

学校法人沖縄科学技術大学院大学学園

基本方針、ルールおよび手続き（PRP）

第 44 章 - 人工知能（AI）の責任ある利用

権限：理事長

ポリシーオーナー：最高情報責任者（CIO）

バージョン：V1.0

施行日：2026 年 11 月 1 日

44.1 方針

本ポリシーは、学術の自由を尊重し、正当な探究、研究手法、または学術的判断を制限することなく、本学の使命を支援するために、人工知能（AI）および生成 AI（GenAI）の責任ある利用を可能にするものです。学校法人沖縄科学技術大学院大学学園および沖縄科学技術大学院大学（以下、総称して「本学」）は、AI を科学的発見、教育およびイノベーションを推進する革新的な技術と認識しています。本学は、責任ある実験、開かれた科学的探究、ならびに透明性が高く再現可能な研究実践の推進を支援します。

本学は、AI および GenAI 技術の安全かつ確実に責任ある利用に尽力し、AI が人間の判断、説明責任および知的責任に取って代わるのではなく、それらを強化することを確保します。

本ポリシーは、本学における AI 利用に関する原則、規則およびガバナンスを定めるものであり、本学の情報セキュリティ、学術的誠実性、研究倫理、データ保護および記録管理に関する枠組みと整合しており、組織、学術および研究の各文脈にわたる一貫性を確保することを目的としています。

本ポリシーは、AI のライフサイクル全体（計画、調達、開発、導入、運用、監視および廃棄）において、透明性、公平性、安全性、科学的誠実性、人による監督および説明責任を促進するものです。

本ポリシーは、本学の活動の一環として AI システムまたは AI 生成コンテンツを使用、開発、調達または管理する、教員、研究者、学生、職員、契約業者および来訪者を含むすべての本学構成員に適用されます。本ポリシーは、学術的判断や研究成果を規制するものではなく、機関のシステム、サービスおよび管理プロセスを含め、本学の活動を支援する AI ツールおよびシステムの責任ある利用を規定するものです。

AI に関する活動は、最高情報責任者（CIO）の権限の下で統括されます。CIO は AI 統括責任者（CAIO）を兼任し、CAIO の職務を遂行する代理人を正式に指名することができます。AI のガバナンスおよび活動は、最高情報セキュリティ責任者（CISO）および AI ガバナンス委員会によって支援されます。同委員会には、教員、研究部門、研究科、事務部門、技術部門および OIST 研究者コミュニティ（ORC）からの代表者が参加しており、学内全体にわたる広範な監督を確保するとともに、学術、研究、事務、技術の各観点からの実践的な意見を取り入れることを目的として、部門横断的な構成となっています。

44.2 一般的な考慮事項

44.2.1 既存の方針との整合性

本ポリシーは、本学の情報技術およびセキュリティに関する枠組み [PRP17]、研究科の「学術的誠実性に関するポリシー」および「生成 AI の利用に関するポリシー」、ならびに関連するすべての PRP 章の一部を構成するものです。

矛盾が生じた場合は、より厳格な規定が適用されます。

本学は、本ポリシーの解釈および日常的な適用を支援するため、付随するガイダンス、事例、運用指針を公表する場合があります。これらの資料は、本ポリシーに優先するものではなく、本ポリシーと整合性を保たなければなりません。

44.2.2 目的および適用範囲

本学は、人工知能（AI）および生成 AI（GenAI）技術が、業務の効率化、新たな知見の発見およびイノベーションの向上に寄与する可能性を認識する一方で、これらの技術に内在するリスクバイアス、誤情報、プライバシー、データの機密性および機微性、著作権侵害のリスク、モデルドリフトならびに倫理に反する利用を含む一についても認識しています。

本ポリシーは、事務、教育、研究の各機能における AI の利用を規定し、AI のライフサイクル全体を通じて、リスクに見合った管理措置が適用されることを確保するものです。

44.2.3 定義

本ポリシーにおいて：

人工知能（AI）とは、入力情報から、物理的または仮想的な環境に影響を与える可能性のある予測、コンテンツ、推奨事項、分類、または決定といった出力を生成する方法を推論する、機械ベースのシステムを意味します。

生成 AI（GenAI）とは、プロンプト、文脈または学習に応じて、テキスト、コード、画像、音声、動画、要約または合成データなどの新しいコンテンツを生成する AI システムを意味します。

外部 AI ツールとは、第三者によって提供され、本学が管理する環境内で完全に運用されていない、AI 機能を備えたソフトウェア、サービス、モデル、プラグイン、API またはクラウドベースの機能を指します。

組み込み AI 機能とは、すでに承認済みの企業向けアプリケーションまたはデジタルサービスに組み込まれた AI 機能を指します。

内部 AI 開発とは、本学の職員または本学に代わって業務を行う契約業者が、開発、ホスティングまたは運用を行う AI 関連のコード、モデル、ノートブック、ワークフロー、プロトタイプまたは統合機能を指します。

研究目的のみの活動とは、調査、実験またはプロトタイプの開発を目的として、もっぱら研究の文脈内で行われる AI の利用を指し、他者向けの共有サービスや運用システムとして展開されるものではないことを意味します。

サンドボックス環境とは、本学が管理する隔離された環境を指し、テスト、実験または限定的な評価を目的としており、明確な利用制限が設けられ、運用上の依存関係がないものです。

導入とは、AI システム、モデル、ワークフローまたはサービスを、本番運用、共有利用、あるいは個人や研究室内での実験的利用を超えた組織的な利用のために提供し、組織が継続的に依存する状態に置くことを指します。

「**実質的な寄与**」とは、著作者、学術的誠実性、または記録の信頼性に関連する形で、文言、分析、コード、画像、解釈またはその他のコンテンツを実質的に形成する AI による支援を意味します。

「**承認済み AI プラットフォーム**」とは、本学における所定の承認プロセスまたはカテゴリー承認プロセスを完了し、付属書 C に基づき登録され、またはその他の方法により認定された AI ツール、サービス、モデルまたはツールのカテゴリーを意味します。

「**カテゴリーレベルの承認**」とは、リスク特性、管理措置および使用制限が同等である特定の低リスク AI ツールまたはモデル環境の区分について、承認された条件の範囲内では個々のツールごとの審査を省略できる CIO による承認をいいます。

44.2.4 AI ライフサイクルにおける役割

説明責任と透明性を確保するため、本学における AI ガバナンスでは、以下の 4 つの機能的役割を区別しています。

- **企画者（Planners）**：AI 導入の目的と要件を定義します。
- **開発者（Developers）**：AI システムまたはコンポーネントを設計、構築およびテストします。
- **提供者（Providers）**：本学のシステムやプロセス内に AI サービスを展開または統合します。
- **利用者（Users）**：業務、研究または学習において AI ツールを適用し、人による監督を行い、出力を検証します。

各役割は、本ポリシーの基準に従い、責任を持って行動しなければなりません。

本ポリシーの実施は、十分な情報に基づく専門的判断および個人の説明責任に依拠します。管理措置、文書化および監督は、リスクに比例し、合理的に実行可能なものでなければなりません。現行の登録簿における承認の範囲内での日常的な利用については、別段の定めがない限り、個別の承認または文書化を必要としません。

44.2.5 AI リスク管理フレームワーク

AI に関する活動は、付属書 A に基づき、そのリスクに応じて適切に管理するものとします。OIST が、新規または重要な変更が加えられた AI ツール、モデル、サービス、またはそのカテゴリーを承認または導入する前に、文書化されたリスク評価を実施する必要があります。ただし、個々の利用者による利用が AI ツール登録簿の有効な登録内容、またはカテゴリー単位で承認された範囲内にとどまる場合には、個別にリスク評価を実施する必要はありません。

承認内容については、少なくとも年 1 回見直すものとし、サービス、モデル、機能、ベンダーの契約条件、データフロー、システム連携、セキュリティ管理策、または適用される要件に重要な変更があった場合には、年次見直しを待たずに見直しを行うものとします。また、提案された利用方法が承認済みの範囲を超える場合、またはリスクレベルが高まる場合には、追加のリスク評価を実施する必要があります。

リスク評価では、以下の双方を考慮する必要があります。

1. セキュリティ、プライバシー、倫理、運用上の影響を含む**固有のリスク**
2. 効率性、研究能力、サービス品質またはイノベーションの向上など、**期待される便益および本学にもたらされる効果**

必要に応じて、文書化された**リスク軽減措置**を適用し、残存リスクを許容可能なレベルまで低減することができます。

CIO は、適切なセキュリティ管理および利用制限が実施されていることを条件として、低リスクの AI ツール（本学が管理する環境、サンドボックス環境またはエアギャップ環境内で動作するオープンソースモデルなど）の分類について、**カテゴリーレベルの承認**を行うことができます。

このようなカテゴリー単位の承認により、組織としての監督と説明責任を維持しつつ、管理上の負担を軽減することができます。

研究の文脈における AI の低リスクな探索的利用は、適切な管理措置を通じて支援され、可能にされるべきです。

44.2.6 情報の分類とデータの利用

すべての AI の利用は、OIST 情報資産の格付けフレームワーク [PRP 17.3.9] に準拠しなければなりません。複数のデータセットを組み合わせる場合は、最も高い分類レベルが適用されます。付属書 B には、AI 特有のデータ取り扱い要件がまとめられています。

個人情報、規制対象情報、輸出管理対象情報、パートナーによる制限がある情報、および契約上の制限がある情報については、PRP 17.3.9 に基づいて分類し、適用される要件に従って取り扱う必要があります。これらは、情報の取扱いに関する属性であり、独立した機密区分を構成するものではありません。

IT 部門は、本学が承認した AI プラットフォームに、必要に応じて契約上の保証や技術的統制を含め、著作権に関連するリスクを軽減するための組織的な保護措置が盛り込まれていることを確保しなければなりません。

44.2.7 協議および報告

本学は、本学の活動に使用される外部提供の AI ツールおよび AI を活用したデジタルサービスについて、既存の承認またはカテゴリーレベルの承認の対象となっている場合を除き、情報セキュリティセクション（IT 部門）が管理する標準的な導入および検証プロセスに従うことを義務付けています。

これにより、本学のデータが処理される前に、セキュリティ、データ保護、著作権関連、契約上および運用上のリスクが評価されることが保証されます。

a) 外部 AI ツール (SaaS、クラウドサービス、プラグイン、API ベースのツール)

- 当該ツールまたはサービスがすでに承認済み AI ツール登録簿に掲載されている場合、その使用目的が承認された範囲、データ分類の制限および明示された条件内にとどまる限り、新たな承認は必要ありません。
- ツールまたは環境がカテゴリー単位の承認の対象となっている場合、モデルごとまたはツールごとに個別の承認を得る必要はありません。通知が必要となるのは、当該承認において通知が必要である旨が定められている場合に限られ、通知は環境、サービス、またはプロジェクト単位で行うものとし、個々の利用ごとに行う必要はありません。
- ツールがまだ承認されておらず、本学のデータが処理される場合は、標準的な導入および検証プロセスを開始できるよう、ユーザーまたは申請部署は使用前に情報セキュリティセクションに相談しなければなりません。

b) 内部 AI 開発 (研究コード、モデル、プロトタイプ、ノートブック)

内部開発活動については、以下の場合を除き、使用前の検証は不要です：

1. 当該システムが他者による利用のために展開される場合（例：内部サービスとして）
2. システムが「機密」または「クリティカル」なデータを処理する場合
3. 展開が研究環境またはサンドボックス環境の範囲を超える場合

このような場合は、「AI 開発および導入基準」（付属書 D）に従う必要があり、情報セキュリティセクションは、研究プロセスそのものではなく、導入の状況を検証します。

当該研究を行う個人、研究室、またはプロジェクトチームの内部に限定された研究目的のみの AI 利用で、公開 (Public) または学内 (Internal) に分類される情報のみを使用し、かつ、導入、共有、または重要な結果を伴う業務に使用されないものについては、正式な承認を必要としません。ただし、この免除は、付属書 A および D に定める規制対象、その他リスクが高いと判断される条件に該当する場合には適用されません。

c) 報告

ユーザーは、いつでも情報セキュリティセクションに連絡し、不明点の確認またはインシデント、プライバシーに関する懸念、セキュリティに関する懸念、未承認のツールに関する懸念、不審な出力、予期せぬモデルの挙動などを報告することができます。善意に基づく報告は奨励され、保護されます。

44.3 規則

44.3.1 AI ツールの調達、承認、および登録

a) 本学での使用を目的とする、外部から提供されるすべての AI システム、クラウドサービス、プラグイン、API および AI 対応アプリケーションは、承認済みリストまたはカテゴリーレベルの承認の対象となっている場合を除き、本学のデータを処理する前に、情報セキュリティセクションによるセキュリティおよびデータ保護の検証を受けなければなりません。すでに承認済みのエンタープライズシステムに組み込まれている AI 機能については、新たなデータフロー、外部のモデルプロバイダー、自動化された意思決定、ベンダー規約の重要な変更またはその他のリスクにおける重要な変更をもたらさない限り、本ポリシーの適用対象となります。

b) 承認されたツールは、AI ツール登録簿（付属書 C）に登録され、定期的に見直されるものとします。

c) 承認は、登録簿に記載されたサービスまたはカテゴリー、アカウントの範囲、情報の機密区分、利用範囲、および適用条件に対して適用されます。これらの要素に重要な影響を及ぼす変更がある場合に限り、追加の審査を行う必要があります。

d) AI システムの調達にあたっては、ベンダーの契約条件、データ取り扱い、セキュリティおよび著作権に関連する管理策について、リスクに応じた適切な審査を実施する必要があります。

e) 機密（Confidential）に分類される情報は、最新の OIST AI 登録簿において、当該機密区分の情報を取り扱うことが承認されているサービスでのみ処理することができます。クリティカル（Critical）に分類される情報は、文書化されたリスク評価を実施した上で、当該サービスおよびユースケースについて CIO の事前承認を得ている場合に限り、処理することができます。

f) CIO は、低リスクの AI ツール（例：本学が管理するサンドボックス環境で動作するオープンソースモデル）のクラスに対して、カテゴリーレベルの承認を発行することができ、承認された条件の範囲内では、モデルごとの審査が不要となります。

g) CIO は、安全でないと判断された AI ツールについて、承認を一時停止または取り消すことができます。

44.3.2 AI の開発および導入基準

a) AI システムの開発または統合はすべて、「AI 開発および導入基準」（付属書 D）に準拠しなければなりません。

b) 機関として実施する AI プロジェクトは、文書化された設計、データガバナンスおよびテストの要件に従わなければなりません。

c) 以下の条件に該当する内部開発システムの導入前には、CISO による審査および CIO の承認が必要です。

- (1) 開発元の研究室または部署以外の者が使用するもの
- (2) 機密データまたはクリティカルデータを処理するもの
- (3) 研究環境またはサンドボックス環境以外で稼働するもの
- (4) 重要な組織運営を支援するもの、または規制対象その他の高リスク条件を伴うもの

d) サードパーティの API および生成モデルの統合については、本学のデータ取り扱い要件への準拠を確保するため、契約およびセキュリティに関する審査を受けなければなりません。

44.3.3 学術的誠実性、研究倫理および AI の学術的利用

a) 学術的責任の原則

人工知能（AI）は、研究、教育、学術サービス、アウトリーチ、広報、商業化および関連する本学の活動を支援することがありますが、人間の知的責任、批判的判断または説明責任に取って代わるものではありません。学術的および学問的な成果物は、その正確性、誠実性、独創性および適切性について全責任を負う人間の著者または責任ある個人に帰属し続けなければなりません。

b) 適用範囲

本節は、以下を含むがこれらに限定されない、学術的および教育的活動に適用されます：

1. 研究の設計、分析、解釈および公表
2. 助成金申請書の作成、助成金の申請、助成金報告および関連する助成機関との連絡
3. 査読、編集業務および学術的評価
4. 教材の作成、授業の実施、学生の評価、指導および論文に関連する業務

5. アウトリーチ、一般向け広報、メディア対応、CPR が支援する業務およびその他の学術的または機関としての広報活動
6. 商業化、イノベーション、技術移転その他の学術的または研究関連の対外的活動

本節に基づく利用は、本ポリシーのその他の適用規定、研究科の規則、研究ガバナンスの要件、知的財産の要件、契約上の義務、助成機関の要件その他本学の適用される方針や手続きに従うものとします。

c) 批判的評価および検証基準

AI によって生成された出力は、学術的、教育的な文脈または影響力の大きい文脈で使用する前に、適切な権威ある情報源と照らし合わせて批判的に評価されなければなりません。利用者は、その使用の文脈、重要度および潜在的な影響に見合った合理的な検証基準を適用しなければなりません。これには、該当する場合、以下が含まれます：

1. 事実に関する主張、参考文献、引用および出典表示の検証
2. 分析の正確性、方法論の妥当性および必要に応じた再現性の確認
3. 潜在的な偏り、捏造、省略または誤解を招くような表現の評価
4. 提出、伝達、公表、評価または依拠を行う前に、文言、解釈、コード、画像、その他の実質的内容の確認

検証が期待される状況において、合理的な検証を適用しなかった場合は、AI 利用上の過失とみなされる可能性があります。

d) 不正行為、過失、および善意の使用

AI に関連する行為は、その行為の性質および行使された注意の程度に応じて評価されるものとします：

1. 意図的な誤用（不正行為）。意図的な剽窃、捏造、改ざん、著者の虚偽表示、誤解を招くことを意図した非開示の AI 生成コンテンツまたはその他の AI の欺瞞的な使用は、不正行為を構成し、懲戒処分やその他の組織的な手続きの対象となる可能性があります。
2. 過失による利用。合理的な検証、レビューまたは修正を行わずに AI の出力を不適切に依拠することは、特に学術、教育、報告、評価、アウトリーチ、商業化またはその他の影響力の大きい文脈において、過失による利用とみなされる可能性があります。
3. 善意かつ十分な注意を払った利用。適切な検証と開示を行い、かつ、該当する機関、学問分野、助成機関、出版社、講座および発表の場における要件を遵守した上で AI を利用することは許可されます。

e) 意図しない誤り

AI によって生成されたコンテンツに起因する意図しない誤り（ハルシネーション、誤った引用、不正確な要約または偶発的な文章の類似性を含む）は、それ自体では不正行為とはみなされません。ただし、ユーザーは、提出、公表、共有、評価またはその他の利用に先立ち、出力結果を確認し、誤りを修正し、最終的な成果物が正確かつ適切であることを確保する責任を負います。

f) 開示と透明性

AI の利用が、学術、教育、教授、アウトリーチ、報告または商業化に係る成果物における文言、分析、コード、画像、解釈その他の実質的な要素に寄与する場合には、その利用を開示しなければなりません。開示は、文脈に応じた適切な方法で行うものとし、これには、方法論の記述、謝辞、脚注、指導教員による指示に基づく開示、助成機関が要求する開示、あるいは学問分野の規範、出版社の要件、助成機関の条件、コースの規則または所属機関の指針に合致するその他の形式が含まれる場合があります。

g) 再現性と科学的誠実性

AI が研究成果に実質的に寄与する場合、合理的に可能な範囲で、透明性、再現性および検証を裏付ける適切な文書を保持しなければなりません。不透明なモデルや外部でホストされているモデルが、研究の結論、解釈または生成されたコンテンツに実質的な影響を与える場合、その使用については、目的、制限および利用可能な代替案を考慮して正当化されなければなりません。

h) 教育および学生の評価

教育、教材の作成、学生の評価、指導および学位論文に関連する業務における AI の利用は、学術的誠実性、公平性および個々の学習に対する有意義な評価を損なわないものでなければなりません。AI は、評価や指導における指導教員の判断に取って代わるものであってはなりません。学生は、授業、試験、評価および学位論文作成における AI の利用に関する研究科の方針および各科目の規定を遵守しなければなりません。

i) 研究代表者、シニアオナーおよび責任者の説明責任

研究、助成金、教育、アウトリーチ、広報、報告または商業化に係る成果物が、チームで作成される場合、または機関の責任の下で作成される場合には、研究代表者、シニアオナー、教員、指導教員、プロジェクトリーダーまたはその他の責任者が、AI の利用が適切であること、必要な開示が行われていることおよびその成果物が正確かつ検証可能であり、適用される義務に準拠していることを確保する責任を負います。

j) 商業化および知的財産

商業化、イノベーション、技術移転、対外的活動または関連活動における AI の利用は、知的財産に関する義務、機密保持要件、契約条件、スポンサーの条件、必要に応じた輸出管理要件およびその他の関連する機関の規則を遵守しなければなりません。

k) 日常的な支援

文法修正、軽微な文章の推敲、書式設定の支援、文字起こしの整理その他これらに類する支援など、成果物の実質的な内容に重大な影響を及ぼさない限定的な AI の利用は、学問分野の規範、指導教官のガイダンス、授業の規則、出版社や発表先の要件およびその他の適用される機関の要件に従って行うことができます。

44.3.4 生成 AI の責任ある利用

AI は、起草、分析または調査を支援することはできますが、学術的または機関としての説明責任が求められる場面において、人間の意思決定、著作者としての役割または評価に取って代わることは許されません。

- a) 利用者は、タスクに必要な情報のみを提出し、AI ツール登録簿に従わなければなりません。機密情報は、明示的に許可されている場合に限り利用できます。クリティカル情報には、第 44.3.1 (e) 項に基づく CIO の事前承認が必要です。
- b) 利用者は、プロンプトが可能な限り正確かつ中立であるよう合理的な注意を払い、倫理および学術的誠実性の原則に整合させなければなりません。
- c) 利用者は、文脈、重要度およびリスクに応じた程度で、AI が生成した出力を確認しなければなりません。
- d) 利用者は、承認された AI プラットフォームおよび適用されるライセンスに関連する、著作権に関する保護措置、制限および開示要件を遵守しなければなりません。
- e) 利用者は、AI の支援を受けて行う意思決定について引き続き責任を負い、リスクの程度に応じた適切な人による確認を実施するものとします。
- f) AI の出力を、自動化された意思決定や、個人に重大な影響を及ぼす意思決定（例：採用、評価、学業成績の評価、懲戒処分、アクセス制御または福利厚生配分）の唯一の根拠として使用してはなりません。AI がそのような意思決定を支援するために使用される場合、文書化された人によるレビューと説明責任が求められ、当該 AI ツールは、該当するデータ分類およびユースケースについて明示的に承認されている必要があります。

- g) CIO または CISO から指示があった場合、ユーザーは安全でないと判断された AI ツールまたはモデルの使用を直ちに中止しなければなりません。

44.3.5 リスクおよびインシデントの報告

- a) **報告対象となるインシデント。** 重大なリスク、ポリシーの不遵守、データ保護上の懸念、セキュリティ上の懸念、重大な結果をもたらす文脈における著しいバイアス、無許可の AI ツール使用、著作権上の懸念または組織への潜在的な損害をもたらす AI 関連のインシデントおよび懸念事項は、付属書 G に記載されたエスカレーション手順を通じて報告しなければなりません。
- b) **通常の軽微なエラー。** 通常の出力の不正確さ、一般的な「ハルシネーション（事実に基づかない出力）」、不十分な要約、軽微な書式上の不具合その他これらに類する影響が軽微なエラーであって、通常の人的確認により検出・修正されるものについては、それ自体では正式なインシデントのエスカレーションする必要はありません。ただし、より広範なリスクの傾向を示す場合、制限付きデータが関与する場合、重要な意思決定に影響を及ぼす場合または法的、コンプライアンス、安全性もしくは評判上の懸念を生じさせる場合は、この限りではありません。
- c) **CIO および CISO の対応。** CIO および CISO は、データ漏えいの評価、情報共有に関する要件ならびに必要な応じたツールの利用停止を含め、封じ込めおよび対応措置を共同で決定します。
- d) **一時停止および是正措置。** 本学は、リスクが是正されるまで、影響を受けたシステムを一時的に停止するか、承認を取り消すことがあります。

44.3.6 研修および啓発

すべての OIST 構成員は、それぞれの役割およびアクセス権限に応じた **Responsible AI**（責任ある AI）に関する基礎的な教育を受けるものとします。正式な研修の受講および再受講に関する要件は、役割、システムへのアクセス権限、取り扱う情報、およびリスクへの関与の程度に応じて適切に定めるものとします。開発者、提供者、および管理者は、付属書 F に定める、それぞれの役割に応じた研修を受講するものとします。

44.3.7 監査およびコンプライアンス

情報セキュリティセクションは、付属書 E に基づくリスクベースの監査を実施するものとします。AI ガバナンス委員会は、重要な監査結果を定期的に審査し、必要に応じて、CIO または CISO に対して是正措置または使用停止措置を提言することができます。

44.3.8 施行

コンプライアンス違反への対応は、その行為の重大性、意図、被害、および再発の有無に応じて、適切に行うものとします。善意による過誤であり、速やかに報告および是正された場合は、原則として、指導または是正対応を通じて対処するものとします。故意、重大な不注意、または繰り返し行われた違反については、アクセス制限、契約の見直し、懲戒処分、または適用される OIST の手続きに基づく上位への報告・対応につながる場合があります。

44.4 責任

役割	責任
最高情報責任者（CIO）／AI 統括責任者（CAIO）	組織の AI ガバナンスを主導し、AI ツールおよびその分類を承認し、使用停止指示を発令し、他の PRP 章との整合性を確保し、理事長／学長に年次報告を行います。

役割	責任
最高情報セキュリティ責任者 (CISO)	セキュリティおよびリスク評価を実施し、外部ツールの検証を行い、研究環境を超えて利用される学内 AI システムの導入を審査し、インシデント対応を監督し、AI セキュリティ基準を徹底します。
プロボスト	研究、教育および学術的誠実性における AI の責任ある利用について学術的なリーダーシップを発揮し、本学の学術的使命および組織戦略との整合性を確保します。部門横断的な実施を支援します。プロボストによる学術上の承認や指針は、AI ツールの承認、セキュリティ検証またはデータ取扱許可に関する CIO/CISO の要件に取って代わるものではありません。
教員担当学監 (DFA)	教員および教員人事に関する業務において、本ポリシーの運用および遵守を確保します。
研究科 (GS) / 研究科長	教育・学術プログラムに「責任ある AI」を組み込み、学生の学術的誠実性に関する要件を徹底させ、学生および TA (ティーチング・アシスタント) への研修を確実に実施するとともに、授業に関連する AI の利用について CIO と連携します。
副プロボスト	OIST の学術・研究活動における責任ある AI 利用を監督し、学術・研究ガバナンスとの整合性を確保するとともに、AI に関するトレーニングおよび基準の導入を支援し、必要に応じてプロボスト、CIO/CAIO、CISO と連携します。
AI ガバナンス委員会	AI ポリシー、高リスクまたはポリシーに影響を及ぼす承認事項、監査の傾向、研修の実施状況および本学への提言について、助言的な監督を行います。
情報セキュリティセクション	セキュリティ検証を実施し、承認済み AI ツール登録簿を管理し、インシデントの記録およびリスク監視を管理し、監査準備を支援します。契約や導入プロセスに盛り込むべきセキュリティおよびデータ保護要件を調達部門へ提供します。
調達課	ベンダー選定および契約内容が付属書 D の要件を満たしていることを確認し、必要に応じて情報セキュリティセクションおよび法令・コンプライアンス部門と連携の上、必須条項（データ処理、非保持、提出データを用いた学習の禁止、セキュリティ管理、該当する場合は保証・監査証拠および著作権関連の保護措置）が含まれていることを確認します。
人事 (HR)	職員の研修を調整し、研修修了記録を管理し、コンプライアンスの追跡を支援します。
OIST イノベーション	産学連携やスタートアップ支援における責任ある AI 利用を監督し、必要に応じて、データ取り扱いおよび契約上の義務が CIO/CISO の要件に整合していることを確保します。
本学構成員（企画者、開発者、プロバイダー、ユーザー）	本ポリシー、付属書 A～H および付属書 D の基準に従い、人的監督を維持し、承認済みのツールまたは該当するカテゴリーの承認対象となるツールのみを使用し、データを保護し、承認された利用条件を遵守し、確立された相談およびエスカレーションの手順を通じて、懸念事項やインシデントを誠実に報告します。

44.5 手続き

44.5.1 承認および調達

a) 最新の登録簿への登録またはカテゴリーレベルの承認の対象となっていない、新規または重要な変更を伴う AI 提案には、AI リスク評価ならびに該当するベンダー情報および契約情報を提出する必要があります。外部ツールについては、「テクノロジーオンボーディング承認プロセス」を完了し、情報セキュリティセク

ションによる確認を受ける必要があります。情報セキュリティセクションは、本ポリシーに基づく AI ツールに関する相談、オンボーディングおよび承認手続きの調整に関する窓口としての役割を担います。

b) 情報セキュリティセクションは、本学での使用が承認された、条件付きで許可されたおよび禁止された AI ツールおよびサービスに関する機関の公式記録として AI ツール登録簿を管理します。登録簿には、ツール固有の条件、アカウント要件、承認された使用範囲ならびに安全かつコンプライアンスに準拠した利用に必要なその他の運用上の指針が含まれる場合があります。登録簿は、ベンダーの利用規約、技術的制御、機関の決定、法的または規制上の考慮事項および運用リスクの変更を反映するため、随時更新される場合があります。利用者および各部門は、登録簿の最新版に従わなければなりません。ツールまたはサービスの承認は、登録簿に記載された範囲、条件、アカウントの種類および機関としての利用条件の範囲内でのみ有効です。

c) 情報セキュリティセクションは、本ポリシーに基づき導入手続きまたは補足的な審査を必要とする外部から提供されるすべての AI システムおよび AI 対応クラウドサービスについて、セキュリティおよびデータ保護の評価を実施します。

d) 調達部門、情報セキュリティセクションおよび法令・コンプライアンス部門は、該当する場合、ベンダー情報、管理措置および利用可能な保証をリスクに応じて評価します。

e) CIO が行うカテゴリーレベルの承認（例：サンドボックス環境で実行されるオープンソースモデル）は、個々のモデルごとの審査を要することなく有効とします。カテゴリー承認済みの環境については、特に指定がない限り、個々の利用ごとではなく、CIO が定める環境、サービスまたはプロジェクト単位で、情報セキュリティセクションへの通知が必要です。

44.5.2 インシデントおよびリスクの報告

a) **報告対象となるインシデント。** データ漏洩、セキュリティ上の懸念、未承認ツールの使用、モデルの反復的または有害な挙動、重大な結果をもたらす状況における著しいバイアス、著作権侵害の疑い、重大なコンプライアンス違反またはその他の重要な組織的リスクを伴う AI 関連のインシデントや懸念事項は、付属書 G に定めるエスカレーション手順に従い、情報セキュリティセクションに報告しなければなりません。

b) **通常の軽微なエラー。** 通常の人的確認により修正される影響が軽微な出力エラーについては、それ自体では正式なインシデントとしてエスカレーションする必要はありません。ただし、より広範な問題を示す場合、または制限付きデータ、規制対象の活動、安全性、重要な意思決定その他リスクの高い状況において発生した場合は、この限りではありません。

c) **CIO および CISO の対応。** CISO と CIO は封じ込め、是正および連絡対応を共同で決定します。

d) **一時停止および是正措置。** 本学は、リスクが対処および軽減されるまで、影響を受けたツール、サービスまたはカテゴリーを一時的に停止することができます。

44.5.3 定期および年次レビュー

a) AI ガバナンスに関する報告は、可能な限り既存の組織内記録を活用して、定期的に作成するものとします。追加の情報や背景の把握が必要な場合には、各ディビジョンに対して必要な情報の提供を求めることがあります。

b) **年次報告書の内容。** AI ガバナンス委員会は、以下の内容を要約した年次報告書を理事長／学長に提出します：

- (1) AI のリスクプロファイルおよび傾向分析
- (2) 監査結果（付属書 E）
- (3) 研修の遵守率（付属書 F）
- (4) ポリシーまたは基準の更新
- (5) 推奨される組織的な措置または投資

44.5.4 研修および記録

- a) **研修記録。** 人事部門および研究科は、付属書 F において必須と指定された研修の修了記録を管理します。
- b) **プロジェクト文書。** 関係部署は、第 44.2.7 項または付属書 D に基づき承認を必要とする AI システムについて、リスクに応じた文書を保存しなければなりません。適用除外の範囲内で行われる通常の研究実験については、通常の研究またはプロジェクト慣行を超える追加文書を必要としません。
- c) **保存および保管。** すべての文書は、PRP 12（法人文書管理）に従って保存され、承認された法人文書ファイルに保管されなければなりません。

44.6 付属書

付属書 A - AI リスクマトリックスおよびリスク事例

A.1 目的

本付属書は、AI 活動をリスクレベル別に分類し、それに伴う影響を特定し、必要なリスク軽減措置、監督および承認プロセスを定める方法を規定するものです。

このリスクマトリックスは、リスクに応じたガバナンスを支援し、本学の構成員が、関連する PRP の規定および倫理的・学術的基準に沿って、AI 技術を安全に活用できるようにすることを目的とします。

A.2 適用範囲

本付属書は、本学の研究、教育、事務および運営において使用されるすべての AI および生成 AI のツール、モデル、サービスに適用されます。これには以下が含まれます：

- 外部の SaaS 型 AI ツールおよびクラウドサービス
- ローカルでホストされるオープンソースモデル
- 研究環境やサンドボックス環境の外に展開された、内部で開発されたシステム
- 商用アプリケーションに組み込まれた AI 機能
- 管理された環境下で動作する、カテゴリ承認済みのオープンソースモデル

本付属書は、作成元の個人、研究室またはプロジェクトチーム内に限定され、「公開」または「学内」の情報のみを使用し、導入、共有または重要な業務運用に使用されない研究目的のみのプロトタイプには適用されません。ただし、規制対象その他の高リスク条件を伴う場合は、この限りではありません。

A.3 リスク評価手法

AI 活動は、以下の 5 つの側面から評価されなければなりません：

1. データの機密性：PRP 17.3.9 に従い、AI 活動によって処理、生成、保存または開示されるすべてのデータの分類および機密性
2. 運用への影響：エラー、ハルシネーション、誤動作、利用不能、遅延または誤用が、本学の学術、研究、事務または業務運営に及ぼす潜在的な影響
3. 倫理的および評判上のリスク：AI 活動またはその成果物から生じる、危害、偏見、不公平、誤用、欺瞞、著作権侵害または評判の毀損の可能性
4. システムの重要度：AI 活動が、学術、研究、運営、安全、法務、規制またはコンプライアンスに関連する重要な機能や意思決定をどの程度支援または影響を与えるか
5. 科学的誠実性リスク：AI の利用により、捏造、誤解を招く、偏ったまたは十分に文書化されていない結果などを通じて、研究や学術的成果の妥当性、再現性、解釈可能性、検証可能性または追跡可能性が損なわれる可能性

これらの側面は、学術、研究、事務、経営管理、ワークフローの自動化およびアシスタントツールを含む、あらゆる形態の AI の利用に適用されます。

各評価項目のうち、最も高い評価が全体のリスクレベルを決定します。

AI のリスクレベルは、情報の機密区分とは異なるものです。個人情報、規制対象情報、輸出管理対象情報、パートナーによる制限がある情報、または契約上の制限がある情報を取り扱う場合、PRP 17.3.9 に基づき付与された機密区分を変更することなく、AI のリスクレベルが高くなる、または追加の要件が適用される場合があります。

レベル	説明	代表的な用途（詳細）	必要な管理措置
低	Public に分類されるデータを使用し、誤りがあった場合の影響が軽微なもの	文法チェック、翻訳、公開コンテンツ向けの会議アシスタント、一般的なスケジュール管理支援ツール	最新の AI ツール登録簿への登録またはカテゴリ単位の承認の範囲内で利用する場合は、通常の利用者による確認を行います。
中	Internal に分類される情報、または限定的な研究・業務データを使用し、風評または業務への影響が中程度のもの	内部メモの要約、講義準備、内部データに基づく業務分析、ワークフロー支援ツール、内部会議の要約ツール	最新の AI ツール登録簿に定める条件の範囲内で利用し、リスクの程度に応じた適切な人による確認を行います。提案された利用が承認済みの範囲を超える場合、またはリスクを著しく高める場合を除き、利用者による個別のリスク評価は必要ありません。
高	Confidential に分類される情報、機密性の高い情報、または重要な研究・業務データを使用するもの	人事・財務関連文書、研究助成金関連資料、事業計画データ、内部の意思決定支援ツール	AI ツール登録簿および適用される承認において、当該情報の機密区分およびユースケースが対象となっている場合に限り利用できます。必要なリスク評価および承認は、所定の承認プロセスを通じて完了するものとし、承認された範囲内での利用について、利用者が重複して個別のリスク評価を行う必要はありません。
クリティカル	個人を特定できる情報、医療情報、契約上の制限がある情報、規制対象情報、またはパートナーによる制限がある情報を使用するもの	臨床情報、パートナーから提供された情報、契約上の制限がある情報、輸出管理対象情報、法的制限のある情報、その他の規制対象情報または AI 関連成果物（該当する場合は、関連する出力結果や技術資料を含みます）	文書化されたリスク評価を実施し、必要な保護措置を講じた上で、特定のサービスおよびユースケースについて CIO の事前承認を得る必要があります。

研究および規制対象利用に関する追加の考慮事項

研究、契約、輸出規制、軍民両用（デュアルユース）、パートナーによる制限またはその他の規制対象となる状況においては、リスク評価において、ソースデータの分類だけでなく、該当する場合は以下の事項も考慮する必要があります：

- モデルの重み、チェックポイントおよび学習成果物
- コード、アルゴリズム、ワークフローおよび技術的手法
- プロンプト、検索コンテンツおよび生成された出力
- みなし輸出または同等の制限が適用され得る場合における、共同研究者、請負業者、来訪者その他の人物によるアクセス
- スポンサー、パートナー、契約、出版または管轄区域固有の制限

研究活動は、それが探索的である、研究のみを目的としているまたはサンドボックス環境で行われているという理由だけで、低リスクとして扱うことはできません。輸出管理、パートナーによる制限、法的、契約上、国家安全保障、安全またはこれらに相当する規制上の懸念が適用される可能性がある場合には、適切な学内の担当部署との協議が必要となります。

著作権リスク

本学は、利用可能なプラットフォームの著作権保護措置および契約上の保証を評価します。利用者は、提供者の学習データまたは出所を検証することまでは求められませんが、出力の適切な利用ならびに適用される引用、出典表示およびライセンス要件の遵守について責任を負います。不審な出力または著作権侵害の疑いがある出力は、情報セキュリティセクションに報告しなければなりません。

A.4 リスク事例

以下の例は、学術、研究、事務、業務および運用といった様々な場面において、AIに関連する一般的なリスクを示しています：

- **ハルシネーションおよび誤った出力**
→ 使用する前に、権威ある情報源や一次資料で検証してください。人による確認なしに、事実の正確性について AI のみに依存しないでください。
- **外部モデルへのデータ漏洩**
→ OIST AI Tool Register で許可されていない限り、非公開情報を外部 AI ツールに入力しないでください。クリティカル情報には、第 44.3.1 (e) 項に基づく CIO の事前承認が必要です。
- **バイアス、公平性および差別**
→ 特に人事、入学審査、評価の文脈における中リスクまたは高リスク活動については、定期的なバイアステストと人による検証を行ってください。
- **シャドウ AI (未承認 AI ツールの使用)**
→ 検証されていない AI ツールの使用は、重大なセキュリティ、プライバシー、契約またはコンプライアンス上のリスクを引き起こす可能性があるため、既定の手順に従い、速やかに報告または相談を行ってください。リスクを評価し、必要に応じて是正または封じ込めを行うため、善意に基づく早期報告が推奨されます。
- **モデルドリフト (性能劣化や挙動変化)**
→ 運用されているシステムは、少なくとも年 1 回および重要な変更後に見直し、必要に応じて管理措置を調整しなければなりません。
- **オープンソースのサンドボックスモデル**
→ 本学が管理する環境で稼働するモデルは、カテゴリーレベルの承認の対象となる場合があり、当該承認に定められた文書化要件に従います。
- **業務プロセスの自動化またはワークフローのエラー**
→ AI アシスタントが、メールの返信、スケジュール設定または業務ロジックを誤って生成する可能性があります。これらについては、人による確認と承認が必要です。
- **意思決定支援における誤解釈**
→ AI が生成した分析結果（予測、KPI ダッシュボード、予算見積もり）を検証なしに信頼した場合、誤った意思決定につながるおそれがあります。
- **評判への悪影響**
→ AI が生成した公開コンテンツ、要約または翻訳は、不適切であったり、誤解を招いたり、本学の発信内容と整合しない可能性があります。
- **著作権およびライセンスに関するリスク**
→ AI の出力結果には、意図せず著作権で保護されたテキストやメディアが複製されている可能性があります。不審な出力結果については、情報セキュリティセクションへ報告する必要があります。
- **セキュリティ上の脆弱性**

→ 敵対的プロンプト、改ざんされたデータセットまたは操作されたモデルにより、有害または不正な出力が生成される可能性があります。

- **外部パートナーまたはベンダーによる不正行為**

→ 外部ベンダーが提供する AI システムは、契約上の義務、データ保持規則または学習データの制限を遵守しない可能性があります。

A.5 見直しサイクル

AI リスクマトリックスは、AI ガバナンス委員会により、少なくとも年 1 回、以下の情報を踏まえて見直しおよび更新されるものとします：

- 監査結果（付属書 E）
- インシデント報告（付属書 G）
- 技術や規制の変更
- 新たに生じている倫理的またはセキュリティ上の懸念
- 学生、ORC、教職員および各部門からのフィードバック

更新内容は CIO によって承認され、必要に応じて本ポリシーまたは付随するガイダンスに反映されます。

付属書 B - 情報の分類およびデータ取り扱いに関する概要

B.1 目的

本付属書は、本学の情報分類フレームワーク（PRP 17.3.9）と、人工知能（AI）および生成 AI（GenAI）の利用に固有の要件との関係を定めるものです。

また、安全で、コンプライアンスに準拠した、責任ある AI 処理を確保するため、各分類レベルに応じた取り扱いルールを定めます。

本付属書は、データ分類およびデータセキュリティの要件のみを扱います。AI に関連する倫理的、運用上、学術的およびガバナンス上の考慮事項については、付属書 A、付属書 D および PRP 44 の他のセクションで別途取り扱います。

B.2 分類レベルと例

レベル	定義	AI 取り扱い要件
公開	一般公開を目的とした情報	適用される登録条件に従い、承認された AI ツールと併用可能です。
学内	本学構成員のみを対象とする情報	オンプレミスまたは審査済みの AI サービスでの使用が可能です。
機密	漏洩した場合、個人や業務に影響を及ぼし得る情報	暗号化し、本学が承認した安全な AI プラットフォーム内でのみ使用してください。
クリティカル	法的、個人的または国家的に機密な情報	文書化されたリスク評価および必要な保護措置を経て、CIO が事前承認した本学承認済みのサービスおよびユースケースでのみ処理できます。

追加の説明

- 異なる分類のデータセットを組み合わせる場合、最も高い分類レベルが適用されます。
- カテゴリーレベルの承認については、登録簿に記載された情報分類の上限および条件の範囲内で情報を利用できます。当該条件内にとどまる場合、個別の審査は必要ありません。
- ベンダーから提供される GenAI ツールについては、契約条件において、許可される処理および学習への利用、保持および削除、保存およびアクセス、ならびに適用される著作権または出所に関する保護措置について明記しなければなりません。本学から提出された情報をベンダーの基盤モデルの学習に使用することは、特別に承認された場合を除き認められません。

B.3 取り扱い規則

- プロンプト、ファインチューニングまたは評価には、必要最小限のデータを使用してください。
- 「学内」、「機密」または「クリティカル」に分類される情報を処理する承認済みのサービスは、サービスのアーキテクチャおよびリスクに応じて、通信時および保存時の暗号化を適切に実施するものとします。
- 可能な限り、個人識別子およびメタデータを削除してください（仮名化または匿名化）。
- 非公開情報は、AI ツール登録簿で許可されている場合に限り、AI サービスに入力することができます。Critical に分類される情報については、44.3.1 (e) に基づき、CIO の事前承認を得る必要があります。

- ユーザーには著作権の出所を確認する義務はありません。その責任は、承認されたプラットフォームおよび組織の安全対策に帰属します。
- 「学内情報」、「機密情報」または「クリティカル情報」は、その使用が明示的に承認された AI ツールでない限り、いかなる AI ツールにもアップロードしないでください。
- 個人データの取り扱いについては、PRP 17、日本のプライバシー関連規制および本学の契約上の義務を遵守しなければなりません。
- 不明な点がある場合は、AI ツールでデータを処理する前に、情報セキュリティセクションに相談してください。

付属書 C - AI ソリューションの承認プロセスおよび承認済みツール登録簿

C.1 目的

本付属書は、本学で使用される、外部から提供された AI システム、サービス、プラグインおよび AI 対応アプリケーションを評価、承認、追跡するための標準化されたプロセスを定めるものです。

これにより、リスク評価、セキュリティ検証、調達管理およびライフサイクル管理の一貫性が確保されます。

新規または重要な変更が加えられた外部提供の AI ツールまたは共有 AI サービスについては、現行の AI ツール登録簿への登録またはカテゴリ単位の承認の対象となっていない場合、あるいは提案された利用が承認済みの条件を超える場合に限り、個別の承認および登録が必要です。内部で開発された研究用プロトタイプについては、付属書 D に従います。

C.2 承認ワークフロー

外部の AI ツールまたは AI 機能を備えたサービスを利用する前に、企画者、提供者、または利用を申請する部署は、当該サービスおよびその利用目的が、AI ツール登録簿または適用可能なカテゴリ単位の承認の対象となっているかを確認しなければなりません。

対象である場合、利用が承認された範囲、条件、および情報分類上の制限の範囲内である限り、新たな承認、検証チケットの作成、または利用者によるリスク評価は必要ありません。

対象となっていない場合、申請部署は情報セキュリティセクションに相談し、以下の検証ワークフローに従ってください。

1. 検証チケットの作成

情報セキュリティセクションは、IT サービス管理システムに検証チケットを作成します。

2. セキュリティおよびリスク評価 (CISO)

1. セキュリティ対策
2. データの取り扱い
3. 著作権に関する管理および保証
4. ベンダーの義務
5. AIリスクマトリックス (付属書 A) との整合性

3. 委員会による審査

高リスク、組織全体に影響を及ぼすもの、新規性のあるもの、判断が分かれるものまたはポリシーに影響を及ぼす事案については委員会の審査を要します。委任されたワークフローにより処理される通常の低リスクまたは中リスクの事案については、審査を省略することができます。

4. 承認 (CIO)

CIOは、承認、条件付き承認または却下を決定します。

承認の結果は、必要に応じて本学の AI ツール登録簿に反映されるものとします。

5. 通知

申請部署には、条件、制限事項および承認されたデータ分類レベルが通知されます。

C.3 カテゴリレベルの承認

ボトルネックを回避し、研究の革新を支援するため、CIO は、以下のような低リスクのツールの分類に対して、カテゴリ単位の承認を行うことがあります。

- 本学が管理するサンドボックス環境またはエアギャップ環境で実行されるオープンソースモデル
- 審査済みのモデルファミリー (例: HF モデルや GGUF モデル) を使用するローカル推論環境

- 同一のリスクプロファイルまたは同一の技術的制御を備えたツール

カテゴリ承認を受けた環境については、特に指定がない限り、個々の利用ごとではなく、CIO が定義する環境、サービスまたはプロジェクトの単位で通知を行う必要があります。

C.4 承認済みツール登録簿

情報セキュリティセクションは、本学における利用が承認されている AI サービス、条件付きで許可されている AI サービス、または禁止されている AI サービスならびにサービスカテゴリについて、利用者が参照する正式な記録として AI ツール登録簿を管理します。

登録簿には、利用者がサービスおよび利用目的の可否を判断するために十分な情報を記載しなければなりません。これには、該当する場合、承認状況、OIST 管理アカウントまたはテナントの使用が必要かどうか、承認されている情報分類の上限、承認済みの利用範囲、および主要な条件または制限が含まれます。

適用される管理措置およびリスク条件が実質的に同等である場合、承認は特定のサービスだけでなく、サービスのカテゴリに対して適用することもできます。製品名、モデルバージョンまたは機能の通常の変更は、データの取扱い、セキュリティ、契約条件、システム連携またはリスクに重要な影響を及ぼさない限り、個別の登録を必要としません。

情報セキュリティセクションは、IT およびその他の関係部門と協力して、技術、契約およびリスク評価に関する記録を維持します。これらの記録は、別個の AI ツール登録簿を構成するものではありません。

本登録簿は、IT 部門が指定する学内のチャネルを通じて、本学構成員がアクセスできるものでなければなりません。

C.5 調達管理

調達課、情報セキュリティセクションおよび法令・コンプライアンス部門は、該当する場合、ベンダーの管理措置および契約上の保護をリスクに応じて評価します。これには、許可される処理および学習利用、保持および削除、保存およびアクセス、セキュリティ上の義務、保証または監査証拠、著作権関連の保護措置ならびにインシデント対応義務が含まれる場合があります。

C.6 審査および有効期限

登録簿の記載事項およびカテゴリレベルの承認は、少なくとも年 1 回、または重要な変更が承認条件もしくはリスクに影響を及ぼす可能性がある場合には、それ以前に見直します。既存の承認条件が引き続き適用可能な場合、通常のベンダー、機能またはモデルの更新について個別の審査は必要ありません。

C.7 登録の取り消しまたは一時停止

CIO は、ツール、バージョン、ベンダーの慣行またはモデルカテゴリについて、以下のいずれかに該当すると判断された場合、承認を一時停止または取り消すことができます：

- 許容できないセキュリティまたはプライバシーのリスクをもたらす場合
- 契約上の義務に違反している場合
- 本学のデータを漏洩または保持している場合
- 有害または偏った挙動を示す場合
- 付属書 D の基準を満たしていない場合

利用者は、使用停止の指示に直ちに従わなければなりません。

付属書 D - AI の開発および導入に関する基準

D.1 目的

本学における AI システムの安全な設計、開発、導入および保守のために必要な最低限の技術的、倫理的および運用上の基準を定めることです。

本基準は、他の PRP の規定、学術的誠実性の原則および契約上の義務への適合を確保するものです。

D.2 適用範囲

本付属書は、AI システムが次のいずれかに該当する場合に適用されます。

- (a) 作成元の個人、研究室またはプロジェクトチームを超えて導入または共有される
- (b) 機密情報またはクリティカル情報を処理する場合
- (c) 重要な業務上の意思決定または組織的サービスを支援する場合
- (d) 組織利用のために請負業者またはパートナーが開発する場合
- (e) 規制対象その他の高リスク条件を伴う場合

作成元の研究室またはプロジェクトチーム内に限定され、「公開」または「学内」の情報のみを使用し、導入、共有または重要な業務運用に使用されない研究目的のみのプロトタイプは、本ポリシーの一般要件に従うことを条件として、正式な承認を免除されます。

D.3 原則

本基準は、以下の原則に基づいています：

- 透明性
- セキュリティ・バイ・デザイン
- 説明責任
- 人による監督 (HITL)
- 公平性および非差別
- データの完全性とプライバシー

これらの原則は、AI のライフサイクルのあらゆる段階に適用されます。

適切な管理：措置は、想定される使用目的、データの機密性、導入状況および本学への潜在的な影響に応じたものでなければなりません。低リスクの研究実験については、より緩やかな管理措置を適用する一方で、運用、規制対象または高リスクの利用については、より強力な設計、検証およびガバナンス措置が必要となります。

D.4 設計および開発に関する要件

本学の正式な「AI 開発および導入基準」は、CIO オフィスが別文書として管理します。当該基準は、本ポリシーを運用し、補完するものであり、本ポリシーとの整合性を保たなければなりません。矛盾が生じた場合には、本ポリシーが正式に改正されるまで、本ポリシーが優先します。

少なくとも、本付属書の対象となる AI システムについては、利用目的および管理責任を文書化するとともに、適切なデータガバナンスに関する管理措置、リスクに応じた検証およびテスト、必要に応じた人による監督体制、リスクに応じたログの記録および監視ならびに変更、インシデント対応および廃止のための管理プロセスを整備しなければなりません。

第三者が提供する API、ホスト型モデルまたは外部の開発パートナーを利用する場合には、申請部門または所有部門は、運用展開に先立ち、契約、セキュリティ、データ取り扱いおよび依存関係に関するリスクへ適切に対応しなければなりません。

D.5 ガバナンス

本基準への準拠は CIO および CISO が監督し、情報セキュリティセクションが管理します。AI ガバナンス委員会は助言的な監督を行います。

第 44.3.2 項に記載のとおり、実験室や研究環境以外の場所での AI システムの導入には、**CISO による審査**および **CIO の承認**が必要となります。

D.2 に定める適用除外の範囲内で行われる通常の学内開発および実験については、正式な導入承認は必要ありませんが、適用される本学のポリシー、データ取り扱いに関する規則および本ポリシーに基づき発行される補足ガイダンスに従い、責任を持って実施しなければなりません。

D.6 施行

CIO は、セキュリティ、法的、契約上、倫理的または運用上のリスクが特定された場合、ベンダーの慣行に重大な変更があった場合、許容できないモデルの挙動が検出された場合またはシステムが付属書 D の要件を満たしていない場合には、AI システム、ツールまたは開発活動を一時停止することができます。一時停止措置は、特定されたリスクが評価され、適切な是正措置が完了し、必要に応じて継続使用が再承認されるまで、継続することがあります。

付属書 E - オペレーショナル・リスク管理および監査チェックリスト

E.1 目的

本付属書は、PRP 第 44 章への準拠を確認するための要件を定めるとともに、情報セキュリティセクションが AI 関連の監査を実施する際に使用する標準チェックリストを提示するものです。

監査では、AI 活動が以下に適合していることを確認します：

- AI リスク管理フレームワーク（付属書 A）
- 情報分類に関するルール（付属書 B）
- ツールの承認および登録（付属書 C）
- 開発および導入基準（付属書 D）
- 研修および啓発の義務（付属書 F）
- インシデントのエスカレーション要件（付属書 G）

E.2 チェックリストの概要

レビュー対象となる活動に該当する場合、監査では以下の項目を確認することがあります。

1. ツール承認の確認

1. 外部 AI ツールが情報セキュリティセクションにより検証されていること
2. ツールまたはモデルが AI ツール登録簿（付属書 C）に記載されていること
3. 低リスクのサンドボックス化されたオープンソースモデルについて、カテゴリー承認が適切に適用されていること

2. リスク分類の文書化

1. 第 44.2.5 項に基づき必要とされる場合、AI リスクマトリックスによる評価が完了していること
2. 最も高い分類次元が明確に記録されていること

3. データ取扱いの検証

1. データの分類が付属書 B の要件に合致していること
2. 「学内」「機密」または「クリティカル」なデータについて、外部ツールが不適切に使用されていないこと
3. 必要に応じて、暗号化および匿名化が適用されていること

4. 開発基準の適用

1. システムが AI 開発および導入基準（付属書 D）に準拠していること
2. 第 44.3.2 項に基づき必要な場合、サンドボックス環境以外への展開については、CISO による審査および CIO による承認が行われていること
3. 文書およびテストの証拠が利用可能であること

5. 研修記録の確認

1. 責任ある AI トレーニングが完了していること（付属書 F）
2. 関連する担当者が開発者／プロバイダー向け研修を完了していること

6. インシデント報告の検証

1. 報告経路が正常に機能しており、職員に周知されていること
2. 過去のインシデントが付属書 G に従って記録およびエスカレーションされていること
3. 是正措置が追跡され、完了していること

7. 文書および記録の準拠

1. 必要なすべての文書が PRP 12（法人文書管理）に基づき保管されていること

E.3 監査プロセス

(a) 情報セキュリティセクションは、年次レビューならびに CIO または CISO の指示による臨時監査を含む、リスクベースの監査プログラムを実施します。監査では、文書レビュー、面談、システムチェック、テストシナリオまたは代表サンプリングを用いる場合があります。

(b) 重要な監査結果は AI ガバナンス委員会に報告され、同委員会は結果を審査し、是正措置を提言するとともに、是正状況の進捗をモニタリングします。正式な是正措置および使用停止指示に関する権限は、本ポリシーに基づき指定された権限者が有するものとします。

(c) 重大なまたは繰り返されるコンプライアンス違反があった場合には、ツールの利用停止、追加審査または理事長／学長への報告が行われることがあります。

付属書 F - 研修および啓発の枠組み

F.1 目的

本付属書は、本学のすべての構成員が「責任ある AI」の原則を理解し、業務、研究、および学修において適切に適用できるようにするために必要な研修および啓発の要件を定めるものです。

研修は、安全なイノベーションを支援し、AI の誤用、データ漏洩または非倫理的な AI の挙動のリスクを低減します。

F.2 研修プログラムの構成

- **基礎認識**：本学のすべての構成員を対象とし、ポリシーの認識、責任ある利用、情報分類およびプロンプトの適切情報管理について取り扱います。
- **開発者向けモジュール**：セキュアコーディング、バイアスの軽減、モデルの文書化を扱います。
- **管理者向けモジュール**：リスク評価、調達基準、監査要件を扱います。
- **必須モジュール**：役割、システムアクセス、取り扱う情報またはリスクへの関与度に応じて、所定のトレーニングの修了が求められる場合があります。

F.3 実施方法と頻度

基礎認識については、該当する場合、オンボーディング時に実施します。リフレッシュ研修は、定期的に、また技術、ポリシー又はリスクに重大な変更が生じた場合に実施します。基準や職責が大きく変更された場合には、役割に応じた専門研修を実施します。

F.4 責任

- **人事部門**は、研修受講記録を一元的に管理します。
- **研究科（GS）**は、学生向けの研修プログラムおよび学術的誠実性との整合性を管理します。
- **IT 部門／情報セキュリティセクション**は、技術的なコンテンツ、リスク認識向上のための資料およびインシデント報告に関する指針を提供します。
- **AI ガバナンス委員会**は、研修結果を毎年検証し、改善策を提言します。

F.5 見直しサイクル

研修内容および実施方法は、以下に基づき毎年見直します：

- 監査結果（付属書 E）
- インシデントの傾向（付属書 G）
- 利用者からのフィードバック
- 新たな AI 機能またはリスク

付属書 G - 相談およびエスカレーションの流れ

G.1 目的

本付属書は、人工知能（AI）活動に関する本学の相談およびエスカレーションの手順を定めるものです。

これにより、ユーザーが適時にガイダンスを受け、インシデントが一貫した方法で報告され、組織的なリスクが適切なガバナンス機関へエスカレーションされることを確保します。

G.2 相談窓口

情報セキュリティセクションが相談窓口を運営します。ツールの承認、プロンプトの設計、データの利用およびインシデントの報告に関する助言を提供します。

G.3 エスカレーション手順

AI に関連するすべての懸念事項は、以下のエスカレーション手順に従います。

1. ユーザー → 相談窓口（情報セキュリティセクション）

ユーザーは、以下について相談窓口に問い合わせることができます。

- 承認済みツールまたはカテゴリーの承認に関する確認
- データ利用、開示またはプロンプト設計に関する質問
- 未承認ツールまたは予期しないツール挙動に関する懸念
- データ漏洩、セキュリティ上の懸念、重要な意思決定に影響する重大なバイアス、未承認ツールの使用、著作権上の疑義その他本学に重大なリスクをもたらす事項に関するインシデントの正式な報告

通常の人的確認により検出・修正される影響が軽微な出力エラーについては、より広範な問題を示す場合またはリスクの高い状況で発生した場合を除き、正式なインシデントとしてエスカレーションする必要はありません。

2. 相談窓口 → 最高情報セキュリティ責任者（CISO）

CISO は、重大性を確認し、セキュリティおよびプライバシーへの影響を評価した上で、封じ込め措置を決定します。

3. CISO → AI ガバナンス委員会（組織的または重大な影響がある場合）

エスカレーションの対象となる条件には、繰り返し発生するインシデント、クリティカルデータの漏洩、ベンダーの不履行、倫理的な影響またはモデルの挙動に関する異常が含まれます。

委員会は、本学への影響を精査し、関係者と調整の上、提言を行います。

4. AI ガバナンス委員会 → CIO

CIO は、一時的な停止、カテゴリーレベルの制限またはポリシーの明確化など、必要な措置を承認します。

必要に応じて、CIO は学内全体への通知を行うことができます。

5. CIO → 理事長／学長（ポリシーレベルまたは組織的なリスクがある場合）

以下の場合には、上層部への報告が行われます：

1. ポリシーの改正が必要な場合
2. 本学のリスクが許容可能な閾値を超える場合
3. 評判または法的リスクが重大である場合

G.4 報告および透明性

- 情報セキュリティセクションは、AI に関する重要な問い合わせ、インシデントおよび軽減措置の概要を、定期的に作成します。
- これらの要約は、AI ガバナンス委員会およびエグゼクティブチームに提供され、監査（付属書 E）、研修の更新（付属書 F）およびポリシーの改正に反映されます。
- 重大なインシデントは直ちに CIO に報告され、理事長／学長への年次報告に含めます。

G.5 善意による報告の保護

- インシデント、懸念事項または未承認ツールの使用について善意で報告したユーザーは、不利益な取扱いから保護されます。
- 本学は、被害を未然に防ぎ、安全対策を強化するため、早期の報告を奨励します。

付属書 H - ガバナンスの役割および委員会設置要綱

H.1 目的

本付属書は、国の CAIO（AI 統括責任者）のベストプラクティスおよび本学の内部ガバナンスの枠組みに整合した、本学における人工知能（AI）の組織的ガバナンス体制を定めるものです。また、AI に関するリスク、コンプライアンス、倫理および組織としての対応能力を監督するために必要な役割、責任および運用手順を定めます。

H.2 構造

本学は、以下の階層的な AI ガバナンスモデルを実施します。

- **最高情報責任者（CIO）が AI 統括責任者（CAIO）を兼務。** AI に関連するすべての活動について、組織的なリーダーシップ、戦略的な方向性を示し、最終的な説明責任を負います。
- **最高情報セキュリティ責任者（CISO）。** セキュリティ対策の実施、技術的リスク評価および各部門間の業務調整を監督します。
- **AI ガバナンス委員会。** 次の代表から成る、部門横断的な諮問・監督機関です：
 - 教員
 - 研究科（GS）
 - 人事（HR）
 - 情報技術（IT 部門）
 - 法令・コンプライアンス担当
 - 副プロボストまたはその代理人（本学の学術・研究上のリーダーシップを代表する者）
 - OIST 研究者コミュニティ（ORC）の代表者（PRP 4.5 に基づき追加）
 - 必要に応じて招へいする専門家

この体制により、組織全体を幅広くカバーするとともに、研究者、事務、技術の各観点が統合されます。

H.3 任務

AI ガバナンス委員会は、以下の役割を担います。

- 本ポリシー、AI リスクマトリックスおよび関連する付属書の見直しおよび更新を提言すること。
- 付属書 C に従い、高リスク、組織全体に影響を及ぼすもの、新規性のあるもの、判断が分かれるものまたはポリシーに影響を及ぼす AI システムおよびカテゴリーについて審査し、その承認を提言すること。通常の低リスクおよび中リスクの事例については、委任されたワークフローにより処理することができます。
- 監査結果（付属書 E）を精査し、コンプライアンスの動向を監視すること。
- 研修および啓発活動（付属書 F）を調整し、監督すること。
- AI に関連する組織的リスク、新興技術および規制動向を特定すること。
- CIO を通じて、理事長／学長に対し、ポリシー、基準またはガバナンスの更新を提言すること。
- AI の利用およびリスクに関する透明性の確保および組織内の情報共有を支援すること。

AI ガバナンス委員会は、諮問および監督を行う機関です。本ポリシーに基づき CIO、CISO、調達セクションその他の組織機能に割り当てられた承認権限に代わるものではありません。

H.4 会議および報告

委員会は定期的に、原則として四半期ごとに開催され、重大なインシデント、システム全体に関わるリスク、または重要な技術的变化が生じた場合には、追加会合を招集することができます。

委員会は、以下のとおり報告を行います。

- 定期的に CIO に対し、インシデント、ツールの導入状況、リスクの動向および監査結果を要約して報告します。
- 年 1 回、理事長／学長に対し、リスク、リスク軽減措置、研修の実施状況およびポリシー改正に関する提言を含む、統合された AI ガバナンス報告書を提出します。

H.5 事務局および記録

IT 部門が事務局を務め、以下の業務を行います：

- 議題および議事録を作成すること
- 委員会の活動に関する公式記録を管理すること
- PRP 12（法人文書管理）に従って文書を管理すること
- 報告およびデータ収集について、人事（HR）、研究科（GS）および各部門と調整すること

H.6 定足数

委員会が正式な決定を行うための定足数は、次の通りとします。

- CIO またはその指名する代理者を含む常任委員の過半数が出席していること
- 定足数の算定にあたっては、遠隔による出席を認めます。

H.7 見直しサイクル

委員会は、その要綱を 2 年ごとに見直すものとします。ただし、規制の変更、内部の事案または国の CAIO ガイドラインの改正により必要とされる場合には、それより早期に見直すものとします。