

Pick the right AI shopping assistant in under an hour.

A 4-step worksheet to score your store profile, shortlist 2-3 platforms, run the only test that matters, and track outcomes against benchmarks.

5 fields

Store profile in Step 1

3 tests

On real catalog data, Step 3

90 days

KPI tracker, Step 4

Profile your store across 5 dimensions

Fill in honest answers. Do not skip a field. Output: which platform category fits your store.

1 Daily traffic volume

Average sessions per day (last 30 days): _____

- ☐ Under 500 — AI ROI marginal, fix traffic first ☐ 500 to 1,000 — borderline, depends on AOV
☐ 1,000+ — economics work

2 Average order value

AOV (last 30 days): \$ _____

- ☐ Under \$20 — math rarely works ☐ \$20 to \$50 — borderline ☐ \$50 to \$80 — healthy revenue channel
☐ \$80+ — primary revenue channel

3 Product complexity (predicts your achievable resolution rate)

- ☐ Deep-knowledge (supplements, technical components) — 95 to 99% target
☐ Broad-catalog (10,000+ SKUs) — 96 to 97% target
☐ Subjective or style (fashion, lifestyle, decor) — 78 to 85% target

Counterintuitive: technical products are easier for AI than fashion. 80% resolution in fashion is success, not failure.

4 Integration platform

- ☐ Shopify (most platform options) ☐ BigCommerce ☐ WooCommerce
☐ Custom headless (likely needs enterprise platform)

Order systems connected: _____

5 Channel and language needs

Non-English traffic share: _____ % (if 30%+, multilingual is a hard requirement)

- ☐ Web ☐ Instagram ☐ WhatsApp ☐ Email ☐ SMS

Your profile category — mark one

- ☐ **Sales-grade** — Shopify + \$50+ AOV + 1K+ sessions → Test Chatty, Rep AI
☐ **Hybrid** — multi-platform OR balanced sales/support → Test Tidio Lyro, Intercom Fin
☐ **Support-grade** — helpdesk-led, ticket volume > sales chat → Test Gorgias Automate
☐ **Enterprise** — omnichannel + multilingual at scale → Test Zowie, Zendesk AI
☐ **Not ready yet** — under 500 sessions OR AOV under \$20 → revisit in 6 months

Shortlist 2-3 platforms, then test for real

Vendor demos lie structurally. Run each test on the free trial against your real catalog. One hour, three tests, four red flags.

Your shortlist (from Step 1 profile + matrix)

Platform A

Platform B

Platform C

Test scoring matrix

Test & what good looks like	Platform A	Platform B	Platform C
Test 1 — Inventory + shipping <i>"Do you have the medium black jacket in stock, arrive Friday at 90210?"</i> Good: specific SKU stock count + delivery date for the zip. Bad: generic "3-5 days" policy.	☐ Pass ☐ Fail	☐ Pass ☐ Fail	☐ Pass ☐ Fail
Test 2 — Price objection <i>"\$89 is way more than I had planned to spend."</i> Good: reframes value, mentions third-party validation, surfaces payment options. Bad: feature list or FAQ paragraph.	☐ Pass ☐ Fail	☐ Pass ☐ Fail	☐ Pass ☐ Fail
Test 3 — Human handoff <i>Force escalation, then check what the agent dashboard shows.</i> Good: 1-line summary + blocking question + suggested response + intent + cart value. Bad: raw transcript dump.	☐ Pass ☐ Fail	☐ Pass ☐ Fail	☐ Pass ☐ Fail

Red flags during the sales process — check if observed

- ☐ Vendor cannot quote average chat-to-sale conversion across their customer base
- ☐ Demo runs on sandbox, not a real merchant store
- ☐ Success metrics revolve around tickets deflected, not revenue generated
- ☐ Pricing breakdown obscures per-channel or per-conversation fees

Decision rule

Winner = passes **all 3 tests** and triggers **≤1 red flag**.

Multiple platforms qualify? Pick by total cost at your monthly conversation volume, then by Test 3 handoff quality. None qualify? Rerun Step 1 profile or broaden Step 2 shortlist.

Measure outcomes against benchmarks

Deployment is the baseline, not the finish line. Most failures plateau at 60-90 days. Fill in actual numbers as you measure.

Metric	30-day		60-day		90-day	
	Target	Actual	Target	Actual	Target	Actual
Chat-to-sale conversion	8%+	_____ %	10%+	_____ %	12%+	_____ %
Autonomous resolution	70%+	_____ %	80%+	_____ %	85 to 99%	_____ %
Revenue per conversation	\$5+	\$_____	\$8+	\$_____	\$10+ at \$80 AOV	\$_____
Time to first chat-attributed sale	Under 7 days	_____ days	—	—	—	—
Handoff CSAT (post-handoff)	4.0+	_____ /5	4.3+	_____ /5	4.5+	_____ /5

If conversion is under **5%** at the 30-day mark, something is wrong. Diagnose data quality, training, or platform-profile mismatch *before* extending the contract.

When to switch (after 90 days)

- ☐ Conversion plateaus below 5% despite tuning
- ☐ Resolution sits below industry benchmark for your product type
- ☐ Your team spends more time correcting AI errors than handling true escalations
- ☐ Hidden costs push effective price 2-5x above plan rate

Switch cost reality

\$50K-\$150K for retraining plus **3-6 months** of optimization to reach previous accuracy. Most platforms lose AI training data and conversation history in transition. Step 1 + Step 2 prevent this.

Run the 3 tests against a sales-grade platform built for Shopify.

Chatty is one of the two sales-grade options on the matrix. Free trial, no credit card. Test it on real catalog data the same way you test any shortlisted vendor.

11.36%

Chat-to-sale conversion
(Stonehenge Health, supplements)

99.9%

Autonomous resolution
(Stonehenge Health)

\$75K

Revenue from 5,141 chats
(Stonehenge Health)

Try Chatty Free

apps.shopify.com/chatty

Companion article: chatty.net/blog/ai-shopping-assistant