CAIO 設置・AI ガバナンス実務マニュアル(案)

2026 年 2 月

AI セーフティ・インスティテュート（AISI）事務局

AISI Japan AI Safety Institute

目次

要約

本マニュアルは、主に民間事業者が CAIO を設置し、AI 戦略の推進、AI ガバナンス、リスクマネジメント、セキュリティ・プライバシー、データガバナンス、人材育成、調達・ベンダー管理、監査・モニタリングを統合的に実現するための実務指針である。生成 AI やエージェント型 AI が事業インフラの一部となる中で、企業は価値創出と権利・安全・公平性・プライバシー等の保護を両立させる体制が求められている。本マニュアルは、こうした要請に応え、AI を経営課題として扱うための統合アーキテクチャを提示する。

CAIO は、単なる技術統括者ではなく、AI に関する戦略、ガバナンス、リスク、人材、調達を結びつける「単一の責任点 (Single Point of Accountability)」として位置付けられる。CEO 直下の独立した C-suite として CAIO を配置し、その直下に AI ガバナンス室 (AI 推進チーム) を設けるとともに、CAIO、CDO、CIO、CTO、CISO、法務・コンプライアンス、DPO、人事、主要事業部等で構成される全社 AI ステアリングコミッティを設置する体制を推奨する。この体制により、高リスク AI ユースケースの採否、投資配分、リスク許容度、重大インシデント対応などを経営と一体で意思決定し、分散的な取り組みや責任の曖昧さを最小化することを狙いとする。

本マニュアルは、AI 事業者ガイドライン、NIST AI Risk Management Framework (AI RMF)、ISO/IEC 42001 (AI マネジメントシステム)、ISO/IEC 23894 (AI リスクマネジメント手法)、EU AI 法、米国 OMB ガイダンス (M-24-10) 等の国内外の枠組みを参照しつつ、AI インベントリ、AI 影響評価 (AIIA)、モデルカード・データシート、リスクレジスター、KPI・ダッシュボード、内部監査・外部認証といった要素を組み合わせた実務的な運用モデルを提示する。特に、人の生命・身体・自由・財産その他の基本的権利や安全性に重大な影響を与えうる「権利・安全影響 AI」については、CAIO および全社 AI ステアリングコミッティによる承認、人間による監督・最終判断 (Human-in-the-loop)、影響評価・公平性評価・救済プロセス等の追加統制を求める。

また、本マニュアルは、CAIO に求められるスキルと資質、役割・責務、組織設計と規模別の体制案 (大企業・中規模企業・スタートアップ)、社内協働および社外ステークホルダー関与のポイント、コンプライアンス・規制対応、セキュリティ・プライバシー、データガバナンス・品質保証、調達・ベンダー管理、教育・人材、KPI・ダッシュボード設計、主要リスクと緩和策、監査・モニタリング・報告、ユースケース別適用例、テンプレート要点、初年度実践計画例までを一貫した構造で整理している。これにより、各社が自社の規模・業種・リスクプロファイル・既存の内部統制との整合を踏まえつつ、CAIO を中心とした AI ガバナンス体制を段階的に設計・導入し、本マニュアルを「Living Document」として参照しながら継続的に改善していくための道筋を示すものである。

1. 総論

1.1 背景

生成 AI やエージェント型 AI は、従来の AI による自動化や分析といった枠を超えて、企業の事業活動に浸透しつつあり、顧客接点では対話型エージェントやパーソナライズされた提案、バックオフィスでは文書作成や分析業務の効率化、製造や物流では予測保全や最適化など、価値創出の可能性を広げている。他方で、法令遵守、セキュリティ、プライバシー、公平性、説明責任などの複合的なリスクも増幅している。アルゴリズムによる意思決定の透明性や差別の予防、個人情報の適正利用、生成物の真正性に関する課題は、単なる IT ガバナンスの延長では対応が難しく、個別部署での試行錯誤だけでは管理が追いつかない状況が生まれており、AI 特有の統治と運用のフレームワークが必要となっている。

国内外では AI 事業者ガイドライン[1]、NIST AI RMF[2]、EU AI 法[3]、ISO/IEC 42001[4]といった枠組みが整備されつつある。例えば米国の連邦政府機関では CAIO の指定（設置）を前提としたガバナンス要件が示されており、日本の国の行政機関でも、デジタル庁の「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」[5]において各府省庁に AI 統括責任者（CAIO）の設置が求められている。これらは、公的機関にとどまらず民間事業者においても、AI を単一の技術テーマとしてではなく、経営課題として統合的に管理することが求められていることを示している。

一方、多くの企業では、各部門の裁量で生成 AI ツールや AI サービスの導入が先行し、投資判断、品質基準、リスク評価、ベンダー選定などが部門ごとにばらばらに行われる傾向がある。その結果、重複投資や過少投資により ROI の最大化が困難になるだけでなく、法令違反やセキュリティインシデント、差別・不公平の顕在化、ブランド毀損などのリスクが、後追いで発覚するおそれが高まっている。AI の利活用が事業の中核インフラとなるにつれ、こうした分散的な取り組みを前提としたガバナンスには限界が生じている。

このような状況を踏まえ、AI 戦略、AI ガバナンス、リスクマネジメント、人材育成、調達・ベンダー管理を全社視点で統合し、経営層に対して継続的に説明責任を果たす専任の統括責任者の必要性が高まっている。国際的にも CAIO 等の役職を設ける動きが広がっており、本マニュアルではこの役割を担う CAIO を中核に据える。CAIO の具体的な必要性とミッションについては、第 3 章「CAIO の必要性と目的」で詳述する。

---

[1] 経済産業省: AI 事業者ガイドライン

[2] NIST: AI Risk Management Framework

[3] The EU Artificial Intelligence Act

[4] ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system

[5] 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

このような専任の責任者を中核に据え、AI の価値創出と権利・安全・公平性・プライバシー等の保護を両立させる体制を構築することは、民間事業者が AI を持続的かつ信頼性高く活用し、国内外の規制動向に対応しながら競争力と社会的信頼を維持・強化していくうえで、不可欠な前提となりつつある。

1.2 本マニュアルの目的

本マニュアルの目的は、民間事業者が CAIO を設置・運用する際の標準的な実務指針を提供し、AI による価値創出（ROI 最大化）と責任ある活用（リスク低減・規制遵守）の両立を支援することである。CAIO の役割を明確化し、組織設計、プロセス、評価、監督、教育、調達などの要素を統合的に示すことで、各社が自社の文脈に合わせて着実に AI 活用を進められるようにする。

あわせて本マニュアルは、AI 事業者ガイドライン、NIST AI RMF、EU AI 法、ISO/IEC 42001 などの国内外の枠組みを、民間事業者の CAIO および関連部門が日々の業務として実装できるように翻訳し、AI ガバナンス・リスクマネジメント・人材育成・調達・監査といった機能を一貫したアーキテクチャとして整理することを目指す。主たる利用者としては、CAIO 本人とその実務スタッフ、CAIO の任命を検討する経営陣、ならびに法務・コンプライアンス・情報システム・データ・人事・事業部門などの関係者を想定する。

1.3 範囲

本マニュアルは、経営、事業、技術、コンプライアンスの各部門を横断する全社的運用に焦点を当てる。具体的には、組織設計、役割・責務、運用プロセス、テンプレート、KPI、成熟度モデル、監査・報告の仕組みなどを包含し、企画から運用、ベンダー調達までのライフサイクル全体を対象範囲とする。単一の部門に閉じた指針ではなく、企業全体で AI を安全かつ効果的に機能させるための基盤作りを目的とする。

主として日本国内の民間事業者を念頭に置いているが、海外に事業展開する企業や、類似の課題を抱える公的機関等にとっても参考となる内容を含む。一方で、個別のアルゴリズム設計やモデルアーキテクチャといった技術仕様そのものを詳細に扱うものではなく、AI を取り巻くガバナンス・組織・プロセス・運用管理の枠組みに主眼を置く。

1.4 前提・制約

本マニュアルは、各企業の規模・業種・法域差を踏まえた一般化を行うものであり、すべての企業にそのまま適用できる「唯一の正解」を示すものではない。適用にあたっては、自社のリスクアペタイト、規制環境、事業優先度、既存の情報セキュリティマネジメントシステム（ISMS）や内部統制の枠組みとの整合を踏まえ、自社にとって実効性のある形に補正することが必要である。

また、本マニュアルは特定法域における法的助言を構成するものではない。具体的な法令・規制への適合性の判断は、自社の法務部門・コンプライアンス部門の監督のもとで行い、必要に応じて外部専門家の助言も併用することが望ましい。参照している国内外のガイドラインや規格、法令は、今後の改正・新設により変更される可能性があるため、本マニュアル自体も CAIO および関係部門による定期的な見直しを前提とした「Living Document」として運用されるべきものである。

2. 用語定義

用語	意味
CAIO (Chief AI Officer)	AI に関する戦略、ガバナンス、リスクマネジメント、人材育成、調達を統合的に率いる最高責任者。CEO 直下等のレイヤーで、事業横断の意思決定・調整を担う。
AI ガバナンス	AI 原則 (公平性、安全性、透明性、説明責任等)、ポリシーの策定と実装、運用の監督、監査を含む体系であり、企業が AI を信頼できる形で活用するためのルールと仕組みの総体。 なお、AI 事業者ガイドラインにおける定義は以下の通りであり、その定義内容は踏襲される。 「AI の利活用によって生じ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正のインパクト (便益) を最大化することを目的とする、ステークホルダーによる技術的、組織的、及び社会的システムの設計並びに運用。」
AIIA (AI Impact Assessment: AI 影響評価)	AI による、権利・安全、公平性、プライバシー等への影響を体系的に評価する手順であり、AI の導入や変更に際して事前に影響と緩和策を明らかにする評価プロセス。
PIA (Privacy Impact Assessment: プライバシー影響評価)	個人情報の取得・利用・共有に伴うプライバシーリスクと影響を事前に洗い出し、対策や責任分担を明確化する評価プロセス。
モデルカード/データシート	AI モデルやデータセットに関する設計意図、訓練データ、評価結果、制約、適用範囲などを標準化して記録する透明性ドキュメント。
AI インベントリ	全社で利用・開発する AI のユースケース、モデル、データ、ベンダーを俯瞰できる総覧であり、管理・監督の土台となるべき情報。
全社 AI ステアリングコミッティ	CAIO を議長とし、AI 戦略および AI ガバナンスに関するハイレベルな意思決定を行う全社横断の合議体。
権利・安全影響 AI	人の生命・身体・自由・財産その他の基本的権利や安全性に重大な影響を与えうるユースケースに用いられる AI。

3. CAIO の必要性とミッション

3.1 必要性

生成 AI やエージェント型 AI は事業価値を生み出す中核インフラになりつつある一方で、法令遵守、セキュリティ、プライバシー、公平性、説明責任等の観点から複雑なリスクを伴う。また、国内外のガイドラインや国際標準により、AI に関する統一的な方針とガバナンス体制の整備が求められている。しかし、これらの要請を実務レベルで統合し、全社の意思決定や日々の運用に落とし込む役割を、既存の Chief Information Officer (以下、CIO)、Chief Technology Officer (以下、CTO)、Chief Data Officer (以下、CDO) 、Chief Information Security Officer (以下、CISO) 等のみで兼務するには限界がある。

多くの企業では、AI 戦略は経営企画、技術戦略は CTO、データガバナンスは CDO、セキュリティは CISO、法令・倫理は法務部門・コンプライアンス部門といったように、AI に関わる機能が縦割りで分散している。その結果、(1) 投資判断やユースケース選定が部門ごとの最適化にとどまり、企業全体でのポートフォリオ管理や ROI 最大化が困難になる、(2) ガバナンスやリスクマネジメントの基準が部門ごとにばらつき、責任の所在が不明確になる、(3) 経営層・取締役会・社外ステークホルダーに対する説明責任を一元的に果たす窓口が存在しない、といった構造的な課題が生じやすい。

こうした課題に対処するためには、AI に関する価値創出とリスクマネジメント双方の観点から全社を俯瞰し、戦略・ガバナンス・リスク・人材・調達の各機能を統合する「単一の責任点」が不可欠である。その役割を担うのが CAIO であり、CEO 直下など経営の最上位レイヤーで意思決定に関与しつつ、AI ガバナンスの標準化、リスクベースの運用プロセス設計、KPI・ダッシュボードを通じた継続的モニタリング、ならびに外部規制・国際標準との整合を横断的にリードすることが期待される。

まとめると、CAIO の設置が必要とされる主な理由は、以下のとおりである。

- 経営戦略と AI 利活用を結びつけ、AI 投資とユースケースを全社ポートフォリオとして最適化するため。
- AI 原則・ポリシー・承認フロー・監査を社内で共通化し、「誰が・何に責任を負うか」を明確にするため。
- 権利・安全・公平性・プライバシー等に関するリスクマネジメントをライフサイクル全体で継続的に運用し、規制変更を含む内部環境変化にも迅速に対応するため。
- 人材育成・文化醸成・ベンダー管理・外部認証等を一体として設計し、AI の価値創出と社会的信頼の両立を図るため。

3.2 ミッション

CAIO のミッションは、AI による事業価値の創出を最大化すると同時に、権利・安全・公平性・プライバシー・セキュリティ等を保護し、社会から信頼される形で AI を活用できるようにすることで

ある。このミッションを実現するために、CAIO は、戦略、ガバナンス、リスク、人材、調達の各機能を結びつける統合アーキテクチャを設計し、全社の意思決定と日々の運用に組み込む役割を担う。このミッションは、次の四つの目的に整理できる。

1) 経営戦略との整合と価値創出

AI 利活用のロードマップと投資ポートフォリオを経営戦略と整合させ、限られた資源を最も価値の高いユースケースに配分する。短期的な効率化と中長期的な競争優位の双方を見据え、AI 投資の ROI を継続的に可視化し、改善する。

2) 信頼できる AI ガバナンスの確立

AI 原則と法令・ガイドラインに基づき、ポリシー、承認フロー、監督・監査の仕組みを設計・更新し、権利・安全・公平性・プライバシー・セキュリティに関するリスクを許容水準内に保つ。AI システムのライフサイクル全体にわたり、一貫した統制と透明性を実現する。

3) 組織能力・人材・文化の形成

必要なロールとスキル標準を定義し、教育・リスキリング・人材配置を通じて、AI を適切に活用できる組織能力を内製化する。現場での実験とガバナンスを両立させる文化を醸成し、得られた知見やベストプラクティスを全社に横展開する。

4) ステークホルダーへの説明責任の遂行

経営層・取締役会、従業員、顧客、規制当局などのステークホルダーに対し、AI 戦略、リスク、対応状況を一元的に説明しうる窓口として機能する。透明性レポートや KPI、ダッシュボード等を通じて、AI 活用の状況と改善プロセスを継続的に共有し、信頼形成を図る。

4. CAIO に求められるスキルと資質

CAIO には、AI の価値創出を事業に結び付けながら、リスクを統制し、組織としての実行力を高めるための幅広いスキルと資質が求められる。以下に、CAIO に求められる主要なスキル・資質を整理する。なお、ここに掲げる要素を CAIO が一人で全て高度に備えることを前提とするものではなく、CAIO 自身の強みを核としつつ、後述する直轄チームとの役割分担によって補完することを想定している。

4.1 技術的理解・適用力

- 機械学習、自然言語処理、生成 AI (大規模言語モデル、検索拡張生成等)、MLOps 等に関する基礎理解を有し、自社ユースケースに適用する際の論点 (データ、評価、運用) を整理できる。
- 生成 AI の特性・リスク (ハルシネーション、バイアス、ブラックボックス性等) を理解し、業務利用における留意点として社内に分かりやすく周知できる。
- サイバーセキュリティ、データガバナンス、プライバシー保護の観点を踏まえ、AI システムの要求事項 (データ取扱い、アクセス制御、監査、ログ、委託先管理等) として落とし込める。

4.2 戦略的思考と事業・規制環境の洞察

- 経営目標・事業課題と AI の特性・利点を接続し、全体最適の観点で AI 利活用ロードマップを策定できる。
- 施策の市場性・実現可能性・リスクを踏まえ、効果測定 (KPI/ROI 等) と改善サイクルを設計・推進できる。
- 官民連携、学術・研究機関、スタートアップ等の外部知見を取り込み、事業とガバナンスの両面で意思決定の質を高められる。

4.3 組織変革推進力・リーダーシップ

- 部門横断型プロジェクトを統括し、関係者間の利害調整を行いながら推進できる。
- AI 活用を単なるツール導入に留めず、業務再設計、人材配置、評価・制度を含む組織文化の変革として実装できる。
- 役員・社員等の AI リテラシー向上、教育、リスキリングを計画的に推進し、人材育成の仕組みを作れる。
- 最新 AI 技術・社会動向を継続的に学び、施策・ルールに俊敏に反映する姿勢を持つ。

4.4 倫理・法務・リスクマネジメント能力

- AI 倫理・法令遵守を高度に理解し、社内ルールや審査・承認プロセスとして運用できる。

- ハルシネーション等の偽・誤情報出力やバイアス等の潜在リスクを特定・評価し、軽減策の策定・実行、継続的なモニタリングまでを一連のリスクマネジメントプロセスとして設計できる。
- 高リスク AI の利活用判断に際し、AI 倫理委員会/リスク委員会等のリスクマネジメントを担う組織・会議体と連携し、適切にエスカレーションできる。
- 説明責任（アカウンタビリティ）を果たす意識を持ち、公正性・透明性・社会的信頼の観点からガバナンスを維持できる。

4.5 適応力・先進的マインド

- 国際標準・最先端技術の動向を継続的に把握し、自社環境へ適応させるためのグローバルな視野を持つ。
- 変化が激しい AI 領域において、アジャイルな思考で柔軟に方針転換し、迅速に意思決定できる。

5. 役割・責務

CAIO の役割・責務は多岐にわたるが、現場で機能するようにするためには、責任の所在と関係者の協力体制を明確にすることが重要である。CAIO はガバナンス設計と監督の最終責任を持つが、事業固有の最終意思決定は事業責任者が保持し、重大リスクがある場合は CAIO が取締役会へエスカレーションする。また、権利・安全への影響が高いユースケースについては、CAIO は本番投入の停止・差し止めを含む否認権を持つことが望ましい。

CAIO は全社 AI ステアリングコミッティの議長等として、部門間の利害調整と全社最適の意思決定をリードする。

5.1 経営・戦略

CAIO は、全社 AI 戦略を策定し、ロードマップと投資配分を定め、ROI の管理を主導する。これには、単なる目標設定だけではなく、具体的なユースケースの選定基準、技術選択の方針、プラットフォーム戦略、ベンダーとの関係方針などを含む。CEO や事業責任者、財務部門と協調し、事業価値への期待、リスク許容度、資源制約を統合して投資判断を下す。リスク許容度は、全社的リスクマネジメント (ERM) の枠組みと整合させ、CFO やリスクマネジメント部門と連携して定期的に見直す。短期の成果と長期の基盤整備をバランスさせ、重複投資と機会損失を AI インベントリの活用などにより最小化することが求められる。事業価値・信頼性等の KPI を定義し、ダッシュボードで可視化することで、経営層・取締役会との対話の基盤を整備する。

5.2 ガバナンス・AI 原則実装

AI 原則の実装に向けたポリシーを整備し、AIIA、AI インベントリ、モデルカード、データシート等の基盤文書の整備と維持をガバナンスの要素として位置付け、承認フローと監査計画を構築する。承認フローには CAIO、事業責任者、CISO、法務部門・コンプライアンス部門等の合議を組み込み、責任と検証のバランスをとる。また法務部門・コンプライアンス部門、CDO と連携して、AI 原則を具体的な運用ルールに落とし、例外や緩和の判断基準も明文化する。これらのポリシーとプロセスは、ISO/IEC 42001 等のマネジメントシステムと一体的に運用し、PDCA サイクルおよび規制や技術の変化に応じて定期的 (例えば半期ごと) に見直すようにする。

5.3 リスクマネジメント

AI のリスクは、誤分類による不利益、バイアスによる差別、誤情報、セキュリティ侵害、プライバシー侵害、法令不適合など多岐にわたる。ISO/IEC 23894[6]等も参考にして、特定 (脅威・脆弱性・影響・関係者の洗い出し)、分析 (定性・定量評価、シナリオ分析、公平性・バイアスによる差別的影

[6] ISO/IEC 213894: 2023 Information technology — Artificial intelligence — Guidance on risk management

響有無の測定)、評価 (許容水準との比較と優先順位付け)、対応 (回避、軽減、受容、移転) を系統立てて行うことを推奨する。各ユースケースのリスク評価には責任者 (事業責任者またはモデルオーナー) を明確に指名する。CAIO は評価設計の標準化と最終承認を担う。権利・安全に重大な影響を及ぼす可能性のあるユースケースについては、5.8 で定める追加統制を適用し、より厳格な評価・承認・モニタリングを行う。

リスクマネジメントは以下の手順で運用する。

1) リスク対応方針、役割・責務、監督・監査計画を確立し、組織としてのリスク許容度を明文化する。
2) AI 利活用のユースケースを特定し、関与するステークホルダーと権利・安全への影響を把握する。
3) 性能、ロバスト性、公平性、説明可能性、セキュリティ、プライバシーに関する評価・テスト計画と指標を設計する。
4) リスク軽減策の実施、コントロールの運用、モニタリング、およびインシデント発生時にはその対応と是正措置を行い、継続的改善を推進する。
5) CISO、法務部門、リスクマネジメント部門と協働し、運用の現場に根差した実効性のある統制を維持する。AI に係るリスクレジスターは、全社的なリスクレジスターと統合し、他の財務・オペレーショナルリスクとの整合を図る。

5.4 セキュリティ・プライバシー

CISO・Data Protection Officer (以下、DPO)・法務部門と連携しつつ、AI 特有の脅威とプライバシーリスク (プロンプトインジェクション、モデル抽出、データ汚染等) が既存の情報セキュリティマネジメントと整合的に管理されるよう、全体方針と統制要件を定める。

セキュリティ・プライバシーは設計段階からの組込み (セキュア・バイ・デザイン、プライバシー・バイ・デザイン等) を原則として評価・監督する。具体的な設計・運用手法については「10. セキュリティ・プライバシー」を参照。

5.5 データガバナンス・品質保証

AI に利用するデータのガバナンスと品質保証、ライセンス適法性の監督は CAIO の重要な任務である。CDO、法務部門と連携し、データライフサイクルの標準化、品質メトリクスの設定、第三者素材のライセンス確認を徹底するようにする。モデル性能の向上と権利侵害の防止を両立させる観点が必要である。データ品質とライセンス適法性は、モデルの性能のみならず、公平性・説明可能性・透明性の前提となるため、データシート等の文書化を通じて一貫して管理する。詳細については「11. データガバナンス・品質保証」を参照。

5.6 調達・ベンダー管理

調達・ベンダー管理は方針・評価・監視の三層で運用する。CAIO は調達部門、法務部門、CISO、DPO 等と連携し、評価基準とベンダーロックインを含むベンダーリスク評価プロセスを設計するとともに、高リスク AI については採用可否の最終判断を担う。詳細な評価観点と契約条項、契約後監視の要件は「12. 調達・ベンダー管理」に、関連するチェックリストについては「18. テンプレート要点」を参照。

5.7 外部適合性評価・認証

認証戦略 (対象スキーム、スコープ、認証機関選定、取得時期)、相互承認・法域別要件の整合、監査対応計画を「外部認証ロードマップ」として策定・更新する。国内の認定動向 (例: ISMS-AC、JAB 等による AI マネジメントシステム認証機関認定) 及び海外の認定動向 (ANAB、IAS、RvA 等) を四半期ごとにレビューし、調達・市場投入計画に反映する。必要に応じ、ISMS 等既存認証との統合監査を図り、重複統制と審査負荷を最小化する。

5.8 権利・安全影響 AI に対する追加の内部統制

権利・安全影響 AI を利用するユースケースについては、CAIO および全社 AI ステアリングコミッティによる承認を必須とし、人間による監督・最終判断 (Human-in-the-loop) をワークフロー上で義務付ける。影響評価、公平性評価、差別緩和、児童性的虐待コンテンツ (CSAM) 対策、利用者通知、そして人間による救済経路の設計を標準化し定期的に見直す。法務部門、事業部門、顧客対応と協力し、具体的な救済手順と問い合わせ窓口を整備するとともに、苦情処理および是正・再発防止に関する KPI を設ける。

5.9 イノベーション・PoC 管理

ユースケース発掘から PoC (Proof of Concept) の進め方、本番移行の基準、撤退基準までを明確化する。ルールの範囲内で迅速に試行できる検証環境の運用ルールを定め、シャドー運用・限定リリース等の段階的導入を経ることを標準とする。事業部門、CTO、データサイエンスチームと協働し、価値仮説とリスク評価を同時に行い、スピードと安全性のバランスを取る。

5.10 教育・人材

AI の利活用に関わるロールを定義し、それに対するスキル標準を定め、教育プログラムを常設化し、評価・報酬への反映により、AI の活用を持続可能な組織能力として根付かせる。人事部門、事業部門と連携し、キャリアパスを整備し学習機会を継続的に提供する。また、取締役会・経営層向けの AI リテラシー研修やリスク・ガバナンスに関するブリーフィングを継続的に実施し、経営レベルでの理解と説明責任の遂行を支援する。詳細については「13. 教育・人材」を参照。

6. ガバナンスプロセスワークフロー

AI の導入・運用を安全に再現可能な形にするためには、ライフサイクルの各段階に「ゲート」を設け、必要な評価・文書作成・承認を明確にすることが不可欠である。これは、ISO/IEC 42001 の AI マネジメントシステム (以下、AIMS) や NIST AI RMF の “文書化・証跡” 要求と整合している。

1) アイデア段階では、事業価値の仮説と影響範囲の初期評価を行い、AIIA の準備に入る。
2) PoC に進む前に AIIA を実施し、権利・安全・公平性・プライバシーの影響を具体化し、必要な設計上の対策 (関係者への通知、人によるレビュー、救済手順、用途制限等) を定義する。
3) PoC では、性能・公平性・ロバスト性・セキュリティ・プライバシーの評価指標と試験計画を実行し、その結果をモデルカード・データシートに記録する。
4) 本番化承認ゲートでは、事業責任者、CAIO、CISO、法務部門・コンプライアンス部門が合議し、評価・対策が要件を満たすことを確認する。
5) 運用段階では、ダッシュボードで性能・公平性・インシデント・苦情をモニタリングし、異常検知から是正・再評価へとつなげる。
6) 所定のレビューと定期的な (例: 四半期ごと)内部監査で、方針遵守・記録の完全性・是正の実効性を点検する。ここでの要は「文書主義」であり、AIIA、モデルカード、データシート、承認記録、監査報告を整備し、意思決定の根拠を残すことで、運用の透明性と説明可能性を担保する。

本ワークフローは、「7. 組織設計と推奨体制」、「14. KPI・測定・ダッシュボード」、「15. 主要リスクと緩和策」、「16. 監査・モニタリング・報告」、「17. ユースケース別の適用具体例」の各章と対応しており、ここでは全体の骨格を示している。

7. 組織設計と推奨体制

本章では、AI ガバナンス体制の設計にあたり、(1) 独立した監督機能、(2) 二線防御、(3) 企業規模・リスクに応じた比例性、(4) 成熟度に応じた段階的拡張、の 4 原則を前提とする。

7.1 推奨体制

推奨する体制としては、CEO 直下の独立 C-suite として CAIO を設置することである。CAIO の独立性を確保することで、事業横断の意思決定や優先順位付け、監督機能を中立に行うことができる。併せて CAIO 直下に AI ガバナンス室（AI 推進チーム）を設置し、AI 戦略策定、標準化、運用監督・自己点検を中核機能として担わせる。さらに、全社 AI ステアリングコミッティを設置し、CAIO、CDO、CIO、CTO、CISO、法務部門・コンプライアンス部門、DPO、人事部門、主要事業部が参加する合議体を通じて、ユースケースの採否、ポリシー改定、事業の優先順位付けや投資配分、重大インシデント対応などを経営直結で迅速に決定する。

全社 AI ステアリングコミッティの開催頻度は少なくとも月 1 回を標準とし、(1) 新規ユースケースの審査、(2) 主要 KPI のレビュー、(3) インシデント・苦情の報告と是正計画、を固定アジェンダとすることが実務上望ましい。

なお全社 AI ステアリングコミッティは CAIO を議長とする合議体であり、原則として以下に関する最終決定権を持つ、または経営会議への答申機能を持つことが想定される。いずれの位置付けとするかは各社のガバナンス構造に応じて選択し、社内規程等で明確にしておくことが望ましい。

- ユースケース採否（AIIA・評価結果を踏まえた承認）
- 資源配分（人・予算・プラットフォーム）
- 重大インシデントの対応方針（全社的な意思決定）

また、AI ガバナンス室には具体的に以下のような機能を持たせる。

- AI ポリシー・ガイドラインの策定・改訂（定期的な改定および規制やリスクの変化に応じて）
- AIIA・モデルカード・データシート等テンプレートの整備・運用支援
- ユースケース審査の事務局（台帳管理・議案整理・記録管理）
- KPI・ダッシュボードの集約・レポーティング
- 自己点検・内部監査との連携窓口

これにより、分散的な取り組みを避けつつ、価値とリスクの両面を一元的に考慮した意思決定が実現する。この体制のメリットをまとめると以下のとおりとなる。

- 経営直結で優先順位と投資配分を迅速に決められる
- ガバナンスの標準化と例外管理が一貫する（部門ごとのばらつきが減る）
- セキュリティ・法務部門・プライバシーと事業部門の利害調整を独立の視点で行える

7.2 代替の体制案

高リスク AI が存在しないなど、AI ユースケースが限定的な場合や規制リスクが中程度以下の場合は、上記推奨体制の代替として、CDO に CAIO 機能を兼務させる体制が考えられる。この体制は、データ戦略との密接な統合が利点であり、データ品質や AI 利用に対するガバナンスとの連動が強化される。他方で、独立監督の確実性が低下し、法務・セキュリティとの横断調整の機動性が弱まる懸念がある。権利・安全への影響に関する優先事項がデータ課題の陰に隠れ、意思決定における重み付けが偏る可能性があるため、体制の選択にあたっては自社のリスクアペタイトと組織文化を踏まえた慎重な検討が必要である。

なお規制産業では、監督の独立性に関する期待水準が高く、監督機能とデータ利活用推進機能を分離しておくことが望ましいため、本代替案は原則非推奨とする。

7.3 各機能との連携

CAIO がすべてを直接管理するのではなく、機能ごとにリードする CxO と役割分担を明確にし、連携の強度に濃淡を付けることが重要である。ただし、社内各部門との連携の強度は検討すべき機能により異なる。

- **戦略策定:** 事業部門、CDO、CIO、CTO との連携が最も強く、事業価値と技術実現性、データ資源の観点を統合する。
- **ガバナンス・リスクマネジメント:** CISO、法務部門・コンプライアンス部門との連携が最も強く、セキュリティ・プライバシー・規制適合の観点を運用に織り込む。
- **イノベーション:** 事業部門、データサイエンス、CIO、CTO と強く連携し、実験と本番化の基準を共有する。
- **人材・文化:** 人事部門、事業部門、データサイエンスが中心となり、ロール定義、スキル標準、教育を進める。

なお挙げられたこれらの連携は固定的ではなく、ユースケース別に連携責任表 (RACI) を作成・四半期ごとに更新し、必要に応じて責任と承認フローの強度を見直すのが実務的である。

RACI の例: 高リスク AI ユースケースの承認においては、事業責任者を Responsible、CAIO を Accountable、CISO・法務・DPO を Consulted、IT 部門を Informed とする、など。

7.4 中規模企業における体制案

中規模企業として、ここでは従業員数数百名程度までを目安としつつ、AI ユースケース数が年間数十件未満の事業者を想定する。

CEO 直下に CAIO を置き、事業横断の承認・監督を一元化するが、専任が困難な場合は CTO または事業企画責任者が兼務し、停止権限を持つセーフティ担当 (情報管理責任者) を独立に配置して二

線防御[7]を確保する。専用のガバナンス室を設けることが難しい場合は、1〜3 名のバーチャル体制で、AIIA、ユースケース台帳、モデル登録・監視を運用する。

会議体は月次のレビュー（AI 責任者、セーフティ担当、業務責任者）と四半期の CEO 決裁に簡素化する。

法務部門・プライバシー部門は外部顧問と連携し、個人データ取り扱い時は必ずレビューを実施する。基盤・ツール選定は IT 責任者が担い、ベンダー利用時はセキュリティ・データ処理契約・越境データのチェックを必須化する。

当該担当部門がない場合は以下の部門・組織を代替として検討する。

- **IT 責任者が不在の場合:** 外部 MSP(Managed Service Provider)/クラウドベンダーのセキュリティチーム
- **法務部門が不在の場合:** 顧問弁護士・DPO 外部サービス
- **セーフティ担当が不在の場合:** 情報セキュリティ責任者または個人情報保護管理者

7.5 少人数のスタートアップ企業における体制案

少人数のスタートアップ企業として、ここでは従業員数数十名程度までのスタートアップを想定する。

二線防御を最小限確保しつつ意思決定を簡素化する。CAIO 機能は CEO または CTO が兼務し、出荷可否の最終責任を負う。プロダクト以外のメンバーをセーフティオーナーに指名し、リリース停止権限を付与する。必要に応じて外部の法務・セキュリティ専門家と連携する。

週次等定期的に開催するプロダクト会議に AI ガバナンスの固定アジェンダを設け、重大論点を確認する。固定アジェンダとしては、(1) 新規 AI 機能のリリース予定と AIIA の状況、(2) 直近のインシデント・苦情、(3) 規制・プラットフォームポリシーの変更点、を確認することが望ましい。

出荷前には、AIIA、モデルカード、データ使用台帳、ベンダー確認等に関するチェックリストを準備してリスクと緩和策を記録・承認する。

インシデント発生時は重大度の定義に基づき速やかに停止・通知し、事後レビューを実施する。

スタートアップ企業としての体制でガバナンスを開始したのち、例えば、高リスク AI の本番運用を開始する場合や、恒常的に大量の個人データを処理することになった場合、または AI 関連売上が一定比率（例: 全売上の 20%以上）を超える場合等には、上記中規模企業向け体制案への移行を検討することが望ましい。

---

[7] ここでいう二線防御とは、AI 活用推進側とは独立した立場から、権利・安全・コンプライアンスの観点で停止・是正を勧告できる機能を指す。

8. 協働体制とステークホルダー関与

8.1 社内協働

AI の運用は、特定の部署だけで完結するものではないため、社内の主要な役職者は、それぞれの役割を担いつつ CAIO と密に連携する必要がある。下表では、「7.3 各機能との連携」でも示した各機能との連携を、CAIO 視点での具体的な連携ポイントとして再整理する。CAIO と各役職者は、ここで示した連携ポイントについて、全社 AI ステアリングコミッティ（月次～四半期）および個別のワーキンググループを通じて協働し、それぞれの活動を相互で最適化できるようにする。特に高リスク AI に関連する重要ユースケースについては、AIIA 審査→ステアリングコミッティ承認→本番化レビューという共通プロセスに則るようにする。

役職例	連携ポイント
Chief Technology Officer (CTO)	◦ AI 技術の選定と実装方針について議論し、技術ロードマップを共有する。 ◦ プロダクトや生産ラインでの AI 活用内容について協議する。 ◦ インフラ・アーキテクチャ・運用面で、AI システムを既存基盤に統合する。 ◦ 技術規程の策定、技術人材育成・AI 関連 R&D 予算配分で連携する。
Chief Data Officer (CDO)	◦ AI プロジェクトに必要なデータ収集・品質管理・ガバナンスを協働で推進する。 ◦ データ活用の倫理やプライバシー保護のルール作成・運用で協力する。 ◦ データ戦略と AI 戦略の整合性を継続的に議論し、優先順位を更新する。
Chief Information Officer (CIO)	◦ AI 導入に伴う IT システム全体の最適化、コスト管理、IT ガバナンス、IT 規程策定で協働する。 ◦ AI 関連システムと基幹業務の連動設計、DX 推進全体計画を共有する。 ◦ AI 活用により増える IT 上の脅威・課題を識別し、対策を検討する。
Chief Information Security Officer (CISO)	◦ AI によるセキュリティ強化やリスク対応策を共同で策定する。 ◦ AI データ保護、モデル悪用防止（不正利用・情報漏えい等）に関する対策で協力する。 ◦ セキュリティ規程や対応策のアップデートについて定期的に意見交換する。

Chief Supply Chain Officer (CSCO)	◦ サプライチェーンにおける AI 導入（予測・最適化・リスクマネジメント等）の戦略立案で協働する。 ◦ サプライチェーンデータ活用や、外部パートナーとの AI・データ連携を議論する。 ◦ AI ツールの現場浸透に向けた教育・チェンジマネジメントを推進する。
Chief Human Resource Officer (CHRO)	◦ AI 人材の採用・育成・リスキリング、AI リテラシー向上施策を共同企画・実行する。 ◦ AI 普及による業務変革と組織設計、ジョブシフトやチェンジマネジメントについて連携して進める。 ◦ 倫理・透明性・公平性など、AI 導入が従業員に与える影響について意見交換する。
Chief Operating Officer (COO)	◦ 業務オペレーション改革における AI 活用の導入計画および KPI 設定を協働する。 ◦ 業務プロセスの再設計（To-Be 設計）と標準化を推進し、AI 活用を日常業務に組み込むための運用設計（役割分担・手順・責任分界）について合意する。 ◦ 本番運用後の品質・安定運用（SLA/SLI、監視・ログ、効果測定）と継続改善のサイクルを整備し、逸脱や不具合発生時のエスカレーション/是正の運用で連携する。
Chief Executive Officer (CEO)	◦ CAIO を任命し、そのミッション・権限・リソース（人員・予算）を明確化する。 ◦ 全社 AI 戦略と経営戦略・リスクアペタイトの整合について CAIO と協議し、最終判断を行う。 ◦ CAIO および全社 AI ステアリングコミッティから定例報告を受け、高リスク AI を利用した案件や重大インシデントの意思決定を行う。 ◦ AI ガバナンス体制や方針の見直しについて、取締役会・経営会議での議論を主宰し、意思決定をリードする。
Chief Financial Officer (CFO)	◦ AI 投資（CAPEX/OPEX）の予算配分、優先順位付けを、全社ポートフォリオの観点から協働する。 ◦ AI 投資の ROI/KPI、コスト、リスクのバランス評価（価値管理）を共同で設計・運用する。

	◦ 財務制約やリスクアペタイトを踏まえた投資判断ルール（承認ゲート、評価タイミング等）の設計・見直しで連携する。
事業責任者	◦ 重点ユースケースの発掘・選定、価値仮説の策定、KPI 設定を協働して行う。 ◦ PoC・評価の進め方、スケール判断（本番化/撤退）を、リスク評価と合わせて合意する。 ◦ 業務要件・運用要件（現場定着、品質、責任分界）を整理し、開発・運用・ガバナンス部門との調整を主導する。
法務/コンプライアンス	◦ AI 関連法令・ガイドラインの解釈を踏まえ、社内ルール（利用ルール、審査・承認プロセス等）の整備・更新で協働する。 ◦ 契約・調達（委託先管理、権利処理、監査可能性、説明責任に関わる条項等）の確認・是正を共同で行う。 ◦ 高リスク AI を利用した案件や重大インシデント発生時のエスカレーション、当局対応、社内調査・是正措置で連携する。
Data Protection Officer (DPO)	◦ PIA を含むデータ保護設計を監督し、プライバシー・バイ・デザインの実装で連携する。 ◦ 個人データの取扱い（目的、収集・利用の最小化、保存期間、本人対応等）の要件を整理し、AI 案件の要件定義・審査に反映する。 ◦ 委託先管理や越境移転等を含むデータ保護上のリスク評価・是正、および重大インシデント時の対応で連携する。

8.2 ステークホルダー関与

社外のステークホルダーとの関与も重要である。様々なステークホルダーとの関与について以下に整理する。

ステークホルダー	関与の内容
顧客・利用者	◦ 期待・苦情の収集（カスタマーサポート、ユーザー調査、NPS:ネット・プロモーター・スコア等） ◦ 通知・救済プロセス（苦情窓口、異議申立て、再審査）
影響を受けるコミュニティ・当事者	◦ 当事者団体との対話、アドバイザリーボード、フォーカスグループ ◦ 高リスクユースケースでは AIIA プロセスの一部として意見聴取を位置付ける

規制当局	◦ 相談・事前協議、ガイダンスの取り込み、重要インシデント時の報告ライン
パートナー/ベンダー	◦ 技術、契約、透明性のレビュー (モデルカード・データシートの共有、監査権行使)
外部監査人・認証機関	◦ ISO/IEC 42001 等の監査で是正状況を確認し、改善計画に反映

得られたフィードバックは AI インベントリ、リスクレジスター、KPI ダッシュボードに反映し、全社 AI ステアリングコミッティで改善策とセットでレビューする。高リスク AI については、通常よりも高頻度のステークホルダー関与 (当事者コミュニティとの定期対話、インパクト評価への参加等) を行う。

さらに、透明性レポート (利用目的、性能/限界、公平性評価結果、更新履歴等) の公開、FAQ やポータルサイトを通じた情報提供を行い、ステークホルダーとの対話の基盤とする。

広報・IR・法務・カスタマーサポート等と連携し、ステークホルダー関与方針とメッセージを統一する。個別の問い合わせや苦情対応は CS や窓口部門が担当し、AI ガバナンスに関わる論点について CAIO にエスカレーションするようにする。

9. コンプライアンス・規制対応

9.1 AI 事業者ガイドライン

AI 事業者ガイドラインは、人間中心、安全性、公平性、プライバシー保護、セキュリティ、透明性、説明責任といった原則を、具体的な業務プロセスや運用ルールに落とし込むための枠組みである。本マニュアル全体も、この AI 事業者ガイドラインの考え方を基本的な参照軸としている。

CAIO および AI ガバナンス室は、AI 事業者ガイドラインの各原則を、自社の AI ポリシー、AIIA テンプレート、モデルカード・データシート等の文書様式にマッピングし、日々の運用に組み込む責任を負う。具体的には、次のような取組が求められる。

- AI 事業者ガイドラインの各原則を、自社の「AI 原則」「行動規範」に対応付け、その関係を明文化する。
- AIIA、モデルカード、データシート等のテンプレートに、安全性、公平性、安全性、プライバシー保護、セキュリティ、説明責任に関する項目を組み込むことで、原則の実装状況がユースケースごとに可視化されるようにする。
- 権利・安全影響 AI については、AI 事業者ガイドラインの該当部分を必須チェック項目として AIIA に反映し、ステアリングコミッティでの審査時に確認する。

また、AI 事業者ガイドライン自体も改訂や補足資料の公表が行われる Living Document であるため、CAIO は法務部門・コンプライアンス部門と連携し、少なくとも年 1 回を目安にガイドラインと自社ポリシー・運用とのギャップ分析を実施することが望ましい。その結果は全社 AI ステアリングコミッティに報告し、必要な是正計画や改訂方針を決定する。

9.2 EU AI 法 [8]

EU AI 法は、人間中心の信頼できる人工知能 (AI) の導入を促進すること、AI システムの有害な影響に対して、健康、安全、民主主義、法の支配、環境保護等の基本的権利の高水準の保護を確保すること、イノベーションを支援することを目的として、2024 年 8 月 1 日に施行され、2026 年 8 月 2 日から本格適用が開始された。

リスクベースアプローチを採用し、 4 つのリスクレベルを設け、各々のリスクに応じた要件・規制を設定するとともに、広範なタスクを学習・実行可能で他の AI システムに統合可能な汎用 AI モデルに関する規律を規定している。

- **容認できないリスク**
  - サブリミナル技術、ソーシャルスコアリング、職場または教育機関での感情推測システム、公共空間における法執行目的でのリアルタイム遠隔生体認証システム 等

[8] 欧州連合日本政府代表部: EU AI 規則の概要

- 原則禁止

- **ハイリスク**
  - 機械、医療機器、生体認証、重要インフラ、教育、雇用、法執行、移民管理 等
  - プロバイダー、輸入者、販売業者、導入者それぞれに対して、リスク管理、データガバナンス、技術文書の作成、人的監視措置、適合性評価手続、ログ保存など厳格な規制
- **限定的なリスク**
  - 生成 AI、自然人とやり取りする AI、感情認識システム 等
  - AI により生成されたコンテンツである旨のマーキングや AI 使用の告知など限定的な透明性義務
- **最小限のリスク**
  - 上記以外
  - 自由に利用可能 (自主的な行動規範の推奨あり)

これらに加えて、汎用目的 AI (GPAI) については、特に汎用目的 AI モデル (GPAI model) の提供者に横断的な義務が課され、一定の基準によりシステミックリスクを有する GPAI モデルに該当する場合は追加義務の対象となる。

また、AI の「提供者 (provider)」「導入者 (deployer)」「輸入者 (importer)」「流通業者 (distributor)」等の主体ごとに要求事項と責務を定め、EU 域外の提供者については場合により認定代理人 (authorised representative) の関与が求められる。

CAIO および AI ガバナンス室は、EU 域内で提供・利用される AI システムについて、EU AI 法との整合性を確保するため、少なくとも次のような取組を行う。

- 自社の AI ユースケースを EU AI 法のリスクカテゴリと対応付けた AI インベントリを作成し、定期的に更新する。
- 各ユースケースについて、自社が「提供者」「導入者」「輸入者」「流通業者」のいずれに該当するかを整理し、それぞれに必要な義務 (技術文書、ログ、品質マネジメント、監視・報告等) を整理する。
- 高リスク用途や GPAI に該当する可能性があるシステムについては、ISO/IEC 42001、ISO/IEC 23894、NIST AI RMF 等の枠組みと整合的な内部プロセスを構築し、要求事項をマネジメントシステムの中で運用できるようにする。
- 技術文書、学習データの記録、ログ、AIIA、モデルカード等を、EU AI 法上の証跡要求を満たす形で体系的に保管・管理する。
- EU AI 法の施行状況や二次法 (実施法・委任法) の動向について、法務部門・コンプライアンス部門と連携してモニタリングし、必要に応じてポリシーや運用を改訂する。

なお、EU AI 法の適用にあたっては、GDPR (個人データ保護)、製品安全関連法制、オンラインプラットフォーム規制等の他の EU 法との関係も無視できない。CAIO は、これらを総合的なコンプラ

イアンス設計の一部として捉え、法務・コンプライアンス部門と協働しながら、個別案件ごとに統合的な判断が行われるよう体制を整える。

9.3 その他法域・関連法

AI の利活用は、専ら AI に関するルールにとどまらず、既存の多様な法令との整合が求められる。典型的には、以下のような領域が論点となる。

- **プライバシー・データ保護:** 個人情報保護法、GDPR 等
- **消費者保護:** 不当表示、誇大広告、説明義務、契約条件の公正性 等
- **労働法制:** 採用・評価・配置における差別防止、就業環境への影響 等
- **知的財産権:** 学習データの権利、生成物に関する権利関係 等
- **公正競争・独禁法:** アルゴリズムによるカルテル・差別的取扱い 等
- **その他の分野別規制:** 金融、医療、教育、交通などの業法・ガイドライン 等

また、国境をまたいだデータ移転、クラウド・AI サービスの利用規約、プラットフォームポリシーなど、いわゆるソフトローや契約上の制約も、実務上はコンプライアンスの一部として機能する。CAIO および AI ガバナンス室は、法務部門・コンプライアンス部門と連携し、次のような形で「レギュレーションマッピング」を行うことが望ましい。

- 主要なユースケースごとに適用される可能性のある法令・ガイドライン・プラットフォームポリシーを一覧化した規制マトリクスを作成し、定期的に更新する。
- 高リスク AI については、該当する法令・ガイドラインの要求事項を AIIA や運用ルールに直接組み込む。
- 新たな事業展開 (新市場・新法域への展開、規制産業への参入等) の検討段階で、CAIO と法務が共同で初期的なリスク・法令調査を行い、Go/No-Go や前提条件の検討材料とする。
- 主要な法令・ガイドラインの改正や重要判例等の情報を、法務部門から AI ガバナンス室に共有し、必要に応じてポリシーやテンプレートの改訂を行う。

これらにより、個別案件の対応に終始するのではなく、法域ごとの要求事項がユースケース設計・運用プロセスに組み込まれた状態を目指す。

9.4 適合性評価制度の基本枠組みとアクター

AI に関する国際標準や規制の多くは、適合性評価の枠組みを前提としている。適合性評価には、一般に次のようなレベルとアクターが存在する。

- **第一者評価 (自己適合宣言):** 事業者自らが、自社のシステムやマネジメントが要求事項に適合していることを評価・宣言する。
- **第二者評価:** 顧客やパートナー等が、取引関係の一環として事業者を評価する。

- **第三者評価・認証:** 認定機関に認定された第三者の認証機関が、国際標準等に基づき事業者を評価・認証する。
- **規制当局・指定機関による評価:** 特定の規制において、主管官庁や指定機関が評価・認証を行うケース。

このほか、適合性評価制度を支えるアクターとして、認証機関を認定する認定機関、認証スキームを運営するスキームオーナー、認証結果を活用する顧客・規制当局等が存在する。

CAIO および AI ガバナンス室は、自社の AI ガバナンスの成熟度や事業戦略を踏まえ、どの領域についてどのレベルの適合性評価を受けるかを設計する役割を担う。具体的には、次のような事項が含まれる。

- どの標準・スキーム (例: ISO/IEC 42001、ISO/IEC 27001、業界固有スキーム等) について第三者認証を取得するかの方針を策定する。
- 権利・安全影響 AI や高リスク用途について、自己適合宣言・内部監査・第三者評価・規制当局への届出等、どの組合せで適合性評価を行うかをユースケース分類と紐付けて設計する。
- 適切な認証機関を選定し、当該機関が当該スキームで有効な認定を保持していることを確認する。
- 認証の有効期限、サーベイランス監査予定、適用範囲、指摘事項・是正状況等を「外部認証・登録台帳」として管理する。
- 認証や外部評価で得られた指摘や推奨事項を、AI ポリシー、プロセス、KPI、監査計画の改善に反映し、全社 AI ステアリングコミッティで共有する。

これらの方針と仕組みは、「5.7 外部適合性評価・認証」および「16.4 外部監査・評価」と整合をとりつつ、全社的なリスクマネジメントと一体として設計することが望ましい。

9.5 ISO/IEC 42001 (AIMS) 認証の活用方針

ISO/IEC 42001 は、AIMS の国際標準であり、AI に固有のリスクやガバナンスをマネジメントシステムの枠組みで管理することを目的とする。ISO/IEC 42001 認証を取得することにより、AI ガバナンスの仕組みが一定水準で整備され、PDCA に基づき継続的に運用されていることを、第三者に対して示すことができる。他方で、認証は「取得すればコンプライアンスが完了する」という意味ではなく、AI ガバナンスの継続的改善を外部的に証明するための一つの手段であることに留意が必要である。

CAIO および AI ガバナンス室が ISO/IEC 42001 認証を活用するにあたっては、少なくとも以下の点を検討する。

- **スコープ設計:** AI に関する活動のうち、どの組織単位・プロセス・システムを AIMS の範囲とするかを、事業戦略とリスクプロファイルに基づき設計する。ISO/IEC 27001 等の既存マネジメントシステムとの統合可能性も考慮する。

- **ギャップ分析:** 現状の AI ガバナンス (ポリシー、AIIA、リスクマネジメント、教育・人材等) が ISO/IEC 42001 の要求事項とどの程度整合しているかを分析し、優先度の高いギャップから改善計画を立案する。
- **マネジメントシステムの統合:** ISMS や QMS 等を所管する部門と連携し、内部監査、マネジメントレビュー、是正・予防措置のプロセスを可能な範囲で統合することで、重複を減らし運用負荷を抑える。
- **認証機関の選定・対応:** 適切な認証機関を選定し、審査計画 (初回審査、サーベイランス、更新審査) を中長期の AI ガバナンス計画と整合させる。
- **他のフレームワークとの統合運用:** AI 事業者ガイドライン、NIST AI RMF、ISO/IEC 23894 等で示される原則・プロセス・管理策を、AIMS の中で統合的に運用し、重複や抜け漏れを減らす。
- **外部発信・ステークホルダー対応:** 認証取得の有無や範囲、運用状況を、顧客・利用者、規制当局、株主等に対する説明の材料として適切に活用しつつ、過度な「免罪符」として扱わないよう、メッセージを整理する。

ISO/IEC 42001 認証の取得を目指すか否かは、事業戦略、規制環境、取引先の要求等によって異なるが、いずれの場合も、AIMS の考え方を AI ガバナンス設計の基盤として活用することは有用である。

10.　セキュリティ・プライバシー

10.1 AI 特有の脅威

AI システム、とりわけ生成 AI や大規模言語モデルを活用するシステムには、従来の情報システムと共通する脅威に加え、AI 特有の脅威が存在する。これらは AI ライフサイクルの各段階で異なる形で顕在化するため、段階ごとに整理して把握することが重要である。以下に脅威の例を示すが、これらに限定されない。

1）データ収集・設計段階

- 不適切なデータ収集・同意取得の不備
- 権利侵害のおそれのあるデータの収集・利用
- バイアスを含むデータセットの設計不備 (特定属性の過小・過大代表等)

2）学習・チューニング段階

- データポイズニング (学習データへの悪意ある改ざん・混入)
- 学習環境への不正アクセス (モデル・データ・コードの窃取)
- ライセンス条件に反するデータ・モデルの利用

3）デプロイ・推論段階

- プロンプトインジェクション、敵対的入力 (プロンプトベースの攻撃)
- モデル抽出攻撃、モデル反転攻撃、メンバーシップ推論攻撃
- 出力を通じた機密情報・個人情報の漏えい
- ハルシネーション (誤情報・虚偽情報の生成) による誤案内
- 敵対的サンプルによる誤分類・誤判断の誘発

4）運用・保守・サプライチェーン段階

- 外部モデル・API・ライブラリ・ツールチェーンの脆弱性
- モデル更新・パラメータ変更に伴う性能劣化・安全性低下
- ベンダー・クラウド事業者のインシデントに伴う影響
- ログ・監視データの不適切な保存・管理による漏えい

CAIO は、CISO および情報セキュリティ部門と連携し、上記のような AI 特有の脅威を一覧化した「AI 脅威カタログ」を整備し、既存の情報セキュリティポリシー・リスク登録簿と統合する責任を負う。また、生成 AI を取り巻く脅威は急速に変化するため、外部の脅威情報・ベンダー情報・コミュニティからの知見を踏まえ、少なくとも年 1 回を目安にカタログと対策方針を見直すことが望ましい。

10.2 セキュリティ設計

AI システムのセキュリティは、「セキュア・バイ・デザイン」と「ゼロトラスト」を前提に、既存の ISMS と整合的に設計されるべきである。そのうえで、AI 特有の脅威に対応した追加の統制を上乗せする。

1） 基本的な設計原則

- **アイデンティティ・アクセス管理**

  ユーザー・サービス・API ごとに認証・認可を適切に設計し、最小権限の原則を徹底する。管理者権限やモデル管理権限へのアクセスは厳格に制限し、多要素認証を適用する。

- **環境分離**

  開発・検証・本番環境を論理的・物理的に分離し、本番データを開発・検証環境で安易に利用しない。

- **秘密情報の管理**

  API キー、トークン、認証情報等を安全なストアで管理し、コードやプロンプト、リポジトリに埋め込まない。

- **ログと監査証跡**

  入力・出力・モデルバージョン・設定変更・アクセス履歴等を適切にログ化し、改ざん防止と保存期間を定める。

- **構成管理と変更管理**

  モデルバージョン、プロンプトテンプレート、設定値の変更を記録し、レビューとロールバック手順を定める。

2） AI 特有の制御

- **入力検証・フィルタリング**

  プロンプトインジェクションや有害入力を検出・ブロックする仕組みを設ける。外部システムへのアクション（ツール実行等）を行う場合は、入力検証とサニタイズを徹底する。

- **出力フィルタリング・サニタイズ**

  個人情報・機密情報・有害コンテンツ等を含む出力を検出し、マスキング・ブロック・レビューに回す仕組みを設計する。

- **セーフティレイヤーの設計**

  大規模 AI モデルに設置される安全制御層であり、モデル単体では保証できない安全性・一貫性・ガバナンス適合性を補完する。単なる入力・出力のフィルタリングに留まらず、ルールベースフィルタや追加モデル等を用いて会話の文脈・ユーザー意図・ツール実行の影響まで解釈した上で、許可・ブロック・要レビューといった判断を体系的に適用し、危険な指示や望ましくない挙動を抑止することを目的とする。

- **RAG 等におけるアクセス制御**

検索拡張生成（RAG）や社内文書検索を組み合わせる場合、ユーザーの権限に応じて参照可能なデータを制御し、権限外情報が引き当てられないよう設計する。

3）データ分類と外部 AI サービスの利用

「11.データガバナンス・品質保証」で言及されたデータ分類（機密・社外秘・公開等）と連動し、分類ごとに外部 AI サービスへの送信可否やマスキング要件を定める。例えば、

- **機密情報・特定個人情報:** 原則として外部 AI API に送信しない、または匿名化・マスキング後に限り送信可とする。
- **社外秘情報:** 利用目的・契約条件・技術的対策を確認したうえで、限定的に送信可とする。
- **公開情報:** 必要な範囲で送信可とする。

CAIO は、CISO・CDO・法務と連携し、こうしたルールを「AI 利用ポリシー」「技術標準」として明文化する。

4）脅威モデリングとレッドチーミング

重要な AI システムについては、設計時に脅威モデリング（想定される攻撃経路・誤用シナリオの洗い出し）を行い、その結果を踏まえて対策を設計する。また、運用開始前および開始後定期的に、プロンプトインジェクションやモデル抽出攻撃等を想定したレッドチーミングを実施し、想定外の挙動や弱点を検証する。

レッドチーミングや脆弱性診断の結果は、リスク登録簿・KPI・改善計画に反映させ、全社 AI ステアリングコミッティおよび情報セキュリティ委員会等で共有することが望ましい。

10.3 インシデント対応

AI に関連するインシデントは、単なるサイバー攻撃や障害にとどまらず、AI 特有の挙動や出力に起因して、権利・安全・公平性・プライバシー等に重大な影響を与える事案を含む。本マニュアルにおいて「AI インシデント」とは、概ね次のような事案を含むものとする。

- モデルや AI サービスを経由した不正アクセス、情報漏えい、サービス妨害等
- AI の誤作動・ハルシネーション等により、利用者に重大な誤案内・損失を与えた事案
- AI による差別的・不公平な判断・出力が顕在化した事案
- 外部 AI サービスへの想定外の機密情報・個人情報の送信
- AI 生成コンテンツが CSAM 等の違法有害情報を含んでいた事案
- 規制当局への報告義務が生じる可能性のある事案　等

インシデント対応にあたっては、既存の CSIRT・情報セキュリティインシデント対応プロセスと統合しつつ、AI 特有の観点を組み込む。基本的なフローは以下のとおりである。

1）検知・通報

ログ監視・アラート、現場からの報告、顧客・利用者からの苦情・問い合わせ等により、AI インシデントの兆候を把握する。AI に関する事案であることが判明した場合は、CAIO に速やかに共有する。

2) 初動対応

追加被害を防ぐための暫定的な措置 (機能停止、アクセス遮断、設定変更等) を講じる。必要に応じて対象システムの一時停止を含む。

3) 影響評価

CISO・CAIO・DPO・関係部門 (事業部門、法務、顧客対応等) が連携し、以下を評価する。

- 影響範囲 (件数・対象者・地理的範囲等)
- 影響の種類 (生命・身体・財産・プライバシー・信用等)
- 法令違反・契約違反の可能性

4) 封じ込め・暫定是正

原因の切り分けが進むにつれて、必要な封じ込め措置 (設定変更、モデルロールバック、特定機能の無効化、アクセス制御強化等) を実施する。

5) 根本原因分析と恒久対策策定

技術的要因・組織的要因・プロセス上の要因を分析し、再発防止のための恒久対策 (設計変更、運用ルール改訂、教育強化等) を策定する。

6) 通知・説明

重大度に応じて、以下への通知・説明を検討し、法令・契約・社内ポリシーに基づき必要な対応を行う。

- 経営層・取締役会
- 顧客・利用者・パートナー
- 規制当局・監督機関

7) 記録・学習・改善

インシデント発生状況、対応内容、再発防止策を記録し、リスク登録簿・AI インベントリ・AIIA・モデルカード等に反映する。定期的なレビュー・監査 (「16. 監査・モニタリング・報告」) を通じて、対応プロセスの継続的な改善を図る。

重大度の基準については、特に人の生命・身体・自由・財産その他の基本的権利に重大な影響を与えうる事案を「高重大度」と定義し、CAIO・CISO・DPO および経営層への即時エスカレーションと、必要に応じた規制当局への報告を行う基準を事前に定めておくことが望ましい。

また、権利・安全影響 AI については、少なくとも年 1 回以上、AI インシデントを想定した机上演習・シミュレーションを実施し、対応体制と連携の実効性を検証する。

10.4 プライバシー

AI 利活用におけるプライバシー保護は、個人データ保護法制 (個人情報保護法、GDPR 等) との整合を前提としつつ、「プライバシー・バイ・デザイン/バイ・デフォルト」の考え方に基づいて設計されなければならない。

1) 設計原則と PIA/AIIA の連携

- 個人データ・プライバシーに関する影響をあらかじめ評価する PIA を、AIIA の一部または並行プロセスとして位置付ける。
- 権利・安全影響 AI については、PIA の実施を必須とし、その結果をユースケース承認プロセスおよび全社 AI ステアリングコミッティで確認する。
- データの収集目的の特定・利用目的の限定、データ最小化、保存期間の制限等の原則を、設計段階の要件として明文化する。

2) 学習データと推論データの区別と管理

- 学習データ (モデルの訓練・チューニングに利用するデータ) と推論データ (運用時にユーザーが入力するデータ) を区別して管理する。
- 推論データについては、デフォルトでモデル再学習には利用しない、または本人の明示的な同意がある場合に限り利用可能とするといった方針を定める。
- 学習・推論のいずれにおいても、可能な範囲で匿名化・仮名化・集約化を行い、識別可能性を低減する。

3) データ主体の権利への対応

- 個人データの学習・推論利用に関し、データ主体からの開示・訂正・削除・利用停止・異議申立て等の請求に応じるためのプロセスを定める。
- 特定のユースケースやモデルにおいて、個人データの削除要求があった場合に、どの範囲・方法で対応可能か (再学習、フィルタリング、マスキング等) を技術的・運用的に検討し、対応方針を整理する。
- 利用者に対しては、AI の利用目的、データの取り扱い、学習への利用有無等について、分かりやすい形で通知・説明を行う。

4) 越境移転・第三者提供・外部サービス利用

- クラウドサービスや外部 AI API へのデータ送信が、国外移転や第三者提供に該当する場合の法的要件 (同意、契約、適切な保護措置の確認等) を整理し、PIA および調達プロセスに反映する。
- ベンダー・クラウド事業者との契約において、学習データ・推論データの利用範囲、再利用の可否、保存期間、サブプロセッサの取扱い等を明確化する (「12. 調達・ベンダー管理」参照)。

5）CAIO と DPO の役割分担

- CAIO は、AI 利活用におけるプライバシー保護の方針や設計原則を統括し、AI システム全体のアーキテクチャ・運用に反映させる責任を負う。
- DPO (または個人情報保護管理者) は、法令要件・ガイドライン・PIA の妥当性を監督し、規制当局との窓口となる。
- 双方は、PIA・AIIA、AI インベントリ、AI インシデント対応、教育・研修等に関して密接に連携し、プライバシー保護と AI 利活用の両立を図る。

以上のように、セキュリティとプライバシーは AI ガバナンスの中核要素であり、CAIO には CISO・DPO・CDO・法務部門等と協働しながら、設計・運用・監査の各段階で一貫した統制を維持することが求められる。

11. データガバナンス・品質保証

11.1 データライフサイクル管理

AI に利用するデータについては、CDO が所管する全社のデータガバナンスの枠組みと整合させつつ、AI 特有の要件を加えたライフサイクル管理を行う必要がある。データの収集・取得、加工・統合、保管、利用・分析、共有、廃棄・アーカイブの各段階について標準を定め、誰がいつ何をどの根拠で実施したかを記録・維持することで、品質の一貫性と追跡可能性を確保する。

特に、AI 利用に関するデータライフサイクル管理では、以下の点が重要となる。

1) 対象データの明確化

- 学習用データ、評価用データ、推論時に入力されるデータ (ログを含む) 等、AI に利用されるデータの範囲を明確化する。
- ユースケース・モデルごとに、どのデータソースがどの目的で利用されているかを整理する。

2) データ分類と取り扱いルール

- 機密情報・特定個人情報・社外秘・公開情報等のデータ分類に基づき、収集方法・保管方法・外部 AI サービスへの送信可否・マスキング要件を定める。
- 特に機密情報・特定個人情報については、原則として外部 AI サービスへの送信を禁止するか、匿名化・仮名化・集約化等の処理後に限定する。

3) 記録とトレーサビリティ (データリネージ)

- 収集元、取得根拠 (同意・契約・法令等)、前処理内容、統合・加工の履歴、保管場所、アクセス権限、利用先 (どのモデル・ユースケースに利用されるか) 等を記録し、追跡できるようにする。
- これらの情報は、AI インベントリやデータシート (「18.3 データシート」参照) と連動させ、ユースケース単位で把握できるようにする。

4) 廃棄・アーカイブの基準

- 学習用データ、評価用データ、ログ等について、保存期間と廃棄・アーカイブの基準を明文化する。
- 保存期間経過後は、復元不能な形で削除するか、必要に応じて匿名化・集約化したうえでアーカイブする。

5) 役割と責任の整理

- データライフサイクル管理の枠組み自体は CDO が主導し、AI 利用に関する追加要件 (外部 AI サービス利用、AI 特有のリスク等) は CAIO が定義する。
- 各事業部門・システムオーナーは、定められた標準に従って運用し、AI ガバナンス室・データガバナンス組織に対して必要な情報を提供する。

このように、AI に利用するデータのライフサイクルを体系的に管理することで、監査や是正に耐えうる透明性の高い運用を実現するとともに、品質・プライバシー・ライセンス適法性に関するリスクを低減する。

11.2 品質保証

AI に利用するデータの品質は、モデルの性能だけでなく、公平性・説明可能性・信頼性に直結する。また、訓練・テストデータに含まれる第三者素材のライセンス適法性は、コンプライアンスおよび レピュテーションリスクの観点から極めて重要である。本節では、「データ品質」と「ライセンス適法性」の二つの観点から品質保証の枠組みを示す。

1) データ品質

データ品質の観点では、少なくとも次のような事項を検討・管理する。

代表性の確認

- 対象とする利用場面・利用者・対象集団を踏まえ、データが偏りなく収集されているか (特定グループの過小・過大代表がないか) を確認する。
- 権利・安全影響 AI や高リスク用途については、代表性の確認を特に厳格に行う。

欠損・ノイズの管理

- 欠損値、不整合データ、外れ値等の扱いについて方針を定め、処理内容を記録する。
- 生データと前処理済みデータの関係を把握し、品質への影響を評価する。

バイアス分析

- 性別、年齢、出自、障がいの有無等の属性に関するバイアスが結果に影響していないかを分析する。
- バイアスが検知された場合の緩和策 (再サンプリング、重み付け、属性別性能の制約等) を検討し、必要に応じて実施する。

更新頻度・データドリフト

- データの更新頻度や有効期間を定義し、古くなったデータが意思決定に悪影響を与えていないかを定期的に確認する。
- データ分布の変化 (データドリフト) をモニタリングし、一定の閾値を超えた場合には、再学習・再評価を検討するトリガーとする。

データ来歴の追跡 (プロベナンス)

- データがどこから来て、どのような加工・統合を経て現在の形になっているか (データリネージ) を把握できるようにする。
- 来歴情報はデータシートや AI インベントリに整理し、モデルの説明・監査に利用できるようにする。

評価用データ・ゴールデンデータの整備

- 主要なユースケースについては、性能評価・回帰テスト用のゴールデンデータセットを定義し、モデルバージョン間の性能比較や品質確認に用いる。
- 評価指標 (精度、再現率、誤検知率等) と閾値をあらかじめ定め、変更時には理由と影響範囲を記録する。

権利・安全影響 AI や高リスク用途では、上記のチェックの頻度・厳格度を引き上げ、再評価の周期も短く設定することが望ましい。

2) ライセンス適法性

ライセンス適法性の観点では、訓練・テストデータに含まれる第三者素材 (コンテンツ、データベース、ソフトウェア等) が、適切な権利処理と利用条件の範囲内で使用されているかを確認する。少なくとも以下の事項を検討する。

権利の種類と権利者の特定

- 著作権・著作隣接権、データベース権、肖像権、パブリシティ権等、関連する権利の種類を整理し、可能な範囲で権利者を特定する。

利用条件・ライセンスの確認

- オープンデータ、オープンソース、商用ライセンス、独自契約等、それぞれの利用条件 (改変の可否、商用利用の可否、再配布条件、表示義務等) を確認する。
- ウェブスクレイピング等により取得されたデータについては、サイトの利用規約・robots.txt・著作権表示等を踏まえた取扱いを検討する。

生成物の権利と責任分担

- 生成 AI による出力の権利帰属、第三者権利侵害時の責任分担、免責・補償の範囲等について、契約・利用規約上の整理を行う (詳細は「12. 調達・ベンダー管理」および「15.6 知的財産・ライセンス」を参照)。
- ユーザーに対する注意喚起や利用条件 (商用利用・再配布等) を明確にし、必要に応じてガイドラインや利用規約に反映する。

文書化と再利用条件の明確化

- 利用するデータセットごとに、権利関係・利用条件・再利用の可否を文書化し、データシート等に反映する。
- 再利用・二次利用・第三者提供の可否に関する条件を整理し、他プロジェクトや外部への展開時に参照できるようにする。

契約面での条項設計は「12. 調達・ベンダー管理」、リスクの整理は「15.6 知的財産・ライセンス」を参照しつつ、本節ではデータセット設計・利用の現場で参照する具体的なチェック項目として運用する。

11.3 透明性ドキュメント

モデルカードやデータシート等の透明性ドキュメントは、AI システムの前提・設計・性能・制約・リスク・利用条件等を整理し、社内外の関係者に対する説明責任を果たすための基盤文書である。形式的な帳票ではなく、意思決定と説明のための実務的な資料として活用されるべきものである。

1）位置付けと他文書との関係

- モデルカードは主としてモデルの目的・前提・性能・限界・リスク・モニタリング方法等を整理する文書であり、AIIA やリスク評価の結果を要約したものとして位置付ける。
- データシートは、学習・評価に用いられたデータセットの来歴・構成・品質・バイアス・ライセンス条件等を整理する文書であり、「11.1 データライフサイクル管理」・「11.2 品質保証」の結果を反映する。
- これらの文書は、AI インベントリから参照できるようにし、ユースケースごとの全体像を把握するための中核情報として扱う。

2）作成・レビューの責任主体

- モデルカードは、当該モデルのオーナー (開発チーム・プロダクトチーム) が作成し、CAIO 配下の AI ガバナンス室がガバナンス・リスク観点からレビューする。
- データシートは、CDO 配下のデータガバナンス組織が中心となって作成し、法務・DPO がライセンス・プライバシーの観点から確認する。
- 必要に応じて、事業部門、セキュリティ部門、品質保証部門等が内容をレビューする。

3）作成・更新のタイミング

- 少なくとも以下のタイミングで作成・更新することが望ましい。
  - PoC 等で一定の結論に至った時点 (本格導入の判断材料として)
  - 本番運用開始前 (承認プロセスの一部として)
  - 重大な変更 (データソースの変更、大規模な再学習、アルゴリズムの変更、利用目的の拡張等) が行われた場合
- 更新履歴とバージョンを記録し、いつの時点でどの前提に基づいて運用されていたかを後から確認できるようにする。

4）保管・アクセスとバージョン管理

- 透明性ドキュメントは、中央リポジトリ (ドキュメント管理システム等) で一元的に管理し、承認済みバージョンとドラフト版を区別して保管する。
- AI インベントリからリンクし、関係者が容易にアクセスできるようにする。
- 監査・インシデント対応・規制当局への説明等に備え、適切な期間保管する。

5）社外開示の方針

- 透明性ドキュメントの社外開示範囲・粒度は、競争上の秘匿と信頼確保のバランスを踏まえ、ユースケース別に方針を定める。例えば、

    - 権利・安全影響 AI については、利用者向けに要約版モデルカード (目的、主要な前提・限界、人間による監督の有無等) を公開することを検討する。
    - 社内向けツールや競争優位性の源泉となるモデルについては、社内限定にとどめるが、規制当局・監査対応に耐えうる詳細な内部版を保持する。
- 公開対象・公開方法 (ウェブサイト、透明性レポート、利用規約・FAQ 等) は、CAIO が広報・IR・法務と連携して決定する。

透明性ドキュメントは、AIIA、リスク登録簿、KPI・ダッシュボード等と並び、AI ガバナンスの「見える化」を支える重要な要素である。CAIO は、これらの文書が形式的な書類にとどまらず、ユースケースの承認・見直し・説明に実際に活用されるよう、プロセスと文化の両面から定着を図る。

12.　調達・ベンダー管理

12.1 調達方針

AI 関連の調達方針としては、単に技術的適合性や価格のみならず、AI ガバナンスの観点を含めた総合的な評価軸を明文化することが重要である。CAIO は調達部門・法務部門・情報セキュリティ部門等と連携し、次のような観点を含むベンダー評価方針を策定する。

- **技術的適合性:** 機能要件・性能要件・既存システムとの連携可能性・拡張性・サポート体制
- **ガバナンス/透明性:** モデルカード・データシート・AIIA 等のベンダーからの提供可否、利用目的・前提・制約の明確さ
- **セキュリティ:** ISMS 等の認証取得状況、脆弱性対応プロセス、AI 特有の脅威への対策方針
- **プライバシー・データ保護:** 個人データ取扱方針、学習データ・推論データの利用条件、PIA の実施状況
- **公平性・説明可能性:** バイアス評価・公平性評価・緩和策の有無、説明可能性に関する機能・文書化
- **コンプライアンス:** AI 事業者ガイドライン、法規制（EU AI 法等）への対応方針、業法・プラットフォームポリシーとの整合
- **運用実績・信頼性:** 稼働実績、主要顧客・ユースケース、インシデント履歴と是正状況
- **サステナビリティ:** 学習・推論に伴うエネルギー消費や環境負荷に関する情報開示と削減の取組
- **ベンダーロックインリスク:** データ/モデルのエクスポート可否、標準技術の採用状況、代替可能性

これらの評価基準は、可能な範囲で事前にベンダーに開示し、比較可能性と再現性を確保する。また、権利・安全影響 AI や高リスク用途に関する調達については、上記のうちガバナンス・セキュリティ・プライバシー・公平性・コンプライアンス・ロックインの観点の重み付けを高め、必要に応じて第三者認証や外部評価の取得を条件とすることが望ましい。

ベンダー評価方針は、全社 AI ステアリングコミッティの承認を得たうえで社内に周知し、案件ごとの例外適用や特別な条件を設定する際には、その理由とリスクを明文化する。

12.2 契約時の必須条項例

AI 関連サービスやモデルの調達に際しては、性能や価格だけでなく、AI 特有のリスクとガバナンス要件を踏まえた契約条項を設計する必要がある。以下は、典型的な必須条項の例であり、具体的な内容は案件ごとに法務部門・DPO 等と協議のうえ決定する。

1）性能・限界・リスクに関する条項

- 提供される AI システムの機能・性能指標（例：精度、応答時間、可用性等）と、その測定方法を文書化する。

- 「できること・できないこと」「想定利用シナリオ・非想定利用シナリオ」「既知のリスク・限界 (ハルシネーション、バイアス等)」を明示させる。
- モデルのアップデート方針 (頻度・通知方法・互換性・検証期間) を定める。

2) 用途制限・禁止用途

- 許容される利用目的と禁止される利用目的 (例: 特定の監視用途、差別的利用、違法コンテンツ生成等) を明記する。
- 権利・安全影響 AI に該当する用途への利用可否、追加条件 (人間の監督要件等) を定める。

3) データ取り扱いと権利 (学習・評価・ログ)

- 訓練/テストデータに含まれる第三者素材の権利 (著作権、データベース権等) とライセンス条件を確認し、適法な利用範囲を契約上明記させる。
- 顧客が提供するデータ (学習用データ・推論時の入力データ・ログ等) について、
  - 利用目的 (サービス提供のため、モデル改善のため 等)
  - 二次利用の有無 (他顧客向けモデル改善への利用の可否)
  - 保管期間・削除方法

  を明確にする。
- 生成物 (出力) の権利帰属、第三者権利侵害時の責任分担・補償条項 (インデムニティ) の範囲を定める。

4) 継続的改善と SLA

- モデル更新、脆弱性対応、バグ修正、性能劣化時の対応等に関する SLA (サービスレベル合意) を定める。
- 性能・品質に関する重要な KPI と閾値を合意し、閾値未達の場合の是正措置・料金調整・契約解除権等を規定する。
- 重大な仕様変更やモデル変更時には、事前通知と検証期間を設ける。

5) 監視・監査権限

- ログの共有範囲 (入力/出力/メタデータ等) と形式・保存期間を定める。
- 顧客側による評価・検証のためのインタフェース (API・ログ・テスト環境等) の提供を求める。
- 必要に応じて、オンサイト/リモート監査、第三者監査報告書 (SOC 報告書等) の提供、是正計画の提出を求める監査権を設定する。

6) セキュリティ要件

- 攻撃耐性評価 (ペネトレーションテスト、レッドチーミング等) への協力と、その実施条件を定める。
- 脆弱性報告・セキュリティインシデント報告のプロセス (報告期限・内容・連絡先等) を明文化する。

- 入力・出力ログの保持とアクセス統制、暗号化、バックアップ等の技術的・組織的対策の水準を定める。
- AI 特有の脅威 (プロンプトインジェクション、モデル抽出等) への基本的な対策方針を確認する。

7) プライバシー・データ保護要件

- データ最小化、目的限定、保存期間制限等の原則を契約上義務付ける。
- 国外移転・第三者提供・サブプロセッサ利用に関する方針と条件 (事前通知・承諾の要否等) を定める。
- データ主体の権利 (開示、訂正、削除、利用停止等) への対応責任と実務的な手順を整理する。
- PIA/DPIA (Data Protection Impact Assessment) が必要となる場合の協力義務を定める。

8) 環境・サステナビリティ

- 学習・推論に伴うエネルギー消費、データセンターの効率指標 (例: PUE)、CO2 排出原単位等に関する情報開示を求める。
- 環境負荷削減の目標や計画がある場合は、その概要の共有を求める。

9) 終了・移行条項 (ベンダーロックイン緩和)

- 契約終了時または重大な条件変更時に、モデル・データ・設定情報・ログ等の返還やエクスポートをどの範囲で行うかを定める。
- 他ベンダーまたは自社システムへの移行支援 (期間・費用・範囲) について合意する。
- ベンダーロックインを緩和するため、標準フォーマット・標準プロトコルの利用等を検討する。

これらを整理したうえで、契約交渉やレビューの場面で参照しやすいよう、以下のように内容をチェックリスト形式でまとめておくと実務上有用である。

契約の必須事項 (例) まとめ

- 性能・限界の文書化 (できること・できないこと/既知のリスク)
- 用途制限 (禁止用途の明記)
- 訓練・テストデータおよび顧客データの権利と取り扱い (ライセンス適法性、再利用条件)
- 生成物の権利帰属と第三者権利侵害時の責任分担
- 継続改善 (更新計画、脆弱性対応、SLA)
- 監視・監査 (ログ共有、評価指標の合意、監査権)
- セキュリティ要件 (攻撃耐性評価、脆弱性報告プロセス、レッドチーミング協力)
- プライバシー要件 (最小化、目的限定、削除権、データ移転・サブプロセッサ管理)
- 環境配慮 (推論・学習のエネルギー消費、PUE、CO2 排出原単位、削減計画の開示)

- 終了・移行 (ベンダーロックイン緩和、モデル・データ返還、移行支援)

なお、本節は一般的な例示であり、具体的な契約条項の策定・解釈は、各事業体の法務部門や外部の専門家による助言に基づき行う必要がある。

12.3 ベンダーリスク評価

AI 関連ベンダーのリスク評価は、調達時点の一度きりの審査ではなく、契約期間を通じて継続的に実施されるべきプロセスである。CAIO は調達部門・法務部門・情報セキュリティ部門・事業部門と連携し、少なくとも次のような観点からベンダーリスク評価の枠組みを設計する。

- **技術的リスク:** 技術成熟度、アーキテクチャの堅牢性、拡張性、依存技術 (特定クラウド・基盤モデル等) に関するリスク
- **運用リスク:** サービス継続性、サポート体制、SLA 遵守状況、障害対応能力
- **セキュリティ・プライバシーリスク:** セキュリティ認証 (例：ISO/IEC 27001)、プライバシー認証、AI 特有の脅威への対策、インシデント履歴と是正状況
- **コンプライアンスリスク:** AI 事業者ガイドラインや法規制 (EU AI 法等) への対応方針、関連法令・業法・プラットフォームポリシーとの整合、適合性評価・認証の有無 (「9.4 適合性評価制度の基本枠組みとアクター」、「9.5 ISO/IEC 42001 (AIMS) 認証の活用方針」参照)
- **倫理・公平性リスク:** バイアス評価と緩和策の有無、説明可能性の確保状況、人権への影響に関する配慮
- **レピュテーションリスク:** 社会的評価、過去の不祥事・訴訟・規制当局からの指摘等
- **戦略・依存リスク:** ベンダー財務基盤、事業継続性、特定ベンダーへの依存度 (ベンダーロックイン)、代替ベンダーの有無
- **環境・サステナビリティ:** エネルギー消費・排出に関する方針と実績、サステナビリティに関するコミットメント

これらの要素を定量・定性の両面から評価し、ユースケースの重要度・リスクレベルに応じて、ベンダーごとのリスクスコアとリスク区分 (例: 高・中・低) を設定する。権利・安全影響 AI、高リスク用途、大規模なデータ処理を伴う案件については、より詳細なデューデリジェンスと高い基準を適用することが望ましい。

ベンダーリスク評価は、調達前だけでなく、以下のようなタイミングで定期的に見直す。

- 契約更新時
- 重大なインシデント発生時
- 規制やガイドラインの大幅な変更時
- ベンダーの経営状況や事業方針に大きな変化があった場合

評価結果は、AI リスクレジスターや AI インベントリに反映し、ユースケースごとのリスク評価や優先順位付け（「5.3 リスクマネジメント」、「15. 主要リスクと緩和策」）に活用する。また、全社 AI ステアリングコミッティやリスクマネジメント委員会等で共有し、高リスクベンダーについては追加の統制（契約条件の強化、代替ベンダーの検討、監査頻度の増加等）を検討する。

このように、ベンダーリスク評価を調達・契約・運用の一連のプロセスに組み込み、AI ガバナンス全体の一部として継続的に運用することが重要である。

13.　教育・人材

13.1 教育プログラムの設計

AI ガバナンスを実践するためには、AI に関連するすべての人材に対して、職務に応じた教育とリスキリングを行うことが必要である。教育プログラムは、以下の要素を含む形で設計され、実行されるべきである。

1）教育の目的と成果

- **AI ガバナンス文化の浸透:** 全社員に AI ガバナンスに関する基本的な理解を提供し、企業全体で AI の倫理的・法的な利用を支える文化を醸成する。
- **専門的知識の向上:** 技術職や管理職に向けて、AI 技術や AI 関連法令、リスクマネジメントに関する専門知識を深め、プロジェクトや業務で実践的に活用できる能力を身につけさせる。
- **能力評価・スキル開発:** 教育プログラム後に能力評価（AI リテラシーの理解度、プロジェクトへの活用事例）を行い、フィードバックとともにリスキリング計画を立てる。

2）対象層ごとのカスタマイズ

- **経営層向け教育:** 経営層には、AI ガバナンス全体の方向性とそのビジネス価値に関する教育を実施する。教育内容には、AI 技術の進化とリスク、ガバナンスの重要性、法規制の最新動向（EU AI 法等）を中心に、リスク・コンプライアンス・社会的責任に関する事項を含める。
- **技術職向け教育:** データサイエンティストやエンジニアには、AI アルゴリズム、リスク評価方法（AIIA、バイアス評価等)、モデルトレーニング・検証・監視方法に関する技術的な教育を行う。実際のモデル開発・運用に役立つツールやベストプラクティスも提供する。
- **非技術職向け教育:** 営業、マーケティング、サポート部門等には、AI の基本的な利用方法、データの取り扱い方、倫理的観点や法的義務（データ保護等）を理解させ、AI システムに対するフィードバック能力を高める教育を行う。

3）スキルセットとトレーニング内容

- **AI ガバナンスの基本スキル（全社員向け):** AI の基本的な理解、AI がビジネスに与える影響、AI のリスク・法的な課題、プライバシー保護の重要性を教育する。
- **専門的知識スキル（専門職向け):** モデル評価・リスクマネジメント（AIIA、バイアス評価、再学習管理等)、法規制（GDPR、EU AI 法等）を実務レベルで習得させる。
- **AI 倫理スキル（全社員向け):** AI の倫理的問題（バイアス、公平性、説明可能性等)、透明性を確保するための実践的な方法論を学ばせる。
- **技術・ツールの使い方（エンジニア向け):** モデル開発・運用におけるツール（TensorFlow、PyTorch 等）や MLOps ツール、セキュリティ対策ツールを使用するスキルを強化する。

4）継続的な教育プラン

AI 技術や法規制は日々進化しているため、社員の教育は一度きりで終わらせることなく、定期的なアップデートやフォローアップが必須である。これにより、最新技術や法規制を実務に反映し、AI ガバナンス体制の成熟度を向上させる。

13.2 リスキリングとキャリアパス

AI の急速な進化に対応するため、リスキリングは単なるスキルの習得にとどまらず、社員がキャリア全体を通じて成長するためのプロセスである。特に、非技術職を含む社員にとっては、AI 関連業務における新たな役割を見出し、次世代のリーダーとして活躍する機会を提供することが重要である。

1）リスキリングの必要性と目的

AI ガバナンスが企業活動に不可欠となる中で、従業員全体に対して必要なスキルを再教育し、AI 技術の理解度と活用度を高める。これにより、企業は AI を活用した新たな価値創造を加速させることができる。AI リテラシーを向上させることで、業務の効率化や意思決定の質向上、さらには競争優位性の強化に繋がる。

2）AI 関連キャリアパスの設計

- **技術職向けキャリアパス**

  エンジニアやデータサイエンティスト向けに、AI の基礎から高度なアルゴリズム設計、リスクマネジメント、AI ガバナンス領域におけるリーダーシップを取れる役割までを明確に設計する。例えば、技術職から AI ガバナンスチームのリーダーへのキャリアパスを描き、専門性を高めるだけでなく、マネジメントスキルを学ぶ機会を提供する。

- **非技術職向けキャリアパス**

  営業、マーケティング、サポート部門等においても、AI リテラシーを高める教育と、AI 導入・活用のスキルを向上させるトレーニングを行い、特定の分野 (AI プロダクト営業、AI 顧客サポート等) に特化したキャリアパスを設計する。

- **全社向けキャリアパス**

  すべての従業員に対して、AI 関連の基礎スキルを習得した後は、各部門における AI 活用推進役として活躍する道を示し、キャリアパスを幅広く提供する。例えば、AI プロジェクトマネージャー、AI コンサルタント等、社内での役割変化を促進する。

3）実務に即したリスキリング

リスキリングの結果が業務に即時に反映されるよう、教育プログラムに実務案件を組み込む。実際の AI プロジェクトに関与しながら学べる機会を提供することで、学習した内容がすぐに実践で試され、スキルの定着を促進する。特に高リスク AI のプロジェクトがある場合は、それに関与させることで、リスクマネジメントやガバナンスの重要性を実務的に学ばせる。

4）リスキリングの効果測定とフィードバック

教育後には必ず効果測定を行い、定期的に知識テスト・ケーススタディ等を実施して、習得したスキルが実務にどう活かされているかを確認する。フィードバックを受けて、次回以降の教育プログラムの内容を調整し、常に最新の技術や法規制に対応したプログラムを提供し続ける。

13.3 評価とフィードバック

AI ガバナンスに関する教育プログラムが効果的であるかを測定するためには、学習の「定量的な成果」を評価することが重要である。教育内容が実務にどれほど活かされ、業務の効率化やリスク低減に繋がっているかを定期的に評価し、フィードバックを行う体制を整える。

1) 教育効果の測定

- **知識評価:** 教育後にオンラインテストやケーススタディを実施し、受講者が AI ガバナンスに関する基礎的な理解を習得したかを確認する。
- **実務適用度:** AI プロジェクトや業務で学んだ内容をどれだけ活かしているかを、実際のプロジェクト評価で確認する。
- **継続的なパフォーマンス指標:** 社員が教育内容を活用して業務改善を実現した場合、KPI (例えば、AI 関連プロジェクトの成功率、リスクマネジメントの精度等) で成果を測定し、教育後の成長を評価する。

2) フィードバックと改善

教育後には必ずフィードバックを実施し、教育内容に対する理解度や実務での活用状況を確認する。受講者からのフィードバックを収集し、次回の教育プログラムに反映させる。フィードバックの内容をもとに、教育プログラムを柔軟に改訂し、常に最新の技術や法規制、AI ガバナンスに対応した内容を提供できるようにする。

3) 教育効果の組織全体への反映

得られたフィードバックや成果をもとに、全社の AI 教育計画を見直し、組織全体の教育レベルの向上を目指す。定期的に全社 AI ステアリングコミッティに報告し、どのように教育が組織全体の AI ガバナンス強化に寄与しているかを示し、次年度の教育戦略に反映させる。

4) キャリアパスと教育成果のリンク

成果を上げた社員に対しては、キャリアアップや新たなプロジェクトリーダーとしての役割を提供し、リスキリングの成果を実務に直接結びつける。AI ガバナンスのリーダーシップポジションに昇進できるようなキャリアパスを用意し、教育とキャリア成長をリンクさせる。

14. KPI・測定・ダッシュボード

本章では、AI ガバナンスの実効性を継続的に把握し、改善していくための KPI 設計・測定・ダッシュボード運用の考え方を示す。本マニュアルでは、AI に関する指標を次の 5 つの領域に整理し、統合ダッシュボード上で一体としてモニタリングすることを推奨する。

- **事業価値:** 売上・収益への貢献、コスト削減効果、生産性向上、顧客満足度等
- **信頼性:** モデル性能・安定稼働状況、データ品質、バイアス・エラーの発生傾向等
- **ガバナンス:** AIIA 実施率、インシデント件数と是正状況、ポリシー遵守状況等
- **セキュリティ・プライバシー:** セキュリティ/プライバシーインシデント、診断・レッドチーミングの実施状況、PIA の実施率等
- **人材・文化:** 教育・研修の受講率、AI プロジェクトへの参画人員、社内からの改善提案・相談件数等

各社は自社の事業特性や規制環境に応じて、上記指標を具体化・追加・修正してよいが、少なくともこれら 5 つの観点をバランスよくカバーすることが望ましい。以降の節では、これらの領域ごとに KPI の設定と測定、ダッシュボードでの可視化・活用の考え方を示す。

14.1 KPI の設定と管理

AI ガバナンスを効果的に運営するためには、適切な KPI を設定し、その進捗を定期的に管理することが重要である。KPI は単にパフォーマンスの測定にとどまらず、改善点の特定や意思決定の支援に活用されるべきである。以下は、AI ガバナンスにおける KPI 設定の基本的な枠組みと管理方法である。

1）KPI の目的と活用方法

- **改善と調整の支援:** AI ガバナンスに関連する業務やプロジェクトのパフォーマンスを測定するだけでなく、問題点を発見し、早期に改善策を講じるために使用する。
- **データ駆動型の意思決定:** KPI は、経営層や CAIO がデータに基づいて迅速かつ正確な意思決定を行うための材料となる。AI リスク、データ品質、AI モデルの性能などの指標を見守り、ガバナンス戦略を柔軟に調整するためのツールとして機能する。

2）KPI 設定の基本原則

- **具体性:** 各 KPI は測定可能で、具体的にどの成果を追跡するかを明確にする。例えば、「AI モデルの精度を 90%以上に保つ」や「AI プロジェクトの期限遵守率を 95%に維持する」など、定量的な目標を設定する。
- **達成可能性:** KPI は現実的で達成可能な範囲で設定すること。過度に野心的な目標は、逆にモチベーションを低下させる可能性があるため、業界標準や過去の実績を参考に設定する。
- **関連性:** AI ガバナンスの目的や戦略と整合性が取れていること。例えば、リスクマネジメントの KPI はリスク評価の結果に基づき、AI の社会的影響の評価は倫理的な観点と関連付ける。

3) KPI の評価と改善サイクル

KPI の評価は、定期的に実施する。通常、四半期ごとに評価を行い、必要に応じて目標を修正する。評価結果は全社 AI ステアリングコミッティで報告し、KPI に基づいてリソース配分や改善施策を決定する。KPI の管理は単なる数値の追跡にとどまらず、進捗に応じた調整や改善活動を支援するためのものとして活用されるべきである。定期的なレビューを通じて、どの指標が効果的に機能しているかを確認し、必要に応じて変更を加える。

14.2 ダッシュボードと視覚化

ダッシュボードは、AI ガバナンスの状況を迅速かつ直感的に把握するための重要なツールである。AI 関連の KPI を視覚的に表示することで、担当者がすぐに問題点を把握し、必要な対応を取ることができる。また、ダッシュボードは、経営層、CAIO、AI ガバナンス室、リスクマネジメント部門等、各関係者が異なる視点で必要なデータを取り出せるように設計されるべきである。

1) ダッシュボードの目的と活用方法

- **情報共有:** 経営層、CAIO、各部門がリアルタイムで AI ガバナンスの状況を把握し、データに基づいて迅速に意思決定を行うために使用する。
- **早期警告システム:** 特定の KPI が閾値を下回った場合に、早期にアラートを発する仕組みを作り、問題を迅速に発見・対応できるようにする。
- **透明性の向上:** AI ガバナンスの状況を社内外の関係者に適切に伝えるため、主要な指標 (データ品質、AI リスク、バイアス評価等) を視覚化し、説明責任を果たす。

2) ダッシュボードの設計と情報の視覚化

- **シンプルで直感的な設計:** ダッシュボードは、担当者が一目で状況を把握できるようにシンプルかつ直感的に設計する。複雑なデータを簡潔に表現するため、グラフ、色分け、アイコン等を効果的に活用する。
- **インタラクティブな要素:** 関係者が詳細なデータを掘り下げて確認できるインタラクティブな要素を取り入れ、より深い分析が可能となるようにする。
- **リアルタイム更新:** KPI や関連データは、リアルタイムで更新される仕組みを組み込むことで、変化に即応できるようにする。

3) ダッシュボードの活用者と役割分担

- **経営層:** 全社的な AI ガバナンスの進捗状況を把握し、必要なリソース配分や調整を行う。経営層向けのダッシュボードには、AI プロジェクトの全体的な成果や主要なリスク指標を中心に表示する。
- **CAIO:** AI ガバナンス全体を管理する立場として、プロジェクトの進捗、リスク、品質管理に関連する KPI を監視し、問題があれば関係者に早急に伝達する。

- **AI ガバナンス室/リスクマネジメント部門:** リスク評価や AI プロジェクトの評価状況を詳細に監視し、ダッシュボードを用いて改善策を決定するためのデータを分析する。

4）ダッシュボードの更新と改善

ダッシュボードは、常に最新のデータを反映し、定期的に見直しを行う必要がある。新たに追加される AI プロジェクトや KPI の変更、リスクマネジメントの新たな指標等に応じて、ダッシュボード内容を更新する。ダッシュボード利用者からのフィードバックを収集し、視覚化方法やデータ提供方法を改善していく。

14.3 定期的なレビューと改善

KPI やダッシュボードは、ただ設定するだけではなく、定期的に評価・改善することで、その有効性を維持する必要がある。AI ガバナンスにおけるパフォーマンスの向上は、継続的な改善のサイクルを通じて実現される。以下のプロセスを通じて、KPI やダッシュボードの運用状況を見直し、必要な改善策を講じる。

1）定期的なレビュー

少なくとも四半期ごとに、全社 AI ステアリングコミッティや関係部門 (CAIO、CISO、事業部門等) で KPI やダッシュボードのレビューを実施する。レビューでは、各 KPI が設定通りに機能しているか、目標達成に向けて進捗があるか、問題点はないかを確認する。

2）フィードバックと改善

KPI やダッシュボードに関するフィードバックを関係者から受け取り、必要に応じて指標や評価方法を調整する。フィードバックをもとに、新たな課題や改善点を特定し、次回のレビューで解決策を提案・実行する。

3）改善アクションの策定

レビュー結果に基づき、必要な改善アクションを策定し、進捗状況をダッシュボードに反映させる。KPI の見直し、改善に向けたリソースの再配分、優先順位の調整等を行い、次回のレビュー時に進捗を報告する。

4）継続的なパフォーマンス向上

目標未達成の場合は、問題を特定し、根本原因に対処するための施策を講じる。定期的なレビューと改善活動を繰り返し実施することで、AI ガバナンスのパフォーマンスを高め、組織全体でのリスクマネジメントや倫理的 AI 利用を強化する。

15. 主要リスクと緩和策

本章では、CAIO の設置および AI ガバナンス運用に伴って特に生じやすいリスクと、その代表的な緩和策を整理する。ここで挙げるリスクは網羅的なカタログではなく、各社が自社のリスクレジスターを設計する際の出発点として位置づけるものである。CAIO は、事業特性や法域ごとの要件に応じてリスク項目を追加・詳細化し、定期的に見直すことが望ましい。

15.1 過度な中央集権化

CAIO 機能の強化は、全社的なガバナンスの整合性やリスクマネジメントの一元化に有効である一方、意思決定が過度に中央集権化されると、事業部門の自律性やスピードが損なわれるリスクがある。また、すべての判断が CAIO に集中すると、ボトルネック化や責任の過度集中が生じる可能性がある。

緩和策として、以下のような設計が考えられる。

- **分権的な運用との両立**
  - 全社 AI ステアリングコミッティにおける事業主導の意思決定を制度化し、ユースケースの提案・優先順位付けは事業側が主導しつつ、CAIO が横断的観点から調整・監督を行う。
  - 「分散的な実験」と「標準化された本番化ゲート」の組み合わせとし、PoC 段階までは各部門の裁量を尊重しつつ、本番化に際しては統一された承認・評価プロセスを適用する。
- **役割と権限の明確化**
  - AI 利活用に関するポリシーに、「最低限の共通統制」(AIIA 実施、透明性ドキュメント整備、インシデント報告等) と「事業裁量の範囲」(UI やワークフローの詳細設計等) を明記する。
  - RACI などを用いて、CAIO、事業責任者、CISO、法務部門・コンプライアンス部門の責任分担と決定権限を文書化する。
- **承認プロセスの機動性確保**
  - AI ユースケース承認や例外承認について、目標リードタイムを設定し、ダッシュボードでモニタリングする。
  - 低リスク・限定スコープの案件については、簡易フローや事後報告を認めるなど、リスクベースで承認プロセスを簡素化する。
- **例外管理の透明化**
  - 例外承認は理由・期間・是正計画を記録し、「例外台帳」として管理する。
  - 一定期間経過後に例外の妥当性を見直し、継続する場合は例外扱いではなく正式な運用となるようルール改訂を検討する。

これにより、CAIO の統合的な統治機能を維持しつつ、各事業の自律性とスピードを損なわないバランスを図る。

15.2 AI 過信による誤用・依存

AI の性能に過度の期待を寄せ、システムからの出力を無条件に信頼してしまう「オートメーション・バイアス」「権威バイアス」により、過度な依存や誤用が生じるリスクがある。とりわけ生成 AI は、もっともらしいが誤った内容 (ハルシネーション) を自信ありげに提示することがあるため、意思決定や対外説明に重大な影響を及ぼしうる。緩和策としては、以下のようなものがある。

- **Human-in-the-loop / Human-on-the-loop 設計**
  - 重要な判断 (採用・人事、与信・審査、医療・安全、契約・法的判断等) では、人間の最終承認を義務付ける。
  - AI 出力をそのまま採用するのではなく、「候補案」「参考情報」として提示し、担当者が根拠を確認したうえで意思決定するワークフローを標準とする。
- **不確実性・根拠の表示**
  - 可能な範囲で、AI の信頼度や不確実性、利用したデータソースをユーザーに提示する。
  - 重要文書や対外情報の起案に AI を用いる場合、出典・根拠の提示と、事実関係の検証ステップ (ファクトチェック) を UI やプロセスに組み込む。
- **利用ルールと教育**
  - 「AI は事実の最終的な出典ではない」「必ず一次情報・公式情報にあたる」ことを明記した利用ガイドラインを整備する。
  - 従業員向け研修で、AI の限界、バイアス、ハルシネーションの事例を共有し、過信・誤用を避ける実践的なチェックリストを提供する。
- **モニタリングと是正**
  - AI 出力の誤用に起因するインシデントやクレームを分類・記録し、どのプロセス・どの UI 設計が誤用を誘発しているかを分析する。
  - 分析結果に基づき、入力制約、出力フィルタリング、追加レビュー等の対策を継続的に見直す。

これらを通じて、AI を「万能な判断者」とみなさず、「人間の判断を補助するツール」として適切に位置付けることが重要である。

15.3 規制の不確実性

AI に関する規制は、国内外ともに整備が進む一方で、詳細要件や運用解釈が変化しやすく、適合性の判断に不確実性が伴う。複数法域で事業を展開する場合、同一の AI システムに対して異なる義務が課される可能性もある。緩和策として、以下のような枠組みを整備する。

- **規制ウォッチと責任者の明確化**

- 法務部門・コンプライアンス部門を中心に、CAIO と連携した「規制ウォッチ」体制を構築し、AI 関連法令・ガイドライン・標準の動向を定期的にレビューする。
- 規制動向の収集・分析と、社内ルールへの反映について、それぞれ責任者を明確化する。

- **ポートフォリオと要件のマッピング**
  - AI インベントリを基盤として、自社の AI ユースケース/モデルと、各法域の規制区分・義務とのマッピング表を作成する。
  - ハイリスク用途や権利・安全に影響するユースケースについては、より保守的な基準を適用する。
- **ポリシー・標準の機動的な更新プロセス**
  - ポリシー・標準・テンプレート (AIIA、モデルカード等) に、規制変更時の迅速な改訂プロセスを組み込み、四半期ごとの見直しを標準とする。
  - 改訂履歴と適用開始日、対象ユースケースを明確に記録し、監査可能な形で管理する。
- **外部専門家との連携**
  - 必要に応じて外部の法律事務所、認証機関、業界団体と連携し、自社の解釈や対応方針についてセカンドオピニオンを得る。
  - 規制当局のガイダンスや Q&A が公表された場合は、速やかに自社ポリシーに反映する。

これにより、不確実性を完全に排除することはできないものの、「変化に追随する能力」を組織的に高めることができる。

15.4 ベンダーロックイン

特定ベンダーの AI プラットフォームや API への過度な依存は、将来的なコスト増加、機能制約、契約条件の不利な変更、技術刷新の困難さにつながるリスクがある。緩和策として、以下を考慮する。

- **アーキテクチャ設計による依存度低減**
  - 標準化されたインタフェースやプロトコルを活用し、AI コンポーネントを疎結合に設計する。
  - モデルやベンダーの切り替えを想定した抽象化レイヤー (アダプタ) を用意し、「出口戦略」をアーキテクチャレベルで組み込む。
- **マルチベンダー戦略とポータビリティ**
  - 重要なユースケースでは、可能な範囲で代替ベンダーの存在を確認し、テスト利用やセカンドソースの確保を検討する。
  - データポータビリティ (エクスポート形式、メタデータ、ログ) の要件を契約条項に明記し、将来の移行に備える。
- **契約上の終了・移行条項**

- 「12. 調達・ベンダー管理」で示したとおり、終了・移行条項を設け、モデル・データの返還、移行支援、移行期間中のサポートレベル等を具体的に定義する。
- 価格改定や機能変更が行われた場合の協議プロセスや契約解除権を整理する。

- **内部能力の維持**
  - 評価・MLOps・ガバナンスの最低限の能力を社内に維持し、ベンダー任せにしない。
  - ベンダーから提供される評価レポートや監査結果を、内部で検証できるだけのスキルを確保する。

これらを通じて、ベンダーとの建設的なパートナーシップを維持しつつ、将来の選択肢を失わないようにする。

15.5 環境負荷

大規模モデルの学習や推論は、電力消費や CO2 排出量の増加を通じて環境負荷を高める可能性がある。他方で、業務プロセスの最適化や省エネルギー制御など、AI が環境負荷低減に寄与する場面も存在する。緩和策として、以下の観点が重要である。

- **測定と KPI 化**
  - 学習・推論に伴うエネルギー消費や CO2 排出の推計を行い、「14. KPI・測定・ダッシュボード」の KPI に組み込む。
  - 重要なユースケースについては、環境負荷と業務効率化効果の双方を指標化し、トレードオフを可視化する。
- **効率的なモデル・インフラ選択**
  - モデル蒸留、量子化、キャッシング、バッチ処理の最適化などにより、推論コストを削減する。
  - 必要以上に大きなモデルを常用するのではなく、用途に応じて軽量モデルやオンデマンド推論を組み合わせる。
- **サステナブルなインフラ活用**
  - データセンターやクラウド事業者のエネルギー効率指標や再生可能エネルギー利用状況を、調達判断の一要素とする。
  - 可能な範囲で、グリーン電力の利用や、環境認証を受けたインフラの採用を検討する。
- **方針と透明性**
  - AI 利用に伴う環境負荷に関する基本方針を策定し、ESG 戦略との整合性を図る。
  - 必要に応じて、主要な AI ユースケースに関する環境負荷評価や削減の取り組みを、サステナビリティ報告等で開示する。

これにより、AI 活用が中長期的なサステナビリティ目標と矛盾しないようにする。

15.6 知的財産・ライセンス

訓練・テストデータや生成物に係る知的財産・ライセンスには、無許諾の学習・テストデータ利用、生成物の著作権・商標侵害、ライセンス不遵守、オリジナル性不足、二次利用制限、機密漏えい等の法的・倫理リスクがある。責任帰属が不明確な場合、クレームや訴訟時の対応が困難になるほか、パブリシティ権・プライバシー侵害、契約違反のリスクも生じうる。緩和策として、以下のような統制が必要である。

- **データ・IP レビューの制度化**
  - 「11. データガバナンス・品質保証」のデータライフサイクル管理・品質保証と連動し、訓練・テストデータ、評価データごとにソース、ライセンス、利用範囲を記録した「データ台帳/AI インベントリ」を整備する。
  - 第三者素材 (画像・テキスト・コード等) のライセンス条件 (商用利用可否、改変可否、帰属表示義務、地域制限等) を確認し、記録するプロセスを標準化する。
- **生成物の利用ポリシー**
  - 生成物 (テキスト、画像、音声等) の利用可能範囲 (社内限定・対外利用可否・商用利用可否・二次利用可否) をユースケース別に定義し、利用規約や社内ガイドラインに反映する。
  - 帰属表示や免責表示が必要な場合は、テンプレートを用意し、利用者が迷わず適用できるようにする。
- **契約上の保証・補償フレームワーク**
  - 「12. 調達・ベンダー管理」に示す調達・契約条項と一貫させ、ベンダーに対して訓練データ・生成物に関する知的財産権・ライセンスの適法性を保証させる。
  - 権利侵害発生時の補償や責任分担 (インデムニティ) について、契約上明確に定める。
- **CAIO と法務の連携**
  - CAIO は、AI インベントリとデータ台帳に基づき、どのユースケースが高リスクな IP/ライセンス課題を内包しているかを俯瞰し、リスクベースでレビュー優先度を決定する。
  - 法務部門と連携し、著作権法・商標法・契約法等に関する解釈や実務対応を定期的にアップデートする。

これにより、知的財産・ライセンスリスクを事後的なトラブル対応ではなく、事前の設計と契約で制御することが可能となる。

15.7 誤情報・虚偽情報・来歴欠落によるブランド・意思決定リスク

生成 AI の普及により、誤情報・虚偽情報や来歴不明コンテンツが社内外に急増している。自社が発信する情報に誤りが含まれればブランド毀損や法的リスクにつながり、外部の偽情報を誤って引用

した場合には業務文書の誤記・誤引用、誤った意思決定を招くおそれがある。緩和策として、インバウンド (外部情報の取り込み) とアウトバウンド (自社発信) の両面から対策を講じる。

- **来歴の付与と検証 (アウトバウンド側)**
  - 重要な対外発信物については、C2PA (Coalition for Content Provenance and Authenticity) 等の標準 [9]も参考にしつつ、来歴情報 (作成者、作成日時、改訂履歴等) の付与を検討する。
  - 自社の公式情報については、真正性を確認できる「公式情報データベース」や署名付きの配信チャネルを整備し、偽情報との区別を明確にする。
- **AI 利用時の検証ゲート**
  - 重要文書 (プレスリリース、IR 資料、規程類、対顧客説明資料等) に AI を利用する場合、根拠提示・引用元の検証・二重承認などの検証ゲートをワークフローに組み込み、AI が生成した文面をそのまま採用しないことを徹底する。
  - AI が要約・翻訳した外部情報についても、一次情報との突合や出典確認を必須ステップとする。
- **インバウンド情報の評価**
  - 社外から流入する情報 (SNS 投稿、フォーラム、生成 AI コンテンツ等) に対しては、出典の信頼性、来歴の有無、情報発信者のプロファイルを確認するルールを定める。
  - 重要な意思決定や方針策定に用いる情報については、「複数ソースでの確認」「一次情報への遡及」を原則とする。
- **危機広報計画との連動**
  - 偽情報・虚偽情報が自社に関係して拡散した場合に備え、危機広報計画に、一次情報の即時提示、来歴証明の公開、是正情報の発信、訂正のタイムラインと責任所在の明示等を含めておく。
  - シミュレーションや演習を通じて、広報・法務・事業部門が連携して迅速に対応できる体制を整備する。
- **社内リテラシー向上**
  - 従業員向け研修で、誤情報・虚偽情報の典型事例と見抜き方、生成 AI コンテンツの取り扱い上の注意点を共有する。
  - AI を用いた情報検索・要約を行う場合にも、「最終的な事実確認は人間が行う」ことを繰り返し周知する。

これらの対策により、生成 AI 時代における情報の真正性とブランド保護を両立させることができる。

[9] Coalition for Content Provenance and Authenticity: Content Credentials : C2PA Technical Specification

16. 監査・モニタリング・報告

ガバナンスが機能し続けるためには、内部・外部監査とモニタリング、経営層・取締役会への報告を組み合わせたフィードバックループが欠かせない。日常的なモニタリングにより逸脱や異常を早期に検知し、内部監査および外部評価により統制の有効性を検証し、その結果を CAIO および全社 AI ステアリングコミッティを通じて経営に報告・反映することで、継続的な改善を実現する。

16.1 内部監査

内部監査は、AI ガバナンスに関する方針・プロセス・統制が設計どおりに運用され、想定されたリスク低減効果を発揮しているかを、独立した立場から評価する機能である。実施主体は原則として内部監査部門等の第 3 線とし、CAIO および AI ガバナンス室は監査対象となる統制の設計・運用に責任を負うが、自らを監査する立場には立たないよう役割を分離する。

内部監査の対象には、少なくとも次のような事項を含めることが望ましい。

- AI ポリシー・ガイドライン、AI 原則の実装状況
- AI インベントリの整備状況と網羅性
- AIIA・PIA、モデルカード、データシート等の実施率と内容の妥当性
- 高リスク/権利・安全影響 AI に対する追加統制（Human-in-the-loop、救済手続等）の実装状況
- セキュリティ・プライバシー・データガバナンスに関する統制の運用状況
- インシデント・苦情対応と是正措置の有効性
- 教育・研修の実施状況と記録

監査計画は年次で策定し、リスクベースで高リスクユースケースおよび権利・安全影響 AI を優先的に対象とする。低リスクユースケースや補助的な統制については、隔年監査やテーマ別監査として計画するなど、事業特性やリスクアペタイトに応じて柔軟に設計する。監査所見は是正計画に紐付け、締め切りと責任者を明記し、完了まで追跡する。

外部認証との整合のため、ISO/IEC 42001 等の第三者認証の年次サーベイランス監査・再認証審査のスケジュールを内部監査計画に統合する。サーベイランス指摘事項は内部監査の是正計画に統合管理し、認証機関への是正報告と同一の根拠・証跡で閉ループ化する。内部監査の結果および是正状況は、AI リスクレジスターや KPI・ダッシュボードの見直しにも反映する。

16.2 モニタリング

モニタリングは、日々の運用の中で AI システムおよび関連プロセスの状態を継続的に観測し、逸脱や異常の兆候を早期に検知して是正につなげる仕組みである。内部監査が一定期間ごとの独立した検証であるのに対し、モニタリングは運用現場および第 2 線（セキュリティ・コンプライアンス・AI ガバナンス室等）が中心となって継続的に実施する。

モニタリングは、運用ログ、性能・公平性指標、インシデント、ユーザー苦情を継続的に収集・分析する仕組みである。ダッシュボードで可視化し、主要な KPI (価値、信頼性、ガバナンス、セキュリティ・プライバシー、人材の 5 領域) に閾値を設定して異常検知を自動化する。モデルのドリフト検知やデータ品質の劣化に対する再学習・是正のワークフローを標準化する。

モニタリング対象には、技術的指標 (精度、再現率、誤検知率、公平性指標、モデルドリフト等) に加え、ユースケース本番化率、AIIA 実施率、例外承認件数、インシデント・苦情件数および対応リードタイム、Human-in-the-loop での差し戻し状況、アクセス権限の適正化率、研修受講率等を含めることが考えられる。指標ごとに閾値とエスカレーション基準を定め、閾値逸脱時には CAIO または AI ガバナンス室に通知し、必要に応じてユースケースの一時停止、設定変更、追加レビュー等の暫定措置を講じる。

モニタリング結果は、四半期ごとを目安に全社 AI ステアリングコミッティでレビューし、ポリシー改訂、統制の追加・簡素化、教育内容の見直し等、継続的な改善につなげる。

16.3 経営層・取締役会への報告

経営層や取締役会への報告は、AI 利活用がもたらす価値とリスクをバランスよく把握し、戦略とリスクアペタイトに沿った意思決定を行うための重要な手段である。CAIO は、全社 AI ステアリングコミッティでの議論結果とモニタリング・監査の結果を踏まえ、経営層および取締役会に対する定期報告をとりまとめる責任を負う。

経営層や取締役会への報告は、四半期ごとに、価値、信頼性、ガバナンス、セキュリティ・プライバシー、人材のそれぞれの KPI 総覧、主要ユースケースの進捗、重大インシデント・苦情と是正状況、規制対応・外部評価の状況、主要リスクと緩和策の実行状況、次期の重点テーマを含めた概要を提示する。特に重大インシデント、高リスクユースケースの導入・変更、規制上の重要な変化は報告の対象とし、意思決定に必要な代替案と影響評価を付す。報告は簡潔で比較可能性のある形式とし、長期トレンドを示すことが望ましい。

取締役会向けには、少なくとも年 1 回以上、AI ガバナンス全体の成熟度評価 (ISO/IEC 42001 等の認証取得状況や内部監査結果)、AI 戦略と全社戦略との整合性、権利・安全への影響リスクおよび社会的信頼・ブランドへの影響等について、より中長期的な観点からの報告・討議の機会を設ける。

例: 四半期スコアカード (前期比・年初来・目標比) で、価値・信頼性・ガバナンス・セキュリティ/プライバシー・人材の 5 領域を固定フォーマットで提出し、主要なユースケース・インシデント・規制対応・監査所見のサマリを添付する。

16.4 外部監査・評価

外部監査や第三者評価は、AI ガバナンスの客観性を確保し、顧客・規制当局・社会に対する信頼を高めるうえで有効である。CAIO は、「9. コンプライアンス・規制対応」で定める適合性評価・認証方針と整合させつつ、どの領域についてどのような外部評価を受けるかの方針を策定する。

外部監査・評価には、例えば以下のようなものが含まれる。

- ISO/IEC 42001 などのマネジメントシステム認証に係る第三者認証機関による審査
- 業界団体・規制当局によるガイドライン適合性評価や登録・認証制度
- 第三者専門機関によるモデルの公平性・ロバスト性評価、レッドチーミング等の技術評価
- 主要顧客によるベンダー監査・セキュリティレビュー

ISO/IEC 42001 などの認証取得を視野に、ギャップ分析と是正を計画する。また顧客や規制当局への説明責任を果たすため、評価結果と是正状況は適切に開示するようにする。外部監査・評価で指摘された事項は、内部監査の是正計画・AI リスクレジスター・KPI/ダッシュボードの見直しに統合し、ポリシー・標準・運用プロセスの改善につなげる。次回以降の外部評価では、前回指摘に対する是正状況を確認できるよう、証跡と説明をあらかじめ整理しておくことが望ましい。

17. ユースケース別の適用具体例

17.1 採用選考 AI

採用選考 AI は、候補者の雇用機会・キャリアに重大な影響を与えるため、高リスクシステムとして管理する。導入前に AIIA および PIA を実施し、職務関連性、影響を受けるグループ、AI の利用目的・範囲を明確化する。多様性・インクルージョンの観点から、影響評価には人事部門・ダイバーシティ&インクルージョン担当部門・法務部門を参加させる。

モデル提供者については、モデルカードおよび評価報告 (訓練データ来歴、既知の偏り・限界、更新計画等) の提供を必須とし、契約で監査権・変更通知・データ利用制限・責任分担を定める。評価指標としては、予測性能に加え、公平性指標を運用 KPI に組み込み、少なくとも四半期ごとにレビューし、逸脱時には是正措置を講じる。しきい値や評価基準の設定は職務要件および関連法令に整合させ、完全自動化を避け、人間による最終判断 (Human-in-the-loop) を必須とする。

候補者に対しては、AI 利用の有無・目的・影響範囲について事前通知と説明を行い、異議申立て窓口と再審査手順、合理的配慮および代替手段 (AI を利用しない選考経路等) を整備する。記録管理を徹底し、入力・出力・判断理由・人間による介入を保持し、保存期間・削除手順・機密管理を明確化する。個人情報の最小化、越境移転の適法性確認、セキュリティ対策、二次利用禁止を徹底する。

運用中は、否認率の偏りや苦情件数の増加など問題を示唆する指標をダッシュボードで監視し、異常があればシステムの一時停止や設定変更、ベンダーへの再評価要求、人間による監督強化等を行う。モデルの更新や用途拡大時には影響評価を再実施する。

CAIO および全社 AI ステアリングコミッティは、採用選考 AI を高リスクユースケースとして台帳管理し、承認・モニタリング・見直しの責任体制を明確にする。

17.2 カスタマーサポート生成 AI

カスタマーサポートに生成 AI を用いる場合、主な目的は応答速度・セルフサービス率・顧客満足度の向上等である一方、誤情報や機密情報漏えい、感情的配慮の不足などのリスクが存在する。AIIA において、対象とするチャネル (チャット、メール、FAQ、音声支援等)、自動応答と人間オペレーターの役割分担、AI が扱う情報の範囲を明確化する。

AI 利用の透明性通知を行い、利用者が AI 応答であることを認識できるようラベリングする。セーフガードとして、レッドチーミングや入力検証、センシティブテーマ回避、プロンプトインジェクション対策等を組み込み、誤情報・安全性リスクを抑制する。個人情報・機密情報は原則として最小限に限定し、ナレッジベースから機密文書・個人情報を分離する。ログと対応記録を一元的に保全し、再学習や改善のための利用範囲を利用規約・プライバシーポリシーに明示する。

運用においては、苦情・誤案内・エスカレーション率・解決率・顧客満足度 (CSAT/NPS 等) のデータをダッシュボードに集約し、閾値を超えた場合には会話フロー見直し・ナレッジ更新・モデル再

評価などの是正策を定義する。Human-in-the-loop として、一定条件（高額取引、解約・クレーム、健康・法的助言に関する相談等）では必ず人間オペレーターにエスカレーションするルールを設ける。

CAIO および AI ガバナンス室は、カスタマーサポート生成 AI に関する標準オペレーション手順（SOP）と教育プログラムを整備し、フロントラインのオペレーターが AI の限界・想定外挙動への対処法を理解していることを定期的に確認する。

17.3 医療支援 AI

医療支援 AI は患者の生命・健康に直結する高リスクユースケースであり、医療機器規制（欧州 MDR/米国 FDA 等）の適用可否を明確化し、該当する場合は規制を前提とした開発・運用を行う。意図された用途・リスククラスを定義し、臨床評価と技術文書の整備、CE マーキング/510(k) 等の承認取得、品質・リスクマネジメント（ISO 13485/ISO 14971）、医療機器ソフトウェア安全性（IEC 62304）への適合を徹底する。性能指標（感度・特異度・PPV/NPV・キャリブレーション等）を設定し、代表性のあるデータによる検証と、人口集団・疾病群別の公平性評価を行う。

運用フェーズでは、医療機関の安全管理体制・倫理委員会と連携し、ポストマーケット監視として性能指標・有害事象・準有害事象を継続的に収集・評価する。AI の提案は医師に対する診療支援にとどめ、診断・治療の最終判断は医師が行うことを原則とし、人間によるダブルチェックと手動介入の手順（AI 提案の無視・修正・再評価）を明文化する。モデル更新や適応再学習を行う場合は、バージョン管理と変更点の影響評価を行い、必要に応じて再承認・再認証プロセスを踏む。

患者への説明責任として、AI の利用有無・役割・限界・リスクについてわかりやすく説明し、インフォームドコンセントにおいて AI 利用を明示する。誤診・不具合が発生した場合の救済手段（再診、セカンドオピニオン、苦情窓口）と責任分担を明確にする。

CAIO は、医療部門・法務・コンプライアンスと協働し、医療支援 AI を高リスクシステムとして台帳管理し、規制遵守・リスクマネジメント・監査対応が AIMS 全体に整合するよう統括する。

17.4 重要インフラ運用 AI

重要インフラ運用 AI は、安全性・可用性・社会経済活動に直結する高リスクユースケースである。攻撃耐性の強化、冗長化、定期演習の実施、監督責任の明確化を徹底し、「失敗しても安全側に倒れる」設計を基本原則とする。用途・システム境界・安全制約を定義し、IT/OT の分離、ゼロトラストを含む多層防御、権限分掌を設計段階から組み込む。

運用においては、監査ログや SBOM (Software Bill of Materials) あるいは AI-BOM、署名付き更新を標準とし、更新前後の検証とロールバック手順を整備する。データ汚染・敵対的攻撃・センサ異常に備え、物理・工程モデルとの整合性チェック、フェイルセーフ設計、バックアップ運転系による

挙動確認を実施する。自律制御の範囲を限定し、危険状態や閾値超過時には安全側への停止や手動運転への切替えを自動で実行するようにする。

展開先に応じて必要な規制 (米国 NERC CIP、欧州 NIS2、EU AI 法等) への適合を確認し、所管当局との連携と報告ラインを明確化する。ログ保全・改ざん検知、最小化・暗号化・アクセス制御を徹底するとともに、異常時のインシデントレスポンスと BCP/災害対策訓練を統合的に計画する。人間による最終判断を必須とし、RACI 等を用いて権限分掌・責任境界を文書化する。

CAIO は、CISO・CIO・CTO・事業部門と協働し、重要インフラ運用 AI を権利・安全影響 AI として全社 AI ステアリングコミッティの承認対象とし、AIIA・レジリエンス訓練・レッドチーミング・外部監査等の結果を踏まえて継続的にリスク水準と投資配分を見直す。

18. テンプレート要点

18.1 AIIA テンプレート

AIIA には以下の各項目を含めるようにする。各項目に記入ガイダンスを付し、例示を示すことで、記載の質と一貫性を高める。更新時の版管理と、承認者・日付の記録を必須とする。

- **基本情報**
  - ユースケース名、ID、担当部門・責任者
  - 作成日・版数・ステータス (ドラフト/承認済み)
  - 関連するシステム・サービス名、対象地域・法域
- **ユースケース概要**
  - 目的・期待される価値 (事業価値・効率化・品質向上等)
  - 利用者・影響を受けるステークホルダー (顧客・従業員・第三者)
  - 利用場面 (業務プロセス、チャネル、時間帯 等)
  - 想定される利用・想定しない利用
- **規制・内部ルールの位置付け**
  - 当該ユースケースが該当しうる規制・ガイドライン (AI 事業者ガイドライン、EU AI 法、業法 等)
  - 社内ポリシー (AI ポリシー、セキュリティポリシー、プライバシーポリシー等) との関係
  - 社内リスク区分 (高リスク/中リスク/低リスク、権利・安全影響 AI 該当性)
- **データとモデルの概要**
  - 利用するデータ種別 (個人データ/機微情報/機密情報の有無、ログ、外部データ 等)
  - データの出所 (自社システム、ベンダー提供データ、オープンデータ 等) と法的根拠・ライセンス
  - モデル種別 (ルールベース/統計モデル/機械学習/生成 AI 等)、提供形態 (自社内製/ベンダー提供)
  - 関連するモデルカード・データシートへのリンク
- **リスク識別と評価**
  - リスクカテゴリごとの記入欄 (例: 安全性・健康、基本的人権・公平性、プライバシー、セキュリティ、コンプライアンス、レピュテーション、環境負荷 等)
  - 各リスクの発生可能性・影響度の評価 (定性的または定量的)
  - 想定シナリオ・最悪ケースの例示
- **対応策と残存リスク**
  - 各リスクに対応する統制・セーフガード (設計上の制約、アクセス制御、HITL、テスト・監視等)

- 実施状況 (予定/実施済み)、責任者、期限
- 残存リスクの評価と、受容可否に関する判断

- **運用・モニタリング設計**
  - Human-in-the-loop/Human-on-the-loop の要否と具体的な介入ポイント
  - モニタリング対象の指標 (性能、公平性、インシデント・苦情、環境負荷 等)
  - 閾値・アラート条件、エスカレーション先、暫定措置 (利用停止、設定変更 等)
- **ステークホルダー関与**
  - レビュー参加部門 (法務、コンプライアンス、セキュリティ、DPO、人事・労務、労使協議体 等)
  - 外部ステークホルダー (顧客団体、専門家、規制当局 等) への説明や意見聴取が必要な場合の方針
- **承認・改訂履歴**
  - 承認者 (CAIO、全社 AI ステアリングコミッティ、関係役員 等)
  - 承認日、適用開始日
  - 改訂履歴 (版数ごとの変更内容、理由、影響範囲)

18.2 モデルカード

モデルカードに含めるべき項目を以下に示す。可能であれば社内詳細版と社外公開用要約版を分け、開示粒度を調整する。運用者向けの注意事項や、再学習・更新のトリガー条件も明確化するようにする。

- **基本情報**
  - モデル名、バージョン、種別 (分類、回帰、生成、推薦 等)
  - モデルオーナー (組織・責任者)、連絡先
  - 作成日・最終更新日・ステータス (PoC/本番 等)
- **目的・利用シナリオ**
  - 想定された用途・適用範囲 (対象業務、対象利用者)
  - 想定されていない用途 (禁止される利用例、非推奨な環境)
  - ビジネス上の目的 (効率化、品質向上、リスク低減 等)
- **入力・出力仕様**
  - 入力データの形式・前処理要件 (必須項目、単位、言語 等)
  - 出力の形式 (ラベル、スコア、テキスト、画像 等) と意味
  - 依存する外部システム・サービス (API、データベース 等)
- **学習データの概要**
  - 主なデータソース (詳細はデータシート参照)

- データの取得期間・地域・対象集団
- サンプル数・ラベル付け方法・アノテーション品質

- **評価方法と結果**
  - 使用した評価指標 (精度、再現率、F1、AUC 等) とその定義
  - テストデータの条件 (代表性、分割方法、クロスバリデーション 等)
  - 属性別・セグメント別の性能 (公平性指標を含める場合はその概要)
- **既知の制約・リスク**
  - 想定していないデータ・環境・ユーザーにおける限界
  - 既知のバイアス、エラー傾向、ドリフトに敏感な条件
  - 利用時に注意すべき事項 (例: 単独での最終判断への利用禁止 等)
- **安全性・公平性・説明可能性**
  - セーフティ対策 (出力フィルタリング、安全側への丸め 等)
  - 公平性評価の結果概要と緩和策の有無
  - 説明可能性の手段 (特徴量重要度、事例ベース説明 等) の有無と利用方法
- **運用・モニタリング・再学習**
  - モニタリングする指標と閾値 (性能、ドリフト、公平性 等)
  - 再学習・チューニングのトリガー条件 (データ変更、規制変更、性能劣化 等)
  - モデル更新時の手順 (検証、ロールバック、ユーザーへの通知 等)
- **ユーザー向け注意事項**
  - 想定利用者 (専門家/一般ユーザー 等) と必要な前提知識
  - 入力時の留意点 (NG な入力、機密情報の扱い 等)
  - 出力の解釈方法・誤用を避けるためのガイドライン
- **バージョン管理・リンク**
  - バージョンごとの差分概要
  - 関連するデータシート・AIIA・運用ランブックへのリンク

18.3 データシート

データシートに含めるべき項目を以下に示す。データリネージの図示と、プライバシー保護措置 (匿名化、差分プライバシー等) も明記する。

- **基本情報**
  - データセット名、バージョン
  - データオーナー (組織・責任者)、連絡先
  - 作成日・最終更新日、利用開始日・終了予定日
- **目的と利用範囲**

- データセットの作成目的 (学習、テスト、評価、チューニング 等)
- 許容される利用範囲 (社内限定/対外提供可否/再利用可否)
- 関連するユースケース・モデル名

- **構成と範囲**
  - データ件数、特徴量・フィールドの概要
  - 対象期間・対象地域・対象集団 (年齢層、産業、製品ライン 等)
  - データ形式 (構造化/非構造化、テキスト/画像/音声 等)
- **出所・収集方法**
  - データソース (自社システム、顧客提供、公開データ、ベンダー提供 等)
  - 収集方法 (ログ収集、アンケート、クローリング 等) と頻度
  - データ収集に関する同意取得・告知内容の有無
- **品質・前処理**
  - 欠損・外れ値の状況と処理方針
  - 前処理・変換・正規化・匿名化等の実施内容
  - ラベリング・アノテーション方法と品質管理
- **代表性・バイアス**
  - 想定する利用場面・対象集団との整合性 (代表性の評価)
  - 特定属性の過小・過大代表の有無 (例: 性別、年齢、地域 等)
  - 実施したバイアス分析の概要と、検出されたバイアス・緩和策
- **プライバシー・法的根拠**
  - 含まれる個人データ・機微情報の有無と種類
  - 処理の法的根拠 (同意、契約、正当な利益等) とプライバシーポリシーとの整合
  - プライバシー保護措置 (匿名化、仮名化、差分プライバシー 等) の内容
- **ライセンス・権利・再利用条件**
  - 含まれる第三者データ・コンテンツのライセンス種別 (OSS ライセンス、商用ライセンス等)
  - 利用条件 (商用利用可否、再配布可否、帰属表示義務 等)
  - 当該データセットの再利用・二次利用条件 (社内他プロジェクト・外部提供 等)
- **セキュリティ・アクセス制御**
  - 保管場所 (データセンター/クラウド/オンプレミス 等) と暗号化の有無
  - アクセス権限 (閲覧・更新・エクスポート権限者) と認可プロセス
  - ログ取得・監査の方針
- **保存期間・廃棄**
  - 保存期間、見直しタイミング

- 廃棄・アーカイブ方法 (完全削除、匿名化して保存 等)
- 保存期間経過後の自動削除・アラートの有無

- **データリネージ・図示**
  - 主なソースシステムから当該データセットまでの処理経路
  - 派生データセット・下流モデルとの関係
  - リネージの図示 (図表) へのリンク

18.4 調達チェックリスト

調達チェックリストに含めるべき項目を以下に示す。ベンダーの透明性と協力義務を明確化し、契約後の運用に必要な情報提供を担保するようにする。チェックリストは、少なくとも「項目」「質問例」「期待される回答」「確認結果」「リスク評価」「対応方針」等の欄で構成するとよい。

- **基本情報・スコープ**
  - 対象サービス/製品名、バージョン
  - 提供形態 (クラウド／オンプレ／ハイブリッド 等)
  - 想定されるユースケースとリスク区分 (高/中/低)
- **技術的適合性**
  - 必要な機能・性能要件との適合状況
  - 既存システムとの連携・拡張性
  - ロードマップ・サポート体制
- **透明性ドキュメント**
  - モデルカード、データシート、AIIA 等の提供可否
  - 既知の制約・バイアス・リスクの開示
  - 更新・変更時の通知方法
- **データ取り扱い・プライバシー**
  - 訓練・テスト・ログデータの利用目的と保存期間
  - 顧客データの二次利用 (他顧客モデル改善への利用) の有無・条件
  - 個人データ・機微情報・機密情報の扱い、匿名化・暗号化・越境移転の有無
- **セキュリティ**
  - 認証取得状況 (ISO/IEC 27001 等)、セキュリティポリシー
  - 脆弱性管理・インシデント対応プロセス
  - 攻撃耐性評価 (ペネトレーションテスト、レッドチーミング等) への協力可否
- **コンプライアンス・規制適合**
  - AI 事業者ガイドライン、法規制 (EU AI 法等) への対応状況
  - 関連業法・プラットフォームポリシーとの整合性

- 適合性評価・認証の有無 (高リスク AI の場合 等)

- **公平性・説明可能性**
  - バイアス評価・公平性評価の実施状況と結果の提供可否
  - 説明可能性の機能 (出力の根拠を示す機構 等) の有無
  - ユーザー向け注意喚起・利用ガイドライン
- **SLA・継続性**
  - 可用性・応答時間・サポート対応時間等の SLA
  - 災害・障害時の復旧目標時間、BCP/DR 計画
  - サービス終了・機能変更時の通知・猶予期間
- **知的財産・ライセンス**
  - 訓練データ・モデル・生成物の権利帰属
  - 第三者権利侵害時の補償 (インデムニティ) の有無・範囲
  - 利用許諾範囲 (商用利用・再販・再配布 等)
- **監視・監査・協力義務**
  - ログ・レポートの提供範囲
  - 顧客による監査・セキュリティレビューへの対応可否
  - 規制当局・顧客監査への協力義務
- **ベンダーロックイン・終了・移行**
  - データ・モデル・設定情報のエクスポート可否と形式
  - 移行支援の内容・費用・期間
  - 価格改定・契約条件変更時の協議・解除権
- **環境・サステナビリティ**
  - エネルギー消費・CO2 排出等に関する情報開示
  - データセンターの環境認証・効率指標
  - 環境負荷削減の取組

19.　留意事項

本マニュアルは一般的な実務指針であり、特定法域における法的助言を構成するものではない。具体的な法令・規制適合性の判断は、自社の法務部門・コンプライアンス部門の監督のもとで行い、必要に応じて外部の専門家の助言と併用することを推奨する。

企業は自社のリスク許容度、規制環境、事業優先度、組織規模や業種の特性を踏まえて本マニュアルを補正し、適用範囲や優先順位を定めたうえで運用すべきである。CAIO は、AI ガバナンスに関する単一の責任点として本マニュアルの活用を主導するが、最終的な経営責任は取締役会および経営陣全体にあり、各ユースケースに関する事業判断は事業責任者が負う点に留意する必要がある。また既存の情報セキュリティ、内部統制、品質マネジメント等の枠組みと矛盾しないよう統合を図ることが望ましい。

技術の進化と規制の変化に応じて、方針・基準・プロセスを継続的に更新し、組織として成熟度を高めていくことが長期的な競争力の鍵である。本マニュアルについても固定的な「完成形」とみなすのではなく、CAIO と全社 AI ステアリングコミッティを中心に、内部監査やモニタリング、外部評価の結果を踏まえて定期的に見直しを行う「Living Document」として運用されるべきである。

20.　おわりに

本マニュアルは、CAIO という役割を通じて、企業が AI の価値創出と責任ある利活用を両立させるための、ひとつの標準的な道筋を示したものである。実務においては、ここで示した体制やプロセスをそのまま機械的に適用するのではなく、自社の課題や制約を踏まえ、関係者間の対話と試行錯誤を通じて最適な形に作り替えていくことが重要である。

CAIO の設置と AI ガバナンスの整備は、一度完了して終わるプロジェクトではなく、技術・規制・社会の変化に応じて見直され続ける長期的な取り組みである。本マニュアルが、そうした取り組みを社内で議論し、合意形成を進める際の共通言語として活用されることを期待する。

付録

A. 参考フレームワーク

国際標準や既存フレームワーク等は、CAIO が AI ガバナンスの全体設計を行う際の共通言語・参照軸となる。以下に示す各フレームワークや国際標準は、CAIO を含む AI ガバナンス構築の参考になる。なお本章は、各フレームワークの概要と特徴を整理したものであり、自社における適用方針や法令遵守の具体的な考え方は「9. コンプライアンス・規制対応」に委ねることとする。

A.1 AI 事業者ガイドライン (日本)

国内の AI 原則を実装に落とすための実務指針。非拘束のソフトローとして、リスクベースで AI の開発・提供・利用の全ライフサイクルを統合的に指針化。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、説明責任などの共通指針と、主体別 (AI 開発者・AI 提供者・AI 利用者) の留意点、広島 AI プロセスに沿う高度 AI 対応、アジャイル・ガバナンスを示す。

本マニュアル全体においても、AI 事業者ガイドラインを AI 原則実装の基本的な参照枠として用いている。主体別の整理は、CAIO が社内ロール分担を設計する際のベースになる。

A.2 NIST AI RMF (米国)

AI の便益を享受しつつリスクを管理するための、任意かつ産業横断的な枠組み。信頼できる AI の七つの特性 (妥当性・信頼性、安全性、セキュリティ・レジリエンス、説明可能性・解釈可能性、透明性・説明責任、プライバシー、公平性・バイアス管理) を定義し、リスクマネジメントを Govern、Map、Measure、Manage の 4 つの反復的機能に構造化する。社会技術的な TEVV (テスト・評価・検証・妥当性確認) を重視し、明確な役割・責任の設定、ライフサイクル全体のカバレッジ、継続的監視、インシデント対応、文書化、サプライチェーンにおけるデューデリジェンス、人による監督、特定の文脈に応じたプロファイルの策定・活用を推奨する。

CAIO は、NIST AI RMF の Govern 機能を中心に、Map/Measure/Manage の各機能を自社のリスクマネジメントプロセス (「5.3 リスクマネジメント」および「15.主要リスクと緩和策」) に統合することで、AI リスクマネジメントの全体設計を行うことができる。

A.3 EU AI 法 (欧州)

リスクベースの枠組みで AI の取扱いを規律する。人間の権利を脅かす利用 (例：社会的格付け、無差別の生体認証監視) は禁止される。重要分野における高リスク AI には、リスクマネジメント、データ品質、技術文書、サイバーセキュリティ、透明性、人間による監督、適合性評価及び CE マーキング等の厳格な要件の遵守が求められる。汎用・基盤モデルには追加的義務が課され、システミックリスクを伴うものには一層厳格な規制が適用される。

CAIO は、自社の AI ユースケースを EU AI 法のリスク区分と整合的にマッピングし、権利・安全影響 AI (「5.8 権利・安全影響 AI に対する追加の内部統制」) との対応関係を明確化することで、統一的な内部統制水準を設定できる。

A.4 OMB M-24-10 (米国)

米国における AI の統治、革新、リスクマネジメントを推進する政府全体方針を定める。各機関は CAIO を任命し、CFO 法対象機関は AI ガバナンス委員会を設置、AI 活用事例を年次公表し、インフラ・データ・セキュリティ・人材の障壁解消の戦略を策定する。安全または権利に影響する AI は、影響評価、実地試験、独立審査、人の監督、公平性確保、異議申立・オプトアウトの最低基準を満たすこととし、調達指針も示す。

公的機関向けの文書だが、民間事業者においても、(1) CAIO の任命、(2) 全社 AI ガバナンス委員会の設置、(3) 安全・権利に影響する AI に対する共通の最低基準の設定といった要素は参考にし得る。

A.5 ISO/IEC 42001

AI マネジメントシステム (AIMS) の認証フレームワークであり、AI に特化したマネジメントシステムを PDCA サイクルで運用する指針である。Plan では、スコープ、方針、目的・KPI、リスク評価、管理策を選定する。Do では、プロセスを運用し、教育を実施し、文書と記録を維持する。Check では、内部監査や管理レビューを通じて適合性を評価する。Act では、是正・予防措置と改善計画を進め、成熟度を高める。これにより、ガバナンスが形骸化せず、継続的に改善される。ISMS 等の既存マネジメントシステムと統合しやすい構造となっており、AI 特有の統制を既存のリスクマネジメントプロセスに組み込むことができる。

CAIO は、ISO/IEC 42001 を全社 AI ガバナンスの「運用 OS」として位置付け、「5. 役割・責務」で定義した役割・責務や「16. 監査・モニタリング・報告」のモニタリング・報告と一貫した AIMS を設計することが望ましい。

A.6 ISO/IEC 23894

AI のリスク同定、分析、評価、対応を体系化する。リスク特定では、脅威、脆弱性、影響範囲、関係者を構造的に洗い出す。リスク分析では、定性・定量評価、シナリオ分析を用いて影響を測る。リスク評価では、許容水準との比較により優先順位を設定する。リスク対応では、回避、軽減、受容、移転から適切な選択を行い、コストと効果のバランスを取る。AIIA のプロセス設計においても、本規格のリスク同定・分析・評価・処置の枠組みを参照することで、評価の抜け漏れを防ぎやすくなる。

CAIO は、ISO/IEC 23894 に基づくリスクマネジメントプロセスを全社 AI リスクの共通フォーマットとして採用し、リスクレジスターや優先順位付けの基準を標準化する役割を担うことが推奨される。

B. 初年度実践計画例

CAIO 就任直後の 1 年は、その後の AI ガバナンスの組織への定着度を大きく左右する期間であり、注意深く実践を進めていく必要がある。以下では、新たに CAIO を設置した企業が、どの順序で何に着手すべきか、時間軸に沿って具体的に説明する。

B.1 0–90 日: 組織立ち上げ・現状把握

最初の 90 日間では、「誰がどこまで決められるのか」という権限と意思決定の枠組みを明文化し、全社で共有された AI 原則と現状把握（インベントリ）を揃えることをゴールとする。これにより、以降の設計・調達・運用に共通の基準と、優先的に手を打つべきリスク・機会の見取り図が整う。

- CAIO 任命と権限・責務の RACI 定義、憲章（チャーター）策定
- AI 原則・禁止用途・ガードレールの草案作成と経営承認、暫定ガイドラインとしての社内告知
- 全社 AI ステアリングコミッティの設置、構成メンバー、開催頻度と議事運用方針の決定
- 現行ユースケース・モデル・データ・ベンダーの棚卸（AI インベントリ初版）と、簡易リスク区分（高・中・低）の付与
- 法規制マッピング（個人情報、著作権、業法、EU AI 法等）と、初期リスクレジスターの作成開始
- 社内周知計画（社内ポータル、FAQ、相談窓口）と「苦情・異議申立て」受け皿の設計
- 相談窓口や社内アンケートを通じた、クイックウィンとなり得るユースケース候補の収集

[成果物]

RACI/チャーター、AI 原則草案、インベントリ v1 (簡易リスク区分付き)、リスクレジスターv1、クイックウィンユースケース候補リスト

B.2 91–120 日: 標準化・パイロット設計

この期間は、AIIA/PIA やモデルカード等の標準テンプレートと、リスクに応じた承認ゲートを整え、「迷わず同じ基準で審査できる体制」を作ることが目的である。同時に、代表的なパイロット案件を選定し、評価軸（KPI/ROI と法令・安全・公平性）を設計する。

- AIIA/PIA テンプレート、モデルカード・データシート、人間関与（HITL）設計指針を確定
- モデルのリスク区分（高・中・低）と、区分に応じた承認ゲート（Go/No-Go）体制・必要な合議者の定義
- 説明・通知・救済の標準（ユーザー通知文、異議申立てフロー、再審査 SLA）の整備
- 調達条項の検討（性能・制限開示、データ取扱い、継続改善、監査権、環境配慮、SBOM 等）および、それらの「12. 調達・ベンダー管理」と「18.4 調達チェックリスト」との整合を確認
- パイロット候補選定と成功基準（KPI/ROI、法令適合、安全・公平性）設定
- パイロット案件ごとの責任者（RACI）とスケジュールの確定

[成果物]

標準テンプレート一式、承認プロセス図 (リスク区分ごとのゲート定義)、調達条項草案、パイロット計画 (案件別 RACI・KPI 付き)

B.3 121–180 日: パイロット実行・基盤整備

この期間では、選定したパイロット案件を通じて、AIIA/HITL/苦情対応 SLA 等のプロセスを一通り動かすとともに、最小限の MLOps/ガバナンス基盤を導入する。目的は、「スケール可能な運用の最小単位」を実証することである。

- MLOps/ガバナンス基盤の初期導入 (モデルレジストリ、実験追跡、データカタログ、アクセス制御など、パイロット運用に必要な範囲から)
- 公平性・ロバストネス評価、レッドチーミング、ドリフト監視の初期版ダッシュボード構築
- 監査ログの設計と取得開始、インシデント対応ランブック、キルスイッチ/ロールバック手順の整備・検証
- パイロット案件におけるユーザー通知・同意、HITL ワークフロー、苦情対応 SLA の運用テスト
- パイロット結果のレビュー (値・リスク双方) と、「本番化/改善/撤退」の判断基準の明確化

[成果物]

パイロット評価レポート (価値とリスクの両面)、モデルカード/データシート (パイロット分)、運用ランブック (インシデント対応・キルスイッチ手順を含む)、初期ダッシュボード

B.4 181–210 日: 人材育成・定着

この期間では、ロール定義・スキル標準・教育プログラムを結び付け、「誰がどの水準までできればよいのか」を明確にしたうえで、評価・認定に組み込むことを目標とする。「13. 教育・人材」で示した方針を、具体的な職種・研修・評価制度に落とし込むフェーズである。

- 役割別ロール・職種定義 (ML エンジニア、MLOps、リスク/コンプライアンス、PM、データ管理、プロダクトオーナー等)
- ロールごとの必要スキルと期待水準の整理 (スキルマップ) と、習熟度評価の方法設計
- 研修カリキュラム (公平性・セキュリティ・プライバシー・規制対応・説明可能性など) の、新任者向け導入/現場向け実践の両方を開始
- 社内認定制度の導入 (一定水準以上のスキルを可視化) と、評価制度と報酬制度との連動
- ベストプラクティス集・パターン (安全設計、通知テンプレ、HITL パターン、AIIA 記載例等) の整備・社内ポータルでの公開

[成果物]

研修計画・受講記録、ロール・スキルマップ、社内認定制度要領、ベストプラクティス集 (更新可能なナレッジベース)

B.5 211–240 日 (評価・是正・規格準備)

この期間は、初期運用を国際規格に照らして評価し、ギャップを是正することで、継続的改善の PDCA を回し始める段階である。ISO/IEC 42001 や ISO/IEC 23894 を、認証取得の有無にかかわらず「チェックリスト」として活用する。

- ISO/IEC 42001 のスコープ確定とギャップ分析、是正計画の策定 (認証取得を当面予定しない場合でも、要求事項を内部基準として参照)
- ISO/IEC 23894 に基づくリスク評価プロセスの文書化と代表的ユースケースへの適用
- KPI の運用開始 (AIIA 完了率、苦情対応 SLA 遵守率、モデル承認リードタイム、公平性指標、インシデント件数等)
- 内部監査 (簡易版でも可) の定期化と、是正・予防措置 (CAPA) と変更管理の運用開始
- 評価結果を AI ステリングコミッティでレビューし、優先度の高い是正テーマを決定

[成果物]

ギャップ分析報告、是正計画、KPI ダッシュボード初版、内部監査計画と CAPA 記録

B.6 241–270 日: 拡大・監査・演習

この期間は、審査済みユースケースを段階的に拡大しつつ、ベンダー監査やレッドチーム演習を通じて、外部依存とインシデント対応力を確認する。目的は、「スケールさせても破綻しない運用」を実証することである。

- 承認ゲートを通過したユースケースの段階的な展開計画を策定実行する。特に高リスクユースケースは、シャドー運用→限定本番→本格本番のステップを明示する
- ベンダー監査・SLA レビューの実施、ソフトウェア等のセキュアな更新 (署名・SBOM) とサプライチェーンリスクマネジメントプロセスの検証
- ゼロトラスト/セグメンテーションの適用確認、プロンプト/データ汚染対策のテスト
- 机上演習・レッドチーム演習によるインシデント対応訓練、当局・内部報告ラインの実地確認
- 監査・演習で得られた指摘事項を CAPA (Corrective Action Preventive Action) に統合し、翌年度以降も継続実施するための年次計画案を作成

[成果物]

展開計画と進捗記録、ベンダー監査レポート、演習報告と是正措置記録、サプライチェーン管理・ゼロトラスト適用状況のサマリ

B.7 271–360 日: 認証準備・次年度計画

最終四半期は、初年度の成果とリスク対応状況を総括し、必要に応じて外部認証の準備を進めるとともに、次年度の投資と体制を経営と合意するフェーズである。初年度で構築した仕組みを「一過性のプロジェクト」から「継続的なマネジメントシステム」に移行させる。

- 外部認証準備 (ISO/IEC 42001 等) のための文書一式、記録、証跡整備 (認証を取得しない場合でも、外部評価に耐えうるレベルまでドキュメンテーションを整理)
- マネジメントレビューの実施と、残存リスクの経営承認プロセスの確立
- 透明性レポート (利用目的、性能/限界、公平性評価、更新履歴、インシデント概要等) の社外公開方針決定
- ROI とリスクの総括、改善ロードマップ (初年度に積み残した課題、2 年目の重点領域、必要な予算・人員・外部パートナー) 策定
- 不要になったモデルの廃止、改善対象モデルの整理と対応計画策定 (モデル棚卸 v2 として反映)
- 継続教育と再認定スケジュールの更新 (年次・ロール別に必要な教育時間・内容を定義)

[成果物]

認証準備パック (もしくは外部評価用ドキュメント一式)、マネジメントレビュー記録と残存リスク承認文書、透明性レポート案、年次ロードマップ (投資・体制計画)、モデル棚卸 v2、継続教育・再認定計画