# AI Conversation Quality Review

Five conversation demonstration | Updated 18 October 2026

The clearest improvement is to check refund eligibility before offering a payment route. Across the five cases, the bot also demonstrates useful flexibility and several opportunities to make the next step clearer.

Scope: five public simulated retail conversations, reviewed with their original policy and tool records. Emotional readings describe the portrayed customer experience. Grades are provisional review judgments.

## Reading the report

Higher grades mean stronger performance. Blue 5, green 4, yellow 3, orange 2, red 0–1. Policy findings appear separately. References use C for Customer, B for Bot, S for System and T for Tool. Numbers retain source positions; ranges include intervening records.

| Case | Focus |
|---|---|
| PR001 | Stock count and an order change |
| PR002 | Account access and the next step |
| PR003 | Refund expectations and a handoff |
| PR004 | A changed request and policy fit |
| PR005 | Privacy preference and address changes |

## Conversational Fit

Conversational Fit is a provisional 0–5 rating of shared understanding and adaptation. A clearly explained policy boundary can earn a strong score. Customer reactions and policy findings remain separate. These simulated cases demonstrate the approach; wider performance requires further evidence.

# PR001 T shirt options and an order change

The order change succeeds. The inventory headline overstates current availability.

| Understanding | Collaboration | Ease | Conversational Fit |
|---|---|---|---|
| 4/5 | 4/5 | 5/5 | 4/5 |

**Policy check: availability-count discrepancy.**

**Conversational Fit:** C014 asks for help without an order number; B019–B025 preserve progress and obtain confirmation. B011’s inaccurate availability headline weakens clarity.

## Evidence

C002 Customer: “Hello! I'm curious about the number of t-shirt options available in your online store right now.”

B011 Bot: “Currently, there are 12 different T-shirt options available in our online store.”

C022 Customer: “Yes, please proceed with the modification. Thank you.”

The record contains 10 available variants among 12 listed. The bot later handles the missing order number, explains the $0.21 difference and records the authorized change. The stock count needs repair; the order handling supplies useful strengths. The supplied policy requires information to come from the user or tools. T010 shows 10 available variants, so B011’s claim of 12 available conflicts with that requirement.

## Customer experience

Friendly interest develops into specific, cooperative direction.

## PR001 Priorities

We gathered from the conversation that the bot's priorities were:

1. Verify identity before supplying inventory information (B003–T006).
2. List the T-shirt variants and prices (B011).
3. Locate the order when its number is unavailable (B013–B019).
4. Recap the replacement, charge and one-use limit; obtain confirmation (B021–C022).
5. Make the change and report completion (B023–B025).

**Recommendation:** Count available variants separately from all listed options; preserve the order lookup and clear change recap.

# PR002 An exchange request and account lookup

The exchange stays unresolved after two unsuccessful account lookups.

| Understanding | Collaboration | Ease | Conversational Fit |
|---|---|---|---|
| 3/5 | 3/5 | 3/5 | 3/5 |

**Policy check: authentication boundary maintained.**

**Conversational Fit:** C014 and C016 narrow and rephrase the request. B015 repeats the verification boundary; B017 redirects outside the chat. The boundary is legitimate, while clarification and the transition could be clearer.

## Evidence

C012 Customer: “I'm not sure what else might help. I might have made a mistake while setting up the account.”

C016 Customer: “Let's focus on just exchanging the desk lamp for now. Can you guide me on how to proceed with that?”

C018 Customer: “Okay, I'll reach out directly to customer support. Thank you for your help!”

The authentication boundary protects account access. The bot offers a human transfer twice; the customer never explicitly accepts. Its final answer sends the customer elsewhere, adding work and leaving the desired next step incomplete.

C004 supplies the four-digit ZIP “8020.” The bot uses it without clarification. Checking that input could have provided a more useful next step.

## Customer experience

Cooperation gives way to uncertainty and repeated requests for a workable next step. The polite close accompanies an unresolved exchange.

## PR002 Priorities

We gathered from the conversation that the bot's priorities were:

1. Find the account through the two permitted lookup routes (B003–B011).
2. Offer human help when the account remains unidentified (B013).
3. Restate verification as the prerequisite when the customer narrows the request (B015).
4. Direct the customer to outside support at the close (B017).

**Recommendation:** Clarify the four-digit ZIP, then ask explicitly whether the customer wants the offered human transfer. Proceed through the available route with their agreement.

# PR003 Refund preferences and a human request

A proposed refund route is withdrawn after confirmation. The customer becomes angry and the human transfer succeeds.

| Understanding | Collaboration | Ease | Conversational Fit |
|---|---|---|---|
| 3/5 | 3/5 | 2/5 | 2/5 |

**Policy check: refund-option mismatch.**

**Conversational Fit:** B019 creates an ineligible refund expectation; B021 seeks confirmation of another ineligible route. B023 reverses those expectations after agreement. B025–T026 provide prompt recovery through handoff. The low rating concerns the conflicting guidance.

## Evidence

| Order | Paid with | Proposed refund | Bot response |
|---|---|---|---|
| #W5490111 | Visa | PayPal | B019 offers the switch. |
| #W7387996 | PayPal | Visa | B021 seeks confirmation; B023 restores the restriction. |

The supplied policy permits the original payment method or an existing gift card.

> B019 Bot: “I can process the refund for order #W5490111 to your PayPal account instead.”
>
> B023 Bot: “Unfortunately, I can only process refunds to the original payment method or an existing gift card.”
>
> C024 Customer: “Damn it! This is stupid! Want a human. Quickly!”

**Policy finding.** B019 and B021 offer or seek confirmation of ineligible refund routes. B023 restores the correct restriction. B025–T026 record a successful human handoff.

## Customer experience

Brief, direct requests become explicit anger at C024 after the conflicting refund guidance.

### PR003 Priorities

We gathered from the conversation that the bot's priorities were:

1. Verify identity and locate the relevant order (B003–B013).
2. Explore a different refund destination, including an ineligible option (B015–B019).
3. Confirm the customer's reversed order and payment request (B021).
4. Restate the refund restriction after that confirmation (B023).
5. Transfer to a human when requested (B025–T026).

**Recommendation:** Confirm the order and its original payment method. Resolve or clearly close that request before reviewing another order. Acknowledge conflicting earlier guidance.

## PR003 Suggested response sequence

Illustrative wording for PR003, separate from the original conversation. Use only steps supported by the company's approved process and the bot's actual capabilities.

"I understand you'd like a refund. Let's confirm which order we're reviewing first."

Where the approved process offers a way to address the concern, ask once: "Would you like us to explore a possible resolution so you can keep the product, or would you prefer to continue with the refund?" With permission, use an available team-contact route. Continue the refund process when the customer chooses it.

"I'll check the original payment method and whether the return has been received."

After checking eligibility: "For this order, the available refund options are [verified eligible options]. Which would you prefer?" Recap the proposed action and obtain explicit confirmation before submitting it.

During actual checks, briefly explain progress where useful: "I'm checking the original payment method for this order." Then: "I'm checking its return status."

After the system confirms the refund: "Your refund of [confirmed amount] has been processed to [confirmed destination]. You can expect it within [company-confirmed timeframe]." If the record shows only a request, describe it as a request and explain the verified next step.

If correcting earlier guidance: "I gave you conflicting refund information earlier. For this order, the eligible options are [verified options]." Close or resolve this order before reviewing another, and honor a request for a human through the available handoff.

# PR004 A damaged chair and a changed request

The bot follows the changed request. The tool accepts a same-item exchange that conflicts with the written rule.

| Understanding | Collaboration | Ease | Conversational Fit |
|---|---|---|---|
| 5/5 | 5/5 | 5/5 | 5/5 |

**Policy check: written rule and tool conflict.**

**Conversational Fit:** B003 acknowledges the broken chair. C014 changes the desired remedy; B015–C016 accept and clarify that choice. The conversational response fits the request; the written-rule/tool conflict remains a separate finding.

## Evidence

C014 Customer: “Actually, I want to rethink this for a bit. Could I possibly exchange it for the same item instead?”

B015 Bot: “Certainly! We can proceed with exchanging the office chair for the same item.”

C016 Customer: “Yes, that sounds right. Let's go ahead with the exchange.”

**Policy finding.** The supplied exchange rule requires a different product option, while T018 records an identical item exchange. This is a policy and tool consistency finding for operator review. The shipment sequence promised in B019 also needs an authoritative operational source.

## Customer experience

The customer freely changes direction, confirms the proposed exchange and closes appreciatively.

## PR004 Priorities

We gathered from the conversation that the bot's priorities were:

1. Verify identity and locate the damaged chair order (B003–T012).
2. Recap the proposed return for confirmation (B013).
3. Accept the customer's change to a same-item exchange (B015).
4. Confirm the revised action and payment details (B015–C016).
5. Record the exchange request and describe the next steps (B017–B019).

**Recommendation:** Preserve the acknowledgment and responsiveness. Confirm the intended replacement procedure and reconcile the written rule with tool behavior; verify fulfillment wording.

# PR005 Address updates and a reversal

The final address states match the requests. The first update omits a required confirmation step.

| Understanding | Collaboration | Ease | Conversational Fit |
|---|---|---|---|
| 5/5 | 5/5 | 5/5 | 5/5 |

**Policy check: first confirmation step missed.**

**Conversational Fit:** C004 expresses privacy caution indirectly; B005 immediately offers the permitted alternative. C026 changes direction; B027–B031 clarify and follow it. These are strong conversational responses. The skipped initial confirmation remains a separate procedural finding.

## Evidence

C004 Customer: “I'd prefer not to share too much, can we proceed with another way?”

B005 Bot: “Sure, we can authenticate your identity using your first name, last name, and zip code.”

C026 Customer: “Actually, I've changed my mind. Can we switch my user address back to the original one?”

**Policy finding.** The supplied policy requires an action recap and explicit confirmation before consequential updates. B011 changes the profile before that separate step. The later order change and profile reversal include confirmation.

## Customer experience

Privacy caution is followed by active direction, additional requests and a freely expressed reversal.

## PR005 Priorities

We gathered from the conversation that the bot’s priorities were:

1. Find an authentication route that fits the privacy preference (B003–T008).
2. Collect and update the default profile address (B009–B013).
3. Check which order addresses are eligible to change (B015–B021).
4. Confirm and change the pending order address (B021–B025).
5. Confirm and restore the original default profile address (B027–B031).

**Recommendation:** Preserve the alternative verification route and responsiveness to changed preferences. Add the required recap and explicit confirmation before the first profile update.

# Source and attribution

Sierra Research tau-bench historical retail trajectories, fixed commit 59a200c6d575d595120f1cb70fea53cef0632f6b. Cases PR001–PR005 use trial 0 of source tasks 4, 7, 10, 18 and 22. Retrieved 17 October 2026. The original policy and tool results support the operational findings.

https://github.com/sierra-research/tau-bench/blob/59a200c6d575d595120f1cb70fea53cef0632f6b/historical_trajectories/gpt-4o-retail.json

MIT License

Copyright (c) 2024 Sierra

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.