

GenAI <> Control <> Humans:

Where motor claims runs today, and the real potential of GenAI and automation

Generative AI can carry more than twice what it does today in motor claims, and it cannot be left alone to full automation – controls are required. This paper works through all twelve processes in a claim, showing where the industry sits today, how far generative AI can move each one, and what the three carriers each do in the end state.

by Axxion Claims Settlement Services L.L.C. | September 2026

Written from administering motor claims for insurers. Published research is cited where it reaches these processes. Method at section 8.

1. The three carriers	3
2. Reading the triangle	4
3. The twelve processes	5
4. What the twelve add up to	13
5. What happens to the humans	14
6. The end state keeps all three	15
7. What to build and UAE compliance	16
8. Method and limits	17
Sources	18
About Axxion	19
Appendix: Where the value sits, and how an insurer would measure it	20



A motor claim is settled through twelve separate decisions, and each one is carried by some mix of three things: a **generative AI model**, a **deterministic control**, or a **person**. Axxion took each of the twelve in turn and asked how much of the work each of the three carries in the industry today, and how much each could carry with the technology that already exists, properly governed. The result is a position and a direction for every process in a claim, and this paper sets out both, along with what has to sit around a model before any of it is safe to run.

Three findings came out of the exercise. Generative AI runs at under half of its potential, and the constraint is the control layer and not the models. The control layer itself does not shrink as generative AI grows: it holds at the same share of the work in both states, which is the clearest single test of whether an operation is deploying AI correctly.

The third finding carries the consequences. Human oversight falls sharply while concentrating into the four processes that carry the most money, the most dispute and the least measurement. That makes the person the largest uncontrolled variable in a modern claims operation, and it is the part almost nobody is instrumenting.

What this paper finds

- **In motor claims, Generative AI is good at one specific thing, and a claim file is full of it.** Turning unstructured input into structured output is the core competence. A claim arrives as photographs, police reports, handwritten estimates, invoices and voice notes, which is the least structured input in financial services.
- **Generative AI does roughly a sixth of the work today and could do about two fifths.** That is roughly two and a half times what the industry currently asks of it, using technology that already exists.
- **The control layer does not shrink.** Generative AI's share more than doubles while the deterministic layer holds steady, and control stays the largest of the three in both states. Generative AI displaces human effort. It does not displace control, and an operation whose control layer thins as its models grow has removed the thing that made the models safe to run.
- **Human oversight drops by about 60%.** Liability, subrogation, settlement approval and repair method account for 40% of all human judgment in a claim today, and half of it once generative AI is deployed. People move out of the routine work and are left holding the contested work.
- **The person is the least inspectable part of the system.** A model can be tested against held-out data, versioned and monitored for drift. A control can be read by anyone and audited against the files it governed. Neither applies to a person, and adjuster-to-adjuster variance on comparable motor files has never been published by any regulator, actuarial body or journal that an August 2026 search reached.
- **The end state is not an extreme.** Generative AI fails by opacity and control answers it. Control fails by rigidity and generative AI answers it. People fail by inconsistency and the other two answer it. Each carrier covers another's blind spot, which is why all three survive.



1. The three carriers

Every step in a motor claim is carried by a mix of three things. Defining them is important, because most writing on this subject uses "AI" to mean all three at once and then argues about the result.

GENERATIVE AI

Probabilistic models: large language models, vision models, and the prediction models alongside them. What they are good at is turning unstructured input into structured output. A photograph of a police report becomes fields. A set of damage images becomes a parts and labor list. Forty pages of correspondence become a summary with the disputed points marked.

What they return is a confidence, never a certainty. A model that is 94% sure a plate reads TRK-4471 is reporting a distribution. That is the entire reason the controls exist.

Generative AI fails by opacity. A confident wrong answer is indistinguishable from a confident right one at the moment it is produced, which separates it from every other tool in a claims operation.

CONTROL

Deterministic systems. Validation rules, written thresholds, benchmark tables, authority matrices, policy wording, and the governance layer around all of it: explainability, reason codes, audit trails, model inventories, version history.

A control returns the same answer to the same input every time, and anyone can read why. It is what makes a machine-assisted outcome defensible to a policyholder, a workshop, a court or a regulator. In practice it is also the layer that decides whether a model's output reaches the claims management system at all.

Control fails by rigidity. A total loss threshold set against one salvage market keeps answering after that market has moved, with exactly the confidence it had when it was right.

HUMANS

Co-human experts, not a fallback and not a rubber stamp. The human takes the contested call, the unusual claim, and the exception the other two surfaced and could not settle between them.

Humans fail by inconsistency. Two competent adjusters reach different answers on the same file, and unlike the other two failure modes, this one has no instrument pointed at it in any claims operation the published record describes.

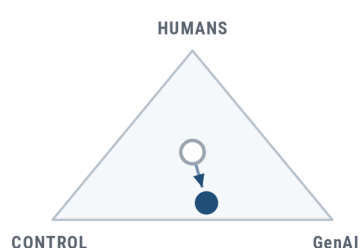
The design rule that follows. Generative AI proposes. Control decides whether the proposal is admissible and records why. Humans settle what the first two cannot. A process that skips the middle carrier has a model writing straight into a system of record. A process that skips the third has no path for the claim the model has never seen.



Capgemini reached a similar picture by survey in 2026, surveying 344 executives and 809 employees including 200 claim adjusters: coverage determination, liability assessment and settlement "still rely heavily on individual adjuster judgment, applied inconsistently across teams, regions, and complexity levels," while AI has been deployed "only at the edges of the claims process." Two different methods, pointing the same way.

3. The twelve processes

Each entry gives where the industry sits, how far generative AI moves it, and what each of the three carriers does in the end state. Where Axxion's own operation illustrates a point, it is marked.



1. FNOL document extraction

A police report, a driving license, an Emirates ID and a workshop invoice arrive as photographs taken at the roadside, at night, at an angle.

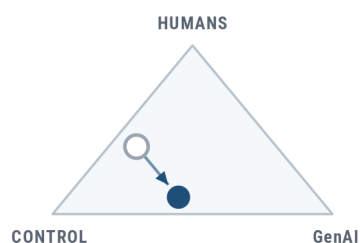
Today. Most operations put a person on every document. That is a trained adjuster performing optical character recognition by hand, and it is the single largest waste of skilled time in the claim.

GenAI in the end state. Reads the images, identifies each document type, and returns structured fields: plate, chassis, date and time of loss, location, parties, police report number, and the officer's account of what happened. This is the competence at its purest.

Control in the end state. Nothing the model produces writes to the claims management system directly. Each field passes a deterministic gate: format validation on chassis and plate, policy match on vehicle and date of loss, a cover check, and a confidence threshold that routes rather than guesses. An audit entry records what was read, from which document, at what confidence.

Humans in the end state. Where model and controls together surface something the system cannot settle: a document that fails validation, a mismatch between police report and policy, or an early leakage or fraud indicator. In a properly built flow that is no more than 10% of notifications received.

At Axxion. The FNOL register runs this shape today: intimated, collecting, submitted, approved, with a 30-day duplicate guard on VIN and plate, and the record frozen read-only once approved. Every write carries an audit entry. The gates exist before the extraction that will sit on top of them, which is the order that makes the extraction safe to trust.



2. Notification valid and complete

Whether what has arrived carries enough to open a claim under this policy.

Today. A checklist confirms fields are populated and a person judges the awkward ones, which means completeness is defined as presence rather than sufficiency.

GenAI in the end state. Reads for sufficiency. A police report saying "minor contact, no injuries" and one saying "vehicle overturned, occupants transported" populate the same field and mean different things for what the file needs next. The model reads the difference and requests the specific missing item.

Control in the end state. Everything with a written answer stays written, and control gives up nothing here. Whether the policy was in force on the date of loss is not a probabilistic question. Cover scope, excess, endorsements, notification deadlines and the required document set stay deterministic, and the model works inside them rather than over them.

Humans in the end state. Late notification, a cover question with a real argument on both sides, and claims where the policy language does not clearly reach the circumstances.



3. Claim assignment

Which handler, which track, which authority level, which service level.

Today. A routing table, which is why this process sits closest to the control corner of anything in the set. Routing tables sort on what is known at intake and cannot sort on what the claim is going to become.

GenAI in the end state. Predicts complexity and ultimate cost from the notification content before anyone has read the file. A rear-end impact on a three-year-old sedan with no injuries and a cooperative counterparty is a different claim from a multi-vehicle incident with a disputed police report, and that difference is legible in the FNOL text long before it is legible in a routing table.

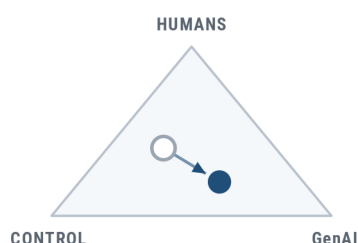
Control in the end state. Holds the authority levels. Who may approve what is a governance question with a written answer, and a prediction never sets it. Control also holds the segregation rules: which claims must reach a named handler, which must never enter an automated track, which require a licensed assessor.



Humans in the end state. Rarely at assignment. The meaningful oversight is retrospective, a supervisor reading the rate at which claims change track after assignment, broken down by original track, which is the only way mis-routing becomes visible at all.

At Axxion. The master SOP runs as data rather than as a PDF: six phases and 28 steps, an urgency-ordered queue, and stage moves that pass a gate asking the SOP's exit questions, with every exception logged against a reason. That is the control corner of this process built explicitly, and it is what a prediction model has to satisfy before it moves anything.

Published anchor. Among US motor insurers that have built such models, the NAIC found 54% of claim assignment models ran with no human intervention on execution, the most automated of the claims model types it measured, across 193 responses.



4. Fraud referral

Whether a file goes to investigation.

Today. Indicator lists and a handler's instinct, which catches the obvious and misses the organized.

GenAI in the end state. Detects the signal no handler can see, because the signal only exists across claims: one phone number behind four unrelated notifications, a workshop whose supplement rate sits three standard deviations above its peers, a damage pattern inconsistent with the described impact, a loss reported eleven days after inception. One file at a time none of that is visible. Across a portfolio it is obvious.

Control in the end state. Records a reason on referral and, harder and rarer, on non-referral. Tracks referral precision instead of referral volume, which is what stops a model drifting into flagging everything. Holds the escalation path and the evidentiary standard the investigation must meet.

Humans in the end state. On every referral. This is an accusatory act against a named customer, and it keeps a human owner even where the model is right. That is a governance choice rather than a capability limit, and it is why human oversight is never designed out of this process.

Published anchor. Fraud detection carries the highest planned generative AI adoption in EIOPA's 2026 survey of 347 undertakings, at 64% within three years, ahead of claims management at 59%.



5. Damage assessment

What is damaged, how badly, and what that implies for parts and labor.

Today. The model proposes and a person concludes, almost everywhere.

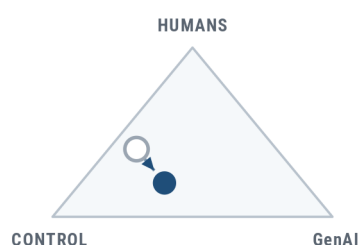
GenAI in the end state. Identifies damaged panels and components from photographs, classifies severity, separates repair from replace, and infers the hidden damage that follows from a visible impact. This is the highest generative AI position in the whole set, because vehicle damage is the most mature computer vision application in insurance and the images are standardized by the process that captures them.

Control in the end state. Less than anywhere else in the claim, and saying so matters. There is a confidence gate, a rule that certain damage types always require physical inspection, and the audit trail. No table tells anyone whether a quarter panel is repairable.

Humans in the end state. Structural damage, suspected water or fire involvement, anything below the confidence gate, and a blind re-inspection sample where a second assessor sees the images without the model output. That sample is the only control that detects automation bias, and it is the one most operations skip.

At Axxion. AI damage assessment is pulled by VIN and shown beside the negotiated estimate, so the machine's view and the actual outcome sit on the same screen and the gap between them is visible rather than assumed. Assessment runs a triage ladder from desktop to guided video to physical inspection, with each attempt kept as history rather than overwritten, and video requests raised and fulfilled through the workshop portal.

Published anchor. In the NAIC survey, 80% of image evaluation models were classified as augmentation against 6% as automation. That balance is set by insurer caution rather than by what the models can do.



6. Repair method, OEM against aftermarket

Which parts go on the vehicle.

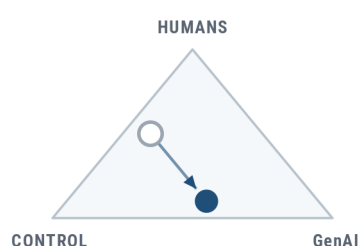
Today. Policy wording and a parts schedule, negotiated against what the workshop can actually source.

GenAI in the end state. Builds the availability and price picture across suppliers, matches part numbers across catalogs that disagree with each other, and flags when the schedule has gone stale against the vehicles actually arriving. It does not choose.



Control in the end state. Everything. Policy wording, parts schedule, warranty position, and the age and value bands that determine what is permissible are all written, and they are what the insurer is contractually bound by. This is a control-corner process and it should stay one.

Humans in the end state. The negotiation with the workshop, and the exceptions where schedule and vehicle disagree. That includes a live and growing category in the UAE: electric and Chinese-brand vehicles entering the fleet faster than parts schedules are revised, where the schedule has no entry and a person has to decide.



7. Repair estimate challenge

Whether the workshop's estimate is right, and which lines to contest.

Today. Benchmark tables and a handler's experience, which means the challenge is only ever as good as the individual running it. This is one of the largest sources of unmeasured variance in a claims operation.

GenAI in the end state. Reads the estimate line by line, including the ones that arrive as a photograph of a printout, normalizes line descriptions against a standard taxonomy, and compares each against the benchmark distribution for that vehicle and that repair. It returns the contested lines with a reason attached: labor hours above the ninetieth percentile, a part billed as new that the damage assessment never called for, a paint time inconsistent with the panel count.

Control in the end state. Holds the benchmark itself, which is a deterministic artifact built from the operation's own history. Holds the tolerance bands that decide what is worth contesting. Records the sustained rate on challenges by reason and by workshop, which is the only way anyone knows whether a challenge is right or merely aggressive.

Humans in the end state. The conversation with the workshop, which is a negotiation between two people who will speak again next week. The model produces the position; a person holds it.

At Axxion. The benchmark is a live store rather than a spreadsheet, carrying 4,043 loaded facts and serving own-versus-pool comparison bands directly on the repair page, with the sample size and confidence level shown every time so nobody argues from a band built on four repairs. Estimate review, the negotiation log and the repair award are all recorded, so every challenge and its outcome is retrievable per workshop.



8. Total loss determination

Repair or write off.

Today. A threshold rule comparing repair cost against vehicle value, with a person handling the files that land near the line.

GenAI in the end state. Improves both inputs and takes none of the decision. It produces a repair cost earlier and more accurately from the damage assessment, and a pre-accident value from the comparables.

Control in the end state. Makes the call, and keeps everything it has. A threshold comparing cost against value is written, auditable and applied identically to every file, which is exactly what a decision of this kind should be. Control also owns the review cadence on the threshold, tested against realized salvage recovery rather than against policy convention.

Humans in the end state. Files landing in a narrow band either side of the threshold, and vehicles where salvage value is genuinely uncertain.

The pattern worth noticing. Better machine inputs make this process more rule-able, not less. Generative AI's largest effect here is to move the decision toward the control corner and away from itself, which is the opposite of what the market expects it to do.



9. Total loss valuation

What the vehicle was worth the moment before the loss.

Today. A comparables model proposes, adjustment rules modify, and a person can override. This is the most disputed number in motor claims.

GenAI in the end state. Widens the comparable set well beyond what a person can assemble, reads condition and specification out of listing text and photographs, and normalizes across listings that describe the same trim four different ways. In a market like the UAE, where specification varies by import route, that normalization is most of the work.

Control in the end state. Gives up nothing, because this is the figure the owner will contest. The adjustment logic is written. The comparable set is disclosed. Every condition deduction is documented against evidence. The audit trail is the entire defense, and a valuation a claimant cannot interrogate is a complaint already in motion.

Humans in the end state. Disputes, modified or unusual vehicles, and any file where the comparable set is thin enough that the distribution is doing more work than the data.



Published anchor. In the NAIC survey, 22% of settlement amount models ran automated and 80% came from third parties, which is why the vendor contract governs this process as much as the model does.



10. Liability apportionment

Who was at fault, and in what proportion.

Today. A police report, a legal framework, and a judgment between two accounts of the same event. The most human process in the claim.

GenAI in the end state. Reconstructs the incident from the report, the images and the parties' accounts. Compresses forty pages of correspondence to the two points still in dispute. Retrieves how comparable circumstances were apportioned before. Everything the person reads becomes machine-produced, and the decision itself barely moves.

Control in the end state. Holds whatever apportionment framework applies, and the escalation path where the parties disagree.

Humans in the end state. Always, and this process has the least headroom of anything in the set.

The risk with no entry in the file. This is the clearest case of a process where the record says a human decided while the human began from a machine's reconstruction. The file preserves the conclusion and nothing about the anchor, which is why section 5 argues for recording the machine's proposal and the person's conclusion as two separate fields.



11. Subrogation viability

Whether recovery is worth pursuing, and against whom.

Today. Rules of thumb about value and counterparty, plus a handler's read of the file.

GenAI in the end state. Reads liability and the counterparty position out of the file, then predicts recoverability and expected net recovery against the cost of pursuing it. This is a clean supervised learning problem, because the label is unambiguous: either the money came back or it did not, and every operation is sitting on years of it.

Control in the end state. Holds limitation periods, pursue thresholds, and reason codes on files not pursued.

Humans in the end state. Recoveries above a written value, and counterparties where the commercial relationship is a factor no model can see.



The quietest leakage in the claim. An unpursued recovery generates no complaint, no dispute and no file. Nobody audits it, which is exactly why a sampled review of files not pursued, scored for what recovery would have returned, belongs in any operation running this process.

At Axxion. Recoveries and salvage are captured against total repair cost rather than tracked separately, so net cost per claim is visible instead of inferred, and same-VIN rework inside 90 days is detected automatically. What was recovered and what was left is a reportable figure rather than a reconstruction exercise at year end.



12. Settlement approval

Whether this amount may be paid, and by whom.

Today. An authority matrix and a person's signature, split roughly evenly.

GenAI in the end state. Almost nothing. Generative AI approves no payments, and the only thing it adds is assembling the approval pack so the approver sees the complete position in one place.

Control in the end state. Nearly everything, and more of it than today. When the figures underneath become reliable, the authority matrix can be widened and written down with confidence, so approvals that used to need a person's judgment become approvals that satisfy a documented rule. The matrix is versioned, and the rate at which approvals are sought outside it is monitored.

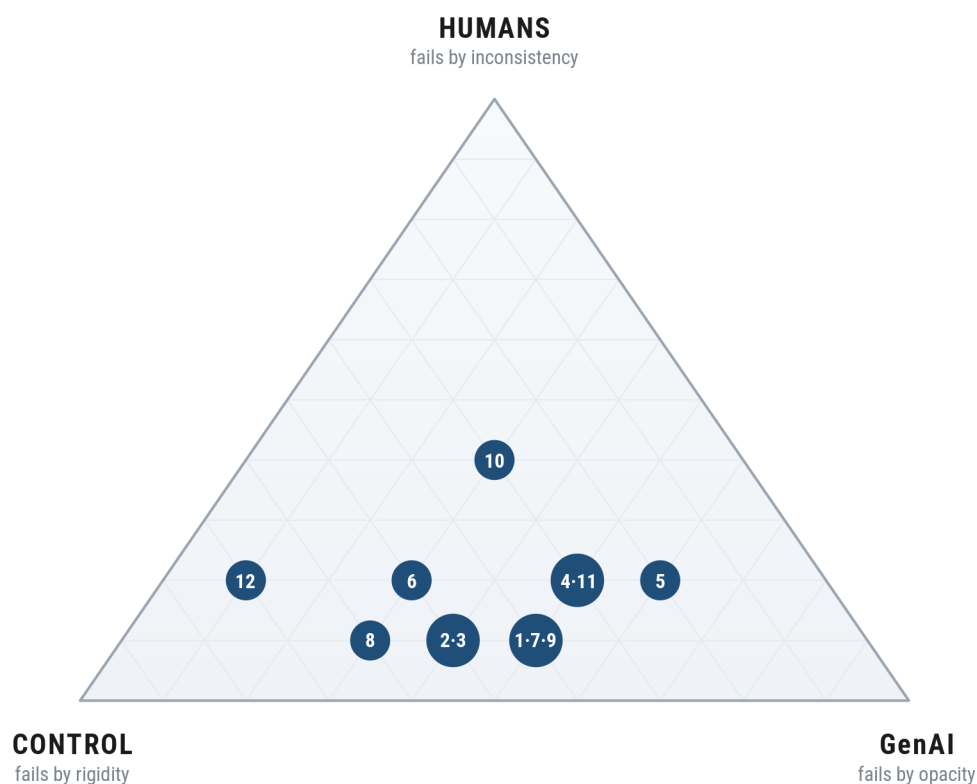
Humans in the end state. Above the authority threshold, on any file with a coverage question, and on anything a regulator or a court might later read.

At Axxion. Payment validation and money bounds are enforced in the platform rather than in a procedure document, and the insurer keeps assessment, approval and settlement authority throughout. The control corner of this process is not an aspiration.



4. What the twelve add up to

Exhibit 2. Where the industry can run with generative AI deployed and properly controlled.



Control does not move. It is the constant. Generative AI's share of the work more than doubles. Human oversight drops by about 60%. The deterministic layer sits at the same share in both states, unchanged. That was not designed into the exercise; it fell out of assessing twelve processes one at a time, and it is the most useful thing in this paper. Generative AI displaces human effort, and it does not displace control. An operation adding generative AI while holding its deterministic layer flat is building this correctly. An operation whose control layer thins as its models grow has removed the thing that made the models safe to run.

The second finding is commercial. The industry runs generative AI at well under half of what existing technology already supports. The constraint is the control layer, which has to be built first, and building one is slower and less interesting than buying a model.



5. What happens to the humans

Human oversight drops by about 60%, which reads as generative AI reducing the human role. The distribution says something more specific.

Four processes hold 40% of all human judgment in a claim today: liability apportionment, subrogation viability, settlement approval and repair method. In the end state the same four hold 50%. People move out of the routine work and are left holding the contested work, and the inputs to all four become machine-produced along the way.

Which makes the human the largest uncontrolled variable in the system. A model can be tested against data it never saw in training, versioned, monitored for drift and inventoried. A control can be read by anyone and audited against the files it governed. Neither applies to a person, and the published record is close to empty: adjuster-to-adjuster variance on comparable motor files has never been measured by any regulator, actuarial body or journal that an August 2026 search reached.

The one occasion anyone measured expert divergence on identical insurance files is a noise audit reported by Kahneman, Sibony and Sunstein, in which adjusters independently assessed written descriptions of the same cases. The median ratio between two adjusters estimating the same claim's ultimate cost was 43%. Most of that insurer's executives had guessed 10% or less, and when 828 CEOs and senior executives across industries were asked the same question, 10% was both the median and the most common answer.

What that figure is. The audit covered commercial underwriting and an industrial injury claim, not motor. The company is not named, and the adjusters were estimating ultimate cost rather than agreeing a settlement. It is quoted for one reason: it is the closest anyone has come to measuring how far two competent assessors diverge on identical insurance files, and the answer was four times what management expected. Nobody has run the equivalent in motor.

There is a second problem, and it has no entry in any claim file. On processes where generative AI supplies the inputs and a person concludes, liability apportionment above all, the record says a human decided. Nothing records that the human began from a machine's reconstruction.

The research on that condition is consistent. In a controlled reader study, very experienced radiologists scored 82.3% accuracy when the AI suggestion shown to them was correct and 45.5% when it was wrong, with the wrong suggestions inserted by the researchers. In a process-control study reported by Parasuraman and Manzey, roughly 43% of operators missed a false automated diagnosis the first time one appeared, and about half of those had already checked every parameter needed to disprove it.

The governance literature arrives at the same place. Ben Green surveyed 41 policies requiring human oversight of government algorithms and found that any nominal human involvement satisfies the requirement, which creates an incentive to introduce oversight that looks like a control and functions as a signature. Riikka Koulu describes human oversight as easily observable and in danger of becoming "an empty procedural shell used as a stand-in justification for algorithmization."



A signature is not oversight. Oversight is the override rate, the reason code, the share of overrides later reversed, and the gap between what the machine proposed and what the person concluded. All four are countable, and none of them is counted in most claims operations.

6. The end state keeps all three

The triangle has three corners and none of them is a destination. A claims operation works because each carrier's failure is answered by one of the other two.

Failure	Answered by	How, in practice
GenAI fails by opacity. A confident wrong answer looks exactly like a right one.	Control	A confidence threshold that routes instead of guessing. A hard authority ceiling above which no model output takes effect. A reason code on every machine output, from a closed list. An audit trail that survives the model being retrained.
Control fails by rigidity. A threshold set against one market keeps answering after the market moved.	GenAI	The model supplies the evidence that the rule has gone stale: realized salvage against the total loss threshold, live parts availability against the schedule, benchmark drift against the estimate tolerance.
Humans fail by inconsistency. Two competent adjusters reach different answers on the same file.	Both	The model narrows the range before the person sees it. The control bounds what the person may conclude without a second signature. The person is spent where judgment is required and nowhere else.

That last row is the operating case for generative AI in claims, and it inverts the usual one. Generative AI is worth deploying because it makes human judgment more consistent, not because it removes human judgment. A handler working from a normalized estimate, a disclosed comparable set and a bounded authority matrix produces a narrower spread of outcomes than the same handler working from a blank file and thirty years of instinct.

The condition is that the human control has to be designed rather than assumed. A review step that nobody measures produces a signature. What turns it into a control is measuring it.

How Axxion builds it. The platform is built on one rule: a machine output is a proposal, and the platform decides whether it becomes a state change. Nothing an agent produces alters a claim by itself. That is why the deterministic layer went in first, and why the audit trail, the stage gates, the money bounds, the benchmark store and the SOP-as-data exist ahead of the models that will use them. The order is the point. A control layer retrofitted around a model already in production is a description of what the model does, not a constraint on it.



7. What to build and UAE compliance

Five things, all answerable from an existing claims operation and none of them waiting on a regulator.

- **Measure divergence on the four that carry the money.** Liability, settlement approval, subrogation and repair method. For each, how often two handlers reach a different answer on comparable files. No published figure exists for motor, so the first insurer to hold one holds something nobody can contest.
- **Separate what the machine proposed from what the person concluded.** Two fields, every file, on every process where generative AI supplies the input and a person signs. Without that split there is no way to detect anchoring, and no way to answer a supervisor asking who decided.
- **Give every override a reason code and a number.** The reason code comes from a closed list. Override rate by reason, and the share later reversed, are the two figures that show whether human review is doing work.
- **Write the authority boundary before the model reaches it.** For each process, the money figure above which no automated output takes effect without a person. Setting it after deployment means setting it around what the system already does.
- **Put the same fields in the vendor contract.** Most of this will be bought rather than built: 80% of settlement amount models and 72% of image evaluation models in the NAIC survey came from third parties. An insurer that cannot obtain these fields from a vendor is holding a responsibility it cannot evidence.

For UAE insurers, February put a supervisory expectation behind that list. On 11 February 2026 the Central Bank of the UAE issued a guidance note on the responsible adoption of artificial intelligence by licensed financial institutions, with insurance providers in scope. Its legal status is principles-based supervisory expectation: the only "shall" sits in the definitions and the rest runs on "should."

Those definitions do something no other published instrument does. A high-impact decision is defined as any determination using AI that materially affects a customer's access to financial products or services, and an insurance claim is one of the two named examples. The regulator has named the insurance claim itself.

What the note expects against that definition is close to the list above: an inventory of every AI model with purpose and risk rating, audit trails and clear data provenance, explainability with greater force on high-impact decisions, meaningful human oversight with human-out-of-the-loop reserved for low-risk processes, and consumers able to request human review or an explanation. Institutions should not employ models they have no control over, and they remain responsible for AI functions they outsource.

Third-party administrators, claims platforms and AI vendors are not named in the note and are not direct addressees. They are captured through its outsourcing section, which leaves the licensed insurer to demonstrate oversight. The binding version sits in the Risk Management and Internal Controls Regulation for Insurance Companies, which requires outsourced material business activity to be overseen in the equivalent manner to work kept in-house and makes the company fully responsible for the risks arising from any process it outsources.



The gap to be closed. Every expectation above points at the machine. No instrument found in the UAE framework, or in any other reviewed for this paper, requires a record of whether a human or an automated system decided a claim outcome, or of what the human was looking at when they decided. The regime governs the corner with two working controls and is silent on the corner with none.

8. Method and limits

Each of the twelve processes was assessed on one question: of the work this process takes, how much is carried by generative AI, how much by deterministic control, and how much by a person. The three shares are allocated across a fixed total, so a gain for one carrier is always a loss for another and no process can improve on every axis at once. Every position in this paper is one of those allocations, and the triangles are that allocation drawn.

The today position is Axxion's read of industry practice. The end state is what the process can run at with technology that exists now, properly controlled, and it carries no date. The underlying allocations for all twelve processes, and the arithmetic behind the aggregate findings, are available to any insurer that asks for them. They are left out of this document because a reader following the argument does not need them, and printing twenty-four sets of figures obscures the point rather than proving it.

Four of the twelve are informed by published data: claim assignment, damage assessment, total loss valuation and fraud referral. The first three come from the automation and augmentation split in the NAIC private passenger auto survey, calculated across all three of its categories. Figures in wider circulation drop the survey's support column from the denominator and run high as a result. That data fixes the ordering of those processes against each other and does not set the allocation.

The remaining eight carry no external reference, because none was found. The NAIC survey excluded linear regression, generalized linear and generalized additive models, and anything a company could realistically have used in the year 2000 or earlier, so it understates machine involvement. It counts models in use or under construction, so none of its figures describes live deployment alone.

The allocations are judgment, not measurement.

The concentration finding in section 5 survives a one-step change to the allocation on any single process. The direction holds and the decimal does not, which is why it is stated as a direction and a rough share rather than to the percentage point.

What is not in the twelve. Coverage determination, reserve setting, hire-car authorization, salvage disposal and complaint handling were left out, to keep the set to processes running inside the claim file itself on facts the file contains. Coverage determination is the most defensible addition and would sit close to the control corner.



Jurisdiction. The CBUAE guidance note and the outsourcing regulation apply to UAE licensed institutions. None of the other instruments or datasets binds a UAE or GCC insurer. The NAIC survey is American, EIOPA covers 25 EU and EEA countries, and the research on human oversight of machine output comes from radiology, process control and government algorithm policy. They are included as evidence of direction and because they are the only published material of their kind. No equivalent GCC dataset was found.

Sources

CBUAE, *Guidance Note on the Consumer Protection and Responsible Adoption and Use of Artificial Intelligence and Machine Learning by Licensed Financial Institutions in the U.A.E.*, issued 11 February 2026. CBUAE, *Risk Management and Internal Controls Regulation for Insurance Companies*, Circular No. 25/2022. NAIC, *Private Passenger Auto Artificial Intelligence/Machine Learning Survey Results*, 8 December 2022, 193 responses across nine states. EIOPA, *Generative AI Market Survey: Outlook, Use Cases and Risk Management*, 2 February 2026, 347 undertakings across 25 countries. Capgemini Research Institute, *World Property and Casualty Insurance Report 2026*, 5 May 2026, 344 executives and 809 employees across 16 markets. Kahneman, Sibony and Sunstein, *Noise: A Flaw in Human Judgment*, 2021, chapter 2. Dratsch et al., *Radiology* 307(4), 2023. Manzey, Reichenbach et al. 2008, reported in Parasuraman and Manzey, *Human Factors* 52(3), 2010. Green, *Computer Law & Security Review* 45, 2022. Koulu, *Maastricht Journal of European and Comparative Law* 27(6), 2020.



About Axxion

Axxion Claims Settlement Services L.L.C. is the UAE's first dedicated motor third-party administrator. From Dubai, Axxion manages the full motor claims lifecycle for insurance partners: first notification of loss, surveying, repair coordination, quality control, recovery, and settlement. The company serves UAE insurers across both large and small-to-medium carriers and is preparing to extend into Saudi Arabia and the wider GCC.

Compliance by design. Axxion was built to operate inside a tightening regulatory environment. The Central Bank of the UAE absorbed insurance regulation in 2020 and consolidated the framework under Federal Decree-Law No. 6 of 2025, which brings TPAs and loss adjusters explicitly inside the CBUAE perimeter. Every claim Axxion handles passes through formal compliance gates covering UAE PDPL data protection, policyholder-consent requirements, settlement-authority bands, sanctions screening, and audit-trail completeness. Compliance is not an overlay; it is the operating substrate.

The Axxion Intelligent Operating System (AIOS). The ClaimOS that Axxion presents to insurer partners runs on the AIOS, a unified operating layer orchestrating a seven-stage claims pipeline across surveying, estimation, repair coordination, quality control, recovery, settlement, and reporting. The AIOS integrates human operators with structured AI-assisted decision points at every stage. Insurers receive cleaner data, faster cycle times, lower per-claim cost, and a complete audit trail, without giving up control over their portfolio.

More information: Managing Director and Co-Founder Frederik Bisbjerg · f@axxion.co · <http://www.axxion.co/>



Appendix:

Where the value sits, and how an insurer would measure it

Twelve processes scored against five benefit dimensions, with the measure that carries each one. The value is spread across the whole claim, and no single process carries the case on its own.

1. The value map

The map on the next page reads in two directions. Across a row: what one process does for the insurer. Down a column: which processes have to move for one benefit to arrive.

How to read it.

- **The color is the strength of the effect.** Dark navy means the process is a primary driver of that benefit. Pale blue means it contributes. Grey means no material effect. Hatched amber means the process makes that dimension worse before the claim gets better.
- **The codes inside each cell name the measure.** The letter is the benefit family and the number identifies the specific KPI within it: **T** for turnaround, **D** for data quality, **S** for satisfaction, **L** for leakage and fraud, **C** for cost. So T2 is the second turnaround measure, receipt to survey, and L1 is the first leakage measure, recovery capture rate. All 22 are defined in section 2.
- **The row total counts effects, not money.** Two points for a primary effect, one for a contributing effect, zero for no effect or a trade-off. It is a count of how many benefits a process touches and how hard, and it says nothing about currency. A process scoring 7 is not worth more than one scoring 4; it is broader, not bigger.
- **The dimension total does the same down each column,** showing how much of the claim has to change for that one benefit to materialize.

Three cells are marked as trade-offs, because two of these processes make a dimension worse before they make the claim better, and an appendix that hides that is worth less than one that shows it.



	Turnaround	Data quality	Satisfaction	Leakage & fraud	Repair cost	Total
01 FNOL extraction	T1 T2	D1 D3 D5	S2	L2	C3	7
02 Notification complete	T1	D1 D4	S3	L2	n/a	6
03 Claim assignment	T5 T6	D4	S2	L2	C3	6
04 Fraud referral	trade-off	D2	trade-off	L2 L3	C4	4
05 Damage assessment	T2	D2	S1	L2	C1 C3	7
06 Repair method	T4	D2	S1 C5	L2	C2 C3	6
07 Estimate challenge	trade-off	D2 D3	n/a	L2 L3	C1 C2 C4	4
08 Total loss decision	T5	D4	S3	L1	C2 C3	7
09 Total loss valuation	T3	D2 D4	S1 S3	L1	C3	6
10 Liability apportionment	T5	D2	S3	L1	n/a	5
11 Subrogation	T6	D5	S3	L1	C2	6
12 Settlement approval	T3 T5	D4	S1	L2	C2	6
Dimension total	15	14	11	16	14	70

◆ Primary effect ◆ Secondary effect ◆ Not material ◆ Gets worse first

KPI codes defined in section 2

No single process carries the full effect. Twelve processes score between four and seven out of ten, and the five dimensions land between 11 and 16. Value is spread across the whole claim. There is no silver-bullet process, and buying one AI product for one step returns a fraction of what is available. The benefit is a property of the operation rather than of any component in it, which is exactly why point solutions disappoint and why this has to be done end to end.

TWO TRADE-OFFS

Fraud referral costs time. A referral stops a claim while an investigation runs, so turnaround gets worse on every referred file, and a referral made against an honest claimant is the most damaging single event in a claim. Both costs are real and neither is a reason not to do it. They are a reason to govern it on precision rather than on volume, because a model rewarded for flagging will flag, and the cost lands on customers who did nothing wrong. The evidence position is honest and thin: Axxion's exception gate has stopped claims pre-authorization on coverage and documentation grounds, which is measured.

Estimate challenge costs time. Contesting an estimate means a conversation with a workshop, and that conversation takes time the vehicle spends off the road. An operation that deploys challenge without watching turnaround will book the saving and lose the service level, and the customer will experience the saving as a delay. The measure that resolves it is the sustained rate: a challenge that does not hold has cost days and returned nothing. Evidence: a 47% reduction against the initial workshop estimate is measured at claim level across a live pilot. Sustained rate by reason and by workshop is defined but not yet reported, which is the gap to close before this is presented as a repeatable capability.



2. Key Performance Indicators

Twenty-two measures carry the five dimensions across all twelve processes. The reference code is the benefit family plus a sequence number, so any cell in the value map can be traced straight to a definition here.

T turnaround · **D** data quality · **S** satisfaction · **L** leakage and fraud · **C** cost

Ref	KPI	Definition
T1	Notification to first touch	Hours from receipt to first substantive action. Median and 90th percentile.
T2	Receipt to survey	Working days, acceptance to inspection complete.
T3	Survey to authorization	Working days, assessment complete to repair authorized.
T4	Authorization to delivery	Working days to handback, banded by damage class.
T5	End-to-end cycle time	Calendar days, notification to handback. Median, never mean.
T6	Touch time against elapsed time	Hours worked over days elapsed. Separates waiting from working.
D1	Field completeness at intake	Share of the required field set populated at acceptance, weighted by downstream dependency.
D2	Structured capture rate	Share of claim facts held as fields rather than free text.
D3	Extraction accuracy	Field-level agreement between machine read and source document, on a sampled re-read.
D4	Downstream analyzability	Binary, per analysis: can severity, frequency, cycle time and recovery rate be computed without a fresh data request.
D5	Linkage integrity	Share of payment transactions linkable to a claim, a policy and a vehicle.
C1	Cost against opening estimate	Reduction from first workshop quote to authorization. Directional; overstates saving.
C2	Cost against matched baseline	Reduction against comparable claims in a control period. The only defensible savings measure.
C3	Cost per claim by severity band	Banded, so mix shift cannot be mistaken for saving.
C4	Challenge sustained rate	Share of contested lines that hold after workshop response, by reason and by workshop.
C5	Rework rate	Same vehicle returning inside 90 days. A saving that produces rework is not a saving.
L1	Recovery capture rate	Recovered against recoverable identified, with collection lag in days.



Ref	KPI	Definition
L2	Exception catch rate	Claims stopped pre-authorization by reason: coverage, duplicate, documentation, damage inconsistency.
L3	Referral precision	Share of fraud referrals substantiated. Guards against a model rewarded for volume.
S1	Post-delivery satisfaction	Score at handback within 48 hours, always reported with response rate.
S2	Update promise adherence	Share of open claims receiving a proactive update inside the promised interval.
S3	Complaint and dispute rate	Per 100 claims, split by cause: cost, delay, quality, communication.

3. The data-quality finding

The same pattern shows up in almost every dataset that reaches Axxion's operation: the fields the analysis needs were rarely captured in the first place. The gap sits primarily at intake, not anywhere downstream.

What that looks like in practice is consistent. Payment transactions arrive without a claim identifier, so several payments cannot be tied to one claim and per-claim severity cannot be produced at all. Exposure is missing, so frequency and burning cost cannot be calculated. Premium is missing, so loss ratio cannot be calculated. Reserves and claim status are missing, so ultimate loss cannot be estimated and averages end up mixing claims at different maturities. Vehicle detail stops at make, so vehicle-age effects cannot be drawn. Often there is no date on the money, which means collection speed cannot be computed even when that is the exact question being asked.

Every field named above is one that generative AI can populate at first notification, from documents that already arrive with the claim, provided a deterministic layer validates it and commits it.

The industry treats data quality as a reporting problem to be solved at the warehouse. **It is an intake problem**, and it is solved at first notification or not at all. An insurer that fixes capture gets the analytics as a by-product, and one that does not will keep commissioning extracts that cannot answer the question they were commissioned for.



4. The 12 motor claims processes

01 · FNOL document extraction

Turnaround	Complete file the first time a person opens it, not after keying
Data quality	Every downstream field is captured here or is never captured
Satisfaction	The claimant is asked once for each document
Leakage & fraud	Duplicate and coverage catches before any cost is incurred
Repair cost	Prevents the re-survey that a missing document causes

T1 T2 · D1 D3 D5 · S2 · L2 · C3

The highest-leverage process in the set. Every field the data-quality finding in section 3 identifies as missing is populated here or is never populated. Extraction stops being a queue, so the file is complete when a person first opens it rather than after they have keyed it.

Evidence: receipt-to-survey inside one working day, measured. Duplicate and coverage catches at intake, measured.

02 · Notification valid and complete

Turnaround	Removes the round trip a missing item causes three days later
Data quality	Decides whether the claim enters the book analyzable at all
Satisfaction	Avoids the late coverage decline, the worst complaint in motor
Leakage & fraud	A coverage failure caught here costs nothing
Repair cost	No direct effect

T1 · D1 D4 · S3 · L2

The gate that decides whether a claim enters the book analyzable or not. Coverage failures caught here cost nothing; caught after authorization they cost the full repair, and they produce the highest-severity complaint category in motor.

Evidence: coverage exceptions stopped pre-authorization, measured on pilot claims.

03 · Claim assignment

Turnaround	Removes the re-assignment cycle and the wrong-desk wait
Data quality	Track and authority as fields make handler performance computable
Satisfaction	Correct routing sets a promise the operation can keep
Leakage & fraud	Workshop concentration becomes visible at allocation
Repair cost	Under-routing turns a repairable vehicle into a total loss

T5 T6 · D4 · S2 · L2 · C3

The largest single source of dead time in a claim is a file on the wrong desk. Routing on predicted complexity rather than on intake attributes removes the re-assignment cycle, and under-routing a complex claim is how a repairable vehicle becomes a total loss through delay.

Evidence: workshop concentration measurable at assignment; volume skew observed on pilot panels.



04 · Fraud referral

Turnaround	Referral adds time to the claim; precision keeps that cost proportionate
Data quality	Reason codes on referral and non-referral make precision computable
Satisfaction	A wrong referral against an honest claimant is the most damaging single event
Leakage & fraud	Catches the cross-claim pattern no handler sees one file at a time
Repair cost	Supplement outliers surface before they arrive as cost

L2 L3 · D2 · C4

Catches what no handler sees one file at a time, because the signal is a correlation across claims rather than anything inside one. Adds days and carries reputational risk on a wrong call, which is why referral precision governs it rather than referral volume.

Evidence: none. No fraud case, value or detection rate measured. The largest claim-versus-evidence gap in the set.

05 · Damage assessment from images

Turnaround	Desktop assessment removes a physical visit worth one to three days
Data quality	Structured damage lines instead of a narrative report
Satisfaction	Fewer visits and an earlier answer
Leakage & fraud	Damage inconsistent with the described impact is detectable here
Repair cost	Sets the ceiling every later negotiation works down from

T2 · D2 · S1 · L2 · C1 C3

The step with the largest available time compression, because desktop assessment removes a physical visit. It also sets the ceiling every later negotiation works down from, so an error here propagates through the whole claim.

Evidence: machine assessment run against actual outcome on live claims; damage-inconsistency catch at physical inspection, measured.

06 · Repair method, OEM against aftermarket

Turnaround	Parts availability, not labor, is the binding constraint on duration
Data quality	Parts decisions as fields feed the benchmark governing the next claim
Satisfaction	Parts choice is the most common cause of a dispute at handback
Leakage & fraud	Billing OEM and fitting aftermarket is only visible against a record
Repair cost	Where policy permits a choice, the choice is worth real money

C2 C3 · T4 · D2 · S1 C5 · L2



Where policy permits a choice, the choice is worth a material share of the repair. Parts availability rather than labor is the binding constraint on most repair durations, and parts choice is the most common cause of a repair disputed at handback.

Evidence: channel severity gap quantified on an insurer book; agency-channel cost ran a multiple of non-agency.

07 · Repair estimate challenge

Turnaround	Challenging an estimate adds time; the trade-off has to be visible, not hidden
Data quality	Line-level normalization is what builds the benchmark
Satisfaction	No direct effect, and negative if a challenge delays without reducing
Leakage & fraud	Labor-hour and supplement outliers by workshop
Repair cost	The single largest cost lever in the claim

C1 C2 C4 · D2 D3 · L2 L3

The single largest cost lever in the claim, and today it is only as good as the individual handler running it. Line-level normalization against a benchmark distribution replaces experience with a position that can be defended, tracked and improved.

Evidence: 47% average reduction against the initial workshop estimate, measured at claim level, with a wide per-claim spread. Measured against a matched baseline rather than the initial quote, the figure would be lower.

08 · Total loss determination

Turnaround	Deciding total loss late is the most expensive delay in motor claims
Data quality	Salvage realization against the threshold makes the threshold testable
Satisfaction	Total loss claims carry the highest dispute rate of any disposition
Leakage & fraud	Salvage under-recovery is leakage and is rarely tracked as such
Repair cost	Repairing a write-off, and writing off a repairable, are both pure loss

T5 · C2 C3 · D4 · S3 · L1

Deciding total loss late is the most expensive delay in motor claims: storage accrues, the vehicle depreciates and the customer waits without a car. Repairing a vehicle that should have been written off, and writing off one that should have been repaired, are both pure loss.

Evidence: total loss disposition tracked on pilot claims; salvage realization not yet tested against the threshold.



09 · Total loss valuation

Turnaround	Valuation disputes are a leading cause of long-tail claims
Data quality	Valuation inputs as fields make the figure defensible and reviewable
Satisfaction	The most disputed number in motor claims
Leakage & fraud	Under-settlement arrives as a complaint rather than as a cost
Repair cost	On these claims the settlement is the cost

S1 S3 · D2 D4 · T3 · L1 · C3

The most disputed number in motor claims. A disclosed comparable set with documented adjustments converts an argument into a conversation, and valuation disputes are a leading cause of long-tail claims.

Evidence: none specific to valuation. Dispute rate is not yet captured by cause.

10 · Liability apportionment

Turnaround	Contested liability is the main driver of claims running past sixty days
Data quality	Reasoning recorded is what makes handler consistency measurable
Satisfaction	Drives premium consequences the claimant cares about more than the repair
Leakage & fraud	Determines whether a recovery exists at all; an error forfeits it silently
Repair cost	Indirect, through recovery

L1 · T5 · D2 · S3

Determines whether recovery is available at all, so an error here forfeits the recovery silently and nobody ever sees it. Contested liability is the main driver of claims running past sixty days, and apportionment drives premium consequences the claimant cares about more than the repair.

Evidence: at-fault versus recovery split quantified on insurer books; handler-to-handler consistency never measured anywhere.

11 · Subrogation viability

Turnaround	Recovery lag runs in months and is almost never on a dashboard
Data quality	Recovery cannot be managed where the money carries no date
Satisfaction	Excess refunds follow recovery, so a slow recovery is a slow refund
Leakage & fraud	The quietest and largest opportunity in the set
Repair cost	Net cost per claim is half recovery

L1 · D5 · T6 · C2 · S3

The quietest and largest opportunity in the set, because an unpursued recovery generates no complaint, no dispute and no file. Recovery cannot be managed where the money carries no date, and on one book analyzed no settlement or recovery date existed at all.

Evidence: gross recovery outstanding stood at a multiple of what was collected over the same period, measured on an insurer extract. The insurer's own view was that most of the balance was collectible.



12 · Settlement approval

Turnaround	Approval queues are pure waiting
Data quality	Who approved what, against which limit, is the first record a supervisor asks for
Satisfaction	Time to money is what the customer experiences as service
Leakage & fraud	Payment validation and duplicate-payment prevention
Repair cost	Authority set against evidence is where the saving is locked in

T3 T5 · D4 · S1 · L2 · C2

Approval queues are pure waiting. A widened, written and versioned authority matrix removes the wait without removing the control, and time to money is what the customer experiences as service.

Evidence: payment validation and money bounds enforced in the platform rather than in a procedure document.

5. Testing this on a live book

Everything in this paper is checkable. The allocation across the twelve processes, the finding that control does not shrink as generative AI grows, the concentration of human judgment into four decisions – none of it has to be taken on trust. It can be tested against an insurer's own claims inside a single quarter.

That is what a pilot is for. A defined slice of claims runs through the governed process described here, alongside the existing operation rather than in place of it. The insurer keeps assessment, approval and settlement authority throughout. Nothing is migrated, no core system is touched, and the arrangement can be stopped at short notice.

What the insurer holds at the end

- **A cost reduction with a method attached.** Measured on the insurer's own claims, line by line, against its own historical baseline, so the figure can be handed to an actuary rather than to a marketing department.
- **Cycle time as a distribution instead of an impression.** Stage by stage, banded by damage class, with waiting separated from working. Most operations have never seen their own claims this way.
- **The first measurement of its own handler variance.** How far two competent people diverge on comparable files. No insurer has published this figure, in any market. The first one to hold it holds something nobody can argue with.
- **A governance record that survives a supervisory question.** Every machine output recorded as a proposal, every commit logged, every override reasoned and counted.

The CBUAE guidance note of February 2026 sets its expectations against AI used in high-impact decisions, and an insurance claim is one of the two examples it names. An operation that can already produce a model inventory, an audit trail and an override record is answering a question that has not formally been asked yet.



What has to be in place for the result to mean anything

A pilot without a baseline can only report what happened, not what changed. These are agreed before the first claim is routed, not fitted around the results afterwards.

- **18 to 24 months of historical repair data, banded by severity.** The highest-value item by a distance. It is the difference between reporting a 47% reduction against the initial workshop estimate, which is measurable from day one, and demonstrating what the insurer saved against how the same claims would otherwise have run. Choosing the comparison period after the results are in invites the suspicion that it was chosen to flatter them.
- **A written definition of the measures that count.** Cycle time, cost per claim by severity band, recovery capture rate, satisfaction. Settling these in advance prevents the argument at the end where both sides measure the same operation and reach different answers.
- **What the machine proposed and what the person concluded, as two separate fields.** This paper's central argument, applied to the pilot itself.
- **Override reason codes from a closed list,** reported with the override rate and the share later reversed.
- **Satisfaction captured at handback within 48 hours,** with the response rate reported beside the score.
- **Authorization-to-delivery timing, banded by damage class.** The data usually already exists and has never been aggregated.
- **Recovery capture rate and collection lag.** The quietest money in the claim, and it appears on almost no dashboard.

Fraud referral sits outside a first pilot until referral precision is tracked. A detection capability without a precision figure invites the obvious question, and the answer has to exist before the claim is made.

Starting point

One working session with the claims data in front of both sides: whether the baseline exists, which slice of claims to use, and what the insurer wants to prove.

***Note on sources.** The KPI design and the baselines in this appendix were informed by claims analytics engagements carried out during 2026. No insurer is identified, and dataset sizes, claim counts and absolute values have been withheld so that no book can be recognized from its own figures. Savings are framed as reduction against the initial workshop estimate rather than as a loss-ratio saving, and require VP of Actuarial sign-off before external use.*