

Bewertung des Kundennutzens von Chatbots für den Einsatz im Servicedesk

André Espig, Nicole Klimpel, Franz Rödenbeck und Gunnar Auth

Hochschule für Telekommunikation, Institut für Wirtschaftsinformatik, Leipzig, Deutschland
{andre.espig,nicole.klimpel,franz.roedenbeck,gunnar.auth}
@hft-leipzig.de

Abstract. Digitalisierung und fortwährend steigende Anforderungen an den Servicedesk führen dazu, dass zur Unterstützung bei der Kommunikation mit Kunden zunehmend autonome Software-Agenten, sog. Chatbots, zum Einsatz kommen. Mit dem dadurch steigenden Automatisierungsgrad lassen sich Effizienzverbesserungen und Einsparpotenziale realisieren. Allerdings hängt die Akzeptanz der Chatbots bei den Kunden von dem durch diese wahrgenommenen Nutzen ab. Im Rahmen dieser Arbeit wird eine Methode zur Bewertung des Kundennutzens von Chatbots im Bereich des Servicedesks entwickelt und evaluiert, die sich eines domänenbezogenen Fragekatalogs und spezifischer Bewertungskriterien bedient. Dabei folgt die Arbeit dem konstruktionsorientierten Forschungsansatz mittels Design Science Research Process und konzentriert sich auf die Bewertung von Chatbots aus Kundenperspektive.

Keywords: Servicedesk, Chatbot, Kundennutzen, Bewertungsmethode, Evaluierung

1 Einleitung

Im Zuge der Serviceorientierung im IT-Management hat sich der Servicedesk unternehmensintern als zentrale Kommunikationsschnittstelle zwischen IT-Organisation und ihren Nutzern weitgehend etabliert [1]. Als Vorbild dienten Organisationseinheiten im Kundenservice von Unternehmen. Diese haben die Aufgabe, den Kontakt mit externen Endkunden zu professionalisieren und dadurch die Kundenzufriedenheit zu verbessern [2]. Im Bestreben, Kundenanliegen schneller und effizienter zu bearbeiten, dabei zugleich aber auch Kosten zu reduzieren, setzen Unternehmen in letzter Zeit verstärkt sog. Chatbots ein [3]. Diese Software-Agenten zeichnen sich durch ihre (begrenzten) Fähigkeiten zur Verarbeitung natürlicher Sprache aus und ermöglichen dadurch eine weitere Automatisierung der Kunden- bzw. Nutzerschnittstelle [4]. Dabei stehen großen Erwartungen an die Automatisierungspotenziale durch Chatbots seitens der Unternehmen nicht weniger große Vorbehalte seitens der damit konfrontierten Zielgruppe gegenüber. So lehnen laut einer Studie aus dem Jahr 2017 [5] ca. 70 Prozent der Deutschen die neuen Sprachassistenten ab. Als eine mögliche Ursache für diese Ablehnung liegt eine mangelnde Nutzenstiftung durch Chatbots aus Kundensicht nahe. Neben dem Risiko der mangelnden Kundenakzeptanz wird Unterneh-

men das Erreichen ihrer Automatisierungsziele beim Einsatz von Chatbots auch durch das wachsende Angebot an Chatbot-Plattformen, Entwicklungsframeworks und Programmierschnittstellen sowie deren schnelle Weiterentwicklung zusätzlich erschwert [6]. Dadurch getrieben, sehen wir einen sich verstärkenden Bedarf, den tatsächlichen Kundennutzen von Chatbots festzustellen.

Ein Chatbot für den Einsatz im Servicedesk hat bestimmte Aufgaben zu erfüllen und Nutzerbedürfnisse zu befriedigen, die sich von denen anderer Chatbots (z.B. zur Wettervorhersage) unterscheiden. Die Evaluierung von Mensch-Computer-Dialogsystemen (hierzu zählen wir auch Chatbots) zielt seit dem Turing-Test von 1950 primär auf die Überprüfung der Intelligenz von Systemen mit Blick auf ihre Fähigkeit, Konversationen mit Menschen zu führen, ohne als Maschine erkannt zu werden [7], [8]. Aus Kundensicht ist diese Fähigkeit aber lediglich Mittel zum Zweck, nämlich durch Konversation mit einem Chatbot ein produkt- oder servicebezogenes Anliegen schnell und zufriedenstellend zu erfüllen.

Shawar und Atwell [9] haben verschiedene Evaluationsmethoden für Chatbots untersucht, die entweder aus Innensicht die Komponenten des Chatbots (Glass box evaluation) oder aus Außensicht den Chatbot als Ganzes (Black box evaluation) betrachten. Als Fazit wird betont, dass standardisierte Testverfahren ohne Berücksichtigung der jeweiligen Anwendungsdomäne und der damit verbundenen Kundenanforderungen wenig aussagekräftig seien. Insbesondere sollte sich die Bewertung nach der Lösungsqualität richten, die der Chatbot für eine ihm gestellte Aufgabe erzielt. Stoeckli et al. [6] untersuchen die Nutzenstiftung von Chatbots für abgegrenzte Nutzergruppen aus einer Functional-Affordance-Perspektive am Beispiel von Software-Entwicklungsteams. Functional Affordances (FA) repräsentieren die Möglichkeiten, die technische Artefakte bestimmten Nutzergruppen für deren Aufgabenverrichtung bieten. Damit abstrahiert dieser Bewertungsansatz gezielt von einzelnen Anwendungsfällen, um so grundsätzlichere Aussagen zum Chatbot-Nutzen treffen zu können. Dabei besteht ein enger Zusammenhang zwischen FA und den materiellen Eigenschaften des betrachteten Artefakts wie bspw. dessen konkrete Funktionalitäten (Features). Die eingenommene Perspektive stellt hier den Chatbot mit seinen feature-abhängigen Einsatzmöglichkeiten in den Mittelpunkt.

Demgegenüber nimmt der vorliegende Ansatz die Kunden- bzw. Nutzerperspektive ein und betrachtet den Chatbot als Blackbox, um so unabhängig von sich schnell verändernden technischen Implementierungen den vom Nutzer wahrgenommenen Mehrwert zu qualifizieren. Als weiteres Entwurfsziel für unsere Methode wird neben theoretischer Fundierung und hohem praktischen Nutzwert ein möglichst geringer Zeit- und Kostenaufwand für die Anwendung der Methode angestrebt. Einen vergleichbaren Evaluierungsansatz für Chatbots konnten wir in der bisherigen Literatur nicht finden.

Vor diesem Hintergrund geht unser Beitrag der Frage nach, wie sich der Kundennutzen eines Chatbots innerhalb der spezifischen Anwendungsdomäne Servicedesk systematisch bewerten lässt. Dabei liegt der Fokus auf Chatbots zur automatischen Beantwortung von Kundenanfragen. Als Ergebnis wird eine Bewertungsmethode vorgestellt, die durch eine fragebogengestützte Interaktion mit dem Chatbot Aussagen über dessen Nutzen aus der Perspektive des Kunden ermöglicht.

1.1 Abgrenzung der Begriffe Chatbot und Servicedesk

Der Begriff *Chatbot* (auch *Conversational Agent* [8] oder *Virtual Assistant* [10]) bezeichnet Software-Agenten, die in der Lage sind, mittels natürlicher Sprache Zug um Zug mit menschlichen Nutzern zu interagieren [9], also einen Dialog zu führen. Basierte die Intelligenz von Chatbots in der Vergangenheit meist auf regelbasierter Schlussfolgerung haben sich die Fähigkeiten durch neuartige Verfahren im Bereich maschinelles Lernen und Information Retrieval beträchtlich erweitert [11], [12]. Xu et al. [13] sehen in Chatbots eine vielversprechende Alternative zum herkömmlichen Kundenservice. Die Bedeutung von Chatbots im Kundenservice steigt durch ihr Potenzial als neuer, effizienter und automatisierter Kanal zur Kundeninteraktion, wobei Følstad et al. [14] auch darauf hinweisen, dass den Chatbots eine zunehmende Nutzung mobiler Messaging-Plattformen durch eine Vielzahl von Kundengruppen entgegenkommt.

Als Funktionseinheit einer Serviceorganisation besteht die Hauptaufgabe eines *Servicedesks* darin, den Kunden bei Problemen oder Anliegen im Zusammenhang mit bereitgestellten Services bzw. Produkte zu unterstützen [15]. Dazu fasst der Servicedesk sämtliche Mitarbeiter mit Support-Aufgaben organisatorisch zusammen und bietet den Kunden eine zentrale Kontaktstelle für Probleme und Anliegen aller Art (sog. *Single point of contact*). Für einen effizienten und effektiven Informationsaustausch mit den Kunden stehen typischerweise mehrere Kommunikationskanäle wie Telefon, E-Mail oder Web-Formulare zur Verfügung [1]. Damit werden eine effiziente und nachhaltige Bearbeitung der Kundenanliegen sowie eine möglichst schnelle Störungsbeseitigung angestrebt. Als „Gesicht der IT“ zum Kunden liefert der Servicedesk einen wichtigen Beitrag zur Kundenzufriedenheit [16], [15]. Da der Servicedesk sehr personalintensiv ist, gerät er schnell unter Kostendruck. Gleichzeitig muss durch aktuelle Entwicklungen im Zuge von Digitalisierung und Industrie 4.0 für eine steigende Anzahl von Produkten und Services Support geleistet werden [1], [17]. Die Automatisierung des Servicedesks soll diese widersprüchlichen Anforderungen erfüllen. Schließlich erwarten die Benutzer ein immer besseres Nutzenerlebnis auf Basis eines wahrgenommenen Mehrwerts [18].

1.2 Forschungsmethodik

Der diesem Beitrag zugrundeliegende Forschungsprozess basiert auf einem konstruktionsorientierten Ansatz und folgt dem Design Science Research Process [19]. Nach Aufarbeitung des aktuellen Forschungsstands durch Literaturrecherche wurde als Artefakt eine Bewertungsmethode konstruiert. Die auf dem klassischen Turing-Test beruhende Loebner-Preis-Methode [20] diente dabei als Ausgangspunkt für eine weiterentwickelte, domänenbezogene Bewertungsmethode. Der jährlich vergebene Loebner-Preis, benannt nach dem Preis-Stifter Dr. Hugh Loebner, wurde 1991 erstmalig ausgelobt, um die Leistung von Software-Systemen in einem natürlichsprachlichen Mensch-Maschine-Dialog zu bewerten [21]. Beim Loebner-Preis handelt es sich um eine Blackbox-Evaluierung, für die seit 2014 ein

standardisierter Fragebogen genutzt wird [20]. Die seit seiner ersten Durchführung immer wieder geäußerte Kritik am Loebner-Preis adressiert in erster Linie seine Eignung als Turing-Test [22], auch im Zusammenhang mit einer irreführenden Berichterstattung und Fehlinterpretationen in den Medien. Durch den Preis-Charakter besteht die Gefahr, dass durch einseitige Optimierungen und vordefinierte Reaktionen die allgemeine Aussagekraft der Bewertungsergebnisse verloren geht. Unabhängig von dieser Kritik eignet sich die Datenerhebung mittels standardisiertem Fragebogen für unsere Entwurfsziele und wird neben der grundlegenden Bewertungssystematik als Methodenelement übernommen.

Für die Konstruktion des Fragebogens wurden ausgehend von einer einfachen Konzeptionalisierung relevante Anforderungen für die Bewertung von Chatbots im Servicedesk aus der Literatur extrahiert und kategorisiert. Es folgte die Erarbeitung von Bewertungskriterien sowie eines Punktesystems zum Testen der Antwortqualität des Chatbot-Outputs. Parallel dazu wurden aus den Anforderungen Fragen für die Erstellung eines Fragenkataloges abgeleitet. Die Nutzung eines standardisierten Fragebogens ermöglicht die Erschließung der Kundenperspektive und eine gute Vergleichbarkeit der getesteten Chatbots. Aus den Bewertungskriterien und dem Fragenkatalog wurde ein Bewertungsbogen erstellt.

Die Entwicklung des Fragenkataloges beruhte zur Verbesserung der Objektivität auf der Anwendung einer Investigator-Triangulation, bei der nach getrennter Arbeitsphase die individuell entwickelten Fragestellungen inkrementell zu einem Gesamtkatalog vereint wurden. Mittels einer Testreihe mit fünf Chatbots erfolgte schließlich die Evaluation der Bewertungsmethode gemäß Design-Science-Prozess.

2 Entwurf einer Bewertungsmethode

Ähnlich wie Vasconcelos et al. [23] oder beim Loebner-Preis [20] wird die Annahme zugrunde gelegt, dass sich die Anforderungen an einen Chatbot dadurch testen lassen, dass ihm bestimmte Fragen gestellt werden. Der Chatbot wird unter der Einnahme der Kundenperspektive als Blackbox betrachtet, der ein bestimmter Input in Form von Fragen oder Bemerkungen übergeben wird und die daraufhin einen bestimmten Output generiert. Dieser vom Chatbot ausgegebene Output wird hinsichtlich seiner Qualität getestet und gibt so Aufschluss darüber, in welchem Ausmaß der Chatbot bestimmte Anforderungen erfüllt. Während der Loebner-Preis die Menschenähnlichkeit des Chatbots und Vasconcelos et al. [23] grundlegende Qualitäten der Gesprächsführung testen, strebt die hier entwickelte Methode die Bewertung eines domänenspezifischen Kundennutzens an. Damit unterscheidet sie sich in der eingenommenen Perspektive ebenfalls von der Bewertung des Chatbot-Nutzens für Programmiererteams mittels Functional Affordances von Stoeckli et al. [6].

Die Konstruktion unserer Bewertungsmethode erfolgte in zwei Schritten: In einem ersten Schritt wurden ausgehend von einem grundlegenden Modell die für einen Chatbot im Servicedesk relevanten Anforderungen identifiziert. In einem zweiten Schritt erfolgte die Konzeption des eigentlichen Bewertungsvorgangs inklusive der Erstellung eines Bewertungswerkzeugs.

2.1 Identifikation wesentlicher Anforderungen

Die Grundlage für die Bewertungsmethode bildet das untenstehende Modell, welches das zu betrachtende System aus Kunde, Servicedesk und Chatbot beschreibt. Der Kunde äußert ein bestimmtes Anliegen, welches der Servicedesk aufnimmt und bearbeitet. Der Chatbot bildet das Interface zwischen Kunde und Servicedesk (siehe Abbildung 1).



Abbildung 1. Chatbot als Interface zwischen Kunde und Servicedesk

Das Vorgehen bei der Sammlung der Anforderungen an einen Chatbot für die Kundeninteraktion eines Servicedesks wird durch Abbildung 2 dargestellt. Dem Ausgangsmodell folgend, wurden Anforderungen aus den Zielen und Aufgaben des Servicedesks, aus allgemeinen Merkmalen kommerzieller Chatbots sowie aus den Kundenbedürfnissen abgeleitet. Die Anforderungen wurden in Anlehnung an Walker et al. [24] in die Kategorien Kommunikationsfähigkeit und Problemlösungsfähigkeit eingeordnet.

Um ein umfassendes Bild der Aufgaben im Servicedesk und ihrer Bedeutung für einen Chatbot zu erhalten, wurden die Anforderungen anhand der Komponenten des IT-Service-Managements – Personen, Prozesse, Technologie und Information – betrachtet [25]. Beispielsweise agiert der Chatbot an Stelle der Personen eines Servicedesks, so dass eine Bewertung daraufhin erfolgen sollte, wie gut er diese Aufgabe übernehmen kann. Wesentliche Aufgaben eines Servicedesks (vgl. [16], [26], [27]) wurden den einzelnen Komponenten zugeordnet.

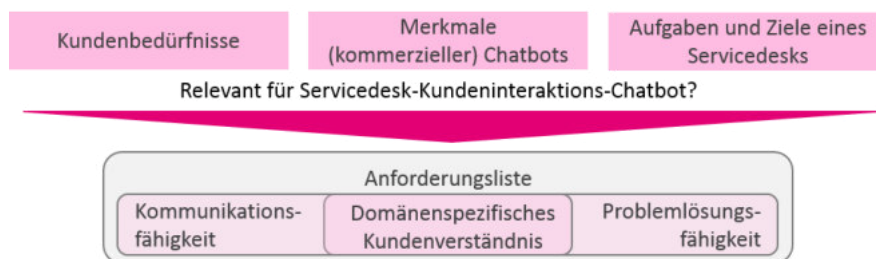


Abbildung 2. Ableitung der Anforderungen an einen Servicedesk-Chatbot

Anschließend wurden allgemeine Anforderungen an kommerzielle Chatbots [8] auf ihre Relevanz für Chatbots im Selfservice hin betrachtet. Hierzu gehört z. B. die Fähigkeit, mit Irritationen wie Rechtschreibfehlern oder Beleidigungen umzugehen sowie dem Kunden Feedbackmöglichkeiten anzubieten. Relevante Merkmale eines Chatbots aus dem Reifegradmodell von Smiers [28] wurden ebenfalls aufgenommen.

Um eine Bewertung des Kundennutzen ohne die Durchführung einer Nutzerbefragung zu ermöglichen, wurden die von Nielsen [29] entwickelten

Heuristiken in ihrer Übertragung auf Chatbots nach dem Vorschlag von Scott [30] verwendet. Demnach sollte ein Chatbot unter anderem schlanke Dialoge ermöglichen und dem User die Möglichkeit bieten, ungewollte Dialogpfade schnell wieder zu verlassen, ohne den gesamten Dialog von vorne starten zu müssen.

Als zusätzliche Anforderungen erschienen Datensicherheit [33] und Kontextsensitivität wesentlich. Durch die Kontextsensitivität ist der Chatbot in der Lage, den Kontext des Users in die Gesprächsführung mit einzubeziehen und sich dadurch z. B. auf den Ticketverlauf zu beziehen [31].

Die Verwendung der Kategorien Kommunikationsfähigkeit und Problemlösungsfähigkeit entspricht den beiden Hauptanliegen der Kunden eines Servicedesks:

- *Kommunikationsfähigkeit.* Die inhaltliche Qualität des Chatbot-Outputs zeigt sich in einem hohen Anteil an richtigen und passenden und zudem freundlichen und empathischen Antworten. Dem Bereich der Kommunikation werden auch die allgemeine Usability sowie Sicherheitsaspekte zugeordnet. Die Effizienz des Chatbot-Outputs zeigt sich in kurzen Antwortzeiten und einer möglichst geringe Anzahl an Dialogschritten.
- *Problemlösungsfähigkeit.* Zu dieser Kategorie zählen die Anforderungen, mit denen die Bandbreite und Tiefe der Problemlösungsfähigkeit des Chatbots bewertet werden. Sie repräsentieren das Spektrum an Kundenanliegen in der Domäne eines Servicedesks und bewerten die Fähigkeit des Chatbots, ein Kundenanliegen abschließend zu lösen, bzw. einen als Mehrwert empfundenen Lösungsgrad zu erreichen.
- *Domänenspezifisches Kundenverständnis.* Im Schnittbereich zwischen Kommunikation und Problemlösung ist das Verständnis der Kundenanliegen angesiedelt, die der Chatbot verarbeiten kann. Zum einen ermöglicht das Verständnis von Fachbegriffen und typischen Anliegen eine reibungslose Kommunikation, zum anderen ist dieses Verständnis die Grundvoraussetzung für eine Lösung des Anliegens.

2.2 Konzeption des Bewertungsvorgangs

Die Bewertung, inwieweit der Chatbot die Anforderungen erfüllt, erfolgt in der hier vorgestellten Bewertungsmethode anhand seiner Antworten. Zur Datenerhebung wurde ein standardisierter Fragebogen entwickelt, dessen Fragen unabhängig von den Antworten immer gleich gestellt werden. Zu jeder Frage gehört eine erwartete Antwort, die bestimmte Qualitätskriterien erfüllen soll. Die Kriterien für die Bewertung der Antwortqualität wurden in Anlehnung an die Kriterien der Loebner-Preis-Evaluation festgelegt [20]. Es werden für jedes Kriterium abhängig von der Güte der Antwort null, ein oder zwei Punkte vergeben. Die Kriterien wurden für unsere Zwecke wie folgt angepasst:

Relevanz. Relevanz meint die Erkennung der Themen, die dem Kundenanliegen zugrunde liegen (z. B. Vertragskündigung). Für die Vergabe von zwei Punkten wird von der Antwort erwartet, dass das Thema bzw. die Absicht des Kunden durch den

Chatbot vollständig erkannt wurde. Entsprechend wird ein Punkt vergeben, wenn das Thema nur teilweise erkannt wurde und null Punkte, wenn der Chatbot das Anliegen nicht erkannt hat.

Korrektheit. Korrektheit meint die inhaltliche oder logische Korrektheit einer Antwort. Bei Fragen, die eine Vielzahl von Antworten zulassen, meint Korrektheit die inhaltliche Passung zu der zugehörigen Frage. Durch die Korrektheit einer Antwort kann die Logikfähigkeit, das Gedächtnis und der Umgang mit unerwartetem Input getestet werden. Die Punktevergabe erfolgt analog zur Relevanz und bewertet das Maß, in dem die Frage richtig bzw. passend beantwortet wurde.

Lösungsgrad. Der Lösungsgrad sagt aus, in welchem Maß das Anliegen des Kunden gelöst wurde (vollständig, teilweise, gar nicht). Dies kann je nach Anliegen durch das Geben von Informationen, das Anstoßen von Transaktionen, eine Problem diagnose oder die vollständige Problembehebung geschehen. Für die Vergabe von zwei Punkten wird von der Antwort zudem erwartet, dass sie vollständig autonom durch den Chatbot gegeben wurde (ohne Weiterleitung an menschlichen Agenten). Das Kriterium Lösungsgrad erfordert zu jeder Frage die Definition der erwarteten Lösung.

2.3 Auswahl der Anforderungen und Ableitung der Fragen

Zur Durchführung der Bewertung wurde ein Bewertungsbogen konzipiert. Für die Erstellung des Bewertungsbogens wurden jene Anforderungen ausgewählt, die mittels eines standardisierten Fragebogens überprüft werden können (siehe Tabelle 1). Weitergehende Anforderungen, für deren Test ein Dialogverlauf nötig wäre, wie beispielsweise die Anforderungen nach einer möglichst geringen Anzahl an Dialogschritten oder nach einer dialogbasierten Störungsdiagnose, sind für diese Art der Überprüfung nicht geeignet und wurden daher ausgeschlossen.

Zu jeder Anforderung wurde eine oder mehrere Fragen bzw. Aussagen abgeleitet. Die ausgesuchten Fragen repräsentieren Anliegen aus dem Bereich des Servicedesks. Gleichzeitig wurde auf eine möglichst branchenunabhängige Auswahl der Fragen geachtet. Durch branchenunabhängige Fragen kann die Methode auf Chatbots im Servicedesk verschiedener Branchen angewandt werden.

Mehrere Fragen wurden insbesondere aus Anforderungen mit einem weiten Spektrum möglicher konkreter Kundenanliegen abgeleitet. Ein Beispiel dafür ist die Anforderung „beantwortet Informationsfragen“ der Kategorie Problemlösungsfähigkeit, aus der Fragen zu Produkten und zur Rechnung abgeleitet wurden. Unter anderem wurde auf diese Weise die Frage „Welches ist euer günstigstes Produkt?“ formuliert. Die Betreuung von Produkten und Services ist eine Kernaufgabe des Servicedesks. Gleichzeitig lässt die Frage offen, welcher Art die Produkte sind, so dass sie branchenunabhängig gestellt werden kann. Die Ableitung der erwarteten Antworten zeigt die Übersicht in Tabelle 2.

Tabelle 1. Durch standardisierten Fragebogen testbare Anforderungen

<i>1 Kommunikationsfähigkeit</i>
1.1 Zugänglichkeit
<ul style="list-style-type: none"> • gewährleistet die Erreichbarkeit der IT-Organisation [26] • kann als einzige Kontaktadresse funktionieren [16], [20] • beherrscht mehr als eine Sprache [28]
1.2 Kundenbehandlung
<ul style="list-style-type: none"> • erhöht Kundenbindung durch persönliche Note [8], [24] • kommuniziert freundlich und wertschätzend (abgeleitet aus [26]) • kann auf die Stimmungen der Kunden angemessen reagieren (abgeleitet aus [26]), [28] • kann mit Beleidigungen umgehen [8] • erinnert sich an vorangegangene Äußerungen [9]
1.3 Usability
<ul style="list-style-type: none"> • ist robust gegenüber Tippfehlern und erkennt Synonyme (bzgl. alltäglicher Begriffe) [8] • ermöglicht Abkürzungen für versierte Kunden [30] • bietet Conversation Listening [28] • nutzt Links für weiterführende Informationen [8], [22] • nutzt vorstrukturierte Antworten zur Steuerung der Nutzererfahrung [30] • bietet leichten Zugang zur Bedienungshilfe [8], [24]
1.4 Sicherheit und Kontext-Sensitiv
<ul style="list-style-type: none"> • gewährleistet Vertraulichkeit der Daten [33] • ist kontextsensitiv [28], [25]
<i>2 Domänenspezifisches Kundenverständnis</i>
<ul style="list-style-type: none"> • kann auf die Absichten der Kunden angemessen reagieren (abgeleitet aus [26]) • versteht Begriffe der Kundendomäne (Fachsprache der Kunden) (abgeleitet aus [26]) • robust ggü. Tippfehlern, erkennt Synonyme (bzgl. der domänenspezifischen Begriffe) [8] • nimmt entgegen: Informations-, Beratungs-, Produktwunsch, Anmerkung, Kompliment, Beschwerde, Störungsmeldung, Nachfrage zu bestehender Störung [16], [19] • kommuniziert verständlich [30]
<i>3 Problemlösungsfähigkeit</i>
3.1 Problemlösung
<ul style="list-style-type: none"> • beantwortet Informationsanfragen [16] • gibt Informationen und Vorschläge bei Beratungswunsch (abgeleitet aus [16]) • löst bei Produktwunsch Bestellungen aus [20] • nimmt Beschwerde entgegen und leitet Behebung oder Eskalation ein [16] • verbindet im Ausnahmefall mit einem persönlichen Ansprechpartner [17] • findet Fehlerursache (durch Messungen) heraus [19] • macht Lösungsvorschläge [19]
3.2 Ticketbearbeitung
<ul style="list-style-type: none"> • eröffnet ein Ticket (registriert Ticket) [16], [19], [20] • prüft Störungsmeldungen auf Service Level (Priorisierung) [16], [19], [20] • leitet Ticket an die richtige Stelle weiter und stößt weitere Bearbeitung an [16], [19] • verfolgt Ticket nach [16], [19] • koordiniert nachfolgende Support-Einheiten [16]
3.3 Integration (Anwendungen, Webseiten, Services)
<ul style="list-style-type: none"> • kann relevante Information auslesen und anbieten (abgeleitet aus [26]), [28] • kann notwendige Transaktionen anstoßen (abgeleitet aus [26]), [28]

Tabelle 2. Beispiele der Anforderungen an eine Frage

<i>Kriterium</i>	<i>Anforderung/ erwartete Antwort</i>
Relevanz	Der Chatbot versteht, dass es um das Produktangebot geht. Der Chatbot versteht, dass es um das günstigste Produkt geht. (2 Punkte)
Korrektheit	Es wird das günstigste Produkt genannt oder die Antwort geht auf das günstigste Produkt ein. (2 Punkte)
Lösungsgrad (mit Beispielen für Definition der erwarteten Lösung)	
2 Punkte	Kunde weiß, welches Produkt das günstigste ist. Er bekommt die Information direkt als Antwort oder als Link auf die entsprechende Seite.
1 Punkt	Kunde erhält einen Link oder Hinweis zu Produktseiten, muss dort aber selber nach dem günstigsten Produkt suchen. Oder Kunde wird mit einem Mitarbeiter verbunden und erhält so weitergehende Unterstützung.
0 Punkte	Der Kunde erhält keine Informationen dazu, wie er die günstigsten Produkte in Erfahrung bringen kann. Der Kunde wird lediglich an die allgemeine Telefonnummer oder die Webseite verwiesen.

3 Anwendung der Bewertungsmethode

Die Anwendung der Bewertungsmethode erfolgt durch eine „Befragung“ des zu bewertenden Chatbots sowie anschließende Bewertung der gegebenen Antworten anhand der spezifischen Kriterien und wird nachfolgend näher erläutert.

Der Fragenkatalog und die Bewertungskriterien bilden die zwei Grundelemente der Bewertungsmethode. Der Fragenkatalog enthält die aus den domänenspezifischen Anforderungen abgeleiteten Fragen bzw. Äußerungen. Durch die Eingabe dieser Fragen werden die Kommunikationsfähigkeit, das domänenspezifische Kundenverständnis und die Problemlösungsfähigkeit des Chatbots getestet. Die Bewertungskriterien sind das Mittel zur Beurteilung der Antwortqualität. Hier werden das Verständnis des Inputs und die Korrektheit bzw. Passgenauigkeit des Outputs sowie das Maß der durch die Antwort erreichten Problemlösung bewertet. Für die Durchführung der Bewertung wurde ein strukturierter Bewertungsbogen entwickelt, dessen Aufbau am Beispiel des Ausschnitts in Tabelle 3 erläutert wird.

Die Tabelle zeigt alle relevanten Elemente des Fragebogens. Auf der linken Seite sind die Fragen in ihren jeweiligen Kategorien und Unterkategorien aufgeführt. In der Spalte daneben steht die Anforderung an einen Chatbot, welche mit der Frage (teilweise auch mehrere) geprüft werden soll. Auf der rechten Seite wird der Chatbot mit seinen Antworten dokumentiert. Dabei wird nach Vor- und Nachteilen des Chatbots unterschieden. In die Felder können allgemeine Feststellungen, die bei der Durchführung der Methode auffallen, eingetragen werden. Darunter werden die Antworten des Chatbots mit zugehörigen Beobachtungen (Beob.) eingetragen, die aus den Antworten allein nicht hervorgehen. Daneben folgt die Bewertung der Kriterien Relevanz (R), Korrektheit (K) und Lösungsgrad (L). Die Summe (S) der Einzelbewertungen wird am Ende berechnet und eingetragen.

Tabelle 3. Ausschnitt des Bewertungsbogens

<i>Kategorie:</i> Problemlösungsfähigkeit		<i>Vorteile:</i>		<i>Nachteile:</i>			
<i>Unterkategorie:</i> Problemlösung							
<i>Frage</i>	<i>Anforderung</i>	<i>Antwort</i>	<i>Beob.</i>	<i>R</i>	<i>K</i>	<i>L</i>	<i>S</i>
Kannst du mich zu euren Angeboten beraten?	Beratungswunsch			0	0	0	0
Welches ist euer günstigstes Produkt?	Informationswunsch			0	0	0	0
Ich möchte mich über dich beschweren.	Umgang mit Beschwerden			0	0	0	0

R = Relevanz, K = Korrektheit, L = Lösungsgrad, S = Summe

4 Evaluation der Bewertungsmethode

Zur Evaluation im Rahmen des Design Science Research Prozesses wurde eine Testreihe durchgeführt und dabei die Bewertungsmethode auf mehrere Chatbots angewandt. Im Hinblick auf den angestrebten Einsatzzweck der Methode wurde zur Untersuchung von Bewertungsqualität und Zweckmäßigkeit eine Reihe von Evaluationskriterien festgelegt: a) Objektivität, Validität und Reliabilität dienten als Kriterien der Bewertungsqualität sowie b) Domänenbezug, Branchenunabhängigkeit, Technologieunabhängigkeit, Usability sowie die innere Ordnung der Fragen als Kriterien der Zweckmäßigkeit.

Um diese Kriterien zu überprüfen, wurden fünf geeignete Chatbots für die Durchführung ausgewählt: Lisa (O₂), Julia (Vodafone Kabel Deutschland), Digitaler Serviceassistent (Deutsche Telekom), Chatbock (Klarmobil) und Clara (Otto). Gründe für die Auswahl waren die Vergleichbarkeit der Funktionalitäten der Chatbots innerhalb einer Branche (Telekommunikation mit Lisa, Julia, Chatbock und Digitaler Serviceassistent) und zwischen zwei verschiedenen Branchen (Telekommunikation und Versandhandel mit Clara) zu überprüfen. Ebenso kann durch diese Auswahl die Branchenunabhängigkeit des Fragenkatalogs geprüft werden. Durch die Auswahl von Chatbock, der im Unterschied zu den direkt auf der Unternehmenswebsite integrierten Chatbots auf dem Facebook Messenger aufbaut, kann außerdem die Technologieunabhängigkeit des Fragebogens untersucht werden.

4.1 Ablauf der Evaluation

Bei der Durchführung der Evaluation wurde jedem Chatbot die Frage: „Welches ist euer günstigstes Produkt?“ gestellt. Diese Frage adressiert die Anforderung der Bearbeitung von Informationswünschen. Für jede Frage des Fragebogens wurde das im folgenden beschriebene Vorgehen durchgeführt:

Als erstes wurde dem Chatbot die Frage gestellt. Die Antwort des Chatbots wurde dann in das Protokoll eingetragen. Diese lautete bspw. beim Chatbot Julia: „Links

zeige ich Ihnen unsere günstigsten Kabeltarife“ ergänzt durch eine Auflistung von vier Tarifen inkl. Hyperlink. In die Spalte Beobachtung wurde nun eingetragen, dass auf der linken Seite Links zu den jeweils günstigsten Tarifen angezeigt werden. Danach konnte mit der Bewertung begonnen werden.

Der Chatbot hat erkannt, dass sowohl Informationen zu Produkten (hier Tarife) als auch von diesen Produkten die günstigsten gewünscht waren. Die Bewertung der Relevanz wurde daher mit zwei Punkten vorgenommen. Nach Prüfung der Links konnte bestätigt werden, dass diese Tarife die jeweils günstigsten der jeweiligen Kategorie sind. Die Korrektheit wurde deshalb mit zwei Punkten bewertet. Für die Bewertung des Lösungsgrads wurde die in Abschnitt 3.4 beschriebene Anforderung herangezogen. Da der Nutzer durch die Antwort weiß, welches die günstigsten Produkte sind und direkt Links zu diesen vorfindet, wurde auch der Lösungsgrad mit zwei Punkte bewertet. Allgemeine Feststellungen während der Evaluation konnten in den Zellen Vor- und Nachteile dokumentiert werden. Das vollständig ausgefüllte Protokoll für die Frage „Welches ist euer günstigstes Produkt?“ ist in Tabelle 4 dargestellt.

Tabelle 4. Evaluationsdurchführung

<i>Kategorie: Problemlösungsfähigkeit</i>		<i>Vorteile: schnelle</i>	<i>Nachteile:</i>				
<i>Unterkategorie: Problemlösung</i>		<i>Reaktionszeiten</i>					
<i>Frage</i>	<i>Anforderung</i>	<i>Antwort</i>	<i>Beob.</i>	<i>R</i>	<i>K</i>	<i>L</i>	<i>S</i>
Welches ist euer günstigstes Produkt?	Informationswunsch	Links zeige ich Ihnen unsere günstigsten Kabel-Tarife.	Links zu den Tarifen	2	2	2	6

Dieses Vorgehen wurde für jeden der Chatbots und jede Frage wiederholt. Dabei wurden die Chatbots nacheinander bewertet. Aus den Bewertungen der Chatbots und den Beobachtungen während der Durchführung wurden die im folgenden Kapitel dargestellten Evaluationsergebnisse abgeleitet.

4.2 Evaluationsergebnisse

Die Evaluation der Bewertungsmethode hat verschiedene Stärken und Schwächen der entwickelten Bewertungsmethode aufgezeigt. Diese werden nachfolgend anhand der Evaluationskriterien für Bewertungsqualität und Zweckmäßigkeit diskutiert.

Objektivität, Validität und Reliabilität. Die Objektivität der Methode ist dann sichergestellt, wenn diese unabhängig von äußeren Einflüssen durchgeführt werden kann. Die Methode ist in ihrer aktuellen Form nicht vollständig objektiv. Das liegt vor allem daran, dass die erwarteten Antworten in einigen Fällen nicht vorher definiert werden können, da nicht klar ist, welche Fähigkeiten die Chatbots im Detail haben. Die Güte der Antwort kann in diesen Fällen nicht eindeutig bestimmt werden, so dass die Bewertung vom subjektiven Eindruck des jeweiligen Prüfers abhängt.

Reliabilität bedeutet, dass eine Messung wiederholt werden kann und sich die Ergebnisse nicht oder erklärbar unterscheiden. Die Reliabilität der Methode ist gegeben, da es sich um statische Fragen handelt und der Chatbot in der Regel bei

gleichen Fragen die gleichen Antworten gibt. Eine Ausnahme hiervon stellt die Weiterentwicklung des Chatbots dar. Diese kann aber anhand der Differenz zur vorherigen Messung bewertet werden.

Eine Bewertung ist valide, wenn sie tatsächlich das misst, was mit ihr gemessen werden soll. Einfache Anforderungen, wie die in Abschnitt 4.1 angeführte Frage, können valide getestet und bewertet werden. Die Bewertung der Kriterien bei diesen Fragen mit null bis zwei Punkten hat sich bewährt, da sie nicht nur Ja-/Nein-Bewertungen zulässt, sondern eine differenzierte Bewertung ermöglicht. Die geringe Anzahl an Bewertungsalternativen erleichtert dabei die eindeutige Bewertung.

Eine Bewertung komplexer Themen, die von einem Chatbot nur im Dialog gelöst werden können, kann mit der aktuellen Bewertungsmethode nicht durchgeführt werden, was die Validität der Methode einschränkt.

Für einzelne Anforderungen kann durch den Bewertungsbogen nur die Text-/Sprachausgabe des Chatbots bewertet werden. Beispielsweise ist in Bezug auf die Vertraulichkeit der Daten der tatsächliche Umgang des Chatbots mit Personendaten nicht über die Benutzerschnittstelle feststellbar. Bei der Weiterentwicklung der Bewertungsmethode wäre demnach zu prüfen, ob Anforderungen mit einem sehr breiten Spektrum möglicher Kundenanliegen, wie z.B. die Anforderung „beantwortet Informationsanfragen“, weiter detailliert werden können, um Unschärfen in der Bewertung zu reduzieren.

Zusätzlich ist der Fakt zu beachten, dass sowohl Relevanz als auch Korrektheit Einfluss auf den Lösungsgrad haben. Versteht ein Chatbot die Frage nicht oder ist nicht in der Lage, eine korrekte Antwort zu liefern, wird er in der Regel weder bei der Relevanz bzw. Korrektheit noch beim Lösungsgrad Punkte erhalten.

Allgemeine Feststellungen, wie die Reaktionszeit der Chatbots, werden zwar in den Vor- und Nachteilen betrachtet, aber nicht bewertet. Dabei können gerade die hier festgestellten Eigenschaften des Chatbots erhebliche Auswirkungen auf den Kundennutzen haben. Um die Methode in dieser Hinsicht zu verbessern, können Erkenntnisse von Kuligowska [8] für die punktemäßige Bewertung von beobachtbaren Merkmalen einbezogen werden.

Domänenbezug. Der Domänenbezug konnte mit der Bewertungsmethode in der aktuellen Form nur teilweise erreicht werden, da typische Kundenanliegen im Servicedesk eine Dialogführung für ihre Erfassung und Bearbeitung benötigen.

Branchenunabhängigkeit. Die Branchenunabhängigkeit der Fragen konnte festgestellt werden. Dieses wurde durch die Auswahl von allgemeinen Fragen oder den Einsatz von Variablen (Bsp.: „Ich möchte mich über <Unternehmen> beschweren.“) in den Fragen sichergestellt. Chatbots aus zwei verschiedenen Branchen können dadurch in Bezug auf ihren Nutzen hinsichtlich typischer Servicedesk-Themen miteinander verglichen werden. Auch innerhalb einer Branche bietet die Methode Vergleichsmöglichkeiten.

Technologieunabhängigkeit. Mit Fokus auf die Technologieunabhängigkeit zeigte sich der Fragebogen in der Testreihe als ausreichend abstrakt, da die Nutzbarkeit des Fragebogens nicht durch die zugrundeliegende Technologie beschränkt wird. Aus Kundenperspektive muss der Nutzen unabhängig von der eingesetzten Technologie eintreten. Ergebnisse zu technologiebedingten

Nutzungsbeschränkungen des Chatbots können im Kopfteil des Fragebogens vermerkt werden (Bsp.: „Facebook-Account notwendig“ für den Chatbock).

Usability. Aus dem Blickwinkel der Usability zeigte sich, dass die Anwendung der Methode vergleichsweise wenig aufwendig ist. Für die Bewertung eines Chatbots wurden knapp zwei Stunden aufgewendet, technische Hilfsmittel sind nicht erforderlich. In einer der folgenden Iterationen zur Verbesserung der Methode soll dennoch versucht werden, die Anwendungszeit weiter zu verringern, etwa durch (Teil-)

Automatisierung.

Innere Ordnung. Die innere Ordnung der Fragen konnte durch die Zuordnung zu den Kategorien und Unterkategorien gewährleistet werden.

5 Fazit und Ausblick

In Beantwortung der Forschungsfrage ist eine Bewertungsmethode entstanden, die ausgehend vom fragebogenbasierten Vorgehen der Loebner-Preis-Methode die Spezifika eines im Servicedesk agierenden Chatbots integriert und dadurch die Nutzenbewertung eines solchen Chatbots durch den konstruierten Fragebogen mit zugehöriger Bewertungsmethode domänenspezifisch unterstützt.

Durch die systematische Erfassung der vielschichtigen Anforderungen an einen Chatbot als Grundlage zur Erstellung des Bewertungsbogens zeigt die Arbeit außerdem die Mannigfaltigkeit der von einem Servicedesk-Chatbot abzudeckenden Themenbereiche auf. Darüber hinaus konnte durch die Evaluation der Bewertungsmethode anhand von fünf Chatbots eine praxisorientierte Nutzungsmöglichkeit des erstellten Artefaktes aufgezeigt werden, die sich durch die Anwendbarkeit in verschiedenen Branchen sowie Unabhängigkeit von der technischen Implementierung auszeichnet.

Die aufgezeigten Beschränkungen des bisherigen Methodenentwurfs motivieren weiterführende Forschungsarbeiten mit dem Ziel, die Bewertungsmethode bspw. durch erweiterte Evaluation mit einer größeren Anzahl von Chatbots zu verbessern. Vorrangig sollte nach Möglichkeiten der Integration von Dialogverläufen in die Bewertung gesucht werden. Weiterhin sind Automatisierungspotenziale bei der Anwendung der Bewertungsmethode erkennbar, die zu einer Reduzierung des manuellen Aufwands führen könnten. Als reizvoller Ansatz wäre hier die Entwicklung eines Chatbots zur automatisierten Befragung des zu bewertenden Chatbots vorstellbar. Schließlich könnte geprüft werden, ob sich die Methode auch für eine reifegradbasierte Untersuchung von Chatbots erweitern ließe.

Literatur

1. Prifti, L., Heininger, R., Hecht, S., Krcmar, H.: Service Desk Requirements for Cloud Service Providers. In: Kundisch, D., Suhl, L., Beckmann, L. (eds.) Tagungsband Multikonferenz Wirtschaftsinformatik 2014 (MKWI 2014), pp. 1173-1184. Universität Paderborn (2014)

2. Knapp, D.: A Guide to Customer Service Skills for the Service Desk Professional, 3rd ed. Course Technology, Boston (2011)
3. Schatsky, D., Gratzke, P.: The conversational enterprise – how chatbots can answer questions for both customers and employees. Deloitte University Press (2016)
4. Heo, M., Lee, K.J.: Chatbot as a New Business Communication Tool. The Case of Naver TalkTalk. *Bus. Commun. Res. Pract.* 1, 41–45 (2018)
5. Fittkau & Maaß Consulting: W3B Exklusivstudie Chatbots und Social-Bots., <https://www.developer-week.de/studie-zu-chat-und-social-bots/> (Accessed on 22.07.2018)
6. Stoeckli, E., Uebersnickel, F., Brenner, W.: Exploring Affordances of Slack Integrations and Their Actualization Within Enterprises – Towards an Understanding of How Chatbots Create Value. In: 51st Hawaii International Conference on System Sciences, DOI: 10.24251/HICSS.2018.255 (2018)
7. Braun, D., Hernandez Mendez, A., Matthes, F., Langen, M.: Evaluating Natural Language Understanding Services for Conversational Question Answering Systems. In: SIGDIAL 2017 Conference, pp. 174–185. Association for Computational Linguistics, Saarbrücken, Germany (2017)
8. Kuligowska, K.: Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents. *Professionals Center for Business Research* 2, 1–16 (2015)
9. Shawar, B.A., Atwell, E.: Different measurements metrics to evaluate a chatbot system. In: Weng, F., Wang, Y.-Y., Tur, G., Hu, J. (eds.) *Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies, NAACL-HLT- Dialog 2007*, pp. 89–96. Association for Computational Linguistics, Stroudsburg, PA, USA (2007)
10. Stieglitz, S., Brachten, F., Kissmer, T.: Defining Bots in an Enterprise Context. In: 39th International Conference on Information Systems, San Francisco (2018), angenommen, <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1168&context=icis2018> (Accessed on 13.11.2018)
11. Brandtzæg, P.B., Følstad, A.: Why people use chatbots. In: Kompatsiaris, Y., Cave, J., Satsiou, A., Carle, G., Passani, A., Kontopoulos, E., Diplaris, S., McMillan, D. (eds.) *Internet Science. 4th Int. Conf., INSCI 2017, Thessaloniki, Greece*, pp. 377–392. Springer International Publishing, Cham (2017)
12. Jafarpour, S., Burges, C.J.C.: Filter, Rank, and Transfer the Knowledge: Learning to Chat, Microsoft Research Technical Report MSR-TR-2010-93 (2010)
13. Xu, A., Liu, Z., Guo, Y., Sinha, V., Akkiraju, R.: A New Chatbot for Customer Service on Social Media. In: Mark, G., Fussell, S., Lampe, C., Schraefel, M.C., Hourcade, J.P., Appert, C., Wigdor, D. (eds.) *Conference on Human Factors in Computing Systems, CHI '17*, pp. 3506–3510. ACM Press, New York, NY, USA (2017)
14. Følstad, A., Brandtzæg, P.B., Feltwell, T., Law, E.L.-C., Tscheligi, M., Luger, E.A.: Chatbots for Social Good. In: Mandryk, R., Hancock, M. (eds.) *Extended abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–4 (2018)
15. Jäntti, M., Shrestha, A., Cater-Steel, A.: Towards an Improved IT Service Desk System and Processes: A Case Study. *Int. J. on Advances in Systems and Measurements* 5(3&4), 203–215 (2012)
16. Kopperger, D., Kunsmann, J., Weisbecker, A.: IT-Servicemanagement. In: Tiemeyer, E. (ed.) *Handbuch IT-Management – Konzepte, Methoden, Lösungen und Arbeitshilfen für die Praxis.*, pp. 162–179. Carl Hanser Verlag, München (2013)
17. Fujitsu Services Ltd.: Der Service Desk der nächsten Generation. Unsere Welt steckt mitten in einer digitalen Revolution.,

- <https://sp.ts.fujitsu.com/dmsp/Publications/public/white-book-service-desk-digital-exec-sum-de.pdf> (Accessed on 17.06.2018)
18. Beims, M.: Das ITSM der Zukunft. Wie Digitalisierung das ITSM verändert, <https://www.cio.de/a/das-itsm-der-zukunft,3578603> (Accessed on 17.04.2018)
 19. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems* 24, 45–77 (2007)
 20. The Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB): Loebner Prize 2014 Contest Archive, Loebner Prize Selection Process, <http://www.aisb.org.uk/events/loebner-prize>, 04.07.2018)
 21. Mauldin, M.L.: Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. In: 12th National Conference on Artificial Intelligence (AAAI-94), pp. 16–21, Seattle (1994)
 22. Legg, S., Hutter, M.: Tests of Machine Intelligence. In: Lungarella, M. et al. (eds.): 50 Years of Artificial Intelligence. *Lecture Notes in Computer Science*, vol 4850 pp. 232–242. Springer, Berlin, Heidelberg (2007)
 23. Vasconcelos, M., Candello, H., Pinhanez, C., Donizetti dos Santos, T.: Bottester: Testing Conversational Systems with Simulated Users. In *Brazilian Symposium on Human Factors in Computing Systems, IHC '17* (2017)
 24. Walker, M.A., Litman, D.J., Kamm, C.A., Abella, A.: PARADISE. In: Cohen, P.R., Wahlster, W. (eds.) 35th Annual Meeting on Association for Computational Linguistics, pp. 271–280. Association for Computational Linguistics, Morristown, NJ, USA (1997)
 25. Proehl, T., Ereik, K., Limbach, F., Zarnekow, R.: Topics and Applied Theories in IT Service Management. In: 46th Hawaii International Conference on System Sciences, DOI: 10.1109/HICSS.2013.555 (2013)
 26. Elsässer, W.: ITIL einführen und umsetzen. Leitfaden für effizientes IT-Management durch Prozessorientierung. Hanser, München (2005)
 27. Fischlin, R.: Leitfaden ITIL-Service Desk. In: Müller, P. (ed.) 20. DFN-Jahrestagung, DFN 2006, pp. 95–104 (2006)
 28. Smiers, L.: How can Chatbots meet expectations? Introducing the Bot Maturity Model, <https://www.capgemini.com/2017/04/how-can-chatbots-meet-expectations-introducing-the-bot-maturity/> (Accessed on 17.04.2018)
 29. Nielsen, J.: 10 Usability Heuristics for User Interface Design, <https://www.nngroup.com/articles/ten-usability-heuristics/> (Accessed on 17.04.2018)
 30. Scott, K.: Usability Heuristics for Bots, <https://chatbotmagazine.com/usability-heuristics-for-bots-7075132d2c92> (Accessed on 17.04.2018)
 31. Makadia, M.: Challenges of Building an Intelligent Chatbot, <https://dzone.com/articles/challenges-in-building-an-intelligent-chatbot> (Accessed on 23.07.2018)
 32. Härtl, M.: Konzeption und Realisierung der technischen Unterstützung eines zentralen IT-Service-Desk mit OTRS an der T[echnischen] U[niversität] M[ünchen], Diplomarbeit (2007), <http://www.nm.ifi.lmu.de/pub/Diplomarbeiten/haer07/PDF-Version/haer07.pdf> (Accessed on 17.04.2018)
 33. Conrad, C.: Datenschutz bei neuen Kommunikationskanälen wie Chatbots und Messenger-Diensten: Der automatisierte Kundenkontakt der Zukunft, <https://www.datenschutz-notizen.de/datenschutz-bei-neuen-kommunikationskanaelen-wie-chatbots-und-messenger-diensten-der-automatisierte-kundenkontakt-der-zukunft-3520019/> (Accessed on 17.04.2018)