Virtual Assistance in Any Context

A Taxonomy of Design Elements for Domain-Specific Chatbots

Antje Janssen · Jens Passlick · Davinia Rodríguez Cardona · Michael H. Breitner

Abstract Several domain-specific assistants in the form of chatbots have conquered many commercial and private areas. However, there is still a limited level of systematic knowledge of the distinctive characteristics of design elements for chatbots to facilitate development, adoption, implementation, and further research. To close this gap, the paper outlines a taxonomy of design elements for chatbots with 17 dimensions organized into the perspectives intelligence, interaction and context. The conceptually grounded design elements of the taxonomy are used to analyze 103 chatbots from 23 different application domains. Through a clustering-based approach, five chatbot archetypes that currently exist for domain-specific chatbots are identified. The developed taxonomy provides a structure to differentiate and categorize domain-specific chatbots according to archetypal qualities that guide practitioners when taking design decisions. Moreover, the taxonomy serves academics as a foundation for conducting further research on chatbot design while integrating scientific and practical knowledge.

Keywords Chatbot taxonomy · Design elements · Domain-specific chatbots · Human computer interaction

A. Janssen (✉) · J. Passlick · D. Rodríguez Cardona · M. H. Breitner
Information System Institute, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany
e-mail: janssen@iwi.uni-hannover.de

1 Introduction

Chatbots as a form of conversational agents have been developed for different applications. This is due to the evolving of artificial intelligence (AI) and natural language processing (NLP), which is changing the way artificial assistants communicate and interact with people (Nguyen and Sidorova 2018; Jain et al. 2018). By improving text-to-speech and speech-to-text communication, the use of chatbots has become more convenient and common (Bittner et al. 2019). For instance, new smart assistants, such as Cortana, Alexa, Google Assistant and Siri, have been designed with the intention of supporting users in everyday life as voice-activated intelligent personal assistants. The proliferation of these assistants has contributed to the popularity of chatbots worldwide (Di Prospero et al. 2017; Gnewuch et al. 2017; Jain et al. 2018; Diederich et al. 2019). This in turn has also led to an increasing use of domain-specific chatbots (Di Prospero et al. 2017). While Di Prospero et al. (2017) argue that there are similarities that unite all chatbots regardless of application purpose or domain, other scientists claim that there are several aspects in which chatbots differ (Følstad et al. 2019; Bittner et al. 2019; Diederich et al. 2019).

Although some elemental chatbot classification frameworks can be found in scientific literature, the research is dispersed into different thematic axes and research areas. Furthermore, the scientific and practical knowledge about chatbots has also grown in a segregated manner given a shortage of integrative perspectives to support chatbot development and design processes (Følstad and Brandtzaeg 2017; Jain et al. 2018; Piccolo et al. 2018). For instance, most scientific studies today concentrate on particular aspects of chatbots, such as the personality of cognitive chatbots, technical capabilities or their specific application

purpose without providing a holistic view (Gnewuch et al. 2017; Di Prospero et al. 2017). Particularly, for domain-specific chatbots there are no classification schemes that integrate scientific and practical knowledge of chatbot design elements through the differentiation and categorization of domain-specific chatbots according to archetypal qualities.

Previous research has shown that the application domains influence the design of chatbots (Bittner et al. 2019). Therefore, it is necessary to determine whether chatbots differ in their structural representation according to their application domain. The development of a classification scheme of domain-specific chatbots is a fundamental milestone to bridge the research to practice gap by providing guidance to practitioners on design options for the construction of chatbots. It would also supply academics with a foundation for further theory building processes regarding chatbot design and engineering. In view of the foregoing, this paper addresses the following research questions:

> *RQ1* What are conceptually grounded and empirically validated design elements for domain-specific chatbots?
> *RQ2* Which chatbot archetypes can be empirically identified across diverse application domains?

To answer these research questions, we develop a taxonomy of design elements for domain-specific chatbots based on the scientific literature on chatbot design and empirical data. For this purpose, the research approach of this paper follows the taxonomy development framework of Nickerson et al. (2013). After five iterations involving a deductive concept modeling approach based on prior research and the iterative classification of 103 real-world domain-specific chatbots, we present a conceptually and empirically derived taxonomic structure of design elements for domain-specific chatbots. We evaluate the proposed taxonomy in terms of both method and content by means of three focus group discussions. Subsequently, in order to demonstrate the applicability of our taxonomy and to analyze the status quo of current chatbots, we further deploy a cluster analysis and identify five chatbot archetypes. Lastly, our results and outline implications, recommendations, limitations and suggestions for further research are discussed.

2 Overview of Related Chatbot Literature

Chatbots are conversational agents (CA) that enable users to access data and services (Følstad et al. 2019) as well as exchange information by simulating a human conversation (Bittner et al. 2019; Diederich et al. 2019). This conversation is conducted in form of a natural language dialogue about a common topic (Følstad et al. 2019; Diederich et al. 2019). The text-based or speech-based conversation resembles a human-to-human conversation in that the chatbot responds to the input and keeps the conversation going by analyzing single words, phrases and sentence constructions (Nguyen and Sidorova, 2018; Følstad et al. 2019; Diederich et al. 2019). Chatbots are used in different commercial and private situations, such as education, food, travel, finance and mobility, which is also reflected in the application purpose (Følstad et al. 2019).

Diverse scientific articles examine the design and engineering of chatbots from different technical perspectives, e.g., emotional intelligence (e.g., Feine et al. 2019) or anthropomorphic features (Kim et al. 2018), and others focus on the study of chatbots in particular application domains (e.g., Bittner et al. 2019). Gnewuch et al. (2017) provide a basic classification of chatbots based on two dimensions named as "context" and "primary mode of communication". The first dimension categorizes chatbots into general-purpose and domain-specific, while the second dimension arranges them into text-based and speech-based. Speech-based chatbots with general purpose, such as Google Assistant, Cortana, Alexa and Siri are the most widespread and frequently used chatbots (Di Prospero et al. 2017; Kepuska and Bohouta 2018). These voice-activated assistant applications are usually installed directly by the smartphone or smart device manufacturer and offer a large variety of functionalities (Kepuska and Bohouta 2018). Speech-based and domain-specific chatbots, on the other hand, can be found, for instance, as in-vehicle assistants in cars (Diederich et al. 2019). Domain-specific and text-based chatbots interact with humans primarily through text messages about a specific topic (Gnewuch et al. 2017; Diederich et al. 2019). These chatbots undertake different tasks in countless application domains, such as customer support, education, travel, finance and mobility (Følstad et al. 2019), which makes domain-specific chatbots challenging for researchers. In the scientific literature, domain-specific chatbots have been analyzed and classified according to certain criteria. For instance, Maedche et al. (2016) state that differences between chatbots can mainly be abstracted on the interaction and the intelligence levels. Bittner et al. (2019) focus on the development of a nine-dimensional classification for CA used in collaborative work, in which they see the role of the CA as the key dimension. Følstad et al. (2019) perform a chatbot classification by concentrating on two typology dimensions "duration of relation" and "locus of control" while classifying 57 chatbots. Diederich et al. (2019) classify 51 platforms of chatbots into eleven dimensions. Feine et al. (2019) concentrate on building a taxonomy of social cues of CA focused on verbal, visual, auditory and invisible aspects. However, a comprehensive and empirically tested

chatbot taxonomy for domain-specific chatbots, integrating scientific and practical knowledge into different classes or groups, is still missing.

3 Research Approach

3.1 Taxonomy Development Procedure

This paper develops a taxonomy of design elements for chatbots based on scientific literature and empirical data in order to provide a systematic representation of existent scientific knowledge on chatbot design and to develop a deeper understanding of the degree to which domain-specific chatbots integrate conceptually grounded characteristics in practice. Therefore, our taxonomy not only provides a structure to differentiate domain-specific chatbots according to archetypal qualities, but also reflects the extent of their current technological development and allows to identify gaps between research and practice.

To develop our taxonomy, we followed the seven-step framework of Nickerson et al. (2013). The first step begins by the determination of a meta-characteristic, which embodies a superordinate and abstract description of the taxonomy's focus (Nickerson et al. 2013). We defined the meta-characteristic as the design elements for domain-specific chatbots. For the purpose of this analysis, the term "design elements" refers to the distinctive technical, situational and knowledge features that frame the structure of chatbots and act as delimiting factors of the extent to which domain-specific chatbots can maintain a human-like interactive communication process with awareness for and understanding of the discussed topic. The second step consists of determining the objective and subjective ending conditions that define when the iterative development process can be considered as completed. To this end, we adopted all the objective and subjective ending conditions (see Table A.5 in the Appendix, available online via https://doi.org/10.1007/s12599-020-00644-1) suggested by Nickerson et al. (2013, p. 344). In the third step, the process provides the possibility to combine conceptual knowledge and empirical findings either through an empirical-to-conceptual or a conceptual-to-empirical path (Nickerson et al. 2013), which can be applied alternately until all end conditions are met. For the development of our taxonomy for design elements of chatbots, we have adopted a conceptual-to-empirical path as a starting point. Hence in the fourth step and through a deductive concept modeling approach based on prior research, we abstracted a preliminary conceptual taxonomic structure, which we subsequently refined in the fifth step through an iterative analysis of existing domain-specific chatbots. After conducting five iterations, we obtained a taxonomic structure. Subsequently, in the sixth step, we evaluated the taxonomic structure using three focus group discussions. Below we provide a description of the procedure executed in each individual iteration.

3.2 Iteration 1

In the first iteration, we conceptualized an initial collection of dimensions and characteristics through deductive reasoning and extraction, using a set of English written, peer-reviewed scientific articles published in high quality academic journals or conference proceedings belonging to the field of information systems (IS). These articles were identified by means of an explorative literature review. We selected the electronic databases EBSCOhost Business Source Premier, AISeL, ScienceDirect and ACM, which cover relevant literature in both IS and computer science.

To consider various terms used to describe chatbots, we first performed an explorative search to identify relevant keywords. This explorative search formed the basis for the creation of the search string ("chatbot*" OR "conversational agent*" OR "dialog system*" OR "computer user communication*" OR "conversational robot*"), which we used to search for relevant literature via titles and abstracts search that yielded a total of 1076 hits in the four databases which we reduced to 72 articles after excluding the literature that contains our search string, but is unrelated to chatbot design. Additionally, through a full-text revision, we further discarded articles that do not match our conception of "design elements" or provide elemental classification frameworks, narrowing our initial set to 24 relevant scientific articles. This set was further reinforced by means of a backward and forward reference search that led to the identification of four additional scientific articles related to the areas of computer science and software engineering (i.e., Mittal et al. 2016; Saravanan et al. 2017; Wei et al. 2018), as well as language technology (i.e., McTear 2016). This procedure led us to identify a final sample of 28 articles (see Table A.1 in online appendix) that concretely deal with specific technical, situational and knowledge structure features of chatbots.

Consistent with Nickerson et al. (2013), we applied a deductive development approach to derive an initial set of conceptually grounded dimensions and characteristics in line with our meta-characteristic from the identified scientific literature on chatbot design. The taxonomy was developed in a way that all characteristics of a dimension are to be regarded as exclusive. This means that for a chatbot only one characteristic can be true within one dimension. A description of each characteristic from the final taxonomy can be found in Table A.2 of the online appendix.

In line with our working definition of "design elements", the dimensions were allocated to three overarching

perspectives: (i) intelligence (knowledge structure features), (ii) interaction (technical features), and (iii) context (situational features) to facilitate the comprehension of the taxonomy. The adoption of these overarching perspectives is in line with the primary aim for the development of chatbots which is to emulate the process of human communication using AI. Here, the perspectives of intelligence, interaction and context are envisioned as natural attributes of the human communication process. As described by Littlejohn and Foss (2010, p. 8), human communication is "[…] the primary process by which human life is experienced; communication constitutes reality. How we communicate about our experience [Intelligence] helps to shape that experience. The many types of experience are the result of many forms of communication [Interaction]. Our meanings change from one group to another, from one setting to another, and from one time period to another because communication itself is dynamic across situations [Context]". In this respect, the notions of interaction and intelligence are two common levels of abstraction that have been widely used in the IS scientific literature to describe the structural characteristics of chatbots (see Maedche et al. 2016; Knote et al. 2019; Stoeckli et al. 2019). On the other hand, the notion of context has been commonly used to frame the extension of the mediated environment (i.e., general-purpose and domain-specific, see Gnewuch et al. 2017; Diederich et al. 2019) in which the chatbot is used and hence has an influence on the chatbot construction (Knote et al. 2018).

3.3 Iteration 2

In this iteration we chose to follow an empirical approach to substantiate our conceptual taxonomic structure (T_1) (Nickerson et al. 2013). We distributed the empirical investigation of all chatbots among the authors. To determine the characteristics of a sample of real-world chatbots, we used the definitions provided in Table A.2 of the online appendix and jointly determined selection criteria for non-self-explanatory dimensions. This empirical chatbot classification was achieved primarily through targeted interaction with the chatbot and secondarily partly through available videos and reports, which we also consulted. To this end, we classified an initial sample of 12 chatbot interfaces (see Table A.3 in online appendix) within the taxonomic structure (T_1). This sample was composed of the most popular chatbots in the areas of communication, cryptocurrency, analytics and education according to the ranking provided by the third-party database BotList.co (2019).

Within this iteration, we removed all dimensions that were important from a conceptual point of view but could not be empirically determined from the outside by testing a

chatbot, as detailed in Fig. A.1 in the online appendix. This includes, e.g., *type of artificial intelligent system (AIS)*, *memory*, and *sequentiality of process structure*. After reviewing the aforementioned chatbots, we systematically readjusted our conceptual taxonomy by (i) removing the characteristics that were not empirically observable in any of the analyzed objects; (ii) merging redundant characteristics (i.e., *conversational chatbots* and *interactive chatbots*, (iii) disjoining characteristics that showed to have individual descriptive power (i.e., the compound characteristic *daily life and family* was divided into the individual characteristics *daily life* and *family*) and; (iv) adding the new characteristics identified during the examination (i.e., *utility* into the dimension *motivation for chatbot use*). Additional to the mentioned adjustments, we proceeded to merge the dimensions of *personality processing* and *sentiment detection* because of their overlapping nature, as well as to add to the taxonomy a new empirically observed dimension named *additional human support* to reflect the interactive design of those chatbots that enable a connection of the digital and physical world by means of integrating human support into its collection of interactive capabilities.

3.4 Iteration 3

To obtain a sample composed of chatbots from different application domains and platforms, we decided to search for a database that allows us to include chatbots from multiple domains. Accordingly, we analyzed five different chatbot databases (botlist.co, chatbottle.co, chatbots.org, 50bots.com, botfinder.io). The most suitable database for our purposes turned out to be the database chatbots.org, given that it allows to filter a total of 1194 chatbots according to 27 application domains. This feature enabled us to view 10% of the chatbots from each area (chatbots.org, 2019). In this iteration we categorized a collection of 66 chatbots (see Table A.3 in online appendix) composed by the ten percent of the total chatbots listed on the third-party database chatbots.org (2019) within the areas finance and legal ($n = 15$), social ($n = 11$), home and living ($n = 5$), body health ($n = 5$), government ($n = 5$), education ($n = 5$), electronics and hardware ($n = 4$), career and education ($n = 3$), cooking ($n = 3$), children ($n = 2$), environmental ($n = 2$), fashion ($n = 2$), sport ($n = 2$), culture ($n = 1$) and beauty ($n = 1$).

During the development of this iteration, we merged the characteristics of *crowd setting* and *two or more humans* of the dimension *number of participants* due to their overlapping nature. Additionally, we identified the feature *only rule-based knowledge* as additional descriptive characteristic of the *intelligence quotient* dimension; likewise the characteristics of *advice* and *customer support* were added

to the dimension *motivation for chatbot use* to enhance its descriptive power.

3.5 Iteration 4

Subsequently, we analyzed additional 13 chatbots relating to the areas of telecommunication and utilities ($n = 6$), mobility ($n = 2$), mental and spirituality ($n = 2$), news and gossip ($n = 2$), and leisure ($n = 1$) from the database Chatbots.org (2019). In this iteration, we added a new dimension named *service provider integration,* consisting of the following characteristics: *none, single integration* and *multiple integration,* to describe the capacity of different chatbots to integrate supplementary services.

3.6 Iteration 5

As the ending conditions were not fulfilled in the last iteration due to the addition of one dimension, we proceeded then to carry out a further empirical iteration path. In this iteration, we integrated into the taxonomy an additional subset of chatbots interfaces consisting of in total 12 chatbots of the areas of travel ($n = 5$), TV, visual entertainment, creation and gaming ($n = 4$) and trade ($n = 3$) indexed as well in Chatbots.org (2019) database. As a result of this iteration, in the dimension *motivation for chatbot use,* we changed the name of the characteristic *work support* to *work and career.* Likewise, to enhance the explanatory power of the taxonomy, we also modified the name of the characteristic *multiple* to *text understanding plus further elements* in the *intelligence quotient* dimension.

3.7 Evaluation

To evaluate the taxonomy, we considered and answered three questions: "who", "what" and "how" within the framework for taxonomy evaluation by Szopinski et al. (2019). With regard to the subject of evaluation (the "who"), we decided to choose individuals who had no previous contact with the development of the taxonomy. For the evaluation of the taxonomy in terms of both method and content, we involved three sets of participants within three separated focus group discussions: practitioners with domain knowledge about chatbots, academics with methodological knowledge about taxonomy development, and academics with chatbot domain knowledge. This heterogeneity is supposed to avoid inconsistencies and to ensure a broad applicability and usefulness for academic and practical purposes. With regard to the object of evaluation (the "what"), we determined "the design of a chatbot" as the real-world problem to be investigated. Focus group discussions were chosen as the method of

evaluation (the "how"), because hereby the taxonomy can be analyzed jointly and new thoughts and ideas can be discussed.

As mentioned above, we conducted three focus group discussions, each of which began with a presentation of the taxonomy and the delivery of a sheet of paper with the taxonomy and all definitions. Then a worksheet was presented in which each participant was asked, as a first step, to note on an individual basis which perspectives, dimensions and characteristics should be deleted, added, merged, relocated or modified in wording, and their rationale for each proposed change. This was followed by a discussion on the fulfillment of the subjective ending conditions (see Table A.5 in online appendix) and the criteria of comprehensiveness, understandability, wording, and extendibility for the individual dimensions and characteristics explored by Szopinski et al. (2019).

Group 1 consisted of five participants with an academic background, all with methodological knowledge and two with chatbot domain knowledge. As a result of the discussion, which lasted 40 min, the characteristic *text understanding and further abilities* and the dimension *intelligence quotient* were renamed. The dimension *socio-emotional behavior* was particularly discussed, since emotional intelligence is currently gaining importance. This dimension was assigned to the intelligence perspective. The descriptions of the dimensions and characteristics were seen as appropriate and understandable.

Group 2 consisted of three participants with doctoral and post-doctoral backgrounds, one with strong methodological taxonomy knowledge and two with knowledge about the introduction of chatbots within the context of a research project on the development of a digital assistant for e-learning. Within this discussion, which lasted 105 min, we debated the results of the first group and placed a special emphasis on the evaluation of the definitions of dimensions and characteristics. The results were to rename dimension D_5 to *service integration*, to rewrite the corresponding definition, to rename D_{14} to *relation duration* and to rename $C_{15,1}$ to *e-customer service*. Furthermore, it was suggested to change the order of the characteristics at D_4, D_{11}, D_{13} and D_{14}.

The third focus group discussion was held in an industrial company with four participants, each with previous experience in the development and implementation of domain-specific chatbots. The discussion lasted 75 min and was aimed at evaluating the taxonomy in terms of its comprehensibility for practitioners as well as the potential applicability and usefulness of the taxonomy in practice. Participants reported that the use of the taxonomy would provide a great added value before and during the development of chatbots. It helps them as an overview, as it can be used as a template for guiding the fundamental

questions that every chatbot developer team should ask itself before starting the process of chatbot design, such as whether a chatbot should be better embodied or disembodied, whether socio-emotional behavior should be incorporated into the chatbot architecture, or what role a chatbot should play within the intended interaction with users. Furthermore, the taxonomy was considered to provide a useful synthesis of design elements that is independent of chatbot design providers and industries. Participants also stated that it would not only be helpful for them to classify their own chatbots in the taxonomy, but also to use this classification to analyze chatbots of competitors in a structured way, which in turn helps as a basis for decision-making.

Since no more dimensions or characteristics were merged, split, added or eliminated during the focus group discussion with group 3, the ending conditions have been fulfilled as shown in Table A.5 of the online appendix; consequently, the taxonomy development process ended after six iterations. The taxonomy development over iterations is shown in Fig. A.1 in the online appendix.

4 Chatbot Taxonomy

The overall results of the present taxonomy-based analysis show that chatbots can be classified and categorized on the basis of three taxonomy layers (see Table 1). Layer 1 comprises the types of design elements, which is divided into three perspectives. Layer 2 comprises the design elements in the form of 17 dimensions. Layer 3 summarizes the conceptually grounded characteristics of the design elements. The division into three perspectives aims to increase the comprehensibility and usability of the taxonomy. In each perspective there are between five and seven dimensions. This fulfils the '7+-2 rule' of Miller (1956), which describes that a person can only grasp a certain amount of information.

4.1 Intelligence

Chaves and Gerosa (2019) describe intelligence as the ability of a chatbot to participate in a dialogue with an awareness of the discussed topic, while Jain et al. (2018) believe that the intelligence of a chatbot can be also deduced from its ability to proactively ask suitable questions and to involve the participant in a meaningful and human-like dialogue. At a holistic level, in line with the proposed final taxonomy, the design elements for chatbots related to specific intelligence features can be described using 15 characteristics, which in turn can be categorized into the following 5 dimensions: The *intelligence framework* D_1 depicts the underlying cognitive system design

delimiting the technical principles under which a chatbot communicates, processes information, and/or selects an action or response (Saravanan et al. 2017; Knote et al. 2018; Diederich et al. 2019). The *intelligence quotient* D_2 indicates whether a chatbot is primarily based on simple 'if-then' pattern-matching rules, whether it understands textual input or whether it has the capability to enhance its responses through math calculation, inference or photo recognition etc. (Wei et al. 2018; Knote et al. 2018). While the intelligence framework classifies the entire conversation process, the intelligence quotient dimension describes the intelligence in evaluating a single response. The *personality processing* D_3 characterizes the capacity of the chatbot to emphatically tailor its notation responses to the specific personality and mood of the user by identifying the personality trait of the counterpart (Di Prospero et al. 2017; Yorita et al. 2019). The chatbot adapts to the real-time identified personality trait of the user (Yorita et al. 2019). The *socio-emotional behavior* D_4 characterizes the resonance capacity of the chatbot to show affection or empathy towards the individual needs and immediate feedback of the user, which the user reveals through resonating emotions within a dialogue (Bittner et al. 2019; Yalçın 2019). This is expressed by "text-based linguistic emotional recognition and expression" (Yalçın 2019, p. 6). While a distinctive personality processing of a chatbot can be recognized mainly by the adaptation of its language, a socio-emotional behavior is shown by the alignment of the chatbot's answers to the user's mood. The *service integration* D_5 states the number of further integrated services enabled by the chatbot, e.g., retrieving information from external data sources (den Boer 2017).

4.2 Interaction

Kiousis (2002, p. 372) defines interaction as *"the degree to which a communication technology can create a mediated environment in which participants can communicate […], both synchronously and asynchronously, and participate in reciprocal message exchanges."* In line with this, chatbots allow people to interact with computer systems via written and/or spoken natural language with the aim of leading the interaction as naturally as possible to resemble a face-to-face dialogue (Diederich et al. 2019). However, the key design challenge at this abstraction level is to create natural interactions with human-like elements to support the interaction experience (Bittner et al. 2019; Gnewuch et al. 2018a, b). By integrating conceptual and empirical insights, we identified 17 characteristics of interactive features which enable chatbots to interact with their users. These characteristics can be represented by means of the next 7 dimensions: The *multimodality* D_6 points toward the capacity of the chatbot to receive input and respond

Table 1 Final taxonomy of design elements for chatbots with all dimensions D_i, characteristics $C_{i,j}$ and perspectives

Layer 1: Perspective	Layer 2: Dimensions D_i	Layer 3: Characteristics $C_{i,j}$	
Intelligence	D_1 Intelligence framework	$C_{1,1}$ Rule-based system	$C_{1,2}$ Utility-based system
		$C_{1,3}$ Model-based system	$C_{1,4}$ Goal-based system
		$C_{1,5}$ Self-learning system	
	D_2 Intelligence quotient	$C_{2,1}$ Only rule-based knowledge	$C_{2,2}$ Text understanding
		$C_{2,3}$ Text understanding and further abilities	
	D_3 Personality processing	$C_{3,1}$ Principal self	$C_{3,2}$ Adaptive self
	D_4 Socio-emotional behavior	$C_{4,1}$ Not present	$C_{4,2}$ Present
	D_5 Service integration	$C_{5,1}$ None	$C_{5,2}$ Single integration
		$C_{5,3}$ Multiple integration	
Interaction	D_6 Multimodality	$C_{6,1}$ Unidirectional	$C_{6,2}$ Bidirectional
	D_7 Interaction classification	$C_{7,1}$ Graphical	$C_{7,2}$ Interactive
	D_8 Interface personification	$C_{8,1}$ Disembodied	$C_{8,2}$ Embodied
	D_9 User assistance design	$C_{9,1}$ Reactive assistance	$C_{9,2}$ Proactive assistance
	D_{10} Number of participants	$C_{10,1}$ Individual human participant	$C_{10,2}$ Two or more human participants
	D_{11} Additional human support	$C_{11,1}$ No	$C_{11,2}$ Yes
	D_{12} Front-end user interface channel	$C_{12,1}$ App	$C_{12,2}$ Collaboration and communication tools
		$C_{12,3}$ Social media	$C_{12,4}$ Website
		$C_{12,5}$ Multiple	
Context	D_{13} Chatbot role	$C_{13,1}$ Facilitator	$C_{13,2}$ Peer
		$C_{13,3}$ Expert	
	D_{14} Relation duration	$C_{14,1}$ Short-term relation	$C_{14,2}$ Long-term relation
	D_{15} Application domain	$C_{15,1}$ E-customer service	$C_{15,2}$ Daily life
		$C_{15,3}$ E-commerce	$C_{15,4}$ E-learning
		$C_{15,5}$ Finance	$C_{15,6}$ Work and career
	D_{16} Collaboration goal	$C_{16,1}$ Non goal-oriented	$C_{16,2}$ Goal-oriented
	D_{17} Motivation for chatbot use	$C_{17,1}$ Productivity	$C_{17,2}$ Entertainment
		$C_{17,3}$ Social/relational	$C_{17,4}$ Utility

through only one or various interaction modalities or communication channels, e.g., text, voice, facial expression, etc. (Knote et al. 2018). The *interaction classification* D_7 typifies the human-computer interaction (HCI) method used by the chatbot (den Boer 2017). The *interface personification* D_8 illustrates the extent to which a chatbot incorporates visual or physical anthropomorphic or personification features in the form of static, animated or reactive avatars (Knijnenburg and Willemsen 2016; Bittner et al. 2019). The *user assistance design* D_9 denotes whether the chatbot interacts with the user in a proactive or reactive way (Sarikaya 2017; Jain et al. 2018; Følstad et al. 2019). The *number of participants* D_{10} identifies whether one or more humans are involved in the interaction (Bittner et al. 2019; Mittal et al. 2016). The *additional human support* D_{11} specifies whether or not the chatbot offers the possibility to contact a human agent in case of open questions

(Zumstein and Hundertmark 2017). The *front-end user interface channel* D_{12} indicates the respective platform which the chatbot has been integrated into. (Sarikaya 2017; Følstad et al. 2019).

4.3 Context

Context in general is the totality of all implicit and explicit situational information about people, objects, time and location within an interaction that can be used to describe a situation (Abowd et al. 1999; Kim et al. 2018). The context shows whether and in which domain the chatbot operates (Gnewuch et al. 2017; Diederich et al. 2019). The characteristics of the environment in which interactions takes place can be classified and categorized into 17 characteristics grouped into the 5 dimensions described below: The *chatbot role* D_{13} designates the role that the chatbot plays

during the interaction (Bittner et al. 2019). The *relation duration* D_{14} describes the ability of the chatbot to remember information from previous conversations to influence future interactions (Wei et al. 2018; Følstad et al. 2019). The *application domain* D_{15} specifies the primary application purpose for which the chatbot has been designed (Zumstein and Hundertmark 2017; Knote et al. 2018). The *collaboration goal* D_{16} determines whether or not the chatbot helps the user to accomplish a common goal or task (Bittner et al. 2019). The *motivation for chatbot use* D_{17} identifies the primary extrinsic motivation for the chatbot use from the user perspective (Deci and Ryan 2000; Brandtzaeg and Følstad 2017).

5 Taxonomy Application

5.1 Distributions of the Analyzed Chatbots

With the 103 analyzed chatbots, which were already used during the creation of the taxonomy, it is possible to show an application of the taxonomy. The classification process was based on the information detailed by the respective chatbot directories and equally divided among the authors. If the assignments were not clear, the respective chatbot was discussed by the entire team of authors. Furthermore, we performed an inter-coder reliability test by classifying ten randomly selected chatbots again. This step was performed independently by all authors. As a result, we were able to calculate the quality of the agreement with the Fleiss' (1971) kappa coefficient which is 0.63. Based on Landis and Koch (1977), a "substantial" agreement can be assumed for this value. We can therefore assume that there was no bias caused by different coders. An overview of the results of the classification process by perspective is shown in Fig. 1.

In the intelligence framework dimension, which is assigned to the perspective *intelligence*, the majority of the chatbots investigated tend to function with a less intelligent rule-based behavior (73%). This can also be seen in the dimension personality processing, where principal self (96%) strongly dominates the rather complex property adaptive self. The presence of socio-emotional behavior was only found in a few chatbots (12%). Most of the examined chatbots use only one service (59%), while 18% integrate multiple services. In the perspective *Interaction* and the dimension number of participants almost all chatbots assume an individual partner (96%). About four fifths (79%) react to the user, while one fifth can also proactively send information to the user. Most of the analyzed chatbots are not embodied (71%) and were categorized as interactive chatbots (77%). A connection to humans can be offered by 20% of the chatbots. In the perspective *Context*

there are some dimensions where one characteristic clearly dominates, but also dimensions that are evenly distributed. Many chatbots are designed for a short-term relationship (84%). 47% of the chatbots have their application domain in the daily life category, whereas 21% are used for customer service issues. Two thirds of the chatbots work towards a specific goal (77%). Furthermore, we identified utility (45%) and entertainment (29%) as the most common motivations for using a chatbot.

5.2 Chatbot Archetypes

Based on the chatbots examined, a further step is to determine whether certain archetypes can be identified. To this end, we applied the Ward (1963) algorithm to the data set. The Ward algorithm is often used for practical applications and is a hierarchical cluster algorithm which calculates the distances between all elements (Gimpel et al. 2017). In contrast to non-hierarchical partitioning algorithms such as the K-means algorithm, this has the advantage that it can be used without having to predefine a certain number of clusters. The combination of hierarchical algorithms like the Wards's (1963) and the k-means algorithm is a recommended approach to exploit advantages of both algorithm types (Balijepally et al. 2011). We used the Sokal and Michener (1958) matching coefficient to determine the distances between the clusters. After running the Ward (1963) algorithm, the question arises which number of clusters is appropriate for the further analysis. Gimpel et al. (2017) have shown that quite different measures can be applied to answer this question, but the number of resulting clusters tend to vary depending on the measure applied. Our data set also shows the completely different results of the measures. The results of the algorithms can be seen in the online appendix Table A.4. Hence, we have graphically analyzed the dendrogram resulting from the Ward (1963) algorithm (Täuscher and Laudien 2018). The dendrogram is shown in Fig. 2. At the height of more than 3 the first splitting is visible. The next splits follow at approx. 1.98 and approx. 1.95. After that the splits are relatively close together. The distance of the groups here is smaller than 1.5. Consequently, we have examined the possibilities of two and five groups.

For the aim of identifying the groups, both hierarchical cluster algorithms and partitioning algorithms can be used. A partitioning algorithm suitable for cluster analysis of taxonomies is the k-means (Täuscher and Laudien 2018). We applied the k-means algorithm for two and five groups to our data set. After examining the results, we concluded that the division into five groups provides more plausible results than a division into two groups. Hence, we analyzed the clusters for five groups more closely.
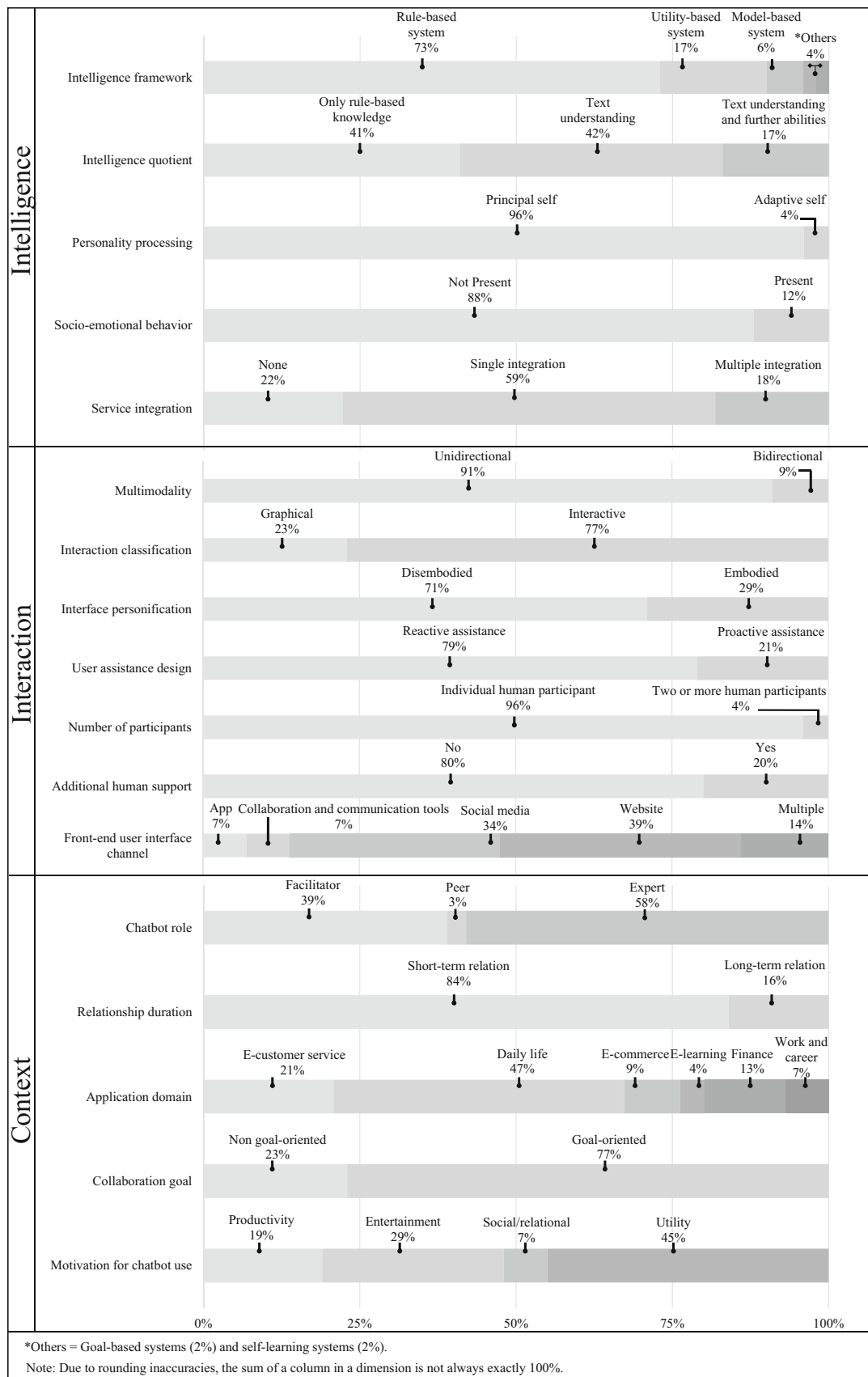
Fig. 1 Distribution of characteristics per perspective
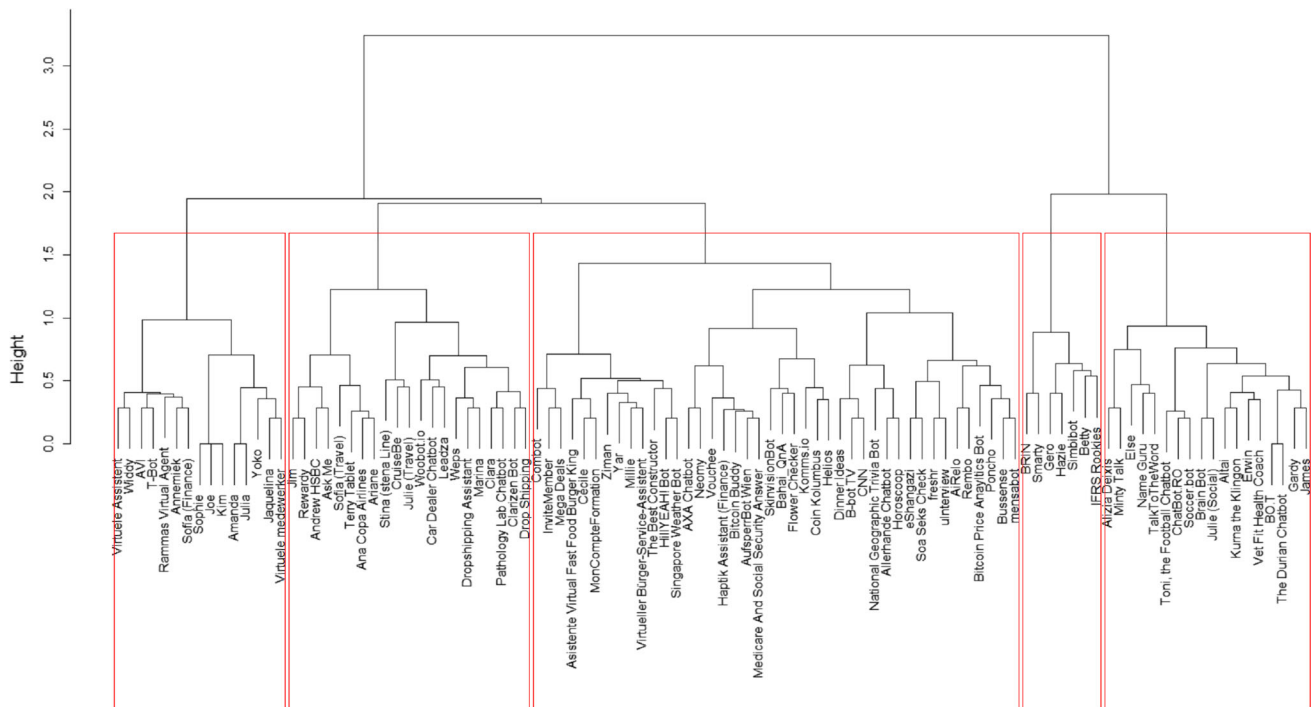
Fig. 2 Result of the Ward clustering visualized by a dendrogram

Table 2 shows the distributions of the characteristics in the five archetypes. We have named the five archetypes *goal-oriented daily chatbot* (A), *non goal-oriented daily chatbot* (B), *utility facilitator chatbot* (C), *utility expert chatbot* (D) and *relationship-oriented chatbot* (E) to represent the focus of each archetype. These five archetypes are intended to help developers to identify the relevant characteristics and derive fields of action based on their problem and area of application. Already in the first dimension intelligence framework, a clear difference between the archetype E and the other four archetypes is recognizable. While 44% of chatbots in archetype E have the ability to adapt to the end-user's behavior during conversation (adaptive self), all other archetypes do not have this ability. The chatbots in archetype E (e.g., Smarty Simple Mind chatbot) are characterized by 89% showing a high socio-emotional behavior and all chatbots being proactive in the human-computer dialogue by asking specific context-relevant questions. This can be associated with the AI-based emotional intelligence of chatbots described in the literature (e.g., Feine et al. 2019).

Here 56% aim at establishing a long-term relationship, where the emotional bond can be helpful. There are two archetypes that primarily unite daily life chatbots (A = 83%; B = 95%). The main difference between these two archetypes is that most of the chatbots in archetype A are goal-oriented (96%), while 89% of the chatbots in archetype B do not have a main goal. While half of A

(54%) (e.g., Dinner ideas chatbot) act as a facilitator and guide to help users reach a certain goal, most of B (79%) (e.g., The Durian chatbot) are experts. It can be observed that the chatbots in archetype A pursue a goal by integrating services, whereas the chatbots in archetype B convince with their own skills like text understanding and other abilities such as photo recognition or math calculation, which is often the purpose of using the chatbot. Archetype C and D mainly include chatbots that pursue a utility purpose (D = 93%; C = 55%), or a productivity purpose (C = 32%). Archetype C (e.g., Pathology Lab chatbot) mainly consists of chatbots who act as facilitators and mostly have rule-based knowledge only. The chatbots in archetype D (e.g., Neomy chatbot) are slightly more interactive in that 86% communicate interactively, of which three-quarters have the ability to read and evaluate conversations while acting as an expert.

6 General Discussion

6.1 Theoretical and Practical Implications

We have developed a domain spanning taxonomy based on the scientific literature and empirical data which allows to classify chatbots according to 17 dimensions and 49 characteristics (i.e., design elements) organized into the perspectives intelligence, interaction and context. While

Table 2 Results of the cluster analysis

	Label	Goal-oriented daily chatbot	Non goal-oriented daily chatbot	Utility facilitator chatbot	Utility expert chatbot	Relationship-oriented chatbot
	n	24	19	22	29	9
	Archetype	A	B	C	D	E
Intelligence	**Intelligence framework**					
	Rule-based systems	88%	79%	77%	76%	
	Utility-based systems	13%	16%	18%	17%	33%
	Model-based systems		5%	5%		44%
	Goal-based systems				7%	
	Self-learning systems					22%
	Intelligence quotient					
	Only rule-based knowledge	46%	11%	95%	24%	11%
	Text understanding	38%	47%		72%	44%
	Text understanding and further abilities	17%	42%	5%	3%	44%
	Personality processing					
	Principal self	100%	100%	100%	100%	56%
	Adaptive self					44%
	Socio-emotional behavior					
	None/low	96%	84%	100%	100%	11%
	High	4%	16%			89%
	Service integration					
	None		68%	32%	7%	11%
	Single	92%	5%	45%	79%	56%
	Multiple	8%	26%	23%	14%	33%
Interaction	**Multimodality**					
	Unidirectional	100%	89%	86%	97%	67%
	Bidirectional		11%	14%	3%	33%
	Interaction classification					
	Graphical	8%	5%	64%	14%	33%
	Interactive	92%	95%	36%	86%	67%
	Interface personification					
	Disembodied	79%	68%	86%	72%	11%
	Embodied	21%	32%	14%	28%	89%
	User assistance design					
	Reactive assistance	75%	95%	91%	86%	
	Proactive assistance	25%	5%	9%	14%	100%
	Number of participants					
	Individual human partner	96%	100%	95%	93%	100%
	Two or more human participants	4%		5%	7%	
	Additional human support					
	No	100%	100%	45%	86%	44%
	Yes			55%	14%	56%
	Front-end user interface					
	App	8%	11%		3%	22%
	Social media	63%	32%	23%	21%	33%
	Collaboration and communication	8%		5%	14%	
	Website	13%	26%	55%	59%	33%
	Multiple	8%	32%	18%	3%	11%
Context	**Chatbot role**					
	Facilitator	54%	5%	77%	21%	33%
	Peer		16%			
	Expert	46%	79%	23%	79%	67%
	Relation duration					
	Short-term relation	75%	89%	86%	100%	44%
	Long-term relation	25%	11%	14%		56%
	Application domain					
	E-customer service	4%		27%	48%	11%
	Daily life	83%	95%	36%		22%
	E-commerce	4%		27%	7%	
	E-learning		5%			33%
	Finance	4%		5%	38%	
	Work and career	4%		5%	7%	33%
	Collaboration goal					
	Non goal-oriented	4%	89%	5%		56%
	Goal-oriented	96%	11%	95%	100%	44%
	Motivation for chatbot use					
	Productivity	17%	11%	32%	7%	56%
	Entertainment	38%	84%	14%		22%
	Social/relational	25%	5%			
	Utility	21%		55%	93%	22%

| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |

Due to rounding inaccuracies, the sum of a column in a dimension is not always exactly 100%

other scientists have so far only focused on the classification of diverse types of CA within specific domains, such as collaborative work (Bittner et al. 2019), CA platforms (Diederich et al. 2019) or customer service (Gnewuch et al. 2017), we have developed a domain-spanning taxonomy. At the practical level, our examination of the degree and

frequency in which the characteristics are distributed throughout the taxonomy dimensions (Table 2) provides insights into the current state of technological development of chatbots that can help practitioners with the conception of chatbots. Likewise the taxonomy, as well as the five chatbot archetypes identified through it provide practitioners with a blueprint of the different design decisions that can be made to develop a domain-specific chatbot. Therefore, as reported by practitioners (Sect. 3), the developed taxonomy can act as a supporting tool to systematically derive design decisions based on the 17 dimensions. In the same manner, the five archetypes help practitioners to streamline the chatbot design process by categorizing a chatbot which they plan to develop according to sets of features described in Sect. 5 and adopting the typical archetype design elements, which in turn further facilitates decision making.

Nevertheless, each of the identified design elements in the taxonomy has its relative advantages and disadvantages, therefore the most suitable combination of the design elements depends on case-specific conditions such as the user target group(s), the boundaries of the project, e.g., financial or other resources, and underlying value proposition behind the particular chatbot to be developed. Accordingly, there is not a one-size-fits-all approach for the design of a chatbot, but the analysis of empirically identified chatbot archetypes embodying real-world combinations of design elements serves practitioners by illustrating overall directions for the chatbot to guide and simplify the decision process. On this basis, e.g., the underlying cognitive system design delimiting the capacity of the chatbot to process user utterances should match the chatbot application purpose. When the purpose of a chatbot is to support the end-users with concise and predefined responses to common questions in a non-complex application domain, the chatbot can be configured as rule-based system using artificial intelligence markup language (AIML) response templates as a basis (Nuruzzaman and Hussain 2018). As shown in our empirical analysis, four of the five identified archetypes were predominantly designed as rule-based systems. The disadvantages linked to this design decision are the inability of such a chatbot to effectively respond to user utterances that are out-of-domain or that contain syntactic or lexical variations such as spelling errors or colloquial language (McTear 2018; Nuruzzaman and Hussain 2018). As a counteracting measure to these limitations on NLP capabilities, a chatbot can integrate graphical elements, such as predefined buttons for selection to interact with users, which enhances interaction efficiency not only by reducing typing effort, but also the input errors (Jain et al. 2018). However, up to this point, the chatbot is yet not able to rationalize textual input or to identify the context during the interaction with the user

(Nuruzzaman and Hussain 2018). To achieve this, the intelligence quotient, instead of only being driven by rule-based knowledge, should incorporate text understanding capabilities through the use of semantics, NLP and deep neural networks (DRN) (Nuruzzaman and Hussain 2018). As can been seen in the configuration of the archetype C, chatbots acting as facilitators can be designed with a rule-base intelligence quotient, however, chatbots with the role of experts, as in the case of archetype D, should incorporate sufficient domain-specific linguistic knowledge to provide more suitable and versatile human-like responses related to a specific subject of a domain (Li et al. 2018).

Taking a look at the developed 17 dimensions, there are large differences in the degree and frequency in which characteristics are distributed within dimensions (Fig. 1). This not only shows the current state of technological development of chatbots, but also allows practitioners and researchers to identify further lines of research, technological trends and areas of improvement for existing chatbots (e.g., within the dimensions of intelligence quotient and multimodality). Additional areas of improvement for existing chatbots can be found in the dimensions of socio-emotional behavior and personality processing, where most of the analyzed chatbots present limited capabilities, while research has progressed significantly. Users expect chatbots to have human-like communication skills, which implies that not only chatbot personality, but also conversational style and socio-emotional skillset need to be adapted to the domain, end-user, and platform for which a chatbot has been designed (Jain et al. 2018; Piccolo et al. 2018). The area of emotional processing is currently being studied in the scientific community from various perspectives. Hu et al. (2018) observed that a passionate and empathetic tone, compared to six other tones of a chatbot, increases the user experience, while Yorita et al. (2019) concluded that the reaction of the user depends strongly on the design of the socio-emotional skillset of the chatbot. Therefore, various researchers (e.g., Yalçın 2019; Yorita et al. 2019; Rouast et al. 2019) focus on automatic affect recognition of chatbot user's personality and emotional state to adapt to it. The two design elements socio-emotional behavior and personality processing are of great importance for the user experience and acceptance of chatbots (Jain et al. 2018). These design elements are particularly essential in domains where emotional awareness is highly important due to the sensitivity of information being disclosed or because the emotions and feelings of the user are a fundamental axis for the interaction. Under these conditions, the integration of socio-emotional behavior and personality processing design principles can lead to a decisive competitive advantage, which is particularly important for practitioners and ultimately for chatbot developers.

6.2 Limitations and Further Research

The limitations of this study are mainly related to the subjective nature of the selection procedure of the dimensions and ending conditions as well as to the reliance of the taxonomy's explanatory power on the comprehensiveness, essence and maturity of the theoretical and empirical knowledge underpinning it. However, these limitations, which we will explain in more detail below, also give rise to many open research directions (RD) which can be addressed by IS and HCI researchers in the future.

To empirically determine the conceptually developed taxonomy, we used two chatbot databases. Because chatbot developers or people responsible for chatbots are free to decide to publish information about a developed chatbot in one of the two databases we consulted, the sample is subject to a certain self-selection bias according to Olteanu et al. (2019). This can also be related to the varying number of chatbots per application area. We chatted with chatbots that are open to the public. Chatbots that are, e.g., exclusively intended for internal use within companies were excluded. However, there is no indication in the scientific literature that we have not considered certain aspects or application areas. Further research can adapt this taxonomy to chatbots that are exclusively for internal usage, e.g., within a company (RD1). We suggest to carry out this chatbot analysis regularly in the future, as the taxonomy can be used to depict precise trends, e.g., in the direction of emotion processing (RD2).

We currently cannot make any statements about the success of the reviewed chatbots. However, these limitations can be mitigated by incorporating insights from qualitative interviews at the users' and experts' level. Further research needs to discuss characteristics which describe a successful chatbot (RD3). In addition, the determination of key performance measures is necessary to make this success quantifiable (RD4).

Although our final sample integrates chatbot interfaces belonging to 23 different application domains, it is not possible to affirm completeness of the taxonomy since the technological possibilities are subject to fast change. In a few years, access to even more data sources will enable much stronger individualization. This can then lead to further dimensions. Therefore, it is recommendable to regularly repeat the empirical examination of chatbots to enhance the integration of conceptual and technological developments into the taxonomy (RD5). Further dimensions show the difference between research and practice. An example of this is the field of socio-emotional behavior which was discussed in 6.1. Additional research can investigate whether scientific literature is perhaps already ahead of practice, sets other priorities that lack practical viability or is largely not relevant in practice (RD6).

Not all dimensions discussed at the conceptual level in Iteration 1 can be empirically surveyed (Table A.1) which does not mean that they do not exist in practice, e.g., *memory* and *sequentiality of process structure*. This circumstance shows that there is a difference between chatbot design elements discussed in scientific research (Iteration 1) and dimensions observed in practice (Iterations 2–5). However, this does not mean that the eliminated dimensions are less relevant. Further research can use additional methods e.g., conducting interviews with chatbot developers, to obtain further expert information to, e.g., determine design principles and frameworks for the development of long-term advance memory capabilities on domain-specific chatbots (RD7). We further suggest developing a taxonomy from a chatbot developer's perspective providing valuable insights on practice-relevant chatbot characteristics (RD8) or compare the result of this study against the insights achieving from a taxonomy emerging from an inductive approach using proof-of-concepts developed by theory-tool-makers in scientific literature, as the real-world object to be examined (RD9).

We conducted three focus group discussions to evaluate the taxonomy, in which the twelve participants first evaluated the taxonomy on an individual basis and then discussed their results with other participants. Since mutual influence cannot be completely discarded, a quantitative survey can also be used to evaluate the taxonomy (RD10). Likewise, a further evaluation of the usefulness of the identified five archetypes in terms of applicability and identifiability in practical settings can provide additional insights for the development of design principles using the identified archetypes as guidance (RD11). In the future, the underlying business model of the chatbots in each archetype can be re-examined to assess the usefulness of the archetype beyond a merely IS perspective (RD12).

Finally, we recommend to investigate the factors driving the technological development of chatbots at the user, organizational and industry level, as well as to reinforce the investigation on chatbot implementation and adoption, for which the dimensions of the proposed taxonomy can provide a common framework for chatbot developers and practitioners to formulate design principles which guide the further development of chatbots (RD13).

7 Conclusions

We have created a taxonomy following the framework of Nickerson et al. (2013) to increase the existent knowledge and conceptual understanding of the distinctive design elements of chatbots across diverse application domains. The overall results of the present taxonomy study indicate that the design elements of chatbots can be classified and

categorized into 17 dimensions and a total of 49 characteristics. Our taxonomy analysis shows that the majority of the analyzed chatbots integrate by far not all the technical possibilities from an intelligence and interaction perspective.

By using the aforementioned taxonomy to analyze 103 chatbots from 23 different application domains, we provided a holistic representation of the degree to which real-world examples of chatbots integrate conceptually grounded design elements, which in turn enabled us to identify five archetypes of chatbots by means of a clustering analysis. Such a classification can be used to provide an integrative base of knowledge for further theory building processes and to guide chatbot developers when designing domain-specific chatbots.

References

Abowd GD, Dey AK, Brown PJ, Davies N, Smith M, Steggles P (1999) Towards a better understanding of context and context-awareness. In: International symposium on handheld and ubiquitous computing. pp 304–307

Balijepally V, Mangalaraj G, Iyengar K (2011) Are we wielding this hammer correctly? A reflective review of the application of cluster analysis in information systems research. J Assoc Inf Syst 12(5):375–413

Bittner E, Oeste-Reiß S, Leimeister JM (2019) Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work. In: HICCS 52 proceedings. Maui, pp 284–293

Botlist.co (2019) https://botlist.co/. Accessed 11 July 2019

Brandtzaeg PB, Følstad A (2017) Why people use chatbots. In: International conference on internet science proceedings. Thessaloniki, pp 377–392

Chatbots.org (2019) https://www.chatbots.org/. Accessed 11 July 2019

Chaves AP, Gerosa MA (2019) How should my chatbot interact? A survey on human-chatbot interaction design. arXiv:1904.02743, pp 1–44

Deci EL, Ryan RM (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 55(1):68–78

den Boer PB (2017) Talking to bots: estimating the environmental impact of human–computer interaction methods. In: 27th Twente student conference on IT proceedings. Enschede, pp 1–7

Di Prospero A, Norouzi N, Fokaefs M, Litoiu M (2017) Chatbots as assistants: an architectural framework. In: Proceedings of the 27th annual international conference on computer science and software engineering. Markham, pp 76–86

Diederich S, Brendel AB, Kolbe LM (2019) Towards a taxonomy of platforms for conversational agent design. In: WI 2019 proceedings. Siegen, pp 1100–1114

Feine J, Gnewuch U, Morana S, Maedche A (2019) A taxonomy of social cues for conversational agents. Int J Hum Comput Stud 132:138–161

Fleiss JL (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76(5):378–382

Følstad A, Brandtzaeg PB (2017) Chatbots and the new world of HCI. Interactions 24(4):38–42

Følstad A, Skjuve M, Brandtzaeg PB (2019) Different chatbots for different purposes: towards a typology of chatbots to understand interaction design. In: International conference on internet science proceedings. St. Petersburg, pp 145–156

Gimpel H, Rau D, Röglinger M (2017) Understanding fintech start-ups–a taxonomy of consumer-oriented service offerings. Electron Mark 28(3):245–264

Gnewuch U, Morana S, Maedche A (2017) Towards designing cooperative and social conversational agents for customer service. In: ICIS proceedings. Seoul, pp 1–14

Gnewuch U, Morana S, Adam MT, Maedche A (2018a) Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. In: ECIS proceedings. Portsmouth, pp 1–16

Gnewuch U, Morana S, Adam MT, Maedche A (2018b) "The Chatbot is typing…"–The role of typing indicators in human-chatbot interaction. In: Pre-ICIS workshop on HCI research in MIS proceedings. San-Francisco, pp 1–5

Hu T, Xu A, Liu Z, You Q, Guo Y, Sinha V, Luo J, Akkiraju R (2018) Touch your heart: a tone-aware chatbot for customer care on social media. In: CHI conference on human factors in computing system. Montreal, pp 1–12

Jain M, Kumar P, Kota R, Patel SN (2018) Evaluating and informing the design of chatbots. In: *Designing interactive system conference*. Hong Kong, pp 895–906

Kepuska V, Bohouta G (2018) Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In: IEEE 8th CCWC proceedings. Las Vegas, pp 99–103

Kim K, Boelling L, Haesler S, Bailenson J, Bruder G, Welch GF (2018) Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In: IEEE international symposium on mixed and augmented reality proceedings. Munich pp 105–114

Kiousis S (2002) Interactivity: a concept explication. New Media Soc 4(3):355–383

Knijnenburg BP, Willemsen MC (2016) Inferring capabilities of intelligent agents from their external traits. ACM Trans Interact Intell Syst 6(4):1–25

Knote R, Janson A, Eigenbrod L, Söllner M (2018) The what and how of smart personal assistants: principles and application domains for IS research. In: MKWI 2018 proceedings. Lüneburg, pp 1083–1094

Knote R, Janson A, Söllner M, Leimeister JM (2019) Classifying smart personal assistants: an empirical cluster analysis. In: 52nd HICSS proceedings. Maui, pp 2024–2033

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174

Li X, Wang Y, Sun S, Panda S, Liu J, Gao J (2018) Microsoft dialogue challenge: building end-to-end task-completion dialogue systems. arXiv:1807.11125, pp 1–5

Littlejohn SW, Foss KA (2010) Theories of human communication. Waveland, Long Grove

Maedche A, Morana S, Schacht S, Werth D, Krumeich J (2016) Advanced user assistance systems. Bus Inf Syst Eng 58(5):367–370

McTear MF (2016) The rise of the conversational interface: a new kid on the block?. In: International workshop on future and emerging trends in language technology proceedings. Seville, pp 38–49

McTear MF (2018) Conversational modelling for chatbots: current approaches and future directions. In: Proceedings of the 29th conference on electronic speech signal. Ulm, pp 175–185

Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63(2):81–97

Mittal A, Agrawal A, Chouksey A, Shriwas R, Agrawal S (2016) A comparative study of chatbots and humans. Int J Adv Res Comput Commun Eng 5(3):1055–1057

Nguyen QN, Sidorova A (2018) Understanding user interactions with a chatbot: a self-determination theory approach. In: AMCIS proceedings. New Orleans, pp 1–5

Nickerson RC, Varshney U, Muntermann J (2013) A method for taxonomy development and its application in information systems. Eur J Inf Syst 22(3):336–359

Nuruzzaman M, Hussain OK (2018) A survey on chatbot implementation in customer service industry through deep neural networks. In: IEEE international conference on e-business engineering proceedings. Shaanxi, pp 54–61

Olteanu A, Castillo C, Diaz F, Kiciman E (2019) Social data: biases, methodological pitfalls, and ethical boundaries. Front Big Data 2(13):1–33

Piccolo LS, Mensio M, Alani H (2018) Chasing the chatbots. In: International conference on internet science proceedings, pp 157–169

Rouast PV, Adam MTP, Chiong R (2019) Deep learning for human affect recognition: insights and new developments. IEEE Trans Affect Comput 14(8):1–20

Saravanan K, Sreedevi E, Subhamathi V (2017) A review of artificial intelligence systems. Int J Adv Res Comput Sci 9(2):397–401

Sarikaya R (2017) The technology behind personal digital assistants: an overview of the system architecture and key components. IEEE Signal Process Mag 34(1):67–81

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Bull 38(2):1409–1438

Stoeckli E, Dremel C, Uebernickel F, Brenner W (2019) How affordances of chatbots cross the chasm between social and traditional enterprise systems. Electron Mark 8:1–35

Szopinski D, Schoormann T, Kundisch D (2019) Because your taxonomy is worth it: towards a framework for taxonomy evaluation. In: ECIS proceedings. Stockholm, pp 1–19

Täuscher K, Laudien SM (2018) Understanding platform business models: a mixed methods study of marketplaces. Eur Manag J 36(3):319–329

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58(301):236–244

Wei C, Yu Z, Fong S (2018) How to build a chatbot: chatbot framework and its capabilities. In: Proceedings of the international conference on machine learning and computing. Macau, pp 369–373

Yalçın ÖN (2019) Evaluating empathy in artificial agents. In: 8th International conference on affective computing and intelligent interaction. Cambridge, pp 1–7

Yorita A, Egerton S, Oakman J, Chan C, Kubota N (2019) Self-adapting chatbot personalities for better peer support. In: IEEE international conference on systems, man and cybernetics. Bari, pp 4094–4100

Zumstein D, Hundertmark S (2017) Chatbots – an interactive technology for personalized communication, transactions and services. IADIS Int J WWW/Internet 15(1):96–109