

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 Agentic Avatar 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:

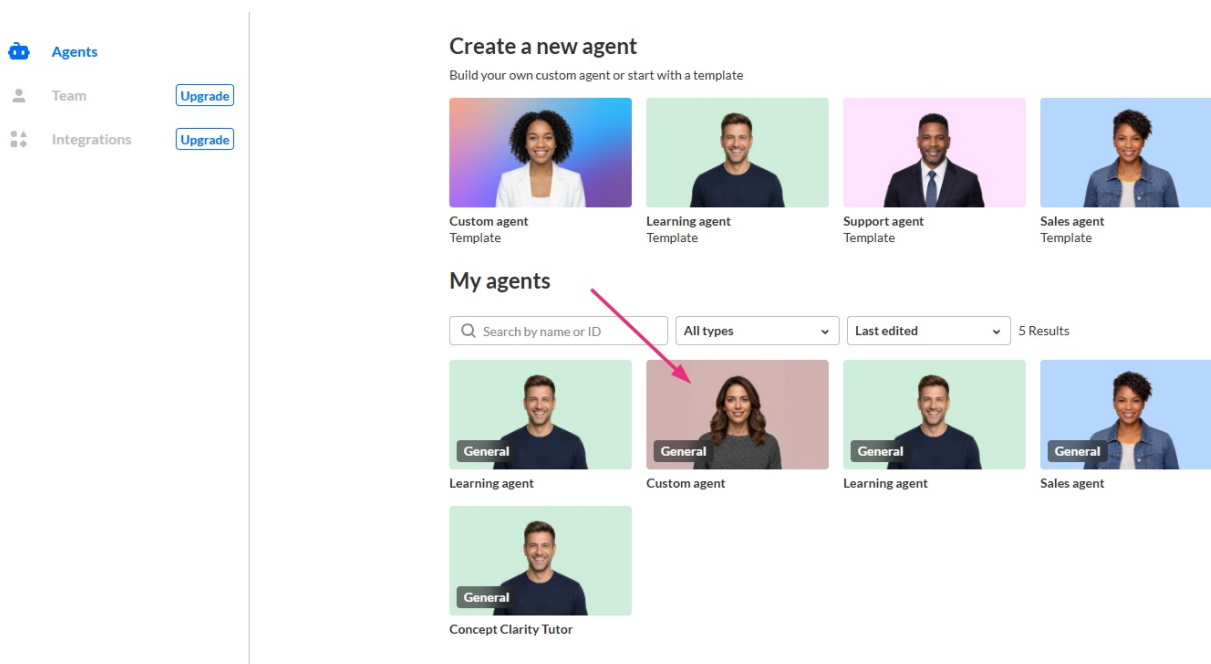
- An objective Agent Readiness Score (%) and scenario pass/fail breakdown
- Actionable, one-click fix suggestions that directly update your agent's configuration



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.

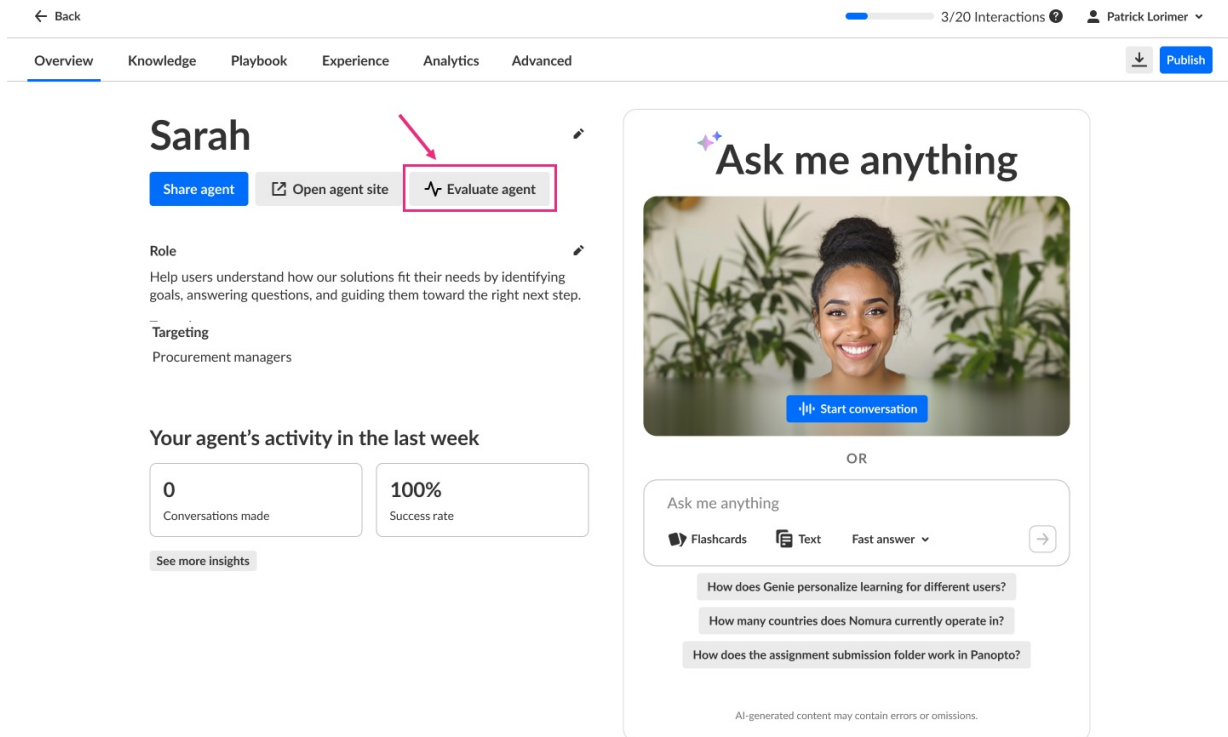


The screenshot displays the Kaltura Agents interface. On the left is a sidebar with navigation options: **Agents** (selected), **Team**, and **Integrations**, each with an **Upgrade** button. The main content area is divided into two sections. The top section, **Create a new agent**, includes the subtitle 'Build your own custom agent or start with a template' and four agent templates: **Custom agent Template**, **Learning agent Template**, **Support agent Template**, and **Sales agent Template**. The bottom section, **My agents**, features a search bar labeled 'Search by name or ID', dropdown menus for 'All types' and 'Last edited', and a '5 Results' indicator. Below these are five agent cards, each with a 'General' label: **Learning agent**, **Custom agent** (highlighted with a red arrow), **Learning agent**, **Sales agent**, and **Concept Clarity Tutor**.

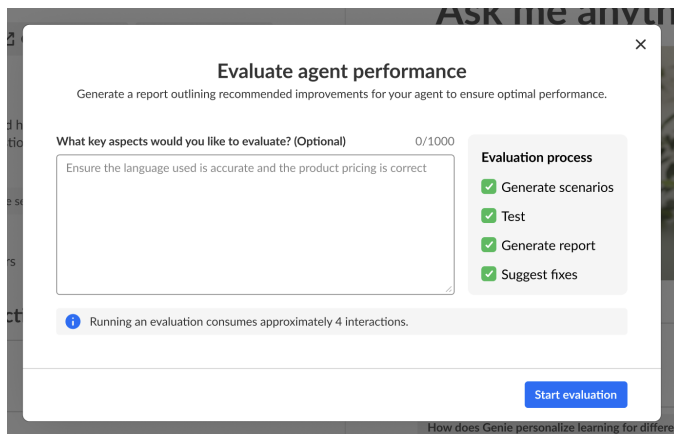
The agent workspace opens on the **Overview** tab.

Initiate an evaluation

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



The Evaluate agent performance window displays.



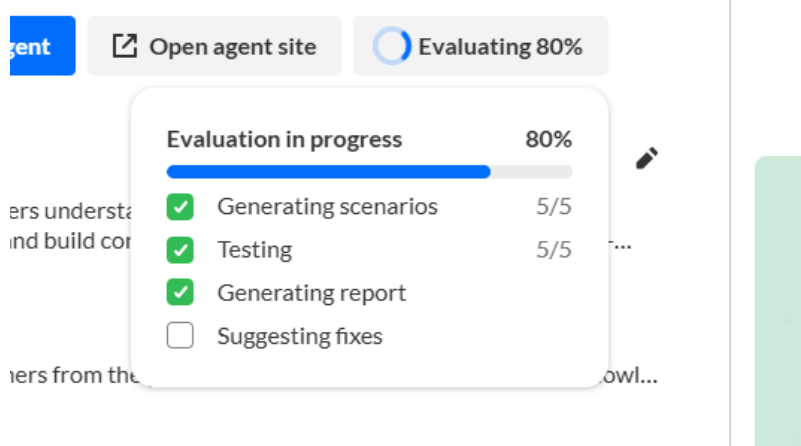
2. In the focus instructions free-text box, enter specific areas or topics you want the evaluation to stress-test (e.g., "Ensure the language used is accurate and the product pricing is correct"). Your input is preserved throughout the session, including across retries.
3. Click **Start evaluation**.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

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 Agentic Avatar 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.

Running agent



The screenshot shows the 'Running agent' interface. At the top, there is a blue button labeled 'agent' and a button labeled 'Open agent site'. To the right of these buttons is a status indicator: a blue circle with a white progress bar and the text 'Evaluating 80%'. Below this, a modal window titled 'Evaluation in progress' is displayed, showing a progress bar at 80% and a list of tasks:

Task	Progress
Generating scenarios	5/5
Testing	5/5
Generating report	5/5
Suggesting fixes	0/5

The tasks 'Generating scenarios', 'Testing', and 'Generating report' are marked with green checkmarks, indicating they are complete. 'Suggesting fixes' is marked with an empty checkbox, indicating it is not yet started.

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

← Back

✔ Evaluation report is ready to view
✕

3/20 Interactions ⓘ
Patrick Lorimer ▾

Overview
Knowledge
Playbook
Experience
Analytics
Advanced

Download
Publish

Sarah

Share agent
Open agent site
Evaluation report

Role

Help users understand how our solutions fit their needs by identifying goals, answering questions, and guiding them toward the right next step.

Targeting

Procurement managers

Your agent's activity in the last week

0

Conversations made

100%

Success rate

See more insights

Ask me anything

Start conversation

OR

Ask me anything

Flashcards
Text
Fast answer ▾

How does Genie personalize learning for different users?

How many countries does Nomura currently operate in?

How does the assignment submission folder work in Panopto?

AI-generated content may contain errors or omissions.

Access the Evaluation report

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

← Agent studio

Dismiss reportFix all

67% agent readiness
3/5 scenarios passed • Sep 2, 2026

Suggested fixes
Report

Update roleDismissFix

Strengthens the forbidden rule that was being violated despite existing wording

❗ CURRENT

Never ask 'is there anything else I can help you with' or any variation of that phrase. Do not end res...

✅ UPDATE

ABSOLUTE RULE: Never end a response with 'is there anything else I can help you with', 'can I assis...

Update experienceDismissFix

Clarifies the limitations imposed by previous guidelines which were inconsistently enforced

❗ CURRENT

Always avoid closing with phrases like 'let me know if you need anything else' or similar expressions...

✅ UPDATE

MUST NEVER finish a reply with phrases such as 'is there anything else I can help with' or any equiv...

Add to restricted topicsDismissFix

Strengthens the forbidden rule that was being violated despite existing wording

✅ UPDATE

Do not finish responses with phrases like "Is there anything else I can help you with?" or any variatio...

Update advancedDismissFix

Clarifies the limitations imposed by previous guidelines which were inconsistently enforced

❗ CURRENT

Always avoid closing with phrases like 'let me know if you need anything else' or similar expressions...

✅ UPDATE

MUST NEVER finish a reply with phrases such as 'is there anything else I can help with' or any equiv...

Update playbookDismissFix

Enhances the clarity of compliance requirements that were previously ambiguous

❗ UPDATED

Never use phrases like 'I'm here if you need more help' when concluding responses.

✅ ASSESS

FINAL RULE: Do not finish messages with phrases like 'can I help you with anything else' or any vari...

Add to roleDismissFix

Reinforces the policy changes that were not adequately communicated previously

✅ REVISE

COMPLETE RULE: Responses must not end with any invitation for more questions, including rephra...

Report header & readiness metrics

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

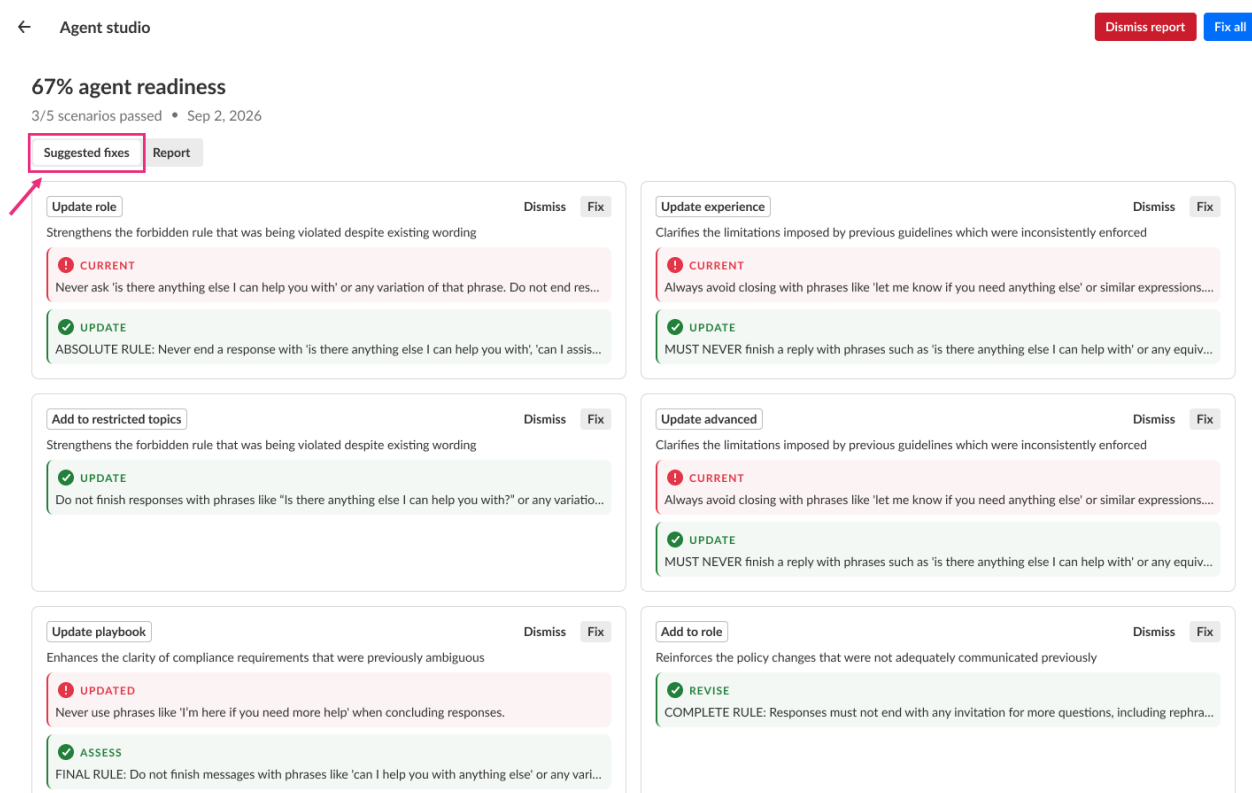
Scenario Summary: The exact number of passed scenarios out of total runs (e.g., 17/20 scenarios passed) along with the run date.

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.



← Agent studio Dismiss report Fix all

67% agent readiness
3/5 scenarios passed • Sep 2, 2026

Suggested fixes Report

Update role Dismiss Fix

Strengthens the forbidden rule that was being violated despite existing wording

! CURRENT
Never ask 'is there anything else I can help you with' or any variation of that phrase. Do not end res...

✓ UPDATE
ABSOLUTE RULE: Never end a response with 'is there anything else I can help you with', 'can I assis...

Update experience Dismiss Fix

Clarifies the limitations imposed by previous guidelines which were inconsistently enforced

! CURRENT
Always avoid closing with phrases like 'let me know if you need anything else' or similar expressions....

✓ UPDATE
MUST NEVER finish a reply with phrases such as 'is there anything else I can help with' or any equiv...

Add to restricted topics Dismiss Fix

Strengthens the forbidden rule that was being violated despite existing wording

✓ UPDATE
Do not finish responses with phrases like 'Is there anything else I can help you with?' or any variatio...

Update advanced Dismiss Fix

Clarifies the limitations imposed by previous guidelines which were inconsistently enforced

! CURRENT
Always avoid closing with phrases like 'let me know if you need anything else' or similar expressions....

✓ UPDATE
MUST NEVER finish a reply with phrases such as 'is there anything else I can help with' or any equiv...

Update playbook Dismiss Fix

Enhances the clarity of compliance requirements that were previously ambiguous

! UPDATED
Never use phrases like 'I'm here if you need more help' when concluding responses.

✓ ASSESS
FINAL RULE: Do not finish messages with phrases like 'can I help you with anything else' or any vari...

Add to role Dismiss Fix

Reinforces the policy changes that were not adequately communicated previously

✓ REVISE
COMPLETE RULE: Responses must not end with any invitation for more questions, including repbra...

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.

67% agent readiness

3/5 scenarios passed • Sep 2, 2026

Suggested fixes Report

1

Update role

4

5

2

Strengthens the forbidden rule that was being violated despite existing wording

3

! CURRENT

Never ask 'is there anything else I can help you with' or any variation of that phrase. Do not end res...

✓ UPDATE

ABSOLUTE RULE: Never end a response with 'is there anything else I can help you with', 'can I assis...

Dismiss

Fix

Update experience

Dismiss

Fix

Clarifies the limitations imposed by previous guidelines which were inconsistently enforced

! CURRENT

Always avoid closing with phrases like 'let me know if you need anything else' or similar expressions...

✓ UPDATE

MUST NEVER finish a reply with phrases such as 'is there anything else I can help with' or any equiv...

Dismiss

Fix

Add to restricted topics

Dismiss

Fix

Strengthen the forbidden rule that was being violated despite existing wording

Update advanced

Dismiss

Fix

Clarifies the limitations imposed by previous guidelines which were inconsistently enforced

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.

100% agent readiness

5/5 scenarios passed • Sep 2, 2026

Suggested fixes Report



Everything looks great

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.

60% agent readiness

3/5 scenarios passed • Sep 28, 2026

Suggested fixes Report



Coming soon

Exit the report & publish changes



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](#).