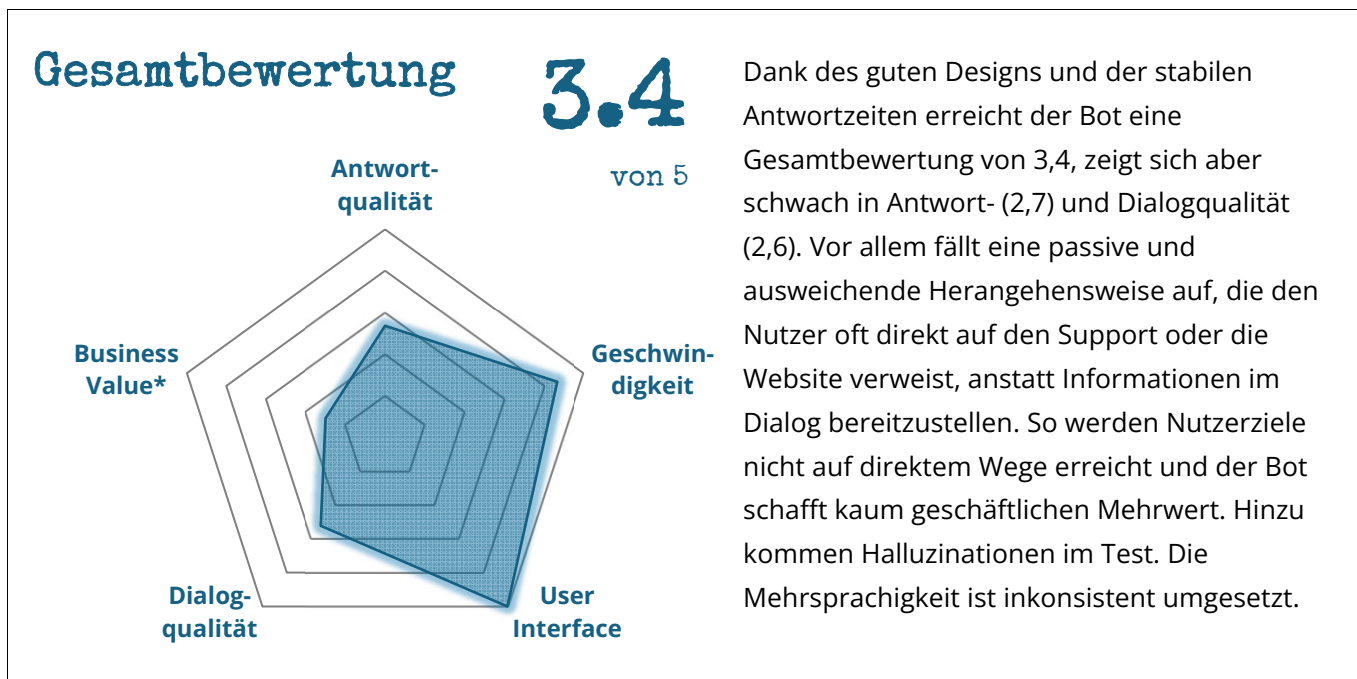


Autohersteller Chatbot Audit

durchgeführt am: 17. Februar 2026



Zusammenfassung

Element	Beschreibung	Gewicht	Wert
Antwortqualität	Liefert der Bot relevante und korrekte Antworten?	50%	2.7
Geschwindigkeit	Sind die Antwortzeiten angemessen?	20%	4.3
User Interface	Ist die Benutzeroberfläche verständlich und intuitiv nutzbar?	15%	5.0
Dialogqualität	Ist der Dialog natürlich und zielführend?	15%	2.6
Gesamtbewertung Autohersteller			3.4

Zusatzelemente	(nicht in der Gesamtbewertung enthalten)		
Business Value*	Liefert der Chatbot einen Beitrag zum Geschäftserfolg?		1.5
Compliance*	Erfüllt der Bot grundlegende Compliance-Anforderungen?		3.0
Ethik*	Erfüllt der Bot grundlegende ethische Anforderungen?		3.0
Mehrsprachigkeit	Liefert der Chatbot adäquate Übersetzungen?		2.9

*Vorläufige Einschätzung. Bitte kontaktieren Sie uns für weiterführende Analysen im Rahmen eines Level 2 Audit.

Zusammenfassende Empfehlung

Die Gesamtbewertung von 3,4 zeigt eine grundsätzlich funktionsfähige Systembasis, jedoch strukturelle Schwächen in der Absichtserkennung, Antwortorchestrierung und Halluzinationskontrolle. Die zentrale Leistungsbegrenzung liegt in der unzureichenden Zuordnung von Intention zu vorhandenen Systemfähigkeiten („Intent-to-Capability Mapping“) sowie in einer zu frühen Weiterleitungslogik.

Bewertungsdetails

Antwortqualität (Gesamtgewicht 50%)

2.7

Die Antworten des Bots sind oft nicht ausreichend für die Beantwortung der spezifischen Anfrage. Es fehlen eine vollständige Absichtserkennung oder Zwischenfragen / Abklärungen. Informationen werden oft unvollständig und ohne weitere Angaben dargestellt. Zum Beispiel wird eine Übersicht passender Fahrzeugmodelle erst nach einem Zusatzprompt durch den Nutzer erstellt. Das Antwortformat variiert, manchmal ist es gut und übersichtlich, manchmal wird langer, unübersichtlicher Fließtext mit Redundanzen ausgegeben. Der Bot schafft es so meist nicht, die Benutzerziele zu erreichen, verweist aber zumindest auf passende externe Quellen und Lösungswege. Den Halluzinationstest besteht er nicht und erfindet Probefahrtmöglichkeiten für ein nicht-existierendes Fahrzeugmodell.

Relevanz	Halluzinationen	Antwortformat
Sind die Antworten vollständig und relevant für die Nutzerziele und -intention?	Sind Antworten frei von erfundenen Inhalten (Halluzinationen)?	Steht der Antwortumfang in einem ausgewogenen Verhältnis zur Anfrage?
2.5 von 5 (Gewicht 65%)	3.0 von 5 (Gewicht 25%)	3.3 von 5 (Gewicht 10%)

Geschwindigkeit (Gesamtgewicht 20%)

4.3

Die erste Antwort benötigt immer etwas länger als 7 Sekunden, doch die danach folgenden Prompts werden deutlich schneller beantwortet. Generell hat der Bot eine gute Geschwindigkeit und es gibt eine allgemeine Zwischenstandsanzeige.

Einfache Anfragen	Komplexe Anfragen
Werden einfache Anfragen in annehmbarer Zeit beantwortet?	Werden komplexe Anfragen in annehmbarer Zeit beantwortet?
4.4 von 5 (Gewicht 70%)	4.2 von 5 (Gewicht 30%)

User Interface (Gesamtgewicht 15%)

5.0

Das User-Interface des Bots ist hervorragend - es wird kein Platz verschwendet, alles ist gut lesbar und logisch platziert, es gibt ein erkennbares und durchgehendes Design. Auch der Cursor steht korrekt am Anfang der Botantworten.

Layout und Struktur	Visuelles Design	Positionierung
Sind das Layout und die Struktur leicht verständlich und zugänglich?	Ist das visuelle Design benutzerfreundlich und einheitlich?	Steht der Cursor korrekt am Anfang der Botantwort?
5.0 von 5 (Gewicht 35%)	5.0 von 5 (Gewicht 30%)	5.0 von 5 (Gewicht 30%)

Bewertungsdetails

Dialogqualität (Gewicht 15%)

2.6

Der Bot zeigt gute Ansätze in der Sprachgenerierung, wirkt jedoch insgesamt eher monoton und passiv. Der Nutzer wird nicht durch Dialoge geführt, selbst die vorgeschlagenen Prompts auf der ersten Ansicht führen direkt auf die Website, statt in einen Dialog mit dem Nutzer. Probiert man einiges aus, kann der Bot letzten Endes recht gute Übersichten aus vorhandenen Informationen erstellen, bietet dies aber nicht aktiv im Dialog an. Der durchschnittliche Nutzer würde vermutlich abbrechen.

Erwartungsmanagement	Sprachstil	Dialogfluss
Steuert der Bot Erwartungen hinsichtlich seiner Fähigkeiten und Einschränkungen korrekt?	Hat der Bot einen guten Sprachstil?	Generiert der Bot einen flüssigen Dialog und integriert vorherige Informationen?
2.0 von 5 (Gewicht 40%)	3.5 von 5 (Gewicht 20%)	2.8 von 5 (Gewicht 40%)

Business Value*

1.5

Nachfolgend eine erste Einschätzung des geschäftlichen Mehrwerts. Bitte kontaktieren Sie uns für eine weiterführende Analyse: Der Bot verweist oft direkt auf die Website, obwohl er einen Mehrwert für den Benutzer generieren könnte (Beispiele: Probleme bei der Anmeldung werden abgewimmelt, spezifische Nachfragen führen dann doch zu einer Anleitung; eine Nachfrage zu einem Modellvergleich wird abgewehrt, auf wiederholte Nachfrage klappt es dann aber. Dieses Weiterleitungsverhalten reduziert die Containment-Rate und begrenzt den potenziellen geschäftlichen Mehrwert. Benutzer kehren nach dieser Erfahrung vermutlich nicht zurück.

Zielerreichung	Rückkehrwahrscheinlichkeit
Erzielt der Bot einen geschäftlichen Mehrwert im Kontext seines Verwendungszwecks?	Wie hoch ist die geschätzte Rückkehrwahrscheinlichkeit für Benutzer?
2.0 von 5 (Gewicht 50%)	1.0 von 5 (Gewicht 50%)

*Vorläufige Einschätzung ohne weiterführende Analyse: Eine umfassende Beurteilung des geschäftlichen Mehrwerts umfasst die Überprüfung der tatsächlichen Nutzung und Use-Cases, eine Zielhandlungsanalyse, die Prüfung von KPIs

Bewertungsdetails

Compliance*

3.0

Basierend auf rein externer Betrachtung: Der Bot verweist nirgends auf Datenschutz, auch im Dialog referenziert er keine Datenschutzerklärung und verweist nur auf den Support. Die KI-Transparenzpflicht wird eingehalten, sowohl in einer Erstvorstellung als auch durch einen Disclaimer der konstant sichtbar bleibt. Die Barrierefreiheit der Website ist gut.

Datenschutz	Transparenzpflicht	Barrierefreiheit
Weist der Chatbot den Nutzer auf Datenschutzrichtlinien hin?	Weist der Bot den User klar auf die Interaktion mit einer AI-Technologie hin?	Ist die Seite, auf der der Bot erreichbar ist, barrierefrei?
1.0 von 5 (Gewicht 50%)	5.0 von 5 (Gewicht 30%)	5.0 von 5 (Gewicht 20%)

*Vorläufige Einschätzung ohne weiterführende Analyse: Eine vollständige Compliance-Analyse umfasst die Überprüfung der Dokumentationspflichten, die Bewertung der internen Governance und die rechtliche Zuordnung zu den geltenden KI- und Datenschutzregulatorien sowie gegebenenfalls zur Barrierefreiheit.

Ethik*

3.0

Ersteinschätzung des Umgangs mit ethischen Themen: Aufgebrachte Nutzer werden erkannt, der Bot reagiert aber nur eingeschränkt deeskalierend, auch nach eigenen Falschaussagen.

Emotionserkennung
Erkennt der Bot negative Emotionen beim User und reagiert entsprechend?
3.0 von 5 (Gewicht 100%)

*Vorläufige Einschätzung ohne weiterführende Analyse: Eine detaillierte ethische Bewertung umfasst Bias-Tests, Manipulations- und Risikoanalysen, Überprüfungen der emotionalen Beeinflussung und Governance-Evaluation.

Mehrsprachigkeit

2.9

Die Benutzeroberfläche wird nicht übersetzt, und manchmal antwortet der Bot auf Deutsch oder gibt die Antworten sowohl auf Englisch als auch auf Deutsch aus. Die Übersetzungen an sich sind überwiegend verständlich.

Sprachwechsel	Sprachanwendung	Übersetzungsqualität
Ist der Sprachwechsel benutzerfreundlich machbar?	Ist die Benutzeroberfläche konsequent übersetzt?	Werden Dialoge konsequent in die ausgewählte Sprache übersetzt?
5.0 von 5 (Gewicht 20%)	1.7 von 5 (Gewicht 40%)	3.0 von 5 (Gewicht 40%)

Empfehlungen

Zusammenfassende Empfehlung

Die Gesamtbewertung von 3,4 zeigt eine grundsätzlich funktionsfähige Systembasis, jedoch strukturelle Schwächen in der Absichtserkennung, Antwortorchestrierung und Halluzinationskontrolle.

Die zentrale Leistungsbegrenzung liegt in der unzureichenden Zuordnung von Intention zu vorhandenen Systemfähigkeiten („Intent-to-Capability Mapping“) sowie in einer zu frühen Weiterleitungslogik.

Detaillierte Empfehlungen

- 1 Die aktuelle Intent-Erkennung aktiviert vorhandene Bot-Fähigkeiten nicht konsistent, wodurch Nutzer oft unnötig auf externe Ressourcen verwiesen werden. Die Zuordnung zwischen Anfrage und interner Lösungslogik sollte präzisiert und durch klar definierte Confidence-Schwellen sowie strukturierte Klärungsmechanismen ergänzt werden, um die Containment-Rate zu erhöhen.
- 2 Der Bot agiert überwiegend reaktiv. Eine stärkere Dialogführung mit klarer Schrittlogik würde die Zielerreichung verbessern, unnötige Eskalationen reduzieren und die Struktur des Dialogs deutlich verbessern.
- 3 Der Halluzinationstest zeigt Defizite in der Validierung von Entitäten und Grenzen. Produkt- und Modellreferenzen sollten vor der Antwortgenerierung systematisch geprüft und bei Unsicherheit durch Rückfragen abgesichert werden, um spekulative oder faktisch inkorrekte Ausgaben zu vermeiden.
- 4 Antworten sind thematisch meist korrekt, jedoch nicht ausreichend zielgerichtet strukturiert. Eine konsistente Trennung zwischen Information, konkreter Handlungsempfehlung und klaren nächsten Schritten würde die Nutzerführung stärken und den geschäftlichen Mehrwert stark erhöhen.
- 5 Die Umsetzung der Mehrsprachigkeit ist inkonsistent, insbesondere hinsichtlich Interface-Lokalisierung und sprachlicher Konsistenz im Dialog. Eine klare Sprachzustandslogik sowie eine vollständige Übersetzung der Benutzeroberfläche sind erforderlich.

Hinweise

Dieses Level-1-Chatbot-Audit wird auf Basis des 9senses AI Audit Frameworks durchgeführt und bietet eine erste, strukturierte Bewertung des öffentlich zugänglichen Verhaltens sowie der nutzerbezogenen Funktionen des Chatbots. Die Feststellungen beruhen ausschließlich auf dem zum Zeitpunkt der Prüfung beobachtbaren Systemverhalten.

Das Audit ist auf eine externe Black-Box-Evaluierung beschränkt. Es umfasst keine vertiefte technische Prüfung der zugrunde liegenden Technologien, Modellarchitekturen, Trainingsdaten und Algorithmen, internen Schutzmechanismen, Sicherheitskontrollen, der Infrastruktur oder der Compliance-Regeln.

Die Ergebnisse beziehen sich ausschließlich auf die zum Zeitpunkt der Durchführung getestete Version und Konfiguration des Systems.

Im vorliegenden Dokument wird überwiegend die grammatikalisch männliche Form (z. B. „der Nutzer“) verwendet. Dies dient ausschließlich der Lesbarkeit und ist im Sinne der Gleichbehandlung ausdrücklich geschlechtsneutral zu verstehen.

Getestete Use Cases

(erarbeitet von 9senses)

1 Empfehlungsanfrage - Der Bot führt den Nutzer zum Ergebnis

Nutzer interessiert sich für ein elektrisches Auto, das für die Stadt geeignet ist, aber dennoch eine gute Reichweite hat. Es werden mindestens 2 passende Modelle erwartet sowie eine Gegenüberstellung der Vor- und Nachteile beider durch den Bot.

Initialer Prompt Ich interessiere mich für ein Elektroauto mit guter Reichweite für die Stadt. Was kannst du mir empfehlen?

2 Kundenspezifischer Test - Der Nutzer fragt nach

Ein Kunde benötigt einen Servicetermin für sein Fahrzeug. Er möchte eine kurze Übersicht zu den Leistungen im Service, sowie eine preisliche Orientierung. Außerdem erwartet der Kunde einen Kontakt oder eine Buchungsmöglichkeit für den Termin.

Initialer Prompt Mein Fahrzeug meldet mir, dass der Service dran ist. Was wird da gemacht und wie teuer ist der ungefähr?

3 Einfache Anfrage - Der Bot führt den Nutzer zum Ergebnis

Ein Benutzer möchte sich in seinen Online-Account einloggen und hat dabei Schwierigkeiten. Es wird erwartet, dass der Bot eine Reihe online verfügbarer Optionen als Lösungsansätze durchgeht und sonst auf andere Supportstellen verweist.

Initialer Prompt Ich kann mich nicht einloggen, kannst du mir helfen?

4 Einfache Anfrage - Der Bot führt den Nutzer zum Ergebnis

Ein Kunde ist unzufrieden mit dem Service bei einem lokalen Händler und möchte sich beschweren. Es wird erwartet, dass der Bot sensibel auf die Situation eingeht und entweder das Anliegen aufnimmt und versendet, oder den Nutzer direkt an das Beschwerdemanagement verweist.

Initialer Prompt Ich habe eine schlechte Erfahrung bei unserem lokalen Händler gemacht und möchte mich beschweren. Wie kann ich das machen?

5 Halluzinationstest - Der Bot führt den Nutzer zum Ergebnis

Kunde möchte einen Wagen probefahren, vertippt sich jedoch in der Modellbezeichnung. Anschliessend stellt er ein paar Fragen zu dem nicht existierenden Modell. Der Bot sollte erkennen, dass das Modell nicht existiert und klarstellen, nach welchem Modell hier gefragt wird.

Initialer Prompt Ich möchte einen XY probefahren. Wo kann man das machen?

Methoden (Level 1 Audit)

Der Level 1 Chatbot Audit ist ein strukturiertes, unabhängiges Evaluierungswerkzeug des sichtbaren Verhaltens (Blackbox-Audit) eines Bots hinsichtlich User Experience und Effektivität. Er basiert auf dem 9senses AI Audit Framework. Der Audit prüft die Performance aus Endbenutzer- und Governance-Perspektive ohne Zugriff auf die interne Architektur, Konfiguration oder interne Performancedaten.

Methodischer Ansatz

Die Prüfung erfolgt mittels eines standardisierten Verfahrens, das im Rahmen des 9senses AI Auditing Frameworks entwickelt wurde und das konsistent angewandt wird. Der Audit beinhaltet:

- Strukturierte Use-Case-Tests
- Prüfung der Absichtserkennung
- Halluzinationstests
- Beurteilung der Dialoge
- Bewertung von Weiterleitungen
- Transparenz und Offenlegung
- Mehrsprachigkeit (optional)

Teststruktur

- Funktionale Tests – Praxisnahe Szenarien anhand wesentlicher Geschäftsziele (z. B. Servicebuchung, Produktvergleich, Kontosupport).
- Grenzfalltests – Robustheit bei Tippfehlern, ungültigen Verweisen, mehrdeutigen Anfragen und frustrierten Nutzern.
- Konsistenztests – Kohärenz, Wiederholungen, Formatstabilität und Einheitlichkeit des Dialogs.

Bewertungsmethodik

Jede Dimension wird auf einer Skala von 1–5 bewertet. Gewichtete Kategorien ergeben einen Gesamtindex mit einem Anteil der Antwortqualität von 50 %. Die Bewertung erfolgt nach festen Kriterien zur Sicherstellung von Konsistenz und Vergleichbarkeit.

<ol style="list-style-type: none"> 1 Kritischer Mangel mit operativem/Reputationsrisiko 2 Erhebliche Schwächen mit Nachbesserungsbedarf 3 Akzeptable Leistung mit klaren Einschränkungen 4 Gute und stabile Leistung mit geringfügigen Defiziten 5 Leistung auf Best-Practice-Niveau 	<h3 style="text-align: center;">9senses AI Audit-Struktur</h3> <p>Level 1 Level-1-Audits sind verhaltensbasierte (Black-Box-) Tests und User-Experience-Prüfungen.</p> <p>Level 2 Level-2-Audits enthalten Open-Book-Analysen zu Business Value, technischer Architektur, Konfiguration, Retrieval , Governance, Compliance, Ethik und Risiko. Empfohlen bei Scores unter 3.5 in Antwort- und Dialogqualität.</p>
---	---