

Travelers - Scaling AI Without Losing the Judgment

Internal ID: S&F-STR-552

Topic: Innovation, Operations Strategy, Customer Strategy

Industry: Property & Casualty Insurance

Source: Strategy & Finance

Publication date: 5 October 2026

License: Not specified

Decision date and information cutoff: 2 October 2026, 4:00 p.m. ET

Decision scope: Next-wave deployment of AI in claims and underwriting

Case convention: This is an author-created management decision exercise. The role and alternatives are hypothetical; company disclosures are historical. No private deliberations are represented.

The decision

By the end of June 2026, Travelers had launched an AI voice assistant for auto claim intake, an AI claims-intelligence feature for commercial risk managers, employee assistants developed with Anthropic, and an insurance-focused large language model trained on internal documents. In its 2025 annual report, the company had already described dozens of generative-AI tools in production, millions of automated transactions, and more than 20,000 employees regularly using AI tools. AI was no longer a question of whether to experiment.

The harder question was what to scale next-and what authority to give the systems. Public disclosures described applications and management-reported adoption, but did not provide controlled product-level results for cost, errors, claim outcomes, underwriting quality, or customer experience. The company's quarterly financial results could not fill that gap: catastrophe losses and reserve estimates moved reported performance by more than any publicly measured AI effect.

Assignment: You are a hypothetical member of Travelers' senior operating committee. Recommend a policy for the next wave of AI in claims and underwriting. Choose between (1) accelerating deployment across more workflows while retaining accountable human approval for consequential decisions, or (2) expanding workflow by workflow through explicit quality and control gates. Specify which tasks may be assisted, where human decision rights remain, what evidence should trigger broader use, and how Travelers should learn from both successes and failures. Use only information available by 2 October 2026.

What Travelers had put in place

The January 2026 Anthropic partnership provided personalized Claude and Claude Code assistants to nearly 10,000 employees in engineering, data science, analytics and product roles. Travelers described productivity benefits, but disclosed no baseline, measurement protocol or independent evaluation. In February, the company launched an AI Claim Assistant for auto-damage first notice of loss. The assistant could handle claim-intake conversations and move customers into digital follow-on actions; customers could ask for a live specialist. The announcement did not say that the assistant adjudicated coverage, set reserves or paid claims. [S1], [S2]

In May, Travelers introduced Claim Insights within e-CARMA, its risk-management platform, to help commercial customers prioritize and analyze claims. In June, it announced TravelersLLM, a model trained on millions of internal company documents and intended to work alongside leading frontier models. The release described potential use in underwriting analysis, research, model development, institutional knowledge and workflows. Travelers said internal testing showed quality, speed and cost advantages over commercial models across tens of thousands of insurance questions. It did not publish the test protocol, comparison set, quantitative results, error rates or production outcomes. [S3], [S4]

Travelers' annual report offered a wider picture of existing automation. Management said more than half of claims were eligible for straight-through processing, customers adopted that option about two-thirds of the time, and another 15% of claims used advanced digital tools. It also reported dozens of generative-AI tools in production and millions of automated transactions. The disclosure did not isolate the contribution of generative AI from other automation, define every denominator in detail, or link these indicators to loss outcomes. [S5]

The distinction mattered. Summarizing a file, retrieving policy language or helping a caller begin a claim differs from recommending a price, interpreting ambiguous coverage, or deciding whether a severe claim is complete. Human review can catch errors, but it can also become a rushed formality if workload, interface design or performance targets discourage challenge.

The operating backdrop

Travelers reported \$21.867 billion of net written premiums for the first half of 2026, down 1% from the prior-year period; excluding the Canadian businesses sold, management reported growth of 1%. The sale of Canadian Personal Insurance and most Canadian Commercial Insurance operations to Definity was completed on 2 January 2026. The company expected about \$0.7 billion of proceeds to fund additional 2026 share repurchases; this was an announced intention, not a completed repurchase result. [S6], [S7]

Exhibit 1. Selected company results

USD billions except ratios; Q2 ended 30 June. Catastrophe losses are pre-tax and net of reinsurance. H1 sums marked "derived" are arithmetic additions of reported Q1 and Q2 amounts.

Measure	Q1 2026	Q2 2026	H1 2026	H1 2025
Net written premiums	10.338	11.529	21.867	22.058
Consolidated combined ratio	88.6%	83.6%	86.1%	96.3%
Underlying combined ratio	85.3%	84.1%	84.7%	84.7%
Catastrophe losses	0.761	0.518	1.279 (derived)	3.193 (derived)
Net favorable prior-year reserve development	0.413	0.578	0.991 (derived)	-

Source: Travelers Q1/Q2 results releases and Q2 Form 10-Q. [S6], [S8]. H1 2025 figures shown for catastrophe losses are likewise the sum of reported Q1 and Q2 values. The underlying combined ratio is a company-defined non-GAAP measure excluding catastrophe losses and net prior-year development.

Q2's 6.7-point improvement in the consolidated combined ratio from a year earlier was attributed in the company's bridge to 3.6 points from lower catastrophe losses, 2.5 points from more favorable reserve development and 0.6 points from underlying underwriting. Catastrophes were substantially lower in H1 2026 than during H1 2025, which included the California wildfires. Prior-year development reflects updated estimates on older claims; it is not a current-year AI saving. The H1 underlying ratio was flat year over year, while segment results and drivers differed. [S6], [S8]

Investment income was another important factor: after-tax net investment income was \$833 million in Q1 and \$883 million in Q2. These release figures differ from pre-tax income-statement amounts. None of these reported results identifies AI's effect. [S6], [S8]

Exhibit 2. Selected segment context, H1 2026

Segment	Combined ratio	Underlying combined ratio	H1 net written premiums	Qualification
Business Insurance	90.2%	89.0%	\$11.770bn, +2%	Q2 premium growth was +5% excluding the Canadian divestiture.
Personal Insurance	81.2%	77.8%	\$7.794bn, -8%	Reported comparison includes the Canadian sale; this figure is not organic contraction. Lower Q2 catastrophe losses and improved auto underwriting contributed to Q2 results.
Bond & Specialty Insurance	83.0%	89.2%*	-	*Underlying ratio derived from the company's combined-ratio, catastrophe and re serve-development bridge. H1 underlying performance weakened, in part from a higher expense ratio and earned-pricing effects.

Source: Q1/Q2 company releases and Q2 Form 10-Q; H1 items are reported unless marked derived. Segment ratios have distinct portfolio mix and company-defined components. [S6], [S8]

The numbers made a tempting story easy to tell: earnings improved, automation was expanding and underwriting was strong. But the reported combined ratio includes catastrophes and prior-year reserve revisions, and H1 catastrophe experience differed sharply across years. Even the underlying measure does not isolate AI. A rollout decision therefore needed operating

evidence that the public financial statements did not provide.

Evidence that could guide the next wave

Research offered reasons both to test aggressively and to preserve task-level controls. In a field study of 5,172 customer-support agents at one software firm, access to a generative-AI assistant increased issues resolved per hour by an estimated 15.2% on average. Gains varied with worker experience and query type; high-skill agents saw little or adverse quality effects on some measures. The study's task and setting differ from insurance claims and underwriting, so the estimate is not a Travelers forecast. [S9]

A randomized experiment with 444 professionals completing short writing tasks found that ChatGPT reduced task time by 10 minutes, or 37% relative to a 27-minute control mean, and improved evaluator-rated quality by 0.45 standard deviations. The authors cautioned that their bounded tasks did not require extensive context-specific knowledge. Neither study measures Travelers' loss costs, coverage accuracy or claim severity. [S10]

The evidence leaves a practical distinction: AI may be useful for bounded, document-heavy steps even when consequential insurance judgments require more evidence. But a slow, fragmented rollout could delay learning and leave useful assistance underused. A broad launch could speed adoption while making errors, rework and responsibility harder to detect.

Two rollout policies

Accelerate across workflows. Extend the AI ecosystem and TravelersLLM into more claim intake, triage, summarization, underwriting research and employee workflows, retaining human sign-off for coverage, pricing and claim outcomes. Set a common governance standard and review performance at regular intervals. This can spread domain knowledge and reduce duplicated effort. It risks scaling uneven quality or correlated errors before the company can identify them, and it may mistake activity measures for business value.

Expand through task gates. Prioritize bounded, reversible tasks such as document retrieval, summarization, claim intake and work prioritization. Pilot by line, task complexity and employee experience; compare against a concurrent control or staggered rollout where practical. Expand only after quality and control measures meet prespecified thresholds. Keep pricing, coverage interpretation and claim disposition under qualified human authority while evidence develops. This protects learning and accountability but may slow useful adoption and create inconsistent workflows.

Either policy requires decisions about exception handling, employee training, escalation routes, system updates and vendor dependence. A speed measure alone is insufficient. Relevant measures could include cycle time, severity-adjusted rework, claim reopening, escalation, coverage errors, complaint rates, customer and agent experience, employee overrides, underwriting referral quality and cost per completed task. Travelers had not disclosed baseline or target values for these measures in the materials cited here. The committee would need to set them rather than present them as existing company standards.

Discussion questions

1. Which activities in claims and underwriting are suitable for assistance now, and which decisions should remain with qualified employees? What makes the boundary defensible?
2. Which rollout policy do you recommend? What is the strongest operational and strategic argument against your choice?
3. What evidence would demonstrate quality and customer value as well as speed? How should results be separated by line, task and employee experience?
4. How should Travelers balance internal model development with frontier-model partnerships, given what is and is not disclosed about TravelersLLM?
5. What thresholds, escalation rules and reversal conditions should govern a rollout? Which public financial measures cannot answer the AI question?

Submit a two-page recommendation memo. State the initial scope, decision rights, evidence gates and conditions for expansion or pause. Do not estimate Travelers-specific ROI from the academic productivity effects.

Sources and conventions

- S1.** Travelers, [Anthropic partnership announcement](#), 15 January 2026.
- S2.** Travelers, [AI Claim Assistant announcement](#), 18 February 2026.
- S3.** Travelers, [Claim Insights announcement](#), 1 May 2026.
- S4.** Travelers, [TravelersLLM announcement](#), 30 June 2026.
- S5.** Travelers, [2025 Form 10-K](#), filed 12 February 2026, CEO letter, printed pp. 19-21. Adoption figures and expectations are management-reported.
- S6.** Travelers, [Q2 2026 results release](#), 17 July 2026, quarter/six months ended 30 June; consolidated and segment results and ratio bridges.
- S7.** Travelers, [Canadian operations sale completion](#), 2 January 2026.
- S8.** Travelers, [Q2 2026 Form 10-Q](#), filed 17 July 2026; MD&A printed pp. 37-50, catastrophe table printed pp. 38-40, and segment disclosures.
- S9.** Erik Brynjolfsson, Danielle Li and Lindsey R. Raymond, “Generative AI at Work,” *Quarterly Journal of Economics* 140(2) (2025), 889-942, [doi:10.1093/qje/qjae044](#). The study reports 5,172 agents and 3,006,395 chats; see pp. 902-903 for sample, pp. 904-927 for design/results and pp. 934-936 for limitations.
- S10.** Shakked Noy and Whitney Zhang, “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence,” *Science* 381(6654) (2023), 187-192, [doi:10.1126/science.adh2586](#). Reported results and design use the accessible author preprint, pp. 2-5; limitations, pp. 11-12. Publication metadata was cross-checked; the preprint is not represented as the final typeset article.

All dollar amounts are U.S. dollars. Company-defined underlying combined ratios are non-GAAP. H1 quarter sums are identified as derived where used. Q3 results and the scheduled 16 October release/call fall after the cutoff and are not part of the case information set.