーナー（責任者）設定、変更管理、レビュー体制。**手続**：リスク評価→承認→段階的解放（feature flag）	“完全自律”を目指すほど監督コストが跳ね上がる。**対応案**：自律度を段階定義し、段階ごとに要件（レビュー頻度・ログ粒度・停止条件）を固定化する。 25
フィジカルAIの定義明確化	AIが物理世界に作用する場合の危害防止を、概念レベルで明示	人身・設備損傷、重要インフラ干渉、連動ロボット制御不能など安全上の重大事故	**技術**：安全停止（E-Stop）、フェイルセーフ、サンドボックス/デジタルツイン検証、センサー異常検知、冗長系。**組織**：安全責任と品質保証の統合、OT/IT横断CSIRT。**手続**：安全認証・現場教育・運用監査	現場では“例外”が多い。**対応案**：想定外環境を前提にしたテスト設計と、稼働時の「停止判断基準」を文書化する。 26
「学習」「推論」定義の追加（RAG等含む）	リスク源を「モデル」ではなく「推論時に参照する外部知識・データ」まで含めて共通理解化	RAGで機密/個人情報が混入し漏えい、外部知識の汚染、誤情報の参照、プロンプト注入経由の不正実行	**技術**：データ参照制御、PII/機密のフィルタリング、参照元のアクセス制御、出典提示。**組織**：データオーナー、利用目的/同意管理。**手続**：プロンプト/データ取扱ルール、DLP、定期棚卸	“推論用データ”の境界が曖昧。**対応案**：RAGインデックス・コネクタ・外部APIを「データ源」として台帳化する。 2
アカウンタビリティにおけるトレーサビリティ中核化	データ出所・意思決定等を追跡可能にし、説明と改善の前提を整える	事故時に原因究明不能、説明不能、是正不能→信用失墜・契約違反・賠償リスク	**技術**：データ来歴、意思決定ログ、モデル/プロンプト/設定のバージョン管理。**組織**：説明窓口、責任者明示。**手続**：文書化（設計/評価/運用）、定期説明	ログが多すぎて運用破綻。**対応案**：リスクベースでログ粒度を決め、重要操作・外部連携・高リスク判断を優先保存。 6

変更点（第1.2版での強調/追加の核）	目的	想定されるリスク	企業が取るべき具体的対策（技術・組織・手続）	実装上の課題と対応案
ステークホルダー（サプライヤー含む）への説明強化	供給網・利用者・委託先まで含む“説明の相手”を明示し、責任分配を可能にする	サプライチェーンのブラックボックス化、委託先事故の連鎖、責任の押し付け合い	**技術**：提供情報パッケージ（SBoM/Model Card相当、運用条件、既知リスク）。**組織**：ベンダマネジメント、共同インシデント対応。**手続**：契約の責任分界、監査権、通知義務	ベンダが情報を出さない。**対応案**：調達要件に「説明可能性/トレーサビリティ提供」を明記し、段階的に義務化。 4
説明可能性・解釈可能性を“説明の設計”として位置づけ	説明主体が「何を説明するべきか」を把握し、受け手と共有する	説明不足による不信、誤使用、クレーム/紛争	**技術**：根拠提示、検証可能性、モデル/出力の制約明示。**組織**：CS/法務/開発の連携。**手続**：説明要求の受付→回答→記録のフロー	何をどこまで説明するかがケース依存。**対応案**：説明レベルを“用途別テンプレ”化（金融与信/採用/医療等）。 3
生成物の識別（透かし・来歴・認証）等の推奨（高度AI含む）	AI生成物の誤認・偽情報・ディープフェイク等の社会リスク低減	偽情報拡散、詐欺・なりすまし、ブランド毀損、法令違反	**技術**：電子透かし、C2PA等の来歴、生成物ラベリング。**組織**：対外発信ルール、承認フロー。**手続**：利用規約・表示・社内教育	透かしは回避され得る。**対応案**：表示＋監視＋削除/通報＋証拠保全の多層防御にする。 27
リスクチェーン/価値連鎖全体でのリスク削減	便益最大化とリスク低減を“複数主体またがり”で考える	システム・オブ・システム（連結）による連鎖障害、責任不在	**技術**：相互接続点の可視化、依存関係図。**組織**：共同リスクレビュー。**手続**：責任分配、共同計画	多主体調整コストが現実的障壁。**対応案**：まず重要取引先（主要ベンダ/顧客）から“共通言語”で合意形成。 28
ユーザビリティ改善（活用の手引き/チャットボット等）	ガイドラインを“使える形”にし、導入企業の初動を支援	解釈のばらつき、形骸化	**技術/組織**：社内教育に組み込み、参照窓口を一本化。**手続**：手引きに沿って導入手順を標準化	手引き・FAQは更新される。**対応案**：社内文書を“Living”にし、改訂点レビューを定例化。 29

重点論点の条文・注釈引用にもとづく解説

本節では、要求された観点（AIエージェント、フィジカルAI、生成物責任、説明可能性、データガバナンス、セキュリティ、サプライチェーン、第三者評価・監査、国際連携）について、ガイドライン本文・別添・チェックリスト等の原文を引用しつつ、実装に落とす。

AIエージェント

AIエージェントは、定義それ自体が「対象の境界」を決める。ガイドラインは、AIエージェントを“環境を感知し自律的に行動するAIシステム”とし、さらに注釈で「エージェンティックAI（複数エージェントで意思決定・アクション）」にも言及している。 22

ここから実務に落とすべき要点は、**「自律的に行動＝権限行使が発生する」**という点である。従来のチャットボットは“助言”で止まるが、エージェントは“実行”する。よって、次の統制が必須となる。

- 実行権限の最小化（業務システム側のスコープ制御、認証・認可、鍵管理）
- 重要操作の人間介在（Human-in-the-Loop / Human-on-the-Loop）
- ツール呼出ログと、**誰が何を実行させたか**の追跡（トレーサビリティ） 6

フィジカルAI

フィジカルAIは、物理環境を取り込み、推論・判断し、アクチュエータ等を介して現実世界へ直接作用するものとして定義される。 22
この定義が意味するのは、サイバー領域の損害だけでなく、**物理的危害（生命・身体・設備・環境）**が主要リスクになるという点である。本編は共通指針として、危害を及ぼさない安全性確保を強調する。 30

実装上は、AI品質保証・安全設計・現場運用が分離すると破綻しやすい。別添では「ロボット制御不能」等を例示し、リスク最小化（ガードレール技術等）を要求する記述が見え消し版に現れている。 31

生成物責任（生成物の社会的責任・法的責任の整理）

ガイドライン本文は、リスクが制裁金や損害賠償責任につながり得ることを明示し、リスクを“経営層で共有・更新”する枠組みを示す。 32
またチェックリスト（広島AIプロセス部分）では、技術的に可能なら「電子透かし」「コンテンツ認証」「来歴メカニズム」を開発・導入することを求めている。 33

> 「電子透かし…信頼できるコンテンツ認証及び来歴のメカニズム…導入しているか？」 33

実務での“生成物責任”は、製造物責任（PL）的な枠組みだけでなく、①誤情報・偽情報による被害、②知財侵害、③なりすまし・詐欺、④差別的出力、⑤個人情報漏えいのように多層で発生する。別添はリスク例として「偽情報」「著作権等の権利」等を挙げる。 32

実務対応は、「生成物の識別」「発信の承認」「証拠保全」の三点セットが最小構成となる。

- **識別**：透かし・メタデータ・来歴（可能な範囲）
- **承認**：対外発信（広告/IR/採用/医療説明等）は人間承認フロー
- **証拠**：プロンプト・参照データ・モデル/設定・出力の保存（後述トレーサビリティ）

説明可能性・解釈可能性

本編のアカウンタビリティ/透明性周辺では、説明可能性を単なる技術論ではなく、**説明主体が必要な説明を把握し、説明を受ける主体と共有する**こととして記述している。 34

> 「説明する主体がどのような説明が求められるかを分析・把握…説明を受ける主体がどのような説明が必要かを共有…」 34

また注釈で、透明性・説明可能性・解釈可能性の区分をNIST等にならって整理している。 [35]

この書きぶりから、実務上の“説明可能性”は最低でも次を含むべきである。

- **利用上の前提/限界**（適切・不適切な使用方法、既知リスク） [36]
- **判断の根拠の説明**（どのデータ・ルール・モデルが関与したか、検証可能性）
- **問い合わせ窓口**（合理的範囲での設置） [37]

データガバナンス（データ来歴・品質・アクセス制御）

本編は、トレーサビリティの対象として**「データの出所」「意思決定」**を明示し、追跡・遡及可能な状態の確保を求める。 [34]

> 「データの出所…意思決定等について…追跡・遡求が可能な状態を確保する」 [34]

別添では、学習時のデータについて、第三者の個人情報や知財に留意し、ライフサイクル全体で適切に扱うこと、さらに**データ管理・制限機能**によるアクセス管理を検討することが述べられる。 [38]

この2点をセットで実装するのが実務上の要諦である。

- **来歴（lineage）**：学習データ・推論参照データ（RAG/外部API）・生成物の出所を追える
- **統制（control）**：誰がどのデータ源をAIに接続したか、アクセス権限は妥当か

セキュリティ（AI特有の脅威、エージェント化による拡大）

本編の共通指針チェックリストでも、「不正操作による意図せぬ変更又は停止が生じない」状態を求めている。 [33]

さらに、総務省の「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月公表）は、プロンプトインジェクションやDoSを明示しつつ、AIエージェントやMCP等で自律性・連携が増す中で新たな脅威が生じ得ることを指摘する。 [7]

> 「AI エージェントや MCP…自律性を増す中で新たに生じる…脅威…」 [39]

したがって、第1.2版対応のセキュリティ実装は「LLM単体」ではなく、**AIシステム（外部連携込み）の脅威モデル**を持つ必要がある。具体策としては、(a)プロンプト/ツール入力の検証、(b)ツール実行前のポリシーチェック、(c)秘密情報の分離、(d)異常検知と停止、(e)インシデント対応を、開発・提供工程に組み込む。 [40]

サプライチェーン（サプライヤーを含む説明責任、責任分配）

本編のアカウンタビリティでは、「共通の指針」対応状況の説明先として**「ステークホルダー（サプライヤーを含む）」**が明示される。 [34]

これは、委託開発、SaaS利用、外部モデル/API利用のいずれでも、調達・契約の時点で「必要情報」と「責任分界」を決めておかないと、後から説明責任が果たせないことを意味する。別添は、AI取引が多様化している点や、従来のベンダ/ユーザー二分論の限界を指摘している。 [41]

第三者評価・監査（独立性の確保、外部規格の活用）

別添では、AIガバナンス・ゴールとの乖離評価において、**当該AIの開発・提供・利用に直接関わっていない者が加わる**ことが期待されるとしており、内部統制的な“独立レビュー”の思想が読み取れる。 [42]

第三者評価・監査を「外部の認証/監査」として組み立てる場合、国際規格に寄せるのが実務上の近道になる。ISO [43] は、AIマネジメントシステム規格としてISO/IEC 42001を提示しており、組織としてAIの責任ある開発・提供・利用の管理体系を構築する枠組みとなる。 [44]
また、日本規格協会 [45] も、ISO/IEC 42001の邦訳版発行等を通じて国内導入を後押ししている。 [46]

第三者評価の実務設計としては、最低限以下を推奨する。

- **内部独立レビュー（必須級）**：開発部門から独立したレビュー（法務・品質・セキュリティ・監査）
- **外部アセスメント（高リスク領域）**：重要業務（金融・医療・物理制御等）では外部評価の実施
- **証跡一式（監査可能性）**：トレーサビリティ、文書化、リスク評価、インシデント履歴を整備 [6]

国際連携（広島AIプロセスとOECD報告枠組み）

チェックリスト（広島AIプロセス部分）では、導入後の脆弱性/インシデントの特定・緩和、組織間の責任ある情報共有と報告などが問われており、国際的期待に沿った運用が求められる。 [47]
またOECD [48] は、広島AIプロセスの成果として**報告枠組み（Reporting　Framework）**を立ち上げたと説明しており、先進AI開発における透明性・説明責任の“国際的実務”が進行している。 [49]

国際比較（EU AI Act、米国ガイダンス、英国、韓国など）

比較の前提

日本のガイドラインは、リスクベースとLiving　Document設計を強調し、各事業者の自主的取組を前提にする。 [14]
これに対し、EUや韓国は法制度として義務・罰則・当局体制を伴う“ハードロー”を整備しており、同じ「リスクベース」でも実務負担の質が異なる。 [50]

条項対応・厳格度・実務影響の対比表

観点	日本：AI事業者GL 1.2	EU AI Act	米国：NIST AI RMF	英国：AI regulation white paper	韓国：AI Basic Act
法的性格	ガイドライン（自主的取組を後押し、リスクベース/更新前提） [14]	規則（Regulation）として域内市場に統一枠組み、罰則・当局体制あり [51]	任意のリスク管理フレームワーク。信頼性とリスク管理のための手引き [52]	技術横断の硬直規制を避け、利用文脈重視の“pro-innovation”枠組み（原則を既存規制当局が適用） [53]	基本法として産業育成と信頼基盤、透明性・安全義務、影響評価等を規定（施行時期も明示） [54]

観点	日本：AI事業者GL 1.2	EU AI Act	米国：NIST AI RMF	英国：AI regulation white paper	韓国：AI Basic Act
リスク分類	自社の事業/用途に応じたリスクベース（個別最適） 55	高リスク等の枠組みで義務内容を規定し、適合性評価・監督 56	“Map-Measure-Manage-Govern”など機能で整理、組織に適用 52	一律の分類を避け、文脈に応じて原則適用（安全性・透明性等） 57	高影響/生成AIを規制対象として定義し、透明性・安全・運用者責務等 54
人間の監督	「人間中心」を土台に、人間判断の介在を求める 58	高リスクAIに人の監督要件（有効に監督できる設計） 59	監督を含むガバナンスを組織能力として整備 60	原則（安全・説明可能性等）として監督を求める方向 57	高リスクAI等に透明性/責務を規定（当局支援も規定） 61
トレーサビリティ/文書化	アカウンタビリティの中核として追跡可能性を要求 6	技術文書・QMS、ポストマーケット監視計画等を要求 62	文書化は推奨だが義務ではなく、実装は組織裁量 60	原則として透明性等を求めるが詳細実装は当局・領域次第 57	透明性義務・影響評価等（詳細は下位法で具体化） 54
生成物の識別	高度AIで透かし等の導入をチェック項目化 33	透明性義務があるが、他法令との関係も示唆 63	フレームワーク内での推奨（要件化ではない） 60	IP/表示等は議論中、原則と業界対応が中心 57	生成AIの透明性/表示等を含む義務の方向が示される 54
供給網/第三者	サプライヤーを含む説明責任・責任分配を明示 4	適合性評価（通知機関）など第三者関与が制度的に組み込まれる 64	組織が自ら統制を設計（第三者は選択） 60	既存規制当局・業界ガイダンスで補完（統一の第三者制度は限定的） 57	民間の検証・認証への政府支援を規定し、制度化を志向 61

補足：EU AI Act は、高リスクAIに対し「リスク管理システム」「品質管理システム」「ポストマーケット監視計画」等の体系的要件を要求している。 65

一方、米国の代表的ガイダンスとしてのNIST 66 AI RMF は、組織がAIリスクを理解・管理するためのフレームワークであり、更新を予定するLiving Documentの性格も示されている（罰則による強制ではなく“実装の共通言語”）。 52

韓国は、Ministry of Science and ICT 67 の公表によれば、AI基本法により「高リスクAI・生成AI」を対象に透明性・安全・運用者責務等を規定し、2026年1月施行とする。 54

実務チェックリスト・テンプレート・業種別影響・利害関係者提言

企業向け実務チェックリスト（短期/中期/長期）

ガイドライン自体が「チェックリスト（全主体向け）」を提供し、人間中心・安全性・公平性・プライバシー・セキュリティ・透明性・アカウンタビリティ等の確認を促している。 [68]
ここでは、企業導入の現実に合わせ、優先順位を付けた実務チェックリストに組み替える。

優先度	期間	目的	やること（成果物）	ガイドライン根拠（例）
最優先	0〜30日	会社としての“AI統制の入口”を作る	①AI台帳（システム/用途/データ/外部連携/責任者）②役割マッピング（開発者/提供者/利用者）③最低限の利用ルール（機密・個人情報・対外発信）	対象範囲・連携・自主的取組の重要性 [69]
高	30〜90日	本番運用に耐える証跡を整備	①トレーサビリティ設計（ログ方針、来歴、変更管理）②説明窓口/回答フロー③ベンダ要件（情報提供/責任分界/通知義務）	トレーサビリティ/説明の要求 [70]
高	30〜90日	エージェント/外部連携のリスクを下げる	①権限最小化②重要操作のHITL③ツール呼出ログ④脅威モデル（prompt injection/DoS等）	AIエージェント定義、セキュリティ脅威 [2]
中	90〜180日	高リスク領域での品質保証・独立レビュー	①独立レビュー（開発から独立）②高リスク用途の外部評価検討③運用監査	乖離評価の独立性 [42]
中	90〜180日	生成物の社会リスク低減	①生成物ラベリング/透かし②対外発信承認③削除/通報/訂正プロセス	透かし・来歴等 [33]
継続	180日〜	Living Document対応（継続改善）	①改訂点レビュー会（四半期）②教育・リスキリング③インシデント演習	Living Document/教育 [69]

テンプレート（表）

リスク評価テンプレート（AIシステム/サービス単位）

項目	記入内容（例）
システム概要	何をするAIか（機能/利用者/業務影響）
立場	開発者/提供者/利用者のどれか（複合可） [55]
AI種別	生成AI、AIエージェント、フィジカルAI等（該当定義参照） [22]
推論時データ源	入力（プロンプト/センサー）、RAG参照元、外部API、DB等 [71]
主要リスク	安全/公平/プライバシー/セキュリティ/透明性/説明責任/競争等（共通の指針に沿って列挙） [68]
重大度×蓋然性	影響（人身/金銭/信用/法務）と発生可能性

項目	記入内容（例）
コントロール	技術（権限/ログ/フィルタ等）、組織（責任者/レビュー）、手続（承認/監査）
トレーサビリティ設計	何をログ化し、どこまで遡及可能か [34]
説明設計	誰に、何を、どの深さで説明するか [3]
供給網	サプライヤー/委託先/外部モデルの情報提供範囲 [72]
承認結果	Go/条件付きGo/No-Go＋理由（証跡保存）

説明責任フロー（Mermaid）

```
flowchart TD
  A[AIユースケース登録] --> B[役割判定: 開発者/提供者/利用者]
  B --> C[リスク評価: 共通の指針で棚卸]
  C --> D{高リスク?}
  D -- Yes --> E[独立レビュー/外部評価検討]
  D -- No --> F[標準レビュー]
  E --> G[説明設計: 誰に何を説明するか合意]
  F --> G
  G --> H[トレーサビリティ設計: データ出所/意思決定/変更履歴]
  H --> I[運用開始: モニタリング/インシデント窓口]
  I --> J[定期説明・改善: Living Document対応]
```

インシデント対応手順テンプレート（AI特化）

フェーズ	目的	手順（最小セット）	証跡
検知/通報	早期発見	異常出力/不正実行/漏えい疑い受付（窓口） [37]	通報チケット、時刻、影響範囲
トリアージ	重大度判定	人身/金銭/規制/信用で分類、関連ログ保全	関連ログ（入出力/外部API/権限） [73]
封じ込め	被害拡大防止	機能停止、権限剥奪、外部連携遮断、モデル/ルールロールバック	変更履歴、停止判断記録
根本原因分析	再発防止	プロンプト注入/DoS/データ汚染等の原因分析 [39]	RCAレポート
復旧/周知	業務復旧と説明	対応状況の説明、必要に応じて供給網と連携 [6]	顧客/取引先向け説明文
振り返り	改善	コントロール更新、教育、改訂点レビュー	改善計画、再発防止策

業種別の影響分析と追加負担（概算レンジ）

試算の根拠と仮定

- 人件費の基準：厚生労働省「令和7（2025）年賃金構造基本統計調査」によれば、一般労働者の賃金（月額）は**340,600円**（男女計）として公表されている（報道発表PDF）。 [74]
- 試算仮定：
- **実務負担＝（追加FTE月）×（月額単価）×（間接費係数）**
- 間接費係数は企業差が大きいため、本稿では**1.3〜2.0**をレンジで置く（社会保険、管理、外部委託の混在を吸収するため）。
- 外部評価や認証は別途費用が発生し得るが、ここでは“追加人件費中心”で概算する（第三者認証等はISO/IEC 42001が参考枠組み）。 [75]

業種別インパクト表

業種	影響が大きいAI形態	追加で重くなる論点	追加負担（目安）	概算コストレンジ（円）
製造（ロボット/検査/最適化）	フィジカルAI、外部連携AI	安全性（人身・設備）、OT/ITセキュリティ、停止基準、変更管理	3〜12 FTE月/年（安全・品質・情シス横断） [2]	約130万〜820万（340,600×3×1.3 〜 ×12×2.0）
医療（診断支援/事務）	高リスク用途の生成AI/推論	説明責任、プライバシー、誤作動の影響、監査可能性	4〜18 FTE月/年（法務・品質・運用含む） [76]	約180万〜1,226万
金融（与信/不正検知/客服）	エージェント化、意思決定AI	公平性、人間判断介在、説明可能性、記録保持	6〜24 FTE月/年（モデルリスク管理＋運用） [77]	約265万〜1,635万
物流（配車/倉庫/配送）	最適化＋一部フィジカル	事故時追跡、外部API連携、権限管理、セキュリティ	3〜10 FTE月/年 [73]	約130万〜681万
自治体（窓口/文書/調達）	生成AI利用（RAG等）	調達/利活用ルール、機密、説明窓口、教育	2〜8 FTE月/年（ガバナンス整備） [78]	約89万〜545万

注意：上記は“最低限の整備”のオーダー感であり、AI利用範囲（部門数・外部連携数・高リスク業務比率）で大きく増減する。特にエージェント化（自動実行）やフィジカルAI（物理行為）は、事故対応・安全設計のコスト弾性が高い。 [2]

主要な利害関係者への提言

規制当局（総務省・経産省・関係省庁）には、ガイドラインがLiving　Documentであることを踏まえ、**改訂点の“実装可能な粒度”の提示**が重要である。特にAIエージェント/外部連携の増加に伴い、セキュリティについては総務省技術ガイドラインのように脅威と対策例を継続アップデートし、横展開すべきである。 [79]

業界団体には、企業間で責任分配・説明責任の共通言語が必要である（サプライヤーを含む説明が明示されているため）。そのため、契約条項例・提供情報パッケージ（最低限の文書）・共同インシデント対応手順の標準化を、業界横断で整備することが望ましい。 [72]

第三者評価機関には、内部独立レビューだけでは不足する領域（医療、金融、物理制御等）で、**評価観点と証跡要件の標準化**が求められる。ISO/IEC 42001等のマネジメントシステム規格と整合する評価メニューを用意すれば、企業側の二重投資を減らせる。 80

ユーザー団体・消費者側には、生成物の識別（透かし/来歴）や問い合わせ窓口の整備が、実害低減に直結する。ガイドラインのチェックリストや説明窓口の要求は、社会的実装の“最低条件”として啓発と監視の対象にし得る。 81

---

1 3 4 5 6 10 14 34 36 37 43 45 48 55 58 69 70 72 73 76 77 78 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf

2 22 23 25 26 31 35 71 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_10.pdf
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_10.pdf

7 39 40 79 https://public-comment.e-gov.go.jp/pcm/download?seqNo=0000310937
https://public-comment.e-gov.go.jp/pcm/download?seqNo=0000310937

8 https://www8.cao.go.jp/cstp/ai/ai_act/main.png
https://www8.cao.go.jp/cstp/ai/ai_act/main.png

9 12 19 20 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20260331_report.html
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20260331_report.html

11 https://www8.cao.go.jp/cstp/ai/ai_guideline/ai_guideline.html
https://www8.cao.go.jp/cstp/ai/ai_guideline/ai_guideline.html

13 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_4.pdf
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_4.pdf

15 18 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/index.html
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/index.html

16 17 21 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20240419_report.html
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20240419_report.html

24 29 https://www.ipa.go.jp/disc/committee/begoj9000000egny-att/20260305_009_03_00.pdf
https://www.ipa.go.jp/disc/committee/begoj9000000egny-att/20260305_009_03_00.pdf

27 30 33 47 68 81 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_5.pdf
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_5.pdf

28 32 38 41 42 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_3.pdf
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_3.pdf

44 75 80 https://www.iso.org/standard/42001
https://www.iso.org/standard/42001

46 66 https://webdesk.jsa.or.jp/common/W10K0620?id=1207
https://webdesk.jsa.or.jp/common/W10K0620?id=1207

49 https://www.oecd.org/en/events/2025/02/launch-of-the-hiroshima-ai-process-reporting-framework.html
https://www.oecd.org/en/events/2025/02/launch-of-the-hiroshima-ai-process-reporting-framework.html

50 51 56 59 62 63 64 65 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ%3AL_202401689
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ%3AL_202401689

52 60 67 https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

53 57 https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

54 61 https://www.msit.go.kr/eng/bbs/view.do?bbsSeqNo=42&mId=4&mPid=2&nttSeqNo=1071&sCode=eng
https://www.msit.go.kr/eng/bbs/view.do?bbsSeqNo=42&mId=4&mPid=2&nttSeqNo=1071&sCode=eng

74 https://www.mhlw.go.jp/toukei/itiran/roudou/chingin/kouzou/z2025/dl/13.pdf
https://www.mhlw.go.jp/toukei/itiran/roudou/chingin/kouzou/z2025/dl/13.pdf