発行者：ISACA 東京支部基準委員会 AI ガバナンスの基準研究グループ

（福田重遠、小林清志）

発行日：2025年2月11日

改訂日：2026年7月 5 日

題名：AI ガバナンスに関する基本的な考え方

前書き

今回は、IT ガバナンスと AI ガバナンスの相違点を明確に示した上で、AI ガバナンスに関する基本的な考え方を整理しました。本稿を執筆するにあたり、複数の LLM を活用し、現在の AI ガバナンスの考え方を整理した上で、AI に関する専門的な知見を有するメンバーが本考察をまとめました。

なお、近年の生成 AI、基盤モデル、エージェント型 AI などの急速な発展に伴い、AI ガバナンスに求められる視点も変化しています。各国の社会環境や規制、国際標準、企業における利用実態も継続的に変化していることから、AI ガバナンスは一度定めれば固定される静的枠組みではなく、技術・リスク・社会的要請の変化に応じて継続的に見直し・更新する動的かつ柔軟な考え方として捉えることが重要です。

本稿は、2026年時点における AI ガバナンスの基本的な考え方を整理したものです。今後の AI の発展や社会的・制度的変化に応じて、内容を継続的に見直す必要がある点を考慮の上、ご一読いただければ幸いです。

第1章　IT ガバナンスと AI ガバナンスの相違

1. はじめに

IT ガバナンスと AI ガバナンスの相違点を明確にすることで AI ガバナンスの特異性を明確にします。

2. IT ガバナンスの概要

IT ガバナンスは日本の経済産業省の定義によると、企業が IT システムを戦略的に活用し、その効果を最大化するための組織的な仕組みです。 ステークホルダーのニーズに基づき、組織の価値を高めるための IT 戦略と方針の策定及びその実現のための活動とされています。主要な構成要素は以下の通りです。

- 戦略と情報システムの整合性
- 組織体制の確立
- 業務内容の把握
- コストの最適化
- 運用体系の構築
- ガイドラインの設定
- リスク管理
- システムの調達方法

3. AI ガバナンスの概要

経済産業省と総務省が策定した「AI 事業者ガイドライン」によると、AI ガバナンスとは、AI の利活用によって生じるリスクを関係者（ステークホルダー）にとって受容可能な水準に管理しつつ、そこからもたらされる正のインパクト（便益）を最大化することを目的とする、技術的、組織的、および社会的システムの設計・運用を指します。

これは、システムの管理を中心とする従来の IT ガバナンスとは異なり、AI 特有のリスク（AI による誤判定、偏見や差別、プライバシー侵害など）や倫理的考慮に対応する点に特徴があります。具体的には、AI の透明性、公正性、安全性を確保し、リスクを特定・評価・軽減し続ける「アジャイルな（継続的かつ機敏な）プロセス」が求められます。

国際的には、法的拘束力を持つ「EU AI 規則（Regulation (EU) 2024/1689、2024年8月1日発効）」が世界初の包括的 AI 法として施行されており、事実上の国際基準として各国の法整備や企業の対応に大きな影響を与えています。

4. IT ガバナンスと AI ガバナンスの基本的な相違点

AI ガバナンスは、IT ガバナンスの枠組みの中で機能すると考えられています。AI システムは IT システムの一部として運用されることが多く、IT ガバナンスで確立されたリスク管理、コンプライアンス、セキュリティ対策などの仕組みを AI システムにも適用する必要があるためです。

基本的な観点の相違点は以下の通りです。

観点	IT ガバナンス	AI ガバナンス
主な目的	IT 投資の最適化とリスク管理	AI 技術の倫理的・責任ある利用
重点領域	システム運用と効率化	倫理的配慮と社会的影響
特徴的なリスク	システム障害、セキュリティ	バイアス、判断の透明性、誤った判断、差別
生成 AI・LLM 特有のリスク	*(技術の性質上、対象外)*	**ハルシネーション** （事実と異なる尤もらしい出力） **プロンプトインジェクション / ジェイルブレイク** （間接インジェクションを含む、悪意ある指示入力による脆弱性） **著作権・知的財産権リスク** （学習データや出力に起因するライセンス違反等） **ディープフェイク・偽/誤情報の拡散** （情報インテグリティの毀損） **エージェンティック AI の挙動** （自律的に考えて行動する AI の挙動）

第 2 章以降では、IT ガバナンスと AI ガバナンスの考え方の相違を念頭においたうえで、AI ガバナンスの

基本的な考え方や AI ガバナンスの策定方針の考え方について解説します。

第2章　AI ガバナンス原則の策定の考え方

1. はじめに

AI ガバナンス原則の構成要素の策定とその考慮点を明確化にしたうえで、AI ガバナンスに関するベストプラクティス例を提示します。

2. AI ガバナンス原則の構成要素

生成 AI を用いて多様な情報を分析した結果、AI ガバナンス原則は、以下の要素から成り立つという結論に至りました。

① **人間中心主義:** AI 開発・利用において、常に人間の幸福と権利を最優先する。経済産業省の「人間中心の AI 社会原則」にも示されるように、倫理的な配慮が不可欠である。また、人間による監督も重要となる。

② **信頼性と安全性:** AI システムの安全性と信頼性を確保する。これは、システムの堅牢性、予測可能性、説明可能性などを含む。ハルシネーション対策（RAG 等）、エージェンティック AI に対する Human-in-the-loop と「人間によるオーバーライド（介入・停止・デコミッショニング）」

③ **透明性と説明可能性:** AI アルゴリズムの動作と意思決定プロセスを明確にし、その結果を説明できる必要がある。IBM の原則にも含まれるように、ブラックボックス化を防ぎ、説明責任を

果たすための重要な要素である。AI の意思決定プロセスの追跡可能性は重要な視点で、今後、技術の進歩とともに説明可能性を高める方法論も発展する。例えば、基盤モデルのブラックボックス性、来歴管理（provenance）、AI 生成物の電子透かし・メタデータ付与（C2PA 等）

④ **公平性と倫理:** AI システムによる差別やバイアスを排除し、公平性を確保する。トレーニングデータの厳密な調査、バイアス検知・軽減技術の活用、プロンプトインジェクションによる有害出力誘導リスクと、レッドチーミングによる事前評価が不可欠となる。

⑤ **説明責任:** AI システムの運用に関する責任を明確化し、責任体制を構築する。重大な変化への対応やリスク管理体制の構築が重要となる。

⑥ **プライバシー保護:** AI システムの開発・利用における個人情報の適切な保護を確保する。データガバナンスと密接に関連する重要な原則である。例えば、学習データの記憶（memorization）による漏洩、入力プロンプト中の機密情報再利用リスク

⑦ **国際協調と標準化:** 国際的なコンセンサスに基づく原則の策定と標準化が求められる。ISO/IEC JTC 1 SC 42などの国際標準化活動や、OECD などの国際的なイニシアティブへの参加が重要となる。また、中国、米国、欧州それぞれで AI の捉え方が異なり、国家の政治的、経済的、文化的背景が大きく影響する。各国で開発する国産 AI の開発については、自国の価値観を反映させることも重要だが、同時にグローバルな相互運用性と倫理的な普遍性も考慮する必要がある。

項目	内容
EU AI Act	Regulation（EU）2024/1689。2024年8月1日発効。段階適用：**2025年2月2日**＝禁止行為（第5条）＋AI リテラシー義務（第4条）、**2025年8月2日**＝GPAI ルール（第5章）・ガバナンス・罰則、**2026年8月2日**＝高リスク AI 含む大部分の全面適用、2027年8月2日＝Annex I 対象。2025年2月の禁止行為ガイドライン公表、7月の GPAI 行動規範（Code of Practice）
日本：AI 推進法	「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）。**2025年5月28日成立、6月4日公布・一部施行**。内閣総理大臣を本部長とする AI 戦略本部の設置、AI 基本計画策定を規定。
AI 事業者ガイドライン	旧3ガイドライン（引用[3]含む）は統合され、第1.0版（2024年4月）→**第1.1版（2025年3月28日）**へ。第1.1版で AI エージェント／フィジカル AI の定義と関連リスクが追記。
広島 AI プロセス	報告枠組み（Reporting Framework）が2025年2月に OECD サイトで運用開始、初回回答が2025年4月公表。フレンズグループの拡大
OECD AI 原則改訂	**2024年5月3日改訂**。情報インテグリティ（誤・偽情報）、環境持続可能性、人間によるオーバーライド、相互運用可能なガバナンスを追加。トレーサビリティを説明責任原則に集約。
米国の政策転換	バイデン政権の EO 14110（2023年10月30日）が**2025年1月20日に撤回**。トランプ政権が EO 14179「Removing Barriers to American Leadership in AI」（2025年1月23日）**を発令し、安全・権利重視から「米国の AI 優位確保・イノベーション重視」へ転換。NIST AI RMF は維持。AISI が** CAISI（Center for AI Standards and Innovation）に再編（2025年6月）
ISO/IEC 標準	ISO/IEC 42001:2023（AI マネジメントシステム、2023年12月）、ISO/IEC 23894:2023（AI リスク管理）、**ISO/IEC 42005:2025（AI システム影響評価、2025年5月）**、ISO/IEC 42006:2025（監査・認証機関要求事項）

その他、**欧州評議会の AI 枠組条約**（初の拘束力ある国際条約、2024年9月署名開

放）、**韓国 AI 基本法**（2024年末成立、2026年施行）、**パリ AI 行動サミット**（2025年2月）、**中国の生成 AI・コンテンツ表示規制**。

3. 原則策定における考慮事項

AI ガバナンス原則の策定にあたっては、以下の点を考慮する必要があると考えます。

① **組織の規模と業種:** 企業規模や業種によって、AI の活用状況やリスク要因が異なるため、原則の内容も調整する必要がある。

② **AI システムの種類と用途:** 使用する AI システムの種類（GPAI・基盤モデル、生成 AI、エージェンティック AI など）を明確に分類した上で、その用途に応じて、求められるガバナンスのレベルや重点事項を（柔軟に）変化させる必要があります。

③ **リスクアセスメント:** AI システム導入による潜在的なリスクを事前に評価し、適切なリスク軽減策を講じる必要がある。評価項目に「プロンプトインジェクション耐性」「ハルシネーションの影響度」「学習データの権利侵害」「ディープフェイク悪用」「エージェンティック AI の連鎖的影響」を追加。**ISO/IEC 42005や EU の FRIA（基本的権利影響評価）と連動した影響評価**の実施を推奨。

④ **コンプライアンス:** 関連法規制や業界標準への準拠を確保する必要があり、以下の規制領域への検討が想定される。なお、各国の法規制等によりコンプライアンスの視点は異なる可能性がある。

＜例示＞

- プライバシーの侵害
- 他者知的財産権の権利侵害
- 人類に危害を与える兵器への使用
- データの公正な使用
- アルゴリズムによる差別の防止
- AI 開発における知的財産権の権利明確化
- サイバーセキュリティ
- 労働市場への影響
- 環境への配慮
- EU AI Act への適合
- AI 推進法・AI 事業者ガイドライン第1.1版への準拠
- GPAI 提供者の著作権・学習データ要約の透明性義務
- ディープフェイクによるなりすまし対策
- プロンプトインジェクション等のセキュリティ対応

一例として、日本の文化庁が策定した「AIと著作権に関する考え方について」（2024年3月）および、実務者向けに作成された**「AIと著作権に関するチェックリスト＆ガイダンス」**

（2024年7月）における「著作権法第30条の4」の解釈において、全体として、「AI の開発・学習段階（原則として許諾不要）」と「生成・利用段階（通常通りの著作権侵害の判断）」を明確に区別し、どのような場合に例外的にアウト（許諾が必要）になるか、実務上どう対応すべきかを具体的に示しています。

⑤ **ステークホルダーとの連携:** 利害関係者（従業員、顧客、社会全体）とのコミュニケーションとその具体的な検討例は以下がその項目となる。また、ステークホルダーとの連携に当たっては、サプライチェーン視点も考慮に入れる必要がある。

＜例示＞

- 従業員：定期的な研修、フィードバックメカニズム
- 顧客：透明性のある情報提供、苦情処理システムやフィードバックメカニズム等
- 政府：規制遵守、政策対話
- 取引先：倫理的、かつ安全な AI 利用の共通ガイドライン、フィードバックメカニズム
- 学術界：共同研究、技術検証
- サプライチェーン上の関係者（提供者／導入者）：EU AI Act 等の役割区分に基づく責任分担、川上～川下での情報伝達

また、将来的な AI のテクノロジー面での発達も考慮し、AI のガバナンス原則は継続的な変更や追加の可能性があることも留意する必要があります。

4. AI ガバナンスの実践におけるベストプラクティス

AI ガバナンス原則の効果的な実践において、国際標準規格である「ISO/IEC 42001:2023」に基づくAI マネジメントシステム（AIMS）の構築と運用を中核に据えるアプローチが極めて重要です。ただし、すべての組織に一律で最適なわけではなく、自社のビジネス特性や AI への業務依存度に応じて、これらのベストプラクティス例から適切に取捨選択・カスタマイズする必要があります。

- **AIMS（AI マネジメントシステム）に基づくガバナンス体制の構築：** ISO/IEC 42001:2023 に準拠した管理体制（AIMS）を組織に確立し、AI ガバナンス責任者や倫理委員会などの設置、役割・責任・権限の明確化を行い、透明性の高い意思決定プロセスを確保する。
- **多様な専門家の起用と「AI リテラシー教育プログラム」の実践：** AI ガバナンス運用にあたり、法務、倫理、技術、ビジネスなどの多様な専門家を参画させて多角的な視野を確保する。また、全社的な「AI リテラシー教育プログラム」を構築・整備し、すべての従業員に対して継続的な教育と啓発活動を恒常的に実施する。

- **目標設定と指標管理:** AIMS の運用方針に則り、AI ガバナンスにおける明確な目標を設定し、各種管理指標（KPI）を設計したうえで、その達成状況を組織的に評価・モニタリングする。
- **最新の多層的リスク管理フレームワークの導入:** AI システムのライフサイクル全般におけるリスクを特定・評価・軽減するため、以下のような最先端のリスク管理フレームワークをプロセスに組み込む。
- **基礎的フレームワーク**:「NIST AI RMF 1.0」を全体の共通基盤とする。
- **生成 AI への特化対応**: 生成 AI に固有のハルシネーションや情報漏洩といった高度なリスクに対処するため、**「NIST AI 600-1（Generative AI Profile、2024年7月26日発行）」**に基づくリスクアセスメントの実施。
- **攻撃・脆弱性への防備**: 敵対的機械学習（Adversarial ML）やデータ汚染といったシステムへの攻撃脅威モデルを明確化するため、高度な分類と語彙を定めた**「NIST AI 100-2e2025（2025年発行）」**を取り入れる。
- **レッドチーミング（Red Teaming）の実践**: 開発・配備前や定期的なプロセスとして、模擬的な攻撃や脱獄（Jailbreak）などを能動的に仕掛け、システムの脆弱性や予期せぬ有害挙動を事前に検出・修正するアプローチ。
- **倫理ガイドラインの策定と遵守:** 国際的な倫理潮流と自社のミッションを整合させ、組織独自の「AI 倫理ガイドライン」を策定するとともに、それが現場の業務プロセスで確実に遵守される仕組みを作る。

- **データガバナンスとの統合:** 入力データの品質、バイアス、機密性、プライバシー保護の観点から、既存のデータガバナンス体系と AI ガバナンスをシームレスかつ統合的に管理する。
- **継続的モニタリングと動的改善:** 本番運用開始後も、AI のデータドリフトやモデルドリフト、システムの出力品質低下、新たなハッキング試行などを検知するため、**継続的なモニタリング**の仕組みを常時稼働させる。運用状況から得られた実データを基に、ガイドラインや AIMS（AI マネジメントシステム）を PDCA サイクルで迅速に、また動的に改善し続ける。
-

※.マイクロソフト、Google および IBM は、ISO/IEC 42001などの国際標準規格への準拠・適合をいち早く進め、また NIST フレームワーク等の最先端の管理手法を自社の監査プロセスと実質的に調和させて活用するなど、極めて先進的な AI ガバナンス体制を自社内で構築・実践している事例とされています。

第3章　まとめと今後

AI の発展に伴い、AI ガバナンスの重要性は一層高まっています。特に生成 AI、基盤モデル、エージェンティック AI などの新技術の登場により、AI システムは従来の固定的な情報システムとは異なり、利用環境やデータ、外部システムとの相互作用を通じて、リスクの現れ方が変化し得るものとなっています。そのため、欧州の「EU AI Act（AI 法）」や、日本国内で2025年9月に全面施行された「AI 推進法（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）」といった法規制への適合はもとより、米国 NIST の「NIST AI RMF（AI リスクマネジメントフレームワーク）」をはじめとするグローバルな枠組みと連動した、より包括的かつ実効性のあるガバナンス体制の構築・運用が求められています。また、AI ガバナンスに関連して IT ガバナンスの重要性も相対的に高まっています。

AI ガバナンス原則の策定と実践は、AI 技術の安全で倫理的な活用に不可欠です。第 1 章で示した IT ガバナンスとの相違、第 2 章で示した原則、考慮事項、ベストプラクティス例を参考に、組織の状況に合わせた適切な AI ガバナンス体制を構築することが重要です。ただし、AI ガバナンスは、導入時点での方針やルールを策定して完了するものではありません。AI 技術の進歩、モデルの利用形態、社会的受容性、法規制や国際標準の変化に応じて、リスク認識や管理策も変化します。そのため、AI ガバナンスを継続的に改善・向上するには、国際標準とベストプラクティスを踏まえた包括的なアプローチを行い、定期的なガバナンス態勢のレビュー、リスクアセスメントの継続的な実施、倫理ガイドラインの見直し、関係者と

の連携、人材育成、技術革新への対応などを通じて、動的に更新し続ける姿勢が求められます。

ISACA 国際本部においても AI に関連するガバナンスや監査の枠組み整備が急速に進んでおり、2024年7月にリリースされた『AI Audit Toolkit[5]』に続き、2025年5月には新たな専門認定資格である「AAIA（Advanced in AI Audit）」がリリースされました。本資格は CISA、CIA、CPA などの専門資格保持者を対象としており、その試験は「AI Governance and Risk（33%）」「AI Operations（46%）」「AI Auditing Tools and Techniques（21%）」の3つのドメインで構成されています。また関連して、CISM や CISSP の保持者を対象とした「AAISM（Advanced in AI Security Management）」（2025年 Q3リリース）などの後続資格も展開されています。今後の ISACA の AI に関するガバナンス、セキュリティ、監査に関するリリースについても、継続してご案内をしていきます。

また、本記事は ISACA 東京支部基準委員会のボランティアによりまとめられたもので、本記事に記載された情報は、収集時点での最善の努力をもって正確性を期しておりますが、その完全性や正確性について必ずしも保証するものではありません。本情報を利用するにあたり、自己の責任において判断していただくようお願いいたします。

以上

<引用>

利用 AI モデル:　Google Gemini シリーズ、Anthropic Claude シリーズ，Perplexity，OpenAI 社各種モデル

[1] システム管理基準

https://www.meti.go.jp/policy/netsecurity/sys-kansa/sys-kanri-2023.pdf

[2] AI ガバナンス

https://www.meti.go.jp/policy/it_policy/ai-governance/index.html

[3] AI 事業者ガイドライン第1.1版（2025年3月28日）

https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20250328_2.pdf

[4] デジタルガバナンス・コード

https://www.meti.go.jp/policy/it_policy/investment/dgc/dgc.html

[5] AI Audit Toolkit

https://store.isaca.org/s/store#/store/browse/detail/a2S4w000007kB9pEAE

[6] EU AI Act（Reg. (EU) 2024/1689）

https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

[7] AI 推進法（令和7年法律第53号）

https://laws.e-gov.go.jp/law/507AC0000000053

[8] OECD AI Principles（2024年改訂版）

https://oecd.ai/en/dashboards/policy-initiatives/oecd-ai-principles-9705

[9] NIST AI 600-1

https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

[10] 米国 EO 14179

https://www.govinfo.gov/content/pkg/DCPD-202500170/pdf/DCPD-202500170.pdf