

OIST School Corporation

Policies, Rules & Procedures (PRP)

Chapter 44 – Responsible Use of Artificial Intelligence (AI)

Authority: CEO

Policy Owner: Chief Information Officer (CIO)

Version: V1.0

Effective Date: 2026/11/01

44.1 Policy

This Policy enables responsible use of Artificial Intelligence (AI) and Generative AI (GenAI) in support of OIST's mission, while preserving academic freedom and not restricting legitimate inquiry, research methodologies, or scholarly judgment. OIST School Corporation and OIST Graduate University (collectively referred to as "OIST") recognize AI as a transformative capability for scientific discovery, education, and innovation. OIST supports responsible experimentation, open scientific inquiry, and the advancement of transparent and reproducible research practices.

OIST is committed to the safe, secure, and responsible use of AI and GenAI technologies, ensuring that AI enhances rather than replaces human judgment, accountability, and intellectual responsibility.

This Policy establishes principles, rules, and governance for AI use at OIST, aligned with OIST's frameworks for information security, academic integrity, research ethics, data protection, and records management, and is intended to ensure consistency across institutional, academic, and research contexts.

This Policy promotes transparency, fairness, safety, scientific integrity, human oversight, and accountability across the AI lifecycle—planning, procurement, development, deployment, operation, monitoring, and retirement.

This Policy applies to all OIST members, including faculty, researchers, students, staff, contractors, and visitors, who use, develop, procure, or manage AI systems or AI-generated content as part of their OIST activities. This Policy does not regulate academic judgment or research outcomes; it governs the responsible use of AI tools and systems supporting OIST activities, including institutional systems, services, and administrative processes.

AI activities are governed under the authority of the Chief Information Officer (CIO), who serves as the Chief AI Officer (CAIO) and may formally designate a delegate to perform CAIO duties. AI governance and activities are supported by the Chief Information Security Officer (CISO) and the AI Governance Committee. The Committee includes representation from Faculty, Research, the Graduate School, administrative divisions, technical units, and the OIST Researcher Community (ORC), with cross-functional participation intended to ensure broad institutional oversight and practical input from academic, research, administrative, and technical perspectives.

44.2 General Considerations

44.2.1 Alignment with Existing Policies

This Policy forms part of OIST's Information Technology and Security framework [PRP 17], the Graduate School Policies on Academic Integrity and the Use of Generative AI, and all related PRP policies.

Where conflicts arise, the stricter provision shall apply.

OIST may publish companion guidance, examples, and operational instructions to support interpretation and day-to-day application of this Policy; such materials do not override this Policy and must remain consistent with it.

44.2.2 Purpose and Scope

OIST recognizes the potential of Artificial Intelligence (AI) and Generative AI (GenAI) technologies to improve efficiency, discovery, and innovation while acknowledging their inherent risks — including bias, misinformation, privacy, data confidentiality, data sensitivity, and copyright exposure, model drift, and ethical misuse.

This Policy governs AI use across administrative, academic, and research functions, ensuring risk-proportionate controls are applied across the full AI lifecycle.

44.2.3 Definitions

For purposes of this Policy:

Artificial Intelligence (AI) means machine-based systems that infer from inputs how to generate outputs such as predictions, content, recommendations, classifications, or decisions that can influence physical or virtual environments.

Generative AI (GenAI) means AI systems that generate new content, including text, code, images, audio, video, summaries, or synthetic data, in response to prompts, context, or training.

External AI tool means any AI-enabled software, service, model, plug-in, API, or cloud-based capability provided by a third party and not fully operated within an OIST-controlled environment.

Embedded AI features means an AI capability built into an already approved enterprise application or digital service.

Internal AI development means AI-related code, models, notebooks, workflows, prototypes, or integrations developed, hosted, or operated by OIST personnel or contractors on behalf of OIST.

Research-only activity means AI use conducted solely within a research context for inquiry, experimentation, or prototype development, and not deployed as a shared service or operational system for others.

Sandbox environment means an OIST-controlled isolated environment intended for testing, experimentation, or limited evaluation, with defined usage restrictions and without operational dependency.

Deployment means making an AI system, model, workflow, or service available for operational use, shared use by others, or ongoing institutional reliance beyond an individual's or lab's internal experimentation.

Material contribution means AI assistance that substantively shapes wording, analysis, code, images, interpretation, or other content in a manner relevant to authorship, academic integrity, or the reliability of the record.

Approved AI platform means an AI tool, service, model, or category of tools that has completed the applicable OIST approval or category-approval process and is listed or otherwise recognized under Annex C.

Category-level approval means CIO approval for a defined class of low-risk AI tools or model environments with equivalent risk characteristics, controls, and usage restrictions, eliminating the need for tool-by-tool review within the approved conditions.

44.2.4 Roles within the AI Lifecycle

To ensure accountability and transparency, AI governance at OIST distinguishes among four functional roles:

- **Planners:** Define purposes and requirements for AI introduction.
- **Developers:** Design, build, and test AI systems or components.
- **Providers:** Deploy or integrate AI services within OIST systems or processes.
- **Users:** Apply AI tools in their work, research, or study, exercising human oversight and validating outputs.

Each role must act responsibly, following the standards in this Policy.

Implementation of this Policy relies on informed professional judgment and individual accountability. Controls, documentation and oversight shall be proportionate to risk and reasonably practicable. Routine

use within a current Register approval does not require separate authorization or documentation unless stated otherwise.

44.2.5 AI Risk Management Framework

AI activities shall be managed proportionately under Annex A. A documented risk assessment is required before OIST approves or deploys a new or materially changed AI tool, model, service or category. Individual users do not require a separate assessment when their use remains within a current OIST AI Tools Register entry or category-level approval.

Approvals shall be reviewed at least annually and earlier following material changes to the service, model, functionality, vendor terms, data flows, integrations, security controls or applicable requirements. A supplementary assessment is required when a proposed use exceeds the approved scope or increases the risk level.

Risk evaluation should consider **both**:

1. **the inherent risks**, including security, privacy, ethical, and operational impacts; and
2. **the expected benefits and institutional gains**, including improvements to efficiency, research capability, service quality, or innovation.

Where appropriate, documented **risk-mitigation measures** may be applied to reduce residual risk to acceptable levels.

The CIO may authorize *category-level approvals* for classes of low-risk AI tools (such as open-source models operating within OIST-controlled, sandboxed, or air-gapped environments), provided that appropriate security controls and usage restrictions are enforced. Such category approvals reduce administrative burden while maintaining institutional oversight and accountability.

Low-risk exploratory use of AI in research contexts should be supported and enabled through proportionate controls.

44.2.6 Information Classification and Data Use

All AI use must comply with the OIST Information Classification Framework [PRP 17.3.9]. When combining multiple datasets, the highest classification level applies. Annex B summarizes AI-specific data handling requirements.

Personal, regulated, export-controlled, partner-restricted and contract-restricted information must be classified under PRP 17.3.9 and handled according to applicable requirements. These are handling attributes, not separate classification levels.

The IT Division shall ensure that OIST-approved AI platforms include institutional safeguards to reduce copyright-related risks, including where appropriate contractual assurances and technical controls.

44.2.7 Consultation and Reporting

OIST requires that externally provided AI tools and AI-enabled digital services used for OIST activities follow the standard onboarding and validation process managed by the Information Security Section (IT Division), unless they are already covered by an existing approval or category-level approval.

This ensures that security, data-protection, copyright-related, contractual, and operational risks are assessed before OIST data is processed.

a) External AI Tools (SaaS, cloud services, plug-ins, API-based tools)

- If the tool or service is already listed in the OIST AI Tools Register, no new approval is required provided the intended use remains within the approved scope, data-classification limits, and stated conditions.
- If a tool or environment is covered by a category-level approval, no separate model-by-model or tool-by-tool approval is required. Notification is required only where stated in the approval and at the environment, service or project level, not for each individual use.
- If the tool is not yet approved and OIST data will be processed, the user or requesting unit must consult the Information Security Section before use so that the standard onboarding and validation process can be initiated.

b) Internal AI Development (research code, models, prototypes, notebooks)

Internal development activities do not require pre-use validation unless:

1. the system will be deployed for use by others (e.g., as an internal service); or
2. the system processes Confidential or Critical data; or
3. deployment moves beyond research or sandbox environment.

Such cases must follow the AI Development and Deployment Standards (Annex D), and the Information Security Section will validate the deployment context rather than the research process itself.

Research-only AI activity confined to the originating individual, lab or project team, using only Public or Internal information, and not deployed, shared or used for consequential operations does not require formal approval. This exemption does not apply to regulated or other elevated-risk conditions identified in Annexes A and D.

c) Reporting

Users may contact the Information Security Section at any time for clarification, or to report incidents, privacy concerns, security concerns, unapproved-tool concerns, suspicious outputs, or unexpected model behavior. Good-faith reporting is encouraged and protected.

44.3 Rules

44.3.1 AI Tool Procurement, Approval, and Registration

- a) All externally provided AI systems, cloud services, plug-ins, APIs, and AI-enabled applications intended for OIST use shall undergo security and data-protection validation by the Information Security Section before processing OIST data, unless they are already covered by an approved listing or category-level approval. AI-enabled features embedded in already approved enterprise systems are covered under this Policy unless they introduce new data flows, external model providers, automated decision-making, materially changed vendor terms, or other material changes in risk.
- b) Approved tools shall be entered into the OIST AI Tools Register (Annex C) and periodically reviewed.
- c) Approval applies to the service or category, account boundary, information classification, use scope, and conditions identified in the Register. Further review is required only where a change materially affects these factors.
- d) Procurement of AI systems must include proportionate review of vendor contractual, data-handling, security and copyright-related controls.

- e) Confidential information may be processed only in services approved for that classification in the current OIST AI Tools Register. Critical information may be processed only where the service and use case have prior CIO authorization following documented risk review.
- f) The CIO may issue category-level approvals for classes of low-risk AI tools (e.g., open-source models running in OIST-controlled sandboxed environments), eliminating the need for model-by-model review within the approved conditions.
- g) The CIO may suspend or revoke approval of any AI tool deemed unsafe.

44.3.2 AI Development and Deployment Standards

- a) All development or integration of AI systems must comply with the AI Development and Deployment Standards (Annex D).
- b) Institutional AI projects must follow documented design, data-governance, and testing requirements.
- c) CISO review and CIO approval are required before deployment of internally developed systems that:
 - (1) will be used by individuals outside the originating lab or unit;
 - (2) process Confidential or Critical data; or
 - (3) operate beyond a research or sandbox environment
 - (4) support consequential institutional operations or involve regulated or otherwise elevated-risk conditions.
- d) Third-party APIs and generative model integrations must undergo contractual and security review to ensure compliance with OIST data-handling requirements.

44.3.3 Academic Integrity, Research Ethics, and Scholarly Use of AI

a) Principle of Scholarly Responsibility.

Artificial Intelligence (AI) may support research, teaching, academic service, outreach, communications, commercialization, and related OIST activities; however, it shall not replace human intellectual responsibility, critical judgment, or accountability. Scholarly and academic outputs must remain attributable to human authors or responsible individuals, who retain full responsibility for their accuracy, integrity, originality, and appropriateness.

b) Scope of Application.

This Section applies to scholarly and academic activities, including but not limited to:

1. research design, analysis, interpretation, and publication.
2. grant writing, grant applications, grant reporting, and related sponsor communications.
3. peer review, editorial activities, and scholarly evaluation.
4. teaching material preparation, teaching delivery, student assessment, supervision, and thesis-related work.
5. outreach, public communication, media engagement, CPR-supported work, and other academic or institutional communications; and
6. commercialization, innovation, technology transfer, and other scholarly or research-related external engagement.

Use under this Section remains subject to other applicable provisions of this Policy, Graduate School rules, research governance requirements, intellectual property requirements, contractual obligations, sponsor requirements, and other applicable institutional policies and procedures.

c) Critical Evaluation and Verification Standard.

AI-generated outputs must be critically evaluated against appropriate authoritative sources before use in scholarly, academic, or high-impact contexts. Users shall apply a reasonable standard of verification proportionate to the context, materiality, and potential impact of the use. This includes, as applicable:

1. validating factual claims, references, citations, and source attribution.
2. confirming analytical correctness, methodological soundness, and reproducibility where relevant.
3. assessing potential bias, fabrication, omission, or misleading presentation; and
4. reviewing wording, interpretation, code, images, or other substantive content before submission, communication, publication, assessment, or reliance.

Failure to apply reasonable verification in circumstances where such verification is expected may constitute negligent use of AI.

d) Misconduct, Negligence, and Good-Faith Use.

AI-related conduct shall be evaluated in accordance with the nature of the conduct and the degree of care exercised:

1. **Intentional Misuse (Misconduct).** Intentional plagiarism, fabrication, falsification, deceptive authorship, undisclosed AI-generated content intended to mislead, or other deceptive use of AI constitutes misconduct and may be subject to disciplinary or other institutional procedures.
2. **Negligent Use.** Inappropriate reliance on AI outputs without reasonable verification, review, or correction may constitute negligent use, particularly in scholarly, teaching, reporting, assessment, outreach, commercialization, or other high-impact contexts.
3. **Good-Faith Use with Due Care.** Use of AI with appropriate verification, disclosure, and adherence to applicable institutional, disciplinary, sponsor, publisher, course, and venue requirements is permitted.

e) Unintentional Errors.

Unintentional errors arising from AI-generated content, including hallucinations, incorrect citations, inaccurate summaries, or accidental textual similarity, shall not by themselves constitute misconduct. However, users remain responsible for reviewing outputs, correcting errors, and ensuring the final work is accurate and appropriate before submission, publication, communication, assessment, or other use.

f) Disclosure and Transparency.

AI use must be disclosed where it materially contributes to wording, analysis, code, images, interpretation, or other substantive elements of scholarly, academic, teaching, outreach, reporting, or commercialization outputs. Disclosure shall be made in a manner appropriate to the context, which may include a methods statement, acknowledgment, footnote, instructor-directed disclosure, sponsor-required disclosure, or other form consistent with disciplinary norms, publisher requirements, sponsor terms, course rules, or institutional guidance.

g) Reproducibility and Scientific Integrity.

Where AI materially contributes to research outputs, appropriate documentation shall be retained to support transparency, reproducibility, and verification to the extent reasonably possible. Where non-transparent or externally hosted models materially affect research conclusions, interpretations, or generated content, their use shall be justified considering the purpose, limitations, and available alternatives.

h) Teaching and Student Assessment.

Use of AI in teaching, teaching material preparation, student assessment, supervision, and thesis-related work must preserve academic integrity, fairness, and meaningful evaluation of individual learning. AI shall not replace instructor judgment in evaluation or supervision. Students must comply with Graduate School

policies and course-specific rules governing AI use in coursework, examinations, assessments, and thesis work.

i) Principal Investigator, Senior Author, and Responsible Person Accountability.

For research, grant, teaching, outreach, communications, reporting, or commercialization outputs prepared in teams or under institutional responsibility, the principal investigator, senior author, instructor, supervisor, project lead, or other responsible person remains accountable for ensuring that AI use is appropriate, that required disclosures are made, and that the resulting work remains accurate, verifiable, and compliant with applicable obligations.

j) Commercialization and Intellectual Property.

AI use in commercialization, innovation, technology transfer, external engagement, or related activities must comply with intellectual property obligations, confidentiality requirements, contractual terms, sponsor conditions, export control requirements where applicable, and other relevant institutional rules.

k) Routine Assistance.

Limited AI assistance, such as grammar correction, minor language polishing, formatting support, transcription cleanup, or similar support that does not materially affect substantive content, may be used in accordance with disciplinary norms, supervisor guidance, course rules, publisher or venue requirements, and other applicable institutional requirements.

44.3.4 Responsible Use of Generative AI

AI may assist with drafting, analysis, or exploration, but must not replace human decision-making, authorship, or evaluation where academic or institutional accountability applies.

- a) Users shall submit only information necessary for the task and follow the OIST AI Tools Register. Confidential information may be used only where specifically permitted. Critical information requires prior CIO approval under section 44.3.1(e).
- b) Users shall take reasonable care to ensure that prompts are accurate and neutral to the extent possible and align with ethical and academic integrity principles.
- c) Users shall review AI-generated outputs to a degree proportionate to the context, materiality and risk.
- d) Users shall comply with any copyright-related safeguards, restrictions, and disclosure requirements associated with approved AI platforms and applicable licenses.
- e) Users remain accountable for AI-assisted decisions and shall apply human review proportionate to the risk.
- f) AI outputs shall not be used as the sole basis for automated decisions or decisions that materially affect individuals (e.g., hiring, evaluation, academic assessment, disciplinary actions, access control, or allocation of benefits). Where AI is used to support such decisions, documented human review and accountability are required, and the AI tool must be explicitly approved for the applicable data classification and use case.
- g) If instructed by the CIO or CISO, users must immediately stop using any AI tool or model deemed unsafe.

44.3.5 Risk and Incident Reporting

- a) **Reportable Incidents.** AI-related incidents and concerns that present material risk, policy non-compliance, data-protection concern, security concern, significant bias in a consequential context, unauthorized AI tool use, copyright concern, or potential institutional harm shall be reported through the escalation path described in Annex G.

- b) **Routine Low-Consequence Errors.** Routine output inaccuracies, ordinary hallucinations, weak summaries, minor formatting defects, or other low-consequence errors that are detected and corrected through normal human review do not, by themselves, require formal incident escalation unless they indicate a broader pattern of risk, involve restricted data, affect consequential decisions, or create legal, compliance, safety, or reputational concerns.
- c) **CIO and CISO Actions.** The CIO and CISO jointly determine containment and response actions, including data-loss assessment, communication requirements, and tool suspension where appropriate.
- d) **Suspension and Remediation.** OIST may temporarily suspend affected systems or revoke approvals until the risk is remediated.

44.3.6 Training and Awareness

All OIST members shall receive Responsible AI awareness appropriate to their role and access. Formal completion and refresher requirements shall be proportionate to role, system access, information handled and risk exposure. Developers, providers and managers shall complete applicable role-specific training as defined in Annex F.

44.3.7 Audit and Compliance

The Information Security Section shall conduct risk-based audits using Annex E. The AI Governance Committee shall review material findings periodically and may recommend corrective actions or stop-use measures to the CIO or CISO.

44.3.8 Enforcement

Responses to non-compliance shall be proportionate to the seriousness, intent, harm and recurrence of the conduct. Good-faith errors that are promptly reported and corrected will normally be addressed through guidance or remediation. Deliberate, reckless or repeated violations may result in access restrictions, contract review, disciplinary action or escalation under applicable OIST procedures.

44.4 Responsibilities

Role	Responsibilities
Chief Information Officer (CIO) / Chief AI Officer (CAIO)	Leads institutional AI governance; approves AI tools and categories; issues stop-use directives; ensures alignment with other PRP provisions; reports annually to the CEO/President.
Chief Information Security Officer (CISO)	Conducts security and risk assessments; validates external tools; reviews deployment of internal AI systems beyond research environments; oversees incident response; and enforces AI security standards.
Provost	Provides academic leadership on the responsible use of AI in research, teaching, and academic integrity, and ensures alignment with the OIST Graduate University's academic mission and institutional strategy. Supports cross-divisional implementation. Academic approvals or guidance issued by the provost do not replace CIO/CISO requirements for tool approval, security validation, or data-handling authorization.
Dean of Faculty Affairs (DFA)	Provides operational policy enforcement within faculty and faculty affairs.

Role	Responsibilities
Graduate School (GS) / Dean of Graduate School	Integrates Responsible AI into educational academic programs; enforces student academic-integrity requirements; ensures training for students and TAs; coordinates with CIO on course-related AI use.
Vice Provost	Provides oversight for responsible AI use in academic and research activities within OIST portfolio; ensures alignment with academic and research governance; supports implementation of AI training and standards; and coordinates with the Provost, CIO/CAIO and CISO as appropriate.
AI Governance Committee	Provides advisory oversight for AI policy, higher-risk or policy-impacting approvals, audit trends, training outcomes, and institutional recommendations.
Information Security Section	Conducts security validations; maintains the OIST AI Tools Register; manages incident logging and risk monitoring; supports audit preparation; provides security and data-protection requirements to Procurement for inclusion in contracts and onboarding.
Procurement Section	Ensures vendor selection and contractual agreements satisfy Annex D requirements; confirms inclusion of required clauses (data-processing, non-retention, no-training-on-submitted-data, security controls, assurance/audit evidence where applicable, and copyright-related safeguards) in coordination with the Information Security Section and Legal/Compliance as needed.
Human Resources (HR)	Coordinates staff training; maintains training-completion records; supports compliance tracking.
OIST Innovation	Oversees responsible AI use in industry collaborations and startup engagement, ensuring data handling and contractual obligations align with CIO/CISO requirements as needed.
OIST Members (Planners, Developers, Providers, Users)	Follow this Policy, Annexes A–H, and Annex D Standards; maintain human oversight; use only approved tools or tools covered by applicable category approvals; protect data; comply with approved-use conditions; and report concerns and incidents in good faith through the established consultation and escalation pathways.

44.5 Procedures

44.5.1 Approval and Procurement

- a) New or materially changed AI proposals not covered by a current Register entry or category-level approval must include an AI Risk Assessment and the applicable vendor and contractual information. External tools require completion of the Technology Onboarding Approval Process and validation by the Information Security Section. The Information Security Section serves as the institutional entry point for AI tool consultation, onboarding, and approval coordination under this Policy.
- b) The Information Security Section maintains the OIST AI Tools Register as the authoritative institutional record of approved, conditionally permitted, and prohibited AI tools and services for OIST use. The Register may include tool-specific conditions, account requirements, approved usage boundaries, and other operational guidance necessary for safe and compliant use. The Register may be updated from time

to time to reflect changes in vendor terms, technical controls, institutional decisions, legal or regulatory considerations, and operational risk. Users and divisions must follow the current version of the Register. Approval of a tool or service applies only within the scope, conditions, account type, and institutional arrangement stated in the Register.

- c) The Information Security Section conducts security and data-protection assessments for externally provided AI systems and AI-enabled cloud services requiring onboarding or supplementary review under this Policy.
- d) Procurement, the Information Security Section and Legal/Compliance, as applicable, shall evaluate vendor information, controls and available assurances proportionate to risk.
- e) Category-level approvals issued by the CIO (e.g., open-source models running in sandboxed environments) are valid without requiring model-by-model review. For category-approved environments, notification to the Information Security Section is required at the level of the environment, service, or project as defined by the CIO, not for each individual use unless otherwise specified.

44.5.2 Incident and Risk Reporting

- a) **Reportable Incidents.** AI-related incidents or concerns that involve data leakage, security concerns, unapproved tool usage, repeated or harmful model behavior, significant bias in a consequential context, suspected copyright infringement, material non-compliance, or other meaningful institutional risk shall be reported to the Information Security Section, following the escalation flow in Annex G.
- b) **Routine Low-Consequence Errors.** Routine low-consequence output errors corrected through normal human review do not require formal incident escalation unless they indicate a broader issue or occur in a context involving restricted data, regulated activity, safety, consequential decision-making, or other elevated risk.
- c) **CIO and CISO Actions.** The CISO and CIO jointly determine containment, remediation, and communication actions.
- d) **Suspension and Remediation.** OIST may temporarily suspend affected tools, services, or categories until risks are addressed and mitigated.

44.5.3 Periodically and Annual Reviews

- a) AI governance reporting shall be prepared periodically using existing institutional records where practicable. Targeted input from divisions may be requested where additional context is needed.
- b) **Annual Report Contents.** The AI Governance Committee produces an annual report to the CEO/President summarizing:
 - (1) AI risk profile and trend analysis,
 - (2) audit findings (Annex E),
 - (3) training compliance rates (Annex F),
 - (4) policy or standards updates, and
 - (5) recommended institutional actions or investments.

44.5.4 Training and Records

- a) **Training Records.** HR and the Graduate School shall maintain completion records for training designated as mandatory under Annex F.
- b) **Project Documentation.** Relevant units shall retain proportionate documentation for AI systems requiring approval under section 44.2.7 or Annex D. Routine research experimentation within the exempt scope does not require separate documentation beyond normal research or project practices.

- c) **Retention and Storage.** All documentation must be retained in accordance with PRP 12 (Corporate Records Management) and stored in approved OIST repositories.

44.6 Annexes

Annex A – AI Risk Matrix and Illustrative Risk Cases

A.1 Purpose.

This Annex defines the method for classifying AI activities by risk level, identifying associated impacts, and prescribing required mitigation, oversight, and authorization pathways.

The matrix supports proportional governance and ensures that OIST members can safely adopt AI technologies in alignment with relevant PRP provisions and ethical and academic standards.

A.2 Scope

This Annex applies to all AI and Generative AI tools, models, and services used in OIST research, teaching, administration, and operations, including:

- External SaaS AI tools and cloud services.
- Locally hosted open-source models.
- Internally developed systems deployed beyond a research or sandbox environment.
- Embedded AI functions in commercial applications.
- Category-approved open-source models operating in controlled environments.

This Annex does not apply to research-only prototypes confined to the originating individual, lab or project team, using only Public or Internal information, and not deployed, shared or used for consequential operations, unless they involve regulated or otherwise elevated-risk conditions.

A.3 Risk Assessment Method

AI activities must be assessed across the following five dimensions:

1. Data Sensitivity: Classification and sensitivity of all data processed, generated, stored, or exposed by the AI activity, in accordance with PRP 17.3.9.
2. Operational Impact: Potential effect of error, hallucination, malfunction, unavailability, delay, or misuse on academic, research, administrative, or business operations.
3. Ethical and Reputational Risk: Potential for harm, bias, unfairness, misuse, deception, copyright infringement, or reputational damage arising from the AI activity or its outputs.
4. System Criticality: Degree to which the AI activity supports or influences essential academic, research, operational, safety, legal, regulatory, or compliance-related functions or decisions.
5. Scientific Integrity Risk: Potential for AI use to compromise the validity, reproducibility, interpretability, verifiability, or traceability of research or scholarly outputs, including through fabricated, misleading, biased, or insufficiently documented results.

These dimensions apply to all forms of AI use, including academic, research, administrative, business-management, workflow automation, and assistant tools.

The highest-rated dimension determines the overall risk level.

AI risk level is distinct from Information Classification. Personal, regulated, export-controlled, partner-restricted or contract-restricted information may increase the AI risk level or trigger additional requirements without changing the classification assigned under PRP 17.3.9.

Level	Description	Typical Use (Expanded)	Required Controls
Low	Uses public data; minimal consequence if incorrect	Grammar-checking, translation, meeting assistants for public content, generic scheduling helpers	Use within a current Register entry or category-level approval; normal user validation.
Medium	Uses Internal or limited research/business data; moderate reputational or operational impact	Summaries of internal memos, course prep, business analytics on Internal data, workflow-assistant tools, internal meeting summarizers	Use within the current Register conditions with proportionate human review. No separate user assessment is required unless the proposed use exceeds the approved scope or materially increases risk.
High	Uses Confidential, sensitive, or important research or business data	HR/Finance documents, grant materials, business planning data, internal decision-support tools	Use only where the Register and applicable approval cover the information classification and use case. The required risk review and authorization shall be completed through the approval process; no duplicate user assessment is required for use within the approved scope.
Critical	Uses personally identifiable, medical, contractual, regulated, or partner-restricted data	Clinical, partner-provided, contract-restricted, export-controlled, legally restricted, or other regulated information or AI-related artifacts, including where applicable associated outputs or technical materials	Prior CIO authorization of the specific service and use case following documented risk review and implementation of required safeguards.

Additional Research and Regulated-Use Considerations

For research, contractual, export-controlled, dual-use, partner-restricted, or otherwise regulated contexts, risk assessment must consider not only source data classification but also, where applicable:

- model weights, checkpoints, and training artifacts;
- code, algorithms, workflows, and technical methods;
- prompts, retrieval content, and generated outputs;
- access by collaborators, contractors, visitors, or other persons where deemed-export or equivalent restrictions may apply; and
- sponsor, partner, contractual, publication, or jurisdiction-specific restrictions.

A research activity shall not be treated as low risk solely because it is exploratory, research-only, or conducted in a sandbox environment if export-control, partner-restriction, legal, contractual, national-security, safety, or equivalent regulated-use concerns may apply. In such cases, consultation with the appropriate institutional functions is required.

Copyright Risk

OIST shall assess available platform copyright safeguards and contractual assurances. Users are not expected to verify a provider's training data or provenance but remain responsible for the appropriate use of outputs and compliance with applicable citation, attribution and licensing requirements. Suspicious or potentially infringing outputs must be reported to the Information Security Section.

A.4 Illustrative Risk Cases

The following examples illustrate common AI-related risks across academic, research, administrative, business, and operational contexts:

- **Hallucination and False Output**
→ Validate against authoritative or primary sources before use; do not rely on AI for factual accuracy without human review.
- **Data Leakage to External Models**
→ Do not enter non-public information into external AI tools unless permitted by the OIST AI Tools Register. Critical information requires prior CIO approval under section 44.3.1(e).
- **Bias, Fairness, and Discrimination**
→ Perform periodic bias testing and human validation for medium- or high-risk activities, especially in HR, admissions, or evaluation contexts.
- **Shadow AI (Unapproved Tool Usage)**
→ Use of unvalidated AI tools may create significant security, privacy, contractual, or compliance risk and must be reported or consulted on promptly through the established pathway. Good-faith early reporting is encouraged so that risk can be assessed and, where appropriate, regularized or contained.
- **Model Drift**
→ Operational systems must be reviewed at least annually and following material changes, with controls adjusted as necessary.
- **Open-Source Sandbox Models**
→ Models operating in an OIST-controlled environment may be covered by a category approval and are subject to the documentation requirements stated in that approval.
- **Business-Process Automation or Workflow Errors**
→ AI assistants may incorrectly generate email replies, scheduling actions, or business logic; these require human review and approval.
- **Decision-Support Misinterpretation**
→ AI-generated analytics (forecasting, KPI dashboards, budget estimates) may influence decisions incorrectly if trusted without validation.
- **Reputational Harm**
→ AI-generated public content, summaries, or translations may be inappropriate, misleading, or misaligned with institutional messaging.
- **Copyright and Licensing Risk**
→ AI outputs may unintentionally reproduce copyrighted text or media; suspicious outputs must be reported to the Information Security Section.
- **Security Vulnerabilities**
→ Adversarial prompts, poisoned datasets, or manipulated models may produce harmful or unauthorized outputs.

- **External Partner or Vendor Misconduct**

→ AI systems provided by external vendors may fail to adhere to contractual obligations, data-retention rules, or training-data restrictions.

A.5 Review Cycle

The AI Risk Matrix shall be reviewed and updated at least annually by the AI Governance Committee, informed by:

- Audit findings (Annex E).
- Incident reports (Annex G).
- Changes in technology or regulation.
- Emerging ethical or security concerns.
- Feedback from Students, ORC, faculty, and divisions.

Updates are approved by the CIO and incorporated into this Policy or companion guidance, as applicable.

Annex B – Information Classification and Data Handling Summary

B.1 Purpose

This Annex links OIST’s Information Classification Framework (PRP 17.3.9) to requirements specific to Artificial Intelligence (AI) and Generative AI (GenAI) use.

It defines the handling rules for each classification level to ensure safe, compliant, and responsible AI processing.

This Annex addresses only data-classification and data-security requirements. Ethical, operational, academic, and governance considerations related to AI are covered separately in Annex A, Annex D, and other sections of PRP 44.

B.2 Classification Levels and Examples

Level	Definition	AI Handling Requirement
Public	Information intended for public release	May be used with approved AI tools in accordance with applicable register conditions
Internal	For OIST members only	May be used in on-premises or vetted AI services
Confidential	Could impact individuals or operations if leaked	Encrypt; use only within OIST-approved secure AI platforms
Critical	Legal, personal, or nationally sensitive	May be processed only in an OIST-approved service and use case with prior CIO authorization following documented risk review and required safeguards.

Additional Clarifications

- When datasets of different classifications are combined, the highest classification level prevails.
- For a category-level approval, information may be used up to the maximum classification and within the conditions stated in the Register. No separate review is required where use remains within those conditions.
- For GenAI tools supplied by vendors, contractual terms must address permitted processing and training use, retention and deletion, storage and access, and applicable copyright or provenance safeguards. Submitted OIST information must not be used to train vendor foundation models unless specifically authorized.

B.3 Handling Rules

- Use the minimum necessary data for prompting, fine-tuning, or evaluation.
- Approved services processing Internal, Confidential or Critical information must provide encryption in transit and at rest, as appropriate to the service architecture and risk.
- Remove personal identifiers and metadata whenever possible (pseudonymization or anonymization).
- Non-public information may be submitted only where permitted by the OIST AI Tools Register. Critical information requires prior CIO approval under section 44.3.1(e).
- Users are not expected to verify copyright provenance; responsibility rests with approved platforms and institutional safeguards.

- Do not upload Internal, Confidential, or Critical information to any AI tool unless the tool is explicitly approved for such use.
- All handling of personal data must comply with PRP 17, Japanese privacy regulations, and OIST contractual obligations.
- If unsure, users must consult the Information Security Section before processing data with AI tools.

Annex C – AI Solution Approval Process and Approved Tools Register

C.1 Purpose

This Annex defines the standardized process for evaluating, approving, and tracking *externally provided* AI systems, services, plugins, and AI-enabled applications used at OIST.

It ensures consistency in risk assessment, security validation, procurement controls, and lifecycle management.

Individual approval and registration are required for a new or materially changed externally provided AI tool or shared AI service only where it is not already covered by a current Register entry or category-level approval, or where the proposed use exceeds those approved conditions. Internally developed research prototypes follow Annex D.

C.2 Approval Workflow

Before using an external AI tool or AI-enabled service, the Planner, Provider or requesting unit shall check whether the service and intended use are covered by the OIST AI Tools Register or an applicable category-level approval.

If covered, no new approval, validation ticket or user risk assessment is required, provided the use remains within the approved scope, conditions and information-classification limits.

If not covered, the requesting unit shall consult the Information Security Section, and the following validation workflow applies.

1. **Validation Ticket Creation.**

The Information Security Section creates a validation ticket in the IT service management system.

2. **Security and Risk Assessment (CISO)**

1. Security controls
2. Data handling
3. Copyright-related controls and assurances
4. Vendor obligations
5. Alignment with AI Risk Matrix (Annex A)

3. **Committee Review**

Committee review is required for high-risk, institution-wide, novel, disputed, or policy-impacting cases, and may be omitted for routine low- or medium-risk cases handled through delegated workflow.

4. **Approval (CIO).**

The CIO issues approval, conditional approval, or rejection. Approval outcomes shall be reflected in the OIST AI Tools Register as applicable.

5. **Notification.**

The requesting unit is notified with conditions, limitations, and authorized data-classification levels.

C.3 Category-Level Approvals

To avoid bottlenecks and support research innovation, the CIO may issue category approvals for classes of low-risk tools, such as:

- Open-source models executed in OIST-controlled sandboxed or air-gapped environments
- Local inference environments using vetted model families (e.g., HF or GGUF models)
- Tools with identical risk profiles or identical technical controls

For category-approved environments, notification is required at the level of the environment, service, or project as defined by the CIO, not for each individual use unless otherwise specified.

C.4 Approved Tools Register

The Information Security Section maintains the authoritative OIST AI Tools Register as the user-facing record of approved, conditionally permitted, and prohibited AI services and service categories for OIST use.

The Register shall provide sufficient information for users to determine whether a service and intended use are permitted. This shall include, as applicable, the approval status, whether an OIST-managed account or tenant is required, the maximum approved information classification, the approved scope of use, and any principal conditions or restrictions.

An approval may apply to a named service or to a category of services where the applicable controls and risk conditions are materially equivalent. Routine changes to product names, model versions, or features do not require separate Register entries unless they materially affect data handling, security, contractual conditions, integrations, or risk.

The Information Security Section, in cooperation with IT and other relevant functions, shall maintain the supporting technical, contractual, and risk-assessment records. These supporting records do not constitute a separate AI Tools Register.

The Register must be accessible to OIST members through an institutional channel designated by the IT Division.

C.5 Procurement Controls

The Procurement Section, the Information Security Section and Legal/Compliance, as applicable, shall assess vendor controls and contractual protections proportionate to risk. These may include permitted processing and training use, retention and deletion, storage and access, security obligations, assurance or audit evidence, copyright-related safeguards and incident obligations.

C.6 Review and Expiration

Register entries and category-level approvals shall be reviewed at least annually and earlier where a material change may affect the approval conditions or risk. Routine vendor, feature, or model updates do not require separate review where the existing approval conditions remain applicable.

C.7 Removal or Suspension

The CIO may suspend or revoke approval if a tool, version, vendor practice, or model category is found to:

- present unacceptable security or privacy risk,
- violate contractual commitments,
- leak or retain OIST data,
- exhibit harmful or biased behavior, or
- fail to meet Annex D standards.

Users must comply immediately with stop-use directives.

Annex D – AI Development and Deployment Standards

D.1 Purpose

To define the minimum technical, ethical, and operational standards required for the secure design, development, deployment, and maintenance of AI systems at OIST.

These Standards ensure compliance with other PRP provisions, academic integrity principles, and contractual obligations.

D.2 Scope

This Annex applies where an AI system:

- (a) is deployed or shared beyond the originating individual, lab or project team;
- (b) processes Confidential or Critical information;
- (c) supports consequential operational decisions or an institutional service;
- (d) is developed by a contractor or partner for institutional use; or
- (e) involves regulated or otherwise elevated-risk conditions.

A research-only prototype confined to the originating lab or project team, using only Public or Internal information, and not deployed, shared or used for consequential operations is exempt from formal approval while remaining subject to the general requirements of this Policy.

D.3 Principles

The Standards are built on the following principles:

- **Transparency**
- **Security-by-Design**
- **Accountability**
- **Human Oversight (HITL)**
- **Fairness and Non-Discrimination**
- **Data Integrity and Privacy**

These principles apply to every stage of the AI lifecycle.

Proportionate Control. Controls should be commensurate with the intended use, data sensitivity, deployment context, and potential institutional impact. Low-risk research experimentation should be supported through lighter controls, while operational, regulated, or higher-risk uses require stronger design, validation, and governance measures.

D.4 Design and Development Requirements

The authoritative OIST AI Development and Deployment Standards are maintained separately by the CIO Office. They operationalize and supplement this Policy and must remain consistent with it. If a conflict arises, this Policy controls until formally amended.

At a minimum, AI systems subject to this Annex must have documented purpose and ownership, appropriate data-governance controls, proportionate validation and testing, defined human-oversight arrangements where applicable, logging and monitoring appropriate to risk, and a managed process for change, incident response, and retirement.

Where third-party APIs, hosted models, or external development partners are used, the requesting or owning unit must ensure that contractual, security, data-handling, and dependency risks are addressed before operational deployment.

D.5 Governance

Compliance with the Standards is overseen by the CIO and CISO and administered by the Information Security Section, with advisory oversight from the AI Governance Committee.

Deployment of AI systems beyond a lab or research environment requires **CISO review** and **CIO approval**, as described in section 44.3.2.

Routine internal development and experimentation that remain within the exempt scope of D.2 do not require formal deployment approval, but must still be conducted responsibly and consistently with applicable OIST policy, data-handling rules, and any companion guidance issued under this Policy.

D.6 Enforcement

The CIO may suspend any AI system, tool, or development activity if security, legal, contractual, ethical, or operational risks are identified; vendor practices materially change; unacceptable model behavior is detected; or the system fails to comply with Annex D requirements. Suspension may remain in place until the identified risk is assessed, appropriate remediation is completed, and continued use is reauthorized where applicable.

Annex E – Operational Risk Management and Audit Checklist

E.1 Purpose

This Annex defines the requirements for verifying compliance with PRP Chapter 44 and provides the standard checklist used by the Information Security Section to perform AI-related audits.

Audits confirm that AI activities comply with:

- the AI Risk Management Framework (Annex A),
- Information Classification rules (Annex B),
- Tool Approval and Registration (Annex C),
- Development and Deployment Standards (Annex D),
- Training and awareness obligations (Annex F), and
- Incident escalation requirements (Annex G).

E.2 Checklist Summary

As applicable to the activity reviewed, audits may verify the following elements:

- 1. Tool Approval Verified**
 1. External AI tools validated by the Information Security Section.
 2. Tool or model listed in the *OIST AI Tools Register* (Annex C).
 3. Category approvals applied correctly for low-risk sandboxed open-source models.
- 2. Risk Classification Documented**
 1. AI Risk Matrix assessment completed where required under section 44.2.5.
 2. Highest classification dimension clearly recorded.
- 3. Data Handling Validated**
 1. Data classification matched to Annex B requirements.
 2. No inappropriate use of external tools with Internal, Confidential, or Critical data.
 3. Encryption and anonymization applied where required.
- 4. Development Standards Applied**
 1. Systems follow the *AI Development and Deployment Standards* (Annex D).
 2. Deployment beyond sandbox environments reviewed by the CISO and approved by the CIO, where required under section 44.3.2.
 3. Documentation and testing evidence available.
- 5. Training Records Verified**
 1. Required Responsible AI training completed under Annex F.
 2. Developer/Provider training completed for relevant personnel.
- 6. Incident Reporting Tested**
 1. Reporting channels functional and known to staff.
 2. Past incidents logged and escalated per Annex G.
 3. Remediation tracked and closed.
- 7. Documentation and Records Compliance**
 1. All required documentation is retained under PRP 12 (Corporate Records Management).

E.3 Audit Process

(a) The Information Security Section shall operate a risk-based audit program, including an annual review and ad hoc audits directed by the CIO or CISO. Audits may use document review, interviews, system checks, test scenarios or representative sampling.

(b) Material findings shall be reported to the AI Governance Committee, which reviews the results, recommends corrective actions and monitors remediation. Formal corrective actions and stop-use directives remain with the authorities assigned under this Policy.

(c) Significant or repeat non-compliance may result in tool suspension, further review, or escalation to the CEO/President.

Annex F – Training and Awareness Framework

F.1 Purpose

This Annex defines the training and awareness requirements necessary to ensure that all OIST members understand Responsible AI principles and can apply them appropriately in their work, research, and study. Training supports safe innovation and reduces the risk of misuse, data leakage, or unethical AI behavior.

F.2 Training Program Structure

- **Baseline Awareness:** All OIST members; covers policy awareness, responsible use, information classification and prompt hygiene.
- **Developer Module:** Secure coding, bias mitigation, model documentation.
- **Manager Module:** Risk assessment, procurement standards, audit expectations.
- **Required Modules:** Formal completion may be required according to role, system access, information handled or risk exposure.

F.3 Delivery and Frequency

Baseline awareness shall be provided during onboarding where relevant. Refresher training shall occur periodically and when material changes in technology, policy or risk arise. Role-specific training shall be provided when standards or responsibilities materially change.

F.4 Responsibility

- **HR Division** maintains centralized training completion records.
- The **Graduate School (GS)** manages student training components and academic integrity alignment.
- The **IT Division / Information Security Section** provides technical content, risk-awareness materials, and incident-reporting guidance.
- The **AI Governance Committee** reviews training outcomes annually and recommends improvements.

F.5 Review Cycle

Training content and delivery methods are reviewed annually based on:

- audit findings (Annex E),
- incident trends (Annex G),
- user feedback, and
- emerging AI capabilities or risks.

Annex G – Consultation and Escalation Flow

G.1 Purpose

This Annex defines OIST’s consultation and escalation pathway for Artificial Intelligence (AI) activities. It ensures that users receive timely guidance, incidents are reported consistently, and systemic risks are escalated to the appropriate governance bodies.

G.2 Consultation

Desk Operated by Information Security Section. Provides advice on tool approval, prompt design, data use, and incident reporting.

G.3 Escalation Process

All AI-related concerns follow the escalation pathway below:

1. User → Consultation Desk (Information Security Section)

Users may contact the Consultation Desk for:

- clarification on approved tools or category approvals;
- questions about data use, disclosure, or prompt design;
- concerns about unapproved tools or unexpected tool behavior; and
- formal reporting of incidents involving data leakage, security concerns, significant bias in consequential contexts, unauthorized tool use, suspected copyright issues, or other material institutional risk.

Routine low-consequence output errors that are detected and corrected through normal human review do not require formal incident escalation unless they indicate a broader issue or occur in a higher-risk context.

2. Consultation Desk → Chief Information Security Officer (CISO)

CISO validates severity, evaluates security and privacy implications, and determines containment steps.

3. CISO → AI Governance Committee *(if systemic or high impact)*

Trigger conditions include repeated incidents, critical data exposure, vendor failures, ethical implications, or model-behavior anomalies.

Committee reviews institutional impacts and provides coordinated recommendations.

4. AI Governance Committee → CIO

CIO approves required actions, including temporary suspension, category-level restrictions, or policy clarification.

CIO may issue campus-wide communications if needed.

5. CIO → CEO/President *(if policy-level or institutional risk)*

Escalation occurs when:

1. policy amendments are required,
2. institutional risk exceeds acceptable thresholds,
3. reputational or legal exposure is significant.

G.4 Reporting and Transparency

- The Information Security Section prepares **periodic summaries of** AI-related inquiries, incidents, and mitigation actions.
- These summaries are provided to the **AI Governance Committee** and **Executive Leadership**, informing audits (Annex E), training updates (Annex F), and policy revisions.
- Significant incidents are reported immediately to the CIO and included in the annual report to the CEO/President.

G.5 Good-Faith Reporting Protection

- Users who report incidents, concerns, or unapproved tool usage in good faith are **protected from adverse action**.
- OIST encourages early reporting to prevent harm and improve safeguards.

Annex H – Governance Roles and Committee Terms of Reference

H.1 Purpose

This Annex defines the institutional governance structure for Artificial Intelligence (AI) at OIST, consistent with national CAIO (Chief AI Officer) best practices and OIST’s internal governance framework. It establishes the roles, responsibilities, and operational procedures required to oversee AI risk, compliance, ethics, and institutional readiness.

H.2 Structure

OIST implements a tiered AI governance model comprising:

- **Chief Information Officer (CIO) acting as Chief AI Officer (CAIO).** Provides institutional leadership, strategic direction, and ultimate accountability for all AI-related activities.
- **Chief Information Security Officer (CISO).** Oversees security implementation, technical risk assessments, and operational coordination across divisions.
- **AI Governance Committee.** A cross-functional advisory and oversight body representing:
 - Faculty.
 - Graduate School (GS).
 - Human Resources (HR).
 - Information Technology (IT Division).
 - Legal and Compliance functions.
 - Vice Provost or delegate, representing institutional academic and research leadership.
 - OIST Researcher Community (ORC) representative (added per PRP 4.5).
 - Additional experts invited as needed.

This structure ensures broad institutional coverage and integrates researcher, administrative, and technical perspectives.

H.3 Mandate

The AI Governance Committee shall:

- Review and recommend updates to this Policy, the AI Risk Matrix, and associated annexes.
- Review and recommend approval of high-risk, institution-wide, novel, disputed, or policy-impacting AI systems and categories in accordance with Annex C. Routine low- and medium-risk cases may be handled through delegated workflow.
- Review audit findings (Annex E) and monitor compliance trends.
- Coordinate and oversee training and awareness activities (Annex F).
- Identify institutional risks, emerging technologies, and regulatory developments relevant to AI.
- Recommend policy, standards, or governance updates to the CEO/President via the CIO.
- Support transparency and institutional communication about AI use and risks.

The AI Governance Committee is an advisory and oversight body. It does not replace the approval authorities assigned to the CIO, CISO, Procurement Section, or other institutional functions under this Policy.

H.4 Meetings and Reporting

The Committee meets periodically, normally quarterly, with additional sessions convened in response to major incidents, systemic risks or significant technological changes.

The Committee reports:

- Periodically to the CIO, summarizing incidents, tool adoption, risk trends, and audit findings.
- Annually to the CEO/President, providing a consolidated AI Governance Report including risks, mitigation actions, training compliance, and policy update recommendations.

H.5 Secretariat and Records

The IT Division serves as the Secretariat and:

- Prepares agendas and minutes.
- Maintains the official records of Committee activities.
- Manages documentation in accordance with PRP 12 (Corporate Records Management).
- Coordinates with HR, GS, and divisions for reporting and data collection.

H.6 Quorum

A quorum for formal Committee actions requires:

- At least 50 percent of standing members, including the CIO or designated delegate.
- Remote attendance is permitted for quorum purposes.

H.7 Review Cycle

The Committee shall review its Terms of Reference every two years, or earlier if required by regulatory changes, internal incidents, or updates to national CAIO guidelines.

End of Policy