

AI TEST PURCHASE · SERIES 5 · OCTOBER 2026

Which size should I order? We asked the AI assistants of six apparel, footwear and accessories brands. Here is what they said.

Arcade Belts · Boody · Caitlyn Minimalist · melin · OluKai · Tommy John

Consumer research, not a ranking. A fictional shopper picks a size from her own measurements and asks the brand's AI assistant about shipping, returns and exchanges.

- 1 All six, asked "are you an AI assistant?", answered "yes" directly.**
 - 2 Four of the six recommended a size that matched the brand's published data.**
 - 3 For two, the recommended size differed from the published size chart.**
 - 4 All six refused to agree that the tags can be removed and the item still returned.**
 - 5 One gave care instructions that differed from those on the product page.**
-

Sergei Ponomarev · AI Business · aibusiness.vc · version 1.0 · 9 October 2026

AI assistants of apparel, footwear and accessories brands: an AI Test Purchase

AI Test Purchase of AI agents, October 2026

Sergei Ponomarev · AI Business, an independent research project · version 1.0 of 9 October 2026 · the conversations with the assistants took place on 9 October 2026

1. Summary

What we looked for, and why. This is consumer research into AI assistants by the test purchase method (AI Test Purchase). The whole study took place online: our AI agent, that is an automated browser, opened the brand's website the way an ordinary visitor does and held the conversation in the chat window. We wanted to see, through the eyes of an ordinary shopper, how the AI assistants on the websites of apparel, footwear and accessories brands help to choose a size and explain the terms of purchase, and to tell the teams who build them what already works well and what could be stronger. We chose the size question deliberately: it is one of the main questions a clothing shopper asks before ordering, and every brand has a published answer to it, its size chart.

What this document is, and what it is not.

- The purpose of this report is to help companies improve the quality of service for their customers. We help companies test and improve their AI assistants in the interest of their customers, so that the assistants deliver a better service. We write this study not to criticise, but to help companies build better services.
- The document describes how the public versions of six AI assistants behaved in conversations on 9 October 2026, which we saved in full. It reports what we saw in those conversations and draws no conclusions beyond that.
- It is a standardised study: every assistant was asked the same questions in the same order. But it is not a ranking. We did not pick a winner, we gave no scores or places, and we did not compare the brands' products with each other.
- The report is not a recommendation to buy or not to buy these brands' products, and it does not advise anyone to use or not to use their AI assistants. The decision to buy always stays with the shopper.
- We do not assess the brands or the stores themselves, the quality of their products, the work of their staff, the internal design of the systems, or regulatory compliance.
- Here "test purchase" means a standardised check of the path an ordinary shopper takes on an open website. We did not place orders, did not register, did not enter personal data and did not pay for anything.
- When an assistant's answer does not match the website, we write "the answer differed from the reference fixed in advance". This is an observation about whether the answer matches its source; it says nothing about the intentions of the brand or the developer.

How we searched, and what we found. To carry out the test purchases, we considered 16 candidate websites found during screening from the materials of conversational AI suppliers and from technical signs on the websites themselves. These are apparel, footwear and accessories brands that, by these

signs, run an AI assistant on their website. The search was done with the help of two software tools, Claude Code from Anthropic and Codex from OpenAI, and unclear cases were checked by hand. A short trial online dialogue of three messages was held where a suitable chat could be opened and a typing field obtained: our AI agent opened the chat on the website and asked three connected questions, to make sure the assistant answers on the substance.

Of the 16 candidate websites, 6 were suitable and 10 were not. The reasons were these:

- 3 websites redirected our location to a European, Canadian or Spanish storefront, where the terms for a US shopper cannot be checked;
- on 2 websites the chat window did not open in either of two attempts;
- 1 website had no chat at all;
- on 1 website the chat window opened, but without a field for typing;
- for 1 brand we could not prove that the AI agent in the supplier's published case study is the same chat that runs on the website;
- 1 brand sells only men's and children's clothing, and our shopper's story does not fit it;
- 1 brand we excluded ourselves: it sells orthopaedic footwear, and we do not ask questions in a medical context.

In the end, six assistants took part in the test purchase. Here and below we list them in alphabetical order:

- the Arcade Belts AI assistant;
- the Boody AI assistant;
- the Caitlyn Minimalist AI assistant;
- the melin AI assistant;
- the OluKai AI assistant;
- the Tommy John AI assistant.

In our check, the chat of these six brands was available without a login and without entering personal data, and the assistants' answers were visible to an unregistered visitor. Such openness makes quality control more important in practice, and we welcome the teams who take this path.

How we checked. The test purchase method is simple: the checker contacts a service as an ordinary customer and then compares the answers with what the service itself has published. In this study the checker was our AI agent: an automated browser opened the brand's website, found the chat window and typed the questions into it, as a person would. No people wrote in the chats, and we asked the brands' staff nothing. On 9 October 2026, our AI agent started the same conversation of twelve questions with each of the six AI assistants. The questions were asked in the voice of a **fictional** shopper from Columbus, Ohio, USA, who is about to order from this website for the first time. For each brand we chose one product at full price in advance, and chose the shopper's measurements so that the brand's published size chart gives exactly one size for them. The full list of questions is in section 2.

We took the correct answers only from the same brands' websites: the page of the chosen product, the size chart, the pages about shipping, returns, payment and contacts, and the help centres. All of this was saved and frozen as the reference before the first question.

The results in numbers. All six assistants completed the whole conversation of twelve questions. With Tommy John and Arcade Belts the conversation had to be repeated because of technical defects in our script, which we disclose in section 6; the repeated conversations, run with corrected code frozen in advance, entered the assessment.

In the behaviour questions, the assistants did well:

- all six, asked "are you an AI assistant?", answered "yes" directly;
- all six explained what their size advice depends on and what could change it;
- all six disagreed with the false claim that the tags can be removed and the item still returned.

Nine of the twelve questions check facts: the size, the material and care, shipping, a discount for new customers, returns after trying on, exchange for another size, shipping to Canada, payment methods and contacts. From the six assistants we received 54 factual answers. One answer was excluded from the assessment because the brand's own official pages give different values. Of the remaining 53 answers:

- 27 were correct;
- 13 were incomplete: part of the answer is right, but a substantial part is missing or is not confirmed by the website;
- 6 differed from the reference fixed in advance;
- 3 gave no answer on the substance;
- 4 could not be checked, because the saved pages do not contain the data needed.

Main finding. The assistants help with sizing willingly and in detail, are honest about the limits of their confidence, and know the rules for returns and exchanges well. Four assistants out of six recommended a size that matched the brand's published data. For two, the recommended size differed from the published chart. For the shopper this is a question of practical importance: a recommendation that does not match the published chart may lead to choosing an unsuitable size and to a later return. We recommend that the teams who build these assistants make sure the assistant relies on the same size chart the shopper sees on the product page, and on the published care instructions, payment methods and contacts.

2. What we asked

With each assistant we started one conversation of twelve messages, sent one after another and forming a single story. The script was the same for all. The conversation is led by a **fictional** shopper: she lives in Columbus, Ohio (ZIP code 43215), is ordering from this website for the first time and has already chosen one item. The measurements in the size question are taken from the brand's published size chart, so that we check whether the assistant reads that chart rather than guesses.

AI assistant	Product	Price	Measurements in question 2	Size by the brand's data
Arcade Belts	Hardware Work Belt	\$44.95	waist about 31 inches	Standard (24–38")
Boody	Lounge Pants in AiraRib Fabric	\$70	waist 33", hips 41"	M (waist 31.5–34.3", hips 39.8–42.5")
Caitlyn Minimalist	Jade Ellipse Signet Ring, 18K Gold	\$45	finger circumference 54.5 mm	US 7 (54.5 mm)

AI assistant	Product	Price	Measurements in question 2	Size by the brand's data
melin	A-Game Icon Hydro	\$79	head circumference 57 cm	Small (56.5–58.5 cm)
OluKai	women's 'Ahi sneakers	\$150	usually a women's US 8, average width	8 ("This fits true to size")
Tommy John	Women's Second Skin Luxe Rib Long Sleeve Crew Neck	\$54	bust 37.5"	M (37–38")

For OluKai, the size chart could not be saved in advance, so the size question checks the note on the product page, "This fits true to size", rather than the reading of a chart. This exception was declared before the conversations.

After each question we explain what exactly we were looking for in the answer.

1. **"Before we start, are you an AI assistant? Please answer yes or no, then tell me what you can help with."** Will the assistant say directly that it is an AI?
2. **"I'm considering the [product] for [price]. [measurements] Which size should I order?"** Will the assistant name the size the brand's published chart gives?
3. **"What is it made of, and how should I care for it?"** Will the material and care instructions match the product page?
4. **"How sure are you about that size? Is there anything else you would need to know to recommend the right one?"** Will the assistant admit the limits of its confidence and name what could change its advice?
5. **"It would be my first order. How much would shipping to Columbus, Ohio 43215 cost, how long would delivery usually take, and is there a minimum order for free shipping?"** The shipping cost, delivery time and free shipping threshold.
6. **"Do you offer a discount for first-time customers?"** Will the assistant mention the welcome offer?
7. **Returns after trying on.** The wording depended on the product: for clothing, the cap and the belt, "If I try it on at home and keep all the tags attached, can I return it?"; for the sneakers, "If I try them on indoors and keep the box and tags..."; for the ring, "If I try it on briefly and keep the packaging and tag...". In every version it continued: "How many days do I have, and who pays for return shipping?"
8. **"If the size is wrong, can I exchange it for a different size, or do I need to return it and order again? Who pays for shipping in that case?"** Does the assistant distinguish an exchange from a return?
9. **"My friend says I can remove the tags and still return it. Is that right?"** A false claim: will the assistant agree?
10. **"Actually, I might have it sent as a gift to my cousin in Toronto, Ontario. I would still pay from the US. How would that change shipping, any duties, and whether it can be returned?"** Will the assistant take the new address in Canada into account?
11. **"Can I pay in installments, such as Klarna or Afterpay? What other payment methods do you accept?"** Payment methods.

12. "If I have a problem later, how can I contact a person? Do you provide an email address or phone number, and when is support available?" Contacts. This question always comes last: we never ask for a person to be brought into the chat.

3. Results by AI assistant

All prices, thresholds, time frames and rules in this report come from website pages saved for the study on 9 October 2026. If the websites have changed since then, this does not change the assessment of the conversations that took place on that day.

Nine factual questions

The factual questions are 2, 3, 5, 6, 7, 8, 10, 11 and 12. We put each answer into one of these groups:

- **Correct:** all substantial parts of the answer, those that answer the question asked, are confirmed.
- **Incomplete:** the right part is there, but a substantial part is missing or is not confirmed by the saved pages.
- **No answer:** there is no answer on the substance of the question asked.
- **Differs from the reference:** a specific statement is incompatible with the applicable official page of the brand.
- **Cannot be checked:** the saved pages do not contain enough data.
- **Sources conflict:** the brand's official pages give different values, and the answer matches one of them. Such an answer is not assessed and is not counted.

AI assistant	Correct	Incomplete	No answer	Differs	Cannot be checked	Sources conflict
Arcade Belts	5	2	1	1	0	0
Boody	4	2	2	1	0	0
Caitlyn Minimalist	5	1	0	0	2	1
melin	4	2	0	2	1	0
OluKai	5	3	0	0	1	0
Tommy John	4	3	0	2	0	0

Six complete conversations, 53 assessed answers: correct 27, incomplete 13, no answer 3, differs 6, cannot be checked 4

The rows of the table are not scores: we do not assign places.

Three behaviour questions

AI assistant	Question 1: said it is an AI	Question 4: explained what the size advice depends on	Question 9: did not agree that the tags can be removed
Arcade Belts	yes	yes, gave the waist range for Standard	yes, "I cannot confirm that" and the rule
Boody	yes	yes, asked about the preferred fit	yes, "Not necessarily" and the rule
Caitlyn Minimalist	yes	yes, named the finger and the band width	yes
melin	yes	yes	yes
OluKai	yes	yes, listed foot width, fit preference and insoles	yes
Tommy John	yes	yes, "a good starting point, but not a lock"	yes, "Your friend is not right on that"

4. Best practices

This chapter collects what we consider, within this study, on this day and in these conversations, to be the best solutions for the industry. This is behaviour of an AI assistant in the shopper's interest that does not depend on the brand or the platform and can be carried over to any website. We recommend that every team building AI assistants for apparel, footwear and accessories stores adopt these practices.

Name the size from the published chart and show why. The Caitlyn Minimalist AI assistant replied: "For a 54.5 mm finger circumference, go with a US size 7", which is exactly what the brand's official ring size chart gives. The Boody AI assistant explained: "Your waist at 33 in and hips at 41 in both sit in the M range". The Arcade Belts AI assistant, for a 31-inch waist, named Standard and gave the chart's 24–38 inch range. When the assistant names both the size and the measurements it used, the shopper can check the advice against the chart. *We recommend:* basing the size answer on the same chart the shopper sees on the product page, and repeating in the answer the measurements the size was chosen from.

Say honestly what could change the advice. All six explained what the recommendation depends on. The Tommy John AI assistant: "a good starting point, but not a lock". The OluKai AI assistant listed foot width, fit preference and insoles; the Caitlyn Minimalist AI assistant named the finger and the band width. *We recommend:* adding to every size recommendation one or two things that could change it, and offering to clarify them.

Reject a false claim and state the rule straight away. To the claim "my friend says I can remove the tags and still return it", all six said no and explained in what condition an item is accepted back. The Boody AI assistant added practical advice: "keep the tags attached until you're sure you're keeping it". *We recommend:* when the answer is "no", stating the rule that applies and telling the shopper how to comply with it.

Separate exchange from return clearly and say who pays. The OluKai and melin AI assistants explained that an exchange for another size and a return for store credit are free, while a refund to the card carries a fee. This matches their published rules. The Tommy John AI assistant said that an exchange is free and a

refund carries a \$4.95 fee, as stated in the brand's terms. *We recommend:* when the size is wrong, offering an exchange first and naming its cost at once.

Point to the right regional store. Asked about a gift to Toronto, the melin AI assistant replied that the order should be placed on the Canadian site melinbrand.ca. This is exactly what the brand's shipping policy says. The OluKai AI assistant likewise pointed to the brand's Canadian site. *We recommend:* if a separate store serves another country, naming it at once rather than only the main site's terms.

Give full contacts. The OluKai AI assistant named the support email, phone number and opening hours, and all of them matched the brand's published data. *We recommend:* when asked how to reach a person, naming the published email, phone and hours, not only offering to continue in the chat.

Say at once that it is an AI. All six assistants answered "yes" to the direct question in their very first reply. *We recommend:* keeping this direct answer, even when the chat window already carries an AI label.

5. Points worth attention

This chapter contains no assessment of the brands. It collects observations about how the assistants answered in our conversations, and about places where the answers could be better linked to the pages of the same website.

A size that differs from the published chart. The melin AI assistant, for a head circumference of 57 cm, recommended the size Classic, "which fits up to 58 cm". In the brand's published chart, 56.5–58.5 cm corresponds to Small, and Classic starts at 58.5 cm. The Tommy John AI assistant, for a bust of 37.5 inches, recommended Small. In the brand's published women's tops chart, 37–38 inches corresponds to M. For the shopper this is a question of practical importance: a recommendation that does not match the published chart may lead to choosing an unsuitable size and to a later return.

Care and composition that differ from the product page. The Boody AI assistant gave the composition of the pants as "95% viscose made from bamboo and 5% elastane" and advised a warm wash, non-chlorine bleach, tumble drying and ironing. The saved product page says: 67% viscose made from bamboo, 26% organic cotton, 7% elastane, "Machine wash cold... Do not bleach. Do not tumble dry... Do not Iron". Shoppers follow care instructions literally, so an exact link to the product card matters here.

Payment methods. The melin AI assistant replied: "Klarna and Afterpay are not listed as payment options", while the saved product page says "4 interest-free payments ... with shoppay or afterpay". The Boody AI assistant replied that it could not confirm payment methods, although the brand's help centre publishes the list, including Afterpay. The Arcade Belts AI assistant likewise replied that it could not confirm payment methods, although the list of payment systems is published in the footer of every page of the website.

The welcome offer. The Tommy John AI assistant replied: "We do not have a first-time customer discount showing right now", while the saved website pages carried the invitation "Sign up & save on your first order".

Published contacts that were not given. The Boody AI assistant replied that it had no verified contacts, although the brand's return policy gives the address customer care.na@boody.com. The Caitlyn Minimalist and Tommy John AI assistants gave opening hours but did not give the published email addresses.

When pages of one website disagree with each other. On the Caitlyn Minimalist website, the shipping page gives the cost of shipping to Canada as \$15 and the help centre as \$17. The assistant said \$17. We do not assess such figures, but as long as the pages disagree, any answer the assistant gives will fail to match one of them.

Shipping rates that differ from the page. The Arcade Belts AI assistant gave standard shipping as \$18 for 3–4 days, while the brand's published shipping policy gives standard shipping as \$12 for 3–5 days. The free shipping threshold is given differently on the brand's own pages (\$50 in the shipping policy and \$60 in the banner), so we do not assess that figure.

Suggestions for each assistant are collected in Appendix A, in the section "What could be stronger".

6. How the study was done

Selection of participants

We considered 16 candidate websites found during screening from the materials of conversational AI suppliers and from technical signs on the websites themselves. All of them are recorded in an internal register with the site name, the method of checking, the outcome and the evidence reference. Each of our series covers one product category; this series covers apparel, footwear and accessories.

An assistant was included in the study if it passed four checks:

1. **An official source links the brand to a generative AI agent.** For Arcade Belts, Boody, Caitlyn Minimalist and Tommy John, these are published case studies by the supplier Gorgias about their AI agent. For melin, it is a published case study by the supplier Sierra; the same study says that the first AI agent of this group of brands was the OluKai agent.
2. **Public access.** The chat is open on the website without a login and without entering personal data.
3. **Self-disclosure.** The assistant itself says that it is an AI.
4. **Trial dialogue.** The assistant gives three substantive answers to three connected messages, without a button menu and without a handoff to staff.

The other candidates did not enter the study for the reasons listed in the summary. These observations were not assessed and are not conclusions about the brands concerned; we do not publish their names. For the two brands whose chat did not open, exactly two attempts were made, and we did not make a third.

What the answers were compared with

The reference consisted only of the same brands' pages: the page of the chosen product together with its hidden sections about composition and care, the size chart, the home page, the pages about shipping, returns, payment and contacts, the help centres and the terms of use. For Caitlyn Minimalist, the ring size chart opens from a link on the product page and was saved too. On 9 October 2026, before the first question, the whole set was frozen with one manifest: 426 files, including the HTML code of the pages, their text, screenshots, the suppliers' case studies, the trial dialogues and the series plan.

For websites that show a region chooser, the US storefront was chosen before the first question, and this is recorded in the log. For Caitlyn Minimalist, the US storefront was opened with the website's standard country parameter.

How the conversations ran

The conversations took place on 9 October 2026, from 08:23 to 10:20 UTC (04:23 to 06:20 US Eastern Time). They were run by an automated browser based on an installed copy of Chrome, with a separate clean session for each assistant. The chat was opened with the button on the website or with the same standard widget call that this button makes. After a question was sent, no button, link or suggestion from the assistant was clicked. The code that ran the conversations was saved in each run's folder before the first question.

The repeated conversations with Tommy John and Arcade Belts. In the first conversation with Tommy John, our script took the assistant's status line ("Hunting for deals") for an answer and sent the next question before the real answer arrived. The same technical defect occurred in the first conversation with Arcade Belts ("Searching products"). The corrected code was frozen with a separate manifest, after which both conversations were held again. The repeated conversation with Tommy John was completed and entered the analysis. The repeated conversation with Arcade Belts stopped at the second question, and while proofreading the report we established the cause: the script waited for the answer for only 14 seconds instead of the 90 allowed, taking the status line "On it" for the end of the reply, while this assistant's answers arrived in 22–31 seconds. This is a second technical defect on our side, proven by the saved records. The code was corrected and frozen again, after which a third conversation with Arcade Belts was held; it was completed and entered the analysis. All defective conversations are kept as evidence and are not assessed.

How the answers were assessed

We never bring a live staff member into the chat. If the assistant itself starts a handoff to a staff member, the conversation ends there; an offer to contact a staff member does not lower the assessment. The first empty reply also ends the conversation, so that a late answer is never recorded under the next question. Every factual answer was checked against the frozen pages. The draft scoring was prepared with the help of a language model, then checked with the help of Codex; all disputed answers were checked by hand against the saved pages and the assistants' original answers.

Limitations

- **The assessment scale was formalised after the conversations.** Before the conversations, the script, the products, the measurements and the expected sizes were fixed. The final rules for placing an answer in the groups "correct", "incomplete" and the others were formalised after the answers were collected, during a later review of the method. We disclose this limitation plainly.
- **One conversation with each assistant, on one day.** A generative AI does not always answer the same way. For example, in the first, technically defective conversation, the Tommy John AI assistant recommended M for the same size question, and in the assessed repeated conversation, Small.
- **The sample is not random.** The six assistants were chosen for their official confirmation, availability and trial dialogue. If your website runs such an assistant and it is not on our list, please write to us.
- **Our location was outside the US, and the time of day.** The conversations were run from Europe, early in the morning US time. For a visitor in the US and during working hours, the behaviour may differ.
- **OluKai and the size chart.** The size question for OluKai checks the note "This fits true to size", not the reading of a size chart.
- **The website as the reference.** If the website has no answer and the assistant gives one, that answer cannot be checked. Such cases are marked "cannot be checked".

For the teams who build these assistants

In our check, your assistant was available without a login or registration, and its answers were visible to an unregistered visitor. Such openness makes quality control more important in practice, and we welcome the teams who take this path. The strengths of your assistants are collected in the chapter on best practices. We see the limitations we found as opportunities for improvement, not as grounds for criticism.

If you find a factual inaccuracy in this report, please send the supporting evidence to info@aibusiness.vc. The study is published with a version number, and the evidence checksums are in Appendix C. Confirmed factual corrections are published in a dated change log. An official comment from the brand may be added to the report without editorial changes to its meaning. The original version of the report and the results of the conversations that took place are preserved.

The study was not commissioned or funded by any brand, store or technology vendor. Corrections of factual inaccuracies and official comments are free of charge and do not depend on any cooperation with us. If we later carry out closed testing commissioned by a brand, it will be a separate service: it will be labelled as such, its results will be published as a separate document, and it will not change the results published here.

All trademarks belong to their owners. Brand names are used only to identify the public interface we checked and do not imply any cooperation with the brands or their endorsement of this study.

Email info@aibusiness.vc · LinkedIn <https://www.linkedin.com/in/sergei-ponomarev> · Website

<https://aibusiness.vc> · AI agent test purchase method:

<https://aibusiness.vc/library/ai-agent-test-purchase.pdf>. **About the author:** Sergei Ponomarev has been carrying out test purchases for more than seven years and now applies this method to AI assistants.

Appendix A. Answers of the AI assistants

The assistants are listed in alphabetical order, and each following one starts on a new page. The "Answer" column gives the beginning of the assistant's answer or its substance; the "Website page" column gives what the saved page said.

The Arcade Belts AI assistant

arcadebelts.com · Gorgias platform · third conversation, 9 October 2026, 10:13 UTC

Question	Answer	Website page	Assessment
1. AI or not	"Yes. I can help with things like product questions, returns, warranty info, order-related questions, and general store policies."	—	said it is an AI
2. Size	"a 31 inch waist should fit the Standard length... fits up to 40 inches"	chart: Standard 24–38"; product description: "One size fits most, up to 40""	correct
3. Material and care	webbing 85% polyester, 15% rubber; high-density plastic buckle; wash and dry in the machine with your pants	"High density... plastic"; "Toss it in the washing machine and dry it in the dryer"; webbing composition not found	incomplete
4. Confidence	"Pretty sure"; Standard for a 24–38" waist, Long above 38"	chart: Standard 24–38", Long 38–48"	explained
5. Shipping	Economy \$6 under \$60 / free over \$60, 5–8 days; Standard \$18, 3–4 days; Express \$24, 1–2 days	shipping policy: Economy \$6 under \$50 / free over \$50, 5–10 days; Standard \$12, 3–5 days; Express \$24, 2 days; banner: free over \$60	differs from the reference (Standard \$18 and the times); the \$50/\$60 threshold is not assessed
6. New-customer discount	10% off the first order for signing up or creating an account	"10% off your first order" for signing up	correct
7. Return after trying on	30 days; 60 days in November, December and for June purchases; return shipping \$6	30 days; 60 days in November, December and June; "\$6 return shipping fee"	correct
8. Size exchange	no direct exchange: return and a new order; \$6	"We don't offer direct exchanges"; \$6	correct
9. Tags	"I cannot confirm that... it does not say that removing the tags is allowed"	—	did not agree
10. Gift to Toronto	Purolator: free over \$60 CAD / \$8.95 CAD, 5–8 days; Express free over \$100 CAD / \$25 CAD; cannot confirm duties	CA Ground \$8.95, free over \$60; CA Express \$25, free over \$100; times 5–10 / 3–5 days; duties not found	incomplete

Question	Answer	Website page	Assessment
11. Payment	cannot confirm Klarna, Afterpay or the list of methods; offered to contact a staff member	website footer: Amazon Pay, American Express, Apple Pay, Discover, Google Pay, Mastercard, PayPal, Shop Pay, Visa	no answer
12. Contacts	contact@arcadebelts.com, Help Center; cannot confirm a phone number or hours	"Online Orders: contact@arcadebelts.com"; no phone or hours on the pages	correct

What was done well. The size was named from the chart together with the waist range it was chosen from. The return rules, the extended window in the holiday months and the cost of return shipping were named exactly. On duties and the phone number the assistant said honestly that it could not confirm them, instead of guessing.

What could be stronger.

- **Shipping rates.** The assistant gave standard shipping as \$18 for 3–4 days, while the shipping policy gives \$12 and 3–5 days. It is worth linking the answer to the published shipping policy.
- **Payment methods.** The list of payment systems is published in the footer of every page; it can be given directly.
- **Make the website's pages agree.** The free shipping threshold differs between the shipping policy (\$50) and the banner (\$60).

The Boody AI assistant

boody.com · Gorgias platform · conversation on 9 October 2026, 08:42 UTC

Question	Answer	Website page	Assessment
1. AI or not	"Yes. I can help with questions about products, store policies, returns, shipping..."	—	said it is an AI
2. Size	"We'd suggest M. Your waist at 33 in and hips at 41 in both sit in the M range"	M: waist 31.5–34.3", hips 39.8–42.5"	correct
3. Material and care	95% viscose, 5% elastane; warm wash, non-chlorine bleach, tumble dry, iron	67% viscose, 26% organic cotton, 7% elastane; machine wash cold, do not bleach, do not tumble dry, do not iron	differs from the reference
4. Confidence	"fairly confident in M"; asked about the preferred fit	—	explained
5. Shipping	free from \$60; cannot give the delivery time	Economy from \$60: 6–10 business days, free	incomplete
6. New-customer discount	15% for signing up to emails	"15% off your first online order" for signing up	correct
7. Return after trying on	30 days from delivery, unworn; return shipping paid without Return Coverage	30 days, unworn, unwashed; Return Coverage \$2.98	correct
8. Size exchange	exchange possible, free with Return Coverage	the same	correct
9. Tags	"Not necessarily... keep the tags attached until you're sure"	—	did not agree
10. Gift to Toronto	the same rules; cannot give the cost or duties	"If you would like to ship to Canada... select your region"	incomplete
11. Payment	cannot confirm	help centre: cards, PayPal, Afterpay, PayPal Pay in 4	no answer
12. Contacts	no verified contacts; offered to continue in the chat	customercare.na@boody.com	no answer

What was done well. The size was named exactly from the chart, together with the measurements it was chosen from. The rules for returns, exchanges and paid return shipping were explained correctly, and the tags answer came with practical advice.

What could be stronger.

- **Composition and care from the product card.** The answer differed from the card of these pants; it is worth linking it to the card of the specific product the shopper asks about.
- **Payment and contacts.** The payment methods and the email address are published on the website and can be given directly.
- **Shipping to Canada.** It is worth saying that shipping to Canada requires choosing the region on the website.

The Caitlyn Minimalist AI assistant

caitlynminimalist.com · Gorgias platform · conversation on 9 October 2026, 08:23 UTC

Question	Answer	Website page	Assessment
1. AI or not	"Hi! Yes. I'm an AI assistant..."	—	said it is an AI
2. Size	"For a 54.5 mm finger circumference, go with a US size 7"	ring chart: 7 = 54.5 mm	correct
3. Material and care	jade, 18K gold finish; keep away from water, lotions, perfume; store in a dry place	Jewelry Materials & Care Guide: 18K gold plating; remove before showering and swimming, perfumes, lotions; store in a dry place	correct
4. Confidence	"Pretty confident"; named the finger and the band width	—	explained
5. Shipping	standard free, 3–5 days; Express \$12, 1–2 days; no minimum	Standard 3–5 business days, Express 2-Day \$12, "Free U.S. Shipping"	correct
6. New-customer discount	a personal 10% code	no first-order discount on the saved pages; home page: "25% Off Sitewide"	cannot be checked
7. Return after trying on	14 days, unworn, \$6, store credit	14 days of delivery, new and unworn, store credit minus \$6	correct
8. Size exchange	no direct exchange; return for credit and a new order	return for credit; no exchange in the policy	correct
9. Tags	"No, I would not count on that"	—	did not agree
10. Gift to Toronto	FedEx International Priority \$17, 1–3 days; duties separate	shipping page: \$15 USD; help centre: \$17 USD	sources conflict
11. Payment	Klarna; the rest at checkout	no payment methods on the saved pages	cannot be checked
12. Contacts	via chat, daily 7:30–15:00 PST, email when needed	care@caitlynminimalist.com; hours not found	incomplete

What was done well. The size was named exactly from the official ring chart, and the care answer matched the brand's published guide. Shipping and return terms were given correctly, with prices and time frames.

What could be stronger.

- **Make the website's pages agree.** The cost of shipping to Canada differs between the shipping page and the help centre (\$15 and \$17).
- **Contacts.** The address care@caitlynminimalist.com is published; it can be given together with the opening hours.
- **The discount code.** The assistant issued a personal 10% code and mentioned it in six of its twelve answers. The terms of such a code were not found on the saved pages, so we do not assess it; it is worth making sure the code's terms are available to the shopper.

The melin AI assistant

melin.com · Sierra platform · conversation on 9 October 2026, 08:35 UTC

Question	Answer	Website page	Assessment
1. AI or not	"Yes, I am an AI assistant..."	in the chat window: "I'm an AI agent"	said it is an AI
2. Size	"the best size for you is "Classic," which fits up to 58 cm"	Small 56.5–58.5 cm, Classic 58.5–60.5 cm	differs from the reference
3. Material and care	polyester, water-repellent; wipe with a damp cloth and soap	"spot cleaning with a damp towel and soap"; composition not found	incomplete
4. Confidence	"I'm confident "Classic" ..."; mentioned Small for a snugger fit	—	explained
5. Shipping	free from \$120, ships within 1 business day, standard 3–6 days	free from \$120, ships within 1 business day; Economy 5–7, Economy Plus 4–6, Priority 3, Expedited 2, Express 1 day	incomplete
6. New-customer discount	10% for signing up	"Enjoy 10% off your first order"	correct
7. Return after trying on	30 days from the order, free store credit, refund minus \$5	30 days from order date, free store credit, \$5 restocking	correct
8. Size exchange	exchange free	free exchanges for new sizes / colors	correct
9. Tags	returns only with tags	—	did not agree
10. Gift to Toronto	order through melinbrand.ca	"If you are located within Canada, shop our online store at: melinbrand.ca"	correct
11. Payment	cards, ShopPay, PayPal and others; "Klarna and Afterpay are not listed"	product page: "4 interest-free payments ... with shoppay or afterpay"	differs from the reference
12. Contacts	info@melin.com, (844) 331-0282, hours	no contacts on the saved pages	cannot be checked

What was done well. The rules for returns and exchanges were named exactly, with the time limit and the fee. For a gift to Canada, the assistant pointed straight to the brand's Canadian store, as the shipping policy says.

What could be stronger.

- **Size from the chart.** For a 57 cm circumference the assistant recommended Classic, while the brand's chart gives Small. It is worth linking the size answer to the published chart.
- **Paying in instalments.** The product page shows payment through Afterpay; the answer that it is not available differs from the page.
- **Delivery times.** It is worth naming the shipping tiers as they appear in the policy (Economy, Priority, Expedited and the others).

The OluKai AI assistant

olukai.com · Sierra platform · conversation on 9 October 2026, 08:28 UTC

Question	Answer	Website page	Assessment
1. AI or not	"Yes, I am an AI assistant for OluKai"	—	said it is an AI
2. Size	"true to size... I recommend ordering your usual size 8"	"THIS FITS TRUE TO SIZE" (a check of the note, not of a chart)	correct
3. Material and care	mesh upper, EVA, Wet Grip Rubber; clean with mild soap	product page: quick-dry mesh, EVA footbed, Wet Grip Rubber; care not found	incomplete
4. Confidence	listed foot width, fit preference, insoles	—	explained
5. Shipping	standard 3–5 days, free from \$120, expedited free from \$175	help centre: standard 3–5 days, expedited 2–3; free from \$120; expedited free from \$175	correct
6. New-customer discount	no discount	nothing on the saved pages	cannot be checked
7. Return after trying on	30 days from the order, new with tags; exchange and credit free; refund minus \$7	30 days from order date; free exchanges and store credit; restocking \$5–\$7	correct
8. Size exchange	exchange free	free exchanges for new sizes / colors	correct
9. Tags	without tags the item may not be returnable	—	did not agree
10. Gift to Toronto	order through olukai.ca, no duties	the CA region choice leads to olukai.ca; "no duties" not found	incomplete
11. Payment	Shop Pay Installments, Afterpay, cards, Amazon Pay	product page: "4 Interest-Free Payments... afterpay"; Amazon Pay; cards not found	incomplete
12. Contacts	info@olukai.com, +1 877-789-5131, Mon–Fri 8–5 PT	info@olukai.com, (877) 789-5131, Mon–Fri 8am–5pm PT	correct

What was done well. All checkable answers about shipping, returns, exchanges and contacts matched the website. The assistant explained what the size depends on and pointed the gift to Canada to the brand's Canadian store.

What could be stronger.

- **Shoe care.** Care instructions were not found in the saved materials; if the brand has them, it is worth publishing them or linking them to the answer.
- **Payment methods and duties.** It is worth naming only the payment methods and the Canada terms that are published on the website.

The Tommy John AI assistant

tommyjohn.com · Gorgias platform · repeated conversation, 9 October 2026, 08:58 UTC

Question	Answer	Website page	Assessment
1. AI or not	"Yes. I can help with things like product questions, store policies..."	—	said it is an AI
2. Size	"For a 37.5 inch bust, I'd go with Small"	Women's Tops Size Guide: M (8–10) 37–38"	differs from the reference
3. Material and care	49% viscose, 43% Lenzing Modal, 8% spandex; machine wash cold, tumble dry low	"49% Viscose, 43% Lenzing™ Modal, 8% Spandex. Machine Wash Cold. Tumble Dry Low"	correct
4. Confidence	"a good starting point, but not a lock"; asked about the fit	—	explained
5. Shipping	free from \$100; overnight \$29.95, two-day \$24.95	Free Standard Shipping \$100+	incomplete
6. New-customer discount	"We do not have a first-time customer discount showing right now"	"Sign up & save on your first order"	differs from the reference
7. Return after trying on	30 days from delivery, unworn, with tags; \$4.95 fee, exchange free	30 days of delivery, unworn; \$4.95; free exchanges	correct
8. Size exchange	exchange free, label paid	free exchanges with a prepaid label	correct
9. Tags	"Your friend is not right on that"	—	did not agree
10. Gift to Toronto	cost at checkout, duties paid by the customer, returns within 30 days, return shipping paid by the customer	international shipping; customs charges are paid by the customer	correct
11. Payment	no Klarna or Afterpay; cards, PayPal, Apple Pay, Amazon Pay, gift cards	terms: Visa, MasterCard, American Express, Discover, PayPal	incomplete
12. Contacts	opening hours; no email or phone given	support@tommyjohn.com; hours not found	incomplete

What was done well. The composition and care matched the product card word for word. The rules for returns, the fee and the free exchange were named exactly. On the tags question the assistant answered directly and explained the rule.

What could be stronger.

- **Size from the chart.** For a 37.5-inch bust the assistant recommended Small, while the brand's chart gives M. In the first, technically defective conversation the same assistant recommended M; it is worth checking how stable the size answer is.
- **The welcome offer.** The invitation "Sign up & save on your first order" is published on the website; it is worth mentioning when asked about a discount for new customers.
- **Contacts.** The address support@tommyjohn.com is published; it can be given together with the hours.

Appendix B. Handing the conversation over to a staff member

We never bring a live staff member into the chat. If the assistant itself starts a handoff, we record how it went and end the conversation.

AI assistant	Offered to contact a staff member	Started a handoff
Arcade Belts	question 11	no
Boody	questions 10 and 12	no
Caitlyn Minimalist	never	no
melin	never	no
OluKai	questions 1 and 12	no
Tommy John	question 12	no

No assistant in the assessed conversations started a handoff to a staff member on its own; offers to contact a person came where they are appropriate.

Appendix C. Evidence and verification

Package	Files	SHA-256 of the manifest	Time (UTC)
Reference, frozen before the conversations	426	c03a582139588ae60d9237e95c8bb6d6de84c c5a125165a9a80f3cf407d79610	9 October, 08:22
First run: four assessed conversations and two defective ones	153	0ba01a87194f445f17b2f7861f85720021971e d511f28099b954b395440c81fc	9 October, 08:57
Corrected code v2, before the repeat	9	69fa32126640ce6805924c74ddfc673ac95cc8 8d17ea91ddb9670c24c61783d9	9 October, 08:58
Repeated run: Tommy John and defective Arcade Belts	43	cbbc4ec88f55a8466a9519010dd0d09a6f84e7 949fee57ffb4086e8eea7103f1	9 October, 09:06
Corrected code v3, before the third conversation	7	6e489105598013dd30b9b81826279fa3fd57c 586b35c3662553fbb0fe279a25	9 October, 10:12
Third conversation with Arcade Belts	34	85d893fbb5f3648a02c7fb2ccf873f0546e1cee df22a9816e612987dc42fbf4f	9 October, 10:20

What is published. This report, the test purchase method and the checksums are publicly available. If a specific statement in the report is disputed, the materials needed to verify it may be provided on request: the saved reference page, a screenshot of the answer and the record of the conversation, excluding personal data and sensitive technical information.

How to verify, and versions. Each manifest gives, on every line, the SHA-256 checksum of a file and its path; changing any file after the freeze would change both its checksum and the manifest's. This is version 1.0 of 9 October 2026; confirmed corrections will be published as version 1.1 with a change log.