



GOVERNANCE INCIDENT BRIEFING

Briefing

Operations & Governance

THE METRIC THAT MADE KLARNA'S AI LOOK LIKE A SUCCESS

THE SYSTEM WASN'T FAILING ON AVERAGE. IT WAS FAILING IN THE TAIL.

In February 2024, Klarna launched an OpenAI-powered customer service assistant, and by every headline number, it looked less like a pilot than a triumph.

2.3M

Conversations handled in the first month

across 35 languages

700

Full-time employees' worth of work, by Klarna's own estimate

hiring paused, headcount fell 22%

<2 min

Average handle time, down from eleven minutes

\$40M projected profit improvement



WHAT THE AVERAGES WERE BUILT NOT TO CATCH

What actually went wrong

By May 2025, CEO Sebastian Siemiatkowski was telling Bloomberg something markedly different. Klarna was rehiring human agents, specifically for disputes, hardship cases, and its higher-value customers. "Investing in the quality of human support is the way of the future for us," he said. "I just think it's so critical that you are clear to your customer that there will always be a human if you want."

What actually went wrong is more specific than "AI wasn't good enough." Klarna's original success metrics, handle time, cost per transaction, aggregate satisfaction, measured speed and average experience. They didn't measure whether genuinely complex problems, disputes, fraud claims, anything requiring judgement rather than a lookup, were actually being resolved. The AI matched human performance on routine queries. On the harder cases, quality quietly degraded in ways none of the headline metrics were built to catch.

The system wasn't failing on average. It was failing in the tail, and the tail is exactly where a customer's trust in a brand actually gets decided.

A scope correction, not a retreat

It's worth being precise about what this case does and doesn't show. Klarna didn't ban AI, and it would be inaccurate to describe this as a clean reversal. The AI assistant remained the front line for the high-volume, routine tier. What changed was a deliberate decision to reintroduce people for the tier where AI parity hadn't held. That distinction, a scope correction rather than a retreat, is the actually useful lesson here, not a simple story of AI failing.

Any business considering an AI chatbot, booking system, or automated phone line faces the identical decision at a smaller scale, and rarely makes it deliberately. A routine booking confirmation is a fine job for AI. A customer disputing a charge, cancelling in anger, or dealing with something genuinely wrong isn't, not because AI can't process words, but because those moments are exactly where a business's reputation actually gets made or lost.

(01) Which interactions AI owns outright

Routine queries, order status, standard payments: simple enough for AI to own completely.

(02) Which interactions default to a person

Disputes, hardship cases, anything requiring genuine judgement.

(03) What "resolved" actually means

Defined for each category, not just how quickly a conversation ended.

(04) A metric for the hardest cases specifically

An average can hide exactly where the damage is happening.

WHERE THIS FITS

This is, in a fairly direct way, the whole case for building an AI integration Blueprint before switching anything on. Propel Growth's Solus Requirements Assessment exists to force this decision early: which interactions genuinely suit AI, which need a person, and what success actually means for each, so a business finds out during a diagnostic, not eighteen months and a public correction later.

BOOK A CONFIDENTIAL CONSULTATION

Map out where AI genuinely fits in your business, and where it doesn't. No obligation, no sales pitch, just a clear read.

↗ [Book a call](#) ↗



propelgrowth.com.au · +61 452 177 673 (04521 PROPEL) ·
enquiries@propelgrowth.com.au
Level 1, 89 Pirie Street Adelaide, SA 5000 Australia

Bloomberg via MarketBeat, "Klarna Is Hiring Customer Service Agents After AI Couldn't Cut It on Calls" (9 May 2025); Perspective AI, "Klarna AI Customer Service: Replacing 700 Agents, A 2026 Case Study"; Bigeye, "Klarna's AI Customer Service Deployment | AI Autopsy 002."