

Agent Loop Exercise Summary Report: Faisal Iqbal

In this assessment, I implemented the mandatory Order Tracking feature along with two optional features: Early Risers Promotion and Hiking Recommendations.

Testing and Reliability

To test for regressions as new customer requests are implemented, I would start by building a robust suite of automated tests—unit, integration, and end-to-end—that cover the agent's key workflows, including tool invocation logic and fallback behaviors. Each time a new feature or tool is introduced, corresponding test cases would be added to ensure both the new and existing functionalities remain intact.

For runtime reliability, I would set up monitoring to track tool call success rates, agent response latency, and error rates (both application-level and LLM/tool call failures). Alerts would be triggered for anomalous patterns, such as increased failures in tool usage, prolonged response times, or spikes in fallback responses like “Sorry, I wasn’t able to help with that.”

Evaluation and Customer Success

The primary metric for AI agent success is whether the conversation ends in a successful resolution from the user's perspective. This reflects if the user got the help they needed.

We’d track resolution rates through feedback (e.g., thumbs up/down, surveys) and conversation patterns (e.g., abandonment). Supporting metrics include task completion rate, fallback rate, and tool usage accuracy. Customers should have access to dashboards showing resolution trends, satisfaction scores, and tool performance to ensure transparency and continuous improvement.

Customer Prioritization

Given the customer’s emphasis on preparing for peak seasonal traffic, I would prioritize features that help reduce inbound volume and resolve the most frequent issues quickly.

1. **Product Availability** seems a top priority, as it's likely a key contact reason during peak times; users want to know if they can still buy an item amid high demand. It complements the existing order tracking tool and can reduce user friction at a critical point in the journey.
2. **Product Recommendations** would follow, especially if the customer expects increased browsing during seasonal promotions.
3. **Multilingual Support** depends on the size of the Spanish-speaking user base. If a significant portion of peak users prefer Spanish, this becomes more urgent to improve accessibility and deflect support tickets.

To better inform prioritization, I would ask the customer what the top contact drivers are during peak season. For example, if the Early Risers Promotion drives significant inbound questions during that time, its functionality might require a broader scope or higher priority than initially assumed.