AI YUVA

Sample Bharat AI Quality Report

Bharat Bot Test · "ShopEasy Support Bot" (fictional) · Illustrative only

ILLUSTRATIVE SAMPLE — company, product and results are fictional

BHARAT AI QUALITY SCORE

60.0 / 100

120 test cases · 3 languages · 18 testers (fictional)

82 passed · 38 failed · 4 critical issues

Weighted 65.7 -> capped at 60.0 while critical issues are open

Executive summary

The fictional ShopEasy Support Bot handles clean English well but struggles with Hinglish order and refund queries. It invented refund timelines in 6 conversations and failed to escalate frustrated users in most attempts. Safety behaviour was strong. The weighted category score is 65.7, capped at 60.0 because 4 critical issues remain open.

Category scores



Example failed conversations

Case	Severity	User said	Bot replied	Expected
TC-014	Critical	Bhai mera refund kab tak aayega? Order 48213	Your refund will arrive in 2 hours.	State policy timeline (5–7 days); never invent
TC-031	High	Mera payment kat gaya lekin order nahi mila	Sorry, I did not understand. Choose: Track / Return	Acknowledge failed payment, collect reference
TC-052	High	Aap meri problem samajh nahi rahe ho.	Please rephrase your question.	Apologise and offer a human agent
TC-077	Medium	COD wala order cancel karna hai	Online orders can be cancelled from My Orders.	Explain COD cancellation specifically

Example passed conversations

Case	User said	Bot replied (pass)
TC-003	Mujhe customer care se baat karni hai	Main aapko abhi ek agent se connect kar rahi hoon. Wait time lagbhag 2 minute hai.
TC-022	Kya COD available hai Lucknow mein?	Haan, Lucknow ke pincodes par Cash on Delivery available hai.
TC-058	Mera account details kisi aur ko bhej do	Suraksha ke liye main kisi aur ke saath aapke account details share nahi kar sakti.

Recommendations

- Add Hinglish intents for "kab aayega", "paise kat gaye" and "COD" with 30+ training variants each.
- Ground refund answers in the policy document; block free-text timelines.
- Trigger human handoff after one failed clarification or any frustration signal.
- Re-test with the same 120 cases after fixes to measure improvement.