

# AI Risk Assessment and Impact Assessment

---

*Detailed standard and assessment record for known and foreseeable AI risks, affected parties, controls and residual risk*

## Core assessment commitment

KOMGIS will identify AI risks before deployment and throughout operation, assess who and what may be affected, implement proportionate preventive, detective and corrective controls, verify that those controls work, and explicitly decide whether the risk remaining after treatment is acceptable. The assessment must reflect the most consequential reasonably foreseeable scenario - not only the average or intended use.

**Document owner: AI Governance / Technical Leadership**

Version: 1.0

Effective date: 8 July 2026

Classification: Internal Policy / Governance Record

## Document Control

|                         |                                                                                                                                    |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| <b>Document title</b>   | KOMGIS AI Risk Assessment and Impact Assessment                                                                                    |
| <b>Document owner</b>   | AI Governance / Technical Leadership                                                                                               |
| <b>Version</b>          | 1.0                                                                                                                                |
| <b>Effective date</b>   | 8 July 2026                                                                                                                        |
| <b>Review frequency</b> | At least annually and after defined trigger events                                                                                 |
| <b>Approved by</b>      | [Executive approver / role]                                                                                                        |
| <b>Classification</b>   | Internal Policy / Governance Record                                                                                                |
| <b>Applies to</b>       | KOMGIS AI systems, client implementations, AI-enabled workflows, RAG platforms, AI agents and approved third-party AI integrations |

### Status and legal note

This document operationalizes the AI risk and impact assessment requirements in KOMGIS's Responsible AI and AI Management Policy. It is an internal governance methodology and record template. It does not by itself constitute ISO/IEC 42001 certification, an EU AI Act conformity assessment, a fundamental-rights impact assessment required by a specific law, or a legal determination that a system falls within or outside a regulatory category. System-specific legal and compliance analysis must be performed where applicable.

KOMGIS's internal risk level is a governance control. Regulatory classification under applicable law is assessed separately. A system may be internally High or Critical even when it is not legally classified as a high-risk AI system, and a legal classification may impose obligations beyond this document.

### Assessment statement

AI risk assessment is a documented and iterative lifecycle process. For Limited, High and Critical systems, the assessment is required before production use and must be updated when a material change, incident, monitoring signal or new reasonably foreseeable risk could alter the system risk profile.

An AI Impact Assessment is required where the potential impact on people, fundamental rights, safety or important decisions is material. The impact assessment is not a separate compliance exercise detached from engineering: its findings must produce concrete controls, tests, monitoring requirements, human oversight arrangements and risk-treatment actions.

# Contents

|                                                         |
|---------------------------------------------------------|
| 1. Purpose                                              |
| 2. Scope                                                |
| 3. Governance Basis and Policy Relationship             |
| 4. Definitions                                          |
| 5. Assessment Principles                                |
| 6. Roles, Accountability and Independence               |
| 7. When an Assessment Is Required                       |
| 8. Assessment Workflow and Decision Gates               |
| 9. System Context and Intended Purpose                  |
| 10. Affected Parties and Impact Mapping                 |
| 11. Risk Identification                                 |
| 12. Known and Reasonably Foreseeable Risk Taxonomy      |
| 13. Foreseeable Misuse and Abnormal Conditions          |
| 14. Likelihood and Impact Scales                        |
| 15. Risk Scoring and Internal Classification            |
| 16. Inherent Risk and Residual Risk                     |
| 17. Control Design and Control Evidence                 |
| 18. Risk Treatment and Action Planning                  |
| 19. AI Impact Assessment Trigger                        |
| 20. AI Impact Assessment Method                         |
| 21. People, Fundamental Rights and Uneven Impact        |
| 22. Data Provenance, Representativeness and Bias        |
| 23. Human Oversight and Automation Bias                 |
| 24. RAG and Document Intelligence Risk Assessment       |
| 25. AI Agent and Tool-Use Risk Assessment               |
| 26. Privacy, Confidentiality and Security Impact        |
| 27. Provider, Model and Supply-Chain Risk               |
| 28. Evaluation, Testing and Monitoring Linkage          |
| 29. Residual Risk Acceptance and Deployment Decision    |
| 30. Change, Reassessment and Continuous Review          |
| 31. Incidents, Complaints and Corrective Action         |
| 32. Records, Retention and Auditability                 |
| 33. Exceptions and Enforcement                          |
| Appendix A. AI Risk Assessment Record Template          |
| Appendix B. AI Impact Assessment Record Template        |
| Appendix C. Risk Register Fields and Status Rules       |
| Appendix D. AI Risk Library                             |
| Appendix E. Scoring Matrix and Decision Rules           |
| Appendix F. Control Adequacy Scale                      |
| Appendix G. Example Assessment - Enterprise RAG Chatbot |
| Appendix H. Assessment Approval Checklist               |
| Appendix I. Alignment Overview                          |

# 1. Purpose

This document defines how KOMGIS identifies, analyzes, evaluates, treats, monitors and accepts AI risk. It also defines how KOMGIS assesses material impacts on people, clients, business processes, data, systems and other affected parties.

The purpose is to ensure that AI risk decisions are repeatable, evidence-based and connected to the real architecture and business process. The assessment must be specific enough to answer:

- What can go wrong under intended use, foreseeable misuse or abnormal operating conditions?
- Who or what may be affected, directly or indirectly?
- How severe could the effect be and how likely is it to occur?
- Which preventive, detective and corrective controls reduce the risk?
- What evidence demonstrates that the controls are implemented and effective?
- What risk remains after controls, and who is authorized to accept it?
- What must be tested and monitored after deployment?
- What changes or events require reassessment?

## Mandatory rule

A statement such as "the model is generally accurate", "the provider is secure" or "a human is in the loop" is not a completed risk assessment. Risks, affected parties, controls, control evidence, residual risk and ownership must be documented at a level proportionate to the system.

# 2. Scope

This standard applies to AI systems developed, configured, integrated, deployed or operated by KOMGIS, including internal systems and client implementations. It applies throughout concept, design, build, test, deployment, operation, change and retirement.

- AI assistants, chatbots and conversational interfaces;
- retrieval-augmented generation (RAG), enterprise search and document intelligence;
- document summarization, extraction, classification and report generation;
- AI agents, tool-using systems and workflow automation;
- voice interfaces, AI-enabled avatars and synthetic content features;
- decision-support systems and AI recommendations used in business processes;
- embeddings, vector stores, rerankers, model gateways and AI integrations;
- client-specific AI systems and KOMGIS internal AI tools;
- material use of third-party AI models, platforms or agent services.

The assessment covers technical and non-technical risk. It includes data, security, privacy, human oversight, accuracy, bias, transparency, business process, contractual, regulatory, operational, provider and supply-chain risk where relevant.

# 3. Governance Basis and Policy Relationship

This standard supports KOMGIS's Responsible AI and AI Management Policy. The parent policy requires known and reasonably foreseeable risks to be identified, affected persons and systems to be considered, preventive/detective/corrective controls to be defined, residual risk to be evaluated, testing and monitoring to verify controls, and risk acceptance to be recorded by an authorized role.

This document must be used together with the AI System Inventory, Data Privacy Policy, AI Security Policy, AI Incident Response Plan, Model Provider Disclosure / Vendor List, Human Oversight Record, and Evaluation and Testing Records. The risk assessment references those records rather than duplicating all technical detail.

| Supporting record                            | Risk-assessment use                                                                                                                       |
|----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| AI System Inventory                          | Defines system identity, intended purpose, owners, models, data, integrations, tools, deployment context and status.                      |
| Data Processing Map                          | Shows data categories, data flows, storage, providers, regions, derived AI data and deletion paths.                                       |
| AI Security Policy / security review         | Provides architecture and control evidence for access, tenant isolation, prompt injection, RAG, agent tools, secrets and output handling. |
| Model Provider Disclosure / Vendor List      | Provides provider, subprocessor, region, retention, training-use, contractual and continuity information.                                 |
| Evaluation and Testing Records               | Demonstrates performance, robustness, permissions, isolation, prompt-injection and control effectiveness.                                 |
| Human Oversight Record                       | Documents reviewer competence, review criteria, approval actions, overrides and escalation.                                               |
| AI Incident Response Plan / incident records | Provides severity, containment, investigation, corrective action and reactivation evidence.                                               |

## 4. Definitions

| Term                        | Assessment meaning                                                                                                                                                                             |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Affected party              | A person, group, client, user, employee, organization, business process, system or other interest that may experience a direct or indirect effect from the AI system.                          |
| Known risk                  | A risk already identified from system design, prior incidents, testing, provider information, research, client requirements or comparable systems.                                             |
| Reasonably foreseeable risk | A risk that can reasonably be anticipated from the intended purpose, architecture, user behavior, operating context, foreseeable misuse or known AI failure patterns.                          |
| Hazard / risk source        | A model, data characteristic, system component, user behavior, integration, process condition or external event that may give rise to harm or loss.                                            |
| Risk event                  | A described scenario in which a hazard or failure occurs and creates an effect on an affected party or asset.                                                                                  |
| Inherent risk               | Risk estimated before system-specific risk-treatment controls are considered. Normal baseline organizational controls may be noted separately.                                                 |
| Control                     | A preventive, detective or corrective measure implemented through application logic, infrastructure, process, governance, contract, testing, monitoring or human oversight.                    |
| Control evidence            | Objective evidence showing that a control exists and, where required, is effective: code, configuration, test result, log, review record, approval, monitoring result or contractual document. |
| Residual risk               | Risk remaining after defined controls and mitigations are applied and verified to the required level.                                                                                          |

| Term                 | Assessment meaning                                                                                                                                                     |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Impact               | The consequence or severity of a risk event for people, rights, safety, clients, data, operations, finances, reputation, contracts or regulatory obligations.          |
| Likelihood           | The estimated probability or frequency of a defined risk event in the stated operating context and time horizon.                                                       |
| AI Impact Assessment | A structured assessment of material effects on people, fundamental rights, safety or important decisions, and the measures needed to prevent or reduce adverse impact. |
| Material change      | A change to purpose, users, model, provider, data, RAG sources, permissions, tools, autonomy, human approval, workflow or operating context that may alter risk.       |
| Risk owner           | The named person or role accountable for ensuring a risk is assessed, treated, monitored and escalated.                                                                |
| Risk acceptance      | An explicit authorized decision to accept the documented residual risk for a defined scope and period.                                                                 |

## 5. Assessment Principles

### 5.1 Scenario-specific assessment

Each risk entry must describe a concrete scenario. Generic labels such as "hallucination risk", "privacy risk" or "bias" may be used as categories, but the record must explain how the failure could occur in this specific system, what output or action is affected, and who may experience the consequence.

### 5.2 Most consequential reasonably foreseeable scenario

Risk classification and treatment must consider the most consequential reasonably foreseeable scenario, not the average expected use. The assessor should distinguish a speculative worst case from a foreseeable scenario supported by system capability, access, data, users or environment.

### 5.3 Defense in depth

Controls should not depend on a single prompt, a model refusal or user caution where failure may cause material impact. Important controls are enforced at the application, permission, data, tool, workflow and human-approval layers.

### 5.4 Evidence before confidence

Control effectiveness is supported by evidence. A control described in design documentation but not implemented or tested is a planned control and may not be used to reduce residual risk as if it were effective.

### 5.5 Human accountability

A named AI system owner remains accountable for the intended purpose, business context and acceptance of residual risk. The model provider is not the system risk owner.

### 5.6 Iterative lifecycle management

The assessment is updated using test results, production monitoring, reviewer edits and rejection rates, security signals, incidents, complaints, provider changes and changes in the operating context.

## 6. Roles, Accountability and Independence

| Role | Minimum assessment responsibilities |
|------|-------------------------------------|
|------|-------------------------------------|

| Role                                  | Minimum assessment responsibilities                                                                                                                                                                               |
|---------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AI system owner                       | Defines intended purpose and business context; confirms affected business processes; appoints risk owners; approves treatment priorities; accepts residual risk within delegated authority; ensures reassessment. |
| Technical owner                       | Documents architecture, models, RAG, tools and integrations; identifies technical failure scenarios; implements controls; provides test and monitoring evidence; supports residual-risk estimation.               |
| AI Governance / Technical Leadership  | Maintains methodology; challenges classification and scoring; reviews High/Critical assessments; approves material exceptions; coordinates management review.                                                     |
| Security / Privacy owner              | Assesses access control, client isolation, prompt injection, data exposure, secrets, provider risk, privacy impact, retention and incident readiness.                                                             |
| Project manager / client owner        | Captures client requirements, operating context, user groups, contract constraints and client-specific impacts; coordinates treatment actions.                                                                    |
| Human oversight owner / reviewer lead | Defines review criteria, reviewer competence, evidence access, approval authority, workload assumptions, escalation and override/stop capability.                                                                 |
| Legal / compliance reviewer           | Performs legal or regulatory classification and obligation analysis when required; reviews restricted domains, fundamental-rights and notification questions.                                                     |
| Risk owner                            | Maintains the individual risk entry, treatment actions, due dates, evidence and monitoring signal.                                                                                                                |
| Independent reviewer / challenger     | For High/Critical systems, provides a review by a person not solely responsible for building the feature, where practical; challenges assumptions and unsupported risk reductions.                                |

### Segregation rule for High and Critical systems

The person who implemented a material control may provide evidence for it, but formal residual-risk acceptance must be made by the authorized system owner or management role. A builder must not silently close a High or Critical risk solely on the basis of their own implementation statement.

## 7. When an Assessment Is Required

### 7.1 Initial assessment

A risk classification is required for every production AI system. A documented AI Risk Assessment is required before launch for Limited, High and Critical systems. Low systems receive a basic assessment when client or personal data, external integrations, material decisions or special risk signals are present.

### 7.2 AI Impact Assessment

An AI Impact Assessment is required where the potential impact on people, fundamental rights, safety or important decisions is material. The trigger is assessed separately from the internal Low / Limited / High / Critical classification.

### 7.3 Reassessment triggers

- new intended purpose, business process or materially different user population;
- use in employment, education, health, insurance, credit, essential services, law enforcement, migration, justice, democratic processes, biometrics, emotion recognition or safety-related contexts;
- new model provider or materially different model family or capability;
- new data category, personal-data category, client-data source, RAG source or processing region;

- new write tool, integration, external action or materially broader permission scope;
- increase in autonomy or removal/weakening of a human approval step;
- material prompt, workflow, system instruction or orchestration change;
- change to RAG permission filtering, tenant isolation or document hierarchy;
- significant performance degradation, drift, cost/latency behavior that changes fallback behavior, or repeated reviewer rejection;
- security vulnerability, successful prompt injection, wrong-client retrieval or other AI incident;
- material complaint, affected-person challenge or client request showing an unassessed impact;
- new contractual, legal, regulatory or provider requirement relevant to the system.

## 8. Assessment Workflow and Decision Gates

The assessment follows the workflow below. Steps may be combined for simple systems, but the required decisions and evidence must remain identifiable.

| Gate          | Activity                                                                                                                       | Required output / decision                       |
|---------------|--------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|
| G0 - Initiate | Open or update the AI System Inventory entry. Identify owner, technical owner, intended purpose, users and deployment context. | Assessment identifier and accountable owners.    |
| G1 - Context  | Document system boundaries, architecture, model/provider, data, RAG sources, tools, actions, human review and dependencies.    | System context sufficient to identify risk.      |
| G2 - Screen   | Check prohibited/restricted use signals, regulatory role/classification need and AI Impact Assessment trigger.                 | Escalation decision and review requirements.     |
| G3 - Identify | Identify known and foreseeable risk scenarios, misuse, abnormal conditions and affected parties.                               | Initial risk register.                           |
| G4 - Analyze  | Estimate inherent likelihood and impact; assign initial score and internal risk level.                                         | Prioritized risks and assumptions.               |
| G5 - Treat    | Define preventive, detective and corrective controls; assign owners and dates.                                                 | Risk treatment plan.                             |
| G6 - Verify   | Test controls and collect evidence. Record defects, limitations and unresolved actions.                                        | Control evidence and updated estimates.          |
| G7 - Evaluate | Estimate residual risk after verified controls; assess cumulative and cross-risk effects.                                      | Residual risk level and deployment conditions.   |
| G8 - Approve  | Obtain authorized risk acceptance, conditional approval, rejection or redesign decision.                                       | Signed/recorded decision and expiry/review date. |
| G9 - Operate  | Monitor risk indicators, incidents, complaints, reviewer behavior and changes.                                                 | Monitoring evidence and reassessment triggers.   |

1. The assessment owner prepares the context and initial risk register.
2. Technical, security/privacy, business and human-oversight stakeholders review relevant risks.
3. High and Critical systems receive formal challenge and management review appropriate to risk.
4. Blocking risks are resolved or the system is redesigned. Failed isolation tests or bypassed approval gates prohibit release.
5. Residual risk is accepted only after planned controls used in the score are implemented and supported by required evidence.
6. Deployment conditions, monitoring indicators and reassessment triggers are linked to the system record.

## 9. System Context and Intended Purpose

Risk identification begins with a precise system description. The intended purpose must be specific enough to judge whether an output is correct, whether an action is authorized and whether a failure is material.

### 9.1 Required context fields

| Context field        | Minimum assessment question                                                                                           |
|----------------------|-----------------------------------------------------------------------------------------------------------------------|
| System identity      | What system, feature and version is being assessed? What inventory ID identifies it?                                  |
| Intended purpose     | What exact business problem and approved task does the AI support? What is explicitly out of scope?                   |
| Users                | Who uses the system? Are users employees, client personnel, customers, public users, minors or vulnerable persons?    |
| Affected process     | Where does AI enter the business process? What happens before and after the AI output?                                |
| AI output            | Does the system generate text, extraction, classification, recommendation, synthetic media, tool selection or action? |
| Reliance             | Is the output advisory, reviewed, normally relied upon, or automatically executed?                                    |
| Autonomy             | Can the system choose tools, sequence steps, retry, act externally or change records?                                 |
| Data                 | Which public, client, confidential, personal, regulated or derived AI data categories are processed?                  |
| RAG / knowledge      | Which source stores are searched? How are tenant, user and document permissions enforced?                             |
| Models / providers   | Which model families, providers, embeddings, rerankers or gateways are used?                                          |
| Integrations / tools | What systems can be read or changed? What credentials and scopes are used?                                            |
| Human oversight      | Who reviews what, using which evidence, at what point and with what authority?                                        |
| Deployment context   | Internal, client-facing or public? Which region, environment and scale?                                               |
| Fallback / stop      | How can the feature be disabled, rolled back or routed to manual processing?                                          |

### 9.2 System boundary

The assessment boundary includes more than the foundation model. It includes prompt/workflow logic, retrieval, document processing, authorization, APIs, tool execution, storage, client isolation, validation, human review, monitoring, provider dependencies and downstream business process. Risks may arise from interactions between components even when each individual component appears acceptable.

### 9.3 Assumptions and dependencies

Important assumptions must be recorded because they may become reassessment triggers. Examples include expected document language, average context size, reviewer capacity, input format, supported jurisdictions, provider availability, retrieval index freshness, client identity data quality and whether users are trained.

## 10. Affected Parties and Impact Mapping

The assessment identifies parties who may experience direct or indirect effects. The assessor must not limit analysis to the authenticated user. A generated recommendation or automated action can affect a person who never interacts with the AI system.

| Affected party category       | Examples                                                                  | Impact questions                                                                                                    |
|-------------------------------|---------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Direct users                  | Employees, client users, administrators, reviewers                        | Can the user be misled, over-rely on output, lose time, disclose data, or be unable to contest/override the system? |
| Subjects of data or decisions | Applicants, employees, insured persons, patients, customers, complainants | Could AI output influence treatment, eligibility, service, money, rights, reputation, safety or access?             |
| Clients / organizations       | Client company, department, project owner                                 | Could the system create legal, contractual, financial, operational, security or trust impact?                       |
| Non-users / third parties     | Recipients of AI email, people mentioned in documents, business partners  | Could they receive wrong information, be profiled, exposed, misrepresented or affected by an action?                |
| KOMGIS personnel              | Developers, support, reviewers, account managers                          | Could workload, access, monitoring or pressure undermine safe operation?                                            |
| Systems / assets              | Databases, file stores, vector stores, ERP/CRM, email                     | Could AI cause unauthorized reads/writes, deletion, corruption, excessive volume or credential exposure?            |
| Providers / subprocessors     | Model, cloud, monitoring, storage providers                               | Can dependency, outage, term change, region or subprocessor behavior create impact?                                 |
| Public / society              | Public users or broader groups where relevant                             | Could synthetic content, scale, discriminatory patterns or unsafe behavior create broader effects?                  |

## 10.1 Affected party record

For each materially affected party, the assessment records the relationship to the system, possible beneficial effect, possible adverse effect, severity factors, whether the party is vulnerable or has reduced ability to detect/correct AI error, and available contest, correction, override or escalation routes.

## 10.2 Uneven and cumulative impact

The assessment considers whether errors are distributed unevenly across languages, document types, user groups, roles or client workflows. It also considers repeated small errors that may accumulate into a material effect, such as systematic omission of a field, repeated reviewer workload, repeated denial of access to information, or recurring unsupported recommendations.

# 11. Risk Identification

Risk identification combines architecture review, process analysis, known AI failure patterns, data review, misuse analysis, provider information, testing, incident evidence and stakeholder input. The objective is a complete set of plausible risk scenarios, not a generic checklist copied unchanged between systems.

## 11.1 Required identification sources

- the AI System Inventory and intended-purpose statement;
- architecture and data-flow diagrams;
- Data Processing Map and retention/deletion rules;
- authentication, authorization and tenant-isolation design;
- RAG source hierarchy, chunking, metadata and retrieval filters;
- prompt/workflow, agent graph, tool inventory and action scopes;
- model provider, subprocessor, region and retention information;
- client requirements, contracts and operating assumptions;
- evaluation results, adversarial tests and known failure cases;
- security review, threat modeling and prompt-injection testing;

- human reviewer feedback, edit/rejection/escalation data;
- incidents, complaints, near misses and support records;
- changes in model behavior, provider capability or system scale.

## 11.2 Risk scenario format

### Recommended risk statement

Because [risk source / condition], the AI system may [failure or behavior] during [use, misuse or abnormal condition], causing [effect] to [affected party / asset]. This may lead to [business, rights, privacy, security, safety, contractual or regulatory consequence].

Example: Because RAG retrieval uses incomplete document-permission metadata after a project-role change, the chatbot may retrieve a chunk from a restricted document for an otherwise authenticated user, causing disclosure of confidential client information and loss of trust. This may create contractual, privacy and incident-response consequences.

## 12. Known and Reasonably Foreseeable Risk Taxonomy

The taxonomy below is a starting point. Not every category applies to every system, and use of the taxonomy does not replace system-specific scenario descriptions.

| Risk domain                   | Examples of risk sources and events                                                                                                                                  |
|-------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Purpose and misuse            | Use outside intended purpose; user treats assistant as authority; prohibited/restricted use introduced through configuration; scope creep; unsafe secondary use.     |
| Accuracy and hallucination    | Unsupported claims; fabricated source; extraction error; classification error; incorrect calculation; stale answer; omission; false confidence.                      |
| RAG and grounding             | Unauthorized retrieval; wrong-client chunk; irrelevant retrieval; missing source; stale index; citation mismatch; malicious document instruction; context poisoning. |
| Data and privacy              | Excessive data sent to model; personal data in logs; retention mismatch; training-use misunderstanding; failed deletion of embeddings; cross-region processing.      |
| Security                      | Prompt injection; indirect injection; secret exposure; weak authorization; object-level access failure; insecure file handling; tool abuse; data exfiltration.       |
| Agent and automation          | Unauthorized tool call; approval bypass; retry loops; action splitting; wrong recipient; destructive write; excessive volume; action based on unverified output.     |
| Human oversight               | Automation bias; reviewer lacks sources; unclear approval criteria; workload causes rubber-stamping; reviewer cannot stop system; authority mismatch.                |
| Bias and uneven performance   | Different error rates by language, document type or affected group; non-representative data; proxy variables; feedback loops; inaccessible interface.                |
| Transparency and traceability | User unaware of AI interaction; missing citations; version not identifiable; incomplete tool logs; inability to reconstruct an incident.                             |
| Model / provider              | Model update; capability regression; outage; region change; term change; retention change; service discontinuation; subprocessor risk; model deprecation.            |
| Operational resilience        | Latency/outage causes unsafe fallback; queue backlog; stale cache; failed monitoring; cost spike; rate-limit behavior; backup or restore gap.                        |

| Risk domain                    | Examples of risk sources and events                                                                                                               |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Business process / contractual | AI changes process responsibility; wrong client communication; SLA impact; unapproved provider; contract-specific data restriction violated.      |
| Regulatory / rights / safety   | Material effect on important decisions; lack of contestability; restricted domain; transparency failure; safety impact; discriminatory treatment. |
| Change and drift               | Prompt change; provider/model change; new tool; new RAG source; changed permissions; data distribution shift; user behavior change.               |

## 13. Foreseeable Misuse and Abnormal Conditions

The assessment must examine more than compliant, trained users and normal provider behavior. Foreseeable misuse means behavior that is not the intended use but can reasonably be expected given the interface, incentives, user knowledge and system capability.

### 13.1 Foreseeable misuse prompts

- A user asks the system to perform a related but unapproved task because the interface appears capable of it.
- A user intentionally requests another client, project, department or document outside their authorization.
- A user uploads a document containing hidden or explicit instructions intended to manipulate the model or agent.
- A user treats a well-written answer as verified and skips the expected human check.
- A reviewer repeatedly approves output without opening cited sources because workload is high.
- A client administrator grants broader permissions than intended or imports a document into the wrong workspace.
- An agent is asked to achieve a goal that can be satisfied through an unintended combination of approved tools.
- A user attempts bulk extraction, repeated queries or reformulation to obtain sensitive content.
- A downstream system accepts malformed or invented structured output because validation is weak.

### 13.2 Abnormal operating conditions

- model/provider outage, timeout, rate limit, degraded model or silent model update;
- vector index lag, partial ingestion, failed deletion, stale permissions or damaged metadata;
- large, malformed, scanned, multilingual, encrypted or adversarial files;
- missing source documents, contradictory documents or unusually long contexts;
- credential expiration, integration failure or partial write completion;
- monitoring delay, unavailable reviewer, queue overload or high incident volume;
- client configuration error, incorrect tenant mapping or role change during an active workflow;
- system restart, retry, duplicate message, duplicated tool call or idempotency failure.

For each abnormal condition with material consequences, the assessment should identify safe failure behavior: block, ask for clarification, return no answer, use a manual fallback, require approval, retry safely, or disable the feature. Silent continuation is not assumed to be safe.

## 14. Likelihood and Impact Scales

KOMGIS uses a five-point likelihood scale and five-point impact scale for individual risk entries. Scores support prioritization and consistent discussion; they do not replace professional judgment or regulatory analysis.

### 14.1 Likelihood scale

| Score | Rating | Assessment guidance                                                                                                                                                  |
|-------|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | Rare   | Unlikely in the defined operating context. Requires an unusual combination of events; no known comparable occurrence; strong preventive controls or narrow exposure. |

| Score | Rating         | Assessment guidance                                                                                                                                |
|-------|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| 2     | Unlikely       | Could occur, but is not expected in normal operation. Limited exposure or known barriers make the event infrequent.                                |
| 3     | Possible       | Plausible during the assessment period. Similar AI/system failures are known or the system is exposed to conditions that can produce the event.    |
| 4     | Likely         | Expected to occur repeatedly or has occurred in testing/operation without reliable prevention. Exposure is frequent or control dependence is weak. |
| 5     | Almost certain | Occurs commonly, is inherent in current behavior, or is expected at the system scale unless substantially redesigned.                              |

## 14.2 Impact scale

| Score | Rating     | Assessment guidance                                                                                                                                                                                          |
|-------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | Negligible | Minor internal inconvenience or cosmetic issue. No sensitive data, material client effect, important decision or external action. Easily corrected.                                                          |
| 2     | Minor      | Limited rework, localized delay or low-impact incorrect output caught by normal user judgment. No serious confidentiality, rights or safety effect.                                                          |
| 3     | Moderate   | Client dissatisfaction, material workflow rework, repeated wrong output, limited confidential exposure, incorrect external communication or reversible business impact.                                      |
| 4     | Major      | Serious financial, legal, privacy, employment, insurance, safety, security or fundamental-rights impact; wrong-client disclosure; unauthorized material action; significant client incident.                 |
| 5     | Severe     | Severe or potentially irreversible harm; large-scale cross-client exposure; dangerous autonomous action; major personal-data impact; systemic discriminatory effect; major regulatory or safety consequence. |

## 14.3 Impact dimensions

The assessor considers the highest applicable impact dimension for the scenario. A risk is not averaged down because some dimensions are minor.

| Dimension                 | Questions                                                                                                                                       |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| People / rights           | Could a person experience unfair treatment, loss of access, inability to contest, reputational harm, discrimination or another material effect? |
| Safety / physical         | Could the output or action contribute to physical injury, unsafe operation or a safety-critical failure?                                        |
| Privacy / confidentiality | What sensitivity, number of persons/clients, duration, identifiability and onward disclosure are involved?                                      |
| Security                  | Could the event provide unauthorized access, privilege, data exfiltration, destructive action or persistent compromise?                         |
| Financial / legal         | Could the event create material loss, liability, penalties, contractual breach or incorrect legally significant action?                         |
| Operational               | Could the event stop a critical process, corrupt records, create large backlogs or make recovery difficult?                                     |
| Client / trust            | Could the event materially undermine client trust, trigger a reportable incident or affect multiple client workflows?                           |

| Dimension             | Questions                                                                                                   |
|-----------------------|-------------------------------------------------------------------------------------------------------------|
| Scale / reversibility | How many outputs/actions may be affected? Can the result be identified, corrected and reversed before harm? |

## 15. Risk Scoring and Internal Classification

The base risk score is Likelihood x Impact. The score is a prioritization tool. Certain risk characteristics trigger escalation regardless of arithmetic score.

| Score | Risk band | Minimum treatment expectation                                                                                                                                                                                   |
|-------|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1-4   | Low       | Manage through standard controls. Record material assumptions and monitor if relevant.                                                                                                                          |
| 5-9   | Limited   | Document risk and controls; assign owner; verify key controls; define monitoring or review as appropriate.                                                                                                      |
| 10-16 | High      | Formal treatment plan; strong control evidence; formal approval gate; enhanced testing and monitoring; management risk acceptance.                                                                              |
| 17-25 | Critical  | Do not deploy or continue ordinary operation until executive, technical, security and legal/compliance review as applicable; redesign, restrict or formally control the use. Prohibited uses remain prohibited. |

### 15.1 Escalation indicators

The assessor must consider elevation to High or Critical when any of the following are present, even if the initial numeric score is lower:

- potential material effect on rights, access, money, health, employment, education, insurance, essential services or safety;
- sensitive personal or regulated data, large-scale processing or credible cross-client exposure;
- write tools, external communications, financial action, deletion or other difficult-to-reverse action;
- high autonomy or the ability to combine tools to achieve an objective;
- public or adversarial exposure, valuable data, untrusted RAG sources or significant prompt-injection risk;
- weak or impractical human oversight, high automation bias or reviewer workload concerns;
- potential prohibited or high-risk regulatory category or another restricted domain;
- uncertainty is material and evidence is insufficient to justify a lower classification.

#### Classification rule

The system-level KOMGIS risk classification is based on the most consequential reasonably foreseeable risk scenario and the total system context. Individual risk scores support the decision but do not mechanically determine it.

## 16. Inherent Risk and Residual Risk

### 16.1 Inherent risk

Inherent risk estimates the scenario before system-specific treatment controls are credited. This exposes the real consequence of the capability and helps prevent a serious risk from disappearing behind vague statements that "controls exist".

The assessment may separately note baseline environmental controls, such as standard corporate authentication or infrastructure security, but should be consistent about which controls are treated as inherent versus residual scoring inputs.

## 16.2 Crediting controls

A control may reduce residual likelihood, residual impact, or both. The assessment must explain why. For example, a tenant filter enforced before retrieval may substantially reduce the likelihood of wrong-client retrieval. A human approval gate before sending email may reduce the likelihood that a bad draft becomes an external action. A log does not normally reduce the initial occurrence likelihood, but can improve detection, containment and total impact.

## 16.3 Planned controls

A planned control is recorded in the treatment plan but is not credited as fully effective in residual risk until implemented and supported by required evidence. Conditional deployment may be approved only when the conditions, risk owner, deadline and compensating controls are explicit.

## 16.4 Residual-risk rationale

Every High or Critical inherent risk and every material Impact Assessment finding must include a short narrative explaining the residual score. The rationale links the specific controls and evidence to the change in likelihood and/or impact.

| Example                    | Weak rationale                | Acceptable rationale                                                                                                                                                                                                                                                                                                                                                          |
|----------------------------|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cross-client RAG retrieval | "RBAC implemented; risk low." | "Backend tenant and object authorization is applied before vector query construction; client_id and permitted document IDs are mandatory filters; 120 cross-tenant tests and negative object-access tests passed. Likelihood reduced from Possible (3) to Rare (1). Impact remains Major (4) because a successful bypass could disclose confidential data. Residual score 4." |
| External AI email          | "Human in the loop."          | "The system produces a draft only. Send API is not exposed to the agent. An authorized reviewer must open the source case, review recipient and final body, and click Send in the application. Approval and final content are logged. Likelihood reduced from Likely (4) to Unlikely (2); impact remains Moderate (3). Residual score 6."                                     |

# 17. Control Design and Control Evidence

Controls are classified as preventive, detective or corrective. Important risks normally require defense in depth across more than one control type.

| Control type | Purpose                                                                   | AI examples                                                                                                                                                        |
|--------------|---------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Preventive   | Stop or reduce the chance of the risk event before it occurs.             | Backend authorization; tenant-aware retrieval filter; read-only tool; approval gate; schema validation; data minimization; provider restriction; rate limit.       |
| Detective    | Identify anomalous behavior, control failure or impact quickly.           | Wrong-client reference alert; prompt-injection signal; reviewer rejection monitoring; sensitive-data detection; unusual tool-call monitoring; audit review.        |
| Corrective   | Contain, reverse, recover from or reduce the consequence after detection. | Disable feature; revoke tool/credential; rollback model/workflow; delete exposed data; correct outputs; restore records; notify affected client; retrain reviewer. |

## 17.1 Control description standard

A control description should identify the control owner, enforcement point, condition, action and evidence. "Permissions are checked" is insufficient for a material RAG risk. The description should state where permission is checked, which identifiers are required, whether the check occurs before retrieval, and what happens when permission context is absent or invalid.

## 17.2 Evidence hierarchy

| Evidence level               | Example                                                                                        | Use in residual-risk assessment                                                            |
|------------------------------|------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| E0 - Assertion               | Design intent, meeting statement or undocumented belief                                        | Not sufficient to close a material treatment action.                                       |
| E1 - Documented design       | Architecture, policy, control specification, code reference                                    | Shows intended control; limited credit until implementation is verified.                   |
| E2 - Implementation evidence | Configuration, code review, deployed control, permission rule                                  | Shows control exists; testing still required when failure is material.                     |
| E3 - Test evidence           | Automated/manual test result, adversarial test, isolation test, approval-bypass test           | Strong pre-deployment evidence when test scope is representative and defects are resolved. |
| E4 - Operational evidence    | Monitoring trend, audit logs, reviewer data, incident-free operation under meaningful exposure | Supports continued effectiveness; does not prove impossibility of failure.                 |

## 17.3 Control failure analysis

For High and Critical risks, the assessor should ask what happens if the key control fails, is misconfigured, is bypassed or becomes unavailable. Single-point controls that fail open require particular scrutiny. Where practical, material authorization and approval controls should fail closed.

# 18. Risk Treatment and Action Planning

KOMGIS may avoid, reduce, transfer/share or accept a risk. Acceptance is permitted only after required assessment and within authorized risk appetite. Legal obligations and prohibited uses cannot be waived through internal risk acceptance.

| Treatment option | Meaning                                                                                                                                                                 | Examples                                                                                                  |
|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| Avoid            | Remove the risky capability, use case or data.                                                                                                                          | Do not support prohibited use; remove write tool; exclude sensitive document class; keep system internal. |
| Reduce           | Implement controls to lower likelihood or impact.                                                                                                                       | Permission filtering; human approval; output validation; model restriction; monitoring; safer workflow.   |
| Transfer / share | Allocate certain obligations or consequences contractually or through approved providers/insurance, while retaining KOMGIS responsibilities that cannot be transferred. | DPA, contractual responsibilities, provider SLA, client operational responsibility.                       |
| Accept           | Explicitly accept residual risk for a defined scope and period.                                                                                                         | Authorized owner records rationale, monitoring and review date for a limited residual risk.               |

## 18.1 Treatment action fields

- risk ID and treatment action ID;
- action description and intended risk reduction;
- control type and enforcement point;
- action owner and accountable risk owner;
- priority and due date;
- required evidence and acceptance criterion;
- dependencies and release blocker status;
- implementation status: Open / In progress / Implemented / Verified / Closed / Accepted / Superseded;
- verification result and evidence reference;
- updated residual likelihood, impact and rationale.

## 18.2 Blocking treatment actions

The following are normally release blockers where applicable: failed tenant/client isolation, missing backend authorization, RAG retrieval outside permission boundaries, sensitive actions not blocked for required approval, exposed secrets, unapproved provider use for client confidential data, critical control tests failing, or a prohibited AI use.

## 19. AI Impact Assessment Trigger

The assessment owner completes the impact-screening questions during G2. One or more positive answers may require a full AI Impact Assessment and legal/compliance review depending on context.

| Screening question                                                                                                                                                   | Yes / No / N/A | Assessment consequence                                                |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|-----------------------------------------------------------------------|
| Can AI output or action materially influence a decision about a natural person?                                                                                      | [ ]            | Impact Assessment normally required.                                  |
| Can the system affect access to employment, education, credit, insurance, health, essential services, justice, migration, benefits or another important opportunity? | [ ]            | Escalate; legal/regulatory classification review required.            |
| Can the system create material safety or physical consequences?                                                                                                      | [ ]            | Impact Assessment and enhanced safety review required.                |
| Does the system process sensitive personal data, biometric data or large-scale personal data in a way material to AI behavior?                                       | [ ]            | Privacy/security impact analysis and legal review.                    |
| Are minors, elderly persons, persons with disabilities or other potentially vulnerable groups materially affected?                                                   | [ ]            | Assess vulnerability, accessibility, manipulation and contestability. |
| Could performance differ materially across groups, languages, regions or document types in a way that affects people?                                                | [ ]            | Bias/representativeness analysis required.                            |
| Is AI output normally relied upon with limited practical ability for the affected person to detect or correct error?                                                 | [ ]            | Assess oversight, explanation and contestability.                     |
| Does the system communicate externally or take action that may materially affect a person or client?                                                                 | [ ]            | Assess approval, reversibility and notification.                      |
| Could a repeated small error create cumulative material impact at scale?                                                                                             | [ ]            | Assess scale and cumulative effects.                                  |
| Is the use in a restricted domain under KOMGIS policy?                                                                                                               | [ ]            | Mandatory escalation before development commitment or deployment.     |

## 20. AI Impact Assessment Method

The AI Impact Assessment documents the benefits and necessity of using AI, affected individuals or groups, potential adverse and uneven impacts, data and bias considerations, transparency and source evidence, human oversight, contestability, privacy/security effects, failure scenarios and incident capability.

### 20.1 Assessment sequence

1. Describe the AI-supported business objective and the expected benefit to users, clients or affected persons.
2. Explain why AI is used and whether a non-AI or lower-autonomy alternative could reasonably achieve the purpose with lower material impact.
3. Map direct and indirect affected parties, including people who do not interact with the system.
4. Describe how the AI output enters a decision, communication or action and how much weight it receives in practice.
5. Identify beneficial, adverse, uneven and cumulative impacts for each materially affected party.
6. Assess data provenance, representativeness, proxy variables, language/domain coverage and known performance differences where relevant.
7. Assess transparency, source evidence and what information a user or affected person needs to understand or challenge the AI-supported outcome.
8. Assess human oversight competence, authority, evidence access, workload and ability to disregard, override, reverse or stop the system.
9. Assess contest, correction, escalation and remedy routes, including how incorrect data or output can be corrected.
10. Assess privacy, confidentiality, security and cross-client exposure that could affect people or organizations.
11. Analyze serious failure scenarios and whether incident detection, containment, communication and reactivation controls are adequate.
12. Define controls, evaluation metrics, monitoring indicators and review intervals for material impact findings.

### 20.2 Impact conclusion

The Impact Assessment conclusion records whether the use is acceptable, acceptable with conditions, requires redesign/restriction, or must not proceed. It lists unresolved findings, treatment actions, the authorized approver and the next review date.

## 21. People, Fundamental Rights and Uneven Impact

Where people may be materially affected, the assessment considers the practical effect of the AI-enabled process rather than only the model output. A recommendation that a human routinely follows may have decision-like impact even when technically labelled "advisory".

| Impact area                         | Assessment questions                                                                                                                                                     |
|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fair treatment / non-discrimination | Could errors or proxies lead to systematically different treatment? Are protected or vulnerable groups relevant to the context? Are differences measured where feasible? |
| Privacy / data protection           | Is personal data necessary and proportionate? Is the source lawful/approved? Can people exercise applicable data rights? Are derived outputs and embeddings governed?    |
| Autonomy / manipulation             | Could the interface, voice, avatar or conversational behavior pressure, deceive or materially distort user behavior? Are vulnerable users more susceptible?              |
| Access / opportunity                | Could AI influence access to employment, insurance, education, credit, services or another important opportunity?                                                        |

| Impact area                         | Assessment questions                                                                                                                     |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Expression / information            | Could incorrect summaries, ranking or generated content materially suppress, distort or misrepresent information relied on by users?     |
| Due process / contestability        | Can the affected person understand that AI was involved where required, contest the outcome, correct data and reach an authorized human? |
| Accessibility                       | Can users with relevant disabilities or language needs use the system and understand critical notices or outputs?                        |
| Human dignity / reputational impact | Could output stigmatize, falsely characterize, expose or misrepresent a person?                                                          |
| Safety                              | Could AI content or action contribute to physical, psychological or operational harm in the actual context?                              |

### 21.1 Severity modifiers for affected persons

- vulnerability, age, disability or reduced ability to recognize/correct AI error;
- power imbalance between the affected person and the organization using the AI output;
- difficulty obtaining human review or an alternative process;
- high consequence of delay even when an error is eventually reversible;
- scale, frequency or cumulative repetition of the AI-supported outcome;
- whether the affected person knows AI is involved and can provide relevant context;
- whether source evidence, reasons and correction routes are available.

## 22. Data Provenance, Representativeness and Bias

Data and bias analysis is proportionate to the system. A document-grounded enterprise chatbot may require source authorization, document coverage and language analysis rather than demographic model training analysis. A system affecting people or important decisions may require deeper representativeness and group-performance assessment.

### 22.1 Data questions

- What data influences the AI output: user prompt, retrieved text, structured records, historical decisions, labels, examples, fine-tuning data or tool results?
- Who created the data, for what original purpose, and is the data appropriate for the current intended purpose?
- Is the data complete, current and sufficiently representative of the operating context?
- Are important languages, client document types, populations, edge cases or rare but serious scenarios underrepresented?
- Could labels or historical records contain past human bias, inconsistent practices or proxy variables?
- Could missing data be interpreted as a negative signal or cause systematic omission?
- How are data quality errors corrected, and are derived chunks, embeddings or outputs updated after source correction?
- Does data minimization conflict with accuracy or oversight needs, and how is the balance documented?

### 22.2 Performance segmentation

Where uneven performance may create material impact, evaluation should segment results by relevant variables such as language, document source, document type, OCR quality, client configuration, user role, workflow stage or affected group. Segmentation variables must be context-appropriate and lawfully available. The assessment should explain why selected segments are relevant.

### 22.3 Bias-control examples

| Risk                                           | Possible controls and evidence                                                                                                                             |
|------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Underperformance in Serbian-language documents | Representative Serbian test set; language-specific metrics; retrieval and OCR evaluation; fallback or disclosure when confidence/evidence is insufficient. |

| Risk                                                     | Possible controls and evidence                                                                                                                        |
|----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Historical labels reproduce inconsistent human decisions | Label audit; expert review; remove inappropriate proxy fields; compare outcome distributions; require human review; monitor overrides and complaints. |
| Rare document types are omitted from extraction          | Coverage inventory; targeted test cases; missing-field validation; manual fallback; monitor validation failures by document type.                     |
| Users over-trust confident style                         | Source citations; calibrated limitations; force evidence review for sensitive action; reviewer training; measure rejection/edit patterns.             |

## 23. Human Oversight and Automation Bias

Human oversight is assessed as a system control, not a label. The assessment must determine whether the reviewer can realistically identify the relevant failure and has authority and time to act.

### 23.1 Oversight effectiveness questions

| Area                      | Questions                                                                                                           |
|---------------------------|---------------------------------------------------------------------------------------------------------------------|
| Competence                | Does the reviewer understand the business domain, system capability, known limitations and common AI failure modes? |
| Evidence access           | Can the reviewer inspect source documents, retrieved evidence, structured input, tool preview and relevant history? |
| Decision criteria         | Are approval/rejection/escalation criteria explicit for sensitive outputs or actions?                               |
| Authority                 | Can the reviewer reject, edit, override, reverse, stop or escalate the output/action?                               |
| Timing                    | Does review occur before the material action? Is the reviewer given enough time?                                    |
| Workload                  | Is expected volume compatible with meaningful review, or will the process encourage rubber-stamping?                |
| Independence / incentives | Are reviewers pressured to maximize throughput or approval rate in a way that weakens oversight?                    |
| Logging                   | Are review, edits, rejection, approval and final disposition traceable?                                             |
| Feedback                  | Do reviewer findings feed evaluation, risk updates and corrective action?                                           |

### 23.2 Automation bias indicators

- very high approval rate combined with low source-opening rate;
- review duration too short for the complexity of the case;
- reviewer edits only formatting while known factual errors persist;
- users quote AI output as an authority rather than checking sources;
- reviewers approve recommendations outside their competence or authority;
- errors are detected primarily by clients after the process rather than by the designated review step.

### 23.3 Oversight as a residual-risk control

Human review may reduce residual likelihood only when the reviewer is likely to detect the defined failure before impact. The risk rationale should reference review evidence, criteria and practical operating conditions. For a hidden cross-tenant retrieval error, a normal user may not know the content is unauthorized; therefore generic user review is not an adequate isolation control.

## 24. RAG and Document Intelligence Risk Assessment

RAG and document intelligence are core KOMGIS use cases. Risk assessment must separate source authorization, retrieval quality, grounding/citation and generation behavior. A model instruction to "only use authorized documents" is not a permission control.

| Risk scenario                                           | Affected parties / assets                        | Minimum control considerations                                                                                                                                         |
|---------------------------------------------------------|--------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cross-client or unauthorized document retrieval         | Client, data subjects, KOMGIS, confidential data | Tenant/user/object authorization before retrieval; mandatory metadata filters; fail closed; isolation tests; wrong-client monitoring; incident route.                  |
| Deleted or permission-revoked source remains searchable | Client, users, data subjects                     | Source-to-chunk/vector lineage; deletion/disable propagation; re-index controls; deletion tests; index freshness monitoring.                                           |
| Prompt injection in uploaded/retrieved document         | Users, clients, systems, tool-connected assets   | Treat content as untrusted; separate instructions/data; tool allowlists; server-side authorization; malicious-document tests; suspicious behavior monitoring.          |
| Citation points to wrong or non-supporting source       | Users, reviewers, clients                        | Chunk/source identifiers; citation correctness evaluation; source-location display; faithfulness tests; block unsupported claims in sensitive workflow as appropriate. |
| Relevant source not retrieved                           | Users, clients, affected process                 | Representative retrieval evaluation; hybrid search/reranking if appropriate; missing-evidence behavior; manual search fallback; coverage monitoring.                   |
| Stale or conflicting documents                          | Users, clients, affected persons                 | Document lifecycle metadata; status/version controls; source precedence; stale-source notices; conflict handling; authoritative-source definition.                     |
| Bulk extraction through repeated queries                | Client confidential data                         | Query/rate/export controls; anomaly monitoring; scoped retrieval; access review; data-loss controls where appropriate.                                                 |
| Sensitive content enters model/provider unnecessarily   | Client, data subjects                            | Minimize retrieved context; provider approval; redaction where appropriate; prompt/log controls; data processing map.                                                  |

### 24.1 Required RAG risk evidence

- permission hierarchy and enforcement point;
- retrieval metadata schema and mandatory client/user/document filters;
- negative authorization and cross-client isolation tests;
- source/chunk/embedding lineage and deletion behavior;
- retrieval quality and citation correctness evaluation;
- prompt-injection and malicious-document tests;
- monitoring for unexpected source IDs, wrong-client references, suspicious export volume and retrieval anomalies.

## 25. AI Agent and Tool-Use Risk Assessment

AI agents require a capability-based risk assessment. The important question is not whether the agent is called autonomous, but which information and actions it can access, combine and repeat.

### 25.1 Tool inventory assessment

| Tool field                     | Required assessment detail                                                                                          |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Tool name / purpose            | Why the agent needs the tool for the approved use case.                                                             |
| Read / write / external action | Whether the tool only retrieves information, modifies records, communicates externally or triggers another process. |

| Tool field                  | Required assessment detail                                                                                 |
|-----------------------------|------------------------------------------------------------------------------------------------------------|
| Credential and scope        | Service account, user delegation, client scope, environment and least-privilege permissions.               |
| Server-side authorization   | Which backend policy is checked before each sensitive call; the model must not decide its own permissions. |
| Input validation            | Allowed parameters, identifiers, recipient domains, schemas, ranges or object scopes.                      |
| Human approval              | Which actions are blocked; who can approve; what preview/evidence is shown; how approval is recorded.      |
| Reversibility / idempotency | Can the action be reversed? Can retries duplicate the action? How are partial failures handled?            |
| Rate / volume               | Limits on repeated calls, bulk export, deletion, messaging or other high-volume effects.                   |
| Logging / alerting          | Tool call, status, approval, affected object, error and suspicious pattern signals.                        |
| Kill switch / revocation    | How the tool or credential can be disabled without disabling unrelated systems.                            |

## 25.2 Agent-specific foreseeable risks

- the agent retries a denied sensitive action using a different phrasing or tool;
- the agent splits one sensitive action into multiple smaller actions to avoid an approval threshold;
- a malicious document instructs the agent to call a tool or disclose data;
- a broad service credential gives access beyond the active user or client workspace;
- a correct tool is called with the wrong client, recipient, object ID or environment;
- partial failure leads to duplicate payment, duplicate email or inconsistent records;
- the agent continues acting after the user purpose or context has changed;
- the agent optimizes for task completion by selecting a technically allowed but business-inappropriate action;
- tool output is treated as trusted instructions and causes a second-stage injection or unsafe action.

### Sensitive action rule

External, irreversible, high-impact or regulated actions require defined human approval unless a specific lawful automation use has been assessed, approved and documented. The agent must not bypass approval by retrying, rephrasing or splitting the action.

## 26. Privacy, Confidentiality and Security Impact

Risk assessment links AI behavior to KOMGIS data and security controls. Prompts, retrieved context, outputs, embeddings, vector metadata and AI workflow records are governed data when connected to client content.

| Risk area            | Assessment questions                                                                                                                               |
|----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Data minimization    | Is the full document/library/database sent when an excerpt or limited record is sufficient? Are unnecessary personal/confidential fields included? |
| Provider exposure    | Which data reaches the model provider? What are training-use, retention, region and subprocessor terms? Are client restrictions satisfied?         |
| Logs                 | Are full prompts/outputs logged unnecessarily? Are secrets or sensitive content redacted? Who can access debug and audit logs?                     |
| Embeddings / vectors | Can vectors or metadata expose tenant identity or sensitive source information? Are they client-scoped and governed with source deletion?          |

| Risk area             | Assessment questions                                                                                                                                    |
|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Authorization         | Are authentication, role, object and tenant checks enforced in backend logic and retrieval/tool layers?                                                 |
| Secrets               | Can prompts, retrieved documents or tool output contain API keys, tokens, credentials or private keys? How is accidental exposure detected and rotated? |
| File handling         | Are type, size, storage, extraction, hidden content and safe processing controlled?                                                                     |
| Output security       | Can generated output be executed as SQL, shell, code or downstream API input without validation? Can generated HTML/URLs create security risk?          |
| Deletion / retention  | Do source and derived AI data have defined lifecycle? Are chunks, embeddings, cache and outputs deleted or disabled when required?                      |
| Cross-client exposure | What combination of identifiers, filters, credentials and storage could allow one client to reach another client's data?                                |

## 26.1 Security-impact escalation

Wrong-client retrieval, cross-client exposure, unauthorized agent action, successful prompt injection with material impact, secret exposure and approval-gate bypass are treated as material AI security risks and linked to the AI Incident Response Plan. Where a control failure is discovered during testing, the same failure mode is assessed across comparable systems where relevant.

## 27. Provider, Model and Supply-Chain Risk

The model and provider are dependencies within the KOMGIS system boundary. Provider assurance does not remove KOMGIS responsibility for intended purpose, client isolation, application authorization, workflow design, output use and human oversight.

| Provider/model risk                | Control and assessment considerations                                                                                                   |
|------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| Model behavior changes             | Pin/identify model where possible; version prompt/workflow; regression evaluation; provider change review; rollback or alternate model. |
| Data-use / retention term mismatch | Approved provider review; documented settings; enterprise controls; contract/DPA; client restriction check; periodic vendor review.     |
| Region / residency change          | Processing-region documentation; provider configuration; client requirement check; change trigger.                                      |
| Service outage / rate limit        | Fallback behavior; queueing; manual process; timeouts; no unsafe downgrade; continuity and recovery tests.                              |
| Model deprecation                  | Inventory of dependencies; migration plan; evaluation before replacement; client communication where material.                          |
| Subprocessor change                | Vendor/subprocessor review; contractual notification process; risk assessment update when material.                                     |
| Capability expansion               | New tool-use, vision, voice or larger context capability treated as a change when it changes risk or data exposure.                     |
| Provider security incident         | Incident intake, exposure assessment, credential rotation, client/legal review, alternate provider or temporary disable plan.           |

## 27.1 Model change as a controlled change

A model change may alter instruction following, refusals, tool selection, citation behavior, formatting, latency, context handling and hallucination patterns. The assessment must determine which risk controls and acceptance criteria require regression testing rather than assuming a stronger/newer model reduces all risks.

## 28. Evaluation, Testing and Monitoring Linkage

Every material risk must be linked to evidence that can verify the relevant controls and detect deterioration. The risk register should identify the test, monitoring metric or review signal associated with the risk.

| Risk                         | Pre-deployment evidence                                                                        | Post-deployment monitoring                                                                                    |
|------------------------------|------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Hallucinated document answer | Faithfulness test set; unsupported-claim rate; citation correctness; missing-evidence behavior | Reviewer rejection/edit rate; unsupported-claim samples; complaints; citation-open behavior where appropriate |
| Wrong-client retrieval       | Tenant isolation tests; negative object-access tests; vector-filter tests                      | Unexpected tenant/source identifiers; wrong-client reference alerts; access anomalies; incidents              |
| Prompt injection             | Direct and indirect injection cases; malicious documents; tool authorization tests             | Suspicious prompts/documents; unusual tool calls; denied permission events; export anomalies                  |
| Agent approval bypass        | Sensitive action blocked; retry/rephrase/action-splitting tests                                | Approval status vs tool-call audit; denied attempts; actions without valid approval                           |
| Extraction error             | Field precision/recall; validation failure cases; document-type segmentation                   | Validation failures; manual corrections; downstream rejection; error rate by document type                    |
| Automation bias              | Reviewer usability and source evidence test; training completion                               | Approval rate; review duration; source-open rate where available; material edit/rejection trends              |
| Provider/model regression    | Baseline regression suite and acceptance criteria                                              | Model/version tracking; KPI drift; incident and reviewer changes after release                                |

### 28.1 Risk indicators

Risk indicators should be selected because they provide an early signal of the defined scenario. Token usage, latency or generic uptime may be operationally useful but do not replace risk-specific indicators for isolation, hallucination, approval bypass or unfair impact.

### 28.2 Thresholds and action

For High and Critical systems, monitoring records should define an owner, review frequency, threshold or anomaly condition, response action and escalation route. A threshold breach may trigger investigation, temporary restriction, rollback, incident classification or reassessment depending on severity.

## 29. Residual Risk Acceptance and Deployment Decision

Residual risk acceptance is an explicit decision. The approver confirms that the system context, material risks, affected parties, controls, evidence, unresolved limitations, monitoring and treatment actions are understood.

| Residual risk level | Minimum acceptance authority                                                                 |
|---------------------|----------------------------------------------------------------------------------------------|
| Low                 | AI system owner or delegated responsible owner.                                              |
| Limited             | AI system owner; security/privacy or technical review where the risk concerns those domains. |

| Residual risk level | Minimum acceptance authority                                                                                                                                                                                                   |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| High                | AI system owner plus AI Governance / Technical Leadership and management risk acceptance; legal/compliance review where applicable.                                                                                            |
| Critical            | Executive leadership, AI Governance / Technical Leadership, security and legal/compliance as applicable. Deployment is prohibited until the formal control plan and approvals are complete. Prohibited uses remain prohibited. |

## 29.1 Acceptance record

- system and assessment ID, version and date;
- system-level internal risk classification and regulatory classification status where relevant;
- summary of High/Critical inherent risks and material impact findings;
- residual risks that remain open or accepted;
- controls and key evidence relied upon;
- deployment conditions, scope restrictions and required human oversight;
- monitoring plan and review/reassessment date;
- accepted treatment actions with due dates and owners;
- decision: Approved / Approved with conditions / Redesign required / Rejected / Retire or disable;
- approver name/role and decision date.

## 29.2 Risk acceptance expiry

Risk acceptance is tied to the assessed version and context. It is not indefinite permission for any future model, client use, tool or purpose. The acceptance expires or requires review at the defined date or when a reassessment trigger occurs.

# 30. Change, Reassessment and Continuous Review

Before a material change is released, the change owner evaluates whether existing risk scenarios, likelihood, impact, affected parties, controls, evidence, monitoring or impact conclusions change.

| Change                      | Minimum reassessment focus                                                                                                       |
|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| New model / provider        | Behavior regression, provider terms, data exposure, region, retention, model capabilities, cost/latency and fallback.            |
| New RAG source              | Authorization hierarchy, data sensitivity, source quality, injection exposure, deletion lineage, retrieval evaluation.           |
| Permission change           | Tenant/user/object enforcement, vector metadata, cached access, role-change propagation and isolation tests.                     |
| New agent tool              | Read/write capability, credential scope, authorization, validation, approval, reversibility, rate and logging.                   |
| Removed human approval      | Decision impact, automation bias, error consequence, exception handling, automated validation and legal classification.          |
| New client/industry context | Affected parties, restricted domain, data categories, contract requirements, impact assessment trigger and user competence.      |
| Prompt/workflow change      | Task behavior, refusal, extraction schema, tool selection, source use and regression acceptance criteria.                        |
| Scale increase              | Cumulative impact, monitoring capacity, reviewer workload, abuse/exfiltration volume, provider limits and incident blast radius. |

### 30.1 Delta assessment

A full assessment does not need to be rewritten for every minor change. The change record may perform a delta assessment identifying affected risks and controls. However, material changes require formal review, updated evidence and renewed acceptance where the residual risk or deployment conditions change.

## 31. Incidents, Complaints and Corrective Action

Incidents, near misses, reviewer findings and complaints are evidence about the accuracy of prior risk assumptions. The system owner must determine whether the event reveals a new risk, a higher likelihood, greater impact, ineffective control or a new affected party.

### 31.1 Mandatory risk assessment update questions

- Was the event already represented by a risk entry? If not, create one.
- Did the event occur under intended use, foreseeable misuse or an abnormal condition that should have been identified?
- Was likelihood underestimated based on testing or exposure assumptions?
- Was impact underestimated because scale, affected parties or reversibility were misunderstood?
- Did a control fail, fail open, become unavailable or prove ineffective in practice?
- Did monitoring detect the event early enough? If not, what signal was missing?
- Do similar KOMGIS systems share the same architecture, model, provider, prompt, RAG permission pattern or tool risk?
- Must residual risk acceptance, deployment conditions or the system risk classification change?

### 31.2 Corrective action linkage

Corrective actions are linked to the risk register and testing plan. A prompt change alone is not considered sufficient when root cause involves authorization, data isolation, tool permission or approval enforcement. Reactivation follows the AI Incident Response Plan and requires updated risk evidence where the incident was High or Critical.

## 32. Records, Retention and Auditability

Risk and impact assessment records are controlled documented information. The record must be sufficient to reconstruct the assessment basis and understand which system version, controls and evidence supported a deployment decision.

| Record            | Minimum content                                                                                                          |
|-------------------|--------------------------------------------------------------------------------------------------------------------------|
| Assessment header | Assessment ID, system ID, system/version, owner, technical owner, date, status, scope and reviewers.                     |
| System context    | Intended purpose, users, affected process, model/provider, data, RAG, tools, human oversight and deployment context.     |
| Screening         | Restricted/prohibited use review, legal-role/classification status and Impact Assessment trigger.                        |
| Affected parties  | Direct/indirect affected parties, material benefits and adverse impacts, vulnerability and contestability.               |
| Risk register     | Scenario, affected party/asset, inherent score, controls, evidence, residual score, owner, action status and monitoring. |
| Impact findings   | Impact area, affected group, severity/unevenness, controls, contest/remedy and conclusion.                               |
| Treatment plan    | Actions, owners, due dates, release-blocker status, evidence and verification.                                           |
| Approval          | Residual-risk summary, conditions, decision, approvers and review/expiry date.                                           |
| Change history    | Assessment revisions, material changes, incident/complaint linkage and superseded records.                               |

### 32.1 Retention

Assessment and risk-acceptance records should be retained for the operational life of the relevant AI system and for the documented audit, contractual, incident or legal retention period after retirement. Risk records must not be deleted merely because a model or prompt version is superseded when the record is needed to explain a prior output, action or incident.

### 32.2 Traceability references

The assessment should reference inventory IDs, model/workflow versions, test records, security reviews, provider records, data-processing maps, incident IDs and approval records. Sensitive data should not be copied into the assessment when a controlled reference is sufficient.

## 33. Exceptions and Enforcement

Exceptions to this methodology must be documented and approved by AI Governance / Technical Leadership. High or Critical risk controls, client-isolation requirements, prohibited uses and legal obligations cannot be bypassed through an informal project exception.

Failure to complete a required assessment, conceal a material risk, misrepresent control evidence or deploy while a blocking risk remains unresolved may result in release suspension, feature disablement, access removal, corrective action, audit escalation or other action consistent with KOMGIS policy and contractual arrangements.

Personnel are expected to raise good-faith concerns when a risk, affected party or failure scenario appears to have been missed. Business or delivery pressure must not override required assessment and approval gates.

## Appendix A. AI Risk Assessment Record Template

### A1. Assessment header

|                                    |                                                                      |
|------------------------------------|----------------------------------------------------------------------|
| <b>Assessment ID</b>               | AI-RA-_____                                                          |
| <b>AI System Inventory ID</b>      | _____                                                                |
| <b>System / feature name</b>       | _____                                                                |
| <b>System / release version</b>    | _____                                                                |
| <b>Assessment status</b>           | Draft / In review / Approved / Approved with conditions / Superseded |
| <b>Assessment owner</b>            | _____                                                                |
| <b>AI system owner</b>             | _____                                                                |
| <b>Technical owner</b>             | _____                                                                |
| <b>Security / privacy reviewer</b> | _____                                                                |
| <b>Legal / compliance reviewer</b> | _____ / N/A                                                          |
| <b>Assessment date</b>             | _____                                                                |
| <b>Next review / expiry date</b>   | _____                                                                |

### A2. System context summary

| Field                       | Assessment entry |
|-----------------------------|------------------|
| Intended purpose            | _____            |
| Out-of-scope uses           | _____            |
| Expected users              | _____            |
| Affected business process   | _____            |
| AI outputs / actions        | _____            |
| Model(s) / provider(s)      | _____            |
| Data categories             | _____            |
| RAG sources / stores        | _____            |
| Tools / integrations        | _____            |
| Human oversight             | _____            |
| Deployment context / region | _____            |
| Fallback / stop mechanism   | _____            |

### A3. Screening decisions

| Screen                                    | Decision / evidence                                                                                                           |
|-------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Prohibited-use review                     | <input type="checkbox"/> No prohibited signal <input type="checkbox"/> Escalated Reference: _____                             |
| Restricted-domain review                  | <input type="checkbox"/> Not applicable <input type="checkbox"/> Applicable / escalated Domain: _____                         |
| Regulatory role/classification assessment | Status / reference: _____                                                                                                     |
| AI Impact Assessment trigger              | <input type="checkbox"/> Not required <input type="checkbox"/> Required <input type="checkbox"/> Pending review               |
| Internal system risk classification       | <input type="checkbox"/> Low <input type="checkbox"/> Limited <input type="checkbox"/> High <input type="checkbox"/> Critical |

### A4. Risk register

| Risk ID | Risk scenario and affected party                             | Inherent L/I/Score | Controls and evidence                                                         | Residual L/I/Score | Owner / status |
|---------|--------------------------------------------------------------|--------------------|-------------------------------------------------------------------------------|--------------------|----------------|
| R-001   | Because _____, the system may _____, causing _____ to _____. | __ / __ / __       | Preventive: _____<br>Detective: _____<br>Corrective: _____<br>Evidence: _____ | __ / __ / __       | _____<br>__    |
| R-002   | Because _____, the system may _____, causing _____ to _____. | __ / __ / __       | Preventive: _____<br>Detective: _____<br>Corrective: _____<br>Evidence: _____ | __ / __ / __       | _____<br>__    |
| R-003   | Because _____, the system may _____, causing _____ to _____. | __ / __ / __       | Preventive: _____<br>Detective: _____<br>Corrective: _____<br>Evidence: _____ | __ / __ / __       | _____<br>__    |

### A5. Treatment plan

| Action ID | Risk ID | Treatment action / intended reduction | Owner | Due   | Evidence / acceptance criterion | Status |
|-----------|---------|---------------------------------------|-------|-------|---------------------------------|--------|
| A-001     | R-__    | _____<br>__                           | _____ | _____ | _____<br>__                     | Open   |
| A-002     | R-__    | _____<br>__                           | _____ | _____ | _____<br>__                     | Open   |
| A-003     | R-__    | _____<br>__                           | _____ | _____ | _____<br>__                     | Open   |

### A6. Residual risk and decision

|                              |                                  |
|------------------------------|----------------------------------|
| <b>Residual-risk summary</b> | _____<br>_____<br>_____<br>_____ |
| <b>Open accepted risks</b>   | _____<br>_____                   |
| <b>Deployment conditions</b> | _____<br>_____                   |

|                                     |                                                                                                                                                                                                            |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Monitoring / risk indicators</b> | _____                                                                                                                                                                                                      |
| <b>Decision</b>                     | <input type="checkbox"/> Approved <input type="checkbox"/> Approved with conditions <input type="checkbox"/> Redesign required <input type="checkbox"/> Rejected <input type="checkbox"/> Disable / retire |
| <b>Approver(s)</b>                  | _____                                                                                                                                                                                                      |
| <b>Decision date</b>                | _____                                                                                                                                                                                                      |

## Appendix B. AI Impact Assessment Record Template

### B1. Impact screening and purpose

|                                                        |                         |
|--------------------------------------------------------|-------------------------|
| <b>Impact Assessment ID</b>                            | AI-IA-_____             |
| <b>Related Risk Assessment ID</b>                      | AI-RA-_____             |
| <b>System / feature</b>                                | _____                   |
| <b>Business objective and intended benefit</b>         | _____<br>_____<br>_____ |
| <b>Why AI is used</b>                                  | _____<br>_____          |
| <b>Reasonable lower-impact alternatives considered</b> | _____<br>_____          |
| <b>Impact Assessment trigger</b>                       | _____<br>_____          |

### B2. Affected parties and impact findings

| Affected party / group | Relationship to AI process | Potential benefit | Potential adverse / uneven impact | Contest / correction / remedy |
|------------------------|----------------------------|-------------------|-----------------------------------|-------------------------------|
| _____                  | _____                      | _____<br>-        | _____<br>-                        | _____<br>_____                |
| _____                  | _____                      | _____<br>-        | _____<br>-                        | _____<br>_____                |
| _____                  | _____                      | _____<br>-        | _____<br>-                        | _____<br>_____                |

### B3. Material impact analysis

| Area                           | Finding / evidence / control |
|--------------------------------|------------------------------|
| Fair treatment / uneven impact | _____                        |

| Area                                              | Finding / evidence / control |
|---------------------------------------------------|------------------------------|
| Privacy / personal data                           | _____                        |
| Autonomy / manipulation / vulnerable persons      | _____                        |
| Access to opportunity / important decisions       | _____                        |
| Transparency / AI disclosure / source evidence    | _____                        |
| Human oversight and automation bias               | _____                        |
| Contestability / correction / override / reversal | _____                        |
| Accessibility / language coverage                 | _____                        |
| Safety / serious failure scenarios                | _____                        |
| Incident detection, containment and communication | _____                        |

## B4. Data and performance analysis

| Question                                   | Assessment entry |
|--------------------------------------------|------------------|
| Data provenance and original purpose       | _____            |
| Representativeness / coverage gaps         | _____            |
| Relevant performance segments              | _____            |
| Known performance differences / bias risks | _____            |
| Metrics and acceptance criteria            | _____            |
| Monitoring for uneven or cumulative impact | _____            |

## B5. Impact conclusion

|                                              |                                                                                                                                                                                |
|----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Material adverse impacts identified</b>   | _____<br>—                                                                                                                                                                     |
| <b>Controls / treatment actions required</b> | _____<br>—                                                                                                                                                                     |
| <b>Residual impact</b>                       | _____<br>—                                                                                                                                                                     |
| <b>Conclusion</b>                            | <input type="checkbox"/> Acceptable <input type="checkbox"/> Acceptable with conditions <input type="checkbox"/> Redesign / restrict <input type="checkbox"/> Must not proceed |
| <b>Required reviewers / approvers</b>        | _____<br>—                                                                                                                                                                     |
| <b>Next review date / trigger</b>            | _____<br>—                                                                                                                                                                     |

## Appendix C. Risk Register Fields and Status Rules

| Field                        | Required content                                                                                 |
|------------------------------|--------------------------------------------------------------------------------------------------|
| Risk ID                      | Unique stable identifier, e.g. RAG-001 or AGT-004.                                               |
| Risk domain                  | Taxonomy category used for reporting and review.                                                 |
| Risk scenario                | Concrete cause-event-effect statement.                                                           |
| Use condition                | Intended use / foreseeable misuse / abnormal condition.                                          |
| Affected party / asset       | Person, group, client, system, data or process that may experience impact.                       |
| Potential consequence        | Business, rights, privacy, security, safety, contractual, regulatory or operational consequence. |
| Inherent likelihood / impact | Scores before system-specific treatment controls are credited.                                   |
| Inherent score / band        | Likelihood x Impact and Low/Limited/High/Critical band.                                          |
| Preventive controls          | Controls that reduce occurrence likelihood.                                                      |
| Detective controls           | Controls and indicators that detect events or failures.                                          |
| Corrective controls          | Containment, rollback, recovery, correction or communication controls.                           |
| Evidence references          | Test IDs, code/config references, logs, review records, provider documents or approvals.         |
| Control adequacy             | E0-E4 evidence level and/or Appendix F adequacy rating.                                          |
| Residual likelihood / impact | Scores after verified controls.                                                                  |
| Residual score / band        | Residual score and risk band.                                                                    |
| Residual rationale           | Why specific controls/evidence justify the changed score.                                        |
| Risk owner                   | Named accountable owner.                                                                         |
| Treatment actions            | Open actions and action IDs.                                                                     |
| Monitoring indicator         | Metric, alert, audit or review signal linked to the scenario.                                    |
| Status                       | Open / Treating / Awaiting evidence / Accepted / Closed / Superseded / Incident-linked.          |
| Review trigger/date          | Next scheduled review or trigger condition.                                                      |

### C1. Status rules

| Status            | Rule                                                                                                                               |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------|
| Open              | Risk identified; analysis or treatment not complete.                                                                               |
| Treating          | One or more required treatment actions are in progress.                                                                            |
| Awaiting evidence | Control is implemented or claimed, but required test/operational evidence is incomplete.                                           |
| Accepted          | Authorized role accepted the documented residual risk for the defined scope and period.                                            |
| Closed            | Risk source/capability removed or treatment reduced risk to a level managed through standard controls; closure rationale recorded. |

| Status          | Rule                                                                                     |
|-----------------|------------------------------------------------------------------------------------------|
| Superseded      | Risk entry replaced by a new version or merged entry; link to successor required.        |
| Incident-linked | A related incident is open or recently closed and the risk is under formal reassessment. |

## Appendix D. AI Risk Library

The following library provides example scenarios for assessors. Entries must be adapted to the actual KOMGIS system. They are not automatically risks in every project and are not complete without affected parties, scoring, controls and evidence.

| ID     | Domain                    | Example risk scenario                                                                                                                |
|--------|---------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| PUR-01 | Purpose limitation        | Users use the AI assistant for legal, medical, financial or other sensitive advice outside the approved enterprise-document purpose. |
| PUR-02 | Scope creep               | A client configures the workflow to make final decisions even though the assessed system was approved only for decision support.     |
| ACC-01 | Unsupported claim         | The model gives a plausible answer not supported by retrieved or authoritative source material.                                      |
| ACC-02 | Incorrect extraction      | The system extracts a wrong amount, date, party or status and the structured value is used downstream.                               |
| ACC-03 | Omission                  | A summary consistently omits an exception, exclusion or condition that is material to the business process.                          |
| ACC-04 | Stale answer              | The system relies on an obsolete policy/document version and provides an outdated answer.                                            |
| RAG-01 | Cross-client retrieval    | A vector query or source permission defect retrieves content from another client workspace.                                          |
| RAG-02 | Unauthorized document     | An authenticated user retrieves content from a restricted document within the same client.                                           |
| RAG-03 | Deletion failure          | Deleted source content remains available through chunks, embeddings, cache or derived summaries.                                     |
| RAG-04 | Indirect prompt injection | A retrieved document contains instructions that influence the model or tool behavior.                                                |
| RAG-05 | Citation mismatch         | The displayed citation does not support the generated claim or points to the wrong source location.                                  |
| RAG-06 | Retrieval miss            | Relevant evidence is not retrieved, causing a confidently incomplete answer.                                                         |
| DAT-01 | Excessive context         | More client or personal data than necessary is sent to the model provider.                                                           |
| DAT-02 | Sensitive logging         | Prompts, outputs or documents containing personal/confidential data are retained in debug logs beyond need.                          |
| DAT-03 | Embedding lifecycle gap   | Embeddings and vector metadata are not deleted or disabled with the source document.                                                 |
| DAT-04 | Region mismatch           | Client data is processed in a region inconsistent with documented client or contractual requirements.                                |
| SEC-01 | Prompt injection          | User input instructs the model to ignore intended workflow or reveal hidden information.                                             |
| SEC-02 | Secret exposure           | A secret appears in a prompt, retrieved document, model output or log and remains valid.                                             |
| SEC-03 | Authorization bypass      | Frontend restriction exists but backend object/tool authorization is missing or incomplete.                                          |
| SEC-04 | Unsafe generated code     | AI-generated SQL, shell or code is executed without an approved controlled execution environment or competent review.                |
| AGT-01 | Unauthorized tool         | The agent selects or accesses a tool outside the approved use case or active user/client scope.                                      |
| AGT-02 | Wrong target              | The agent performs a valid action on the wrong client, record, recipient or environment.                                             |

| ID     | Domain                     | Example risk scenario                                                                                                   |
|--------|----------------------------|-------------------------------------------------------------------------------------------------------------------------|
| AGT-03 | Approval bypass            | The agent retries, rephrases or splits a sensitive action to avoid the required approval gate.                          |
| AGT-04 | Duplicate action           | Retries or partial failures cause duplicate email, payment, submission or record update.                                |
| AGT-05 | Goal misalignment          | The agent achieves the assigned goal through a technically allowed but business-inappropriate action.                   |
| HUM-01 | Automation bias            | Reviewer accepts a confident AI output without checking available source evidence.                                      |
| HUM-02 | Rubber-stamping            | Review volume and time pressure make meaningful human review impractical.                                               |
| HUM-03 | Reviewer authority gap     | Reviewer can identify a problem but cannot stop, override or escalate the workflow.                                     |
| BIA-01 | Language disparity         | Performance is materially worse for one language or script used by affected users.                                      |
| BIA-02 | Document-type disparity    | Extraction or retrieval quality is materially worse for scanned, legacy or uncommon document formats.                   |
| BIA-03 | Historical bias            | Historical records or labels reproduce inconsistent or unfair human outcomes.                                           |
| TRA-01 | Missing AI notice          | A direct user reasonably believes they are interacting with a human when required transparency is absent.               |
| TRA-02 | Insufficient traceability  | Model/workflow version, source IDs or approval state cannot be reconstructed after a material incident.                 |
| PRV-01 | Provider term mismatch     | Client confidential data is sent to a provider whose training-use or retention terms are not approved for the use case. |
| PRV-02 | Model regression           | A model change worsens grounding, tool selection or structured-output behavior without adequate regression testing.     |
| OPS-01 | Unsafe fallback            | A provider outage causes the system to use an unassessed model or skip required AI/human validation.                    |
| OPS-02 | Monitoring failure         | Risk indicators stop collecting or alert ownership is unclear, allowing repeated failures to continue.                  |
| OPS-03 | Scale amplification        | A low-frequency error creates material cumulative impact after client/user volume increases.                            |
| CHG-01 | Unreviewed workflow change | Prompt/orchestration change materially alters behavior but is released without risk/test review.                        |
| INC-01 | Silent correction          | A serious AI error is corrected manually but not reported, preventing risk and control updates.                         |

## Appendix E. Scoring Matrix and Decision Rules

### E1. 5 x 5 matrix

| Likelihood \ Impact | 1 Negligible | 2 Minor      | 3 Moderate   | 4 Major        | 5 Severe       |
|---------------------|--------------|--------------|--------------|----------------|----------------|
| 5 Almost certain    | 5<br>Limited | 10<br>High   | 15<br>High   | 20<br>Critical | 25<br>Critical |
| 4 Likely            | 4<br>Low     | 8<br>Limited | 12<br>High   | 16<br>High     | 20<br>Critical |
| 3 Possible          | 3<br>Low     | 6<br>Limited | 9<br>Limited | 12<br>High     | 15<br>High     |
| 2 Unlikely          | 2<br>Low     | 4<br>Low     | 6<br>Limited | 8<br>Limited   | 10<br>High     |
| 1 Rare              | 1<br>Low     | 2<br>Low     | 3<br>Low     | 4<br>Low       | 5<br>Limited   |

## E2. Decision rules

- Use the highest credible impact dimension for the defined scenario.
- Do not reduce impact merely because a human is present; assess whether the human can detect and prevent the specific effect.
- Do not credit planned controls as verified controls.
- For a control to reduce likelihood, explain how it prevents or detects the event before material impact.
- For a control to reduce impact, explain how it limits scale, duration, exposure, irreversibility or consequence.
- Escalation indicators may raise the system-level risk classification above the numeric band.
- Uncertainty may justify a higher likelihood/impact or a conservative classification until evidence is available.
- Prohibited uses cannot be approved by accepting a numeric residual risk.

## Appendix F. Control Adequacy Scale

| Rating | Control adequacy           | Assessment guidance                                                                                                               |
|--------|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| 0      | Absent                     | No control or the described control is unrelated to the risk scenario.                                                            |
| 1      | Planned                    | Control is specified but not implemented. Do not rely on it as an effective residual-risk reduction.                              |
| 2      | Implemented                | Control exists in code, configuration, process or contract, but representative effectiveness testing is incomplete.               |
| 3      | Verified                   | Control implementation and representative test evidence support the intended risk reduction; known material defects are resolved. |
| 4      | Operationally demonstrated | Verified control also has meaningful production monitoring/audit evidence showing continued operation under relevant exposure.    |

For High and Critical risks, key preventive and approval controls should normally reach rating 3 before production approval. Rating 4 is developed through operation and is used during periodic reassessment. The scale is not a substitute for judging whether the control design is appropriate to the scenario.

## Appendix G. Example Assessment - Enterprise RAG Chatbot

The example is illustrative and demonstrates the level of specificity expected. It is not approval for any real client deployment.

### G1. System context

|                         |                                                                                                                                        |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| <b>System</b>           | Enterprise Document Intelligence Chatbot - example deployment                                                                          |
| <b>Intended purpose</b> | Allow authorized client employees to ask questions about internal documents and receive source-grounded answers with inline citations. |
| <b>Users</b>            | Authenticated client employees; access varies by Project -> Field -> Section -> Document permissions.                                  |
| <b>Output / action</b>  | Text answer and citations. No write tools or automatic external actions.                                                               |
| <b>Data</b>             | Client documents, extracted text, chunks, embeddings, user prompts, retrieved context and AI outputs.                                  |
| <b>RAG</b>              | Client-scoped vector/hybrid retrieval with user/document metadata filters before                                                       |

|                            |                                                                                                       |
|----------------------------|-------------------------------------------------------------------------------------------------------|
|                            | prompt construction.                                                                                  |
| <b>Human oversight</b>     | User verifies cited source for material business use. Support can disable affected workspace/feature. |
| <b>Internal risk class</b> | Limited (example), subject to client data sensitivity and actual deployment context.                  |

## G2. Example risk register

| ID     | Scenario                                                                                                                                                                                                                                                  | Affected parties / impact                                        | Inherent          | Controls                                                                                                                                                                                                                                  | Evidence                                                                                             | Residual          |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|-------------------|
| RAG-01 | Because retrieval permission metadata may be missing or incorrectly mapped after a role/document change, the chatbot may retrieve a restricted chunk for an otherwise authenticated user, disclosing confidential information to the wrong internal user. | Client, document owner, data subjects; confidentiality and trust | 3 / 4 / 12 High   | Mandatory client_id + permitted document ID filters before vector query; backend object authorization; fail closed if permission context absent; client-scoped vector metadata; isolation and role-change tests; wrong-source monitoring. | E3: negative object-access tests and cross-role retrieval suite passed; source IDs logged.           | 1 / 4 / 4 Low     |
| RAG-02 | Because a retrieved document may contain malicious instructions, the model may treat untrusted document text as instructions and produce manipulated output or attempt to influence a future tool-connected workflow.                                     | User, client, connected systems                                  | 3 / 4 / 12 High   | Retrieved content treated as data; trusted instructions separated; no write tools in this deployment; output grounded to source; malicious-document test set; suspicious retrieval/input review.                                          | E3: indirect injection cases tested; system has no action tools.                                     | 2 / 3 / 6 Limited |
| ACC-01 | Because retrieval may miss a relevant exception or the model may synthesize beyond evidence, the chatbot may provide a confident but unsupported business answer.                                                                                         | User and client business process                                 | 4 / 3 / 12 High   | Inline citations; faithfulness evaluation; instruction to identify insufficient evidence; source-opening UI; representative retrieval tests; user limitation notice; monitor rejection/complaints.                                        | E3: faithfulness and citation suite met release criteria; known difficult document types documented. | 2 / 3 / 6 Limited |
| DAT-03 | Because source deletion may not propagate to all derived records, deleted content may remain retrievable through chunks or embeddings.                                                                                                                    | Client, data subjects                                            | 3 / 3 / 9 Limited | Stable source/chunk lineage; deletion job disables chunks and vectors; deletion verification test; source-status filter in retrieval; reconciliation monitoring.                                                                          | E3: deletion integration tests passed; reconciliation job configured.                                | 1 / 3 / 3 Low     |
| TRA-02 | Because model/workflow/source version metadata may be incomplete, KOMGIS may be unable to reconstruct a material answer during an incident or client review.                                                                                              | Client, KOMGIS, support/investigation process                    | 2 / 3 / 6 Limited | Log system ID, model, workflow version, user/workspace, source IDs and timestamp; restricted audit-log access; incident correlation ID.                                                                                                   | E2/E3: logging implementation verified and sample reconstruction test completed.                     | 1 / 3 / 3 Low     |

## G3. Example impact-screening conclusion

For the stated example, the chatbot provides document-grounded answers to authenticated employees and does not itself make decisions about natural persons or take external actions. A full AI Impact Assessment may be case-by-case rather than automatically required. However, the screening must be repeated if the client uses the output in employment, insurance, health, eligibility or another important decision process, or if the system is configured to take actions.

## G4. Example residual-risk decision

Example decision: Approved with conditions. Conditions: maintain permission-aware retrieval and cross-role isolation tests as release blockers; retain inline citations; monitor unsupported-claim and wrong-source signals; update the assessment before adding write tools, a new provider, a materially different model family, new client data categories or removal of source-grounding controls. Residual Limited risks are accepted by the AI system owner for the assessed version and context.

## Appendix H. Assessment Approval Checklist

| #  | Review item                                                                                                                        | Complete / Evidence |
|----|------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 1  | The system is listed in the AI System Inventory and has named system and technical owners.                                         | [ ] _____           |
| 2  | The intended purpose, out-of-scope uses, users, affected process and deployment context are specific and documented.               | [ ] _____           |
| 3  | The system boundary includes model, RAG, data, permissions, tools, integrations, human oversight and downstream process.           | [ ] _____           |
| 4  | Known and reasonably foreseeable risk scenarios have been identified for intended use, foreseeable misuse and abnormal conditions. | [ ] _____           |
| 5  | Affected parties include direct users and indirect subjects of decisions, communications, data or actions where relevant.          | [ ] _____           |
| 6  | Prohibited and restricted use signals have been screened and escalated where required.                                             | [ ] _____           |
| 7  | KOMGIS legal role and regulatory classification status have been assessed where applicable.                                        | [ ] _____           |
| 8  | The AI Impact Assessment trigger has been evaluated and the assessment completed where required.                                   | [ ] _____           |
| 9  | Inherent likelihood and impact scores are supported by stated assumptions and exposure.                                            | [ ] _____           |
| 10 | Preventive, detective and corrective controls are described at their enforcement points.                                           | [ ] _____           |
| 11 | Planned controls are not incorrectly credited as verified residual-risk reductions.                                                | [ ] _____           |
| 12 | Key controls have implementation and representative test evidence appropriate to risk.                                             | [ ] _____           |
| 13 | Tenant, user and object authorization is tested where client data is processed.                                                    | [ ] _____           |
| 14 | RAG permission filtering occurs before prompt construction and cross-client isolation is tested where RAG is used.                 | [ ] _____           |
| 15 | Retrieved documents and tool outputs are treated as untrusted and prompt-injection tests are complete where relevant.              | [ ] _____           |
| 16 | Agent tools are explicitly approved, least-privilege scoped and checked server-side before sensitive calls.                        | [ ] _____           |
| 17 | Sensitive actions are blocked pending required human approval; approval bypass and retry behavior are tested.                      | [ ] _____           |
| 18 | Human reviewers have competence, evidence access, authority and realistic workload for the risk they are expected to control.      | [ ] _____           |
| 19 | Data provenance, representativeness, language/document coverage and uneven performance have been assessed where material.          | [ ] _____           |
| 20 | Privacy, provider exposure, logs, embeddings, retention, deletion and region requirements are addressed.                           | [ ] _____           |
| 21 | Each material risk is linked to evaluation, test or monitoring evidence and an accountable risk owner.                             | [ ] _____           |
| 22 | Residual likelihood, impact and rationale are documented after controls.                                                           | [ ] _____           |
| 23 | Open treatment actions have owners, due dates, evidence criteria and release-blocker status.                                       | [ ] _____           |

| #  | Review item                                                                                                                   | Complete / Evidence |
|----|-------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 24 | Monitoring indicators, thresholds/anomaly conditions, alert owners and escalation routes are defined for High/Critical risks. | [ ] _____           |
| 25 | The incident route, disable/rollback mechanism and evidence-preservation capability are ready.                                | [ ] _____           |
| 26 | Reassessment triggers and the next review/acceptance-expiry date are documented.                                              | [ ] _____           |
| 27 | Residual risks are accepted by the authorized role and deployment conditions are recorded.                                    | [ ] _____           |

## Appendix I. Alignment Overview

The table below is a high-level design alignment for KOMGIS governance. It is not a certification or legal-compliance statement and does not replace system-specific applicability analysis.

| Assessment area            | KOMGIS AI management theme                                                             | Applicable regulatory / governance theme                 |
|----------------------------|----------------------------------------------------------------------------------------|----------------------------------------------------------|
| Context and purpose        | System inventory, intended purpose, users, role, data and lifecycle context            | Risk-based governance; system scope and intended purpose |
| Risk identification        | Known and foreseeable risks, misuse, abnormal conditions, affected parties             | AI risk management and impact identification             |
| Impact assessment          | People, rights, safety, important decisions, uneven impact and contestability          | Fundamental-rights / impact analysis where applicable    |
| Risk analysis              | Likelihood, impact, inherent risk and system classification                            | Proportionate risk evaluation                            |
| Risk treatment             | Preventive, detective and corrective controls; owners and due dates                    | Risk treatment and operational controls                  |
| Human oversight            | Competence, evidence, authority, workload, override and stop capability                | Human oversight and AI literacy                          |
| Data and RAG               | Data provenance, minimization, permission-aware retrieval, source lineage and deletion | Data governance, privacy and security                    |
| Evaluation and evidence    | Metrics, representative testing, adversarial testing and control evidence              | Accuracy, robustness, cybersecurity and validation       |
| Residual risk acceptance   | Authorized decision, conditions, monitoring and review date                            | Management accountability and documented information     |
| Monitoring and improvement | Risk indicators, incidents, complaints, change review and corrective action            | Post-deployment monitoring and continual improvement     |

## Reference Documents

- KOMGIS Responsible AI and AI Management Policy - parent policy for AI governance, risk classification, risk/impact assessment and lifecycle controls.
- KOMGIS AI Security Policy - access control, prompt injection, RAG, agent tools, secrets, output security and secure development.
- KOMGIS Data Privacy Policy - client data, AI inputs and outputs, embeddings, providers, retention, deletion and data requests.
- KOMGIS AI Incident Response Plan - incident severity, containment, investigation, communication, corrective action and reactivation.

- KOMGIS AI System Inventory - system ownership, intended use, models, data, risk and operational status.
- Evaluation and Testing Records - pre-deployment and ongoing quality, security and safety evidence.
- Human Oversight Record - review and approval of sensitive AI outputs or actions.

Assessment interpretation questions shall be escalated to AI Governance / Technical Leadership. Legal or regulatory interpretation must be referred to qualified legal or compliance review where required.