SUTHERLAND®

When You Can't Trust the Claim File

How Synthetic Evidence Is Changing P&C Fraud Detection

The claim file used to be the evidence. Now it may also be the attack surface. A practical framework for establishing claim integrity.

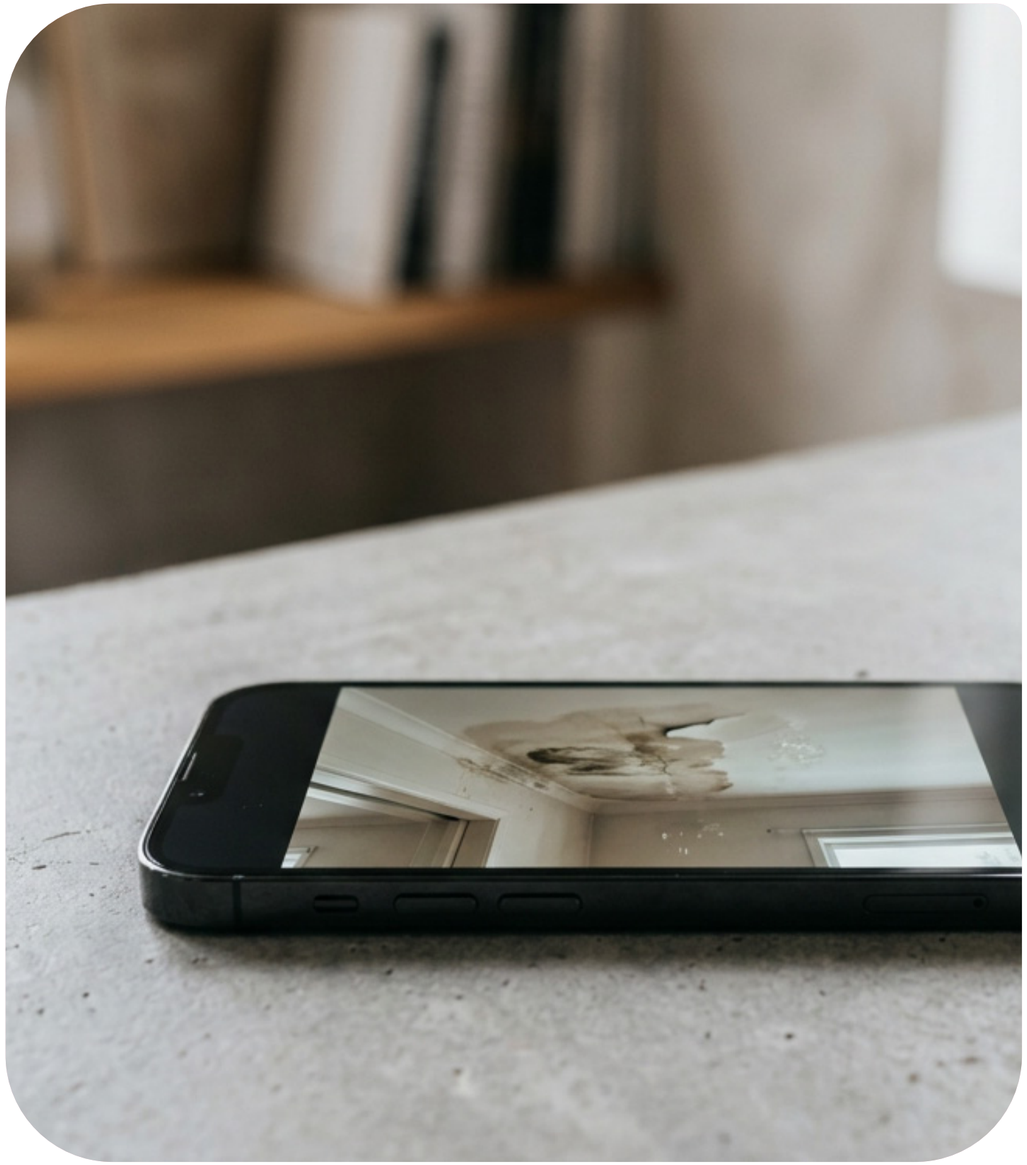
1. When the Evidence Becomes the Fraud

Over 10 days in November 2025, a North Carolina woman allegedly submitted five insurance claims for damaged items. The accompanying photographs appeared to support her account. It all looked real enough. But investigators say some had been altered using AI, while others had been taken from the internet. Seven months later, she was charged with 10 felony counts.

The claims were for $4,876 in total. The amount is relatively modest, and that is precisely what makes the case significant. It did not require an organized fraud ring, specialist software, or an elaborate high-value scheme. Consumer AI tools allegedly enabled a person to quickly manufacture initially convincing evidence.

Of course, the risk of fraud is not new. Claims teams have always dealt with the possibility of fabricated evidence. But generative AI changes the scale and economics of the problem, making credible-looking fakes faster, cheaper, and easier to produce repeatedly.

A genuine loss can be exaggerated with manipulated evidence. A real person or asset can be attached to an event that never happened. At the furthest extreme, almost every component of the claim file can be generated: the photograph, invoice, repair estimate, medical record, or police report.



1 https://www.ncdoi.gov/news/press-releases/2026/06/19/guilford-county-woman-faces-insurance-fraud-related-charges-using-ai-alter-photos

NICB estimates that at least 10% of P&C claims in the US may be fraudulent.[2] In the UK, Aviva identified more than 18,400 suspected fraudulent claims worth £233 million across its brands in 2025.[3] In 2026, the concern is not simply that fraud will continue, but that AI-generated evidence will make fraudulent claims faster to create and harder to distinguish from legitimate ones.

The lower the barrier to fabrication, the less confidence any individual document or image can provide.

That places greater pressure on established fraud controls, many of which focus on suspicious behavior around the claim: unusual loss patterns, inconsistencies in the claimant's account, links to previous claims, or known fraud networks. Those signals remain important, but they cannot establish the authenticity of the underlying evidence on their own. Yet closing that gap through blanket manual review would add cost, delay, and friction to legitimate claims.

2 https://www.nicb.org/prevent-fraud-theft
3 https://www.aviva.com/newsroom/news-and-research-overview/news-releases/2026/06/aviva-stops-record-levels-of-claims-fraud-as-scams-become-more-sophisticated/

2. Every Claim Can’t Become an Investigation

If fraudsters can use AI to manufacture claim evidence at speed and scale, insurers need an equally scalable way to assess it. This is driving the use of AI and machine learning to analyze evidence, identify anomalies, connect claims and prioritize cases for investigation. In effect, insurers are fighting fire with fire.

Among these capabilities are tools designed specifically to detect AI-generated or manipulated content. But their performance can vary under real-world conditions, and no detector can establish claim integrity on its own. AI detector findings, therefore, need to be considered alongside the evidence’s provenance, the wider claim, and independent signals about the physical event.

AI detection gets harder in real-world conditions

A photograph may contain signs of manipulation without proving that the claim is fraudulent. Equally, a convincing fake may pass through undetected.

- NIST’s work on evaluating deepfake-analysis systems highlights the importance of testing against realistic conditions rather than relying solely on controlled benchmarks.[4]
- Researchers benchmarked 15 top-performing deepfake detectors across 12,832 videos captured using different screens, smartphones, lighting conditions, and camera angles. Moiré artifacts, the visual interference patterns created when one screen is recorded through another device, degraded detector performance by up to 25.4%.[5]

Although these studies were not insurance-specific, they demonstrate why a detector score cannot be treated as definitive proof that claim evidence is genuine or fabricated.

4 https://www.nist.gov/publications/guardians-forensic-evidence-evaluating-analytic-systems-against-ai-generated-deepfakes
5 https://arxiv.org/abs/2510.23225

3. The Claim Integrity Framework

Synthetic evidence cannot be addressed through a single detector or fraud score.

Claims organizations need to build confidence progressively, combining signals about where evidence came from, whether it has been manipulated, and whether it makes sense within the wider claim.

The operating principle behind **Sutherland's Claim Integrity Framework** is risk-based orchestration. Different integrity checks fire at the stage where they add value, using the signals already available and escalating only when confidence falls below a defined threshold. Straightforward claims continue. Ambiguous claims gather more corroboration. High-risk or consequential decisions reach the right specialist with the relevant evidence assembled.

The objective is to establish whether the claim evidence is sufficiently trustworthy for the next decision. This makes the claim's progression conditional on confidence, while retaining an auditable record of why it moved, paused, or escalated.

Sutherland's Claim Integrity Framework

Capture provenance at source

Where it applies: ***FNOL and evidence submission***

Collect evidence through controlled channels where possible, retaining available capture signals such as time, location, device information, and content credentials. This gives the carrier a clearer record of where evidence originated and how it entered the claim.

Test the media

Where it applies: ***Intake and triage***

Assess photographs, videos, and documents for signs of generation, manipulation, or reuse. Relevant signals may include editing artifacts, inconsistent metadata, duplicate imagery, and discrepancies between multiple views.

Validate the physical event

Where it applies: ***Triage and investigation***

Compare the claim with independent sources such as weather data, telematics, geolocation data, repair histories, IoT records, satellite imagery, or police information. The objective is to determine whether the reported event is supported by evidence beyond the claimant's submission.

Validate internal consistency

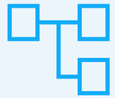
Where it applies: ***Triage and investigation***

Assess whether the evidence agrees with the wider claim. Images, damage descriptions, estimates, asset details, injury narratives, and accident mechanics should form a coherent account of what happened.

Analyze the network

Where it applies: ***Investigation***

Connect identities, devices, addresses, bank accounts, repairers, witnesses, and previous claims. An individual claim may appear plausible while still belonging to an implausible or organized pattern.



Use synthetic adversarial testing

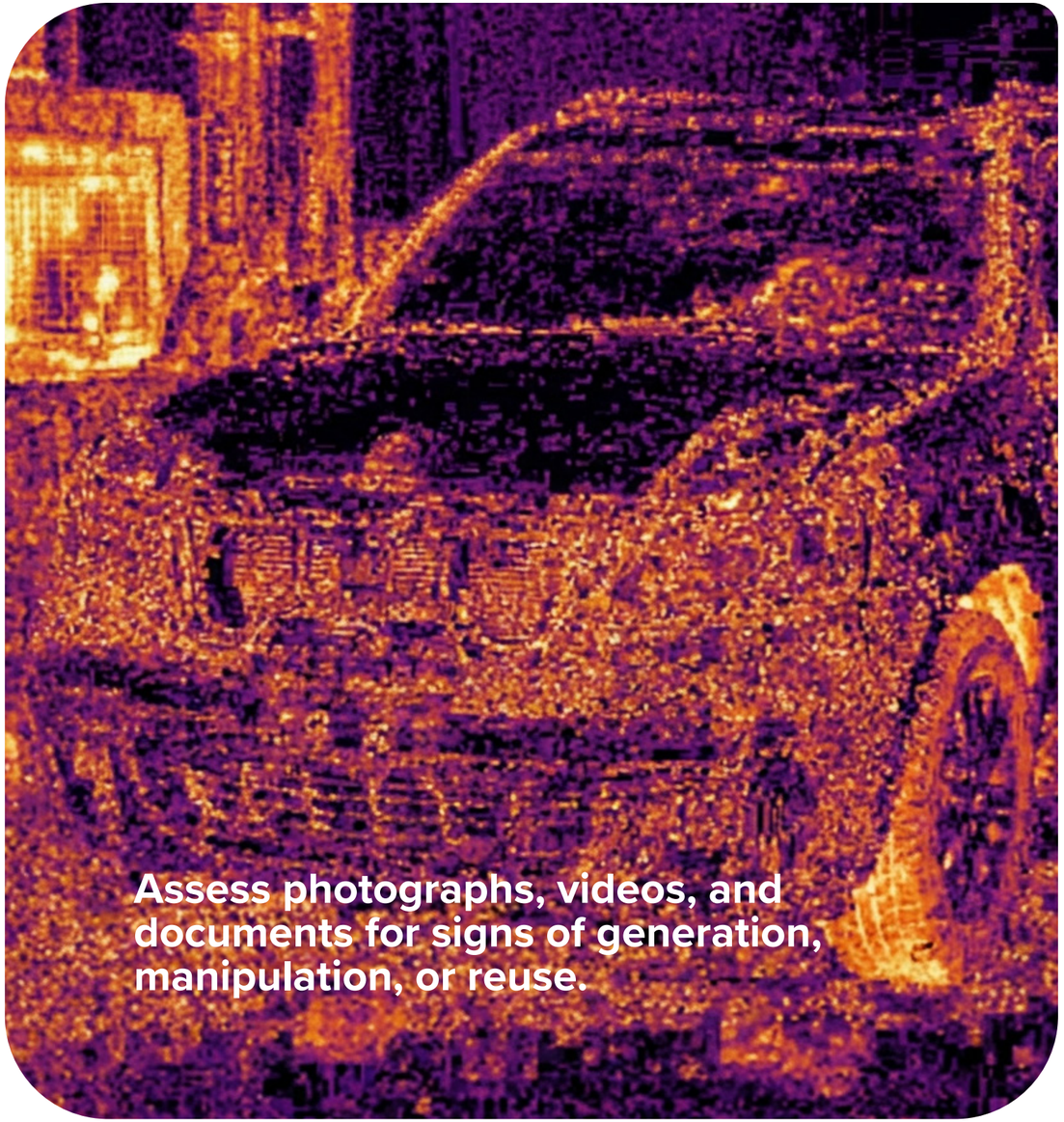
Where it applies: ***Investigation***

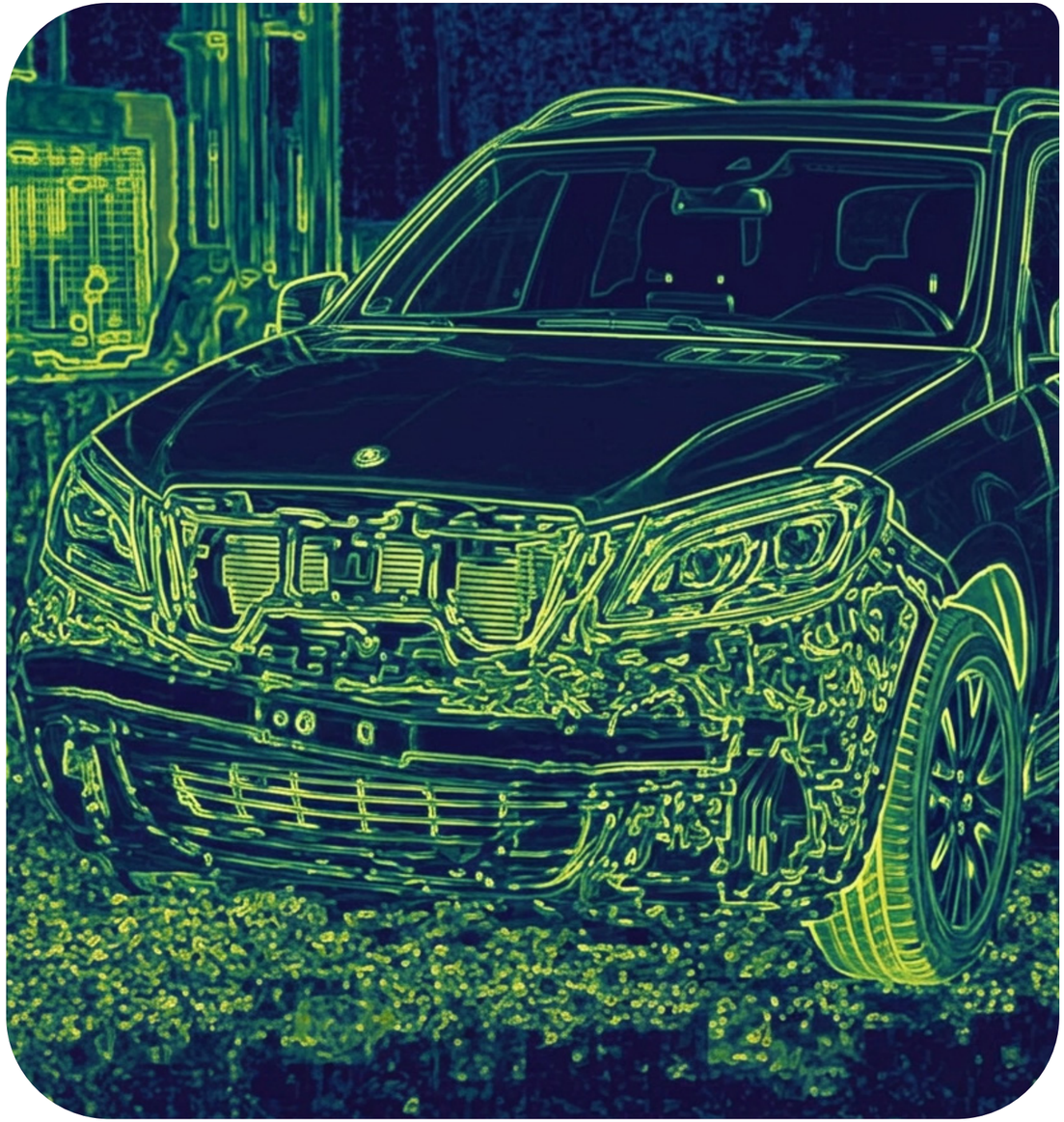
Use controlled synthetic evidence and variations of known fraud scenarios to test where automated workflows and detection models fail. This helps claims teams identify vulnerabilities before fraudsters exploit them and adapt controls as tactics evolve.

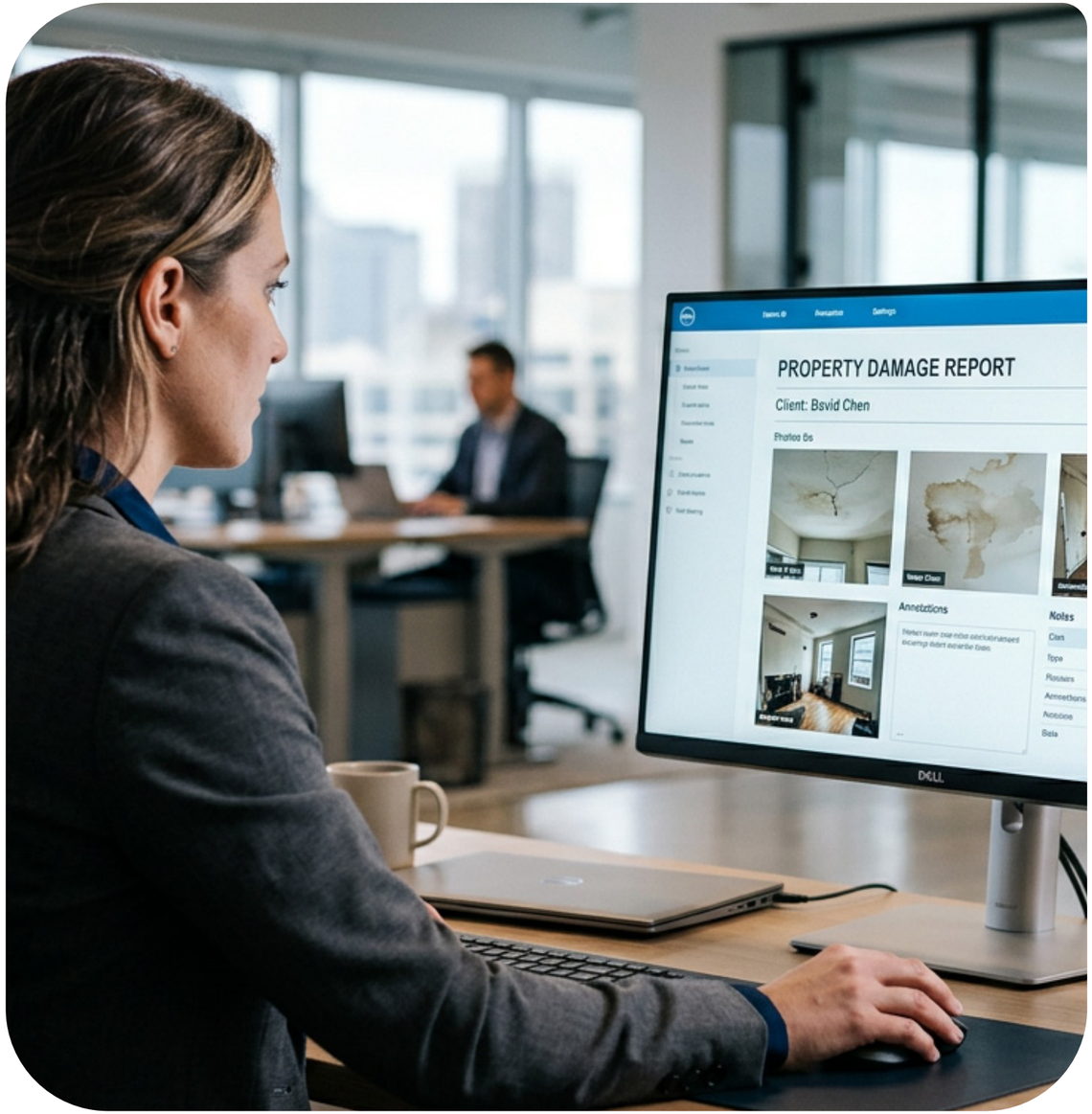
Preserve human escalation

Where it applies: ***Defined exception and decision points across the claim***

Route the claim to the appropriate licensed or specialist reviewer when confidence remains below the required threshold, the case falls outside established parameters, or human judgment is required by policy or regulation. Provide the reviewer with the signals, discrepancies, and evidence already gathered, and retain an auditable record of the referral and resulting decision.


Assess photographs, videos, and documents for signs of generation, manipulation, or reuse.





A control plane across every layer

Explainability, auditability, and decision rights sit across the framework rather than forming an eighth layer. For every automated action, the carrier should be able to see what signals were used, what conclusion was reached, what happened next, and when a human was required.

This matters operationally and regulatorily. As of August 2026, 24 US states plus the District of Columbia had adopted the NAIC Model Bulletin on insurers' use of AI; California, Colorado, New York, and Texas had separate insurance-specific regulation or guidance. The bulletin expects a written AI governance program proportionate to risk and makes clear that insurers remain responsible for third-party systems.

6 https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-ai-model-bulletin.pdf.pdf

4. Where to Start: Putting the Claim Integrity Framework to Work

Most carriers already have some of these capabilities in place, spread across systems, teams, and stages of the claim. The practical challenge is to orchestrate them so that each integrity check fires when it can inform the next decision. Claims leaders can begin by mapping how confidence is established today.

Five questions can help expose the most important gaps in your current operating model:

1 At intake:

Do we know where the evidence came from, how it was captured, and whether its origin can be verified?

2 Before the next decision:

Can we corroborate the file against the physical event, the wider claim, and connected entities?

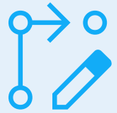
3 In the flow:

Can additional scrutiny be triggered selectively without slowing every credible claim?

4 At escalation:

Does the specialist receive the evidence, contradictions, and reasoning needed to decide, rather than another undifferentiated queue item?

5 After the decision:



Can we reconstruct why automation progressed, paused, or referred the claim, including the role of any third-party model?

5. Building Claims Processes for an Era of Synthetic Evidence

Synthetic evidence does not replace traditional fraud. It compounds it. Staged events, inflated costs, identity misuse, and organized networks remain; generative AI makes the supporting story cheaper to construct, easier to vary, and harder to judge from a single artifact. For claims leaders, the strategic response is therefore not a race to buy the best fake-image detector. It is to redesign how confidence is established and acted upon across the claim.

Sutherland's Insurance AI Hub for Claims turns the Claim Integrity Framework from a design principle into an operating model. More than 60 purpose-built agents span intake, triage, document and fraud intelligence, decisioning, compliance, and communication, with Robility coordinating which capability acts next.

Sutherland Extract structures and validates incoming documents and images, while CognilinkClaims and SUL connect the workflow and claim data across existing Guidewire, Duck Creek, EIS, and FINEOS environments. This allows carriers to introduce new integrity checks without replacing the claims core.

Explainability and Compliance agents record the rationale behind each progression, pause, or escalation. Where confidence falls below the required threshold, the claim can be routed to the appropriate human reviewer, backed by licensed US adjudication, FCA-certified UK operations and more than 200 pre-litigation adjusters.

Together, these capabilities allow carriers to establish confidence earlier, keep credible claims moving and direct specialist attention toward the evidence and decisions that genuinely require it.

The result is better-directed scrutiny — and a claim that can keep moving.

From Orchestration to Outcomes

60+
purpose-built agents spanning intake, triage, document validation, fraud, decisioning, compliance and communication.

20%
NPS uplift

Zero Touch
Claims process

70%+
FNOL Automation

40%
Cost reduction

35%
Cycle time reduction

15%
Leakage reduction

The technology is backed by licensed US adjudication, FCA-certified UK operations and more than 200 pre-litigation adjusters.

Explore Sutherland's Insurance AI Hub for Claims

Sutherland is an AI-driven business transformation partner to enterprises, including several Fortune 500 clients. We design, run, and automate business operations at scale, helping clients improve efficiency, strengthen customer experiences, and deliver measurable outcomes.

Most companies either build AI-led capabilities or run operations. Sutherland does both. We run complex, mission-critical processes while engineering the AI-led solutions that sit within them. This closed loop helps clients move beyond AI pilots to reliable, governed outcomes that improve and scale over time.

With deep operational expertise, advanced products, platforms, and engineering, we help clients modernize operations and build future-ready businesses.

SUTHERLAND®
driving outcomes.

sutherlandglobal.com