ChatGPT

日本政府AI事業者ガイドライン改定案の分析

エグゼクティブサマリ

本改定案は、従来の「生成AIを含む一般的なAI利活用」中心のガイドライン運用を、**自律的に意思決定し外部へ“行動”するAI（AIエージェント）**と、**現実世界へ作用するAI（フィジカルAI／ロボAI）**に適合させる方向で、ガバナンス要件を一段具体化しようとするものと位置づけられる。会議資料の公開範囲（後述）が限定されるため逐条レベルの確定はできないが、報道・二次情報が一致して示す中心軸は、①新概念（AIエージェント／フィジカルAI等）のガイドライン上の位置づけ明確化、②リスクベースで「特に留意すべきユースケース」を前に出す構成、③**「人間の判断を必須とする仕組み（Human-in-the-Loop）」の明記**と、それに付随する最小権限・ログ等の統制設計の要求、である。 [1]

法的には、ガイドライン自体は（少なくとも従来版と同様に）**直接の法的拘束力や直罰を伴う規制ではない**と整理される一方、①契約・調達・監査（対顧客／対当局／対投資家）での「説明責任の共通言語」化、②事故・漏えい・差別等が生じた際の「相当な注意義務（標準的安全配慮）」評価、③2025年に全面施行された日本のAI法が想定する「国際規範に即した指針整備」との連動、により、**実務上の準規範（ソフトロー）としての実効性は高まる**可能性が高い。 [2]

「人間の判断必須」は、単なる“目視確認”のスローガンではなく、（少なくとも二次情報上は）**外部アクション／重要変更の前に承認を挟むワークフロー、最小権限（Least Privilege）、監査ログ、リスクレベル別制御**といった、システム設計・運用設計へ落とし込める形で語られている。 [3]
企業負担は一律ではなく、①「エージェントが触る資産（資金・個人情報・制御系）」、②「自律度（単発提案→連続行動→マルチエージェント）」、③「物理安全の有無」、でコストが急増する。特にフィジカルAI（自動運転・自律ロボ等）は、ソフトウェア統制に加え安全工学（フェイルセーフ／緊急停止／安全境界）の実装・検証が必要になり、相対的に高コスト化しやすい。 [4]

海外比較では、EUは法規制（EU AI Act）として高リスク用途に**人間監督・リスク管理・文書化**等を義務付け、違反に行政罰（罰金）を伴うのに対し、日本は（少なくとも本件ガイドラインは）**非拘束の指針**として、当面は支援策・普及策を軸に「統制設計の標準化」を狙う構図が近い。米国はNIST AI RMFのような**任意フレームワーク**が中心で、英国は「5原則」を既存規制当局が適用する**非法定（当初）**の横断原則アプローチを採る。 [5]

政策提言としては、短期は①AIエージェント／フィジカルAIの定義と境界、②「人間判断必須」の適用対象（何が外部アクション／重要変更か）と例示、③最小権限・ログの実装要件（最低限）の明文化が最優先。中期は、監査・アシュアランスのエコシステム（評価手順、チェックリスト、参照実装、育成）整備が鍵となる。長期は、AI法に基づく指針群との一体運用（高リスク領域の実効性確保）と国際整合（EU/NIST/英国等とのクロスウォーク）を進め、国内企業の越境コンプライアンスを“先回りで楽にする”設計が望ましい。 [6]

調査範囲と不確実性

本レポートは、ユーザー指定の「AIネットワーク社会推進会議／AIガバナンス検討会　第29回」関連として、公開情報上で確認できる会議アジェンダ（配布資料の項目）と、報道・解説記事の要旨を突合して分析した。第29回の配布資料項目として、（1）今後の活動スケジュール、（2）国際・国内動向報告、（3）事業者アンケート結果概要、（4）活用事例、（5）令和7年度更新内容、が列挙されていることは確認できる。 [7]

一方で、**一次資料（総務省サイト上の会議資料PDF／本文案）について、閲覧環境上の取得制限があり**、逐条案・条文案（見え消し版を含む）を直接参照しての引用・逐条対比はできなかった。このため、本稿で「改定案の変更点」として扱う内容は、（a）会議資料項目として存在が示されるもの、（b）複数の報道・解説が一致して述べる要旨、（c）企業実装例として“ガイドラインの方向性を踏まえた”と明示している公開情報、を根拠にしている。 [8]

また、指定媒体である日本経済新聞 [9] およびLedge.ai [10] の記事本文は、閲覧制限により全文確認ができず、検索結果として取得できた要旨範囲に依拠している（内容の一部欠落可能性あり）。本稿では、その不確実性を明示し、断定が困難な箇所は「未記載／不特定／要確認」として扱う。 [11]

改定案の具体的な変更点と法的・運用上の意味

公開情報から抽出できる主な変更点

逐条（条文案）ではなく、「公開要旨ベースの変更点」である点に注意されたい。

- **AIエージェントの定義・便益・リスク・対策の追記**（検討）
- 併せて「エージェンティックAI（強い自律度、マルチエージェント）」のリスク追記が検討されている。 [12]
- **フィジカルAI（自動運転、自律型ロボット等）の定義・便益・リスク・対策の追記**（検討） [12]
- **「人間の判断を必須とする仕組み」**の明記（方針）
- 二次情報上は、特に「外部アクションや重要変更」の前に人間承認を挟む設計が想定される。 [13]
- **最小権限設定（Least Privilege）**の明記（方針） [14]
- **監査ログ／記録（説明可能な運用）**の重視（方針）
- 企業実装例としては「判断過程・参照データ・出力結果の保存」「完全監査ログ」等が挙げられている。 [15]
- **データ管理の徹底**（方針）と、**ハードウェア残存データへの配慮**（要旨上の記載） [14]
- **リスク評価手法の記載追加**と、**特に留意すべきユースケースの追加**（方向性） [12]
- **リスクベースアプローチ導入**、および **RAGと機械学習の概念整理**を含む「複数論点での包括見直し」 [16]
- **第1.2版を2026年3月末に公開予定**（スケジュール） [16]

法的意味

内閣府 [17] のAI政策文脈では、2025年に全面施行されたAI法が、国の司令塔（AI戦略本部 [18] ）と基本計画、そして国際規範に即した「指針」整備を制度的に位置付けている。 [19]
この構造上、AI事業者ガイドラインは「法律そのものの罰則」ではなくても、国が整備する指針群の一部として、官民の説明責任・適正実施の期待値を引き上げる作用を持つ（＝実務上のソフトローとして作用し得る）。 [20]

また、既存の整理として、AI事業者ガイドラインには法的拘束力がなく、準拠していなくても直ちに制裁が科されるわけではない、という見解が示されている。 [21]
ただし、改定案が「人間判断必須」や最小権限といった**設計要件に近い表現**へ踏み込むほど、事故・漏えい時の事後評価（注意義務、説明義務、製造物責任・契約責任の争点等）で参照される可能性が上がる点は、運用上の重要な意味となる（本項は法的評価の一般論としての推論）。 [22]

運用上の意味

運用面では、従来の「AIを使う際の望ましい行動」から、**“自律的に動くAIを、どう統制設計として組み込むか”**へ主眼が移る。企業のAI活用が「導入」から「統制設計」へ移行している、という整理も示される。 [23]

この変化は、単なるコンプライアンス対応ではなく、顧客・利用者・取引先に対する信頼構築（監査可能性、制御可能性の提示）を競争力要因に変換しようとする政策意図とも整合する。 24

新規概念の定義と範囲

AIエージェント

公開要旨では「AIエージェント」の定義追加が検討されていることは確認できるが、**改定案の定義文そのものは（本稿の確認範囲では）明記を検証できないため不特定**である。 25
したがって以下は、**一般的に通用している技術的・運用的な作業定義（ワーキング定義）**であり、ガイドライン上の確定定義ではない。

- ワーキング定義（非公式）：
  「目標（タスク）達成のために、環境・情報を観測し、計画を立て、ツール（API／システム）を操作して、複数ステップの行動を自律的に実行するAIシステム」
  ※“生成するだけ”ではなく“実行する”点がガバナンス上の差分となる。 26

さらに、エージェンティックAI（強い自律度、マルチエージェント）のリスク追記が検討されているため、単一エージェントに留まらず、複数エージェントが役割分担して意思決定・実行する構成まで射程に入る可能性が示唆される。 12

フィジカルAI（ロボAI）

公開要旨では、フィジカルAIを自動運転や自律型ロボット等として例示しつつ、定義・便益・リスク・対策の追記が検討されている。 27
ただしこの場合も、**定義文言自体は確認できないため不特定**である。

- ワーキング定義（非公式）：
  「センサー入力（視覚・距離・位置など）に基づき、物理デバイス（車両、産業ロボ、家庭用ロボ、ドローン等）の挙動へ影響を与える意思決定・制御を行うAI（ソフトウェアとハードウェアを含むシステム）」
  ※結果が物理安全へ直結するため、ヒューマン・オーバーライド（停止・介入）設計が重要になる。 28

「定義が曖昧なまま導入が先行する」ことの実務リスク

定義が確定しないまま「AIエージェント」「フィジカルAI」をプロダクト／社内規程で用いると、対象範囲（どこまでが規制対応・監査対象か）が組織ごとにズレ、監査・調達・契約で摩擦が生じる。したがって、改定版の公開時には、①境界事例（例：RPA＋LLMは？、倉庫内AGVは？）の例示、②自律度レベルの層別、③「外部アクション／重要変更」の定義、が不可欠になる。 29

「人間の判断必須の仕組み」の実装例、課題、企業負担・コスト推定

仕組みの基本設計

公開情報が示す方向性は、「AIが外部アクションや重要変更を実行する前に、人間の承認プロセスを必須化」「最小権限」「監査ログ」「リスクレベル別制御」である。 30
これを実装に落とすと、最小構成でも次の4要素が必要になる。

- **アクション分類**：
  「外部アクション」（例：送金、発注、メール一斉送信、ユーザー権限変更、ロボ稼働開始）と、「重要変更」（例：モデル更新、プロンプト／ポリシー改変、権限昇格、学習データ差し替え）を定義する。 31
- **リスクレベル判定**：
  影響度×発生可能性×可逆性（ロールバック可能性）で、承認必須か自動実行かを決める（リスクベース）。 27
- **承認ワークフロー**：
  「誰が」「何を根拠に」「どの条件で」承認できるか（職務分掌、責任分解）を定義し、システム上で強制する。 26
- **監査ログ（改ざん耐性のある記録）**：
  AIの提案、参照データ、実行アクション、承認者、時刻、結果を記録し、後から追跡できるようにする。 32

Mermaid：Human-in-the-Loopの典型フロー（エージェント型）

```
flowchart TD
  A[Agent receives goal] --> B[Plan generation]
  B --> C{Action impacts external world or critical assets?}
  C -- No --> D[Execute with least-privilege token]
  D --> E[Write audit log]
  C -- Yes --> F[Risk scoring & policy check]
  F --> G{Approval required by policy?}
  G -- No --> D
  G -- Yes --> H[Create approval request packet]
  H --> I[Human reviewer UI / 4-eyes if needed]
  I --> J{Approve?}
  J -- Deny --> K[Block action + log denial]
  J -- Approve --> L[Issue time-bounded token]
  L --> D
  E --> M[Continuous monitoring]
  K --> M
```

Mermaid：統制データモデル（ER図の例）

```
erDiagram
  AGENT ||--o{ ACTION_REQUEST : creates
  ACTION_REQUEST ||--o{ APPROVAL : requires
  USER ||--o{ APPROVAL : performs
  ACTION_REQUEST ||--o{ AUDIT_LOG : produces
  POLICY ||--o{ ACTION_REQUEST : evaluates
```

```
PERMISSION_SET ||--|| AGENT : binds
TOOL ||--o{ ACTION_REQUEST : targets

AGENT {
  string agent_id
  string owner_org
  string purpose
}
ACTION_REQUEST {
  string request_id
  string action_type
  string risk_level
  datetime created_at
}
APPROVAL {
  string approval_id
  string decision
  datetime decided_at
}
AUDIT_LOG {
  string log_id
  string request_id
  string artifacts_hash
  datetime timestamp
}
POLICY {
  string policy_id
  string rule_set_version
}
PERMISSION_SET {
  string permission_set_id
  string scope
  string expiry
}
TOOL {
  string tool_id
  string system_name
}
USER {
  string user_id
  string role
}
```

技術的実装例（代表例）

以下は「ガイドラインの方向性を踏まえた」と明示する公開情報に現れる実装要素であり、政府の必須実装を網羅するものではない。

- **人間承認必須フロー**：外部アクション・重要変更の前に承認を挟む。 15
- **最小権限アーキテクチャ**：業務単位でアクセス範囲を制御し、AIに付与する権限を最小化する。 33
- **監査ログの自動保存**：判断過程・参照データ・出力結果の保存を想定。 15

- **リスクレベル別制御**：業務内容・データ機密度で承認や制御強度を動的に変える。 33

運用上の課題

- **承認ボトルネックと“承認疲れ”**：
  エージェントが多数のアクション提案を行うと、人間が形式的に承認するだけになり、実質的統制が失われる。対策として、リスクスコアリング、バッチ承認、二段階（自動＋抜取監査）、4-eyes（ダブルチェック）などの運用設計が必要になる（推論）。 34
- **「重要変更」の定義が難しい**：
  例：プロンプト変更、ツール追加、モデル更新、学習データ差替え、権限変更。どこからが“重要”かは業務・業種で異なるため、ユースケース例示が不可欠。 35
- **ログの機密性・プライバシー**：
  ログに個人情報・機密が含まれる可能性がある。ログ保存は「説明可能性」を上げる一方で新たな漏えい面（攻撃面）を生むため、アクセス制御・暗号化・保存期間・マスキングが必要になる（推論）。 36
- **フィジカルAI特有の安全課題**：
  物理挙動は可逆性が低く、誤作動が直ちに危害になり得る。ヒューマン・オーバーライド（停止・退避）設計は、OECD原則が述べる「望ましくない挙動時に上書き／停止できる仕組み」と整合するが、実装・検証は重い。 37

企業への負担・コスト推定（高・中・低）

ここでは「追加開発・運用設計・監査対応」の相対評価として整理する（定量は企業規模・既存基盤に依存するため提示しない）。

対象	コスト推定	理由（要点）
社内業務限定のAIエージェント（参照のみ／外部アクションなし）	低〜中	承認対象が少なく、既存のアクセス制御・ワークフローを流用しやすい（推論）。 38
外部アクションを行うAIエージェント（発注・送金・顧客連絡等）	中	承認フロー、職務分掌、監査ログ、最小権限の統合が必須になりやすい。 23
規制産業（医療・金融等）で意思決定に影響するAIエージェント	中〜高	既存規制との整合、説明責任、ログ保全、第三者監査が重くなる（推論）。 39
フィジカルAI（自動運転・自律ロボ等）	高	物理安全要求（フェイルセーフ、停止、検証）、事故時の責任分界が重く、欧州等の高リスク要件（人間監督等）とも近い統制が必要になりやすい。 40

影響業種・ユースケース別の具体例とリスク評価

影響が大きい業種（例示）

改定案は、リスク評価手法と「特に留意すべきユースケース」を前に出す方向性が示されるため、**“自律的に動く／現実世界へ作用する”**ほど影響が大きい。 41

業種	代表ユースケース（AIエージェント／フィジカルAI）	主なリスクカテゴリ	概括リスク
医療	検査・診療支援、問診エージェント、院内搬送ロボ	安全（誤判断）、説明責任、個人情報、サイバー	高（特に安全・個人情報）
交通	自動運転、運行管理エージェント、倉庫・港湾の自律搬送	物理安全、サイバー、責任分界	高（物理安全）
金融	融資審査支援、取引監視、資産運用エージェント	公平性、説明責任、詐欺・不正、モデルリスク	中〜高
製造	工場の自律制御、予防保全エージェント、協働ロボ	労働安全、停止設計、サイバー・OT	高（物理安全＋OT）
家庭用ロボ／IoT	見守りロボ、家事支援ロボ、スマートホーム自動化	プライバシー（カメラ等）、誤作動、安全	中〜高

（注）上表は、本改定案がAIエージェント／フィジカルAIを追記対象とするという公開要旨を前提にしたリスク整理であり、個別ユースケースの適用要件は改定版公開後に要確認。 42

リスクの焦点：誤作動・プライバシー侵害・攻撃面の増大

報道・要旨レベルでは、誤作動やプライバシー侵害リスクを念頭に「人間判断必須」を求める、という筋が繰り返し現れる。 43
また、カメラ等との連携によるプライバシー侵害可能性、攻撃対象や攻撃手法の増加、複雑化による保守困難化、などが論点として言及されている（要旨上）。 44

このため、リスク評価の実務では「AIモデル品質」だけでなく、次の観点が重みを増す。

- **“実行権限”を持つことによる被害上限（Blast Radius）**：最小権限とセグメンテーション。 23
- **“観測（センサー・個人情報）”の高度化**：カメラ・音声・位置情報の取り扱いとログの副次リスク。 45
- **“停止できること”の設計**：望ましくない挙動時のオーバーライド／フェイルセーフ。 46

海外ガイドライン・法規制との相違点と整合性

国際比較の軸

比較の観点を「拘束力」「リスク分類の仕方」「人間監督の設計要求」「監督・罰則」「国際展開への影響」に割り付ける。

枠組み	拘束力	中心アプローチ	人間監督（Human oversight）	監督・罰則	国際影響
日本：AI事業者ガイドライン（改定案）	任意（ソフトロー）	役割別（開発者・提供者・利用者）＋リスク評価強化（方向性）	「人間判断必須」の明記（要旨）	ガイドライン自体の直罰は想定しにくい	国内調達・契約で準拠要求が広がる可能性

枠組み	拘束力	中心アプローチ	人間監督（Human oversight）	監督・罰則	国際影響
EU：欧州連合[47] AI Act	法規制（規則）	リスクベース（高リスク等）	高リスクで人間監督を要求（条文上）	行政罰（罰金）を含む	域外企業にも影響（EU市場アクセス要件）
NIST[48] AI RMF	任意	リスク管理フレームワーク（Govern/Map/Measure/Manage）	組織が文脈に応じて統制を設計	罰則なし	国際的な参照枠として利用されやすい
英国[49]政府方針（White Paper）	当初は非法定	既存規制当局が5原則を適用	原則としての安全・透明性等を求める	既存法で対応、原則は当初非法定	規制の俊敏性重視、企業負担を抑制
OECD[50] AI原則	非拘束	価値原則＋リスク管理	望ましくない挙動時の上書き／停止等の仕組みを提起	罰則なし	各国政策の共通土台

日本の改定案（要旨）に見える「人間判断必須」「リスク評価手法」「特に留意すべきユースケース」は、EU（高リスク要件）やNIST（リスク管理）、英国（原則ベース）と“目的”の方向性は一致しやすいが、**拘束力と罰則の設計が異なる**点が決定的な差分になる。[51]

EU AI Actとの整合・差分

EU AI Act（規則）は、高リスクAIに対し、人間監督が健康・安全・基本権リスクの予防・最小化を目的とすること、また監督措置がリスク・自律度・文脈に整合すること等を条文レベルで定める。[52]
また、EUは高リスク要件の適用開始時期を段階的に示しており、高リスク規則は2026年・2027年の適用が見込まれる。[53]
罰則（行政罰）も条文で規定され、違反類型に応じて高額制裁金があり得る。[54]

対して日本のガイドライン改定案は、少なくとも現時点では「罰則付き規制」ではなく、採用促進・共通言語化を狙った設計に見える（推論）。ただし日本のAI法をめぐっては、当初「罰則のない法案」である点が論点化した経緯も報じられており、社会的関心（消費者側）は“より強い実効性”を求める方向に振れ得る。[55]

米国（NIST）・英国との整合

NIST AI RMFは、任意でありつつも、組織がAIリスクを体系的に扱うための枠組みとして設計されている。[56]
英国のWhite Paperは、5原則（安全・透明性等）を掲げつつ、当初は法定化せず既存規制当局が適用する方針を明示している。[57]
日本の改定案が「リスク評価」「ユースケース」「人間判断必須（運用設計）」へ踏み込むほど、NIST・英国の“原則→運用”系アプローチと整合しやすくなる一方、定義・例示が不足すると各社実装がばらつく（＝原則の実装ギャップが拡大する）懸念がある。[58]

（参考）米国の行政措置として知られたEO 14110は、NIST情報によれば2025年1月に撤回された。米国が一枚岩の規制枠組みを持たない状況は今後も続き得るため、日系企業は「EU向けはEU法、米国向けは任意枠＋州法等」といった二重（多重）対応が必要になり、日本ガイドラインがその橋渡し（社内統制の共通基盤）になる余地がある（推論）。[59]

ステークホルダー反応、実施スケジュール、監督・罰則、遵守支援策と提言

主要利害関係者の想定反応と論点

- 政府（総務省[60]・経済産業省[61]、内閣府、関係府省）
- 論点：イノベーション促進とリスク対応の両立、国際競争力の確保、官民の共通言語化。[62]
- 大企業・AI提供事業者
- 反応（想定）：既にIAM（権限管理）や監査基盤を持つ企業は適合しやすい一方、エージェントの外部アクションが製品価値の中核である場合、UX低下（承認待ち）とのトレードオフを強く意識。[63]
- 中小企業・スタートアップ
- 反応（想定）：統制設計・ログ基盤の整備が「人手不足をさらに逼迫させる」懸念。支援策（テンプレ、参照実装、簡易監査）への需要が高い。[64]
- 消費者団体・市民社会
- 反応（想定）：誤作動やプライバシー侵害が顕在化するほど、任意指針では不十分として、より強い実効性（調査権限・罰則等）を求める議論が起こり得る。AI法案段階で「罰則なし」が論点化したことは、この方向性を示唆する材料になる。[55]
- 研究者・専門家
- 論点：定義の厳密化（AIエージェント／フィジカルAI／自律度）、評価手法（リスク評価、検証、レッドチーミング等）、説明可能性とプライバシーの両立。[65]
- 労働組合・労働政策関係
- 論点：監視・評価へのAI適用、職務再設計、事故時責任が現場に転嫁されない設計（人間承認が“形式責任”にならないこと）。OECD原則が「労働権」も含むことは論点の方向性を裏付ける。[66]

実施スケジュール（確認できる範囲）

- 2025年9月1日：AI法の全面施行、人工知能戦略本部の設置（内閣府説明）。[67]
- 2025年12月23日：人工知能基本計画の閣議決定（内閣府発表）。[68]
- 2026年2月16日：第29回AIガバナンス検討会で「令和7年度更新内容（案）」提示（解説要旨）。[16]
- 2026年3月末：AI事業者ガイドライン第1.2版 公開予定（解説要旨）。[16]

（注）会議資料本文に基づく確定日程は未確認のため、改定版公開後に要再検証。[69]

監督・罰則の有無・可能性

- **AI事業者ガイドライン**：従来整理として法的拘束力はない（直罰なし）とされる。[70]
- **AI法（日本）**：制度設計としては国の司令塔・指針整備・情報収集等を位置づけるが、少なくとも法案段階で「協力しなくても罰則はない」点が論点化している（報道要旨）。[55]
- **EU AI Act**：高リスク分野で義務・監督が制度化され、罰則（行政罰）も条文で規定される。[71]

「可能性」としては、日本側も重大インシデントが続発した場合、（ガイドラインの実効性を補う形で）特定高リスク領域に限った規律強化が議論され得る。ただし現時点で、改定案＝直罰付き規制化、と断定できる一次根拠はないため、本稿では“将来議論の射程”に留める。[20]

遵守支援策の提案（政策・実装の両面）

短期に効果が出やすいのは、「要件の具体化」と「実装コストの逓減」である。

- **参照実装（リファレンスアーキテクチャ）の公開**：
  Human-in-the-Loop、最小権限、監査ログ、リスク別制御を、クラウド／オンプレ／ロボ等の代表パターンで示す。[72]

- **“外部アクション／重要変更”の類型表と例示**：
  業種別に、承認必須の境界事例を提示し、過剰統制（UX低下）と過小統制（事故）の双方を抑制。[35]
- **中小企業向け簡易チェックリスト／テンプレ**：
  既存の「使い方ガイド」「チャットボット」等支援策が示唆されているため（解説要旨）、そこへ統制設計の雛形を統合する。[16]
- **公共調達・行政利用ガイドラインとの整合**：
  デジタル庁[73]は行政向け生成AI調達・利活用ガイドラインを策定しているため、官側の要求水準（ログ、データ管理等）と民間ガイドラインの整合を取ると、調達・監査の摩擦が減る。[74]
- **国際クロスウォークの整備**：
  EU AI Act（人間監督・文書化）とNIST AI RMF（任意枠）の対応表を整備し、輸出企業の二重負担を削減。NISTはクロスウォーク類を継続更新している。[75]

結論と政策提言（短期・中期・長期）

- 短期（～2026年3月末～直後）
  1) AIエージェント／フィジカルAIの定義と境界事例を明文化（不特定状態の解消）。[27]
  2) 「人間判断必須」を“どの行為に”適用するか（外部アクション／重要変更）を、例示ベースで具体化。[31]
  3) 最小権限・監査ログの**最低限要件**（最低限のログ項目、権限付与の原則、保存期間の考え方）を提示。[76]

- 中期（2026～2027）
  1) “統制設計”の社会実装支援（参照実装、監査・保証プロセス、教育）を整備し、企業の遵守コストを下げる。[77]
  2) AI法に基づく指針群とAI事業者ガイドラインの役割分担（汎用／高リスク／行政利用等）を明確化し二重規範化を避ける。[78]
  3) EU AI Actの施行ステップ（2026/2027適用）を見据え、越境対応のガイダンスを用意する。[79]

- 長期（2027以降）
  1) 高リスク領域（生命・身体安全、重大な権利侵害、重要インフラ等）に関して、必要なら“任意指針＋実効性確保策”の組合せ（監督、報告、事故調査）を検討。AI法案で「罰則なし」が論点になった経緯も踏まえ、社会的合意形成を重視。[20]
  2) OECD等の国際原則に整合する形で、停止可能性・説明責任・救済（redress）を、エージェント／フィジカルAI時代の“標準コントロール”として定着させる。[80]

【脚注（参照URL一覧）】
1. https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02tsushin06_04000136.html
2. https://b.hatena.ne.jp/entry/s/www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02tsushin06_04000136.html
3. https://ledge.ai/articles/government_ai_guideline_revision_human_judgment_required_x_debate
4. https://innovatopia.jp/ai/ai-news/80408/
5. https://kouhaku-st.com/mailmagazine/setting-the-table-20251202/
6. https://note.com/novi__1988/n/nb22d38fca4da
7. https://prtimes.jp/main/html/rd/p/000000687.000099810.html
8. https://www8.cao.go.jp/cstp/ai/ai_act/ai_act.html
9. https://www.cao.go.jp/press/new_wave/20251003.html
10. https://www8.cao.go.jp/cstp/stmain/20250901ai.html
11. https://www8.cao.go.jp/cstp/stmain/20251205ai.html

12. https://www.cao.go.jp/press/new_wave/20260206.html
13. https://www.digital.go.jp/news/3579c42d-b11c-4756-b66e-3d3e35175623
14. https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10
15. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
16. https://www.oecd.org/en/topics/sub-issues/ai-principles.html
17. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
18. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
19. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-99
20. https://www.ipa.go.jp/disc/committee/expert-group-on-aigfb.html
21. https://www.meti.go.jp/press/2023/01/20240119002/20240119002.html
22. https://www.meti.go.jp/press/2024/04/20240419004/20240419004.html

---

1 11 61 https://ledge.ai/articles/government_ai_guideline_revision_human_judgment_required_x_debate
https://ledge.ai/articles/government_ai_guideline_revision_human_judgment_required_x_debate

2 21 22 70 https://www.pwc.com/jp/ja/knowledge/column/ai-governance/ai-guideline.html
https://www.pwc.com/jp/ja/knowledge/column/ai-governance/ai-guideline.html

3 9 10 13 14 15 17 18 23 26 29 30 31 32 33 35 36 38 43 63 72 76 https://prtimes.jp/main/html/rd/p/000000687.000099810.html
https://prtimes.jp/main/html/rd/p/000000687.000099810.html

4 12 25 27 34 40 41 42 65 https://kouhaku-st.com/mailmagazine/setting-the-table-20251202/
https://kouhaku-st.com/mailmagazine/setting-the-table-20251202/

5 28 39 48 50 51 52 71 https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

6 78 https://www8.cao.go.jp/cstp/stmain/20251205ai.html
https://www8.cao.go.jp/cstp/stmain/20251205ai.html

7 8 69 https://b.hatena.ne.jp/entry/s/www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02tsushin06_04000136.html
https://b.hatena.ne.jp/entry/s/www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02tsushin06_04000136.html

16 24 49 https://innovatopia.jp/ai/ai-news/80408/
https://innovatopia.jp/ai/ai-news/80408/

19 62 https://www.cao.go.jp/press/new_wave/20251003.html
https://www.cao.go.jp/press/new_wave/20251003.html

20 55 https://www.asahi.com/articles/AST2W4J6LT2WULFA011M.html
https://www.asahi.com/articles/AST2W4J6LT2WULFA011M.html

37 46 66 80 https://www.oecd.org/en/topics/sub-issues/ai-principles.html
https://www.oecd.org/en/topics/sub-issues/ai-principles.html

44 45 47 73 https://note.com/novi__1988/n/nb22d38fca4da
https://note.com/novi__1988/n/nb22d38fca4da

53 75 79 https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

54 https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-99
https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-99

56 58 60 77 https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10
https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10

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

59 https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence
https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence

64 https://www.nist.gov/news-events/news/2023/01/nist-risk-management-framework-aims-improve-trustworthiness-artificial
https://www.nist.gov/news-events/news/2023/01/nist-risk-management-framework-aims-improve-trustworthiness-artificial

67 https://www8.cao.go.jp/cstp/stmain/20250901ai.html
https://www8.cao.go.jp/cstp/stmain/20250901ai.html

68 https://www.cao.go.jp/press/new_wave/20260206.html
https://www.cao.go.jp/press/new_wave/20260206.html

74 https://www.digital.go.jp/news/3579c42d-b11c-4756-b66e-3d3e35175623
https://www.digital.go.jp/news/3579c42d-b11c-4756-b66e-3d3e35175623