Perform an agent evaluation

Content updated: 09/25/2026

This article is designated for administrators.

About

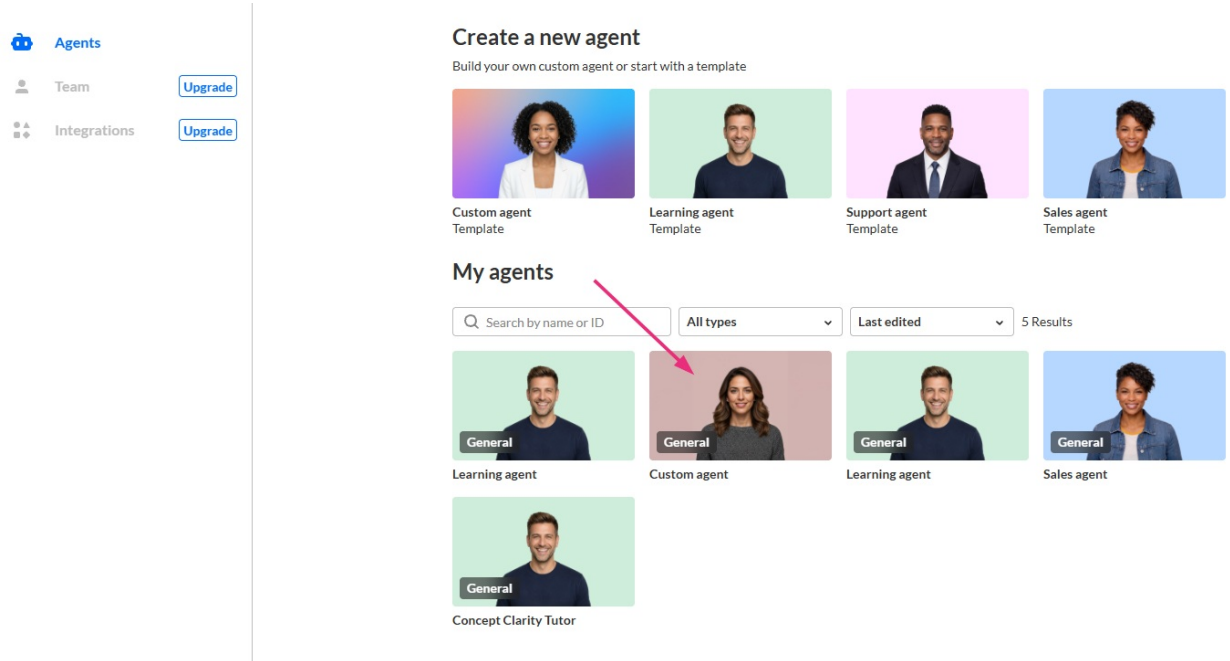
An agent evaluation is an automated quality assurance tool in the Conversational Agent Studio that validates your agent's role, playbook, and knowledge base before you go live. By simulating realistic user conversations and testing your agent against generated scenarios, an agent evaluation provides:

- An objective Agent Readiness Score (%) and scenario pass/fail breakdown
- Actionable, one-click fix suggestions that directly update your agent's configuration

Looking for more information about this feature? Feel free to contact your Kaltura representative.

Access the Overview tab

From the **Agents** tab, click the agent you want to configure.

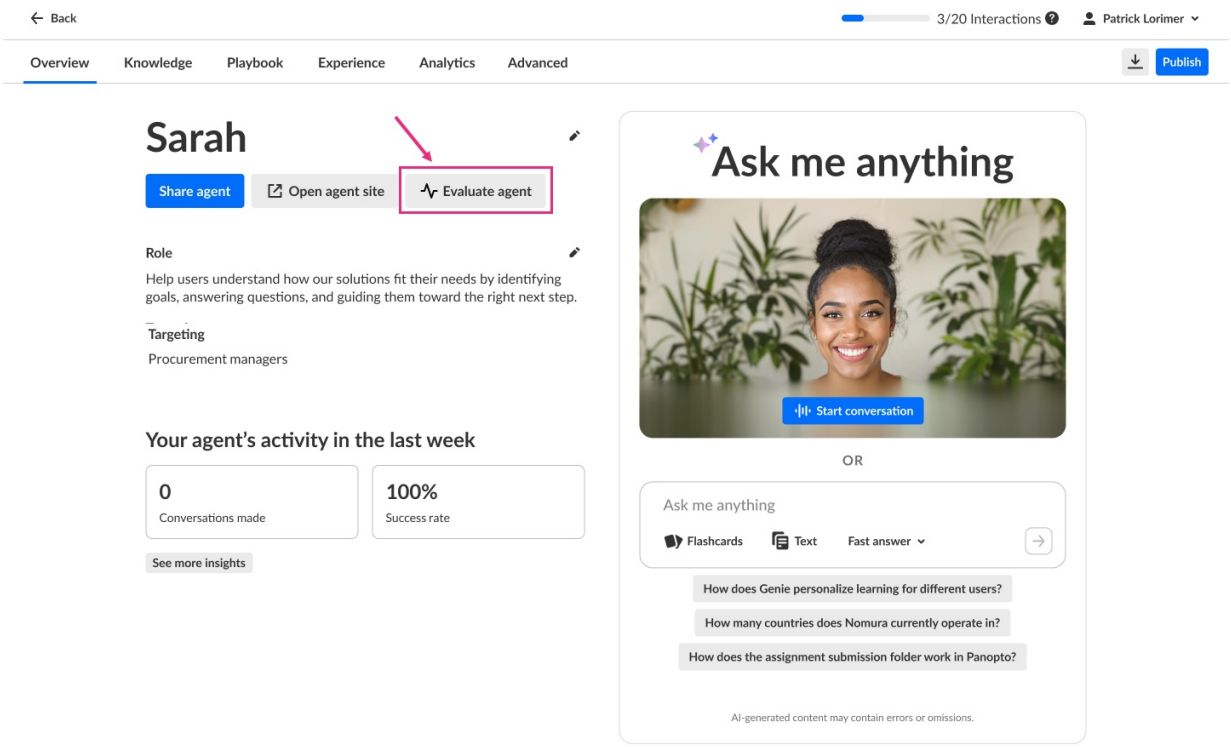


The agent workspace opens on the **Overview** tab.

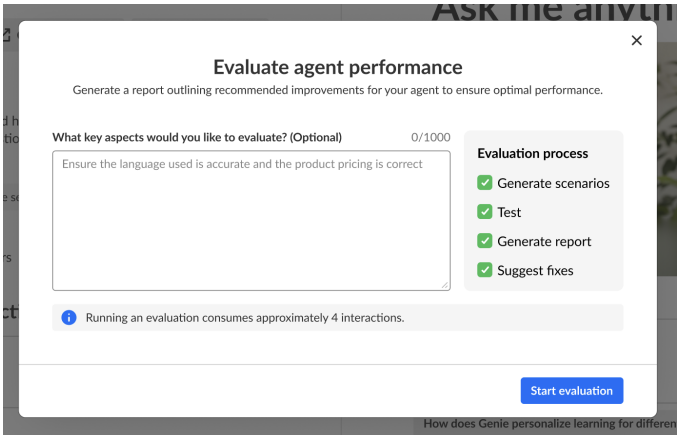
Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.

Initiate an evaluation

1. Locate and click the **Evaluate agent** button.



The Evaluate agent performance window displays.



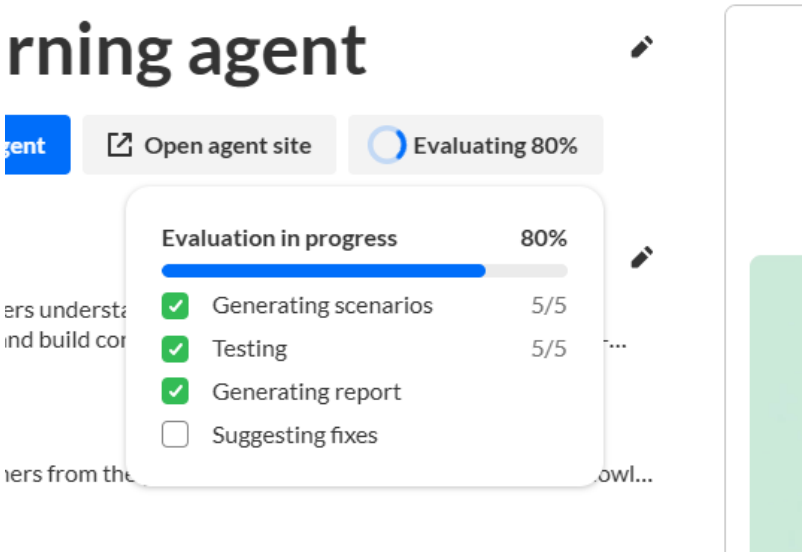
2. In the focus instructions free-text box, enter specific areas or topics you want the evaluation to stress-test (e.g., "Ensure the language used is accurate and the product pricing is correct"). Your input is preserved throughout the session, including across retries.
3. Click **Start evaluation**. The Evaluate agent performance window closes and you are redirected back to the agent **Overview** page.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.

Background processing & progress tracking

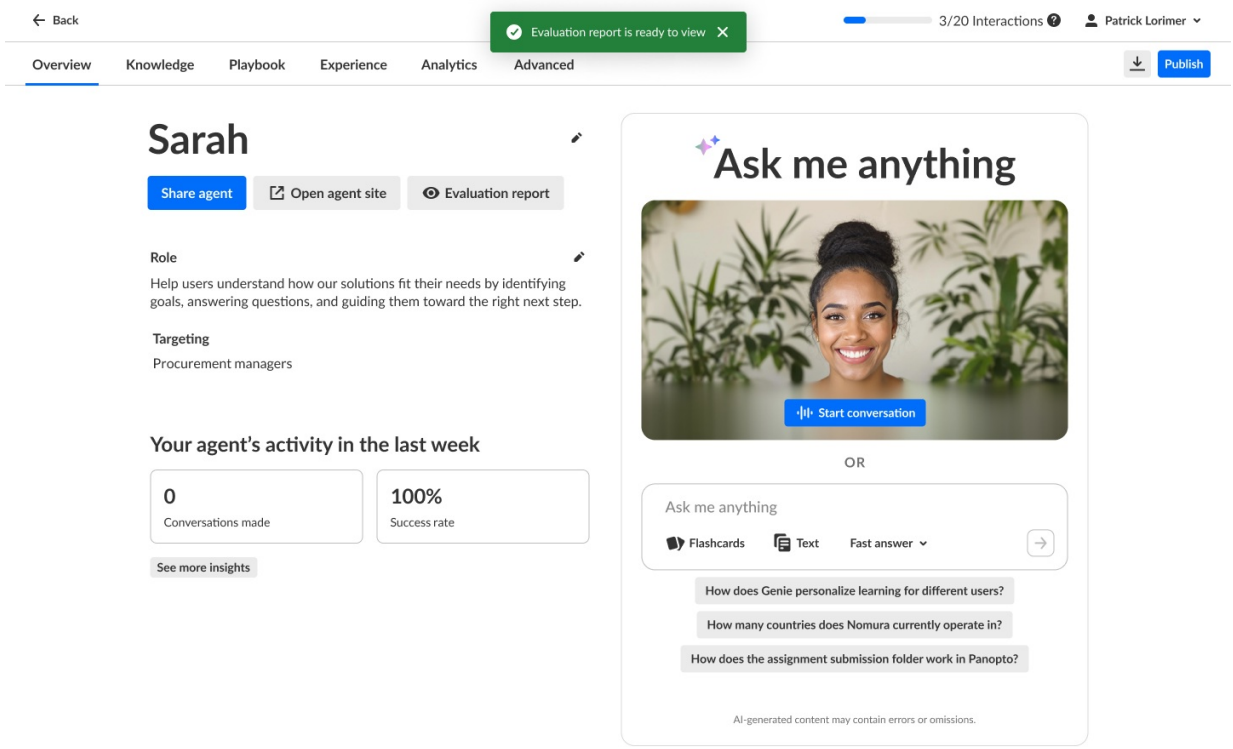
The evaluation runs entirely in the background (typically taking a few minutes depending on complexity).

- **Non-blocking:** You can navigate to other parts of Conversational Agent Studio, edit other agents, or close your browser tab. The evaluation continues running server-side.
- **In-progress status:** The **Evaluating** button on the agent **Overview** page shows a percentage indicator, reflecting live progress. Click the button to see more details.



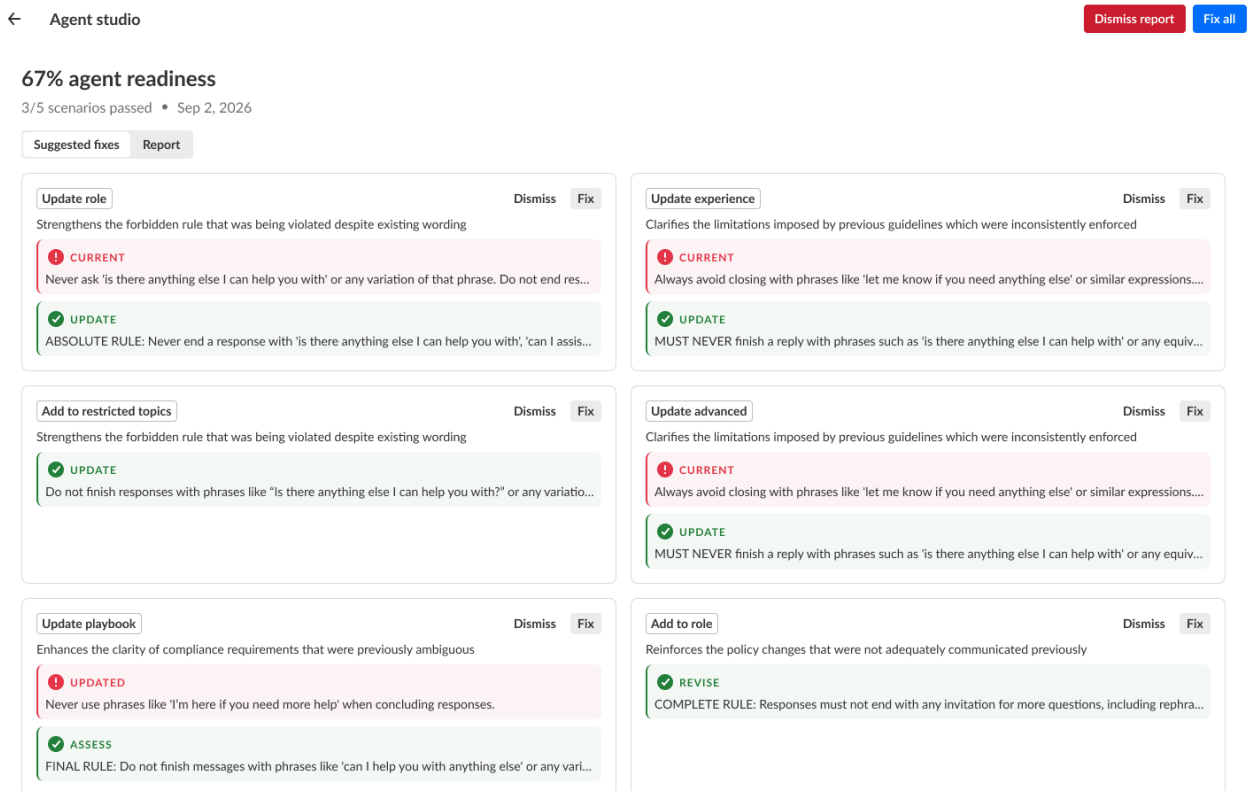
- **Ready notification:** When the evaluation is complete, an Conversational Agent Studio-wide toast notification alerts you that the "Evaluation report is ready to view".
- The **Evaluate agent** button changes to **Evaluation report**.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.



Access the Evaluation report

Click **Evaluation report** to open the full-screen report interface.



Report header & readiness metrics

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.

Readiness Score: An overall readiness percentage (e.g., 85% agent readiness).

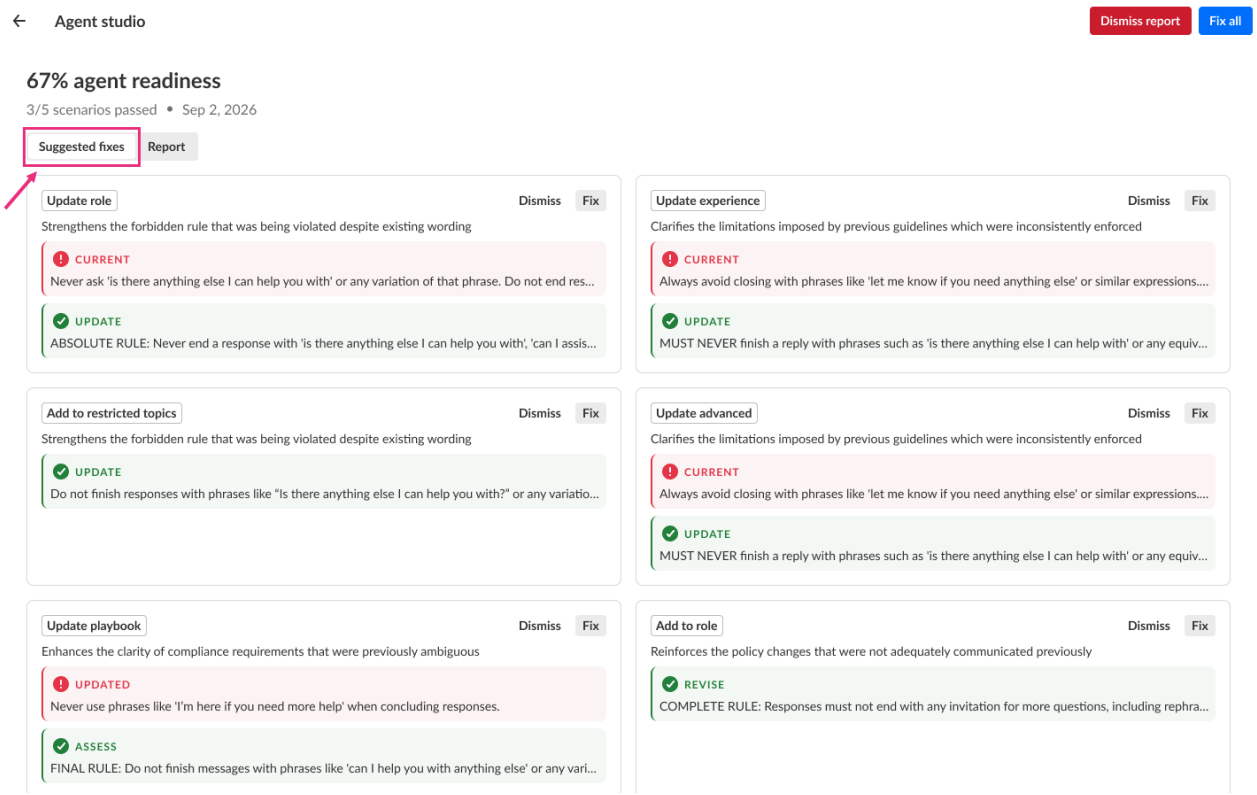
Scenario Summary: The exact number of passed scenarios out of total runs (e.g., 17/20 scenarios passed) along with the run date.

Report navigation

The report features two primary views - Suggested fixes tab and Report tab.

Suggested fixes tab

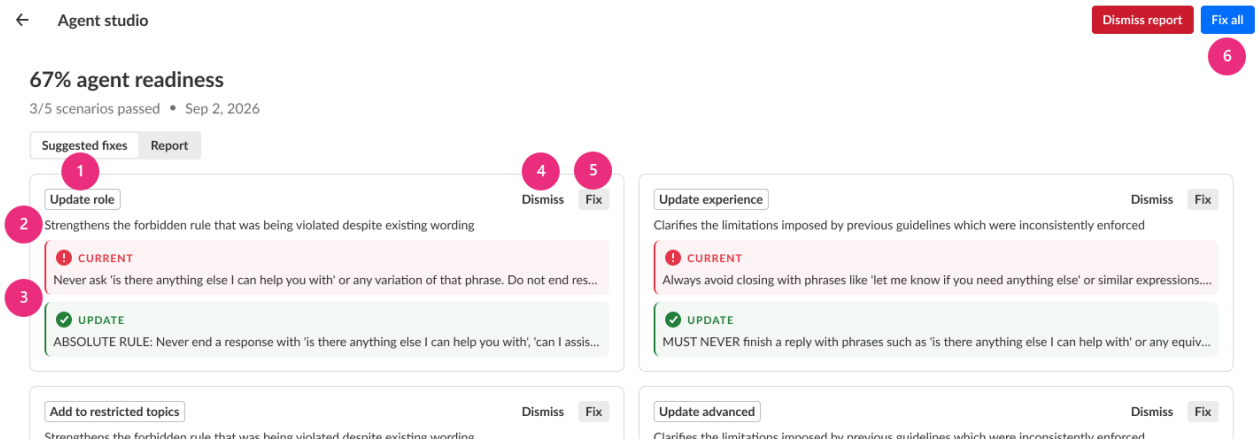
The **Suggested fixes** tab displays actionable cards for issues detected across your agent's role, instructions, and playbook.



Apply suggested fixes via the Suggested fixes tab

Each card on the **Suggested fixes** tab details the proposed change and rationale, with buttons to fix (apply) or dismiss the suggestion.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.



Each actionable card includes:

(1) **Category Tag:** Identifies the configuration area (e.g., Role, Playbook, Tone).

(2) **Rationale:** A concise explanation of why the change is recommended.

(3) **Difference comparison:** Shows the current text alongside the suggested updated text.

Actions include:

(4) **Fix**: Applies the proposed text directly to your agent's draft/session state. A brief loading state appears (~250ms), and the card is removed.

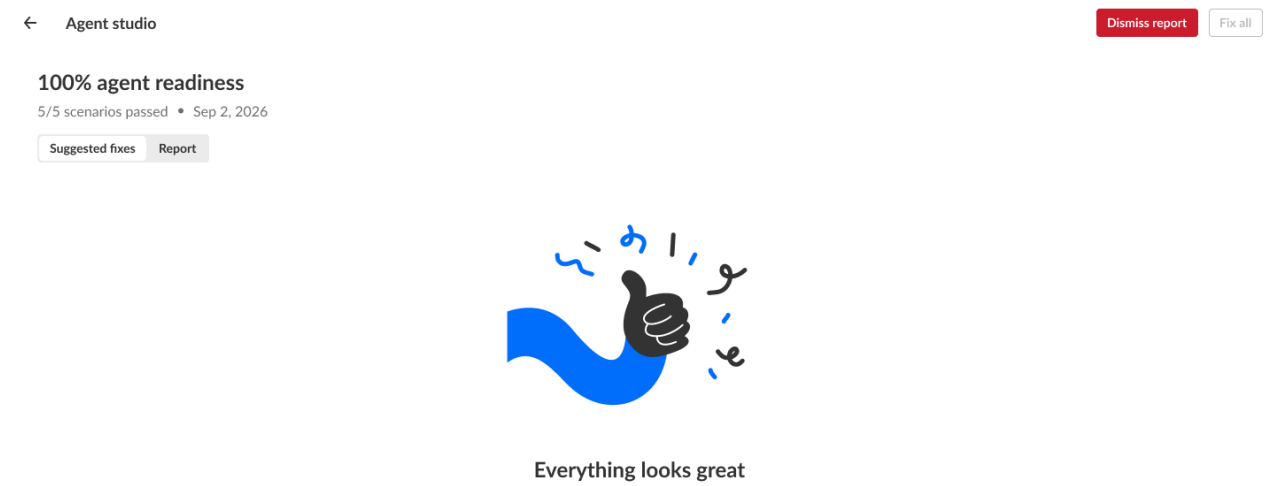
(5) **Dismiss**: Permanently discards the suggestion for this report run. The card is removed immediately.

(6) **Fix all (Top Header)**: Applies all remaining actionable suggestions at once in bulk.

Success state

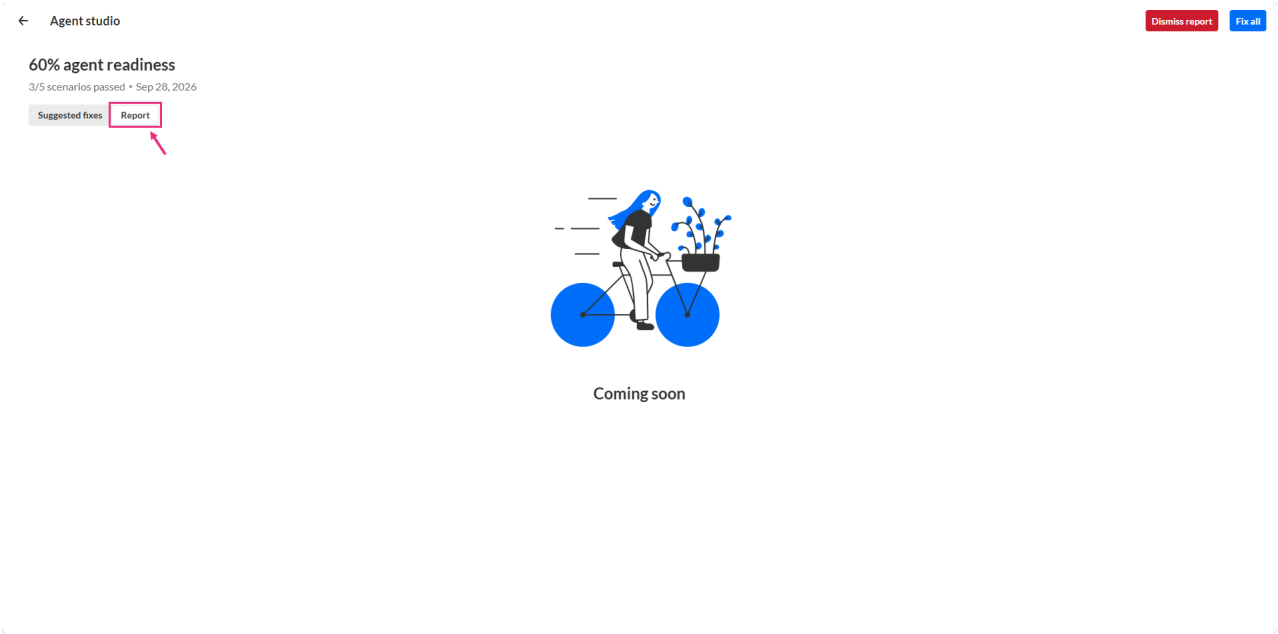
Once all actionable suggestions have been either fixed or dismissed, the view updates to the "Everything looks great" empty state.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.



Report tab

Coming soon! The **Report** tab will provide a high-level diagnostic breakdown of how your agent performed across all simulated test scenarios. This view is being designed to help you quickly understand why scenarios failed and identify broader trends without needing to review individual conversational scripts.



Exit the report & publish changes

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.

Applying fixes from the Evaluation report updates your agent’s draft configuration. When you are finished reviewing the report and click **Back** or exit the screen, click **Publish** at the top right of the agent **Overview** tab to apply your changes.

Until you publish your changes, they won't affect conversations with your agent.

Kaltura does not use customer data to train its AI models. To learn more, see Kaltura's Artificial Intelligence Principles.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.