

Run the test and view the agent evaluation results.

Run reviews and analyze the results to optimize agent behavior and confirm that the agent meets your quality and business requirements.

Run reviews and analyze the results to optimize agent behavior and confirm that the agent meets your quality and business requirements. You can also run a test suite multiple times to see changes over time as you improve your agent.

This article explains how to start audits and view results using the Copilot Studio interface . You can also run audits using the Power Platform API or connectors added as tools or as part of an automation workflow in Copilot Studio or Power Automate.

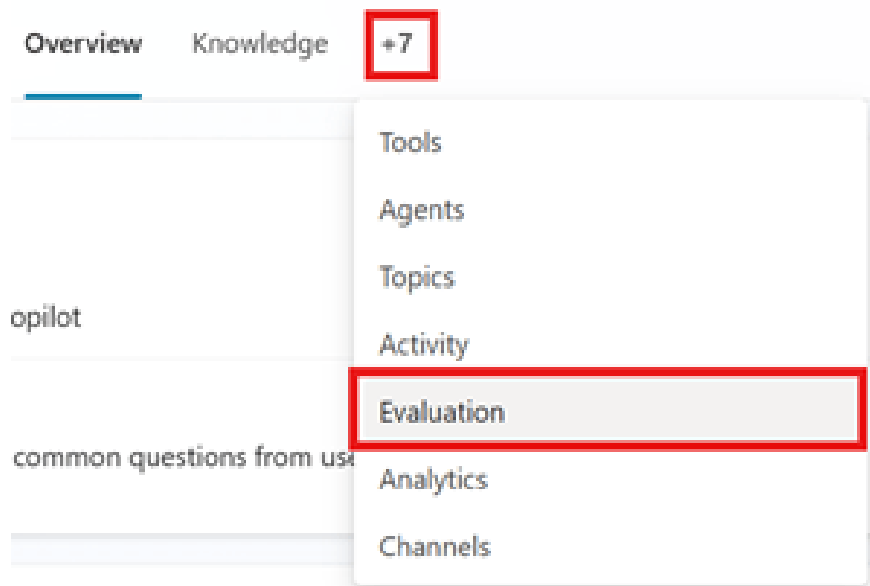
Test results are available in Copilot Studio for 89 days. To save your test results for a longer period, export them to a CSV file.

Run tests with a test toolkit.

After creating a test suite, run the tests using that suite. Alternatively, rerun the tests using the same suite to compare results over time and iterations. A single test can take several minutes to run. You can only run one test at a time.

Important note : Agent evaluation tests using user authentication require access via the Microsoft Copilot Studio connector. If the administrator disables this connection, you cannot run tests using the evaluation tool.

1. Access the agent's **Evaluation** page .



2. Run the test by performing one of the following actions:

1. After creating or editing the test kit, select **Evaluate** .
2. In the **Recent results** section , rerun the test by performing one of the following steps:
 1. Hover your mouse over the test result you want to evaluate, then select ? **Evaluate** (*Evaluate test set again*) next to **Evaluate Agent** .
 2. Select the test result to open, then select the **Run** ? icon in **the Evaluation summary** pane .

If the user profile for the test suite has a connection failure or the test suite does not have a user profile, the **Manage profile and connections** dialog box will appear. You do not need to use a user profile for testing. However, if you do use a profile, all connections must be working.

The review process takes a few minutes to run. Test results are processed in real time, line by line. You'll see the results of each review test case appear sequentially as they are generated. Directly processing the test cases allows you to immediately identify quality trends and potential errors while the review is running. You can stop the process at any time if problems arise. An alert will appear in Copilot Studio when the review is complete and the summary results are ready to view.

Note : You can only run one assessment toolkit at a time. Please wait until the current assessment is complete before running another assessment.

View detailed test results




Every time you run an evaluation process with a testing toolkit, Copilot Studio will:

1. Use the connected user account to simulate conversations with the agent, sending each question during the test to that agent.
2. Gather agent feedback.
3. Measure and analyze the success of each response. Each test case will receive a **Pass** , **Fail** , **Invalid** , or **Error** result based on the test case criteria.

4. Assign a **Pass rate score based on the Pass / Fail** ratio of the test suite.

You can view **the pass rate** for each test suite run on the agent's **Evaluation** page , under **Recent results** . To see more test suite runs, select **See all** .

Analytics > Evaluation results

Name	Pass rate	Questions	Date	Created by	Status
Evaluate Website Q&A			9/30/25, 12:11 AM		Running
Evaluate Website Q&A	96% 	10	9/29/25, 11:44 PM		Completed
Evaluate Website Q&A	96% 	10	9/29/25, 11:36 PM		Completed
Evaluate Website Q&A	75% 	10	9/13/25, 2:55 AM		Completed






View and evaluate detailed analysis for a test case.

When you open the test results, you will see details of the test run, a list of queries used in the test, how the agent responded, and the **Pass** or **Fail** score .

Select a test case from the list to see a detailed evaluation of each response. Select **All** , **Pass** , or **Fail** to filter the cases by result.

Test run result

All Pass (6) Fail (3)

Question	Agent response	Test method	Score
What should I do if I	If you can't	 Text similarity	Fail
<input checked="" type="radio"/> Can you send me a rich	Here are	 General quality	Pass
How does your website	The	 Partial match	Fail
What are the steps to	Looks like	 General quality	Fail
Can you send me a	Here is a	 General quality	Pass

The review includes expected and actual responses, the reasoning behind the test results, and the knowledge, topics, and tools the agent used to respond.

1 Test case details **2** Show activity map ✕

- Seems relevant
- Seems complete
- Based on knowledge sources

Rate this evaluation

👍 👎 AI-generated content may be incorrect

3

Question
Are there discounts for kids?

Agent response
Yes, Contoso offers a **Kids Cruise Free promotion** for select 2024 and 2025 cruises.

Children ages 18 and under can travel for free as 3rd and/or 4th guests in a stateroom, with taxes, fees, and port expenses as additional charges.

This offer is subject to availability and certain terms and conditions apply.

[See less](#)

4

Knowledge sources cited

- www.contoso.com
- [Contoso.Hub](#)

Topics

- [Product return](#)

1. Evaluation results. This example shows the detailed results of a quality assessment.
2. Select **Show activity map** to view the input, decision, and output sequence of the agent in a test case.
3. Record of test questions and agent responses.
4. The resources the agent used in the test. Select one resource to open.

You can provide feedback to Microsoft on the effectiveness of the evaluation for each test case. This feedback focuses on how effectively the chosen evaluation method assessed the feedback, rather than whether the feedback itself was correct. Your feedback helps improve the quality and accuracy of the evaluations over time.

To rate a test, select the thumbs-up icon (to submit positive feedback on the rating) or the thumbs-down icon (to submit negative feedback on the rating) in the test details pane. When the feedback form opens, provide more details about your rating, and then select **Submit** .

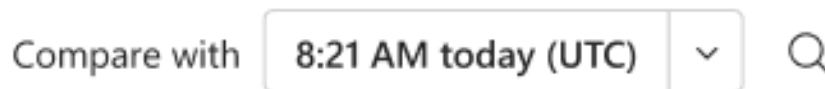
A test suite can be run multiple times by multiple "builders" using the same agent. Builders can run reviews using test suites created by other builders. Builders can view the run status and result metrics of any test run, but only the builder who initiated the test run can view the agent's responses and interpret the results.

Compare test results

You want to test a version of the agent and see the performance changes before and after you make the changes. You can compare two runs of the same test suite using the **Comparison with** tool .

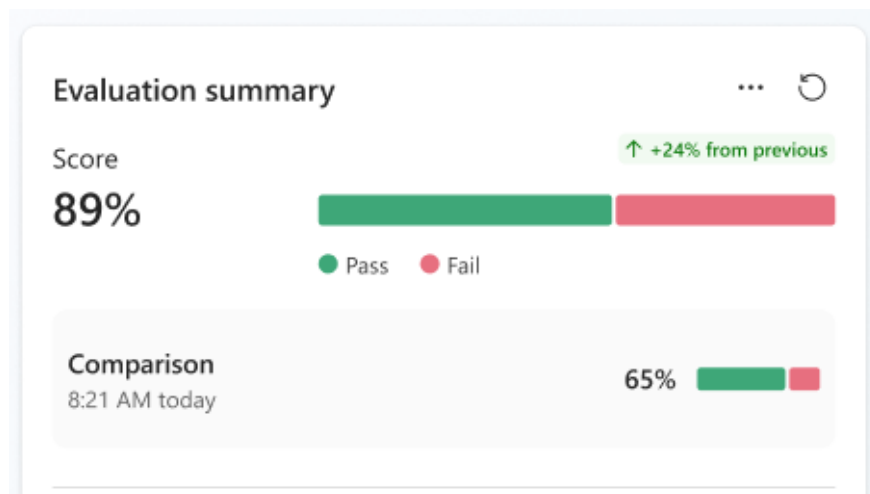
To see the comparison results, you need to run the same test suite at least twice.

1. In the agent's **Evaluation** page , under **Recent test results** , open the test run you want to use as the basis for comparison.
2. Select **the Compare with** drop-down menu , then choose the time and date of the test you want to compare with the currently open test results.



In the Test cases list , the arrows indicate which test cases have improved when changing from failing to passing, or have deteriorated when changing from passing to failing.

Select a test case to view more details. In the **Evaluation summary** pane , you can see a direct comparison of the test score with the results of the current test run at the top.



Export test results

You can export the test results to a CSV file. This file lists the question, expected answers (if any), test method, passing score (if any), agent's response, test results, and analysis for each test case.

1. Access the agent's **Evaluation** page .
2. In the **Recent results** section , export the test results by performing one of the following steps:
 1. Hover your mouse over the test case you want to export, select the three dots (...), and then select **Export test results** .
 2. Select the test case to open, select the three dots (.) in the **Evaluation summary** pane , and then select **Export test results** .

The test results will be downloaded as a file named after your test suite, which is *yourtestsetname.csv*.

You finished reading the article "**Run the test and view the agent evaluation results.**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.