CASE STUDY · INSURANCE · COST DISCIPLINE & INTELLIGENT ROUTING

A national insurer cut agent spend 71% without losing a point of claims accuracy

Every claims directive was going to the most expensive model available. Stakes-based routing moved 82% of them to a cheaper sufficient tier and left the hard ones where they belonged.

Partner
National property & casualty insurer · 4.2m policies

Published
2026-08-26

Orcher components
Cost Governance, Observability, Logic Scrubber

Summary

A national P&C; insurer was spending frontier-model rates on triage tasks that a mid-tier model handled identically. Orcher's Cost Governance routed by the stakes of the action rather than the habit of the developer, escalated only where verification demanded it, and halted two runaway agent loops that had been quietly burning budget for weeks.

71%

reduction in monthly agent spend

measured over two full billing cycles

82%

of directives served by a cheaper sufficient tier

±0.4pp

change in claims decision accuracy

within sampling error against the human-reviewed baseline

19 days

of runaway loop spend eliminated on detection

The situation

— Eleven claims workflows all defaulted to a single frontier model, because that was the model the first prototype used and nothing forced the question again.

— Monthly agent spend had grown 340% in two quarters while claim volume grew 11%.

— Two recursive agent loops in subrogation had run for nineteen days before anyone noticed; neither produced a committed outcome.

What Orcher enforced

— Directive classes were scored by stakes — reserve impact, regulatory exposure and reversibility — and each class was given a routing floor and ceiling instead of a fixed model.

— Cost Governance enforced budget ceilings per role, per department and per directive class, with automatic escalation only where the Logic Scrubber's verification demanded a stronger model.

— Loop detection halted recursive execution at a hard concurrency and depth limit, and recorded the halt in the ledger as an outcome like any other.

— Observability gave finance a single cross-provider view of spend by role, workflow and model, replacing three provider dashboards that never reconciled.

The data

Directive class	Model before	Model after	Cost per 1k directives	Accuracy delta
First-notice triage	Frontier	Mid-tier	$412 → $58	–0.2pp
Coverage lookup	Frontier	Small + retrieval	$381 → $19	+0.1pp
Damage narrative	Frontier	Mid-tier	$466 → $121	–0.3pp
Reserve recommendation	Frontier	Frontier (held)	$503 → $503	0.0pp
Subrogation referral	Frontier	Mid-tier + escalation	$489 → $147	+0.4pp

Blended cost per thousand directives across two billing cycles. Reserve recommendations were deliberately held at the frontier tier — high stakes, low volume.

Outcomes

— Finance now forecasts agent spend per workflow with a variance under 8%, where the previous quarter's forecast missed by 3.4×.

— The saved budget funded two additional workflows rather than being returned — the programme expanded on the same envelope.

— Escalation is now a recorded, reviewable event, so 'why did this cost more?' has an answer in the ledger.

— The insurer adopted the routing floors as a written standard for all new agent development.

> “The uncomfortable finding was that the expensive model was not making us more accurate on the easy work. It was just making us slower to notice the cost.”
>
> — VP of Claims Technology, national P&C; insurer (partner declined attribution by name)

Read further

— Cost governance and enforced routing — dippai.com/research/cost-governance-enforced-routing

— Vol. I — Enterprise Superintelligence — dippai.com/research/vol-1-enterprise-superintelligence

— Full case study — dippai.com/case-studies/national-insurer-claims-routing-cost

Disclosure. Design-partner engagement, anonymised at the partner's request. Figures are measured by Orcher's own observability and ledger instrumentation over the stated period and have not been independently audited. Market figures carry their own source line.