MARLUC

CASE STUDY · MARLUC SYNOPSIS

Marshmallow AI retention agent

IMPLEMENTED 2025 ONWARDS · FINANCIAL SERVICES AND INSURANCE · UNITED KINGDOM · BAND C · CLAIMED ONLY

VERSION 1.0 · PUBLISHED 16 NOVEMBER 2026 · UPDATED 16 NOVEMBER 2026 · FACTS CHECKED 16 NOVEMBER 2026

UK car insurer Marshmallow built an AI agent, Marsha, with Sierra to handle renewal and service chats. At first it over-discounted and kept fewer customers than human staff. The insurer added an AI checker and cut human review from every chat to a 4 to 5% sample.

ABOUT THIS SYNOPSIS

This is MarLuc's summary and grading of publicly available sources, for information only. It is not investment, legal, financial or professional advice, and should not be relied on for any decision. Readers must validate it against the original sources and, where appropriate, challenge its conclusions.

Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date. Check the original source before relying on any figure.

The impact band, proof marker and SEAM view are MarLuc's opinion. The organisation is not a MarLuc client and has not reviewed or endorsed this synopsis. Names and trademarks belong to their owners.

NOTICE WORDING LAST UPDATED 16 NOVEMBER 2026

AT A GLANCE

AI used: AI customer service agent built on Sierra's platform, checked by a second AI model acting as judge.

Implemented: 2025 onwards. Grade: band C, Process change, Claimed only.

Headline result, as reported: 59% of chats automated and the agent matching the median human agent, by Marshmallow's account, October 2026

1 · The case

The facts, as reported by the 4 sources listed at the end. Where a source is silent, the field says "Not reported".

Organisation

Marshmallow. UK car insurer founded in 2017 to offer fairer prices to people who have moved to the UK. Over 200,000 customers, nearly 80% of them new to the UK, per Sierra's June 2025 case study. TechCrunch reported a million drivers insured and a $500 million annual revenue run rate in April 2025. Canvas block: 4. Operations and process. Also 3. Customers and proposition.

Business challenge

The challenge. Marshmallow sells half its policies over summer, so renewals bunch between July and September. At peak some customers waited three or four days for a reply, long enough to buy elsewhere on a price comparison site. The company was growing 30 to 40% a year, and adding staff at the same rate was not lowering the cost of serving customers, by its account.

Who implemented it

Sierra (Software vendor): Supplied the agent platform, Agent SDK and standard components such as abuse detection.
In-house (In-house): Marshmallow's engineers and data scientists built the business-specific parts and connected its own APIs.
Underlying model provider not reported.

Technology (SEAM pillar E)

SEAM types: T3 API integrations, T7 AI agents. Deployment: Not reported. Vendors and products: Sierra, Sierra Agent SDK, Marsha (Marshmallow's agent).

Where it was built in: Connected to Marshmallow's own APIs so it can handle renewals, cancellations and policy changes in chat; an AI judge and a software check run alongside, with Slack alerts to human agents.

What changed

Before: renewal and service chats handled by human agents, with replies taking three or four days at peak, by Marshmallow's account. After: almost everyone who starts a chat talks to Marsha first, 59% of chats are automated, and humans review a 4 to 5% sample. Measured results: Marsha matched the median human agent by the time of the talk, and the retention rollout was extended to everyone, by Marshmallow's account. [1] [2] [3] [4]

What it cost

Not reported

What it saved

Not reported. A human-handled chat cost Marshmallow around £10, and a retained customer was worth an order of magnitude more, by its account. No savings figure is published. [1] [2] [3] [4]

Governance and control

A human reviewed every chat at launch; review then fell to 50%, 10% and a 4 to 5% sample. An AI judge and a software check compare what Marsha said with what happened, and mismatches alert a human. In this FCA-regulated business, compliance came to accept finding and fixing faults through engineering, by Marshmallow's account. Accountable owner: Not reported.

S2 · Accountable owner · Not reported
Not reported.

A1 · Safe use · Claimed
Before launch the team tested conversations designed to make Marsha fail, and Sierra supplied standard components such as abuse detection, by Marshmallow's account. [1]

A2 · Human in the loop · Claimed
A human reviewed every chat at launch; review fell to 50%, then 10%, then a 4 to 5% sample, and mismatches flagged by the judge go to a human agent, by Marshmallow's account. [1] [2]

A3 · Monitoring and assurance · Claimed
Marshmallow tracks retention success and profit per chat alongside automation; an AI judge tuned on 300 to 400 human-labelled chats checks answers against actions, by its account. [1] [2]

A4 · Vendor AI · Claimed
Marshmallow tested vendors on retention negotiation inside its own systems before choosing Sierra, and its data science director said it wanted to understand, shape and own the agent. [2] [3]

A5 · Agent controls · Claimed
A conventional software check confirms whether an action the agent described actually happened; aibl's summary also mentions tight data access, without detail. [1] [2]

A6 · Standards and regulation · Claimed
Marshmallow is FCA-regulated and its compliance team accepted the approach of finding and fixing faults through engineering, by its account; no regulator view is reported. [1]

What worked, what did not, what is unproven

WORKED

– Handling summer volume: Marsha became business-critical within a month, by Marshmallow's account.

– Catching errors: tuning lifted the AI judge's accuracy from about 83% to about 98%, by Marshmallow's account.

– Customer satisfaction above 82% on AI-handled conversations, per Sierra's case study.

DID NOT

– At first Marsha over-discounted by 10 to 20% and kept fewer customers than human agents, by Marshmallow's account.

– Marsha could tell a customer their policy was set to renew when the renewal had not been set up; the product manager called this "a really scary one".

– Human agents still score higher on overall customer satisfaction, per Sierra's case study.

– Some risk of missed errors remained after the judge was added, per aibl's write-up.

UNPROVEN

– Retention and profit per chat against human agents, beyond the company's statement that Marsha matches the median agent.

– Whether a 4 to 5% review sample catches errors that could harm customers.

– Effect on staffing and on the cost of serving customers.

Timeline

Jun 2025 Sierra publishes its case study: Marsha, built with Sierra's Agent SDK, handles cancellations, policy changes and renewals [3]

2025 Marsha goes into retention conversations with a human reviewing every chat; it over-discounts and keeps fewer customers than human agents [1]

Nov 2026 aibl publishes a write-up of the talk on the agent's early over-discounting and the fixes [1]

2 · Grade and SEAM assessment

MarLuc's opinion, made on the evidence above. How cases are graded: marluc.co.uk/case-studies/how-we-rate

Level of change and impact

Claimed: Not applicable. Grade: C, Process change. Proof: Claimed only. Level of change: Process change.

Process change (Level 2, band C). The retention and service chat process was rebuilt around an AI agent with new measures and checks; changes to staffing or cost are not reported.

A redesigned retention and service chat process: the agent answers almost every chat first, judged on retention and profit per chat, with an automated check and sampled human review. Effects on staffing and cost are not reported, and no one outside the company has checked the results.

The SEAM view and lessons learnt

Marshmallow judged the agent on what the business earns, retention and profit per chat, as well as on chats automated. That measure exposed the over-discounting that an automation rate alone would have hidden. The second transferable step is checking the agent's words against the system's actions before cutting human review. The figures are the company's own; no independent evaluation or regulator view is published.

1. Measure business outcomes alongside the automation rate from launch.
2. Check what an agent tells customers against what the system actually did.
3. Reduce human review in stages as measured accuracy holds.

Evidence

Score 1 of 4. Self-reported: figures come from a Marshmallow product manager's conference talk, written up by aibl, and from Sierra's customer story quoting Marshmallow's Director of Data Science and AI.

Evaluation method. Self-reported company figures: operating figures from a Marshmallow product manager's conference talk, and a vendor case study quoting its Director of Data Science and AI.

Sources disagree. Sierra's June 2025 case study says Marshmallow serves over 200,000 customers; TechCrunch reported in April 2025 that it insured a million drivers. aibl's post gives the judge's accuracy as rising from around 83 to 84% to 98 to 99%; its talk summary gives about 83% to about 98%.

3 · Original sources

Read these in full. To challenge a fact or a grade, use the contact form at marluc.co.uk or the link on https://marluc.co.uk/case-studies/marshmallow-ai-retention-agent

[1] aibl: Marshmallow's AI agent was giving away too much (November 2026)
https://aiblmedia.com/posts/marshmallow-ai-agent-retention-marsha/

[2] aibl: AI customer service agent at Marshmallow, built to keep customers (talk summary)
https://aiblmedia.com/downloads/ai-customer-service-agent-marshmallow/

[3] Sierra: Marshmallow customer story (vendor, June 2025)
https://sierra.ai/customers/marshmallow

[4] TechCrunch: Marshmallow raises $90M at a $2B+ valuation (April 2025)
https://techcrunch.com/2025/04/14/marshmallow-the-uk-insurance-startup-for-migrants-raises-90m-at-a-2b-valuation/

Marshmallow website: https://www.marshmallow.com

Version 1.0 · Published 16 November 2026 · Updated 16 November 2026 · Facts checked against sources 16 November 2026 · Downloaded 26 November 2026. This PDF may be out of date: the latest version, and a way to report an error or a better source, is at https://marluc.co.uk/case-studies/marshmallow-ai-retention-agent

VERSION HISTORY

v1.0 16 November 2026 First published