REAL TIME
RISK SOLUTIONS

Claims Data as Infrastructure

How AI-Driven Claims Normalization Is Transforming Loss Intelligence

A Research Paper on the Future of Claims Analytics in Insurance

Executive Summary

Insurance organizations sit on mountains of claims data, yet much of it remains locked in incompatible formats and legacy systems. As much as 80% of large insurers' data is unstructured, with 88% of it going unanalyzed[1]. Loss run reports arrive as PDFs, spreadsheets vary by carrier, and critical insights hide in adjuster notes and scanned documents. The result is a claims ecosystem that is labor-intensive, slow to react, and backward-looking. Analysts spend up to 80% of their time searching for and cleaning data instead of analyzing it, while business leaders make decisions on barely 12% of available information. This gap between data abundance and decision-grade intelligence is a structural inefficiency and a durable opportunity.

AI-driven claims normalization offers a way out. By combining OCR, data harmonization, and natural-language analysis, insurers can transform fragmented documents into a continuously updated, structured dataset. Adoption is moving from experimentation to execution. 76% of US insurance firms had implemented generative AI in at least one business function by 2024[2], with claims among the most practical use cases. Early adopters of AI in claims have seen 30%+ reductions in processing costs and cycle times[3,4]. Leading carriers report 3-5% improvements in loss ratios by using data-driven insight to manage risk earlier[5].

Over five years, AI leaders achieved 6.1× the total shareholder return of laggards[6].

These outcomes are not driven by a single algorithm. They come from building the data foundation that makes analytics trustworthy and scalable.

Claims normalization also produces measurable hard-cost savings across labor, external fees, and claim outcomes. An analysis estimates the industry wastes $17-32 billion annually on administrative inefficiencies such as rekeying, reconciliation, and repetitive reporting[7]. Normalization reduces this friction, reallocating skilled time from data cleanup to risk management. The largest impact sits inside claims themselves. Early intervention can reduce total incurred loss by 10-25% on affected claims. One global carrier reported $414 million in annual savings by deploying AI across underwriting and claims to identify outliers and fraud earlier. These are recurring, compounding advantages when the operating system for claims becomes data-first.

AI's greatest impact on claims is not automation alone. It is clarity, continuity, and cost control.

1. https://insuranceblog.accenture.com/heres-why-insurers-struggle-to-get-value-from-their-data
2. https://www.insurancethoughtleadership.com/ai-machine-learning/ai-insurance-2025-predictions
3. https://www.ust.com/en/insights/the-ai-advantage-in-claims-reimagining-the-heart-of-the-insurance-experience
4. https://datagrid.com/blog/ai-agent-for-insurance-statistics
5. https://www.hyperexponential.com/blog/ai-for-p-c-insurers
6. https://www.mckinsey.com/industries/financial-services/our-insights/the-future-of-ai-in-the-insurance-industry
7. https://blog.talli.ai/claims-payout-statistics/

1. The Claims Data Paradox

Insurance organizations generate vast volumes of claims data, yet decision-making within claims management remains heavily manual. The underlying constraint is structural. The majority of claims data exists in unstructured or semi-structured formats, including carrier-specific PDF loss runs, inconsistent spreadsheets, scanned adjuster notes, widely different nomenclature and narrative medical reports. Approximately 97% of claims data falls into these unstructured categories[8].

Each carrier or third-party administrator defines fields differently. Terminology varies. Even foundational measures such as paid, reserved, and incurred losses are not consistently aligned across sources. As a result, organizations with years of historical claims data often struggle to answer basic questions with confidence.

Persistent challenges include separating frequency trends from severity trends, identifying statistically meaningful outliers, comparing performance across facilities or carriers, and determining whether loss performance is improving or merely shifting between categories. Traditional RMIS platforms and spreadsheet-driven workflows were not designed for multi-carrier, multi-format, continuously changing datasets.

Legacy technology compounds the issue. 74% of insurers continue to rely on legacy systems for core functions such as claims processing[9]. Maintenance of these systems absorbs as much as 70% of IT budgets, limiting investment in modern data infrastructure. Cultural inertia further slows change. As noted by leadership within large carriers, the technical challenge is often secondary to altering established workflows and trust models around data.

The operational consequence is a widening gap between data availability and data usability.

Accenture research shows that analysts in large organizations spend roughly 80% of their time locating, cleaning, and preparing data rather than analyzing it.

In claims functions, this dynamic manifests as repeated reconciliation exercises, brittle spreadsheets, and delayed insight. The paradox is not data scarcity but data inaccessibility.

2. The Hidden Cost of Unstructured Claims Data

The cost of unstructured claims data is pervasive and largely invisible in financial statements.

Operationally, claims teams devote substantial time to manual extraction from PDFs, reformatting spreadsheets, rebuilding reports, and validating totals across reporting cycles. A mid-sized organization may spend between 120 and 600 hours annually on data preparation alone, depending on the number of carriers and reporting cadence. At a fully loaded labor cost of approximately $75 per hour, this equates to $9,000-$45,000 per year in direct labor expense dedicated to data preparation rather than analysis.

At an industry level, these inefficiencies scale dramatically. Estimates place annual administrative waste in insurance between $17 and $32 billion. Underwriters spend up to 40% of their time on administrative work, and similar patterns exist across claims and analytics functions.

[8] https://www.theclm.org/Magazine/articles/mining-for-information-gold-in-unstructured-claims-data/112

[9] https://www.insurancebusinessmag.com/us/news/technology/are-legacy-systems-weighing-down-the-insurance-industry-514574.aspx

Strategic costs are more consequential. Fragmented data delays identification of emerging loss drivers, interrupts continuity during carrier transitions, and undermines confidence in reported results. Poor data quality has been estimated to cost organizations 15-25% of revenue[10]. Within insurance, this manifests as reserve volatility, delayed corrective action, and inconsistent communication with leadership.

Timing amplifies financial impact. Delayed detection of abnormal claim trajectories increases the likelihood of litigation, prolonged disability, and excessive medical spend. AI-based detection systems have identified suspicious claims within two weeks of filing, while manual processes often surface issues only after payments have escalated[11]. Each week of delay narrows the opportunity for cost-effective intervention[12].

Unstructured claims data functions as an implicit tax on the organization. It inflates overhead, obscures risk signals, and increases the probability of adverse claim development.

3. What Claims Normalization Actually Means

Claims normalization is frequently misunderstood as basic digitization. In practice, it is a multi-layer process that converts documents into operational infrastructure.

Layer 1: Document Ingestion

Loss runs, PDFs, spreadsheets, scanned reports, and other claim artifacts are ingested regardless of source format or carrier origin.

Layer 2: Optical Character Recognition (OCR)

Text and tabular data are extracted from documents, including those with inconsistent layouts and low-quality scans. Insurance-specific OCR models address variability in formatting and terminology.

Layer 3: Data Harmonization

Carrier-specific fields are mapped into a unified schema. Dates, currencies, classifications, and terminology are standardized. Duplicate records across reporting periods are resolved, and updates are applied to existing claims rather than creating redundant entries.

Layer 4: Structured Intelligence

Claims become comparable across carriers, facilities, departments, and time periods. At this stage, analytics, dashboards, and AI summaries operate on consistent, reliable inputs.

This is the critical distinction.

Normalization creates a claims dataset overtime that behaves like a system of record.

That dataset can feed reporting, analytics, reserving review, litigation monitoring, and executive summaries without rebuilding the data layer each cycle. It turns recurring work into one-time configuration, then compounds value as more data flows through the same framework.

[10] https://www.insurancethoughtleadership.com/our-partners/true-cost-big-bad-data

[11] https://claraanalytics.com/news/clara-analytics-study-reveals-ai-as-early-warning-system-for-insurance-fraud/

[12] https://www.bakertilly.com/insights/2025-insurance-industry-outlook

4. From Static Reports to Living Data

Traditional loss runs are snapshots. They are prepared artifacts in a specific carrier format at a moment in time. The operational reality becomes version confusion, lagged insight, and repeated reconciliation at renewal.

Normalized claims data is living data. When new loss runs are uploaded, existing claims update automatically, historical trends recalculate in real time, and metrics remain consistent across periods. This eliminates version confusion, reduces lag between claim activity and insight, and reduces repeated reconciliation at renewal time.

Living data changes the operating cadence. Instead of quarterly data marathons, teams monitor loss drivers continuously. This supports faster action on emerging trends, better internal reporting, and more credible conversations with underwriting partners. It also improves governance. A single source of truth reduces disputes about totals and shifts attention toward decisions.

5. Why Normalization Enables Predictive Insight

Unnormalized data restricts organizations to backward-looking metrics. Normalized data enables predictive insight.

With a unified dataset, teams can perform frequency vs severity decomposition, facility-level and department-level benchmarking, injury-type trend detection, and lag analysis for reporting, treatment, and closure. These are not sophisticated on their own. Their value is that they become reliable across carriers and over time.

When paired with natural-language AI, normalized data becomes more accessible. Executives receive plain-language summaries. Analysts query trends conversationally. Insights become usable beyond technical users. This matters because claims outcomes depend on timely decisions by claims leaders, risk managers, and operations leadership, not only by analysts.

Early identification is the economic hinge. One study found that reducing attorney involvement by addressing issues early cut legal involvement by 15% and reduced claim costs by about 5% in impacted cases[13]. AI systems have detected patterns in claims within two weeks of filing, accelerating intervention windows. These timing advantages compound because preventing escalation on a small set of claims can shift the entire loss distribution.

[13] https://www.gradientai.com/news_ai-reduces-costs-for-most-expensive-types-of-workers-comp-claims

6.1 Administrative Labor Savings

Typical annual effort for a mid-sized organization can range from 120 to 600 hours spent on data preparation. With normalized claims data, reporting time can be reduced by 60-80%, since preparation occurs once and the dataset updates continuously. Example annual savings:

Scenario	Hours Saved	Fully Loaded Cost ($75/hr)
Conservative	150 hrs	~$11k
Moderate	300 hrs	~$22.5k
Aggressive	500+ hrs	$37.5k+

These savings recur annually and scale with program complexity.

6.2 Reduced External Consulting and Broker Costs

Organizations often rely on brokers or consultants to reconcile multi-carrier claims data, prepare renewal loss summaries, and identify trends manually. Normalized data reduces dependence on these services by eliminating repetitive cleanup, shortening renewal cycles, and enabling internal teams to self-serve analysis[14]. Example annual savings:

Scenario	Hours Saved	Fully Loaded Cost ($75/hr)
Conservative	Partial consulting reduction	$10k-20k
Moderate	Elimination of recurring analysis projects	$25k-50k
Aggressive	Reduced reliance across multiple programs	$75k+

These ranges are consistent with the broader trend toward insourcing enabled by automation in insurance operations.

6.2 Reduced External Consulting and Broker Costs

The largest savings occur inside the claims themselves. Normalized data surfaces claims with abnormal cost trajectories, medical-only claims trending toward indemnity, and facilities or job classes driving disproportionate losses.

Industry studies consistently show early intervention can reduce total incurred by 10-25% on affected claims.

Example annual impact (illustrative):

Program Size	Claims Affected	Avg Reduction	Savings
Conservative	3–5 claims	$15k/claim	$45k-75k
Moderate	5–10 claims	$25k/claim	$125k-250k
Aggressive	10+ claims	$50k+/claim	$500k+

Even modest improvements in timing produce outsized financial returns.

[14] https://www.insurancethoughtleadership.com/ai-machine-learning/ai-insurance-2025-predictions

6.4 Cumulative Annual Impact

When combined, savings compound:

Category	Conservative	Moderate	Aggressive
Labor Savings	$10k	$22k	$35k+
Consulting Reduction	$15k	$40k	$75k+
Claim Severity Reduction	$50k	$175k	$500k+
Total Annual Impact	$75k+	$235k+	$600k+

These are recurring savings driven by structural change, not one-time efficiencies.

7. Implications Across the Insurance Ecosystem

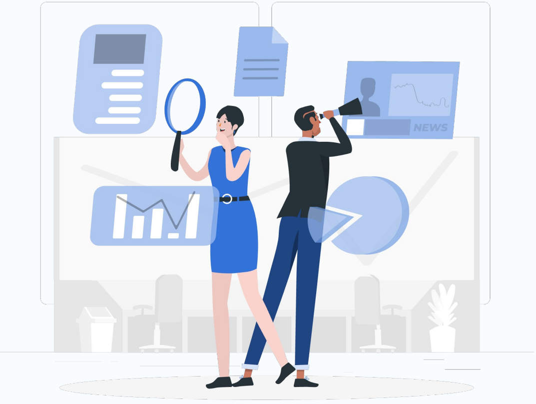
Brokers benefit from faster, more accurate renewals and cross-carrier visibility for clients, enabling higher-value advisory conversations.

Carriers and TPAs gain earlier identification of high-risk claims, improved reserve accuracy, and portfolio-level insight when normalized data supports consistent monitoring and collaboration.

Risk managers gain visibility across facilities and operations, continuity during carrier transitions, and data-driven prevention strategies. Claims normalization elevates claims from administrative necessity to strategic intelligence.

Conclusion: From Efficiency to Economics

The transition from PDFs to predictions is an economic shift. AI-driven claims normalization reduces labor and consulting costs, improves claim outcomes through earlier insight, and creates a durable, reusable data foundation. Organizations that treat claims data as infrastructure, not paperwork, gain sustained advantages in cost control, underwriting performance, and risk management effectiveness.



The future of claims is not more reports.
It is better understanding, earlier action, and lower cost.

Contact us:

info@realtimerisksolutions.com

www.realtimerisksolutions.com