Universität Regensburg

Methoden und Ansätze für die zielgerichtete Gestaltung der Interaktionen mit modernen Informations- und Kommunikationssystemen

DISSERTATION

zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaft

Eingereicht an der Fakultät für
Wirtschaftswissenschaften der
Universität Regensburg

Vorgelegt von:
M. Sc. Wirtsch.-Inf. Daniel Konadl

Berichterstatter:
- Prof. Dr. Susanne Leist -
- Prof. Dr. Gregor Zellner -

Tag der Disputation: 03.06.2025

Inhaltsverzeichnis

Inhaltsverzeichnis I
Abbildungsverzeichnis VI
Tabellenverzeichnis VII
Abkürzungsverzeichnis VIII
Genderhinweis IX

1. Einleitung 1

1.1. Motivation, Problemstellungen und Zielstellungen 1
1.1.1. Gestaltung von Interaktionen im Rahmen von Social Media 2
1.1.2. Gestaltung von Interaktionen mit intelligenten, digitalen Kommunikationsagenten 9
1.2. Forschungsfragen 15
1.3. Aufbau der Dissertation 23

2. Wissenschaftliche Beiträge 25

2.1. Beitrag 1: Identifying Sentiment Influences Provoked by Context Factors – Results from a Data Analytics Procedure Performed on Tweets 28

1. Motivation 29
2. Conceptional basics 30
3. Related Work 30
4. Methodology 31
5. Preparing the analysis 32
5.1. Particularities of tweets 32
5.2. Data extraction, data preprocessing and data reduction 32
5.3. Selection of the methods to be applied 33
6. Performing the analysis 33
7. Reporting the results 34
7.1. Results of univariate analysis 34
7.2. Results of bivariate analysis 34
7.3. Interpretation and implications 36
8. Conclusion 36
9. References 37

2.2. Beitrag 2: Comparison of Topic Modelling Techniques in Marketing – Results From an Analysis of Distinctive Use Cases 39

1. Introduction 40
2. Theoretical Background 41
2.1. Social media 41
2.2. Topic modelling 42
2.3. Corporate use cases and requirements of marketing 43
3. Procedure of the Research 46
4. Selection of the Topic Modelling Techniques 46
5. Data Analysis 47
5.1. Data collection 47
5.2. Data cleansing and analysing 48
6. Discussion 50
7. Conclusion and Outlook 52
8. References 53

2.3. Beitrag 3: MANTRA: A Topic Modeling-Based Tool to Support Automated Trend Analysis on Unstructured Social Media Data 58

1. Introduction 59
2. Conceptual Foundations and Related Work 60
2.1. Conceptual Foundations 60
2.2. Distinctive Use Cases of Topic Modeling-Based Trend Analysis 61
2.3. Design Requirements for a Topic Modeling-Based Trend Analysis Tool 62
2.4. Assessment of Tools for Automated Trend Analysis on Social Media 64
3. Research Procedure 64
4. Design and Development of the Artifact 66
4.1. Design Principles for a Trend Analysis Tool 66
4.2. Technical Realization 67
5. Demonstration and Evaluation of MANTRA 68
6. Discussion and Contributions 70
7. Conclusion, Limitation, and Future Research 72
8. References 73

2.4. Beitrag 4: Supporting Product Development by a Trend Analysis Tool Applying Aspect-Based Sentiment Detection 76

1. Motivation 77
2. Foundations and Related Work 78
2.1. Conceptual Background 78
2.2. Design Requirements and Available Tools on the Market 79
3. Research Procedure 80

4. Design and Development....81
5. Demonstration and Discussion of the Artifact....83
5.1. Demonstration of the Artifact....83
5.2. Discussion of the Demonstration....84
6. Conclusion, Contribution and Outlook....85
7. References....87

2.5. Beitrag 5: Artificial Intelligence in Augmentative and Alternative Communication Systems – A Literature-Based Assessment and Implications of Different Conversation Phases and Contexts....90

1. Motivation....91
2. Conceptional Basics and Related Work....93
2.1. Conceptional Basics....93
2.2. Design Requirements of AAC Systems....94
2.3. AI Approaches and AI-based Potentials in the Context of AAC Systems....95
3. Research Approach....95
3.1. Collection and Evaluation of the Literature....95
3.2. Literature Analysis....96
3.2.1. Deriving the AI-based Potentials in the context of AAC Systems....96
3.2.2. Assigning AI-based Potentials to Conversation Phases and Conversational Contexts....97
4. Results of the Literature Analysis....97
5. Discussion and Future Research....99
5.1. Implications for Future Research....100
5.2. Limitations....103
6. Conclusion....103
7. References....104

2.6. Beitrag 6: A Generative AI-Based Approach to Support Automated Utterance Generation for Different Conversational Contexts within AAC Systems....109

1. Introduction....110
2. Conceptual Foundations and Related Work....111
2.1. Conceptual Foundations....111
2.2. Design Requirements of AI-Based Utterance Composition for AAC Systems....113
2.3. Assessment of Available AAC Systems on the Market....115
3. Research Procedure....115
4. Design and Development of the Artifact....117
4.1. Design of the AI-Based Utterance Composition for AAC Systems....117

4.2. Technical Realization 118
5. Demonstration and Evaluation of the Artifact 119
6. Discussion 122
6.1. Contributions to Theory and Practice 122
6.2. Limitations and Opportunities for Future Research 123
7. Conclusion 123
8. References 124

2.7. Beitrag 7: Comparison of Commercial Chatbot Solutions for Supporting Customer Interaction 127

1. Motivation 128
2. Foundations 129
2.1. Background on chatbots 129
2.2. Requirements for chatbot usage 130
3. Survey on current providers of chatbot solutions 131
3.1. Selection of chatbot solution providers for the study 132
3.2. Functionalities investigated 132
3.3. Questionnaire and collection of data 136
4. Results and Discussion 136
4.1. Results 136
4.2. Discussion and reflection against the requirements 138
5. Conclusion and Contribution 140
6. References 141

2.8. Beitrag 8: Chatbot Design Features to Increase Productivity 145

1. Introduction 146
2. Research Procedure 147
3. Results: Chatbot Design Features and Categories 148
4. Next Steps and Limitations 150
5. References 150

3. Schlussbetrachtung und Fazit 152
3.1. Zusammenfassung der Forschungsergebnisse 152
3.2. Beiträge für Wissenschaft und Praxis 156
3.2.1. Beiträge der Dissertation für die Wissenschaft 157
3.2.2. Beiträge der Dissertation für die Praxis 160
3.3. Kritische Würdigung und Ausblick auf weitere Forschungsfelder 164
3.3.1. Limitationen und weitere Forschungsarbeiten innerhalb der Themenfelder 164

3.3.2. Forschungsansätze aus der themenübergreifenden Betrachtung und Integration der Forschungsergebnisse 170

Literaturverzeichnis 176

Abbildungsverzeichnis

Abbildung 1: Aufbau und Überblick über die Dissertation.....23
Abbildung 2: Forschungssynergien aus den Themengebieten der Dissertation.....170

Tabellenverzeichnis

Tabelle 1: Zielstellungen und Forschungsfragen der Dissertation..... 24
Tabelle 2: Status, Ranking und Autorenanteile zu den wissenschaftlichen Beiträgen.....26
Tabelle 3: Fact Sheet zu Beitrag 1.....28
Tabelle 4: Fact Sheet zu Beitrag 2.....39
Tabelle 5: Fact Sheet zu Beitrag 3.....58
Tabelle 6: Fact Sheet zu Beitrag 4.....76
Tabelle 7: Fact Sheet zu Beitrag 5.....90
Tabelle 8: Fact Sheet zu Beitrag 6.....109
Tabelle 9: Fact Sheet zu Beitrag 7.....127
Tabelle 10: Fact Sheet zu Beitrag 8.....145
Tabelle 11: Limitationen und zukünftige Forschungsmöglichkeiten.....165

Abkürzungsverzeichnis

AAC	Augmentative and Alternative Communication
B2C	Business to Consumer
bzw.	beziehungsweise
CLR	Computational Literature Review
DSR	Design Science Research
DMR	Dirichlet Multinomial Regression
et al.	et alia
ggf.	gegebenenfalls
GPT	Generative Pretrained Transformer
GQM	Goal Question Metric
GUI	Graphical User Interface
inkl.	inklusive
IS	Information Systems
IT	Information Technology
KI	Künstliche Intelligenz
LDA	Latent Dirichlet Allocation
Mio.	Millionen
Mrd.	Milliarden
NLP	Natural Language Processing
NLU	Natural Language Understanding
OS	Operating System
PAM	Pachinko Allocation Model
sog.	sogenannt
u. a.	unter anderem
v. a.	vor allem
vgl.	vergleiche
VoC	Voice of the Customer
WoM	Word of Mouth
z. B.	zum Beispiel

Genderhinweis

Zur besseren Lesbarkeit wird in dieser Arbeit auf die gleichzeitige Verwendung der weiblichen, männlichen und diversen Sprachformen verzichtet. Alle personenbezogenen Bezeichnungen gelten gleichermaßen für alle Geschlechter.

Die gewählte Form dient ausschließlich der sprachlichen Vereinfachung und impliziert keine Benachteiligung oder Bevorzugung eines bestimmten Geschlechts.

1. Einleitung

Dieses Kapitel definiert den grundlegenden Rahmen und den Kontext der vorliegenden Dissertation. In Kapitel 1.1 werden zunächst die Themen der Forschungsarbeiten motiviert, die zugrundeliegenden Problemstellungen erläutert und die Zielstellungen formuliert, welche mit der Dissertation erreicht werden sollen. Darauf aufbauend werden in Kapitel 1.2 insgesamt elf Forschungsfragen entwickelt, die im weiteren Verlauf der Arbeit adressiert und beantwortet werden. Abschließend bietet Kapitel 1.3 einen Überblick über den strukturellen Aufbau der Dissertation und das Zusammenwirken der Zielstellungen und Forschungsfragen für die jeweiligen Veröffentlichungen.

1.1. Motivation, Problemstellungen und Zielstellungen

Die Entstehung neuer Technologien stellt sowohl Unternehmen als auch die Gesellschaft vor große Herausforderungen. Gleichzeitig bieten moderne, digitale Informations- und Kommunikationssysteme in diesen beiden Bereichen vielfältige Chancen und Potenziale. Es können dadurch Lösungen generiert werden, welche bestehende Probleme noch besser lösen können und gleichzeitig eng auf die Bedürfnisse und Herausforderungen ihrer Nutzer abgestimmt sind (vgl. Maedche et al., 2019; Van der Aalst et al., 2018; Vom Brocke et al., 2018).

Vor diesem Hintergrund verfolgt die vorliegende Dissertation das zentrale Ziel, ausgewählte Ansätze und Technologien im Rahmen konkreter Anwendungsszenarien zu bewerten und zielgerichtete Adaptionen und Empfehlungen für deren Einsatz zu formulieren. Mit verschiedenen gestaltungsorientierten Forschungsbeiträgen wird dadurch die Lösung von unterschiedlichen operativen und gesellschaftlichen Problemen an der Schnittstelle der Interaktionen von Individuen mit modernen und digitalen Informations- und Kommunikationssystemen bewerkstelligt.

Im Folgenden sollen nun zunächst die beiden zentralen Themenbereiche, Social Media und digitale intelligente Kommunikationsagenten, vorgestellt werden, welche im Rahmen der kumulativen Dissertation tiefergehend behandelt worden sind.

1.1.1. Gestaltung von Interaktionen im Rahmen von Social Media

Die vorliegende Dissertation greift einerseits den Themenbereich Social Media auf, um einen Beitrag zur Gestaltung von Interaktionen mit diesen modernen digitalen Informations- und Kommunikationssystemen zu leisten.

In der wissenschaftlichen Literatur wird der Begriff *„Social Media“* als Gruppe von internetbasierten Anwendungen definiert, welche auf den ideologischen und technologischen Grundlagen des Web 2.0 aufbauen und die Erstellung sowie den Austausch von nutzergenerierten Inhalten ermöglichen (Kaplan und Haenlein, 2010). Auf diese Weise ermöglicht es Social Media seinen Nutzern, sowohl bidirektional über Chatfunktionalitäten als auch öffentlich über das Posten von Beiträgen mit anderen Nutzern in Austausch zu treten (Hanna et al., 2011; Kietzmann et al., 2011).

Aktuellen Statistiken zufolge nutzen weltweit immer mehr Menschen soziale Netzwerke. Die Nutzerzahlen sind seit 2018 mit 3,2 Mrd. Nutzern bis zum Jahr 2023 mit 4,76 Mrd. Nutzern um fast 50 % angestiegen (Statista, 2024). Social Media dient dabei nicht nur als Kommunikationskanal zwischen Privatpersonen, sondern wird sowohl von Kunden als auch von Unternehmen dazu genutzt, um mit der jeweils anderen Seite in Kontakt und Austausch zu treten (Kaplan und Haenlein, 2010, Kietzmann et al., 2011). Online Communities als themenspezifische Interessensgruppen in sozialen Netzwerken bieten in diesem Kontext für Unternehmen eine effektive Möglichkeit, um mit ihren Zielgruppen zu kommunizieren. Auf diese Weise können Unternehmen (potenzielle) Kunden ansprechen und Marketingbotschaften sowie Informationen über Produkte, Dienstleistungen oder Sonder- und Rabattaktionen bei diesen direkt platzieren (Hanna et al., 2011; Kietzmann et al., 2011). Die große Reichweite von sozialen Medien ermöglicht dabei, dass Unternehmen eine Vielzahl von Personen erreichen können (vgl. Gabriel und Röhrs, 2017; Kaplan und Haenlein, 2010; Pein, 2014).

Jedoch können sich im Rahmen der Kundenkommunikation über Social Media auch Probleme für Unternehmen ergeben (Gabriel und Röhrs, 2017; Pein, 2014). Diskussionen, in denen kontroverse Ideen und kritische Themen (z. B. Kritik über Produkte oder die Unternehmenspolitik) diskutiert werden, müssen nicht zwangsläufig immer ruhig und gesittet verlaufen. Im schlimmsten Fall können daraus Shitstorms und eine

Vielzahl negativer Reaktionen und schädlicher Social Media Posts für ein Unternehmen resultieren (Einwiller und Steilen, 2015; Grégoire et al., 2015; Kim et al., 2016; Pfeffer et al., 2014). Durch solche unternehmenskritischen Social Media Posts beeinflusst werden Kunden die Leistungen des Unternehmens möglicherweise nicht mehr in Anspruch nehmen und stattdessen zur Konkurrenz wechseln. Deshalb müssen Unternehmen in der Gestaltung ihrer Social Media Posts zwingend darauf achten, dass diese nicht zu Reputationsschäden und zu einer negativen Unternehmenswahrnehmung bei (potenziellen) Kunden führen.

Selbst wenn Unternehmen aber alle Aspekte bei der Formulierung von Social Media Posts berücksichtigen (z. B. die Stimmung und den Wahrheitsgehalt für die zu übermittelnde Botschaft), gibt es noch weitere, kontextuelle Einflussfaktoren (vgl. Dey, 2001), welche eine wichtige Rolle spielen. Sogenannte Kontextfaktoren, (z. B. Ort, Zeit, personenbezogene und demographische Merkmale wie das Geschlecht, das Alter, oder der soziale Status einer Person) (vgl. Cao et al., 2018; Cruz-Albrecht et al., 2017; Dodds et al., 2011; Golder und Macy, 2011; Lim et al., 2018; Lin, 2014; Plunz et al., 2019; Schwartz et al., 2019) beeinflussen maßgeblich die Wahrnehmung von Social Media Posts. So könnte ein Kunde am Montagmorgen aufgrund der typischerweise schlechteren Stimmung zu dieser Tagesphase und an diesem Wochentag negativ auf einen Social Media Post zu einem Produktrelease reagieren, während am Wochenende am Samstag oder Sonntag derselbe Kunde auf dieselbe Markenbotschaft mit einem positiven Kommentar reagieren würde (vgl. Cao et al., 2018; Golder und Macy, 2011; Hu et al., 2019; Wang et al., 2016). Ebenso könnte eine Reaktion auf einen Social Media Post negativer ausfallen, wenn sich ein Nutzer gerade am Arbeitsplatz befindet und deswegen gestresst und negativer gestimmt ist. Im Gegensatz dazu würde eine Person, welche sich gerade im Urlaub befindet und deshalb grundsätzlich positiver gestimmt ist, auf denselben Post auch positiver gestimmt reagieren (vgl. Golder und Macy, 2011; Hu et al., 2019; Mitchell et al., 2013).

In diesem Zusammenhang sind neben den Einflüssen einzelner Kontextfaktoren (z. B. die Einflüsse der Faktoren Ort und Zeit werden isoliert voneinander betrachtet) insbesondere diejenigen Effekte von Bedeutung, welche aus der Kombination von mehreren Kontextfaktoren resultieren (z. B. Reaktion auf einen Post zu einem Produktrelease

am Montagmorgen am Arbeitsplatz). Diese feingranularen Einflüsse der Kontextfaktoren können für Unternehmen ein verlässliches Instrument darstellen, um in der Interaktion mit (potenziellen) Kunden über Social Media Reputationsschäden, negative Kundenwahrnehmungen und Shitstorms zu verhindern. Um die Effekte von Kontextfaktoren tiefergehend zu untersuchen und das gewonnene Analysewissen als ein Planungswerkzeug für die Gestaltung der Social Media Kommunikation für Unternehmen bereitstellen zu können, wird die folgende Zielstellung aufgestellt:

Zielstellung 1 (ZS 1):
Identifikation der Einflüsse, sowohl von einzelnen als auch von kombinierten Kontextfaktoren, im Hinblick auf die in Social Media Posts zum Ausdruck gebrachten Stimmungen.

Für Kunden ergibt sich mit Social Media neben dem Austausch im Ladengeschäft, via Telefon oder via E-Mail ein weiterer Kommunikationskanal, um die sog. *„Voice of the Customer“* (vgl. Pande et al., 2014) zu artikulieren. Damit können sie ihre Anforderungen, Erwartungen, Meinungen und Wünsche zu bestimmten Produkten, Services oder dem Unternehmen im Allgemeinen öffentlichkeitswirksam mitteilen (Einwiller und Steilen, 2015; Hicks et al., 2012; Jansen et al., 2009).

Die Identifikation dieser VoC ist für Unternehmen unerlässlich, um mit ihrem Leistungsangebot auf die Bedürfnisse, sowohl von aktuellen als auch von potenziellen, noch nicht erschlossenen Kundengruppen, eingehen zu können (Hong et al., 2012; Lozano et al., 2017). Die Interaktion von Unternehmen mit sozialen Medien und die darauf basierende frühzeitige Erkennung von neuen und vielversprechenden Ideen und Trends in den Bereichen Produktentwicklung, Kundenverhaltensanalyse und Markt-/Markenbeobachtung kann zu entscheidenden Wettbewerbsvorteilen führen (Bhor et al., 2018; Jeong et al., 2019). Trends sind komplexe, mehrdimensionale Phänomene, welche sich im Laufe der Zeit (weiter-)entwickeln und, sofern sie sich durchsetzen, einen erheblichen Einfluss auf die Gesellschaft ausüben können. In der heutigen, schnelllebigen und wettbewerbsorientierten Geschäftswelt ist es für Unternehmen daher von entscheidender Bedeutung, Trends kontinuierlich zu beobachten (Endo et al., 2015; Lozano et al., 2017). Soziale Medien bieten in diesem Kontext für Trendsetter

sowie für Kunden eine ideale Möglichkeit, um aufkommende Trends (z. B. Technologie-, Mode- oder Lebensmitteltrends) offen zu diskutieren und diesbezüglich ihre ungefilterten Meinungen zu äußern (Jeong et al., 2019; Park et al., 2021; Yan et al., 2014). Auf dieser Grundlage können Unternehmen wertvolle Einblicke darüber gewinnen, was aktuell als Trend diskutiert wird, welche möglichen Entwicklungen der Trends zu erwarten sind und wie das zukünftige Leistungsangebot unter Berücksichtigung der geäußerten Meinungen, Bedürfnisse und Vermutungen konkret ausgestaltet werden könnte (Tucker und Kim, 2011; Zhong und Schweidel, 2020).

Um Trends auf der Grundlage von Social Media Beiträgen und dabei insbesondere ihrer textuellen Bestandteile zu identifizieren, werden von den am Markt verfügbaren Softwarelösungen vor allem schlagwortbasierte Frequenzanalysen angewendet (Wörner et al., 2023). Diese Ansätze ermitteln Trends durch die Auftrittshäufigkeiten von vorab definierten Begriffen oder Hashtags (Lau et al., 2012). Die damit generierten Ergebnisse erweisen sich jedoch als äußerst problematisch, weil Trends nicht zwangsläufig durch im Vorhinein bekannte oder besonders häufig auftretende Begriffe bestimmbar sind. Es deuten vielmehr bestimmte Kookkurrenzen von Begriffen, die nicht ungewöhnlich aktuell oder häufig zusammen erwähnt werden, auf einen Trend hin (Khan et al., 2021; Reisenbichler und Reutterer, 2019). Vor diesem Hintergrund wurde in der Forschungsliteratur das Topic Modelling als automatisierter Ansatz zur Themenextraktion umfassend untersucht (Chinnov et al., 2015; Eickhoff und Neuss, 2017; Hong und Davison, 2010). Topic Modelling Techniken ermöglichen es, große Mengen textueller Daten zu analysieren und die enthaltenen Themen basierend auf den Kookkurrenzen der in den Texten enthaltenen Begriffen zu extrahieren (Chinnov et al., 2015; Debortoli et al., 2016; Eickhoff und Neuss, 2017).

Die Interaktion mit den Textbestandteilen von Social Media Beiträgen für die Informationsextraktion mithilfe von Topic Modelling stellt für Unternehmen eine komplexe Herausforderung dar. Die Qualität der extrahierten Themen hängt dabei maßgeblich von der angewendeten Topic Modeling Technik ab. In vielen theoretischen und praktischen Untersuchungen wurde häufig auf LDA (vgl. Blei et al., 2003) als Standardtechnik zurückgegriffen. Allerdings erweist sich die LDA Technik für verschiedene marketingbezogene Anwendungsfälle und ihre spezifischen Anforderungen nicht

immer als die optimale Analysetechnik. In der einschlägigen Forschungsliteratur wurden deshalb verschiedene Erweiterungen zur LDA Technik vorgeschlagen (vgl. Liu et al., 2016; Vayansky und Kumar, 2020), um den Anforderungen verschiedener Anwendungsfälle gerecht werden zu können. Trotzdem bringen auch diese Topic Modelling Techniken immer noch verschiedene Herausforderungen mit sich.

Eine Vielzahl dieser Topic Modelling Techniken ist ausschließlich theoretisch beschrieben und nicht praktisch anwendbar, weil es keine lauffähige Implementierung dieser Techniken gibt (Vayansky und Kumar, 2020). Zudem erfordert die auf diesen Techniken basierte Informationsextraktion aus textuellen Social Media Inhalten oftmals fortgeschrittene und technikbezogene Kenntnisse. Diese Umstände erschweren in erheblicher Weise die Interpretation und zielführende Anwendung der Ergebnisse, insbesondere für Personen wie z. B. Marketingexperten, welche diese technikbezogenen Kenntnisse in der Regel nicht aufweisen (vgl. Wörner et al., 2021).

Neben der Auswahl einer zielführenden Technik stellt auch deren konkrete und praktische Umsetzung im Hinblick auf verschiedene Einsatzszenarien, v. a. im Bereich der Trendanalyse, eine Herausforderung dar. Die auf dem Markt existierenden Trendanalysetools (z. B. Brandwatch, Meltwater, Symanto) weisen erhebliche Nachteile im Hinblick auf die Extraktion von Trends aus Social Media Posts auf (Wörner et al., 2023). Diese Softwaretools decken dabei nicht umfassend diejenigen Anforderungen ab, welche für Trendanalysen in Bezug auf marketingbezogene Anwendungsfälle (z. B. Produktentwicklung, Analyse von Kundenverhalten und Marktanalyse) essenziell wichtig sind. Insbesondere sind die vorhandenen Trendanalysetools nicht in der Lage, verschiedene externe Parameter wie Geolokationen sowie zeitgrößen- und personenbezogene Faktoren flexibel und in unterschiedlicher Granularität bei der Extraktion von Trends einzubeziehen (Wörner et al., 2022; Wörner et al., 2023).

In der Tat aber unterscheiden sich die Wahrnehmungen von Kunden im Hinblick auf Produkte und Dienstleistungen teilweise erheblich über verschiedene Kontinente, Länder und Bundesländer hinweg aufgrund global und lokal spezifischer Erwartungen und Wertevorstellungen (vgl. Bae et al., 2018; Ding et al., 2020; Endo et al., 2015; Ha et al., 2017; Moro et al., 2020; Park et al., 2021). Das Verhalten von Kunden variiert ebenfalls in Abhängigkeit von verschiedenen, personenbezogenen Merkmalen wie z.

B. Alter, Geschlecht sowie gruppenbezogenen Einstellungen und Wertevorstellungen (vgl. Bhor et al., 2018; Shi et al., 2018; Zhang et al., 2016). Auch die Analyse von Trends auf Basis verschiedener zeitlicher Granularitäten (z. B. Wochentage, Tagesphasen, Jahreszeiten) unter zusätzlicher Berücksichtigung von Geolokationen und Kundenmerkmalen bietet für Marketingabteilungen wertvolle und feingranulare Erkenntnisse aus der Interaktion mit Social Media Inhalten.

Damit Marketingabteilungen die VoC effektiv nutzen können, müssen neben der Funktionalität zur Trendextraktion noch weitere Anforderungen durch ein Trendanalysetool erfüllt werden (Wörner et al., 2022; Wörner et al., 2023). Besonders zentral ist dabei die Analyse der in den Social Media Beiträgen zum Ausdruck gebrachten Sentimente (Irawan et al., 2020; Jeong et al., 2019; Mastrogiacomo et al., 2021; Moro et al., 2020; Tucker und Kim, 2011). Erst dadurch wird für Marketingverantwortliche ersichtlich, welche Themen von Kunden in einer positiven oder einer negativen Tonalität diskutiert worden sind. Für zielgerichtete Entscheidungen im Hinblick auf Verbesserungen von Produkten und Services ist dabei wichtig, dass die Sentimente feingranular für die Bestandteile und Merkmale von Produkten und Services ermittelt werden (vgl. Mirtalaie et al., 2018; Tucker und Kim, 2011; Tuarob und Tucker, 2015; Vo et al., 2018). Hierzu bietet sich eine Kombination von Topic Modelling (Identifikation der Features und Merkmale von Produkten und Services) mit der aspektbasierten Sentimentanalyse (Bestimmung der Sentimente gegenüber den Features und Merkmalen) an.

Das Potenzial der aspektbasierten Sentimentanalyse zur detaillierten Nachverfolgung von Trends im Zeitverlauf wurde auch bereits in Wissenschaft und Praxis erkannt (z. B. Schouten und Frasincar, 2015; Tuarob und Tucker, 2015; Tucker und Kim, 2011). Dies hat zur Entwicklung von Trendanalysetools geführt, welche Funktionen der aspektbasierten Sentimentanalyse integrieren (vgl. Wörner et al., 2022). Allerdings weisen die auf dem Markt erhältlichen Trendanalysetools dahingehend erhebliche Defizite auf. Insbesondere können diese Softwaretools verschiedene, externe Parameter (z. B. Geolokationen, zeitbezogene und personenbezogene Merkmale) ebenfalls nicht umfassend und flexibel in die Identifikation von Trends einbeziehen. In Konsequenz wird daher die Interaktion mit den Textbestandteilen von Social Media Posts zur Extraktion kontextualisierter Trends durch die vorhandenen Trendanalysetools unvollständig und

unzureichend unterstützt. Darüber hinausgehend ermöglichen es diese Trendanalysetools nicht, das Fach- und Domänenwissen von Marketingverantwortlichen in die Trendanalyse einzubeziehen (z. B. Tuarob und Tucker, 2015; Tucker und Kim, 2011; Vo et al., 2018). Jedoch bietet genau dieses Wissen, das z. B. in Form von Ontologien, Taxonomien oder Spezifikationen über die Bestandteile von Produkten und Services formalisiert sein kann, die Möglichkeit, interessierende Aspekte sowie thematische Zusammenhänge effektiv und gezielt in die Trendanalyse einzubeziehen.

Die hier beschriebenen Problemstellungen sind den Bereichen Social Media und Social Media Analytics sowie dem Bereich automatisierter Trendanalysen, insbesondere auf Basis von Topic Modelling und der aspektbasierten Sentimentanalyse, zuzuordnen. Die vorliegende Dissertation leistet dabei verschiedene Beiträge, um existierende Ansätze, Techniken und Softwaretools für die automatisierte Informationsextraktion basierend auf den Textbestandteilen von Social Media Beiträgen zu untersuchen und weiterzuentwickeln. Einerseits ergibt sich daraus eine Verbesserung der Interaktionen von Marketingverantwortlichen bei der Informationsextraktion aus Social Media Netzwerken als modernen digitalen Informationssystemen. Andererseits eröffnen die Ergebnisse der Dissertation eine Verbesserung in Form automatisierter Analysetools, die in der Lage sind, die Anforderungen verschiedener marketingbezogener Anwendungsfälle umfassend und unter Einbezug zielführender Analysetechniken abzudecken. Dem zugrundeliegend werden die folgenden drei Zielstellungen abgeleitet:

Zielstellung 2 (ZS 2):
Vergleich von verschiedenen Topic Modelling Techniken mit dem Ziel, für ausgewählte, marketingbezogene Anwendungsfälle Handlungsempfehlungen zur Auswahl von zielführenden Topic Modelling Techniken zu geben.

Zielstellung 3 (ZS 3):
Ableitung von Anforderungen mit dem Ziel, Trendanalysetools für verschiedene, marketingbezogene Anwendungsfälle zu konzipieren. Insbesondere sollen dabei als zentrale Designprämissen das Topic Modelling und die aspektbasierte Sentimentanalyse als Analysetechniken untersucht und einbezogen werden.

Zielstellung 4 (ZS 4):
Entwicklung und Evaluation von Trendanalysetools zur automatisierten Analyse von textuellen Social Media Inhalten unter Berücksichtigung verschiedener anwendungsfallbezogener Anforderungen.

1.1.2. Gestaltung von Interaktionen mit intelligenten, digitalen Kommunikationsagenten

Neben den Beiträgen zur Gestaltung von Interaktionen mit Social Media beschäftigt sich die vorliegende Dissertation andererseits auch mit der Untersuchung, Weiterentwicklung und Gestaltung von Interaktionen mit intelligenten, digitalen (KI-basierten) Kommunikationsagenten.

KI zählt zu den wegweisenden Themen des 21. Jahrhunderts und hat sowohl in der wissenschaftlichen Literatur (vgl. Lu et al., 2018; Nilsson, 2014; Russell et al., 2015) als auch in der Gesellschaft (vgl. Fast und Horvitz, 2017) immense Aufmerksamkeit erlangt. Diese wachsende Relevanz zeigt sich insbesondere in den Investitionen, welche Unternehmen und Regierungen in diesem Themenbereich tätigen. Im Jahr 2016 haben Unternehmen schätzungsweise 26 bis 36 Mrd. US-Dollar in KI investiert (vgl. Bughin et al., 2017). Die Tendenz dabei ist deutlich steigend, und auch Regierungen sind bestrebt, ihre Ausgaben für KI-Forschung aufgrund der sich eröffnenden Potenziale weiter zu steigern. Der Gesamtmarkt für KI-Anwendungen wird Schätzungen zufolge bis 2025 ein Volumen von 127 Mrd. US-Dollar erreichen (Barton et al., 2017).

Bezüglich des Begriffs der *„Künstlichen Intelligenz“* existieren in der Literatur verschiedene Definitionen. Winston (1992) beschreibt KI dabei als das Bestreben, dass ein Computer in die Lage versetzt wird, über grundlegende menschliche Fähigkeiten wie Wahrnehmen, Denken und Handeln zu verfügen (vgl. Winston, 1992). Dieses Verständnis von KI wird in der vorliegenden Dissertation verfolgt und angewendet, um verschiedene Gestaltungsmöglichkeiten für intelligente Kommunikationsassistenten zu untersuchen. Auf diese Weise sollen zunächst bestehende, technische Unterstützungsmöglichkeiten für sprachbehinderte Menschen zielgerichtet verbessert werden.

Laut Hourcade et al. (2004) ist der Austausch von Gedanken, Ideen, Bedürfnissen, Wünschen und Gefühlen mit anderen Personen eine zentrale Eigenschaft des Menschseins (Hourcade et al., 2004). In diesem Kontext wird die lautsprachliche Kommunikation als das wichtigste Instrument eines Individuums angesehen, um mit seiner Umgebung in Kontakt zu treten (vgl. Fritzell, 1996; Kane et al., 2017; Pollak und Gallagher, 1989). Personen, welche an schweren körperlichen Behinderungen oder Hirnverletzungen leiden, sind jedoch nicht in der Lage, ihre oral-respiratorische Muskulatur für das Sprechen ausreichend zu kontrollieren (Allen, 2005; Vanderheiden, 1983). Diese Einschränkung der lautsprachlichen Kommunikationsfähigkeit und damit die Ausgrenzung aus der *„gesellschaftlichen Mitte“* (Hourcade et al., 2004) trifft ebenso auf Menschen mit Schädigungen des Vokaltrakts oder anderen Behinderungen zu (z. B. Aphasie, Autismus), welche die Sprachfähigkeit einschränken (Allen, 2005).

Damit sich Menschen mit Sprachbehinderungen mit ihrer Umgebung austauschen können, müssen sie auf alternative Kommunikationsformen zurückgreifen, die verschiedene Strategien und Techniken zur Förderung zielgerichteter Kommunikation kombinieren (American Speech Language Hearing Association, 1991). Die *„Augmentative and Alternative Communication“* umfasst einerseits analoge Kommunikationsformen wie die Gebärdensprache, bei der Ausdrücke durch Mimik und Gestik gebildet werden. Diese Ausdrücke können jedoch leicht übersehen oder missverstanden werden, wenn Gesprächspartner wenig oder gar keine Kenntnisse in der Gebärdensprache besitzen (Boyes-Braem et al., 1994). Im Vergleich zur Lautsprache fehlt es der Gebärdensprache zusätzlich an Ausdrucksmöglichkeiten, weil sie nur wenige Präpositionen und Konjunktionen und auch keine Genusmarker zur Unterscheidung zwischen Maskulinum und Femininum bei Substantiven bereitstellt (Louis-Nouvertné, 2001).

Um Menschen mit Sprachbehinderungen die Möglichkeit einer lautsprachlichen Kommunikation zu eröffnen, sind digitale AAC-Systeme entwickelt worden. Diese AAC-Systeme ermöglichen als Softwareanwendungen die Übersetzung der von einer Person beabsichtigten Ausdrücke in Sprache und eine anschließende Sprachausgabe ihrer Gedanken, Bedürfnisse, Wünsche und Gefühle (vgl. Bradshaw, 2013). Der positive Einfluss dieser sprachgenerierenden Systeme auf das Interaktionsverhalten von sprachbehinderten Menschen konnte bereits in zahlreichen Studien nachgewiesen werden (z.

B. Brady, 2000; Desai et al., 2014; Schepis et al., 1996). Bestehende AAC-Systeme weisen jedoch verschiedene, gravierende Probleme auf, welche Menschen mit Sprachbehinderungen in verschiedenen Gesprächssituationen im Alltag vor erhebliche Herausforderungen stellen. Dazu gehören niedrige Gesprächsraten, ein hoher Aufwand bei der Erstellung der Gesprächsbeiträge für die sprachbehinderten Personen und die unzureichende Anpassung der Gesprächsbeiträge auf die Persönlichkeit einer sprachbehinderten Person (vgl. Dempster et al., 2010; Laxmidas et al., 2021; Obiorah et al., 2021). Daneben werden wesentliche Aspekte von Gesprächen wie die Unterstützung von zentralen Gesprächsphasen (u. a. Gesprächsbeginn, Gesprächsmitte, Gesprächsende) (vgl. Henne und Rehbock, 2012; Spiegel und Spranz-Fogasy, 2001) sowie von unterschiedlichen Gesprächskontexten (z. B. formelle und informelle Gesprächssituationen) (Bedrosian et al., 2003; Hoag et al., 2004; Todman und Alm, 1997) unzureichend durch bestehende AAC-Systeme unterstützt (vgl. Konadl et al., 2023; Konadl, 2024). Vor diesem Hintergrund verspricht der Einsatz von passgenauen, KI-basierten Ansätzen aussichtsreiche Lösungsmöglichkeiten für AAC-Systemen.

Obwohl sich die wissenschaftliche Literatur mit dem Einsatz von KI-basierten Ansätzen in AAC-Systemen bereits intensiv auseinandergesetzt hat, fehlt es an einem strukturierten und konsolidierten Überblick darüber, welche Potenziale durch KI-basierte Ansätze im Rahmen von AAC-Systemen bislang realisiert worden sind. Es bleibt dabei insbesondere unbeantwortet, ob und in welcher Weise diese KI-basierten Ansätze diese Potenziale in zentralen Gesprächsphasen und Gesprächskontexten realisiert haben. Ein literaturbasierter Überblick über den Einsatz und die Gestaltung von KI-basierten Ansätzen für AAC-Systeme bietet die Möglichkeit, bestehende Lücken zu identifizieren, Vorschläge zu deren Schließung zu unterbreiten und damit die Interaktionen von sprachbehinderten Menschen mit AAC-Systemen entscheidend verbessern zu können.

Eine Verbesserung der Interaktionen von sprachbehinderten Menschen mit AAC-Systemen könnte sich auch durch das Einbeziehen neuer, digitaler Technologien ergeben. ChatGPT als eine solche moderne, digitale Technologie hat in den letzten Jahren mit seinen Fähigkeiten zur Generierung textueller Beiträge für alltägliche Aufgaben und Situationen in unterschiedlichen Formulierungsstilen und Ausdrucksweisen große

Aufmerksamkeit erlangen können (Banh und Strobel, 2023; Dwivedi et al., 2023; Feuerriegel et al., 2023; Fui-Hoon et al., 2023; Teubner et al., 2023). Die generierten Textbeiträge sind dabei von denen, welche von Menschen verfasst worden sind, kaum mehr zu unterscheiden. Dies ist darauf zurückzuführen, dass das bei ChatGPT verwendete KI-Sprachmodell auf Texten basiert, die von Menschen zu einer Vielzahl verschiedenster Themen (u. a. zu Inhalten alltäglicher Gespräche und Begebenheiten) geschrieben wurden (vgl. Banh und Strobel, 2023; Feuerriegel et al., 2023; Radford et al., 2019; Teubner et al., 2023). Diese Fähigkeit von Generativer KI erscheint im Hinblick auf die oben genannten Probleme für AAC-Systeme als eine vielversprechende Lösung, um basierend auf einzelnen Wörtern und Begriffen, wie sie in AAC-Systemen eingegeben werden, komplette und passgenaue Gesprächsbeiträge für verschiedene Gesprächssituationen zu generieren.

Angesichts der aufgezeigten Probleme und Herausforderungen bestehender AAC-Systeme sowie der festgestellten Forschungslücken in der wissenschaftlichen Literatur werden in der vorliegenden Dissertation für den Themenbereich der AAC-Systeme die folgenden drei Zielstellungen formuliert:

Zielstellung 5 (ZS 5):
Identifikation der Potenziale von KI-basierten Ansätzen im Rahmen von AAC-Systemen sowie der Abdeckung dieser KI-bezogenen Potenziale in verschiedenen, ausgewählten Gesprächsphasen und Gesprächskontexten.

Zielstellung 6 (ZS 6):
Konzeption einer Generativen KI-basierten AAC-Lösung für sprachbehinderte Menschen. Insbesondere sollen dabei die zentralen Gestaltungsanforderungen an eine solche Lösung aus der Literatur identifiziert und zentrale Gestaltungsprinzipien für die Implementierung der Lösung abgeleitet werden.

Zielstellung 7 (ZS 7)
Entwicklung und Evaluation einer Generativen KI-basierten AAC-Lösung für sprachbehinderte Menschen unter Berücksichtigung der identifizierten zentralen Gestaltungsanforderungen und Gestaltungsprinzipien.

Neben den Beiträgen zur Lösung des gesellschaftlichen Problems der Inklusion von sprachbehinderten Menschen in alltägliche Gesprächssituationen liefert die vorliegende Dissertation auch wichtige Erkenntnisse für die Gestaltung von Interaktionen mit Chatbots als digitale, intelligente Kommunikationsagenten.

Unter einem Chatbot wird ein Computerprogramm verstanden, welches über die Fähigkeit verfügt, ein Gespräch mit Menschen unter Anwendung natürlicher Sprache zu führen (vgl. Abdul-Kader und Woods, 2015; Stucki et al., 2018). Der Dialog beginnt, sobald ein Nutzer eine Anfrage in die Eingabezeile im GUI des Chatbots eingegeben und diese abgeschickt hat (Stucki et al., 2018). Nach diesem initialen Interaktionsschritt löst der Chatbot in der Verarbeitung der Nutzereingabe einen Datenbereinigungsschritt aus, der sich u. a. mit der Korrektur von Tippfehlern und der Entfernung von Satzzeichen beschäftigt (Stucki et al., 2018). Die bereinigte Nutzereingabe wird dann für die weitere Verarbeitung mit Techniken aus dem syntaktischen NLP (z. B. Stemming oder Lemmatisierung der Wörter in der Nutzereingabe) und dem semantischen NLP (z. B. Word-Sense Disambiguation, Bag-of-Words Repräsentation der Nutzereingabe) vorbereitet (Stucki et al., 2018). Anschließend kann mittels NLU und dem sog. *„Intent Matching“* die Intention des Nutzers auf Basis seiner Eingabe identifiziert werden (Stucki et al., 2018). Hierfür können verschiedene Techniken wie einfache Pattern-Matching-Ansätze, aber auch KI-basierte Ansätze wie künstliche neuronale Netze zum Einsatz kommen (Stucki et al., 2018). Anschließend wird eine geeignete Reaktion ausgewählt und die entsprechende Antwort gegenüber dem Nutzer im GUI des Chatbots präsentiert.

Der Einsatz von Chatbots in der Kundeninteraktion bringt für Unternehmen vielfältige Vorteile und Potenziale mit sich. In der externen Kommunikation tragen Chatbots dazu bei, dass die Arbeitslast von Mitarbeitern im Kundenservice reduziert werden kann (Chakrabarti und Luger, 2015; Luger und Sellen, 2016; Meuter et al., 2005). Gleichzeitig ermöglicht ein Chatbot einen 24/7-Kundenservice, sodass Kunden ihre Serviceanfragen unabhängig von den Geschäftszeiten und somit auch an Wochenenden stellen können. Darüber hinaus sorgt der Einsatz von Chatbots für eine gleichbleibende Servicequalität, weil Chatbots im Vergleich zu menschlichen Angestellten keinen

schwankenden Leistungen und Gemütszuständen unterliegen (Braun, 2013). Insgesamt betrachtet können sich Chatbots und Menschen im Kundenservice ergänzen, indem der Chatbot standardisierte Kundenanliegen mit geringer Variabilität übernimmt und sich die menschlichen Angestellten stattdessen auf aufwändigere Aufgaben und Fälle konzentrieren können (Dale, 2016).

Gleichzeitig finden Chatbots zunehmende Anwendung in den operativen Prozessen innerhalb von Unternehmen, um Angestellte bei der Ausführung und Erledigung von Aufgaben zu unterstützen (Meyer von Wolff et al., 2019; Meyer von Wolff et al., 2020). Ein wesentlicher Anreiz für den Einsatz von Chatbots ist dabei die Steigerung der Produktivität der Mitarbeiter. Chatbots können hierbei bei Aufgaben in verschiedenen operativen Prozessen innerhalb von Unternehmen, z. B. im Rahmen von Innovationsprozessen (Johannsen et al., 2021), Buchhaltungsprozessen (Henrich, 2017) oder IT-Servicemanagement-Prozessen (Kottorp und Jäderberg, 2017), unterstützen und den Mitarbeitern z. B. für die jeweiligen Arbeitsschritte aufgabenrelevante Informationen unmittelbar aus den angebundenen Quellsystemen zur Verfügung stellen (vgl. Brandtzaeg und Folstad, 2017). Infolgedessen müssen die Mitarbeiter nicht selbst in den verschiedenen Unternehmenssystemen nach den benötigten Informationen suchen, sodass die Aufgabenerfüllung durch die unmittelbar verfügbaren Informationen schneller und bequemer abgeschlossen werden kann. Gleichzeitig können Chatbots über die Anbindung zu Unternehmenssystemen auch Operationen auf Datensätzen ausführen und z. B. eine Eintragung von Terminen in die Kalender von an bestimmten Aufgaben und Arbeitsschritten beteiligten Mitarbeitern veranlassen.

Chatbot-Projekte verlaufen jedoch nicht immer erfolgreich. Von Unternehmen eingeführte Chatbot-Lösungen führen bei Kunden oftmals zu Frust und Unzufriedenheit. Statt den Chatbot zu nutzen, werden weiterhin die herkömmlichen Kontaktkanäle wie der persönliche Kontakt im Ladengeschäft oder der telefonische Kontakt zum Kundenservice genutzt. Auch Mitarbeiter sind mit der Unterstützung durch Chatbot-Lösungen in operativen Prozessen oftmals unzufrieden und greifen daher in ihrer Aufgabenerfüllung nicht auf deren Unterstützung zurück (vgl. Akhtar et al., 2019; Ashktorab et al., 2019; Van der Goot et al., 2021). Aus diesen Gründen wurden zahlreiche Chatbot-Projekte von Unternehmen wieder eingestellt (vgl. Ashktorab et al., 2019; Janssen

et al., 2021). Ein häufiger Grund für das Scheitern ist, dass die eingeführten Chatbots nicht alle notwendigen Anforderungen und Funktionalitäten abdecken, welche für eine solche Lösung zwingend erforderlich sind.

Entscheidungsträgern in Unternehmen fehlt es in dieser Hinsicht an einer Entscheidungshilfe, welche Anforderungen und Funktionalitäten Chatbots, sowohl im Kontext der Kundenkommunikation als auch zur Förderung der Produktivität ihrer Mitarbeiter, erfüllen sollten (vgl. Johannsen et al., 2018; Ramming et al., 2022). Gleichzeitig fehlt es dahingehend an einer Marktübersicht, welche kommerziellen Chatbot-Lösungen welche dieser Anforderungen und Funktionalitäten wie gut erfüllen. Darauf basierend werden für die Dissertation die folgenden beiden Zielstellungen für den Themenbereich der Chatbots formuliert:

Zielstellung 8 (ZS 8):
Identifikation von Anforderungen, welche Chatbots zur Unterstützung der Kundeninteraktion und zur Förderung der Produktivität bei der Aufgabenerfüllung im Rahmen von operativen Prozessen erfüllen sollten.

Zielstellung 9 (ZS 9):
Durchführung einer Marktstudie, um am Markt verfügbare, kommerzielle Chatbot-Lösungen im Hinblick auf zentrale anwendungsfallbezogene Anforderungen zu vergleichen.

1.2. Forschungsfragen

Basierend auf den in Kapitel 1.1 beschriebenen Problemstellungen und den daraus resultierenden Zielstellungen werden im Folgenden die in der Dissertation behandelten Forschungsfragen aufgestellt und diese jeweils erläutert. Die Zusammenhänge der Zielstellungen zu den Forschungsfragen werden in Kapitel 1.3 in Abbildung 1 sowie in Tabelle 1 nochmals konsolidiert aufgezeigt.

Zielstellung 1: Damit der aktuelle Stand der Forschung zu den Einflüssen von Kontextfaktoren auf die Stimmungen in Tweets als einer Form von Social Media Posts

abgebildet werden kann, bedarf es zunächst einer umfassenden Aufarbeitung der einschlägigen Forschungsliteratur. Es ergibt sich dadurch eine konsolidierte Übersicht über Kontextfaktoren und ihre Wirkungsweisen. Diese Übersicht soll Unternehmen dabei unterstützen, die Veröffentlichung von Social Media Posts an die Kontextfaktoren angelehnt so vorzunehmen zu können, damit negative Reaktionen vermieden werden können. Darüber hinaus zeigt diese Übersicht, welche Kontextfaktoren in der Forschung bislang unzureichend untersucht worden sind. Dabei haben die beiden zentralen Kontextfaktoren „Kontinente“ und „mobile Betriebssysteme“ bislang sehr wenig Beachtung in der Forschungsliteratur erhalten (Konadl et al., 2021).

Während bereits die Einflüsse einzelner Kontextfaktoren für Unternehmen eine nützliche Entscheidungshilfe für die Planung zur Veröffentlichung ihrer Social Media Posts darstellt, verspricht die Untersuchung der Effekte verschiedener Kombinationen von Kontextfaktoren noch präzisere Handlungsempfehlungen. Besonders aussagekräftig wird die Identifikation der Effekte der Kontextfaktoren, wenn diese auf Basis der Analyse einer großen Menge an Social Media Posts erfolgt. Dadurch können Schwankungen in den Effekten durch die immense Menge der Daten geglättet und infolgedessen verlässliche Effekte identifiziert werden. Basierend auf Zielstellung 1 wird für die vorliegende Dissertation die folgende, erste Forschungsfrage abgeleitet:

Forschungsfrage 1 (FF 1):
Welche Einflüsse können für die Kontextfaktoren „Kontinente“ und „mobile Betriebssysteme“, sowohl einzeln als auch in Kombination, für das in Tweets zum Ausdruck gebrachte Sentiment identifiziert werden?

Zielstellung 2: Um mithilfe einer bestmöglich geeigneten Analysetechnik die in textuellen Social Media Daten zugrundeliegenden Themen extrahieren zu können, wird zunächst ein Vergleich von verschiedenen Topic Modelling Techniken angestrebt. Ausgehend von der Topic Modelling Technik LDA, die in der IS-Forschung aufgrund ihrer leichten Anwendbarkeit und ihrer guten Analyseergebnisse häufig verwendet wird (Eickhoff und Neuss, 2017), sollen mit DMR (Mimno und McCallum, 2008) und PAM (Li und McCallum, 2006) zwei weitere LDA-basierte Techniken in den Ver-

gleich einbezogen werden. Dieser Vergleich soll vor dem Hintergrund der drei zentralen marketingbezogenen Anwendungsfälle (a) Themenextraktion, (b) Trendanalyse und (c) Themenstrukturierung erfolgen, welche im Hinblick auf das Topic Modelling im Marketing in der einschlägigen Forschungsliteratur häufig diskutiert worden sind.

Damit ein einheitlicher und strukturierter Vergleich der Topic Modelling Techniken gewährleistet werden kann, müssen verschiedene Vergleichskriterien identifiziert und diese den marketingbezogenen Anwendungsfällen (a)-(c) sowie den dazugehörigen Anforderungen zugeordnet werden. Durch die Anwendung der Topic Modelling Techniken an einem Realwelt-Datensatz können anschließend die Unterschiede der Techniken anhand dieser Vergleichskriterien herausgestellt werden. Auf Basis dieses Vergleichs können dann Handlungsempfehlungen für die Auswahl der am besten geeignetsten Topic Modelling Technik für die drei Anwendungsfälle (a)-(c) abgeleitet werden. Die beiden Forschungsfragen, die sich daraus ergeben, lauten wie folgt:

Forschungsfrage 2 (FF 2)
Anhand welcher Kriterien lassen sich Topic Modelling Techniken miteinander vergleichen?

Forschungsfrage 3 (FF 3)
Welche Topic Modelling Technik sollte basierend auf den identifizierten Vergleichskriterien für die marketingbezogenen Anwendungsfälle Themenextraktion, Trendanalyse und Themenstrukturierung verwendet werden?

Zielstellungen 3 und 4: Aufgrund seiner Vorteile soll das Topic Modelling als effektive und zielgerichtete Trendanalysetechnik für textuelle Social Media Beiträge für Unternehmen nutzbar gemacht werden. Zu diesem Zweck soll im Rahmen der Dissertation die Entwicklung von entsprechenden Softwaretools erfolgen. Dazu müssen zunächst aus der Forschungsliteratur die zentralen Anforderungen abgeleitet werden, welche vor dem Hintergrund der jeweiligen Zielstellungen für Topic Modelling basierte Trendanalysetools essenziell wichtig sind.

Die frühzeitige Identifizierung von neuen und vielversprechenden Trends und Ideen in den Bereichen (a) Produktentwicklung, (b) Kundenverhaltensanalyse und (c) Marken-

/Marktbeobachtung eröffnet für Unternehmen die Möglichkeit, entscheidende Wettbewerbsvorteile zu erlangen. Die am Markt verfügbaren Softwaretools (sowohl Trendanalysetools als auch dedizierte Topic Modelling-basierte Tools) weisen dahingehend jedoch entscheidende Defizite auf. Insbesondere können diese Softwaretools externe Parameter wie z. B. Geolokationen, Zeit oder nutzerbezogene Informationen bei der Identifikation von Trends nicht umfassend und flexibel berücksichtigen. Dies führt dazu, dass Trends nur oberflächlich identifiziert werden und tiefergehende Analysen für den jeweiligen Trend nicht möglich sind. Gerade aber diese Kontextualisierung von Trends mit externen Parametern führt dazu, dass Unternehmen diese tiefergehenden Einblicke über Produkte, Kunden und Wettbewerber in der notwendigen Granularität erhalten können, um ihre Produkte und Services zielführend anzupassen.

Bei der technischen Realisierung des Artefakts zur automatisierten Trendanalyse soll der Fokus auf die Kombination von verschiedenen Social Media Analysetechniken in Bezug auf die Anforderungen der marketingbezogenen Anwendungsfälle (a)-(c) gelegt werden. Dabei soll die in dem zu realisierenden Artefakt eingesetzte Topic Modelling Technik insbesondere in der Lage sein, verschiedene externe Parameter in unterschiedlichen Granularitäten flexibel einzubeziehen. Die Umsetzung dieses Trendanalysetools wird dabei durch die nachfolgende vierte Forschungsfrage geleitet:

Forschungsfrage 4 (FF 4):
Wie sollte ein auf Topic Modelling basierendes Trendanalysetool aussehen, welches die Anforderungen der marketingbezogenen Anwendungsfälle (a) Produktentwicklung (b) Kundenverhaltensanalyse und (c) Markt- /Markenbeobachtung umsetzt?

Für die Bewertung von Trends ist es eine zentrale Anforderung an ein Trendanalysetool, das Stimmungsbild und die Tonalitäten gegenüber den Trendthemen und ihren Bestandteilen zu identifizieren. Hierfür kann die Sentimentanalyse angewendet werden, um als Unternehmen nachvollziehen zu können, ob ein Trendthema, wie z. B. ein bestimmtes Produkt oder ein bestimmter Service, positiv oder negativ in den Social Media Posts diskutiert wird. Dabei ist es aber entscheidend, in welcher Granularität die Sentimente identifiziert werden. Eine zielgerichtete Anpassung von Produkten und Services an die Vorstellungen von Kunden ist nur dann zielführend möglich, wenn bei

der Trendanalyse ermittelt wird, welche konkreten Aspekte und Eigenschaften eines Produkts oder eines Service positiv oder negativ beurteilt worden sind.

Die Anwendung der aspektbasierten Sentimentanalyse, welche Sentimente feingranular gegenüber einzelnen Aspekten bestimmt, kann diese geforderte Granularität des Sentiments im Rahmen eines Trendanalysetools bewerkstelligen. Jedoch werden, so wie es eine Analyse der am Markt verfügbaren Trendanalysetools ergeben hat (vgl. Wörner et al., 2022), wichtige Anforderungen an eine entsprechende Lösung von keinem der verfügbaren Softwaretools ausreichend gut erfüllt. Unter anderem ermöglicht es keines dieser Softwaretools, domänenspezifisches Wissen in die Identifikation von Trends einzubeziehen. Dieses Wissen kann aber entscheidend sein, um für einen spezifischen Kontext zielführende Trends zu identifizieren. Daher soll ein Trendanalysetool entwickelt werden, welches die Defizite bestehender Softwaretools behebt und unter Anwendung der aspektbasierten Sentimentanalyse die notwendige Granularität der Sentimente für (inkrementelle) Produktinnovationen und Produktverbesserungen bietet. Um zu diesem Softwaretool zu gelangen, lautet die fünfte Forschungsfrage:

Forschungsfrage 5 (FF 5):
Wie sollte ein auf aspektbasierter Sentimentanalyse basierendes Trendanalysetool aussehen, welches die Trendanalyse für den Anwendungsfall der Produktentwicklung unterstützt und welche Anforderungen sollte ein solches Trendanalysetool dabei erfüllen?

Mit den ersten fünf Forschungsfragen wird die Gestaltung von Interaktionen mit sozialen Medien als moderne, digitale Informations- und Kommunikationssysteme untersucht. Es sollen dadurch maßgebliche Beiträge geleistet werden, um die in Social Media Beiträgen enthaltenen Informationen und Zusammenhänge zielgerichteter extrahieren und für Unternehmen nutzbar machen zu können. Im Folgenden werden nun die weiteren Forschungsfragen aufgezeigt, die sich mit der Gestaltung von Interaktionen mit intelligenten, digitalen Kommunikationsagenten befassen.

Zielstellung 5: Um den aktuellen Stand der Forschung zum Einsatz von KI-Ansätzen in AAC-Systemen umfassend darzustellen, ist eine systematische Aufarbeitung der einschlägigen, wissenschaftlichen Literatur erforderlich. AAC-Systeme ermöglichen

sprachbehinderten Menschen lediglich geringe Gesprächsraten und die vorgeschlagenen Gesprächsbeiträge passen oft nicht zu der konkreten Situation, an der die sprachbehinderte Person teilnimmt. In diesem Zusammenhang ist es ein zentrales Ziel der Forschung zu AAC-Systemen, die Kommunikationslücke zwischen sprachbehinderten Menschen und ihren *„sprachfähigen“* Gesprächspartnern zu verringern.

Um dieses Ziel zu erreichen, wurden verschiedene KI-Ansätze für AAC-Systeme konzipiert. Jedoch sind diese KI-Ansätze wenig zielführend, wenn diese nicht die Anforderungen der konkreten Gesprächssituationen erfüllen, in denen sie eingesetzt werden. Grundlegend können Gespräche in verschiedene zentrale Gesprächsphasen und Gesprächskontexte strukturiert werden, welche jeweils unterschiedliche Anforderungen an einen KI-Ansatz stellen. Es fehlt allerdings an einem strukturierten Überblick darüber, welche Potenziale durch KI-Ansätze für zentrale Gesprächsphasen und Gesprächskontexte in der wissenschaftlichen Literatur realisiert wurden. Vor diesem Hintergrund und der Zielstellung, die Entwicklung von passgenauen KI-Ansätzen für diese Gesprächsphasen und Gesprächskontexte anzustoßen, sollen die folgenden beiden Forschungsfragen beantwortet werden:

Forschungsfrage 6 (FF 6):
Inwieweit werden durch die in der Literatur für AAC-Systeme vorgeschlagenen KI-Ansätze für sprachbehinderte Menschen Potenziale in den Gesprächsphasen „Gesprächseröffnung“, „Gesprächsmitte“ und „Gesprächsabschluss“ sowie für die Gesprächskontexte „Formell“ und „Informell“ realisiert?

Forschungsfrage 7 (FF 7):
Was sind darauf basierend mögliche Implikationen für die zukünftige Forschung zu KI-Ansätzen im Hinblick auf diese Gesprächsphasen und Gesprächskontexte?

Zielstellungen 6 und 7: Die am Markt verfügbaren AAC-Systeme unterstützen lediglich eine begrenzte Anzahl an alltäglichen Gesprächssituationen, weil die zugrundeliegenden KI-Sprachmodelle auf einer limitierten Auswahl von Alltagsgesprächen trainiert wurden (Dempster et al., 2010; Heo und Kang 2019; Kristensson et al., 2020). Diese Sprachmodelle haben deshalb Probleme dabei, vollständige Gesprächsbeiträge

für verschiedene Domänen und Kommunikationsaktivitäten zu generieren, an denen Menschen ohne Sprachbehinderung ohne Probleme teilnehmen können. Zudem bringt die schnellste Weise sich auszudrücken wenig, wenn die von einem AAC-System vorgeschlagenen Gesprächsbeiträge von sprachbehinderten Menschen und ihren Gesprächspartnern als unpassend empfunden werden. Insbesondere könnte der Stil einer Äußerung nicht der erwarteten Ausdrucksweise in einem formellen oder informellen Gesprächskontext entsprechen (Bedrosian et al., 2003; Hoag et al., 2004; Todman und Alm 1997). Dies kann dazu führen, dass Gesprächspartner eine negative Wahrnehmung gegenüber sprachbehinderten Menschen entwickeln und ein Gespräch mit ihnen vermeiden. Auch die sprachbehinderten Menschen als Nutzer dieser AAC-Systeme selbst könnten Gespräche scheuen, da die Äußerungen, die ihre AAC-Systeme ihnen vorschlagen, nicht ihre Persönlichkeit und ihren individuellen Artikulationsstil widerspiegeln (Shen et al., 2022; Tintarev et al., 2016; Valencia et al., 2023).

Als Generative KI kombiniert ChatGPT umfangreiche Sprachmodelle mit textuellen Trainingsdaten, die verschiedene Alltagsgespräche, Kommunikationsaktivitäten und Gesprächsstile abdecken (einschließlich formeller und informeller Gesprächskontexte). Dennoch ist die direkte Integration von ChatGPT-4.0 in ein AAC-System ohne umfassende Anpassungen nicht zielführend und ChatGPT allein ist nicht ausreichend, um alle an einen Generativen KI-basierten Kompositionsansatz gestellten Designanforderungen zu erfüllen. Vor diesem Hintergrund soll in der Dissertation die folgende, achte Forschungsfrage untersucht werden, um diesen auf Generativer KI basierenden Kompositionsansatz zu entwickeln:

Forschungsfrage 8 (FF 8):
Wie könnte ein umfassender Kompositionsansatz für ein AAC-System technisch realisiert werden, der auf Generativer KI beruht und es einem sprachbehinderten Menschen ermöglicht, sowohl formelle als auch informelle Gesprächsbeiträge zu generieren?

Zielstellungen 8 und 9: Damit sowohl Unternehmen als auch die einschlägige Forschung einen umfassenden Überblick über wichtige Anforderungen und Funktionalitäten von Chatbots zur Unterstützung der B2C-Kundenkommunikation erhalten, sind diese zunächst in der einschlägigen Literatur zu identifizieren. Zudem sollen am Markt

verfügbare Chatbot-Lösungen im Rahmen einer Marktstudie bewertet werden. Dabei soll herausgefunden werden, ob und inwieweit am Markt verfügbare Chatbot-Lösungen diese identifizierten Anforderungen und Funktionalitäten erfüllen. Hierfür soll ein Evaluationskatalog entwickelt werden, welcher an Anbieter von Chatbot-Lösungen zur Bewertung ihrer Lösungen verteilt wird.

Die Frage, wie Chatbot-Lösungen für einen effektiven und zielgerichteten Einsatz gestaltet sein sollten, stellt sich für Unternehmen aber nicht nur im Kontext der Kundenkommunikation. Im Unternehmensumfeld ist ein häufiger Beweggrund für den Einsatz eines Chatbots die Steigerung der Produktivität. Die Gestaltung eines Chatbots beeinflusst dabei maßgeblich, inwieweit die Produktivität von Mitarbeitern unterstützt und gefördert werden kann. Deshalb soll in der vorliegenden Dissertation eine systematische Konsolidierung von Funktionalitäten und Designanforderungen aus der wissenschaftlichen Literatur vorgenommen werden, die zur Förderung der Produktivität bei der Nutzung eines Chatbots in operativen Prozessen innerhalb eines Unternehmens beitragen. Für den Themenbereich der Chatbots soll basierend auf den obigen Ausführungen daher den folgenden drei Forschungsfragen nachgegangen werden:

Forschungsfrage 9 (FF 9):
Welche Anforderungen sollten Chatbot-Lösungen für den Einsatz im B2C-Kontext zur Unterstützung der Kundenkommunikation erfüllen?

Forschungsfrage 10 (FF 10):
Welche Gestaltungsmerkmale sollte ein Chatbot erfüllen, um Nutzer in einem produktivitätsorientierten Kontext bei ihrer Aufgabenerfüllung zu unterstützen?

Forschungsfrage 11 (FF 11):
Wie umfassend werden diese Anforderungen und Funktionalitäten von den Anbietern von Chatbot-Lösungen abgedeckt?

1.3. Aufbau der Dissertation

Um das Thema der Dissertation „Methoden und Ansätze für die zielgerichtete Gestaltung der Interaktionen mit modernen Informations- und Kommunikationssystemen“ zu adressieren und somit die vorgestellten Zielstellungen und Forschungsfragen zu behandeln, ist die Dissertation wie folgt aufgebaut: In Kapitel 1 *„Einleitung“* werden das Thema der Dissertation motiviert und die zugrundeliegenden Problemstellungen aufgezeigt, die sowohl auf der Forschung als auch der Praxis basieren. Darauf aufbauend werden die Zielstellungen formuliert und konkrete Forschungsfragen abgeleitet, die im Rahmen der vorliegenden Arbeit beantwortet werden. Im zweiten Kapitel *„Wissenschaftliche Beiträge“* werden die Ergebnisse der kumulativen Dissertation vorgestellt. Dieses Kapitel umfasst insgesamt acht Veröffentlichungen, wobei jeweils (Kapitel 2.1 bis 2.8) eine oder mehrere Zielstellungen sowie damit zusammenhängende Forschungsfragen behandelt werden. Auf Basis der vorgestellten Ergebnisse werden abschließend in Kapitel 3 *„Schlussbetrachtung und Fazit“* die Ergebnisse dieser acht Veröffentlichungen zusammengefasst, diese weiter diskutiert und ihre Beiträge zu Wissenschaft und Praxis herausgearbeitet. Zusätzlich erfolgt ein Ausblick auf weitere mögliche Forschungsarbeiten und Forschungsfelder. Abbildung 1 gibt nochmals einen komprimierten Überblick über den Aufbau der Dissertation, während Tabelle 1 das Zusammenwirken der verschiedenen Zielstellungen und Forschungsfragen für die jeweiligen Beiträge nochmals aufzeigt.

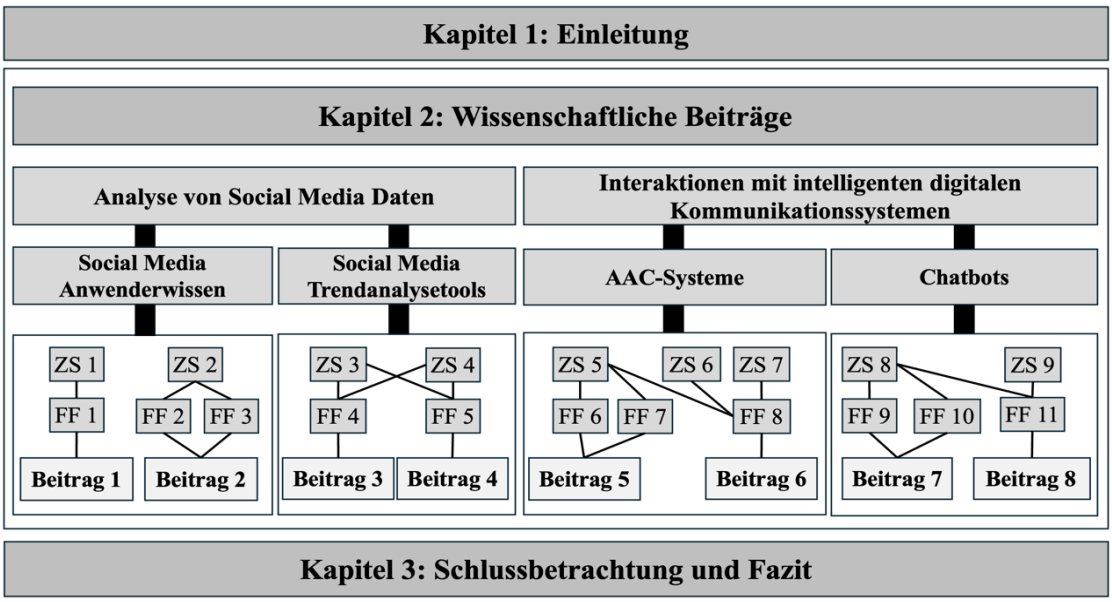

Abbildung 1: Aufbau und Überblick über die Dissertation

Zielsetzungen	Forschungsfragen	Veröffentlichungen
ZS1. Identifikation der Einflüsse, sowohl von einzelnen als auch von kombinierten Kontextfaktoren, im Hinblick auf die in Social Media Posts zum Ausdruck gebrachten Stimmungen.	**FF1.** Welche Einflüsse können für die Kontextfaktoren „Kontinente“ und „mobile Betriebssysteme“, sowohl einzeln als auch in Kombination, für das in Tweets zum Ausdruck gebrachte Sentiment identifiziert werden?	**1**
ZS2. Vergleich von verschiedenen Topic Modelling Techniken mit dem Ziel, für ausgewählte, marketingbezogene Anwendungsfälle Handlungsempfehlungen zur Auswahl von zielführenden Topic Modelling Techniken zu geben.	**FF2.** Anhand welcher Kriterien lassen sich Topic Modelling Techniken miteinander vergleichen?	**2**
	FF3. Welche Topic Modelling Technik sollte basierend auf den identifizierten Vergleichskriterien für die marketingbezogenen Anwendungsfälle Themenextraktion, Trendanalyse und Themenstrukturierung verwendet werden?	
ZS3. Ableitung von Anforderungen mit dem Ziel, Trendanalysetools für verschiedene, marketingbezogene Anwendungsfälle zu konzipieren. Insbesondere sollen dabei als zentrale Designprämissen das Topic Modelling und die aspektbasierte Sentimentanalyse als Analysetechniken untersucht und einbezogen werden.	**FF4.** Wie sollte ein auf Topic Modelling basierendes Trendanalysetool aussehen, welches die Anforderungen der marketingbezogenen Anwendungsfälle (a) Produktentwicklung (b) Kundenverhaltensanalyse und (c) Markt-/Markenbeobachtung umsetzt?	**3 + 4**
ZS4. Entwicklung und Evaluation von Trendanalysetools zur automatisierten Analyse von textuellen Social Media Inhalten unter Berücksichtigung verschiedener anwendungsfallbezogener Anforderungen.	**FF5.** Wie sollte ein auf aspektbasierter Sentimentanalyse basierendes Trendanalysetool aussehen, welches die Trendanalyse für den Anwendungsfall der Produktentwicklung unterstützt und welche Anforderungen sollte ein solches Trendanalysetool dabei erfüllen?	
ZS5. Identifikation der Potenziale von KI-basierten Ansätzen im Rahmen von AAC-Systemen sowie der Abdeckung dieser KI-bezogenen Potenziale in verschiedenen, ausgewählten Gesprächsphasen und Gesprächskontexten.	**FF6.** Inwieweit werden durch die in der Literatur für AAC-Systeme vorgeschlagenen KI-Ansätze für sprachbehinderte Menschen Potenziale in den Gesprächsphasen „Gesprächseröffnung“, „Gesprächsmitte“ und „Gesprächsabschluss“ sowie für die Gesprächskontexte „Formell“ und „Informell“ realisiert?	**5**
	FF7. Was sind darauf basierend mögliche Implikationen für die zukünftige Forschung zu KI-Ansätzen im Hinblick auf diese Gesprächsphasen und Gesprächskontexte?	

	FF8. Wie könnte ein umfassender Kompositionsansatz für ein AAC-System technisch realisiert werden, der auf Generativer KI beruht und es einem sprachbehinderten Menschen ermöglicht, sowohl formelle als auch informelle Gesprächsbeiträge zu generieren?	**6**
ZS6. Konzeption einer Generativen KI-basierten AAC-Lösung für sprachbehinderte Menschen. Insbesondere sollen dabei die zentralen Gestaltungsanforderungen an eine solche Lösung aus der Literatur identifiziert und zentrale Gestaltungsprinzipien für die Implementierung der Lösung abgeleitet werden.		
ZS7. Entwicklung und Evaluation einer Generativen KI-basierten AAC-Lösung für sprachbehinderte Menschen unter Berücksichtigung der identifizierten zentralen Gestaltungsanforderungen und Gestaltungsprinzipien.		
ZS8. Identifikation von Anforderungen, welche Chatbots zur Unterstützung der Kundeninteraktion und zur Förderung der Produktivität bei der Aufgabenerfüllung im Rahmen von operativen Prozessen erfüllen sollten.	**FF9.** Welche Anforderungen sollten Chatbot-Lösungen für den Einsatz im B2C-Kontext zur Unterstützung der Kundenkommunikation erfüllen?	**7**
	FF10. Welche Gestaltungsmerkmale sollte ein Chatbot erfüllen, um Nutzer in einem produktivitätsorientierten Kontext bei ihrer Aufgabenerfüllung zu unterstützen?	
ZS9. Durchführung einer Marktstudie, um am Markt verfügbare, kommerzielle Chatbot-Lösungen im Hinblick auf zentrale anwendungsfallbezogene Anforderungen zu vergleichen.	**FF11.** Wie umfassend werden diese Anforderungen und Funktionalitäten von den Anbietern von Chatbot-Lösungen abgedeckt?	**8**

Tabelle 1: Zielstellungen und Forschungsfragen der Dissertation

2. Wissenschaftliche Beiträge

In diesem Kapitel werden die einzelnen Forschungsbeiträge vorgestellt. Dazu wird zunächst ein Überblick über alle Beiträge gegeben (siehe Tabelle 2). Anschließend werden die Beiträge in den Kapiteln 2.1 bis 2.8 jeweils präsentiert.

Kapitel	Titel	Autoren	Status	Ranking[1]	Autorenanteil
2.1	Identifying Sentiment Influences Provoked by Context Factors – Results from a Data Analytics Procedure Performed on Tweets	**Daniel Konadl** Janik Wörner Susanne Leist	**Veröffentlicht** in den Proceedings der 54. Hawaii International Conference on System Sciences.	**B**	**60 %**
2.2	Comparison of Topic Modelling Techniques in Marketing – Results From an Analysis of Distinctive Use Cases	Janik Wörner **Daniel Konadl** Isabel Schmid Susanne Leist	**Veröffentlicht** in den Proceedings der 29. European Conference on Information Systems.	**A**	**25 %**
2.3	MANTRA: A Topic Modeling-Based Tool to Support Automated Trend Analysis on Unstructured Social Media Data	Janik Wörner **Daniel Konadl** Isabel Schmid Susanne Leist	**Veröffentlicht** in den Proceedings der 44. International Conference on Information Systems, Hyderabad, Indien.	**A**	**35 %**
2.4	Supporting Product Development by a Trend Analysis Tool Applying Aspect-Based Sentiment Detection	Janik Wörner **Daniel Konadl** Isabel Schmid Susanne Leist	**Veröffentlicht** in den Proceedings der 17. International Conference on Design Science Research in Information Systems and Technology, St. Petersburg, Florida, USA.	**B**	**30 %**
2.5	Artificial Intelligence in Augmentative and Alternative Communication Systems – A Literature-Based Assessment and Implications of Different Conversation Phases and Contexts	**Daniel Konadl** Janik Wörner Lucas Luttner Susanne Leist	**Veröffentlicht** in den Proceedings der 31. European Conference on Information Systems, Kristiansand, Norwegen.	**A**	**50 %**
2.6	A Generative AI-Based Approach to Support Automated Utterance Generation for Different Conversational Contexts within AAC Systems	**Daniel Konadl**	**Veröffentlicht** in den Proceedings der 45. International Conference on Information Systems, Bangkok, Thailand.	**A**	**100 %**
2.7	Comparison of Commercial Chatbot Solutions for Supporting Customer Interaction	Florian Johannsen Susanne Leist **Daniel Konadl** Michael Basche Bastian de Hesselle	**Veröffentlicht** in den Proceedings der 26. European Conference on Information Systems, Portsmouth, Großbritannien.	**A**	**20 %**

[1] Die Angabe des Rankings erfolgte jeweils nach dem VHB JQ 4 Ranking von 2024.

2.8	Chatbot Design Features to Increase Productivity	Lukas Ramming **Daniel Konadl** Susanne Leist	**Veröffentlicht** in den Proceedings der 17. International Conference on Design Science Research in Information Systems and Technology, St. Petersburg, Florida, USA.	**B**	**40%**
					360 %

Tabelle 2: Status, Ranking und Autorenanteile zu den wissenschaftlichen Beiträgen

2.1. **Beitrag 1:** Identifying Sentiment Influences Provoked by Context Factors – Results from a Data Analytics Procedure Performed on Tweets

Adressierte Forschungsfrage	Welche Einflüsse können für die Kontextfaktoren „Kontinente“ und „mobile Betriebssysteme“, sowohl einzeln als auch in Kombination, für das in Tweets zum Ausdruck gebrachte Sentiment identifiziert werden?
Zielstellungen	• Identifikation der in der Literatur untersuchten Kontextfaktoren im Hinblick auf das in Tweets zum Ausdruck gebrachte Sentiment. • Identifikation der Einflüsse der Kontextfaktoren „Kontinente“ (Afrika, Asien, Nordamerika, Europa, Ozeanien und Südamerika) und „mobile Betriebssysteme“ (Apple OS und Android OS). • Identifikation des Einflusses der Kombination aus den beiden Kontextfaktoren „Kontinente“ und „mobile Betriebssysteme“.
Forschungsmethode	**Social Media Analytics und Text Mining** • Extraktion eines Realweltdatensatzes mit 358.923.210 Tweets. • Bereinigung des Rohdatensatzes und Ermittlung der Einträge für die Kontinente und die mobilen Betriebssysteme aus den Rohdaten. • Bestimmung des Sentiments (245.077.312 Tweets im bereinigten Datensatz) der Tweets mit der VADER Sentimentanalyse von Hutto und Gilbert (2014).
Kernergebnisse (Überblick)	• Die Sentimente unterscheiden sich bei der univariaten Analyse der Tweets signifikant für die einzelnen Kontinente. Das Sentiment, welches den Tweets aus Asien zugeordnet werden kann, ist am positivsten. Das negativste Sentiment kann in den Tweets festgestellt werden, denen der Kontinent Nordamerika zugeordnet wurde. • Die Sentimente unterscheiden sich bei der univariaten Analyse der Tweets signifikant für die beiden untersuchten mobilen Betriebssysteme. Das Sentiment derjenigen Tweets, welche von Android-Geräten abgesetzt worden sind, ist positiver als das Sentiment der von Apple-Geräten abgesetzten Tweets. • Bei der bivariaten bzw. kombinierten Analyse von Kontinenten und mobilen Betriebssystemen zeigt sich, dass bei Android die Reihenfolge und die Ergebnisse der univariaten Analyse für den Kontextfaktor bei den Kontinenten weitestgehend bestätigt werden kann. Im Vergleich dazu kann bei Apple-Geräten das positivste Sentiment nun für diejenigen Tweets beobachtet werden, welche aus Europa stammen, während Nordamerika weiterhin das negativste Sentiment aufweist.
Publikationsort	**Veröffentlicht** in den Proceedings der 54. Hawaii International Conference on System Sciences, Online, 2022.
Ranking VHB JQ 4	**B**
Autoren und Anteile	**Daniel Konadl** **60 %** Janik Wörner 30 % Susanne Leist 10 %

Tabelle 3: Fact Sheet zu Beitrag 1

Identifying Sentiment Influences Provoked by Context Factors – Results from a Data Analytics Procedure Performed on Tweets

Daniel Konadl
University of Regensburg
daniel.konadl@ur.de

Janik Wörner
University of Regensburg
janik.woerner@ur.de

Susanne Leist
University of Regensburg
susanne.leist@ur.de

Abstract

Context factors have lasting impacts on people's sentiments. Exploring impacts that different contexts have on sentiments can be crucial for managing the increasing number of communications companies nowadays maintain with customers via social media channels. To help companies prevent impacts of negative word of mouth, we provide an overview about sentiment-influential contexts for tweets as one kind of social media texts previously discussed within the literature. We collected an overall amount of 358.923.210 tweets and performed analysis to uncover the effects of continents, mobile devices' operating systems (OS) and the combination of both on sentiments expressed within tweets. Our results show remarkable differences for tweets originating from North America and Apple devices, which turned out to be the tweets with the lowest sentiments compared to the other continents and the mobile OS Android.

1. Motivation

Over the last decade, social media have reached an immense widespread [1, 2]. In private settings, social media are used for connecting with friends, for communicating with each other, presenting oneself to other users, sharing personal experiences or achieving social standing and reputation within communities [3]. Textual social media (e.g. tweets on Twitter or posts in Facebook) have proven to be a channel for complaint articulation for customers, e.g. about inadequate behaviors of companies resulting from negative experiences with products and services [4, 5]. Negative perceptions and experiences can prompt consumers to articulate negative sentiments online in written social media [4-7]. As social media texts are widely visible within social media channels, negative sentiments articulated in social media texts can potentially infect huge masses of social media users and provoke negative sentiments [5-7]. Additionally, social media texts containing negative sentiments tend to spread more quickly [8] so that companies have to react fast to prevent damages. Companies reacting adequately and quickly to customers' concerns can turn negative experiences into positive perceptions [5, 9, 10]. Otherwise, negative perceptions can provoke other users to contribute their own negative experiences [5]. This so-called negative word of mouth (nWoM) can lead to the far-reaching dissemination of negative perceptions towards a company in the online as well as in the offline sphere [5, 7]. Beyond keeping potential customers away from buying companies' products and consuming their services, nWoM can furthermore lead to the termination of existing customer relationships [11-13]. Eventually, churning customers and the absence of new customers lead to decreasing sales [11-13], high costs for acquiring new customers [14] and thereafter even to companies' existence being threatened.

For monitoring the opinions expressed in social media texts, sentiment analysis offers a solution to automatically identify opinion polarities from huge volumes of textual data (cf. [15]). The results of sentiment analysis can inform companies about negative sentiments expressed in social media channels, to make companies try to meet and overcome customers' concerns. While sentiment analysis enables companies to identify negative customers' perceptions identified from the texts' contents, companies require knowledge about the circumstances provoking certain sentiments. In addition to the contents of the texts captured by sentiment analysis tools, there are other factors that influence the sentiments of people.

For example, the days during a week have been shown to predominantly provoking negative sentiments compared to weekends (cf. [16-18]). Furthermore, places related to spare-time activities (e.g. parks and green spaces [19-21]) are associated with more positive sentiments than workplaces where sentiments are on average less positive (e.g. [22]). Companies that are aware of contexts as sentiment-influential factors can tailor their customer communication according to the effects of contexts. For strengthening the effects of positive messages propa-

URI: https://hdl.handle.net/10125/70922
978-0-9981331-4-0
(CC BY-NC-ND 4.0)

gated, companies can also rely on contexts provoking positive sentiments. Companies aiming at positive Word of Mouth (pWoM) spread (cf. [4, 11, 23]) for evoking positive perceptions towards their products, services or the company itself, should rather communicate related messages e.g. on weekends.

Depending on the locations (e.g. country or continent), customers may be more or less likely expressing negative sentiments within social media posts. To counter the sentiment tendencies prevalent in certain locations (e.g. countries or continents), companies could thereafter adjust their customer communication. We address this research gap and perform analysis to identify the influences mobile devices' operating systems (OS) have on sentiments expressed within tweets as one kind of social media posts. As a second context, we include continents. To address the lack of investigations considering at least two contexts, we perform analysis combining mobile devices' OS and continents as contextual factors.

The rest of the paper is organized as follows: In the next chapter, conceptional basics of social media, word-of-mouth, sentiment analysis and contexts are introduced. Afterwards, within the chapter related work, we give an overview of sentiment-influential context factors and corresponding effects. Subsequently, we describe the steps performed within our investigation and report the results of our analysis. After interpreting and discussing the said results, we draw on implications of our findings. The paper is rounded off with a conclusion including limitations and an outlook on future research.

2. Conceptional basics

In literature, the term "social media" is often described as "*a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content (UGC)*" ([24], p. 61). UGCs represent "*the sum of all ways in which people make use of social media*" ([24], p. 61). Social media tools provide users with functionalities to connect with friends, presenting themselves to other users and communicating with each other [3]. Social media posts are suitable to conduct word-of-mouth propaganda and to lead communication that includes personal experiences and opinions about a product, service or promotion [4, 5] with consumers, friends, colleagues or other acquaintances [23, 25]. Within these communications, messages as well as therein expressed sentiments are spread [23, 25] and widely noticeable by other users within a social media channel [4-7]. To help companies take notice and control the sentiments expressed within the ever-increasing amount of social media posts, sentiment analysis proposes algorithmic approaches to identify the polarity of texts [15]. In terms of sentiment analysis, there are different approaches, amongst others dictionary- and sentence-based sentiment analysis [15]. When performing dictionary-based sentiment analysis, the sentiment of each entity (e.g. each word) from a text is classified into a positive or negative class using the dictionaries. These dictionaries annotate opinion carrying words, and the sentiment of the whole sentence is determined by considering the sum of the combined scores of all its entities [26].

However, not only the content or the formulation of messages or opinions of customers on products, services or the company itself, but also contexts play an important role in provoking sentiments. Depending on the manifestations of contexts, people experience different sentiments [27]. Context can be defined as "*any information that can be used to characterize the situation of an entity*" ([28], p. 5) whereby "*an entity is a person, place, or object that is considered relevant*" ([28], p. 5). Contexts such as time, location or mobile OS as well as combinations of contexts can be investigated as sentiment-influential factors in the offline world. While time describes the temporal contexts (e.g. time of day, day of week, month of year) in which a tweet is posted, location relates to the spatial properties (e.g. county, country, continent) the user is surrounded by, when tweeting. Mobile devices' OS (e.g. Google Android or Apple iOS) responsible for operating essential system functions on mobile devices, are associated with different personality traits [29] and can thus be seen as another sentiment-influential context. Additionally, contexts are as well experienced, while being in the situation of writing social media posts (e.g. tweets). Therefore, contexts also act as sentiment-influential factors within social media spheres (e.g. Twitter).

3. Related Work

The idea of observing contextual factors and their influence on sentiments apparent within tweets is not a new one. Efforts have been made to uncover the effects of temporal factors such as time of day [16, 18, 30] and day of week [16-18, 31] on sentiments within tweets. Tuesdays, Wednesdays and Thursdays have been identified as days characterized by negative emotions [18], with Wednesdays having even been identified as being the most negative days of the week [16]. Throughout the week and towards weekends, the sentiments become more positive, with Fridays and Saturdays being the most positive days [16, 17]. Sundays are associated with peaks in positive sentiment [18]. However, literature also reports de-

clining sentiments on Sundays [26]. Regarding weekends, the results suggest that both positive and negative sentiments are more present as opposed to the working week [31]. Regarding the times of a day, there are contradictory findings within the literature. While [30] state that the time between 5 a.m. and 6 a.m. is the happiest hour of a day, [18] identified the most negative sentiments for this phase and the most positive sentiments in the evenings. [16] showed that there is a maximum of positive sentiments from 11 a.m. to 1 p.m. as well as from 5 p.m. to 8 p.m.

As for spatial influences, efforts have been made to analyze the effects of different kinds of locations and the impact of their properties on sentiments [16, 17, 19-21, 32-37]. Staying in green spaces such as parks has been identified as inducing positive sentiments within tweets [19-21]. In line with that, [16] identified that how a location is used, influences the sentiments expressed in tweets. Farmland and places associated with public transportation or industry bring up more negative than positive sentiments. Public places are almost equally likely to provoke positive and negative sentiments. Commercial areas tend to generate more positive than negative sentiments [16]. Places where time is shared with friends and family also induce predominantly positive sentiments [17]. The weather prevailing at a location is another sentiment-influential context [33-37]. People are happier when temperatures drop slowly rather than rapidly [35]. Extremely hot and cold temperatures [34], cloudy weather [34], high humidity [34], excessive snowfall [35], hailstorms [35], and extreme weather events such as hurricanes [33, 36] or earthquakes [37] have negative influences on the sentiment expressed within tweets.

There is further context information that rely on characteristics of individuals. People with high incomes articulate themselves more positively within tweets as people living in poorer neighborhoods [32, 38, 39]. Older people express more fear compared to middle-aged people whose tweets contain more joyful terms [38]. Neighborhoods with higher proportions of White, Asian and Hispanic populations also share predominantly more joy-related expressions [38]. However, as [39] showed, Hispanic residents can also be associated with less positive and sadder emotions. People with an African background tend to use more dimness-related terms in their tweets [38]. Higher degrees of education shape and higher-earning populations share happier and more positive emotions [39]. Tweets that are sent from mobile devices tend to be more negative in terms of expressed sentiments than those that are sent from desktops [40, 41]. As for mobile devices' OS, more positive posts are more likely to stem from blackberry devices [42].

Within the literature, there are already approaches relating to the effects of contexts on sentiments expressed within tweets. There are many investigations concerning the identification of temporal, spatial and person-related contexts as well as corresponding effects provoking positive and negative sentiments. Nevertheless, related work mostly focuses on one context solely or considers them isolated from each other. [16, 17, 19-21, 32-37] have dealt with spatial factors and focused on certain countries (e.g. USA (cf. [16, 17, 19, 21, 32-36]), Haiti (cf. [37]) or Australia (cf. [20])) instead of whole continents. [42] analyzed the influence of mobile OS on sentiments. However, the authors do not match these influences with continents (cf. [42]) as we did within our paper. To address this research gap, we performed analysis that combine mobile devices' OS and continents.

4. Methodology

To identify the effects of mobile devices' OS, continents and the combination of both on sentiments expressed within tweets, we aligned our approach to the steps proposed within the text-mining procedure of [43]. As we aim at identifying the effects of contexts on sentiments, we also describe how we proceeded this task. We split our approach (*cf. figure 1*) into **(1)** preparing the analysis, **(2)** conducting the analysis and **(3)** reporting the results.

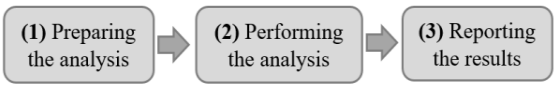

Figure 1: Steps applied within the approach

Within the **(1)** preparatory steps, we first familiarized with the particularities of tweets. Then, we extracted the tweets and performed exploratory analysis followed by conducting data reduction. Because mobile devices' OS and continents as the context factors to be investigated were not directly provided with the extracted tweets, we had to perform further steps to transform the provided data attribute values to obtain the said context factors. Then, we appended the continents and mobile devices' OS to the corresponding tweets and **(2)** determined sentiment values for the extracted tweets by applying an existing sentiment analysis approach. As with the continents and mobile devices' OS, sentiment values were appended to the corresponding tweets. Within the next step, summarization techniques are applied to the sentiment values for tweets of the investigated context factors. Hereby, the mean sentiment values for each context factor and

the share of tweets at a certain sentiment level are determined. The last step **(3)** is about reporting the results. This includes the results to be presented and interpreted. The obtained results may contradict previous findings so that resolving these conflicts can be necessary. In the next chapters, we describe in more detail how we approached within these three steps.

5. Preparing the analysis

5.1. Particularities of tweets

As tweets are the research objects within our investigation, we firstly familiarized ourselves with the particularities of Twitter and tweets as the corresponding social media texts. Hereby, Twitter is a microblogging application "*stand[ing] halfway between traditional blogs and social networking sites, and are characterized by a high degree of self-presentation/self-disclosure and a medium to low degree of social presence/media richness*" ([44], p. 106). As a specific type of social media, it allows users to "*exchange small elements of content such as short sentences, individual images or video links*" ([44], p. 106). Tweets are amongst others characterized by their shortness (e.g. [31, 36, 42, 44, 45]). Since November 2017, the maximum number of characters to be used within a tweet is set to 280 (cf. [46, 47]). This shortness in text length must be taken into consideration when identifying sentiments expressed within tweets [36, 44]. Furthermore, our analysis has to cope with the huge number of users around the world that post tweets (cf. [31, 39, 42, 44, 48]). Therefore, we assume that the methods for processing the tweets need to be fast and performant and must deal with the shortness of tweets to obtain accurate and reliable results [45].

5.2. Data extraction, data preprocessing and data reduction

Using Twitter's sampling Application Programmable Interface (API) in the "*Spritzer*" version, we sampled approximately uniformly from all messages being posted via Twitter in 2019. The data collection procedure resulted in 358.923.210 English tweets in the time range from January 1st, 2019 to December 31st, 2019. For our investigation, we only collected tweets written in English. To that purpose, we filtered the provided language field to determine only tweets with the value "EN", which indicates that a tweet is written in English language. By this means, we were able to omit the complications of multiple languages (cf. [49]). Further restrictions beyond the language restriction as sampling by only incorporating tweets containing certain hashtags have not been applied.

Fluctuations in collective public emotions and sentiments can occur due to a multitude of competing effects (cf. [50]) and can influence the results of this investigation. By examining a whole year of tweets, seasonal influences and corresponding distortions resulting from deviating levels of sentiments expressed in certain months, such as higher temperature and more positive sentiments in summer months than in winter months, can be omitted. Therefore, because all twelve months of a year are included in our investigation, the influences of the incorporated contexts regarding sentiments are not distorted by a month being not included. As our data collection comprises 358.923.210 tweets, we assume that influences of external events and competing effects are smoothed to a high degree. We further regard the sample as a representative collection of tweets appropriate for our investigations because Twitter's "*Spritzer*" API provides 1 % of all tweets posted with a maximum margin of error of 0.06 at a confidence level of 99 % [51]. Each tweet delivered by Twitter comes as a JSON (Java Script Object Notation) object that contains tweet text and meta data characterizing both tweets and the situation in which it has been posted. These meta data, such as timestamp, language, source, geolocation and the device used for tweeting (cf. [19]), are logged and provided by Twitter.

Subsequently, we performed exploratory analysis to identify noise and outliers within the obtained data. By this means, we noticed that not every tweet is provided with location or device information. Thus, we agreed on performing analysis concerning influences of location and mobile devices' OS contexts only with tweets containing this context information. Filling up missing values by applying any of the commonly proposed strategies (cf. [52]) could have distorting effects as there were many data instances with missing attributes. To ensure the dataset to be without any redundant tweets, we additionally applied redundancy detection using tweet text and the creation date as the properties of a redundancy. The following steps of data reduction were carried out as part of the transformation of the JSON objects into CSV (Comma Separated Values) files. In this step, we excluded entries that don't match the goals of our investigation. Many of the provided meta data are not necessary for our analysis so that we only included the device field, the self-reported location field and the tweet text to identify sentiments expressed within the tweets. Only a small fraction of the tweets has geolocation coordinates that can be mapped directly to locations in terms of latitude and longitude (cf. [48]). Therefore, we relied on parsing the free-

response location field that accompanies a tweet. By applying the methods from the Pycountry library (*https://pypi.org/project/pycountry/*), we mapped the self-reported location information to continents and appended the continent to the corresponding tweet within the CSV files.

5.3. Selection of methods to be applied

Identifying the influences of contexts on sentiments expressed within tweets requires a method to assign sentiment values to tweets. Considering the findings from *subchapter 5.1.*, we decided to apply the freely accessible "*Valence Aware Dictionary for sEntiment Reasoning*" (VADER) approach [45]. VADER was specifically developed for sentiment analysis in microblog-like texts and has achieved remarkable results compared to other sentiment analysis approaches [45]. The approach is also fast enough and can deal with huge numbers of tweets as required for our investigation. Because VADER identifies sentiment values using a built-in sentiment dictionary [45], there is no need for labeled training data. VADER provides a compound sentiment score that combines positive and negative sentiments into one single value. This score can take values in the range of -1 to +1 [45] and is calculated separately for each tweet contained in the data set. Furthermore, we applied summarization measures and decided to calculate mean values in terms of sentiments for all contexts being investigated. As the calculation of mean sentiment values consolidates several sentiment values to one single value, we additionally decided to determine the number of tweets whose sentiments are at certain intervals. This included counting the number of tweets that have certain sentiment values regarding the investigated context factors. In doing so, determining the sentiment intervals is independent from calculating the mean sentiment values.

To test the statistical significance of our findings, we additionally carried out t-tests (cf. [53]). In our case, there are independent samples since one sample selected from one population is not related in any way to the sample from another population [53]. This is because the assignment of tweets and the respective sentiment to a continent and a mobile devices' OS is exclusive and does not consider an assignment to more than one continent or mobile devices' OS. To be applicable, the tests require the dependent variable to be at least interval scaled. We see this requirement fulfilled as the dependent variable (sentiment value) of each tweet is numeric. Additionally, the independent variables (contextual factors) are at least nominal-scaled [53]. In the following chapters, every time the term "significant" or "significantly" is used, the differences in mean sentiment values or proportions within sentiment intervals showed to be significant by the pairwise calculated t-tests.

6. Performing the analysis

We observed a high proportion of neutral tweets by filtering for tweets with a sentiment value of "0" within the appended sentiment value field. As this huge number of neutral tweets can have remarkable influences on the results, we decided to exclude neutral tweets from further analysis. Therefore, we proceeded with the remaining 245.077.312 tweets being either positive or negative. Then, we filtered the tweets (*cf. table 1*) along with the corresponding sentiment values. For the analysis regarding one context, either continents or devices, we applied one filter criterion. E.g. by filtering the tweets with the value (a) "Europe" or (b) "Android", all tweets originating from (a) Europe *(cf. table 1 - IDs 1, 2 and 4)* or tweets sent from an Android OS *(cf. table 1 - IDs 1 and 4)* and their sentiment values are retrieved. When combining the contexts of continents and mobile devices' OS, we simultaneously set two filter criteria. We retrieve e.g. all tweets sent from Android powered devices originating from Europe together with the corresponding sentiment values by filtering with "Android" and "Europe" *(cf. table 1 - IDs 1 and 4)*.

Thereafter, the mean values of the sentiment values regarding the selected singular and combinatorial contexts were calculated and the numbers of occurrences of sentiment values in the respective sentiment intervals were counted. The results of determining the mean sentiment values and the sentiment intervals are more closely described in the next chapter.

Table 1: Example Tweets and Results

IDs	***Example Tweets***		
1	*I pray that your August will be full of good news, positivity and blessings.*		
2	*Hope you're having a great week.*		
3	*I wish you all the best :-)*		
4	*I have been on hold with you for 40 minutes and then the call hangs up. Poor customer service!!!*		
…	**…**		
IDs	***Results***		
	Continent	***Mobile OS***	***Sentiment***
1	Europe	Android	0.8481
2	Europe	Apple	0.7732
3	Asia	Apple	0.8481
4	Europe	Android	-0.4767
…	**…**	**…**	**…**

7. Reporting the results

7.1. Results of univariate analysis

Users of Apple devices turned out to be less positive compared to users of Android devices (*cf. table 2*). Tweets posted by Android devices have on average higher sentiments than tweets sent by Apple devices. Although the differences for mean sentiment values between Apple and Android users seem to be comparably low in our analysis, they showed to be significant. We additionally determined the number of tweets whose sentiments are at certain intervals (*cf. table 4*). Regarding the mobile devices' OS, the distributions reflect that 12.89 % of the tweets sent by Apple devices are strongly negative (within [-1; -0.66[), compared to 9.32 % for tweets sent by Android devices. As regards the strongly positive tweets (within [+0.66 to +1]), the proportion of Android tweets is about 5 % higher than the proportion of tweets sent by Apple devices.

In terms of location as the second context, we obtained results for six continents, namely *Africa (AF), Asia (AS), Europe (EU), North America (NA), Oceania (OC)* and *South America (SA)*. Tweets originating from AS showed to be the most positive, followed by tweets from SA, EU and AF (*cf. table 2*). Interestingly, the average sentiment values for tweets from OC and NA show significant differences. In addition, we notice that tweets in the range of +0.66 to +1.0 originate most frequently from SA, followed by AS, EU and OC (*cf. table 4*). Considering the mean sentiment values, it was assumable for NA having a low proportion of tweets within this interval. Based on the mean sentiment value of AF, it is remarkable that tweets are strongly positive (within [+1; +0.66]) with a proportion of only 23.19 %. However, tweets with mean sentiment values between +0.33 and +0.66 occur most often for AF, while the other continents have at least 7.85 % fewer proportions in this interval. When investigating the intervals for negative sentiments, it is significant that NA consistently has the highest proportions, followed by OC. Asia having the highest mean sentiment value, interestingly shows as well comparably high proportions in the strongly negative and the negative intervals (within [-1; -0.66[and [-0.66; -0.33[).

Table 2: Sentiments of singular contexts

Singular Contexts		*Mean Sentiment Values*
Continents	***Asia***	0.21452
	South America	0.19716
	Europe	0.19327
	Africa	0.18906
	Oceania	0.16195
	North America	0.12869
Mobile OS	***Android***	0.20179
	Apple	0.13223

7.2. Results of bivariate analysis

The observations of the univariate analysis (*cf. tables 2 and 4*) are partly reflected within the bivariate analysis (*cf. table 3 and 5*). Combining each continent with Android gives on average always more positive sentiment values as if the same continents are combined with Apple. Hereby, the highest difference can be observed for Asian tweets, where Android achieves on average significantly higher sentiment values compared to Apple (Δ 0.10366). South American and African tweets from Android devices are also significantly more positive. Differences observed for NA, EU and OC are however comparably marginal. Tweets posted from AF, EU and OC using Apple devices have consistently higher proportions within the negative sentiment intervals and consistently lower proportions for the positive sentiment intervals (*cf. table 5*). South American tweets having a sentiment value in the range of]0; +0.33[occur slightly more often for Apple devices than Android. In Asia, there are also comparably many positive tweets that are sent from Apple devices compared to Android. However, strongly positive tweets (within [+0.66; +1.0]) originate significantly more often from Android devices (Δ 9.33 %). The observations for North America are also remarkable, as one would expect to find a higher number of positive tweets from Apple devices because these devices originate from NA. However, NA does not show the expected higher sentiment values due to a possible connection of this continent to the brand Apple that is based there. But Android powered devices predominantly originating from AS provoke more positive sentiment scores for this continent.

Table 3: Sentiments of combined contexts

Combined Contexts		*Mean Sentiment Values*
Android	***Asia***	0.25694
	South America	0.22781
	Africa	0.21481
	Europe	0.19987
	Oceania	0.16436
	North America	0.13537
Apple	***Europe***	0.17091
	Asia	0.15328
	Africa	0.15323
	South America	0.14674
	Oceania	0.14323
	North America	0.10668

Table 4: Proportions of sentiment intervals for continents and mobile devices' OS

Singular Contexts	[-1; -0.66[	[-0.66; -0.33[	[-0.33; 0[	]0; 0.33[	[0.33; 0.66[	[0.66; 1]
Africa	8.46 %	13.05 %	8.94 %	10.02 %	36.34 %	23.19 %
Asia	9.32 %	14.36 %	9.65 %	11.15 %	28.11 %	27.41 %
North America	12.03 %	16.92 %	10.77 %	11.41 %	26.49 %	22.38 %
Europe	9.02 %	15.12 %	10.02 %	11.63 %	28.49 %	25.72 %
Oceania	10.31 %	15.93 %	10.61 %	11.66 %	27.74 %	23.75 %
South America	8.12 %	13.89 %	8.86 %	10.28 %	26.33 %	32.52 %
Android	9.32 %	14.55 %	10.01 %	10.89 %	27.57 %	27.66 %
Apple	12.89 %	17.21 %	10.89 %	11.19 %	25.80 %	22.02 %

Table 5: Proportions sentiment intervals for the combinations of continents and mobile devices' OS

Combined Contexts	*Mobile OS*											
	Apple						*Android*					
Continents	[-1; -0.66[	[-0.66; -0.33[	[-0.33; 0[	]0; 0.33[	[0.33; 0.66[	[0.66; 1]	[-1; -0.66[	[-0.66; -0.33[	[-0.33; 0[	]0; 0.33[	[0.33; 0.66[	[0.66; 1]
Africa	10.73 %	16.19 %	10.50 %	11.55 %	27.08 %	23.95 %	8.34 %	14.04 %	10.19 %	11.64 %	28.13 %	27.66 %
Asia	10.16 %	16.55 %	10.68 %	11.13 %	29.43 %	22.05 %	7.77 %	13.21 %	8.80 %	11.21 %	27.63 %	31.38 %
North America	12.96 %	19.38 %	11.13 %	10.81 %	25.94 %	19.78 %	10.71 %	17.27 %	10.87 %	11.01 %	26.40 %	23.74 %
Europe	9.42 %	16.44 %	10.35 %	11.42 %	28.25 %	24.12 %	7.98 %	15.66 %	10.30 %	11.64 %	27.84 %	26.58 %
Oceania	8.26 %	18.15 %	11.75 %	12.50 %	27.53 %	21.81 %	9.46 %	16.45 %	10.58 %	11.96 %	27.57 %	23.98 %
South America	10.26 %	16.82 %	10.76 %	12.35 %	28.90 %	20.91 %	7.79 %	14.29 %	9.37 %	11.34 %	29.14 %	28.07 %

7.3. Interpretation and implications

Actions, decisions and sentiments expressed within social media texts are influenced by the contexts in which people act. Certain manifestations of contexts thereby provoke more likely positive or negative sentiments. While reacting is only possible after certain conditions have already occurred, including contexts into decision-making enables companies to perform preventive actions. They can benefit by adapting the way they communicate within certain markets by aligning to the sentiments provoked by contexts. E.g. as customers from certain countries or continents may be more or less likely expressing negative sentiments within social media posts, companies could thereafter adjust their customer communication to counter sentiment tendencies prevalent there. Regarding the results of our analysis, companies should especially adapt messages for customers originating from NA, and those North Americans that send tweets from their Apple devices. Although the contexts NA and Apple provoke negative sentiments, the combination of NA and Apple devices turned out to provoke the most negative sentiments according to our results (*cf. tables 2 and 3*). Therefore, we suppose that there is the most potential for this combination when companies want to include contexts into decision-making within the activities and tasks of customer communication. By providing tailored contents and messages for these customers, the negative attitudes caused by the corresponding context factors of NA (continent) and Apple (mobile devices' OS) could be countered. Instead, when companies intentionally aim at provoking pWoM, they should better concentrate on AS and Android users.

Our findings retrieved for the mobile devices' OS are supported by the findings of [29] who report their results from a psychological investigation. Users of Apple devices are associated with more negative traits. They are perceived as less honest, less humble and are considered to manipulate others more often to gain personal advantages [29]. Hereby, the results of [42] are contradicting the findings of [29] and our results. In this work, tweets sent from Blackberry devices are associated with more positive sentiments compared to tweets sent from Apple or Android devices (cf. [42]). As the data set of [42] comprises the time range of May 1st, 2012 to April 30th, 2014, we assume that this data no longer reflect the current situation of mobile devices being used. The observed differences of Android users being more positive within our results could therefore be justified by Blackberry users that switched to Android devices. The market share of Blackberry has continuously dropped so that there are nowadays predominantly two major mobile OS, Android and Apple's iOS, that dominate the mobile devices market [54].

8. Conclusion

Our paper provides an overview about sentiment-influential contexts within tweets, which is followed by the identification of the influences of continents, mobile devices' OS and the combination of both. Our approach is structured into preparatory steps, the execution of the analysis and the reporting of the corresponding results. The results of our analysis have implications for the management of customer communications within social media channels because companies strongly build on social media to foster the external communication with customers (e.g. [55, 56]). NWoM expressed by disappointed or angry customers within social media channels (e.g. Twitter or Facebook) has the potential to negatively impact the perception of (potential) customers towards a company (cf. [5-7, 11-13]). As customers reporting negative experiences within Twitter await replies within one to three hours, companies have to react quickly [10] or even better take preventive actions. Hereby, the contexts inducing negative sentiments can support the corresponding decision-making. In addition, we see contributions of contexts for companies intentionally aiming at provoking pWoM. Companies that use Twitter for customer communication benefit most from our results and findings about sentiment-inducing contexts due to the required higher reaction speed [10]. Our results can support this task as our investigation is based on tweets as research objects. Companies apply Facebook as well as a customer communication channel where customers also articulate negative perceptions and experiences (cf. [5, 6, 10]). Therefore, we recommend as a possible step for future research to give an overview of sentiment-influential contexts and corresponding effects within Facebook posts.

The paper on hand is however not free of limitations. First, the analysis performed in our investigation was only performed with the VADER sentiment approach. Secondly, our results can be seen rather as initial findings derived from a data analytics procedure. Therefore, it is a necessary step in future research to deduce more sound recommendations from our results. Thirdly, the results of our work are based on the tweets of the year 2019. For this reason, we propose to apply the described analysis procedure to the data of previous years and to figure out observed similarities and differences regarding the influences of context factors.

9. References

[1] Dickey, I.J., and Lewis, W.F., "The evolution (revolution) of social media and social networking as a necessary topic in the marketing curriculum: a case for integrating social media into marketing classes", Advances in Marketing: Embracing Challenges and Change-A Global Perspective. 2010.

[2] Chaffey, D., "Global social media research summary 2016", Smart Insights: Social Media Marketing. 2016.

[3] Kietzmann, J.H., Hermkens, K., McCarthy, I.P., and Silvestre, B.S., "Social media? Get serious! Understanding the functional building blocks of social media", Business horizons. 54(3), 2011, pp. 241-251.

[4] Jansen, B.J., Zhang, M., Sobel, K., and Chowdury, A., "Twitter power: Tweets as electronic word of mouth", Journal of the American society for information science and technology. 60(11), 2009, pp. 2169-2188.

[5] Grégoire, Y., Salle, A., and Tripp, T.M., "Managing social media crises with your customers: The good, the bad, and the ugly", Business Horizons. 58(2), 2015, pp. 173-182.

[6] Einwiller, S.A., and Steilen, S., "Handling complaints on social network sites–An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies", Public Relations Review. 41(2), 2015, pp. 195-204.

[7] Pfeffer, J., Zorbach, T., and Carley, K.M., "Understanding online firestorms: Negative word-of-mouth dynamics in social media networks", Journal of Marketing Communications. 20(1-2), 2014, pp. 117-128.

[8] Naveed, N., Gottron, T., Kunegis, J., and Alhadi, A.C., "Bad news travel fast: A content-based analysis of interestingness on twitter", in Proceedings of the 3rd international web science conference, 2011.

[9] Joireman, J., Grégoire, Y., Devezer, B., and Tripp, T.M., "When do customers offer firms a “second chance” following a double deviation? The impact of inferred firm motives on customer revenge and reconciliation", Journal of Retailing. 89(3), 2013, pp. 315-337.

[10] Istanbulluoglu, D., "Complaint handling on social media: The impact of multiple response times on consumer satisfaction", Computers in Human Behavior. 74, 2017, pp. 72-82.

[11] Chevalier, J.A., and Mayzlin, D., "The effect of word of mouth on sales: Online book reviews", Journal of marketing research. 43(3), 2006, pp. 345-354.

[12] Mizerski, R.W., "An attribution explanation of the disproportionate influence of unfavorable information", Journal of Consumer Research. 9(3), 1982, pp. 301-310.

[13] Kim, S.J., Wang, R.J.-H., Maslowska, E., and Malthouse, E.C., "“Understanding a fury in your words”: The effects of posting and viewing electronic negative word-of-mouth on purchase behaviors", Computers in Human Behavior. 54, 2016, pp. 511-521.

[14] Greenberg, P., "CRM at the speed of light: Capturing and keeping customers in Internet real time": Elsevier. 2001.

[15] Liu, B., "Sentiment analysis and opinion mining": Morgan & Claypool Publishers. 2012.

[16] Cao, X., MacNaughton, P., Deng, Z., Yin, J., Zhang, X., and Allen, J.G., "Using twitter to better understand the spatiotemporal patterns of public sentiment: a case study in Massachusetts, USA", International journal of environmental research and public health. 15(2), 2018.

[17] Hu, T., She, B., Duan, L., Yue, H., and Clunis, J., "A Systematic Spatial and Temporal Sentiment Analysis on Geo-Tweets", IEEE Access. 8, 2019, pp. 8658-8667.

[18] Golder, S.A., and Macy, M.W., "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures", Science. 333(6051), 2011, pp. 1878-1881.

[19] Plunz, R.A., Zhou, Y., Vintimilla, M.I.C., Mckeown, K., Yu, T., Uguccioni, L., and Sutto, M.P., "Twitter sentiment in New York City parks as measure of well-being", Landscape and urban planning. 189, 2019, pp. 235-246.

[20] Lim, K.H., Lee, K.E., Kendal, D., Rashidi, L., Naghizade, E., Winter, S., and Vasardani, M., "The grass is greener on the other side: Understanding the effects of green spaces on Twitter user sentiments", in Companion Proceedings of the The Web Conference 2018.

[21] Schwartz, A.J., Dodds, P.S., O'Neil-Dunne, J.P., Danforth, C.M., and Ricketts, T.H., "Visitors to urban greenspace have higher sentiment and lower negativity on Twitter", People and Nature. 1(4), 2019, pp. 476-485.

[22] Sandstrom, G.M., Lathia, N., Mascolo, C., and Rentfrow, P.J., "Putting mood in context: Using smartphones to examine how people feel in different locations", Journal of Research in Personality. 69, 2017, pp. 96-101.

[23] Kaplan, A.M., and Haenlein, M., "Two hearts in three-quarter time: How to waltz the social media/viral marketing dance", Business horizons. 54(3), 2011, pp. 253-263.

[24] Kaplan, A.M., and Haenlein, M., "Users of the world, unite! The challenges and opportunities of Social Media", Business horizons. 53(1), 2010, pp. 59-68.

[25] Hennig-Thurau, T., Wiertz, C., and Feldhaus, F., "Does Twitter matter? The impact of microblogging word of mouth on consumers’ adoption of new movies", Journal of the Academy of Marketing Science. 43(3), 2015, pp. 375-394.

[26] Kundi, F.M., Ahmad, S., Khan, A., and Asghar, M.Z., "Detection and scoring of internet slangs for sentiment analysis using SentiWordNet", Life Science Journal. 11(9), 2014, pp. 66-72.

[27] Mehrabian, A., " An approach to environmental psychology", 1974, Cambridge, MA: MIT Press, 1974.

[28] Dey, A.K., "Understanding and using context", Personal and ubiquitous computing. 5(1), 2001, pp. 4-7.

[29] Shaw, H., Ellis, D.A., Kendrick, L.-R., Ziegler, F., and Wiseman, R., "Predicting smartphone operating system from personality and individual differences",

Cyberpsychology, Behavior, and Social Networking. 19(12), 2016, pp. 727-732.
[30] Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A., and Danforth, C.M., "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter", PloS one. 6(12), 2011.
[31] Wang, W., Hernandez, I., Newman, D.A., He, J., and Bian, J., "Twitter analysis: Studying US weekly trends in work stress and emotion", Applied Psychology. 65(2), 2016, pp. 355-378.
[32] Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S., and Danforth, C.M., "The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place", PloS one. 8(5), 2013.
[33] Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., and Rodrigue, J., "A demographic analysis of online sentiment during hurricane irene", in Proceedings of the second workshop on language in social media, 2012: Association for Computational Linguistics.
[34] Baylis, P., Obradovich, N., Kryvasheyeu, Y., Chen, H., Coviello, L., Moro, E., Cebrian, M., and Fowler, J.H., "Weather impacts expressed sentiment", PloS one. 13(4), 2018.
[35] Li, J., Wang, X., and Hovy, E., "What a nasty day: Exploring mood-weather relationship from twitter", in proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, 2014.
[36] Neppalli, V.K., Caragea, C., Squicciarini, A., Tapia, A., and Stehle, S., "Sentiment analysis during Hurricane Sandy in emergency response", International journal of disaster risk reduction. 21, 2017, pp. 213-222.
[37] Oh, O., Kwon, K.H., and Rao, H.R., "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010", in ICIS, 2010.
[38] Lin, Y.-R., "Assessing sentiment segregation in urban communities", in Proceedings of the 2014 international conference on Social Computing, 2014.
[39] Lerman, K., Arora, M., Gallegos, L., Kumaraguru, P., and Garcia, D., "Emotions, demographics and sociability in Twitter interactions", in Tenth International AAAI Conference on Web and Social Media, 2016.
[40] Murthy, D., Bowman, S., Gross, A.J., and McGarry, M., "Do we tweet differently from our mobile devices? A study of language differences on mobile and web-based Twitter platforms", Journal of Communication. 65(5), 2015, pp. 816-837.
[41] Groshek, J., and Cutino, C., "Meaner on mobile: Incivility and impoliteness in communicating contentious politics on sociotechnical networks", Social Media+ Society. 2(4), 2016.
[42] Cruz-Albrecht, L., Xu, J., Ni, K.-Y., and Lu, T.-C., "Characterizing Regional and Behavioral Device Variations Across the Twitter Timeline: A Longitudinal Study", in Proceedings of the 2017 ACM on Web Science Conference, 2017.
[43] Aggarwal, C.C., and Zhai, C., "Mining text data": Springer Science & Business Media. 2012.
[44] Kaplan, A.M., and Haenlein, M., "The early bird catches the news: Nine things you should know about micro-blogging", Business horizons. 54(2), 2011, pp. 105-113.
[45] Hutto, C.J., and Gilbert, E., "Vader: A parsimonious rule-based model for sentiment analysis of social media text", in Eighth international AAAI conference on weblogs and social media, 2014.
[46] Newton, C., "Twitter just doubled the character limit for tweets to 280", The Verge. 2017.
[47] Busby, M., "Twitter to introduce expanded 280-character tweets for all its users", The Guardian. 2017.
[48] Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Lucas, R.E., Agrawal, M., Park, G.J., Lakshmikanth, S.K., Jha, S., and Seligman, M.E., "Characterizing geographic variation in well-being using tweets", in Seventh International AAAI Conference on Weblogs and Social Media, 2013.
[49] Thelwall, M., Buckley, K., and Paltoglou, G., "Sentiment in Twitter events", Journal of the American Society for Information Science and Technology. 62(2), 2011, pp. 406-418.
[50] Lansdall-Welfare, T., Dzogang, F., and Cristianini, N., "Change-point analysis of the public mood in UK Twitter during the Brexit referendum", in 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), 2016: IEEE.
[51] Gerlitz, C., and Rieder, B., "Mining one percent of Twitter: Collections, baselines, sampling", M/C Journal. 16(2), 2013.
[52] Bramer, M., "Principles of data mining", Vol. 180: Springer. 2007.
[53] Sakai, T., "Two sample t-tests for ir evaluation: Student or welch?", in Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, 2016.
[54] Statista. "Marktanteile der führenden mobilen Betriebssysteme an der Internetnutzung mit Mobiltelefonen weltweit von September 2009 bis Mai 2020", visited on 06.07.2020, available from: https://de.statista.com/statistik/daten/studie/184335/umfrage/marktanteil-der-mobilen-betriebssysteme-weltweit-seit-2009/.
[55] Stobbe, A., "Enterprise 2.0–Wie Unternehmen das Web 2.0 für sich nutzen. Deutsche Bank Research", Economics. 78, 2010, pp. 1-16.
[56] Heidemann, J., Klier, M., and Probst, F., "Online social networks: A survey of a global phenomenon", Computer networks. 56(18), 2012, pp. 3866-3878.

2.2. **Beitrag 2:** Comparison of Topic Modelling Techniques in Marketing – Results From an Analysis of Distinctive Use Cases

Adressierte Forschungsfragen	**(1)** Anhand welcher Kriterien lassen sich Topic Modelling Techniken miteinander vergleichen? **(2)** Welche Topic Modelling Technik sollte basierend auf den identifizierten Vergleichskriterien für die marketingbezogenen Anwendungsfälle Themenextraktion, Trendanalyse und Themenstrukturierung verwendet werden?
Zielstellungen	• Identifikation von Anforderungen an eine Topic Modelling Technik hinsichtlich der marketingbezogenen Anwendungsfälle (a) Themenextraktion, (b) Trendanalyse und (c) Themenstrukturierung. • Vergleich der Topic Modelling Techniken LDA, PAM und DMR anhand der Metriken Berechnungszeit („Model Building Time"), Log-Likelihood, Coherence, Perplexity, Word Intrusion und Topic Intrusion. • Identifikation von Unterschieden dieser Topic Modelling Techniken, um Handlungsempfehlungen für die Auswahl einer bestmöglich geeignetsten Technik für die drei Anwendungsfälle (a)-(c) abzuleiten.
Forschungsmethode	**Goal Question Metric nach Basili et al. (1994)** • Goal: Vergleichsparameter definieren, um die ausgewählten Topic Modelling Techniken miteinander vergleichen zu können. • Question: Inwieweit erfüllen die drei Topic Modelling Techniken (LDA, PAM, DMR) die Anforderungen der marketingbezogenen Anwendungsfälle (a)-(c). • Metric: Metriken werden aus der Forschungsliteratur identifiziert, um die Vergleichsparameter zu operationalisieren und die Leistungsfähigkeit der Topic Modelling Techniken vor dem Hintergrund der drei Anwendungsfälle quantitativ zu erfassen.
Kernergebnisse (Überblick)	• Identifikation von insgesamt elf Anforderungen an Topic Modelling Techniken sowie von sechs verschiedenen Vergleichsmetriken. • Zuordnung der Anforderungen und Metriken zu den in der Literatur identifizierten, marketingbezogenen Anwendungsfällen (a)-(c). • Vergleich von LDA, PAM und DMR anhand eines Facebook Datensatzes mit ca. 4,1 Mio. Posts unter Anwendung der Metriken. • Ableitung von Anwendungsempfehlungen für die drei Techniken: o Themenextraktion: LDA (v. a. bei kleineren Dimensionen) oder DMR (wenn eine Vielzahl an Themen extrahiert werden soll) o Trendanalyse: DMR o Themenstrukturierung: PAM
Publikationsort	**Veröffentlicht** in den Proceedings der 29. European Conference on Information Systems, A Virtual AIS Conference 2021.
Ranking VHB JQ 4	**A**
Autoren und Anteile	Janik Wörner 40 % **Daniel Konadl** **25 %** Isabel Schmid 25 % Susanne Leist 10 %

Tabelle 4: Fact Sheet zu Beitrag 2

COMPARISON OF TOPIC MODELLING TECHNIQUES IN MARKETING - RESULTS FROM AN ANALYSIS OF DISTINCTIVE USE CASES

Research paper

Janik Wörner, University of Regensburg, Regensburg, Germany, Janik.Woerner@ur.de
Daniel Konadl, University of Regensburg, Regensburg, Germany, Daniel.Konadl@ur.de
Isabel Schmid, University of Regensburg, Regensburg, Germany, Isabel.Schmid@ur.de
Susanne Leist, University of Regensburg, Regensburg, Germany, Susanne.Leist@ur.de

Abstract

Currently, topic modelling is an effective analytical tool for the automated investigation of text data. However, identifying the underlying topics is still a challenging task that is dependent on the selection of the proper technique. Moreover, due to the considerable number of topic modelling techniques reported in the literature, uncertainty about the application of the techniques arises for both researchers and practitioners. Therefore, we conducted a comparison of three different topic modelling techniques (LDA, PAM, DMR) to give recommendations for three use cases identified in the literature: content extraction, trend analysis and content structuring. For each of them, we identified several requirements and by conducting the method 'Goal Question Metric', we derived several comparison metrics. We applied these metrics to a real-world Facebook data set (4,155,992 posts) to highlight the differences between the three topic modelling techniques and to give recommendations for our defined use cases.

Keywords: topic modelling, social media analysis, text analysis, marketing use cases

1 Introduction

Topic modelling is a prevalent kind of probabilistic generative model for extracting latent variables from large unstructured data sets (Liu et al., 2016). This can be applied to analyse different data such as bioinformatics data (Coelho et al., 2010), environmental data (Girdhar et al., 2013), and text data (Vayansky and Kumar, 2020). Thus, topic modelling has been studied in different disciplines and is also prevailing in information systems (IS) research, mainly focusing on Latent Dirichlet Allocation (LDA) (Blei et al., 2003) because of its simple applicability and good analysis results (Debortoli et al., 2016, Eickhoff and Neuss, 2017). With the continuous growth of social media and the consequential transformation of the way individuals interact with each other, increasing amounts of written data are created that can be analysed i.a. to support marketing related decision-making (Ghosh and Guha, 2013). Thus, social media data are increasingly used to enrich marketing tasks, such as complaint management (Einwiller and Steilen, 2015, Grégoire et al., 2015), innovation management (Mount and Martinez, 2014, Piller et al., 2012), or sales (Guesalaga, 2016, Marshall et al., 2012). However, the huge amount of written social media contents (Statista, 2020) complicates manual content analysis.

To solve this problem, automated topic discovery techniques and – in particular – topic modelling have been widely investigated (Chinnov et al., 2015, Eickhoff and Neuss, 2017, Hong and Davison, 2010). Topic modelling enables the analysis of a large amount of written social media data to extract embedded topics. Therefore, topic modelling has facilitated addressing marketing related questions and problems that have exceeded the feasibility of in-depth qualitative analysis (Eickhoff and Neuss, 2017). Thus, marketing related problems that refer to (1) content extraction, (2) trend analysis and (3) content structuring have often been discussed in the literature. Companies are required to base their products and services on customer requirements. Therefore, (1) content extraction with topic modelling is an appropriate application to extract customers' praise and criticism for product planning purposes (Irawan et

al., 2020, Rathore et al., 2018). To be aware of evolving trends concerning their own products and services, marketing departments conduct (2) trend analysis. Tracking evolving and changing requirements of customers is imperative to fulfil customers' wishes (Hong et al., 2012, Lozano et al., 2017). Moreover, (3) content structuring can help marketing departments to gain deeper insights into topics and their interrelatedness (Anoop et al., 2015, Srijith et al., 2017). Extracted hierarchical structures can reveal relationships between topics (e.g. price and product quality) and support more coordinated and sounder decision-making. Nevertheless, identifying the underlying topics of these documents is still a challenging task as the reasonable extraction of significant statistics and features from a dataset is dependent on the selection of the proper technique (Vayansky and Kumar, 2020).

As mentioned above, a growing number of IS-related investigations are currently using LDA. However, the basic LDA cannot represent all use cases (e.g. mapping hierarchies) for marketing related tasks so that extensions of LDA are essential. Therefore, not only do the numerous existing techniques for topic modelling hinder practical applications, but also the necessity of advanced technique-related knowledge. Liu et al. (2016) have divided various extensions of LDA into three areas: (I) extension of topic attributes (II) extension of document attributes and (III) extension of word attributes. Although numerous techniques are presented in these three areas, such as the Partially Labelled LDA (PLLDA) (Ramage et al., 2011), the Dirichlet Multinomial Regression (DMR) (Mimno and McCallum, 2008), or the Pachinko Allocation Model (PAM) (Li and McCallum, 2006), these extensions are scarcely applied. Due to the large number of topic modelling techniques in the current research literature, uncertainty about the selection of the right technique can arise. Vakansky and Kumar (2020) addressed this problem by conducting a theoretical comparison of different topic modelling techniques based on a structured literature review. Although this serves as a good overview of various topic modelling approaches and as a starting point for selecting a technique, differences only become obvious when applying them to a real-world data set. Furthermore, the results do not give clear suggestions which problem should be addressed with which technique. We addressed these problems by conducting a comparison between three different topic modelling techniques to give recommendations regarding the three use cases of (1) content extraction, (2) trend analysis, and (3) content structuring. We contribute to close this identified gap by comparing the practical application of the three techniques. This leads to the following research questions:

RQ1: Which criteria can be used to compare the different topic modelling techniques with each other?
RQ2: Which topic modelling technique can be recommended for the marketing related use cases (1) content extraction, (2) trend analysis, and (3) content structuring?

Topic modelling, both in general and especially regarding the analysis of companies´ social media posts, represents an important area for IS research. Accordingly, against the background of marketing we uncover various corporate use cases in the context of social media. By applying and comparing the three topic modelling techniques LDA, DMR, and PAM and by using build time, log-likelihood, coherence, word and topic intrusion as evaluation measures, we want to show differences between these techniques, identify advantages, disadvantages, and various potentials to enhance topic modelling techniques. Hence, we apply LDA, DMR, and PAM to a real-world data set. The remainder of this paper is as follows: section 2 provides a theoretical background. Then, we refer to the derivation of the three use cases and the respective requirements from literature. The transformation of them into topic modelling related metrics is also described here. Next, the procedure of the research 'Goal Question Metric' (cf. Basili, 1994) is described in section 3. Section 4 deals with the selection of the topic modelling techniques that are used for our comparison. The data analysis in section 5 achieves this comparison and further explains the data collection, the preparation of the data and the data analysis. Afterwards, in section 6, we present and discuss our results. Finally, section 7 draws an overall conclusion.

2 Theoretical Background

2.1 Social media

Social media can be defined as '*a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated*

Content (UGC)' (Kaplan and Haenlein, 2010, p.61). Social media connects people with the same interests, activities, backgrounds, or friendships (AlFalahi et al., 2014, Schneider et al., 2009). The active utilisation of social media enjoys particularly great popularity in private use. However, companies have also adopted social media to support value-creation (Hanna et al., 2011, McDonald and Aron, 2012). In particular, many companies apply these Internet-based applications such as content communities (e.g. YouTube), blogs, or social networks (e.g. Facebook or Twitter) to enable communication mainly with external stakeholders (Kietzmann et al., 2011). Thus, companies adopt social media to achieve different business objectives such as branding, including advertising, marketing, and content delivery for creating brand awareness (Culnan et al., 2010, Di Gangi et al., 2010, Kietzmann et al., 2011). Therefore, social media such as Facebook or Twitter serve as an important interface between companies and customers. This interface generates large amounts of data that need to be analysed and interpreted, as a company can strongly benefit from these data. In addition to structured social media data (e.g. timestamps, like counts, etc.), especially unstructured text data contain interesting contents for companies. Posts and comments often include an user´s major wishes, ideas, and expectations towards products, services, or a company in general (Hienerth et al., 2011, Sigala, 2012a, Sigala, 2012b). The so called 'Voice of the Customers' can be used to adjust marketing campaigns, to identify and support the position in the market and to adjust product features on customers´ favourability. However, to uncover this useful information from the large amount of data requires considerable effort (Dahal et al., 2019, Kumar and George, 2007, Womack and Jones, 1996). To avoid this problem, automated analysis of social media data such as social network analysis, sentiment analysis, and topic modelling can be conducted. Especially through the latter one, valuable information for companies can be extracted, as this technique is able to identify various (discussion) topics, perceptions, and opinions (Dahal et al., 2019, Lozano et al., 2017). However, automated analysis such as topic modelling are often complicated, as many companies are not familiar with the applied techniques, its implementation, and its purpose (Dai et al., 2011).

2.2 Topic modelling

Topic modelling aims to determine content structures in underlying document collections. Hereby, topic modelling refers to the use of generative probability models for determining latent relationships within a corpus of text data. The dataset under investigation is to be seen as a mixture of individual documents, where each document affects several corpus-wide topics, that in turn consist of frequently occurring words within the dataset (Blei, 2012). LDA can be considered as one of the most fundamental works in the topic discovery research area, wherefore a growing number of investigations currently uses this technique proposed by Blei et al. (2003). The authors describe their probabilistic model as *'a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics'* (p. 993). This means that LDA is a generative model that is based on the assumption that documents are represented by a collection of different, latent topics. Each topic will be represented as a probability distribution over all words of a corpus (Krestel et al., 2009). LDA is frequently used in marketing to identify important issues for the adaptation of marketing campaigns or to identify product and service features currently being discussed (Chae and Park, 2018, Gao et al., 2012, Jeong et al., 2019, Ko et al., 2017, Luo et al., 2015, Xu and Xiong, 2020, Yu et al., 2019). However, the LDA as proposed by Blei et al. (2003) cannot represent all use cases (e.g. mapping hierarchies).
Thus, there are adaptations and extensions that are based on the probabilistic model of Blei et al. (2003). Generally, it can be differentiated between (I) topic-based extensions, (II) document-based extensions and (III) word-based extensions (cf. Liu et al., 2016). (I) Topic-based extensions derive structures and dependencies within the latent topics of a document (cf. Rathore et al., 2018, Rathore and Ilavarasan, 2017, Tuarob and Tucker, 2015). (II) Document-based extensions have the ability to incorporate an additional parameter into the model building (cf. Cheng et al., 2020, Lozano et al., 2017, Zhang et al., 2016). (III) The word-based extensions compute n-grams instead of Bag of Words (BoW) to incorporate the order of words in a document within the model generation procedure (cf. Wallach, 2006).
As can be concluded from previous research, the LDA approach is the predominantly applied technique (cf. Eickhoff and Neuss, 2017). Vakansky and Kumar (2020) provide a good overview of existing topic modelling techniques and accordingly develop a decision tree model supporting the selection of a

technique. These authors also include adaptations and extensions of the basic LDA for recommending an appropriate technique. However, their recommendations are theoretical in nature and not derived from practical applications of the techniques. There are numerous topic modelling techniques in the literature, but the number of papers comparing them is scare. Further, it is not clear if the theoretical overview and the decision tree of Vakansky and Kumar (2020) can withstand empirical investigations. The authors only conditionally contribute to the operational application of topic modelling as recommendations should ideally be deduced from the analysis of real-world social media data.

2.3 Corporate use cases and requirements of marketing

Within the prevailing literature, we identified papers that deal with utilising topic modelling techniques in marketing. These applications of topic modelling described in the identified papers cover a wide range of marketing tasks. In particular, three main use cases could be identified: (1) content extraction is concerned about consolidating insights of topics discussed in written social media data. By investigating brand-related content from social media, (1) content extraction enables marketing representatives to develop an understanding of topics and themes, e.g. sustainability or product feature favourability, that are discussed by customers and parties of interest. Thereby, companies can improve their external presentation by, e.g. emphasising the ecological superiority of the own products in brand communications (Chae and Park, 2018), and guide future product planning initiatives by putting more focus on product features that are appreciated by customers (Cirqueira et al., 2017, Irawan et al., 2020, Ko et al., 2017, Rathore et al., 2018). Providing a current overview of customers' major wishes, ideas, and thoughts is therefore a central requirement for these techniques as the ever-increasing amount of social media data along with the breadth of the user base may hinder marketing departments to focus on the essential aspects (cf. Cirqueira et al., 2017, Ko et al., 2017, Lee et al., 2016, Liu et al., 2017, Rathore et al., 2018, Rathore and Ilavarasan, 2017). Hereby, (1) content extraction is of a retrospective nature and is less concerned with speed. Because of the abundance of information provided by social media and the necessity to cover as much useful information as possible, the main aim is to cover the relevant and most frequent topics embedded in social media texts (cf. Gao et al., 2012, Ibrahim and Wang, 2019, Irawan et al., 2020, Wang et al., 2016, Yang et al., 2016). Therefore, we assume training time of the topic modelling techniques as secondary, because there are the two central quality dimensions, relevance and dominance of topics, as the basis for well-funded decisions. Furthermore, techniques for (1) content extraction need to support comparisons with competitors, assessments of a company´s position in the market (Aiello et al., 2013) and effectively support product and service opportunities generation (Ko et al., 2017, Rathore et al., 2018, Rathore and Ilavarasan, 2017).

The second use case we could identify is (2) trend analysis. It has in common with (1) content extraction that it deals with extracting topics from large amounts of written social media data. However, (2) trend analysis focuses on keeping track of emerging trends and their development, while the results of (1) content extraction, are instead point-in-time snapshots of the contents that do not illustrate the dynamic courses of the topics. Marketing representatives applying (2) trend analysis related techniques strive to track how certain topics (e.g., product favourability or customer satisfaction) evolve geographically and temporally (Hong et al., 2012, Jeong et al., 2019, Lozano et al., 2017, Zhang et al., 2017). Thereby, marketing departments can enhance the effectiveness of brand message placement and the allocation of appropriate resources to marketing campaigns depending on geographical and temporal developments. Furthermore, linking topics and contents to groups of interested parties and customers enables companies to adapt brand messages to meet the respective target groups´ expectations and attitudes (Zhang et al., 2016). Therefore, topic modelling techniques for (2) trend analysis need to flexibly incorporate different parameters like authors (Zhang et al., 2016), locations (Cheng et al., 2020, Hong et al., 2012, Lozano et al., 2017, Wang et al., 2007), or time (Cheng et al., 2020, Wang et al., 2012, Zhang et al., 2017) into the model building procedure. Ideally, it should even be possible to include implicit parameters such as places or times mentioned in the texts (Lozano et al., 2017). Since trends describe current issues that influence customers' decision-making, companies want to respond to the resulting customer demands to adapt e.g. marketing campaigns (cf. Luo et al., 2015, Rathore et al., 2018, Zhang et al., 2015, Zhong and Schweidel, 2020). Compared to (1) content extraction, especially with fast sequenced social

media data and topics of interests changing quickly, it is necessary to keep track of trends and topic transitions (cf. Wang et al., 2012, Zhang et al., 2017). The requirement to be able to react as quickly as possible is further intensified by reduced product lifecycles and globalised business environments that have made customer needs more dynamic (Jeong et al., 2019). For this reason, the applied topic modelling techniques must provide short training times (cf. Cheng et al., 2020, Wang et al., 2012, Zhang et al., 2017) and as well support quick comprehension of the extracted topics (cf. Jeong et al., 2019, Lozano et al., 2017) in order to keep up with the speed of the trends.
Compared to (3) content structuring, (2) trend analysis captures the dynamic courses of topics and do not collect hierarchical relationships and correlations. Topic modelling techniques for (3) content structuring enable deeper insights from textual social media by extracting not only the topics (cf. (1) content extraction) but also relationships and connections between them. In this way, (3) content structuring supports decisions that need to connect different aspects with each other (e.g. identifying the influence of different product features on customers´ favourability) (Rathore et al., 2018). Mining the inter-relatedness of individual topics can help to detect subtopics (Anoop et al., 2015, Nolasco and Oliveira, 2019, Park et al., 2015, Rathore and Ilavarasan, 2017, Shahbaznezhad, 2016, Srijith et al., 2017) that can be investigated more closely as driving or inhibiting factors. Hereby, it is also possible to identify niche topics at a finer level of granularity of the topical structure. In line with that, recognising the properties of subevents can enrich the understanding of the main event and to create a powerful knowledge about the scenario (Nolasco and Oliveira, 2019, Srijith et al., 2017). In general, corresponding techniques need to extract topics and at the same time establish connections between them. In line with that, the relationships identified between the topics should be understandable and applicable.
In the next step, the identified requirements (cf. tab. 1) are transformed into corresponding metrics that are necessary for evaluating and comparing the topic modelling techniques in section 5.3.

	Requirements	Sources	Metrics
(1) Content Extraction	(a) Cover all relevant and the most frequent topics embedded in textual social media data	Gao et al. 2012, Ibrahim and Wang, 2019, Irawan et al. 2020, Lee et al., 2016, Liu et al., 2017, Wang et al. 2016, Yang et al. 2016	log-likelihood
	(b) Provide a current overview of events and insights about customers' wishes and complaints	Aiello et al., 2013, Chae and Park, 2018, Cirqueira et al., 2017, Ibrahim and Wang, 2019, Ko et al., 2017, Lee et al., 2016, Liu et al., 2017, Rathore et al., 2018, Rathore and Ilavarasan, 2017	
	(c) Support comparisons with competitors and assessments of one's position in the market	Aiello et al., 2013, Ko et al., 2017, Rathore et al., 2018	
(2) Trend Analysis	(d) Contextualise the extracted topics with additional parameters	Cheng et al., 2020, Hong et al., 2012, Lozano et al., 2017, Luo et al., 2015, Wang et al., 2007, Wang et al., 2012, Zhang et al., 2016, Zhang et al., 2017, Zhong and Schweidel, 2020	build time, coherence, word intrusion
	(e) Support a flexible inclusion of different parameters (e.g. authors, locations or time)	Cheng et al., 2020, Lozano et al., 2017	
	(f) Support quick information provision	Cheng et al., 2020, Wang et al., 2012, Zhang et al., 2017	
	(g) Support quick comprehension of contents	Jeong et al., 2019, Lozano et al., 2017	
	(h) Support continuous tracking of trends and developments	Wang et al., 2012, Zhang et al., 2016, Zhang et al., 2017, Zhong and Schweidel, 2020	
(3) Content Structuring	(i) Identify niche topics at a finer level of granularity of the topical structure	Anoop et al., 2015, Nolasco and Oliveira, 2019, Park et al., 2015, Rathore et al., 2018, Rathore and Ilavarasan, 2017, Shahbaznezhad, 2016, Srijith et al., 2017	topic intrusion
	(j) Identify meaningful relationships and the inter-relatedness of topics		
	(k) Cover all aspects that are semantically related to the extracted topics		

Table 1. Identified corporate use cases and their requirements

When evaluating a model with respect to (1) content extraction, the ability of the respective technique to (a) cover all relevant and the most frequent topics embedded in textual social media data is required. An excessive number of topics leads to the generation of not only relevant but also irrelevant topics. If the number of topics is too small, however, the given overview of topics will lack relevant content (cf. Liu et al., 2017, Yang et al., 2016). Therefore, the researcher must ensure that all relevant topics within an underlying dataset are considered within the analysis and thus integrated within the resulting extraction, which (b) provides a comprehensive and current overview of customers' wishes and complaints. This results in a comprehensive, decision driven base of information which supports (c) comparisons with competitors and assessments of one's position in the market. In order to evaluate this descriptive ability, the metric of log-likelihood is used. Using this, it is possible to quantify how accurately a model can represent the underlying data and thus models all relevant information (cf. Daud et al., 2010, Wallach et al., 2009). Furthermore, the evaluation has to consider different circumstances regarding the number of topics to be identified. Therefore, the evaluation of each technique takes place multiple times, with a continuously increasing number of topics. Thus, the strengths and weaknesses in modelling low (high) numbers of topics and thus lowly (highly) differentiated contextual information can be assessed.
When evaluating a model with regard to the described use case of (2) trend analysis, (d) the ability to contextualise the extracted topics and (e) hereby flexibly include different additional parameters (e.g. authors, location or time) is mentioned. Furthermore, corresponding techniques need to enable time-critical reactions to emerging circumstances so that (f) the provision of the topics should be as quick as possible. Therefore, the analysis of the different techniques is twofold. On the one hand, authors refer to the coherence measure (Dahal et al., 2019, Paul and Girju, 2009, Wang et al., 2007), which describes the property of the respective technique to generate topics that correlate well with the human understanding of semantically coherent topics. This results in a semantically meaningful and sound analysis output that (g) supports time-critical decisions and does not need further investigations to be applicable. On the other hand, to validate the calculated reasonability of the respective analysis output, we incorporate word intrusion (Chang et al., 2009). Thus, multiple subjects evaluate the consistency of the extracted topics by analysing the associated words. The objective is to identify the so-called 'intruder' within the topic, which is represented by a single word without contextual relevancy regarding the intruded topic. If the subjects are able to identify the respective intruder, the evaluated topic is consistent with the human understanding of a meaningful and sound topic. Besides the unhindered applicability of the analysis results, we also classified the build time as a time-critical evaluation metric. So, a short build time results in a timely output supporting faster decision-making. Beyond the aforementioned time-critical requirements, the ability to contextualise the extracted topics and (h) to continuously track their development is also required. To assess the possibility of accounting for further contextual information such as geological or time-based data, we qualify the ability to incorporate external information.
Concerning an evaluation of a model with regard to the described use case of (3) content structuring, the ability of the techniques to reveal the underlying structure within the data is focused. Thus, (i) to be able to extract hidden niche topics to identify relationships between the discussed topics, the ability of the respective technique to identify relationships within the data is qualified. Besides the pure ability to identify relationships, the assessment of the (j) meaningfulness of the identified relations is also required. In this regard, we opted to use the topic intrusion approach, which measures how well a topic model's decomposition of a document as a mixture of topics agrees with human associations of topics related within a document (Chang et al., 2009). Using these techniques, the (k) coverage of all aspects within the extracted topics as well as their interrelationships in the relevant document are analysed. The procedure of the analysis is similar to that of word intrusion. Specifically, the subjects are presented the document title alongside a short extract thereof. In addition to the document information, the subject receives four topics, of which three are the most probable topics assigned to the document and the remaining topic embodies the intruder topic to be identified. If the subjects are able to identify the erroneously listed topic, the topics and their contextual relationship to one another are meaningful and sound. A more detailed insight into the tasks of word and topic intrusion can be found in section 5.2.
As our review shows, different use cases and corresponding requirements have been reported in line with the literature that provide insights into the application of topic modelling techniques in marketing. However, the used data sets and the applied topic modelling techniques vary across the different papers,

so that recommendations made on this basis may not be sufficiently reliable. Therefore, within the research at hand, three different topic modelling techniques are applied to the same data set to give recommendations that are not only theory-driven but also based on the results of a data analysis.

3 Procedure of the Research

Our investigation follows the '*Goal Question Metric*' (GQM) approach outlined by Basili et al. (1994) for a systematic development of metrics for conducting a comparison of topic modelling techniques. The GQM approach is based on the idea that measurements in an organisational context depend on a thoroughly defined goal firstly operationalised by relevant enterprise data, which are then interpreted regarding the goal (Basili et al., 1994). Thus, this approach focuses on which informational needs a company exhibit in order to quantify them and consequently examine if the quantified information meets the goals or not. Especially for our investigation, the GQM approach is well qualified as it assures a systematic research procedure in which reproducible results are achieved. According to Basili et al. (1994) the GQM approach is divided into three different levels:

1. Conceptual level (**GOAL**): We set our goal as the development of means to compare different topic modelling techniques with each other regarding different use cases. Therefore, with this investigation we seek to highlight differences in the organisational application between the three different topic modelling techniques LDA, DMR, and PAM.
2. Operational level (**QUESTION**): In order to characterise how the assessment of our goal is performed, we formulate a question. In our study, the requirements of the different use cases for the techniques identified in the literature form the basis for our questions of the GQM approach: 'To what extent do the three topic modelling techniques (LDA, DMR, PAM) meet the requirements of the different use cases (a-k) (cf. tab. 1)?'
3. Quantitative level (**METRIC**): To answer this question, we conducted a quantitative analysis of the topic modelling techniques, that helps us to evaluate to what extent a selected topic modelling technique can meet the requirements. Subsequently, in order to meet these requirements, that we have already identified in a previous step, we now need proper metrics. Therefore, we consulted the research literature and established specific metrics for the requirements to enable a comprehensible comparison. Consequentially, we will be able to formulate recommendations for the allocation of the topic modelling techniques to the use cases.

4 Selection of the Topic Modelling Techniques

As mentioned above, the field of topic modelling has many different techniques, which all try to identify specific topics within large sets of text data by reducing the dimensionality and attaching different weights to the specific data set (Crain et al., 2012). In order to optimally meet the identified use cases, the selection of the techniques to be used is critical to success. Besides the aforementioned LDA with its extensions, a variety of different categories of topic modelling techniques like Latent Semantic Analysis, Probabilistic Latent Semantic Analysis, Correlated Topic Models, Dynamic Topic Models, or Topic Evolution Model exist (cf. Alghamdi and Alfalqi, 2015). Although they potentially could offer benefits in terms of different applications, most approaches lack a ready-to-use implementation or require an advanced technique-related knowledge and therefore suitable applicability for companies is not given. Differently LDA, where a clear dominance in the use has become apparent, as it offers simple applicability and good analysis results (Eickhoff and Neuss, 2017). Due to the multifaceted challenges to be mastered in the analysis of text data, different extensions of the basic LDA procedure have been developed over time which are suitable for the solution of different scenarios depending on their extending characteristics. Generally, a distinction is made between three expanding properties: (I) extension of topic attributes, (II) extension of document attributes, and (III) extension of word attributes (cf. Liu et al., 2016). In order to answer **RQ2**, the identification of topic modelling approaches for processing certain corporate use cases, we decided to compare LDA as well as selected extensions with respect to the identified use cases. Therefore, to achieve optimal coverage of different techniques, we selected one specific technique for each extension class.

The (I) topic-based extensions deal with the mapping of relations within the latent topics of a corpus. In this context, a variety of techniques is highlighted in literature (cf. Griffiths et al., 2004, Liu et al., 2016). The focus is on the identification of relationships between the inferred topics allowing a hierarchical representation of them. Mimno et al. (2007) compared the ability of their PAM algorithm to represent a hierarchical data structure and to predict a topic distribution for new data not included in the training set with a variety of techniques of the same extension class (cf. Mimno et al., 2007). Since PAM was characterised by better evaluation results, we decided to choose PAM as the topic-related extension. PAM represents the relationships among the topics as directed graphs, which allows the representation of a hierarchical structure within the topics.
The extension based on documents (II) enables the consideration of document-specific meta-information such as authors, document titles, points in time, or geographical information (cf. Liu et al., 2016). In this context, approaches such as the author-topic model (cf. Rosen-Zvi et al., 2012), Topics over Time (cf. Wang and McCallum, 2006), embedded topic model (Dieng et al., 2020) as well as DMR (cf. Mimno and McCallum, 2008) are highlighted in the current research literature. Because the technique presented by Mimno and McCallum (2008) is more flexible with respect to the incorporation of additional information as well as performs remarkable in terms of information quality, DMR is chosen for the implementation of (II) extension of document attributes. DMR is an upstream topic model with a particularly attractive technique for integrating any document features. Instead of defining specific random variables in the graphical model for each new document feature, DMR treats the document annotations as features in a log-linear model. The log-linear model parameterises the Dirichlet before the document's topic distribution, making the Dirichlet's hyperparameter document-specific. Since no assumptions are made about the model structure of new random variables, DMR is flexible to include various types of features, resulting in a flexible use of DMR (Benton and Dredze, 2018).
The above-mentioned topic modelling techniques, which are all based on the BoW approach, do not consider the order of words within a document. This resulted in extension (III), attempting to eliminate the interchangeability of words. Therefore, Wallach (2006) argued that the consideration of word orders in the form of bi-grams can lead to improved results when using a topic modelling approach. Since the consideration of word orders in the form of bi-grams did not show any difference with regard to the generated topics and the underlying topic quality compared to LDA, it will be equated with the use of the basic technique in the following. By choosing these techniques, a selection was made which considers each extension class of the basic approach, whereby a broad spectrum of different techniques is compared with regard to their applicability against diverse use cases.

5 Data Analysis

5.1 Data collection

To identify the potential of the different techniques with regard to the applicability to different use cases, an existing data set of Facebook posts was used (cf. Martinchek, 2017). This comprises 4,155,992 documents from the 15 most popular news services in the United States of America for the period from 2012 to 2016. The raw text dataset contains information such as the respective picture URL or the like count, which are not relevant for the application of the implemented topic modelling approaches. Therefore, to reduce the dimensions of the data, a custom converter was developed and applied to the data. The resulting data set contains three parameters after conversion: the ID of the respective document, the respective year – which serves as feature input to determine the topic relevance at different points in time within the analysis via DMR – and the description – which reflects the actual text of the contribution. The following excerpt from the training data set gives an insight into the data (cf. tab. 2).

ID	Year	Description
52921	2016	Dow Drops More Than 300 Points Following Market Rout...
21049	2014	How to Greet People During Flu Season: Handshake, ...

Table 2. Structure of training data

5.2 Data cleansing and analysing

In order to compare the topic modelling techniques empirically on the basis of their analysis results, the data must first be prepared. With respect to this, we applied tokenisation, stopwords removal and case folding as proposed by Boyd-Graber et al. (2014). As the use of stemming procedures does not improve the interpretability of the results, but can potentially even deteriorate the topic stability (Schofield and Mimno, 2016), we did not incorporate stemming.

By conducting this comparison between LDA, DMR, and PAM based on the mentioned evaluation measures, we aim to reveal the strengths and weaknesses of the different techniques in terms of identifying embedded topics within written social media data. As it is necessary to provide similar conditions for a comparison to be valid, all techniques were configured with their default parameters and trained with iteratively increasing numbers of topics. Accordingly, the selected topic range includes 10, 30, 50, 100, and 300 topics (*k*). Furthermore, all evaluation metrics were validated by cross-validation to eliminate the choice of a potentially non-representative test dataset (Bramer, 2007). The evaluation approach is further distinguishing between intrinsic and extrinsic measures. Intrinsic evaluations measure the performance of a component on its defined subtask, usually against a defined standard in a reproducible laboratory setting. Extrinsic evaluations focus on the component's contribution to the performance of a complete application, which often involves the participation of a human in the loop (Resnik et al., 2006).

#	log-likelihood measurements ($\cdot 10^6$)			coherence measurements			build time (min)		
k	LDA	DMR	PAM	LDA	DMR	PAM	LDA	DMR	PAM
10	- 4.86	- 5.03	- 4.85	- 197.96	- 198.50	- 208.34	342	363	832
30	- 4.49	- 4.61	- 4.41	- 232.84	- 220.75	- 236.81	571	634	1,264
50	- 4.29	- 4.40	- 4.19	- 244.29	- 217.90	- 251.17	912	1,083	1,992
100	- 4.06	- 4.10	- 3.88	- 241.74	- 213.03	- 254.92	1,407	1,732	3,481
300	- 3.91	- 3.62	- 3.67	- 215.02	- 186.95	- 241.52	3,180	3,821	6,984

Table 3. Intrinsic evaluation measurements

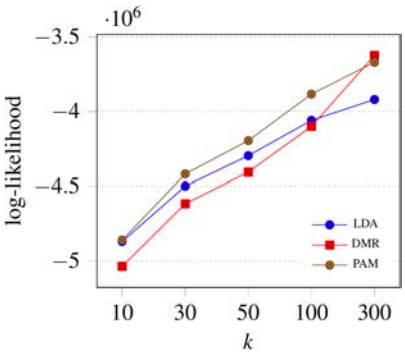

Figure 1. log-likelihood

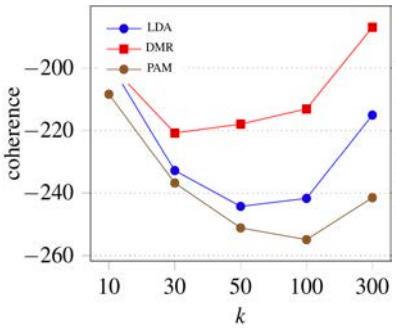

Figure 2. coherence

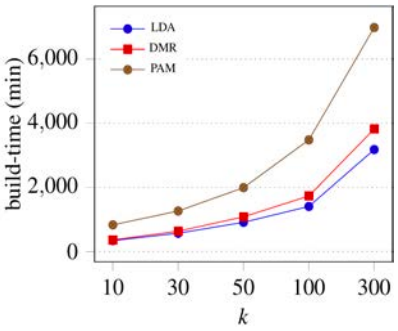

Figure 3. build time

In the first step of the evaluation, the models are compared according to their ability to represent the underlying data. Therefore, the metric of log-likelihood was applied (cf. tab. 3, fig. 1). A higher value represents a better model with regard to the ability to adequately represent the underlying data. The best model here is DMR, with k = 300 (-3.62). LDA performs better than DMR (e.g. k = 30 ≙ -4.86 > -5.03) and worse than PAM (e.g. k = 30 ≙ -4.49 < -4.41) until the number of topics exceeds 100. PAM and LDA have in common that, as the number of topics increases, the log-likelihood improves iteratively (cf. fig. 1). The increase in the number of topics for DMR results in a continuous improvement of the log-likelihood as well. Nevertheless, DMR always outperforms PAM and LDA when k >~ 300. This result implies that DMR is the technique that can be recommended when tasks require many different topics to be identified. However, within tasks that require a lower number of topics to be generated, our results indicate that PAM and LDA may be superior to DMR.

Besides the general ability of the models to adequately represent the underlying data, the semantic quality of the generated topics was evaluated. For that reason, the coherence measure was applied (cf. fig.

2). Therefore, we conclude that DMR generates topics, that are more coherent in general. LDA has a slightly better coherence for k = 10. However, any modelling of k > 10 is outperformed by DMR. Further, PAM performed worst in terms of coherence. Here, all different circumstances of topics are dominated by the other techniques. Nevertheless, PAM generated its most relatable topics for k = 10, which leads to the conclusion, that PAM, similar to LDA, exhibits its strength in modelling low topic dimensions.

To react in a time-critical manner to changing circumstances, it is also necessary to acquire supporting information as quickly as possible. Therefore, to measure the time an approach needs to extract the required information (cf. fig. 3), the build time of each technique is tracked. The respective build time includes data preprocessing, the actual model building and cross-validation. A lower build time indicates a quicker model training, which results in faster information provision. In this regard, LDA outperforms the other two techniques in terms of information extraction time. The difference between LDA and DMR in the lower range of topics is negligible (21 min), but the higher the value of topics, the larger the difference. The difference in the context of k = 300 amounts to 641 minutes. Because of the hierarchies to be modelled by PAM, it generally takes much longer to extract the respective information in comparison to LDA and DMR. Here, PAM needs at least twice as much time under almost all circumstances.

Besides the intrinsic measurements of log-likelihood, coherence and build time, the topic modelling techniques were further assessed by humans to evaluate the semantic quality of the generated topics as well as their interrelationships. Therefore, word and topic intrusion procedures were performed. The respective survey was undertaken by two researchers and administered to 18 participants, all of whom evaluated the semantic coherence of three randomly selected topics (word intrusion) and the decomposition of a single document into its topics and the corresponding relationships (topic intrusion). Each survey ranged between 38 and 51 minutes. The topics as well as the respective excerpts of a document were extracted randomly for each trained model. Further, to account the inter-rater reliability of the results, all participants evaluated the same set of topics or documents respectively for each trained model. Regarding the word intrusion task, the subjects had to identify the intruder within the topics that did not cohere to the semantics of the other presented words. The corresponding results for word intrusion are calculated as the sum of the correctly classified intruders by the test subjects in relation to the total number of tests per model. By analysing three topics per trained model, a total of 810 individual observations were carried out. Regarding the topic intrusion task, the participants had to identify the intruding topic by reading a document title alongside a short extract thereof. The respective document was randomly extracted for each trained model. All participants evaluated the decomposition of the same documents. By this, a total of 270 individual observations are accomplished. The respective results of the topic intrusion task are calculated as the amount of correctly classified intruder topics in relation to the total number of observations per model.

	Word intrusion (%)			Topic intrusion (%)		
k	LDA	DMR	PAM	LDA	DMR	PAM
10	74.0	85.1	53.7	83.3	72.2	83.3
30	72.2	79.6	57.4	66.6	61.1	72.2
50	62.9	75.9	51.8	72.2	72.2	72.2
100	64.8	70.3	48.1	55.5	66.6	55.5
300	55.5	77.7	46.2	55.5	72.2	44.4

Table 4. Extrinsic evaluation measurements

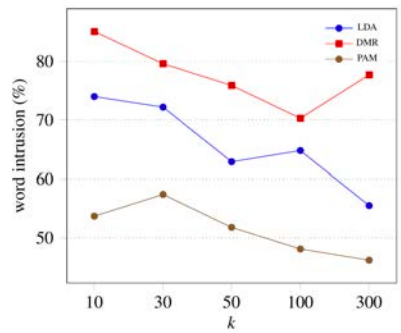

Figure 4. Word intrusion

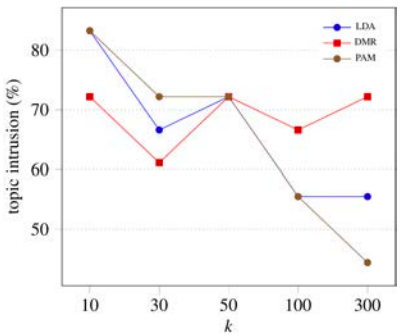

Figure 5. Topic intrusion

Regarding the word intrusion task, it becomes apparent that the previously determined values of coherence are in line with human understanding so DMR generates the most semantically coherent topics. Specifically, the best topics are generated by DMR with k = 10 (cf. fig. 4). The achieved model precision is 85.1% for k = 300. The evaluated minimum precision of DMR is 70.3%. As the results for word intrusion do not show any distinct extremes for different values of k, every model of DMR seems to produce a coherent word-topic distribution. Besides that, LDA also proves to generate the most coherent topics for k = 10 (74.0%). The higher the number of topics to be generated, the more difficult it becomes for the subjects to determine the intruder. This leads to the assumption, that the topics generated by LDA will be increasingly difficult to interpret with a rising number of k, thereby losing their semantic coherence and therefore their meaningfulness. The minimal precision of LDA is recorded for k = 300. Here, only 30 of 54 measurements are classified correctly, resulting in a model precision of 55.5%. There is a clear discrepancy in the quality of topics generated by PAM. The semantically most coherent topics are determined by the model with k = 30. The model precision achieved is 57.4%. The minimum semantic coherence of the topics is with k = 300, with a precision of 46.2%. Thus, it can be concluded that PAM, similar to LDA, has its strength in modelling a smaller number of topics but is clearly inferior to DMR for larger numbers. Besides the evaluation of the topic quality, their decomposition and interrelationships were evaluated. Therefore, topic intrusion tasks were conducted, where the best results were obtained with 83.3% for LDA and PAM with k = 10. Further, it is remarkable, that all techniques achieve the same score with k = 50 (cf. fig. 5), which leads to the conclusion, that all techniques generate consistent topic decompositions for a medium number of topics. The higher the number of topics to be generated, the worse LDA and PAM perform. The worst result is achieved by PAM with k = 300. That decrease of PAM for larger numbers of topics to be generated can be traced back to the formation of the many hierarchical levels, since a high number of k also means that a correspondingly large number of relationships between the different topics must be inferred.

6 Discussion

By evaluating the different techniques, the strengths and weaknesses (cf. tab. 5) of them were identified, that provide information about the applicability of the techniques related to the identified corporate use cases (1) content extraction, (2) trend analysis as well as (3) content structuring settled in marketing.

	Requirements	LDA	DMR	PAM
(1) Content Extraction	(a) Cover all relevant and the most frequent topics embedded in textual social media data	●	●	◗
	(b) Provide a current overview of events and insights about customers' wishes and complaints	●	●	◗
	(c) Support comparisons with competitors and assessments of one's position in the market	●	●	◗
(2) Trend Analysis	(d) Contextualise the extracted topics with additional parameters	○	●	○
	(e) Support a flexible inclusion of different parameters (e.g. authors, locations or time)	○	●	○
	(f) Support quick information provision	●	●	○
	(g) Support quick comprehension of contents	●	●	○
	(h) Support continuous tracking of trends and developments	◗	●	○
(3) Content Structuring	(i) Identify niche topics at a finer level of granularity of the topical structure	○	◗	●
	(j) Identify meaningful relationships and the inter-relatedness of topics	○	○	●
	(k) Cover all aspects that are semantically related to the extracted topics	○	◗	●

Table 5. Results of the comparison ●: applies fully ◗: applies partly ○: applies not

For the first use case – (1) content extraction – the models´ ability to provide an adequate overview of customers' major wishes, ideas, and thoughts to cover all relevant topics embedded in a collection of written social media data was evaluated. Therefore, we applied the metric of log-likelihood, which describes the ability of a model to represent the underlying data as appropriately as possible. This guarantees that the generated output contains all necessary information and that no relevant topics are missing. The standard procedure of LDA as well as that of PAM present their strength compared to DMR in the extraction of topics in the lower range of generated topics. DMR, in contrast, has its strength in the representation of large numbers of topics. This leads to the conclusion that the selection of the best fitting technique depends on the needs to be fulfilled within the extraction task:

- If the content must be very specific, e.g. to support a comparison with competitors and assessments of one's position in the market, a high number of topics is required. So, DMR should be considered.
- If the task requires an extraction on a more abstract level, a low number of topics will mostly be satisfying. If so, the usage of LDA and PAM could be considered. Due to the higher semantic coherence within the topics generated by LDA (cf. tab. 3, tab. 4, fig. 2, fig. 4), the use of LDA is recommended with respect to (1) content extraction for small dimensions.
- The use of PAM is considered as partly applicable, but not recommended, as it is outperformed by LDA and DMR for the criteria being evaluated.

Regarding the second use case (2) trend analysis, the ability of the techniques to generate immediately meaningful and sound output was analysed, which results e.g., in a quick decision supporting information base. This is indispensable regarding the need of time-critical actions within the volatile characteristic of trends. Hereby, DMR shows an advantage leading to the conclusion, that it generates the most reasonable topics (cf. tab. 3, fig. 2). To validate the collected intrinsic evaluation results, the techniques were further evaluated by humans regarding their semantic quality and soundness in an extrinsic way. Here, DMR could be confirmed to generate the most comprehensible output (cf. tab. 4, fig. 4). Further, LDA and PAM show their strength for modelling low numbers of topics as both techniques achieve their best results in the range of 10 (LDA) and 30 (PAM) topics. In addition to direct applicability, the time required for a technique to provide the necessary information was also accounted for. In this respect, LDA provides the fastest output, followed by DMR. The measured discrepancy between these two techniques is negligible for a small number of topics ($k = 10 \triangleq 21$ min.), but the larger the number of topics, the larger the gap ($k = 300 \triangleq 641$ min.). Since the focus of trend analysis is on identifying individual trends and tracking their development, the number of topics will not be that high. Therefore, LDA and DMR are considered capable of reacting to rapidly changing circumstances. PAM, however, requires at least twice as much time for each condition and is therefore not suitable. The strengths and weaknesses of the techniques in the context of (2) trend analysis are represented as follows:

- DMR provides the most reasonable and meaningful topics and is further able to provide them quickly. Additionally, DMR has the advantage of taking external parameters into account. This allows, e.g., tracking the development of topics over a certain period of time or based on geolocation data. Therefore, DMR should be considered regarding (2) trend analysis.
- If the contextualisation of topics does not apply, LDA can also be used for tracking trends and their development, as it can be used to quickly identify meaningful and sound analysis results. Therefore, LDA can be seen as partly applicable regarding trend analysis.
- PAM is not suitable due to the amount of time required and the lack of ability to contextualise topics.

Regarding the last identified use case – (3) content structuring - the ability of the techniques to reveal meaningful relationships and the inter-relatedness of topics was evaluated. Therefore, the extrinsic evaluation metric of topic intrusion was applied. By doing so, it could be guaranteed that all semantically related aspects were extracted. Here, PAM shows a slightly better result than the two remaining techniques for a small number of topics to be generated (cf. tab. 4, fig. 5). Further, DMR underlined its strength in the representation of a high number of topics. A large advantage of using PAM is represented by its ability to model hierarchical structures within the topics themselves. Thus, it is possible to extract general topics as well as their respective subtopics, whereby the topics can be divided into different,

thematically consistent groupings that can support the identification of niche topics at a finer level of granularity. The elicited strengths and weaknesses of the techniques are listed in the following:

- As the best evaluation results regarding the topic intrusion task were achieved by PAM, we recommend this technique as an approach to support (3) content structuring.
- Besides that, DMR can also be applied if the number of topics to be generated reaches a comparatively high number. Thus, the applicability is considered partial.
- Since LDA is surpassed by the two techniques here, the application is not considered suitable.

In summary, all the investigated techniques have different strengths and weaknesses in their applicability to the identified use cases. Nevertheless, LDA showed its strength in modelling low-dimensional topics. In comparison, DMR showed to be superior in representing high-dimensional topics. Regarding trend analysis, DMR showed its strength within the generation of semantically meaningful results. For content structuring tasks, PAM showed superior results in extracting meaningful relationships compared to LDA and DMR.

7 Conclusion and Outlook

Analysing written social media data with automated techniques has massively gained in importance as being aware of customers' wishes is no longer sustainable with manual analysis due to the sheer volume of available posts. Topic modelling has shown to be an adequate instrument to support these tasks by extracting the topics discussed within documents (e.g. Eickhoff and Neuss, 2017, Vayansky and Kumar, 2020). However, it can be observed that especially LDA has been given particular attention for marketing-specific applications. Furthermore, since LDA cannot cover all fields of the identified use cases, marketing tasks may only be supported to a limited extent by automated approaches.

Within this work, the use cases (1) - (3) were identified from the literature due to the frequency that the identified papers related to them. For each of these use cases, corresponding requirements were identified and assigned to different metrics (log-likelihood, coherence, build time, word and topic intrusion) for evaluating these topic modelling techniques (cf. **RQ1**). Thus, LDA, DMR and PAM were applied to a real-world data set, evaluated and compared with each other. Thereafter, this work gives recommendations regarding which topic modelling technique could be applied for which use case (cf. **RQ2**).

Through our comparison of topic modelling techniques, practitioners are given means to select a technique that can best support their daily business activities. Decision makers in marketing can classify their concrete task into one of the three identified use cases and derive a recommendation for a suitable technique. Tasks in marketing, which can be enriched by topic modelling, can thus be supported more optimally and thus the performance of this division, which is so important for companies, can be increased. Beyond creating value for practitioners, theoretical contributions in the research area of IS are also provided. First, based on the use cases we derived several requirements for topic modelling techniques and assigned several evaluation criteria to each of them. Second, in order to provide recommendations, we compared LDA, DMR, and PAM with each other regarding five different evaluation metrics by analysing a real-world data set. This comparison is based on the GQM approach and assures therefore a systematic research procedure in which reproducible results are achieved. Based on this comparison potentials for further enhancements (e.g. considering a faster build time for PAM when $k > 50$) could be evolved. Third, we further deduced strengths (such as DMR should be considered regarding trend analysis, cf. section 6) and weaknesses (such as PAM is not suitable for trend analysis due to the amount of time required and the lack of ability to contextualise topics) of the three topic modelling techniques which is valuable for both researchers and practitioners.

There are some limitations to this study: first, the number of papers we incorporated in identifying use cases and related requirements for topic modelling within marketing is limited. Nevertheless, these requirements enabled the central metrics of topic modelling techniques to be assigned and generally valid recommendations for appropriate procedures to be derived. Second, the number of topic modelling techniques being compared is limited to three. Although the extensions proposed by Liu et al. (2016) could thus be covered to a large extent, we plan to include further techniques in future analysis for each extension and thereby further refine our recommendations.

References

Aiello, L. M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I. and Jaimes, A. (2013). Sensing trending topics in Twitter. *IEEE Transactions on Multimedia,* 15 (6), 1268-1282.

AlFalahi, K., Atif, Y. and Abraham, A. (2014). Models of influence in online social networks. *International journal of intelligent systems,* 29 (2), 161-183.

Alghamdi, R. and Alfalqi, K. (2015). A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6 (1).

Anoop, V., Asharaf, S. and Alessandro, Z. (2015). Generating and visualizing topic hierarchies from microblogs: An iterative latent dirichlet allocation approach. *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 824-828.

Basili, G., Caldiera, V. and Rombach, H. D. (1994). The goal question metric approach. *Encyclopedia of software engineerin*g, 528-532.

Benton, A. and Dredze, M. (2018). Deep Dirichlet multinomial regression. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 365-374.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM,* 55 (4), 77-84.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 993-1022.

Boyd-Graber, J., Mimno, D. and Newman, D. (2014). Care and feeding of topic models: Problems, diagnostics, and improvements. *Handbook of mixed membership models and their applications*, 225255.

Bramer, M. (2007). Principles of data mining. *Springer*, ISBN: 978-1-84628-765-7.

Chae, B. K. and Park, E. O. (2018). Corporate social responsibility (CSR): A survey of topics and trends using Twitter data and topic modeling. *Sustainability,* 10 (7), 2231.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L. and Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 288-296.

Cheng, L., Li, J., Candan, K. S. and Liu, H. (2020). Tracking Disaster Footprints with Social Streaming Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 370-377.

Chinnov, A., Kerschke, P., Meske, C., Stieglitz, S. and Trautmann, H. (2015). An Overview of Topic Discovery in Twitter Communication through Social Media Analytics. *Americas Conference On Information Systems (AMCIS) 2015.*

Cirqueira, D., Pinheiro, M., Braga, T., Jacob Jr, A., Reinhold, O., Alt, R. and Santana, Á. (2017). Improving relationship management in universities with sentiment analysis and topic modeling of social media channels: learnings from ufpa. *Proceedings of the International Conference on Web Intelligence*, 998-1005.

Coelho, L. P., Peng, T. and Murphy, R. F. (2010). Quantifying the distribution of probes between subcellular locations using unsupervised pattern unmixing. *Bioinformatics,* 26 (12), 7-12.

Crain, S. P., Zhou, K., Yang, S.-H. and Zha, H. (2012). Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. *Mining text data*, Springer, 129-161.

Culnan, M. J., McHugh, P. J. and Zubillaga, J. I. (2010). How large US companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive,* 9 (4).

Dahal, B., Kumar, S. A. and Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Minin*g, 9 (1), 1-20.

Dai, Y., Kakkonen, T. and Sutinen, E. (2011). MinEDec: a decision-support model that combines text-mining technologies with two competitive intelligence analysis methods. *International Journal of Computer Information Systems and Industrial Management Applications,* 3 (3), 165-173.

Daud, A., Li, J., Zhou, L. and Muhammad, F. (2010). Knowledge discovery through directed probabilistic topic models: a survey. *Frontiers of computer science in China,* 4 (2), 280-301.

Debortoli, S., Müller, O., Junglas, I. and vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems,* 39 (1), 7.

Di Gangi, P. M., Wasko, M. M. and Hooker, R. E. (2010). GETTING CUSTOMERS' IDEAS TO WORK FOR YOU: LEARNING FROM DELL HOW TO SUCCEED WITH ONLINE USER INNOVATION COMMUNITIES. *MIS Quarterly Executive,* 9 (4).

Dieng, A. B., Ruiz, F. J. and Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8, 439-453.

Eickhoff, M. and Neuss, N. (2017). Topic modelling methodology: its use in information systems and other managerial disciplines. *European Conference On Information Systems (ECIS) 2017.*

Einwiller, S. A. and Steilen, S. (2015). Handling complaints on social network sites–An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies. *Public Relations Review,* 41 (2), 195-204.

Gao, W., Li, P. and Darwish, K. (2012). Joint topic modeling for event summarization across news and social media streams. *Proceedings of the 21st ACM international conference on Information and knowledge management*, 1173-1182.

Ghosh, D. and Guha, R. (2013). What are we 'tweeting'about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and geographic information science,* 40 (2), 90-102.

Girdhar, Y., Giguere, P. and Dudek, G. (2013). Autonomous adaptive underwater exploration using online topic modeling. *Experimental Robotics*, 789-802.

Grégoire, Y., Salle, A. and Tripp, T. M. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons,* 58 (2), 173-182.

Griffiths, T. L., Jordan, M. I., Tenenbaum, J. B. and Blei, D. M. (2004). Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems*, 17-24.

Guesalaga, R. (2016). The use of social media in sales: Individual and organizational antecedents, and the role of customer engagement in social media. *Industrial Marketing Management,* 54, 71-79.

Hanna, R., Rohm, A. and Crittenden, V. L. (2011). We're all connected: The power of the social media ecosystem. *Business Horizons,* 54 (3), 265-273.

Hienerth, C., Keinz, P. and Lettl, C. (2011). Exploring the nature and implementation process of user-centric business models. *Long Range Planning,* 44 (5-6), 344-374.

Hong, L., Ahmed, A., Gurumurthy, S., Smola, A. J. and Tsioutsiouliklis, K. (2012). Discovering geographical topics in the twitter stream. *Proceedings of the 21st international conference on World Wide Web*, 769-778.

Hong, L. and Davison, B. D. (2010). Empirical study of topic modeling in twitter. *Proceedings of the first workshop on social media analytics*, 80-88.

Ibrahim, N. F. and Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from Twitter. *Decision Support Systems,* 121, 37-50.

Irawan, M., Wijayanto, R., Shahab, M., Hidayat, N. and Rukmi, A. (2020). Implementation of social media mining for decision making in product planning based on topic modeling and sentiment analysis. *Journal of Physics: Conference Series*, 012068.

Jeong, B., Yoon, J. and Lee, J.-M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management,* 48, 280-290.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons,* 53 (1), 59-68.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P. and Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons,* 54 (3), 241-251.

Ko, N., Jeong, B., Choi, S. and Yoon, J. (2017). Identifying product opportunities using social media mining: application of topic modeling and chance discovery theory. *IEEE Access,* 6, 1680-1693.

Krestel, R., Fankhauser, P. and Nejdl, W. (2009). Latent dirichlet allocation for tag recommendation. *Proceedings of the third ACM conference on Recommender systems*, 61-68.

Kumar, V. and George, M. (2007). Measuring and maximizing customer equity: a critical analysis. *Journal of the Academy of Marketing Science,* 35 (2), 157-171.

Lee, T.-H., Sung, W.-K. and Kim, H.-W. (2016). A text mining approach to the analysis of information security awareness: Korea, United States, and China. *Pacific Asia Conference On Information Systems (PACIS) 2016*.

Li, W. and McCallum, A. (2006). Pachinko allocation: DAG-structured mixture models of topic correlations. *Proceedings of the 23rd international conference on Machine learning*, 577-584.

Liu, L., Tang, L., Dong, W., Yao, S. and Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus,* 5 (1), 1608.

Liu, X., Burns, A. C. and Hou, Y. (2017). An investigation of brand-related user-generated content on Twitter. *Journal of Advertising,* 46 (2), 236-247.

Lozano, M. G., Schreiber, J. and Brynielsson, J. (2017). Tracking geographical locations using a geo-aware topic model for analyzing social media data. *Decision Support Systems,* 99, 18-29.

Luo, J., Pan, X. and Zhu, X. (2015). Identifying digital traces for business marketing through topic probabilistic model. *Technology Analysis & Strategic Management,* 27 (10), 1176-1192.

Marshall, G. W., Moncrief, W. C., Rudd, J. M. and Lee, N. (2012). Revolution in sales: The impact of social media and related technology on the selling environment. *Journal of Personal Selling & Sales Management,* 32 (3), 349-363.

Martinchek, P. (2016). 2012-2016 Facebook Posts. URL: *https://data.world/martinchek/2012-2016-facebook posts*.

McDonald, M. and Aron, D. (2012). Amplifying the enterprise: the 2012 CIO agenda. *Gartner Executive Programs*, Gartner Inc.

Mimno, D., Li, W. and McCallum, A. (2007). Mixtures of hierarchical topics with pachinko allocation. *Proceedings of the 24th international conference on Machine learning*, 633-640.

Mimno, D. M. and McCallum, A. (2008). Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. *UAI*, 411-418.

Mount, M. and Martinez, M. G. (2014). Social media: A tool for open innovation. *California Management Review,* 56 (4), 124-143.

Nolasco, D. and Oliveira, J. (2019). Subevents detection through topic modeling in social media posts. *Future Generation Computer Systems*, 93, 290-303.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H. and Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology,* 108 (6), 934-952.

Paul, M. and Girju, R. (2009). Cross-cultural analysis of blogs and forums with mixed-collection topic models. *Proceedings of the 2009 conference on empirical methods in natural language processing*, 1408-1417.

Piller, F. T., Vossen, A. and Ihl, C. (2012). From social media to social product development: the impact of social media on co-creation of innovation. *Die Unternehmung,* 65 (1).

Ramage, D., Manning, C. D. and Dumais, S. (2011). Partially labeled topic models for interpretable text mining. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 457-465.

Rathore, A. K., Das, S. and Ilavarasan, P. V. (2018). Social media data inputs in product design: Case of a smartphone. *Global Journal of Flexible Systems Management,* 19 (3), 255-272.

Rathore, A. K. and Ilavarasan, P. V. (2017). Social media analytics for new product development: case of a pizza. *International Conference on Advances in Mechanical, Industrial, Automation and Management Systems (AMIAMS)*, 213-219.

Resnik, P., Niv, M., Nossal, M., Schnitzer, G., Stoner, J., Kapit, A. and Toren, R. (2006). Using intrinsic and extrinsic metrics to evaluate accuracy and facilitation in computer-assisted coding. *Perspectives in Health Information Management Computer Assisted Coding Conference Proceedings*, 2006.

Rosen-Zvi, M., Griffiths, T., Steyvers, M. and Smyth, P. (2012). The author-topic model for authors and documents. *arXiv preprint,* 4169.

Schneider, F., Feldmann, A., Krishnamurthy, B. and Willinger, W. (2009). Understanding online social network usage from a network perspective. *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, 35-48.

Schofield, A. and Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics,* 4, 287-300.

Shahbaznezhad, H. (2016). Fan page management via content generation and feedback strategies. *SSRN*, 2769428.

Sigala, M. (2012a). Exploiting web 2.0 for new service development: Findings and implications from the Greek tourism industry. *International Journal of Tourism Research*, 14 (6), 551-566.

Sigala, M. (2012b). Social networks and customer involvement in new service development (NSD). *International Journal of Contemporary Hospitality Management*.

Srijith, P., Hepple, M., Bontcheva, K. and Preotiuc-Pietro, D. (2017). Sub-story detection in Twitter with hierarchical Dirichlet processes. *Information Processing & Management,* 53 (4), 989-1003.

Statista (2020b). Number of social network users worldwide from 2017 to 2025. URL: *https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/* (visted on 25.10.2020).

Tuarob, S. and Tucker, C. S. (2015). Quantifying product favorability and extracting notable product features using large scale social media data. *Journal of Computing and Information Science in Engineering*, 15 (3).

Vayansky, I. and Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94, 101582.

Wallach, H. M. (2006). Topic modeling: beyond bag-of-words. *Proceedings of the 23rd international conference on Machine learning*, 977-984.

Wallach, H. M., Murray, I., Salakhutdinov, R. and Mimno, D. (2009). Evaluation methods for topic models. *Proceedings of the 26th annual international conference on machine learning*, 1105-1112.

Wang, C., Wang, J., Xie, X. and Ma, W.-Y. (2007). Mining geographic knowledge using location aware topic model. *Proceedings of the 4th ACM workshop on Geographical information retrieval*, 65-70.

Wang, S., Chen, Z., Fei, G., Liu, B. and Emery, S. (2016). Targeted topic modeling for focused analysis. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1235-1244.

Wang, X. and McCallum, A. (2006). Topics over time: a non-Markov continuous-time model of topical trends. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 424-433.

Wang, Y., Agichtein, E. and Benzi, M. (2012). TM-LDA: efficient online modeling of latent topic transitions in social media. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 123-131.

Womack, J. P. and Jones, D. T. (1996). Beyond Toyota: how to root out waste and pursue perfection. *Harvard business review,* 74 (5), 140-172.

Xu, S. and Xiong, Y. (2020). Setting socially mediated engagement parameters: A topic modeling and text analytic approach to examining polarized discourses on Gillette's campaign. *Public Relations Review,* 46 (5), 101959.

Yang, L., Lin, H., Lin, Y. and Liu, S. (2016). Detection and extraction of hot topics on chinese microblogs. *Cognitive Computation,* 8 (4), 577-586.

Yu, D., Xu, D., Wang, D. and Ni, Z. (2019). Hierarchical topic modeling of Twitter data for online analytical processing. *IEEE Access*, 7, 12373-12385.

Zhang, H., Kim, G. and Xing, E. P. (2015). Dynamic topic modeling for monitoring market competition from online text and image data. *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 1425-1434.

Zhang, P., Gu, H., Gartrell, M., Lu, T., Yang, D., Ding, X. and Gu, N. (2016). Group-based Latent Dirichlet Allocation (Group-LDA): Effective audience detection for books in online social media. *Knowledge-Based Systems,* 105, 134-146.

Zhang, X., Zhao, L., Chen, Z., Boedihardjo, A. P., Dai, J. and Lu, C.-T. (2017). Trendi: Tracking stories in news and microblogs via emerging, evolving and fading topics. *2017 IEEE International Conference on Big Data*, 1590-1599.

Zhong, N. and Schweidel, D. A. (2020). Capturing changes in social media content: a multiple latent changepoint topic model. *Marketing Science,* 39 (4), 827-846.

2.3. Beitrag 3: MANTRA: A Topic Modeling-Based Tool to Support Automated Trend Analysis on Unstructured Social Media Data

Adressierte Forschungsfrage	Wie sollte ein auf Topic Modelling basierendes Trendanalysetool aussehen, welches die Anforderungen der marketingbezogenen Anwendungsfälle (a) Produktentwicklung (b) Kundenverhaltensanalyse und (c) Markt- /Markenbeobachtung umsetzt?
Zielstellungen	• Identifikation von Anforderungen eines auf Topic Modelling basierenden Trendanalysetools für die drei Anwendungsgebiete (a)-(c). • Entwicklung und Evaluation eines Softwaretools, das alle drei Anwendungsgebiete und deren Anforderungen durch die Kombination von verschiedenen Social Media Analysetechniken erfüllt.
Forschungsmethode	**Design Science Research** • DSR-Prozess nach Peffers et al. (2007), der basierend auf einer Problem- und Lösungsidentifikation die Schritte Entwicklung, Demonstration und Evaluation für das auf Semi-Supervised Topic Modelling basierende Trendanalysetool umgesetzt hat. • Anlehnung der Forschung an Hevner et al. (2004) durch die Ausgestaltung des Design Cycle (Demonstration und Evaluation des Artefakts), Relevance Cycle (Beitrag zur Praxis) und Rigor Cycle (Beitrag zur Kernel Theorie und (Nascent) Designtheorie (vgl. Gregor und Hevner, 2013)).
Kernergebnisse (Überblick)	• Ableitung von Designanforderungen für die strukturierte Entwicklung des Artefakts, sowohl auf Grundlage der wissenschaftlichen Literatur als auch auf Basis einschlägiger Theorien (Customer Focus Theory (Gulati und Oldroyd, 2005; Tseng und Piller, 2003), Innovation Theory (Axtell et al. 2000; Ye und Kankanhalli, 2018) und Social Media Analytics Theory (Fan und, Gordon 2014; Stieglitz et al., 2018)). • Entwicklung und Validierung von insgesamt sechs Designprinzipien für ein auf Topic Modelling basierendes Trendanalysetool. • Identifikation von Trendthemen durch Anwendung des Trendanalysetools „MANTRA“ auf ca. 1,03 Mio. Yelp Reviews. • Vergleich der beiden Trends „Vegane Cuisine“ und „Global Cuisine“ über den Zeitverlauf inkl. Unterscheidung von Sentiment und Geolokationen (Massachusetts, Texas und Oregon). • Aufzeigen der Bedeutung des Einbeziehens des Sentiments (zur Identifikation der Stimmung gegenüber den Trends) und der flexiblen Einbindung verschiedener externer Parametern (zur Kontextualisierung der Trendthemen) für die Aussagekraft der identifizierten Trends.
Publikationsort	**Veröffentlicht** in den Proceedings der 44. International Conference on Information Systems, Hyderabad, Indien.
Ranking VHB JQ 4	**A**
Autoren und Anteile	Janik Wörner 35 % **Daniel Konadl 35 %** Isabel Schmid 20 % Susanne Leist 10 %

Tabelle 5: Fact Sheet zu Beitrag 3

MANTRA: A Topic Modeling-Based Tool to Support Automated Trend Analysis on Unstructured Social Media Data

Completed Research Paper

Janik Wörner
University of Regensburg
Universitätsstraße 31,
93053 Regensburg, Germany
Janik.Woerner@ur.de

Daniel Konadl
University of Regensburg
Universitätsstraße 31,
93053 Regensburg, Germany
Daniel.Konadl@ur.de

Isabel Schmid
Regensburg, Germany
Isabel.m.schmid@gmail.com

Susanne Leist
University of Regensburg
Universitätsstraße 31,
93053 Regensburg, Germany
Susanne.Leist@ur.de

Abstract

The early identification of new and auspicious ideas leads to competitive advantages for companies. Thereby, topic modeling can serve as an effective analytical approach for the automated investigation of trends from unstructured social media data. However, existing trend analysis tools do not meet the requirements regarding (a) Product Development, (b) Customer Behavior Analysis, and (c) Market-/Brand-Monitoring as reflected within extant literature. Thus, based on the requirements for each of these common marketing-related use cases, we derived design principles following design science research and instantiated the artifact "MANTRA" (MArketiNg TRend Analysis). We demonstrated MANTRA on a real-world data set (~1.03 million Yelp reviews) and hereby could confirm remarkable trends of vegan and global cuisine. In particular, the importance of meeting all specific requirements of the respective use cases and especially flexibly incorporating several external parameters into the trend analysis is exemplified.

Keywords: Social Media Analytics, Trend Analysis, Topic Modeling-Based Tool, Design Science, Marketing

Introduction

In the fast-paced and competitive world of business today, it is crucial for companies to keep track of trends (e.g., Endo et al. 2015; Lozano et al. 2017). Trends are complex, multidimensional phenomena that evolve over time and that can exert a significant impact on society in case they prevail (cf. Burmeister et al. 2002). In this context, social media plays a pivotal role as it allows trendsetters to openly discuss emerging trends across various industries (e.g., technology, fashion, or food trends) (e.g., Jeong et al. 2019; Park et al. 2021) and enables customers to express their unfiltered opinions (Yan et al. 2014). Thus, based thereon, companies can get an impression of what customers discuss as trends and what they think that future products and services should look like. In this respect, there are three focal points in the extant marketing research literature that deal with identifying and implementing trends: Developing a sound comprehension

of feasible development paths of products and services (Puranam et al. 2017), customers' cultural norms and behaviors (Humphreys and Wang 2018), and market structures (Netzer et al. 2012). Our research considers these focal points by three different use cases and aims to support the early exposure of trends.

To identify trending topics from social media, marketing departments can apply different automated analysis techniques (cf. Batrinca and Treleaven 2015; Fan and Gordon 2014; Stieglitz et al. 2018). There are quantitative approaches that identify trends based on the frequency by which certain keywords appear within social media posts over time. However, this is insufficient, as trends, whose constituent elements are not necessarily known in advance, cannot be found in this way. Supervised learning could also be applied to trend analysis, but this requires ex-ante knowledge about trend characteristics and trend impulses. Unsupervised learning techniques can autonomously extract disruptive and previously unknown trends from large text volumes. Trending topics can thereby include combinations of two or more words that are not unusually topical or frequently mentioned together (Khan et al. 2021; Reisenbichler and Reutterer 2019). In addition, identifying meaningful trends can benefit from previously gained domain knowledge (cf. Burmeister et al. 2002). Therefore, semi-supervised topic modeling turns out to be a suitable automated trend analysis technique as it determines latent relationships between the individual components of a trend (Crain et al. 2012; Hong and Davison 2010) and allows its users to provide seed information as semantic anchors to guide the topics to be discovered (Jagarlamudi et al. 2012).

The potential that automated trend analysis techniques propose has already been recognized in practice and science. Nevertheless, available solutions show remarkable drawbacks, as they do not cover important Design Requirements (DRs) that are deemed crucial for a trend analysis concerning the common marketing-related use cases (e.g., incorporation of different geolocation data and customer characteristics as external parameters) that we could derive. Hence, the capabilities for comprehensively assessing trends from different perspectives (i.a., products, customers, markets) are restricted. Companies need to operate different trend analysis approaches to adequately support these common marketing-related use cases, leading, i.a., to the burden of tediously compiling individual results to gain a holistic view of trends. Moreover, the available tools proposed for automated trend analysis of marketing-related use cases do not offer marketing representatives to incorporate prior domain knowledge. As a result, the extracted trends are not necessarily semantically tailored to the company and its fields of activity, which severely impedes the integration of the gained insights into decision-making processes within marketing departments.

We address these drawbacks of existing solutions and propose the design and a concrete instantiation of a software tool as an automated artifact for trend analysis. By including a topic modeling technique that particularly enables the incorporation of previously gained domain knowledge, we show the proficiency of this technique in automated trend analysis. Regarding the DRs that a corresponding automated trend analysis tool needs to meet for common marketing-related use cases, we focus especially on the combination of different machine learning techniques, which leads to the trend analysis being purposeful. Against this background, the following two Research Questions (RQs) are posed:

- ***RQ1:*** *Which DRs should a topic modeling-based trend analysis tool meet that supports the marketing-related use cases Product Development, Customer Behavior Analysis, and Market-/Brand-Monitoring?*
- ***RQ2:*** *How can such a topic modeling-based trend analysis tool be designed and implemented, and which contributions can be derived?*

The remainder of this paper is structured as follows: the next section provides the conceptual foundations and the related work. Subsequently, the research procedure, following the Design Science (DS) approach (Hevner et al. 2004; Peffers et al. 2007), is described. The next section deals with the compilation of the DPs that we have derived from the DRs. Then, we report the technical realization of the tool, demonstrate its application on about 1.03 million Online Customer Reviews (OCRs), and present the resulting outcomes. The paper closes with concluding remarks, potential limitations, and future research ideas.

Conceptual Foundations and Related Work

Conceptual Foundations

Social media entails *"a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation of User Generated Content"* (Kaplan and Haenlein 2010, p.61). Online social platforms (e.g., Yelp) as a form of social media thereby serve as an important

medium for the exchange between companies and customers by unstructured and opinionated content (e.g., textual social media posts). Trendsetters (e.g., domain luminaries or pioneers in certain fields) can share their thoughts on trending topics and discuss what future products and services should look like (e.g., Jeong et al. 2019; Park et al. 2021; Tucker and Kim 2011). In this respect, OCRs constitute *"peer-generated product evaluations posted on company or third party websites"* (Mudambi and Schuff, 2010, p.186), enabling users to evaluate companies' current products and services by disseminating their experiences. Thus, OCRs are also beneficial for companies as they include customers' expectations and opinions concerning products, services, or the company in general (cf. Hicks et al. 2012; Sigala 2012). As OCRs need to be online, social media platforms have shown effective for their dissemination in practice. Therefore, companies can harness OCRs to gain insights about central marketing-related dimensions unfiltered and in real-time (Yan et al. 2014): i.a., opinions about products and services (Ko et al. 2018; Moro et al. 2020), customers' behaviors (Humphreys and Wang 2018), and market structure and competition (Netzer et al. 2012; Tirunillai and Tellis 2014). Keeping track of the evolution of topics and these marketing-related dimensions is central to trend analysis. In this way, marketing departments can enhance the effectiveness of brand message placement and the allocation of appropriate resources to marketing campaigns depending on geographical and temporal developments (Hong et al. 2012; Lozano et al. 2017).

To reveal trending topics, marketing representatives can apply topic modeling, which refers to the use of generative probability models for determining latent relationships within a corpus of text data. Hereby, semi-supervised topic modeling allows the user to guide the topic discovery process by providing seed information as semantic anchors (Jagarlamudi et al. 2012). In this way, different contextual information (e.g., geolocation, time, user characteristics, a company's field of activity) can be incorporated to tailor the trends to be derived. Despite these potentials, semi-supervised topic modeling has yet not comprehensively been harnessed for trend analysis within marketing use cases.

Distinctive Use Cases of Topic Modeling-Based Trend Analysis

To identify common and distinctive marketing-related use cases of topic modeling-based trend analysis, we have comprehensively searched and consolidated the extant research literature. Thereby, we have aligned with the procedure for conducting a systematic literature review as proposed by Vom Brocke et al. (2015). In the **first step**, we applied a keyword-based search on *ACM Digital Library, AIS Electronic Library, EBSCOhost, Emerald Insight, Google Scholar, IEEE Xplore, SAGE Publications, SpringerLink, and the Wiley Online Library (keywords: "topic model", "topic modeling", "trend analysis", "trend tracking", "marketing", "social media", and several combinations thereof)*. Within the keyword-based search, we conducted iterative keyword refinements and could thereby derive the following further keywords: *"online customer review", "product development", "customer behavior", "customer behavior change", "market surveillance", "market monitoring", and "brand monitoring"*. We only kept papers dealing with topic modeling within the domain of marketing for automated trend analysis on unstructured social media data. Based thereon, the three involved researchers independently analyzed the title, abstract, and thematic focus of the initially derived *682 papers*. It turned out that there are papers that mention topic modeling but do not apply it to trend analysis. Some papers used social media and marketing-related terms, but these concepts were not reflected in the application of topic modeling *(remaining: 52)*. In the cases where assessments of the researchers showed ambiguities, the relevance and fit of the respective paper to our investigation were determined by clearing up the ambiguities through discussions and justifications among the involved researchers *(remaining: 38)*. Duplicates were eliminated, only papers written in English and only those that have passed a peer-review assessment were kept *(remaining: 26)*. In the **second step**, we conducted a forward and a backward search, resulting in another *3 relevant papers* that meet the defined inclusion criteria. Therefore, in total *29 papers* could be derived.

Subsequently, we have investigated the contents of the derived papers. Thereby, the wording of the statements in the examined sources often differed considerably, and the logical consolidation of the individual statements offered room for interpretation. Using qualitative content analysis and inductive category development, we could carry out an abstraction from the basic data, while, at the same time, deriving a coherent image of these data (cf. Mayring 2000; Mayring 2002). In this way, in particular, three distinctive marketing-related use cases could be identified for automated topic modeling-based trend analysis: (a) **Product Development** (e.g., Bae et al. 2018; Irawan et al. 2020; Jeong et al. 2019), (b) **Customer Behavior Analysis** (e.g., Bhor et al. 2018; Chen et al. 2012; Zhang et al. 2016) and (c) **Market-/Brand-Monitoring** (e.g., Lozano et al. 2017; Qu et al. 2015; Zhao et al. 2019).

Design Requirements for a Topic Modeling-Based Trend Analysis Tool

After having derived the distinctive marketing-related use cases, we next aimed to derive DRs for the respective use cases. Because different papers pointed to the same requirements but in (slightly) different wordings, we applied the qualitative content analysis (cf. Mayring 2000; Mayring 2002) here as well. In this way, we could systematically condense the available data and consolidate the contents that relate to the respective DRs. In total, we could derive 14 DRs that we will describe in more detail in the following:

By applying topic modeling for (a) **Product Development**, marketing representatives can develop an understanding of how customers perceive their products, services, and corresponding features. Topics identified from OCRs may, e.g., indicate that customers perceive the product packaging as unecological. Harnessing OCRs may bring up the idea of replacing the existing plastic packaging with paper packaging so that customers' expectations are subsequently met (cf. Park et al. 2021). As the prevailing literature shows for the use case (a), customers' product and service expectations can differ across geographical markets. To be successful and competitive in a targeted market, marketing representatives need to know which features to design in which way to meet the global and local customers' expectations. Therefore, topic modeling needs to **(DR1)** incorporate different geolocation data as an external parameter (e.g., Bae et al. 2018; Ding et al. 2020; Ha et al. 2017) that match the generated topics with the corresponding geographical locations. Customers' self-reported opinions of products and services also play an important role. Marketing representatives aim at retaining features evoking positive customer perceptions, while features evoking negative sentiments need to be improved. Thus, a trend analysis tool needs to **(DR2)** incorporate the sentiment of social media posts (e.g., Irawan et al. 2020; Jeong et al. 2019). Thereby, numerous opportunities for improving products and services can be determined, and certain features are given greater importance than others. Yet, it is not feasible to address all identified issues because of limited resources in marketing departments. Thus, it is necessary to support marketing representatives **(DR3)** to assess the product and service feature favorability (e.g., Ding et al. 2020; Moro et al. 2020; Tucker and Kim 2011). Regarding use case (a), e.g., Production Theory and Innovation Theory can serve as theoretical foundations. Innovation Theory proposes harnessing external and internal sources for identifying new and auspicious ideas of what future products may look like. These ideas are assessed and refined so that the most promising ideas are integrated as input factors into the product refinement and development process (cf. Axtell et al. 2000; Ye and Kankanhalli 2018). So, we argue that it is imperative for a trend analysis tool to support the refinement of the identified trends and to enable marketing representatives to **(DR4)** link the anticipated customer preferences with new features (e.g., Jeong et al. 2019; Ko et al. 2018; Tucker and Kim 2011).

In contrast to (a), the focus in (b) **Customer Behavior Analysis** is on the authors of the social media posts. Marketing representatives responsible for customer targeting must know how customer behaviors and associated trends differ for certain customer characteristics like age or gender (e.g., Bhor et al. 2018; Chen et al. 2012). Customer Behavior Analysis can for example reveal that older people perceive comfort as an essential product property. To increase favorability among older customers, a company may communicate its superiority regarding comfort to this respective target group (cf. Duncan and Moriarty 1998). For certain customer characteristics, there are as well different customer behaviors. As such, the needs for online shopping do distinguish between males (e.g., the accuracy of product descriptions) and females (e.g., fair pricing) (cf. Ulbrich et al. 2011). In this vein, Customer Focus Theory can be harnessed as it proposes companies to align their actions to customers' characteristics and expectations (Gulati and Oldroyd 2005; Tseng and Piller 2003). To establish the topic-person connections, topic modeling techniques, therefore, need to **(DR5)** incorporate different user-related information and customer characteristics (e.g., age, gender, group-related attitudes, and values) as external parameters (e.g., Bhor et al. 2018; Shi et al. 2018; Zhang et al. 2016). Based thereon, marketing representatives can tailor marketing messages with proposals that the individual customer groups appreciate most. As (b) is about customer behavior, **(DR6)** a functionality to filter posts is essential to exclude company posts before generating customer-related topics (Iwata et al. 2009; Saha and Sindhwani 2012). Fast reactions **(DR7)** are important regarding topic modeling techniques for this use case because customers' perceptions can change rapidly (Bhor et al. 2018; Iwata et al. 2009; Sasaki et al. 2014) and, while, identified concerns about products or services require a certain time for refactoring, customers need to be rapidly contacted and appeased.

The third use case (c) **Market-/Brand-Monitoring** is about investigating how brands and companies act and which strategies they deploy. Therefore, (c) is concerned with the companies and brands instead of (a), their products and services. Applying trend analysis to the social media posts of competitors could reveal

that competing brands include reporting about their sustainable acting within public communications. To support decisions on the market-/brand level, the topics must be generated for the respective companies or brands. Beyond generating topics according to geographical markets, the topic modeling technique **(DR8)** must also be capable of incorporating brands as external parameters (e.g., Endo et al. 2015; Lozano et al. 2017; Qu et al. 2015). Based thereon, a company may, e.g., adapt its market strategy to focus on niche markets that other brands have yet paid little attention to. Innovation Theory (cf. Sundbo 1995) thereby serves as a foundation as it recommends how to successfully place new innovative products and services within a market. Hereby, the Strategic Innovation Paradigm points out that companies need to assess how competitors shape their strategic communications concerning innovations. Thus, a trend analysis tool **(DR9)** must support the identification of competitors and their market orientations (e.g., Lozano et al. 2017; Park et al. 2021; Zhao et al. 2019). Moreover, marketers need not only support in identifying emerging competitors but also in understanding the behaviors of current ones (e.g., communication patterns and discount strategies). Therefore, **(DR10**) filter criteria like time periods, geolocations, and brands should be proposed to narrow down the dataset (e.g., Sohn et al. 2019; Valdez et al. 2018; Zhao et al. 2019).

Beyond use case-specific DRs, we could identify DRs that apply to all three use cases. One theory posing fundamental implications across all use cases is Social Media Analytics Theory (e.g., Fan and Gordon 2014; Stieglitz et al. 2018). When applying social media analytics for trend analysis on unstructured social media data, a central necessity is **(DR11)** dealing with a huge number of social media posts because of the ever-increasing speed of social media posts emergence (e.g., Jeong et al. 2019; Lozano et al. 2017; Stieglitz et al. 2018). Further, it is essential to provide means to include **(DR12)** historical data so that past developments of trends can be considered when assessing the current state of a trending topic (e.g., Bhor et al. 2018; Lozano et al. 2017). Because trends constitute dynamic developments and not solely static points in time, the topic modeling technique needs to integrate different temporal parameters into the topic generation. Marketers require **(DR13)** support in assessing the course of topics over time at different levels of granularity and in identifying emerging and declining trends (e.g., Ha et al. 2017; Zhong and Schweidel 2020). In general, applying the results of topic modeling requires advanced technique-related knowledge. Thus, **(DR14)** the results need to be visualized (by charts and plots) so that marketing representatives are supported in making sense of the derived topics (Bae et al. 2018; Ha et al. 2017; Yang et al. 2017).

Use Cases	Design Requirements	Sources
(a) Product Development	**DR1.** Incorporation of different geolocation data as external parameters	Bae et al. 2018; Ding et al. 2020; Endo et al. 2015; Ha et al. 2017; Moro et al. 2020; Park et al. 2021
	DR2. Consideration of sentiment	Irawan et al. 2020; Jeong et al. 2019; Ko et al. 2018; Moro et al. 2020
	DR3. Providing means to assess the importance of a product or service feature	Bai et al. 2023; Ding et al. 2020; Jeong et al. 2019; Mastrogiacomo et al. 2021; Moro et al. 2020; Tucker and Kim 2011
	DR4. Linking of newly discovered preferences with new features	Axtell et al. 2000; Bai et al. 2023; Ding et al. 2020; Jeong et al. 2019; Ko et al. 2018; Puranam et al. 2017; Tucker and Kim 2011; Ye and Kankanhalli 2018
(b) Customer Behavior Analysis	**DR5.** Incorporation of different user-related information and customer characteristics as external parameters	Bhor et al. 2018; Chen et al. 2012; Gulati and Oldroyd 2005; Puranam et al. 2017; Qu et al. 2015; Shi et al. 2018; Tseng and Piller 2003; Zhang et al. 2016
	DR6. Provide means to filter out social media posts being posted by the company itself	Gulati and Oldroyd 2005; Iwata et al. 2009; Saha and Sindhwani 2012; Tseng and Piller 2003
	DR7. Support fast reactions to changing customer perceptions	Bhor et al. 2018; Iwata et al. 2009; Sasaki et al. 2014
(c) Market-/ Brand-Monitoring	**DR8.** Incorporation of different brands and geolocation data as external parameters	Endo et al. 2015; Lozano et al. 2017; Qu et al. 2015; Sohn et al. 2019; Zhao et al. 2019
	DR9. Identify differences of how enterprises communicate and place themselves in markets	Park et al. 2021; Sundbo 1995; Tirunillai and Tellis 2014; Valdez et al. 2018; Zhao et al. 2019
	DR10. Apply time periods, geolocations, and brands as filter criteria	Sohn et al. 2019; Sundbo 1995; Valdez et al. 2018; Zhao et al. 2019
Use Case Independent Design Requirements	**DR11.** Deal with large numbers of social media posts	Fan and Gordon 2014; Jeong et al. 2019; Lozano et al. 2017; Puranam et al. 2017; Stieglitz et al. 2018; Yan et al. 2015
	DR12. Capability of analyzing historical data	Bai et al. 2023; Bhor et al. 2018; Ha et al. 2017; Lozano et al. 2017; Puranam et al. 2017; Sohn et al. 2019; Zhong and Schweidel 2020
	DR13. Identification of trends over time (topic evolution) at different levels of granularity (e.g., phases of a day, days of a week, seasons)	Ha et al. 2017; Lozano et al. 2017; Puranam et al. 2017; Tucker and Kim 2011; Zhang et al. 2015; Zhong and Schweidel 2020
	DR14. Visualization of the results	Bae et al. 2018; Ha et al. 2017; Yang et al. 2017

Table 1. Identified Use Cases and Corresponding Design Requirements

Assessment of Tools for Automated Trend Analysis on Social Media

To assess the capabilities of automated trend analysis tools for unstructured social media data, we first identified corresponding tool providers. The Google search engine and social media analytics groups within online social networks (i.a., LinkedIn) were drawn upon for this purpose. Tools identifying trends based on ex-ante knowledge (i.a., approaches that assess trends by the frequency by which certain keywords appear) were disregarded, as this tremendously hinders the extraction of disruptive and previously unknown trends.

By three researchers conducting the identification of tool providers independently, the possibility of overlooking established tool providers could be reduced. The identified tool providers were then consolidated and potential disagreements about the tools were resolved. In this way, we could cover a broad spectrum of providers and take an up-close look at the most established tools for automated trend analysis *(Brandwatch, Hootsuite, IBM, Keyhole, Meltwater, MineMyText, NetBase Quid, Pulsar, Q-Markets, Symanto, Synthesio, Trend-Sonar)*. Based on the DRs that we could derive *(see Table 1)*, a tool survey was designed to assess whether the available trend analysis tools cover the posed requirements. The tool survey was validated by the involved researchers and subsequently applied to the data collection procedure. For this reason, we interviewed sales representatives from these tool providers on the specific features and additionally installed demo versions of the tools. Based on the interview results and by independently testing the functionalities of the trend analysis tools against the posed DRs, we could gain detailed and comprehensive insights into the capabilities as well as into the drawbacks of those trend analysis tools.

Solution Prov.	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
# Employees	~ 10	~ 30	~ 60	~ 70	~ 70	~ 120	~ 300	~ 1.3k	~ 1.4k	~ 2.2k	~2.5k	~ 290k
General Approach	NLP	NLP	AI	AI	NLP	NLP	NLP	AI	AI	NLP	NLP	AI
DR1	-	-	○	-	-	-	-	-	-	-	-	○
DR2	-	●	-	●	-	-	●	●	●	-	●	●
DR3	●	●	○	●	●	●	●	●	●	●	●	●
DR4	-	○	-	-	●	○	○	-	○	-	-	-
DR5	-	-	-	-	-	-	-	-	-	-	-	-
DR6	●	○	●	●	-	●	-	●	●	●	●	●
DR7	*DR7 was **not** included in the comparison because a fast execution is highly dependent on the underlying operating hardware.*											
DR8	-	-	○	-	-	-	-	-	-	-	-	-
DR9	-	-	-	○	○	○	●	-	-	-	-	●
DR10	●	●	●	●	●	●	●	●	●	●	●	●
DR11	●	●	●	●	●	●	●	●	●	●	●	●
DR12	●	○	●	●	●	●	●	●	●	●	●	●
DR13	○	○	○	○	○	○	○	○	○	○	○	○
DR14	●	●	●	●	●	●	●	●	●	●	●	●

Table 2. Results of Assessing Solutions for Automated Trend Analysis
AI: Artificial Intelligence (i.a., neural networks, support vector machines), **NLP:** Natural Language Processing (i.a., unsuperv. topic modeling)

● = applicable
○ = partly applicable

As our assessment could reveal, there is no automated tool that offers marketing representatives to unveil the latent relationships between the individual components of a trend and that enables them at the same time to incorporate their previously gained domain knowledge. The tools lack to meet various specific DRs of the marketing-related use cases (a)-(c), none of the tools thereby manages to cover all of these DRs, and, in particular, they cannot flexibly integrate the various external parameters at different granularity. Nonetheless, the perceptions of products and services differ across continents, countries, and federal states because of globally and locally differing customer expectations. Customer behaviors do differ for certain customer characteristics (e.g., age, gender, group-related attitudes, and values). Without the ability to flexibly integrate different external parameters like time (e.g., days of a week, phases of a day, seasons), geolocations, and user-related information at different granularity, an automated trend analysis tool does not sufficiently support trend analysis and the marketing-related use cases (a)-(c). Beyond the revealed drawbacks of the available tools, extant literature clearly points out the need of such a comprehensive topic modeling-based tool that supports all three use cases. With this research, we aim to close this gap.

Research Procedure

To accomplish the development of a tool for automated trend analysis regarding the common marketing-related use cases that we could derive from the extant research literature, we applied Design Science (DS) research. A widely recognized suggestion on how to conduct DS projects was introduced by March and Smith (1995) and Peffers et al. (2007). In this respect, DSR represents a synthesis of the activities "build/development" and "justify/evaluate" with the main goal of designing an IT artifact to address an organizational problem *(see Figure 1)* (cf. Gregor and Hevner 2013; Hevner et al. 2004; Peffers et al. 2007).

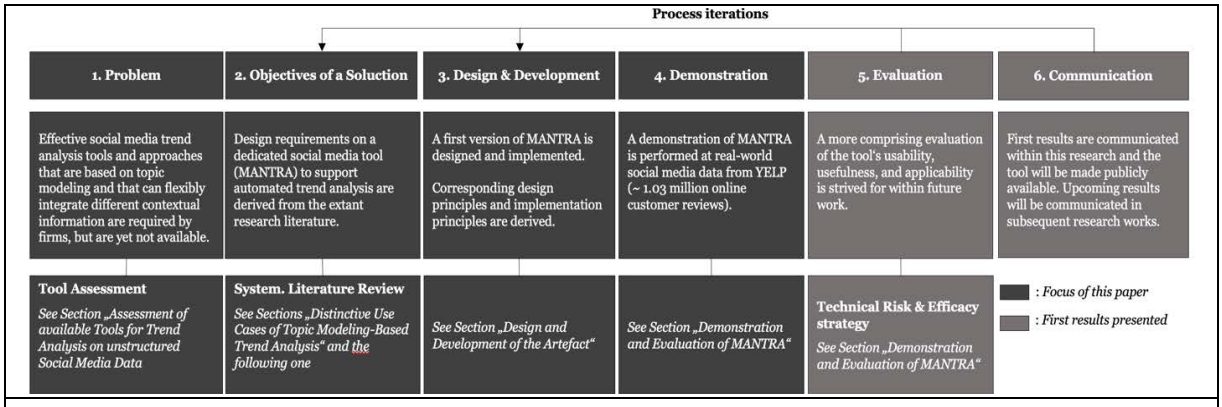

Figure 1. Procedure of the Research, own Representation Based on Peffers et al. (2007)

As a first step, **(1) corresponding problems** of existing software tools for trend analysis were identified *(see Section "Introduction" and Table 2)*. Automated tools for trend analysis neither do meet the requirements that are indispensable to the successful application of trend analysis nor do provide flexibly incorporating external parameters (e.g., by means of semi-supervised topic modeling). Thus, our **(2) objective** is to provide and combine a set of machine learning techniques, based on the DRs, to facilitate trend analysis regarding common marketing-related use cases based on semi-supervised topic modeling.

The third step contains the **(3) design and development** *(see Section "Design and Development of the Artifact")* of the solution. By conducting a systematic literature review (cf. Vom Brocke et al. 2015), we could identify DRs an automated trend analysis tool needs to meet to be comprehensive for common marketing-related use cases *(see Section "Distinctive Use Cases of Topic Modeling-Based Trend Analysis")*. To fill the gaps identified within step (1), we focus on the design of the technical realization of the tool by establishing DPs that are founded on the identified DRs. Thus, our approach was established to support a comprehensive trend analysis for marketing-related use cases and to eliminate the drawbacks of existing solutions.

Steps **(4)** and **(5)** deal with the **demonstration** and **evaluation** of the implemented artifact. To define a proper setting for rigorously demonstrating and evaluating the utility of the artifact, we related to Venable et al. (2016) who have defined four evaluation strategies for DSR projects. We aligned our DSR-project-specific evaluation strategy to the "Technical Risk & Efficacy strategy" because the focal aim of our investigation was to rigorously demonstrate that the evaluand "MANTRA" (MArketiNg TRend Analysis) achieves its purpose (i.e., identify meaningful and sound trends from unstructured social media data) and that the incorporation of semi-supervised topic modeling and meeting several DRs are thereby imperative (cf. March and Smith 1995; Peffers et al. 2012). Thus, we decided to apply one single summative evaluation episode (cf. Venable et al. 2016) in which trends that have already occurred are reproduced retrospectively. Thereby, the results produced by MANTRA are verified and confirmed with two trend reports. Beyond, we plan to investigate the tool within a controlled environment through a laboratory experiment where participants (i.a., experts from the food industry) identify the soundness of the trends and provide feedback on how to improve the tool before we will conduct a more extensive evaluation in a natural setting (e.g., within a field study). In step **(6)**, we aim to **document** and **communicate** the results of our research.

With this procedure, we align our research also with the guideline of Hevner et al. (2004). Regarding the design cycle, we present our artifact as the result that has gone through the process of demonstration (application to a Yelp dataset) and evaluation by a survey of an expert panel from the National Restaurant Association Board (NRAB) *(see Section "Demonstration and Evaluation of MANTRA")*. Regarding the relevance cycle, we identified several DRs from research literature and theory *(see Table 1)* that guided the design of the artifact, and so the practical application of our artifact brought up several contributions for practice *(see Section "Discussion and Contributions")*. In view of the rigor cycle, we used several techniques (i.a., semi-supervised topic modeling, sentiment analysis) to rigorously construct and evaluate our artifact. We have derived initial findings as contributions to theory, both kernel theory (e.g., Innovation Theory, Social Media Analytics Theory) and design theory. To contribute to a rather general and abstract knowledge base – "Nascent design theory" (Gregor and Hevner 2013) – and to design a purposeful artifact in a comprehensible way, we deduced Meta Requirements (MRs) and DRs (Gregor and Jones 2007; Heinrich and Schwabe 2014) for a trend analysis tool grounded in seminal works, resulting in the DPs.

Design and Development of the Artifact

Design Principles for a Trend Analysis Tool

The composition of MRs that describe "*what the system is for*" (Gregor and Jones 2007, p.325) is based on the purpose and scope of the trend analysis tool as discussed in the introduction. Thus, we define the solution objectives based on the problems of the investigation that our paper addresses. These MRs to be suitable for a class of artifacts are based on the current research literature (Gregor and Jones 2007; Heinrich and Schwabe 2014; Walls et al. 1992). Subsequently, we have synthesized DPs based on the DRs we have derived from the related work (i.e., extant literature) as well as based on DRs that are inspired by theory (cf. Möller et al. 2020; Purao et al. 2020). These DPs are defined as prescriptive statements, showing how to do something to achieve a certain goal (Gregor et al. 2020). The DPs we dispose fall into the category of action and materiality-oriented DPs that describe what an artifact should enable users to do and how the artifact should be built in order to do so (Chandra et al. 2015). For companies (= users) who are interested in conducting trend analysis based on unstructured social media data (= boundary conditions), and keeping our DRs for our artifact in mind, we derive six DPs for trend analysis tools:

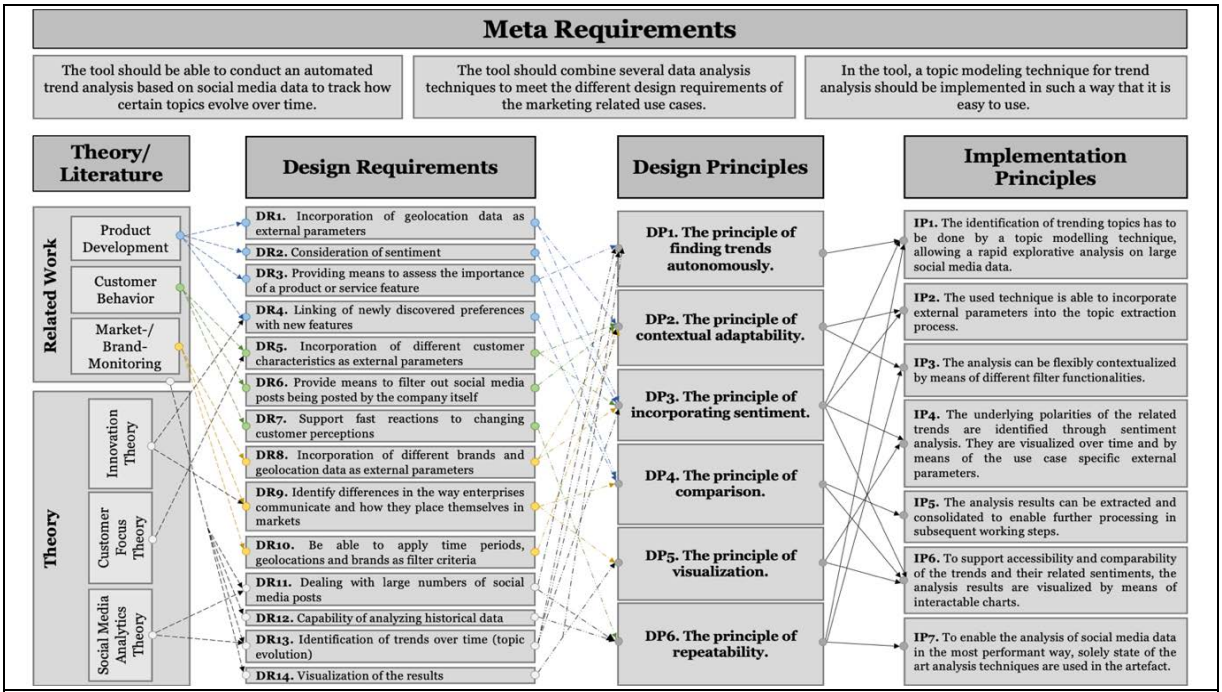

Figure 2. Design of an Automated Trend Analysis Tool for Unstructured Social Media Data

- **DP1: The principle of finding trends autonomously.** To track customers' evolving and changing requirements, it is necessary to retrace the evolution of topics. The tool should be able to find trends in large amounts of unstructured social media data and provide its users to apply autonomous trend analysis by including domain knowledge (i.a., seed words) to refine the identification of trending topics.
- **DP2: The principle of contextual adaptability.** Since external parameters (e.g., geolocations or customer characteristics) directly influence how a trend is pronounced, they must be included in the trend analysis tool. Therefore, the tool needs to provide the possibility to flexibly incorporate different external parameters and filters, so that the users can tailor the trend identification to their peculiarities.
- **DP3: The principle of incorporating sentiment.** To meet customers' expectations, it is necessary to identify issues evoking positive customer perceptions to retain them and to identify issues evoking negative customer perceptions to be improved. Thus, the tool needs to assign polarities to the individual trends and differentiate them over time and for the corresponding use case-specific parameters.
- **DP4: The principle of comparison.** To be competitive in the future, comparisons are necessary to identify the extent to which existing products and services meet customers' expectations (internal

comparison) and to compare a company's actions with those of competitors (external comparison). Thus, the tool should provide the retrieved analysis results to enable both kinds of comparisons.

- **DP5: The principle of visualization.** To obtain added value from the results and to benefit from the derived trends, they must be refined approachable. In doing so, the tool should be able to visualize the trends and sentiment values so that users are aided in interpreting and applying the derived results.
- **DP6: The principle of repeatability.** As a trend can change over time, the identification process should be executable often, and in a resource-saving way. Therefore, the tool should allow repetition at any time (including both historical and current data) to react quickly to changing circumstances.

These DPs are deduced from the DRs that are based on current research literature and theory. Gregor and Jones (2007) state that this reference to theory and literature discloses "*an explanation of why an artifact is constructed as it is and why it works*" (p.328). Based on related work and theory, we derive DRs our tool should meet. These DRs offer guidance by designing the artifact and advising the DPs (Böckle et al. 2021; Gregor and Jones 2007). These DPs refer to at least one DR and serve as an abstract "blueprint" of our artifact (Böckle et al. 2021; Gregor and Jones 2007; Heinrich and Schwabe 2014). By establishing these DPs, we made sure that they follow the value grounding (reference to the DRs) and the explanatory grounding (DPs are based on current literature and kernel theories) (Heinrich and Schwabe 2014). Furthermore, based on the instantiation of the DRs and DPs within our artifact MANTRA, we have defined Implementation Principles (IPs). In this way, we support "*the implementation in practice of an abstract, generic design method or development approach*" (Gregor and Jones 2007, p.329) in view of such an artifact for automated (topic modeling-based) trend analysis on unstructured social media data.

Technical Realization

To address the drawbacks of prevailing research and existing trend analysis tools, we have designed and developed the artifact *MANTRA* in the programming language *Python*. In the following, the instantiation of the DRs is described as they depict the required features of an automated topic modeling-based trend analysis tool in the most detailed way. Since the DPs and IPs depict a generic and prescriptive statement of how something should be done, they capture design-related knowledge and support the development of further IS (design) theories and new artifacts focusing on automated trend analysis. To reveal the association of the derived DPs/IPs to the DRs, their relations are pointed out as well.

Regarding the autonomous identification of trends **(DP1)**, the selection of the topic modeling technique to be applied is critical to success. As trends are time-dependent constructs, a rapid and explorative analysis must be provided. With this respect, unsupervised topic modeling techniques (e.g., LDA) are conceivable to enable an explorative analysis. However, given **DP2**, the incorporation of external parameters **(DR1, DR5, DR8)** is not feasible using unsupervised topic modeling. Beyond that, as stated by Chang et al. (2009) the potential of unsupervised topic modeling is stymied by their purely unsupervised nature, which often leads to topics that are neither meaningful nor effective at extrinsic tasks such as conducting a marketing campaign. Semi-supervised topic modeling techniques retain the flexibility of unsupervised techniques and can identify trends in an autonomous and performant way (Jagarlamudi et al. 2012). In doing so, they can incorporate external parameters such as geolocation information **(DR1, DR8)**, brands or products, and user-related external information **(DR5)**. As a semi-supervised topic modeling technique, *GuidedLDA* was chosen as it has achieved convincing analysis results within previous research (e.g., Ahadh et al. 2021; Chandrasekaran et al. 2020; Toubia et al. 2019). By providing seed information, such as a specific product/brand name or words related to an area such as the global food industry (e.g., burger sauce or burger patty), the topics generated by *GuidedLDA* will converge to gravitate into the contextual direction of those seed words, fostering a successful integration of identified trends into marketing decision-making.

Concerning use case (a) **Product Development**, the identification of trends and their matching with corresponding geolocations are required. Besides the incorporation of geolocation information through respective seed words, we, therefore, enabled MANTRA to pre-define different geolocations by applying a filter function, gaining maximum flexibility regarding the incorporation of external geolocation information **(DR1, DP2, IP3)**. This use case stipulates further the inclusion of sentiments within the analysis to determine the positive or negative tonality of the social media posts **(DR2, DP3, IP4)**. Therefore, we related to the "Valence Aware Dictionary for sEntiment Reasoning" (VADER) (Hutto and Gilbert 2014) technique. *VADER* is a lexicon- and rule-based sentiment analysis technique that is specifically attuned to sentiments expressed in social media and has achieved remarkable results (Hutto and Gilbert 2014). It can

extract positive, neutral, and negative inflections in the range of -1 to +1. Additionally, the importance of a respective product or service feature has to be taken into account **(DR3, DP1, IP1)**. As probabilistic topic modeling techniques such as *GuidedLDA* infer their resulting topics based on various probabilistic distributions, depicting the relevance of the underlying topic words and thus the resulting topics (Crain et al. 2012), the assessment of the importance of the specific topic is met through the nature of topic modeling itself. Moreover, the linking of newly discovered preferences with new features **(DR4, DP4, IP5)** is compulsory. To develop an artifact that is suitable for various applications in marketing, including the processing of heterogeneous datasets, an automated linking of newly discovered preferences will not be feasible. Nevertheless, to take **DR4** into account, a function was implemented to extract the respective results into a Microsoft Excel file. In this way, the results are consolidated and can be used to manually map the identified preferences to new product or service features. Use case (b) **Customer Behavior Analysis** deals, i.a., with identifying behavior changes regarding different kinds of customers. Besides considering customer-related information by biasing the generated topics by means of seed words, users can pre-define customer groups based on attributes such as gender or age **(DR5, DP2, IP2)**. Further, a function to filter out posts created by the company itself to inhibit a bias in the analysis results **(DR6, DP2, IP3)** is required. Thus, MANTRA provides the possibility to dynamically define an identifier that is used to filter out associated data entries. **DR7 (DP6, IP7)** reflects a fast execution time when analyzing data concerning customer behavior. Here, MANTRA implements multi-threading, resulting in the parallel processing of independent tasks. Besides incorporating different brands and geolocation data with respective seed words **(DR8, DP2, IP2)**, (c) **Market-/Brand-Monitoring** requires identifying differences in the way companies communicate and place themselves within markets **(DR9, DP4, IP6)** by applying filter criteria such as time periods, geolocations, or specific brands **(DR10, DP2, IP3)**. Therefore, a filter function was implemented, handling the generation of different subsets of data concerning a specific time period or brand names. This allows comparisons based on diverse dimensions such as topics of interest or specific products and services. To enable the processing of large amounts of unstructured data **(DR11, DP6, IP7)**, only performant analysis techniques settled in the field of text mining are considered. Regarding the capability to view historical data **(DR12, DP6, IP7)**, flexible incorporation of different datasets including different time periods is feasible. The identification of trends over time **(DR13, DP2, IP2)** is considered by the possibility to determine specific time periods, resulting in a flexible identification of time-specific trends and their evolution over time. Finally, the trends, their use case-specific peculiarities, and all customizabilities are visualized **(DR14, DP5, IP6)** by a developed graphical user interface (GUI).

Demonstration and Evaluation of MANTRA

To assess MANTRA's capabilities, we followed our defined evaluation strategy *(see Steps "demonstration" and "evaluation" in Section "Research Procedure")*. In general, MANTRA is designed as an artifact to identify trends from various unstructured social media data. We decided to conduct the demonstration and evaluation on OCRs of the academic Yelp dataset, as it is a publicly available real-world dataset and has proven appropriate for verifying the capabilities of machine learning techniques (e.g., Binder et al. 2019; Heinrich et al. 2022). The academic Yelp dataset (Yelp 2021) comprises about 8.6 million OCRs of 160,585 businesses in different fields (i.a., restaurants, cultural sites, sports facilities) and contains different context information (i.a., creation time, federal state, author information) concerning these OCRs. Because this dataset has already been subject to trend analysis within the food and restaurant industry (e.g., Heinrich et al. 2022; Wörner et al. 2022), we decided to narrow our demonstration to this domain, resulting in a dataset of about 1.03 million OCRs. According to a 2018 survey by a panel of experts from the NRAB, *vegan cuisin*e and *global cuisine* are likely to gain popularity in the US until 2030 (Statista 2019). By means of explorative analysis, we have next identified that restaurants with these two orientations and corresponding food offerings are actually included in the dataset. Together with the results from the previous trend analysis for the food and restaurant domain, this ensured that the possibility of MANTRA producing so-called false positives is considerably reduced (Pries-Heje et al. 2007). Consequently, MANTRA identifies topics for the two trends *vegan cuisine* and *global cuisine* because they are actually contained within the Yelp dataset.

The configuration *(see Figure 3)* represents the view when starting the tool and can be used to customize the trend analysis. The layout is aligned with the GUI design suggestions of Garrett (2010). *Section (i)* presents an excerpt of the dataset used, including the actual reviews and the accompanying business data (e.g., federal state or business name). To consider a representative analysis period concerning our demonstration, the analysis uses data from 2015 to 2021 *(ii)*, as the related NRAB survey conducted in 2018.

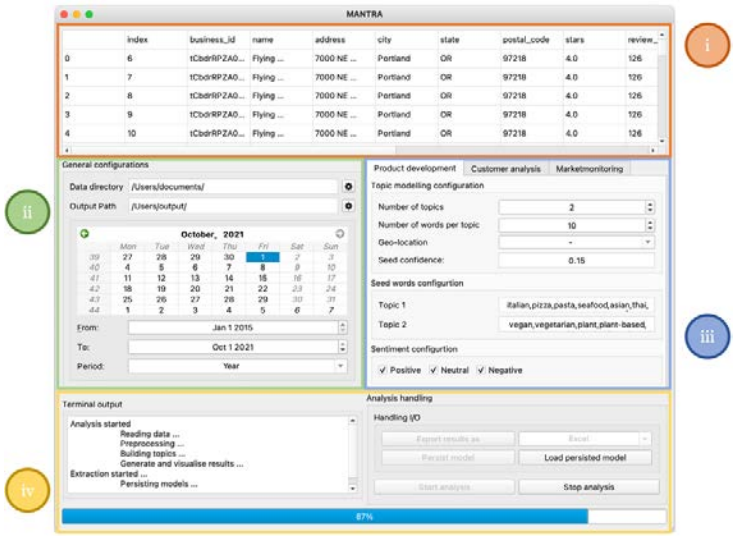

Figure 3. MANTRA – Configuration View

To incorporate the identified use cases, they were implemented in *section (iii)* using modular, dynamic tabs to enable a distinctive configuration. Here, regarding the use case (a) Product Development, the number of topics (*2*), words per topic (*10*), the respective external parameters of geolocation (*none*) and contextualized seed words (e.g., topic 1: Italian, Asian; topic 2: vegan, vegetarian) were defined. Moreover, all sentiment levels (*pos, neg, neu*) were considered. The start of the analysis, the handling of input and output (I/O) operations, and the monitoring take place in *section (iv)*. Once the process has been finished, the results are consolidated and displayed by responsive, interactive charts *(see Figures 4, 5, and 6)*.

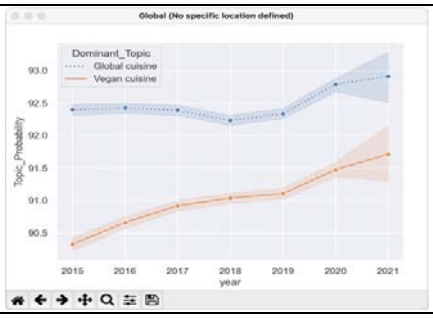

Figure 4. MANTRA – Identified Trends

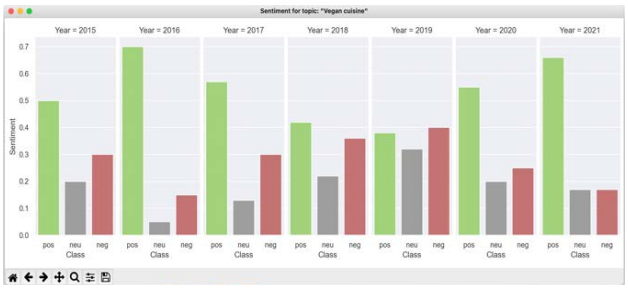

Figure 5. Corresponding Sentiments (Vegan)

Figure 4 represents the determined topic probabilities with respect to the years, resulting in the tracking of the topic evolution based on their relevance. Here, both identified trends are continuously increasing in terms of their relevance. The trend *global cuisine* appears to slightly decrease from 2015 to 2018, but then continuously increases. The trend *vegan cuisine* increases steadily within the analyzed period. These results provide convincing evidence that both trends are on an upward trajectory and exhibit consistency with the findings described in the 2018 NRAB forecast. Moreover, the 2020 and 2022 Food & Health surveys by the International Food Information Council (IFIC 2020, 2022) show a significant percentage of US residents turned to a vegetarian/vegan diet relative to the previous years. As our tool detects a significant increase compared 2019 and 2020, convincing evidence is provided that our tool can identify even small deviations within the evolution of a trend. In addition, *Figure 5* comprises the trend-specific sentiments, by which a trend and the customer demands can be further deconstructed. All in all, our results confirm the general feasibility of MANTRA, as the elicited trends are consistent with the results of the two aforesaid surveys.

Trend topics	Products/Features (noun)	Descriptive characteristics (adj., verb)
vegan cuisine	food, restaurant, vegan, service, gluten, ...	good, great, delicious, best, unfriendly, free, ...

Table 3. Excerpt of the Resulting Excel

The analysis also comprises an Excel file to link the elicited results to new products/features. *Table 3* shows an excerpt of the resulting Excel file, containing the trend-related words and their classification into the part of speech for the topic *vegan cuisine*. This breaks down the trends into constituent parts, enabling the identification of specific products/features and related characteristics. As MANTRA depicts co-occurrences within the elicited trends by topic modeling, identifying sub-trends is feasible. Here, our analysis could, e.g., reveal that customers eating vegan also refer to gluten-free food in a vast frequency *(see Table 3)*.

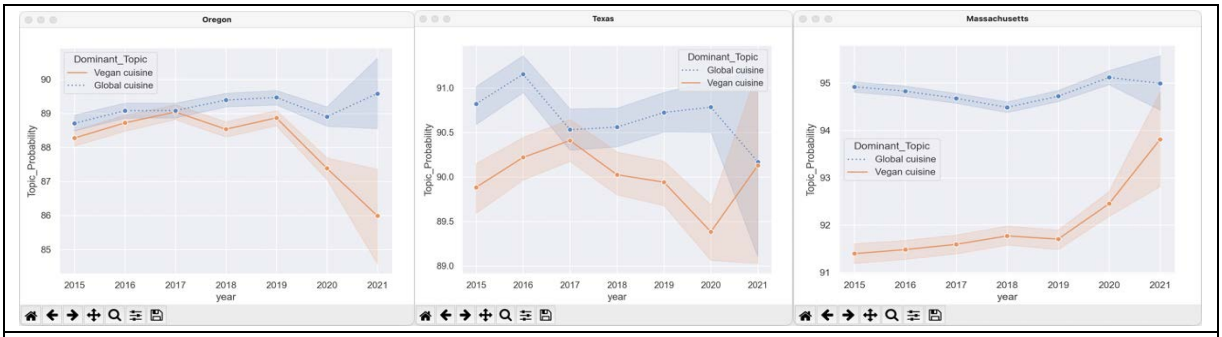

Figure 6. MANTRA - Identified Trends (From Left to Right: OR, TX, MA)

Since (a) **Product Development** refers to examining geolocation-based differences regarding the needs and expectations of customers, we further tested our artifact on sub-datasets referring to three different states in the US: *Massachusetts (MA), Texas (TX), and Oregon (OR)*. The results provide convincing evidence that MANTRA is able to identify geolocation-based discrepancies within the evolution of trends *(see Figure 6)*. It can be clearly seen that the topic *vegan cuisine* differs across the locations, as it consistently increases in MA but remains steady or even decreases in TX and OR. *Global cuisine* has declined since 2015 in TX, while slightly increasing in MA and OR. The sentiment distributions further reveal that – although the relevance of the vegan trend is vigorously increasing in TX from 2020 to 2021 – the respective reviews are negatively connotated *(pos: 0.27, neg: 0.61, neu: 0.12)*. In MA, the trend also increases sharply from 2019 to 2021 but is predominantly characterized by positive reviews *(pos: 0.56, neg: 0.3, neu: 0.14)*.

Generally, the development of our tool was based on the DRs, which are all met and technically realized *(see Figure 2 and the DPs and IPs)*. The 2018 and 2019/2020 trends could be verified by applying our tool to a representative real-world dataset, validating its functionality and therefore its practical applicability.

Discussion and Contributions

The implementation of the DRs and DPs has provided interesting results. As we have included the sentiment in our trend analysis tool, we could identify positive and negative inflections about the trends and how the sentiment values evolve over time. *Figure 5* shows that users spoke positively about the vegan trend in 2016. This positive resonance decreased until 2019, but in 2020 and 2021 the positive comments predominate again. Thus, our demonstration reveals that, although the vegan trend is consistently rising (in terms of topic probability) *(see Figure 4)*, the sentiment evolves distinctively. Without including the sentiment, one would not know that especially in 2019 ambivalent opinions were prevalent about the vegan trend and opening a vegan restaurant would not have yielded the success implicated by the upward trend.

To be successful and competitive in a targeted market, marketing representatives need to know which product and service features to design in which way to meet the global and local customers' expectations (Bae et al. 2018; Endo et al. 2015; Ha et al. 2017). This becomes particularly evident when comparing the three states of OR, TX, and MA *(see Figure 6)*. While both trends in MA have been growing steadily since 2015, the situation in TX and OR is different. In TX, the vegan trend declined between 2017 and 2020. However, in 2020, it started to gain importance. This is in contrast to the global cuisine trend which, peaked in 2016 but has dropped sharply since 2020. OR exhibits a completely different picture of the two trends. While the vegan trend initially remains stable and dropped sharply in 2019, the global cuisine trend remains stable over the years. The comparison of the three figures clearly indicates that flexibly including geolocation information at different levels of granularity (such as federal states) can have a considerable influence on whether and how to harness the trends. While aligning a restaurant to the vegan trend is (based

on these analyses) advisable in MA, this is not the case in OR. To better assess these trends in TX and MA, we have combined the two alignments – geolocation and sentiment – because, as mentioned above, disclosing an upcoming trend does not directly mean that the trend exhibits necessarily positive connotations. Whereas the vegan trend in TX has gained in importance since 2020, the sentiment is predominantly characterized by negative contributions *(neg: 0.61)*. This indicates that either the vegan food offered so far does not meet the customers' expectations, or that vegan food is not well received in TX in general. In contrast to TX, the vegan trend in MA is predominantly characterized by positive reviews on average *(pos: 0.56)*. Without the inclusion of the geolocation data and the sentiment values, these aforementioned discrepancies in the results would not have been revealed.

Our investigation contributes to research and practice alike. As a contribution to practice, first companies can benefit from our comprehensive and modular artifact. MANTRA meets several DRs and, in contrast to the available solutions, does additionally provide a semi-supervised topic modeling technique *(GuidedLDA)*. Thus, MANTRA is capable of integrating several external parameters, which is deemed important in the extant research literature (Bae et al. 2018; Endo et al. 2015; Ha et al. 2017). Beyond that, MANTRA allows marketing representatives to integrate the use case-specific external parameters even at different granularity (e.g., federal states as geolocation) and therefore, i.a., to identify fine-grained discrepancies in customer perceptions. Moreover, we have implemented a functionality to link future customer preferences with new features (cf. DR4). By extracting the results of the trend analysis into an Excel file and classifying the trend-related words into the part of speech, trends and their constituent parts can be analyzed in more detail. For example, a company that has analyzed the trends that emerged in Section "*Demonstration and Evaluation of MANTRA*" could recognize that customers eating vegan also refer to gluten-free food with a vast frequency. As our results indicate, this functionality of a trend analysis tool is eminent as it supports companies, on the one hand, to identify fine-grained aspects of the elicited trends and, on the other hand, to support that the gained insights can be seamlessly included into the steps that are following the actual trend analysis (i.a., trend assessment and integration into the product and service portfolio of a company). In general, MANTRA has been developed in a modular way, enabling its users to define various settings through the corresponding GUI elements. For example, the time period for a specific trend can be set to conduct the analysis resource-efficiently. Several use case-specific DRs were implemented by offering modular and dynamic tabs to enable quick and almost effortless processing of common marketing-related use cases. As we have combined different machine learning approaches and designed our tool modularly, companies can adapt the analysis to their individual circumstances.

Furthermore, as an outcome of our DSR project, we achieved theoretical contributions to research that go beyond the technical contribution (i.e., the artifact). As we have designed a tool that is able to identify the voice of the customer, even about unknown customer demands and design features, our results also contribute to research theories such as Innovation Theory. Here, our approach can be included, e.g., in the stage-gate model of the innovation process that describes a conceptual and operational model for moving new product projects from idea to launch (Cooper 1996). Thereby, our results show that it is important to distinguish trends based on external parameters (e.g., geolocation), as customer perceptions may differ across global and local sites. By incorporating specific trend-related information within the different stages of the stage-gate model, the rigid specifications of stages and gates can be loosened up. Since social media enables customers to express their unfiltered opinions (Yan et al. 2014), thereon identified trends reflect customer perceptions, which can ensure that customers' needs are considered. Thus, by integrating the trend information in this way, the assessments at the go/kill checkpoints (i.e., gates) become less elaborate as the alignment with external requirements (e.g., customer wishes and demands) is maintained constantly. Moreover, Social Media Analytics Theory (especially concerning the research of automated trend analysis) states that the automated analysis of social media content still holds various challenges when it comes to its actual application (Fan and Gordon 2014; Stieglitz et al. 2018). With the design and development of our automated trend analysis tool, we purposefully combined various machine learning techniques (i.a., semi-supervised topic modeling, and sentiment analysis) and ensured that the DRs and the corresponding marketing-related use cases are technically realized. In this way, we contribute to Social Media Analytics Theory by proposing a need-fitted automated trend analysis for unstructured social media data.

The discussed results have shown the advantages of our tool "MANTRA" compared to existing ones, placing our DSR project in the group of improvements (development of new solutions for known problems) in terms of the DSR knowledge contribution framework of Gregor and Hevner (2013). DSR improvement projects contribute to both prescriptive theory, i.e., design theory (Gregor 2006), and descriptive theory, i.e., kernel

theories such as Innovation- or Social Media Analytics Theory as described above (Gregor and Hevner 2013). By including GuidedLDA as a semi-supervised topic modeling technique within MANTRA that can flexibly integrate different external parameters as deemed crucial by our marketing-related use cases, our research as well affects the quadrant of exaptation (cf. Gregor and Hevner 2013). We have shown the proficiency of this technique and in this way contribute to IS and DSR research alike, as we propose semi-supervised topic modeling as a suitable technique for automated trend analysis (tools) to tailor the derived trends towards context factors and semantically towards the company and its fields of activity.

Based on the DRs derived from literature and kernel theories, DPs were formulated and proposed. By applying them during the design and development of the artifact followed by the demonstration and evaluation, an implicit empirical grounding of the DPs was achieved here (Heinrich and Schwabe 2014). Our DPs capture design-related knowledge and can thus support the development of further IS (design) theories and new artifacts. By considering e.g., DP2 Contextual adaptability, the importance of the contextualization (e.g., location) is highlighted. Since context directly impacts trend analysis results, the alignment with context leads to more meaningful insights. With the DPs, we made a first step towards contributing to design theory in terms of theory for design and action (Gregor 2006) as we comply with the conditions of March and Smith (1995) and Hevner et al. (2004), who pointed out under which a contribution to knowledge in DS has occurred: utility to a community of users, the novelty of the artifact and the persuasiveness of claims that it is effective.

Conclusion, Limitation, and Future Research

The early identification of new and auspicious ideas as well as trending topics regarding the development of products and services, the analysis of customers' behaviors, and the monitoring of markets and competing brands lead to competitive advantages for companies. However, prior literature and existing solutions for automated trend analysis on unstructured social media data do not incorporate semi-supervised topic modeling and do not sufficiently cover the different external use case-specific parameters as well as the specific requirements of the common marketing-related use cases (a) Product Development, (b) Customer Behavior Analysis, and (c) Market-/Brand-Monitoring.

Therefore, we identify several DRs and derive DPs. To technically realize the DPs, we combine several machine learning techniques and transfer these in an intuitive GUI to close the revealed research gaps *(see RQ1 and RQ2)*. Especially with the demonstration to about 1.03 million OCRs, we show that covering all DRs is essential for a target-oriented and feasible trend analysis. MANTRA also supports unsupervised trend analysis when no seed words are provided. Hence, also marketing representatives who have not gained prior knowledge of underlying trends can also conduct an explorative trend analysis using MANTRA.

Our investigation proposes contributions to practice and research alike *(see RQ2)*. Companies can benefit from our comprehensive and modular artifact, with which large amounts of unstructured social media data can be analyzed in a way best suited to the company's individual circumstances. We have also highlighted how our investigation made a first step towards contributing to design theory and kernel theories. Regarding Innovation Theory, the rigid sequence of stages and gates in the stage-gate model (Cooper 1996) can be loosened up by examining external requirements constantly. In this vein and according to the demonstration of MANTRA, it is exemplified that the stages should incorporate activities to assess the influences of the different external parameters (e.g., continents, phases of a day, personal characteristics; in the demonstration: federal states) on the perception of trends. By incorporating the DRs of a trend analysis tool, we enrich Social Media Analytics Theory with a need-fitted automated approach for analyzing social media content regarding the automated identification of trends.

It is worth mentioning, that the initial assessments on the practical applicability of MANTRA regarding (b) Customer Behavior Analysis and (c) Market-/Brand-Monitoring revealed promising results. Therefore, we will investigate these two use cases in more detail in the future and examine in more detail the potential of combining the three identified use cases with each other. In addition, to take the next steps toward a more mature design theory, we will first evaluate our artifact in a formative and artificial environment (e.g., a laboratory experiment). This allows us to improve our tool (whereby our DPs can be confirmed or adapted) before conducting a more elaborate evaluation in a more natural setting as a further part of the design cycle. There are also limitations to this research: Although we included a large set of investigations, we could identify probably even more use cases in further literature. Additionally, researchers from other fields could identify other use cases. Nevertheless, the identified use cases are undoubtedly important for marketing.

References

Ahadh, A., Binish, G., and Srinivasan, R. 2021. "Text mining of accident reports using semi-supervised keyword extraction and topic modeling," *Process Safety and Environmental Protection* (155).

Axtell, C., Holman, D., Unsworth, K., Wall, T., Waterson, P., and Harrington, E. 2000. "Shopfloor innovation: Facilitating the suggestion and implementation of ideas," *J. Occup. Psychol.* (73:3).

Bae, J., Havsol, J., Karpefors, M., Karlsson, A., and Mathiason, G. 2018. "Short Text Topic Modeling to Identify Trends on Wearable Bio-Sensors in Different Media Type," in *Proceedings of the 6th International Symposium on Computational and Business Intelligence*.

Bai, S., He, H., Han, C., Yang, M., Yu, D., Bi, X., and Panigrahi, P. 2023. "Exploring thematic influences on theme park visitors' satisfaction: An empirical study on Disneyland China," *J. Consum. Behav.* (22:2).

Batrinca, B., and Treleaven, P. 2015. "Social media analytics: a survey of techniques, tools and platforms," *AI & Society* (30), pp. 89-116.

Bhor, H., Koul, T., Malviya, R., and Mundra, K. 2018. "Digital Media Marketing Using Trend Analysis on Social Media," in *Proceedings of the 2nd International Conference on Inventive Systems and Control*.

Binder, M., Heinrich, B., Klier, M., Obermeier, A., and Schiller, A. 2019. "Explaining the stars: Aspect-based sentiment analysis of online customer reviews," in *Proceedings of the ECIS 2019*.

Böckle, M., Bick, M., and Novak, J. 2021. "Toward a Design Theory of User-Centered Score Mechanics for Gamified Competency Development," *Information Systems Management* (40:1), pp. 1-27.

Burmeister, K., Neef, A., Albert, B., and Glockner, H. 2002. "Zukunftsforschung und Unternehmen," *Praxis, Methoden, Perspektiven*. Essen.

Chandra, L., Seidel, S., and Gregor, S. 2015. "Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions," in *Proceedings of the 48th Hawaii International Conference on System Sciences*.

Chandrasekaran, R., Mehta, V., Valkunde, T., and Moustakas, E. 2020. "Topics, trends, and sentiments of tweets about the COVID-19 pandemic," *Journal of Medical Internet Research* (22:10).

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., and Blei, D. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models," *Advances in Neural Information Processing Systems*, pp. 288-296.

Chen, J., Yu, J., and Shen, Y. 2012. "Towards Topic Trend Prediction on a Topic Evolution Model with Social Connection," in *Proceedings of the International Conferences on WI-IAT*.

Cooper, R. 1996. "Overhauling the new product process," *Industrial Marketing Management* (25:6).

Crain, S., Zhou, K., Yang, S., and Zha, H. 2012. "Dimensionality Reduction and Topic Modeling: From Latent Semantic Indexing to Latent Dirichlet Allocation and Beyond," *Mining Text Data*, pp. 129-161.

Ding, K., Choo, W., Ng, K., Ng, S. 2020. "Employing structural topic modelling to explore perceived service quality attributes in Airbnb accommodation," *International Journal of Hospitality Management* (91).

Duncan, T., and Moriarty, S. 1998. "A Communication-Based Marketing Model for Managing Relationships," *Journal of Marketing* (62:2), pp. 1-13.

Endo, Y., Toda, H., and Koike, Y. 2015. "What's Hot in the Theme: Query Dependent Emerging Topic Extraction from Social Streams," in *Proceedings of the 24th International Conference on the WWW*.

Fan, W., and Gordon, M. 2014. "The power of social media analytics," *Communications of the ACM* (57:6).

Garrett, J. 2010. "The Elements of User Experience: User-Centered Design for the Web and Beyond".

Gregor, S. 2006. "The nature of theory in information systems," *MIS Quarterly* (30:3), pp. 611-642.

Gregor, S., Chandra Kruse, L., and Seidel, S. 2020. "Research Perspectives: The Anatomy of a Design Principle," *Journal of the AIS* (21:6).

Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.

Gregor, S., and Jones, D. 2007. "The Anatomy of a Design Theory," *Journal of the AIS* (8:5).

Gulati, R., and Oldroyd, J. 2005. "The Quest for Customer Focus," *Harvard Business Review* (83:4).

Ha, T., Beijnon, B., Kim, S., Lee, S., and Kim, J. 2017. "Examining User Perceptions of Smartwatch through Dynamic Topic Modeling," *Telematics and Informatics* (34:7), pp. 1262-1273.

Heinrich, B., Hollnberger, T., Hopf, M., and Schiller, A. 2022. "Long-term Sequential and Temporal Dynamics in Online Consumer Ratings," in *Proceedings of the ECIS 2022*.

Heinrich, P., and Schwabe, G. 2014. "Communicating Nascent Design Theories on Innovative Information Systems through Multi-Grounded Design Principles," in *Proceedings of DESRIST 2014*.

Hevner, A., March, S., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75–105.

Hicks, A., Comp, S., Horovitz, J., Hovarter, M., Miki, M., and Bevan, J. 2012. "Why People Use Yelp.com: An Exploration of Uses and Gratifications," *Computers in Human Behavior* (28:6), pp. 2274-2279.
Hong, L., Ahmed, A., Gurumurthy, S., Smola, A., and Tsioutsiouliklis, K. 2012. "Discovering Geographical Topics in the Twitter Stream," in *Proceedings of the 21st international conference on WWW*.
Hong, L., and Davison, B. D. 2010. "Empirical Study of Topic Modeling in Twitter," in *Proceedings of the 1st Workshop on Social Media Analytics*.
Humphreys, A., and Wang, J.-H. R. 2018. "Automated text analysis for consumer research," *Journal of Consumer Research* (44:6), pp. 1274-1306.
Hutto, C., and Gilbert, E. 2014. "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in *Proceedings of the International AAAI Conference on Web and Social Media*.
IFIC (2020). "2020 Food & Health Survey. 9th June 2020," *https://foodinsight.org/2020-food-and-health-survey/*. Lastly accessed on 6th April 2023.
IFIC (2022). "2022 Food & Health Survey. 18th May 2022," *https://foodinsight.org/2022-food-health-survey/*. Lastly accessed on 6th April 2023.
Irawan, M., Wijayanto, R., Shahab, M., Hidayat, N., and Rukmi, A. 2020. "Implementation of Social Media Mining for Decision Making in Product Planning Based on Topic Modeling and Sentiment Analysis," *Journal of Physics: Conference Series*.
Iwata, T., Watanabe, S., Yamada, T., and Ueda, N. 2009. "Topic tracking model for analyzing consumer purchase behavior," in *Proceedings of the 21st International Joint Conference on AI*.
Jagarlamudi, J., Daumé III, H., and Udupa, R. 2012. "Incorporating Lexical Priors into Topic Models," in *Proceedings of the 13th Conference of the European Chapter of the ACL*.
Jeong, B., Yoon, J., and Lee, J.-M. 2019. "Social Media Mining for Product Planning: A Product Opportunity Mining Approach Based on Topic Modeling and Sentiment Analysis," *Int. J. Inf. Manag.* (48).
Kaplan, A., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business Horizons* (53:1), pp. 59-68.
Khan, H. U., Nasir, S., Nasim, K., Shabbir, D., and Mahmood, A. 2021. "Twitter trends: a ranking algorithm analysis on real time data," *Expert Systems with Applications* (164).
Ko, N., Jeong, B., Choi, S., and Yoon, J. 2018. "Identifying product opportunities using social media mining: application of topic modeling and chance discovery theory," *IEEE Access* (6), pp. 1680-1693.
Lozano, M., Schreiber, J., and Brynielsson, J. 2017. "Tracking Geographical Locations Using a Geo-Aware Topic Model for Analyzing Social Media Data," *Decision Support Systems* (99), pp. 18-29.
March, S., and Smith, G. 1995. "Design and Natural Science Research on Information Technology," *Decision Support Systems* (15:4), pp. 251-266.
Mastrogiacomo, L., Barravecchia, F., Franceschini, F., and Marimon, F. 2021. "Mining quality determinants of product-service systems from user-generated contents," *Quality Engineering* (33:3), pp. 425-442.
Mayring, P. 2000. "Qualitative Content Analysis," *Forum Qualitative Social Research* (1:2).
Mayring, P. 2002. "Introduction to qualitative social research," *Weinheim: Beltz*.
Möller, F., Guggenberger, T. M., and Otto, B. 2020. "Towards a method for design principle development in information systems," in *Proceedings of DESRIST 2020*.
Moro, S., Lopes, R., Esmerado, J., and Botelho, M. 2020. "Service quality in airport hotel chains through the lens of online reviewers," *Journal of Retailing and Consumer Services* (56).
Mudambi, S. M., and Schuff, D. 2010. "Research note: What makes a helpful online review? A study of customer reviews on Amazon.com," *MIS Quarterly* (34:1), pp. 185-200.
Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. 2012. "Mine your own business: Market-structure surveillance through text mining," *Marketing Science* (31:3), pp. 521-543.
Park, E., Kwon, J., and Kim, S. 2021. "Green marketing strategies on online platforms: a mixed approach of experiment design and topic modeling," *Sustainability* (13:8).
Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3).
Peffers, K., Rothenberger, M., Tuunanen, T., and Vaezi, R. 2012. "Design science research evaluation," in *Proceedings of DESRIST 2012*.
Pries-Heje, J., Baskerville, R., and Venable, J. 2007. "Soft design science research: Extending the boundaries of evaluation in design science research," in *Proceedings of DESRIST 2007*, pp. 18-38.
Puranam, D., Narayan, V., and Kadiyali, V. 2017. "The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors," *Marketing Science* (36:5).
Purao, S., Kruse, L. C., and Maedche, A. 2020. "The Origins of Design Principles: Where do… they all come from?, " in *Proceedings of DESRIST 2020*, pp. 183-194.

Qu, Q., Chen, C., Jensen, C., and Skovsgaard, A. 2015. "Space-Time Aware Behavioral Topic Modeling for Microblog Posts," *IEEE Data Engineering Bulletin* (38:2), pp. 58-67.
Reisenbichler, M., and Reutterer, T. 2019. "Topic modeling in marketing: recent advances and research opportunities," *Journal of Business Economics* (89), pp. 327-356.
Saha, A., and Sindhwani, V. 2012. "Learning Evolving and Emerging Topics in Social Media: A Dynamic NMF Approach with Temporal Regularization," in *Proceedings of the Conference on WSDM 2012*.
Sasaki, K., Yoshikawa, T., and Furuhashi, T. 2014. "Online Topic Model for Twitter Considering Dynamics of User Interests and Topic Trends," in *Proceedings of the Conference on Empirical Methods in NLP*.
Shi, L., Wu, Y., Liu, L., Sun, X., and Jiang, L. J. 2018. "Event Detection and Identification of Influential Spreaders in Social Media Data Streams," *Big Data Mining and Analytics* (1:1), pp. 34-46.
Sigala, M. 2012. "Social Networks and Customer Involvement in New Service Development (Nsd): The Case of www.Mystarbucksidea.com," *International Journal of Contemporary Hospitality* (24:7).
Sohn, B., Lim, H., and Choi, J. 2019. "The Prediction of Sales Volume and WoM Effect through Network Structure and Topic Modeling on Social Media," in *Proceedings of the PACIS 2019*.
Statista. 2019. "Ranking der prognostizierten Restaurant-Trends im Bereich 'Food & Menu'," *https://de.statista.com/statistik/daten/studie/1125284*. Lastly Accessed on 6th April 2023.
Stieglitz, S., Mirbabaie, M., Ross, B., and Neuberger, C. 2018. "Social Media Analytics – Challenges in Topic Discovery, Data Collection, and Data Preparation," *Int. J. Inf. Manag.* (39), pp. 156-168.
Sundbo, J. 1995. "Three paradigms in innovation theory," *Science and Public Policy* (22:6), pp. 399-410.
Tirunillai, S., and Tellis, G. 2014. "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," *Journal of Marketing Research* (51:4).
Tseng, M., and Piller, F. 2003. "The customer centric enterprise," *Springer Science & Business Media*.
Toubia, O., Iyengar, G., Bunnell, R., and Lemaire, A. 2019. "Extracting features of entertainment products: A guided latent dirichlet allocation approach informed by the psychology of media consumption," *Journal of Marketing Research* (56:1), pp. 18-36.
Tucker, C., and Kim, H. 2011. "Predicting Emerging Product Design Trend by Mining Publicly Available Customer Review Data," in *Proceedings of the 18th International Conference on Engineering Design*.
Ulbrich, F., Christensen, T., and Stankus, L. 2011. "Gender-Specific on-Line Shopping Preferences," *Electronic Commerce Research* (11:2), pp. 181-199.
Valdez, D., Padon, A., Barry, A., and Russell, A. 2018. "Alcohol Advertising on Twitter - a Topic Model," *American Journal of Health Education* (49:4), pp. 256-263.
Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: a framework for evaluation in design science research," *European Journal of Information Systems* (25:1), pp. 77-89.
Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. "Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research," *Communications of the AIS* (37:1), pp. 205-224.
Walls, J., Widmeyer, G., and El Sawy, O. 1992. "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research* (3:1), pp. 36-59.
Wörner, J., Konadl, D., Schmid, I., and Leist, S. 2022. "Supporting Product Development by a Trend Analysis Tool Applying Aspect-Based Sentiment Detection," in *Proceedings of DESRIST 2022*.
Yan, X., Guo, J., Lan, Y., Xu, J., and Cheng, X. 2015. "A Probabilistic Model for Bursty Topic Discovery in Microblogs," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence*.
Yan, Z., Xing, M., Zhang, D., Ma, B., and Wang, T. 2014. "A Context-Dependent Sentiment Analysis of Online Product Reviews Based on Dependency Relationships," in *Proceedings of the ICIS 2014*.
Yang, Y., Yao, Q., and Qu, H. 2017. "Vistopic: A Visual Analytics System for Making Sense of Large Document Collections Using Hierarchical Topic Modeling," *Visual Informatics* (1:1), pp. 40-47.
Ye, H. J., and Kankanhalli, A. 2018. "User service innovation on mobile phone platforms: Investigating impacts of lead userness, toolkit support, and design autonomy," *MIS Quarterly* (42:1), pp. 165-188.
Yelp. 2021. "Yelp Open Dataset," *https://www.yelp.com/dataset*. Accessed in April 2022.
Zhang, H., Kim, G., and Xing, E. P. 2015. "Dynamic Topic Modeling for Monitoring Market Competition from Online Text and Image Data," in *Proceedings of the 21th KDD Conference*.
Zhang, P., Gu, H., Gartrell, M., Yang, D., Ding, X., Gu, N. 2016. "Group-based Latent Dirichlet Allocation (Group-LDA): Effective audience detection for books in online social media," *Knowl. Based Syst.* (105).
Zhao, K., Cong, G., Chin, J.-Y., and Wen, R. 2019. "Exploring Market Competition over Topics in Spatio-Temporal Document Collections," *The VDLB Journal* (28:1), pp. 123-145.
Zhong, N., and Schweidel, D. 2020. "Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model," *Marketing Science* (39:4), pp. 827-846.

2.4. Beitrag 4: Supporting Product Development by a Trend Analysis Tool Applying Aspect-Based Sentiment Detection

Adressierte Forschungsfrage	Wie sollte ein auf aspektbasierter Sentimentanalyse basierendes Trendanalysetool aussehen, welches die Trendanalyse für den Anwendungsfall der Produktentwicklung unterstützt und welche Anforderungen sollte ein solches Trendanalysetool dabei erfüllen?
Zielstellungen	• Identifikation von Designanforderungen an ein Trendanalysetool unter Berücksichtigung der aspektbasierten Sentimentanalyse. • Erstellung des auf aspektbasierter Sentimentanalyse basierenden Trendanalysetools durch die technische Umsetzung und Kombination verschiedener Social Media Analysetechniken.
Forschungsmethode	**Design Science Research** • DSR-Prozess nach Peffers et al. (2007), der basierend auf einer Problem- und Lösungsidentifikation die Schritte Entwicklung, Demonstration und Evaluation für das auf aspektbasierter Sentimentanalyse basierenden Trendanalysetools umgesetzt hat. • Anlehnung der Forschung an Hevner et al. (2004) durch die Ausgestaltung des Design Cycle (Demonstration und Evaluation des Artefakts), Relevance Cycle (Beitrag zur Praxis) und Rigor Cycle (Beitrag zur (Nascent) Designtheorie durch Designprinzipien (vgl. Gregor und Hevner, 2013)).

Kernergebnisse (Überblick)	• Ableitung von Designprinzipien auf Basis der in der wissenschaftlichen Literatur identifizierten Designanforderungen.
	• Umsetzung und Kombination von verschiedenen Analyseverfahren (aspektbasierte Sentimentanalyse, Topic Modelling, Neuronale Netzwerke) zur technischen Umsetzung der Designanforderungen.
	• Einbezug externer Parameter (z. B. Geolokationen) und Berücksichtigung der Identifikation von Trends mit und ohne Domänenwissen der Toolanwender (insb. Marketingverantwortliche in Unternehmen).
	• Implementierung des auf aspektbasierter Sentimentanalyse basierenden Trendanalysetools zur Produktverbesserung.
	• Demonstration an einem Yelp Datensatz mit 37.638 Online Customer Reviews aus dem Restaurantbereich:
	  o Identifikation von verschiedenen Aspekten (Burgerbun, Fleisch, Sauce), die aus Sicht der Verfasser der Online Customer Reviews am wichtigsten sind.
	  o Betrachtung der Sentimente zu den Aspekten im Zeitverlauf.
	  o Unterscheidung der Sentimente hinsichtlich der Geolokationen, um ortsabhängige Diskrepanzen zu identifizieren.

Publikationsort	**Veröffentlicht** in den Proceedings der 17. International Conference on Design Science Research in Information Systems and Technology, St. Petersburg, Florida, USA.
Ranking VHB JQ 4	**B**
Autoren und Anteile	Janik Wörner 30 % **Daniel Konadl 30 %** Isabel Schmid 30 % Susanne Leist 10 %

Tabelle 6: Fact Sheet zu Beitrag 4

Supporting Product Development by a Trend Analysis Tool Applying Aspect-Based Sentiment Detection

Janik Wörner(✉), Daniel Konadl, Isabel Schmid, and Susanne Leist

University of Regensburg, Regensburg, Germany
{Janik.Woerner,Daniel.Konadl,Isabel.Schmid,Susanne.Leist}@ur.de

Abstract. Incorporating product trends into innovation processes is imperative for companies to meet customers' expectations and to stay competitive in fiercely opposing markets. Currently, aspect-based sentiment analysis has proven an effective approach for investigating and tracking towards products and corresponding features from social media. However, existing trend analysis tools on the market that offer aspect-based sentiment analysis capabilities, do not meet the requirements regarding the use case Product Development. Therefore, based on these requirements, we implemented an artifact by following the design science research. We applied our tool to real-world social media data (37,638 Yelp reviews) from one major fast-food restaurant in the US, and thereby demonstrated that our tool is capable of identifying remarkable and fine-grained product trends.

Keywords: Trend analysis tool · Aspect-based sentiment · Product development

1 Motivation

Social media such as Yelp or Twitter have evolved rapidly over the last years. These platforms have become increasingly important for interaction in both private and business contexts [1, 2]. As social media is a channel for the exchange of user-generated content and unfiltered voices about products, services and the company in general, social media data contain the so-called "Voice of the Customer" (VoC). Thus, the VoC provides deep insights into customers' current expectations. To meet customers' expectations, marketing representatives need to identify and continuously track trending topics regarding product and service features and incorporate the VoC into product innovation processes. For example, identified product features and correspondingly mentioned opinions may indicate shortcomings (e.g., low battery capacity of a smartphone) and which improvements to be made to meet customers' requirements (e.g., [3]). One possibility to identify these shortcomings in an automated way from social media texts is to conduct aspect-based sentiment analysis [4].

The potential of aspect-based sentiment analysis for tracking fine-grained trends over time has already been recognized in practice and in theory (e.g., [3, 5–9]). This has led to the emergence of trend analysis tools that include aspect-based sentiment analysis

© Springer Nature Switzerland AG 2022
A. Drechsler et al. (Eds.): DESRIST 2022, LNCS 13229, pp. 68–80, 2022.
https://doi.org/10.1007/978-3-031-06516-3_6

functionalities. However, trend analysis tools available on the market have remarkable drawbacks as they do not cover the comprehensive requirements that are deemed essential within the extant literature for the use case Product Development (e.g., [3, 5–9]).

With this work at hand, we make practical as well as theoretical contributions. We address drawbacks of existing software tools by suggesting a comprehensive artifact for automated trend analysis that allows marketing representatives to conduct aspect-based sentiment analysis. To meet several use case-specific requirements, we focus especially on the combination of different data analysis methods regarding the particular requirements, leading to a constructive trend analysis. By this, we aim to propose an automated solution for identifying ideas as the basis of (incremental) product innovation. Summing up, the research at hand is guided by the following research question:

What could an aspect-based sentiment analysis tool that supports trend analysis for Product Development purposes look like, and which requirements should such a tool meet?

The remainder of this paper is structured as follows: In the next section, we provide conceptual basics and related work. Following on this, we turn to the Design Requirements (DRs) and Design Principles (DPs) for implementing our tool, and to the shortcomings of trend analysis tools on the market. Next, we show the research methodology. After a description of the tool's design and development as well as its demonstration, the paper concludes with a discussion and its contributions to theory and practice.

2 Foundations and Related Work

2.1 Conceptual Background

Social media serves as an important interface between companies and customers. In content communities, users can evaluate products by disseminating their opinions in form of online customer reviews (OCR). In doing so, customers not only rate products as a whole but express their opinions and attitudes towards different features of the rated items (e.g., service quality in a Yelp restaurant review). In this way, OCR not only help customers to make informed decisions but are also beneficial for driving innovations of products within companies. As OCR include customers' experiences and expectations of product features [10], unfiltered and in real-time [11], they can serve as a valuable resource for product innovations. Thus, OCR can be harnessed to identify ideas, to either develop new value propositions (i.e., disruptive innovation) or to improve the performance of existing products (i.e., incremental innovation) [12, 13].

To identify ideas for product improvements as well as product development and therefore to drive incremental product innovations, marketing representatives can conduct aspect-based sentiment analysis. The first step of an aspect-based sentiment analysis deals with extracting aspects from OCR. For this purpose, unsupervised as well as supervised techniques can be applied. Topic modeling techniques (e.g., LDA [14]) suggest a possibility to identify aspects without prior knowledge (i.e., unsupervised) [14]. Compared to that, supervised techniques (e.g., artificial neural networks) need first to be trained on training data (e.g., ontologies) to extract the proper aspects (cf. [4]). Subsequently, the expressed tonalities can be identified for each of the aspects [4] by means of automated sentiment analysis techniques [4, 15]. Therefore, aspect-based sentiment

analysis offers benefits in terms of Product Development. For marketing representatives that lack the ability to implement aspect-based sentiment analysis themselves, trend analysis tools on the market offer this functionality in a ready-to-use way. However, these tools show remarkable drawbacks as they do not cover the comprehensive requirements that are deemed essential within the extant literature for the use case of Product Development.

2.2 Design Requirements and Available Tools on the Market

In a first step, we have comprehensively searched and consolidated literature (cf. [16]) to identify DRs of a trend analysis tool that applies aspect-based sentiment analysis for Product Development purposes. Based on the attention and importance received, we could derive several DRs (DR1)-(DR10) (see Fig. 2) from the extant literature.

Concerning the identification of product features, the tool (DR1) should be capable of extracting the aspects autonomously from social media posts (e.g., [6–8, 17]). However, if marketing representatives have already knowledge documented about a domain problem (e.g., domain ontologies or product trees), the tool (DR2) should provide the option to include this prior knowledge into the automated identification of aspects (cf. [6, 7, 17, 18]). Beyond that, customers' self-reported opinions of product features play an important role for Product Development. Marketing representatives aim to retain the features that evoke positive perceptions, while features evoking negative perceptions need to be improved. Thus, (DR3) determining the polarity as well as the intensity of the opinions expressed about respective aspects is mandatory [3, 6, 8, 19, 20]. Subsequently, the product features can be adapted so that customers' requirements are met (e.g., increasing smartphone screen size). However, adapting features may influence perceptions of the features customers currently appreciate (e.g., high battery capacity of the smartphone). Therefore, to support informed improvement decisions, the tool (DR4) needs to identify the dependencies between product features [7, 9, 17]. Furthermore, to be successful and competitive in a targeted market, marketing representatives must decide which product features to propose in which way to meet customers' expectations within geographical markets. Product trends do also converge over time as they are dynamic developments and not solely static points in time. It is therefore essential that the tool (DR5) can flexibly match aspect-sentiment relations to different geographical (e.g., continents, countries, federal states) and temporal (e.g., days of a week, phases of a day) parameters [3, 5, 17–20]. The huge volume of available social media posts requires the incorporated techniques (DR6) to deal with vast amounts of textual data [8, 9, 18–20]. As past developments of trends are essential for assessing the current state of trending topics, the tool (DR7) needs to allow the user to consider historical data [3, 17–19]. To support Product Development in prioritizing product improvement decisions, the tool (DR8) needs to output aggregated sentiment values for the identified aspects [3, 6, 8, 19, 20] and (DR9) illustrate the frequency of the identified aspects [3, 8, 20, 21]. To immediately identify the most important aspect-sentiment relations, the tool (DR10) should provide means to rank the results in either descending or ascending order [3, 21].

In the second step, we searched the market for available trend analysis tools. We took an up-close look at the most popular tools (e.g., Brandwatch, Meltwater, Symanto) that offer trend analysis by means of aspect-based sentiments. We thoroughly analyzed

the functionalities by installing and applying demo versions of these tools. The analysis process offered within the demo versions and the provided diagrams and charts that are offered to illustrate the results enabled us to apply the DRs to the existing tools, and to assess them both conceptually and visually. Based on our observations, we could conclude that there is no software tool that meets all the specific requirements for the use case Product Development, as we could derive them from extant literature. While we could agree that existing tools are able to meet most of the requirements, there are nevertheless remarkable drawbacks regarding DR1, DR2 and DR5. To confirm these observations, we subsequently turned to sales representatives from these companies. As it turned out, there are trend analysis tools that indeed provide pre-defined possibilities to contextualize the identified aspect-sentiment relations (e.g., time within a day or continents). However, the ability to flexibly match further temporal (e.g., phases of a day) and geographical parameters (e.g., countries or federal states) to aspect-sentiment relations is missing. Nonetheless, perceptions of product features may differ across countries or federal states and trends are temporal developments (e.g., days of a week, phases of a day). As reaching out to the sales representatives also could confirm, existing trend analysis tools that apply aspect-based sentiment analysis extract aspects, either with or without incorporating prior knowledge. Literature unveils the need for a comprehensive trend analysis tool that meets all the requirements for the use case Product Development. With this research, we aim to close this gap.

3 Research Procedure

In order to systematically develop an artifact for the automated trend analysis in marketing, we followed the Design Science (DS) approach [22, 23] and aligned our research activities with the DS procedure as proposed by [23] (see Fig. 1).

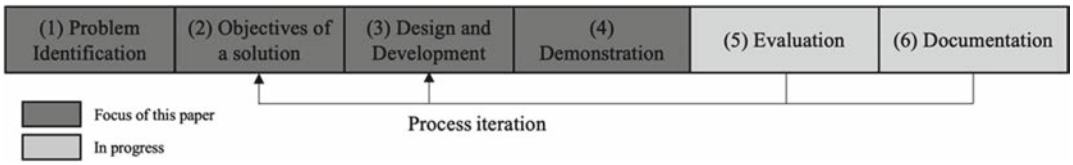

Fig. 1. Design science research (DSR) procedure

As a first step, (1) corresponding problems and drawbacks of previously existing approaches regarding the automated trend analysis using aspect-based sentiment analysis were identified (see Sects. 1 and 2.2). Hence, the revised tools supporting trend analysis by aspect-based sentiment detection do not meet the requirements that are indispensable to the successful application. Consequently, our (2) objective is to address drawbacks of existing software tools by suggesting a comprehensive artifact for automated trend analysis that allows marketing representatives to conduct aspect-based sentiment analysis (see Sects. 2.2 and 4). The third step of our DS process model contains the (3) design and development (see Sect. 4) of an artifact. To fill the gaps identified within phase (1), we focus on the design of the technical realization of the tool by combining different machine learning techniques, following our derived DPs. Thus, our approach

was established to support the trend analysis and to eliminate the existing disadvantages. By (4) demonstrating our artifact (see Sect. 5.1), we highlight the application of our tool on 37,638 Yelp reviews [24]. Thus, we showed the implementation of the requirements identified in literature. In Step 5 the usefulness, applicability and usability of the tool are to be analyzed in a larger field study. Finally, the tool will be further enhanced before it is provided to marketing departments of large companies (6).

4 Design and Development

First, the composition of Meta Requirements (MRs) that describe "*what the system is for*" ([26], p. 325) is based on the purpose and scope of the tool that was discussed in the motivation. Thus, we define the solution objectives based on the investigations´ problems and present them in Fig. 2. Besides the MRs, the Design Principles (DPs) are synthesized in a next step. DPs are defined as prescriptive statements that show how to do something to achieve a goal [27]. These DPs are deduced from the design requirements (DRs) that are further influenced by prior theories and current research literature [25]. Gregor and Jones (2007) state that this foundation in form of theories disclose "an explanation of why an artifact is constructed as it is and why it works" (p. 328). Thus, these DR offer guidance by designing the artifact and advise the DPs [26]. So, we derived the DRs from prior research literature to develop supportive DPs a priori to any instantiation of our artifact. These principles refer to at least one requirement and serve as an abstract "blueprint" of our artifact [26, 27]. By establishing these design principles, we made sure that they follow the value grounding (reference to the requirement) and the explanatory grounding (design principles are based on the current literature) [27, 28]. The DPs we derived fall into the category of "action and materiality-oriented design principles", describing what an artifact should enable users to do and how the artifact should be built to do so [28]. The development of the DPs follows the guidelines of [28] and [27]. Furthermore, we take one step further and append another layer by including the design features respectively the implementation of the DPs [28].

Since no particular machine learning technique is capable of accurately representing all DRs, a combination of them was essential. Regarding *DP1*, the autonomous extraction of aspects, unsupervised techniques (e.g., topic modeling) are required to enable an explorative analysis without prior domain knowledge. However, as stated by [29] the potential of totally unsupervised techniques is stymied by their purely unsupervised nature. Thus, semi-supervised techniques have arisen, facilitating an effective way to guide the analysis specific to a user by manipulating the analysis process even without structured prior domain knowledge [30]. Therefore, to take advantages of semi-supervised techniques while maintaining the flexibility of unsupervised ones, the known semi-supervised topic modeling technique GuidedLDA found application as it achieves convincing analysis results [30]. Besides the explorative analysis, the artifact must provide the ability to incorporate prior domain knowledge (*DP2*). Thus, the artifact provides a supervised aspect extraction using deep learning. Specifically, it applies a convolutional neural network (CNN) as proposed by [31], using two types of pre-trained embeddings for the aspect extraction: a general-purpose embedding and a domain specific embedding, containing domain related information used by the CNN

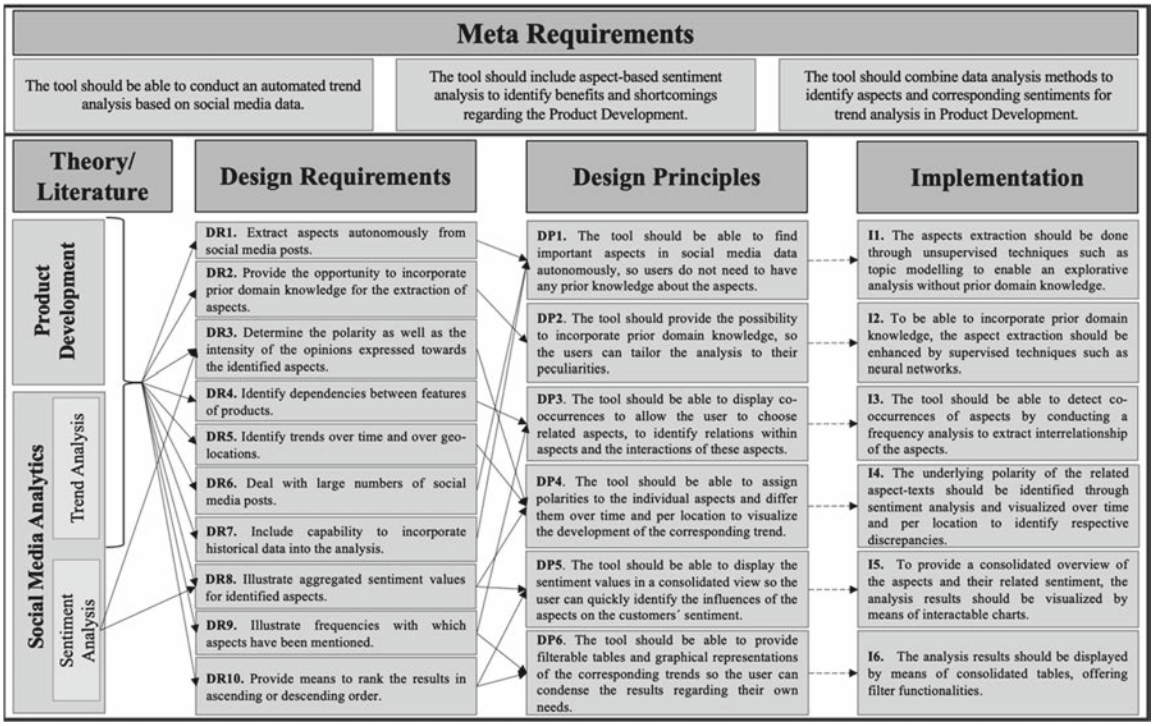

Fig. 2. Design of the artifact

to learn the domain specific peculiarities. Hence, the analysis can easily be adapted and tailored by changing the underlying domain embedding, resulting in a highly generic and customizable artifact. With respect to *DP3*, the artifact must depict co-occurrences of the related aspects to detect their interrelationship. As probabilistic topic modeling techniques such as GuidedLDA infer the resulting topics based on various probabilistic distributions, depicting the relations of the underlying topic words (and thus the resulting aspects) [32], the identification of their interrelationships is met through the nature of topic modeling itself. Considering the CNN, the artifact depicts the co-occurrences of the aspects by conducting a frequency analysis. Here, sub-aspects are identified for each extracted (main-)aspect by analyzing their respective occurrence in the context of the related main-aspect, resulting in an n-dimensional occurrence-tree. To determine the tonality of each aspect (*DP4*), the "Valence Aware Dictionary for sEntiment Reasoning" (VADER) [33] technique (a lexicon and rule-based sentiment analysis technique specifically attuned to sentiments expressed in social media) has been implemented. To further consolidate and visualize (*DP5*) the underlying aspects' sentiment as well as its evolution over time, corresponding line-charts are implemented using the python library matplotlib [34]. Finally, to ensure an adequate illustration of the analysis results (*DP6*), the extracted aspects, the corresponding sentiment values and their means, but also specific references to the extracted aspects are displayed using filterable tables and lists as demonstrated in the following section.

5 Demonstration and Discussion of the Artifact

5.1 Demonstration of the Artifact

To examine the tool's ability to identify meaningful and sound trends (including related aspects and their sentiment), we applied it to a real-world dataset. Therefore, we consulted the academic Yelp dataset [24], represented by a subset of real-world reviews and businesses in the US. The dataset comprises ~8.6 million OCR concerning 160,585 businesses and spans the period from October 13th, 2004 to January 28th, 2021. To demonstrate our tool, we narrowed the analysis to the reviews of a fast-food restaurant with multiple franchises in various locations to extract the relevant aspects and associated customer perceptions, resulting in 37,638 reviews.

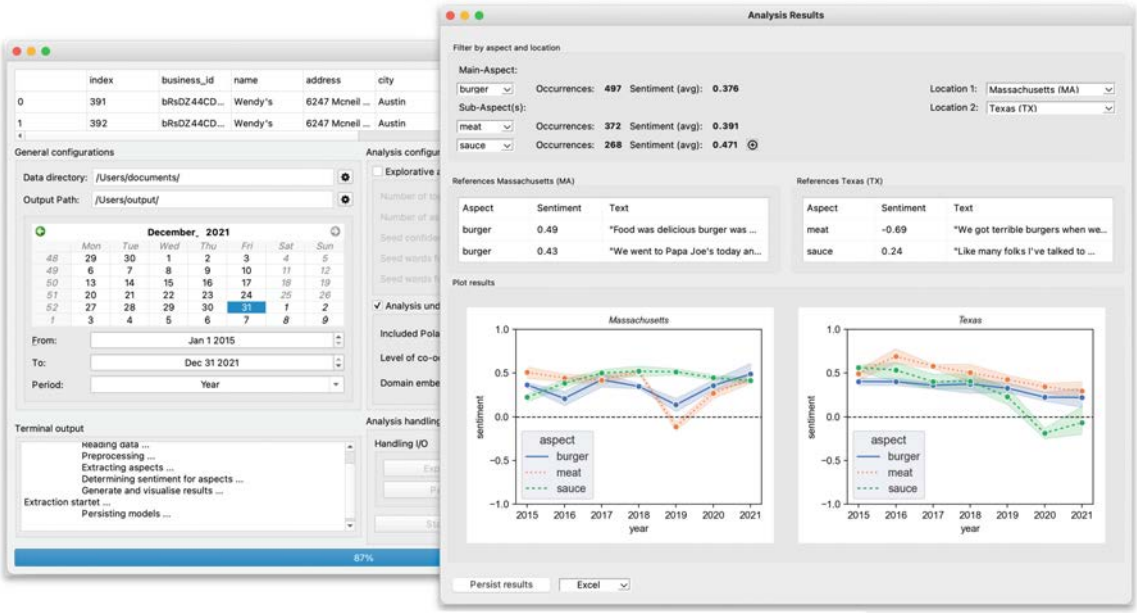

Fig. 3. Configuration and result view of the artifact

Figure 3 represents the tool's configuration view (left) and results view (right). By the configuration view, the underlying analysis can be customized to the one's own needs. For the purposes of our demonstration, the analysis was conducted using data spanning the years 2015 to 2021, prior domain knowledge to extract the aspects and all sentiment levels (*positive, neutral, negative*). The domain knowledge used was extracted from the renowned dataset of the 2016 SemEval task [35]. By using this word embedding tailored to the use case of restaurant reviews, the CNN is trained in the domain of the food and restaurant industry. The level of co-occurrences has been set to two, resulting in a two-dimensional occurrence-tree. The result view represents the sentiments of the extracted aspects with respect to the years, resulting in the monitoring of their evolution based on their customer perceptions.

Table 1 presents a detailed excerpt of the resulting Excel file, containing the elicited sentiment values per aspect and year. Here, both locations refer to the same aspects (main-aspect: burger; sub-aspects: meat, sauce), facilitating a comparison of temporal

Table 1. Analysis results - sentiments per year and location

Aspect	Location	Sentiment per Year						
		2015	2016	2017	2018	2019	2020	2021
Burger	*MA*	0.373	0.214	0.435	0.381	0.131	0.353	0.459
	TX	0.413	0.40	0.378	0.419	0.312	0.216	0.222
Meat	*MA*	0.515	0.447	0.421	0.521	−0.109	0.291	0.422
	TX	0.495	0.682	0.55	0.514	0.433	0.348	0.283
Sauce	*MA*	0.241	0.382	0.509	0.524	0.503	0.427	0.414
	TX	0.582	0.588	0.39	0.416	0.212	−0.186	0.12

and location-based discrepancies, thus enabling conclusions about different customer perceptions. In Massachusetts (MA), the main-aspect *burger* exhibits slight oscillation across the years but generally remains stable (see Fig. 3, Table 1). The worst average customer perception occurs in 2019 and is represented by a slightly positive sentiment score of 0.131. Furthermore, the two sub-aspects *meat* and *sauce* differ strongly in their course in 2019. Here, the course of *meat* collapses drastically (−0.109), while the course of *sauce* remains nearly constant (0.503) compared to the previous year, leading to the assumption that the negative reflections of *meat* may influence the main-aspect *burger*. Moreover, this assumption is supported by the development of the respective aspects. Here it becomes apparent that the significant improvement of the *meat* also potentially causes an improvement of the *burger*, supporting the conclusion that our tool can identify meaningful and sound trends based on the corresponding aspects and customer perceptions. Moreover, our tool has revealed discrepancies across the respective locations, as the perception of *sauce* overall has remained quite constant in MA but consistently decreases in Texas (TX) (see Fig. 3, Table 1).

Generally, the development of our tool was based on the DPs (see Fig. 2). Accordingly, all expectations posed were technically realized. In addition, several trends and their evolution could be identified by applying our tool to a real-world dataset, validating its functionality. To subsequently evaluate its practical applicability by identifying meaningful and sound trends, an evaluation will be conducted in future work.

5.2 Discussion of the Demonstration

The implementation of the DPs enabled us to design and develop a tool which has provided interesting results. As we establish the opportunity to extract the aspects in either a supervised or an unsupervised way (*DP1, DP2*), we can identify in the first instance aspects and/or product features which are discussed in the OCR. This allows us to identify those aspects which are most important from the customer's point of view and, above all, which must be considered in Product Development [3, 8, 20, 21]. Figure 4 shows that the most frequently named aspects are *burger, meat* and *sauce* (497, 372 and 268 occurrences). In the case of Product Development, the restaurant can start screening these aspects as they are particularly important for the customers. The

extraction of aspects is possible on the one hand via a CNN including word embedding, providing the restaurant to incorporate prior domain knowledge (*DP2*) such as aspects about food, drinks, and processes in the restaurant. So, the extraction of the aspects is tailored precisely to the company. But, on the other hand, if a restaurant wants to extract aspects without exerting any influence, the tool can also identify them autonomously.

Moreover, the demonstration of the artifact has also shown that the customers have spoken differently about the three aspects. Here, our results show that users spoke about *sauce* more positively (0.471) over the years than about the aspects *burger* (0.376) and *meat* (0.391). Thus, numerous opportunities for improving products can be identified, and certain features can be given greater importance [5, 19]. Figure 4 shows that for MA the values of the individual aspects can change considerably: While in 2015–2018 the sentiment values of *meat* range between 0.40 and 0.52, in 2019 it slipped down into the negative range with −0.12. Simultaneously, the sentiment value of *burger* has fallen too. In this context, it is therefore possible that the negative sentiment about the *meat* had also influenced the customers' opinion about *burger*. If a company did not have this fine-grained information gained through the aspect-based sentiment analysis and wanted to adjust the product in 2019 based on the negative reviews, they potentially would have changed product features which were actually rated positively. In TX, considering these aspects is inevitable as the results show that *sauce* needs to be changed so the customers' opinions about it and thus of the whole burger can be improved again. Already existing trend analysis tools often include sentiment analysis to show the general tonality about the company or the product over time – without considering that different aspects influence the product's evaluation. However, this leads to a distortion of the results and the benefit for companies is no longer a given.

Furthermore, including geolocations and therefore considering location-based discrepancies can also create significant benefits. To be competitive in a targeted market, companies need to know which features should be designed in which way to meet the local customers' expectations [17, 36]. This becomes particularly evident when comparing MA and TX (see Fig. 4). While in MA the *burger* exhibits a positively connotated trend over the years, the sentiment over time in TX shows worse values. Especially in 2020 and 2021, the sentiment of the *burger* differs immensely (MA: 0.35, 0.45; TX: 0.28, 0.22). In both locations, there are different reasons for the burger's better (MA) or worse (TX) rating. While in TX the *sauce* must be adjusted to the taste of the people, this is not necessary in MA due to the steady positive values. Without including geolocations and also sentiment values, these discrepancies would not have been revealed, which clearly is a benefit in comparison to other existing tools.

6 Conclusion, Contribution and Outlook

Assessing and identifying people's opinions about a particular aspect and its future impact (thus a social media trend), is difficult, especially given the vast amount of social media data. Thus, techniques for analyzing textual social media data, e.g., aspect-based sentiment analysis, topic modeling and neural networks, have gained in importance, as companies need to be aware of customers' expectations regarding products. However, prior literature and existing tools do not include external parameters and do not

cover specific requirements crucial in the field of Product Development (see Sect. 2.2). However, the early identification of new and auspicious ideas and trends means a competitive advantage for companies [19]. Thus, we provide a comprehensive tool by combining several machine learning techniques and transfer them in a highly responsive and platform-independent GUI. By demonstrating our tool on 37,638 Yelp OCR we have shown that considering all identified requirements is necessary to analyze trends.

Besides creating value for practitioners, theoretical contributions in the research area of IS are also provided. To acknowledge the importance of different DSR perspectives, we have related our DSR contribution to the category "design artifacts" according to [37], including both: the demonstration of the artifacts practical benefit and design theory contributions [38]. Therefore, by providing a tool for automated trend analysis that can identify aspects that are discussed within social media, we enable companies to gain deep insights into customers' current opinions and future expectations to tailor their products. Hence, to meet these customers' needs, a company has to identify and continuously track product features by incorporating the VoC into internal Product Development processes. Thus, as tracking evolving and changing customer requirements is imperative to meet customers' wishes [36, 39], companies can respond to them quickly and with minimal effort. Compared to existing trend analysis tools, our tool meets all the specific requirements set out within the literature regarding Product Development. In particular, our tool can flexibly match different temporal and geographical parameters to identify aspect-sentiment relations and it provides users the possibility to extract the aspects either with or without prior domain knowledge. Thus, customer perceptions for specific periods or geolocations can be displayed to track fine-grained variabilities. On the one hand, this makes it possible to visualize influences affecting the sentiment. On the other hand, downward trends can be counteracted and upward trends can be strengthened. Hence, this combined with the integration of geolocations can especially support large companies with multiple branches in their efforts to easily perceive location-specific sentiment changes and explicitly react to them. As we have combined different techniques and designed our tool modularly, companies can adapt the analysis to their specific needs. A further contribution is that the tool can be supportive in identifying the aspects of the products to be changed to meet the customer's expectations (remove existing aspects, others need to be enhanced) with the realization of (*DP4*)–(*DP6*). In summary, companies can benefit from our comprehensive and modular artifact by analyzing large amounts of data in a way best suited to their circumstances, aiming to analyze trends regarding their aspect-based sentiment values.

Besides our technical contribution (i.e., the artifact), we achieved prescriptive theoretical contributions as a further outcome of our DSR project. Therefore, we formulated and proposed DPs based on the DRs derived from current research literature. By applying them an implicit empirical grounding of the DPs was achieved. Our DPs capture design-related knowledge and can therefore support the development of further IS (design) theories and new artifacts. For designing further (trend analysis) tools in related areas, our DPs can be applied as we have formulated them in a prescriptive and abstract manner by generally describing what the artifact should enable users to do and how it should be built. Furthermore, the formulation of DPs allows us to abstract away from individual settings and DRs, and rather align with generalizing prescriptive knowledge whereby

they contribute to "nascent design theory" [38, 40]. For example, by DP4, the importance of including time, geolocation and/or further external parameters (e.g., customers' characteristics) in a trend analysis tool is highlighted. As these external parameters have a direct impact on the customers' sentiment and therefore on the analysis results (cf. [41]), the alignment to them will lead to a more targeted trend analysis tool. Thus, for researchers that intend to design such a tool, we suggest considering the influences that are evoked by external factors. So, with the compilation of the DPs, we made a first step towards contributing to nascent design theory. To take a next step towards a more mature design theory, we intend to verify our DPs by further evaluating our tool. Therefore, we will first evaluate our tool in a formative and artificial environment (i.e., a laboratory experiment). Here, participants will use it to identify relevant aspects and associated customer perceptions. Subsequently, they will complete a questionnaire to indicate their perceptions of the analysis quality and tool usability. This allows us to identify difficulties and improve our tool (whereby our DPs can be confirmed or adapted) before conducting a more elaborate evaluation in a more natural setting. This second evaluation will be a field study with a large restaurant chain that plans to integrate a tool to support its marketing departments.

However, there are also some limitations to this research: Although we included a large set of investigations, we could identify probably even more requirements the tool should meet in further literature.

References

1. Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of Social Media. Bus. Horiz. **53**(1), 59–68 (2010)
2. Kim, Y.K., Lee, D., Lee, J., Lee, J.-H., Straub, D.W.: Influential users in social network services: the contingent value of connecting user status and brokerage. ACM SIGMIS: The Database Adv. Inform. Syst. **49**(1), 13–32 (2018)
3. Tuarob, S., Tucker, C.S.: Quantifying product favorability and extracting notable product features using large scale social media data. J. Comput. Inf. Sci. Eng. **15**(3), 031003 (2015)
4. Schouten, K., Frasincar, F.: Survey on aspect-level sentiment analysis. IEEE Trans. Knowl. Data Eng. **28**(3), 813–830 (2015)
5. Tucker, C.S., Kim, H.: Predicting emerging product design trend by mining publicly available customer review data. In: 18th International Conference on Engineering Design (2011)
6. Mirtalaie, M.A., Hussain, O.K., Chang, E., Hussain, F.K.: Sentiment analysis of specific product's features using product tree for application in new product development. In: Barolli, L., Woungang, I., Hussain, O.K. (eds.) INCoS 2017. LNDECT, vol. 8, pp. 82–95. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-65636-6_8
7. Mirtalaie, M.A., Hussain, O.K., Chang, E., Hussain, F.K.: Extracting sentiment knowledge from pros/cons product reviews: discovering features along with the polarity strength of their associated opinions. Expert Syst. Appl. **114**, 267–288 (2018)
8. Mirtalaie, M., Hussain, O.: Sentiment aggregation of targeted features by capturing their dependencies: making sense from customer reviews. Int. J. Inform. Manag. **53**, 102097 (2020)
9. Vo, A.-D., Nguyen, Q.-P., Ock, C.-Y.: Opinion–aspect relations in cognizing customer feelings via reviews. IEEE Access **6**, 5415–5426 (2018)
10. Hicks, A., Comp, S., Horovitz, J., Hovarter, M., Miki, M., Bevan, J.L.: Why people use Yelp.com: an exploration of uses and gratifications. Comput. Human Behav. **28**(6), 2274–2279 (2012)

11. Yan, Z., Xing, M., Zhang, D., Ma, B., Wang, T.: A context-dependent sentiment analysis of online product reviews based on dependency relationships. In: 35th ICIS (2014)
12. Ye, H.J., Kankanhalli, A.: User service innovation on mobile phone platforms: investigating impacts of lead userness, toolkit support, and design autonomy. MIS Quart. **42**(1), 165–188 (2018)
13. Axtell, C.M., Holman, D.J., Unsworth, K.L., Wall, T.D., Waterson, P.E., Harrington, E.: Shopfloor innovation: facilitating the suggestion and implementation of ideas. J. Occup. Organ. Psychol. **73**(3), 265–285 (2000)
14. Blei, D.M., Ng, A.Y., Jordan, M.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
15. Liu, B.: Sentiment analysis and opinion mining. In: Synthesis Lectures on Human Language Technologies, vol. 5, 1st edn. Morgan & Claypool Publisher (2012)
16. Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., Cleven, A.: Standing on the shoulders of giants. Commun. Assoc. Inform. Syst. **37**(1), 205–224 (2015)
17. Han, Y., Moghaddam, M.: Analysis of sentiment expressions for user-centered design. Expert Syst. Appl. **171**, 114604 (2021)
18. Cheng, L.-C., Chen, K., Lee, M.-C., Li, K.-M.: User-defined SWOT analysis – a change mining perspective on user-generated content. Inform. Process. Manag. **58**(5), 102613 (2021)
19. Jeong, B., Yoon, J., Lee, J.-M.: Social media mining for product planning: a product opportunity mining approach based on topic modeling and sentiment analysis. Int. J. Inform. Manag. **48**, 280–290 (2019)
20. Zhang, J., Zhang, A., Liu, D., Bian, Y.: Customer preferences extraction for air purifiers based on fine-grained sentiment analysis of online reviews. Knowledge-Based Syst. **228**, 107259 (2021)
21. Yu, J., Zha, Z.-J., Wang, M., Chua, T.-S.: Aspect ranking: identifying important product aspects from online consumer reviews. In: 49th Annual Meeting of the ACL, pp. 1496–1505 (2011)
22. Hevner, M., Park, R.: Design science in information systems research. MIS Quart. **28**(1), 75–105 (2004)
23. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inform. Syst. **24**, 45–77 (2007)
24. Yelp open dataset (2021). https://www.yelp.com/dataset, Accessed on 15 Dec 2021
25. Purao, S., Kruse, L.C., Maedche, A.: The origins of design principles: where do… they all come from? In: Hofmann, S., Müller, O., Rossi, M. (eds.) DESRIST 2020. LNCS, vol. 12388, pp. 183–194. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64823-7_17
26. Jones, D., Gregor, S.: The anatomy of a design theory. J. Assoc. Inform. Syst. **8**(5), 312–335 (2007)
27. Heinrich, P., Schwabe, G.: Communicating nascent design theories on innovative information systems through multi-grounded design principles. In: Tremblay, M.C., VanderMeer, D., Rothenberger, M., Gupta, A., Yoon, V. (eds.) DESRIST 2014. LNCS, vol. 8463, pp. 148–163. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06701-8_10
28. Möller, F., Guggenberger, T.M., Otto, B.: Towards a method for design principle development in information systems. In: Hofmann, S., Müller, O., Rossi, M. (eds.) DESRIST 2020. LNCS, vol. 12388, pp. 208–220. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64823-7_20
29. Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J.L., Blei, D.M.: Reading tea leaves: How humans interpret topic models. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A. (eds.) Advances in neural information processing systems, pp. 288–296. Curran Associates, Inc. (2009)
30. Jagarlamudi, J., Daumé III, H., Udupa, R.: Incorporating lexical priors into topic models. In: 13th Conference of the European Chapter of the ACL, pp. 204–213 (2012)

31. Xu, H., Liu, B., Shu, L., Philip, S.: Double embeddings and CNN-based sequence labeling for aspect extraction. In: 56th Annual Meeting of the ACL, pp. 592–598 (2018)
32. Crain, S.P., Zhou, K., Yang, S.-H., Zha, H.: Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. In: Aggarwal, C.C., Zhai, C.X. (eds.) Mining Text Data, pp. 129–161. Springer US, Boston, MA (2012). https://doi.org/10.1007/978-1-4614-3223-4_5
33. Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: International AAAI Conference on Web and Social Media (2014)
34. Hunter, J.: Matplotlib: a 2D graphics environment. Comput. Sci. Eng **9**(3), 90–95 (2007)
35. Pontiki, M., et al.: Semeval-2016 task 5: aspect based sentiment analysis. In: International Workshop on Semantic Evaluation, pp. 19–30 (2016)
36. Lozano, M.G., Schreiber, J., Brynielsson, J.: Tracking geographical locations using a geo-aware topic model for analyzing social media data. Decis. Support Syst. **99**, 18–29 (2017)
37. Baskerville, R., Baiyere, A., Gergor, S., Hevner, A., Rossi, M.: Design science research contributions: finding a balance between artifact and theory. J. Assoc. Inform. Syst. **19**(5), 358–376 (2018)
38. Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. MIS Quart. **37**(2), 337–355 (2013)
39. Hong, L., Ahmed, A., Gurumurthy, S., Smola, A.J., Tsioutsiouliklis, K.: Discovering geographical topics in the twitter stream. In: 21st International Conference on WWW, pp. 769–778 (2012)
40. Gregor, S., Kruse, L., Seidel, S.: Research perspectives: the anatomy of a design principle. J. Assoc. Inform. Syst. **21**, 1622–1652 (2020)
41. Konadl, D., Wörner, J., Leist, S.: Identifying sentiment influences provoked by context factors–results from a data analytics procedure performed on tweets. In: 54th HICSS (2021)

2.5. **Beitrag 5:** Artificial Intelligence in Augmentative and Alternative Communication Systems – A Literature-Based Assessment and Implications of Different Conversation Phases and Contexts

Adressierte Forschungsfragen	**(1)** Inwieweit werden durch die in der Literatur für AAC-Systeme vorgeschlagenen KI-Ansätze für sprachbehinderte Menschen Potenziale in den Gesprächsphasen „Gesprächseröffnung“, „Gesprächsmitte“ und „Gesprächsabschluss“ sowie für die Gesprächskontexte „Formell“ und „Informell“ realisiert? **(2)** Was sind darauf basierend mögliche Implikationen für die zukünftige Forschung zu KI-Ansätzen im Hinblick auf diese Gesprächsphasen und Gesprächskontexte?
Zielstellungen	• Konsolidierung und Strukturierung der KI-Ansätze, welche im Hinblick auf AAC-Systeme in der wissenschaftlichen Literatur zur Unterstützung von sprachbehinderten Menschen untersucht worden sind. • Ableitung von entsprechenden KI-Potenzialen und Identifikation des Abdeckungsgrades dieser KI-Potenziale in ausgewählten, zentralen Gesprächsphasen und Gesprächskontexten.
Forschungsmethode	**Systematischer Literature Review** • Anlehnung des Literature Reviews an die methodischen Vorgehensweisen nach Cooper (1988), Cooper und Hedges (1994), Vom Brocke et al. (2015) sowie Webster und Watson (2002). • Beschreibung der Literatursuche, des Reduktionsvorgangs und der Literaturauswahl. • Literaturanalyse unter Anwendung der qualitativen Inhaltsanalyse nach Mayring (2004): Ableitung von KI-Potenzialen, Ableitung von Zuordnungskriterien der KI-Potenziale zu den Gesprächsphasen und zu den Gesprächskontexten und Identifikation von Forschungslücken mithilfe einer Konzeptmatrix.
Kernergebnisse (Überblick)	• Ableitung von insgesamt vier Kategorien von Potenzialen von KI-Ansätzen im Rahmen von AAC-Systemen: (a) Aufrechterhaltung des Gesprächsverlaufs, (b) Einbezug von natürlichen Kontextfaktoren, (c) Anpassung gegenüber dem Gesprächspartner und (d) Anpassung gegenüber dem Nutzer (= sprachbehinderte Person) des AAC-Systems. • Strukturierte Übersicht über die Abdeckung der KI-Potenziale in zentralen Gesprächsphasen (Gesprächseröffnung, Gesprächsmitte und Gesprächsabschluss) und Gesprächskontexten (formelle und informelle Gesprächssituationen). • Aufzeigen des Forschungsbedarfs hinsichtlich der in den verschiedenen Gesprächsphasen und Gesprächskontexten eingesetzten KI-Ansätze. Es wurden hierbei weiterhin mögliche Ansatzpunkte zur Schließung der festgestellten Forschungslücken abgeleitet und diskutiert.
Publikationsort	**Veröffentlicht** in den Proceedings der 31. European Conference on Information Systems, Kristiansand, Norwegen.
Ranking VHB JQ 4	**A**
Autoren und Anteile	**Daniel Konadl** **50 %** Janik Wörner 30 % Lucas Luttner 10 % Susanne Leist 10 %

Tabelle 7: Fact Sheet zu Beitrag 5

ARTIFICIAL INTELLIGENCE IN AUGMENTATIVE AND ALTERNATIVE COMMUNICATION SYSTEMS - A LITERATURE-BASED ASSESSMENT AND IMPLICATIONS OF DIFFERENT CONVERSATION PHASES AND CONTEXTS

Research Paper

Daniel Konadl, University of Regensburg, Regensburg, Germany, Daniel.Konadl@ur.de

Janik Wörner, University of Regensburg, Regensburg, Germany, Janik.Woerner@ur.de

Lucas Luttner, University of Regensburg, Regensburg, Germany, Lucas.Luttner@ur.de

Susanne Leist, University of Regensburg, Regensburg, Germany, Susanne.Leist@ur.de

Abstract

Even though AAC systems and corresponding AI approaches have been investigated within the extant research, they still show remarkable drawbacks, resulting in a low prevalence among speech-impaired individuals. As the suggestions and adaptions proposed by AI within AAC systems may show insufficiencies in certain situations (e.g., unreliable suggestions, low conversational rates, unauthentic adaptions towards users), we take an up-close look, especially at the conversation phases and contexts in which the supporting AI is applied. Therefore, we have conducted a systematic literature review as well as a literature analysis. Thereby, we could reveal that there are indeed several gaps within the extant research on AI regarding the coverage of the conversational context "informativeness" and the conversation phases "beginning" and "closing". To dismantle the existing communication barriers that speech-impaired individuals suffer from, several implications for investigating AI in the context of AAC systems are derived and proposed for future (IS) research.

Keywords: Augmentative and Alternative Communication Systems, Artificial Intelligence, Conversation Phases, Conversational Contexts, Systematic Literature Review.

1 Motivation

"Perhaps the single quality most central to humanness is the ability to exchange thoughts, ideas, and feelings with others" (Hourcade et al., 2004, p. 235). In this context, speech is seen as an individual's most important instrument to be in contact with its human surroundings (cf. Fritzell, 1996; Kane et al., 2017; Pollak and Gallagher, 1989). However, individuals that suffer from severe physical disabilities or brain injury are not able to control their oral-respiratory musculature sufficiently for speech (Allen, 2005; Vanderheiden, 1983). This restriction of verbal communication, and hence, the "separation from the mainstream of society" (Hourcade et al., 2004, p. 235) holds true for people with damages to the vocal tract or other impairments (e.g., aphasia, autism) affecting speech as well (Allen, 2005).

To get in exchange with their surroundings, speech-impaired individuals need to apply Augmentative and Alternative Communication (AAC) which combines strategies, symbols, and techniques to promote goal-directed communication (American Speech Language Hearing Association, 1991). AAC comprises different analogue forms of communication (i.a., sign language, picture boards). In the context of sign language, expressions are formed using mimics and gestures. But these expressions can easily be overlooked or misinterpreted when the interlocutors have little or no experience in sign language (Boyes-Braem et al., 1994). Compared to vocal language, sign language lacks expressiveness because it provides

only a few prepositions and conjunctions. Furthermore, it contains no genus markers to differentiate between masculinum and femininum for nouns (Louis-Nouvertné, 2001).

To give individuals with speech impairments a voice and enable them to express themselves using vocal language, digital AAC solutions have emerged that enable the translation of an individual's intended meaning into speech and subsequent voice outputs (cf. Bradshaw, 2013). The positive influence of these voice-generating AAC systems on the interaction behaviour of speech-impaired individuals has already been proven in numerous studies (e.g., Brady, 2000; Desai et al., 2014; Schepis et al., 1996). Despite the positive influence and although these AAC systems manage to meet several design requirements *(see Section 2.2)*, they still show drawbacks that research has attempted to address by various Artificial Intelligence (AI) approaches (e.g., Dempster et al., 2010; Neamtu et al., 2019; Obiorah et al., 2021). While it is undeniable that these AI approaches could increase the practicability of AAC systems (e.g., Klauer et al., 2021; Laxmidas et al., 2021; Obiorah et al., 2021), these systems nevertheless have only a low prevalence among the target group as they still face different problems.

For example, AAC systems propose only low conversation rates as the suggestions generated, and the utterances composed by AI often do not fit the concrete situation in which the AAC user takes part. As a consequence of unreliable suggestions, speech-impaired individuals have to exert effort to produce an utterance or have to adjust the suggestions to generate an appropriate contribution to the conversation (e.g., Dempster et al., 2010; Laxmidas et al., 2021; Obiorah et al., 2021). The possibilities AI approaches propose concerning personalisation towards the user of the AAC system show limitations as well (e.g., within situation-depending adaptions of the speech output tonality). The adaptions proposed by the AI are often not perceived as authentic by the conversation participants given a certain situation because certain situational factors (e.g., time progression and changes of involvement and attitude towards a conversational topic) are not incorporated into the adaption (e.g., Ascari et al., 2020; Mills et al., 2014; Shen et al., 2022). Because the suggestions and adaptions proposed by AI approaches within AAC systems show insufficiencies in certain situations, this suggests that the AI approaches in particular need to be investigated and adapted concerning the conversations in which they are applied. Indeed, conversations are performed in different conversation phases ("beginning", "middle" and "closing" of a conversation) and within different contexts (e.g., formal conversations within meetings at the workplace vs. informal conversations with friends within spare-time activities).

So, while the beginning of a conversation with an acquaintance in an informal context does not require the assessment of the conversation partner (i.a., personality, attitudes, conversational behaviour), the conversational beginning with a business customer in a formal negotiation within a meeting requires the assessment of him or her. In the same vein, while the beginning of a conversation is comparably standardised (cf. Henne and Rehbock, 2012; Spiegel and Spranz-Fogasy, 2001), the middle of conversations may be more complex as the involved conversation participants may aim at diverging conversational goals, requiring mutual and coordinated conversational acting until a targeted and mutually accepted state of the conversation is reached. As can be concluded, the different conversation phases and conversational contexts consequently lead to different demands and requirements of a supporting AI approach that need to be met to comprehensively and optimally support a speech-impaired individual.

With the work at hand, we address these issues by conducting a systematic literature review (Cooper, 1988; Vom Brocke et al., 2015; Webster and Watson, 2002) to establish an initial basis for the underlying research topic and thus develop a deeper understanding of the research field of AI in the context of AAC systems as well as to identify existing research problems and gaps and justify the relevance of addressing them (Vom Brocke et al., 2015). In this context, we identify AI-based potentials in view of AAC systems and assign them to the phases "beginning", "middle" and "closing" (cf. Henne and Rehbock, 2012) of a conversation as well as to the two conversational contexts "socialness" and "informativeness" (cf. Bedrosian et al., 2003; Hoag et al., 2004; Todman and Alm, 1997). Against this background, the following two research questions (RQs) are posed: ***(RQ1)*** *Do the applications of Artificial Intelligence that are proposed within the extant research literature for AAC systems cover the conversation phases "beginning", "middle" and "closing" and the conversational contexts "socialness" and "informativeness" appropriately?* ***(RQ2)*** *What are possible implications for AI in the context of AAC systems regarding the consideration of conversation phases and conversational contexts?*

We contribute to the inclusion of speech-impaired individuals into societal life and accelerate the corresponding Information Systems (IS) research so that these individuals are proposed comprehensive and individualised AI-supported AAC. The remainder of this paper is structured as follows: Section 2 provides the theoretical background and the related work. In Section 3, we describe the procedure of the literature review and literature analysis. Afterwards, in Section 4, we present our results and discuss the reasonable implications of the research in Section 5. Finally, Section 6 draws an overall conclusion.

2 Conceptional Basics and Related Work

2.1 Conceptional Basics

To investigate the problems of current AI-based approaches within the context of AAC systems in more detail, we take an up-close look at the conversations in which AI is applied. In this context, extant communication research has investigated structures and components (i.a., communication activities, contents, sequences) within daily conversations. According to Henne and Rehbock (2012), a conversation can be structured into an opening, a middle and a closing part. Each of these conversation phases has specific goals and comprises individual communication activities the conversation participants need to perform (cf. Spiegel and Spranz-Fogasy, 2001). The opening part concerns, i.a., the recognition of incoming communication requests, the identification, assessment and greeting of the conversation partner as well as the introduction of the conversation topics (cf. Henne and Rehbock, 2012). The closing part deals with closing the conversation, thanking the conversation partners and farewelling them (cf. Schegloff and Sacks, 1973; Spiegel and Spranz-Fogasy, 2001). In comparison to the opening and closing part, the middle part aims to fulfil specific communication activities that are related to the conversation goal and the purpose framed within the opening part (Henne and Rehbock, 2012).

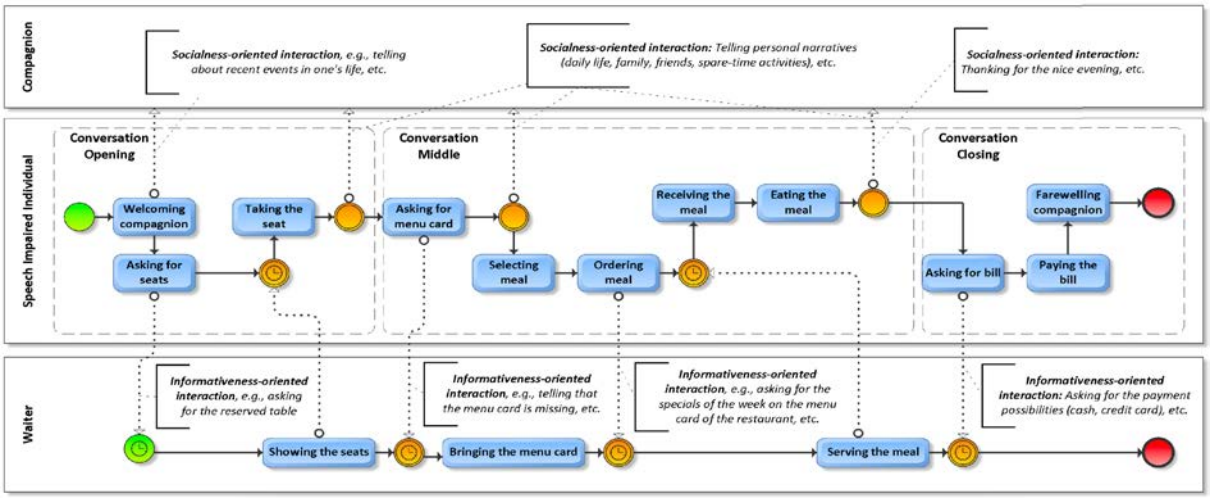

Figure 1. Illustration of different types of interactions (in particular "socialness-" and "informativeness"-oriented) in the example of visiting a restaurant with an acquaintance, own illustration related to Ehlich and Rehbein (1972) and Schindler (2013).

The conversation activities being performed within the respective conversation phases are further determined by different factors, i.a., the number of conversation participants involved (one-to-one conversations vs. group conversations) (Ehlich and Rehbein, 1972), the relationships that exist between the conversation partners (acquaintances vs. strangers) (Spiegel and Spranz-Fogasy, 2001) and the conversational context ("socialness"-oriented vs. "informativeness"-oriented conversations (cf. Bedrosian et al., 2003; Hoag et al., 2004; Todman and Alm, 1997)). Accordingly, group conversations require more complex activities to synchronise the conversation flow and the conversation utterances being exchanged. Further, interactions with friends or acquaintances, e.g., within a restaurant visit, are rather

"socialness"-oriented and informal while the interactions with the waiter at the restaurant are rather "informativeness"-oriented and formal (cf. Ehlich and Rehbein, 1972; Schindler, 2013). As a consequence, humans communicate differently and apply different articulations depending on whom they interact with *(see Figure 1)*. Nonetheless, speech-impaired individuals lack the fundamental capability to apply vocal language and therefore need to apply Augmentative and Alternative Communication.

Augmentative and Alternative Communication encapsulates strategies, techniques, and tools that support individuals with speech impairments in expressing their thoughts, desires, feelings, and ideas (Hourcade et al., 2004). In this context, AAC systems can be described as VOCAs (Voice Output Communication Aids) or SGDs (Speech Generating Devices) providing speech-impaired individuals with the ability to express themselves using vocal language. Currently available high-tech AAC systems encompass the use of electronic devices (e.g., smartphones, tablet PCs) and the devices' built-in peripherals (e.g., camera, microphone) (e.g., Baldassarri et al., 2014; Laxmidas et al., 2021). The board within AAC systems contains the vocabulary the users can employ. To contribute utterances to conversations, users of AAC systems can either relate to pre-stored sentences (e.g., Alm and Higginbotham, 2008), can compose single words within the so-called speech line (Klauer et al., 2021) or can make use of keyboard capabilities and a combination of letters to express words and sentences (e.g., Kristensson et al., 2020). The utterance formed and to be spoken towards the conversation partner is then processed by the speech synthesising module that generates the voice output that is subsequently emitted towards the conversation partner by the loudspeaker of the device. This voice output can in turn be interpreted and reciprocated by the conversation partner so that the conversation can progress using mutual and coordinated utterances until a targeted state of the conversation is reached.

2.2 Design Requirements of AAC Systems

To be purposeful for speech-impaired individuals, research has investigated several design requirements for AAC systems. There have been considered general requirements, such as the support of multiple languages (e.g., Guasch et al., 2019; Kane et al., 2017), availability for different operating systems (e.g., iPad OS, Android) and devices (e.g., smartphones, tablet PCs) (e.g., Baldassarri et al., 2014; Dattilo et al., 2008), reliability and confidentiality of the AAC system concerning generated communication data (e.g., Dattilo et al., 2008; McNaughton and Bryen, 2007). Beyond that, an AAC system needs to propose a core vocabulary that enables users to exchange needs, wants and thoughts within various daily life situations (cf. McNaughton and Bryen, 2007; Van Tilborg and Deckers, 2016). In this connection, the AAC systems must offer a dynamic expansion of the vocabulary (e.g., Van Tilborg and Deckers, 2016; Wang et al., 2018). To open up the variability of speech, the storage of synonymous outputs for respective the same utterance needs to be enabled (Alm and Higginbotham, 2008). For individuals hindered to edit the vocabulary on their own because of severe cognitive disorders accompanying their speech impairment, family members, caregivers, and nursing staff need to be given access to the AAC system (Baldassarri et al., 2014). Cognitive disorders also require assigning pictograms to the respective utterances to support the understanding of the concept related to an utterance (e.g., Guasch et al., 2019; Klauer et al., 2021). To support efficiency in selecting vocabulary, it is necessary to propose means to structure the utterances, e.g., with topic- and concept-specific folders or grids (e.g., Klauer et al., 2021; Wang et al., 2018). Further, there are design requirements for the individualisation of the AAC system towards the users' personalities. To customise the way an utterance is outputted, AAC systems propose the user means to adjust the speed and the volume of the voice output (e.g., via sliders in the GUI) (e.g., Arnott and Alm, 2013; Kane et al., 2017). In this vein, maintaining individual user profiles that include user-related preferences enables the customisation of the speech output (e.g., gendering the voice output, storing the user's personal life history) (Alm and Higginbotham, 2008; Guasch et al., 2019).

Although these several design requirements have been investigated within the extant literature, AAC systems still show drawbacks, i.a., low communication rates (cf. Arnott and Alm, 2013; Kane et al., 2017; Klauer et al., 2021) and low user- and situation-specific individualisation (e.g., Arnott and Alm, 2013; Kane et al., 2017; Murphy, 2004; Van Tilborg and Deckers, 2016). Therefore, research has started to investigate the potentials of AI approaches within AAC systems.

2.3 AI Approaches and AI-based Potentials in the Context of AAC Systems

Artificial intelligence relates to the ability of a machine to apply characteristics of human intelligence to solve specific problems and adapt to changing environments (Boden, 2018). The capabilities of AI have recently been incorporated within AAC systems, leading to advances in the support of speech-impaired individuals. Corresponding AI approaches consider for example the user's past selection data, including the object of interest, time of use or, related GPS locations, to predict what a user wants to express in a specific context (e.g., Klauer et al., 2021; Neamtu et al., 2019; Obiorah et al., 2021). To further enhance the users' communication competency, approaches to model conversational topics that could be of interest in a conversation take advantage of Natural Language Generation (NLG) (e.g., Dempster et al., 2010; Vertanen, 2017). AI-based speech recognition and parsing of the interlocutors' verbal expressions propose need-fitting information that could improve both rate and relevancy of the utterances (e.g., Neamtu et al., 2019; Wisenburn and Higginbotham, 2009). Besides the consideration of the interlocutors' voice, facial features can be useful when retrieving relevant information regarding the conversation (e.g., Ascari et al., 2020; Kane et al., 2012).

As can be derived, there are AI approaches that can support speech-impaired individuals within AAC systems. Nevertheless, it remains unclear whether the conversation phases and conversational contexts have so far been sufficiently addressed. However, it is necessary to cover several conversation phases and conversational contexts to effectively reduce the communication gap speech-impaired individuals suffer from. Thus, within the research at hand, we conduct a differentiated investigation of the extant AI-based potentials and thereby aim to contribute to closing this communication gap.

3 Research Approach

3.1 Collection and Evaluation of the Literature

Within our investigation, we performed two systematic literature reviews (Cooper, 1988; Vom Brocke et al., 2015; Webster and Watson, 2002). With the **first** literature review, we aimed at identifying design requirements posed for AAC systems within the extant literature. Thereby, it could be revealed that, although several design requirements are being investigated, there are still drawbacks these AAC systems show, i.a., low communication rates (e.g., Arnott and Alm, 2013; Kane et al., 2017) and low user- and situation-specific individualisation (e.g., Arnott and Alm, 2013; Murphy, 2004; Van Tilborg and Deckers, 2016). To address these drawbacks, the prevailing research has proposed various AI approaches. Because these AAC systems nevertheless show only a low prevalence among speech-impaired individuals, with the **second** literature review, we focused on the AI approaches and AI-based potentials for AAC systems that are proposed within the extant research literature and that we subsequently aim to investigate in more detail regarding the conversation in which the AI is applied.

Considering the taxonomy of Cooper (1988), the focus of this literature review was to investigate the potentials of AI in the context of AAC systems. Thereby, we could identify that applying AI for supporting individuals suffering from speech impairments by means of AAC systems across different phases of a conversation and within different conversational contexts is yet an under-researched topic *(see Section 2.3)*. To address this gap, firstly we identified current AI approaches and corresponding potentials for AAC systems and, secondly within the analysis of the literature, assigned the identified AI-based potentials to the three conversation phases "beginning", "middle" and "closing" as well as to the two conversational context "informativeness" and "socialness". As the differentiated investigation of AI approaches and their potentials in the context of AAC systems has so far not been addressed sufficiently within the extant research, it is nonetheless essential in order to design and develop more suitable AI approaches for AAC systems and their users.

The literature review followed a five-step procedure (Cooper and Hedges, 1994): **(I)** Problem Statement, **(II)** Collection of the Data, **(III)** Data Evaluation, **(IV)** Analysis and Interpretation of the Data, and **(V)** Presentation of the Results. The problem statement has been introduced above in *Sections 1, 2.2 and 2.3*, while the remaining aspects **(II)** - **(V)** are described in detail in the upcoming sections. At first, a

topic-based search was performed to query scholarly databases (Vom Brocke et al., 2009; Webster and Watson, 2002). The following databases were analysed: ACM Digital Library, AISeL, Emerald Insight, Google Scholar, IEEE Xplore, SpringerLink, Science Direct and Web of Science. These databases cover leading journals of the IS discipline, proceedings of highly renowned IS conferences, as well as technical reports and dissertations. By means of these selected databases, we not only cover the extant IS research but also research communities that explicitly address the investigation of AAC systems and their support potentials for speech-impaired individuals (e.g., the "Augmentative and Alternative Communication" Journal within ScienceDirect). Next, by following an iterative refinement process (cf. Kitchenham et al., 2009; Wohlin et al., 2012), we derived keywords for our topic-based search which focuses on AI applications and AI-based potentials to support speech-impaired individuals by means of AAC systems. In general, there are numerous techniques that can potentially be applied in the realm of AI (e.g., for artificial neural networks: convolutional neural network or backpropagation neural network). Since it is difficult to cover all the specific AI techniques within the literature search, we took an initial look at the extant literature on AI in the context of AAC. Thereby, we could reveal that these papers also relate to the categories of AI techniques (e.g., "machine learning", "deep learning", "supervised learning", "unsupervised learning"). Thus, we applied these and additionally "AI" and "artificial intelligence" as keywords to cover the AI aspect within the literature collection. In the context of VOCAs and SGDs that we focused on as solutions for supporting people with speech impairments, it could be revealed that the relevant literature persistently referred to those in connection with the term "augmentative and alternative communication" or the abbreviation "AAC". Overall, the following keywords, and combinations of them, were therefore applied: "augmentative and alternative communication", "AAC", "artificial intelligence", "AI", "machine learning", "supervised learning", "unsupervised learning" and "deep learning". As a result of this literature search procedure, **107 papers** could be derived.

To assess the appropriateness of these 107 publications, we first scanned the titles of the papers. If the title hinted at the application of AI in the context of AAC to support speech-impaired individuals, in the second step, we read the abstract and the conclusion of the respective paper to verify its appropriateness. Regarding the purpose of our research, we applied the following preconditions within the literature assessment: ***(1)*** *Does the publication deal with the application of AI in the context of AAC?* ***(2)*** *Does the publication reveal how these AI applications propose support for individuals suffering from a speech impairment?* ***(3)*** *Does the publication limit its application to the end device of the user and does not include further external devices, hindering the portability of the AAC systems within daily life situations, e.g., EEG sensors, sensors for acquisition of nasal pressure?* (cf. Elsahar et al., 2019).

After applying these inclusion criteria to the identified literature, 25 papers remained. For these **25 papers**, we additionally and iteratively applied backward and forward searches (cf. Webster and Watson, 2002). This resulted in **three** additional references that could meet the above-mentioned inclusion criteria. Regarding the distribution of the finally resulting **28 papers** used for the upcoming analysis across the searched databases, we found the following (with duplicates and papers not in the English language already sorted out): **14** references were extracted via Google Scholar, **three** via SpringerLink, **three** via Web of Science**, two** via Science Direct, **two** via the ACM Digital Library, **two** via IEEE Xplore, **one** via Emerald Insight and **one** via the AISeL.

3.2 Literature Analysis

After evaluating the literature collection, we analysed the relevant publications. Here, qualitative content analysis (cf. Mayring, 2004) was applied, which helps to filter out those fragments of information from larger texts that are suitable for answering the RQs.

3.2.1 Deriving the AI-based Potentials in the context of AAC Systems

The AI-based potentials being harnessed in the context of AAC systems were derived as follows: First, the publications were screened regarding any text statement that hinted at a realisable benefit (potential) of AI approaches to support speech-impaired individuals in the course of a conversation by means of an AAC system. On the one hand, some papers directly related to the potentials of applying the described

AI approaches, i.a., within the problem statements or the application and evaluation of the AI approaches (e.g., increase the speed of response or the individualisation of conversation contributions). On the other hand, in cases when there was no AI-based potential immediately observable, we related to the notion of key performance indicators in the context of AAC systems (e.g., number of clicks to produce a conversation contribution, flexibility of the utterances) and to the already derived potentials. In this way, we could derive a broad range of potentials for AI approaches for AAC systems. For the subsequent assignment to conversation phases and conversational contexts, we aimed to derive a manageable number of categories. Therefore, an abstraction of the AI-based potentials into coherent categories was carried out (cf. Mayring, 2004). Based on the attention and importance received within the literature, and led by the entities and structures that constitute a conversation (cf. Henne and Rehbock, 2012; Schank and Schoenthal, 2016), in particular, four categories could be derived *(see Section 4)*.

3.2.2 Assigning AI-based Potentials to Conversation Phases and Conversational Contexts

Subsequently, we aimed at assigning the identified AI-based potentials to the phases "beginning", "middle" and "closing" as well as to the conversational contexts "informativeness" and "socialness". The following assignment criteria could be derived from the extant literature and were applied: The publications were screened regarding any text statement that hinted at the phases "beginning", "middle" and "closing" of a conversation or the conversational context "informativeness" and "socialness". Hereby, the text statements were observed whether communication activities that are commonly performed within the respective phases are described within the paper (e.g., recognition of incoming communication requests, the identification, assessment and greeting of the conversation partner as well as the introduction of the topics of the conversation in the "beginning", recognition of the fulfilment of the conversational goals, closing the conversation, thanking the involved conversation partners and farewelling them in the "closing"). Regarding "informativeness" and "socialness", the assignments related to the description of the situation in which the conversations in the respective papers were performed. Regarding the criteria for an assignment, "informativeness"-oriented interactions are rather formal and aim at conciseness, clarity and relevance of the contents (cf. Bedrosian et al., 2003; Hoag et al., 2004). In comparison, "socialness"-oriented interactions are rather informal, aim at keeping pace with the speed of the conversational flow and are rather observed in social contexts (cf. Todman and Alm, 1997). In those cases when none of the conversation phases or conversational contexts could be observed within the respective paper, an assignment to "Not specified" was applied. If any assignment faced ambiguities, these were cleared within discussions and justifications among the involved researchers.

4 Results of the Literature Analysis

To effectively synthesise the extant research literature, we derived four coherent categories, representing the potentials of AI in the context of AAC systems, by applying inductive category development (Mayring, 2004). These categories represent the potentials of AI for AAC systems, distinguished by four dimensions: *1. Formal Course of a Conversation, 2. Natural Context Factors, 3. Interlocutor, and 4. Speech-Impaired Individual (i.e., the user of the AAC system).* In the following, exemplary potentials of the derived categories are presented.

Maintaining the Formal Course of a Conversation

The first category deals with the potential of AI in terms of the formal course of a conversation. The main objective is to ensure the grammatical correctness of the utterances, supporting their comprehensiveness and thus maintaining the overall conversation course. From the perspective of Natural Language Processing (NLP), speech errors from speech-impaired individuals can be interpreted as grammatical errors at the morphological and syntactic levels. Therefore, Park et al. (2021) applied neural Grammatical Error Correction (GEC) to ensure the formal correctness of the utterances. More precisely, they applied a deep learning-based speech-to-text approach, taking the voice of the user as input and converting it into text. Subsequently, the GEC algorithm revises various linguistic errors to replace missing or wrong words and to enable the users to practice the vocal language. To support users with

autism, Hervás et al. (2020) proposed an application to automatically compose messages based on pictograms that include AI predictive capabilities. Since predictive systems and orthographic correctors can sometimes persistently provide incorrect or undesirable options, the authors implemented a functionality whereby users can correct messages using NLP-based part-of-speech decomposition, thereby ensuring the semantic and formal correctness of the message.

Incorporating Natural Context Factors

To further support AAC systems through the potential of AI, researchers attempted to incorporate natural context factors to improve the proposed user responses by recommending situationally appropriate utterances. In this regard, numerous scholars have recognised the importance of environmental-related (e.g., GPS coordinates) or time-related (e.g., daytime) information (e.g., Klauer et al., 2021; Neamtu et al., 2019; Reddington and Tintarev, 2011; Waller, 2019).

In this context, Emil et al. (2020) developed an AI approach facilitating communication by analysing users' past selection data, including the object of interest, time of use, touch duration, GPS location, and pictogram location on the screen. Similar to this, Laxmidas et al. (2021) proposed an approach including a recommendation engine based on nearest-neighbour clustering. Each time an utterance is selected, the engine updates its internal ranking and takes into account the given time of the day and the number of selections. Subsequently, these features are partitioned into ideal clusters so that an optimal collection of relevant utterances is available for any given time of day. Also, Neamatu et al. (2019) turned their attention to the use of AI to make context-based recommendations for utterances. By analysing historical usage data (e.g., the object of interest, time of usage, GPS data, touch time) the developed machine learning algorithm predicts what a person with a speech impairment may want to express in a specific context. Further, an approach to model conversational topics that would be of interest in social conversations is presented by Dempster et al. (2010), which takes advantage of NLG to automatically create and adopt a user-specific ontology (cf. Karakatsiotis et al., 2008). To do so, several contextual information (location, time) are turned into useful conversational utterances through a template-driven utterance generation system. In addition to looking at contextual factors in general, several investigations took a step further and tried to determine how contextual factors can support specific daily activities. Obiorah et al. (2021) focused on supporting a specific leisure activity (dining in a certain restaurant) by facilitating dynamic language productions by incorporating location-related web content to support the comprehension of novel vocabulary. Similarly, Tintarev et al. (2016) focused on supporting children with complex communication needs in school to facilitate interactive narratives about personal experiences. To achieve this, the authors developed an AI-based clustering algorithm considering several locations, time and voice recording data to create concrete utterances, reflecting holistic personal narratives. To personalise the respective utterances, the approach identifies different event boundaries in terms of location changes and events specified in the children's timetables.

Tailoring towards the Interlocutor

Apart from the consideration of natural context factors such as geolocation and time of day, consideration of the interlocutor has also gained attention in the research field of AAC. In this regard, the approaches to integrate the interlocutor can be divided into two areas - speech recognition and image recognition - based on the technology used.

Wisenburn and Higginbotham (2009) demonstrated a computerised communication program utilising an NLP strategy consisting of speech recognition and a parser that proposes utterances incorporating the noun phrases spoken by the interlocutor as modular speech keys. In addition, Neamtu et al. (2019) enable speech-impaired individuals to respond more quickly by integrating a natural language conversation feature, which continuously listens to the surroundings and searches for specific trigger words and activates appropriate utterances. Therefore, the obstacle to conversation initiation and thus the communication gap are reduced. Further, the authors implemented a natural language sentence classifier to examine the raised question of the interlocutor and to classify whether the question is dichotomous. If so, the AAC system will solely present the answers of "Yes" and "No" to further enhance the communication exchange. Similarly, Emil et al. (2020) could reduce the time and effort required to communicate using

a text-to-speech module in conjunction with an NLP-based classifier. The algorithm detects specific questions and presents the most relevant utterances to users based on the interlocutors' statements. In contrast, Vertanen (2017) developed an approach to initially extract the mentioned conversation topics and subsequently generate holistic utterances by using a neural network specifically tuned to the conversation topic. In this way, the cohesion between individual parts of the conversation can be increased, resulting in a reduction of the communication gap.

Besides the consideration of the interlocutors' voice, also facial features or gestures can be useful when retrieving relevant information regarding the course of the conversation. Accordingly, Kane et al. (2012) developed an approach to generate person-associated utterances based on facial image recognition of the interlocutor, resulting in a more personalised and convenient system. Ascari et al. (2020) further applied AI in the form of a gesture recognition classification algorithm to identify non-verbal expressions of the interlocutor and transform them into associated emotions. Subsequently, they will be included in the generation of the users' responses, allowing them to respond to the mood or emotions of the interlocutor and thus strengthen their ability to communicate.

Tailoring towards the Speech-Impaired Individual (AAC System User)

In addition, numerous researchers have harnessed the potential of AI to better tailor the conversation towards the user (e.g., characteristics, attitudes and preferences of the user). Shen et al. (2022) proposed a multi-turn dialogue system that applies a language model with bag-of-keywords and personality-related information to generate high-quality responses based on keyword entries. By the generation of meaningful utterances with very little user input, additional keystrokes were saved, leading to a significant reduction in the communication gap. Further, Mancilla et al. (2015) developed an ontology personalisation approach for an AAC system, which automatically extends a given domain ontology into a user-related one, including user-specific information. By including personal information in the communication process, the message quality improves as it becomes more complete and personal. According to Hernández et al. (2014), user behaviour is essential to establish personalised recommendations. Thus, the authors implemented a memory-based recommender system that tracks users by user archetypes, representing their behaviour patterns, goals and needs. Subsequently, the archetypes are used in conjunction with AI and statistical methods to predict and suggest need-fitted utterances. Further, Heo and Kang (2019) proposed a system to incorporate user-related information by applying a slot-filling method. By employing decision trees, it automatically learns slot priority strategies to enable the efficient selection of information for the user and thus an efficient way to prevent unnecessary inputs and minimise the number of keystrokes. Another method for integrating user-related information by generating context-related vocabulary from photographs of personally relevant events is presented by De Vargas and Moffatt (2021), who aimed at supporting people with speech impairments in recounting their past experiences. Here, NLP-based photo-to-story vocabulary generation reduced irritation and navigation errors and generated improved lexical retrieval during activity retelling. Johnson et al. (2020) demonstrated the usage of various sensors for measuring the voice and video recordings of non-verbal users. By using speech and image recognition, the user's moods, emotions and sounds can be measured and assigned to different categories (e.g., laughter, crying, or protest), thereby clarifying the underlying intention and supporting the user's expressiveness. Finally, as stated by Mills et al. (2014), several individuals suffering from speech impairment use the same synthetic voice, diminishing their expressiveness. To enable a more personalised voice, the authors developed an approach that extracts prosodic properties from the target talker's voice and applies these features to a surrogate talker's database, generating a synthetic voice with the vocal identity of the target talker and the clarity of the surrogate talker.

5 Discussion and Future Research

To identify existing gaps in the coverage of AI-based potentials concerning the conversational phases and contexts and thus deduce related research prospects, the AI potentials were further assigned to the appropriate conversation phases and contexts as described in Section 3.2.2. The respective assignments are shown in Table 1.

Authors / Publications	AI-based Potentials of AAC Systems				Conversation Contexts			Conversation Phases			
	Maintaining the Formal Course of a Conversation	Incorporating Natural Context Factors	Tailoring towards the Interlocutor	Tailoring towards the Speech Impaired	Socialn.	Inform.	Not Spec.	Beg.	Mid.	Clos.	Not Spec.
Ascari et al. (2020)			X				X				X
Dempster et al. (2010)		X	X	X	X				X		
De Vargas and Moffatt (2021)		X		X	X				X		
Emil et al. (2020)		X	X	X			X		X		
Heo and Kang (2019)				X			X				X
Hernández et al. (2014)				X	X						X
Hervás et al. (2020)	X			X		X					X
Hossain et al. (2018)		X					X				X
Johnson et al. (2020)				X	X				X		
Kane et al. (2012)		X			X						X
Klauer et al. (2021)		X					X				X
Kristensson et al. (2020)		X				X					X
Laxmidas et al. (2021)		X		X			X				X
Li et al. (2022)				X			X				X
Mancilla et al. (2015)		X		X	X				X		
Mills et al. (2014)				X			X				X
Neamtu et al. (2019)		X	X	X	X			X	X		
Obiorah et al. (2021)		X	X			X		X	X		
Park et al. (2021)	X	X		X			X		X		
Reddington and Tintarev (2011)		X			X				X		
Roy et al. (2021)	X		X	X			X	X	X		
Shen et al. (2022)		X		X	X				X		
Tintarev et al. (2016)		X		X	X						X
Vertanen (2017)			X				X		X		
Waller et al. (2013)				X	X				X		
Waller (2019)		X	X		X				X		
Wandmacher et al. (2007)		X					X				X
Wisenburn and Higginbotham (2009)			X		X				X		

Explanation of the Abbreviations: *Socialn.: Socialness, Inform.: Informativeness; Beg.: Beginning; Mid.: Middle; Clos.: Closing; Not Spec.: Not Specified.*

Table 1. Assignment of AI-related potentials to conversation phases and conversational contexts.

The general picture that emerges from the analysis is that some studies have already addressed specific contexts and phases of conversations. However, significant gaps in the literature could be revealed, which indicate shortcomings in the research on the examined subject of AAC and AI. To guide research into the future, we derived several implications for research below.

5.1 Implications for Future Research

The findings deduced from the assignment of the AI-based potentials to the respective conversation phases and conversational contexts are intended to help researchers and practitioners to make sense of the accumulated knowledge in this emerging research field. Based thereon, we discuss implications concerning the observations throughout this investigation and outline novel directions for future research that may yield promising potentials in the field of AAC.

First, as a cursory glance at Table 1 reveals, there is no prevailing research that leverages the potentials of all derived categories to support speech-impaired individuals. However, to support communication holistically using an AAC system, the consideration of all conversation-related dimensions (categories) and thus the use of their associated potentials is indispensable. Thus, studies considering all relevant dimensions of a conversation simultaneously could be promising avenues for further research.

- The prevalent literature on the first category is sparse, with only three articles specifically addressing the use of AI to maintain a formal course within a conversation. In comparison, 17 articles deal with the integration of natural context factors as well as nine and 17 articles deal with the consideration of the conversation partner and the speech-impaired individual, respectively. This leads to the

assumption, that there is a necessity for research concerning maintaining the formal course of the conversation by means of tailored AI approaches. In this respect, additionally to neural GEC, word prediction engines could be leveraged to influence the choices of spell checkers or to automatically correct minor spelling errors. Moreover, prediction engines that use syntactic or semantic components could similarly be used for grammar correction and disambiguation of meaning. For example, Widgit's Communicate software (Widgit, 2022) uses syntactic disambiguation to provide the appropriate symbol when an ambiguous word such as "can" occurs in a sentence. In general, probability-based prediction engines can be used in many situations where textual ambiguity exists, and support both conversational comprehension and users' communicative competence.

- Moreover, AI holds several potentials for text simplification and summarisation and thus supporting the maintenance of the formal course of conversation. Text simplification converts the associated text message into simpler sentence structures and vocabulary while retaining the meaning of the original (Al-Thanyyan and Azmi, 2021). This could also support the formal course of the conversation by breaking down more complex issues into simpler elements of communication, and thus reducing the likelihood of errors.
- Similar to text simplification, text summarisation aims to provide a brief overview of what is contained within large texts. Using an automatic summarisation approach, particularly important sentences in a document can be identified and included in the summary (Adhikari, 2020). Although comparable technologies are occasionally used for AAC systems (e.g., Obiorah et al., 2021), the potential of automatic text summarisation has yet to be exhausted. In this way, complex text messages could be simplified for individuals with cognitive-linguistic challenges. Newspaper or magazine articles, website content, or emails could be automatically processed and inserted into the AAC system to transform them into modules for expressive communication. With a simplified and appropriately segmented set of text modules, the user could select from these to discuss current events or personal narratives with grammatically and orthographically correct sentences.

Second, the consideration of conversational contexts is worth mentioning. As previously indicated, considering specific conversational contexts is one way to address the associated requirements (e.g., formal vs. informal) to specifically support the conversational process. However, as Table 1 reveals, solely three articles refer to the conversational context of "informativeness", while 13 articles take into account the conversational context of "socialness". Since the differentiation between these conversational contexts is indispensable to holistically support AAC-based communication *(see Sections 1 and 2.1)* and thus achieve a high level of communication competence in speech-impaired individuals, future research will need to shed light on how the conversational contexts can be supported by appropriate AI methods. Especially concerning the context of "informativeness", our results implicate research gaps, since the prevailing literature to date mainly focuses on the context of "socialness".

- Since "informativeness"-oriented communication is more concerned with the formal exchange of relevant information that focuses on conciseness, clarity, and relevance of content rather than the speed of exchange as in "socialness"-oriented communication (cf. Bedrosian et al., 2003; Hoag et al., 2004), "informativeness"-oriented communication aims at a holistic view of a topic, including the coverage of all topic-specific relationships. In contrast to "socialness"-oriented conversations, the focus is thus mainly on the meticulous elaboration of a specific topic and less on the superficial but more flexible coverage of a broad spectrum of differentiated topics.
- One conceivable approach for the holistic mapping of a discourse domain is the AI-based generation of an ontology that can be applied to map all topic-related concepts and their relationships. Ontologies model particular knowledge through a representative vocabulary determined by the formal definition of a set of concepts associated with a particular domain and the meta-relationships between them (Mancilla et al., 2015). In this way, criteria of "informativeness"-oriented communication such as clarity, coherence, and extensibility are further met (Gruber, 1995). To enable the holistic coverage of a conversation topic with all associated concepts, deep learning approaches (e.g., neural networks) can be applied in conjunction to dynamically identify and integrate topic-relevant concepts. Once a specific topic has been identified, probabilistic neural networks can allocate (potentially)

relevant concepts and dynamically extend the ontology with the topic-specific concepts. Therefore, semantically related concepts in various degrees of detail can be identified, resulting in a holistic coverage of the discussed topic and thus enabling, e.g., the formation of logical conversation chains supporting the formal exchange of information. As an example, when ordering a dish in a restaurant *(see Figure 1, Section 2.1)*, the user would also like to receive information about certain aspects that might be relevant to ordering the dish. If the user wants to ask about the origin of certain ingredients or possible allergens, the relevant topic-specific concepts must be available within the ontology. To achieve the dynamic extension of the given ontology, the associated concepts (e.g., the origin of specific ingredients) could be determined by a neural network and dynamically integrated into the ontology. By dynamically integrating topic-specific concepts, all relevant facets of a conversation can be taken into account, ensuring conciseness as well as the relevance of the respective content on the one hand and supporting clarity through a logical structure of the topics on the other.

- In this vein, studies additionally have to examine which event boundaries can be defined to dynamically switch between the conversational contexts to adapt the proposed utterances in a target-oriented way. Automatic speech recognition could be utilised to analyse the spoken words of the interlocutor regarding specific keywords to determine whether a communication shifts from "socialness" to “informativeness”. If a shift in the communication context is identified, the generation of utterances can be specifically geared to the requirements of the currently identified context, thereby increasing both the prospect of conversational success and the conversational competence of the user.

Third, in view of the conversation phases, each phase has to meet specific goals and therefore comprise individual communication activities the conversation participants need to perform *(see Section 2.1)*. Taking into account the assignment of the AI-based potentials to the respective communication phases, it becomes apparent, that the middle part is densely covered (15 articles), while the opening part (three articles) and the closing part (null articles) are not sufficiently considered yet. Nonetheless, since each conversation phase is associated with specific activities such as recognising incoming communication requests or closing the conversation by thanking the interlocutors, the consideration of each conversation phase and its associated requirements is essential to support a conversation holistically.

- In view of the conversation initiation, image recognition techniques and deep learning-based computer vision propose promising potentials. In this context, action and event recognition could identify an emerging conversation beginning by analysing, from a sequence of images, different actions that a human performs (Leo et al., 2018). Common examples of actions are, e.g., “answering the phone” or “shaking hands”, but also facial actions like “smiling” or “crying”. By identifying certain actions that indicate an upcoming conversation situation (e.g., a greeting gesture such as waving), a conversation situation could be initiated, whereupon the corresponding utterances would be suggested automatically. For illustration purposes, if a speech-impaired user enters a restaurant and is guided to his seat by a waiter, the AAC system could automatically identify, based on the waiter's gesture, that the user is guided to take a seat. Accordingly, the system can generate appropriate utterances that support the respective conversation situation and thus enable the generation of targeted utterances. Likewise, the facial expression of the conversation partner could be included in the automated adaptation of the vocabulary. Thus, in the case of an angry interlocutor, positively connotated vocabulary could be used to appease in.
- In the context of conversation closing, the focus is on farewelling the interlocutor (cf. Schegloff and Sacks, 1973; Spiegel and Spranz-Fogasy, 2001). Thereby, deep learning-based computer vision approaches could be used to enable automated recognition of the closing of the conversation by identifying gestures of farewell and adapting the proposed utterances accordingly. Further potential in the conversation phase could be represented by the automated evaluation of the conversation, the verification of the achievement of conversation goals, and the extraction and evaluation of discussed conversation topics. In this context, machine learning approaches such as automatic topic extraction (cf. Jelodar et al., 2019) could be used to extract the discussed topics and subsequently have them evaluated by the user. If the user rates a topic positively, it can be prioritised in the automated identification of future conversation topics concerning upcoming conversations.

Overall, our results emphasise that AI techniques are ubiquitous in current AAC technologies and propose great potentials for development in the future. Therefore, we encourage research to address the consideration of phase- and context-specific requirements of an AAC conversation to develop appropriate techniques to support speech-impaired individuals and thus better integrate them into social life.

5.2 Limitations

The investigation at hand has several limitations. Although we conducted a broad and structured database search covering major outlets, other researchers might have uncovered additional search terms and additional relevant papers. Yet, this structured literature review allows for a transparent, replicable, and comprehensive overview of AI-based potentials in the context of AAC systems. The assignments of the derived AI-based potentials to conversation phases and conversational contexts bear to a certain extent subjectivity. However, several assignments were performed by three researchers individually. In case of disagreements, the article in question was analysed by a fourth researcher, followed by a discussion until a consensus was reached. Despite these limitations, we are confident that our findings help to get a first overview and a better understanding of the current AI approaches in the context of AAC systems and the insufficiencies of the AI decisions that can be related to the underlying conversations.

6 Conclusion

Digital systems of Augmentative and Alternative Communication for supporting speech-impaired individuals have gained tremendous importance as a means of the inclusion of speech-impaired individuals into societal life. Although the extant research has investigated several design requirements and AI approaches, there are certain circumstances where AAC systems deduce recommendations and apply adaptions that are insufficient given the specific situation in which they are applied. This is due to the divergent requirements imposed by the conversation context (e.g., formal vs. informal conversations) and its structure (e.g., conversation phases). Thus, as the AI approaches may need to be adapted concerning the conversations in which they are applied, we took an up-close look thereon within our investigation.

To establish a first basis for the underlying research topic and to develop a deeper understanding of the research field of AI in the context of AAC systems, we have therefore conducted a systematic literature review *(see Section 3)*. In this way, we have derived corresponding AI approaches and deduced related AI-based potentials. For their subsequent assignment to conversation phases and conversational contexts, we could derive four categories that are based on the attention and importance they have received within the extant literature and are additionally led by the entities and structures that constitute a conversation (e.g., conversational partners, the course of the conversation). Subsequently, based on the determinants of the phases and contexts that are described in the corresponding literature, we have assigned the categories of AI-based potentials to the conversation phases “beginning“, “middle“, “closing“ and to the two conversational contexts “informativeness“ and “socialness“. As this assignment revealed, there are indeed several gaps within the extant research on AI regarding the coverage of the conversational context “informativeness” and the conversation phases “beginning” and “closing” *(see RQ1)*.

To this extent, this paper proposes several contributions. First, based on the extant literature, it provides a comprehensive and structured overview of AI-based potentials in the context of AAC systems. Moreover, we have consolidated these AI-based potentials into an abstraction of four coherent categories *(see Section 4)* and could establish synergies of these research areas. Second, our results propose possibilities for AI research to investigate the identified AI-based potentials within conversation phases and conversational contexts that have received less attention yet. Based on the contributions of the work at hand, AI approaches that are established within other fields (e.g., within voice assistants for voice commerce, e.g., Reinkemeier et al., 2022) and current generative AI-based language models could be investigated in specific for the case of AAC systems. To guide future research, we contribute several implications for investigating AI in the context of AAC systems, especially regarding the conversation in which the supporting AI is applied *(see RQ2)*. We hope future IS research strives to address the revealed gaps to dismantle communication barriers speech-impaired individuals suffer from.

References

Adhikari, S. (2020). NLP-based machine learning approaches for text summarization. *Proceedings of the 4th International Conference on Computing Methodologies and Communication*, 535-538.

Allen, J. (2005). Designing desirability in an augmentative and alternative communication device. *Universal Access in the Information Society*, 4(2), 135-145.

Alm, N., and Higginbotham, J. (2008). Whole utterance approaches in AAC. *Augmentative and Alternative Communication*, 24(3), 235-254.

Al-Thanyyan, S. S., and Azmi, A. M. (2021). Automated text simplification: a survey. *ACM Computing Surveys*, 54(2), 1-36.

American Speech Language Hearing Association (1991). Report: augmentative and alternative communication. *ASHA*, 33, 9–12.

Arnott, J. L., and Alm, N. (2013). Towards the improvement of Augmentative and Alternative Communication through the modelling of conversation. *Computer Speech & Language*, 27(6), 1194-1211.

Ascari, R. E., Pereira, R., and Silva, L. (2020). Computer vision-based methodology to improve interaction for people with motor and speech impairment. *ACM Transactions on Accessible Computing*, 13(4), 1-33.

Baldassarri, S., Marco, J., Cerezo, E., and Moreno, L. (2014). Accessibility evaluation of an alternative and augmentative communication (AAC) tool. *Proceedings of the 2014 International Conference on Universal Access in Human-Computer Interaction*, 529-540. Springer, Cham.

Bedrosian, J. L., Hoag, L. A., and McCoy, K. F. (2003). Relevance and speed of message delivery trade-offs in augmentative and alternative communication. *Journal of Speech, Language, and Hearing Research*, 46, 800–817.

Boden, M. A. (2018). Artificial Intelligence: A very short introduction. *Oxford, England: Oxford University Press*.

Boyes-Braem, P., Erting, C. J., Johnson, R. C., Smith, D. L., and Snider, B. D. (1994). An overview of current sign language projects in Switzerland. *The Deaf Way: Perspectives from the International Conference on Deaf Culture*, 382-387. Washington, DC: Gallaudet University Press.

Bradshaw, J. (2013). The use of augmentative and alternative communication apps for the iPad, iPod and iPhone: an overview of recent developments. *Tizard Learning Disability Review*, 18(1), 31-37.

Brady, N. (2000). Improved comprehension of object names following voice output communication aid use: Two case studies. *Augmentative and Alternative Communication*, 16(3), 197-204.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1), 104-126.

Cooper, H. M., and Hedges, L. V. (1994). Research synthesis as a scientific enterprise. *The Handbook of Research Synthesis*, 285-299.

Dattilo, J., Estrella, G., Estrella, L. J., Light, J., McNaughton, D., and Seabury, M. (2008). "I have chosen to live life abundantly": Perceptions of leisure by adults who use augmentative and alternative communication. *Augmentative and Alternative Communication*, 24(1), 16-28.

Dempster, M., Alm, N., and Reiter, E. (2010). Automatic generation of conversational utterances and narrative for Augmentative and Alternative Communication: a prototype system. *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, 10-18.

Desai, T., Chow, K., Mumford, L., Hotze, F., and Chau, T. (2014). Implementing an iPad-based alternative communication device for a student with cerebral palsy and autism in the classroom via an access technology delivery protocol. *Computers & Education*, 79, 148-158.

De Vargas, M. F., and Moffatt, K. (2021). Automated Generation of Storytelling Vocabulary from Photographs for use in AAC. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics,* 1353-1364.

Ehlich, K., and Rehbein, J. (1972). Zur Konstitution pragmatischer Einheiten in einer Institution: Das Speiserestaurant. *Linguistische Pragmatik*, 209-254.

Elsahar, Y., Hu, S., Bouazza-Marouf, K., Kerr, D., and Mansor, A. (2019). Augmentative and alternative communication (AAC) advances: A review of configurations for individuals with a speech disability. *Sensors*, 19(8).

Emil, Z., Robbertz, A., Valente, R., and Winsor, C. (2020). Towards a more inclusive world: Enhanced augmentative and alternative communication for people with disabilities using AI and NLP. *Major Qualifying Projects, Digital Worcester Polytechnic Institute*.

Fritzell, B. (1996). Voice disorders and occupations. *Logopedics Phoniatrics Vocology*, 21(1), 7-12.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing?. *International Journal of Human-Computer Studies*, 43(5-6), 907-928.

Guasch, D., Martín-Escalona, I., Macías, J. A., Moreno, L., Hervás, R., and Bautista, S. (2019). Expert-based assessment of an augmentative and alternative communication tool. *Proceedings of the 20th International Conference on Human Computer Interaction*, 1-8.

Henne, H., and Rehbock, H. (2012). Einführung in die Gesprächsanalyse. *Einführung in die Gesprächsanalyse*. de Gruyter.

Heo, Y., and Kang, S. (2019). Augmentative and Alternative Communication System Using Information Priority and Retrieval. *Proceedings of the 2019 IEEE International Conference on Big Data and Smart Computing*, 1-4.

Hernández, S. S., Mancilla, D., Medina, J. M., and Iregui, M. (2014). User-centric Recommendation Model for AAC based on Multi-criteria Planning. *Proceedings of the 2014 International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies and Services*.

Hervás, R., Bautista, S., Méndez, G., Gavlán, P., and Gervás, P. (2020). Predictive composition of pictogram messages for users with autism. *Journal of Ambient Intelligence and Humanized Computing*, 11(11).

Hoag, L. A., Bedrosian, J. L., McCoy, K. F., and Johnson, D. E. (2004). Trade-offs between informativeness and speed of message delivery in augmentative and alternative communication. *Journal of Speech, Language, and Hearing Research*, 47, 1270–1285.

Hossain, M. S., Takanokura, M., Sakai, H., and Katagiri, H. (2018). Using context history and location in context-aware AAC systems for speech-language impairments. *Proceedings of the International Multiconference of Engineers and Computer Scientists*, 1.

Hourcade, J., Everhart Pilotte, T., West, E., and Parette, P. (2004). A history of augmentative and alternative communication for individuals with severe and profound disabilities. *Focus on Autism and other Developmental Disabilities*, 19(4), 235-244.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., and Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169-15211.

Johnson, K. T., Narain, J., Ferguson, C., Picard, R., and Maes, P. (2020). The ECHOS Platform to Enhance Communication for Nonverbal Children with Autism: A Case Study. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-8.

Kane, S. K., Linam-Church, B., Althoff, K., and McCall, D. (2012). What we talk about: designing a context-aware communication tool for people with aphasia. *Proceedings of the 14th International ACM Conference on Computers and Accessibility*, 49-56.

Kane, S. K., Morris, M. R., Paradiso, A., and Campbell, J. (2017). "At times avuncular and cantankerous, with the reflexes of a mongoose" Understanding Self-Expression through Augmentative and Alternative Communication Devices. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1166-1179.

Karakatsiotis, G., Galanis, D., Lampouras, G., and Androutsopoulos, I. (2008). NaturalOWL: Generating texts from OWL ontologies in Protégé and in Second Life. *Proceedings of the 18th European Conference on Artificial Intelligence*.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and Software Technology*, 51(1), 7-15.

Klauer, A., Johannsen, F., and Leist, S. (2021). "BeMyVoice"–An iPad-Based Helper for Speech Impaired Persons. *Proceedings of the International Conference on Design Science Research in Information Systems and Technology*, 40-51. Springer, Cham.

Kristensson, P. O., Lilley, J., Black, R., and Waller, A. (2020). A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-11.

Laxmidas, K., Avra, C., Wilcoxen, C., Wallace, M., Spivey, R., Ray, W., Polsley, S., Kohli, P., Thompson, J., and Hammond, T. (2021). CommBo: Modernizing Augmentative and Alternative Communication. *International Journal of Human-Computer Studies*, 145.

Leo, M., Furnari, A., Medioni, G. G., Trivedi, M., and Farinella, G. (2018). Deep learning for assistive computer vision. *Proceedings of the European Conference on Computer Vision Workshops*.

Li, W., Qiu, X., Li, Y., Ji, J., Liu, X., and Li, S. (2022). Towards a novel machine learning approach to support augmentative and alternative communication (AAC). *International Journal of Speech Technology*, 25(2), 331-341.

Louis-Nouvertné, U. (2001). Was sind Gebärdensprachen?: Eine Einführung in die wichtigsten Ergebnisse der linguistischen Gebärdensprachforschung. *Sprache und Literatur*, 32(2), 3-20.

Mancilla, D., Sastoque, S., and Iregui, M. (2015). Towards ontology personalization to enrich social conversations on AAC systems. *Proceedings of the 10th International Symposium on Medical Information Processing and Analysis*, 9287, 212-217.

Mayring, P. (2004). Qualitative content analysis. *A Companion to Qualitative Research*, 1(2), 159-176.

McNaughton, D., and Bryen, D. N. (2007). AAC technologies to enhance participation and access to meaningful societal roles for adolescents and adults with developmental disabilities who require AAC. *Augmentative and Alternative Communication*, 23(3), 217-229.

Mills, T., Bunnell, H. T., and Patel, R. (2014). Towards personalized speech synthesis for augmentative and alternative communication. *Augmentative and Alternative Communication*, 30(3), 226-236.

Murphy, J. (2004). "I Prefer Contact This Close": Perceptions of AAC by People with Motor Neurone Disease and their Communication Partners. *Augmentative and Alternative Communication*, 20(4), 259-271.

Neamtu, R., Camara, A., Pereira, C., and Ferreira, R. (2019). Using artificial intelligence for augmentative alternative communication for children with disabilities. *Proceedings of the 2019 IFIP Conference on Human-Computer Interaction*, 234-243. Springer, Cham.

Obiorah, M. G., Piper, A. M., and Horn, M. (2021). Designing AACs for people with aphasia dining in restaurants. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-14.

Park, C., Jang, Y., Lee, S., Seo, J., Yang, K., and Lim, H. (2021). PicTalky: Augmentative and Alternative Communication Software for Language Developmental Disabilities.

Pollak, V. A., and Gallagher, B. (1989). A fast communication aid for non-verbal subjects with severe motor handicaps. *Journal of Medical Engineering & Technology*, 13(1-2), 23-27.

Reddington, J., and Tintarev, N. (2011). Automatically generating stories from sensor data. *Proceedings of the 16th International Conference on Intelligent User Interfaces*, 407-410.

Reinkemeier, F., Gnewuch, U., and Toporowski, W. (2022). Can Humanizing Voice Assistants Unleash the Potential of Voice Commerce?. *Proceedings of the 43rd International Conference on Information Systems*.

Roy, Q., Berlioux, S., Casiez, G., and Vogel, D. (2021). Typing Efficiency and Suggestion Accuracy Influence the Benefits and Adoption of Word Suggestions. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-13.

Schank, G., and Schoenthal, G. (2016). Gesprochene Sprache. *Gesprochene Sprache*. Max Niemeyer Verlag.

Schegloff, E. A., and Sacks, H. (1973). Opening up closings. De Gruyter.

Schepis, M. M., Reid, D. H., and Behrman, M. M. (1996). Acquisition and functional use of voice output communication by persons with profound multiple disabilities. *Behavior Modification*, 20(4), 451-468.

Schindler, E. (2013). Kommunikative Ritualität im Gastronomiebereich. *Freiburger Arbeitspapiere zur germanistischen Linguistik*, 11. Freiburg: Germanistische Linguistik, Universität Freiburg.

Shen, J., Yang, B., Dudley, J. J., and Kristensson, P. O. (2022). Kwickchat: A multi-turn dialogue system for aac using context-aware sentence generation by bag-of-keywords. *Proceedings of the 27th International Conference on Intelligent User Interfaces*, 853-867.

Spiegel, C., and Spranz-Fogasy, T. (2001). Aufbau und Abfolge von Gesprächsphasen. *Text-und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. 2. Halbband, 1241-1252. De Gruyter.

Tintarev, N. Reiter, E., Black, R., Waller, A., and Reddington, J. (2016). Personal storytelling: Using Natural Language Generation for children with complex communication needs, in the wild …. *International Journal of Human-Computer Studies*, 92, 1-16.

Todman, J., and Alm, N. (1997). Pragmatics and AAC approaches to conversational goals. In Natural language processing for communication aids. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Vanderheiden, G. C. (1983). Non-Conversational Communication Technology Needs of Individuals with Handicaps. *Rehabilitation World*, 7(2), 8-12.

Van Tilborg, A., and Deckers, S. R. (2016). Vocabulary selection in AAC: Application of core vocabulary in atypical populations. *Perspectives of the ASHA Special Interest Groups*, 1(12), 125-138.

Vertanen, K. (2017). Towards improving predictive aac using crowdsourced dialogues and partner context. *Proceedings of the 19th International ACM Conference on Computers and Accessibility*, 347-348.

Vom Brocke, J. Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., and Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. *Proceedings of the 17th European Conference on Information Systems*.

Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, 37(1).

Waller, A., Menzies, R., Herron, D., Prior, S., Black, R., and Kroll, T. (2013). Chronicles: Supporting conversational narrative in alternative and augmentative communication. *Proceedings of the 2013 IFIP Conference on Human-Computer Interaction*, 364-371. Springer, Berlin, Heidelberg.

Waller, A. (2019). Telling tales: unlocking the potential of AAC technologies. *International Journal of Language & Communication Disorders*, 54(2), 159-169.

Wandmacher, T., Antoine, J. Y., and Poirier, F. (2007). SIBYLLE: a system for alternative communication adapting to the context and its user. *Proceedings of the 9th International ACM Conference on Computers and accessibility*, 203-210.

Wang, E. H., Zhou, L., Chen, S. H., Hill, K., and Parmanto, B. (2018). Development and evaluation of a mobile AAC: a virtual therapist and speech assistant for people with communication disabilities. *Disability and Rehabilitation: Assistive Technology*, 13(8), 731-739.

Webster, J., and Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, xiii-xxiii.

Widgit (2022). SymWriter 2 – Easy symbol writing for everyone. URL: https://www.widgit.com/products/symwriter/index.htm (visited on November 01, 2022).

Wisenburn, B., and Higginbotham, J. (2009). Participant evaluations of rate and communication efficacy of an AAC application using natural language processing. *Augmentative and Alternative Communication*, 25(2), 78-89.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). Systematic literature reviews. *Experimentation in Software Engineering*. Springer, Berlin-Heidelberg. 45-54.

2.6. **Beitrag 6:** A Generative AI-Based Approach to Support Automated Utterance Generation for Different Conversational Contexts within AAC Systems

Adressierte Forschungsfrage	Wie könnte ein umfassender Kompositionsansatz für ein AAC-System technisch realisiert werden, der auf Generativer KI beruht und es einem sprachbehinderten Menschen ermöglicht, sowohl formelle als auch informelle Gesprächsbeiträge zu generieren?
Zielstellungen	• Identifikation von Designanforderungen für den Generativen KI-basierten Kompositionsansatz auf Basis der wissenschaftlichen Literatur und basierend auf der „Cognitive Load Theory“ nach Sweller (1988) sowie den „Cooperative Principles of Conversation“ nach Grice (1975). • Entwicklung des Generativen KI-basierten Kompositionsansatzes und Integration in einen ebenfalls implementierten AAC-Prototyp. Der Kompositionsansatz soll dabei sämtliche der identifizierten Designanforderungen und Designprinzipien erfüllen.
Forschungsmethode	**Design Science Research** • DSR-Prozess nach Peffers et al. (2007), der basierend auf einer Problem- und Lösungsidentifikation die Schritte Entwicklung, Demonstration und Evaluation für den auf Generativer KI basierenden Kompositionsansatz für AAC-Systeme umgesetzt hat. • Anlehnung der Forschung an Hevner et al. (2004) durch die Ausgestaltung des Design Cycle (Demonstration und Evaluation des Artefakts), Relevance Cycle (Beitrag zur Praxis) und Rigor Cycle (Beitrag zur (Nascent) Designtheorie durch Designprinzipien (vgl. Gregor und Hevner, 2013)).
Kernergebnisse (Überblick)	• Konsolidierung von Designanforderungen für die strukturierte Entwicklung des Generativen KI basierten Kompositionsansatz. • Ableitung und Evaluation von insgesamt fünf Designprinzipien. • Im Hinblick auf die Generierung von Gesprächsbeiträgen: Identifizierung und Umsetzung von Adaptionen, damit ChatGPT-4.0 als effektive und zielgerichtete Generative KI Technik zur Komposition von formellen und informellen Gesprächsbeiträgen im Rahmen eines AAC-Systems funktioniert. • Im Hinblick auf die Möglichkeit zum Einbeziehen individueller Nuancen und Ausdrücke zur Artikulation: Umsetzung eines Personalisierungs-Tabs, in welchem individuelle Nuancen und Ausdrücke eines Nutzers eingegeben werden können, sodass diese als feste Bestandteile in der Generierung der Gesprächsbeiträge berücksichtigt werden.
Publikationsort	**Veröffentlicht** in den Proceedings der 45. International Conference on Information Systems, Bangkok, Thailand.
Ranking VHB JQ 4	**A**
Autoren und Anteile	**Daniel Konadl** **100 %**

Tabelle 8: Fact Sheet zu Beitrag 6

A Generative AI-Based Approach to Support Automated Utterance Generation for Different Conversational Contexts within AAC Systems

Completed Research Paper

Daniel Konadl
University of Regensburg
Universitätsstraße 31,
93051 Regensburg, Germany
Daniel.Konadl@ur.de

Abstract

To foster their full integration into the mainstream of society, speech-impaired individuals need to be empowered to use vocal language in various daily life situations. AI-based AAC systems enable speech-impaired individuals to generate completed utterances for use in conversations. However, available solutions exhibit remarkable drawbacks. They do not meet all posed requirements and cannot provide completed sentences that match the expected styles of formal and informal conversational contexts. Therefore, design requirements for a Generative AI-based utterance composition approach were identified. It has been followed design science research to derive design principles and an instantiation of an AAC prototype that enables context-specific and user-individualizable articulation. The artifact was demonstrated and evaluated with formal and informal interactions commonly performed in the daily routine of visiting a restaurant. In particular, the importance of meeting all specific requirements and employing Generative AI, such as ChatGPT-4.0, to generate both formal and informal utterances, is exemplified.

Keywords: Augmentative and Alternative Communication, Generative AI, Conversational Contexts, Automated Utterance Composition, Design Science Research

Introduction

"Perhaps the single quality most central to humanness is the ability to exchange thoughts, ideas, and feelings with others" (Hourcade et al. 2004, p. 235). Speech is perceived as an individual's most important instrument to get in contact with their surroundings (Fritzell 1996; Kane et al. 2017). However, individuals suffering from severe physical disabilities or brain injuries are not able to control their oral-respiratory musculature sufficiently for speech (Allen 2005; Vanderheiden 1983). This restriction to verbal communication and the *"separation from the mainstream of society"* (Hourcade et al. 2004, p. 235) holds true for people with damage to the vocal tract or other impairments (e.g., aphasia, autism, dyslexia) (Allen 2005).

To give speech-impaired individuals a voice and enable them to express themselves using vocal language, digital solutions for Augmentative and Alternative Communication (AAC) have emerged. These AAC systems enable the translation of an individual's intended meaning into speech and subsequent voice outputs (Bradshaw 2013; Desai et al. 2014). One primary goal has thereby been to reduce the communication gap between speech-impaired individuals and their speech-capable interlocutors. To reach this goal, the extant AAC research has investigated different AI-based techniques that can automatically compose completed sentences from the single words and expressions speech-impaired individuals provide within their AAC systems (cf. Konadl et al. 2023). In this way, the communication gap could indeed be reduced by increasing the rate of words that speech-impaired individuals can contribute to conversations using AAC systems.

Nevertheless, AI-based AAC solutions still exhibit remarkable drawbacks regarding the utterances they generate. There is no comprehensive AAC system or tool available that meets all the requirements outlined in the scientific literature *(see Table 2)*. However, extant literature points out the necessity of an AAC system that provides speech-impaired individuals with comprehensive support in various daily life conversations (Kane et al. 2017; Konadl et al. 2023). Beyond the extant research literature, the importance of such a comprehensive AI-based AAC system has been further underlined in this research through exchanges with speech-impaired individuals and representatives of the "Professional Association for the Hearing and Speech Impaired in Bavaria". Concerning that, available AAC systems can provide completed utterances only for a limited number of situations as the underlying language models are trained on a limited number of daily life conversations with limited diversity (e.g., Dempster et al. 2010; Heo and Kang 2019; Kristensson et al. 2020). As a result, these language models struggle to generate completed utterances for domains and communication activities that speech-capable individuals usually can take part in. Moreover, even the fastest way of expressing oneself is of little help if the utterances proposed by an AAC system are perceived as distracting by speech-impaired individuals and their interlocutors. In particular, the style of an utterance may not fit the expected way of articulation in a formal or informal conversational context (Bedrosian et al. 2003; Hoag et al. 2004; Todman and Alm 1997). For example, an AAC system might suggest an utterance in which an authority figure like a doctor at the doctor's office is addressed with colloquial language (e.g., "bro" or "buddy") that is more suited to informal contexts. Hence, interlocutors may form negative perceptions towards speech-impaired individuals and avoid engaging in a conversation with them. The style of the utterances may also lead to speech-impaired individuals themselves avoiding conversations, as the utterances their AAC systems provide them do not reflect their personality and individual style of articulation (Shen et al. 2022; Tintarev et al. 2016; Valencia et al. 2023).

Therefore, this research aims to close these gaps. Based on Design Requirements (DRs) derived from the extant AAC research and theory, an instantiation of a functional AAC system prototype is proposed that integrates comprehensive functionalities to tackle the aforementioned drawbacks. The AI-based utterance composition approach is designed and constructed to combine single words and expressions into completed utterances in a wording that is appropriate for either formal or informal conversational contexts. Thereby, ChatGPT-4.0 as a Generative AI technique is employed because of its impressive capabilities in formulating human-like texts (cf. Feuerriegel et al. 2023; Radford et al. 2019; Teubner 2023). As a Generative Pretrained Transformer (GPT), it combines large-scale architectures with vast amounts of textual training data, resembling contributions of humans on diverse daily life conversations, communication activities, and conversational styles (i.a., for formal and informal situations). However, the integration of ChatGPT-4.0 into an AAC system is not purposeful without additional and extensive adaptions, which have been realized in the design and implementation of the artifact. Furthermore, ChatGPT-4.0 alone is insufficient to meet all the DRs. Consequently, further AI and machine learning techniques were integrated and purposefully combined in order to create a comprehensive Generative AI-based utterance composition within the AAC system prototype. This proposed AI-based approach was evaluated with several communication activities commonly performed in the daily life routine of visiting a restaurant with a companion (cf. Ehlich and Rehbein 1972; Schindler 2013). Against this background, the following two Research Questions (RQs) are posed:

- ***RQ1:*** *How to design a Generative AI-based approach that can compose within an AAC system single expressions into completed utterances in formal and informal wordings?*
- ***RQ2:*** *How to instantiate and technically realize the Generative AI-based approach?*

The paper unfolds as follows: The following section provides the conceptual foundations and related work. Next, the research procedure following the design science approach (Hevner et al. 2004; Peffers et al. 2007) is described. The next section deals with the compilation of the Design Principles (DPs) that have been derived from the DRs. Then, the technical realization of the proposed AI-based approach, its demonstration to interactions in visiting a restaurant, and the results are presented. The paper proceeds with its contributions, potential limitations, and future research ideas. Afterwards, it closes with concluding remarks.

Conceptual Foundations and Related Work

Conceptual Foundations

Communication research and extant theories like the Gricean Maxims of Communication (Grice 1975) have investigated structures and components of daily life conversations (i.a., conversation phases, communica-

tion activities, and communication strategies) (Henne and Rehbock 2012). The way communication activities are performed is determined by different factors. In this regard, one of the most important factors is the conversational context (Bedrosian et al. 2003; Hoag et al. 2004; Todman and Alm 1997). Therefore, interlocutors use different articulations depending on where a conversation takes place and who they are speaking with. Articulations can differ in conversations with friends in restaurants compared to meetings with superiors at work. Furthermore, articulations are adjusted based on the complexity and cognitive demands of the conversation (cf. Cognitive Load Theory, Sweller 1988) and based on whether a social relationship has already been established with the conversation partner (social closeness to friends vs. social distance to strangers or authority figures).

Interactions with a friend (e.g., talking about recent events in one's life, telling personal stories, or thanking for a nice evening in a restaurant) are rather socialness-oriented and informal (cf. Todman and Alm 1997). Compared to that, interactions with waiters as authority figures in the restaurant (e.g., asking for the booked table, asking for the specials of the week, asking for payment) are rather informativeness-oriented and formal (cf. Bedrosian et al. 2003; Hoag et al. 2004). These differences are also reflected in the way interlocutors accommodate speech and the concrete wordings they apply within their utterances for the respective contexts (cf. Giles et al. 1991). In the German language, for example, interlocutors in an informal context tend to address each other by their first names and are more liable to apply contractions to words. Thereby, they may incorporate dialect, slang, and rather unsophisticated expressions into their utterances. In contrast, in a formal context, interlocutors are more likely to use the formal term of address (i.a., using last names) and employ more elaborated language constructs (i.a., euphemisms, precise terminology, forms of courtesy) (cf. Bedrosian et al. 2003; Budde 2012; Elspaß 2018; Hoag et al. 2004; Todman and Alm 1997). Speech-impaired individuals lack the fundamental capability to perform communication activities using their own vocal language in a style that fits the respective conversational contexts. Nonetheless, vocal language is an essential human capability for interacting with other individuals in various daily life routines like visiting a restaurant with a companion.

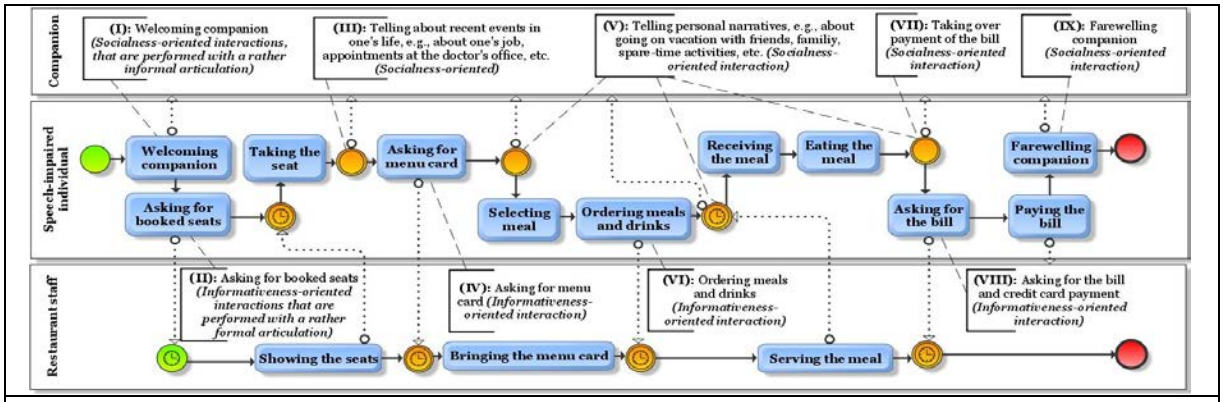

Figure 1. Socialness- and Informativeness-Oriented Interactions when Visiting a Restaurant with a Companion (own Representation, based on Ehlich and Rehbein 1972; Schindler 2013)

AAC encapsulates strategies, techniques, and tools that support individuals with speech impairments in expressing their thoughts, ideas, and feelings (Hourcade et al. 2004). In this context, AAC systems can be described as voice output communication aids or speech generating devices that provide speech-impaired individuals with the ability to express themselves using vocal language. High-tech AAC systems encompass the use of electronic devices (e.g., smartphones, tablet PCs) and their built-in peripherals (e.g., cameras, microphones) (Baldassarri et al. 2014; Laxmidas et al. 2021). The board within an AAC system contains the vocabulary that users can employ. Cognitive and language comprehension disorders may also require assigning pictograms to utterances or words to support the understanding of the related concepts. To contribute utterances to conversations, users of AAC systems can relate to pre-stored sentences (Alm and Higginbotham 2008), manually compose single words into sentences within the so-called speech composition line, or make use of keyboard capabilities and combinations of letters to express words and sentences (cf. Klauer et al. 2021; Kristensson et al. 2020). The utterance formed is then processed by the speech synthesis module which generates the voice output that is emitted to the conversation partner by the loudspeaker of the device. Voice outputs can then be interpreted and reciprocated by the interlocutor, allowing the conversation to progress by mutual and coordinated utterances until it reaches a targeted and desired state.

To help speech-impaired individuals compose completed sentences from single expressions, there have been proposed different AI-based approaches (i.a., abbreviation expansion, sentence prediction) (e.g., Dempster et al. 2010; Heo and Kang 2019) and recently also Generative AI (e.g., Cai et al. 2022; Valencia et al. 2023). As elaborated in the next two sections, Generative AI has yet not been harnessed to construct an AAC system with a comprehensive Generative AI-based utterance composition. Crucial aspects regarding the style of the utterances (e.g., matching expected and common styles of articulation in formal and informal contexts, including personal nuances and styles of articulation into the utterances) are thereby not met. Thus, in the following section, DRs derived from related work and extant theory are outlined, specifying what a Generative AI-based utterance composition needs to meet to be comprehensive and purposeful.

Design Requirements of AI-Based Utterance Composition for AAC Systems

To derive the DRs for the Generative AI-based utterance composition approach, an extensive search and consolidation of the research literature was conducted. Thereby, it was aligned with the procedure for conducting a systematic literature review as proposed by Vom Brocke et al. (2015). In the **first step**, a keyword search was performed on *ACM Digital Library, AIS Electronic Library, EBSCOhost, Emerald Insight, Google Scholar, IEEE Xplore, SAGE Publications, and SpringerLink*. According to Snyder (2019), the search strings for conducting a literature review should refer to words and concepts that are directly linked to the research questions. Thus, *"augmentative and alternative communication", "artificial intelligence", "sentence generation", "generative artificial intelligence",* and several combinations and abbreviations thereof were derived as search terms. To obtain even further keywords, iterative keyword refinement was performed *(e.g., "natural language generation", "sentence composition", "large language model")*. Additionally, it was related to the systematic literature review by Konadl et al. (2023), who have identified and assessed the current state of AI approaches applied within AAC systems.

As a result, *1,129 papers* could be initially collected whose titles, abstracts, and thematic foci were then investigated. In general, only papers dealing with the generation of completed utterances based on single words or expressions using an AI-based approach were retained. Papers that included further external devices, which hinder the portability of the AAC systems in daily life situations, were removed. In applying these two inclusion criteria to the *1,129 papers*, several decisions about the inclusion and exclusion of the papers were documented and validated with two additional researchers. To assess the coding and the agreement between the three raters, interrater reliability has been calculated using Fleiss' Kappa (Fleiss et al. 2013). The rating procedure yielded Fleiss' Kappa values of 0.8772 for the first inclusion criterion and 0.9072 for the second inclusion criterion, both indicating high levels of agreement (Fleiss et al. 2013). In the very few cases of disagreement, the coding was discussed to come to a consensus about including or excluding the respective paper. Next, duplicates and preprints were eliminated, and only papers written in English were retained. In the **second step**, forward and backward searches have been conducted, resulting in two additional relevant papers. As follows, a total of *19 relevant papers* could be derived.

Subsequently, the contents of these papers were investigated. Using qualitative content analysis and inductive category development, it could be carried out an abstraction from the basic data, while, at the same time, deriving a coherent image of these data (cf. Mayring 2000; Mayring 2002). To inductively extract the DRs from the papers, the notion of what constitutes a DR was applied. In doing so, a DR is considered as any specific condition, capability, or functionality that the intended artifact needs to implement in order to achieve its intended goals and purposes (cf. Möller et al. 2020; Walls et al. 1992). Based on that, the 19 relevant papers have been screened regarding any text statement that hinted at a DR. Several text statements hinting at a DR were consolidated, and based on them, corresponding DRs were deduced. These resulting DRs were then verified with two additional researchers. Thereby, a Fleiss' Kappa value of 0.8871 could be obtained, indicating a high level of interrater agreement (Fleiss et al. 2013). In case of any disagreement about a DR, the respective DR was analyzed again, followed by discussions until a consensus was reached. Based on the attention and importance that the respective DRs have received within extant research literature, **nine DRs** have been derived that will be described in more detail in the following:

AAC systems enable speech-impaired individuals to articulate their thoughts, ideas, and feelings within daily life conversations. Thereby, if speech-impaired individuals must sense on their own for signs of the respective conversational context, this puts additional cognitive efforts on them. This, in turn, may result in decelerations of the conversational flow. This is underlined by Cognitive Load Theory in that the amount of information that can be processed within the working memory is limited (cf. Sweller 1988). Exceeding the cognitive capacity of speech-impaired individuals could hinder their full concentration on the ongoing

conversation. Therefore, **(DR1)** the AI-based utterance composition approach for AAC systems needs to incorporate an automated identification of formal and informal conversational contexts (e.g., Arnott and Alm 2013; Black et al. 2012; Fang et al. 2023). In this vein, **(DR2)** the applied language model must resemble formal and informal conversational styles to produce sentences for both contexts (e.g., Dempster et al. 2010; Shen et al. 2022; Valencia et al. 2023). There are individual styles of articulation in formal and informal contexts. For example, one may tend to apply in an informal context rather unsophisticated language and expressions (e.g., applying slang and dialect, calling the interlocutor "buddy" or "bro"). Compared to that, others may tend to apply a more reserved wording in an informal conversation (e.g., Cai et al. 2022; Tintarev et al. 2016; Valencia et al. 2023). Therefore, the utterance composition approach **(DR3)** needs to provide means to tailor the utterances towards the personal style of articulation. Furthermore, as Grice (1975) states in the "Maxim of Relation", it is expected by interlocutors that several conversational contributions are appropriate at each stage. So, it is essential that **(DR4)** the utterance composition is adaptive based on the conversation history and that it incorporates the preceding turns of a conversation (e.g., Kristensson et al. 2020; Mitchell and Sproat 2012; Shen et al. 2022). This also requires **(DR5)** to generate statement and question sentences so that the speech-impaired individuals can apply appropriate conversational acting at different stages of a conversation (i.a., sharing ideas and information with others, asking for information from others) (cf. Grice 1975, and e.g., Black et al. 2012; Garcia et al. 2016; Tintarev et al. 2016). The utterance composition approach **(DR6)** must generate completed utterances that are grammatically correct and semantically tailored to the user input (e.g., Adhikary et al. 2021; Garcia et al. 2016; Mitchell and Sproat 2012). This is necessary to maintain the impression of communicative competency of the speech-impaired individuals towards their interlocutors because these errors transmit to the speech synthesis and are reflected within the voice output. Beyond that, these are functionalities that relieve speech-impaired individuals in composing completed utterances. In this way, they can reduce the mental efforts of speech-impaired individuals and increase their level of participation in conversations (cf. Sweller 1988).

Within AAC systems, it ultimately needs to remain the speech-impaired individual whose thoughts, ideas, and feelings are communicated. Thus, the generated utterances must be semantically tailored towards the words and expressions that the users provide. To ensure that speech-impaired individuals maintain complete control over the generated utterances, **(DR7)** the utterance composition approach must allow them to modify the utterances before they are outputted by means of voice (e.g., Mitchell and Sproat 2012; Shen et al. 2022; Valencia et al. 2023). Speech-impaired individuals using a digital AAC system must be able to keep up with the pace of daily conversations. Therefore, **(DR8)** the generated conversational contributions must be provided immediately. Since conversations are central actions in daily life that consist of multiple interaction steps, **(DR9)** the approach needs to be repeatable at any time (e.g., Arnott and Alm 2013; Cai et al. 2022; Dempster et al. 2010; Fang et al. 2023; Walsh 2010).

Design Requirements	Sources
DR1. Support the automated identification of the conversational context.	Arnott and Alm 2013; Bedrosian et al. 2003; Black et al. 2012; Fang et al. 2023; Hoag et al. 2004; Sweller 1988; Todman and Alm 1997
DR2. Incorporate a language model that can produce utterances that match the style of articulating in the respective conversational contexts.	Bedrosian et al. 2003; Dempster et al. 2010; Hoag et al. 2004; Shen et al. 2022; Tintarev et al. 2016; Todman and Alm 1997; Valencia et al. 2023
DR3. User-adaptive tailoring of the utterances to be generated towards individual styles of articulation and nuances of the speech-impaired individual.	Black et al. 2012; Cai et al. 2022; Grice 1975; Shen et al. 2022; Tintarev et al. 2016; Valencia et al. 2023
DR4. Adaptive sentence generation based on the conversation history.	Black et al. 2012; Dempster et al. 2010; Grice 1975; Kristensson et al. 2020; Mitchell and Sproat 2012; Shen et al. 2022; Sweller 1988
DR5. Support the automated generation of statement sentences and question sentences.	Black et al. 2012; Garcia et al. 2016; Grice 1975; Reddington and Tintarev 2011; Tintarev et al. 2016; Walsh 2010
DR6. Automated generation of completed sentences that are grammatically correct and semantically tailored to the provided user input.	Adhikary et al. 2021; Alm and Higginbotham 2008; Cai et al. 2022; Dempster et al. 2010; Garcia et al. 2016; Grice 1975; Mitchell and Sproat 2012; Shen et al. 2022; Sweller 1988; Valencia 2023
DR7. Provide flexible modification of the generated sentences before they are outputted by voice.	Adhikary et al. 2021; Alm and Higginbotham 2008; Black et al. 2012; Kristensson et al. 2020; Mitchell and Sproat 2012; Shen et al. 2022; Tintarev et al. 2016; Valencia et al. 2023
DR8. Provide generated sentences immediately.	Alm and Higginbotham 2008; Cai et al. 2022; Dempster et al. 2010; Fang et al. 2023; Heo and Kang 2019; Kristensson et al. 2020; Reddington and Tintarev 2011; Shen et al. 2022; Tintarev et al. 2016; Valencia et al. 2023
DR9. Be repeatable and to be triggered by the speech-impaired individual at any time.	Arnott and Alm 2013; Cai et al. 2022; Dempster et al. 2010; Fang et al. 2023; Shen et al. 2022; Valencia et al. 2023; Walsh 2010

Table 1. Design Requirements of an AI-based Utterance Composition Approach

Assessment of Available AAC Systems on the Market

To assess the capabilities of the AAC systems available on the market against the background of the derived DRs, corresponding AAC system providers were first identified. For this purpose, the Google search engine as well as the search engines of the Google Play Store and the Apple App Store were drawn upon. By verifying several steps in the identification of the AAC systems with two additional researchers, the possibility of overlooking established tool providers whose systems offer an automated or an AI-based utterance composition could be tremendously reduced. The resulting set of tool providers was then consolidated and any disagreements about the tools were resolved. In this way, a broad spectrum of providers could be covered, and it could be taken an up-close look at the most established AAC systems in the market. Based on the derived DRs *(see Table 1)*, a tool survey was designed to assess whether the available AAC systems cover the posed requirements. The tool survey was then validated and subsequently applied to the data collection procedure. To gather the data for assessing the AAC systems, sales representatives of the identified AAC system providers were contacted regarding the specific features. Additionally, demo versions of the AAC systems were installed to test them out. Based on the providers' responses and by independently testing the functionalities of the AAC systems against the posed DRs, detailed and comprehensive insights into the capabilities and drawbacks of these AAC systems could be gained. The table below presents the results of the tool assessment for the resulting set of eleven AAC system providers: *Assistiveware, Attainment Company, Avaz, Coughdrop, Livox, Prentke Romich Company, Smartbox, Symbotalker, Therapybox, Tobii Dynavox, and Willowtree*. Due to reasons of confidentiality posed by the providers, it was necessary to anonymize their answers. Thus, the providers are labeled below with #1–#11 and their order is randomized.

Design Requirements	Results of the Tool Assessment										
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11
DR1	-	-	-	-	-	-	○	-	-	-	○
DR2	○	-	-	-	○	○	-	○	-	-	○
DR3	○	-	○	○	-	-	-	-	-	-	-
DR4	-	-	-	-	●	-	-	●	-	-	●
DR5	○	○	○	○	○	-	-	○	-	○	●
DR6	○	○	○	○	●	○	○	●	○	○	●
DR7	●	●	●	●	●	●	●	●	●	●	●
DR8	DR8 has **not** been included in the comparison because an immediate provision of the completed utterance is also dependent on the respective device the AAC system is applied on.										
DR9	●	●	●	●	●	●	●	●	●	●	●

Table 2. Results of Assessing the Capabilities of Available AAC Systems against the Background of the Derived Design Requirements

Legend:
● = DR is fully met
○ = DR is partly met
- = DR is not met

As has been revealed, there are AAC systems that provide an automated composition of completed sentences. However, essential requirements like context- and user-specific styles of articulation as well as the automated identification of the current conversational context are thereby not met *(see DR1–DR3 in Table 2)*. Without AAC systems that offer speech-impaired individuals completed utterances in both situational and personal styles, while also being morphologically and grammatically correct, their separation from the mainstream of society will continue. Research literature (cf. Konadl et al. 2023) as well as the affected individuals, caregivers, and representatives *(see Section "Introduction")* clearly point out the necessity of AAC systems that support both formal and informal contexts with appropriately-styled utterances. To contribute to dismantling the communication barriers that speech-impaired individuals still suffer from across various daily life conversations, this research aims to close this gap.

Research Procedure

To accomplish the development of the approach that enables speech-impaired individuals to generate completed utterances in both formal and informal styles within an AAC system, Design Science Research (DSR) was applied (cf. Gregor and Hevner 2013; Hevner et al. 2004). In doing so, the procedure outlined by Peffers et al. (2007) was followed *(see Figure 2)*.

As a first step, **(1) corresponding problems** of existing software tools and approaches were identified *(see Section "Introduction" and Table 2)*. Available AAC systems are not capable of generating completed sentences for formal and informal conversational contexts. However, covering both contexts within AAC systems is indispensable for paving the way for speech-impaired individuals into the mainstream of society and for enabling them context- and user-specific articulation. Thus, the **(2) objective** is to combine, based

on the DRs, a set of machine learning techniques. Thereby, the Generative AI technique ChatGPT-4.0 is purposefully integrated to facilitate a comprehensive composition of completed utterances from single expressions as speech-impaired individuals provide them within AAC systems.

The third step contains the **(3) design and development** of the solution *(see Section "Design and Development of the Artifact)*. By relating to extant theory (Grice 1975; Sweller 1988) and by conducting a systematic literature review (cf. Vom Brocke et al. 2015), there could be identified the DRs the AI-based approach needs to meet. To close the gaps identified within step (1), it was focused on the design and technical realization of the Generative AI-based utterance composition. Beyond that, it was developed an AAC system prototype that employs this proposed AI-based approach. Regarding the design of the approach, there have been established DPs that are founded on the identified DRs. Thus, the artifact was established to provide speech-impaired individuals with a comprehensive utterance composition, enabling user-specific articulation and utterances that fit the expected style of articulation in formal and informal contexts.

Steps (4) and (5) deal with the **demonstration and evaluation** of the generated artifact. First, a proper setting was defined for rigorously demonstrating and evaluating the utility of the artifact. In this regard, it was related to Venable et al. (2016) who defined four evaluation strategies for DSR projects. In line with that, the DSR-project-specific evaluation strategy was aligned to the "Technical Risk & Efficacy Strategy" to rigorously demonstrate that the evaluand achieves its intended purpose (i.e., composing single expressions into completed utterances in a style that fits formal and informal conversational contexts and that reflects personal nuances of articulation). In addition, the aim is to show that meeting all the derived DRs *(see Table 1)* and purposefully harnessing ChatGPT-4.0 as a Generative AI technique are thereby imperative (cf. March and Smith 1995; Peffers et al. 2012). Thus, it was decided to apply one summative evaluation episode (cf. Venable et al. 2016). In doing so, the developed AAC system prototype equipped with the AI-based approach was applied to 61 communication activities of the common daily life routine of visiting a restaurant with a companion (cf. Ehlich and Rehbein 1972; Schindler 2013) *(see Figure 1)*. The composed utterances were then assessed based on characteristics for the formal and informal context of the German language (cf. Budde 2012; Elspaß 2018) and whether the provided individual nuances of articulation are reflected within the generated sentences. Beyond that, it is planned to apply the "Human Risk & Effectiveness Strategy" (cf. Venable et al. 2016) and to investigate the constructed artifact within a large-scale empirical study in a controlled environment (e.g., within a laboratory experiment). In **step (6)**, the aim is to **document and communicate** the results of the research at hand.

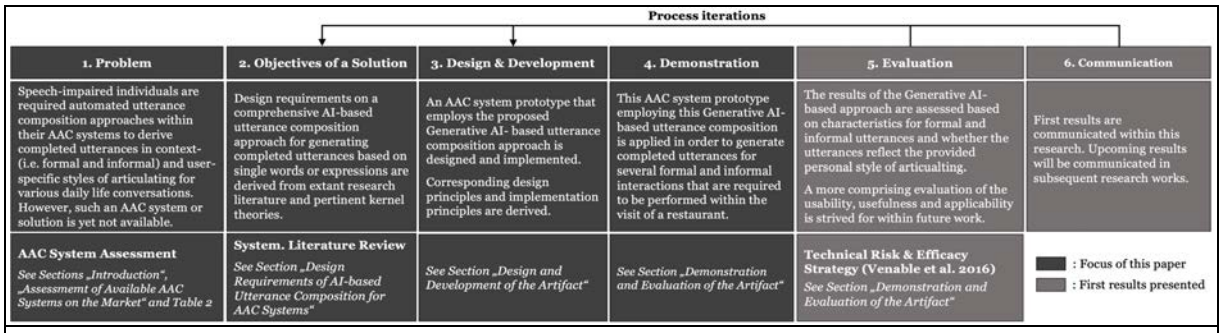

Figure 2. Procedure of the Research (own Representation, based on Peffers et al. 2007)

With this procedure, the research also aligns with the guideline of Hevner et al. (2004). Regarding the design cycle, the artifact is presented as the result that has gone through the process of demonstration and evaluation *(see Section "Demonstration and Evaluation of the Artifact")*. Regarding the relevance cycle, several DRs from research literature and theory are identified *(see Table 1)* that guided the design of the artifact. In this way, the practical application of the generated artifact brought up several contributions for practice *(see Section "Contributions to Theory and Practice")*. In view of the rigor cycle, several techniques (ChatGPT-4.0, automated speech recognition, natural language processing) were applied to rigorously construct the artifact. There have been derived initial findings as contributions to theory, including both kernel theory (Gricean Maxims of Communication, Cognitive Load Theory) and design theory. To contribute to a rather general and abstract knowledge base – "Nascent design theory" (Gregor and Hevner 2013) – and to design a purposeful artifact in a comprehensible way, there have been deduced meta requirements and DRs (Gregor and Jones 2007; Heinrich and Schwabe 2014) for an automated (AI-based) utterance composition for AAC systems grounded in seminal works, which resulted in the DPs.

Design and Development of the Artifact

Design of the AI-Based Utterance Composition for AAC Systems

The meta requirements describe *"what the system is for"* (Gregor and Jones 2007, p. 325). In this way, meta requirements constitute solution objectives that are based on the clearly defined purpose and scope of the artifact and thus, are based on the problems the investigation addresses as discussed in the introduction. Therefore, it is for example identified as one meta requirement the objective of the proposed AI-based approach to empower speech-impaired individuals to express their needs, wishes, and wants by means of voice. Thereby, these meta requirements to be suitable for a class of artifacts are derived based on the current research literature (Gregor and Jones 2007; Heinrich and Schwabe 2014; Walls et al. 1992).

Subsequently, DPs have been synthesized based on the DRs that have been derived from related work (i.e., extant literature) and based on DRs that are inspired by theory (cf. Möller et al. 2020; Purao et al. 2020). These DRs constitute specific conditions, capabilities, and functionalities that an artifact needs to implement in order to enable its intended users to fulfill certain tasks within a specific problem class (cf. Möller et al. 2020; Walls et al. 1992). The DRs that semantically correspond to each other constitute conceptual influences that need to be considered when proposing a DP. Thus, the derived DRs were carefully examined and DRs that are semantically related to each other have been consolidated into a respective DP (cf. Möller et al. 2020; Purao et al. 2020). For example, DP3 was deduced based on DR2, DR6, and DR8, as it is necessary to automatically provide sentences that are both completed and correct. Hence, the applied language model needs to comprise both capabilities. In addition, because it is essential for the composition approach to provide the completed sentences immediately, this aspect was also considered within this DP.

These DPs are defined as generic and prescriptive statements, showing how to do something to achieve a certain goal (Gregor et al. 2020). In this way, the proposed DPs foster the transfer and the application of the design-related knowledge to similar problem classes beyond the specific problem class of Generative AI-based utterance composition within AAC systems. The disposed DPs fall into the category of action and materiality-oriented DPs that describe what an artifact should enable users to do and how the artifact should be built in order to do so (Chandra et al. 2015). Speech-impaired individuals (= users) are required to compose single expressions into completed sentences to communicate and collaborate within formal and informal contexts in daily life situations (= boundary conditions). Keeping the DRs of the artifact in mind, there have been derived **five DPs** regarding the proposed Generative AI-based approach.

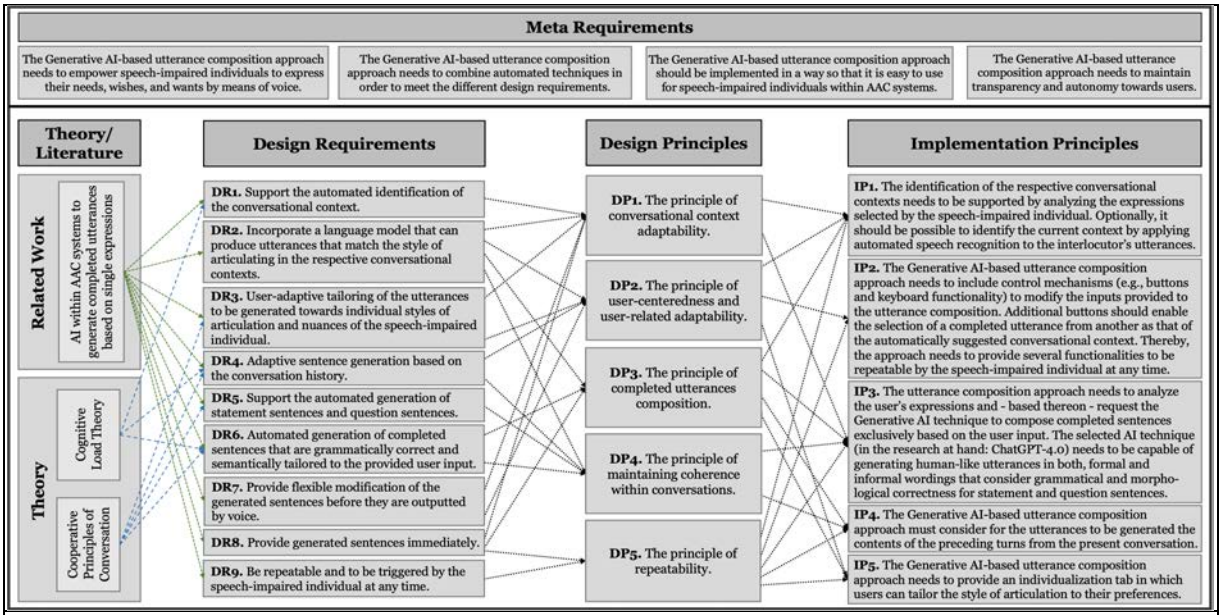

Figure 3. Derived Design of the Generative AI-Based Utterance Composition Approach

- **DP1: The principle of conversational context adaptability.** To enable speech-impaired individuals to apply situation-adaptive wordings within conversations, the Generative AI-based utterance com-

position for an AAC system needs to support the tailoring towards the respective conversational context by sensing and proposing conversational contributions for the current context.

- **DP2: The principle of user-centeredness and user-related adaptability.** To consider individual styles of articulation and to give speech-impaired individuals control regarding the AI-generated conversational contributions, it is necessary to enable speech-impaired individuals to include personal nuances and expressions into the Generative AI-based utterance composition. Thereby, the approach needs to provide flexible means to modify a generated sentence before it is outputted by means of vocal language.
- **DP3: The principle of completed utterances composition.** The Generative AI-based utterance composition approach needs to include the user-selected, single expressions and form them into completed sentences by harnessing the relationships from a (large) language model. By that, these models need to cover various daily life situations, daily communication activities, and conversational contexts.
- **DP4: The principle of maintaining coherence within conversations.** To foster communicative effectiveness, the Generative AI-based utterance composition needs to take account of correctness (i.a., morphological, grammatical, spelling) and situational fit of the generated contributions (i.a., relatedness to previous utterances in a conversation, proposing the required sentence type).
- **DP5: The principle of repeatability.** As conversations consist of multiple interaction steps, the Generative AI-based utterance composition needs to provide repetition of several functionalities at any time and to enable speech-impaired individuals to immediately act and react within conversations.

These DPs are deduced from the DRs that are based on current research literature and theory. Gregor and Jones (2007) state that this reference to theory and literature discloses *"an explanation of why an artifact is constructed as it is and why it works"* (p. 328). Based on related work and theory, DRs were derived that an AI-based utterance composition approach to be employed in AAC systems should meet. These DRs offer guidance by designing the artifact and advising the DPs (Böckle et al. 2021; Gregor and Jones 2007). The DPs refer to at least one DR and serve as an abstract "blueprint" of the artifact (Böckle et al. 2021; Gregor and Jones 2007; Heinrich and Schwabe 2014). By establishing the DPs, it is ensured that they follow the value grounding (reference to the DRs) and explanatory grounding (DPs are based on current literature and kernel theories) (Heinrich and Schwabe 2014). Furthermore, based on the instantiation of the DRs and DPs within the generated artifact, Implementation Principles (IPs) are defined. In this way, it is supported *"the implementation in practice of an abstract, generic design method or development approach"* (Gregor and Jones 2007, p. 329) in view of an artifact for (Generative) AI-based utterance composition for AAC systems.

Technical Realization

To address the drawbacks of prevailing research and existing AAC systems, the artifact was designed and developed in the programming language Python. There has resulted the Generative AI-based utterance composition approach. This proposed AI-based approach has been integrated into a more comprehensive AAC system prototype, enabling users to apply and interact with the approach. In the following, the instantiation of the DRs is described, as they depict the required features of the proposed AI-based approach in the most detailed way. Since the DPs and IPs depict a generic and prescriptive statement of how something should be done, they capture design-related knowledge. In this way, these IPs support the development of further IS (design) theories and new artifacts focusing on AI-based utterance composition for AAC systems. To reveal the association of the derived DPs and IPs to the DRs, their relations are pointed out as well.

Regarding the adaption of utterances to the current context within a conversation, it is essential to include a functionality within the AI-based approach that automatically identifies the conversational context. A technique that has been applied for this task within extant AAC research is automated speech recognition (e.g., Higginbotham et al. 2012; Kane et al. 2017; Neamtu et al. 2019). The Google cloud speech-to-text engine was chosen because of its impressive capabilities in correctly recognizing German expressions and sentences (Filippidou and Moussiades 2020). It records the utterances of the interlocutor and analyzes them for signs of formal (i.a., formal terms of addressing the interlocutor, addressing him/her with the last name) and informal expressions (i.a., informal terms of addressing the interlocutor, addressing him/her with the first name) (cf. Budde 2012; Elspaß 2018). Thereby, a manually defined list of common German expressions for the formal and informal contexts is applied. Automated speech recognition is yet not purposeful in every situation (e.g., noisy locations). Hence, a second functionality for identifying the context is implemented that applies common German expressions for formal and informal contexts to the expressions that the speech-impaired individual has selected **(DR1/DP1/IP1)**. In this vein, it is mandatory that the harnessed language model can generate contributions for the respective conversational contexts. ChatGPT-

4.0 (OpenAI 2024) was chosen as its large language model comprises data on conversations from various daily life situations. Prompt engineering was applied to create and iteratively refine the input given to ChatGPT-4.0 until eliciting the desired task behavior (Teubner et al. 2023). ChatGPT-4.0 is thereby employed in the few-shot learning mode (Feuerriegel et al. 2023), as it is provided with examples of characteristics for formal and informal utterances of the German language (cf. Budde 2012; Elspaß 2018). In this way, it is aimed that the artifact reflects the provided characteristics for the formal and informal contexts within the generated utterances. As a result, it is derived a prompt that is sent to GPT-4.0 via its API, which makes ChatGPT-4.0 compose a completed utterance from the provided expressions in the wording for the respective conversational context **(DR2/DP1/IP1, DR2/DP3/IP3)**. Conversational styles can contain specific nuances and expressions that vary for different contexts and individuals. For that reason, the generated artifact includes a separate tab area in which the users can enter expressions they are keen on applying given certain conversational contexts and situations. These expressions are incorporated into the prompts which are sent to GPT-4.0 for composing the utterances **(DR3/DP2/IP5)**.

Low coherence of conversational contributions could lead to deviations in the course of a conversation. Speech-impaired individuals could be required to provide information and to ask questions when they lack information. Spelling errors as well as morphological and grammatical errors are directly reflected within the voice output, hindering the perception of communicative competency by others towards the speech-impaired individuals. To reflect their thoughts, ideas, and feelings, the generated utterances must be semantically tailored to the inputs that the speech-impaired individuals provide. These three requirements can all be met by the capabilities of ChatGPT-4.0. It maintains the coherence of the generated contributions as it automatically relates the contents of the utterances to be generated to the themes of the preceding turns within the ongoing conversation **(DR4/DP4/IP4)**. For generating questions, the GUI includes a question mark symbol that adds via click a question mark to the speech composition line **(DR5/DP4/IP3)**. The large language model applied within ChatGPT-4.0 is trained on large amounts of text data that have comprised statement and question sentences. In addition, the training procedure has adhered to the rules of correct spelling, grammar, and morphology (cf. Radford et al. 2019). Regarding semantically tailoring the generated sentences to the user input, the user-selected expressions are directly integrated to be essential units of the utterance to be composed within the prompt for requesting the utterance composition **(DR6/DP3/IP4)**. In case that the generated sentences digress from the user input even after multiple repetitions of the generation procedure, the artifact provides the user with flexible and multiple means to adapt the contributions before voice output **(DR7/DP2/IP2)**.

The generated utterances must be provided immediately to the speech-impaired individuals. Here, the proposed artifact implements multi-threading and immediate processing of information, resulting in an instant provision of the generated sentences. Furthermore, the generated conversational contributions for the detected context are directly inserted into the speech output line, fostering their immediate application **(DR8/DP1/IP1)**. Conversations consist of multiple interaction steps so that several functionalities can be triggered and repeated by the speech-impaired individual any number of times **(DR9/DP5/IP1–5)**.

The derived DPs and IPs as prescriptive design-related knowledge have all been considered within a concrete instantiation. Consequently, a developed AAC system prototype has resulted that employs the Generative AI-based utterance composition. Thus, the artifact is capable of providing for speech-impaired individuals context- and user-specific styles of articulation across various daily life conversations. To show the applicability of this instantiation, the next section proceeds with its demonstration and evaluation.

Demonstration and Evaluation of the Artifact

The demonstration and evaluation were aligned to the defined evaluation strategy *(see Section "Research Procedure")* and to the routine of visiting a restaurant with a companion *(as outlined in Figure 1)*. This situation was chosen as it is a commonly performed daily life activity, and collaborating and engaging within it requires individuals to perform several formal and informal interactions (cf. Ehlich and Rehbein 1972; Schindler 2013). Beyond that, it has also received considerable attention within extant AAC research (e.g., Kane et al. 2017; Obiorah et al. 2021; Valencia et al. 2023). In the following, there is described **(a)** the handling of the proposed AI-based approach. Then **(b)**, the applied metrics are elaborated, before **(c)** relating to the two iterations back to the phase "Design & Development" and **(d)** to the reporting of the results.

(a). To cover a comprehensive set of communication activities, 121 German vocabulary entries were defined (i.a., conversation openers and closers, foods and drinks, daily activities and situations). Clicking the

button of a vocabulary entry leads to its insertion into the speech composition line. To detect the conversational context, the user can apply the speech recognition functionality to analyze the interlocutor's utterance *(see Figure 4)*. The *"Sprachpersonalisierung" (speech personalization)* tab *(see Figure 5)*, enables the provision of personal nuances and expressions of articulation. After making any modifications to the selected expressions within the speech composition line *(see Figure 4)*, the proposed AI-based approach can be triggered by clicking the *"Satz generieren" (generate sentence)* button. As a result, a completed sentence is generated and proposed for the current conversational context within the speech output line, so that it can be immediately applied. If another context is considered more suitable in a certain situation, the corresponding utterance for this context can be flexibly chosen. By clicking the speech output button, the sentence in the speech output line is voiced in the user-selected volume and speed.

Figure 4. Speech Generation and Speech Output View

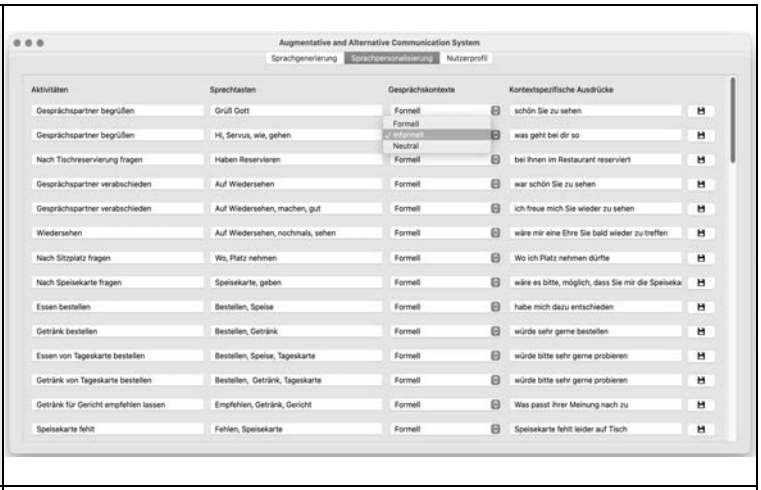

Figure 5. Speech and Articulation Personalization View within the proposed Artifact

(b). Based on the 121 German vocabulary entries, 61 communication activities were defined for the daily life routine of visiting a restaurant with a companion (e.g., Ehlich and Rehbein 1972; Kane et al. 2017; Obiorah et al. 2021; Schindler 2013; Valencia et al. 2023). These 61 communication activities served as the basis for assessing the capabilities of the proposed AI-based approach *(see Table 3 which illustrates the capabilities of the approach on a subset of nine communication activities from the total set of 61)*. On the one hand, it was investigated whether the peculiarities for articulating in formal and informal contexts are reflected within the composed sentences. For that reason, for both conversational contexts, the true positive and false positive rates are reported. The true positive rate counts the number of times in which the respective context is correctly reflected within the generated sentences. In contrast, the false positive rate reports the number of times in which the Generative AI-based approach generates sentences in the formal or informal context, despite it is expected to generate a sentence for the opposite context. The generated sentences were analyzed to determine whether they contained any characteristics of the formal or informal context (cf. Budde 2012; Elspaß 2018). Based on the expected and the determined contexts, the number of times the sentences reflected the correct or incorrect context was counted. On the other hand, it was investigated whether the generated sentences comprised a completed structure (i.a., correct word orders for statement and question sentences) and whether they were morphologically and grammatically correct (i.a., correct conjugation of verbs, correctly inserted filling words, and prepositions). Consequently, the completeness of the structure of the composed sentences and whether they adhered to the themes of previous conversational turns were also investigated. If any errors regarding these criteria have been observed within the utterances composed by the Generative AI-based approach, the corresponding number of errors is reported in (d).

Communication Activities	User Input (German)	GPT-Generated Sentences (German)	Translation to English
(I) Welcoming companion	Hallo, wie, gehen	Hi, wie geht's dir denn so? (Informal)	Hi, how are you doing?
(II) Asking for the booked seats	Wir, haben, reservieren	Grüß Gott, wir haben bei Ihnen im Restaurant einen Tisch reserviert. (Formal)	Excuse me, we have booked a table at your restaurant.
(III) Telling about recent events in one's life	Arbeit, stressig	Ich hab' aktuell echt gut zu tun in der Arbeit. (Informal)	There is currently much stress for me at work.
(IV) Asking for the menu card	Speisekarte, geben, ?	Entschuldigen Sie, aber könnten Sie mir bitte eine Speisekarte bringen? (Formal)	Excuse me, but could you please bring me a menu card?

(V) Telling personal narratives	Ich, fahren, bald, Urlaub, mit, Freunden	Ich werde bald mit meinen Kumpels in den Urlaub fahren. (Informal)	I will soon go on vacation with my buddies.
(VI) Ordering meals and drinks	Ich, Bestellen, Fleisch, Tageskarte	Ich würde bitte gerne das Fleischgericht von der Tageskarte bestellen. (Formal)	I would like to order the meat dish from the daily menu.
	Nehmen, Beilage, Pommes	Als Beilage würde ich bitte die Pommes nehmen. (Formal)	As a side I would appreciate to take the fries.
(VII) Taking over the payment of the bill	Ich, heute, übernehmen, Rechnung	Die Rechnung geht heut auf mich. (Informal)	Today's bill is on me.
(VIII) Asking for the bill and for credit card payment	Rechnung, bringen, ?, Ich, bezahlen, mit, Kreditkarte	Entschuldigen Sie, könnten Sie mir bitte die Rechnung bringen? Ich würde bitte gerne mit Kreditkarte bezahlen. (Formal)	Excuse me, could you please bring me the bill? I would like to pay by credit card please.
(IX) Farewelling companion	Tschüss, machen, gut, nochmals, sehen	Tschüss, mach's gut ich freu' mich dich bald wieder zu sehen. (Informal)	Bye, take care, I look forward to seeing you again soon.

Table 3. Utterances Generated by the Generative AI-Based Composition Approach

(c). During the demonstration and evaluation, there were observed cases where the prompt to compose a completed utterance based on the provided expressions was not followed correctly. Therefore, **two iterations back to the phase "Design & Development" within the DSR procedure** were performed. It turned out that, among other things, for the communication activities of greeting the conversation partner or ordering food or drinks, templates for writing a letter or instructions on how to perform these tasks have been harnessed. Instead, the prompt was expected to lead to the composition of a sentence to be uttered towards an interlocutor. In these cases, the extensive scope of the large language model employed by ChatGPT-4.0 seems leading to other relationships from the large language model being harnessed instead of strictly following this prompt. To mitigate this, it was implemented a semantical framing to limit the scope that is harnessed from the large language model. For each of the 61 communication activities, there was defined a specific prompt that is selected based on the expressions and single words that are commonly applied in the respective communication activity (e.g., the verb "order" within the selected expressions leads to a specific prompt for the respective communication activity and prepares ChatGPT-4.0 that the utterance to be generated is about ordering something). Since there may be communication activities that are not previously framed, it was added a default prompt to provide rudimentary utterance composition also for these cases. Proceeding with the *61 framed communication activities*, there were tested several variants of individual nuances and expressions of articulation in formal and informal conversational contexts. In doing so, it turned out that when individual nuances are semantically close to already selected vocabulary entries, this leads to duplicates within the generated utterances. Thus, it was implemented an automated mechanism that clears these duplicates. Employing these two iterations, the Generative AI-based approach could be improved, achieving accurate operation and properly meeting the posed DRs.

(d). In the following, this improved utterance composition approach was applied to the set of the *61 framed communication activities*. It was thereby found that in utterances for the formal context, the interlocutor is generally addressed by a formal term of address (cf. "Grüß Gott", "Ihnen", "Sie" in entries (II), (IV) and (VIII) in Table 3). These reflect more elaborated language and constructs like euphemisms, forms of courtesy, as well as a clear and precise terminology (cf. the German entries (VI) and (VIII) in Table 3). In contrast, the utterances for the informal context consistently address the interlocutor by an informal term (cf. "Hi", "Tschüss" in entries (I) and (IX) in Table 3), contain contractions (cf. "geht's", "mach's" in entries (I) and (IX) in Table 3), and in general a rather unsophisticated style of articulation (cf. "Ich hab' aktuell echt gut zu tun in der Arbeit", "Die Rechnung geht heut auf mich" in entries (III) and (VII) in Table 3). Regarding the formal context, three sentences contained characteristics of the informal context and were therefore incorrect. This results in a true positive rate of 58/61 and a false positive rate of 3/61. For the informal context, four generated sentences contained characteristics of the formal context, although these sentences were expected to be informal. This results in a true positive rate of 57/61 and a false positive rate of 4/61. Beyond that, it was assessed whether the proposed AI-based approach mixes both contexts. The Generative AI-based utterance composition was found to consistently adhere to the conversational style it was requested. Furthermore, all the provided individual nuances and expressions of articulation have been correctly considered within the generated sentences.

All in all, these results provide convincing evidence that the Generative AI-based approach can generate utterances for both conversational contexts, thereby confirming the general feasibility of the approach. Several utterances comprised a completed structure for the respective sentence types and were morphologically and grammatically correct. Only two minor grammatical errors could be observed as two verbs were not correctly conjugated. Adding a question mark to the speech composition line leads to generating questions instead of statement sentences (cf. entries (IV) and (VIII) within the "User Input" column of Table 3).

Except for three minor errors, the utterances consistently adhered to the themes of previous conversational turns. For example, when ordering food (cf. entry (VI) in Table 3), the meaning of the verb "take" is correctly considered given the previous utterance. The Generative AI-based approach automatically concludes that this means ordering fries as a side dish and not taking the sides in the sense of grabbing them.

Generally, the development of the artifact was based on the DRs, which are all met and technically realized *(see Figure 3)*. The proposed AI-based approach was verified with 61 formal and informal communication activities of visiting a restaurant with a companion, validating its functionality and practical applicability.

Discussion

Contributions to Theory and Practice

The implementation of the DRs and DPs has provided interesting results. ChatGPT-4.0 as a Generative AI technique turned out to be powerful to generate completed utterances from single words and expressions as speech-impaired provide them using their AAC systems. This not only leads to conversational contributions that are more quickly available within a conversation, but also relieves speech-impaired individuals of cognitive efforts when producing completed and grammatically correct sentences (i.a., inserting prepositions, conjugating verbs). As the demonstration and evaluation has shown, the generated conversational contributions consider the content and semantics of preceding turns within a conversation. Beyond that, the proposed AI-based approach can flexibly generate different sentence types (statement and question sentences). This empowers speech-impaired individuals to convey information when they are required and to ask questions when they need information within a conversation.

Even the fastest way of expressing is of little help if the utterances that an AAC system proposes are perceived as distracting by speech-impaired individuals and their interlocutors. The style of an utterance may not fit the expected way of articulation in a formal or an informal conversational context. As a contribution to practice, the proposed AI-based approach can suggest utterances for socialness-oriented (i.e., informal) as well as informativeness-oriented (i.e., formal) contexts (cf. Bedrosian et al. 2003; Hoag et al. 2004; Todman and Alm 1997). As Table 3 shows, the generated sentences clearly reflect wordings that match the common style of articulation of the German language (Budde 2012; Elspaß 2018). Moreover, the approach enables reflecting individual nuances one is keen on applying in formal and informal situations. This enables speech-impaired individuals to implement within utterances for formal situations the demanded level of contenance (i.a., non-familiar individuals or superiors may attach importance to maintaining distance because of being unfamiliar or due to their position). At the same time, they can demonstrate within their utterances for the informal contexts their closeness to familiar people like friends or family members.

As a further outcome of this DSR project, the investigation achieved theoretical contributions that go beyond the technical contribution (i.e., the artifact). The Generative AI-based approach and its capability of proposing completed utterances in a situation-specific articulation style empower speech-impaired individuals to comply with the Maxim of Modality. By complying with this Maxim, misunderstandings within a conversation can be reduced, leading to maintaining the flow of the conversation. Consequently, speech-impaired individuals will experience fewer breakdowns, making it easier for them to convey their thoughts, to be understood, and boosting their confidence in engaging with others using their AAC systems. Additionally, the compilation of the Generative AI-based utterance composition points out that there are deeply intertwined relationships of the Modality Maxim with the other Maxims of Grice (1975) in the AAC system context. The increased expressiveness provided by the proposed artifact could lead to speech-impaired individuals (deliberately) violating the Maxim of Relevancy or the Maxim of Quality with utterances that are not meaningful or not correct. Thus, investigating the opposing and amplifying effects of the Gricean Maxims in empirical settings could suggest refinements to context-specific communication styles, and insights into how speech-impaired individuals apply the Maxims (i.a., within negotiations with interlocutors).

Moreover, the research offers promising contributions to Cognitive Load Theory (cf. Sweller 1988). This theory addresses, among other things, the design of software system functionalities (cf. Hollender et al. 2010; Westphal et al. 2023). In the design and development of the proposed Generative AI-based approach, various AI and machine learning techniques (ChatGPT-4.0, automated speech recognition, natural language processing) were purposefully combined. In this way, speech-impaired individuals can be relieved in obtaining completed utterances that reflect situation- and user-specific styles of articulation. However, in light of the Cognitive Load Theory, it is not obvious in which circumstances speech-impaired individuals

will perceive this as a relief. The implemented supportive functionalities can take over several tasks typically performed during a conversation. But AAC systems that propose speech-impaired individuals context- and user-specific utterances exert for them also mental effort and cognitive load. Investigating the interaction effects of relief and cognitive load in empirical settings could suggest when the proposed AI-based approach should provide how much support. For example, there could be conversations where less support is deliberately preferred. Manually composing utterances can be harnessed as a means to demonstrate knowledge about certain topics to interlocutors (Valencia et al. 2023). Considering different situation-specific levels of support within the proposed AI-based approach could hence foster even more need-fitting communication.

The discussed results have shown the advantages of the proposed Generative AI-based approach for AAC systems compared to existing ones. Therefore, the DSR project is positioned in the group of improvements (new solutions for known problems) in terms of the DSR knowledge contribution framework of Gregor and Hevner (2013). DSR improvement projects contribute to both prescriptive theory, i.e., design theory (Gregor 2006), and descriptive theory, i.e., kernel theories such as the Gricean Maxims and Cognitive Load Theory as described above (cf. Gregor and Hevner 2013). By including ChatGPT-4.0 as a Generative AI technique within the utterance composition approach for AAC systems, this research also affects the quadrant of exaptations (known solutions extended to new problems) (Gregor and Hevner 2013). It is shown the proficiency of this technique for composing single expressions into completed utterances in different (formal and informal) context-specific styles. In this way, contributions to IS and DSR research are made alike. GPTs and Generative AI are proposed and proven as a suitable technique that enables situation- and user-specific styles of articulation for speech-impaired individuals.

Based on the DRs derived from literature and kernel theories, DPs were proposed. By applying them during the design and development of the artifact, followed by the demonstration and evaluation, an implicit empirical grounding of the DPs was achieved (Heinrich and Schwabe 2014). The DPs capture design-related knowledge and can thus support the development of further IS (design) theories and new artifacts. For example, by considering the principle of user-centeredness and user-related adaptability (DP2), the importance of including user-specific nuances into an AI-based utterance composition is highlighted. Since these expressions are directly reflected within the generated utterances, the alignment with user-centeredness leads to the communication being more personal and individualized for speech-impaired individuals. With the DPs, a first step is made towards contributing to design theory in terms of theory for design and action (Gregor 2006). It is complied with the conditions of March and Smith (1995) and Hevner et al. (2004), pointing out conditions in which a contribution to knowledge in design science has occurred: utility to a community of users, the novelty of the artifact and the persuasiveness of claims that it is effective.

Limitations and Opportunities for Future Research

There are also limitations to this research. Although this research encompasses a large set of investigations, researchers from other fields could probably identify even more DRs. However, the derived DRs are undoubtedly important for AI-based utterance composition within AAC systems. While this research proposes promising contributions to the Gricean Maxims and to Cognitive Load Theory, more in-depth investigations will suggest even more mature contributions.

To take the next steps toward a more mature design theory, the proposed AI-based approach will be evaluated in a formative environment (e.g., a laboratory experiment). Thereby, a homogenous sample of speech-impaired individuals will perform different formal and informal interactions with their companions and the staff of a restaurant. This allows for observing how speech-impaired individuals collaborate with the approach. Thereby, improvements can be made to it (whereby the DPs can be confirmed or adapted), before conducting a more elaborate evaluation in a natural setting (i.a., a field study in a restaurant) as a further part of the design cycle. There are several avenues for future research, for which the proposed AI-based approach constitutes the first essential step. Future studies can investigate the specific circumstances that lead to exceeding the cognitive load of speech-impaired individuals and avoiding context-specific articulation. Additionally, investigating in more detail the transferability of the Gricean Maxims to AAC-supported communication could shape daily life conversations for speech-impaired individuals even more inclusively.

Conclusion

To enable speech-impaired individuals to plenary participation within daily life conversations, they must be empowered to communicate in formal and informal conversational contexts in an appropriate, situation-

and user-specific style. However, prior literature and existing solutions for automated utterance composition do not sufficiently cover the informal and formal contexts. Beyond that, the specific requirements deemed crucial within extant AAC literature for AI-based utterance composition have not been comprehensively met. Hence, this research has derived several DRs and DPs. To technically realize the DPs, several AI and machine learning techniques (ChatGPT-4.0, automated speech recognition, natural language processing) were purposefully combined into an automated utterance composition approach to close the revealed gaps *(see RQ1 and RQ2)*. Demonstrating and evaluating a developed AAC system prototype employing this approach to 61 formal and informal communication activities of visiting a restaurant has indicated that covering all the derived DRs is essential. This coverage is crucial for ensuring that the proposed Generative AI-based approach is comprehensive and purposeful. To enable speech-impaired individuals to use context- and user-specific styles at the lowest possible level of effort, the approach integrates further supportive functionalities like an automated identification of the current conversational context. In this way, it fosters the immediate application of the utterances that match the current context of a conversation.

The investigation contributes to practice and research alike. The generated artifact empowers speech-impaired individuals to increase their expressiveness using formal and informal conversational contributions from their AAC systems and thus dismantling communication barriers that speech-impaired individuals suffer from. The research also highlights first contributions to design theory and kernel theory. Regarding the Gricean Maxims of Communication (Grice 1975), the compilation of the Generative AI-based utterance composition could reveal for AAC-supported communication intertwined relationships of the Modality Maxim with the other Maxims. Moreover, the research provides the basis to enrich Cognitive Load Theory based on the design and the concrete instantiation of the Generative AI-based approach. In this way, the proposed artifact is expected to relieve speech-impaired individuals from the cognitive load by obtaining coherent and grammatically correct utterances in context- and user-specific wordings.

References

Adhikary, S., Berger, J., and Vertanen, K. 2021. "Accelerating text communication via abbreviated sentence input," *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*.

Allen, J. 2005. "Designing desirability in an augmentative and alternative communication device," *Universal Access in the Information Society* (4:2), pp. 135–145.

Alm, N., and Higginbotham, J. 2008. "Whole utterance approaches in AAC," *Augmentative and Alternative Communication* (24:3), pp. 235–254.

Arnott, J., and Alm, N. 2013. "Towards the improvement of augmentative and alternative communication through the modelling of conversation," *Computer Speech & Language* (27:6), pp. 1194–1211.

Baldassarri, S., Marco, J., Cerezo, E., and Moreno, L. 2014. "Accessibility evaluation of an alternative and augmentative communication (AAC) tool," *Proceedings of the 2014 International Conference on Universal Access in Human-Computer Interaction*, pp. 529–540.

Bedrosian, J., Hoag, L., and McCoy, K. 2003. "Relevance and speed of message delivery tradeoffs in augmentative and alternative communication," *Journal of Speech Language and Hearing Research* (46).

Black, R., Waller, A., Turner, R., and Reiter, E. 2012. "Supporting personal narrative for children with complex communication needs," *ACM Transactions on Computer-Human Interaction* (19:2), pp. 1–35.

Böckle, M., Bick, M., and Novak, J. 2021. "Toward a Design Theory of User-Centered Score Mechanics for Gamified Competency Development," *Information Systems Management* (40:1), pp. 1–27.

Bradshaw, J. 2013. "The use of augmentative and alternative communication apps for the iPad, iPod and iPhone: an overview of recent developments," *Tizard Learning Disability Review* (18:1), pp. 31–37.

Budde, M. 2012. "Über Sprache reflektieren," *Kassel University Press*.

Cai, S., Venugopalan, S., Tomanek, K., Narayanan, A., Morris, M., and Brenner, M. 2022. "Context-Aware Abbreviation Expansion Using Large Language Models," *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics*.

Chandra, L., Seidel, S., and Gregor, S. 2015. "Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions," *Proceedings of the 48th Hawaii International Conference on System Sciences*.

Dempster, M., Alm, N., and Reiter, E. 2010. "Automatic generation of conversational utterances and narrative for Augmentative and Alternative Communication: a prototype system," *Proceedings of the 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pp. 10–18.

Desai, T., Chow, K., Mumford, L., Hotze, F., and Chau, T. 2014. "Implementing an iPad-based alternative communication device for a student with cerebral palsy and autism in the classroom via an access technology delivery protocol," *Computers & Education* (79), pp. 148–158.
Ehlich, K., and Rehbein, J. 1972. "Zur Konstitution pragmatischer Einheiten in einer Institution: Das Speiserestaurant," *Linguistische Pragmatik*, pp. 209–254.
Elspaß, S. 2018. "Sprachvariation und Sprachwandel," *Sprache in sozialen Gruppen*, pp. 87–107.
Fang, Y., Xu, Y., Teng, Z., Jiang, Z., and Xiang, W. 2023. "SocializeChat: A GPT-Based AAC Tool for Social Communication Through Eye Gazing," *Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing,* pp. 128–132.
Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. 2023. "Generative AI," *Business & Information Systems Engineering* (66), pp. 111–126.
Filippidou, F., and Moussiades, L. 2020. "A benchmarking of IBM, Google and Wit automatic speech recognition systems," *Proceedings of the 16th IFIP WG 12.5 International Conference*.
Fleiss, J., Levin, B., and Paik, M. 2013. "Statistical methods for rates and proportions," *John Wiley & Sons*.
Fritzell, B. 1996. "Voice disorders and occupations," *Logopedics Phoniatrics Vocology* (21:1), pp. 7–12.
Garcia, L., de Oliveira, L., and de Matos, D. 2016. "Evaluating pictogram prediction in a location-aware augmentative and alternative communication system," *Assistive Technology* (28:2), pp. 83–92.
Giles, H., Coupland, N., and Coupland, J. 1991. "Accommodation theory: Communication, context, and consequence," *Cambridge University Press*.
Gregor, S. 2006. "The nature of theory in information systems," *MIS Quarterly* (30:3), pp. 611–642.
Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337–355.
Gregor, S., and Jones, D. 2007. "The Anatomy of a Design Theory," *Journal of the AIS* (8:5), pp. 312–335.
Gregor, S., Chandra Kruse, L., and Seidel, S. 2020. "Research Perspectives: The Anatomy of a Design Principle," *Journal of the AIS* (21:6), pp. 1622–1652.
Grice, H. 1975. "Logic and conversation," *Speech Acts, New York: Academic Press*, pp. 41–58.
Heinrich, P., and Schwabe, G. 2014. "Communicating Nascent Design Theories on Innovative Information Systems through Multi-Grounded Design Principles," *Proceedings of the 9th International Conference on Design Science Research in Information Systems and Technology (DESRIST)*.
Henne, H., and Rehbock, H. 2012. "Einführung in die Gesprächsanalyse," *de Gruyter*.
Heo, Y., Kang, S. 2019. "Augmentative and Alternative Communication System Using Information Priority and Retrieval," *Proceedings of the 2019 International Conference on Big Data and Smart Computing*.
Hevner, A., March, S., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75–105.
Higginbotham, D., Lesher, G., Moulton, B., and Roark, B. 2012. "The application of natural language processing to augmentative and alternative communication," *Assistive Technology* (24:1), pp. 14–24.
Hoag, L., Bedrosian, J., McCoy, K., and Johnson, D. 2004. "Trade-offs between informativeness and speed of message delivery in augmentative and alternative communication," *Journal of Speech, Language, and Hearing Research* (47), pp. 1270–1285.
Hollender, N., Hofmann, C., Deneke, M., and Schmitz, B. 2010. "Integrating cognitive load theory and concepts of human–computer interaction," *Computers in Human Behavior* (26:6), pp. 1278–1288.
Hourcade, J., Everhart Pilotte, T., West, E., and Parette, P. 2004. "A history of augmentative and alternative communication for individuals with severe and profound disabilities," *Focus on Autism and other Developmental Disabilities* (19:4), pp. 235–244.
Kane, S., Morris, M., Paradiso, A., and Campbell, J. 2017. "At times avuncular and cantankerous, with the reflexes of a mongoose - Understanding Self-Expression through AAC Devices," *Proceedings of the 2017 Conference on Computer Supported Cooperative Work and Social Computing*.
Klauer, A., Johannsen, F., and Leist, S. 2021. "BeMyVoice - An iPad-Based Helper for Speech Impaired Persons," *Proceedings of the 16th International Conference on DESRIST*.
Konadl, D., Wörner, J., Luttner, L., and Leist, S. 2023. "Artificial Intelligence in Augmentative and Alternative Communication Systems - A Literature-Based Assessment and Implications of Different Conversation Phases and Contexts," *Proceedings of the 31st European Conference on Information Systems*.
Kristensson, P., Lilley, J., Black, R., and Waller, A. 2020. "A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities," *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

Laxmidas, K., Avra, C., Wilcoxen, C., Wallace, M., Spivey, R., Ray, W., Polsley, S., Kohli, P., Thompson, J., and Hammond, T. 2021. "CommBo: Modernizing Augmentative and Alternative Communication," *International Journal of Human-Computer Studies* (145).

March, S., and Smith, G. 1995. "Design and natural science research on information technology," *Decision Support Systems* (15:4), pp. 251–266.

Mayring, P. 2000. "Qualitative Content Analysis," *Forum Qualitative Social Research* (1:2).

Mayring, P. 2002. "Introduction to qualitative social research," *Weinheim: Beltz*.

Mitchell, M., and Sproat, R. 2012. "Discourse-Based Modeling for AAC," *Proceedings of the 3rd Workshop on Speech and Language Processing for Assistive Technologies*.

Möller, F., Guggenberger, T., and Otto, B. 2020. "Towards a method for design principle development in information systems," *Proceedings of the 15th International Conference on DESRIST*.

Neamtu, R., Camara, A., Pereira, C., Ferreira, R. 2019. "Using artificial intelligence for augmentative alternative communication for children with disabilities," *Proceedings of the 2019 IFIP Conference on HCI*.

Obiorah, M., Piper, A., and Horn, M. 2021. "Designing AACs for people with aphasia dining in restaurants," *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

OpenAI. 2024. "ChatGPT-4.0," Accessed on 11.02.2024 *via this Link*.

Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S. 2007. "A design science research methodology for information systems research," *Journal of Management Information Systems* (24:3), pp. 45–77.

Peffers, K., Rothenberger, M., Tuunanen, T., and Vaezi, R. 2012. "Design science research evaluation," *Proceedings of the 7th International Conference on DESRIST*.

Purao, S., Kruse, L., and Maedche, A. 2020. "The Origins of Design Principles: Where do… they all come from?," *Proceedings of the 15th International Conference on DESRIST*.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. 2019. "Language models are unsupervised multitask learners," *OpenAI Blog* (1:8).

Reddington, J., and Tintarev, N. 2011. "Automatically generating stories from sensor data," *Proceedings of the 16th International Conference on Intelligent User Interfaces*.

Schindler, E. 2013. "Kommunikative Ritualität im Gastronomiebereich," *Freiburger Arbeitspapiere zur germanistischen Linguistik, 11. Freiburg: Germanistische Linguistik, Universität Freiburg*.

Shen, J., Yang, B., Dudley, J., and Kristensson, P. 2022. "Kwickchat: A multi-turn dialogue system for AAC using Context-Aware Sentence Generation by Bag-of-Keywords," *Proceedings of the 27th International Conference on Intelligent User Interfaces*.

Snyder, H. 2019. "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research* (104), pp. 333–339.

Sweller, J. 1988. "Cognitive load during problem solving: Effects on learning," *Cognitive Science* (12:2).

Teubner, T., Flath, C., Weinhardt, C., van der Aalst, W., and Hinz, O. 2023. "Welcome to the era of chatgpt et al.," *Business & Information Systems Engineering* (65:2), pp. 95–101.

Tintarev, N., Reiter, E., Black, R., Waller, A., and Reddington, J. 2016. "Personal storytelling: Using Natural Language Generation for children with complex communication needs, in the wild …," *International Journal of Human-Computer Studies* (92), pp. 1–16.

Todman, J., and Alm, N. 1997. "Pragmatics and AAC approaches to conversational goals in natural language processing for communication aids," *Proceedings of the 35th Annual Meeting of the ACL*.

Valencia, S., Cave, R., Kallarackal, K., Seaver, K., Terry, M., and Kane, S. 2023. "The less I type, the better: How AI Language Models can Enhance or Impede Communication for AAC Users," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.

Vanderheiden, G. 1983. "Non-Conversational Communication Technology Needs of Individuals with Handicaps," *Rehabilitation World* (7:2), pp. 8–12.

Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: a framework for evaluation in design science research," *European Journal of Information Systems* (25:1), pp. 77–89.

Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. "Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research," *Communications of the AIS* (37:1), pp. 205–224.

Walls, J., Widmeyer, G., and El Sawy, O. 1992. "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research* (3:1), pp. 36–59.

Walsh, T. 2010. "Utterance-based systems: organization and design of AAC interfaces," *Proceedings of the 12th International Conference on Computers and Accessibility*.

Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. 2023. "Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance," *Computers in Human Behavior* (144).

2.7. **Beitrag 7:** Comparison of Commercial Chatbot Solutions for Supporting Customer Interaction

Adressierte Forschungsfragen	**(1)** Welche Anforderungen sollten Chatbot-Lösungen für den Einsatz im B2C-Kontext zur Unterstützung der Kundenkommunikation erfüllen? **(2)** Wie umfassend werden diese Anforderungen und Funktionalitäten von den Anbietern von Chatbot-Lösungen abgedeckt?
Zielstellungen	• Identifikation von Anforderungen und Funktionalitäten aus der wissenschaftlichen Literatur für den Einsatz eines Chatbots zur Unterstützung der Kundenkommunikation im B2C-Kontext. • Ermittlung des Abdeckungsgrades der identifizierten Anforderungen durch eine international angelegte Marktstudie unter Anbietern von Chatbot-Lösungen.
Forschungsmethode	Marktstudie mithilfe einer Online-Umfrage
Kernergebnisse (Überblick)	• Konsolidierung von Anforderungen und Funktionalitäten für den Einsatz von Chatbots in der Kundenkommunikation im B2C-Kontext. • Übersicht über den Abdeckungsgrad dieser Anforderungen und Funktionalitäten durch am Markt verfügbare Chatbot-Lösungen. Infolgedessen ergibt sich eine Entscheidungsgrundlage für Unternehmen, welche Chatbot-Lösung vor dem Hintergrund der eigenen Anforderungen ausgewählt werden sollte. • Übersicht über Bereiche, in denen die Anbieter bestimmte Anforderungen und Funktionalitäten noch besser abdecken könnten.
Publikationsort	**Veröffentlicht** in den Proceedings der 26. European Conference on Information Systems, Portsmouth, Großbritannien.
Ranking VHB JQ 4	**A**
Autoren und Anteile	Florian Johannsen 40 % Susanne Leist 20 % **Daniel Konadl** **20 %** Michael Basche 10 % Bastian de Hesselle 10 %

Tabelle 9: Fact Sheet zu Beitrag 7

COMPARISON OF COMMERCIAL CHATBOT SOLUTIONS FOR SUPPORTING CUSTOMER INTERACTION

Research paper

Johannsen, Florian, University of Bremen, Bremen, Germany, fjohanns@uni-bremen.de

Leist, Susanne, University of Regensburg, Regensburg, Germany, Susanne.Leist@ur.de

Konadl, Daniel, University of Regensburg, Daniel.Konadl@stud.uni-regensburg.de

Basche, Michael, University of Regensburg, Michael.Basche@stud.uni-regensburg.de

de Hesselle, Bastian, PricewaterhouseCoopers, Germany, bastian.de.hesselle@pwc.com

Abstract

Considering the advancing digitalization and the rise of new technologies, the way people interact with one another but also approach companies has fundamentally changed. In this respect, chatbots have experienced a "comeback" recently and many large companies but also small and medium-sized enterprises (SMEs) in the B2C sector search for ways to support the conversation with customers via chatbots. Nevertheless, chatbots must fulfil particular functionalities (e.g., in terms of language processing or quality of communication) to be purposefully applied in an entrepreneurial setting, and particularly, at SMEs. The investigation at hand analyses the solutions of 14 technology providers for realizing a chatbot in terms of (1) quality of the communication, (2) reaction to inappropriate input, (3) personalization, (4) language processing, (5) security, (6) interfaces & mobile device support, (7) training, (8) implementation and (9) costs. Thereby, the peculiarities of human-machine communications and requirements of firms are taken into account. The results were derived on base of an online survey among the solution providers. It turned out that a high percentage of the functionalities, which are decisive for a valuable entrepreneurial chatbot application, are covered by the providers, however, with nuanced differences being observed regarding particular features.

Keywords: chatbot, human-machine interaction, small and medium-sized enterprise

1 Motivation

Digitalization is one of the major topics CEOs of modern enterprises currently have to deal with (CapGemini, 2016; Matt et al., 2015). The term "digitalization" addresses the spread of digital technologies within society and the changing communication behaviour of individuals and the way people connect with one another (Gimpel and Röglinger, 2015). In this context, the rise of new digital technologies (e.g., social networks, cloud computing, etc.) (Patil et al., 2017) has led to an increase of market transparency while the information asymmetry between sellers and buyers is constantly getting smaller (Laudon and Laudon, 2014). Consequently, the establishment of long-term customer loyalty is of high importance for companies (Lovelock et al., 2015; Mukerjee, 2013) and customer relationship management (CRM) activities have experienced a "second spring" within many firms recently (Orenga-Roglá and Chalmeta, 2016; Faed et al., 2014; Greenberg, 2010). In this regard, the rise of new technologies also affected the way private persons communicate with companies and more and more consumers preferably approach firms via digital channels (e.g., online forms, social media) to utter service requests or to settle transactions amongst others (Berthon et al., 2012; Hanna et al., 2011; Dale, 2016).

Considering this, the use of chatbots to support the external communication with customers (e.g., Gallaugher and Ransbotham, 2010) in business-to-consumer (B2C) settings is ever more paid attention to by companies (Kuligowska, 2015; Xu et al., 2017). Thereby, a chatbot is a *"computer program"* that has *"the ability to hold a conversation"* with human beings *"using Natural Language Speech"* (Abdul-

Kader and Woods, 2015, p. 72). Building on chatbots to interact with consumers brings about several value-propositions (cf. Lanning and Phillips, 1991) for companies. Toshiba for instance could reduce the number of support calls by 30-50% decreasing service employees' workload (cf. Living Actor, 2017). At the same time, customer service via the chatbot is available 24hrs a day and, hence, customers can post their service request independent from usual business hours, which increases consumer satisfaction (cf. Gorelov, 2016). More, chatbots may also contribute to users' "customer experience" during online shopping (Horzyk et al., 2009). The Bank of Amerika for instance uses a chatbot that analyses customers' behaviour to provide individual and tailored product offerings (Kusber, 2017). Generally, chatbots are applied in various branches (e.g., financial services, production industries) (e.g., Deshpande et al., 2015) and the technology has been constantly further developed in recent years, e.g., by an integration with messengers of social media platforms or speech recognition (Dale, 2016; Yan et al., 2016).

Against this background, the use of chatbots to support CRM activities in B2C markets is not only a topic for large firms but is increasingly becoming interesting for small and medium-sized enterprises (SMEs) as well. Generally, SMEs play a decisive role considering the economy of most countries (Newby et al., 2014). However, SMEs have limited budgets (Javaid et al., 2017), buy and sell in small quantities and multiple tasks are done by the same departments or employees, respectively, due to limited personal resources (Bharati and Chaudhury, 2009). At the same time, fostering close relationships with the customer base is decisive for SMEs to maintain a competitive edge (Newby et al., 2014). Considering this, SMEs have recognized the value of IT to effectively and efficiently strengthen the customer relationships (Newby et al., 2014; Bull, 2003). Nevertheless, SMEs still lack behind in terms of digitalization (McKinsey, 2014; Javaid et al., 2017). This is because decision makers are unsure whether potential benefits of IT usage will outweigh the necessary IT investments (Barba-Sánchez et al., 2007).

Incidentally, chatbots are a promising technology to automatize the customer interaction to a large degree, strengthen customer relationships by offering a 24hrs customer service and, consequently, to set free personal resources for the core business activities. However, firms will only invest into the introduction of chatbots in case the aspired value-propositions prevail the costs of investment (e.g., Barba-Sánchez et al., 2007). Against this background, the following question arises: *What are the functionalities current chatbot solutions should meet to be purposefully applied in a B2C setting and do current solution providers on the market cover these appropriately?*

To answer this question an online survey among commercial providers of chatbot solutions was performed. Whereas qualitative evaluations of concrete chatbot implementations (e.g., the chatbot "Ania" realized via the "Artifical Solution" chatbot solution) exist in literature (cf. Kuligowska, 2015), the study at hand abstracts from concrete realizations and analyses the general functionalities of chatbot solutions corresponding technology providers offer for establishing a chatbot. The paper unfolds as follows: in the subsequent section, a short introduction to chatbots is given and requirements on software solutions to realize chatbots are shown. In section 3, the procedure of the investigation is explicated and the functionalities of chatbot solutions – mandatory for their entrepreneurial use at SMEs – are derived. The results of the survey are described and discussed in section 4. The paper is rounded off with benefits of the study as well as limitations and an outlook (section 5).

2 Foundations

2.1 Background on chatbots

The initial purpose of chatbots was to mimic human conversations for user amusement (Shawar and Atwell, 2007). In this context, ELIZA, which was created by Joseph Weizenbaum in the 60's (cf. Weizenbaum, 1966), was the first attempt of building a chatbot and – although its language processing abilities were rather limited – inspired the development of many modern chatbots (Shawar and Atwell, 2007; Dale, 2016). Recently, a new hype arose regarding the chatbot technology triggered by the spread of mobile applications and social media, which fundamentally changed the way people connect and communicate with each other (Dale, 2016; Kuligowska, 2015). Thereby, a chatbot comprises the components "user interface", "responder", "classifier" and "graphmaster" according to Stoner et al. (2003).

The "user interface" presents the interface between the human and the computer and the "responder" controls the input provided by the human and processes it to the "classifier" (Abdul-Kader and Woods, 2015; Stoner et al., 2003). The input is then split into singular units by the "classifier" (e.g., extraction of keywords by eliminating unwanted words) and the syntax of the markup language used (usually the Artificial Markup Language) is checked (Abdul-Kader and Woods, 2015). The "graphmaster" then performs pattern matching, i.e., data is identified on base of the underlying structure (Stoner et al., 2003; Abdul-Kader and Woods, 2015). Then, an answer to be handed back is created on base of the content stored in the "graphmaster", which is then transferred to the user via the "classifier" over the "responder" and the corresponding "user interface" (Stoner et al., 2003). Today's application fields for chatbots comprise "e-government" (e.g., Sandoval-Almazán and GutiérrezAlonso, 2009), "education" (e.g., Kerry et al., 2009; Goda et al., 2014), "healthcare" (e.g., Comendador et al., 2015), "e-commerce" (e.g., Horzyk et al., 2009) or "law & finance" (e.g., Dole et al., 2015; Gorelov, 2016; Gibbs, 2016) amongst others. In "e-government" for instance, Sandoval-Almazán and GutiérrezAlonso (2009) describe a chatbot, which was run by the Public Information Access Unit (IAU) of the Guanajuato State in Mexico to facilitate the interaction with users in terms of administration-related issues. Regarding "education", Goda et al. (2014) analyse to what extent chatbots can be purposefully used for teaching foreign languages. "Pharmabot" is an example for a chatbot in the healthcare industries and serves as a "virtual pharmacist" to consult patients during the shopping of medicaments (Comendador et al., 2015). Bickmore et al. (2010) developed a computer-animated conversational agent for hospital patients providing them with discharge information. More, an automated health counsellor agent for promoting physical activity and fruit/vegetable consumption was created, which was proven to be effective for changing peoples' health behaviour (Bickmore et al., 2013). Dole et al. (2015) give an example for a chatbot in the banking sector with the purpose to answer customer questions about the bank's offerings. In the study of Kuligowska (2015), six Polish-speaking commercial chatbots, applied in B2C commerce (e.g., Ania – "IKEA", Zosia – "Villa Tan Tadeusz", Karen – "WSHiFM", etc.), are evaluated in terms of their visual look, presentation of knowledge as well as conversation abilities amongst others.

2.2 Requirements for chatbot usage

To be purposefully used for entrepreneurial purposes, chatbots have to meet specific requirements. On the one hand, these are general requirements that can be defined considering the specifics of the communication between humans and chatbots. On the other hand, certain requirements arise regarding the chatbot use at SMEs as set forth in the following.

General requirements emerge by the peculiarities considering the communication between humans with chatbots as they were identified by Hill et al. (2015) for instance, but also by the implications for human-computer interaction (HCI) that are caused by the increasing dissemination of natural language interfaces (cf. Følstad and Brandtzæg, 2017). In enterprise settings, it needs to be considered that chatbots play an increasing role in the knowledge creation process of individuals as stipulated by the theory of social constructionism (cf. Andrews, 2012; Schwandt, 2000) because they gradually replace humans in service encounters and become partners for interacting with customers, e.g., in the context of handling service requests. Thereby, the conversation with chatbots should not only help to solve a customer problem but a chatbot's language may also carry emotions and feelings (cf. Andrews, 2012; Burr, 2006). In this respect, messages sent to chatbots are rather short (text economy) which leads to the fact that people tend to send more than twice as many messages to chatbots in comparison to a conversation with other humans (Hill et al., 2015). In this respect, people often feel they have to alter their communication to match the one of chatbots similar to people changing their language when talking to children (Hill et al., 2015). Whereas this circumstance may be acceptable when communicating with chatbots for amusement purposes, a professional business appearance requires enterprises to explain facts (e.g., contractual terms, service responses) using consumers' language. Accordingly, a chatbot must be adjustable to the language of a specific target-audience and ensure a high quality of the communication in addition ***(requirement 1 – Rq 1)***. That way, it can be assured that language and conversation become the objects of design (Følstad and Brandtzæg, 2017) during chatbot implementation.

Furthermore, the communication with chatbots is imprinted by a huge share of profanity but also swear, negative or sexual words respectively (Hill et al., 2015). Accordingly, a chatbot must allow to realize chatbots that are able to react to such inappropriate input in a quick-witted and polite way ***(Rq 2)***. In case chatbots have a human-like appearance it turned out that people do not necessarily feel more uncomfortable when interacting with a chatbot instead of a human even if they are aware that the interaction is performed with a machine (Hill et al., 2015). This circumstance is also in accordance with media equation (cf. Reeves and Nass, 1996) stating that people tend to treat technology and instances of media as people (Heller et al., 2005). Hence, means to strengthen the perception on the users' side to actually communicate with a human need to be offered by a chatbot ***(Rq 3)***. This can be done by the help of individually configurable avatars for instance (cf. Braun, 2003; Berry et al., 2004; Gennermann and Hack, 2011) or by the ability to conduct "small talk".

Further general requirements on chatbot solutions can be derived by considering the specifics of text messages on the internet. Thus, abovementioned text economy (cf. Hill et al., 2015) increases when interacting with machines, i.e., whitespaces are spared and numbers with similar pronunciation are used instead of words (Laboreiro et al., 2010; Petz et al., 2013). Generally, language on the internet is imprinted by the use of internet slang (e.g., the expression "HB2U"), multiple languages within one text message, or spelling errors (e.g., "Hellllllooooo") (Laboreiro et al., 2010; Petz et al., 2013; Schwaiger et al., 2016) that need to be taken into account considering the communication with chatbots. More, as people use a highly emotional language when interacting with chatbots (cf. Hill et al., 2015) capitalization is often used to express emotions or to express irony or sarcasm (Petz et al., 2013). Consequently, means to deal with slang expressions ***(Rq 4)***, multiple languages ***(Rq 5)***, spelling errors and case sensitivity ***(Rq 6)*** as well as irony or sarcasm ***(Rq 7)*** need to be implemented by a chatbot. More, recent trends such as the use of chatbots in branches processing sensible data (e.g., banking) (cf. Kusber, 2017) as well as the rise of mobile technologies and social networks (cf. Statista, 2017) pose further requirements on the entrepreneurial chatbot use. Thus, the handling of sensitive data requires means to enable encoded data transmission and storage ***(Rq 8)***. Furthermore, as social media are increasingly used to get in contact with firms (cf. Hanna et al., 2011), which also holds true for SMEs in special (cf. Schwaiger et al., 2017), the integration of chatbots with new communication channels (e.g., social network messengers) needs to be considered to purposefully handle consumer requests uttered via these (cf. Dale, 2016; Chaykowski, 2016) ***(Rq 9)***. Additionally, an integration with other IT-systems (e.g. to trigger workflows) as well as the ability to be run on mobile devices is required these days (cf. Dale, 2016) ***(Rq10)***.

Besides, specific peculiarities must be considered for an entrepreneurial use of chatbots at SMEs. Hence, SMEs usually have a limited regional presence (Durkin et al., 2013), their customers come from few different communities in the same region (Bharati and Chaudhury, 2009) and, thus, conversations may refer to specific products or local events hosted by the corresponding firms accordingly (Schwaiger et al., 2016). Addressing this peculiarity falls under abovementioned ability to adjust the chatbot to the language of a target-audience (see Rq1). However, due to firms' interaction within a limited geographical area (Bharati and Chaudhury, 2009) the language of customers of SMEs is imprinted by regional dialect and colloquial language (Schwaiger et al., 2016). Therefore, chatbots must be able to recognize these language styles and still reply appropriately ***(Rq11)***. More, SMEs often have limited IT skills (cf. Bull, 2003; Newby et al., 2014; Bharati and Chaudhury, 2009). Accordingly, the training of the chatbot must be possible without technical skills being required ***(Rq12)***. In this respect, the chatbot should ideally learn from interactions with different users, which presents a major topic for HCI regarding natural language interfaces (cf. Følstad and Brandtzæg, 2017). The initial installation of a chatbot by a firm should not expect sophisticated technical skills either ***(Rq13)***. More, due to limited buying powers of SMEs (Bharati and Chaudhury, 2009) and their restraint of investing into IT (cf. Newby et al., 2014), affordable prizing models are required when it comes to chatbot introduction ***(Rq14)***.

3 Survey on current providers of chatbot solutions

The procedure of our research was as follows: in a first step, providers of chatbot solutions were identified and asked for their willingness to participate in this study. Then, functionalities of chatbot solutions

to enable their purposeful application in B2C settings considering abovementioned requirements were derived from literature. Based on that, a questionnaire was designed to assess as to which degree current solutions cover these functionalities. The data was collected via a corresponding online survey.

3.1 Selection of chatbot solution providers for the study

To find participants for the study, commercial providers of chatbot solutions applicable for B2C settings were searched for in a first step. For that purpose the "Google" search engine, the website "chatbots.org", studies about the evaluation of commercial chatbots in literature (cf. Kuligowska, 2015) as well as the expertise of two practitioners working at an internationally operating consultancy firm and dealing with entrepreneurial chatbot application amongst others, were drawn upon. Particular attention was paid to the applicability of the solutions in B2C settings but also to the fact that the companies are still active in further distributing and developing their solutions. In total, 48 technology providers for commercial chatbots were identified that way and asked for their willingness to take part in our investigation. The providers mainly came from North America, South America and Europe but also from Australia, Africa and Asia. Finally, 14 commercial providers of chatbot solutions decided to participate in the study seeing the investigation as promising for gaining a detailed market overview.

Among these participants were "Creative Virtual", "Do You Dream Up", "Synthetix", "Artificial Solutions" and "Virtual Spirits" for instance. However, the majority of participants wanted to stay anonymous in this report due to reasons of confidentiality. An online survey was seen as appropriate for collecting the data since providers' knowledge is decisive to identify the functionalities that are actually offered by the corresponding solutions. In contrast, users of these solutions realizing a chatbot might not be aware of all functionalities that are provided. Their focus will most likely be on those features they particularly require for implementing a chatbot meeting their demands. More, testing chatbots from a customer perspective may give an incomplete picture of their functionalities as well since the capabilities of the underlying solution may not have been fully utilized for implementing the chatbot.

3.2 Functionalities investigated

To effectively support the communication with customers in B2C settings, diverse functionalities of chatbot solutions are decisive. In the following, we present the functionalities of chatbot solutions that were investigated in this study more closely. To support the general validity of our investigation the functionalities were derived from literature (e.g., Klopfenstein et al., 2017; Gennermann and Hack, 2011; Braun, 2003; Kuligowska, 2015) and assigned to corresponding requirements as presented *(section 2.2)*. We clustered the functionalities according to the superordinate categories (1) quality of communication, (2) reaction to inappropriate input (3) personalization, (4) language processing, (5) security, (6) interfaces & mobile device support, (7) training, (8) implementation and (9) costs.

The **categories 1** and **2** focus on the conversation skills, the adaption to the language of a target-audience as well as the reaction to inappropriate user input such as profanity, swear, negative or sexual words (cf. Hill et al., 2015). They thus consider functionalities required for fulfilling the first and second requirement of section 2.2 *(Rq1 and Rq2)*. At the same time, the consideration of regional dialect, slang and colloquial language is a matter concerning the "quality of communication" and, hence, *Rq 4* and *Rq 11* are covered by **category 1** as well (see Table 1). The **categories 3** and **4** deal with functionalities concerned with the personalization of a chatbot and its language processing abilities. Accordingly, the requirements *Rq3, Rq5, Rq6* and *Rq7* are addressed. Security-related functionalities are subject of **category 5** covering *Rq 8* consequently. Interfaces and mobile device support (*Rq9* and *Rq10*) are considered by **category 6**. The SME-specific requirements dealing with training *(Rq12)*, implementation efforts *(Rq13)* and cost-related issues *(Rq14)* are focused by the **categories 7** to **9** (see Table 1). Table 1 shows the aforementioned categories, the functionalities these comprise as well as their assignment to the requirements. Details on each functionality are outlined in the following.

No.	Category	Detailing of functionalities		Rq (section 2.2)		
1	*Quality of the communication*	1.1	The option to adjust the chatbot to the language of a specific target-audience.	**Rq1**		
		1.2	The functionality to individually configure the response time of the chatbot from the user input to the response given by the chatbot.			
		1.3	The provision of a memory function to refer to previous statements of a conversation during interaction.			
		1.4	The ability to insert internet-links to an answer provided.			
		1.5	The functionality to obtain customer feedback considering the advisory skills of the chatbot.			
		1.6	The option of an automated handover to a human agent.			
		1.7	The provision of a "history" function (conversation logs).			
		1.8	The option to identify synonyms within a conversation.			
		1.9	The provision of means to ensure freedom of barriers.			
		1.10	The functionality to provide voice input and output.			
		1.11	The consideration of language packages supporting lexical variation, slang or dialect.	**Rq 1**	**Rq 4**	**Rq 11**
2	*Reaction to inappropriate input*	2.1	The functionality to react to inappropriate user input (e.g., insults).	**Rq2**		
		2.2	The functionality to protect the chatbot from manipulating user input.			
		2.3	The functionality to react to unspecific user input.			
3	*Personalization*	3.1	The option to generate an avatar.	**Rq3**		
		3.2	The capability to individualize the avatar.			
		3.3	The functionality to perform "small talk".			
		3.4	The functionality to support diversified conversations.			
4	*Language processing*	4.1	The support of different languages.	**Rq5**		
		4.2	The offering of a spell-checking functionality.	**Rq6**		
		4.3	The functionality of considering case sensitivity.			
		4.4	The functionality to detect irony and sarcasm.	**Rq7**		
5	*Security*	5.1	The functionality of encoded data storage.	**Rq8**		
		5.2	The functionality of encoded data transmission.			
		5.3	The functionality of authentication towards the chatbot.			
		5.4	The option to implement a role model for administration purposes of the chatbot.			
		5.5	The functionality to provide a data backup.			
		5.6	The support of PCI standards.			
6	*Interfaces & mobile device support*	6.1	The functionality of integrating the chatbot with a website.	**Rq9**		
		6.2	The functionality of integrating the chatbot with messengers.			
		6.3	The provision of interfaces to other IT-systems.	**Rq10**		
		6.4	The functionality to be run on mobile devices.			
7	*Training*	7.1	The use of self-learning algorithms.	**Rq12**		
		7.2	The provision of a graphical user interface (GUI) for a manual training of the chatbot.			
		7.3	The functionality to train the chatbot without technical skills.			
8	*Implementation*	8.1	The provision of initial and prepared language packages to enable a direct implementation.	**Rq13**		
		8.2	The option to initially host the chatbot without requiring any programming skills.			
		8.3	The availability of supportive (discussion) forums for the software product.			
9	*Costs*	9.1	The availability of attractive license/billing models.	**Rq14**		
		9.2	The option to run the chatbot in form of a "software as a service" (SaaS).			
Key: Rq1 – Rq10: general requirements; Rq11 – Rq14: SME-specific requirements						

Table 1. Functionalities investigated

(1) Quality of communication: As mentioned in section 2.2, the success of a chatbot largely depends on whether its language can be adapted to the one of the target-audience or not (Gennermann and Hack, 2011) *(functionality 1.1; see Table 1)*. Hence, in a business setting, customers need to have the feeling of meeting the conversation partner on equal terms, which requires the chatbot to attune to consumers' language. Hence, this represents a key functionality technological solutions for realizing a chatbot must enable. A further success factor for the acceptance of chatbots by customers is the provision of an immediate answer to user input (Gennermann and Hack, 2011). Receiving an answer to a specific question immediately (approx. within 0,1 seconds) gives users the "illusion" to communicate with a real person and, hence, largely contributes to the quality of the communication (Miller, 1968; Gennermann and Hack, 2011). Nevertheless, this illusion is reinforced if the response times slightly vary to simulate the cognitive thought processes of humans before giving an answer. Because of that, the option to configure the response times of chatbots represents an important functionality to evaluate chatbot solutions *(functionality 1.2)*. More, a memory function is required for chatbots to imitate a human being convincingly (Vetter, 2003). That way, information – such as the "name" of the user – can be stored and referenced

later on throughout the conversation (Vetter, 2003) *(functionality 1.3)*. More, certain information customers search for may be found on particular webpages, and, hence, the insertion of corresponding internet links to a chatbot's response is a valuable functionality contributing to the perceived quality of a communication (cf. Kuligowska, 2015; Shawar and Atwell, 2007) *(functionality 1.4)*. For the purpose of assessing users' satisfaction with the chatbot, obtaining feedback about the chatbot's advisory skills is helpful (Klostermann, 2006; Klostermann, 2007; Gennermann and Hack, 2011). Ideally, giving customers the opportunity to submit feedback in the follow-up of a conversation represents a functionality of the chatbot solution itself (Kuligowska, 2015; Callejas et al., 2011) *(functionality 1.5)*. Sometimes, customer concerns and requests may be too complex and individual and, thus, require the intermediation of an employee. In such cases, the customer inquiries should be handed over to the corresponding employee by the chatbot directly (Gennermann and Hack, 2011) *(functionality 1.6)*. This functionality also facilitates the implementation of chatbots as customer touchpoints in the light of holistic service processes (cf. Følstad and Brandtzæg, 2017). Hence, chatbots can be installed by a company to acquire the relevant information that is further processed by employees in the back-office, e.g., in the course of complaint management. In this respect, the chatbot should also provide its users with a protocol of the conversation upon request *(functionality 1.7)* (Gennermann and Hack, 2011). More, the detection of synonyms within a conversation is desirable for being able to adequately respond to users (Bradeško and Mladenić, 2012) *(functionality 1.8)*. Also the technical and language barriers of a chatbot must be kept low to enable people with language problems or disabilities the use of the chatbot (Gennermann and Hack, 2011; Klostermann, 2007) *(functionality 1.9)*. In this context, voice input and output becomes increasingly important as well (Dale, 2016; Kuligowska, 2015) *(functionality 1.10)*. The wish for voice support of chatbots is largely boosted by the upcoming of virtual assistants such as Apple's Siri, Microsoft's Cortana, Google's Assistant and Amazon's Alexa for instance (Armstrong, 2017). Finally, language packages provided for a particular chatbot solution should consider slang, dialect or lexical variations (e.g., Gray and Hansen, 2005; Petz et al., 2013) that may be used by a certain target-audience to assure a high quality of the conversation *(functionality 1.11)*. As mentioned, the language of customers of SMEs is usually characterized by colloquial language or regional dialects (Schwaiger et al., 2016).

(2) Reaction to inappropriate input: A chatbot must react to unspecific user input (e.g., by asking the customer again) or to inappropriate input (e.g., swear, negative or sexual words) in a polite but quick-witted way to maintain users' respect *(functionalities 2.1 and 2.3)* (Pilato et al., 2011; Vetter, 2003; Kuligowska, 2015). In this regard, a chatbot should not be passive but try to bring close those contents it is supposed to provide (Vetter, 2003). Additionally, there is also the danger that users willingly manipulate chatbots as was the case with Microsoft's bot "Tay" (Steiner, 2016; Wakefield, 2016). Hence, means to mitigate a corresponding "re-education" of a chatbot solution is required *(functionality 2.2)*.

(3) Personalization: Avatars are used in human-machine interaction to further strengthen users' impression to actually communicate with human beings (cf. Angga et al., 2015; Lindner, 2003; Sheth, 2003). Therefore, they are a popular means for simulating human communication behaviour *(functionality 3.1)*. Chatbots providing an avatar are less seen as a computer program but much more as a real communication partner (cf. Braun, 2003; Berry et al., 2004). However, the chatbot and the avatar, respectively, should be individually adapted to the domain they are applied for *(functionality 3.2)* (Gennermann and Hack, 2011). Otherwise, there is the danger of distracting users from the communication as was the case with Microsoft's "Clippit" in the past for instance (Sheth, 2003). Furthermore, chatbots with "small talk" abilities are felt to be more trustful and entertaining (Kopp et al., 2005; Klüwer, 2011; Spierling et al., 2006). Hence, the functionality to perform small talk is desirable for chatbot solutions to optimize the personal "customer experience" (cf. Schmitt, 2009) when interacting with the chatbot (Kuligowska, 2015) *(functionality 3.3)*. More, to shape the conversation in a lively way, a chatbot should provide different but correct answers to identical user input (e.g., introductory information) (Geeb, 2007; Gennermann and Hack, 2011) *(functionality 3.4)*.

(4) Language processing: In terms of language processing, different languages should be supported by a chatbot because firms may operate on an international level or customers might switch languages within text messages (Schwaiger et al., 2016) *(functionality 4.1)*. More, a spell-checking functionality should be provided (Polatidis, 2014; Kuligowska, 2015). That way, chatbots are more fault tolerant

considering user input which might not be grammatically correct in each case (Gennermann and Hack, 2011; Schwaiger et al., 2016) *(functionality 4.2)*. Indeed, spelling errors (e.g., "Helllllooooo") are observed above-average considering human-machine interaction on the internet (Laboreiro et al., 2010; Petz et al., 2013; Schwaiger et al., 2016). More, while texts are often written in lowercase, capitalization is used to express emotions in some cases, and, hence, dealing with case sensitivity is a valuable feature (Gennermann and Hack, 2011) *(functionality 4.3)*. More, the detection of irony and sarcasm within a conversation is desirable for being able to adequately respond to users (González-Ibánez et al., 2011; Davidov et al., 2010) *(functionality 4.4)*. Although the automatic detection of irony and sarcasm is highly challenging, González-Ibánez et al. (2011) show that it can be nearly as good as by human assessment in case appropriate mechanisms are applied. Those applied by the participants will be shown hereafter.

(5) Security: In terms of security, the encoded data storage and data transmission are highly relevant for branches transferring sensitive data, e.g., in the financial service industries (e.g., Armstrong, 2017) and should thus be supported by chatbot solutions *(functionalities 5.1 and 5.2)* (Klopfenstein et al., 2017). In this context, also the authentication of the communication partner towards the chatbot (or a registration) may be desired to enable the use of the chatbot for a specific target-audience only *(functionality 5.3)* (Klopfenstein et al., 2017). To prevent unrestricted modifications to the chatbot, a role concept (cf. Bachman and Daya, 1977) for administration and configuration purposes is reasonable *(functionality 5.4)*. Hence, the opportunity to establish a corresponding role concept is a looked-for functionality by firms. More, data backups and the support of the PCI security standards *(https://www.pcisecuritystandards.org/) (functionalities 5.5 and 5.6)* are functionalities that are highly relevant for firms processing sensitive data, e.g., account data.

(6) Interfaces & mobile device support: Chatbots must be easily integrated into websites to be directly found and used by visitors (Kuligowska, 2015; Gennermann and Hack, 2011) *(functionality 6.1)*. However, recently, the integration of chatbots with messengers has become the centre of attention (Dale, 2016; Chaykowski, 2016). Accordingly, firms see massive opportunities in providing an automatized customer service via messengers in social media platforms such as Facebook for instance (Iftene and Vanderdonckt, 2016). Therefore, the ability to be integrated with messengers is more and more becoming a central requirement on chatbot solutions *(functionality 6.2)*. Furthermore, user input may be directly transferred into other IT-systems, e.g., CRM-systems, for the purpose of further processing, which, though, requires chatbot solutions to offer corresponding interfaces *(functionality 6.3)* (e.g., Klopfenstein et al., 2017). Additionally, with the spread of mobile devices within the population, chatbots must be runnable on corresponding devices just as they are executable on work stations *(functionality 6.4)* (Dale, 2016; Price and Lewis, 2017).

(7) Training: Especially SMEs with limited personal capacities neither have the resources to devote themselves to a laborious training of the chatbot nor do they have profound IT skills (cf. Bull, 2003; Newby et al., 2014). So, to constantly improve the conversation skills of a chatbot, it should build on self-learning algorithms (Guo et al., 2014). That way, chatbots may revert to customer data or input to automatically enhance the response spectrum for consumers (Kühl, 2016) *(functionality 7.1)*. However, also a manual training of chatbots may be strived for. For that purpose, corresponding GUIs are necessary *(functionality 7.2)*. What is decisive for the rapid introduction of chatbots at firms is the ability to train them without specific technical skills being required *(functionality 7.3)*.

(8) Implementation: To enable a quick implementation of the chatbot, an initial language package (e.g., Baccianella et al., 2010) should be delivered by chatbot providers straight away *(functionality 8.1)*. Furthermore, due to the lack of IT skills at SMEs (cf. Bull, 2003; Newby et al., 2014), an initial hosting or implementation of the chatbot, respectively, without programming skills being required should be possible *(functionality 8.2)*. Additionally, open accessible discussion forums on the internet, to immediately get help in terms of running the chatbot, would be helpful for SMEs in addition *(functionality 8.3)*.

(9) Costs: As previously mentioned, license/billing conditions should consider the limited buying powers of SMEs (Bharati and Chaudhury, 2009) *(functionality 9.1)*. Further, the running of the chatbot as a "software as a service (SaaS)" solution would be attractive for SMEs, considering scarce financial budgets but also the lack of IT know-how within firms (cf. Bharati and Chaudhury, 2009) *(functionality 9.2)*.

3.3 Questionnaire and collection of data

The above-described functionalities were then transferred to a corresponding questionnaire. The nine categories as shown presented corresponding sections of the questionnaire and we captured each functionality in form of a question. We then asked whether a chatbot provider offered the corresponding functionality or not. Wherever possible, the answers were specified as a 3-point Likert scale (degree of fulfilment – "0" not offered; "1" partly offered; "2" offered). In other cases the items ask for a "yes/no" answer. To achieve additional information, we attached free text fields to each item, asking participants to specify the assessment made in more detail (e.g., the languages supported by the chatbot solution). As an additional help, we explained all terminology and gave short explanations for each item to reduce subjective interpretations of a question. Four researchers designed the questionnaire and validated it in a workshop with two practitioners of a large consulting company who had long years' experience in conducting surveys in practice. The questionnaire was designed as an interactive Excel file and sent to the 14 participants of our study via email in the period from November 2016 to January 2017.

Each participant was given a timeframe of two weeks to complete and return the questionnaire. During this period, the researchers could be asked for assistance in case ambiguities occurred. We then screened the responses for completeness as well as consistency. In case of incomplete information (regarding specific items) the free text fields were drawn upon to see whether conclusions on the associated Likert scales could be made or not. In cases of ambiguity or lack of information the affected items were excluded from the calculation regarding the corresponding chatbot provider. Consistency checks referred to the relation between the Likert scale assessment of an item and the additional information provided via the free text fields. Inconsistencies would be given in case a provider claimed to support different languages but only mentioned one language in the free text field for instance. In such cases, we could either resolve the ambiguities based on the information provided (which was done by three researchers to avoid subjectivity) or the corresponding data was not considered for the calculation.

4 Results and Discussion

4.1 Results

Table 2 summarizes the results of the investigation and presents as to what degree the commercial chatbot providers cover the functionalities as posed in Table 1. However, Table 2 does not show the results for the category *(9) costs*, because a universally valid assessment of the suitability of the licence/billing models for all kinds of firms is not possible. Though, we shortly summarize the results for costs in the following. Generally, the providers saw their chatbot solutions as means to support (online) customer care, the sales conversation, the internal helpdesk, the customer engagement in marketing or even as a tool to serve as a virtual tour guide of smart city concepts. In this respect, the functionalities – as specified in section 3.2 – were largely covered by the chatbot solutions investigated.

Considering the **(1) quality of the communication**, a high coverage of the corresponding functionalities was observed in general (total average coverage of 89%). Thereby, five solutions support all the functionalities as mentioned whereas others have drawbacks (see Table 2). In this respect, all providers submitting answers for this category fully covered *functionalities 1.4, 1.5, 1.7* as well as *1.8*. Accordingly, the providers enabled chatbots to insert internet links to a response given and ask for users' satisfaction with the advisory skills. Further, they also provide a "history" function (conversation logs) and offer the capability to identify synonyms within a conversation. Nearly all providers met the *functionalities 1.1, 1.3, 1.9* and *1.10* with singular solutions representing statistical outliers in terms of particular items. For instance, only one provider (**P2**) claimed to not offer a "memory function" whereas another one (**P5**) did not support the adaption of a chatbot's language to the target-audience. In our study the lowest importance was ascribed to the ability to individually configure a chatbot's response times *(functionality 1.2)*, which was not supported by three providers providing answers for this category (**P1, P2, P5**). Additionally, dealing with slang, lexical variations or dialect seems to pose major problems because

only six providers (**P3, P4, P5, P6, P8, P9**) stated to fully consider these issues via corresponding language packages *(functionality 1.11)*. In summary, providers undertake enormous efforts for establishing a high quality of the communication, but especially in terms of lexical variations, slang or dialect potentials for future refinements of the solutions exist.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Total avg.
(1) Quality of the communication	82%	64%	100%	100%	82%	91%	82%	100%	100%	91%	-	82%	*100%	82%	*******89%
(2) Reaction to inappropriate input	100%	**100%	100%	100%	100%	100%	100%	100%	100%	67%	-	100%	100%	100%	*******97%
(3) Personalization	100%	75%	100%	100%	100%	50%	100%	100%	100%	100%	-	100%	100%	100%	*******94%
(4) Language processing	75%	100%	100%	75%	50%	75%	75%	100%	100%	50%	-	50%	100%	75%	*******79%
(5) Security	83%	83%	100%	***100%	83%	67%	83%	100%	100%	83%	50%	83%	***80%	100%	85%
(6) Interfaces & mobile device support	100%	****100%	100%	****100%	****100%	100%	100%	100%	100%	100%	75%	100%	****100%	100%	98%
(7) Training	100%	33%	100%	100%	67%	100%	33%	100%	100%	67%	-	100%	100%	67%	*******82%
(8) Implementation	67%	33%	67%	67%	67%	100%	67%	67%	100%	33%	*****100%	0%	100%	67%	67%
Total avg. for categories 1-8	88%	74%	96%	93%	81%	85%	80%	96%	100%	74%	******-	77%	98%	86%	

Key:
*: No answer to functionality 1.11 given by the provider; percentage refers to the coverage of the other functionalities
**: No answer to functionality 2.2 given by the provider; percentage refers to the coverage of the other functionalities
***: No answer to functionality 5.6 given by the providers; percentage refers to the coverage of the other functionalities
****: No answer to functionality 6.3 given by the providers; percentage refers to the coverage of the other functionalities
*****: No answer to functionality 8.1 given by the provider; percentage refers to the coverage of the other functionalities
******: No average value calculated due to lack of data regarding categories 1 to 4 and 7
*******: Total average value only considers providers for which percentage values could be calculated

100% fulfillment
<=50% fulfillment

Table 2. Results overview (percentage of functionalities met per chatbot provider)

The functionalities *2.1, 2.2,* and *2.3* related to the **(2) reaction to inappropriate input** were largely covered by the solutions investigated (average coverage of 97%). Only, the ability to reply to inappropriate user input (e.g., swear, negative or sexual words) *(functionality 2.1)* was reported not to be supported by one provider (**P10**). So, means to react to unspecific or inappropriate input as well as to protect the chatbot from being manipulated can be seen as standard functionalities offered by the providers.

The providers make various propositions in terms of **(3) personalization** and, accordingly, we observed a total average coverage of the associated functionalities of 94%. Holding "lively" conversations via "small talk" and communicating in a diversified way was claimed to be supported by all chatbot providers who gave information on that item in the questionnaire *(functionalities 3.3 and 3.4)*. Contrary, one provider did not support the generation and individualization of an avatar for instance, while one provider did not give any information on that category at all *(functionalities 3.1 and 3.2)*. Asked for reasons to not support an avatar, "**P6**" stated that the *"use of avatars was discontinued"* because this feature was believed to be *"counterproductive"* for chatbot use in a business setting. However, from a general perspective, the personalization of chatbots is a feature current providers largely support.

Generally, diversified observations were made regarding the **(4) language processing** abilities of the solutions (total average coverage of 79%) and five providers fully supported the corresponding functionalities. Thereby, all chatbot providers considered multiple languages, whereas one solution even supported 35 languages (**P9**) *(functionality 4.1)*. Two providers (**P10 and P12**) stated not to support spell-checking while this number increases to six providers (**P1, P5, P6, P7, P12, P14**) regarding the consideration of case sensitivity *(functionalities 4.2 and 4.3)*. Three providers claimed their solution not to support sarcasm and irony detection (**P4, P5 and P10**) *(functionality 4.4)*. To identify irony and sarcasm, the other providers either equipped their chatbots' dictionaries with "keywords" or "phrases" that indicate ironic or sarcastic expressions or offered a sophisticated natural language processing (NLP) engine (e.g., **P3**). To sum up, differences in the capabilities of the solutions regarding language processing abilities were given and the detection of sarcasm and irony seems to be challenging and a topic future initiatives to develop the technologies further may focus on.

There were also differences concerning **(5) security** aspects, which is relevant for the processing of sensitive data (total average coverage of 85%). At first, none of the solution providers fully covered all functionalities. However, encoded data storage and transmission, the option to implement a role model for administration purposes as well as the provision of data backups were those functionalities covered to the highest degree (13 providers each) *(functionalities 5.1, 5.2, 5.4 and 5.5)*. A more nuanced differ-

entiation was given regarding the support of PCI regulations *(functionality 5.6)*. Only five solution providers (**P3, P8, P9, P11, P12 and P14**) explicitly claimed to support corresponding standards (e.g., ISO 27001). In summary, considering chatbot usage for application fields processing sensitive data, companies must carefully look at the single providers and choose one meeting their particular needs. This is because only five providers showed a coverage of 100% regarding security-related functionalities.

Considering the proper integration of chatbots with a company's business processes and the wide spread of mobile devices the functionalities related to **(6) interfaces & mobile device support** are gaining importance (total average coverage of 98%). Not much difference was observed for the providers in this regard. For instance, all providers stated to support the integration of their solutions with messengers and their execution on mobile devices *(functionalities 6.2 and 6.4)*. All providers except for one supported the integration into websites **(P11)** *(functionality 6.1)*. Further, ten providers (**P1, P3, P6-P12 and P14**) claimed to facilitate the processing of input data by other IT-systems via the offering of corresponding interfaces (e.g., XML, JSON, etc.) *(functionality 6.3)*.The remaining four providers made no indications in this respect. It can be acknowledged that providers undertake considerable efforts for assuring chatbots to be integrated into a firm's IT landscape and to be executed on mobile devices.

Differences were given in terms of the **(7) training** functionalities (as posed in Table 1) (total average coverage of 82%). Whereas one provider gave no indication on the learning ability of his solution (**P11**), the majority of providers (twelve) indicated that no technical skills are required for training the chatbot (**P1-P9 and P12-P14**) *(functionality 7.3)*. However, only nine providers stated that their products build on self-learning algorithms to improve the conversation skills (**P1, P3, P4, P6, P8, P9, P10, P12 and P13**) *(functionality 7.1)*, whereas eleven providers (**P1, P3-P6, P8-P10 and P12-P14**) enhanced their solutions with GUIs to facilitate a manual training of the chatbot *(functionality 7.2)*. As becomes obvious from these results particular solutions perform significantly better in terms of training than others (see Table 2), a fact firms may consider when selecting a provider. Especially for SMEs with limited IT skills, a high coverage of the training functionalities is of central importance.

Considering the functionalities supporting an easy **(8) implementation** of a chatbot differences between the solution providers got evident (total average coverage of 67%). Whereas only four providers claimed to support all functionalities **(P6, P9, P11 and P13)**, most providers lack a supportive discussion forum *(functionality 8.3)* assisting during the implementation. Contrary, twelve providers stated that no programming skills are required for implementing a chatbot *(functionality 8.2)* and twelve offer prepared language packages for a direct start of the chatbot *(functionality 8.1)*. Hence, single providers stood out in terms of an easy implementation **(P6, P9, P11 and P13)**, which is especially relevant for chatbot applications at SMEs being characterized by limited IT skills. Accordingly, the implementation functionalities can clearly be seen as a major differentiator between the solutions.

Finally, in terms of **(9) costs**, almost all providers (**P1-P9 and P11-P14**) said that they enable to run a chatbot (realized via their software) as a SaaS solution *(functionality 9.2)*. Licence models *(functionality 9.1)* range from country-specific pricing models (e.g., **P12**), session licences (e.g., **P1**), annual licences (e.g., **P3**), licences depending on the number of user requests (e.g., **P14**) to individual pricing models considering the use case for which chatbots are to be realized (e.g., **P6, P9**). Accordingly, the providers offer license models for different companies of various sizes and branches.

In summary, the solutions for realizing chatbots all have particular strengths and, hence, a general suggestion on which tool to use cannot be given. Much more firms have to prioritize the categories or functionalities, respectively, for being able to select a solution meeting their demands.

4.2 Discussion and reflection against the requirements

By reflecting the results against the requirements (see section 2) several conclusions can be drawn. At first, the solutions investigated undertake enormous efforts to assure a high quality of the communication ***(see Rq1)***. In this respect, some solutions (**P3, P4, P8 and P9**) fulfil all the functionalities as described in Table 1. Nevertheless, dealing with lexical variations, slang or dialect seems to pose challenges because only six providers claimed to be fully able to cope with these language peculiarities (see section 4.1). Though, this aspect is particularly relevant for the chatbot application at SMEs or (large) firms

acting within a limited regional region ***(see Rq1, Rq4, Rq11)***. As mentioned, in such settings the language of customers is imprinted by regional dialect, slang expressions and colloquial language (cf. Schwaiger et al., 2016). Considering this, potentials for a further development of software solutions can be acknowledged. Second, the market of software solutions for chatbot realization seems to have a high degree of maturity when it comes to coping with inappropriate user input ***(see Rq2)***. The functionalities used to operationalize this requirement (see Table 1) where largely covered by the providers investigated. This concerns the reaction to inappropriate or unspecific user input as well as the provision of mechanisms to avoid user manipulations (see Table 2). Third, the generation of the illusion to actually communicate with a real person ***(see Rq3)*** is a topic most providers intensively deal with. Only one provider does not offer the opportunity to generate and individualize avatars, though the said provider has particular reasons for not doing so as described in section 4.1 (**P6**). At the same time, holding small talk and diversified conversations are issues the software solutions largely cover. Fourth, while the support of multiple languages is guaranteed by all providers investigated ***(see Rq5)***, spell-checking (supported by eleven providers) and case sensitivity (supported by seven providers) ***(see Rq6)*** are not considered without exceptions. Differences between the providers also get evident when it comes to irony and sarcasm detection ***(see Rq7)***. Thereby, only ten providers claimed to fully consider irony and sarcasm detection.

Fifth, security-related issues play an increasingly important role for the solution providers analysed ***(see Rq8)***, making them applicable for branches processing sensitive data. Only in singular cases, providers said not to support particular functionalities such as encoded data storage and transmission or authentication towards the chatbot amongst others (see section 4.1). Nevertheless, in terms of supporting PCI regulations there seem to be drawbacks with only six providers explicitly stating to fully consider these in their solutions. Sixth, the use of messengers and mobile devices has risen tremendously in recent years (Patil et al., 2017; Statista, 2016). This trend is considered by all of our study participants because the integration and compatibility with messengers ***(see Rq9)*** and mobile devices is explicitly supported ***(see Rq10)***. More, ten providers said to offer interfaces to other IT-systems ***(see Rq10***) allowing an easy further processing of data. Hence, the ground is prepared for an integration of chatbots with a firm's IT-landscape. Seventh, regarding an easy training of the chatbot ***(see Rq12)*** it turned out that self-learning algorithms are applied by many providers, however, with the general danger of users trying to manipulate a chatbot (e.g., Microsoft's "Tray") (e.g., Steiner, 2016; Wakefield, 2016). Nonetheless, all providers offering self-learning algorithms (nine providers) stated to have implemented techniques to prevent user manipulations. The manual training of the chatbot without technical skills being required was explicitly confirmed by twelve providers. Hence, current solutions on the market are suitable to be applied at SMEs or firms lacking IT skills at large. Eighth, the implementation of chatbots ***(see Rq13)*** via the software solutions investigated may be facilitated in future versions. As mentioned, particularly SMEs have limited IT expertise and, hence, in-depth programming knowledge is not available within the workforce. Twelve providers explicitly claimed not to require programming skills to run the chatbot and eleven offer prepared language packages to guarantee a quick installation. However, more supportive discussion forums could be helpful to foster the spreading of chatbots (see section 4.1). Ninth, flexible pricing models make the realization of chatbots affordable, even for SMEs with limited IT budgets ***(see Rq14)***. In this respect, 13 providers stated to support the realization of chatbots as "SaaS".

Summing up, the rise of chatbots and natural language interfaces brings about three major implications for HCI according to Følstad and Brandtzæg (2017). (I) It is no longer the visual layout of the user interface but the conversation itself that becomes the object of design (Følstad and Brandtzæg, 2017). Considering this, functionalities associated with the "quality of the communication" as well as the "reaction to inappropriate user input" that enable to design the conversation with respect to a target-audience were largely covered by all providers of our investigation (see Table 2). Some challenges existed in terms of language processing abilities (e.g., recognition of slang, etc.). However, this offers opportunities for further research and the development of means to cope with language peculiarities to adjust conversations to particular users even more. (II) Additionally, companies need to move from interface design to service design (Følstad and Brandtzæg, 2017). This development is paid attention to by the

providers allowing companies to implement the chatbot as a customer touchpoint and enabling employees to actively intervene in chatbot interactions or use the acquired information to offer services accordingly. (III) Finally, chatbots need to be designed for interaction in networks of human and intelligent machine actors (Følstad and Brandtzæg, 2017). In this respect, particularly self-learning abilities are of utmost importance (Følstad and Brandtzæg, 2017). It became evident that some providers refrain from using self-learning algorithms to prevent user manipulations as have been performed on Microsoft's chatbot Tay for instance (cf. Steiner, 2016; Wakefield, 2016). Hence, the further development of mechanisms supporting the systematic and automatized learning in networks of humans and machines seems to be a promising field for chatbot providers but also for research efforts.

5 Conclusion and Contribution

In the research at hand, the abilities of chatbot providers to cover functionalities relevant for an entrepreneurial use of chatbots at SMEs (in particular) were investigated. Hence, our research differentiates from current evaluations of concrete chatbots in literature (e.g., Kuligowska, 2015) because we focus on the abilities of existing solutions to realize a chatbot.

The research brings beneficial insights for practice and research alike. Practitioners receive a profound overview of the current capabilities of chatbot solutions offered. In this respect, it turned out that a general suggestion on which tool is superior to the others cannot be made. Much more, the providers emphasize different functionalities of their solutions, which are characterized by particular strengths correspondingly. Hence, companies seeking to implement a chatbot must carefully prioritize their requirements for being able to choose the solution that meets their expectations best. Our findings support this selection process since the coverage of certain functionalities – that are relevant for firms – by single providers is explicated in detail. More, providers of chatbot solutions receive valuable indications on which functionalities may help to differentiate from other products. Especially the handling of dialect and colloquial language, the consideration of security standards as well as the support during training and implementation seem to be promising in this respect.

Picking up that point, research may focus on the development of dictionaries considering dialect and colloquial language that are adapted for particular geographical regions or niches, respectively. That way, the perceived quality of the communication can be increased on the customer side and a handover of easy-to-solve requests to human agents reduced. Additionally, creating mechanisms to support an automatized learning in networks of humans and machines (cf. Følstad and Brandtzæg, 2017) is a promising field for research. More, research may focus on the question of how PCI regulations can be better integrated with chatbot solutions and how efficient training opportunities and implementation support (e.g., via web-based-trainings) for chatbots may look like. Finally, the research brought beneficial insights on how requirements on chatbot usage in firms can be operationalized in form of particular functionalities to be met by solutions designed for realizing chatbots.

However, there are limitations: at first, the results build on the answers as they were given by the providers of chatbot solutions. Hence, an embellishment of the answers cannot be fully excluded. However, as mentioned, providers' knowledge is decisive for identifying all functionalities of a solution, and hence, we collected the data via a corresponding online survey. Furthermore, the number of participants is restricted to 14. In total, this makes a participation rate of 29% (14 out of 48 providers contacted) which is still considerable. Another restriction is that the survey was conducted online and not in a personal face-to-face meeting, which would have allowed to ask questions for further clarification. Though, the design of the study as an online survey enabled firms to participate independently of time and location, contributing to the willingness to participate in the investigation. Finally, the results represent a snapshot of the corresponding survey period. The solutions are continuously further developed by the providers and new functionalities may have been added since the initial data collection.

In future work, we will complement the insights gained from this study with findings from an extensive test of chatbots that have been realized with the aforementioned solutions. That way, the results will be validated and an additional perspective on the usability of implemented chatbots taken.

References

Abdul-Kader, S. A. and Woods, J. (2015). *Survey on chatbot design techniques in speech conversation systems*. Int. J. Adv. Comput. Sci. Appl. (IJACSA), 6 (7).

Andrews, T. (2012). *What is social constructionism*. Grounded theory review, 11 (1), 39-46.

Angga, P. A., Fachri, W. E., Elevanita, A. and Agushinta, R. D. (2015). *Design of chatbot with 3D avatar, voice interface, and facial expression*. International Conference on Science in Information Technology 2015 (ICSITech), pp. 326-330.

Baccianella, S., Esuli, A. and Sebastiani, F. (2010). *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*. LREC, pp. 2200-2204.

Bachman, C. W. and Daya, M. (1977). *The role concept in data models*. Proceedings of the third international conference on Very large data bases – Volume 3, pp. 464-476.

Barba-Sánchez, V., Martínez-Ruiz, M. d. P. and Jiménez-Zarco, A. I. (2007). *Drivers, benefits and challenges of ICT adoption by small and medium sized enterprises (SMEs): a literature review*. Problems and Perspectives in Management, 5 (1), 104-115.

Berry, D., Butler, L., de Rosis, F., Laaksolahti, J., Pelachaud, C. and Steedman, M. (2004). *Embodied Believable Agents*. University of Edinburgh.

Berthon, P. R., Pitt, L. F., Plangger, K. and Shapiro, D. (2012). *Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy*. Business Horizons, 55 (3), 261-271.

Bharati, P. and Chaudhury, A. (2009). *SMEs and competitiveness: The role of information systems*. Management Science and Information Systems Faculty Publication Series. Paper 15.

Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K., Pfeifer, L. M. and O'Donnell, J. (2010). *Response to a relational agent by hospital patients with depressive symptoms*. Interacting with computers, 22 (4), 289-298.

Bickmore, T. W., Schulman, D. and Sidner, C. (2013). *Automated interventions for multiple health behaviors using conversational agents*. Patient education and counseling, 92 (2), 142-148.

Bradeško, L. and Mladenić, D. (2012). *A survey of chatbot systems through a loebner prize competition*. Proceedings of Slovenian Language Technologies Society, Eighth Conference of Language Technologies, pp. 34-37.

Braun, A. (2003). *Chatbots in der Kundenkommunikation*. Springer-Verlag.

Bull, C. (2003). *Strategic issues in customer relationship management (CRM) implementation*. Business Process Management Journal, 9 (5), 592-602.

Burr, V. (2006). *An introduction to social constructionism*. Routledge.

Callejas, Z., Griol, D. and López-Cózar, R. (2011). *Predicting user mental states in spoken dialogue systems*. EURASIP Journal on Advances in Signal Processing, 2011 (1), 6.

CapGemini (2016). *Studie IT-Trends 2016*. https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2016/02/it-trends-studie-2016.pdf.

Comendador, B. E. V., Francisco, B. M. B., Medenilla, J. S. and Mae, S. (2015). *Pharmabot: a pediatric generic medicine consultant chatbot*. Journal of Automation and Control Engineering 3 (2).

Dale, R. (2016). *The return of the chatbots*. Natural Language Engineering, 22 (5), 811-817.

Davidov, D., Tsur, O. and Rappoport, A. (2010). *Semi-supervised recognition of sarcastic sentences in twitter and amazon*. Proceedings of the fourteenth conference on computational natural language learning, pp. 107-116.

Deshpande, A., Shahane, A., Gadre, D., Deshpande, M. and Joshi, P. M. (2015). *A SURVEY OF VARIOUS CHATBOT IMPLEMENTATION TECHNIQUES*. International Journal of Computer Engineering and Applications, XI (Special Issue).

Dole, A., Sansare, H., Harekar, M. R. and Athalye, M. S. (2015). *Intelligent Chat Bot for Banking System*. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 4 (5).

Durkin, M., McGowan, P. and McKeown, N. (2013). *Exploring social media adoption in small to medium-sized enterprises in Ireland*. Journal of Small Business and Enterprise Development, 20 (4), 716-734.

Faed, A., Hussain, O. and Chang, E. (2014). *A methodology to map customer complaints and measure customer satisfaction and loyalty*. Service Oriented Computing and Applications, 8 (1), 33-53.

Følstad, A. and Brandtzæg, P. B. (2017). *Chatbots and the new world of HCI*. Interactions, 24 (4), 38-42.

Gallaugher, J. and Ransbotham, S. (2010). *Social Media and Customer Dialog Management at Starbucks*. MIS Quarterly Executive, 9 (4), 197-212.

Geeb, F. (2007). *Chatbots in der praktischen Fachlexikographie/Terminologie*. LDV Forum, pp. 51-70.

Gennermann, H. and Hack, S. (2011). *Qualitätsstandards für Chatbots in der bibliothekarischen Auskunft in Deutschland*. BIT online - innovativ, 111.

Gimpel, H. and Röglinger, M. (2015). *Digital Transformation: Changes and Chances–Insights based on an Empirical Study*. Fraunhofer Report.

Goda, Y., Yamada, M., Matsukawa, H., Hata, K. and Yasunami, S. (2014). *Conversation with a Chatbot before an Online EFL Group Discussion and the Effects on Critical Thinking*. The Journal of Information and Systems in Education, 13 (1), 1-7.

González-Ibánez, R., Muresan, S. and Wacholder, N. (2011). *Identifying sarcasm in Twitter: a closer look*. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers – Volume 2, pp. 581-586.

Gorelov, Z. (2016). *Introducing KAI Banking on Messaging and MyKAI*. http://kasisto.com/introducing-kai-banking-on-messaging-and-mykai/.

Gray, S. and Hansen, J. H. (2005). *An integrated approach to the detection and classification of accents/dialects for a spoken document retrieval system*. IEEE Workshop on Automatic Speech Recognition and Understanding 2005, pp. 35-40.

Greenberg, P. (2010). *The impact of CRM 2.0 on customer insight*. Journal of Business & Industrial Marketing, 25 (6), 410-419.

Guo, Y., Bai, L., Lao, S., Wu, S. and Lew, M. S. (2014). *A Comparison between Artificial Neural Network and Cascade-Correlation Neural Network in Concept Classification*. Pacific Rim Conference on Multimedia, pp. 248-253.

Hanna, R., Rohm, A. and Crittenden, V. L. (2011). *We're all connected: The power of the social media ecosystem*. Business Horizons, 54 (3), 265-273.

Heller, B., Proctor, M., Mah, D., Jewell, L. and Cheung, B. (2005). *Freudbot: An investigation of chat-bot technology in distance education*. EdMedia: World Conference on Educational Media and Technology, pp. 3913-3918.

Hill, J., Ford, W. R. and Farreras, I. G. (2015). *Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations*. Computers in Human Behavior, 49, 245-250.

Horzyk, A., Magierski, S. and Miklaszewski, G. (2009). *An Intelligent Internet Shop-Assistant Recognizing a Customer Personality for Improving Man-Machine Interactions*. Recent Advances in intelligent information systems, 13-26.

Iftene, A. and Vanderdonckt, J. (2016). *MOOCBuddy: a chatbot for personalized learning with MOOCs*. RoCHI – International Conference on Human-Computer Interaction.

Javaid, A., Kurjakovic, S., Masuda, H. and Kohda, Y. (2017). *Enabling Digital Transformation in SMEs by Combining Enterprise Ontologies and Service Blueprinting*. International Conference on Serviceology, pp. 224-233.

Kerry, A., Ellis, R. and Bull, S. (2009). *Conversational agents in E-Learning*. Applications and innovations in intelligent systems XVI, pp. 169-182.

Klopfenstein, L. C., Delpriori, S., Malatini, S. and Bogliolo, A. (2017). *The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms*. Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 555-565.

Klostermann, J. (2006). *Digitale Auskunft in Bibliotheken*. Working Paper University of Cologne.

Klostermann, J. (2007). *Digitale Auskunft in Bibliotheken. Eine vergleichende Evaluation zwischen ausgewählten digitalen Auskunftsverbünden*. Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft, Band 52.

Klüwer, T. (2011). *"I Like Your Shirt"-Dialogue Acts for Enabling Social Talk in Conversational Agents*. Intelligent Virtual Agents, pp. 14-27.

Kopp, S., Gesellensetter, L., Krämer, N. C. and Wachsmuth, I. (2005). *A conversational agent as museum guide–design and evaluation of a real-world application*. International Workshop on Intelligent Virtual Agents, pp. 329-343.

Kuligowska, K. (2015). *Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents*. Professionals Center for Business Research, 2 (2).

Kusber, R. (2017). *Chatbots – Conversational UX Platforms*. Innovationen und Innovationsmanagement in der Finanzbranche. Springer, pp. 231-244.

Laboreiro, G., Sarmento, L., Teixeira, J. and Oliveira, E. (2010). *Tokenizing micro-blogging messages using a text classification approach*. Proceedings of the fourth workshop on Analytics for noisy unstructured text data, pp. 81-88.

Lanning, M. and Phillips, L. (1991). *Building market-focused organizations*. White paper, Gemini Consulting.

Laudon, K. C. and Laudon, J. P. (2014). *Essentials of MIS*. Pearson Education Limited.

Lindner, C. (2003). *Wer braucht wofür Avatare? Konzeption und Implementierung natürlichsprachlicher Systeme – Zur Einführung*. Avatare. Springer, pp. 3-24.

Living Actor (2017). *Clients All over the World.* https://www.livingactor.com/corp/en/clients.

Lovelock, C., Patterson, P. G. and Wirtz, J. (2015). *Services marketing*. Pearson, Australia.

Matt, C., Hess, T. and Benlian, A. (2015). *Digital transformation strategies*. Business & Information Systems Engineering, 57 (5), 339-343.

McKinsey (2014). *Big business in small business: Cloud services for SMBs*. Report No. 25 Recall.

Miller, R. B. (1968). *Response time in man-computer conversational transactions*. Proceedings of the fall joint computer conference, part I, December 9-11, 1968, pp. 267-277.

Mukerjee, K. (2013). *Customer-oriented organizations: a framework for innovation*. Journal of Business Strategy, 34 (3), 49-56.

Newby, M., H. Nguyen, T. and S. Waring, T. (2014). *Understanding customer relationship management technology adoption in small and medium-sized enterprises: An empirical study in the USA*. Journal of Enterprise Information Management, 27 (5), 541-560.

Orenga-Roglá, S. and Chalmeta, R. (2016). *Social customer relationship management: taking advantage of Web 2.0 and Big Data technologies*. SpringerPlus, 5 (1), pp. 1462.

Patil, A., Marimuthu, K. and Niranchana, R. (2017). *Comparative study of cloud platforms to develop a Chatbot*. International Journal of Engineering & Technology, 6 (3), 57-61.

Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stříteský, V. and Holzinger, A. (2013). *Opinion mining on the web 2.0 – characteristics of user generated content and their impacts*. Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. Springer, pp. 35-46.

Pilato, G., Augello, A. and Gaglio, S. (2011). *A modular architecture for adaptive chatbots*. Fifth IEEE International Conference on Semantic Computing (ICSC 2011), pp. 177-180.

Polatidis, N. (2014). *Chatbot for admissions*. arXiv preprint arXiv:1408.6762.

Price, T. and Lewis, J. (2017). *Efficient e-commerce on virtual personal assistants*. Technical Disclosure Commons.

Reeves, B. and Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Sandoval-Almazán, R. and GutiérrezAlonso, M. A. (2009). *Virtual Assistants for E-Government Interaction*. Social and Political Implications of Data Mining: Knowledge Management in E-Government. IGI Global, pp. 255-266.

Schmitt, B. (2009). *Customer Experience Management*. (Bruhn, M. et al. Ed.). Handbuch Kommunikation: Grundlagen – Innovative Ansätze – Praktische Umsetzungen. Gabler, Wiesbaden, pp. 697-711.

Schwaiger, J., Lang, M., Johannsen, F. and Leist, S. (2017). *"WHAT DOES THE CUSTOMER WANT TO TELL US?" AN AUTOMATED CLASSIFICATION APPROACH FOR SOCIAL MEDIA POSTS AT SMALL AND MEDIUM-SIZED ENTERPRISES*. 25th European Conference on Information Systems (ECIS 2017), Guimarães.

Schwaiger, J. M., Lang, M., Ritter, C. and Johannsen, F. (2016). *Assessing the accuracy of sentiment analysis of social media posts at small and medium-sized enterprises in Southern Germany*. 24th European Conference on Information Systems (ECIS 2016), Istanbul.

Schwandt, T. A. (2000). *Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism*. Handbook of qualitative research, 2, 189-213.

Shawar, B. A. and Atwell, E. (2007). *Chatbots: are they really useful?* LDV Forum, pp. 29-49.

Sheth, R. (2003). *Avatar technology: giving a face to the e-learning interface*. The eLearning Developers' Journal, 1-10.

Spierling, U., Weiß, S. A. and Müller, W. (2006). *Towards accessible authoring tools for interactive storytelling*. International Conference on Technologies for Interactive Digital Storytelling and Entertainment, pp. 169-180.

Steiner, A. (2016). *Zum Nazi Und Sexisten in 24 Stunden*. Frankfurter Allgemeine Zeitung, http://www.faz.net/aktuell/wirtschaft/netzwirtschaft/microsofts-bot-tay-wird-durch-nutzer-zum-nazi-und-sexist-14144019.html, 2016.

Stoner, D. J., Ford, L., Ricci, M. and Williams, M. (2003). *Simulating military radio communications using chat-bot technology*. Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Orlando.

Vetter, M. (2003). *„Ich habe Ihre Eingabe leider nicht verstanden" – Qualitätskriterien für Online-Tests von Bots*. Avatare. Springer. 71-92.

Wakefield, J. (2016). *Microsoft Chatbot Is Taught to Swear on Twitter*. BBC News, http://www.bbc.com/news/technology-35890188.

Weizenbaum, J. (1966). *ELIZA – a computer program for the study of natural language communication between man and machine*. Communications of the ACM, 9 (1), 36-45.

Xu, A., Liu, Z., Guo, Y., Sinha, V. and Akkiraju, R. (2017). *A New Chatbot for Customer Service on Social Media*. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3506-3510.

Yan, R., Song, Y., Zhou, X. and Wu, H. (2016). *Shall I Be Your Chat Companion?: Towards an Online Human-Computer Conversation System*. Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp. 649-658.

2.8. **Beitrag 8:** Chatbot Design Features to Increase Productivity

Adressierte Forschungsfrage	Welche Gestaltungsmerkmale sollte ein Chatbot erfüllen, um Nutzer in einem produktivitätsorientierten Kontext bei ihrer Aufgabenerfüllung zu unterstützen?
Zielstellungen	Identifikation von Gestaltungsmerkmalen, welche sich förderlich auf die Produktivität der Nutzer von Chatbots im Rahmen operativer Prozesse innerhalb von Unternehmen auswirken.
Forschungsmethode	**Design Science Research** Durchführung eines systematischen Literature Reviews nach Webster und Watson (2002) sowie nach Vom Brocke et al. (2009) im Rahmen des Rigor Cycles für ein DSR-Projekt (vgl. Hevner, 2007).
Kernergebnisse (Überblick)	• Identifikation von insgesamt 72 Artikeln zu Chatbots und Conversational Agents, welche sich mit der Förderung der Produktivität ihrer Nutzer im Rahmen von operativen Aufgaben sowie Abläufen innerhalb eines Unternehmens befassen. • Ableitung von insgesamt 14 Gestaltungsmerkmalen, welche bei Chatbots und Conversational Agents zur Förderung der Produktivität ihrer Nutzer beitragen können. • Basierend auf der qualitativen Inhaltsanalyse nach Mayring (2002): Systematisierung der identifizierten Gestaltungsmerkmale durch die Konsolidierung der Gestaltungsmerkmale in insgesamt drei Kategorien: (a) “Design features augmenting the textual input paradigm with GUI elements”, (b) “Design features supporting the conversational process” und (c) “Design features personifying the chatbot”.
Publikationsort	**Veröffentlicht** in den Proceedings der 17. International Conference on Design Science Research in Information Systems and Technology, St. Petersburg, Florida, USA.
Ranking VHB JQ 4	**B**
Autoren	Lukas Ramming 50 % **Daniel Konadl** **40 %** Susanne Leist 10 %

Tabelle 10: Fact Sheet zu Beitrag 8

Chatbot Design Features to Increase Productivity

Lukas Ramming[1[0000-0003-3415-1029]], Daniel Konadl[1[0000-0002-3703-8024]] and Susanne Leist[1[0000-0002-2376-1027]]

[1] University of Regensburg, Germany
{Lukas.Ramming, Daniel.Konadl, Susanne.Leist}@ur.de

Abstract. In recent years, chatbots have become a growing presence in our everyday lives. Companies have identified various potential use-cases posing opportunities for reducing costs and providing services through automating processes with the help of chatbots. However, although an increasing number of chatbots are developed, user expectations often cannot be met, leading to frequent discontinuation of the bots. Research suggests that for users, one of the main reasons to use a chatbot is to help them increase their productivity. The literature base so far provides little prescriptive knowledge guiding implementation of chatbots specifically for use-cases where productivity is the main purpose. This short paper is the first step within a Design Science Research project to close this gap. We conducted a systematic literature review and gathered chatbot Design Features that were covered in respective publications. In the further course of our project, we intend to implement a prototype chatbot and formulate Design Principles to provide prescriptive knowledge for designing productivity oriented chatbots.

Keywords: Chatbots, Design Science Research, Rigor Cycle, Systematic Literature Review, Design Features, Productivity.

1 Introduction

Chatbots have become a well-known presence in the digital realm, and we encounter them in various contexts. Companies use chatbots for automating certain customer service tasks, providing users instant responses to their requests [1]. To clarify terms, with chatbots we refer to software programs that mimic human-to-human dialogues by receiving users' utterances (e.g., via text messages or via button clicks), processing the utterances and subsequently presenting a corresponding response to the user [1]. According to [2], one of the most frequent motivational factors for using chatbots is increased productivity, with chatbots providing task assistance and access to information in a timely manner. However, in spite of technological advancements, chatbots still fail to meet users' expectations, leading to dissatisfaction among users and discontinuation of chatbot projects [3, 4]. While the amount of research on chatbot design has increased tremendously in recent years [5], there is only little research on chatbot design that is aimed specifically at increasing productivity. One of the few studies about the topic is by [6] who measured productivity of users who completed different tasks with the chatbot assisting them. The productivity measure used by the authors is based on [7] and will also be used to operationalize productivity for our research project. Accordingly,

productivity is the ratio of output to input. In the context of chatbots, the output can be defined as the task or the number of tasks completed and the input as the amount of time needed for the completion [6]. In our project, we focus specifically on the potentials of chatbot design to increase productivity. We do not consider further types of solutions (e.g., self-service portals) within this research endeavour. Moreover, we do not yet focus on a specific application domain or specific types of tasks as we intended to gather a broad range of DFs based on a comprehensive analysis of literature. We formulate the following research question: *Which chatbot Design Features should be implemented in a context where increased productivity is a desired outcome?* In the next section, we describe the research procedure. We then present the results from our systematic literature review and conclude with an outlook on next steps within the Design Science Research (DSR) project.

2 Research Procedure

To investigate the research question, we follow the DSR approach laid out by [8]. While our overall DSR project will cover all three cycles of this approach, in this paper, we cover the Rigor Cycle as a first step. Thereby, we ground our project in the existing knowledge base by reviewing literature regarding the general topic of chatbot productivity. Following this paper, in the second step, we advance to the Relevance Cycle. With the help of practitioners in the field, we firstly want to identify a specific application and task domain and secondly, extend and prioritize the list of Design Features (DF) that we identified throughout this short paper for the specified domain. In the third step, we enter the Design Cycle and implement a productivity increasing chatbot for the specified domain and with identified DFs to derive Design Principles (DP) [9], providing prescriptive knowledge [10] to chatbot designers in a setting where high productivity is desired. Further details of the next steps following this short paper are described in section 4. To move forward in the Rigor Cycle, we conducted a systematic literature review following the methodological approaches of [11] and [12]. We focused our research on contributions that discuss chatbot design and provide recommendations or present empirical evidence of effects of DFs on goal variables related to productivity. The brackets accompanying the searched databases contain details on the respective keyword-based search queries as follows: (Hits per database | Investigated hits per search query | Papers identified as potentially relevant). Thereby, the following six databases and search engines were applied: *ScienceDirect* (4.358 | 250 | 22)*, SpringerLink* (3.579 | 250 | 15)*, ACM Digital Library* (1,040 | 250 | 24)*, AIS Electronic Library* (18,067 | 250 | 21)*, IEEE Xplore* (228 | 228 | 17) and *Google Scholar* (21,700 | 250 | 58). We used the following keywords to find potential hits for our literature review: *((“(conversational OR digital OR virtual) AND (agent OR assistant)”) OR “chatbot”) AND (“efficient” OR “efficiency” OR “success” OR “successful” OR “effective” OR “effectiveness” OR “productive” OR “productivity”)*. This initial literature research resulted in a total of *157 papers*. Regarding our research question, we only kept articles relating to productivity-related DFs of chatbots, based on their title, the provided keywords and their abstract (*remaining: 86*). Moreover, duplicates found within the

keyword-based searches were eliminated (remaining: 78). We also applied language (English), thematic focus (covering chatbot design) and quality (peer-reviewed) as inclusion criteria (*remaining: 64*). In a second step, we also conducted a forward as well as a backward search that resulted in another 8 articles. Therefore, a total of *72 papers* were selected for in-depth analysis in connection with our literature review.

3 Results: Chatbot Design Features and Categories

The systematic literature review uncovered a broad range of chatbot DFs potentially influencing productivity measures. To delimit terms, a DF describes a specific chatbot design element that provides a certain functionality. A DP on the other hand, is a concept in DSR and specifies prescriptive design knowledge according to a specific schema [9]. Studies that directly investigated the effect of certain DFs on productivity as defined in the introductory chapter could not be identified. Therefore, we selected DFs that respective literature argues to be beneficial to the chatbot interaction in more general terms - either quantitatively by positively affecting certain target variables or qualitatively, identified as favourable through user feedback. A list of DFs extracted from the identified papers was created. To reduce overlaps, similar or identical DFs were merged independently by two of the authors resulting in a consolidated list of DFs. During our analysis, we found that the features mainly fall into only a few categories. To define these categories, we followed the inductive category development approach as proposed by [13]. Thereby, two researchers involved within this DSR project independently derived the categories described in Tables 1 - 3. DFs mapped to category 1 are summarized in Table 1. These are DFs that augment the textual input paradigm of a chatbot by implementing features that are well-know from applications based on a graphical user interface (GUI) like, for example, quick responses (e.g., buttons). We argue that these elements, through speeding up the conversation [14, 15], can increase productivity by decreasing the required input (e.g., the amount of time, by reducing user's typing efforts). In the following, the sources provided in Tables 1 – 3 are to be seen as exemplary sources that cover respective DFs. For a full list of the 72 papers that the DF identification is based on, see [16].

Table 1. Design Features augmenting the textual input paradigm with GUI elements

DF #	Description	Source
DF1	Presenting quick responses (e.g., buttons, carousels) for faster navigation and less typing effort by the users.	[14, 15, 17]
DF2	Providing conversational context through separate GUI elements (e.g., buttons, text fields) that allow manipulation of the context.	[17, 18]
DF3	Providing dynamic quick responses taking into account the conversational context and showing the user possible next steps.	[6, 15, 19]
DF4	Using rich media elements like, for example, images, GIFs and videos in the chat widget to convey information in a visual way.	[15, 17, 20]
DF5	Allow users to edit previous messages by clicking and altering them, thus possibly updating the current intent and the conversational context.	[19, 20]
DF6	Decrease typing efforts by implementing a text prediction feature (auto-completion) providing word and intent suggestions.	[6, 17]
DF7	Provide a persistent menu with chatbot capabilities that is always available.	[15, 17]

The second category of DFs we could identify support the conversational process. They are intended to decrease conversational breakdowns that research shows can lead to lower user satisfaction and users abandoning the specific task [4, 17]. Although a direct effect of this category of DFs on productivity may not be as obvious as for DFs from the first category and secondary effects are to be expected, we assume they have an overall positive effect on productivity by reducing conversational breakdowns and therefore increasing the number of completed tasks. Taking DF11 as an example, a misunderstanding between user and chatbot can lead to repeated questions and answers without reaching a conclusion. DF11 can help to reformulate the input by providing specific words leading to the misunderstanding and therefore increase the chance of a successful task completion leading to increased productivity. On the other hand, user may find calls for reformulation disturbing and they could also slow down the conversation leading to lower productivity measures. Overall, although the DF can increase productivity, the exact effect depends on the specific context and implementation of the DF. In the further course of our research, we will address this line of questioning.

Table 2. Design Features supporting the conversational process

DF #	Description	Source
DF8	Providing information about the conversational context in chatbot responses to bring the user and the chatbot on a common ground of understanding. (e.g., tell the current intent and entities the bot has identified).	[17, 18]
DF9	Identify different signs indicating that the user might abandon the interaction soon (e.g., repeated reformulations) and provide alternatives (e.g., human handover).	[19, 21]
DF10	In case of misunderstandings, provide the user the chatbot capabilities and the intents that have the highest probability based on the user's input so far.	[4, 17, 21, 22]
DF11	In case of misunderstandings, reflect the user message back to the user and highlight signal words (e.g., words that could not be understood or that led to a certain classification of input) and ask for reformulation of the request.	[4, 21, 22]
DF12	When starting the conversation, introduce the chatbot as artificial and provide the user the chatbot capabilities to help manage user expectations.	[17, 21]

There is a broad literature base covering anthropomorphic chatbot design and social cues in chatbot conversations. Anthropomorphic chatbot design can elicit social responses in users, positively affecting dependent variables like service satisfaction [23, 24]. This could also have a positive effect on productivity if users are more inclined to use a chatbot. However, a humanized chatbot interacting in a social way (e.g., by engaging in small talk) [25], may also distract users from completing their task. Moreover, there has been research suggesting that DFs like buttons, that make a chatbot interface more like a common GUI application, negatively influence the social presence of the chatbot [14], which could also have an effect on productivity measures. Therefore, we also want to consider the effect of this category of DFs (Table 3) on productivity.

Table 3. Design Features personifying the chatbot

DF #	Description	Source
DF13	Using social cues (e.g., emojis, response delays, name etc.) to achieve a certain level of humanness.	[20, 26]
DF14	Choosing a consistent conversation style that supports the context of task orientation.	[17, 25]

4 Next Steps and Limitations

As outlined in the second chapter, the next step of our DSR project, will cover the Relevance Cycle [8]. In this cycle we focus on two goals. Firstly, we want to narrow the scope - so far chatbot productivity in general – to a specific application domain and a specified set of tasks to be solved. Secondly, we will prioritize and extend the list of DFs we identified based on the requirements of the specified application domain. To achieve both goals we will conduct interviews with practitioners in the field. Firstly, to ensure that our DSR project contributes to solving an important and relevant problem and secondly, to gather DF that have not been covered in literature yet and rank DFs in respect to their expected effects on productivity, based on practitioners' experience from actual chatbot projects. Concerning the Design Cycle, we will instantiate a productivity oriented chatbot for the specified domain, implementing the most promising DFs. Conducting experiments with a between-subject design (cf. [27]), we want to extract the effect of specific features on productivity. Eventually, through an iterative process of going through the three cycles, we want to formulate DPs [9] to provide prescriptive knowledge to designers implementing a chatbot in a specific domain where productivity is desired. These DPs can serve as the bases for future research about increasing chatbot productivity in other domains. The following limitations must be mentioned: Our definition of productivity introduced in chapter 1 includes input and output as determinants of productivity. However, there are further measures that can influence productivity and could be included in a more comprehensive productivity definition (e.g., cognitive load). Future research can identify which measure are connected to chatbot productivity and examine their effect. Another direction for future research is to study interaction effects of DFs from the different categories (see Tables 1 – 3) for a deeper understanding of their potential impact on productivity. The contribution of this short paper is the systematic literature review with the resulting overview of chatbot DFs having the potential to increase productivity.

References

1. Stucki, T., D'Onofrio, S., Portmann, E.: Chatbot – Der digitale Helfer im Unternehmen: Praxisbeispiele der Schweizerischen Post. HMD (2018)
2. Brandtzaeg, P.B., Følstad, A.: Why People Use Chatbots. INSCI (2017)
3. Janssen, A., Grützner, L., Breitner, M.: Why do Chatbots fail? A Critical Success Factors Analysis. ICIS (2021)
4. Ashktorab, Z., Jain, M., Liao, Q.V., Weisz, J.D.: Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. CHI (2019)
5. Diederich, S., Brendel, A.B., Morana, S., and Kolbe, L.: On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. JAIS 23(1), (2022)
6. Schmidhuber, J., Schlogl, S., Ploder, C.: Cognitive Load and Productivity Implications in Human-Chatbot Interaction. ICHMS (2021)
7. Fried, H.O., Schmidt, S.S., Lovell, C.A.K.: The Measurement of Productive Efficiency. Techniques and Applications. Oxford University Press (1993)

8. Hevner, A.: A three cycle view of design science research. Scandinavian Journal of Information Systems 19(2), (2007)
9. Gregor, S., Kruse, L., Seidel, S.: Research Perspectives: The Anatomy of a Design Principle. JAIS 21(6), (2020)
10. Gregor, S., Hevner, A.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37(2), (2013)
11. Webster, J., Watson, R.: Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly 26(2), (2002)
12. vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., and Cleven, A.: Reconstructing the giant: On the importance of rigour in documenting the literature search process. ECIS (2009)
13. Mayring, P.: Einführung in die qualititative Sozialforschung. Eine Anleitung zu qualitativem Denken. Beltz Verlag (2002)
14. Diederich, S., Brendel, A.B., Lichtenberg, S., Kolbe, L.M.: Design for fast request fulfillment or natural interaction? Insights from an experiment with a conversational agent. ECIS (2019)
15. Valério, F.A.M., Guimarães, T.G., Prates, R.O., Candello, H.: Chatbots Explain Themselves: Designers' Strategies for Conveying Chatbot Features to Users. JIS 9(3), (2018)
16. Full Reference List Available on GitHub:. https://github.com/chatbot-research/desrist22
17. Jain, M., Kumar, P., Kota, R., Patel, S.N.: Evaluating and Informing the Design of Chatbots. DIS (2018)
18. Jain, M., Kota, R., Kumar, P., Patel, S.N.: Convey: Exploring the Use of a Context View for Chatbots. CHI (2018)
19. Li, C.-H., Yeh, S.-F., Chang, T.-J., Tsai, M.-H., Chen, K., Chang, Y.-J.: A Conversation Analysis of Non-Progress and Coping Strategies with a Banking Task-Oriented Chatbot. CHI (2020)
20. Klopfenstein, L.C., Delpriori, S., Malatini, S., Bogliolo, A.: The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms. DIS (2017)
21. Toxtli, C., Monroy-Hernández, A., Cranshaw, J.: Understanding Chatbot-mediated Task Management. CHI (2018)
22. Benner, D., Elshan, E., Schöbel, S., Janson, A.: What do you mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents. ICIS (2021)
23. Diederich, S., Brendel, A.B., Kolbe, L.M.: Designing Anthropomorphic Enterprise Conversational Agents. BISE 62(3), (2020)
24. Gnewuch, U., Morana, S., Adam, M., and Maedche, A.: Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. ECIS (2018)
25. Chattaraman, V., Kwon, W.-S., Gilbert, J.E., Ross, K.: Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. CHB 90, (2019)
26. Feine, J., Gnewuch, U., Morana, S., Maedche, A.: A Taxonomy of Social Cues for Conversational Agents. IJHC 132, (2019)
27. Boudreau, M.-C., Gefen, D., Straub, D.W.: Validation in Information Systems Research: A State-of-the-Art Assessment. MIS Quarterly 25(1), (2001)

3. Schlussbetrachtung und Fazit

Dieses Kapitel gibt zunächst eine Zusammenfassung der Forschungsergebnisse. Die wesentlichen Beiträge der Dissertation für die Wissenschaft und Praxis werden dann in Kapitel 3.2 konsolidiert vorgestellt. Eine kritische Würdigung sowie ein Ausblick auf zukünftige Forschungsarbeiten in Kapitel 3.3 schließen die vorliegende Dissertation ab.

3.1. Zusammenfassung der Forschungsergebnisse

Das erste Themengebiet der Dissertation umfasst die Gestaltung von Interaktionen mit sozialen Medien. Dabei wurden zunächst die Einflüsse von Kontextfaktoren auf das in Tweets geäußerte Sentiment untersucht. **Veröffentlichung 1** hat die Einflüsse der beiden Kontextfaktoren „Kontinente“ und „mobile Betriebssysteme“ untersucht, welche beide in der einschlägigen Literatur noch nicht tiefergehend untersucht worden sind. Zur Beantwortung der zugrunde gelegten **Forschungsfrage 1** kam der „Social Media Analytics“ Forschungsansatz (vgl. Stieglitz et al., 2014; Stieglitz et al., 2018) zum Einsatz. Die Einflüsse wurden nicht nur für die beiden Kontextfaktoren einzeln, sondern auch für deren Kombination, ermittelt. Die Ergebnisse der Untersuchung zeigen, dass sich verglichen mit der Analyse der Kontinente als einzelnem Kontextfaktor, deren Reihenfolge durch die Kontextualisierung mit den mobilen Betriebssystemen an einigen Stellen verschiebt. Unternehmen sollten ihre Kommunikationsstrategie in sozialen Medien daher nicht nur an geographischen Gesichtspunkten allein ausrichten. Ebenso wichtig ist es zu berücksichtigen, dass die von ihren (potenziellen) Kunden genutzten Endgeräte und deren mobile Betriebssysteme eine Rolle für die Kundenresonanz auf Unternehmensposts spielen.

Marketingabteilungen können umfassend von den in der Masse von Social Media Posts enthaltenen Informationen und Zusammenhängen profitieren. Insbesondere die Textinhalte von Social Media Posts bieten wertvolle Einblicke in die ungefilterten Meinungen, Gedanken und Wünsche ihrer (potenziellen) Kunden. Um diese großen Mengen an textbasierten Daten effektiv und effizient analysieren zu können, ist die Anwendung automatisierter Analyseansätze unerlässlich. Eine bewährte Analysetechnik, welcher sich Unternehmen dabei zur Extraktion von kohärenten Inhalten und Themen bedienen können, ist das Topic Modelling.

Im Rahmen von **Veröffentlichung 2** wurde eine Gegenüberstellung von drei ausgewählten Topic Modelling Techniken vor dem Hintergrund der drei zentralen Anwendungsfälle von Topic Modelling im Marketing („Themenextraktion“, „Trendanalyse“ und „Themenstrukturierung“) durchgeführt. Dabei wurden mit „Build Time“, „Perplexity“, „Coherence“, „Log-Likelihood“, „Word Intrusion“ und „Topic Intrusion“ sechs Vergleichskriterien von Topic Modelling Techniken angewendet (**Forschungsfragen 2 und 3**). Wie der Vergleich der Topic Modelling Techniken zeigen konnte, stellt LDA nicht für jeden der drei marketingbezogenen Anwendungsfälle die zwangsläufig beste Analysetechnik dar. Vielmehr sollten Mitarbeiter in Marketingabteilungen vor dem Hintergrund ihrer jeweiligen Aufgabenstellungen und den damit zugrundeliegenden Anwendungsfällen für die Trendanalyse DMR und für die Themenstrukturierung PAM in Betracht ziehen.

Es stellt aber nicht nur die Auswahl einer geeigneten Topic Modelling Technik eine Herausforderung für Unternehmen dar. Insbesondere müssen Angestellte in Marketingabteilungen auch in die Lage versetzt werden, Topic Modelling und die dazugehörigen Funktionalitäten für eine Trendanalyse anzuwenden. Eine praktische Anwendung der Trendanalyse wird insbesondere durch den Einsatz von Softwaretools ermöglicht. Vor diesem Hintergrund wurde im Rahmen von **Veröffentlichung 3** ein Topic Modelling-basiertes Softwaretool (MANTRA) erstellt. Hierfür wurde die DSR-Methodik, angelehnt an die Phasen des DSR-Prozesses nach Peffers et al. (2007), angewendet (**Forschungsfrage 4**). Durch die Demonstration und Evaluation des entwickelten Softwaretools für den Anwendungsfall der Produktenwicklung konnte gezeigt werden, dass MANTRA in der Lage ist, relevante Trends zuverlässig zu identifizieren. Vor allen Dingen ist das Artefakt im Vergleich zu bestehenden Softwaretools in der Lage, externe Parameter wie z. B. Geolokation, Zeit und verschiedene Kundenmerkmale flexibel und in unterschiedlicher Granularität in die Trendanalyse einzubeziehen. Insgesamt können Unternehmen damit Trendthemen anhand verschiedener externer Parameter kontextualisieren und sowohl divergierende als auch detailliertere Trends identifizieren.

Eine weitere effektive Analysetechnik im Rahmen der Trendanalyse ist die aspektbasierte Sentimentanalyse. Basierend auf den Merkmalen von Produkten und Services können die zu diesen Aspekten geäußerten Sentimente ermittelt werden. Zur Beantwortung von **Forschungsfrage 5** wurden deshalb im Rahmen von **Veröffentlichung 4** zunächst anhand

der einschlägigen Forschungsliteratur Designanforderungen identifiziert und ebenfalls unter Anwendung des DSR-Ansatzes ein entsprechendes Softwaretool realisiert. Anhand der Untersuchung konnte verdeutlicht werden, dass die Möglichkeit, Vorwissen und Domänenwissen über bestehende Trends in die Trendanalyse einbeziehen zu können, eine essenzielle Funktionalität für ein solches aspektbasiertes Sentimentanalysetool darstellt.

Neben der Gestaltung von Interaktionen mit sozialen Medien beschäftigt sich die vorliegende Dissertation zudem mit der Untersuchung und Gestaltung von Interaktionen mit intelligenten, digitalen Kommunikationsagenten. Dabei wurden einerseits Forschungsarbeiten zu sog. *„Augmentative and Alternative Communication“* Systemen durchgeführt. In **Veröffentlichung 5** wurde ein systematischer Literature Review angelehnt an Cooper (1988), Cooper und Hedges (1994), Vom Brocke et al. (2015) sowie Webster und Watson (2002) durchgeführt, um zu identifizieren, welche Potenziale bislang durch KI-Ansätze in der einschlägigen Literatur zu AAC-Systemen zur Unterstützung von sprachbehinderten Menschen innerhalb verschiedener, zentraler Gesprächsphasen und Gesprächskontexte realisiert wurden **(Forschungsfrage 6)**. Während für informelle Gespräche bereits einige KI-Ansätze konzipiert und untersucht worden sind, ist auf Basis der Ergebnisse eine Forschungslücke für den formellen Gesprächskontext feststellbar. Die Literatur offenbart zusätzlich Defizite im Hinblick auf eine ganzheitliche Abdeckung von Gesprächen, weil KI-Ansätze bislang v. a. für die Gesprächsmitte, jedoch äußerst spärlich für den Gesprächsbeginn und Gesprächsabschluss konzipiert und untersucht worden sind.

Darüber hinausgehend wurden basierend auf den festgestellten Forschungslücken Implikationen für die zukünftige Forschung zu AAC-Systemen und den in ihnen eingesetzten KI-Ansätzen abgeleitet **(Forschungsfrage 7)**. Durch die Konzeption von auf neuronalen Netzen basierenden Ansätzen könnten z. B. Rechtschreib- und Grammatikfehler automatisiert korrigiert und sprachbehinderte Personen bei der Wahrung des formalen Verlaufs in Gesprächen besser unterstützt werden. Für eine ganzheitliche und damit eine noch umfassendere Kommunikationsunterstützung durch AAC-Systeme ist insbesondere auch die Abdeckung aller gesprächsrelevanten Kategorien in einem AAC-System unerlässlich.

Während mit diesem systematischen Literature Review verschiedene Möglichkeiten zur Schließung der Forschungslücken im Hinblick auf die Gesprächsphasen und Gesprächskontexte aufgezeigt werden konnten, wurde mit **Veröffentlichung 6** ein direkter Beitrag

aus der Dissertation zur Schließung der Forschungslücken bei den Gesprächskontexten geleistet. Dabei wurde ChatGPT als Generative KI-Technik herangezogen und spezifisch für den Einsatz im Rahmen von AAC-Systemen angepasst, weil diese Technik in der Lage ist, Texte zu verschiedenen alltäglichen Gesprächen in verschiedenen Stilen (u. a. formell und informell) bereitzustellen. Zur Erzeugung dieses KI-basierten Ansatzes, welcher in einen ebenfalls entwickelten Prototyp eines AAC-Systems integriert wurde, ist wiederum der DSR-Ansatz verfolgt worden (**Forschungsfrage 8**). Durch die Anwendung des KI-basierten Ansatzes auf 61 Kommunikationsaktivitäten, welche im Rahmen eines Restaurantbesuchs mit einer Begleitung häufig ausgeführt werden, konnte gezeigt werden, dass der Ansatz in der Lage ist, formelle und informelle Gesprächsbeiträge zuverlässig zu generieren. Das resultierende Artefakt eröffnet sprachbehinderten Personen damit die Möglichkeit, vollständige und korrekte Gesprächsbeiträge in alltäglichen Gesprächssituationen mit einem situationsspezifisch adäquaten Artikulationsstil anzuwenden.

Neben den AAC-Systemen wurden weitere Forschungsarbeiten zu Chatbots als einem weiteren Beispiel für intelligente, digitale Kommunikationsagenten durchgeführt. Der Einsatz von Chatbots in der Kundenkommunikation verspricht für Unternehmen verschiedene Potenziale, wie z. B. die Entlastung von Mitarbeitern im Kundenservice oder die Reduktion von Kosten für die Abwicklung von Kundenanfragen. Die Auswahl einer geeigneten Chatbot-Lösung war aber erschwert, da kein Marktüberblick über Chatbot-Lösungen zur Unterstützung der Kundenkommunikation im B2C-Kontext existiert hat.

Deshalb wurde in **Veröffentlichung 7** ein Vergleich von am Markt verfügbaren Chatbot-Lösungen durchgeführt. Anhand einer systematischen Literaturrecherche wurden dafür zunächst Evaluations- und Vergleichskriterien identifiziert (**Forschungsfrage 9**). Darauf basierend wurde ein Evaluationskatalog entworfen, mit dem bei Chatbot-Anbietern erhoben wurde, wie gut und umfassend die identifizierten Funktionalitäten und Anforderungen von den am Markt verfügbaren Chatbot-Lösungen zur Unterstützung der Kundenkommunikation im B2C-Kontext abgedeckt werden (**Forschungsfrage 10**).

Wie die Ergebnisse der Marktstudie zeigen konnten, wird die Qualität der Kommunikation gut von den Chatbot-Lösungen umgesetzt. Die Reaktion des Chatbots auf unangemessene Eingaben der Nutzer, die Personalisierung des Chatbots, IT-Sicherheit, Training und Anforderungen an das Chatbot-GUI sowie die Nutzbarkeit des Chatbots auf mobilen

Endgeräten decken die Anbieter ebenfalls gut und umfassend ab. Die Anforderungen und Funktionalitäten hinsichtlich der Sprachverarbeitung und der Implementierung des Chatbots werden im Vergleich dazu am schlechtesten erfüllt und bieten damit umfangreiche Potenziale für Verbesserungen der am Markt angebotenen Chatbot-Lösungen.

Chatbots können darüber hinausgehend auch die Erfüllung von Aufgaben im Rahmen von innerbetrieblichen Prozessen unterstützen. Ein wesentliches Gestaltungsziel ist hierbei die Förderung der Produktivität von Mitarbeitern. Vor diesem Hintergrund wurde in **Veröffentlichung 8** ein systematischer Literature Review, angelehnt an die Methodik nach Webster und Watson (2002), durchgeführt, um Gestaltungsmerkmale zu identifizieren, die bei einem Chatbot im Rahmen eines Anwendungsfalls umgesetzt sein sollten, in dem die Produktivität seiner Nutzer eine wichtige Rolle spielt (vgl. **Forschungsfrage 11**).

Neben Gestaltungsmerkmalen, welche den Input von Text reduzieren (z. B. Buttons, Rich Response Elemente oder automatisierte Textvervollständigung), spielen auch Funktionalitäten eine Rolle, die den Chatbot menschlicher erscheinen lassen, weil menschlich erscheinende Chatbots im Vergleich zu maschinell erscheinenden Chatbots bei Nutzern weniger Widersprüche auslösen. Die Ergebnisse deuten zusätzlich auf gegenläufige Effekte bei den drei identifizierten Kategorien hin. So könnten menschliche Züge eines Chatbots einerseits die Nutzungsintention und damit die Wahrscheinlichkeit erhöhen, dass der Chatbot für die Steigerung der Produktivität im Rahmen von operativen Prozessen im Unternehmen von Mitarbeitern auch tatsächlich genutzt wird. Andererseits verlangsamen bestimmte menschliche Eigenschaften eines Chatbots (z. B. ausschließlich textbasierte Kommunikation) im Vergleich zu einer rein Button-basierten Interaktion die operative Aufgabenerfüllung. Die Ergebnisse des Beitrags stellen somit einen ersten wichtigen Schritt dar, um sowohl gegenläufige als auch synergetische Effekte der Gestaltungsmerkmale für die Produktivitätssteigerung durch Chatbots tiefergehend zu untersuchen.

3.2. Beiträge für Wissenschaft und Praxis

Durch die Forschungsergebnisse der vorliegenden Dissertation werden verschiedene Beiträge sowohl für die Wissenschaft als auch für die Praxis geleistet. Nachfolgend werden in Kapitel 3.2.1 zunächst die Beiträge für die Wissenschaft aufgezeigt. Anschließend wird dann in Kapitel 3.2.2 auf die Beiträge für die Praxis näher eingegangen.

3.2.1. Beiträge der Dissertation für die Wissenschaft

Veröffentlichung 1 leistet durch die Übersicht über die in der Literatur behandelten Kontextfaktoren und deren Effekte auf die in Social Media Posts geäußerten Sentimente einen ersten wichtigen wissenschaftlichen Beitrag. Durch diese Übersicht konnte weiterhin gezeigt werden, dass einerseits die Effekte von Kombinationen von Kontextfaktoren und andererseits die beiden Kontextfaktoren Kontinente und mobile Betriebssysteme tiefergehender Untersuchungen bedürfen. Forscher können auf diesen Ergebnissen aufsetzen und weitere Kombinationen von Kontextfaktoren sowie deren Einflüsse auf das Sentiment in Social Media Posts untersuchen. Um einen ersten Beitrag zur Schließung der festgestellten Forschungslücken zu leisten, wurden im Rahmen der Dissertation die Effekte von Kontinenten und mobilen Betriebssystemen, sowohl einzeln als auch in Kombination, unter Anwendung des Social Media Analytics Forschungsansatz identifiziert.

Die drei Topic Modelling Techniken LDA, DMR und PAM wurden im Rahmen von **Veröffentlichung 2** mithilfe eines umfassenden Realweltdatensatzes hinsichtlich sechs verschiedener Vergleichsmetriken evaluiert. Darauf basierend konnten sowohl Stärken (z. B. DMR eignet sich besonders für den Anwendungsfall der Trendanalyse) als auch Schwächen (z. B. für den Anwendungsfall der Themenextraktion ist DMR zu empfehlen, wenn eine große Anzahl an Themen extrahiert werden soll, ansonsten ist LDA besser geeignet) abgeleitet werden. Für den Vergleich wurde der GQM-Ansatz verfolgt, welcher ein systematisches Forschungsvorgehen darstellt und somit reproduzierbare Ergebnisse gewährleistet. Auf dieser Grundlage ergeben sich für die einschlägige Forschung, insbesondere an der Schnittstelle des maschinellen Lernens und des Marketings, Möglichkeiten und Ansatzpunkte zur Weiterentwicklung der Topic Modelling Techniken (z. B. Berücksichtigung einer schnelleren Erstellungszeit für PAM bei $k > 50$, Gewährleistung einer hohen Kohärenz für verschiedene Anzahlen an Themen bei PAM).

Weiterhin wurde das Thema Topic Modelling tiefergehend im Hinblick auf den marketingbezogenen Anwendungsfall der Trendanalyse und die daran geknüpften Anforderungen untersucht (siehe **Veröffentlichungen 3 und 4**). So konnte ein Überblick über den aktuellen Stand der Forschung über das Thema Trendanalyse unter Anwendung von Topic Modelling im Marketing herausgearbeitet werden. Darüber hinaus leisten die Unter-

suchungen u. a. wertvolle Beiträge für die Social Media Theorie, indem zielführende Datenanalysemethoden bei der Konzeption und Entwicklung automatisierter Trendanalysetools kombiniert und auf deren spezifischen Anforderungen angepasst wurden. Da die Social Media Theorie in der praktischen Anwendung bei der automatisierten Analyse von Social Media Daten immer noch vor großen Herausforderungen steht, konnten mithilfe der Untersuchung erste Ansatzpunkte dahingehend aufgezeigt werden, wie z. B. Anforderungen in der Trendanalyse konkret implementiert werden können. Im Rahmen von Veröffentlichung 3 wurden weiterhin Beiträge zur Innovationstheorie geleistet. Durch das entwickelte Trendanalysetool MANTRA und die in der Demonstration und Evaluation erzielten Ergebnisse ließen sich Adaptionen für das Stage-Gate-Modell ableiten. MANTRA kann hierbei externe Parameter (z. B. Zeit, Geolokation, personenbezogene Merkmale) flexibel und in unterschiedlichen Granularitäten in die Trendanalyse einbeziehen. Durch die unmittelbare Integration der Trendinformationen in die Analysen können die Bewertungen an den verschiedenen Stellen des Innovationsprozesses gelockert werden, weil der Abgleich sowohl mit den externen Parametern als auch mit den Kundenanforderungen als fester Bestandteil der automatisierten Trendanalyse ständig stattfindet.

Zusätzlich zu den Beiträgen zu den Kernel Theorien eröffnen MANTRA sowie das Trendanalysetool mit aspektbasierter Sentimentanalyse verschiedene Potenziale im Rahmen der Designtheorie. Bei der Gestaltung der beiden Trendanalysetools wurden basierend auf Designanforderungen und einschlägigen Theorien Designprinzipien abgeleitet. Durch die technische Realisierung sämtlicher Designanforderungen sowie durch die anschließende Demonstration der Artefakte wurde eine implizite empirische Fundierung der Designprinzipien erreicht. Die aufgestellten Designprinzipien erfassen designbezogenes Wissen und können daher die Entwicklung vergleichbarer Artefakte unterstützen. Für die Konzeption und Entwicklung neuer Trendanalysetools (u. a. in verwandten Anwendungsbereichen, wie z. B. im Nachhaltigkeitsmanagement, wo Trendanalysen basierend auf Umweltregulierungen und Verbrauchergewohnheiten soziale und ökologische Trends aufdecken sollen) könnten diese Designprinzipien übertragen und herangezogen werden.

Neben den Ergebnissen der Dissertation zur Gestaltung der Interaktionen mit sozialen Medien liefert auch der zweite Teil der Dissertation – die Untersuchung und Gestaltung

von Interaktionen mit intelligenten, digitalen Kommunikationsagenten – verschiedene theoretische Beiträge für die Wissenschaft.

So wurden umfangreiche Forschungsarbeiten zu AAC-Systemen durchgeführt. Im Rahmen von **Veröffentlichung 5** wurde ein systematischer Literature Review ausgeführt. Dadurch konnte eine Übersicht darüber erlangt werden, welche Potenziale bislang durch KI-Ansätze in der einschlägigen Literatur für AAC-Systeme zur Unterstützung von sprachbehinderten Menschen für verschiedene Gesprächsphasen und Gesprächskontexte realisiert wurden. Auf Basis dieser Übersicht konnten verschiedene Forschungslücken identifiziert werden. Dabei konnte gezeigt werden, dass insbesondere das Aufrechterhalten des formalen Verlaufs eines Gesprächs und die Anpassung gegenüber den Gesprächspartnern im Vergleich zu den anderen Kategorien unzureichend erfüllt werden. Während für informelle Gespräche zahlreiche KI-Ansätze konzipiert und untersucht wurden, sind auf Basis der Ergebnisse v. a. für den formellen Gesprächskontext signifikante Forschungslücken feststellbar. Die Literatur offenbart zusätzliche Defizite im Hinblick auf eine ganzheitliche Abdeckung von Gesprächen, weil insbesondere für die Gesprächsmitte, jedoch vergleichsweise wenige KI-Ansätze für den Gesprächsbeginn und Gesprächsschluss identifiziert werden konnten. Die dazu abgeleiteten Forschungsmöglichkeiten (u. a. Erforschung von künstlichen neuronalen Netzen zur Wahrung des formalen Verlaufs innerhalb von Gesprächen) stellen erste Ausgangspunkte zur Schließung der Forschungslücken und damit einen weiteren, wertvollen Beitrag für die Wissenschaft dar.

Veröffentlichung 6, welche ebenfalls dem Themenbereich der AAC-Systeme zuzuordnen ist, liefert weitere, wichtige Beiträge für die Wissenschaft. Basierend auf der einschlägigen Forschungsliteratur sowie auf Basis der Cognitive Load Theorie (Sweller, 1988) und den Gesprächsmaximen nach Grice (1975) wurden Designanforderungen für einen auf Generativer KI-basierten Kompositionsansatz für die Generierung, sowohl vollständiger als auch semantisch kohärenter Gesprächsbeiträge, abgeleitet. Durch die technische Realisierung sämtlicher Designanforderungen sowie durch die Demonstration und Evaluation des entwickelten Artefakts konnte eine implizite empirische Fundierung der Designprinzipien erreicht werden, welche wertvolle Beiträge zur einschlägigen Designtheorie darstellen. Durch das Softwareartefakt ergeben sich ebenfalls erste Beiträge zu den beiden Kernel Theorien. Im Hinblick auf die Gesprächsmaxime nach Grice (1975)

könnte die durch den Generativen KI-basierten Kompositionsansatz eröffnete und gesteigerte Ausdrucksfähigkeit dazu führen, dass sprachbehinderte Personen (absichtlich) mit nicht sinnvollen oder nicht korrekten Gesprächsbeiträgen gegen die Relevanz- oder die Qualitätsmaxime verstoßen. Um basierend auf dieser Überlegung im Hinblick auf AAC-Systeme tiefergehende Beiträge zu den Gesprächmaximen nach Grice (1975) leisten zu können, gilt es den konzipierten Generativen KI-basierten Kompositionsansatz mit sprachbehinderten Menschen in empirischen Settings anzuwenden und zu validieren.

Auch für die Gestaltung von Chatbots als die zweite untersuchte Form von intelligenten, digitalen Kommunikationsagenten konnten durch die Dissertation verschiedene, wertvolle Beiträge für die Wissenschaft erbracht werden. **Veröffentlichung 7** leistet einerseits einen wertvollen Beitrag durch den Marktüberblick, der für die Wissenschaft einschlägige Forschungslücken aufzeigt. Um z. B. die Potenziale von Chatbots auch für kleine und mittelständische Unternehmen zu eröffnen, könnte sich die einschlägige Forschung auf die Entwicklung von Wörterbüchern konzentrieren, die Dialekt und Umgangssprache berücksichtigen und für bestimmte geographische Regionen bzw. Nischen angepasst sind. Auf diese Weise kann die Qualität der Kommunikation auf der Kundenseite erhöht und eine unnötige Übergabe von einfach zu lösenden Anfragen an menschliche Agenten reduziert werden. Darüber hinaus ist die Schaffung von Mechanismen zur Unterstützung eines automatisierten Lernens in Netzwerken aus Mensch und Maschine (vgl. Folstad und Brandtzaeg, 2017) ein vielversprechendes Forschungsfeld. Schließlich liefern sowohl **Veröffentlichung 7** als auch **Veröffentlichung 8** einen wichtigen Ausgangs- und Anhaltspunkt dafür, wie Chatbot-Prototypen für empirische Untersuchungen und Between-Subject Studien (vgl. Boudreau et al., 2001) im Rahmen der Human-Computer Interaktionsforschung gestaltet werden können. Einschlägige Forschungsarbeiten müssen diese Designanforderungen daher nicht noch erst erheben bzw. konsolidieren, sondern können sich direkt der Designanforderungen und Funktionalitäten aus den beiden Veröffentlichungen bedienen und ihre Forschungsprototypen darauf basierend implementieren.

3.2.2. Beiträge der Dissertation für die Praxis

Neben den vielfältigen theoretischen Beiträgen ergeben sich aus der vorliegenden Dissertation ebenso wertvolle Beiträge für die Praxis. Unternehmen können sich der Ergebnisse

von **Veröffentlichung 1** bedienen, um basierend auf den verschiedenen Kontinenten und den von ihrer Zielgruppe hauptsächlich genutzten mobilen Endgeräten (z. B. Apple oder Android) ihre Social Media Kommunikation auszurichten. Dies ermöglicht es ihnen, die Auslösung negativer Stimmungen und Reaktionen in Social Media Posts durch (potenzielle) Kunden eindämmen oder gar verhindern zu können.

Weiterhin werden Unternehmen durch die Ergebnisse aus der Dissertation in die Lage versetzt, für zentrale marketingbezogene Anwendungsfälle eine passende Topic Modelling Technik auszuwählen (**Veröffentlichung 2**). So können Entscheidungsträger im Bereich Marketing ihr konkretes Anliegen in eines der drei Anwendungsgebiete (Themenextraktion, Trendanalyse oder Themenstrukturierung) einordnen und darauf basierend eine Empfehlung für die Auswahl einer bestmöglich geeignetsten Technik ableiten. Die Anwendung von Topic Modelling Techniken zur Extraktion von Themen aus textuellen Social Media Inhalten erfordert von Anwendern zudem fortgeschrittene und technikbezogene Kenntnisse. Ein erheblicher Mehrwert für die Praxis ergibt sich deshalb durch die beiden auf Topic Modelling basierenden Trendanalysetools, welche im Rahmen der Dissertation entstanden sind. Im Vergleich zu bestehenden Softwaretools kann das Trendanalysetool MANTRA (**Veröffentlichung 3**) externe Parameter, die für die Anwendungsfälle Produktentwicklung, Kundenverhaltensanalyse und Marktbeobachtung essenziell wichtig sind, in verschiedenen Granularitäten flexibel in die Trendanalyse integrieren. Infolgedessen können Trends kontextualisiert unter Berücksichtigung von Zeit, Geolokationen und kundenbezogenen Merkmalen und somit zielgerichteter auf die konkreten Fragestellungen von Marketingabteilungen in Unternehmen extrahiert werden. Die Unternehmen müssen somit im Vergleich zu den am Markt weit verbreiteten Social Listening Tools keine Vorkenntnisse über einen bestimmten Trend haben, weil das Setzen von Startwörtern bei MANTRA nicht zwingend erforderlich ist. Diese Social Listening Tools ermitteln Trends basierend auf der Häufigkeit bestimmter Begriffe, was im Vergleich zu MANTRA einen entscheidenden Nachteil bedeutet. Trends, deren Bestandteile nicht unbedingt im Voraus bekannt sind, können nicht auf diese Weise identifiziert werden. Trendthemen deuten sich insbesondere durch Kombinationen mehrerer Begriffe an, die nicht ungewöhnlich aktuell oder häufig zusammen erwähnt werden (Khan et al., 2021; Reisenbichler und Reutterer, 2019). MANTRA ermittelt Trends basierend auf den Kook-

kurrenzen von Wörtern in textbasierten Social Media Daten und ermöglicht es Unternehmen damit, auch bislang unbekannte und unbestimmte Trends explorativ ohne Vorwissen und automatisiert zu identifizieren. Dabei ermöglicht es das Softwaretool jedem Unternehmen, Trendanalysen an seine individuellen Gegebenheiten anzupassen, weil in MANTRA verschiedene Machine Learning-Ansätze kombiniert wurden und die von diesem Softwaretool angebotenen Analysemöglichkeiten modular aufgebaut sind.

Diese positiven Aspekte lassen sich auch bei dem zweiten Trendanalysetool (**Veröffentlichung 4**) erkennen, das Unternehmen die Ausführung einer aspektbasierten Sentimentanalyse für den Anwendungsfall der Produktentwicklung ermöglicht. Insbesondere große Unternehmen mit mehreren Niederlassungen an verschiedenen Orten können davon profitieren, dass feingranulare Schwankungen des Sentiments für bestimmte Zeiträume und Geolokationen dargestellt werden können. Dadurch können standortspezifische Stimmungsänderungen aufgezeigt und zielgerichtete Aktionen zur Anpassung der Produkte und Services an die Kundenwünsche durch ein Unternehmen initiiert werden.

Darüber hinaus resultieren auch vielfältige praktische Beiträge aus dem zweiten Themenbereich der Dissertation – der Gestaltung von Interaktionen mit intelligenten, digitalen Kommunikationsagenten. Die aus dem systematischen Literature Review aus **Veröffentlichung 5** hervorgehende Übersicht zu KI-Ansätzen und KI-Potenzialen kann insbesondere für Anbieter von AAC-Systemen eine entscheidende Hilfestellung darstellen. Diese können darauf basierend prüfen, ob z. B. eine KI-basierte Unterstützung in formellen Gesprächskontexten für die intendierten Zielgruppen ihres AAC-Systems wirklich sinnvoll ist. Eine Einbindung in formelle Gesprächssituationen ist nur dann zielführend, wenn die kognitiven Beeinträchtigungen der sprachbehinderten Personen eher geringfügiger Natur sind. Vielmehr könnten gravierende kognitive Einschränkungen dazu führen, dass eine sprachbehinderte Person bei der Kommunikation in einem formellen Gesprächskontext überfordert ist. Sollte eine Unterstützung formeller Gesprächskontexte für die jeweilige Zielgruppe aber als wichtig und sinnvoll erscheinen, so könnten sich die Anbieter der in den Ergebnissen konsolidierten KI-Ansätze bedienen, um entsprechende Funktionalitäten zur Unterstützung in den verschiedenen Gesprächsphasen und Gesprächskontexten für

ihre AAC-Systeme zu realisieren. Auf diese Weise fungiert die Übersicht aus Veröffentlichung 5 als ein wichtiges Instrument, das die Ableitung von vielfältigen, zielgruppenspezifischen Weiterentwicklungsmöglichkeiten für AAC-Systeme in der Praxis eröffnet.

Der entwickelte KI-basierte Kompositionsansatz (**Veröffentlichung 6**) zeigt den Anbietern ebenfalls auf, wie Generative KI im Rahmen von AAC-Systemen zielführend eingesetzt werden kann. Eine Integration von z. B. ChatGPT ohne weitere Adaptionen reicht nicht aus und es sind für eine zielführende und umfassende Lösung weitere wichtige Funktionalitäten umzusetzen. Diese Adaptionen und die Realisierung eines ganzheitlichen Generativen KI-basierten Kompositionsansatzes wurden allesamt technisch im Rahmen eines funktionsfähigen Prototyps eines AAC-Systems realisiert. Durch das dabei generierte design- und umsetzungsbezogene Wissen wird den Anbietern von AAC-Systemen aufgezeigt, wie die notwendigen Adaptionen bei der Einbindung von Generativer KI in ein AAC-System und die verschiedenen daran geknüpften Designanforderungen konkret implementiert werden können.

Neben den Anbietern von AAC-Systemen profitieren aber insbesondere die sprachbehinderten Personen. Das generierte Artefakt verhilft sprachbehinderten Personen zu einer gesteigerten Ausdrucksfähigkeit. Bestehende AAC-Systeme schlagen für formelle Situationen oftmals Gesprächsbeiträge in einem unpassenden Artikulationsstil vor. Als Konsequenz könnten deshalb Gesprächspartner die Teilnahme an alltäglichen Gesprächssituationen mit sprachbehinderten Personen vermeiden. Auch die sprachbehinderten Personen selbst könnten Gespräche vermeiden, weil die von ihrem AAC-System unterbreiteten Gesprächsbeiträge nicht ihre Persönlichkeit und keinen individuellen, situationsadäquaten Artikulationsstil widerspiegeln. Das generierte Artefakt ist im Vergleich zu den am Markt verfügbaren AAC-Systemen in der Lage, vollständige und korrekte Gesprächsbeiträge, sowohl für formelle als auch für informelle Gesprächskontexte, zu generieren. Auf diese Weise leistet die Dissertation einen wichtigen sozialpolitischen Beitrag, um die Kommunikationsbarrieren von sprachbehinderten Personen abzubauen und diese Personengruppe besser in alltägliche Gesprächssituationen integrieren zu können.

Praktische Beiträge für die Gestaltung von intelligenten, digitalen Kommunikationsagenten im Unternehmenskontext ergeben sich durch die **Veröffentlichungen 7 und 8**. Auf

der einen Seite erhalten Unternehmen durch den literaturbasierten Überblick einen Anhaltspunkt dafür, wie eine Chatbot-Lösung für operative Prozesse in der Kundenkommunikation (vgl. Veröffentlichung 7) und wie zur Förderung der Produktivität ihrer Mitarbeiter (vgl. Veröffentlichung 8) gestaltet werden könnte. Im Rahmen von Veröffentlichung 7 wurde über die Konsolidierung von entsprechenden Anforderungen und Funktionalitäten hinausgehend eine Marktstudie in Form einer Onlineumfrage durchgeführt. Unternehmen, welche planen, einen Chatbot einzuführen, können ihre spezifischen Anforderungen sorgfältig konsolidieren und priorisieren, um basierend auf den Ergebnissen der Marktübersicht diejenige Chatbot-Lösung auszuwählen, welche ihre Erwartungen und Anforderungen am besten erfüllen kann. Auf der einen Seite unterstützen die Ergebnisse diesen Auswahlprozess, weil die Abdeckung bestimmter, für Unternehmen relevanter Funktionalitäten, durch einzelne Chatbot-Anbieter besser erfolgt als durch andere. Auf der anderen Seite erhalten die Anbieter von Chatbot-Lösungen wertvolle Hinweise darauf, welche Funktionalitäten dabei helfen können, um sich gezielt von anderen Chatbot-Lösungen am Markt abzugrenzen. Insbesondere der Umgang mit Dialekt und Umgangssprache, die Berücksichtigung von branchenspezifischen IT-Sicherheitsstandards sowie die Unterstützung bei der Schulung und Integration der Chatbot-Lösung in die Unternehmensprozesse erscheinen in diesem Zusammenhang als vielversprechende Differenzierungsmöglichkeiten.

3.3. Kritische Würdigung und Ausblick auf weitere Forschungsfelder

Die in der Dissertation vorgestellten Forschungsergebnisse sollen nachfolgend kritisch reflektiert werden. Auf dieser Grundlage ergeben sich innerhalb der jeweiligen Themenfelder verschiedene Möglichkeiten für zukünftige Forschungsarbeiten (siehe Kapitel 3.3.1). Weitere Synergien für zukünftige Forschungsansätze ergeben sich aus der Dissertation durch die themenübergreifende Betrachtung und Integration der Forschungsergebnisse aus den beiden behandelten Themenfeldern (siehe Kapitel 3.3.2).

3.3.1. Limitationen und weitere Forschungsarbeiten innerhalb der Themenfelder

Nachfolgend sollen zunächst die Limitationen und die Möglichkeiten für daran anknüpfende Forschungsarbeiten für die beiden in der Dissertation behandelten Themenbereiche – Social Media und digitale, intelligente Kommunikationsagenten – erläutert werden.

Veröffentlichungen	Limitationen	Zukünftige Forschungsmöglichkeiten
1	Ermittlung der Sentimentwerte erfolgte ausschließlich mit der VADER Sentimentanalysetechnik nach Hutto und Gilbert (2014).	Validierung der ermittelten Effekte der Kontextfaktoren könnte mit weiteren Sentimentanalysetechniken erfolgen.
	Analyse von Tweets, welche ausschließlich aus dem Jahr 2019 stammen.	Validierung der Einflüsse von Kontextfaktoren auf die Stimmungen könnte mit Tweets aus weiteren Jahren (vor und nach 2019) erfolgen.
	Ermittlung der Einflüsse der Kontextfaktoren erfolgte ausschließlich für die Sentimente der Tweets.	Ermittlung der Einflüsse von Kontextfaktoren könnte auch auf die Themen und Inhalte von Tweets mit bestimmten Sentimenten ausgeweitet werden.
2	Betrachtung von ausschließlich drei ausgewählten etablierten Topic Modelling Techniken (LDA, DMR und PAM).	Der Methodenvergleich könnte um weitere Topic Modelling Techniken erweitert werden.
	Subjektivität bei der Identifikation der Anforderungen für die drei marketingbezogenen Anwendungsfälle aus der Literatur.	Anforderungserhebung könnte auf Basis von weiteren Quellen (u. a. Befragung im intendierten Nutzerkreis) und unter Einbezug von Forschern aus anderen Themengebieten erfolgen.
3 + 4	Subjektivität bei der Identifizierung der Anforderungen für die implementierten Social Media Analytics Tools aus der Literatur.	Anforderungserhebung könnte auf Basis von weiteren Quellen (u. a. Befragung im intendierten Nutzerkreis) und unter Einbezug von Forschern aus anderen Themengebieten erfolgen.
	Durchführung einer lediglich technischen Evaluation der beiden Trendanalysetools.	Durchführung einer über einen längeren Zeitraum angelegten Evaluation mit Mitarbeitern aus Marketingabteilungen von Unternehmen.
5	Ergebnisse des systematischen Literature Reviews stellen lediglich Momentaufnahme dar.	Aktualisierung der Ergebnisse durch eine Wiederholung des Literature Reviews nach einem gewissen zeitlichen Versatz.
	Identifizierte KI-Potenziale stammen ausschließlich aus der einschlägigen wissenschaftlichen Literatur.	Durchführung einer Marktstudie zum Abgleich der KI-Potenziale mit den Möglichkeiten von am Markt angebotenen AAC-Systemen.
6	Durchführung einer lediglich technischen Evaluation des Generativen KI-basierten Kompositionsansatzes.	Durchführung einer Evaluation des erstellten Artefakts mit sprachbehinderten Personen im Rahmen von Realweltsituationen.
7	Ergebnisse der Marktstudie zu Chatbots zur Unterstützung der Kundeninteraktion stellen lediglich Momentaufnahme dar.	Wiederholung der Marktstudie mit einem gewissen zeitlichen Versatz und mit einer größeren Anzahl an Chatbot-Anbietern.
8	Ausschließliche Identifikation von Designanforderungen, welche eine Produktivitätssteigerung bei den Nutzern eines Chatbots in innerbetrieblichen Prozessen eines Unternehmens fördern.	Identifikation synergetischer und gegenläufiger Effekte im Hinblick auf die Produktivitätssteigerung durch entsprechend gestaltete Chatbots.
		Ableitung von Design- und Implementierungsprinzipien für Chatbots zur Steigerung der Produktivität von Mitarbeitern in innerbetrieblichen Prozessen.
		Instanziierung und Evaluation eines produktivitätsorientierten Chatbots für eine ausgewählte Domäne mit den vielversprechendsten Designanforderungen.

Tabelle 11: Limitationen und zukünftige Forschungsmöglichkeiten

Im Hinblick auf **Veröffentlichung 1** ergeben sich zwei Limitationen. Erstens wurden die Sentimentwerte für die Identifikation der Einflüsse der Kontextfaktoren ausschließlich mit dem Python-Package „VADER“ von Hutto und Gilbert (2014) ermittelt. Es handelt sich dabei um einen Ansatz, der speziell für den spezifischen Kontext der Sentimentanalyse von kurzen, textuellen Social Media Posts wie Tweets entwickelt wurde. Nichtsdestotrotz könnte die Anwendung anderer Ansätze für die Sentimentanalyse zu geringfügig anderen Beobachtungen bei den Effekten der Kontextfaktoren führen. Zweitens wurden in der Untersuchung ausschließlich die Sentimente von Tweets aus dem Jahr 2019 untersucht. Zukünftige Forschungsarbeiten könnten daher eine Ausweitung des Analysezeitraums vornehmen, um die Robustheit der Ergebnisse zu überprüfen. Darüber hinausgehend könnte der konzipierte Untersuchungsansatz um eine für kurze Texte wie Tweets geeignete Topic Modelling Technik wie dem „Biterm Topic Model“ (vgl. Yan et al., 2013) erweitert werden. Dadurch könnte zusätzlich untersucht werden, inwieweit Kontextfaktoren auch die Themen und Inhalte von Tweets mit bestimmten Sentimenten beeinflussen.

Bei **Veröffentlichung 2** ist die Auswahl der verglichenen Topic Modelling Techniken eine Limitation. Mit LDA, DMR und PAM wurden drei ausgewählte Topic Modelling Techniken in den Vergleich einbezogen. Wenngleich damit die drei wesentlichen Kategorien von Erweiterungen zur LDA Technik abgedeckt werden konnten (vgl. Liu et al., 2016), könnten durch das Einbeziehen noch weiterer Techniken noch präzisere Empfehlungen für den Einsatz von Topic Modelling Techniken für die drei identifizierten Anwendungsfälle im Marketing abgeleitet werden. Zudem ist denkbar, dass Forscher aus anderen Disziplinen weitere Anforderungen identifizieren könnten, die bei den drei marketingbezogenen Anwendungsfällen bislang nicht berücksichtigt wurden.

In den **Veröffentlichungen 3 und 4** wurden jeweils Softwareartefakte für marketingbezogene Anwendungsfälle zur Durchführung von Trendanalysen entwickelt. Eine Limitation bei beiden Beiträgen ist wiederum, dass durch die Analyse weiterer Literatur und durch Forscher aus anderen Forschungsgebieten weitere Anforderungen identifiziert werden könnten. Weiterhin wurden beide Softwareartefakte ausschließlich technisch evaluiert. Daran anschließend könnte die Benutzerfreundlichkeit der Softwareartefakte mithilfe einer SUMI-Studie (vgl. Kirakowski und Corbett, 1993) evaluiert und ggf. verbes-

sert werden. Andererseits ist das *„Grounding“* der abgeleiteten Design- und Implementierungsprinzipien noch ausbaufähig. Durch eine Evaluation mit Mitarbeitern in Marketingabteilungen könnten die Softwareartefakte tiefergehend durch den intendierten Nutzerkreis getestet und dadurch die abgeleiteten Design- und Implementierungsprinzipien validiert werden. Insbesondere eine Evaluation über einen längeren Zeitraum mit mehreren Kooperationspartnern könnte zur Weiterentwicklung und daher zu einer höheren Verlässlichkeit des abgeleiteten gestaltungs- und umsetzungsorientierten Wissens führen.

Bei dem in **Veröffentlichung 5** durchgeführten systematischen Literature Review stellt der identifizierte Stand der Literatur eine Momentaufnahme dar. Trotz dieser umfangreich angelegten Literaturrecherche und einer größtmöglichen Sorgfalt bei der Durchführung sämtlicher Schritte in der Untersuchung (u. a. durch das Einbeziehen von drei weiteren Forschern) besteht zudem die Möglichkeit, dass nicht alle für das Thema relevanten Forschungsartikel identifiziert und einbezogen wurden. Weitere Literatur könnte zu kleineren Veränderungen der Ergebnistabelle führen, sodass die identifizierten Forschungslücken geringfügig anders aussehen könnten. Die abgeleiteten Potenziale von KI-Ansätzen für AAC-Systeme basieren ausschließlich auf der Forschungsliteratur. Auf Grundlage dieser Ergebnisse könnte ein Abgleich mit den Funktionalitäten am Markt verfügbarer AAC-Systeme erfolgen. In diesem Zuge würde eine Marktstudie die Möglichkeit bieten, die in AAC-Systemen umgesetzten KI-Funktionalitäten zu erheben und Abdeckungslücken zu identifizieren. Eine Gegenüberstellung mit den in der Forschungsliteratur erläuterten Möglichkeiten und Potenzialen könnte darauf aufbauend für die Anbieter von AAC-Systemen umfangreiche Weiterentwicklungsvorschlägen aufzeigen.

Im Rahmen von **Veröffentlichung 6** wurde ein AAC-Prototyp mit einem Generativen KI-basierten Kompositionsansatz entwickelt. Dieser Prototyp wurde in Anlehnung an Venable et al. (2016) bislang ausschließlich technisch evaluiert. Eine Limitation stellt daher die bislang noch nicht durchgeführte empirische Validierung im Rahmen der von Venable et al. (2016) vorgeschlagenen *„Human Risk & Effectiveness“* Evaluationsstrategie dar. Um die nächsten konsequenten Schritte hin zu einer noch ausgereifteren Designtheorie zu gehen, soll der Prototyp zunächst in einer formativen Umgebung (z. B. im Rahmen eines simulierten Laborexperiments) evaluiert werden. Dabei soll eine Stich-

probe von sprachbehinderten Personen, die alle dieselbe Sprachbehinderung bzw. Spracheinschränkung aufweisen, verschiedene formelle und informelle Interaktionen mit ihren Begleitungen und dem Personal eines Restaurants durchführen. Auf diese Weise soll beobachtet werden, wie sprachbehinderte Personen den Prototyp und insbesondere den Generativen KI-basierten Kompositionsansatz anwenden. Darauf basierend können Verbesserungsmöglichkeiten für diesen Prototyp identifiziert werden, die dann abgewogen und umgesetzt werden sollen, bevor eine Evaluierung im Rahmen einer Feldstudie in einer Realweltsituation in einem Restaurant durchgeführt wird. In diesen empirischen Untersuchungen könnten zudem die spezifischen Umstände untersucht werden (z. B. bestimmte Kommunikationsaktivitäten oder Gesprächspartner), welche zu einer zu großen kognitiven Belastung für sprachbehinderte Personen führen, sodass diese ggf. keine kontextspezifische Artikulation wünschen. Die Untersuchung der Wechselwirkungen zwischen Entlastung und kognitiver Belastung in empirischen Settings könnte damit Hinweise darauf geben, wann der vorgeschlagene KI-basierte Ansatz wie viel Unterstützung für sprachbehinderte Personen bieten sollte. So könnte es bspw. Gespräche geben, in denen bewusst weniger Unterstützung seitens der sprachbehinderten Personen gewünscht wird. Das manuelle und nicht durch die Generative KI unterstützte Erstellen von Gesprächsbeiträgen könnte z. B. von sprachbehinderten Personen als Mittel genutzt werden, um Gesprächspartnern gegenüber ihr Wissen zu bestimmten Themen unter Beweis zu stellen (vgl. Valencia et al., 2023). Die Berücksichtigung verschiedener situationsspezifischer Unterstützungsniveaus im Rahmen des vorgeschlagenen KI-basierten Ansatzes könnte daher eine noch bedarfsgerechtere Kommunikation mit AAC-Systemen eröffnen.

Die Untersuchung sowohl der gegensätzlichen als auch der verstärkenden Wirkungen der Maxime von Grice (1975) in empirischen Kontexten könnte weitere Verfeinerungen der kontextspezifischen Artikulationsstile sowie weitere Erkenntnisse dahingehend aufbringen, wie sprachbehinderte Personen diese Konversationsmaxime im Rahmen konkreter Gesprächssituationen (z. B. in Verhandlungen mit Gesprächspartnern) anwenden. In diesem Zusammenhang könnten zukünftige Forschungsarbeiten auch darauf abzielen, ein ganzheitliches AAC-System zur Abdeckung aller relevanten Dimensionen eines Gesprächs zu entwickeln. Um die Gesprächseröffnung und den Gesprächsabschluss zu unterstützen, bieten sich als KI-Ansätze z. B. Techniken aus den Bereichen der Bilderken-

nung und Spracherkennung an. Dadurch könnten automatisiert Bewegungen (z. B. Handreichen oder Winken) und Floskeln (z. B. Begrüßung oder Verabschiedung) des Gegenübers erkannt werden, die auf eine beginnende oder eine unmittelbar endende Konversation hindeuten.

Der Vergleich der Chatbot-Lösungen in **Veröffentlichung 7** weist die Limitation auf, dass die erhaltenen Antworten durch die Chatbot-Anbieter bewusst und gezielt beschönigt worden sein könnten. Nichtsdestotrotz waren für die Marktstudie die Antworten und Einschätzungen der Chatbot-Anbieter unerlässlich. Weiterhin stellt die zeitlich begrenzte Gültigkeit des Marktüberblicks, welcher im Jahr 2018 durchgeführt wurde, eine Einschränkung dar. Insgesamt lag die Rückläuferquote der Marktstudie bei ca. 30 %, sodass von 48 kontaktierten Anbietern 14 einen vollständig ausgefüllten Fragebogen zurückgesendet haben. Es könnten sich im zeitlichen Verlauf neue bzw. veränderte Anforderungen und Funktionalitäten ergeben haben, die für einen Chatbot in der Kundenkommunikation im B2C-Kontext von essenziell wichtiger Bedeutung sind. Eine Wiederholung der Marktstudie mit einer größeren Anzahl an Anbietern könnte daher ein aktualisiertes Bild über die Fähigkeiten der am Markt angebotenen Chatbot-Lösungen liefern.

Veröffentlichung 8 präsentiert vielfältige Ansätze für zukünftige Forschungsarbeiten. Im Rahmen dieses DSR-Projekts könnte der nächste konsequente Schritt gegangen werden, indem die identifizierten Gestaltungsmerkmale mit Experten aus Wissenschaft und Praxis diskutiert werden. Auf diese Weise könnten synergetische und gegenläufige Effekte im Hinblick auf die Produktivitätssteigerung durch diese Gestaltungsmerkmale identifiziert und Designprinzipien für die Steigerung der Produktivität durch Chatbot-Lösungen für die Unterstützung in operativen Prozessen innerhalb von Unternehmen abgeleitet werden. Daran anschließend könnte in Bezug auf den Design Cycle nach Hevner (2007) ein produktivitätsorientierter Chatbot für eine ausgewählte Domäne mit den vielversprechendsten Gestaltungsmerkmalen instanziiert werden. Durch die Durchführung von Experimenten mit solchen Chatbot-Prototypen unter Anwendung des Between-Subject-Designs nach Boudreau et al. (2001) könnten die aufgestellten Designprinzipien evaluiert und noch tiefgreifendere Effekte und Wechselwirkungen spezifischer Gestaltungsmerkmale auf die Produktivität bei der Ausführung operativer Aufgaben ermittelt werden.

3.3.2. Forschungsansätze aus der themenübergreifenden Betrachtung und Integration der Forschungsergebnisse

Neben den aufgezeigten Forschungsmöglichkeiten innerhalb der beiden Themenbereiche der Dissertation ergeben sich weitere vielversprechende Forschungsansätze aus der themenübergreifenden Betrachtung und Integration der jeweiligen Forschungsergebnisse. Nachfolgend werden exemplarisch die vielversprechendsten Synergien aus den beiden Themengebieten der Dissertation dargelegt und erläutert.

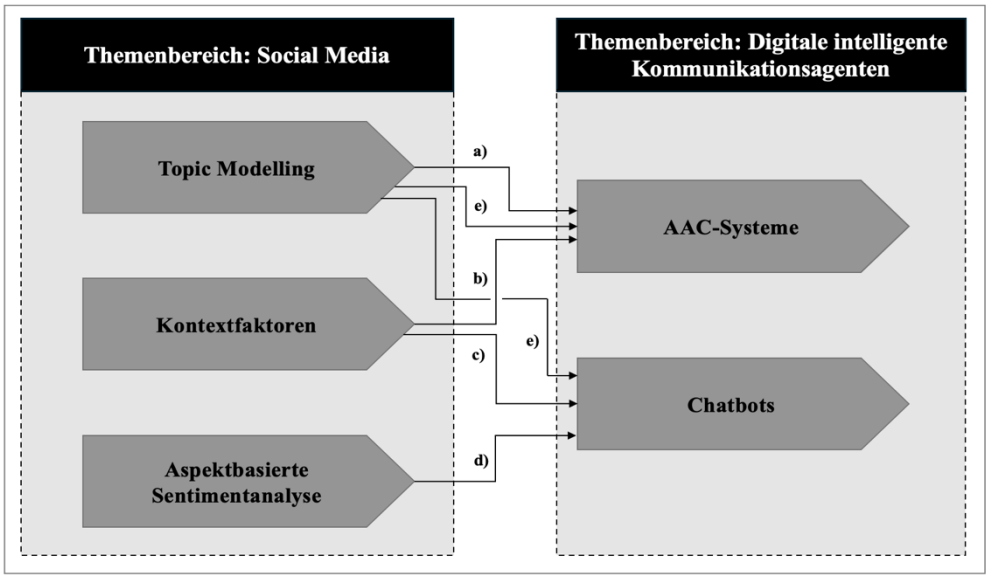

Abbildung 2: Forschungssynergien aus den Themengebieten der Dissertation

a) Das Topic Modelling als Analysetechnik kann neben den aufgezeigten Anwendungsmöglichkeiten für die Trendanalyse für textbasierte Social Media Daten auch gewinnbringend im Rahmen von AAC-Systemen angewendet werden. Topic Modelling ermöglicht es, die Themen in den unter Anwendung von AAC-Systemen geführten Gesprächen zu identifizieren. Dadurch können u. a. für Betreuer, Angehörige oder auch Sprachtherapeuten und Logopäden Zusammenfassungen über die Inhalte und die Sprachvielfalt der von den sprachbehinderten Personen geführten Gespräche generiert werden (Garcia et al., 2016; Hervás et al., 2020; Mitchell und Sproat, 2010).

Im Rahmen von AAC-Systemen werden bislang aber ausschließlich Unsupervised Learning Techniken eingesetzt. Im Vergleich zu Semi-Supervised Learning Techniken wie dem GuidedLDA oder dem SeededLDA lassen diese Unsupervised Learning

Techniken keine Kontextualisierung der identifizierten Themen mit externen Parametern zu (vgl. Ahadh et al., 2021; Chandrasekaran et al., 2020; Toubia et al., 2019). Diese Kontextualisierung ermöglicht es bei Trendanalysen im Rahmen von marketingbezogenen Anwendungsfällen, dass die extrahierten Trendthemen semantisch auf ein Unternehmen und seine Tätigkeitsfelder zugeschnitten sind (Wörner et al., 2023).

Eine solche Kontextualisierung erscheint auch für AAC-Systeme als sinnvoll und könnte speziell für diese Systeme der unterstützten Kommunikation konzipiert und implementiert werden. Eine Semi-Supervised Learning Topic Modelling Technik könnte z. B. dabei helfen, basierend auf Gesprächsverläufen aus Alltagsgesprächen diejenigen Sprechtasten zu identifizieren, welche im Hinblick auf konkrete personenbezogene Merkmale einer sprachbehinderten Person und konkrete Gesprächssituationen semantisch besonders relevant sind. Die personenbezogenen Merkmale der sprachbehinderten Person (z. B. Beruf, Hobbies oder Interessen aus dem im AAC-System hinterlegten Nutzerprofil) und die Charakteristika einer konkreten Gesprächssituation könnten der Semi-Supervised Learning Topic Modelling Technik als externer Parameter übergeben werden, um damit für den jeweiligen Kontext geeignete Sprechtasten im Hinblick auf die Persönlichkeit oder die jeweilige Gesprächssituation zu identifizieren. Die dadurch verbesserte Vorselektion von Sprechtasten könnte dazu beitragen, die durch AAC-Systeme eröffneten Gesprächsraten zu erhöhen und um die kognitive Belastung für die sprachbehinderten Personen bei der Auswahl von situations- und personenspezifisch passenden Sprechtasten zu reduzieren.

b) Kontextfaktoren nehmen auch im Rahmen von AAC-Systemen eine wichtige Rolle ein (vgl. Konadl et al., 2023). Sprachbehinderten Personen können damit z. B. basierend auf bestimmten Geolokationen (z. B. Arztpraxis, Restaurant oder Zuhause bei der Familie bzw. Angehörigen) oder Tagesphasen (Morgen, Mittag oder Abend) situationsspezifisch passende Sprechtasten vorgeschlagen werden (vgl. Dempster et al., 2010; Hossain et al., 2018; Klauer et al., 2021; Laxmidas et al., 2021; Neamtu et al., 2019; Obiorah et al., 2021). Für AAC-Systeme können damit ebenfalls höhere Gesprächsraten ermöglicht und die kognitive Belastung für sprachbehinderte Personen bei der Auswahl von passenden Sprechtasten reduziert werden.

Über die zu den AAC-Systemen bereits betrachteten Kontextfaktoren hinausgehend könnten auch die Kontextfaktoren aus dem ersten Themenbereich der Dissertation für den konkreten Anwendungsfall der AAC-Systeme untersucht werden. So lässt der Kontextfaktor „mobile Betriebssysteme“ (z. B. Android oder Apple) Aussagen über die Persönlichkeit einer Person und über seine Verhaltensweisen zu. Basierend auf einer psychologischen Untersuchung wurden Nutzer von Apple-Geräten demnach z. B. als weniger bescheiden und eher auf die eigenen Vorteile bedacht charakterisiert (Shaw et al., 2016). So könnten auch sprachbehinderte Menschen, welche ihr AAC-System auf einem Apple-Gerät installiert haben, in bestimmten Alltagssituationen in ihren Gesprächsbeiträgen weniger Kompromissbereitschaft gegenüber ihren Gesprächspartnern zeigen wollen. Diese Erkenntnis könnte in den in Veröffentlichung 6 (vgl. Konadl, 2024) vorgeschlagenen Generativen KI-basierten Kompositionsansatz einbezogen werden, um im Rahmen von Verhandlungen (z. B. in Geschäftsterminen im Arbeitsleben, bei Verhandlungen über den Kaufpreis beim Erwerb eines Autos) weniger Entgegenkommen und ein Beharren auf den eigenen Standpunkten in den automatisiert generierten Gesprächsbeiträgen zu verankern.

c) Die Einflüsse von Kontextfaktoren könnten auch für Chatbots im Kundenservice tiefergehend untersucht werden. Kontextfaktoren wie z. B. Geolokation, Zeit oder demographische Merkmale von Personen (Alter, Geschlecht, etc.) beeinflussen maßgeblich, welche Tonalitäten in Social Media Posts zum Ausdruck gebracht werden (vgl. Cao et al., 2018; Cruz-Albrecht et al., 2017; Dodds et al., 2011; Golder und Macy, 2011; Lim et al., 2018; Lin, 2014; Plunz et al., 2019; Schwartz et al., 2019). Auch Konversationen, welche (potenzielle) Kunden mit einem Chatbot eines Unternehmens führen, könnten maßgeblichen Einflüssen verschiedener Kontextfaktoren unterliegen. Um hierfür robuste Effekte ableiten zu können, gilt es in einem ersten Schritt zunächst einen umfangreichen Datensatz mit Chatbot-Konversationen aus dem Bereich der Kundenkommunikation (z. B. Kundenanfragen im Kundenservice) zu extrahieren. Hierfür könnte z. B. ein Chatbot für einen Kundenberatungsprozess mit hoher Anfragehäufigkeit wie z. B. dem IT-Servicedesk der Universität Regensburg implementiert werden. Im Rahmen dieses Prozesses könnte der Chatbot die Erbringung von Serviceleistungen und die Beantwortung von häufig gestellten Fragen

(z. B. Schritte zum Zurücksetzen des Passworts) gegenüber den Angehörigen der Universität als Kunden übernehmen. In einem zweiten Schritt könnten dann die einzelnen Kontextfaktoren (Ort, Zeit, Alter, Geschlecht des Nutzers, etc.) den Gesprächsverläufen zugeordnet (z. B. Ermittlung verschiedenster Zeitmaße auf Basis der Zeitstempel der gesendeten Nachrichten) und deren Einflüsse auf das Sentiment der jeweiligen Nachrichten ermittelt werden. Während die Geolokation über die IP-Adresse zugeordnet werden könnte, könnten die demographischen Merkmale durch den Chatbot im Dialog abgefragt werden und diese für die Zwecke der Untersuchung verschlüsselt in einer entsprechenden Datenbank gespeichert werden.

Abhängig von den ermittelten Effekten könnten je nachdem, welche Kontextfaktoren zu Beginn einer Konversation mit dem Chatbot vorliegen (z. B. montagmorgens im Vergleich zu samstagnachmittags) unterschiedliche Konversationspfade durch den Chatbot ausgespielt werden. Auch könnten bei Kontextfaktoren, bei denen negative Sentimente in einem Gesprächsablauf tendenziell häufiger auftreten (z. B. bei Kundenanfragen montagmorgens) durch adaptierte Tonalitäten und Entgegnungen in den Chatbot-Nachrichten negative Kundenreaktionen und Kundeneindrücke von vornherein verhindert werden (vgl. Diederich et al., 2019).

d) Der in Veröffentlichung 4 konzipierte Ansatz basierend auf der aspektbasierten Sentimentanalyse (vgl. Wörner et al., 2023) könnte weiterhin in Chatbots integriert und spezifisch im Rahmen dieses Anwendungskontexts untersucht werden. Bei der aspektbasierten Sentimentanalyse werden zunächst in einem Text die Bestandteile von z. B. Produkten oder Services identifiziert, um anschließend gegenüber diesen Aspekten die Tonalität (positive oder negative Wahrnehmung des Aspekts durch den Verfasser) zu ermitteln (Binder et al., 2019; Schouten und Frasincar, 2015). Eine satzbasierte Sentimentanalyse kann die dafür notwendige Granularität nicht gewährleisten, weil die Granularität dieser Sentimentanalyse keine Aussage darüber zulässt, gegenüber welchen konkreten Aspekten (z. B. Burgerbun, Sauce oder Fleisch als Aspekte eines Burgers als Produkt) sich der Nutzer positiv oder negativ geäußert hat (Feldman, 2013; Liu, 2012).

Einerseits könnte durch die Integration dieses aspektbasierten Sentimentanalysensatzes in einen Chatbot eine verbesserte Auswertung und Kontrolle der Konversationsverläufe erfolgen, weil für Unternehmen dadurch unmittelbar ersichtlich werden würde, worüber sich ein Kunde positiv oder negativ gegenüber dem Chatbot geäußert hat. Wenn sich ein Kunde gegenüber bestimmter kritischer Aspekte, wie z. B. Produktfehler mit Gefährdungspotenzial für einen Menschen, negativ äußert, könnte eine Übergabe an einen menschlichen Kundenberater frühzeitig erfolgen, um beschwichtigen und negativen Kundeneindrücken und Reputationsschäden unmittelbar entgegenwirken zu können. Durch die feine Granularität der ermittelten Sentimente könnten die Ergebnisse andererseits auch unmittelbar in (inkrementelle) Produktinnovationen und Produktverbesserungen einbezogen werden, ohne dass Unternehmen die Konversationsverläufe ihrer Chatbots nochmals gesondert analysieren müssen. Die Auswertung aus dem Chatbot stellt damit und auf diese Weise eine weitere, effektive Anreicherungsmöglichkeit für die unternehmensinternen Innovationsprozesse dar.

Da die korrekte Funktionsweise des konzipierten Sentimentanalyseansatzes aus Veröffentlichung 4 (Wörner et al., 2023) für den Kontext von Burgern (Speisen) in Restaurants nachgewiesen wurde, könnte eine tiefergehende Untersuchung im Rahmen von Fallstudien in Anlehnung an Yin (2015) zusammen mit Restaurants erfolgen, die bereits einen Chatbot als Kundenkontaktkanal etabliert haben.

e) Weiterhin könnte auf Veröffentlichung 3 (vgl. Wörner et al., 2023) aufgesetzt und mithilfe der im Analysetool „MANTRA“ integrierten Semi-Supervised Topic Modelling Technik für spezifische Themenbereiche ein sogenannter *„Computational Literature Review“* (vgl. Antons et al., 2023; Bandara et al., 2015) durchgeführt werden. Für einen CLR, welcher bedingt durch den Einsatz automatisierter Analysetechniken große Anzahlen an Papern einbeziehen und damit eine große thematische Breite abdecken kann, wird häufig auf Unsupervised Learning oder Semi-Supervised Learning Techniken und insbesondere auf das Topic Modelling zurückgegriffen. Abhängig von der konkreten Zielsetzung des CLR könnten auch anwendungsfallspezifische Topic Modelling Techniken (z. B. LDA für eine reine Extraktion und Konsolidierung der Themen aus Papern, PAM für eine Strukturierung und Hierarchisierung der identifi-

zierten Themen) vor dem Hintergrund eines konkreten Untersuchungsziels bzw. Untersuchungszwecks herangezogen werden (Wörner et al., 2021). Mithilfe evolutionärer Topic Modelling Techniken wie dem Dynamic Topic Modelling oder BERTopic (vgl. Raman et al., 2024) könnten auch Veränderungen der Themen und Forschungstrends im Zeitablauf identifiziert werden.

Vor allem in den Themenbereichen der Chatbots und der AAC-Systeme, in denen solche CLRs bislang wenig verbreitet sind, könnten damit die im Zeitverlauf behandelten Forschungsthemen sowie deren thematische Entwicklungen nachvollzogen werden. Perspektivisch lässt sich damit ein Verständnis über das Zustandekommen bestimmter Forschungsschwerpunkte entwickeln, worauf basierend Forschungsagenden und Forschungsframeworks entwickelt werden könnten. Die Ergebnisse eines CLRs könnten auf diese Weise als Vorstrukturierungen für diese Forschungsgebiete dienen, um daran anschließend in Forschergruppen systematische Literature Reviews als thematische Tiefenbohrungen anstoßen zu können.

Literaturverzeichnis

Hinweis: Die hier angegebene Literatur wird in den Kapiteln 1 und 3 referenziert. In Kapitel 2 wird für die jeweiligen Forschungsbeiträge die referenzierte Literatur im Anschluss an den jeweiligen Beitrag gelistet.

Abdul-Kader, S., und Woods, J. (2015). "Survey on chatbot design techniques in speech conversation systems." *International Journal of Advanced Computer Science Applications*, 6(7).

Ahadh, A., Binish, G., und Srinivasan, R. (2021). "Text mining of accident reports using semi-supervised keyword extraction and topic modeling." *Process Safety and Environmental Protection*, 155.

Akhtar, M.,. Neidhardt J, und Werthner, H. (2019). "The Potential of Chatbots: Analysis of Chatbot Conversations." *Proceedings of the 21st IEEE Conference on Business Informatics*, S. 397–404.

Allen, J. (2005). "Designing desirability in an augmentative and alternative communication device." *Universal Access in the Information Society*, 4(2), S. 135–145.

American Speech Language Hearing Association (1991). "Report: augmentative and alternative communication." *ASHA*, 33, 9–12.

Antons, D., Breidbach, C., Joshi, A., und Salge, T. (2023). "Computational Literture Reviews: Method, Algorithms, and Roadmap." *Organizational Research Methods*, 26(1), S. 107–138.

Ashktorab, Z., Jain, M., Liao, Q., und Weisz, J. (2019). "Resilient chatbots: Repair strategy preferences for conversational breakdowns." *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-12.

Axtell, C., Holman, D., Unsworth, K., Wall, T., Waterson, P., und Harrington, E. (2000). "Shopfloor innovation: Facilitating the suggestion and implementation of ideas." *Journal of Occupational and Organizational Psychology*, 73(3).

Bae, J., Havsol, J., Karpefors, M., Karlsson, A., und Mathiason, G. (2018). "Short Text Topic Modeling to Identify Trends on Wearable Bio-Sensors in Different Media

Type." *Proceedings of the 6th International Symposium on Computational and Business Intelligence*.

Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., und Beekhuyzen, J. (2015). "Achieving rigor in literature reviews: Insights from qualitative data analysis and tool-support." *Communications of the Association for Information systems*, 37(1).

Banh, L., und Strobel, G. (2023). "Generative artificial intelligence." *Electronic Markets*, 33(1), 1-17.

Barton, D., Woetzel, J., Seong, und Tian, Q. (2017). "Artificial Intelligence: Implications for China." *McKinsey Global Institute*.

Basili, G., Caldiera, V., und Rombach, H. (1994). "The goal question metric approach." *Encyclopedia of Software Engineering*, 528-532.

Bedrosian, J., Hoag, L., und McCoy, K. (2003). "Relevance and speed of message delivery trade- offs in augmentative and alternative communication." *Journal of Speech, Language, and Hearing Research*, 46, 800–817.

Bhor, H., Koul, T., Malviya, R., und Mundra, K. (2018). "Digital Media Marketing Using Trend Analysis on Social Media." *Proceedings of the 2nd International Conference on Inventive Systems and Control*.

Binder, M., Heinrich, B., Klier, M., Obermeier, A., und Schiller, A. (2019). "Explaining the stars: Aspect-based sentiment analysis of online customer reviews." *Proceedings of the 27th European Conference on Information Systems*.

Blei, D., Ng, A., und Jordan, M. (2003). "Latent dirichlet allocation." *Journal of Machine Learning Research*, 3, 993-1022.

Boudreau, M.-C., Gefen, D., und Straub, D. (2001). "Validation in Information Systems Research: A State-of-the-Art Assessment." *MIS Quarterly*, 25(1).

Boyes-Braem, P., Erting, C., Johnson, R., Smith, D., und Snider, B. (1994). "An overview of current sign language projects in Switzerland." *The Deaf Way: Perspectives from the International Conference on Deaf Culture*, 382-387.

Bradshaw, J. (2013). "The use of augmentative and alternative communication apps for the iPad, iPod and iPhone: an overview of recent developments." *Tizard Learning Disability Review*, 18(1), 31-37.

Brady, N. (2000). "Improved comprehension of object names following voice output communication aid use: Two case studies." *Augmentative and Alternative Communication*, 16(3), 197-204.

Brandtzaeg, P., und Folstad, A. (2017). "Why people use chatbots." *Proceedings of the 4th International Conference of Internet Science,* 377-392.

Braun, A. (2013). "Chatbots in der Kundenkommunikation." *Springer Verlag*.

Bughin, J., Hazan, E., Ramaswamy, M., Chui, T., Allas, P., Dahlström, N., Henke, M., und Trench, M. (2017). "Artificial intelligence: The next digital frontier." *McKinsey Global Institute*.

Cao, X., MacNaughton, P., Deng, Z., Yin, J., Zhang, X., und Allen, J. (2018). "Using twitter to better understand the spatiotemporal patterns of public sentiment: a case study in Massachusetts, USA." *International Journal of Environmental Research and Public Health*, 15(2).

Chakrabarti, C., und Luger, G. (2015). "Artificial Conversations for Customer Service Chatter Bots: Architecture, Algorithms, and Evaluation Metric." *Expert Systems with Applications,* 42(20), 6878–6897.

Chandrasekaran, R., Mehta, V., Valkunde, T., und Moustakas, E. (2020). "Topics, trends, and sentiments of tweets about the COVID-19 pandemic." *Journal of Medical Internet Research*, 22(10).

Chinnov, A., Kerschke, P., Meske, C., Stieglitz, S., und Trautmann, H. (2015). "An Overview of Topic Discovery in Twitter Communication through Social Media Analytics." *Proceedings of the 21st Americas Conference on Information Systems*.

Cooper, H. (1988). "Organizing knowledge syntheses: A taxonomy of literature reviews." *Knowledge in Society*, 1(1).

Cooper, H., und Hedges, L. (1994). "Research synthesis as a scientific enterprise." *The Handbook of Research Synthesis*, 285-299.

Cruz-Albrecht, L., Xu, J., Ni, K.-Y., und Lu, T.-C. (2017). "Characterizing Regional and Behavioral Device Variations Across the Twitter Timeline: A Longitudinal Study." *Proceedings of the ACM Conference on Web Science*.

Dale, R. (2016). "The return of the chatbots." *Natural Language Engineering*, 22(5), 811-817.

Debortoli, S., Müller, O., Junglas, I., und Vom Brocke, J. (2016). "Text mining for information systems researchers: An annotated topic modeling tutorial." *Communications of the Association for Information Systems*, 39(1).

Dempster, M., Alm, N., und Reiter, E. (2010). "Automatic generation of conversational utterances and narrative for Augmentative and Alternative Communication: a prototype system." *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, 10-18.

Desai, T., Chow, K., Mumford, L., Hotze, F., und Chau, T. (2014). "Implementing an iPad-based alternative communication device for a student with cerebral palsy and autism in the classroom via an access technology delivery protocol." *Computers & Education*, 79, 148-158.

Dey, A. (2001). "Understanding and using context." *Personal and Ubiquitous Computing*, 5(1), S. 4-7.

Diederich, S., Janssen-Müller, M., Brendel, A., und Morana, S. (2019). "Emulating Empathetic Behavior in Online Service Encounters with Sentiment-Adaptive Responses: Insights from an Experiment with a Conversational Agent." *Proceedings of the 40th International Conference on Information Systems*.

Ding, K., Choo, W., Ng, K., Ng, S. (2020). "Employing structural topic modelling to explore perceived service quality attributes in Airbnb accommodation." *International Journal of Hospitality Management*, 91.

Dodds, P., Harris, K., Kloumann, I., Bliss, C., und Danforth, C. (2011). "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter." *PLOS One*, 6(12).

Dwivedi, Y., Kshetri, N., Hughes, L., Slade, E., Jeyaraj, A., Kar, A., ..., und Wright, R. (2023). "'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy." *International Journal of Information Management*, 71, 102642.

Eickhoff, M., und Neuss, N. (2017). "Topic Modelling Methodology: Its Use in Information Systems and Other Managerial Disciplines." *Proceedings of the 25th European Conference on Information Systems*.

Einwiller, S., und Steilen, S. (2015). "Handling complaints on social network sites–An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies." *Public Relations Review*, 41(2), 195-204.

Endo, Y., Toda, H., und Koike, Y. (2015). "What's Hot in the Theme: Query Dependent Emerging Topic Extraction from Social Streams." *Proceedings of the 24th International Conference on the World Wide Web*.

Fan, W., und Gordon, M. (2014). "The power of social media analytics." *Communications of the ACM*, 57(6).

Fast, E., und Horvitz, E. (2017). "Long-term trends in the public perception of artificial intelligence." *Proceedings of the 31st AAAI Conference on Artificial Intelligence*.

Feldman, R. (2013). "Techniques and applications for sentiment analysis." *Communications of the ACM*, 56(4), S. 82-89.

Feuerriegel, S., Hartmann, J., Janiesch, C., und Zschech, P. (2023). "Generative AI." *Business & Information Systems Engineering,* 66, 111-126.

Folstad, A., und Brandtzaeg, P. (2017). "Chatbots and the new world of HCI." *Interactions*, 24(4), 38- 42.

Fritzell, B. (1996). "Voice disorders and occupations." *Logopedics Phoniatrics Vocology*, 21(1), 7-12.

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., und Chen, L. (2023). "Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration." *Journal of Information Technology Case and Application Research*, 25(3), 277-304.

Gabriel, R., und Röhrs, H. (2017). "Trends, Chancen und Risiken von Social-Media-Anwendungen – eine kritische Betrachtung." *Springer Gabler, Berlin, Heidelberg*.

Garcia, L., de Oliveira, L., und de Matos, D. (2016). "Evaluating pictogram prediction in a location-aware augmentative and alternative communication system. " *Assistive Technology*, 28(2), 83-92.

Golder, S., und Macy, M. (2011). "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures." *Science*, 333, 1878-1881.

Gulati, R., und Oldroyd, J. (2005). "The Quest for Customer Focus." *Harvard Business Review*, 83(4).

Grégoire, Y., Salle, A., und Tripp, T. M. (2015). "Managing social media crises with your customers: The good, the bad, and the ugly." *Business Horizons*, 58(2), 173-182.

Gregor, S., und Hevner, A. (2013). "Positioning and Presenting Design Science Research for Maximum Impact." *MIS Quarterly*, 37(2), 337-355.

Grice, H. (1975). "Logic and conversation." *Speech Acts, New York: Academic Press*.

Ha, T., Beijnon, B., Kim, S., Lee, S., und Kim, J. (2017). "Examining User Perceptions of Smartwatch through Dynamic Topic Modeling." *Telematics and Informatics*, 34(7), 1262-1273.

Hanna, R., Rohm, A., und Crittenden, V. (2011). "We‘re all connected: The power of the social media ecosystem." *Business Horizons*, 54(3), 265-273.

Henne, H., und Rehbock, H. (2012). "Einführung in die Gesprächsanalyse."*De Gruyter*.

Henrich, O. (2017). "Chatbots auf dem Vormarsch: Der künstlich–intelligente Buchhalter kommt." *Wirtschaftsinformatik & Management*, 9, 72–75.

Heo, Y., und Kang, S. (2019). "Augmentative and Alternative Communication System Using Information Priority and Retrieval." *Proceedings of the 2019 International Conference on Big Data and Smart Computing*.

Hervás, R., Bautista, S., Méndez, G., Galván, P., und Gervás, P. (2020). "Predictive composition of pictogram messages for users with autism." *Journal of Ambient Intelligence and Humanized Computing*, 11, 5649-5664

Hevner, A., Salvatore, M., Jinsoo, P., und Sudha, R. (2004). "Design science in information systems research." *MIS Quarterly*, 28(1), 75-105.

Hevner, A. (2007). "A three cycle view of design science research." *Scandinavian Journal of Information Systems*, 19(2).

Hicks, A., Comp, S., Horovitz, J., Hovarter, M., Miki, M., und Bevan, J. (2012). "Why People Use Yelp.com: An Exploration of Uses and Gratifications." *Computers in Human Behavior*, 28(6), 2274-2279.

Hoag, L., Bedrosian, J., McCoy, K., und Johnson, D. (2004). "Trade-offs between informativeness and speed of message delivery in augmentative and alternative communication." *Journal of Speech, Language, and Hearing Research*, 47, 1270–1285.

Hong, L., und Davison, B. (2010). "Empirical Study of Topic Modeling in Twitter." *Proceedings of the 1st Workshop on Social Media Analytics*.

Hong, L., Ahmed, A., Gurumurthy, S., Smola, A., und Tsioutsiouliklis, K. (2012). "Discovering Geographical Topics in the Twitter Stream." *Proceedings of the 21st International Conference on the World Wide Web*.

Hourcade, J., Everhart Pilotte, T., West, E., und Parette, P. (2004). "A history of augmentative and alternative communication for individuals with severe and profound disabilities." *Focus on Autism and other Developmental Disabilities*, 19(4), 235-244.

Hossain, M., Takanokura, M., Sakai, H., und Katagiri, H. (2018). "Using context history and location in context-aware AAC systems for speech-language impairments." *Proceedings of the International Multiconference of Engineers and Computer Scientists*, 1.

Hu, T., She, B., Duan, L., Yue, H., und Clunis, J. (2019). "A Systematic Spatial and Temporal Sentiment Analysis on Geo-Tweets." *IEEE Access*, 8, 8658-8667.

Hutto, C., und Gilbert, E. (2014). "VADER: A parsimonious rule-based model for sentiment analysis of social media text." *Proceedings of the International AAAI Conference on Web and Social Media*, 216-225.

Irawan, M., Wijayanto, R., Shahab, M., Hidayat, N., und Rukmi, A. (2020). "Implementation of Social Media Mining for Decision Making in Product Planning Based on Topic Modeling and Sentiment Analysis." *Journal of Physics: Conference Series*.

Jansen, B., Zhang, M., Sobel, K., und Chowdury, A. (2009). "Twitter power: Tweets as electronic word of mouth." *Journal of the American Society for Information Science and Technology*, 60(11), 2169-2188.

Janssen, A., Grützner, L., und Breitner, M. (2021). "Why do chatbots fail? A critical success factors analysis." *Proceedings of the 42nd International Conference on Information Systems*.

Jeong, B., Yoon, J., und Lee, J.-M. (2019). "Social Media Mining for Product Planning: A Product Opportunity Mining Approach Based on Topic Modeling and Sentiment Analysis." *International Journal of Information Management,* 48, 280-290.

Johannsen, F., Leist, S., Konadl, D., Basche, M., und de Hesselle, B. (2018). "Comparison of Commercial Chatbot Solutions for Supporting Customer Interaction." *Proceedings of the 26th European Conference on Information Systems*.

Johannsen, F., Schaller, D., und Klus, M. (2021). "Value propositions of chatbots to support innovation management processes." *Information Systems and e-Business Management*, 19, 205-246.

Kane, S., Morris, M., Paradiso, A., und Campbell, J. (2017). "At times avuncular and cantankerous, with the reflexes of a mongoose. Understanding Self-Expression through Augmentative and Alternative Communication Devices." *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1166-1179.

Kaplan, A., und Haenlein, M. (2010). "Users of the world, unite! The challenges and opportunities of Social Media." *Business Horizons,* 53(1), 59-68.

Khan, H., Nasir, S., Nasim, K., Shabbir, D., und Mahmood, A. (2021). "Twitter trends: a ranking algorithm analysis on real time data." *Expert Systems with Applications*, 164.

Kietzmann, J., Hermkens, K., McCarthy, I., und Silvestre, B. (2011). "Social media? Get serious! Understanding the functional building blocks of social media." *Business Horizons,* 54(3), 241-251.

Kim, S., Wang, R., Maslowska, E., und Malthouse, E. (2016). "'Understanding a fury in your words': The effects of posting and viewing electronic word-of-mouth on purchase behaviors." *Computers in Human Behavior*, 54, 511-521.

Kirakowski, J., und Corbett, M. (1993). "SUMI: The Software Usability Measurement Inventory." *British Journal of Educational Technology*, 24(3), 210–212.

Klauer, A., Johannsen, F., und Leist, S. (2021). "BeMyVoice – An iPad-Based Helper for Speech Impaired Persons." *Proceedings of the International Conference on Design Science Research in Information Systems and Technology*, 40-51.

Konadl, D., Wörner, J., und Leist, S. (2021). " Identifying Sentiment Influences Provoked by Context Factors – Results from a Data Analytics Procedure Performed on Tweets." *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Konadl, D., Wörner, J., Luttner, L., und Leist, S. (2023). "Artificial intelligence in Augmentative and Alternative communication systems – a literature-based assessment and implications of different conversation phases and contexts." *Proceedings of the 31st European Conference on Information Systems*.

Konadl, D. (2024). "A Generative AI-Based Approach to Support Automated Utterance Generation for Different Conversational Contexts within AAC Systems." *Proceedings of the 45th International Conference on Information Systems.*

Kottorp, M., und Jäderberg, F. (2017). "Chatbot as a potential tool for businesses: a study on chatbots made in collaboration with Bisnode." *Royal Institute of Technology Stockholm*.

Kristensson, P., Lilley, J., Black, R., und Waller, A. (2020). "A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities." *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

Lau, J., Collier, N., und Baldwin, T. (2012). "On-Line Trend Analysis with Topic Models:# Twitter Trends Detection Topic Model Online." *Proceedings of the 2012 International Conference on Computational Linguistics.*

Laxmidas, K., Avra, C., Wilcoxen, C., Wallace, M., Spivey, R., Ray, W., Polsley, S., Kohli, P., Thompson, J., und Hammond, T. (2021). "CommBo: Modernizing Augmentative and Alternative Communication." *International Journal of Human-Computer Studies*, 145.

Li, W., und McCallum, A. (2006). "Pachinko allocation: DAG-structured mixture models of topic correlations." *Proceedings of the 23rd International Conference on Machine Learning*, 577-584.

Lim, K., Lee, K., Kendal, D., Rashidi, L., Naghizade, E., Winter, S., und Vasardani, M. (2018). "The grass is greener on the other side: Understanding the effects of green spaces on Twitter user sentiments." *Proceedings of the 'The Web' Conference*.

Lin, Y.-R. (2014). "Assessing sentiment segregation in urban communities." *Proceedings of the 2014 International Conference on Social Computing*.

Liu, B. (2012). "Sentiment analysis and opinion mining." *Morgan & Claypool Publishers*.

Liu, L., Tang, L., Dong, W., Yao, S., und Zhou, W. (2016). "An overview of topic modeling and its current applications in bioinformatics." *SpringerPlus*, 5(1).

Louis-Nouvertné, U. (2001). "Was sind Gebärdensprachen?: Eine Einführung in die wichtigsten Ergebnisse der linguistischen Gebärdensprachforschung." *Sprache und Literatur*, 32(2), 3-20.

Lozano, M., Schreiber, J., und Brynielsson, J. (2017). "Tracking Geographical Locations Using a Geo-Aware Topic Model for Analyzing Social Media Data." *Decision Support Systems*, 99, 18-29.

Lu, H., Li, Y., Chen, M., Kim, H., und Serikawa, S. (2018). "Brain intelligence: go beyond artificial intelligence." *Mobile Networks and Applications*, 23(2), 368-375.

Luger, E., und Sellen, A. (2016). "Like having a really bad PA: The gulf between user expectation and experience of conversational agents." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*.

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., ..., und Söllner, M. (2019). "AI-based digital assistants: Opportunities, threats, and research perspectives." *Business & Information Systems Engineering*, 61, 535-544.

Mastrogiacomo, L., Barravecchia, F., Franceschini, F., und Marimon, F. (2021). "Mining quality determinants of product-service systems from user-generated contents." *Quality Engineering*, 33(3), 425-442.

Mayring, P. (2002). "Introduction to qualitative social research." *Weinheim: Beltz*.

Mayring, P. (2004). "Qualitative content analysis." *A Companion to Qualitative Research*, 1(2), 159-176.

Meuter, M., Bitner, M., Ostrom, A., und Brown, S. (2005). "Choosing among alternative service delivery modes: An investigation of customer trial of self-service technologies." *Journal of Marketing*, 69(2), 61–83.

Meyer von Wolff, R., Hobert, S., und Schumann, M. (2019). "How may i help you?–state of the art and open research questions for chatbots at the digital workplace." *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Meyer von Wolff, R., Hobert, S., Masuch, K., und Schumann, M. (2020). "Chatbots at digital workplaces–a grounded-theory approach for surveying application areas and objectives." *Pacific-Asia Journal of the Association for Information Systems*, 12(2).

Mimno, D., und McCallum, A. (2008). "Topic models conditioned on arbitrary features with Dirichletmultinomial regression." *UAI*, 411–418.

Mirtalaie, M., Hussain, O., Chang, E., und Hussain, F. (2018). "Extracting sentiment knowledge from pros/cons product reviews: Discovering features along with the polarity strength of their associated opinions." *Expert Systems with Applications*, 114, 267-288.

Mitchell, M., und Sproat, R. (2012). "Discourse-based modeling for AAC." *Proceedings of the 3rd Workshop on Speech and Language Processing for Assistive Technologies*, 9–18.

Mitchell, L., Frank, M., Harris, K., Dodds, P., und Danforth, C. (2013). "The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place." *PLOS One*, 8(5).

Moro, S., Lopes, R., Esmerado, J., und Botelho, M. (2020). "Service quality in airport hotel chains through the lens of online reviewers." *Journal of Retailing and Consumer Services*, 56.

Neamtu, R., Camara, A., Pereira, C., und Ferreira, R. (2019). "Using artificial intelligence for augmentative alternative communication for children with disabilities." *Proceedings of the 2019 IFIP Conference on Human-Computer Interaction*, 234-243.

Nilsson, N. (2014). "Principles of artificial intelligence." *Morgan Kaufmann*.

Obiorah, M., Piper, A., und Horn, M. (2021). "Designing AACs for people with aphasia dining in restaurants." *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-14.

Pande, P., Neuman, R., und Cavanagh, R. (2014). "The Six Sigma Way: How to maximize the impact of your change and improvement efforts." *McGraw Hill Professional*.

Park, E., Kwon, J., und Kim, S. (2021). "Green marketing strategies on online platforms: a mixed approach of experiment design and topic modeling." *Sustainability*, 13(8).

Peffers, K., Tuunanen, T., Rothenberger, M., und Chatterjee, S. (2007). "A design science research methodology for information systems research." *Journal of Management Information Systems*, 24(3), 45-77.

Pein, V. (2014) "Der Social Media Manager." *Handbuch für Ausbildung und Beruf*.

Pfeffer, J., Zorbach, T., und Carley, K. (2014). "Understanding online firestorms: Negative word-of-mouth dynamics in social media networks." *Journal of Marketing Communications*, 20(1-2), 117-128.

Plunz, R., Zhou, Y., Vintimilla, M., Mckeown, K., Yu, T., Uguccioni, L., und Sutto, M. (2019). "Twitter sentiment in New York City parks as measure of well-being." *Landscape and Urban Planning*, 189, 235-246.

Pollak, V., und Gallagher, B. (1989). "A fast communication aid for non-verbal subjects with severe motor handicaps." *Journal of Medical Engineering & Technology*, 13(1-2), 23-27.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., und Sutskever, I. (2019). "Language models are unsupervised multitask learners." *OpenAI Blog*, 1(8).

Raman, R., Pattnaik, D., Hughes, L., und Nedungadi, P. (2024). "Unveiling the dynamics of AI applications: A review of reviews using scientometrics and BERTopic modeling." *Journal of Innovation & Knowledge*, 9(3).

Ramming, L., Konadl, D., und Leist, S. (2022). "Chatbot Design Features to Increase Productivity." *17th International Conference on Design Science Research in Information Systems and Technology*.

Reisenbichler, M., und Reutterer, T. (2019). "Topic modeling in marketing: recent advances and research opportunities." *Journal of Business Economics*, 89(3), 327-356.

Russell, S., Dewey, D., und Tegmark, M. (2015). "Research priorities for robust and beneficial artificial intelligence." *AI Magazine*, 36(4), 105-114.

Schepis, M., Reid, D., und Behrman, M. (1996). "Acquisition and functional use of voice output communication by persons with profound multiple disabilities." *Behavior Modification*, 20(4), 451- 468.

Schouten, K., und Frasincar, F. (2015). "Survey on aspect-level sentiment analysis." *IEEE Transactions on Knowledge and Data Engineering*, 28(3), 813-830.

Schwartz, A., Dodds, P., O'Neil-Dunne, J., Danforth, C., und Ricketts, T. (2019). "Visitors to urban greenspace have higher sentiment and lower negativity on Twitter." *People and Nature*, 1(4), 476-485.

Shaw, H., Ellis, D., Kendrick, L.-R., Ziegler, F., und Wiseman, R. (2016). "Predicting smartphone operating system from personality and individual differences." *Cyberpsychology, Behavior, and Social Networking*, 19(12), 727-732.

Shen, J., Yang, B., Dudley, J., und Kristensson, P. (2022). "KwickChat: A multi-turn dialogue system for AAC using Context-Aware Sentence Generation by Bag-of-Keywords." *Proceedings of the 27th International Conference on Intelligent User Interfaces*.

Shi, L., Wu, Y., Liu, L., Sun, X., und Jiang, L. (2018). "Event Detection and Identification of Influential Spreaders in Social Media Data Streams." *Big Data Mining and Analytics*, 1(1), 34-46.

Spiegel, C., und Spranz-Fogasy, T. (2001). "Aufbau und Abfolge von Gesprächsphasen." *Text-und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 2. Halbband, De Gruyter*.

Statista. (2024). "Anzahl der Social-Media-Nutzer weltweit in den Jahren 2012 bis 2023." *https://de.statista.com/statistik/daten/studie/739881/umfrage/monatlich-aktive-social-media-nutzer-weltweit/* (Zugriff erfolgt am 30.08.2024).

Stieglitz, S., Dang-Xuan, L., Bruns, A., und Neuberger, C. (2014). "Social media analytics: An interdisciplinary approach and its implications for information systems." *Business & Information Systems Engineering*, 6, 89-96.

Stieglitz, S., Mirbabaie, M., Ross, B., und Neuberger, C. (2018). "Social media analytics–Challenges in topic discovery, data collection, and data preparation." *International Journal of Information Management*, 39, 156-168.

Stucki, T., D' Onofrio, S., und Portmann, E. (2018). "Chatbot - Der digitale Helfer im Unternehmen: Praxisbeispiele der Schweizerischen Post." *HMD - Praxis der Wirtschaftsinformatik*, 55, 725–747.

Sweller, J. (1988). "Cognitive load during problem solving: Effects on learning." *Cognitive Science*, 12(2).

Teubner, T., Flath, C., Weinhardt, C., van der Aalst, W., und Hinz, O. (2023). "Welcome to the era of chatgpt et al. the prospects of large language models." *Business & Information Systems Engineering*, 65(2), 95–101.

Tintarev, N., Reiter, E., Black, R., Waller, A., und Reddington, J. (2016). "Personal storytelling: Using Natural Language Generation for children with complex communication needs, in the wild …" *International Journal of Human-Computer Studies,* 92, 1–16.

Todman, J., und Alm, N. (1997). "Pragmatics and AAC approaches to conversational goals." *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*.

Toubia, O., Iyengar, G., Bunnell, R., und Lemaire, A. (2019). "Extracting features of entertainment products: A guided latent dirichlet allocation approach informed by

the psychology of media consumption." *Journal of Marketing Research*, 56(1), 18–36.

Tseng, M., und Piller, F. (2003). "The customer centric enterprise." *Springer Science & Business Media*.

Tuarob, S., und Tucker, C. (2015). "Quantifying product favorability and extracting notable product features using large scale social media data." *Journal of Computing and Information Science in Engineering*, 15(3).

Tucker, C., und Kim, H. (2011). "Predicting Emerging Product Design Trend by Mining Publicly Available Customer Review Data." *Proceedings of the 18th International Conference on Engineering Design*.

Valencia, S., Cave, R., Kallarackal, K., Seaver, K., Terry, M., und Kane, S. (2023). "The less I type, the better: How AI Language Models can Enhance or Impede Communication for AAC Users." *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.

Van der Aalst, W., Becker, J., Bichler, M., Buhl, H., Dibbern, J., Frank, U., ..., und Zdravkovic, J. (2018). "Views on the past, present, and future of business and information systems engineering." *Business & Information Systems Engineering*, 60, 443-477.

Van der Goot, M., Hafkamp, L., und Dankfort, Z. (2021). "Customer Service Chatbots: A Qualitative Interview Study into the Communication Journey of Customers." *Proceedings oft he 2021 Conference on Chatbot Research and Design*, 190–204.

Vanderheiden, G. (1983). "Non-Conversational Communication Technology Needs of Individuals with Handicaps." *Rehabilitation World*, 7(2), 8-12.

Vayansky, I., und Kumar, S. A. (2020). "A review of topic modeling methods." *Information Systems*, 94.

Venable, J., Pries-Heje, J., und Baskerville, R. (2016). "FEDS: a framework for evaluation in design science research." *European Journal of Information Systems*, 25(1), 77-89.

Vo, A., Nguyen, Q., und Ock, C. (2018). "Opinion–aspect relations in cognizing customer feelings via reviews." *IEEE Access*, 6, 5415-5426.

Vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., und Cleven, A. (2009). "Reconstructing the giant: On the importance of rigour in

documenting the literature search process." *Proceedings of the 17th European Conference on Information Systems*.

Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., und Cleven, A. (2015). "Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research." *Communications of the Association for Information Systems*, 37(1).

Vom Brocke, J., Maaß, W., Buxmann, P., Maedche, A., Leimeister, J., und Pecht, G. (2018). "Future work and enterprise systems." *Business & Information Systems Engineering*, 60, 357-366.

Wang, W., Hernandez, I., Newman, D., He, J., und Bian, J. (2016). "Twitter analysis: Studying US weekly trends in work stress and emotion." *Applied Psychology*, 65(2), 355-378.

Webster, J., und Watson, R. (2002). "Analyzing the past to prepare for the future: Writing a literature review." *MIS Quarterly*, 26(2), 13-23.

Winston, P. (1992). "Artificial Intelligence." *3rd Edition: Addison-Wesley*.

Wörner, J., Konadl, D., Schmid, I., und Leist, S. (2021). "Comparison of Topic Modelling Techniques in Marketing - Results From an Analysis of Distinctive Use Cases." *Proceedings of the 29th European Conference on Information Systems*.

Wörner, J., Konadl, D., Schmid, I., und Leist, S. (2022). "Supporting Product Development by a Trend Analysis Tool Applying Aspect-Based Sentiment Detection." *Proceedings of the 17th International Conference on Design Science Research in Information Systems and Technology*.

Wörner, J., Konadl, D., Leist, S., und Schmid, I. (2023). "MANTRA: A Topic Modeling-Based Tool to Support Automated Trend Analysis on Unstructured Social Media Data." *Proceedings of the 44th International Conference on Information Systems*.

Yan, X., Guo, J., Lan, Y., und Cheng, X. (2013). "A biterm topic model for short texts." *Proceedings of the 22nd International Conference on World Wide Web*, 1445-1456.

Yan, Z., Xing, M., Zhang, D., und Ma, B. (2014). "A Context-Dependent Sentiment Analysis of Online Product Reviews Based on Dependency Relationships." *Proceedings of the 35th International Conference on Information Systems*.

Ye, H., und Kankanhalli, A. (2018). "User service innovation on mobile phone platforms: Investigating impacts of lead userness, toolkit support, and design autonomy." *MIS Quarterly*, 42(1), 165-188.

Yin, R. (2015). "Case study research: Design and methods." 5th Edition. *Thousand Oaks, CA: Sage*.

Zhang, P., Gu, H., Gartrell, M., Yang, D., Ding, X., und Gu, N. (2016). "Group-based Latent Dirichlet Allocation (Group-LDA): Effective audience detection for books in online social media." *Knowledge-Based Systems*, 105, 134-146.

Zhong, N., und Schweidel, D. (2020). "Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model." *Marketing Science*, 39(4), 827-846.