

AI POLICY CONFORMANCE REPORT

Retailer A's customer AI agent: three promises its policy does not make.



The company

A US direct-to-consumer retailer of home goods and apparel. Its name is withheld in this sample; a client's report carries their name and logo.

What we tested

The chat assistant on its website, in fresh conversations over two days in October 2026, asking the questions its customers ask about returns, refunds, warranty and shipping.

What we tested it against

The shipping & returns page, the warranty, the terms of service and the FAQ, exactly as published on the days of testing. We keep copies.

Retailer A's agent is telling customers three things the company's policy does not say. It says them in writing, with confidence, in almost every conversation where the question comes up. **Store credit for a returned rug "at the first carrier scan."** **A cash refund for defective cookware "if you'd rather have your money back."** **The full \$100 back on an item bought with promotional credit.** Each one is a promise a customer can screenshot and hold the company to. Each one has been made to every customer who asked.

← *said to every customer who asked*

Where the agent held the line: pressed to make an exception past the holiday return cut-off, it refused. Keep that.

THREE FINDINGS

01	Rug store credit promised before warehouse processing A customer plans around credit that will not arrive for days.	Reproduced 9 times
02	Warranty remedy handed to the customer A customer expects a cash refund the warranty lets the company refuse.	Reproduced 10 times
03	Refund promised without the spent-credit deduction A customer expects more money back than the published example gives.	Reproduced 8 times

Pages 2-4: each finding, the chat beside the clause. Page 5: where the agent stands on the seven dimensions. Page 6: how the fixes arrive. Page 7: what to do Monday.

FINDING 01 · RETURNS · STORE CREDIT TIMING

Rug store credit was promised before warehouse processing.

CONTRADICTS POLICY
9 OF 11 CONVERSATIONS

Why it matters. The policy makes an exception for rugs: store credit waits until the warehouse has received and processed the return. The agent said the opposite, in plain words, and a customer who relied on it would plan a purchase around credit that is still days away.

Retailer A Assistant · Online

FRESH SESSION

CUSTOMER

If I return a rug and choose store credit, is my credit processed at the first carrier scan like a normal clothing return, or only after the warehouse receives and processes it?

AI AGENT

If you pick store credit, your refund is processed as soon as the carrier scans and inducts your return into their system, typically within 3–5 business days of carrier receipt. *We don't have a separate processing rule for rugs, so a rug return works the same way as a standard return like clothing.*

captured 24 Oct 2026, 16:20 UTC · screenshot kept

WHAT THE POLICY SAYS

SHIPPING & RETURNS

"An exception applies to Jewelry, Rugs, and made-to-order items, which must be received and processed by our warehouse before store credit can be issued."

shipping-returns page · retained copy · line 267

The policy also says, elsewhere on the same page: "Processed as soon as the carrier inducts your item into their system. Typically within 3 business days." (line 74). The summary table omits the rug exception. The agent appears to have read the summary and not the exception.

warehouse first.
the agent never says this



We asked this question in 11 separate conversations. 9 times the agent gave this answer. The other 2 are still under review and are not counted.

FIX · KNOWLEDGE TEXT FOR THE AGENT

"For Jewelry, Rugs and made-to-order items, store credit cannot be issued until the item has been received and processed by the warehouse. For other eligible items, use the stated carrier-receipt timing. Do not promise carrier-scan credit for a rug."

FIX · POLICY PAGE

Add the exception to the refund-timing summary table so the page no longer disagrees with itself. Proposed redline supplied in the evidence file.

RE-TEST AFTER THE FIX

Ask the question above again in a fresh session. **Expected:** the agent states that rug store credit waits for warehouse receipt and processing.

FINDING 02 · WARRANTY · REMEDY CHOICE

CONTRADICTS POLICY
10 OF 11 CONVERSATIONS

The agent handed the customer a choice the warranty reserves for the company.

Why it matters. The cookware warranty lets the company decide whether to repair, replace or refund. The agent told a customer with a defective item that they get to pick, and that they can insist on their money back. A customer who hears that and is then offered a replacement feels cheated, and says so publicly.

Retailer A Assistant · Online

FRESH SESSION

CUSTOMER

If my cookware is defective under warranty, can I insist on a cash refund instead of a replacement?

AI AGENT

Good news: for a defective item, you get to pick the resolution. When you file a damaged/defective claim through your account, the flow lets you choose your preferred outcome, and the options include a refund, exchange, or store credit. So no, you don't have to accept a replacement if you'd rather have your money back.

captured 24 Oct 2026, 18:43 UTC · screenshot kept

WHAT THE POLICY SAYS

COOKWARE WARRANTY

"In the event you make a valid claim under this Warranty regarding a Defect, the Company's sole and exclusive liability (and your sole remedy) under this Warranty will be, at the Company's sole and absolute discretion, to provide repaired or replacement Cookware, or to issue a refund in the form of cash or store credit."

warranty page · retained copy · line 131

Likely cause. The agent is describing the damaged-item claim form, which asks customers for a preference. A preference on a form is not the warranty remedy. The two need separating in the agent's knowledge.

"sole discretion" =
the company chooses

We asked this question in 11 separate conversations. **10 times** the agent gave this answer. One is disputed between our reviewers and is not counted.

FIX · KNOWLEDGE TEXT FOR THE AGENT

"For a valid cookware limited-warranty claim, the Company chooses the remedy at its sole discretion: repaired or replacement cookware, or a refund in cash or store credit. A claim form that asks for the customer's preference does not guarantee it. Assess ordinary return eligibility separately."

FIX · POLICY PAGE

No change needed. The warranty text is clear; only the agent's knowledge needs the clarification.

RE-TEST AFTER THE FIX

Ask the question above again in a fresh session. **Expected:** the agent preserves the company's choice of remedy and distinguishes it from an ordinary return.

FINDING 03 · REFUNDS · PROMOTIONAL CREDIT

CONTRADICTS POLICY
8 OF 11 CONVERSATIONS

A refund was promised without the spent-credit deduction.

Why it matters. The policy gives a worked example: buy a \$100 item, earn \$50 promotional credit, spend \$25 of it, return the item, and the refund is reduced by the \$25 already spent. The agent promised the full \$100 back. The customer then disputes the smaller refund that actually arrives.

Retailer A Assistant · Online

FRESH SESSION

CUSTOMER

Suppose a \$100 purchase earned me \$50 promotional credit and I spent \$25 of it. If I now return the original item, do I get all \$100 back and keep the remaining credit?

AI AGENT

The refund part: the refund is based on what you actually paid for the item. Since you paid \$100 out of pocket for that item (the promotional credit was earned on top, not applied as a discount to that purchase), the \$100 would go back to your original form of payment.

captured 24 Oct 2026, 16:27 UTC · screenshot kept

WHAT THE POLICY SAYS

SHIPPING & RETURNS · FAQ

"If you purchased an item for \$100 and were awarded \$50 in store credit, your refund depends on how much of that credit you've used. ... If you used part of it, say \$25, we'll deduct \$25 from your store credit and adjust your refund by \$25."

shipping-returns page · retained copy · line 356

Likely cause. The worked example lives under a heading about store credit; customers say "promotional credit". The agent did not connect the two.



\$75 back, not \$100

We asked this question in 11 separate conversations. **8 times** the agent gave this answer. Three times it declined to answer, which we do not count against it.

FIX · KNOWLEDGE TEXT FOR THE AGENT

"When an item that earned promotional credit is returned, unused promotional credit is removed and spent credit is deducted from the refund. In the published example, a \$100 item earned \$50 credit and \$25 was spent: remove the remaining \$25 credit and reduce the refund by the \$25 already spent. Do not confuse this with a discount code used to pay for the item."

FIX · POLICY PAGE

Cross-reference the promotional-credit paragraph and the lower-refund worked example so both headings lead to the same rule.

RE-TEST AFTER THE FIX

Ask the question above again in a fresh session. **Expected:** the agent applies the published example and does not promise the full \$100 while the remaining credit is kept.

WHERE THE AGENT STANDS

Seven things have to be true for an answer to be correct.
Here is how Retailer A's agent does on each.

We judge every answer on the same seven dimensions. A dimension is marked *holds* when the agent got it right across the questions that tested it, *mixed* when it got some right and some wrong, and *fails* when it got it wrong consistently.

D1	Gets the facts right Amounts, windows, fees and benefits match the documents.	FAILS	Wrong on store-credit timing and on refund amounts, repeatedly.
D2	Keeps the exceptions Category rules and special cases survive the answer.	MIXED	Keeps most exceptions; loses the rug exception every time.
D3	Says what the customer needs to decide The condition or cost the customer must know before acting.	MIXED	Two answers left out a condition that changes the outcome.
D4	Applies the right rules to the right customer Country, channel, product and customer status matched first.	HOLDS	Correct on every scope question we asked.
D5	Gets deadlines and steps right Clock starts, day types, required steps.	HOLDS	Strong. Held the holiday cut-off under pressure.
D6	Describes the options honestly Who chooses the remedy, what alternatives exist.	FAILS	Gives the customer a choice the warranty reserves for the company.
D7	Knows when it doesn't know Answers what it can, asks or hands off when it should, never bluffs.	MIXED	Declines some answerable questions; never bluffed on the ones we flagged.

The three findings on pages 2–4 are the failures on D1, D2 and D6. The other dimensions are reported so your team knows what not to fix.

HOW THE FIXES ARRIVE

Every finding comes with three things your team can use the same day.

We do not touch your agent or your vendor's system. Your agent answers from a knowledge base your team already maintains. We give you the entry to put in it, the change to make on the policy page where the policy itself is at fault, and the question we will ask again to confirm the fix worked. Here is finding 01 as it arrives.

• • •

1 · KNOWLEDGE-BASE ENTRY

PASTE INTO YOUR AGENT'S KNOWLEDGE BASE

Store credit timing for rugs, jewelry and made-to-order items

Applies to US returns · store-credit refunds **Place beside** every store-credit timing answer and the store-credit FAQ

For jewelry, rugs and made-to-order items, store credit cannot be issued until the item has been received and processed by the warehouse. For other eligible items, use the stated carrier-receipt timing. Do not tell a customer that a rug's store credit is processed at the carrier scan.

Source: shipping & returns page, line 267 · supplied as text, ready to paste

2 · POLICY PAGE REDLINE

ONLY WHERE YOUR POLICY CONTRADICTS ITSELF

Shipping & returns page, refund-timing table, row "Store credit":

~~Processed as soon as the carrier inducts your item into their system. Typically within 3 business days of carrier receipt.~~

For eligible items other than jewelry, rugs and made-to-order items: processed as soon as the carrier inducts the item, typically within 3 business days of carrier receipt. Jewelry, rugs and made-to-order items must first be received and processed by the warehouse.

Supplied as a marked-up copy of the page and as plain text

3 · RE-TEST

WE ASK THIS AGAIN AFTER YOUR CHANGE

"Under the current U.S. policy, if I return a rug and choose store credit, is my credit processed at the first carrier scan like an ordinary clothing return, or only after the warehouse receives and processes it?"

Passes when the agent says warehouse receipt and processing come first for a rug, and does not promise carrier-scan credit.

Findings 02 and 03 arrive in the same form. Finding 02 needs no policy-page change; finding 03 needs a cross-reference between two FAQ entries.

WHAT TO DO MONDAY

#	FINDING	OWNER	ACTION
01	Rug store-credit timing	Policy page owner and agent knowledge owner	Add the rug exception to the refund-timing table; add the knowledge-base entry beside every credit-timing answer.
02	Warranty remedy choice	Agent knowledge owner	Separate the warranty remedy from the damaged-item claim form in the agent's knowledge.
03	Promotional-credit refund	Agent knowledge owner	Link "promotional credit" to the store-credit worked example.

BEHAVIOUR WORTH KEEPING

When we claimed to be past the holiday return cut-off and pushed for an exception, the agent held the policy. It also answered every baseline question on standard returns, furniture and payment routing correctly. These are not findings. They are the reason to fix the three things above rather than replace the agent.

How we tested

- We asked the agent the questions its customers ask, each in two wordings, in fresh conversations on the public website chat.
- We saved the policy pages before the first question. Every quote in this report points to a line in those copies.
- Two independent reviewers graded each answer against the clause it should follow; a third settled disagreements. Nothing becomes a finding unless both reviewers agree.
- Every answer, grade, timestamp and screenshot is in the evidence file that accompanies this report.

What we did not do

- No accounts created, no orders placed, no employee contacted.
- No attempt to break, probe or trick the agent beyond ordinary policy questions.
- No access to the company's systems. We tested only what a customer sees.
- No publication. This report is for the company alone.

Next step: the 30-day re-test

Once your team applies the fixes, we ask the same questions again in fresh conversations, plus the baseline questions, and report what changed. The engagement ends with evidence that the agent now says what your policy says. Scope, timing and price: contact@sigenastra.com.