

Human-Oversight Decision Matrix

A printable assessment for selecting human command, pre-action approval, active monitoring, periodic review, or redesign.

10	6	5
inherent-risk factors	mandatory overrides	operating modes

Version 1.0 | Printable assessment workbook

Assess the exact AI system, insurance use, affected decision, population, data, autonomy, workflow, and reviewer capacity. Rate inherent exposure before planned controls.

Instructions and decision rules

Rate all ten factors from 0 to 3. Add the ratings to determine the score-based mode. Then complete all mandatory overrides. The final recommendation is the stricter result. An override requiring redesign takes precedence over every oversight label.

Score	Score-based mode	Operating principle
0-6	Automated with periodic review	Narrow, low-impact operation with sampling, monitoring, exceptions, and change review.
7-12	Human on the loop	A qualified person monitors bounded operation and can intervene promptly.
13-20	Human in the loop	The system pauses before the defined action and a qualified person approves, rejects, or modifies it.
21-30	Human in command	A qualified person determines the outcome and uses AI as bounded support.
Override	Redesign before use	Change scope, autonomy, evidence, workflow, or controls before pilot or production.

A named person is not automatically meaningful oversight. The person needs relevant evidence, enough time, training, independence, authority, manageable workload, and a practical way to disagree, stop, correct, and escalate.

Assessment factors 1 through 5

Use the highest defensible inherent-risk anchor. Record evidence and explain material uncertainty.

ID	Risk factor and question	0	1	2	3	Rating and evidence
HO-01	Decision consequence How consequential is the decision or action influenced by the AI?	Administrative support only	Advisory with limited effect	Material influence on an insurance outcome	Determines or executes a consequential outcome	Rating: ____ Evidence:
HO-02	Consumer or rights impact What harm could a wrong, delayed, or unfair result cause?	Negligible and easily corrected	Limited inconvenience or internal rework	Material financial, coverage, service, privacy, or fairness harm	Severe loss of rights, benefits, coverage, safety, or timely remedy	Rating: ____ Evidence:
HO-03	Autonomy and action authority What can the system do without a person approving the specific action?	Draft or analyze only	Recommend or prioritize	Initiate bounded action with confirmation or rapid reversal	Execute binding, external, financial, coverage, or multi-step action	Rating: ____ Evidence:
HO-04	Reversibility How completely and quickly can an error be contained and corrected?	Fully reversible before effect	Reversible with routine correction	Difficult, costly, delayed, or externally visible reversal	Irreversible or remedy cannot restore the person fully	Rating: ____ Evidence:
HO-05	Time to intervene How much time exists for effective review before harm or a deadline?	Days or no time pressure	Hours with a manageable queue	Minutes or high-volume processing	Seconds or action occurs before practical intervention	Rating: ____ Evidence:

Assessment factors 6 through 10

Use the highest defensible inherent-risk anchor. Record evidence and explain material uncertainty.

ID	Risk factor and question	0	1	2	3	Rating and evidence
HO-06	Review evidence and explainability Can a reviewer understand the recommendation and see the evidence needed to challenge it?	Clear evidence and reason	Mostly understandable with bounded uncertainty	Material opacity or incomplete evidence	Reviewer cannot reconstruct or meaningfully challenge the basis	Rating: ____ Evidence:
HO-07	Error detectability How likely is a material failure to be detected before impact?	Obvious and automatically checked	Usually detectable through routine review	Subtle, delayed, subgroup-specific, or dependent on complaints	Silent, systemic, or likely to evade available monitoring	Rating: ____ Evidence:
HO-08	Scale and propagation How broadly or quickly can one error pattern affect people or operations?	Isolated low-volume use	Limited team or portfolio segment	High volume, multiple jurisdictions, or material concentration	Enterprise, systemic, or rapidly cascading effect	Rating: ____ Evidence:
HO-09	Data and vulnerability Does the use involve sensitive data or exposure to manipulation and untrusted content?	Low-sensitivity controlled data	Confidential data in controlled workflow	Sensitive consumer data, external data, or untrusted inputs	Highly sensitive data plus material security, injection, fraud, or abuse exposure	Rating: ____ Evidence:
HO-10	Operational review capacity Can qualified reviewers perform the required review consistently under real workload?	Adequate time, training, authority, and coverage	Minor constraints with tested backup	Material workload, skill, evidence, independence, or escalation gaps	Review cannot be performed reliably in the intended workflow	Rating: ____ Evidence:

Mandatory overrides

Answer every override after scoring. Apply the stricter outcome. Do not average or offset an override.

ID	Mandatory override question	Effect if Yes	Yes or No and rationale
OV-01	Human decision required Does law, regulation, filed practice, contract, company policy, or risk appetite require a qualified person to make or approve the final decision?	Human in command	Response: _____ Rationale: _____
OV-02	Consequential adverse decision Can the use deny, terminate, materially restrict, price, settle, or delay coverage, benefits, payment, appeal, or another consequential consumer outcome?	Human in command	Response: _____ Rationale: _____
OV-03	Binding or irreversible action Can the system execute an external, financial, coverage, legal, safety, or difficult-to-reverse action?	At least human in the loop	Response: _____ Rationale: _____
OV-04	Unbounded agent authority Can an agent select tools, expand its plan, exceed defined limits, or act across systems without reliable confirmation and containment?	Redesign before use	Response: _____ Rationale: _____
OV-05	Ineffective review design Would the reviewer lack necessary evidence, time, training, independence, authority, or a practical way to disagree and stop the process?	Redesign before use	Response: _____ Rationale: _____
OV-06	Safety or urgent rights Could action before qualified human review cause physical safety risk, loss of legal rights, missed deadlines, or harm that later correction cannot repair?	Human in command	Response: _____ Rationale: _____

Minimum operating requirements

The recommendation sets a minimum. Add stricter controls when the use, jurisdiction, consumer impact, residual risk, or company policy requires them.

Control	Periodic review	On the loop	In the loop	In command
Authority	Narrow approved limits.	Bounded operation with stop authority.	Human approves each defined action.	Human determines outcome.
Timing	Periodic and event-driven.	Continuous or near-real-time.	Before execution.	Throughout judgment and before action.
Evidence	Logs, samples, validation, changes.	Live status, alert context, exposure.	Recommendation, source, reason, uncertainty.	Complete case record and bounded AI contribution.
Testing	Sampling, exceptions, drift, stop.	Alert latency, coverage, containment, rollback.	Workload, override, queue, outage, human factors.	Automation bias, second review, explanation, appeal.
Failure	Route exception and suspend at threshold.	Stop, contain, reverse, preserve evidence.	Hold action and route to qualified reviewer.	Human resolves; never default to AI outcome.

Review-design test

Question	Documented answer and evidence
What exactly must the person decide, approve, monitor, stop, or correct?	
What evidence, reason, uncertainty, and alternatives are visible?	
How much time and workload does realistic review require?	
How does the person disagree, override, escalate, and preserve rationale?	
How are absence, surge, outage, fatigue, and conflict handled?	
Which metrics prove the control works in production?	

Illustrative insurance scenarios

These are starting points, not classifications. Complete the assessment for the exact system, configuration, population, jurisdiction, and workflow.

Insurance use	Starting mode	Key intervention	Increase oversight when
Internal meeting summary	Periodic review	Sample, correct, protect confidentiality.	Sensitive content or operational reliance.
Claim document classification	On the loop	Queue alerts, exception routing, service threshold.	Deadlines, vulnerable claimants, poor detectability.
Fraud referral score	In the loop	Reviewer sees evidence and may reject referral.	Opaque evidence, disparity, or default acceptance.
Coverage recommendation	In command	Qualified adjuster interprets policy and facts.	Any execution or inability to explain and correct.
Settlement range assistant	In the loop	Adjuster approves and records rationale.	Binding authority, time pressure, vulnerable claimant.
Autonomous low-value payment	On the loop	Hard limits, anomaly alerts, rollback and audit.	Coverage judgment, broad authority, weak containment.
Underwriting eligibility	In command	Qualified underwriter owns final decision.	External data, opacity, unfair discrimination risk.
Customer communication agent	In the loop	Approval for coverage, claim, cancellation, or rights content.	Direct sending, account changes, weak escalation.

Recommendation and approval record

Decision element	Record
AI system and use	
Business owner and reviewer	
Inherent-risk score	/ 30
Score-based mode	
Overrides triggered	
Final recommendation	
Intervention point and decision authority	
Evidence presented to reviewer	
Workload, coverage, escalation, and stop authority	
Monitoring and test evidence	
Approval, conditions, and review date	

Official references

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