



AI for insurance operations

Supporting underwriting, claims, service and compliance work with controlled, explainable AI.

WORKS WITH THE ASSISTANTS TEAMS ALREADY USE



EDITION 43 / EMBER & INDIGO

EXECUTIVE WORKING EDITION - 08 OCT 2026

43

AT A GLANCE

Supporting underwriting, claims, service and compliance work with controlled, explainable AI.

Insurance runs on documents, judgment and promises made to customers at their worst moments. AI can take on the reading, sorting, summarizing and drafting that surrounds those judgments, so skilled people spend their time on decisions rather than on retyping. It must also operate inside a regulated setting where decisions about coverage and claims need to be explainable, fairly applied, recorded and owned by an accountable person. This playbook covers where AI fits across submission intake, claims triage, document handling, service and quality review, how to design human decision rights, how to measure value honestly, and how to move from a narrow pilot to a controlled rollout.

WHO IT IS FOR Heads of underwriting, claims, policy service and operations, compliance and risk leads, and the technology and data owners at insurers, MGAs, brokers and third-party administrators who are deciding where AI belongs.

TIME TO USE 75 minutes to read; two workshops to apply

What you will have at the end

- A shortlist of insurance workflows where AI assists and a clear list where it must not decide
- A human decision-rights map for coverage, claims and pricing-related outcomes
- A control set covering explainability, fair treatment, record keeping and vendor oversight
- A baseline and metric set that separates speed, quality and customer outcomes
- A 90-day plan from one assisted workflow to a documented decision to scale

Contents

01	What AI changes in insurance operations, and what it must not	03
02	Where AI fits in insurance, and where it does not	04
03	How these systems work, in plain language	06
04	Operating model, roles and decision rights	08
05	Risks, controls and human review	10
06	Measuring quality and value	12
07	Implementation guide: from one workflow to controlled rollout	14
08	Pitfalls and a hypothetical scenario	16
09	Questions to ask before you buy or build	17
10	A first quarter that produces evidence	19
11	Terms used in this publication	20

EXECUTIVE BRIEF

What AI changes in insurance operations, and what it must not

01



AI is most valuable in insurance as a reader, sorter and drafter that prepares work for people, and least defensible as the party that decides a customer's coverage or claim.

An insurance operation is a chain of information handling. A broker sends a submission with a dozen attachments. An underwriter reads, extracts, compares to appetite and pricing guidance, and asks questions. A claim arrives as a phone call, an email or a form, with photos, estimates, medical or repair documents and policy wording that must be matched to the facts. Service teams answer the same questions about documents, billing and endorsements all day. In each link, a large share of the time goes to reading and rekeying rather than to the judgment that the role exists to apply.

That is where current AI is useful. Language models can read unstructured documents, pull out fields, summarize long files, draft correspondence and compare a claim narrative against policy language. Used well, they shorten the time between receipt and a person being able to make a good decision. Used carelessly, they create new risks: a fluent summary that misstates a policy condition, a triage score that quietly disadvantages a group of customers, or an adverse outcome that nobody can explain to the customer or the regulator.

This playbook takes a deliberately conservative line on decision rights. AI may prepare, recommend and draft. A named, trained, accountable person decides anything that denies, reduces, delays on substantive grounds, rescinds, cancels or non-renews, or that otherwise adversely affects a customer. The reason is practical as well as ethical. Regulators in most jurisdictions expect insurers to explain and justify decisions, to treat customers fairly and to keep records. A system that cannot show its reasoning to a reviewer is hard to defend.

- Start with reading and preparation work: intake, extraction, summarization, drafting and quality review
- Keep adverse decisions with accountable people who can see the evidence and give reasons in plain language
- Treat explainability, fair treatment and record keeping as design requirements, not later paperwork

- Measure cycle time together with accuracy, complaints and reversals, so speed never hides harm
- Check current requirements in each jurisdiction and line of business before deployment; this playbook is not legal advice



PLAIN TEST

Could the person who made the decision explain it to the customer, a complaints reviewer and a regulator using only the file? If the AI output is part of the reasoning, it must be visible in that file, with its sources.



WORKING QUESTION

Which two or three steps in your claims and underwriting chains are mostly reading and rekeying, and who owns the decision that follows each one?



WHERE IT FITS

Where AI fits in insurance, and where it does not

02

Fit depends on two questions: how much of the work is reading and drafting, and how serious the consequence is if the output is wrong.



Sort candidate uses along those two axes. High reading volume with low consequence, such as summarizing a broker email thread, is an easy start. High reading volume with high consequence, such as reviewing a complex liability claim, can still benefit from AI assistance, but the person must verify against source documents and own the outcome. Low reading volume with high consequence, such as a final coverage denial, is not an automation candidate at all. It is a human decision that AI may inform.

USE CASE	FIT	REASON AND CONDITION
Submission intake and data extraction	Strong	Repetitive reading; extracted fields shown beside source for confirmation
Underwriting file summary and gap list	Strong with review	Saves reading time; underwriter decides appetite, terms and price
Claims first notice triage and routing	Strong with controls	Routing by complexity helps speed; test for uneven treatment across customer groups
Claim file summarization for adjusters	Strong with review	Long files benefit; adjuster verifies key facts and policy conditions
Coverage decision or claim denial	Weak for automation	Adverse outcome needs accountable human decision with reasons
Customer service answers on policy documents	Strong with grounding	Answers from approved wording with citations; hand-off for disputes
Fraud indicators	Moderate, cautious	Indicators may prompt review; they should not decide or accuse
Compliance and quality review of calls and letters	Strong	Sampling at scale; findings reviewed by compliance staff

Where to be most careful

Three areas deserve extra care. The first is anything that feeds pricing, eligibility or claims outcomes through a model trained on historical data, because historical patterns can reflect past practices that would not be acceptable today, and some variables act as proxies for protected characteristics. The second is fraud detection, where a false flag can delay or damage a legitimate customer. The third is customer-facing generative text about coverage, where a plausible but wrong sentence can be read as a promise or a misrepresentation.

COMMON MISTAKE

Starting with the most exciting use, such as automated claim decisions, instead of the most boring, such as intake extraction. The boring use builds data discipline, test habits and trust that the exciting use will need later.

WORKING QUESTION

For each candidate use, can you state the worst realistic consequence of a wrong output and the human step that would catch it?

HOW IT WORKS

How these systems work, in plain language

03

Most insurance AI combines four parts: reading documents, finding the right policy or guideline text, producing a draft or recommendation, and showing its evidence.

Document reading converts PDFs, scans, emails and photos into text and fields. Accuracy varies with scan quality, handwriting and layout, so every extraction needs a confidence signal and an easy path for a person to check the original. A language model then structures that text, for example identifying the insured, location, coverage requested, prior losses and dates of loss, and flags what is missing.

Retrieval connects the work to approved reference material: policy wordings, endorsements, underwriting guidelines, claims handling manuals and regulatory notices. Instead of asking the model what it remembers about a policy form, the system finds the relevant passage in your current document set and asks the model to answer from it, with a citation. This matters because wording differs by product, state and version, and general knowledge about insurance is often wrong for a specific contract.

01 Receive and read

Ingest the submission, claim notice or message and convert attachments to usable text and fields.

02 Ground in approved sources

Retrieve the correct policy version, guideline or procedure for this product and jurisdiction.

03 Prepare

Produce an extraction, summary, gap list or draft reply, with each statement linked to its source passage.

04 Review

A trained person confirms or corrects the output, and the correction is stored.

05 Decide and record

The accountable person decides, writes the reason, and the file keeps inputs, outputs, edits and the final decision.

Assist, recommend, decide

MODE	WHAT AI DOES	WHO DECIDES
Assist	Extracts, summarizes, drafts, searches	The person, using their own judgment
Recommend	Proposes a route, a priority or a next step with reasons	The person, who may accept or override and records why
Decide	Executes an outcome without a person	Reserved for low-risk, non-adverse, reversible actions with a documented approval

Most insurance work should sit in assist and recommend. Where a decide mode is used, such as automatically confirming receipt of a claim or sending a standard acknowledgment, it should be limited to actions that are favorable or neutral to the customer, reversible, and covered by a written approval from the control owner.

WORKING QUESTION

For your first workflow, which mode are you building, and could you show a reviewer exactly which source passage supports each line of the output?

OPERATING MODEL

Operating model, roles and decision rights

04

The control that matters most is a clear owner for every outcome, with AI placed visibly inside the process rather than beside it.

Insurers already have strong accountability concepts: authority limits, referral rules, peer review, complaints handling and model risk governance. The operating model for AI should reuse them. An AI-prepared claim summary is a work product like any other. The adjuster who relies on it carries the same responsibility as if they had written it. This keeps accountability where it already lives and avoids a parallel structure that nobody understands.

ROLE	RESPONSIBILITY	TYPICAL OWNER
Business process owner	Defines the workflow, decision rights and success measures	Head of claims, underwriting or service
Accountable decision maker	Makes and records every adverse or material decision	Adjuster, underwriter, claims manager
AI product owner	Manages scope, backlog, user feedback and releases	Operations or digital lead
Compliance and risk partner	Sets fair treatment, explainability and record rules; reviews changes	Compliance, legal, model risk
Data and security owner	Controls data access, retention, vendor terms and logging	CISO, data protection lead
Quality reviewers	Sample outputs and decisions on schedule, feed findings back	Claims quality, underwriting audit
Frontline champions	Test in real work, report failures, shape training	Experienced adjusters and underwriters

Decision rights to write down

- Which outcomes AI may never decide, such as denial, rescission, cancellation, non-renewal, material reduction of a claim and referral of a customer for fraud action
- Which actions require a second person, based on size, complexity or vulnerability indicators

- How a person records disagreement with an AI recommendation, and what happens to that record
- Who may change prompts, rules, thresholds and sources, and what testing they must pass first
- Who can pause the system and under what triggers



VULNERABLE CUSTOMERS

Define how the process recognizes indications of vulnerability, such as bereavement, serious illness or financial hardship, and make sure those cases route to trained people with discretion. Do not rely on AI to infer sensitive conditions from text.

Training matters as much as design. Reviewers need to know the typical failure modes: omitted exclusions, dates mixed up between documents, a summary that sounds certain where the file is ambiguous, and automation bias, the habit of accepting a confident suggestion. A useful practice is to seed known errors into training exercises and into periodic live samples, then track whether reviewers catch them.



WORKING QUESTION

If a customer complains about a decision tomorrow, can you name the person who made it, show what the AI contributed and show what they changed?



RISKS AND CONTROLS

Risks, controls and human review

05

Design controls around the concerns regulators and customers care about: explainability, fair treatment, record keeping, privacy and third-party oversight.

Insurance regulators and supervisors in many jurisdictions have published expectations or guidance on the use of AI and data analytics, and professional bodies have issued principles. Requirements differ by country, state and line of business, and they change. Some set expectations on governance, testing for unfair discrimination, consumer notices, and the documentation of models used in underwriting and pricing. Check current requirements with counsel and your supervisor before deployment. The controls below are practical design habits that tend to support most of them.

RISK	WHAT IT LOOKS LIKE	CONTROL
Unexplainable outcomes	A decision depends on a score nobody can interpret	Require stated reasons from the decision maker; avoid opaque inputs to adverse decisions
Unfair treatment	Routing, pricing inputs or fraud flags affect groups unevenly	Test outcomes by relevant segments; remove proxy variables; review regularly
Wrong policy statement	Draft misstates a condition or exclusion	Ground in approved wording, cite sources, require review of coverage statements
Automation bias	Reviewers accept suggestions without checking	Seeded errors, sampling, time for review, override tracking
Privacy and health data	Sensitive data exposed to wrong users or vendors	Minimize data, mask where possible, role-based access, contractual limits on vendor use
Weak records	Cannot reconstruct what the system showed on the decision date	Log inputs, source versions, outputs, edits and final decision

RISK	WHAT IT LOOKS LIKE	CONTROL
Vendor opacity	Supplier changes a model without notice	Change notification clauses, version pinning, re-test on change



Minimum record for every AI-assisted decision file

- Input documents and their versions
- Policy wording or guideline version used
- AI output as shown to the user, with timestamp
- Edits and overrides by the user, with the reason
- Final decision, decision maker and date
- Customer communication sent and its source of approval

Fair treatment in practice

Fair treatment is not a single test. It starts with the choice of use: avoid applications where historical data encodes practices you would not defend. It continues in design, by checking which inputs a triage rule or model uses and whether any act as stand-ins for characteristics that law or policy protects. It continues in monitoring, by comparing handling times, referral rates, outcomes and complaints across customer segments where you are lawfully able to do so, and by investigating differences rather than explaining them away.



ADVERSE DECISIONS

Never automate an adverse decision. An AI recommendation to deny, reduce or cancel can be shown to the decision maker, but the person must review the evidence, apply the policy, consider the customer's circumstances and give their own reasons.



WORKING QUESTION

Which of your seven control areas is weakest today, and what evidence would convince an internal auditor that it is working?



MEASURING VALUE

Measuring quality and value

06

Measure speed, accuracy and customer outcome together, against a baseline taken before the first pilot.

A common failure is to report time saved and stop. In insurance, a faster process that increases rework, complaints, leakage or reversals is a worse process. Build a measurement set with three families: efficiency, quality and customer or compliance outcomes. Capture a baseline for at least four to eight weeks before switching on assistance, using the same definitions after launch. Where possible, run assisted and unassisted groups side by side for a period, with comparable work allocated by a simple rule rather than by user preference.

METRIC	DEFINITION	WHY IT MATTERS
Touch time per file	Active handling minutes from sampled time studies	Shows real effort saved, not elapsed time
Cycle time	Receipt to decision or response, by file type	Customer experience and capacity
Extraction accuracy	Share of fields correct against a verified sample	Catches silent errors at the input stage
Summary fidelity	Reviewer score on whether the summary preserves material facts and conditions	Prevents confident misstatements
Override rate with reasons	Share of AI suggestions changed or grouped by reason	Reveals weak spots and over-trust
Rework and reopen rate	Files returned for correction or reopened	Detects hidden cost
Complaint and reversal rate	Complaints and overturned decisions per thousand files	Guards customer outcomes
Segment comparison	Outcome and handling difference across lawful segments	Supports fair treatment monitoring

Illustrative baseline

The figures here are illustrative only, not benchmarks. Suppose a property claims team handles 400 first notices a week, and sampled time studies show 22 minutes of reading and data entry before an adjuster starts assessing. If assisted intake reduced that to 14 minutes while the error rate on key fields stayed flat and reopen rates did not rise, the team would have recovered about 53 hours a week. If the error rate rose, the saving would be partly or wholly offset by rework. Your own baseline, not these numbers, should decide.

- Use sampled, observed time rather than self-reported time
- Evaluate quality blind: reviewers should not know which file used assistance
- Set a pass mark before launch, such as no loss of accuracy on material fields and no rise in complaints
- Convert freed time into a stated use, such as faster response on complex claims, or the benefit stays invisible
- Include cost of review, licenses, integration and monitoring in the value case

HONEST REPORTING

Report ranges and the sample size. A result from forty files is a signal to continue testing, not a figure to publish in a board deck.

WORKING QUESTION

What baseline have you captured for your first workflow, and which metric would make you stop the pilot even if time savings looked good?

IMPLEMENTATION

Implementation guide: from one workflow to controlled rollout

07

Pick one workflow, build a test set from real files, run assisted work in parallel, then widen only as evidence allows.

01 Choose a narrow workflow

Select one line of business and one task, such as extracting fields from commercial property submissions. Name the process owner and the accountable decision maker.

02 Prepare data and permissions

Gather current policy wordings, guidelines and a set of past files with known correct outcomes. Confirm lawful use of data, masking needs, retention rules and vendor terms before anything is uploaded.

03 Build a test set

Assemble 100 to 200 representative files including hard and unusual cases. Have experienced staff produce the correct answers. Score extraction, retrieval and drafting separately.

04 Pilot in shadow mode

Run the AI alongside normal work without it affecting outcomes. Compare its output with what staff did and study the differences.

05 Go live with assistance

Release to a small trained group, with mandatory review, logging and a one-click way to report errors. Review samples weekly.

06 Decide on scale

Compare results to the baseline and the pass mark. Scale, change or stop, and record the decision and reasons.

Data, systems and integration

Most of the effort sits in connections: policy administration, claims systems, document stores and email. Start with a read-only integration or a side-by-side screen before writing back to core systems. Where write-back is needed, such as populating fields, require a person to

confirm before saving and keep the audit trail. Avoid copying sensitive data into tools outside your approved environment, including personal accounts on public AI services, and tell staff plainly why.

Go-live readiness

- Named business owner, decision maker and compliance partner
- Written decision-rights map approved
- Test set results meet the agreed pass mark
- Logging captures inputs, sources, outputs, edits and decisions
- Staff trained on failure modes, with seeded-error exercise completed
- Privacy and security review completed, vendor terms checked
- Pause procedure and escalation path tested
- Customer-facing wording approved by compliance

Change management is part of the build. Adjusters and underwriters will ask whether the tool is there to replace them or to monitor them. Answer honestly and early: state what the tool will and will not be used for, including whether its data will feed individual performance management. Trust collapses fast if staff discover an unannounced use, and with it the quality of the feedback you need.

WORKING QUESTION

Which 100 files would you use to test your first workflow, and who is expert enough to say what the right answers are?

PITFALLS

Pitfalls and a hypothetical scenario

08

Most problems come from over-reach, thin testing and unclear ownership rather than from the technology.

- Letting a score or recommendation drift into a decision because reviewers lack time or confidence to disagree
- Testing on clean sample files rather than on messy real ones with poor scans, mixed languages and missing pages
- Ignoring policy version control, so the assistant cites last year's wording
- Using customer-facing generative text without approved templates and review
- Treating the vendor demo as evidence, with no testing on your own files
- Reporting time saved without rework, complaints or reversal data
- Skipping segment monitoring because the data is hard to assemble

HYPOTHETICAL SCENARIO (ILLUSTRATIVE, NOT A REAL CASE)

A regional insurer adds AI summaries to auto glass and minor property claims. Adjusters approve most summaries quickly. Three weeks in, a reviewer sampling files notices summaries omit a policy excess variation on one product. Because logs record the wording version used, the team finds a stale document in the index, corrects it, re-runs affected files and adds a freshness check.

The scenario shows why records and sampling matter. The error was small, found by a routine control and traced through logs. Without those controls, the same fault could have run for months, and the insurer would have been unable to say which customers were affected. The remedy was not to abandon the tool but to treat source freshness as a first-class requirement with an owner.

When to stop

Define stop conditions before launch. Examples include a drop in extraction accuracy below the pass mark, a complaint pattern linked to AI-prepared communication, evidence of uneven

outcomes across segments, a vendor model change that has not been re-tested, or a data incident. Stopping is part of a controlled process, and a team that can pause calmly is more likely to be allowed to continue.

WARNING SIGN	LIKELY CAUSE	FIRST RESPONSE
Override rate near zero	Over-trust or time pressure	Seed errors, review workload, check sample files
Override rate very high	Weak retrieval or poor fit	Inspect misses, narrow scope, fix sources
Rising reopen rate	Silent errors in summaries	Sample reopened files, tighten review points
Complaints cite unclear reasons	Reasoning not captured	Require written reasons, adjust templates

WORKING QUESTION

What are the three stop conditions for your first deployment, and who has authority to invoke them without asking permission?

BUYING QUESTIONS

Questions to ask before you buy or build

09

Suppliers can describe capability easily. Your questions should test evidence, controls and the terms that protect your customers and your regulatory position.

Insurance buyers should ask suppliers to demonstrate on their own files rather than on vendor samples, and should expect straight answers on limits. A supplier that cannot say where the system performs poorly has not tested it enough. Involve compliance, security, data protection and the business owner in the evaluation from the start, because late objections from any of them are expensive.

Questions for any supplier

- Can you run our own anonymized files through the product in a supervised trial and share the error analysis?
- How does each output show its source, and can reviewers see the original passage in one click?
- What is logged, for how long, and can we export records for audit, complaints and regulators?
- Where is our data processed and stored, who at the supplier can see it, and is it used to train models for others?
- How are model or version changes notified, and can we pin a version and re-test before adopting an update?
- What testing exists for uneven outcomes across customer groups, and what support do you provide for our own monitoring?
- Which decisions does the product block from automation, and can we configure additional blocks?
- What happens to our data and configurations if we leave?
- What does the total cost include: integration, review time, training, monitoring and support?

Build, buy or combine

OPTION	FITS WHEN	WATCH FOR
Buy a specialist product	The workflow is common, such as submission extraction	Fit with your lines of business, lock-in, change control
Configure a general platform	You have varied internal documents and want flexibility	Your team must own testing, prompts and sources
Build custom	Process is distinctive and you have engineering depth	Ongoing monitoring and maintenance burden
Combine	Core vendor for extraction, own layer for guidelines and review	Clear ownership across the boundary

Whichever route you choose, keep the control points inside your organization: decision rights, source documents, test sets, logs and the ability to switch off. Those are the assets that stay valuable if you change supplier or model, and they are what regulators and customers will ask you about. Halden & Company helps insurers and intermediaries scope the first workflow, build the test set and design the control framework. Free assessments are available at halden.company/assessments, and you can reach the team at halden.company/contact.

WORKING QUESTION

If your supplier changed the underlying model next month, what would you re-test, who would sign off, and how quickly could you roll back?



30 / 60 / 90

A first quarter that produces evidence

10



Move in three stages: prove one workflow on real files, add controls and monitoring while assisted work runs, then decide on scale from evidence.

First 30 days: ground it

- Pick one workflow and name the process owner, decision maker and compliance partner
- Write the decision-rights map, including outcomes AI may never decide
- Capture a four-to-eight week baseline for time, accuracy, rework and complaints
- Assemble a 100 to 200 file test set with answers from experienced staff
- Complete privacy, security and vendor terms review



Days 31 to 60: test it

- Run shadow mode and score extraction, retrieval and drafting separately
- Fix source freshness and version control problems found in testing
- Train reviewers with seeded-error exercises
- Set pass marks and stop conditions with compliance
- Release assisted mode to a small group with full logging



Days 61 to 90: decide

- Review results against baseline, including segment comparisons and complaints
- Hold a decision meeting to scale, change or stop, and record reasons
- Publish a short internal note on what the tool does and does not do
- Plan the second workflow using the same controls
- Schedule quarterly review of sources, thresholds and vendor changes

GLOSSARY

Terms used in this publication

11

TERM	MEANING
Accountable decision maker	The named person who makes and records a decision, and who can explain it, whatever tools helped prepare it.
Adverse decision	An outcome that is unfavorable to a customer, such as denial, reduction, cancellation, rescission or non-renewal.
Explainability	The ability to state, in understandable terms, why a decision was reached and which evidence supported it.
Fair treatment	Handling customers consistently and without unjustified discrimination, tested in design and monitored in operation.
Proxy variable	An input that is not itself a protected characteristic but correlates strongly with one and may produce uneven outcomes.
Retrieval	Finding the relevant policy or guideline passage so the model answers from approved text rather than from memory.
Shadow mode	Running the AI alongside normal work without affecting outcomes, so its output can be compared with human results.
Automation bias	The tendency to accept a confident machine suggestion without checking it.
Override rate	The share of AI suggestions that a person changes, ideally recorded with reasons.
First notice of loss	The initial report of a claim by the policyholder or a representative.

TERM	MEANING
Leakage	Claims cost lost through avoidable error, such as overpayment or missed recovery.
Model risk governance	The policies and reviews used to approve, monitor and retire models used in important decisions.

A note on use

This playbook is general guidance, not legal or regulatory advice. Insurance requirements vary by jurisdiction and line of business and change over time. Check current requirements with qualified counsel and your supervisor.

TAKE THE NEXT STEP

Take the free Halden assessment at halden.company/assessments, or talk to us at halden.company/contact about turning this into a plan for your organization.