

AI Risk-Tiering Worksheet

A structured insurance classification of inherent risk, mandatory overrides, and minimum governance controls.

16	4	6
inherent-risk factors	risk tiers	mandatory overrides

Version 1.0 | Printable assessment workbook

Classify the exact AI system and insurance use before approving a pilot or production deployment. Reassess when the use, data, model, autonomy, jurisdiction, or control environment materially changes.

How to use this worksheet

Define the review unit: the exact model or AI system, version, insurance purpose, users, affected people, decisions, data, integrations, tools, jurisdictions, and level of autonomy. Rate inherent risk before planned controls.

Scoring method

Score	Meaning	Use
0	Limited	Low exposure or consequence.
1	Low	Bounded exposure with routine controls.
2	Material	Meaningful harm or control challenge is credible.
3	Severe	Consequential, autonomous, difficult-to-reverse, or high-scale exposure.

Mandatory overrides

ID	Override question	Effect
OV-01	Is the proposed use prohibited, not legally supportable, outside risk appetite, or inconsistent with approved company policy?	RESTRICTED
OV-02	Can the system determine or execute an adverse or consequential consumer outcome without meaningful human review before the outcome?	TIER 4
OV-03	Can the system autonomously bind coverage, change policy terms, approve or deny a claim, set or transfer material funds, execute settlement, or make another legally binding commitment?	TIER 4
OV-04	Could failure create physical safety risk, prevent a critical insurance service, or produce severe correlated impact across a large portfolio or population?	TIER 4
OV-05	Does the use materially influence underwriting, pricing, coverage, claims, fraud, employment, or another consequential decision using sensitive data, external consumer data, inferred attributes, or potential proxies?	MINIMUM TIER 3
OV-06	Can a generative or agentic system use external tools, write to production records, send external communications, or act on untrusted content?	MINIMUM TIER 3

If OV-01 is Yes, classify the use as Restricted and do not deploy. Other overrides establish a minimum tier. Use the higher of the score-based tier and the override tier.

Consumer and business impact

Rate each factor from 0 to 3 using the anchors shown. Record the evidence and rationale. When uncertainty is material, use the higher defensible rating until evidence resolves it.

IB-01	Consumer impact How directly can the use affect an applicant, policyholder, claimant, beneficiary, producer, or employee?	Rating anchors 0 No individual impact; internal support only 1 Indirect or informational impact 2 Material influence on service, prioritization, or decision 3 Determines or executes an eligibility, pricing, coverage, claim, fraud, employment, or other consequential outcome	Rating: Evidence:
IB-02	Decision materiality How important is the output to the final insurance or business decision?	Rating anchors 0 Output is optional and not used for decisions 1 One low-weight input with independent review 2 Substantial input that normally changes decisions 3 Primary or determinative basis for the decision	Rating: Evidence:
IB-03	Severity of plausible harm What is the most credible harm if the system is wrong, biased, unavailable, manipulated, or misused?	Rating anchors 0 Negligible inconvenience with no lasting effect 1 Limited delay, rework, or small financial effect 2 Material financial, coverage, privacy, legal, or reputational harm 3 Severe or widespread consumer harm, major loss, loss of rights, or physical safety impact	Rating: Evidence:
IB-04	Scale and concentration How many decisions, transactions, users, or business processes can be affected before detection?	Rating anchors 0 Small pilot or infrequent use 1 One team or limited population with bounded volume 2 Enterprise or multi-state use with significant volume 3 Mass-scale, concentrated, correlated, or rapidly propagating impact	Rating: Evidence:

Autonomy and human control

Rate each factor from 0 to 3 using the anchors shown. Record the evidence and rationale. When uncertainty is material, use the higher defensible rating until evidence resolves it.

AH-01	Level of autonomy What can the system do without a person initiating or approving each material action?	Rating anchors 0 Generate drafts or analysis only 1 Recommend an action for human approval 2 Execute bounded workflow steps or communications 3 Act autonomously, transact, bind, approve, deny, settle, or change records	Rating: Evidence:
AH-02	Human review effectiveness How meaningful and reliable is human review before a material outcome?	Rating anchors 0 Qualified reviewer has time, evidence, authority, and routinely challenges output 1 Reviewer approves with adequate context and tested override process 2 Review is rushed, sampled, inconsistent, or prone to automation bias 3 No meaningful review before the outcome, or reviewer cannot understand or change it	Rating: Evidence:
AH-03	Reversibility How easily can an affected outcome be stopped, corrected, or reversed?	Rating anchors 0 Immediate, complete, and low-cost reversal 1 Reversible through a routine process with limited residual effect 2 Correction requires escalation or may leave material residual harm 3 Irreversible, time-critical, legally binding, or difficult to remediate	Rating: Evidence:
AH-04	Time to intervene How much time exists to identify and control a harmful output or action?	Rating anchors 0 Days or longer, with review before use 1 Hours, with effective alerts and intervention 2 Minutes or high-volume same-day activity 3 Seconds or autonomous propagation before meaningful intervention	Rating: Evidence:

Data and model characteristics

Rate each factor from 0 to 3 using the anchors shown. Record the evidence and rationale. When uncertainty is material, use the higher defensible rating until evidence resolves it.

DM-01	Data sensitivity What data does the system access, infer, generate, retain, or expose?	Rating anchors 0 Public, synthetic, or low-sensitivity operational data 1 Confidential business data with limited personal information 2 Personal, claims, health, financial, geolocation, biometric, or similarly sensitive data 3 Highly sensitive data at scale, protected-class data or proxies, or data whose misuse could cause severe harm	Rating: _____ Evidence: _____ _____ _____
DM-02	External data and proxy risk How dependent is the use on external consumer data, inferred attributes, or variables that may proxy protected characteristics?	Rating anchors 0 No external consumer data or inferred personal attributes 1 Limited verified external data with low proxy potential 2 Material external data or inferred attributes requiring proxy testing 3 Opaque, unverified, or highly correlated data used in consequential decisions	Rating: _____ Evidence: _____ _____ _____
DM-03	Model complexity and change How difficult is the system to validate, explain, reproduce, and control as it changes?	Rating anchors 0 Deterministic rule or simple stable model 1 Interpretable model with controlled periodic updates 2 Complex model, ensemble, continuous data change, or limited transparency 3 Adaptive, self-modifying, opaque, or externally controlled model with material behavior changes	Rating: _____ Evidence: _____ _____ _____
DM-04	Generative and agentic behavior Does the system generate open-ended content, retrieve untrusted information, call tools, retain memory, or pursue goals?	Rating anchors 0 No generative or agentic functionality 1 Closed drafting or summarization with approved sources 2 Open-ended generation, retrieval, memory, or limited tools in a controlled environment 3 Agentic planning, external tools, write access, untrusted content, or autonomous multi-step action	Rating: _____ Evidence: _____ _____ _____

Operational and control exposure

Rate each factor from 0 to 3 using the anchors shown. Record the evidence and rationale. When uncertainty is material, use the higher defensible rating until evidence resolves it.

OC-01	System and tool access What systems, records, communications, funds, or external services can the AI reach?	Rating anchors 0 Isolated environment with read-only low-risk access 1 Read access to internal systems or bounded sandbox 2 Write access to material records, communications, or workflows 3 Privileged, production, financial, policy, claim, or externally facing access across systems	Rating: Evidence:
OC-02	Reliability and resilience What happens when the model, data, vendor, integration, or control fails?	Rating anchors 0 Failure is obvious and work stops safely 1 Manual fallback exists and recovery is routine 2 Degraded behavior can create material backlog, error, or inconsistent outcomes 3 Failure can silently propagate, prevent critical service, or cause severe correlated loss	Rating: Evidence:
OC-03	Transparency and contestability Can the insurer explain the use and outcome, preserve evidence, and support correction, complaint, appeal, audit, and examination?	Rating anchors 0 Complete trace and clear explanation are readily available 1 Adequate records and explanation with minor manual work 2 Material explanation, trace, retention, or appeal limitations 3 Outcome cannot be reconstructed or explained sufficiently for affected people or regulators	Rating: Evidence:
OC-04	Security and misuse potential How attractive and capable is the system as a target or mechanism for fraud, manipulation, exfiltration, or unauthorized action?	Rating anchors 0 Minimal attack surface and no meaningful action capability 1 Standard application exposure with mature controls 2 Sensitive data, external content, valuable decisions, or material tool access 3 High-value or privileged target with scalable fraud, exfiltration, evasion, or destructive-action potential	Rating: Evidence:

Minimum control requirements

The final tier sets the minimum governance posture. Apply stricter requirements when law, company policy, jurisdiction, consumer impact, vendor limitations, or residual risk warrants them.

Control	Tier 1	Tier 2	Tier 3	Tier 4
Accountable business owner and inventory record	Required	Required	Required	Required
Documented purpose, boundaries, users, data, and prohibited uses	Basic	Detailed	Detailed	Executive-approved
Legal, compliance, privacy, and security review	As needed	Targeted	Required	Required plus executive review
Independent validation and challenge	Optional	Risk-based	Required	Required with enhanced independence
Fairness, proxy, and consumer-outcome testing	If relevant	If relevant	Required for consequential use	Required before and during use
Human oversight and override testing	Basic	Documented	Tested	Demonstrably effective and monitored
Security, abuse, and prompt-injection testing	Baseline	Targeted	Required	Independent red team or equivalent
Production monitoring and incident response	Basic	Defined metrics	Thresholds and escalation	Continuous or near-real-time
Change approval and reassessment	Material changes	Defined changes	Formal approval	Formal approval plus rollback test
Approval authority	Business owner	Business and control owners	AI governance committee	Executive risk committee or delegated governing body
Reassessment cadence	At material change	At least annually	At least semiannually	At least quarterly and after triggers

Controls should address the actual system and use, not only the underlying model. Document who owns each control, the evidence of operation, open conditions, and the approval authority.

Classification and approval record

Preserve this record with the AI inventory entry, assessment evidence, validation, approval, conditions, and monitoring plan.

Decision element	Record
AI system or use case	
Business owner	
Purpose, users, decisions, and jurisdictions	
Total inherent-risk score	/ 48
Score-based tier	Tier 1 / Tier 2 / Tier 3 / Tier 4
Overrides triggered	
Final classification	Restricted / Tier 1 / Tier 2 / Tier 3 / Tier 4
Key inherent risks and planned controls	
Residual risk and conditions	
Approver and decision date	
Next review date and reassessment triggers	

Official reference foundations

NAIC Model Bulletin on the Use of Artificial Intelligence Systems by Insurers; NAIC Artificial Intelligence resources; NIST AI Risk Management Framework and AI RMF Core; New York DFS Insurance Circular Letter No. 7 of 2024. Confirm current requirements in every applicable jurisdiction.

Use note: This worksheet is an operating aid, not a legal determination. A lower tier does not waive applicable insurance, consumer-protection, privacy, cybersecurity, recordkeeping, anti-discrimination, or market-conduct obligations.