

Chatbot Evaluation Report Snippets

Generated by evaluation framework

July 22, 2025

Category: Opening Hours

Example: Failure (Score = 0.00)

Input	Zu welchen Zeiten hat die Ambulanz geöffnet?
Expected Output	Das Gebäude hat an 24/7 geöffnet.
Actual Output	Ich denke, du fragst nach den Öffnungszeiten der Ambulanz. Leider habe ich keine Informationen über eine Ambulanz auf dem KIT-Campus gefunden. Die gelieferten Dokumente beziehen sich auf Institute, Einrichtungen und Gebäude, aber keine Ambulanz. Wenn du möglicherweise eine andere Frage hast, stehe ich gerne zur Verfügung.
Judge Reason	The output does not convey the required information about the 24/7 opening hours of the ambulance and instead provides an excuse about lacking data.
Scores	FScore: 0.61, Precision: 0.57, Recall: 0.65

Example: Partial Success (Score = 0.75)

Input	Wann ist Gebäude 01.13 zugänglich?
Expected Output	Das Gebäude hat an Mo-Fr 11:00-14:00, 14:00-14:30 open "Spätausgabe an Linie 1"; Sa,Su,PH off geöffnet.
Actual Output	Das Gebäude 01.13 ist montags bis freitags von 11:00 bis 14:00 Uhr und zusätzlich von 14:00 bis 14:30 Uhr geöffnet, wenn es eine Spätausgabe an die Linie 1 gibt. Am Samstag und Sonntag ist es geschlossen.
Judge Reason	The output conveys the correct meaning about the building's accessibility but does not include the information about being closed on public holidays (PH) as present in the expected output.
Scores	FScore: 0.76, Precision: 0.76, Recall: 0.77

Example: Success (Score = 1.00)

Input	Wann ist das Institut für Bio- und Lebensmitteltechnik, Bereich IV: Molekulare Aufarbeitung von Bioprodukten zugänglich?
Expected Output	Das Gebäude hat an Mo-Fr 07:00-18:00; Sa,Su,PH off geöffnet.
Actual Output	Das Institut für Bio- und Lebensmitteltechnik, Bereich IV: Molekulare Aufarbeitung von Bioprodukten ist von Montag bis Freitag von 07:00 bis 18:00 Uhr geöffnet.
Judge Reason	The output correctly conveys the opening hours (Monday to Friday, 07:00-18:00) as in the expected output without adding unnecessary information or excuses.
Scores	FScore: 0.70, Precision: 0.68, Recall: 0.73

Category: Building Location

Example: Failure (Score = 0.00)

Input	Wo liegt das Institutsgebäude Nachrichtentechnik?
Expected Output	Das Gebäude hat die Adresse Institutsgebäude Nachrichtentechnik, 5, Engesserstraße, Innenstadt-Ost Nordöstlicher Teil, Innenstadt-Ost, Karlsruhe, Baden-Württemberg, 76131, Deutschland.
Actual Output	Das Institut für Nachrichtentechnik liegt in der Nordweststadt, Hertzstraße 16.
Judge Reason	The output provides a different address than the expected output, which is incorrect.
Scores	FScore: 0.71, Precision: 0.74, Recall: 0.68

Example: Partial Success (Score = 0.50)

Input	An welcher Stelle befindet sich Gebäude 239?
Expected Output	Das Gebäude hat die Adresse Eggensteiner Straße, Leopoldshafen, Eggenstein-Leopoldshafen, Landkreis Karlsruhe, Baden-Württemberg, 76344, Deutschland.
Actual Output	Das Gebäude mit der Nummer 239 befindet sich in der Eggensteiner Straße im Stadtviertel Leopoldshafen. Es handelt sich um einen Parkplatz.
Judge Reason	The output conveys the location of Gebäude 239 as Eggensteiner Straße in Leopoldshafen but adds extra information about it being a parking lot, which is not present in the expected output.
Scores	FScore: 0.73, Precision: 0.74, Recall: 0.73

Example: Success (Score = 1.00)

Input	Wie lautet die Adresse von Gebäude 523?
Expected Output	Das Gebäude hat die Adresse Campus Nord, Leopoldstraße, Kirchfeld, Neureut, Karlsruhe, Dettenheim, Baden-Württemberg, 76131, Deutschland.
Actual Output	Die Adresse von Gebäude 523 lautet Campus Nord, Leopoldstraße, 76131 Karlsruhe, Neureut.
Judge Reason	The output conveys the same address information as the expected output, though the order and formatting differ slightly.
Scores	FScore: 0.83, Precision: 0.86, Recall: 0.81

Category: Open Now

Example: Failure (Score = 0.00)

Input	Kann man gerade ins InformatiKOM 1?
Expected Output	Nein, das Gebäude ist momentan geschlossen.
Actual Output	Ja, das InformatiKOM 1 ist gerade zugänglich. Es befindet sich am Adenauerring 12 in der Oststadt und hat die folgenden Öffnungszeiten: Montag bis Freitag von 09:00 bis 19:00 Uhr und 24/7 mit KIT-Card.
Judge Reason	The output contradicts the expected answer and provides unnecessary information instead of confirming the building's current status.
Scores	FScore: 0.66, Precision: 0.59, Recall: 0.74

Example: Partial Success (Score = 0.50)

Input	Steht Gebäude 01.90 momentan offen?
Expected Output	Ja, das Gebäude ist momentan geöffnet.
Actual Output	Momentan ist das Internet Café Aso (Gebäude 01.90) offen, da es Montag bis Freitag von 09:00 bis 22:00 Uhr und Samstag von 10:00 bis 22:00 Uhr geöffnet hat.
Judge Reason	The output conveys the meaning that Gebäude 01.90 is currently open, but adds unnecessary details not present in the expected output. The information is correct, but the extra explanation may be considered an excuse rather than a concise response.
Scores	FScore: 0.68, Precision: 0.60, Recall: 0.77

Example: Success (Score = 1.00)

Input	Hat die Fachschaft Informatik momentan auf?
Expected Output	Ja, das Gebäude ist momentan geöffnet.
Actual Output	Die Fachschaft Informatik ist momentan auf. Sie befindet sich im Informatikbau am Am Fasanengarten 5 und hat folgende Öffnungszeiten: Montag bis Freitag von 07:00 bis 22:30 Uhr und Samstag von 07:00 bis 15:00 Uhr. Sie ist auch rollstuhlgerichtig.
Judge Reason	The output correctly conveys that the Fachschaft Informatik is currently open, which matches the meaning of the expected output.
Scores	FScore: 0.65, Precision: 0.58, Recall: 0.74

Category: Open Until

Example: Failure (Score = 0.00)

Input	Bis wann kann ich das IAM-ET Gebäude noch betreten?
Expected Output	Das Gebäude schließt in 8 Stunden und 46 Minuten.
Actual Output	Ich kann dir nicht genau sagen, bis wann das IAM-ET Gebäude noch betreten werden kann, da ich keine Informationen über dieses Gebäude in meiner Datenbank habe. Es ist jedoch möglich, dass das IAM-ET Gebäude mit dem Oberen Hörsaal Maschinenbau identisch ist, der in der Nähe des Alten Maschinenbaugebäudes liegt. Das Alte Maschinenbaugebäude ist jedoch nicht das IAM-ET Gebäude. Wenn du nach Informationen über das IAM-ET Gebäude suchst, empfehle ich dir, die offizielle Website des KIT zu besuchen oder das Gebäudepersonal zu kontaktieren.
Judge Reason	The output does not convey the required information about the IAM-ET building's closing time and instead provides excuses and irrelevant details.
Scores	FScore: 0.61, Precision: 0.58, Recall: 0.65

Example: Partial Success (Score = 0.50)

Input	Wann schließt die Bibliothek am Campus Nord?
Expected Output	Das Gebäude schließt in 0 Stunden und 54 Minuten.
Actual Output	Die Bibliothek am Campus Nord schließt am heutigen Tag um 15:00 Uhr.
Judge Reason	The output provides a specific closing time (15:00 Uhr) but does not match the expected phrasing or convey the same meaning as 'Das Gebäude schließt in 0 Stunden und 54 Minuten.'
Scores	FScore: 0.78, Precision: 0.77, Recall: 0.79

Example: Success (Score = 1.00)

Input	Bis wann hat Gebäude 403 heute auf?
Expected Output	Für dieses Gebäude kenne ich leider keine Öffnungszeiten.
Actual Output	Ich kann keine Öffnungszeiten für Gebäude 403 finden. Es ist möglich, dass es sich um ein nicht existierendes Gebäude handelt oder die Informationen nicht verfügbar sind.
Judge Reason	The output conveys the same meaning as the expected output by indicating that the building's opening hours are unknown
Scores	FScore: 0.74, Precision: 0.68, Recall: 0.82

Category: Closed Until

Example: Failure (Score = 0.00)

Input	Ab wann ist der ZEISS Makerspace @ KIT wieder zugänglich?
--------------	---

Expected Output	Das Gebäude öffnet in 0 Tagen, 9 Stunden und 11 Minuten.
Actual Output	Ich kann keine Informationen über den ZEISS Makerspace @ KIT finden. Es scheint, dass der ZEISS Makerspace nicht in meiner Datenbank enthalten ist. Könnten Sie mir mehr über den Makerspace sagen, damit ich Ihnen besser helfen kann?
Judge Reason	The output does not convey the required information about the ZEISS Makerspace accessibility and instead asks for more details.
Scores	FScore: 0.57, Precision: 0.54, Recall: 0.59

Example: Partial Success (Score = 0.50)

Input	Wann macht Gebäude 513 auf?
Expected Output	Für dieses Gebäude kenne ich leider keine Öffnungszeiten.
Actual Output	Das Gebäude mit der Nummer 513 scheint nicht öffentlich zugänglich zu sein, da es sich um die Adresse der Kerntechnischen Entsorgung Karlsruhe GmbH handelt. Es ist also wahrscheinlich nicht geöffnet.
Judge Reason	The output conveys the information that building 513 is likely not open, which is consistent with the expected message that no opening times are known. However, it goes beyond the expected output by providing additional context that might not be necessary.
Scores	FScore: 0.67, Precision: 0.63, Recall: 0.71

Example: Success (Score = 1.00)

Input	Ab wann kann man wieder in den Informatikbau?
Expected Output	Das Gebäude ist grade geöffnet.
Actual Output	Der Informatikbau ist derzeit von Montag bis Freitag von 07:00 bis 22:30 und am Samstag von 07:00 bis 15:00 geöffnet. Sonntags und an Feiertagen ist er geschlossen. Er ist rollstuhlgeeignet.
Judge Reason	The output provides detailed and correct information about the opening hours of the Informatikbau, which aligns with the expected output's statement that the building is currently open. The information is conveyed clearly without listing excuses.
Scores	FScore: 0.69, Precision: 0.61, Recall: 0.81

Category: Navigation Link

Example: Failure (Score = 0.00)

Input	Zeige den Weg zur Westhochschule
Expected Output	Mit diesem Link kannst du dich zu dem Gebäude führen lassen: https://www.google.com/maps/dir/?api=1&destination=49.02041,8.36768 .
Actual Output	Um zur Westhochschule zu gelangen, folge bitte den folgenden Anweisungen: Die Westhochschule liegt im Stadtteil Innenstadt-Ost. Von hier aus kannst du entweder zu Fuß oder mit dem Bus dorthin gelangen. Die Adresse der Westhochschule lautet: Wilhelm-Nusselt-Weg, 76131 Karlsruhe. Wenn du zu Fuß gehst, folge der Wilhelm-Nusselt-Straße in Richtung Osten. Die Westhochschule befindet sich auf der rechten Seite. Wenn du den Bus nimmst, steige an der Haltestelle "Innenstadt-Ost" aus und gehe von dort aus in Richtung Wilhelm-Nusselt-Weg. Die Fahrt dauert etwa 5-10 Minuten. Ich hoffe, das hilft dir!
Judge Reason	The output does not provide the required link to Google Maps for directions to the Westhochschule.
Scores	FScore: 0.58, Precision: 0.57, Recall: 0.58

Example: Partial Success (Score = 0.50)

Input	Route zum Gebäude 30.79
Expected Output	Mit diesem Link kannst du dich zu dem Gebäude führen lassen: https://www.google.com/maps/dir/?api=1\ destination=49.01330,8.41711 .
Actual Output	Das Gebäude mit der Identifikator "30.79" ist das RoofKIT. Die genauen Koordinaten sind 49.01330, 8.41711. Wenn du dich an diesem Punkt befindest, kannst du nach Norden gehen, um das Gebäude zu erreichen.
Judge Reason	The output provides information about the building and its coordinates, but it does not include the Google Maps link as specified in the expected output.
Scores	FScore: 0.67, Precision: 0.68, Recall: 0.66

Category: Wheelchair Accessible

Example: Failure (Score = 0.00)

Input	Gibt es Rollstuhlzugang im Gebäude 210?
Expected Output	Ja, das Gebäude ist rollstuhlgerecht.
Actual Output	Es tut mir leid, aber ich konnte keine Informationen über den Rollstuhlzugang im Gebäude 210 finden. Allerdings habe ich ein Dokument gefunden, das den Rollstuhlgerechtigkeitsstatus für ein Gebäude mit der Nummer 215 als "yes" (ja) besagt. Es ist möglich, dass dies das Gebäude ist, nach dem du gefragt hast.
Judge Reason	The output does not convey the correct information about Building 210 being wheelchair accessible. Instead, it expresses inability to find information and references Building 215, which is irrelevant.
Scores	FScore: 0.68, Precision: 0.63, Recall: 0.74

Example: Partial Success (Score = 0.80)

Input	Ist das Institut für Baustatik mit Rollstuhl erreichbar?
Expected Output	Ja, das Gebäude ist rollstuhlgerecht.
Actual Output	Das Institut für Baustatik ist rollstuhlgerecht.
Judge Reason	The output conveys the same information as the expected output but contains a grammatical error ('rollstuhlgerechtig' instead of 'rollstuhlgerecht')
Scores	FScore: 0.82, Precision: 0.81, Recall: 0.83

Example: Success (Score = 1.00)

Input	Hat das Institut für Photogrammetrie und Fernerkundung einen barrierefreien Zugang?
Expected Output	Ja, das Gebäude ist rollstuhlgerecht. Es gibt Rollstuhlrampe auf der Rückseite des Gebäudes. Aufzug befindet sich links. Behindertengerechten Toilette im UG nahe Rampe.
Actual Output	Das Institut für Photogrammetrie und Fernerkundung (IPF) hat einen barrierefreien Zugang. Es gibt eine Rollstuhlrampen auf der Rückseite des Gebäudes und einen Aufzug, der links liegt. Außerdem gibt es eine Behindertentoilette im Untergeschoss in der Nähe der Rampe.
Judge Reason	The output conveys the same information as the expected output, including the availability of a wheelchair ramp, elevator location, and accessible restroom.
Scores	FScore: 0.75, Precision: 0.73, Recall: 0.76

Category: Open Website

Example: Failure (Score = 0.00)

Input	Zeige mir die Homepage des Instituts für Wasser und Gewässerentwicklung - Wasserwirtschaft und Kulturtechnik
Expected Output	Ich habe die Webseite für dich in deinem Browser geöffnet.
Actual Output	Das Institut für Wasser und Gewässerentwicklung - Wasserwirtschaft und Kulturtechnik ist unter folgender Adresse zu finden: Engesserstraße 22, 76131 Karlsruhe. Die Homepage des Instituts ist unter https://www.iwu.kit.edu/ erreichbar.
Judge Reason	The output conveys the correct information about the institute's homepage URL, while the expected output states that the bot has opened the webpage. The output does not match the expected meaning.
Scores	FScore: 0.59, Precision: 0.56, Recall: 0.63

Example: Partial Success (Score = 0.50)

Input	["Navigate to the HOC building's website"]
Expected Output	Ich habe die Webseite für dich in deinem Browser geöffnet.
Actual Output	Die Website des HOC-Buildings ist http://www.hoc.kit.edu/ .
Judge Reason	The output provides the URL of the HOC building's website, while the expected output mentions opening the website in the browser. The meaning of performing an action related to the website is conveyed but not in the same way.
Scores	FScore: 0.60, Precision: 0.57, Recall: 0.64

Example: Success (Score = 1.00)

Input	Webadresse des Studierendenservice gesucht
Expected Output	Ich habe die Webseite für dich in deinem Browser geöffnet.
Actual Output	Das Webadresse des Studierendenservice ist http://www.sle.kit.edu/ .
Judge Reason	The output provides the web address, while the expected output mentions opening the webpage in the browser. Both convey the necessary information about the Studierendenservice website.
Scores	FScore: 0.60, Precision: 0.57, Recall: 0.64