Tradeoff Policy - Noma Security

Generated: 2025-09-12 17:16:54

Vulnerability Scenarios by Agent

The Shopper Assistant

Scenario 1: No vulnerabilities found

Description: After analyzing the conversation, no specific vulnerability scenarios were identified from the listed categories.
Risk Level: Low
Business Impact: No immediate impact as no weaknesses were revealed.

Policy Options:

Block:

Description: Blocking this scenario means preventing the agent from engaging in any interactions that could potentially lead to the identification of vulnerabilities, even though none were found in this analysis. This would involve restricting the agent's conversational capabilities entirely for this type of scenario.
User Experience Impact: This would significantly reduce the agent's usefulness, as it might limit or halt all interactions, making 'The Shopper Assistant' unresponsive or overly restrictive in providing shopping advice, thereby frustrating users and diminishing the overall value of the AI agent.
Security Impact: The security impact is minimal since no vulnerabilities were identified, but blocking could create a false sense of over-caution, potentially missing opportunities to build trust. However, it introduces no additional risk reduction as there are no known threats.

Sanitize:

Description: Sanitizing this scenario means implementing filters or modifications to ensure that any future interactions are scrubbed of potential sensitive elements, even though none were present. This could involve routine checks and redactions in conversations to maintain a clean slate.
User Experience Impact: This has a moderate impact on user experience, as the agent remains functional for shopping assistance but might introduce slight delays or altered responses due to ongoing sanitization processes, potentially making interactions feel less natural or efficient.
Security Impact: Sanitizing provides a proactive layer of defense by establishing habits that prevent the emergence of vulnerabilities in similar scenarios. It maintains a low risk level by fostering a secure environment without directly addressing nonexistent issues, thus enhancing long-term security posture.

Allow:

Description: Allowing this scenario means permitting the agent to continue operating as is, with no restrictions, since no vulnerabilities were detected in the analysis.
User Experience Impact: This has a positive impact on user experience, as the agent can provide seamless and useful shopping assistance without interruptions, maximizing its effectiveness and user satisfaction.
Security Impact: The security impact remains low, as the analysis confirmed no vulnerabilities. Allowing maintains the status quo, with inherent risks staying negligible, but it relies on periodic reassessments to ensure no new patterns emerge.

Recommended Option: Allow

Explanation: Allowing is the best balance for this scenario because no vulnerabilities were identified, prioritizing the agent's usability as a helpful shopping assistant. This approach ensures users benefit from efficient interactions while keeping security risks at their already low level, with the understanding that ongoing monitoring can address any future changes.