

A PRACTICAL COMPANION FOR INSTITUTIONAL TEAMS

AI Operations and Assurance

Turn policy into arrangements people can follow. These HumanSkills recommendations help institutional owners decide whether a use may begin, what evidence is needed and how it will remain accountable. Adapt the controls to the actual use and existing institutional authority.

Risk and approval

What exactly are we authorizing?

AI security controls

Can the boundaries hold when the system behaves unexpectedly?

Vendor evidence and commitments

What can the supplier demonstrate and commit to?

Testing and release

What evidence justifies a live pilot or wider use?

Monitoring and incident response

Who notices a problem and who can act?

Use the guidance to establish the arrangements. Use the working records to document local decisions. Each chapter connects to the model policy; the institution retains its authority and responsibility.

The editable Word companion contains five working records. Store completed documents in your own approved systems.

PUT THE ARRANGEMENT IN PLACE

Risk and approval

Classify the use in its actual setting. The same tool can support routine drafting or a consequential decision. A vendor subscription is not approval for every possible use.

Responsible roles: Service owner with the responsible approval authority and specialist reviewers.

Describe the complete use

Record the purpose, people affected, tool and model, information sources, connected systems, intended outputs and any actions the system can take. Include AI already embedded in existing services.

Screen for consequences

Ask whether an error could affect admission, financial aid, employment, grades, health, safety, access to a service or a person's rights. Consider scale, vulnerable groups, sensitive information, difficulty of correction and the ability to act without a person.

Choose the review route

Routine assistance within an already approved tool, information class and use may follow ordinary staff or faculty authority. New institutional deployments, protected information or integrations need the appropriate operational review. Consequential uses and autonomous actions need the designated institutional authority plus relevant specialists before a live pilot.

Separate experimentation from live service

Use an isolated test environment and authorized test information. Use synthetic information where suitable and check that it does not reproduce personal records. Keep production credentials and actions disabled. A pilot with real people or records needs its own scoped approval.

Record the decision

Approve, approve a bounded pilot, request changes or decline. Name the conditions, prohibited actions, human reviewer, expiry or review date, stop authority and evidence needed for wider use. Legal or institutional prohibitions cannot be waived through a risk acceptance form.

Leave with: A scoped approval with conditions, an owner and a review date.

Model policy sections 3, 4, 6, 9.

Risk and approval record

Illustrative campus example

An admissions team may draft a public event invitation with an approved tool under ordinary rules. Ranking applicants changes the review route because it can affect opportunity. Connecting the system to send decisions or change records adds further authority that must be separately considered.

Decide locally

- Name the actual offices and delegated authorities for each route.
- Define which routine uses are covered by standing approval and which require registration.
- Set a review timetable, escalation path and treatment of urgent requests without bypassing required controls.

Institution _____ Owner _____

Date and version _____ Decision authority _____

Purpose and scope

Identify the use, owner, people affected, systems, data and intended benefit.

Risk and review route

Record the consequential effects, action authority, review category and the reasons for it.

Conditions and evidence

List permitted and prohibited uses, specialist findings, unresolved issues and pilot limits.

Approval and renewal

Record the decision, authorized approver, stop authority, review date and change triggers.

PUT THE ARRANGEMENT IN PLACE

AI security controls

Protect the service around the model: accounts, information stores, interfaces, connected tools and logs. Match the testing and controls to what the service can actually reach and do. Apply the institution's existing security standards and establish the requirements below before release.

Responsible roles: Information security and IT with the service and data owners.

Map and limit access

Document information flows, storage, retrieval sources and recipients. Enforce access separately for each user and service identity. Restrict retrieval and connected actions to the approved task. Recheck permissions when roles or integrations change.

Protect credentials and interfaces

Use institution-managed service identities and secret storage. Keep passwords, tokens and keys out of prompts, shared documents and source repositories. Assign credential owners, rotation and revocation arrangements. Authenticate interfaces, enforce authorization, validate tool inputs and set rate and spending limits.

Treat external content as untrusted

Retrieved pages, email, documents and tool results may contain instructions intended to redirect the system. Test whether those instructions can change its task, obtain protected information or trigger unauthorized actions. Keep critical permissions in system controls and authorized approval steps.

Test information boundaries

Use controlled test records to attempt cross-user access, unauthorized retrieval, leakage through outputs and inappropriate storage in logs. Verify deletion and retention settings, access revocation and the behavior of connected services. Do not use live sensitive records merely to test a boundary.

Monitor and contain

Monitor unusual access, attempted privilege changes, repeated tool calls, unexpected recipients and spending or usage spikes. Name who receives alerts and can disconnect the service. Protect logs with limited access and a justified retention period; avoid unnecessary prompt or record content.

Leave with: A verified control record for the approved environment.

Model policy sections 5, 9, 11.

AI SECURITY CONTROLS

Minimum security requirements

These requirements supplement the existing control checks. Apply them to the actual deployment and document which controls are provided by the institution and which by the supplier. Required protections cannot be assumed from a tool approval.

Apply the institutional baseline

For services handling protected information or acting within institutional systems, require multifactor authentication for human access, managed task-limited service identities and timely removal of access. Encrypt protected information in transit and at rest. Control credential and key storage, rotation and revocation. Apply institutional vulnerability-remediation requirements, protected logging and tested recovery or fallback arrangements. Record evidence and any permissible alternative under the established exception process.

Handle model outputs safely

Treat outputs as untrusted input. Validate tool arguments, encode displayed content for its context and use parameterized database operations. Execute generated code or commands only within an approved, constrained process with appropriate isolation. Test attempts to turn an answer into unauthorized code, a database operation or a connected action. Human approval does not replace application security controls.

Protect knowledge sources and changes

Restrict who and what can change source documents, retrieval indexes, training data and model components. Record provenance and versions; review changes for unauthorized or malicious content. Preserve each user's access limits during retrieval. Test removal of a compromised source and restoration of an approved version, including indexes and caches. Verify removal from retrieval, and reassess affected models and downstream uses before restoring service.

Record the control, responsible owner, applicable institutional standard, test or supplier evidence, reviewer and result. Where a required control cannot be demonstrated, resolve the finding or restrict the use before release.

AI security controls record

Illustrative campus example

A document in an advising knowledge base instructs an assistant to export student information to an external address. The test succeeds only if the system preserves access and action limits and the incident can be detected and contained. A polite refusal in one demonstration is insufficient evidence.

Decide locally

- Set data-specific controls, logging limits and security test depth with responsible specialists.
- Identify the approved identity, secret-management and monitoring services.
- Define the failures that prevent release and the evidence needed to close each finding.

Institution _____ Owner _____

Date and version _____ Decision authority _____

Environment and boundaries

List identities, information sources, recipients, tools and actions permitted.

Controls and owners

Record the institutional baseline, authentication, access removal, encryption, credentials, keys, remediation, logging and recovery controls, with owners.

Tests and findings

Link evidence for access boundaries, safe outputs, controlled knowledge-source changes and recovery. Record failures, reviewer and corrections.

Release conditions

Identify mandatory blockers, the reviewer, alert recipients and the authority to disconnect.

PUT THE ARRANGEMENT IN PLACE

Vendor evidence and commitments

Ask for evidence specific to the product, configuration and service tier being purchased. A corporate assurance statement does not establish how your institution's information or use will be handled.

Responsible roles: Procurement with security, privacy, accessibility, counsel and the service owner.

Trace the information

Request a data-flow description naming processing and storage locations, subcontractors, information categories, retention, backup treatment and deletion arrangements. Establish whether inputs, outputs or feedback may be reused for training or other purposes, and which settings and contractual terms govern this.

Examine the relevant assurance

Use the institution's established vendor assessment, such as HECVAT or an equivalent, and supplement it for the AI use. Verify product, configuration, service tier, subcontractors, assessment scope and evidence date. Examine security controls, test findings and vulnerability handling. A completed questionnaire or assurance report is not approval. Test accessibility in the actual workflow with the responsible office.

Confirm operating responsibilities

Identify notification contacts and commitments for incidents, material model or feature changes, subcontractor changes and service interruptions. Determine how the institution can investigate, export relevant records, obtain support, challenge an output and disable new capabilities.

Plan continuity and exit

Establish export formats, termination support, deletion evidence, treatment of backups and the ability to continue essential services. Examine dependencies on model providers and third-party components, along with rights to institutional content and outputs.

Resolve findings before approval

Record each requested item, supplied evidence, review finding, condition and owner. Put necessary commitments in the agreement. Time-limit permissible exceptions through the authorized process; missing mandatory protections block the affected use. Reopen review when the actual service changes.

Leave with: A documented supplier review tied to the actual product and contract.

Model policy sections 4, 5, 10, 12.

Vendor evidence and commitments record

Illustrative campus example

A supplier says customer data are not used for training. The institution checks whether that statement covers its exact subscription, feedback features and subcontractors, then records the relevant contract term and configuration. Marketing language alone does not close the question.

Decide locally

- Use institutional procurement, legal and security requirements as the approval baseline.
- Specify which evidence must be independently verified or contractually enforceable.
- Set renewal checkpoints and the events that require a fresh supplier review.

Institution _____ Owner _____

Date and version _____ Decision authority _____

Supplier and service

Record the product, subscription, configuration, use and accountable institutional owner.

Evidence and finding

Link the institutional vendor assessment or HECVAT, product-specific supporting evidence, scope and date, reviewer, findings and unresolved items.

Contract and conditions

Identify binding commitments, unresolved items, permitted exceptions and limits.

Change and exit

Record notification contacts, renewal date, transition arrangements and deletion evidence required.

PUT THE ARRANGEMENT IN PLACE

Testing and release

Define success and unacceptable failure before testing. Evaluate the complete service in the conditions where people will use it, including human review, connected actions and recovery.

Responsible roles: Service owner and designated approver with independent challenge proportionate to consequence.

Set acceptance criteria

Choose realistic tasks, the existing service baseline and measurable outcomes. Include accuracy, unsupported answers, accessibility, harmful disparities, time to human help, privacy, cost and workload where relevant. Set thresholds locally with the people responsible for the consequences.

Build representative cases

Cover ordinary use, ambiguous requests, missing or outdated information, relevant languages and accessibility needs. Include mistakes people are likely to make. Keep test data authorized, and document limitations in population or workflow coverage.

Challenge the boundaries

Test misleading external instructions, unauthorized data requests and attempts to exceed connected permissions. For agents, include wrong recipients, duplicate actions, partial completion, delegated activity, tool failure, spending limits and the ability to stop and reconcile effects.

Test the human arrangements

Check that reviewers have enough information, time and authority to disagree. Walk through an appeal, service outage and correction with actual role holders. Confirm an accessible alternative and fallback for essential services.

Authorize a bounded release

Record results, failures, corrections, remaining uncertainty and the configuration tested. Apply the deployment-blocking conditions in this guide. A pilot decision names users, duration, limits, monitoring and stop conditions. Expansion requires pilot evidence and an explicit decision; material changes trigger retesting.

Leave with: A release decision supported by results and known limitations.

Model policy sections 4, 6, 9, 12.

TESTING AND RELEASE

Conditions that block deployment

Apply these conditions before a live pilot, wider deployment or material expansion. A favorable average test result does not offset a failed mandatory safeguard. Keep any experiment isolated from affected people, protected records and production actions until the conditions are met.

Unauthorized access or action

Do not release when testing shows access to information or actions outside approved authority. Restrict the scope, fix the boundary and retest.

Unresolved information or contract protections

Do not release while required data-handling protections, processing authority or contractual commitments remain unresolved. Obtain evidence and the necessary approval for the actual use.

No accountable stop authority

Do not release without a named owner, an effective way to stop or disconnect the service, and a person authorized and prepared to use it. Demonstrate the stop before live use.

No workable essential-service fallback

Do not release an essential service without a tested fallback and a responsible operator. Include the effects of partial completion and recovery of affected records.

Other critical security findings

Resolve other critical security findings before release, or document a permissible, authorized and time-limited exception with demonstrated compensating safeguards. Exceptions cannot waive applicable requirements or the blocking conditions above.

For each applicable condition, record the evidence, result, reviewer and date. Use the release decision field in the existing working record. The designated authority records the approved scope; missing evidence leaves the affected condition open.

Testing and release record

Illustrative campus example

An advising assistant answers common questions well but confidently gives outdated financial-aid deadlines. The team restricts those answers until authoritative sources, escalation and test results meet its release conditions. An average success rate cannot cancel an unacceptable high-consequence failure.

Decide locally

- Agree measurable acceptance criteria and non-negotiable failure conditions before seeing results.
- Choose who provides independent challenge and who can accept remaining risk within their authority.
- Set pilot size, duration, expansion evidence and the configuration changes requiring retesting.

Institution _____ Owner _____

Date and version _____ Decision authority _____

Use and acceptance criteria

State the version, environment, intended users, baseline, thresholds and unacceptable failures.

Cases and results

Record representative and adversarial cases, expected outcomes, actual results and evidence locations.

Readiness and limitations

Document human review, accessibility, fallback, corrections and remaining uncertainty.

Release decision

Record each applicable deployment-blocking condition and evidence of closure, any permissible exception, the approver, scope, monitoring and retest triggers.

PUT THE ARRANGEMENT IN PLACE

Monitoring and incident response

Approval begins an operating responsibility. Establish how the institution will notice changing performance, protect affected people and decide whether to continue, change, restrict or retire the use.

Responsible roles: Service owner with existing incident, security, privacy and specialist teams.

Choose signals that matter

Track the outcomes and failure conditions defined for release. Add complaints, appeals, access barriers, disparate effects, human overrides, unusual actions, workload, cost and service interruptions as appropriate. Name each signal's owner, review frequency, threshold and response.

Review changes and exceptions

Maintain the service inventory, current configuration and approval conditions. Review material model, data, feature, vendor or permission changes before expanding exposure. Trend exceptions and unresolved issues, and require renewed evidence when the use or its effects change.

Contain and protect

Make reporting easy for staff and affected people. Route AI failures through existing incident arrangements with the relevant specialists. Pause harmful actions, revoke affected permissions, preserve necessary evidence and activate the fallback. Address urgent service and safety needs without waiting for a complete diagnosis.

Investigate and repair

Identify affected people, records and downstream actions. Determine what was authorized, what happened and what failed. Correct records and services, provide an accessible human route and assess required notifications with responsible offices. Protect confidentiality and avoid copying sensitive material into an unrestricted incident log.

Authorize recovery and learn

Test the correction and reconcile partial or duplicate actions. The designated authority records restart conditions and remaining limits. Complete a review of causes, control changes, training and related uses. Report material risk and remediation to leadership; route board assurance through the president and established channels.

Leave with: A working monitoring and response plan with accountable restart authority.

Model policy sections 6, 9, 11, 12.

Monitoring and incident response record

Illustrative campus example

A scheduling agent changes appointments before a connected service fails. Stopping it prevents further actions, but recovery also means identifying which appointments changed, restoring or confirming them and informing affected people. Restart follows a tested correction and an authorized decision.

Decide locally

- Assign monitoring intervals to the consequence and rate of change, with immediate alerts for critical triggers.
- Name stop, notification, correction and restart authorities and their deputies.
- Schedule a tabletop exercise, retain only necessary evidence and review the process after a material incident.

Institution _____ Owner _____

Date and version _____ Decision authority _____

Monitoring arrangement

Record signals, thresholds, owners, review intervals and alert destinations.

Incident and containment

Capture what happened, affected uses, immediate protections, evidence location and incident lead.

People and correction

Identify affected people or records, specialist decisions, communications and remedial actions.

Restart and learning

Record test evidence, restart authority, remaining conditions and follow-up owners and dates.

SOURCES AND COMPANION RESOURCES

Sources for the guidance

These are HumanSkills implementation recommendations. The sources inform the operating approach; they do not establish a universal compliance checklist or certify a deployment. Confirm current local requirements with the responsible specialists.

NIST AI Risk Management Framework Generative AI Profile

Voluntary risk guidance informing evaluation, third-party management and ongoing review.

OWASP Prompt Injection

Technical guidance informing tests of untrusted content and permission boundaries.

OWASP Excessive Agency

Technical guidance informing limits on connected tools, permissions and autonomous actions.

NIST Cybersecurity Framework 2.0

Outcome-based guidance for connecting AI services to institutional cybersecurity controls; it does not prescribe a universal implementation.

SOURCES AND COMPANION RESOURCES

Sources and related guides

OWASP Improper Output Handling

Guidance on validating model outputs, safe database operations and context-appropriate output encoding.

OWASP Data and Model Poisoning

Guidance on provenance, controlled changes and malicious content affecting data and model components.

EDUCAUSE HECVAT guidance

Higher-education vendor assessment, including privacy and AI questions. Evaluate responses and product-specific evidence through the existing institutional process.

Work from a common foundation

Adaptable model policy

Stakeholder playbooks

Security privacy and compliance

How to use this resource

For help applying these materials: hello@humanskills.ai